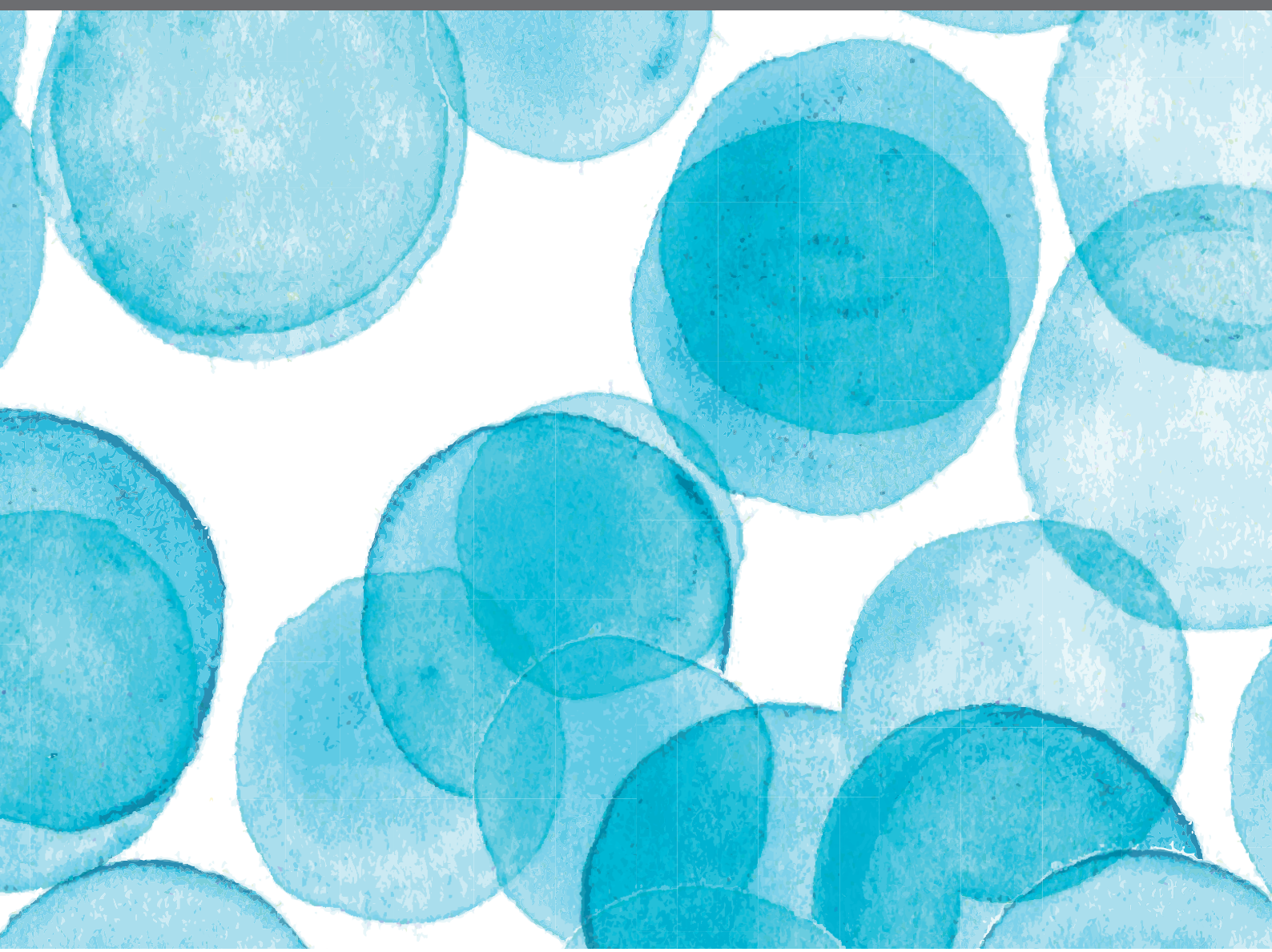
GENETICS ARCHITECTURE AND UNDERLYING MOLECULAR MECHANISMS IN HOST-PATHOGEN INTERACTIONS

EDITED BY: Dong Xia, Androniki Psifidi, Diego Robledo and Benjamin Makepeace

PUBLISHED IN: Frontiers in Genetics and Frontiers in Microbiology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-460-5
DOI 10.3389/978-2-88971-460-5

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

GENETICS ARCHITECTURE AND UNDERLYING MOLECULAR MECHANISMS IN HOST-PATHOGEN INTERACTIONS

Topic Editors:
Dong Xia, Royal Veterinary College (RVC), United Kingdom
Androniki Psifidi, Royal Veterinary College (RVC), United Kingdom
Diego Robledo, University of Edinburgh, United Kingdom
Benjamin Makepeace, University of Liverpool, United Kingdom

Citation: Xia, D., Psifidi, A., Robledo, D., Makepeace, B., eds. (2021). Genetics Architecture and Underlying Molecular Mechanisms in Host-Pathogen Interactions. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-460-5

Table of Contents

frontiers in Genetics

EDITORIAL
published: 18 August 2021
doi: 10.3389/fgene.2021.695109

Editorial: Genetics Architecture and Underlying Molecular Mechanisms in Host-Pathogen Interactions

Benjamin L. Makepeace[1†], *Androniki Psifidi*[2†], *Diego Robledo*[3†] *and Dong Xia*[4*†]

[1] Department of Infection Biology and Microbiomes, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, United Kingdom, [2] Department of Clinical Science and Services, Royal Veterinary College, University of London, London, United Kingdom, [3] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom, [4] Department of Comparative Biomedical Science, Royal Veterinary College, University of London, London, United Kingdom

Keywords: host-pathogen interactions, "omics" and systems biology, genetics, epigenetics, one health

Editorial on the Research Topic

Genetics Architecture and Underlying Molecular Mechanisms in Host-Pathogen Interactions

The study of host-pathogen interactions has been accelerating rapidly in both medical and veterinary contexts due to methodological advances and increasing infectious disease threats. This Research Topic attracted a wide breadth of submissions spanning analyses of host-pathogen interactions using *in vitro* and *in vivo* models of human disease; through to a range of studies in livestock species, comprising cattle, pigs, chickens, and farmed fish. An impressive breadth of technical approaches was also in evidence, from GWAS, SNP analysis, transcriptomics, RNAi, proteomics, and reverse genetics; often deployed in combination in a single study. Overall, there was as a strong bias toward "host-centric" articles ($n = 12$) compared with "pathogen-centric" ($n = 6$) studies, perhaps reflecting the worldwide drive to improve disease resistance in livestock.

Three "host-centric" studies were concerned with pathogens that are primarily of medical importance. Tan and Xia studied herpes simplex virus-1 (HSV-1) infection of the eye, a leading cause of infectious blindness, using a mouse model. Tripartite motif-containing protein-21 (TRIM21) was identified as a key factor in HSV epithelial keratitis. Upregulation of this host protein enhances the replication of HSV-1 in corneal epithelial cells *via* inhibition of stimulator of interferon genes/interferon regulatory factor-3 signalling and subsequent suppression of type I interferon production. TRIM21 also promotes the secretion of interleukin-6 and tumour necrosis factor-α, thereby aggravating the severity of HSV epithelial keratitis. Switching to bacterial pathogens, Wang et al. studied *Staphylococcus aureus* osteomyelitis in mice using both *in vivo* and *in vitro* systems. Based on an analysis of publicly available microarray data, they demonstrated the essential role of twist-related protein 1 (TWIST1) and NF-κB signalling in the early response to *S. aureus* infection in bone. The authors used a range of molecular techniques, including siRNA-mediated suppression of TWIST1, which impaired the migration and phagocytotic ability of macrophages in response to *S. aureus*. The third study on a human pathogen was that of Hsieh et al., who studied the role of intracellular NAD^+ content in the response to Group A streptococcus in a human microvascular endothelial cell line-1 (HMEC-1). The authors discovered that treatment with nicotinamide increases intracellular NAD^+ content, which enhances autophagy-mediated clearance of group A streptococcus in these cells. In contrast, nicotinamide did not have any impact on the infection by other intracellular bacteria.

Among the livestock host studies, two were focused on disease resistance in pigs. Sanglard et al. investigated the genomic architecture underlying the antibody response to porcine reproductive

Edited by:
Matteo Barberis,
University of Surrey, United Kingdom

Reviewed by:
Artur Yakimovich,
Roche, United Kingdom

****Correspondence:***
Dong Xia
dxia@rvc.ac.uk

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Biology Archive, a section of the journal Frontiers in Genetics

Received: *14 April 2021*
Accepted: *29 July 2021*
Published: *18 August 2021*

Citation:
Makepeace BL, Psifidi A, Robledo D and Xia D (2021) Editorial: Genetics Architecture and Underlying Molecular Mechanisms in Host-Pathogen Interactions. Front. Genet. 12:695109. doi: 10.3389/fgene.2021.695109

and respiratory syndrome virus (PRRSV) vaccination and its relationship to reproductive performance in commercial pigs. The trait was moderately heritable ($h^2 = 0.34$) and one major locus was identified on chromosome 7 (in the major histocompatibility region, MHC). The antibody response was highly correlated with several reproductive traits, suggesting it has great potential as an indicator to improve reproductive performance in commercial pigs. In the other study, Bai X. et al. studied phenotypes related to disease resilience using complete blood count (CBC) data from a wean-to-finish natural disease challenge model in F1 crossbred (Landrace × Yorkshire) pigs. Resilient animals were found to be primed to initiate a faster adaptive immune response and recovered earlier following infection, with greater increases in lymphocyte concentration, haemoglobin concentration, and haematocrit values, but exhibited a lower neutrophil concentration than in susceptible animals, including those that died. The CBC traits were found to be heritable and genetically correlated with growth and treatment, which may indicate that this phenotype can be used as part of developing predictions for disease resilience.

Chickens were represented by three host-focused studies. Deist et al. explored the genomic factors that lead to increased resistance to Newcastle disease virus (NDV). The authors studied the transcriptome of three different tissues upon challenge with the virus in two chicken genetic lines showing differential resistance to NDV. The study revealed time- and line-dependent responses to the virus and the authors used a network correlation analysis to integrate all the results and identify key factors driving resistance, such as eukaryotic translation initiation factor 2-alpha kinase-2, macrophage-expressed gene 1 protein, and tumour necrosis factor ligand superfamily member 13B. Meanwhile, Liu R. et al. used transcriptomics and proteomics to study the response of chicken synovial fibroblasts to *Mycoplasma synoviae*, an important poultry pathogen that can cause respiratory disease and arthritis. The authors used a primary cell culture of synovial tissues from pathogen-free chickens. While transcriptomic and proteomic data showed relatively little overlap, the authors identified potentially important players in *M. synoviae*-induced arthritis, including proliferation and apoptosis related factors, inflammatory mediators, or proangiogenic factors. A different approach was taken by Banos et al., who performed a joint analysis of two distinct Ethiopian indigenous chicken ecotypes to investigate the genomic architecture of important health and productivity traits and explore the feasibility of conducting genomic selection across ecotypes. A GWAS identified several significant genomic associations with health and productivity traits, while whole genome sequencing and pathway analysis revealed putative candidate genes and mutations. Reliability of genomic estimated breeding values across ecotype increased compared to within-ecotype calculations, but accuracy of genomic prediction did not, probably because the genetic distance between the two ecotypes offset the benefit from increased sample size.

The increasing importance of aquaculture for protein production was reflected in three articles on the challenges of infectious diseases in farmed fish. Fraslin et al. presented a review on different methods used in fish research to dissect host-pathogen interactions. Identification of QTL underlying resistance to pathogenic microbes, transcriptomics studies, and functional assays that describe host–pathogen interactions are discussed. The utility of fish isogenic lines as well as the need to combine multiple approaches to better understand host-pathogen dynamics is also highlighted. In a primary research article, Mastrochirico-Filho et al. generated a *de novo* transcriptome for a neotropical fish, the small-scaled pacu (*Piaractus mesopotamicus*), which is of major importance for South American aquaculture. The authors use this transcriptome to study the immune responses of pacu to *Aeromonas hydrophila* during the acute mortality phase of the infection and after it decreases and plateaus. Suppression of genes of the complement system and coagulation factor cascades may contribute to the pathogenesis of *A. hydrophila*. Moving to the northern hemisphere, Król et al. explored the molecular mechanisms associated with the progression of multifactorial gill disease in Atlantic salmon. Gills from three different aquaculture sites in Scotland were sampled and examined *via* gross morphology, RNA sequencing and histopathology. Gene expression or histopathology did not directly correlate with the changes in morphology (gill damage). Nonetheless, the authors did identify common expression patterns associated with multifactorial gill disease, mainly connected to two processes: immune responses driven by pro-inflammatory cytokines, and tissue damage and repair driven by caspases and angiogenin.

The final livestock host study was that of Holder et al., who investigated the presence of SNPs in the bovine *SLC11A1* gene and any resulting functional differences in natural resistance-associated macrophage protein 1 expression (NRAMP1; controlled by *SLC11A1*) that might be correlated with resistance/susceptibility to *Mycobacterium bovis* infection. One such SNP was identified and further analysis using ELISA revealed that the presence of an alternative G allele was associated with increased expression of NRAMP1 in bovine macrophages. Since NRAMP1 has been shown to influence the survival of *M. bovis* through the sequestering of iron, it is possible that cattle expressing the alternative G allele might have an increased resistance to infection.

Despite their small number, the "pathogen-centric" studies were diverse in subject matter, including representatives of bacterial, protozoal, and helminth pathogens and an arthropod ectoparasite. Cui et al. performed a genomic analysis of 31 *Cronobacter* (Enterobacteriaceae) isolates representing multiple species to identify putative virulence factors. Genome sequences of these species were compared with virulence genes in a virulence factor database, and those identified exhibited species specificities. Two novel gene clusters associated with higher cytotoxicity and stronger haemolysis capacity were identified. Another Gramme-negative pathogen featured in the study of Park et al., who demonstrated the function of a polycistronic mRNA leader in *Salmonella* that secures the growth of the bacteria. The *mgtCBR* mRNA encodes the virulence protein MgtC and the Mg^{2+} transporter MgtB, while a mutant with leaderless *mgtCBR* mRNA dysregulated the expression of MgtC and MgtB, reducing ATP to abnormal levels in a process that did not require F_1F_0 ATP synthase. Growth of the

mutant was shown to be impaired in both extracellular and intracellular (macrophage) environments. In another *in vitro* study, Peirasmaki et al. investigated the transcriptome of a protozoan, *Giardia intestinalis*, during its interaction with the human intestinal epithelial cell line, Caco-2. They identified members of the high cysteine membrane protein (HCMP) family to be among the most abundant upregulated genes. The authors subsequently located these HCMPs to the plasma membrane and peripheral vesicles and demonstrated that the expression level of these genes was affected by histone acetylation and the level of iron in the growth medium, linking them to potentially important roles during *Giardia*-host cell interactions.

The studies on multicellular parasites included Shi et al.'s proteomic characterisation of excretory-secretory products (ESP) and tegument proteins in the foodborne zoonotic liver fluke, *Clonorchis sinensis*. Enolase and cathepsin C were among the proteins identified in the ESP fraction, whereas annexins, actin, and tetraspanins were identified in tegument fraction as potential immunomodulators and promising vaccine antigens. A subset of tegument proteins labelled with biotin was localised to the surface membrane of the tegument. Different approaches were applied by Bai Y. et al., who employed RNA-seq and miRNA-seq to characterise the bi-directional development of protoscoleces in the dog tapeworm, *Echinococcus granulosus*. Comparisons between profiles of larval cysts and strobilar adult worms revealed that bi-directional development is controlled by miRNAs and genes likely associated with nervous system development and carbohydrate metabolic process. This information provides avenues to develop novel interventions and therapeutics for controlling the important zoonotic disease, cystic echinococcosis. Finally, continuing with the theme of small RNAs, Liu W-G. et al. characterised a novel tick-specific miRNA, hlo-miR-2, in *Haemaphysalis longicornis* that regulates moulting events; a process that is essential for the lifecycle of this three-host tick. The hlo-miR-2 and its predicted target gene *CPR1*, encoding a cuticular protein, exhibited tissue and temporal specificity; whereas overexpression or depletion of hlo-miR-2 led to a delayed or accelerated moulting process through the corresponding changes to *CPR1* expression.

In conclusion, this Research Topic attracted a broad range of host- and pathogen-centric studies employing a wide variety of methodological approaches. It highlights how sequencing technologies are rapidly "levelling the playing field" with respect to high-throughput investigations of host-pathogen interactions in livestock species, while previously these were largely the preserve of studies on humans or model organisms. Combined with continues improvements in both host and pathogen genome annotations and assemblies, as well as recently developed domain and expression based host-pathogen protein interaction predictions (Kshirsagar et al., 2013) and agent-based and dynamic optimization at system level (Pollmacher and Thilo Figge, 2015; Balsa-Canto et al., 2016), a genomic view of host-pathogen interaction models is emerging. A major theme was improving our understanding of disease resistance in livestock and its heritability, but genetic mechanisms underpinning phenotypic regulation in pathogens was also the topic of several papers. As we enter a new era with a belated appreciation of infectious disease threats, the study of host-pathogen interactions and disease susceptibility in wildlife, domestic animals and humans should take on a new urgency.

AUTHOR CONTRIBUTIONS

All authors contributed to this Editorial and co-edited the corresponding Research Topic.

REFERENCES

Balsa-Canto, E., Henriques, D., Gábor, A., and Banga, J. R. (2016). AMIGO2, a toolbox for dynamic modeling, optimization and control in systems biology. *Bioinformatics* 32, 3357–3359. doi: 10.1093/bioinformatics/btw411

Kshirsagar, M., Carbonell, J., and Klein-Seetharaman, J. (2013). Multitask learning for host–pathogen protein interactions. *Bioinformatics* 29, i217–i226. doi: 10.1093/bioinformatics/btt245

Pollmacher, J., and Thilo Figge, M. (2015). Deciphering chemokine properties by a hybrid agent-based model of *Aspergillus fumigatus* infection in human alveoli. *Front. Microbiol.* 6:503. doi: 10.3389/fmicb.2015.00503

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 06 December 2019
doi: 10.3389/fmicb.2019.02831

The *mgtCBR* mRNA Leader Secures Growth of *Salmonella* in Both Host and Non-host Environments

Myungseo Park[1†], Hyunkeun Kim[2†], Daesil Nam[2†], Dae-Hyuk Kweon[1] and Dongwoo Shin[2,3*]*

[1] Department of Integrative Biotechnology, College of Biotechnology and Bioengineering, Sungkyunkwan University, Suwon, South Korea, [2] Department of Molecular Cell Biology, Sungkyunkwan University School of Medicine, Suwon, South Korea, [3] Samsung Medical Center, School of Medicine, Sungkyunkwan University, Suwon, South Korea

Upon intracellular cues, bacterial mRNA leaders often form secondary structures that determine expression of a downstream protein-coding region(s), thereby providing bacteria with a mechanism to control the amounts of necessary proteins in the right locales. Here we describe a polycistronic mRNA leader that secures bacterial growth by preventing dysregulated expression of the protein-coding regions. In *Salmonella*, the *mgtCBR* mRNA encodes the virulence protein MgtC and the Mg^{2+} transporter MgtB. A mutant designed to produce leaderless *mgtCBR* mRNA induced MgtC and MgtB in conditions that promote *mgtC* transcription. The dysregulated expression of MgtC and MgtB impaired bacterial growth under all such non-host environments. While MgtC, but not MgtB, normally reduces ATP levels in a process requiring the F_1F_0 ATP synthase, dysregulated MgtC and MgtB reduced ATP levels independently of the F_1F_0 ATP synthase, which correlated with the mutant's growth defect. The mutant showed dysregulated MgtC expression and attenuated survival inside macrophages. While MgtB normally does not affect the phenotype, MgtB impaired intramacrophage survival of the mutant in the presence of MgtC. We provide an example showing that a polycistronic mRNA leader prevents the dysregulated function of protein-coding regions to allow bacteria to proliferate across complex niches.

Keywords: mRNA leader, MgtC, MgtB, ATP, *Salmonella*

OPEN ACCESS

Edited by:
Benjamin Makepeace, University of Liverpool, United Kingdom

Reviewed by:
Calvin A. Henard, University of North Texas, United States
Audrey Chong, National Institute of Allergy and Infectious Diseases (NIAID), United States

***Correspondence:**
Dae-Hyuk Kweon
dhkweon@skku.edu
Dongwoo Shin
shind@skku.edu

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *04 September 2019*
Accepted: *21 November 2019*
Published: *06 December 2019*

Citation:
Park M, Kim H, Nam D, Kweon D-H and Shin D (2019) The mgtCBR mRNA Leader Secures Growth of Salmonella in Both Host and Non-host Environments. Front. Microbiol. 10:2831. doi: 10.3389/fmicb.2019.02831

INTRODUCTION

The bacterium *Salmonella enterica* serovar Typhimurium (hereafter referred to as *Salmonella*) is a facultative intracellular pathogen that can proliferate both inside and outside host cells. To survive under both conditions, *Salmonella* must tightly control the amounts of necessary proteins in each locale.

The PhoP/PhoQ two-component system in *Salmonella* consists of the response regulator PhoP and the sensor kinase PhoQ (Groisman, 2001). The PhoQ protein senses multiple signals, such as low (i.e., micromolar concentrations) Mg^{2+}, acidic pH, and certain antimicrobial peptides, and phosphorylates the PhoP protein, rendering it functional as a transcriptional regulator (Garcia Vescovi et al., 1996; Bader et al., 2005; Prost et al., 2007). The gene expression programs regulated by PhoP/PhoQ enable *Salmonella* to adapt to low Mg^{2+} (Garcia Vescovi et al., 1996), survive in acidic pH (Foster and Hall, 1990), and acquire resistance against antimicrobial peptides (Fields et al., 1989). Moreover, activation of PhoP/PhoQ is a key event enabling *Salmonella* to survive inside the macrophage phagosome (Fields et al., 1989),

which contains antimicrobial factors, including acidic pH and antimicrobial peptides.

The virulence protein MgtC contributes to the growth of *Salmonella* inside macrophages as well as in low Mg^{2+} conditions (Blanc-Potard and Groisman, 1997; Rang et al., 2007). These dual roles of MgtC are associated with its ability to reduce cytoplasmic ATP levels by directly acting on the bacterium's own F_1F_0 ATP synthase (Lee et al., 2013; Pontes et al., 2015). MgtC also prevents hyperpolarization of *Salmonella* membrane (Lee et al., 2013). Given that MgtC plays this role independently of the F_1F_0 ATP synthase, this result suggests that MgtC could affect ATP levels by acting on a protein(s) other than the F_1F_0 ATP synthase (Lee et al., 2013).

The MgtC protein, together with the Mg^{2+} transporter MgtB (Snavely et al., 1991b) and the regulatory peptide MgtR (Alix and Blanc-Potard, 2008), are produced from the *mgtCBR* operon, which is regulated at multiple levels (Lee and Lee, 2015). When *Salmonella* is placed in low Mg^{2+} or in acidic pH environments, activated PhoP directly promotes *mgtC* transcription (Shin and Groisman, 2005; Choi and Groisman, 2016). However, this event alone does not ensure production of full length *mgtCBR* mRNA. The *mgtCBR* mRNA contains a 296 nucleotides long leader that harbors two short open reading frames (ORFs), *mgtM* and *mgtP* (**Figure 1A** and **Supplementary Figure S1**). In response to intracellular levels of ATP or an amino acid, these ORFs can promote formation of alternative structures of the *mgtCBR* mRNA leader that determine the degree of *mgtC* expression (Lee and Groisman, 2012a,b; Lee et al., 2014). For instance, an increase in ATP levels and a decrease in proline levels in the cytoplasm affects the coupling/uncoupling of transcription of the *mgtCBR* mRNA leader with translation of *mgtM* and *mgtP*, respectively, to induce structures of the leader that allow transcription elongation into the *mgtC*-coding region (Lee and Groisman, 2012a,b; Lee et al., 2014). Such *mgtM*- and *mgtP* regulation enables *Salmonella* to induce *mgtC* transcription at levels that ensure its survival inside macrophages (Lee and Groisman, 2012a,b; Lee et al., 2014).

MgtC expression is also negatively controlled at the posttranscriptional level. In low Mg^{2+} conditions, PhoP binds to and activates the *amgR* promoter located in the *mgtC–mgtB* intergenic region (Lee and Groisman, 2010). This event produces the AmgR antisense RNA that promotes degradation of the *mgtC* mRNA in a process requiring the RNase E (Lee and Groisman, 2010). Inactivation of *amgR* transcription increases MgtC production and renders *Salmonella* more virulent in mice (Lee and Groisman, 2010). The regulatory peptide MgtR encoded by the *mgtCBR* operon binds to the MgtC protein and promotes MgtC degradation in a process dependent on the FtsH protease (Alix and Blanc-Potard, 2008). Although the *mgtR* mutant produces MgtC at much higher levels than the wild-type (WT) strain, the mutant is slightly attenuated for survival inside macrophages (Alix and Blanc-Potard, 2008).

In the present study, we designed a *Salmonella* mutant that produced leaderless *mgtCBR* mRNA. We found that in both host and non-host environments where the PhoP regulator is activated, removal of the *mgtCBR* mRNA leader caused *Salmonella* to induce MgtC and MgtB in a dysregulated manner, which impaired bacterial growth in such conditions. The dysregulated expression of MgtC and MgtB reduces ATP to abnormal levels in a process that does not require the F_1F_0 ATP synthase, and this reduction in ATP levels seems to be a cause for the *Salmonella* growth defect seen. Our study reveals that the role of the *mgtCBR* mRNA leader to prevent dysregulated expression of MgtC and MgtB is important for the growth of *Salmonella* both inside and outside host cells.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Growth Conditions

Bacterial strains and plasmids used in this study are listed in **Table 1**. All *S. enterica* serovar Typhimurium strains were derived from strain 14028s. Phage P22-mediated transductions were performed as described (Davis et al., 1980). Bacteria were grown at 37°C in N-minimal medium (Snavely et al., 1991a), pH 7.5 or pH 5.5, supplemented with 0.1% Casamino Acids, 38 mM glycerol, and the indicated concentrations of $MgCl_2$. Ampicillin, kanamycin, and C18G peptide (AnaSpec) were used at 50, 30, and 5 μg/ml, respectively. For induction of genes from plasmids, isopropyl 1-thio-β-D-galatoside (IPTG) was used at the indicated concentrations.

Construction of Chromosomal Mutants and Plasmid

Salmonella enterica strains carrying a chromosomal gene deletion were constructed using the one-step gene inactivation method (Datsenko and Wanner, 2000) with necessary modifications. DN557 is a strain, in which the sequences encoding the *mgtCBR* mRNA leader regions were deleted from the chromosome. To construct DN557, the *tetRA* fragment was amplified from strain MS7953s using the primer pair Del701/Del702 and integrated into the chromosome of WT strain 14028s harboring the pKD46 plasmid (Datsenko and Wanner, 2000). The resulting strain HK514 keeps the *tetRA* genes in the *mgtCBR* mRNA leader-encoding region and pKD46 at 30°C. To obtain the engineered DNA fragment that is deleted for *mgtCBR* mRNA leader-encoding regions, the left and right arms of the DNA fragment were amplified from strain 14028s using primer pairs Del703/Del704 and Del705/Del706, respectively. Then, the second step of PCR was conducted on the mixture of the first step of PCR products using the primer pair Del703/Del706. The resulting PCR product was used to electroporate the HK514 strain, and the bacterial suspension was plated on medium containing fusaric acid and incubated at 42°C to select against the *tetRA* genes (Maloy and Nunn, 1981).

Strain HK111, in which the sequences corresponding to *mgtCBR* mRNA leader and *mgtC*-coding regions were removed from the chromosome, was similarly constructed. In the first step of PCR, two DNA fragments were amplified from strain 14028s using primer pairs Del703/Del704 and Del705/Del708. The second step of PCR was conducted on the mixture

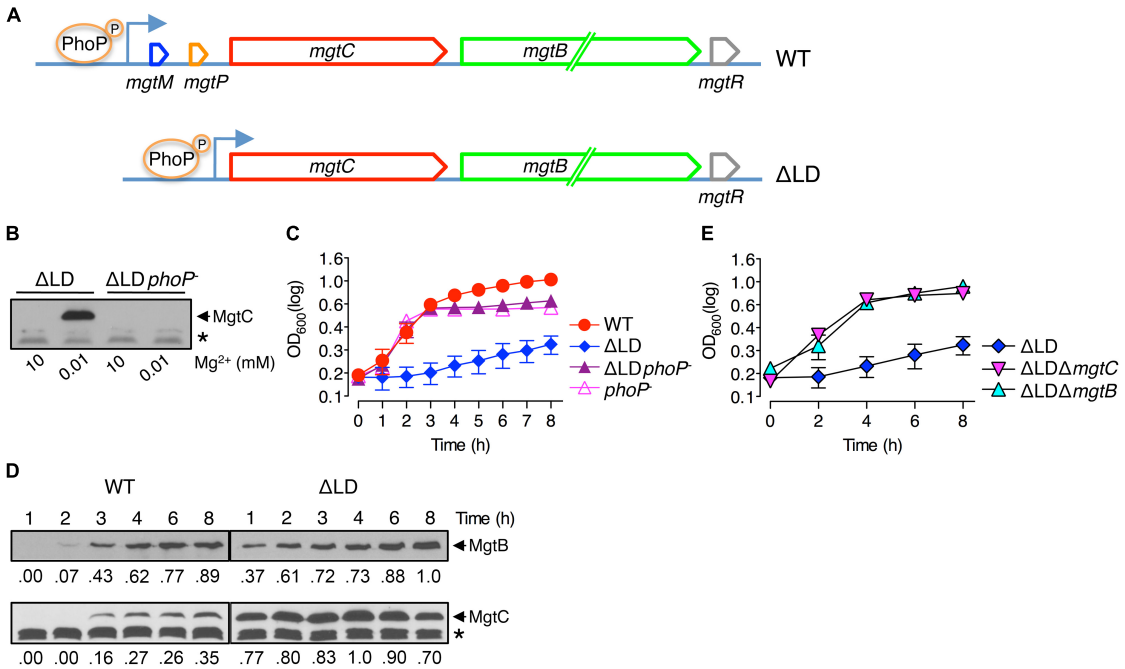


FIGURE 1 | Dysregulated induction of MgtC and MgtB impairs growth of *Salmonella* in low Mg^{2+}. **(A)** Schematic of the *mgtCBR* operon in the wild-type *Salmonella* (WT, 14028s) and ΔLD mutant (DN557). **(B)** Immunoblot analysis using anti-MgtC antibodies of crude extracts prepared from ΔLD and ΔLD *phoP*$^-$ (DN558) strains. Bacteria were grown in N-minimal medium with 10 mM or 10 μM Mg^{2+} at pH 7.5 for 4 h. **(C,E)** Growth curves of wild-type (WT), ΔLD, ΔLD *phoP*$^-$, *phoP*$^-$ (MS7953s), ΔLD Δ*mgtC* (HK111), and ΔLD Δ*mgtB* (DN581) strains. Bacteria were grown in N-minimal medium with 10 μM Mg^{2+} at pH 7.5, and OD_{600} values were determined at the indicated time points. Means and standard deviations from three independent experiments are shown. **(D)** Immunoblot analysis using anti-MgtC and anti-MgtB antibodies of crude extracts prepared from wild-type (WT) and ΔLD strains grown in N-minimal medium with 10 μM Mg^{2+} at pH 7.5 and harvested at the indicated time points. The band indicated with an asterisk (*) corresponds to a protein displaying cross-reactivity against the anti-MgtC antibody and serves as an internal loading control. Numbers below the blots correspond to relative levels of MgtC and MgtB at a given time point.

of the first step of PCR products using the primer pair Del703/Del708. The resulting PCR product was introduced into the strain HK514, and tetracycline-sensitive colonies were selected. Deletion of the sequences in the strains DN557 and HK111 was verified by nucleotide sequencing. To delete the *mgtB* gene in the DN557 and HK111 strains, the kanamycin resistance gene (Km^R) cassette from the pKD4 plasmid (Datsenko and Wanner, 2000) was amplified using the primer pair Del711/Del712. The resulting PCR product was used to electroporate the DN557 and HK111 strains carrying pKD46. To delete the *mgtC*, *mgtB*, and *mgtCB* genes in strain EG19307, the Km^R cassette from plasmid pKD4 was amplified using primer pairs Del713/Del714, Del711/Del712, and Del713/Del712, respectively. The respective PCR products were used to electroporate the EG19307 strain carrying pKD46. To delete the *mgtR* gene in strain DN557, the Km^R cassette from plasmid pKD4 was amplified using the primer pairs Del715/Del716. The respective PCR products were used to electroporate the DN557 strain carrying pKD46. The Km^R cassette was removed using plasmid pCP20 (Datsenko and Wanner, 2000). To delete the *atpB* gene in the 14028s strain, the Km^R cassette from plasmid pKD4 was amplified using the primer pair Del717/Del718, and the PCR product was used to electroporate the 14028s strain carrying pKD46. Deletion of the corresponding genes was verified by colony PCR. The sequences of all primers used for strain construction are listed in **Supplementary Table S1**.

Plasmid p*mgtC*(ORF) expresses the *mgtC* ORF from the *lac* promoter. The *mgtC* ORF was amplified using the primer pair Ex301/Ex302 and chromosomal DNA from the 14028s strain, and the *mgtC* ORF was then introduced between the *Eco*RI and *Pst*I restriction sites of pUHE21-2*lacI*^{q} (Soncini et al., 1995). Recombinant plasmid sequences were confirmed by nucleotide sequencing. The sequences of primers used for plasmid construction are listed in **Supplementary Table S1**.

Immunoblot Analysis

Bacteria were grown in N-minimal medium for the indicated amounts of time. Equivalent amounts of bacterial cells normalized by OD_{600} values were collected, washed with phosphate-buffered saline (PBS), suspended in 0.15 ml SDS sample buffer (Laemmli sample buffer), and boiled. Whole-cell lysates were resolved on 12% SDS polyacrylamide gels, transferred to nitrocellulose membranes, and analyzed by immunoblot using anti-MgtC (Park et al., 2018), anti-MgtB (Choi et al., 2017), or anti-RpoA (NeoClone) antibodies. Membranes were incubated with anti-rabbit IgG horseradish peroxidase-linked antibodies (GE Healthcare), and bands were visualized using the ECL detection system (GE Healthcare). Protein levels were quantified using the ImageJ program (NIH).

TABLE 1 | Bacterial strains and plasmids used in this study.

Strain or plasmid	Genotype or relevant characteristics	References
***S. enterica* serovar Typhimurium**		
14028s	Wild-type	Fields et al. (1986)
MS7953s	14028s *phoP7953*:Tn10	Fields et al. (1986)
DN552	14028s Δ*mgtR*	Choi E. et al. (2012)
DN557	14028s Δ*mgtCBR* mRNA leader (ΔLD)	This study
DN558	14028s ΔLD *phoP7953*:Tn10	This study
DN575	14028s ΔLD Δ*atpB*:Km^R	This study
DN581	14028s ΔLD Δ*mgtB*	This study
DN582	14028s ΔLD Δ*mgtR*	This study
DN608	14028s ΔLD Δ*mgtCB*	This study
DN612	14028s *mgtM*(UAG) Δ*mgtB*	This study
DN649	14028s *mgtM*(UAG) Δ*mgtC*	This study
DN652	14028s *mgtM*(UAG) Δ*mgtCB*	This study
EG19307	14028s *mgtM*(UAG)	Lee and Groisman (2012a)
EN397	14028s Δ*mgtC*	Park et al. (2018)
EN481	14028s Δ*mgtB*	Park et al. (2018)
HK111	14028s ΔLD Δ*mgtC*	This study
HK468	14028s Δ*atpB*:Km^R	This study
MS575	14028s ΔLD Δ*mgtC* Δ*atpB*:Km^R	This study
MS576	14028s ΔLD Δ*mgtB* Δ*atpB*:Km^R	This study
DN686	14028s ΔLD Δ*mgtCB* Δ*atpB*:Km^R	This study
Plasmids		
pUHE21-2*lacI*^{q}	P_{lac} rep_{pMBI} Ap^R *lacI*^{q}	Soncini et al. (1995)
p*mgtC*(ORF)	pUHE21-2*lacI*^{q} *mgtC* ORF	This study
pKD4	$repR_{6K\gamma}$ Ap^R FRT Km^R FRT	Datsenko and Wanner (2000)
pKD46	$rep_{pSC101}{}^{ts}$ Ap^R P_{araBAD} γ β exo	Datsenko and Wanner (2000)
pCP20	$rep_{pSC101}{}^{ts}$ Ap^R Cm^R *cI857* λP_R*flp*	Datsenko and Wanner (2000)

RNA Isolation and Quantitative Real-Time RT-PCR (qRT-PCR) Analysis

Bacteria were grown in N-minimal medium for the indicated amounts of time. The culture (0.5 ml) was removed and mixed with 1 ml RNAprotect Bacteria Reagent (Qiagen). RNA was isolated using the RNeasy Mini Kit (Qiagen) and treated with RNase-free DNase (Ambion). cDNA was synthesized from 0.1 μg of template RNA using the PrimeScript RT reagent Kit (Takara) and random primers (Life Technologies). Amounts of cDNA were quantified by real-time PCR using the SYBR Green Realtime PCR Master Mix (Toyobo) with an ABI7300 Sequence Detection System (Applied Biosystems). The following primer pairs were used for the detection of cDNA corresponding to *mgtC*, *mgtB*, and *gyrB* mRNAs: Q-mgtC-F/Q-mgtC-R, Q-mgtB-F/Q-mgtB-R, and Q-gyrB-F/Q-gyrB-R, respectively. Transcription levels of each gene were calculated from a standard curve obtained by PCR with the same primers and serially diluted genomic DNA. mRNA levels of target genes were normalized to *gyrB* mRNA levels. The sequences of primers are listed in **Supplementary Table S1**.

Measurement of Intracellular ATP Levels

ATP levels were determined as previously described (Lee et al., 2013). Bacteria were grown in N-minimal medium for 4 h. Equivalent amounts of bacterial cells ($0.5 \times OD_{600}$) were removed, washed with PBS, and suspended in 0.5 ml of PBS. Nucleic acids were extracted by adding 100 ml of ice-cold 3 M perchloric acid. After incubation for 5 min, extracts were neutralized with 225 ml of neutralization buffer (1 M KOH, 0.5 M Tris, 0.5 M KCl) and centrifuged at 13,000 rpm and 25°C for 10 min. Fifty microliters of the supernatant were diluted with 50 ml of L buffer (25 mM KCl, 50 mM $MgSO_4$, 100 mM HEPES pH 7.4), and ATP levels were measured using an ATP Determination Kit (Life Technologies) according to the manufacturer's instruction. Statistical analysis of the data was conducted using the GraphPad Prism program (version 5.0).

Macrophage Survival Assay

The experiment was conducted as previously described (Choi Y. et al., 2012). J774A.1 macrophage cells were grown in Dulbecco modified Eagle medium (DMEM) (Life Technologies) supplemented with 10% fetal bovine serum (FBS) and 1% Antibiotic-Antimycotic (Life Technologies) at 37°C under 5% CO_2. Prior to bacterial infection, a monolayer of 1×10^5 J774A.1 cells was prepared in a 24-well tissue culture plate and incubated in DMEM with 10% FBS for 1 h. Bacteria were cultured in Luria-Bertani (LB) medium for 18 h with aeration and opsonized in 20% normal mouse serum for 25 min at 37°C. Opsonized bacteria were diluted in DMEM with 10% FBS and added to the cell monolayer at a multiplicity of infection (MOI) of 10. After

1 h of incubation, the wells were washed three times with pre-warmed PBS to remove extracellular bacteria and then incubated for another 1 h with the pre-warmed medium with 100 μg/ml gentamicin to kill extracellular bacteria. The wells were washed three times with PBS and incubated in pre-warmed medium with 10 μg/ml of gentamicin. At the desired time points, the wells were washed three times with PBS and treated with 1% Triton X-100 for 10 min. The suspension was diluted in PBS and plated on LB agar plates to enumerate colony-forming units.

Determination of MgtC Levels Inside Macrophages

J774A.1 macrophage and bacterial cells were cultured as described above. A monolayer of 1×10^6 J774A.1 cells was prepared in a 6-well tissue culture plate. Opsonized bacteria were added to the cell monolayer at a MOI of 10. The wells were then treated as described above to remove extracellular bacteria. At the desired time points, the wells were washed three times with PBS and treated with 1% Triton X-100 for 30 min. The cell lysis mixture was centrifuged at 13,000 rpm for 10 min, and the bacterial cell pellet was suspended in 0.15 ml SDS sample buffer and boiled. MgtC levels in bacterial cell lysates were analyzed by immunoblot as described above.

Measurement of Bacterial Growth Using a Plate Reader

Bacteria were grown in N-minimal medium with 10 mM $MgCl_2$ at pH 7.5 to saturation. One milliliter of bacterial cells were collected, washed twice with medium not supplemented with $MgCl_2$, and diluted 1:100 into wells of a 24-well plate containing 1 ml of medium with 10 μM or 1 mM $MgCl_2$ at pH 7.5 or pH 5.5. The C18G peptide was added to medium at 5 μg/ml. A plate was covered with a Breathe-Easy sealing membrane (Sigma) to prevent evaporation. By using an xMark^{TM} Microplate Spectrophotometer (Bio-Rad), bacteria were cultivated at 37°C with shaking, and the OD_{600} values were measured every 5 min up to 8 h.

RESULTS

A *Salmonella* Strain Designed to Produce Leaderless *mgtCBR* mRNA Is Severely Impaired for Growth in Low Mg^{2+}

We investigated the behaviors of *Salmonella* that produces leaderless *mgtCBR* mRNA. We deleted the sequences specifying the *mgtCBR* mRNA leader regions that participate in forming regulatory structures from the chromosome without leaving any heterologous sequences behind (**Figure 1A** and **Supplementary Figure S1**). The resulting strain (ΔLD) still possessed PhoP-dependent Mg^{2+} regulation of *mgtC* expression. The MgtC protein was detected in the ΔLD strain grown in defined medium supplemented with 10 μM Mg^{2+} but not in medium with 10 mM Mg^{2+} for 4 h, and the low Mg^{2+}-induced MgtC production did not occur in the strain carrying a *phoP*$^-$ allele (**Figure 1B**). (Note that the medium pH was adjusted to 7.5, unless otherwise stated.)

While preparing bacterial cultures, we repeatedly observed that the ΔLD strain did not grow well in low Mg^{2+}. During experiments, *Salmonella* strains were initially grown to saturation in medium with 10 mM Mg^{2+} and then diluted to OD_{600} values of ∼0.15 in medium with 10 μM Mg^{2+}. After a growth lag for the first 1 h, the WT strain grew logarithmically up to 3 h and then displayed slow linear growth for the subsequent 5 h (**Figure 1C**), generating a typical growth curve in 10 μM Mg^{2+} (Soncini et al., 1996; Blanc-Potard and Groisman, 1997; Park et al., 2018). However, the ΔLD strain did not achieve logarithmic growth and only grew in a slow linear manner during the entire 8 h (**Figure 1C**).

The ΔLD Strain Expresses MgtC and MgtB in Low Mg^{2+} in a Dysregulated Manner

To understand the molecular basis underlying the growth defect seen in the ΔLD strain, we compared expression of *mgtC* and *mgtB* between the WT and ΔLD strains. When *Salmonella* is placed in low Mg^{2+}, PhoP directly promotes transcription initiation of the *mgtC* gene (Shin and Groisman, 2005). However, *Salmonella* can produce MgtC protein at detectable levels after cytoplasmic Mg^{2+} levels drop to a certain threshold (Yeom et al., 2017; Park et al., 2018). Given that a secondary structure of the *mgtCBR* mRNA leader prevents transcription elongation from reaching the *mgtC*-coding region (Lee and Groisman, 2012a,b; Lee et al., 2014), we reasoned that the expression kinetics of *mgtC* and *mgtB* must be different between the WT and ΔLD strains.

When placed in medium with 10 μM Mg^{2+}, the WT strain produced *mgtC* mRNA at very low levels for the first 2 h (**Supplementary Figure S2A**). The *mgtC* mRNA levels then increased at 3 h by ∼10-fold, and this high-level production continued for the following 5 h (**Supplementary Figure S2A**). By contrast, under the same growth condition, the ΔLD strain produced *mgtC* mRNA at levels similar to the maximum levels of the WT as early as at 1 h and maintained these levels for the following 7 h (**Supplementary Figure S2A**). Moreover, consistent with the notion that the *mgtC* and *mgtB* mRNAs constitute a polycistronic mRNA (Alix and Blanc-Potard, 2008; Lee and Groisman, 2010), the WT and ΔLD strains highly induced *mgtB* mRNA at the delayed (i.e., 3 h) and the early (i.e., 1 h) time points, respectively (**Supplementary Figure S2B**).

We also determined levels of the MgtC and MgtB proteins over time during bacterial growth in 10 μM Mg^{2+}. Consistent with previous findings (Yeom et al., 2017; Park et al., 2018) and the *mgtC* mRNA levels (**Supplementary Figure S2A**), MgtC was detected in the WT strain only after 3 h (**Figure 1D**). In contrast, and as observed with *mgtC* mRNA production (**Supplementary Figure S2A**), the ΔLD strain produced detectable levels of MgtC as early as 1 h of growth (**Figure 1D**). Notably, despite the finding that the maximum levels of *mgtC* mRNA were comparable between the two strains (**Supplementary Figure S2A**), MgtC levels were ∼4-fold higher in the ΔLD strain than in the WT strain at 4 h (**Figure 1D**). As observed with *mgtB* mRNA production (**Supplementary Figure S2B**), the WT and ΔLD strains produced detectable levels of MgtB in parallel

with MgtC (**Figure 1D**). However, in contrast to MgtC levels, MgtB levels detected after 4 h were similar between the two strains (**Figure 1D**).

In medium with 10 μM Mg^{2+}, WT *Salmonella* grew logarithmically for 3 h by consuming Mg^{2+} in the medium (**Figure 1C**). During this time period, the ΔLD strain produced the Mg^{2+} transporter MgtB at much higher levels (**Figure 1D**) with even lower growth yields than the WT strain (**Figure 1C**), suggesting that cytoplasmic Mg^{2+} levels in the mutant are at least not lower than the WT levels. Taken together, these results suggest that removal of the leader from the *mgtCBR* mRNA causes *Salmonella* to induce MgtC and MgtB in a dysregulated manner even prior to its experiencing low cytoplasmic Mg^{2+} stress.

MgtC and MgtB Impair Growth of the ΔLD Strain in Low Mg^{2+}

As previously observed (Soncini et al., 1996; Blanc-Potard and Groisman, 1997), in medium with 10 μM Mg^{2+}, the *phoP*$^-$ strain grew logarithmically in a manner similar to the WT strain but showed defective growth in the slow-growth phase (**Figure 1C**). The ΔLD *phoP*$^-$ strain, which failed to induce MgtC expression (**Figure 1B**), showed logarithmic growth, with a growth curve similar to that of the *phoP*$^-$ strain (**Figure 1C**). This result suggests that the growth defect of the ΔLD strain may be associated with dysregulated MgtC induction. Consistent with this prediction, the ΔLD Δ*mgtC* strain recovered the ability of logarithmic growth (**Figure 1E**).

Because the ΔLD strain produced MgtB in a dysregulated manner as well as MgtC (**Figure 1D**), we next explored whether the *mgtB* gene was also responsible for the growth defect. Indeed, in medium with 10 μM Mg^{2+}, *mgtB* deletion enabled the ΔLD strain to grow logarithmically (**Figure 1E**). In the ΔLD background, *mgtC* deletion did not affect MgtB production, and vice versa (**Supplementary Figure S3**), suggesting that the growth phenotypes of the ΔLD Δ*mgtC* and ΔLD Δ*mgtB* strains are due to lack of MgtC- and MgtB function, respectively.

The *mgtR* gene in the *mgtCBR* operon (**Figure 1A**) encodes the regulatory peptide MgtR that controls MgtC levels (Alix and Blanc-Potard, 2008). Because the ΔLD strain is also predicted to produce MgtR in a dysregulated manner, we further examined whether *mgtR* deletion affects growth of the ΔLD strain. However, the ΔLD Δ*mgtR* strain grew in a manner similar to the ΔLD strain (**Supplementary Figure S4**). Together these results indicate that when produced in a dysregulated manner, MgtC and MgtB impair the growth of *Salmonella* under low Mg^{2+} conditions.

In the ΔLD Strain, MgtC and MgtB Contribute to ATP Reduction to Abnormal Levels in a Process That Does Not Require F_1F_0 ATP Synthase

Given that MgtC functions to reduce cytoplasmic ATP levels by acting on the F_1F_0 ATP synthase (Lee et al., 2013) and that the ΔLD strain produced higher levels of MgtC (**Figure 1D**), we reasoned that ATP levels may be different between the ΔLD and WT strains. We thus determined ATP levels in *Salmonella* at 4 h after growth in medium with 10 μM Mg^{2+}, in which both WT and ΔLD strains produced MgtC at detectable levels (**Figure 1D**). Similar to previous finding (Lee et al., 2013), ATP levels were higher in the Δ*mgtC* strain than in the WT strain (**Figure 2A**). In contrast, the ΔLD strain exhibited ∼2-fold lower ATP levels than the WT (**Figure 2A**). The ATP reduction in the ΔLD strain was dependent on MgtC, as deleting *mgtC* increased the ATP levels by ∼3-fold (**Figure 2A**).

Despite the findings that MgtB does not impact ATP levels in the WT strain (Lee et al., 2013; **Figure 2A**) and that MgtB levels were comparable between the WT and ΔLD strains (**Figure 1D**), we examined if MgtB affected ATP levels in the ΔLD strain. Interestingly, deletion of the *mgtB* gene elevated ATP levels in the ΔLD strain by ∼2-fold (**Figure 2A**). Moreover, additional

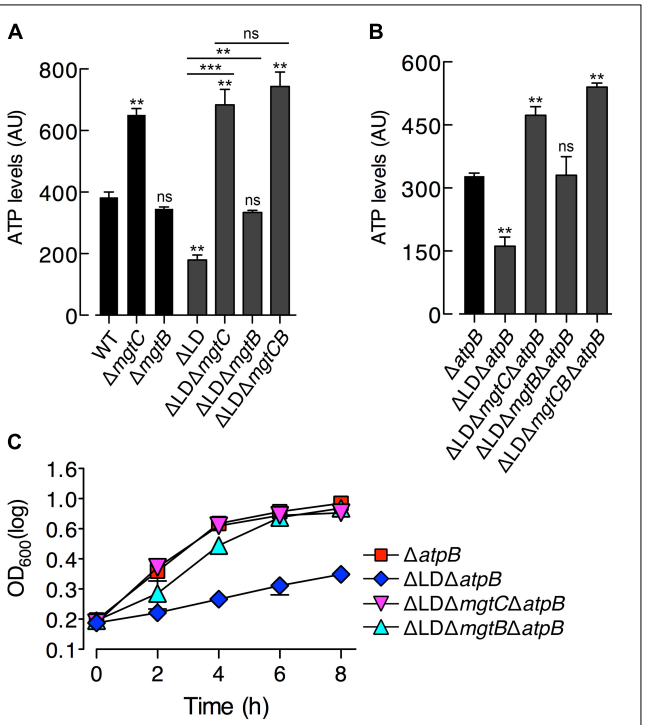


FIGURE 2 | Dysregulated expression of MgtC and MgtB reduces ATP to abnormal levels in a process that does not require F_1F_0 ATP synthase. ATP levels were determined in wild-type (WT, 14028s), Δ*mgtC* (EN397), Δ*mgtB* (EN481), ΔLD (DN557), ΔLD Δ*mgtC* (HK111), ΔLD Δ*mgtB* (DN581), and ΔLD Δ*mgtCB* (DN608) strains **(A)** as well as in Δ*atpB* (HK468), ΔLD Δ*atpB* (DN575), ΔLD Δ*mgtC* Δ*atpB* (MS575), ΔLD Δ*mgtB* Δ*atpB* (MS576), and ΔLD Δ*mgtCB* Δ*atpB* (DN686) strains **(B)**. Bacteria were grown in N-minimal medium with 10 μM Mg^{2+} at pH 7.5 for 4 h. Data depicted in arbitrary units (AU) are means and standard deviations from three independent experiments. $^{**}P < 0.01$, $^{***}P < 0.001$, two-tailed *t*-test with each sample vs. WT **(A)** and with each sample vs. Δ*atpB* **(B)**, ns, not significant. **(C)** Growth curves of *Salmonella* strains, Δ*atpB* (HK468), ΔLD Δ*atpB* (DN575), ΔLD Δ*mgtC* Δ*atpB* (MS575), and ΔLD Δ*mgtB* Δ*atpB* (MS576). Bacteria were grown in N-minimal medium with 10 μM Mg^{2+} at pH 7.5, and OD_{600} values were determined at the indicated time points. Data are representative of three independent experiments, and means and standard deviations from three independent experiments.

deletion of *mgtC* further increased ATP levels of the ΔLD Δ*mgtB* strain to levels of the ΔLD Δ*mgtC* strain (**Figure 2A**). Together these results suggest that in the ΔLD strain, MgtB might contribute to ATP reduction in a manner dependent on MgtC.

The MgtC protein inhibits ATP synthesis by directly binding to the F_0 *a* subunit of the ATP synthase (Lee et al., 2013). Consistent with this observation, MgtC-dependent ATP reduction is no longer observed in the absence of the *atpB* gene, which encodes the F_0 *a* subunit (Lee et al., 2013). To explore whether ATP reduction in the ΔLD strain requires the F_1F_0 ATP synthase, we determined ATP levels in a set of Δ*atpB* strains grown in 10 μM Mg^{2+} for 4 h. We found that ATP levels were still ∼2-fold lower in the ΔLD Δ*atpB* strain than in the Δ*atpB* strain (**Figure 2B**). Moreover, deleting *mgtC* and *mgtB* increased ATP levels in the ΔLD Δ*atpB* strain by ∼3- and ∼2-fold, respectively (**Figure 2B**). Additional deletion of *mgtC* further increased ATP levels of the ΔLD Δ*mgtB* Δ*atpB* strain to levels of the ΔLD Δ*mgtC* Δ*atpB* strain (**Figure 2B**). Together these results suggest that dysregulated expression of MgtC and MgtB reduces ATP to abnormally low levels in a process that does not require F_1F_0 ATP synthase.

ATP Reduction Impairs Growth of the ΔLD Strain in Low Mg^{2+}

In medium with 10 μM Mg^{2+}, the ΔLD strain, which displayed lower ATP levels than the WT (**Figure 2A**), showed impaired growth (**Figure 1C**). In contrast, the ΔLD Δ*mgtC* strain, which displayed ATP levels as high as the Δ*mgtC* strain (**Figure 2A**), recovered growth to levels of the Δ*mgtC* strain (**Supplementary Figure S5A**). A similar relationship between ATP level and growth behavior was also observed between the ΔLD Δ*mgtB* and Δ*mgtB* strains (**Figure 2A** and **Supplementary Figure S5B**).

Growth behaviors of the strains carrying an *atpB* deletion also correlated with their ATP levels. In medium with 10 μM Mg^{2+}, the Δ*atpB* strain grew in a manner similar to the WT strain (**Figure 2C**). However, the ΔLD Δ*atpB* strain, which exhibited lower ATP levels than the Δ*atpB* strain (**Figure 2B**), showed severely impaired growth (**Figure 2C**). Additionally, deleting *mgtC* and *mgtB*, which each increased ATP levels in the ΔLD Δ*atpB* strain (**Figure 2B**), recovered growth of the ΔLD Δ*atpB* strain (**Figure 2C**). Together these results suggest that the ATP reduction led by MgtC and MgtB independent of F_1F_0 ATP synthase caused impaired growth of the ΔLD strain in low Mg^{2+}.

Dysregulated Induction of MgtC and MgtB Impairs *Salmonella* in High Mg^{2+} Environments With Acidic pH or an Antimicrobial Peptide

The PhoP regulator can be activated even in high (i.e., millimolar concentrations) Mg^{2+} by mildly acid pH (Prost et al., 2007; Choi and Groisman, 2016) or certain antimicrobial peptides (Bader et al., 2005). Given that PhoP promotes *mgtC* transcription at acidic pH (Choi and Groisman, 2016), we speculated about the behavior of the ΔLD strain under such conditions. We first determined MgtC levels in *Salmonella* grown in medium supplemented with 1 mM Mg^{2+} adjusted to pH 5.5. The ΔLD strain produced detectable levels of MgtC at 1 and 6 h after growth in a PhoP-dependent manner (**Figure 3A**). However, in the WT strain, MgtC was not detected at either time point (**Figure 3A**).

In the acidified medium, the ΔLD strain grew in a linear, slow fashion during the entire 8 h, which was in contrast to the WT strain that grew logarithmically for the first 4 h and continued to grow slowly for the remaining 4 h (**Figure 3B**). The growth defect seen in the ΔLD strain was also dependent on MgtC and MgtB, as evidenced by the finding that deleting *mgtC* and *mgtB* restored growth to the ΔLD strain (**Figure 3B**). Growth behaviors of the strains correlated with their ATP levels. ATP levels were ∼4-fold lower in the ΔLD strain than in the WT strain (**Figure 3C**). Deleting *mgtC* and *mgtB* increased ATP levels in the ΔLD strain by ∼6- and ∼2.5-fold, respectively (**Figure 3C**).

We next examined growth in medium containing 1 mM Mg^{2+} and sublethal concentrations of C18G, an antimicrobial

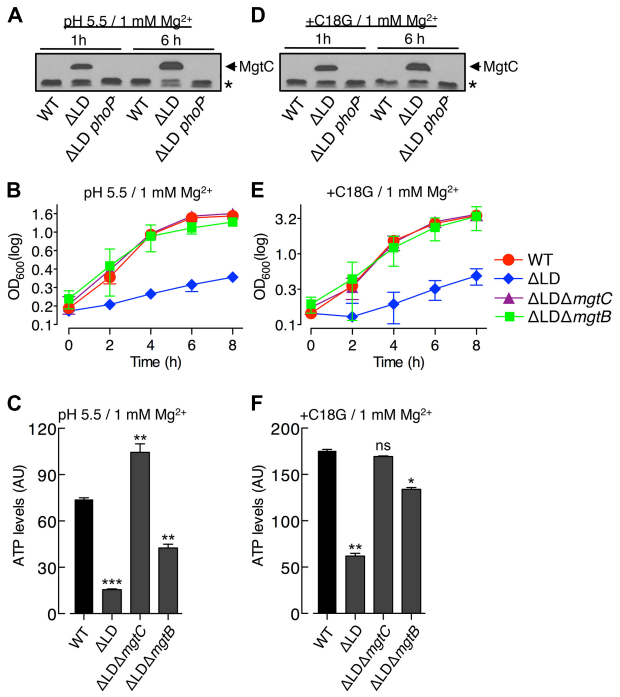


FIGURE 3 | Dysregulated induction of MgtC and MgtB impairs growth of *Salmonella* in high Mg^{2+} with acidic pH or an antimicrobial peptide. Bacteria were grown in N-minimal medium with 1 mM Mg^{2+} at pH 5.5 **(A–C)** or with 1 mM Mg^{2+} and 5 μg/ml C18G peptide at pH 7.5 **(D–F)**. **(A,D)** Immunoblot analysis of MgtC levels in wild-type (WT), ΔLD (DN557), and ΔLD *phoP*$^-$ (DN558) strains at 1 and 6 h after growth. The band indicated with an asterisk (*) corresponds to a protein displaying cross-reactivity against the anti-MgtC antibody and serves as an internal loading control. **(B,E)** Growth curves of wild-type, ΔLD, ΔLD Δ*mgtC* (HK111), and ΔLD Δ*mgtB* (DN581) strains. OD_{600} values were determined at the indicated time points. Means and standard deviations from three independent experiments are shown. **(C,F)** ATP levels were determined in wild-type (WT), ΔLD, ΔLD Δ*mgtC*, and ΔLD Δ*mgtB* strains at 4 h after growth. Data depicted in arbitrary units (AU) are means and standard deviations from three independent experiments. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, two-tailed *t*-test with each sample vs. WT; ns, not significant.

peptide that activates the PhoP regulator (Bader et al., 2005). Again, only the ΔLD strain induced detectable MgtC (**Figure 3D**) and displayed a growth defect dependent on MgtC and MgtB (**Figure 3E**). Moreover, the ΔLD strain displayed reduced ATP levels in a manner associated with MgtC and MgtB (**Figure 3F**). Together, these results indicate that even if *mgtC* transcription is activated in high Mg^{2+} by acidic pH or an antimicrobial peptide, the *mgtCBR* mRNA leader prevents dysregulated induction of MgtC and MgtB, which is necessary for *Salmonella* to normally grow in these environments.

MgtC and MgtB Impair Growth of the ΔLD Strain Inside Macrophages in a Manner Not Associated With the F_1F_0 ATP Synthase

The notion that MgtC induction is necessary for *Salmonella* to survive inside macrophages (Blanc-Potard and Groisman, 1997) led us to explore behaviors of the ΔLD strain inside macrophages. We first determined MgtC levels in cell extracts prepared from *Salmonella* engulfed by macrophages. After entry into macrophages, the WT strain induced detectable MgtC at 6 h but not at 1 h, whereas the ΔLD strain induced MgtC at both time points (**Figure 4A**). This result indicates that the *mgtCBR* mRNA leader prevents earlier induction of MgtC during *Salmonella* infection of macrophages.

We then examined the growth of *Salmonella* inside macrophages. Intracellular numbers of WT *Salmonella* increased at 6 h after engulfment by ∼8-fold and peaked at 12 h by ∼25-fold, and these numbers were maintained at 18 h (**Figure 4B**). By contrast, during the entire experiment, the ΔLD strain was unable to grow inside macrophages, resembling the phenotype of the Δ*mgtC* strain (**Figure 4B**).

We also determined intramacrophage survival of *Salmonella* mutants relative to the WT strain at 18 h post infection. The ΔLD strain displayed only ∼5% survival compared with the WT strain (**Figure 4C**). The ΔLD Δ*mgtC* strain showed a low survival similar to the ΔLD strain, whereas the ΔLD Δ*mgtB* strain exhibited WT levels of survival (**Figure 4C**). These findings suggest that the phenotype of the ΔLD strain is dependent on MgtB and that MgtC alone supports survival of the ΔLD strain unless MgtB is present. Furthermore, the similarly low survival between the ΔLD Δ*mgtC* and the ΔLD Δ*mgtCB* strains (**Figure 4C**) excludes the possibility that MgtB alone might impair survival of the ΔLD strain in the absence of MgtC. Together these results suggest that, when induced inside macrophages in a dysregulated manner, MgtC and MgtB impair intramacrophage growth of *Salmonella*.

The role of MgtC in supporting survival of *Salmonella* inside macrophages is due to its ability to inhibit the F_1F_0 ATP synthase (Lee et al., 2013). Consistent with this notion, *mgtC* deletion did not further impair the Δ*atpB* strain, which

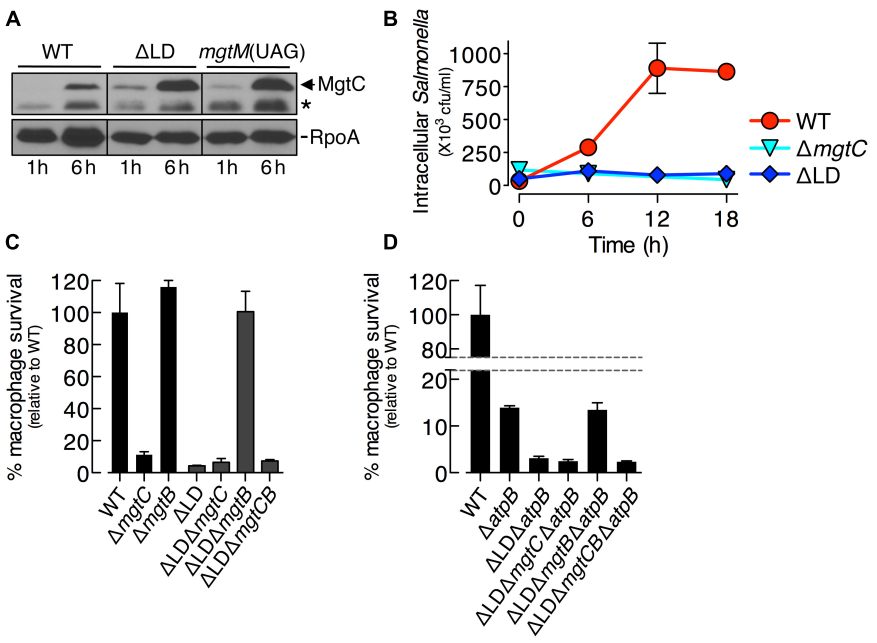


FIGURE 4 | Dysregulated induction of MgtC and MgtB impairs *Salmonella* inside macrophages. **(A)** Immunoblot analysis of MgtC and RpoA determined in the wild-type (WT, 14028s), ΔLD (DN557), and *mgtM*(UAG) (EG19307) strains at 1 and 6 h after engulfment by J774A.1 macrophages. The band indicated with an asterisk (*) corresponds to a protein displaying cross-reactivity against the anti-MgtC antibody. **(B)** Growth behaviors of wild-type, Δ*mgtC* (EN397), and ΔLD strains inside macrophages. After infection of macrophages, intracellular numbers of bacteria were determined at the indicated time points. **(C,D)** Intramacrophage survival of wild-type, Δ*mgtC*, Δ*mgtB* (EN481), ΔLD, ΔLD Δ*mgtC* (HK111), ΔLD Δ*mgtB* (DN581), ΔLD Δ*mgtCB* (DN608), Δ*atpB* (HK468), ΔLD Δ*atpB* (DN575), ΔLD Δ*mgtC* Δ*atpB* (MS575), ΔLD Δ*mgtB* Δ*atpB* (MS576), and ΔLD Δ*mgtCB* Δ*atpB* (DN686) strains. After infection of macrophages, the intracellular numbers of bacteria at 18 h were divided by those at 1 h. The percentage of survival of each mutant relative to the wild-type strain is presented. Means and standard deviations from three independent experiments are shown.

showed much lower survival than the WT strain (**Figure 4D**; Lee et al., 2013). By contrast, the ΔLD Δ*atpB* strain displayed more attenuated survival than the Δ*atpB* strain (**Figure 4D**). The ΔLD Δ*mgtC* Δ*atpB* strain showed a low survival similar to the ΔLD Δ*atpB* strain, whereas survival of the ΔLD Δ*mgtB* Δ*atpB* strain increased to levels of the Δ*atpB* strain (**Figure 4D**). Intramacrophage survival was similarly low between the ΔLD Δ*mgtC* Δ*atpB* and the ΔLD Δ*mgtCB* Δ*atpB* strains (**Figure 4D**). These results suggest that MgtC and MgtB impair intramacrophage growth of the ΔLD strain in a manner not associated with F_1F_0 ATP synthase. Moreover, given that MgtC and MgtB reduced ATP to abnormal levels independently of the F_1F_0 ATP synthase in the ΔLD strain (**Figure 2**), this result suggests that such ATP reduction could also inhibit growth of *Salmonella* in host environments.

A *Salmonella* Mutant Lacking *mgtM* Translation Displays Behaviors Resembling Those of the ΔLD Strain

The *mgtCBR* mRNA leader harbors the two short ORFs *mgtM* and *mgtP* (**Figure 1A**). Translation of *mgtM* and *mgtP* affects *mgtC* expression by affecting conformational changes in the leader (Lee and Groisman, 2012a,b; Lee et al., 2014). Thus, we hypothesized mutations in the leader that cause dysregulated induction of MgtC and MgtB might confer ΔLD strain-like phenotypes to *Salmonella*. We focused on the role of *mgtM*, as mutations preventing normal *mgtM* translation greatly increased β-galactosidase activity produced by *Salmonella* carrying a chromosomal *mgtC–lacZ* fusion (Lee and Groisman, 2012a).

We determined levels of MgtC and MgtB in the *mgtM*(UAG) strain, in which the start codon UUG of *mgtM* was replaced with the amber stop codon UAG (Lee and Groisman, 2012a). When the *mgtM*(UAG) strain was placed in medium with 10 μM Mg^{2+}, both MgtC and MgtB were induced to detectable levels as early as at 1 h, and these levels were maintained for the following 7 h (**Figure 5A**), resembling the temporal production of MgtC and MgtB in the ΔLD strain (**Figure 1D**). Four hours after growth, the *mgtM*(UAG) strain produced MgtC at levels ∼2-fold higher than the WT strain and ∼2-fold lower than the ΔLD strain (**Figure 5B**). However, MgtB levels were similar among the three strains (**Figure 5B**).

We also found that the *mgtM*(UAG) strain exhibited growth defects in all host and non-host environments that promote *mgtC* transcription. When placed in medium with 10 μM Mg^{2+}, the *mgtM*(UAG) strain entered logarithmic growth later and displayed lower growth yields during logarithmic growth than the WT strain (**Figure 5C**). However, deleting *mgtC* and *mgtB* corrected the logarithmic growth of the *mgtM*(UAG) strain (**Figure 5C**). The growth defect was also observed in medium with 1 mM Mg^{2+} and pH 5.5 as well as in medium with 1 mM Mg^{2+} and the C18G antimicrobial peptide (**Figures 5D,E**). Again, deletions of *mgtC* and *mgtB* restored growth of the *mgtM*(UAG) strain under these conditions (**Figures 5D,E**).

Survival of the *mgtM*(UAG) strain inside macrophages was only ∼10% relative to the WT strain (**Figure 5F**). While *mgtB* deletion increased survival of the *mgtM*(UAG) strain to WT levels, deletion of *mgtC* or both *mgtC* and *mgtB* slightly decreased survival of the strain to similar levels (**Figure 5F**). These results suggest that MgtC and MgtB impair growth of the *mgtM*(UAG) strain inside macrophages. Taken together, our results emphasize that the function of the *mgtCBR* mRNA leader in preventing dysregulated induction of MgtC and MgtB is necessary for *Salmonella* to proliferate in both host and non-host environments.

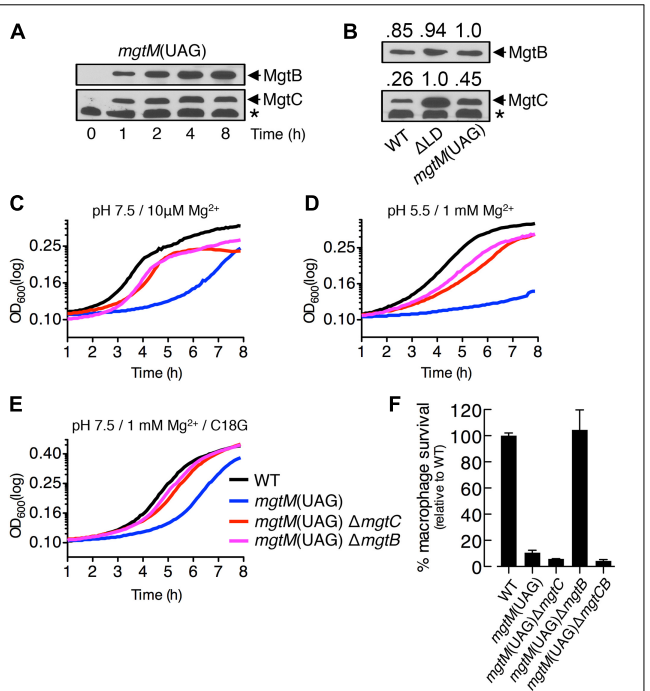


FIGURE 5 | A conformation of the *mgtCBR* mRNA leader that causes dysregulated induction of MgtC and MgtB impairs *Salmonella* in host and non-host environments. **(A,B)** Immunoblot analysis of MgtC and MgtB in wild-type (WT, 14028s), ΔLD (DN557), and *mgtM*(UAG) (EG19307) strains. Bacteria were grown in N-minimal medium with 10 μM Mg^{2+} at pH 7.5 and harvested at the indicated time points **(A)** or at 4 h **(B)**. The band indicated with an asterisk (*) corresponds to a protein displaying cross-reactivity against the anti-MgtC antibody. In panel **(B)**, numbers above the blots correspond to relative levels of MgtC and MgtB in each lane. **(C–E)** Growth curves of wild-type, *mgtM*(UAG) (EG19307), *mgtM*(UAG) Δ*mgtC* (DN649), and *mgtM*(UAG) Δ*mgtB* (DN612) strains. Bacteria were grown in N-minimal medium with 10 μM Mg^{2+} at pH 7.5 **(C)**, with 1 mM Mg^{2+} at pH 5.5 **(D)**, or with 1 mM Mg^{2+} and 5 μg/ml C18G peptide at pH 7.5 **(E)**, and OD_{600} values were determined every 5 min using a plate reader. Data are representative of three independent experiments, which gave similar results. **(F)** Intramacrophage survival of wild-type, *mgtM*(UAG), *mgtM*(UAG) Δ*mgtC*, *mgtM*(UAG) Δ*mgtB*, and *mgtM*(UAG) Δ*mgtCB* (DN652) strains was determined as described in the legend of **Figure 4**. Means and standard deviations from three independent experiments are shown.

DISCUSSION

The *mgtCBR* mRNA leader, which contains two short ORFs for *mgtM* and *mgtP*, possesses tandem attenuators that sense two distinct signals (Lee and Groisman, 2012a,b; Lee et al., 2014). In response to acidic pH, an increase in ATP levels

in the cytoplasm affects coupling/uncoupling of transcription of the leader with *mgtM* translation, inducing a conformation of the leader that promotes expression of the *mgtC*-coding region (Lee and Groisman, 2012a). Likewise, a decrease of intracellular proline promotes *mgtC* expression by affecting the events between leader transcription and *mgtP* translation (Lee and Groisman, 2012b; Lee et al., 2014).

In this study, we sought to further understand the physiological significance of the *mgtCBR* mRNA leader. We investigated two *Salmonella* mutants: the ΔLD strain, in which the sequences specifying the *mgtCBR* mRNA leader were removed (**Figure 1A**), and the *mgtM*(UAG) strain, in which the start codon of *mgtM* was replaced with a stop codon, resulting in the locking of the leader in a conformation allowing expression of the *mgtC*-coding region (Lee and Groisman, 2012a). Of note, MgtC expression was not constitutive but still inducible in these two strains; the ΔLD strain produced detectable levels of MgtC in 10 μM Mg^{2+} but not in 10 mM Mg^{2+} (**Figure 1B**), and MgtC was not detected in the *mgtM*(UAG) strain immediately at transfer from 10 mM to 10 μM Mg^{2+} (**Figure 5A**).

When *Salmonella* is placed in 10 μM Mg^{2+}, the PhoP regulator promotes transcription initiation of *mgtC* (Shin and Groisman, 2005). However, this event alone does not ensure MgtC production. *Salmonella* logarithmically grows for the first few hours in 10 μM Mg^{2+} by consuming Mg^{2+} in the medium and then displays slow linear growth (**Figure 1C**; Soncini et al., 1996; Blanc-Potard and Groisman, 1997; Park et al., 2018). MgtC production is detectable after *Salmonella* enter into the slow-growth phase (**Figures 1C,D**; Yeom et al., 2017; Park et al., 2018). The decrease of cytoplasmic Mg^{2+} levels causes the onset of slow-growth phase (Pontes et al., 2016). The observation that MgtC induction is further delayed when *Salmonella* is grown in 50 μM Mg^{2+} (Park et al., 2018) further supports the relation between decreased cytoplasmic Mg^{2+} levels and MgtC induction. In contrast, both the ΔLD and *mgtM*(UAG) strains produced MgtC at detectable levels even during the time period corresponding to the logarithmic growth phase of the WT strain (**Figures 1C,D**, **5A**). During this time period, growth yields of the two mutants, which both highly produced the Mg^{2+} transporter MgtB (Snavely et al., 1991b; **Figures 1D**, **5A**), were lower than that of the WT (**Figures 1C**, **5C**), suggesting that cytoplasmic Mg^{2+} levels in the mutants could be higher than the WT levels. Together, these results suggest that the ΔLD and *mgtM*(UAG) strains induce MgtC and MgtB in a dysregulated manner under conditions in which low cytoplasmic Mg^{2+} stress does not exist.

The F_1F_0 ATP synthase inhibitory protein MgtC (Lee et al., 2013) contributes to growth of WT *Salmonella* in low Mg^{2+} (Blanc-Potard and Groisman, 1997; Rang et al., 2007). MgtC directly acts on the F_1F_0 ATP synthase and reduces ATP levels, while MgtB has no effect (**Figure 2A**; Lee et al., 2013). Abnormally high ATP levels cause a growth defect of the *mgtC* mutant in low Mg^{2+} (Pontes et al., 2015). By contrast, in the ΔLD strain, MgtC and MgtB reduced ATP at levels lower than the WT levels in a process that did not require F_1F_0 ATP synthase (**Figure 2**), which appeared to impair bacterial growth (**Figures 1E**, **2C** and **Supplementary Figure S5**).

The ΔLD and *mgtM*(UAG) strains produced ∼4- and ∼2-fold higher levels of MgtC than the WT, respectively, whereas MgtB levels were similar among the strains (**Figure 5B**). However, growth of the ΔLD strain was more impaired than the *mgtM*(UAG) strain (**Figures 2A**, **5**). These findings suggest that overproduction of MgtC might impair growth of *Salmonella* in whatever environment it finds itself. However, this scenario is unlikely because of the following observations. The regulatory peptide MgtR specified by the *mgtCBR* operon binds to and promotes MgtC degradation (Alix and Blanc-Potard, 2008). Although the Δ*mgtR* strain produced ∼4-fold higher levels of MgtC than the WT (**Supplementary Figure S6A**), the mutant grew in a manner similar to the WT strain in 10 μM Mg^{2+} (**Supplementary Figure S6B**). The Δ*mgtR* strain also overproduced MgtC only in the slow-growth phase (**Supplementary Figure S6A**), resembling the temporal production of MgtC in the WT strain (**Figure 1D**). Moreover, even if MgtC was produced in a dysregulated manner, its inhibitory effect was not observed unless MgtB is produced in a dysregulated manner. We engineered the Δ*mgtC* and ΔLD Δ*mgtC* strains to express the *mgtC* ORF from the plasmid-linked *lac* promoter. When placed in medium with 10 μM Mg^{2+} and the same concentrations of IPTG, the two strains produced comparable levels of MgtC after 1 h (**Supplementary Figure S7A**). However, the Δ*mgtC* and ΔLD Δ*mgtC* strains produced MgtB in manners similar to the WT and ΔLD strains, respectively (**Supplementary Figure S7A**). Furthermore, ATP reduction and growth defect resembling the phenotypes of ΔLD strain were observed in the ΔLD Δ*mgtC* strain but not in the Δ*mgtC* strain (**Supplementary Figures S7B,C**). Together, all of these lines of evidence reinforce that dysregulated expression of MgtC and MgtB leads to ATP reduction, which impairs *Salmonella* growth.

The molecular basis of how the dysregulated induction of MgtC and MgtB reduces ATP levels independently of the F_1F_0 ATP synthase is currently unclear. MgtC affects membrane potential in a manner not associated with the ability to inhibit the F_1F_0 ATP synthase, suggesting that MgtC also affects ATP levels by acting on a protein(s) other than the F_1F_0 ATP synthase (Lee et al., 2013). The Δ*mgtC* strain harbors a hyperpolarized membrane, whereas *mgtC* overexpression causes membrane depolarization in WT *Salmonella* (Lee et al., 2013). Interestingly, *mgtB* overexpression in WT *Salmonella* also depolarizes a membrane (Lee et al., 2013). Moreover, the MgtB transporter is one of proteins that are crosslinked to the MgtC protein (Lee et al., 2013), suggesting direct interaction between these two proteins. Based on these lines of evidence, we hypothesize that when produced in a dysregulated manner, MgtC and MgtB together act on the protein(s) that affects membrane potential to cause membrane depolarization, which in turn reduces ATP to abnormal levels.

To survive inside the macrophage phagosome, *Salmonella* induces *mgtC* expression (Blanc-Potard and Groisman, 1997). MgtC acts on the F_1F_0 ATP synthase to reduce intracellular ATP, otherwise, ATP levels increase upon acidic pH inside the phagosome (Lee et al., 2013). The coupling/uncoupling of transcription of the *mgtCBR* mRNA leader with the translation of *mgtM* and *mgtP* responds to distinct intracellular cues (i.e., an

increase in ATP and a decrease in proline, respectively), inducing a conformation of the leader that allows transcription elongation into the *mgtC*-coding region (Lee and Groisman, 2012a,b; Lee et al., 2014). These events occur independently but additively to each other and enable *Salmonella* to achieve *mgtC* expression at levels that ensure its survival inside macrophages (Lee et al., 2014). In addition to these findings, our data suggest that preventing dysregulated induction of MgtC by the *mgtCBR* mRNA leader is another important determinant for intramacrophage survival of *Salmonella*. During infection of macrophages, WT *Salmonella* induced detectable MgtC at 6 h but not at 1 h (**Figure 4A**), which is in good agreement with the previous detection of MgtC at 20 h but no detection at 5 h post-infection (Yeom et al., 2018). By contrast, the ΔLD and *mgtM*(UAG) strains induced MgtC at the early time point and showed greatly attenuated survival inside macrophages (**Figures 4**, **5**).

The role of *cis*-acting regulatory RNAs in preventing dysregulated gene expression also impacts the pathogenesis of another bacterial species. In *Yersinia pseudotuberculosis*, the *yscW–lcrF* operon specifies the transcriptional regulator LcrF (Skurnik and Toivanen, 1992). The intergenic region of the *yscW–lcrF* mRNA functions as a thermosensor, forming a secondary structure that permits translation of *lcrF* at host body temperature (Bohme et al., 2012). Mutations that destabilize this structure enable LcrF production at lower temperatures, resulting in LcrF-dependent expression of virulence genes and attenuated bacterial virulence in mice (Bohme et al., 2012). However, the molecular basis underlying the virulence attenuation remains unknown. We determined that the Mg^{2+} transporter MgtB, which is not normally associated with the virulence phenotype (**Figure 4C**; Blanc-Potard and Groisman, 1997), impairs intramacrophage survival of the ΔLD and *mgtM*(UAG) strains in the presence of MgtC (**Figures 4C**, **5F**). Moreover, given that the ΔLD strain exhibited the defective phenotype in the absence of normal F_1F_0 ATP synthase function (**Figure 4D**), these results suggest that the unusual ATP reduction by dysregulated expression of MgtC and MgtB could impair growth of *Salmonella* inside host cells.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MP, HK, DN, and DS conceived the research. MP, HK, and DN performed the research. MP, HK, DN, D-HK, and DS analyzed the data. MP, HK, D-HK, and DS wrote the manuscript.

FUNDING

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01057723 and NRF-2018R1D1A1B07042525).

ACKNOWLEDGMENTS

We thank Dr. Eun-Jin Lee for kindly providing us the *S. enterica* strain EG19307 and anti-MgtB antibodies.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02831/full#supplementary-material

REFERENCES

Alix, E., and Blanc-Potard, A. B. (2008). Peptide-assisted degradation of the *Salmonella* MgtC virulence factor. *EMBO J.* 27, 546–557. doi: 10.1038/sj.emboj.7601983

Bader, M. W., Sanowar, S., Daley, M. E., Schneider, A. R., Cho, U., Xu, W., et al. (2005). Recognition of antimicrobial peptides by a bacterial sensor kinase. *Cell* 122, 461–472. doi: 10.1016/j.cell.2005.05.030

Blanc-Potard, A. B., and Groisman, E. A. (1997). The *Salmonella* selC locus contains a pathogenicity island mediating intramacrophage survival. *EMBO J.* 16, 5376–5385. doi: 10.1093/emboj/16.17.5376

Bohme, K., Steinmann, R., Kortmann, J., Seekircher, S., Heroven, A. K., Berger, E., et al. (2012). Concerted actions of a thermo-labile regulator and a unique intergenic RNA thermosensor control Yersinia virulence. *PLoS Pathog.* 8:e1002518. doi: 10.1371/journal.ppat.1002518

Choi, E., Choi, S., Nam, D., Park, S., Han, Y., Lee, J. S., et al. (2017). Elongation factor P restricts *Salmonella*'s growth by controlling translation of a Mg2+ transporter gene during infection. *Sci. Rep.* 7:42098. doi: 10.1038/srep42098

Choi, E., Lee, K. Y., and Shin, D. (2012). The MgtR regulatory peptide negatively controls expression of the MgtA Mg2+ transporter in *Salmonella enterica* serovar Typhimurium. *Biochem. Biophys. Res. Commun.* 417, 318–323. doi: 10.1016/j.bbrc.2011.11.107

Choi, J., and Groisman, E. A. (2016). Acidic pH sensing in the bacterial cytoplasm is required for *Salmonella* virulence. *Mol. Microbiol.* 101, 1024–1038. doi: 10.1111/mmi.13439

Choi, Y., Choi, J., Groisman, E. A., Kang, D. H., Shin, D., and Ryu, S. (2012). Expression of STM4467-encoded arginine deiminase controlled by the STM4463 regulator contributes to *Salmonella enterica* serovar Typhimurium virulence. *Infect. Immun.* 80, 4291–4297. doi: 10.1128/IAI.00880-12

Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A* 97, 6640–6645. doi: 10.1073/pnas.120163297

Davis, R. W., Bolstein, D., and Roth, J. R. (1980). *Advanced Bacterial Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Fields, P. I., Groisman, E. A., and Heffron, F. (1989). A *Salmonella* locus that controls resistance to microbicidal proteins from phagocytic cells. *Science* 243(4894 Pt 1), 1059–1062. doi: 10.1126/science.2646710

Fields, P. I., Swanson, R. V., Haidaris, C. G., and Heffron, F. (1986). Mutants of *Salmonella typhimurium* that cannot survive within the macrophage are avirulent. *Proc. Natl. Acad. Sci. U.S.A.* 83, 5189–5193. doi: 10.1073/pnas.83.14.5189

Foster, J. W., and Hall, H. K. (1990). Adaptive acidification tolerance response of *Salmonella typhimurium*. *J. Bacteriol.* 172, 771–778. doi: 10.1128/jb.172.2.771-778.1990

Garcia Vescovi, E., Soncini, F. C., and Groisman, E. A. (1996). Mg2+ as an extracellular signal: environmental regulation of *Salmonella* virulence. *Cell* 84, 165–174. doi: 10.1016/s0092-8674(00)81003-x

Groisman, E. A. (2001). The pleiotropic two-component regulatory system PhoP-PhoQ. *J. Bacteriol.* 183, 1835–1842. doi: 10.1128/jb.183.6.1835-1842.2001

Lee, E. J., Choi, J., and Groisman, E. A. (2014). Control of a *Salmonella* virulence operon by proline-charged tRNAPro. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3140–3145. doi: 10.1073/pnas.1316209111

Lee, E. J., and Groisman, E. A. (2010). An antisense RNA that governs the expression kinetics of a multifunctional virulence gene. *Mol. Microbiol.* 76, 1020–1033. doi: 10.1111/j.1365-2958.2010.07161.x

Lee, E. J., and Groisman, E. A. (2012a). Control of a *Salmonella* virulence locus by an ATP-sensing leader messenger RNA. *Nature* 486, 271–275. doi: 10.1038/nature11090

Lee, E. J., and Groisman, E. A. (2012b). Tandem attenuators control expression of the *Salmonella* mgtCBR virulence operon. *Mol. Microbiol.* 86, 212–224. doi: 10.1111/j.1365-2958.2012.08188.x

Lee, E. J., Pontes, M. H., and Groisman, E. A. (2013). A bacterial virulence protein promotes pathogenicity by inhibiting the bacterium's own F1Fo ATP synthase. *Cell* 154, 146–156. doi: 10.1016/j.cell.2013.06.004

Lee, J. W., and Lee, E. J. (2015). Regulation and function of the *Salmonella* MgtC virulence protein. *J. Microbiol.* 53, 667–672. doi: 10.1007/s12275-015-5283-1

Maloy, S. R., and Nunn, W. D. (1981). Selection for loss of tetracycline resistance by *Escherichia coli*. *J. Bacteriol.* 145, 1110–1111.

Park, M., Nam, D., Kweon, D. H., and Shin, D. (2018). ATP reduction by MgtC and Mg2+ homeostasis by MgtA and MgtB enables *Salmonella* to accumulate RpoS upon low cytoplasmic Mg2+ stress. *Mol. Microbiol.* 110, 283–295. doi: 10.1111/mmi.14105

Pontes, M. H., Lee, E. J., Choi, J., and Groisman, E. A. (2015). *Salmonella* promotes virulence by repressing cellulose production. *Proc. Natl. Acad. Sci. U.S.A.* 112, 5183–5188. doi: 10.1073/pnas.1500989112

Pontes, M. H., Yeom, J., and Groisman, E. A. (2016). Reducing ribosome biosynthesis promotes translation during low Mg2+ stress. *Mol. Cell.* 64, 480–492. doi: 10.1016/j.molcel.2016.05.008

Prost, L. R., Daley, M. E., Le Sage, V., Bader, M. W., Le Moual, H., Klevit, R. E., et al. (2007). Activation of the bacterial sensor kinase PhoQ by acidic pH. *Mol. Cell.* 26, 165–174. doi: 10.1016/j.molcel.2007.03.008

Rang, C., Alix, E., Felix, C., Heitz, A., Tasse, L., and Blanc-Potard, A. B. (2007). Dual role of the MgtC virulence factor in host and non-host environments. *Mol. Microbiol.* 63, 605–622. doi: 10.1111/j.1365-2958.2006.05542.x

Shin, D., and Groisman, E. A. (2005). Signal-dependent binding of the response regulators PhoP and PmrA to their target promoters in vivo. *J. Biol. Chem.* 280, 4089–4094. doi: 10.1074/jbc.m412741200

Skurnik, M., and Toivanen, P. (1992). LcrF is the temperature-regulated activator of the yadA gene of *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*. *J. Bacteriol.* 174, 2047–2051. doi: 10.1128/jb.174.6.2047-2051.1992

Snavely, M. D., Gravina, S. A., Cheung, T. T., Miller, C. G., and Maguire, M. E. (1991a). Magnesium transport in *Salmonella typhimurium*. Regulation of mgtA and mgtB expression. *J. Biol. Chem.* 266, 824–829.

Snavely, M. D., Miller, C. G., and Maguire, M. E. (1991b). The mgtB Mg2+ transport locus of *Salmonella typhimurium* encodes a P-type ATPase. *J. Biol. Chem.* 266, 815–823.

Soncini, F. C., Garcia Vescovi, E., Solomon, F., and Groisman, E. A. (1996). Molecular basis of the magnesium deprivation response in *Salmonella typhimurium*: identification of PhoP-regulated genes. *J. Bacteriol.* 178, 5092–5099. doi: 10.1128/jb.178.17.5092-5099.1996

Soncini, F. C., Vescovi, E. G., and Groisman, E. A. (1995). Transcriptional autoregulation of the *Salmonella typhimurium* phoPQ operon. *J. Bacteriol.* 177, 4364–4371. doi: 10.1128/jb.177.15.4364-4371.1995

Yeom, J., Pontes, M. H., Choi, J., and Groisman, E. A. (2018). A protein that controls the onset of a *Salmonella* virulence program. *EMBO J.* 37:e96977. doi: 10.15252/embj.201796977

Yeom, J., Wayne, K. J., and Groisman, E. A. (2017). Sequestration from protease adaptor confers differential stability to protease substrate. *Mol Cell* 66, 234–246.e5. doi: 10.1016/j.molcel.2017.03.009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 February 2020
doi: 10.3389/fgene.2020.00011

Novel Combined Tissue Transcriptome Analysis After Lentogenic Newcastle Disease Virus Challenge in Inbred Chicken Lines of Differential Resistance

Melissa S. Deist[1], *Rodrigo A. Gallardo*[2], *Jack C. M. Dekkers*[1], *Huaijun Zhou*[3] *and Susan J. Lamont*[1]*

[1] Department of Animal Science, Iowa State University, Ames, IA, United States, [2] Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, Davis, CA, United States, [3] Department of Animal Science, University of California, Davis, Davis, CA, United States

OPEN ACCESS

Edited by:
Androniki Psifidi,
Royal Veterinary College,
United Kingdom

Reviewed by:
Georges Nemer,
American University of Beirut,
Lebanon
Priyanka Baloni,
Institute for Systems Biology (ISB),
United States

***Correspondence:**
Susan J. Lamont
sjlamont@iastate.edu

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *19 October 2019*
Accepted: *06 January 2020*
Published: *04 February 2020*

Citation:
Deist MS, Gallardo RA, Dekkers JCM, Zhou H and Lamont SJ (2020) Novel Combined Tissue Transcriptome Analysis After Lentogenic Newcastle Disease Virus Challenge in Inbred Chicken Lines of Differential Resistance. Front. Genet. 11:11. doi: 10.3389/fgene.2020.00011

Disease has large negative impacts on poultry production. A more comprehensive understanding of host–pathogen interaction can lead to new and improved strategies to maintain health. In particular, host genetic factors can lead to a more effective response to pathogens, hereafter termed resistance. Fayoumi and Leghorn chicken lines have demonstrated relative resistance and susceptibility, respectively, to the Newcastle disease virus (NDV) vaccine strain and many other pathogens. This biological model was used to better understand the host response to a vaccine strain of NDV across three tissues and time points, using RNA-seq. Analyzing the Harderian gland, trachea, and lung tissues together using weighted gene co-expression network analysis (WGCNA) identified important genes that were co-expressed and associated with parameters including: genetic line, days post-infection (dpi), challenge status, sex, and tissue. Pathways and driver genes, such as *EIF2AK2*, *MPEG1*, and *TNFSF13B*, associated with challenge status, dpi, and genetic line were of particular interest as candidates for disease resistance. Overall, by jointly analyzing the three tissues, this study identified genes and gene networks that led to a more comprehensive understanding of the whole animal response to lentogenic NDV than that obtained by analyzing the tissues individually.

Keywords: chickens, RNA-seq, weighted gene co-expression network analysis, trachea, lung, Harderian gland

INTRODUCTION

Newcastle disease virus (NDV) is a single-stranded, negative-sense RNA virus. The clinical signs associated with NDV infection in chickens range from subclinical to 100% mortality depending on the virulence of the infection strain (Miller and Koch, 2013). NDV is classified into five pathotypes based on clinical signs: asymptomatic enteric, lentogenic, mesogenic, velogenic viscerotropic, and velogenic neurotropic (Afonso et al., 2012). Velogenic strains, also known in the United States as exotic NDV (END virus), have the most detrimental impacts on the poultry industry. Infection with

velogenic neurotropic NDV can result in neurological clinical signs, whereas velogenic viscerotropic strains cause intestinal lesions (Afonso et al., 2012). Both velogenic strains result in high morbidity and mortality (Afonso et al., 2012). The lentogenic strains are the least virulent, and because all NDV strains belong to the same serotype, lentogenic strains are used for vaccination (Dimitrov et al., 2017). Lentogenic NDV is limited in its replication to host cells that produce a trypsin-like protease to cleave the fusion protein (Nagai et al., 1976). This restricts lentogenic NDV replication to epithelial cells where this protease is expressed (Nagai, 1995). Hence, lentogenic strains have the largest impact near sites of infection such as the trachea, lung, and Harderian gland, all of which contain epithelial cells and are also involved in mucosal immunity. These tissues, in which the virus and host have direct contact, are ideal for examining host–pathogen interaction. Increased understanding of the host–pathogen interaction creates the potential for improving strategies to curb the negative impacts of NDV.

Two inbred chicken lines have been shown to differ in their susceptibility to lentogenic La Sota NDV (Deist et al., 2017b). The Leghorn line (GHs 6) originated from a cross of two commercial lines in the 1950s and the Fayoumi line (M 15.2) originated from the Fayoum region of Egypt (Lamont and Chen, 1992). The Fayoumi and Leghorn lines are each greater than 99.95% homozygous (Fleming et al., 2016) and have different MHC types 15.2 and 6, respectively. These lines have also shown differential responses to avian influenza (Wang et al., 2014), Marek's disease (Lakshmanan et al., 1996), *Salmonella* (Cheeseman et al., 2007), and *Eimeria* (Pinard-van der Laan et al., 2009). Inbred lines that differ in their relative susceptibility offer an excellent tool to study mechanisms of pathogen response.

The RNA sequencing (RNA-seq) approach shows the genes and pathways impacted by a treatment; however, it only shows a snapshot of response in a particular tissue and time. Because the transcriptome is not stagnant, a more comprehensive physiological picture can be obtained by an integrative analysis of the transcriptomes of multiple tissues at multiple time points after a treatment. Previously, the trachea epithelial cells, lung, and Harderian gland of Fayoumis and Leghorns at 2, 6, and 10 days post-infection (dpi) with lentogenic NDV have been individually analyzed with RNA-seq (Deist et al., 2017a; Deist et al., 2017b; Deist et al., 2018). The three tissues had very distinct responses to NDV (Deist et al., 2017a; Deist et al., 2017b; Deist et al., 2018). Taking a systems biology approach and analyzing the three tissues together will provide a more comprehensive picture of how the whole animal responded to NDV. Prior joint-tissue analyses using microarray or RNA-seq to study the impact of avian pathogenic *Escherichia coli* on broiler chickens have been performed (Sandford et al., 2012; Sun et al., 2016). Improvements in the technologies and tools for analyzing expression data now enable a more thorough examination of these complex data.

The objective of this study was to gain a more comprehensive understanding of the transcriptomic response of these two chicken lines to lentogenic NDV. Utilizing co-expression analysis across all three tissues identifies clusters of co-expressed genes that are associated with factors of interest such as NDV challenge, line, and dpi. We expect to identify novel genes and gene clusters that were not previously identified when the tissues were analyzed individually. Clusters of genes that are associated with NDV challenge and line may help identify genes potentially associated with resistance to lentogenic NDV. These data from multiple key tissues after a lentogenic challenge lay the foundation for future comparison of networks and pathways that may differ in experimental parameters including use of velogenic NDV challenge.

RESULTS

Principal Component Analysis of the Trachea, Lung, and Harderian Gland

The transcriptomes of three tissues from birds infected with lentogenic NDV were jointly analyzed to more comprehensively characterize the host response to NDV. Principal component analysis showed a clear separation between the tissues, as expected (**Figure 1**). Principal component (PC) 1 (PC1 = 51.91%) separated all three tissues and PC2 (32.11%) separated the trachea and Harderian gland from the lung (**Figure 1**). No clear clustering was observed by any other parameter in all tissues.

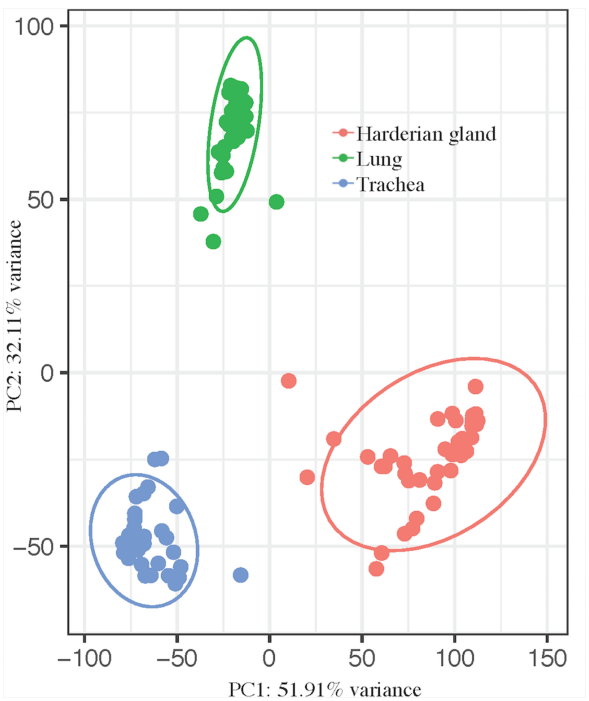


FIGURE 1 | Principal component analysis shows clear separation by tissue. pcaExplorer generated PCA plot using the top 1,000 most variant transcripts. Ellipses were drawn around groups with 95% confidence interval. Groups separated by tissue, lung (green), trachea (blue), and Harderian gland (pink).

Differential Expression Analysis

A differential expression analysis was performed to determine if there was a difference in transcript expression levels between tissues after challenge with lentogenic NDV. The numbers of differentially expressed genes (DEG) for the challenge by tissue interaction, within each line and dpi, are shown in **Table 1**. Large numbers of DEG were identified by the interaction between the lung and Harderian gland in the Fayoumis at 10 dpi (**Table 1**), and separately the large numbers of DEG in the Leghorns at 6 dpi (**Table 1**) may have resulted from large numbers of DEG due to challenge within an individual tissue (Deist et al., 2017a; Deist et al., 2018). With the exception of the Fayoumis at 10 dpi and the Leghorns at 6 dpi, there were few DEG between the lung and Harderian gland. Interestingly, there were no DEG at 10 dpi in the Leghorns and few DEG in the Fayoumis at 6 dpi (**Table 1**). Overall, there were differences in the number of DEG from each interaction contrast, depending on genetic line and time. Of the 18 contrasts examined, *CD8B* and *LY6D* were differentially expressed most frequently in seven and six contrasts, respectively. Full lists of DEG (FDR < 0.05) for each contrast are provided in **Table S1**.

TABLE 1 | Comparing the tissues' response to lentogenic challenge within each line and time.

Line	Time	Interaction contrast[a]	DEG[b]
Fayoumi	2 dpi	Lung vs. Harderian gland	10
		Trachea vs. Harderian gland	151
		Trachea vs. lung	135
	6 dpi	Lung vs. Harderian gland	0
		Trachea vs. Harderian gland	1
		Trachea vs. lung	3
	10 dpi	Lung vs. Harderian gland	1,523
		Trachea vs. Harderian gland	54
		Trachea vs. lung	20
Leghorn	2 dpi	Lung vs. Harderian gland	0
		Trachea vs. Harderian gland	141
		Trachea vs. lung	115
	6 dpi	Lung vs. Harderian gland	743
		Trachea vs. Harderian gland	877
		Trachea vs. lung	226
	10 dpi	Lung vs. Harderian gland	0
		Trachea vs. Harderian gland	0
		Trachea vs. lung	0

[a]*Within each line and dpi: (Tissue1, Challenged–Tissue1, Non-challenged)–(Tissue2, Challenged–Tissue2, Non-challenged).*
[b]*Differentially Expressed Genes (FDR < 0.05).*

Gene Co-Expression Analysis in the Trachea, Lung, and Harderian Gland After Lentogenic Challenge

The module–trait relationship analysis resulted in several associations between the modules generated by weighted gene co-expression network analysis (WGCNA) based on co-expression and factors of the study: line, dpi, challenge status, sex, and tissue (**Figure 2**). The lightyellow module strongly positively correlated with line (0.99 for Fayoumi = 1 versus Leghorn = 0, p = 1e−120), but did not significantly correlate with any other factor, suggesting that for the 161 transcripts within this module, there was no association between line and any other factor. The lightyellow module could help better characterize the inherent differences between these two inbred genetic lines. The orange module was strongly positively correlated with sex (0.97

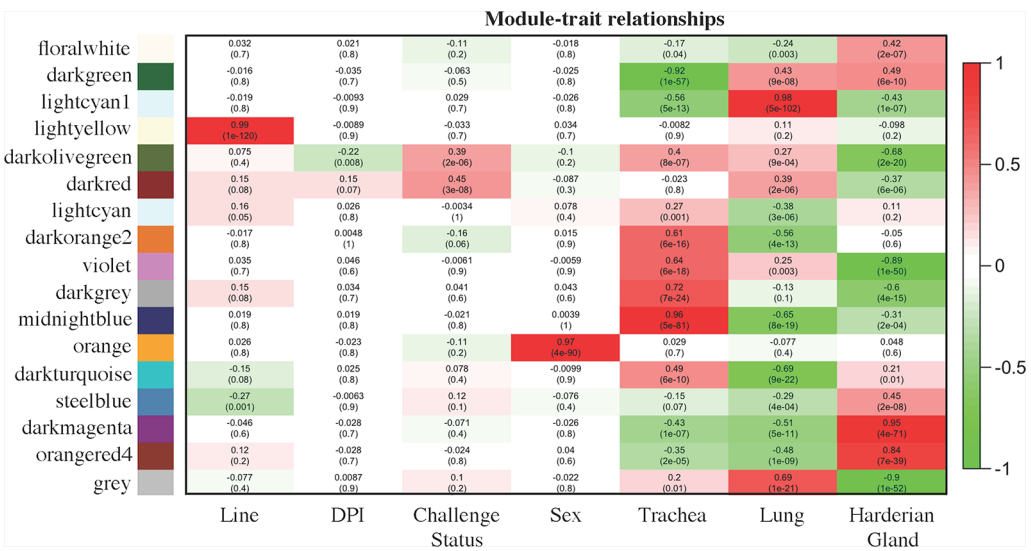


FIGURE 2 | Module–trait relationships calculated by WGCNA. Factors of interest (x-axis) were correlated to each module (y-axis). The first value within each square is the correlation and the second value in parenthesis is the p value of the association. The more positively correlated the module and the factor, the more red the square; the more negatively correlated, the more green. Coding for the factors was as follows: Line [Fayoumi (1), Leghorn (0)], days post-infection [dpi; 2 dpi (0), 6 dpi (1), 10 dpi (2)], NDV [challenged (1), non-challenged (0)], sex [male (1), female (0)], Trachea [Trachea (1), Lung (0), Harderian gland (0)], Lung [Trachea (0), Lung (1), Harderian gland (0)], and Harderian gland [Trachea (0), Lung (0), Harderian gland (1)].

for male = 1 versus female = 0, p = 1e−90). Within this module, 94 are on the Z chromosome, 7 are on the W chromosome, and 17 transcripts have not been assigned to a chromosome. Lack of correlations with other modules indicates that these transcripts are only impacted by sex.

Three modules, midnightblue, lightcyan1, and darkmagenta, may be related to the specific functions of each tissue. The 1,308 transcripts in the midnightblue module were strongly positively correlated and more highly expressed in the trachea (0.96 for trachea = 1 versus lung and Harderian gland = 0, p = 5e−81). The 1,720 transcripts in the lightcyan1 module were strongly positively correlated with and more highly expressed in the lung (0.98 for lung = 1 versus trachea and Harderian gland = 0, p = 5e−102). The large darkmagenta module (3,154 transcripts) strongly positively correlated with and more highly expressed in the Harderian gland (0.95 for Harderian gland = 1 versus trachea and lung = 0, p = 4e−71).

The darkolivegreen module was negatively correlated with dpi (−0.22 for 2 dpi = 0, 6 dpi = 1, versus 10 dpi = 2, p = 0.0008), positively correlated with challenge status (0.39 for challenged = 1 versus non-challenged = 0, p = 2e−06), and negatively correlated with the Harderian gland (−0.68 for Harderian gland = 1 versus trachea and lung = 0, p = 2e−20). The 80 transcripts in this module were likely important for the chicken's early response to NDV because the negative correlation with dpi indicates they were more highly expressed at early time points. The darkred module was positively correlated with challenge status (0.45 for challenged = 1 versus non-challenged = 0, p = 3e−08), positively correlated with the lung (0.39 for lung = 1 versus trachea and Harderian gland = 0, p = 2e−0.6), and negatively correlated with the Harderian gland (−0.37 for Harderian gland = 1 versus trachea and lung = 0, p = 6e−06). Although not significant, this module showed a weak correlation with line (0.15 for Fayoumi = 1 versus Leghorn = 0, p = 0.08). The darkred module potentially included genes related to response to NDV because of its associations with challenge status and line. The 333 transcripts in the steelblue module negatively correlated with line (−0.27 for Fayoumi = 1 versus Leghorn = 0, p = 0.001) and positively correlated with the Harderian gland (0.45 for Harderian gland = 1 versus trachea and lung = 0, p = 4e−71). Although not significant, this module weakly correlated with challenge status (0.12 for challenged = 1 versus non-challenged = 0, p = 0.1). The steelblue module may also include transcripts potentially associated with response to NDV.

Genes that are co-expressed are likely to be regulated by similar mechanisms. Within each module, the transcripts with high gene significance (GS) and module membership (MM), i.e., driver genes, were likely to have the most influence on the expression values of the entire module. The top driver genes were identified for the module that was most highly correlated with each factor (**Table 2**). Several long intervening/intergenic non-coding RNAs (lincRNAs) and unnamed protein coding genes were identified as driver genes (**Table 2**). The *NKX2-1* was among the top 10 bottom-loading genes for PC1, and *CLDN18* among the top 10 top-loading genes for PC2 calculated by pcaExplorer (Marini, 2016).

TABLE 2 | Top 3 driver genes from the module most highly correlated with each factor for birds infected with lentogenic NDV.

Factor / Module	Transcript ID / Gene Name	Gene significance	Module membership
Line/lightyellow	ENSGALT00000082473/ lincRNA	0.97	0.96
	ENSGALT00000065772/ lincRNA	−0.97	−0.97
	ENSGALT00000087010/ protein coding	0.96	0.97
DPI/darkolivegreen	ENSGALT00000017183/ *EIF2AK2*	−0.37	0.83
	ENSGALT00000016499/ *CMTR1*	−0.37	0.74
	ENSGALT00000066953/ protein coding	−0.35	0.90
Challenge status/ darkred	ENSGALT00000027228/ *TNFSF13B*	0.53	0.82
	ENSGALT00000044004/ *MPEG1*	0.52	0.82
	ENSGALT00000003281/ protein coding	0.51	0.74
Sex/orange	ENSGALT00000077789/ *Nipped-B homolog-like*	−0.98	−0.95
	ENSGALT00000047883/ protein coding	−0.98	−0.96
	ENSGALT00000080994/ protein coding	−0.97	−0.95
Trachea/ midnightblue	ENSGALT00000021346/*akr*	0.97	0.96
	ENSGALT00000049178/ *DNAL4*	0.97	0.95
	ENSGALT00000009017/*F3*	0.97	0.94
Lung/lightcyan1	ENSGALT00000089244/ protein coding	0.99	0.98
	ENSGALT00000010644/ *CLDN18*	0.99	0.97
	ENSGALT00000057049/ lincRNA	0.98	0.97
Harderian gland/ dark magenta	ENSGALT00000059647/ *LRRC30*	0.98	0.98
	ENSGALT00000006028/ *PABPC4*	0.97	0.94
	ENSGALT00000059800/ *NKX2-1*	−0.97	−0.94

Functional Analysis of Modules of Interest for Birds Infected With Lentogenic NDV

The darkolivegreen, darkred, and steelblue modules were of particular interest because of their correlations with either challenge status or line. The darkolivegreen module's correlations with dpi and challenge status suggest that these genes were important in the early response to NDV. Of the 80 transcripts, 59 had an associated gene name. No modules were significantly correlated with both the challenge status and line; however, the darkred and steelblue modules may include genes associated with response to NDV because of their correlations with challenge status or line. Gene ontology (GO) term and network analysis was performed on these three modules.

Panther identified significant top-level GO terms for the darkolivegreen, darkred, and steelblue modules (**Figure 3**). The significant GO terms for the darkolivegreen and darkred modules, both significantly correlated with challenge status,

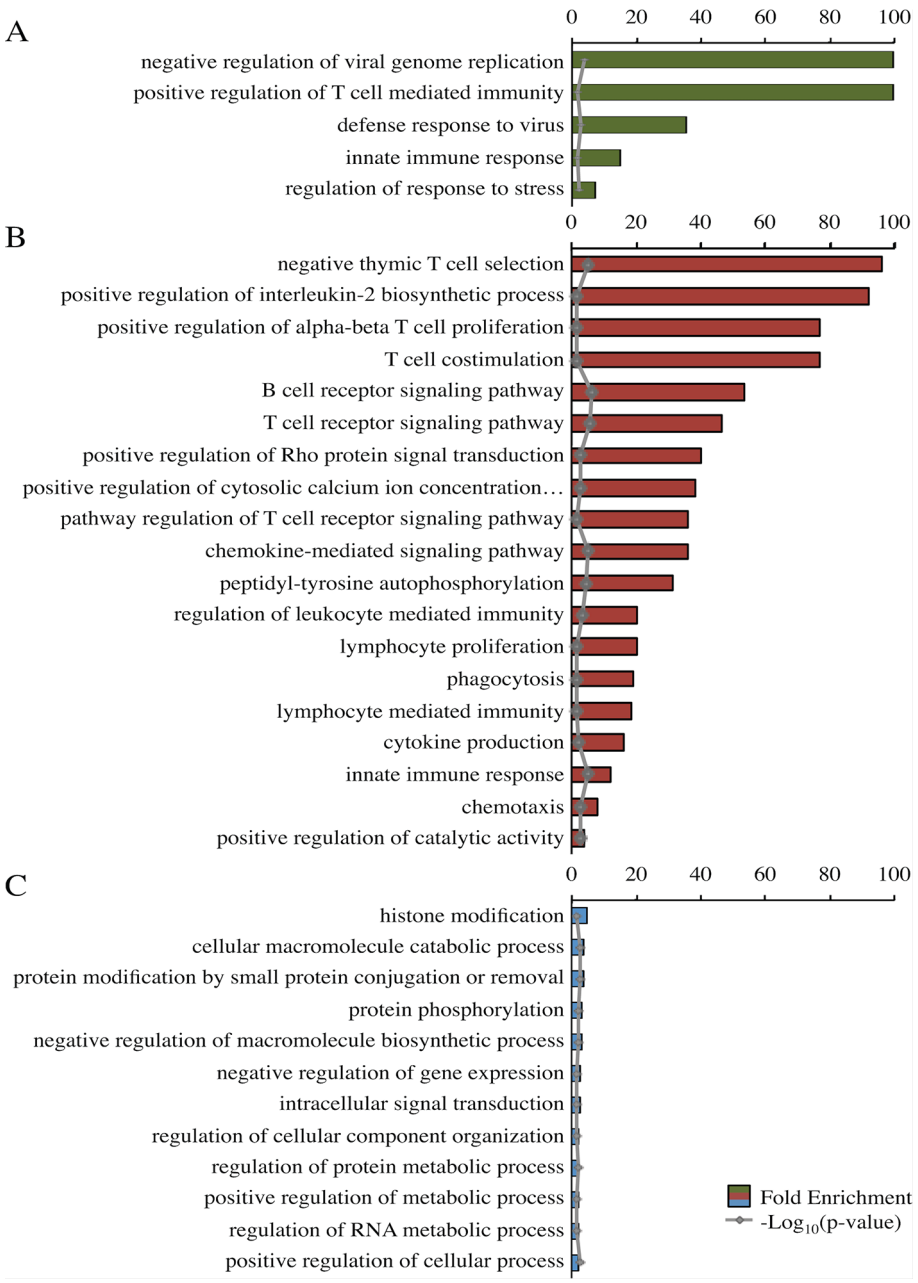


FIGURE 3 | Fold enrichment of top-level overrepresented GO terms within modules of interest. Panther overrepresentation test using associated gene names for each of three modules: darkolivegreen module **(A)**, darkred module **(B)**, and steel blue module **(C)**. Fold enrichment was capped at 100. –Log10(p value) shown in gray. All p values <0.05. Ellipses indicate those involved in phospholipase C-activating G-protein-coupled signaling.

were clearly immune related. The significant GO terms for the steelblue module were more general in nature (**Figure 3**).

To generate protein interaction networks for the transcripts in each module of interest, STRING was utilized (**Figure 4**). The darkolivegreen, darkred, and steelblue networks had significantly more connections than would be expected by chance. The darkolivegreen module included 55 nodes and 31 edges and was significantly associated with several KEGG pathways that are related to response to virus, including influenza A, Herpes simplex infection, and the RIG-I-like receptor signaling pathway (**Figure 4A**). Within the darkolivegreen module, *STAT1* had the second highest gene significance for challenge status and a high module membership (GS = 0.50, MM = 0.88), making the highly connected protein a potential driver gene. The darkred module network included 110 nodes and 60 edges and was significantly associated with the following KEGG pathways:

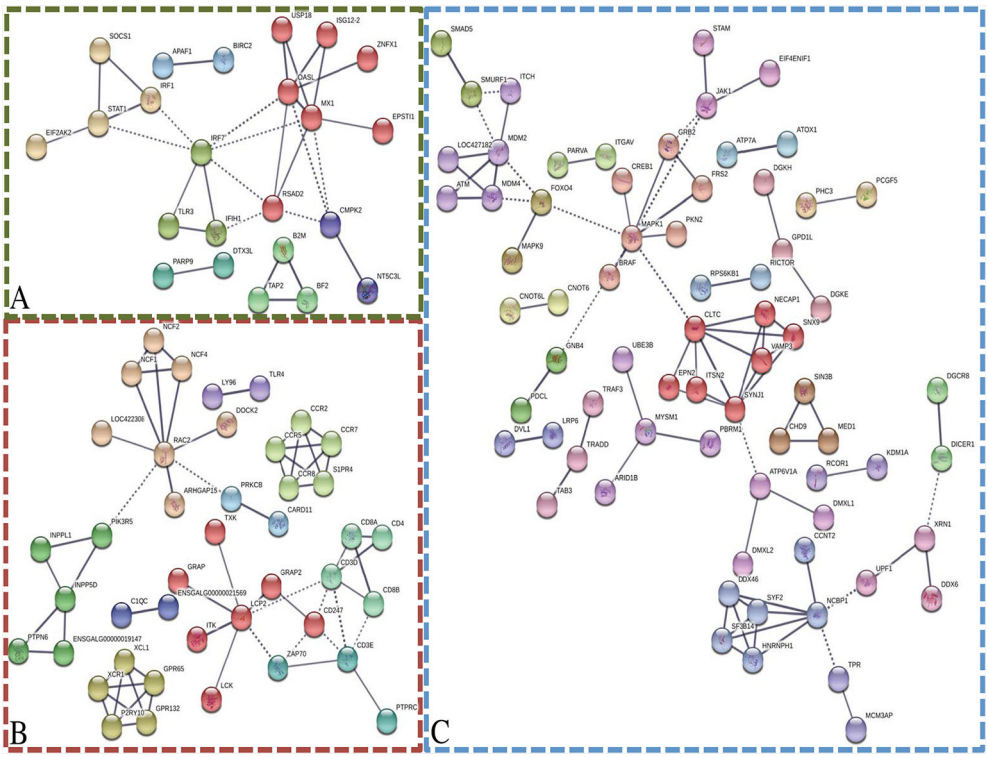


FIGURE 4 | Network analysis of modules of interest. STRING was utilized to generate networks based on the associated gene name of the transcripts within the modules of interest: darkolivegreen module **(A)**, darkred module **(B)**, and steelblue module **(C)**. Disconnected nodes were removed, high confidence (0.700), and MCL clustering (inflation = 3). Solid lines indicate connections within cluster and dotted lines indicate connections between clusters.

cytokine–cytokine receptor interaction, cell adhesion molecules (CAMs), phagosome, Jak-STAT signaling pathway, intestinal immune network for IgA production, and phosphatidylinositol signaling system. The fourth highest driver gene, *INPP5D* (GS = 0.49, MM = 0.84), was moderately well connected (**Figure 4B**). The network generated based on the steelblue module included 251 nodes and 78 edges and was significantly associated with only one KEGG pathway: endocytosis. *DGKE* and *RICTOR* were among the top 4 driver genes in the steelblue module based on gene significance for line (**Figure 4C**).

DISCUSSION

The combined tissue analysis of the current study provided a comprehensive picture of the Fayoumi and Leghorn transcriptomic response to lentogenic NDV challenge in tissues where the NDV vaccine replicates. The Harderian gland, lung, and tracheal epithelial cells come into direct contact with the lentogenic virus when infected *via* the eyes and/or airways and provide the best opportunity to examine the host–pathogen interaction. Previously reported analysis of each individual tissue's transcriptome resulted in unique findings for each tissue, as briefly summarized here. The trachea epithelial cells had the highest amount of detectable virus (Deist et al., 2017b). As time progressed, the number of DEG decreased between the challenged and non-challenged birds in the trachea, suggesting the trachea was consistently active in host response and in recovery from the virus in both lines (Deist et al., 2017b). The transcriptomes of the lung and Harderian gland were more impacted by line than by the challenge (Deist et al., 2017a; Deist et al., 2018). The Leghorn lung appeared to be non-responsive to the virus, whereas the Fayoumi lung produced a large number of DEG at 10 dpi when comparing the challenged to non-challenged birds (Deist et al., 2017a). The Harderian gland of the Leghorns showed a large number of DEG between the challenged and non-challenged birds at 6 dpi, and the Fayoumis had very few DEG at all time points (Deist et al., 2018). The challenge by tissue interaction resulted in more DEG if there were more DEG between challenged and non-challenged birds within a tissue. It is difficult to interpret the results of the challenge by tissue interaction because the tissues had time- and line-dependent responses to the virus as measured by DEG and had different amounts of detectable virus (Deist et al., 2017a; Deist et al., 2017b; Deist et al., 2018).

The complex design of the current combined, three-tissue experiment did not readily lend itself to a differential expression analysis because with three tissues, three time points, two genetic lines, and two challenge statuses, there was a very large number of possible contrasts. Using WGCNA for co-expression analysis does

not rely on specific contrasts (differential expression) and can identify correlations between co-expressed genes and factors of importance in the study design. Using WGCNA to analyze multiple tissues is a novel approach that was successful. Previously, WGCNA was used to analyze the lung tissue individually (Deist et al., 2017a), which had some results in concordance with the combined tissue analysis. ENSGALT00000087010 was among the top driver genes for genetic line in the lung analysis as well as the joint analysis. In addition, *Nipped-B homolog-like* and ENSGALT00000080994 were among the top driver genes for sex in both analyses.

The darkolivegreen module was composed of genes that had a clear functional role in host defense. The unannotated transcripts within the darkolivegreen module should be explored further to identify their functions because they were co-expressed with genes important to the host response to NDV. GO term analysis of the genes in the darkolivegreen module showed they were involved in the host response to virus. *EIF2AK2* (also known as *PKR*) was the top driver gene for dpi for the darkolivegreen module. *EIF2AK2* is interferon-inducible and has important antiviral effects (Sadler and Williams, 2008). The EIF2 signaling pathway is known to inhibit NDV replication (Zhang et al., 2014). This pathway was shown to be significantly different between the two lines in both the trachea (Deist et al., 2017b) and spleen transcriptomes (Zhang et al., 2018). *EIF2AK2* phosphorylates *EIF2α*, which regulates protein synthesis (Deb et al., 2001). Also, treatment with poly(I:C) led to expression changes of *EIF2AK2* in chickens (Kim and Zhou, 2015). The negative correlation of the darkolivegreen module with dpi suggests *EIF2AK2* is likely more highly expressed at earlier times post-infection. Differential expression confirmed that the challenged birds had significantly higher *EIF2AK2* expression levels at 2 dpi when compared to 6 or 10 dpi across all tissues (data not shown). At 2 dpi, the highest levels of viral transcripts were detected in the trachea (Deist et al., 2017b) and Harderian gland (Deist et al., 2018), as measured by RNA-seq; high expression of *EIF2AK2* at 2 dpi would aid in viral clearance.

The functions of the genes in the darkred module involve both the innate and adaptive immune system, as determined by the GO term analysis and the top 2 driver genes. It is well known that TNF genes play an important role in response to pathogens. *TNFSF13B* (also known as *BAFF*) specifically aids in survival and proliferation of chicken B cells (Schneider, 2004), and was the top driver gene for challenge status in the darkred module. *TNFSF13B* receptors are expressed predominantly on chicken B cells (Schneider, 2004). One receptor for *TNFSF13B*, *TNFRSF13B*, was among the genes significant in the trachea challenge by line interaction at 6 dpi (Deist et al., 2017b) and was differentially expressed between the Fayoumi and Leghorn chickens in the lung in both non-challenged birds and birds challenged with avian influenza (Wang et al., 2014). Antibody production is required for NDV clearance (Reynolds and Maraqa, 2000; Kapczynski et al., 2013), and expression of *TNFSF13B* may critically influence B cells and their antibody production in response to NDV.

Another driver gene for challenge status in the darkred module was *MPEG1*. *MPEG1* (also known as *Perforin-2*) is expressed highly in macrophages and is an important component of the innate immune system (McCormack and Podack, 2015). The protein encoded by *MPEG1* works with complement to form membrane attack complexes and create holes in a target cell's membrane (McCormack and Podack, 2015). *MPEG1* was 1 of 19 DEG in the Harderian gland (Deist et al., 2018) and 1 of 16 DEG in the lung (Deist et al., 2017a) between challenged and non-challenged Leghorns at 2 dpi. The infiltration of macrophages into tissues infected with NDV may result in an increase in RNA expression of MPEG1.

The two chicken lines have previously shown differences in the expression levels of *TNFSF13B* and *MPEG1* in response to viral pathogens, which may suggest a potential role in disease resistance (Deist et al., 2017b; Deist et al., 2018). *MPEG1* and *TNFSF13B*, however, were not included in the darkred network. This suggests that these proteins may have unknown or low confidence connections in STRING to the other proteins in the darkred module, or that co-expression of these genes was due to their functional similarity and that their proteins may not interact.

The steelblue module significantly correlated with chicken line and weakly correlated with challenge status. Although no immune-related GO terms were significant, network analysis revealed immune-related genes in this module, including *JAK*, *TRAF3*, *DICER1*, and *MAPK1*. This network was also associated with the endocytosis pathway, which is a mechanism by which NDV enters the host cell. It may be possible for a host to modify gene expression of this pathway to minimize viral entry. These inbred chicken lines have been used as models in multiple studies; further exploration of the lightyellow module will help to identify differences between the two lines.

The combined analysis of three target tissues generated a more extensive understanding of response to infection with lentogenic NDV at the whole organism level. Other studies have compared differences between the expression of key immune genes in these genetic lines after an NDV challenge in embryo (Schilling et al., 2018; Schilling et al., 2019) and at 3 weeks of age under heat stress conditions (Saelao et al., 2018). Also, genetic markers of interest have been identified in a commercial line (Rowland et al., 2018; Saelao et al., 2019) and African ecotypes (Walugembe et al., 2019) after infection with this same lentogenic strain. Data are being generated at a fast rate; new tools and types of analyses are needed to analyze the data in a more comprehensive way. The current study attempted to identify critical genes associated with response to NDV based on RNA-seq data from three different tissues. This new bioinformatics approach was limited by number of tissues and number of individuals per tissue. Further studies to confirm these findings are necessary. Differences between the lines demonstrated how host genetics impacts vaccine response. The darkred and steelblue modules were the modules most likely to contain genes that relate to pathogen response, and *EIF2AK2*, *MPEG1*, and *TNFSF13B* are promising candidate genes for lentogenic NDV resistance. These were not identified as candidate genes in the individual tissue analyses. Future studies are required to determine how the expression of these genes is impacted by velogenic challenge in the Fayoumi and Leghorn. This study demonstrated how WGCNA could be used to efficiently combine transcriptome data from different tissues, genetic lines, challenge statuses, and dpi

to identify important gene clusters associated with factors of interest. Novel clusters and driver genes were identified in this analysis that can be used as a reference for comparison with future studies that use different NDV strains, as well as serving as a basis for investigations of host modulation of vaccine efficacy.

MATERIALS AND METHODS

This study used data from experiments approved by the Iowa State University Institutional Animal Care and Use Committee (IACUC log number 1-13-7490-G). All experiments were performed in accordance with the committee's relevant guidelines and regulations. The experimental design for the lentogenic NDV challenge has been previously described in detail (Deist et al., 2017b). Briefly, 3-week-old chicks were either inoculated with 10^7 EID_{50} La Sota NDV (challenged) or given PBS (non-challenged) *via* an ocular–nasal route. The relatively resistant Fayoumi and relatively susceptible Leghorn were intermixed within the challenged and non-challenged groups. Blocking for challenge status and line, at each time of 2, 6, and 10 dpi, one-third of the birds were euthanized for tissue collection. This design resulted in a total of 12 treatment groups: two challenge groups (challenged and non-challenged), two lines (Fayoumi and Leghorn), and three time points (2, 6, and 10 dpi).

Using the RNAqueous kit Total RNA Isolation Kit (Thermo Fisher Scientific, Waltham, MA, USA), RNA was isolated from the trachea epithelial cells, lungs, and Harderian glands of about four birds per treatment group. The RNA was treated with DNAse and the quality of each sample was confirmed (RQN > 8). The Illumina TruSeq RNA Library Prep Kit v2 (Illumina, San Diego, CA, USA) was used to generate cDNA libraries. Libraries were sequenced on the HiSeq2500 for 100-bp, single-end reads at the Iowa State University DNA Facility (Ames, IA, USA). The sequencing data was analyzed with a standard pipeline and the *Gallus-gallus* 5 reference genome (Gal5; GCA_000002315.3): FASTX, TopHat2 (Kim et al., 2013), and HTSeq (Anders et al., 2015). Transcripts with less than four counts across all samples were removed prior to normalization. The RNA-seq data are publicly available from ArrayExpress at EMBL-EBI under accession numbers: E-MTAB-5431, E-MTAB-5859, and E-MTAB-6038.

For data visualization of the 143 sequenced samples, pcaExplorer (Marini, 2016) was used to generate PCA plots based on count data normalized using DESeq2 (Love et al., 2014), accounting for tissue and treatment group. The variance associated with each principal component was calculated with the 1,000 most variant transcripts (Marini, 2016).

Differential expression analysis was performed to determine how the tissues responded differently to the NDV challenge using the GLM in edgeR (Robinson et al., 2010), accounting for every combination of tissue, line, dpi, and challenge status resulting in 36 levels. Within each line and dpi, the following contrasts were written to identify transcripts that were significantly impacted (FDR < 0.05) by the interaction between challenge and tissue: (Tissue i, Line j, dpi k, Challenged–Tissue i, Line j, dpi k, Non-challenged)–(Tissue i', Line j, dpi k, Challenged–Tissue i', Line j, dpi k, Non-challenged). With three tissues, two lines, and three time points, a total of 18 contrasts were analyzed.

Co-expression analysis was performed using WGCNA (Langfelder and Horvath, 2008). WGCNA clusters genes into modules based on expression levels and correlates a module's eigengene to traits of interest. All transcripts with more than four counts across all samples were included. Transcript counts were normalized using the variance stabilizing transformation by DESeq2 (Love et al., 2014). The soft power threshold was set to 20 and a minimum module size of 30 was used (Langfelder and Horvath, 2008). The module's eigengenes were correlated to factors of interest. Coding for the factors was as follows: Line [Fayoumi (1), Leghorn (0)], days post-infection [dpi; 2 dpi (0), 6 dpi (1), 10 dpi (2)], NDV [challenged (1), non-challenged (0)], sex [male (1), female (0)], Trachea [Trachea (1), Lung (0), Harderian gland (0)], Lung [Trachea (0), Lung (1), Harderian gland (0)], and Harderian gland [Trachea (0), Lung (0), Harderian gland (1)]. For each factor, the driver genes for a module were based on the transcripts with the highest absolute gene significance and module membership calculated by WGCNA (Langfelder and Horvath, 2008). Gene significance was estimated by correlating a transcript's expression profile with a sample trait. A correlation of a module eigengene and the expression profile of each transcript was used to determine a transcript's module membership score. All reported p values were taken directly from the WGCNA output.

GO term and network analyses were performed to ascertain the predicted function of specific modules. All transcript IDs were converted to their associated gene name using Ensembl BioMart, then NCBI and Uniprot were used to find gene names for those transcripts unidentified by Ensembl. Panther (Mi et al., 2017) was utilized for GO term overrepresentation analysis using a Bonferroni correction. Only the top-level GO terms, based on the hierarchical labeling by Panther, were presented (**Figure 3**). Associated gene names were used as input into STRING (Szklarczyk et al., 2017) for network analysis based on protein–protein interactions described in the literature. Disconnected nodes were removed and a high confidence (0.700) was required for connection of the nodes. The nodes were colored based on MCL clustering (inflation = 3) (Szklarczyk et al., 2017).

DATA AVAILABILITY STATEMENT

The RNA-seq data are publicly available from ArrayExpress at EMBL-EBI under accession numbers: E-MTAB-5431, E-MTAB-5859, and E-MTAB-6038.

ETHICS STATEMENT

This study used data from experiments approved by the Iowa State University Institutional Animal Care and Use Committee (IACUC log number 1-13-7490-G). All experiments were performed in accordance with the committee's relevant guidelines and regulations.

AUTHOR CONTRIBUTIONS

MD: Conceptualization of RNA-seq experiment, investigation, collected samples, isolated RNA, constructed cDNA libraries, formal analysis of RNA-seq, methodology, data visualization, writing—original draft preparation. RG: Conceptualization of experiment, experimental design, investigation—prepared viral isolate for inoculation, writing—review and editing. JD: Conceptualization of experiment, experimental design, writing—review and editing. HZ: Conceptualization of experiment, experimental design, funding acquisition, project administration, writing—review and editing. SL: Conceptualization of experiment, experimental design, provided resources (chicken genetic lines), investigation, funding acquisition, project administration, supervised analysis of data, writing—review and editing. All authors have read and approved the manuscript.

FUNDING

This study was made possible by the generous support of the American people through the United States Agency for International Development (USAID) Feed the Future Innovation Lab for Genomics to Improve Poultry (cooperative agreement number AID-OAA-A-13-00080). The contents are the responsibility of the Feed the Future Innovation Lab for Genomics to Improve Poultry and do not necessarily reflect the views of USAID or the United States Government. Melissa Deist was partially supported by USDA NIFA 2013-38,420-20,496.

ACKNOWLEDGMENTS

We thank the S. J. Lamont lab group for their help with sample collection and bird monitoring, especially Michael Kaiser and Kaylee Rowland. The authors thank Claudio Afonso, Patti Miller, Kiril Dimitrov, Tim Olivier, and Dawn Williams-Coplin for reviewing and editing a previous version of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00011/full#supplementary-material

SUPPLEMENTARY TABLE S1 | The S1 Table (table 1.xlsx) includes the Ensembl transcript ID and corresponding log fold change and FDR value for all transcripts for all 18 challenge*line contrasts.

REFERENCES

Afonso, C. L., Miller, P. J., Grund, C., Koch, G., Peeters, B. P., Selleck, P. W., et al. (2012). *Manual of diagnostic tests and vaccines for terrestrial animals* (Paris, France: World Organisation for Animal Health).

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31 (2), 166–169. doi: 10.1093/bioinformatics/btu638

Cheeseman, J. H., Kaiser, M. G., Ciraci, C., Kaiser, P., and Lamont, S. J. (2007). Breed effect on early cytokine mRNA expression in spleen and cecum of chickens with and without Salmonella enteritidis infection. *Dev. Comp. Immunol.* 31 (1), 52–60. doi: 10.1016/j.dci.2006.04.001

Deb, A., Haque, S. J., Mogensen, T., Silverman, R. H., and Williams, B. R. G. (2001). RNA-dependent protein kinase PKR is required for activation of NFKB by IFN-γ in a STAT1-independent pathway. *J. Immunol.* 166 (10), 6170–6180. doi: 10.4049/jimmunol.166.10.6170

Deist, M. S., Gallardo, R. A., Bunn, D. A., Dekkers, J. C. M., Zhou, H., and Lamont, S. J. (2017a). Resistant and susceptible chicken lines show distinctive responses to Newcastle disease virus infection in the lung transcriptome. *BMC Genomics* 18 (1), 989. doi: 10.1186/s12864-017-4380-4

Deist, M. S., Gallardo, R. A., Bunn, D. A., Kelly, T. R., Dekkers, J. C. M., Zhou, H., et al. (2017b). Novel mechanisms revealed in the trachea transcriptome of resistant and susceptible chicken lines following infection with Newcastle disease virus. *Clin. Vaccine Immunol.* 24 (5), 1–17. doi: 10.1128/CVI.00027-17

Deist, M. S., Gallardo, R. A., Bunn, D. A., Kelly, T. R., Dekkers, J. C. M., Zhou, H., et al. (2018). Novel analysis of the Harderian gland transcriptome response to Newcastle disease virus in two inbred chicken lines. *Sci. Rep.* 8 (1), 6558. doi: 10.1038/s41598-018-24830-0

Dimitrov, K. M., Afonso, C. L., Yu, Q., and Miller, P. J. (2017). Newcastle disease vaccines-A solved problem or a continuous challenge? *Vet. Microbiol.* 206, 126–136. doi: 10.1016/j.vetmic.2016.12.019

Fleming, D. S., Koltes, J. E., Fritz-Waters, E. R., Rothschild, M. F., Schmidt, C. J., Ashwell, C. M., et al. (2016). Single nucleotide variant discovery of highly inbred Leghorn and Fayoumi chicken breeds using pooled whole genome resequencing data reveals insights into phenotype differences. *BMC Genomics* 17 (1), 812. doi: 10.1186/s12864-016-3147-7

Kapczynski, D. R., Afonso, C. L., and Miller, P. J. (2013). Immune responses of poultry to Newcastle disease virus. *Dev. Comp. Immunol.* 41 (3), 447–453. doi: 10.1016/j.dci.2013.04.012

Kim, D., Pertea, G., Trapnel, C., Pimental, H., Kelley, R., and S. L, S. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, 13. Kim et al., 2013 doi: 10.1186/gb-2013-14-4-r36

Kim, T. H., and Zhou, H. (2015). Functional analysis of chicken IRF7 in response to dsRNA analog poly(I:C) by integrating overexpression and knockdown. *PloS One* 10 (7), e0133450. doi: 10.1371/journal.pone.0133450

Lakshmanan, N., Kaiser, M. G., and Lamont, S. J. (1996). "Marek's disease resistance in MHC-congenic lines from Leghorn and Fayoumi breeds," in *Current Research on Marek's Disease*. Eds. R. F. Silva, H. H. Cheng, P. M. Coussens, L. F. Lee and L. F. Velicer (Kennett Square, PA: American Association of Avian Pathologists), 57–62.

Lamont, S. J., and Chen, Y. (1992). Endogenous viral genes in thirteen highly inbred chicken lines and lines selected for immune response traits. *Poult. Sci.* 71, 530–538. doi: 10.3382/ps.0710530

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. doi: 10.1186/s13059-014-0550-8

Marini, F. (2016). "pcaExplorer: interactive visualization of RNA-seq data using a principal components approach.")

McCormack, R., and Podack, E. R. (2015). Perforin-2/Mpeg1 and other pore-forming proteins throughout evolution. *J. Leukocyte Biol.* 98 (5), 761–768. doi: 10.1189/jlb.4MR1114-523RR

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., et al. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* 45 (D1), D183–D189. doi: 10.1093/nar/gkw1138

Miller, P. J., and Koch, G. (2013). "Newcastle disease, other avian paramyxoviruses, and avian metapneumovirus infections," in *Diseases of Poultry, 13th ed.* Eds. D. E. Swayne, J. R. Glisson, L. R. McDougald, L. K. Nolan, D. L. Suarez and V. Nair. (Ames, IA. :John Wiley & Sons, Ltd), 89–138.

Nagai, Y., Klenk, H., and Rott, R. (1976). Proteolytic cleavage of the viral glycoproteins and its significance for the virulence of Newcastle disease virus. *Virology* 72, 494–508. doi: 10.1016/0042-6822(76)90178-1

Nagai, Y. (1995). Virus activation by host proteinases. a pivotal role in the spread of infection, tissue tropism and pathogenicity. *Microbiol. Immunol.* 39 (1), 1–9. doi: 10.1111/j.1348-0421.1995.tb02161.x

Pinard-van der Laan, M. H., Bed'hom, B., Coville, J. L., Pitel, F., Feve, K., Leroux, S., et al. (2009). Microsatellite mapping of QTLs affecting resistance to coccidiosis (Eimeria tenella) in a Fayoumi x White Leghorn cross. *BMC Genomics* 10, 13. doi: 10.1186/1471-2164-10-31

Reynolds, D. L., and Maraqa, A. D. (2000). Protective immunity against Newcastle disease: the role of cell-mediated immunity. *Avian Dis.* 44 (1), 145–154. doi: 10.2307/1592518

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616

Rowland, K., Saelao, P., Wang, Y., Fulton, J. E., Liebe, G. N., McCarron, A. M., et al. (2018). Association of candidate genes with response to heat and newcastle disease virus. *Genes (Basel)* 9 (11), 1–14. doi: 10.3390/genes9110560

Sadler, A. J., and Williams, B. R. (2008). Interferon-inducible antiviral effectors. *Nat. Rev. Immunol.* 8 (7), 559–568. doi: 10.1038/nri2314

Saelao, P., Wang, Y., Chanthavixay, G., Yu, V., Gallardo, R. A., Dekkers, J. C. M., et al. (2018). Integrated proteomic and transcriptomic analysis of differential expression of chicken lung tissue in response to NDV infection during heat stress. *Genes (Basel)* 9 (12), 1–16. doi: 10.3390/genes9120579

Saelao, P., Wang, Y., Chanthavixay, G., Gallardo, R. A., Wolc, A., Dekkers, J. C. M., et al. (2019). Genetics and genomic regions affecting response to newcastle disease virus infection under heat stress in layer chickens. *Genes (Basel)* 10 (1), 1–17. doi: 10.3390/genes10010061

Sandford, E. E., Orr, M., Li, X., Zhou, H., Johnson, T. J., Kariyawasam, S., et al. (2012). Strong concordance between transcriptomic patterns of spleen and peripheral blood leukocytes in response to avian pathogenic Escherichia coli infection. *Avian Dis.* 56 (4), 732–736. doi: 10.1637/10261-060512-Reg.1

Schilling, M. A., Katani, R., Memari, S., Cavanaugh, M., Buza, J., Radzio-Basu, J., et al. (2018). Transcriptional innate immune response of the developing chicken embryo to newcastle disease virus infection. *Front. Genet.* 9, 61. doi: 10.3389/fgene.2018.00061

Schilling, M. A., Memari, S., Cavanaugh, M., Katani, R., Deist, M. S., Radzio-Basu, J., et al. (2019). Conserved, breed-dependent, and subline-dependent innate immune responses of Fayoumi and Leghorn chicken embryos to Newcastle disease virus infection. *Sci. Rep.* 9 (1), 7209. doi: 10.1038/s41598-019-43483-1

Schneider, K. (2004). Chicken BAFF–a highly conserved cytokine that mediates B cell survival. *Int. Immunol.* 16 (1), 139–148. doi: 10.1093/intimm/dxh015

Sun, H., Bi, R., Liu, P., Nolan, L. K., and Lamont, S. J. (2016). Combined analysis of primary lymphoid tissues' transcriptomic response to extra-intestinal Escherichia coli (ExPEC) infection. *Dev. Comp. Immunol.* 57, 99–106. doi: 10.1016/j.dci.2015.12.013

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45 (D1), D362–D368. doi: 10.1093/nar/gkw937

Walugembe, M., Mushi, J. R., Amuzu-Aweh, E. N., Chiwanga, G. H., Msoffe, P. L., Wang, Y., et al. (2019). Genetic analyses of tanzanian local chicken ecotypes challenged with newcastle disease virus. *Genes* 10 (7), 1–16. doi: 10.3390/genes10070546

Wang, Y., Lupiani, B., Reddy, S. M., Lamont, S. J., and Zhou, H. (2014). RNA-seq analysis revealed novel genes and signaling pathway associated with disease resistance to avian influenza virus infection in chickens. *Poult. Sci.* 93 (2), 485–493. doi: 10.3382/ps.2013-03557

Zhang, S., Sun, Y., Chen, H., Dai, Y., Zhan, Y., Yu, S., et al. (2014). Activation of the PKR/eIF2α signaling cascade inhibits replication of Newcastle disease virus. *Virol. J.* 11 (62), 1–11. doi: 10.1186/1743-422X-11-62

Zhang, J., Kaiser, M. G., Deist, M. S., Gallardo, R. A., Bunn, D. A., Kelly, T. R., et al. (2018). Transcriptome analysis in spleen reveals differential regulation of response to newcastle disease virus in two chicken lines. *Sci. Rep.* 8 (1), 1278. doi: 10.1038/s41598-018-19754-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 07 February 2020
doi: 10.3389/fmicb.2019.03104

Genomic Analysis of Putative Virulence Factors Affecting Cytotoxicity of *Cronobacter*

Jinghua Cui[1,2], Jinrui Hu[2], Xiaoli Du[2], Chao Yan[1], Guanhua Xue[1], Shaoli Li[1], Zhigang Cui[2], Hua Huang[3] and Jing Yuan[1]**

[1] Department of Bacteriology, Capital Institute of Pediatrics, Beijing, China, [2] State Key Laboratory of Infectious Disease Prevention and Control, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Disease, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China, [3] Beijing Products Quality Supervision and Inspection Institute, Beijing, China

Cronobacter spp. can cause systemic infections, such as meningitis, sepsis, and necrotizing enterocolitis, in immunocompromised patients, especially neonates. Although some virulence factors have been reported previously, the pathogenesis of *Cronobacter* remains unclear. In this study, we compared genome sequences from different *Cronobacter* species, sequence types, and sources, with the virulence genes in the virulence factor database. The results showed that *Cronobacter* has species specificity for these virulence genes. Additionally, two gene clusters, including *sfp* encoding fimbriae and *hly* encoding hemolysin, were discovered. Through cell adhesion, cytotoxicity, and hemolysis assays, we found that the isolates possessing the two gene clusters had higher cytotoxicity and stronger hemolysis capacity than those of other isolates in this study. Moreover, analysis of type VI secretion system (T6SS) cluster and putative fimbria gene clusters of *Cronobacter* revealed that T6SS have species specificity and isolates with high cytotoxicity possessed more complete T6SS cluster construction than that of the rest. In conclusion, the two novel gene clusters and T6SS cluster were involved in the mechanism underlying the cytotoxicity of *Cronobacter*.

Keywords: *Cronobacter* spp., cytotoxicity, hemolysin, fimbria (pilus), species specificity

OPEN ACCESS

Edited by:
Dong Xia,
Royal Veterinary College (RVC), United Kingdom

Reviewed by:
Ozan Gundogdu, University of London, United Kingdom
Angelika Lehner, University of Zurich, Switzerland

****Correspondence:***
Hua Huang
MarsHV@163.com
Jing Yuan
Yuanjing6212@163.com

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *28 September 2019*
Accepted: *23 December 2019*
Published: *07 February 2020*

Citation:
Cui J, Hu J, Du X, Yan C, Xue G, Li S, Cui Z, Huang H and Yuan J (2020) Genomic Analysis of Putative Virulence Factors Affecting Cytotoxicity of Cronobacter. Front. Microbiol. 10:3104. doi: 10.3389/fmicb.2019.03104

INTRODUCTION

Cronobacter spp. can cause systemic infections in immunocompromised patients, especially neonatal meningitis, sepsis, and necrotizing enterocolitis in neonates. In 2018, the journal *Emerging Infectious Disease* published four reports on *Cronobacter*, which showed new research topic focusing on *Cronobacter* (Zeng et al., 2015; Chaves et al., 2018; McMullan et al., 2018; Morato-Rodríguez et al., 2018). To date, seven different species of *Cronobacter* have been identified, namely *Cronobacter sakazakii*, *Cronobacter malonaticus*, *Cronobacter turicensis*, *Cronobacter muytjensii*, *Cronobacter dublinensis*, *Cronobacter universalis*, and *Cronobacter condimenti* (Joseph et al., 2012b). These bacteria were observed to adhere to different epithelial cell lines and even persist in macrophage cells (Mange et al., 2006; Townsend et al., 2008). An outer membrane protein called OmpA is essential in the adhesion process of *C. sakazakii* (Singamsetty et al., 2008; Mittal et al., 2009), and the genes *ompX* and *inv* work synergistically with *ompA* contributing the virulence (Kim et al., 2010; Chandrapala et al., 2014). Enterotoxin production by the bacteria was initially evaluated

by the suckling mouse assay, and some strains were found to possess the capacity to produce enterotoxins. Moreover, these bacteria have been found to cause cytotoxicity in CHO, Vero, and Y-1 cells (Pagotto et al., 2003). However, after the study conducted by Pagotto et al. (2003), no study analyzing the cytotoxicity of *Cronobacter* was published, until 2017, when Singh et al. (2017) reported the cytotoxicity of *C. sakazakii* N81 in Caco-2 cells. Cytotoxicity is an important virulence factor for bacteria, and toxins and exoenzymes play an important role in bacterial cytotoxicity.

Hemolysins are cytotoxic proteins that target cell membranes, and their mechanisms of damaging membrane integrity can be classified into the following three major groups: enzymatic activity, pore-forming cytolysin, or surfactant (Leclercq et al., 2016). Cytotoxicity induced by *Streptococcus agalactiae*, also known as group B streptococcus, was mediated entirely by the bacterial β-hemolysin/cytolysin (Leclercq et al., 2016). In *Staphylococcus aureus*, α-hemolysin cytotoxic to HL60 human promyelocytic leukemia cells was purified (Tsuiji et al., 2019). Hemolysin in *Vibrio vulnificus* exhibits powerful hemolytic and cytolytic activities and contributes to bacterial invasion from the intestine to the blood stream (Elgaml and Miyoshi, 2017). Although it is known that a gene encoding hemolysin exists in *Cronobacter* (Cruz et al., 2011; Singh et al., 2017), whether this hemolysin is active and associated with cytotoxicity, has not yet been elucidated.

In addition, secretion system and fimbriae also assist the cytotoxicity of bacteria. Type VI secretion system (T6SS) genomic island in *Citrobacter freundii* strain CF74 is involved in the adherence to host cells and induces cytotoxicity in host cells (Liu et al., 2015). The genome construction of T6SS in *Cronobacter* has been revealed previously (Wang et al., 2018), but further studies evaluating its cytotoxicity are required. In *Pseudomonas aeruginosa*, type IV pili are required for promoting close contact between bacteria and the host cell so that exolysin (ExlA) can exert its cytotoxic activity. The interaction of ExlA with membranes results in pore formation, followed by lactic dehydrogenase (LDH) release, and eventually death of infected eukaryotic cells (Basso et al., 2017). *Proteus mirabilis* MR/P fimbriae and flagella mediate genotoxic and cytotoxic effects on eukaryotic cells, at least *in vitro* (Scavone et al., 2015). Purified cbl pili were found to directly induce cytotoxicity in *Burkholderia cenocepacia* strain A549 and activate cysteine proteinases involved in apoptosis (Cheung et al., 2007). Although 10 putative fimbria gene clusters were identified in *Cronobacter*, these genes were not present or functional in any *Cronobacter* species (Joseph et al., 2012a). For example, curli fimbriae in *Cronobacter* were reported to be involved in biofilm formation and cell–cell aggregation (Hu, 2018), whereas *C. sakazakii* and *C. muytjensii* strains were devoid of curli fimbriae genes.

Virulence factor database (VFDB[1]) is an integrated and comprehensive online resource for curating information about virulence factors of bacterial pathogens. In the present study, we compared genome sequences of *Cronobacter* from different species, sequence types (STs), and sources with the virulence genes in VFDB, to determine species specificity and elucidate the association between the virulence-related genes and phenotype.

MATERIALS AND METHODS

Strain Isolation

The stains of *Cronobacter* in this study were isolated from food, water, spice, anus swab, blood, and cerebrospinal fluid (CSF) supplied by 12 province Centers for Disease Control and Prevention of China. The study about clinical strains isolated from blood and CSF has been published (Cui et al., 2017), and ethics approval had been submitted to the Journal. This research was approved by Ethics Committee of National Institute for Communicable Disease Control and Prevention. Thirty-one strains included 15 *C. sakazakii*, 7 *C. malonaticus*, 5 *C. dublinensis*, 2 *C. turicensis*, and 2 *C. muytjensii*. The detailed information is listed in **Table 1**. The strain *Escherichia coli* HB101 was as negative control. For cell assays, bacterial cultures were prepared by inoculating tryptone soya broth (TSB) medium (Oxiod, United Kingdom) with single colonies grown on tryptone soya agar plates and incubating them aerobically for 16–18 h at 37°C. For logarithmic subcultures, the cultures were diluted 1:100 into fresh TSB and incubated for 4 h at 37°C with shaking (200 rpm). For the growth phase experiments, subcultures were obtained as described above except that the cultures were incubated up to 24 h.

DNA Preparation and Sequencing

The genomes of *Cronobacter* strains were sequenced by BGI Tech Solutions Co. Ltd., Beijing, China. Sequences were generated with an Illumina HiSeq2000 (Illumina Inc., San Diego, CA, United States). Quality trimming of 150-nucleotide (nt) paired-end reads, produced from a 500-bp genomic library, and subsequent assembly were performed using SOAP *de novo* v1.05 (Li et al., 2010). Raw data were processed in four steps involving the removal of reads with ambiguous bases (1–90 bp), 20 bp of low quality reads (≤Q20), adapter contamination, and duplicated reads.

Bioinformatics Analysis

DNA sequences were obtained in this study and were compared with reference sequences in NCBI database, using BLAST software 2.2.22 of tblastx program with amino acid sequence. The sequences were then screened by identity value (>50%), and heat maps constructed with coverage value. The heat maps were prepared with pheatmap package in R 3.5.1 and cluster analysis performed with ward.D method. Sequences in all the strains were compared with those of the virulence genes in the virulence factor database (VFDB, see text footnote 1), which is an integrated and comprehensive online resource for curating information about virulence factors of bacterial pathogens. The respective virulence genes from five species of *Cronobacter* were analyzed with Venn diagram[2].

[1]http://www.mgc.ac.cn/VFs/

[2]http://www.ehbio.com/ImageGP/index.php/Home/Index/VennDiagram.html

TABLE 1 | The information of *Cronobacter* isolates in this study.

Strain no.	Species	Source	Isolation time	ST	CC	Crispr 1 no.	Genome accession no.
AH02	*C. sakazakii*	Food	2012	21	21	2	RPAZ00000000
CQ04	*C. sakazakii*	Stool	2012	13	13	1	RPBC00000000
HA03	*C. sakazakii*	Food	2012	209		1	RPBD00000000
HB04	*C. sakazakii*	CSF	2014	83	83	1	LGRL00000000
HE35	*C. sakazakii*	Food	2012	64	64	1	RPBE00000000
LN01	*C. sakazakii*	Food	2006	23	23	1	RPBF00000000
LN02	*C. sakazakii*	Food	2006	235	4	1	RPBG00000000
SC16	*C. sakazakii*	Food	2011	4	4	1	RPBJ00000000
SC25	*C. sakazakii*	Food	2010	1	1	1	RPBK00000000
SC26	*C. sakazakii*	Stool	2015	42		1	RPBI00000000
SD04	*C. sakazakii*	Spice	2010	144	40	1	RPBL00000000
SD19	*C. sakazakii*	Water	2010	151		1	RPBP00000000
SD45	*C. sakazakii*	Stool	2011	40	40	1	RPCB00000000
SD47	*C. sakazakii*	Food	2011	155	155	1	RPBZ00000000
XJ02	*C. sakazakii*	Food	2012	12		1	RPBV00000000
BJ15	*C. malonaticus*	Stool	2013	7	7	1	RPBB00000000
HB03	*C. malonaticus*	Blood	2014	60		1	LGRM00000000
SC01	*C. malonaticus*	Food	2010	201	7	1	RPBH00000000
SD11	*C. malonaticus*	Water	2010	169		2	RPBN00000000
SD16	*C. malonaticus*	Water	2010	7	7	1	RPBO00000000
SD26	*C. malonaticus*	Water	2010	159	7	1	RPBR00000000
SD43	*C. malonaticus*	Water	2010	160	62	1	RPCA00000000
BJ09	*C. dublinensis*	Food	2007	239		1	RPBA00000000
SD05	*C. dublinensis*	Plant	2010	167		2	RPBM00000000
SD28	*C. dublinensis*	Water	2010	173		1	RPBS00000000
SD69	*C. dublinensis*	Spice	2016	43		0	RPBY00000000
SX10	*C. dublinensis*	Food	2016	80	80	5	RPBW00000000
SD21	*C. turicensis*	Water	2010	171	147	1	RPBQ00000000
SH11	*C. turicensis*	Stool	2013	309		1	RPBX00000000
SD83	*C. muytjensii*	Spice	2016	489		0	RPBU00000000
SD92	*C. muytjensii*	Spice	2016	493		1	RPBT00000000

Nucleotide Sequence Accession Numbers

This Whole Genome Shotgun project has been deposited at GenBank under the BioProject PRJNA287482 and PRJNA498360, with accession numbers LGRM00000000, LGRL00000000, RPAZ00000000, RPBA00000000-RPBZ00000000, RPCA00000000, and RPCB00000000.

Growth Curve

Isolates were incubated overnight and then transferred into 20 ml of Dulbecco's modified Eagle's medium (DMEM) at ratio of 1:100. The isolates were then grown at 37°C with shaking at 180 rpm/min. OD 600 was measured every hour for each isolate.

HEp-2 Adhesion and Cytotoxicity Assay

The human epidermoid laryngocarcinoma (HEp-2) cell line was obtained from the National Institute for Viral Disease Control and Prevention, China CDC. The cells were maintained in flasks containing DMEM (Gibco, United States) supplemented with 10% fetal bovine serum (Gibco, United States), at 37°C in 5% CO_2. A mixture with a multiplicity of infection (number of bacteria per number of mammalian cells) of 100:1 was added to the HEp-2 monolayers in a 1-ml culture medium. Bacteria were allowed to adhere for 3 h at 37°C in 5% CO_2. Non-adherent cells were removed by washing three times with phosphate-buffered saline (Gibco, United States), and adherent cells were collected using cell scraper and were then transferred to 1.5-ml tubes with 1 ml of phosphate-buffered saline. Using the plate dilution method, the bacterial cells were measured. The interaction rate was expressed as (interaction cell number)/(inoculated cell number) × 100%. *E. coli* HB101 and enteropathogenic *E. coli* 2348/69 were used as negative and positive controls for adhesion assay, respectively.

For cell cytotoxicity assay, bacteria were incubated in cells with a multiplicity of infection of 100:1 for 8 h at 37°C in 5% CO_2. The mixture was collected and centrifuged at 4,000 rpm for 5 min. The supernatant was used to detect the release of LDH using the CytoTox96 kit (Promega, United States) according to the manufacturer's instructions. Lysed cells and cells with *E. coli* strain HB101 were used as positive and negative controls, respectively. The relative amount of cytotoxicity was expressed

as (experimental release − spontaneous release)/(maximum release − spontaneous release) × 100%, in which the spontaneous release was the amount of LDH activity in the supernatant of uninfected cells and the maximum release was that when cells were lysed with the lysis buffer provided by the manufacturer. All experiments were performed thrice.

Hemolysis Assay

Hemolytic activity was evaluated using Columbia Blood Agar Plate (Oxiod, United Kingdom) containing 5% of erythrocytes from sheep. Bacteria were grown in 1 ml TSB, at 37°C with shaking at 180 rpm/min overnight and centrifuged at 4,000 rpm for 5 min. The supernatant was discarded, and the pellet was resuspended with residual supernatant. Five microliter of the above mixture was dropped onto blood agar plates and incubated at 37°C for 24 h. The isolates were evaluated according to the cleared zone produced.

Statistical Analysis

Data were analyzed using the SPSS 20.0 statistical software packages. All values are presented as the means ± standard deviation (mean ± SD). The level of statistical significance was determined using a *t*-test. The statistical significance was set at $P < 0.05$.

RESULTS

Genome Analysis of Virulence Genes for Different Species of *Cronobacter* Isolates

In this study, a total of 319 virulence genes were analyzed after comparison with the virulence genes of VFDB (**Supplementary Table S1**). Using the ward.D method, we found that the distribution of genes in the same species of all the isolates in this study was similar and clustered. As shown in **Figure 1A**, all the virulence genes of the isolates were clustered in five groups: group I was *C. sakazakii*, group II was *C. malonaticus*, group III was *C. dublinensis*, group IV was *C. turicensis*, and group V was *C. muytjensii*. Therefore, *Cronobacter* has species specificity for these virulence genes. However, no difference in gene content seemed to contribute to the differences in ST or source. Among these virulence genes, 131 genes were collectively present in the five species of *Cronobacter*. Furthermore, there were 28 virulence genes in *C. sakazakii*, 11 in *C. malonaticus*, 15 in *C. dublinensis*, 9 in *C. turicensis*, and 10 in *C. muytjensii* possessed by different species (**Figure 1B**).

Analysis of Virulence Genes Related to Adhesion, Toxins, and Secretion System

We selected the genes encoding virulence factors related to adhesion, toxins, and secretion system and compared them with the virulence genes in VFDB. Most of the isolates in this study harbored the genes *acpXL*, *galE*, *gmhA*/*lpcA*, *lpxA*, *lpxC*, *kdsA*, *rfaD*, *rfaE*, and *orfM*, which encode both adhesion and toxins (**Figures 2A,B**). The genes *htpB*, *flaB*, and *hofB/hcpB* associated with adhesion existed in most of the isolates analyzed in this study (**Figure 2A**), and the gene *hlyA* encoding hemolysin was detected in all the isolates except the isolate SD 83 (**Figure 2B**). Notably, some particular genes belonged only to a few specific isolates. A gene cluster consisting of the genes *sfpC*, *sfpD*, and *sfpH* was only found in the isolates *C. sakazakii* SC26, *C. malonaticus* SD16, *C. malonaticus* SD26, and *C. muytjensii* SD83, and was similar to the genes encoding bundle-forming pilus protein in sorbitol-fermenting enterohaemorrhagic *E. coli* O157:H (−) (**Figure 2A**). Interestingly, the four above-mentioned isolates harbored the gene *hlyB* encoding an inner membrane protein with a cytoplasmic ATP-binding cassette domain, which can pump out HlyA protein (Benabdelhak et al., 2003). All the isolates harbored two genes (PLES_00841 and EC042_4545) encoding a hypothetical protein belonging to the T6SS (**Figure 2C**). Some isolates possessed genes such as EC55989_3320, EC55989_3321, EC55989_3335, and EC55989_3339, which also encoded hypothetical proteins belonging to the T6SS.

HEp-2 Cell Adhesion and Cytotoxicity of *Cronobacter* Isolates

All isolates in this study were tested for cell adhesion and cytotoxicity to cultured HEp-2 cells. The average interaction rate of all the isolates was 7.5%, and eight of those isolates were above an average value (**Figure 3**). The interaction rates of the isolates from different sources were compared, and the results showed that there were no significant differences among them ($P > 0.05$). The same results were obtained for different STs and species of *Cronobacter*. The four isolates with *sfp* gene cluster did not show more higher adherence capacity than that of other isolates. Therefore, whether *sfp* gene cluster in *Cronobacter* is associated with adhesion warrants further research. Among the isolates, difference in cytotoxicity in HEp-2 cells was determined by measuring the amount of LDH released by HEp-2 cells when cocultured with *Cronobacter*. We first cultured isolates *C. sakazakii* SC26 and *C. sakazakii* CQ04 in DMEM and detected the OD 600 value hourly. The isolates *C. sakazakii* SC26 and *C. sakazakii* CQ04 exhibited a slow growth in the media, and no difference was observed between these isolates including HB101 (**Figure 4A**). However, the release of LDH was statistically significantly different at 8 h (**Figure 4B**). We then tested the remaining isolates at 8 h and found only four isolates (*C. sakazakii* SC26, *C. malonaticus* SD16, *C. malonaticus* SD26, and *C. malonaticus* SD83) that could cause cell rounding and death among the 31 strains (**Figure 4D**). The average LDH released was 14.12%, ranging from 0.95 to 60.64%. Most isolates showed a rather small range of LDH release, from 0.95 to 22.76% with an average of 8.85%, whereas in the four above-mentioned isolates, LDH release ranged between 41.52 and 60.64%, with an average of 49.66%, which was much higher than that of other isolates ($P < 0.05$, **Figure 4C**). There was no obvious trend in the difference in cytotoxicity between STs and clusters.

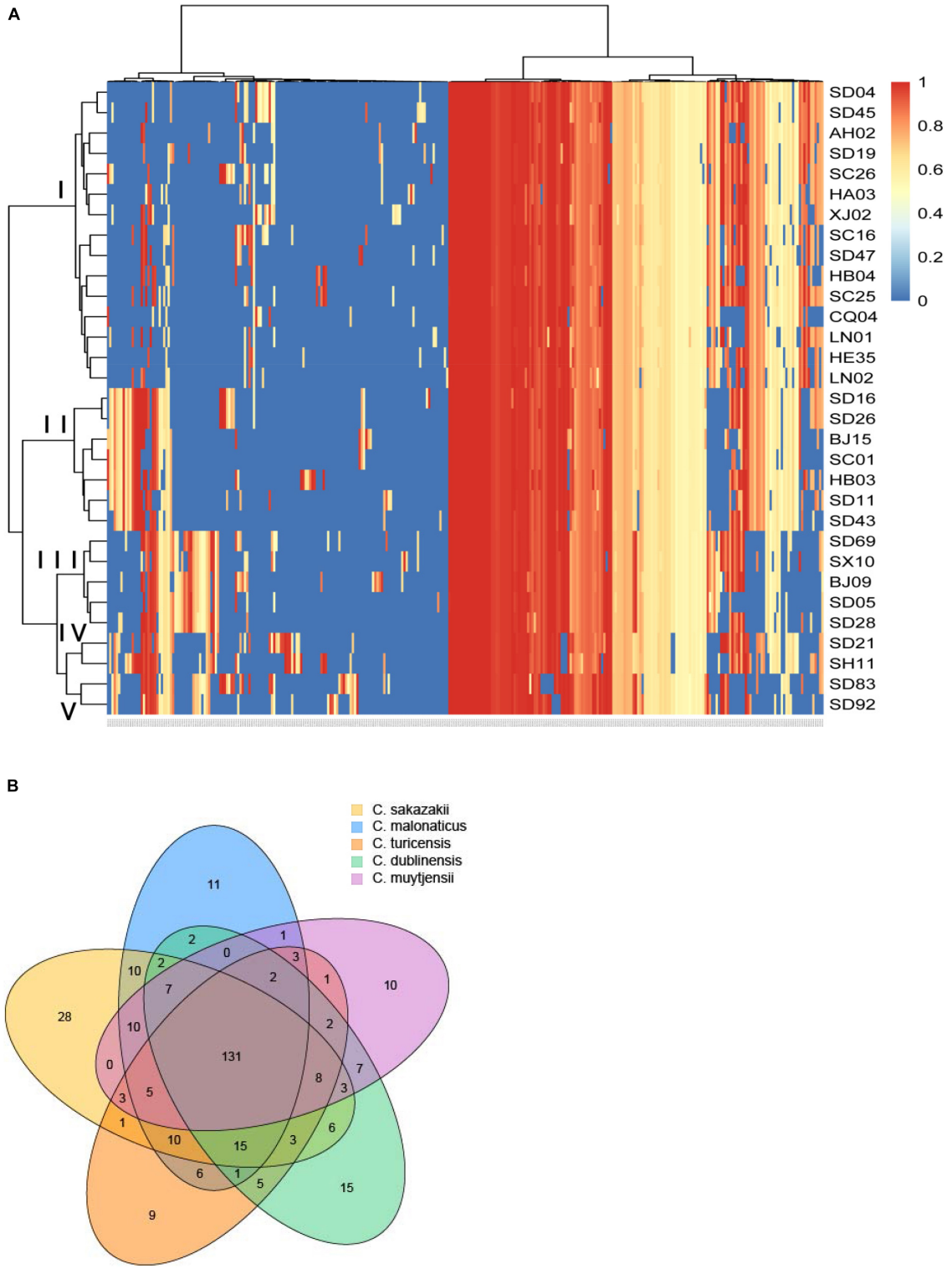


FIGURE 1 | Genome analysis of virulence genes for different species of *Cronobacter* isolates. **(A)** Hierarchically clustered heatmap of the virulence genes distribution of 31 *Cronobacter* isolates. The heatmap plot depicts the sequences coverage value of virulence genes (horizontal axis clustering) within each isolate (variables clustering on the vertical axis). The sequences coverage value of virulence genes are indicated by color intensity with the legend indicated at the right top of the figure. **(B)** Sequence diversity and functional enrichment across clustered genes. Virulence genes from each species were clustered using OrthoFinder v.0.7.1. This yielded 319 genes, shown here using draw quad. venn from the VennDiagram package of Rstudio v.3.0.2. Ovals with no cross indicate singleton genes possessed by every species of *Cronobacter*, and ovals with cross represent clusters of genes that have orthologs and paralogs for different species of *Cronobacter*.

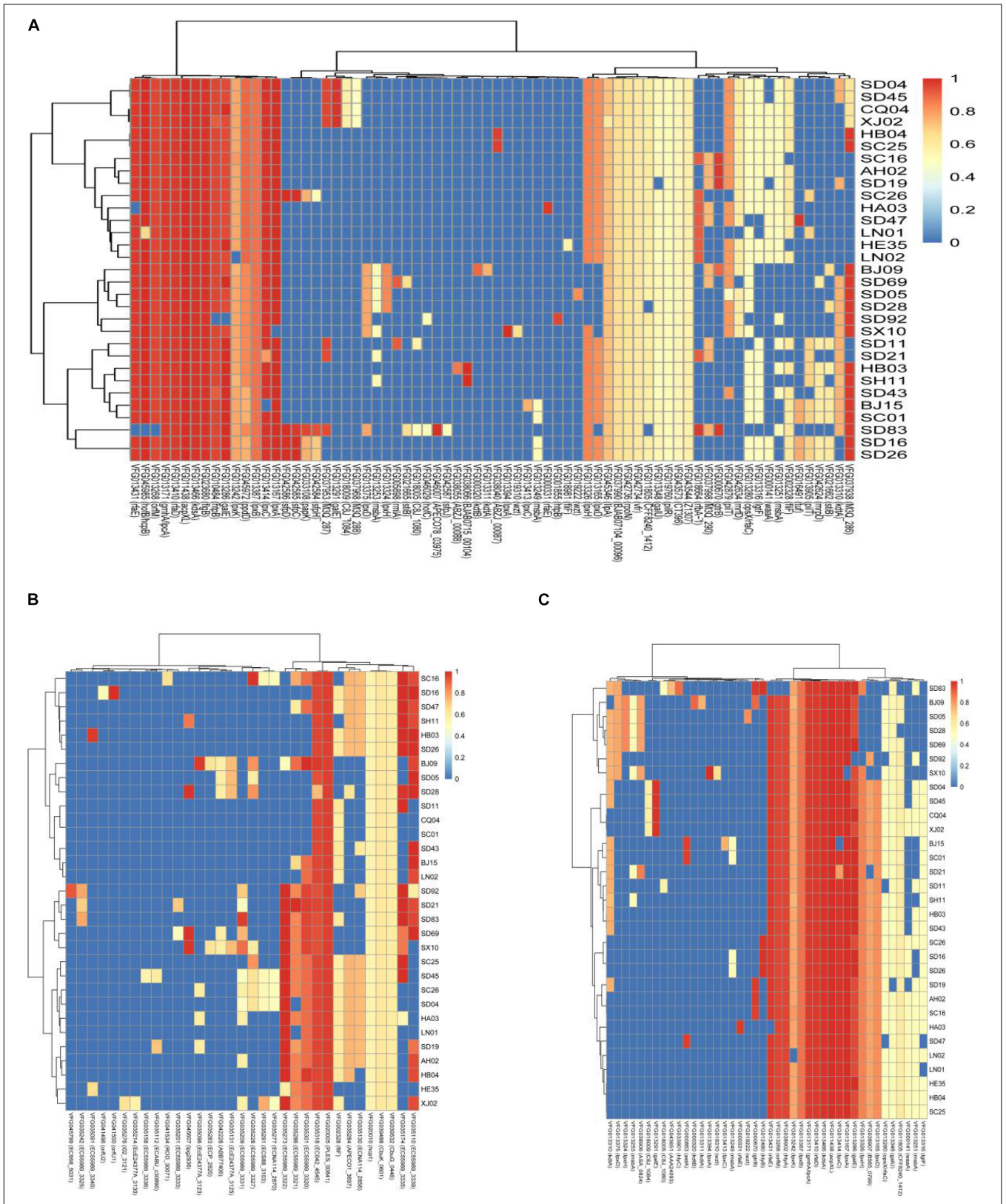


FIGURE 2 | Hierarchically clustered heatmap of the genes encoding adhesion **(A)**, toxin **(B)**, and secretion system **(C)** for 31 *Cronobacter* isolates in this study. The heatmap plot depicts the sequences coverage value of genes encoding adhesion, toxin, and secretion system (horizontal axis) with in each isolate (vertical axis).

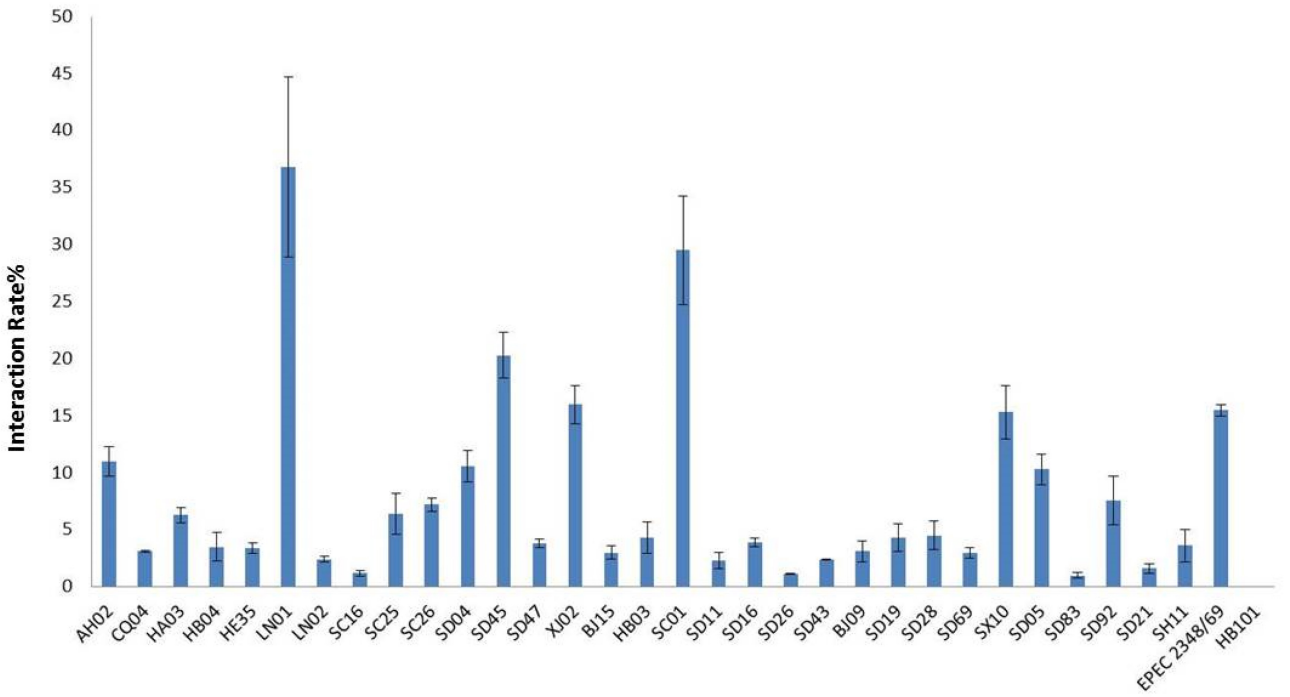


FIGURE 3 | *Cronobacter* attached to the Hep-2 epithelial cell surface was enumerated after 3 h by subtracting the number of bacteria recovered following. Results are presented as the percentage of the inoculum associated. Data are means ± standard errors of two independent experiments performed in triplicate. *E. coli* HB101 and enteropathogenic *E. coli* (EPEC) 2348/69 were used as negative and positive controls, respectively.

Hemolysis and Gene Cluster Analysis of *Cronobacter* Isolates

Hemolysis was tested in all the isolates in this study, and the results showed that most of isolates were negative, except for the isolates *C. sakazakii* SC26, *C. malonaticus* SD16, *C. malonaticus* SD26, and *C. muytjensii* SD83. As shown in **Figure 5A**, the four isolates formed a hemolysis ring around these isolates. The isolates *C. malonaticus* SD16, *C. malonaticus* SD26, and *C. sakazakii* SC26 showed high hemolysis capacity, followed by *C. muytjensii* SD83. Next, we analyzed the flanking region sequences of *hlyB* and found a gene cluster, comprising the genes *hlyC*, *hlyA*, *hlyB*, and *hlyD*, similar to *E. coli* UTI89 (**Figure 5B**). By comparing the amino acid sequences of *E. coli* UTI89, we found that the coverage of *hlyB* and *hlyD* was 99%, *hlyC* was 89%, and *hlyA* was 70%, and the identity of *hlyB*, *hlyC*, *hlyA*, and *hlyD*, was 69, 63, 54, and 47%, respectively. The length of whole gene cluster in *E. coli* UTI89 (7,380 bp) was longer than that in *Cronobacter* (6,771 bp); therefore, the function of the gene cluster in *Cronobacter* needs to be elucidated.

Known Virulence Genes of *Cronobacter* Isolates in Connection With Cytotoxicity

The chromosomal T6SS clusters of different species were downloaded from GeneBank and compared (data not shown). Since two isolates (*C. malonaticus* SD16 and *C. malonaticus* SD26) belonged to *C. malonaticus*, the T6SS cluster in chromosome of *C. malonaticus* NC_023032 was used as the reference sequence in this study (**Figure 6A**). Since there were reports of T6SS cluster in plasmid, the sequences of T6SS cluster in *C. sakazakii* ATCC BAA-894 pESA3 (**Figure 6B**) and *C. dublinensis* LMG 23823 pCDU1 (**Figure 6C**) were used as references. After comparing T6SS clusters of all the isolates with their reference sequences, we found the four above-mentioned strains to possess more complete T6SS clusters than other isolates. Therefore, the complete content of T6SS clusters might be helpful for the release of toxin or protein enzyme causing cytotoxicity.

Ten putative fimbriae gene clusters, reported by Joseph et al. (2012a), were as reference sequences. Sequences of all the isolates in this study were compared with those of the reference genes, and results showed the genes encoding fimbriae, for different species of *Cronobacter*, to be species specific. As in **Figure 6D**, 15 *C. sakazakii* isolates were found to be more similar to the reference than other isolates. Except for the isolate *C. sakazakii* LN01, gene cluster named ESA_01970–ESA_01976 of other 14 strains were absent in *C. sakazakii*. For the other species *C. malonaticus*, *C. dublinensis*, *C. turicensis*, and *C. muytjensii*, sequence similarity with the reference was <50%. The seven *C. malonaticus* isolates possessed only two complete gene clusters, namely, ESA_02342–ESA_02345 and ESA_03812–ESA_03815, whereas the isolates belonging to *C. dublinensis*, *C. turicensis*, and *C. muytjensii* had no complete gene cluster encoding fimbriae. These putative fimbrial gene clusters seemed to have no bearing on cytotoxicity, since the four above-mentioned isolates had no common characteristics in terms of these genes.

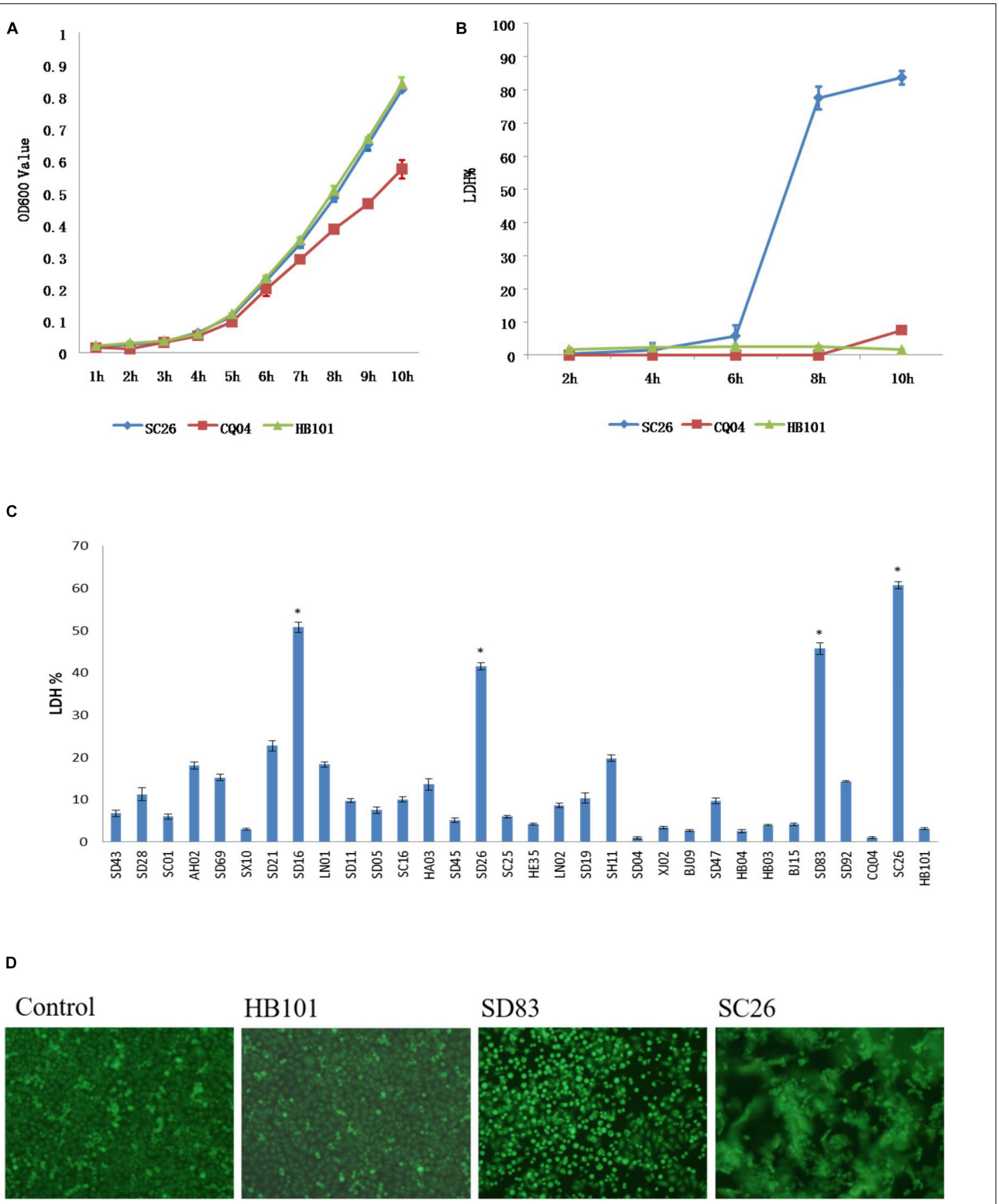


FIGURE 4 | Lactate dehydrogenase (LDH) released by HEp-2 cells after growth with *Cronobacter* and growth curve of these isolates in Dulbecco's modified Eagle's medium (DMEM) media. **(A)** LDH released by HEp-2 cells after growth with isolates *C. sakazakii* SC 26, *C. sakazakii* CQ04, and *E. coli* HB101 at 2, 4, 6, 8, and 10 h. **(B)** The growth curve of isolates *C. sakazakii* SC 26, *C. sakazakii* CQ04, and *E. coli* HB101 in DMEM at 1-h intervals within 10 h. **(C)** LDH released by HEp-2 cells after growth with 31 *Cronobacter* isolates. Negative control was isolate *E. coli* HB101. *LDH value of four isolates that could cause cell rounding and death were much higher than other isolates ($P < 0.05$). **(D)** Morphological changes of HEp-2 cells after growth with *Cronobacter* isolates. Cell rounding, detachment, and death can be seen after 8 h of exposure to *C. muytjensii* SD83 and *C. sakazakii* SC26. *E. coli* HB101 was used as negative control.

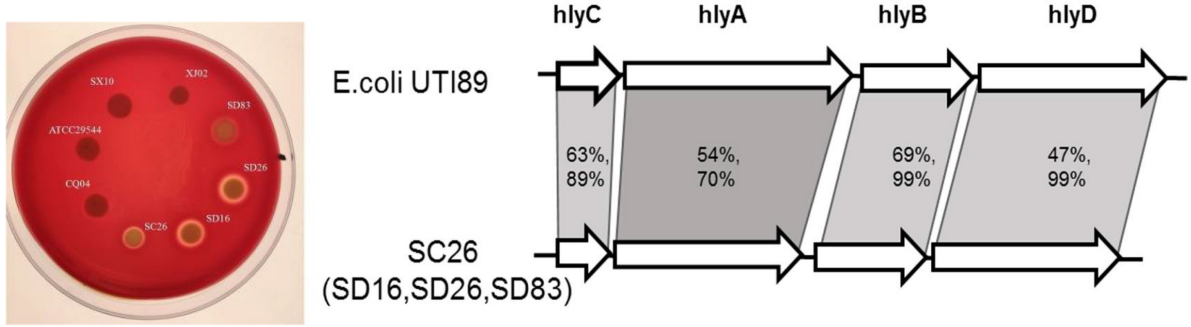


FIGURE 5 | Hemolysis phenotype and gene constrion for *Cronobacter*. **(A)** Different isolates showed different hemolysis capicity on blood agar plate. **(B)** The gene cluster comparision of the four isolates with strong hemolysiscapicity with *E. coli* UTI89.

DISCUSSION

Although some virulence genes have been reported previously, including two iron acquisition system loci (eitCBAD and iucABCD/iutA), a two-partner secretion system/filamentous hemagglutinin gene (fhaB), a transporter gene (fhaC), associated putative adhesins (FHA) locus, and a T6SS locus, hemolysin III (*hly*), and 10 putative fimbriae (pilus) gene clusters (Cruz et al., 2011; Franco et al., 2011; Joseph et al., 2012a), the genes were found only in the subsets of genomes, and all contributed considerably to the variation in gene content. By comparing the genes of all the isolates in this study with those in VFDB, the notable features of the different species of *Cronobacter* and a new virulence factor were discovered. Cluster analysis was performed with ward.D method on all the virulence genes aligned with that of VFDB, and results revealed the isolates from same species to be clustered together, thereby demonstrating that the isolates of same species possessed similar virulence genes. *Cronobacter* being a highly clonal organism, the genes are coinherited owing to the clonality (restricted genome variation) and not due to function, and the difference in these virulence genes might be caused by evolution. The isolates belonging to the clonal complex, such as *C. malonaticus* SD16, SD26, BJ15, and SC01, were clustered together; however, little difference existed between these virulence genes. The cas protein-coding gene (CRISPR-cas) array profiling has been applied to *Cronobacter* genomes of strains from the same ST. Moreover, it was demonstrated that strains in the same ST are distinguishable according to their spacer arrays (Ogrodzki and Forsythe, 2016). The presence and differentiated activity of CRISPR appear in different *Cronobacter* species (Zeng et al., 2018). The genome sequences of the isolates in this study were submitted to CRISPRCasFinder[3], and the results showed that most of the isolates, except for *C. dublinensis* SD69 and *C. muytjensii* SD92, possessed Crispr 1 (**Table 1**). It demonstrated that no significant variation existed between these isolates.

[3]https://crispr.i2bc.paris-saclay.fr

Many new gene clusters encoding virulence factors in these isolates were discovered. Here, the two gene clusters were analyzed especially because they existed in the same isolates and may have some relevance. As shown in **Figure 2A**, the four isolates (*C. sakazakii* SC26, *C. malonaticus* SD16, *C. malonaticus* SD26, and *C. muytjensii* SD83) harbored a specific gene cluster, which was similar to the *sfp* gene cluster encoding the bundle-forming pilus protein. The *sfp* gene cluster was first reported in a plasmid of sorbitol-fermenting enterohemorrhagic *E. coli* O157:H(−) and mediates mannose-resistant hemagglutination and fimbrial expression (Brunder et al., 2001). The *sfp* cluster included six genes, namely *sfpA*, *sfpH*, *sfpC*, *sfpD*, *sfpJ*, and *sfpG*. Expression of Sfp fimbriae in sorbitol-fermenting *E. coli* O157: NM strain is induced under conditions resembling those at the natural site of infection, and it may contribute to the adherence of the organisms to human intestinal epithelium (Müsken et al., 2008). Although the gene content of *sfp* in *Cronobacter* was not completed, we speculated the gene cluster to be similar to that in enterohemorrhagic *E. coli* O157: H(−), which had the same function. Another important gene cluster encoding hemolysin was found in these four isolates. The sequences of gene *hlyB* and flanking region were compared with the ones of *E. coli* UTI89, following which a new and completed gene cluster, including *hlyA*, *hlyB*, *hlyC*, and *hlyD*, was identified (**Figure 5B**). However, notably, the gene *hlyA* of the new cluster was different from the gene *hlyA* (VFG038902) that existed in most of the isolates in this study. It has been reported that *hlyA* and *hlyB* are coexpressed in a non-hemolytic *E. coli* strain, and this strain showed enhanced hemolytic activity on blood agar plates. In addition, the synergistic hemolytic activity of HlyA and HlyB has been detected by liquid hemolytic assay. Moreover, the gene (*hlyIII*) encoding type III hemolysin had been reported to be associated with the hemolytic phenotype of *Cronobacter* (Cruz et al., 2011; Singh et al., 2017), although it was different from the gene cluster discovered in this study. Accordingly, we speculated that the new gene cluster encoding hemolysin was involved in the mechanism underlying hemolysis in *Cronobacter*.

To assess the function of the two gene clusters, we performed cell adhesion, cytotoxicity, and hemolysis assays for the isolates

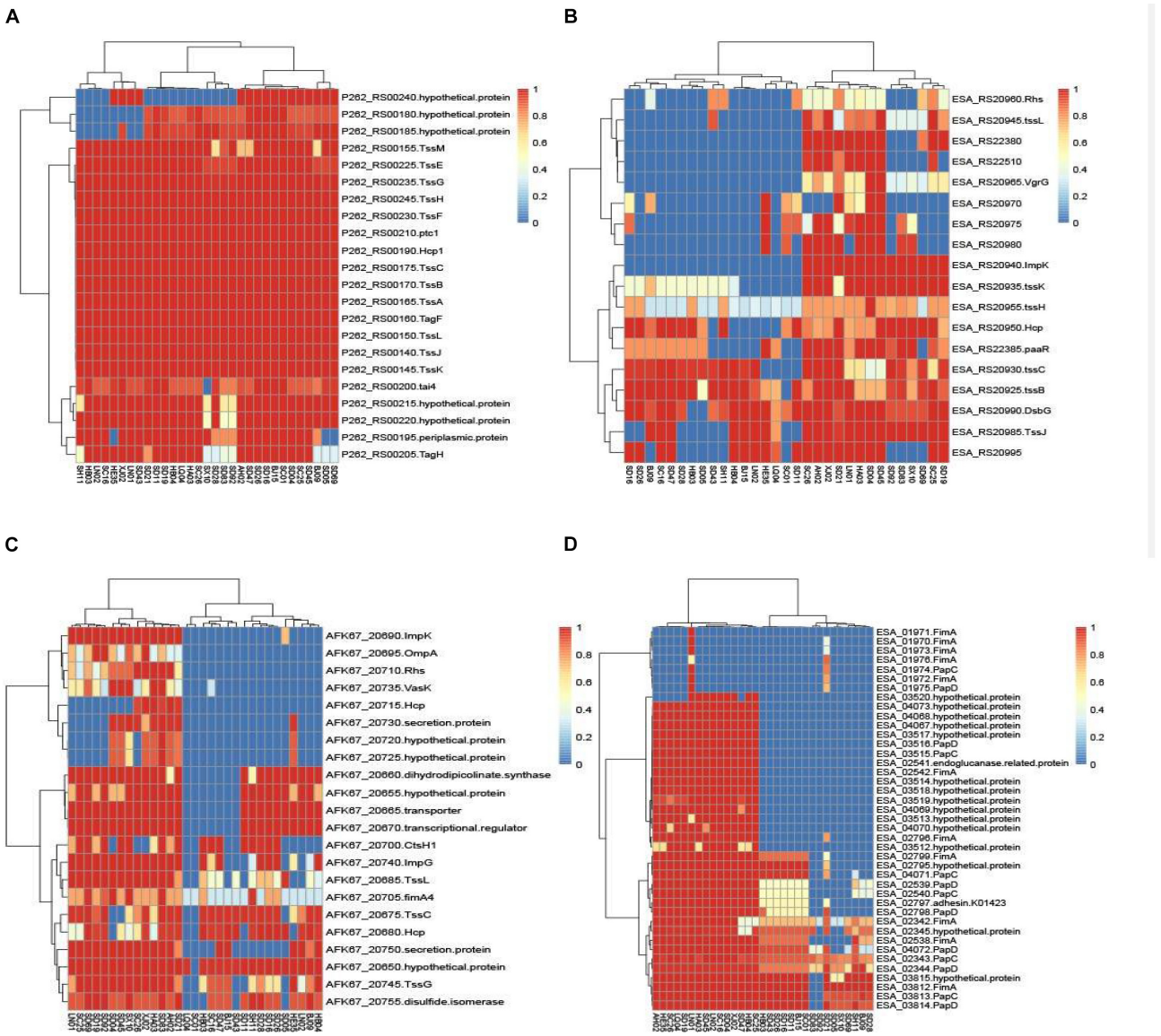


FIGURE 6 | The heatmap was generated based on 31 *Cronobacter* isolates for genes encoding T6SS and fimbriae. The heatmap plot depicts the sequences coverage value of genes encoding T6SS and fimbriae (vertical axis) with in each isolate (horizontal axis). **(A)** T6SS cluster in chromosome of *C. malonaticus* NC_023032 was used as the reference sequences; **(B)** sequences of T6SS cluster in *C. sakazakii* ATCC BAA-894 pESA3 was used as the reference; **(C)** sequences of T6SS cluster in *C. dublinensis* LMG 23823 pCDU1 was used as references; and **(D)** 10 putative fimbriae gene clusters were as reference sequences.

in this study. Most of the *Cronobacter* isolates induced the low release of LDH from cells, but only four isolates (*C. sakazakii* SC26, *C. malonaticus* SD16, *C. malonaticus* SD26, and *C. muytjensii* SD83) induced the cell death in a well after their incubation for 8 h, and LDH values were found to be much higher than in other isolates ($P < 0.05$). Simultaneously, the four isolates showed strong hemolysis capacity, while other isolates were negative. These results were consistent with the gene content characteristics discovered by genome sequencing analysis. Therefore, it was speculated that the two gene clusters, including *sfp* and *hly*, play an important role in the pathogenesis of *Cronobacter*.

Secretion systems are the major weapons required for colonization, survival, cytotoxicity, and evasion of the host innate immune system (Bleves et al., 2010). T6SS is a novel and complex multicomponent secretion system, which is often involved in the interaction with eukaryotic hosts, irrespective of it being a pathogenic or a symbiotic relationship (Filloux et al., 2008). *In silico* analysis of pESA3, harbored by *C. sakazakii* ATCC BAA-894, revealed that this T6SS gene cluster consists of 16 open reading frames (Franco et al., 2011). Besides the plasmid, in the chromosome of *C. sakazakii* ATCC12868, the two phylogenetically distinct T6SS gene clusters (T6SS-1 and T6SS-2) were also investigated (Wang et al., 2018).

We compared the T6SS cluster of strains with the chromosome of *C. malonaticus* (NC_023032), plasmid pESA3 of *C. sakazakii* ATCC BAA-894, and pCDU1 of *C. dublinensis* subsp. *dublinensis* LMG 23823, and found T6SS cluster to have species specificity and that isolates with high cytotoxicity possessed more complete T6SS cluster construction than the rest. However, there was no obvious difference between the isolates with high cytotoxicity and other isolates. Therefore, although T6SS cluster might be involved in the mechanism underlying the cytotoxicity of *Cronobacter*, it may not be a key element.

Fimbriae (pili) were considered to be another important element for cytotoxicity of bacteria. Few reports have described the fimbrial characteristics of *Cronobacter*, except for the 10 putative fimbrial gene clusters identified by Joseph et al. (2012a). Although an individual gene cluster may be absent (for example, curli fimbriae genes were absent in *C. sakazakii*), most gene clusters were present in the 12 strains analyzed in the study. In the current study, *C. sakazakii* isolates had gene sequences more similar to the reference sequence, and sequences of other species were <50% similar to that of reference gene cluster. Nevertheless, results showed the gene distribution to have species specificity characteristic.

CONCLUSION

By comparing with the virulence genes of VFDB, we found that *Cronobacter* has species specificity for these virulence genes. Besides, the two gene clusters, including *sfp* encoding fimbriae and *hly* encoding hemolysin, were analyzed only because they existed in the same isolates (*C. sakazakii* SC26, *C. malonaticus* SD16, *C. malonaticus* SD26, and *C. muytjensii* SD83). Moreover, these four isolates showed higher cytotoxicity and stronger hemolysis capacity than those of other isolates in this study. Therefore, it was considered that the *hly* gene cluster discovered in this study was associated with hemolysis and cytotoxicity of *Cronobacter*. Although the *sfp* gene cluster seemed to not be associated with adhesion of *Cronobacter*, it might mediate cytotoxic effects on cells. The mechanism of cytotoxicity induced by hemolysin and fimbriae in *Cronobacter* should be studied in the future.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Whole Genome Shotgun project at GenBank under the BioProject PRJNA287482 and PRJNA498360, with accession numbers LGRM00000000, LGRL00000000, RPAZ00000000, RPBA00000000-RPBZ00000000, RPCA00000000, and RPCB00000000.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of National Institute for Communicable Disease Control and Prevention. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

JC, HH, and JY conceived and designed the experiments. JH, XD, GX, and SL performed the experiments. ZC and CY analyzed the data. JC, JH, XD, and ZC prepared the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (NSFC) Program (81501799). CAMS Innovation Fund for Medical Sciences (CIFMS) Program (2016-12M-1-008).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.03104/full#supplementary-material

TABLE S1 | The identification and coverage of virulence genes of *Cronobacter* in this study compared with VFDB.

REFERENCES

Basso, P., Ragno, M., Elsen, S., Reboud, E., Golovkine, G., Bouillot, S., et al. (2017). *Pseudomonas aeruginosa* pore-forming exolysin and type iv pili cooperate to induce host cell lysis. *MBio.* 8:e02250-16;. doi: 10.1128/mBio.02250-16

Benabdelhak, H., Kiontke, S., Horn, C., Ernst, R., Blight, M. A., Holland, I. B., et al. (2003). A specific interaction between the NBD of the ABC-transporter HlyB and a C-terminal fragment of its transport substrate haemolysin A. *J. Mol. Biol.* 327, 1169–1179. doi: 10.1016/s0022-2836(03)00204-3

Bleves, S., Viarre, V., Salacha, R., Michel, G. P., Filloux, A., and Voulhoux, R. (2010). Protein secretion systems in *Pseudomonas aeruginosa*: a wealth of pathogenic weapons. *Int. J. Med. Microbiol.* 300, 534–543. doi: 10.1016/j.ijmm.2010.08.005

Brunder, W., Khan, A. S., Hacker, J., and Karch, H. (2001). Novel type of fimbriae encoded by the large plasmid of sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:H (−). *Infect. Immun.* 69, 4447–4457. doi: 10.1128/iai.69.7.4447-4457.2001

Chandrapala, D., Kim, K., Choi, Y., Senevirathne, A., Kang, D. H., Ryu, S., et al. (2014). Putative inv is essential for basolateral invasion of Caco-2 Cells and acts synergistically with OmpA to Affect *in vitro* and *in vivo* virulence of *Cronobacter sakazakii* ATCC 29544. *Infect. Immun.* 82, 1755–1765. doi: 10.1128/IAI.01397-13

Chaves, C. E. V., Brandão, M. L. L., Lacerda, M. L. G. G., Rocha, C. A. B. C., Leone de Oliveira, S. M. D. V., Parpinelli, T. C., et al. (2018). Fatal *Cronobacter sakazakii* sequence type 494 meningitis in a Newborn, Brazil. *Emerg. Infect. Dis.* 24, 1948–1950. doi: 10.3201/eid2410.180373

Cheung, K. J. jr, Li, G., Urban, T. A., Goldberg, J. B., Griffith, A., Lu, F., et al. (2007). Pilus-mediated epithelial cell death in response to infection with *Burkholderia cenocepacia*. *Microbes. Infect.* 9, 829–837. doi: 10.1016/j.micinf.2007.03.001

Cruz, A., Xicohtencatl-Cortes, J., González-Pedrajo, B., Bobadilla, M., Eslava, C., and Rosas, I. (2011). Virulence traits in *Cronobacter* species isolated from different sources. *Can. J. Microbiol.* 57, 735–744. doi: 10.1139/w11-063

Cui, J. H., Yu, B., Xiang, Y., Zhang, Z., Zhang, T., Zeng, Y. C., et al. (2017). Two Cases of Multi-antibiotic Resistant *Cronobacter* spp. Infections of Infants in China. *Biomed. Environ. Sci*. 8, 601–605. doi: 10.3967/bes2017.079

Elgaml, A., and Miyoshi, S. I. (2017). Regulation systems of protease and hemolysin production in *Vibrio vulnificus*. *Microbio. Immunol.* 61, 1–11. doi: 10.1111/1348-0421.12465

Filloux, A., Hachani, A., and Bleves, S. (2008). The bacterial type VI secretion machine: yet another player for protein transport across membranes. *Microbiology.* 154, 1570–1583. doi: 10.1099/mic.0.2008/016840-0

Franco, A. A., Hu, L., Grim, C. J., Gopinath, G., Sathyamoorthy, V., Jarvis, K. G., et al. (2011). Characterization of putative virulence genes on the related RepFIB plasmids harbored by *Cronobacter* spp. *Appl. Environ. Microbiol.* 77, 3255–3267. doi: 10.1128/AEM.03023-10

Hu, L. (2018). Prevalence of curli genes among *Cronobacter* species and their roles in biofilm formation and cell-cell aggregation. *Int. J. Food. Microbiol.* 265, 65–73. doi: 10.1016/j.ijfoodmicro.2017.10.031

Joseph, S., Desai, P., Ji, Y., Cummings, C. A., Shih, R., Degoricija, L., et al. (2012a). Comparative analysis of genome sequences covering the seven cronobacter species. *PLoS. One.* 7:e49455. doi: 10.1371/journal.pone.0049455

Joseph, S., Sonbol, H., Hariri, S., Desai, P., McClelland, M., and Forsythe, S. J. (2012b). Diversity of the *Cronobacter* genus as revealed by multilocus sequence typing. *J. Clin. Microbiol.* 50, 3031–3039. doi: 10.1128/JCM.00905-12

Kim, K., Kim, K. P., Choi, J., Lim, J. A., Lee, J., Hwang, S., et al. (2010). Outer membrane proteins A (OmpA) and X (OmpX) are essential for basolateral invasion of *Cronobacter sakazakii*. *Appl. Environ. Microbiol.* 76, 5188–5198. doi: 10.1128/AEM.02498-09

Leclercq, S. Y., Sullivan, M. J., Ipe, D. S., Smith, J. P., Cripps, A. W., and Ulett, G. C. (2016). Pathogenesis of *Streptococcus* urinary tract infection depends on bacterial strain and β-hemolysin/cytolysin that mediates cytotoxicity, cytokine synthesis, inflammation and virulence. *Sci. Rep.* 7:29000. doi: 10.1038/srep29000

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., et al. (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome. Res.* 20, 265–272. doi: 10.1101/gr.097261.109

Liu, L., Hao, S., Lan, R., Wang, G., Xiao, D., Sun, H., et al. (2015). The Type VI secretion system modulates flagellar gene expression and secretion in *Citrobacter freundii* and contributes to adhesion and cytotoxicity to host cells. *Infect. Immun.* 83, 2596–2604. doi: 10.1128/IAI.03071-14

Mange, J. P. R., Borel, N., Wild, P., Kim, K. S., Pospischil, A., and Lehner, A. (2006). Adhesive properties of *Enterobacter sakazakii* to human epithelial and brain microvascular endothelial cells. *B.M.C. Microbiol.* 6:58.

McMullan, R., Menon, V., Beukers, A. G., Jensen, S. O., van Hal, S. J., and Davis, R. (2018). *Cronobacter sakazakii* infection from expressed breast milk. *Australia. Emerg. Infect. Dis.* 24, 393–394. doi: 10.3201/eid2402.171411

Mittal, R., Wang, Y., Hunter, C. J., Gonzalez-Gomez, I., and Prasadarao, N. V. (2009). Brain damage in newborn rat model of meningitis by *Enterobacter sakazakii*: a role for outer membrane protein A. *Lab. Invest.* 89, 263–277. doi: 10.1038/labinvest.2008.164

Morato-Rodríguez, M. D. R., Velandia-Rodríguez, D., Castañeda, S., Crosby, M., and Vera, H. (2018). Cronobacter spp. in Common Breast Milk Substitutes, Bogotá, Colombia. Emerg. . *Infect. Dis* 24, doi: 10.3201/eid2410.172021

Müsken, A., Bielaszewska, M., Greune, L., Schweppe, C. H., Müthing, J., Schmidt, H., et al. (2008). Anaerobic conditions promote expression of Sfp fimbriae and adherence of sorbitol-fermenting enterohemorrhagic *Escherichia coli* O157:NM to human intestinal epithelial cells. *Appl. Environ . Microbiol.* 74, 1087–1093. doi: 10.1128/aem.02496-07

Ogrodzki, P., and Forsythe, S. J. (2016). CRISPR-cas loci profiling of Cronobacter sakazakii pathovars. *Future. Microbiol.* 11, 1507–1519. doi: 10.2217/fmb-2016-0070

Pagotto, F. J., Nazarowec-White, M., Bidawid, S., and Farber, J. M. (2003). Enterobacter sakazakii: infectivity and enterotoxin production *in vitro* and *in vivo*. *J. Food. Prot.* 66, 370–375. doi: 10.4315/0362-028x-66.3.370

Scavone, P., Villar, S., Umpiérrez, A., and Zunino, P. (2015). Role of *Proteus mirabilis* MR/P fimbriae and flagella in adhesion, cytotoxicity and genotoxicity induction in T24 and Vero cells. *Pathog. Dis.* 73:ftv017. doi: 10.1093/femspd/ftv017

Singamsetty, V. K., Wang, Y., Shimada, H., and Prasadarao, N. V. (2008). Outer membrane protein A expression in *Enterobacter sakazakii* is required to induce microtubule condensation in human brain microvascular endothelial cells for invasion. *Microb. Pathog.* 45, 181–191. doi: 10.1016/j.micpath.2008.05.006

Singh, N., Raghav, M., Narula, S., Tandon, S., and Goel, G. (2017). Profiling of virulence determinants in *cronobacter sakazakii* isolates from different plant and environmental commodities. *Curr. Microbiol.* 74, 560–565. doi: 10.1007/s00284-017-1219-9

Townsend, S., Hurrell, E., and Forsythe, S. J. (2008). Virulence studies of *Enterobacter sakazakii* isolates associated with a neonatal intensive care unit outbreak. *B.M.C. Microbiol.* 18:64. doi: 10.1186/1471-2180-8-64

Tsuiji, M., Shiohara, K., Takei, Y., Shinohara, Y., Nemoto, S., Yamaguchi, S., et al. (2019). Selective Cytotoxicity of Staphylococcal α-Hemolysin (α-Toxin) against Human Leukocyte Populations. *Biol. Pharm. Bull.* 42, 982–988. doi: 10.1248/bpb.b18-01024

Wang, M., Cao, H., Wang, Q., Xu, T., Guo, X., and Liu, B. (2018). The roles of two type VI secretion systems in *Cronobacter sakazakii* ATCC 12868. *Front. Microbiol.* 22:2499. doi: 10.3389/fmicb.2018.02499

Zeng, H., Lei, T., He, W., Zhang, J., Liang, B., Li, C., et al. (2015). Novel Multidrug-Resistant *Cronobacter sakazakii* Causing Meningitis in Neonate. *China. Emerg. Infect. Dis.* 24, 2121–2124. doi: 10.3201/eid2411.180718

Zeng, H., Zhang, J., Wu, Q., He, W., Wu, H., and Ye, Y. (2018). Reconstituting the history of *Cronobacter* Evolution Driven by differentiated CRISPR activity. *Appl. Environ. Microbiol.* 84:e00267-18.. doi: 10.1128/AEM.00267-18

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 February 2020
doi: 10.3389/fmicb.2020.00117

Nicotinamide Increases Intracellular NAD^+ Content to Enhance Autophagy-Mediated Group A Streptococcal Clearance in Endothelial Cells

*Cheng-Lu Hsieh[1], Shu-Ying Hsieh[2], Hsuan-Min Huang[3], Shiou-Ling Lu[4], Hiroko Omori[5], Po-Xing Zheng[6], Yen-Ning Ho[7], Yi-Lin Cheng[7], Yee-Shin Lin[1,6,8], Chuan Chiang-Ni[9,10], Pei-Jane Tsai[1,3], Shu-Ying Wang[1,8], Ching-Chuan Liu[1,6,11], Takeshi Noda[4] and Jiunn-Jong Wu[1,7]**

[1] Institute of Basic Medical Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan, [2] Institute of Molecular Medicine, College of Medicine, National Cheng Kung University, Tainan, Taiwan, [3] Department of Medical Laboratory Science and Biotechnology, College of Medicine, National Cheng Kung University, Tainan, Taiwan, [4] Center for Frontier Oral Science, Graduate School of Dentistry, Osaka University, Osaka, Japan, [5] Research Institute for Microbial Diseases, Osaka University, Osaka, Japan, [6] Center of Infectious Disease and Signaling Research, College of Medicine, National Cheng Kung University, Tainan, Taiwan, [7] Department of Biotechnology and Laboratory Science in Medicine, School of Biomedical Science and Engineering, National Yang-Ming University, Taipei, Taiwan, [8] Department of Microbiology and Immunology, College of Medicine, National Cheng Kung University, Tainan, Taiwan, [9] Department of Microbiology & Immunology, College of Medicine, Chang Gung University, Taoyuan, Taiwan, [10] Molecular Infectious Disease Research Center, Chang Gung Memorial Hospital, Taoyuan, Taiwan, [11] Department of Pediatrics, College of Medicine, National Cheng Kung University and Hospital, Tainan, Taiwan

OPEN ACCESS

Edited by:
Diego Robledo, The University of Edinburgh, United Kingdom

Reviewed by:
Blanca Estela Garcia-Perez, National Polytechnic Institute (Mexico), Mexico
Maghnus N O'Seaghdha, Suffolk University, United States

***Correspondence:**
Jiunn-Jong Wu
jjwu1019@ym.edu.tw

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *10 October 2019*
Accepted: *20 January 2020*
Published: *11 February 2020*

Citation:
Hsieh C-L, Hsieh S-Y, Huang H-M, Lu S-L, Omori H, Zheng P-X, Ho Y-N, Cheng Y-L, Lin Y-S, Chiang-Ni C, Tsai P-J, Wang S-Y, Liu C-C, Noda T and Wu J-J (2020) Nicotinamide Increases Intracellular NAD^+ Content to Enhance Autophagy-Mediated Group A Streptococcal Clearance in Endothelial Cells. Front. Microbiol. 11:117. doi: 10.3389/fmicb.2020.00117

Group A streptococcus (GAS) is a versatile pathogen that causes a wide spectrum of diseases in humans. Invading host cells is a known strategy for GAS to avoid antibiotic killing and immune recognition. However, the underlying mechanisms of GAS resistance to intracellular killing need to be explored. Endothelial HMEC-1 cells were infected with GAS, methicillin-resistant *Staphylococcus aureus* (MRSA) and *Salmonella* Typhimurium under nicotinamide (NAM)-supplemented conditions. The intracellular NAD^+ level and cell viability were respectively measured by NAD^+ quantification kit and protease-based cytotoxicity assay. Moreover, the intracellular bacteria were analyzed by colony-forming assay, transmission electron microscopy, and confocal microscopy. We found that supplementation with exogenous nicotinamide during infection significantly inhibited the growth of intracellular GAS in endothelial cells. Moreover, the NAD^+ content and NAD^+/NADH ratio of GAS-infected endothelial cells were dramatically increased, whereas the cell cytotoxicity was decreased by exogenous nicotinamide treatment. After knockdown of the autophagy-related ATG9A, the intracellular bacterial load was increased in nicotinamide-treated endothelial cells. The results of Western blot and transmission electron microscopy also revealed that cells treated with nicotinamide can increase autophagy-associated LC3 conversion and double-membrane formation during GAS infection. Confocal microscopy images further showed that more GAS-containing vacuoles were colocalized with lysosome under nicotinamide-supplemented

conditions than without nicotinamide treatment. In contrast to GAS, supplementation with exogenous nicotinamide did not effectively inhibit the growth of MRSA or *S.* Typhimurium in endothelial cells. These results indicate that intracellular NAD^+ homeostasis is crucial for controlling intracellular GAS infection in endothelial cells. In addition, nicotinamide may be a potential new therapeutic agent to overcome persistent infections of GAS.

Keywords: Group A streptococcus (GAS), nicotinamide (NAM), NAD^+ homeostasis, intracellular survival, endothelial cells (ECs)

INTRODUCTION

Group A streptococcus (GAS) is recognized as one of the major human pathogens that not only provokes mild skin and throat infections such as impetigo or pharyngitis, but also occasionally causes serious infections such as streptococcal toxic shock syndrome and necrotizing fasciitis (Cunningham, 2008; Walker et al., 2014). Although antibiotics are effective for treating GAS infections, accumulating evidence has shown that invading host cells could be one of the strategies used by GAS to avoid killing by host immune responses and antibiotics, which may contribute to persistent infection seen in 5–30% of individuals (Osterlund et al., 1997; Marouni et al., 2004; Kaplan et al., 2006).

Nicotinamide adenine dinucleotide (NAD^+) and its reduced form (NADH) have been known as important cofactors that are coupled to the redox reactions for energy production, or used as enzymatic cofactors involving in hundreds of biochemical functions within living cells (Canto et al., 2015). Nicotinamide (NAM) is a vitamin B3 derivative that has been used as a metabolic precursor of NAD^+ to raise intracellular NAD^+ level for improvement of metabolic diseases involving energy production (Fricker et al., 2018). Autophagy, the cellular degradation of ubiquitinated proteins in lysosomes under stress conditions, is one of the most tightly regulated homeostatic processes involving intracellular NAD^+ (Klionsky et al., 2016; Kocaturk and Gozuacik, 2018). Moreover, autophagy also plays a crucial role in innate immune defenses against invading pathogens including *Listeria monocytogenes*, *Salmonella* Typhimurium, and GAS (Castrejón-Jiménez et al., 2015; Kimmey and Stallings, 2016; Bah and Vergne, 2017). In order to survive in host cells, GAS expresses various virulence factors to impair autophagic clearance, including streptococcal cysteine protease SpeB, streptolysin O (SLO), and NAD-glycohydrolase (NADase) (Sakurai et al., 2010; Barnett et al., 2013; Mestre and Colombo, 2013; O'Seaghdha and Wessels, 2013; O'Neill et al., 2016; Sharma et al., 2016). NADase is a potent hydrolase involved in the consumption of NAD^+ that leads to intracellular energy collapse and programmed necrosis of infected cells (Chandrasekaran and Caparon, 2015, 2016; Pajuelo et al., 2018). In addition, several studies have indicated that NADase is involved with the structural and functional stabilization of SLO, which contributes to enhance GAS pathogenesis and global dissemination of serotype M1 and M89 GAS, indicating that NADase plays an important role during GAS infection (Michos et al., 2006; Turner et al., 2015; Zhu et al., 2015; Velarde et al., 2017; Barnett et al., 2018). However, the mechanisms of NAD^+ homeostasis controlling GAS survival in the host are complicated and need to be explored.

Previously, we have found that defective acidification of autophagosomes allows GAS growth in endothelial cells (Lu et al., 2015). NADase is responsible for the depletion of intracellular NAD^+ and inhibition of autophagosomal acidification, which results in the multiplication of GAS in endothelial cells (Hsieh et al., 2018). In this study, we demonstrate that supplementation with exogenous NAM significantly restores the intracellular NAD^+ content and NAD^+/NADH ratio, which enhances the acidification of GAS-containing autophagosomes and clearance of intracellular GAS within endothelial cells.

MATERIALS AND METHODS

Cell Culture

Human microvascular endothelial cell line-1 (HMEC-1) cells were cultured in endothelial growth medium M200 with low serum growth factors (Gibco Life Technologies, Grand Island, NY, United States) and 10% fetal bovine serum (FBS) at 37°C in a humidified incubator with 5% CO_2. When the cell confluence reached 80%, cells were detached with trypsin-EDTA (Gibco Life Technologies) and seeded at the density of 0.75×10^6 cells/dish in 10-cm dishes for maintenance or 3×10^5 cells/well in 6-well plates for the intracellular bacteria survival assay and confocal microscopy.

Bacteria and Cultural Conditions

Group A streptococcus strains SF370 (M1 serotype) and NZ131 (M49 serotype) were purchased from the American Type Culture Collection (Manassas, VA, United States). GAS strain A20 (M1 serotype) was isolated from the blood of a patient with necrotizing fasciitis (Zheng et al., 2013). Methicillin-resistant *Staphylococcus aureus* (MRSA) and *Salmonella* Typhimurium were isolated from patients with bacteremia. All strains were susceptible to gentamicin and cultured on tryptic soy agar containing 5% defibrinated sheep blood or tryptic soy broth (Becton Dickinson, Sparks, MD, United States) supplemented with 0.5% yeast extract (TSBY).

Intracellular Bacterial Survival Assay

The cell infection was described in the previous study with modifications (Hsieh et al., 2018). In brief, the overnight

bacterial cultures were transferred and grown to the exponential growth phase, and then resuspended in endothelial growth medium M200. HMEC-1 cells were seeded in 6-well plates at a density of 3×10^5 cells/well and infected with bacteria at different multiplicities of infection (M.O.I.) to ensure equivalent intracellular bacterial load at 1 h of infection. Plates were then centrifuged at 500 *g* for 5 min to synchronize the infection and incubated in humidified 5% CO_2 at 37°C for 30 min. After infection, the infected cells were treated with gentamicin (125 μg/mL) to kill extracellular bacteria. Subsequently, cells were maintained in M200 medium with or without β-NAD, NADH, or NAM for additional 2 and 4 h. At indicated times post infection, cells were lysed by deionized water and plated on TSBY agar plates to evaluate the intracellular GAS growth within endothelial cells.

Determination of NADase Activity

The NADase activity in GAS culture supernatants was measured by the qualitative fluorescence assay according to the method described above by Bricker et al. (2002). Briefly, the bacterial supernatants from different culture treatments were harvested by centrifugation at 2,330 *g* for 10 min. The supernatants were reacted with 1 mM of β-NAD (Sigma-Aldrich, St. Louis, MO, United States) at 37°C in 5% CO_2 for 1 h. After incubation, sodium hydroxide (PanReac AppliChem, Barcelona, Spain) was added to develop reactions at room temperature in the dark for 1 h. The fluorescence intensity of β-NAD was determined by Tecan M200 Pro Infinite plate reader (Tecan, Crailsheim, Germany) at excitation wavelength 360 nm. The NADase activity of each sample was expressed as a relative percentage compared with uninoculated culture medium.

Quantification of Intracellular NAD^+ Level

The change of intracellular NAD^+ content was determined by the NAD^+/NADH quantification kit (Sigma-Aldrich), according to the manufacturer's instructions. Briefly, a total of 2×10^5 cells were harvested by trypsinization and suspended in NAD^+/NADH extraction buffer to lyse the cells by freezing and thawing three times. The 10-kDa cut-off spin column (GE Healthcare, Buckinghamshire, United Kingdom) was used to remove NADH degrading enzymes by centrifugation at 21,000 *g*, 4°C for 60 min. The intracellular NAD^+ of each sample was decomposed in 60°C for 30 min and then transferred to microtiter plates to determine the concentration of intracellular NADH. In addition, total intracellular NADH of each sample was reacted with NAD cycling enzyme to convert NAD^+ to NADH. Subsequently, all samples were loaded into microtiter plates and mixed with NADH developer and incubated at room temperature for 1 h. After reaction, the absorbance of each sample was measured by ELISA reader (Tecan, infinite M200) at the wavelength of 450 nm. The NAD^+/NADH ratio was calculated to evaluate the metabolic status of cells following a formula: $[(NADH_{total} - NADH)/NADH]$.

Cell Cytotoxicity Assay

To determine cell cytotoxicity of GAS-infected cells under NAM treatment, the activity of dead-cell protease was measured by CytoTox-Glo cytotoxicity assay kit according to the manufacturer's instructions (Promega, Madison, WI, United States). Briefly, cell were infected with GAS as described above. The cell culture supernatants were harvested at indicated times post infection, and co-incubated with luminogenic substrate at room temperature for 15 min. The luminescent intensity was measured by Tecan M200 Pro Infinite plate reader (Tecan). The cell cytotoxicity of each sample was expressed as a relative percentage compared with non-infected cells.

RNA Interference

The expression of ATG9A in endothelial HMEC-1 cells was downregulated through lentiviral expression of a short hairpin RNA (clone 81, TRCN0000244081 containing target sequence 5′-AGTCACCTTGGCACCATATTG-3′; clone 82, TRCN0000244082 containing target sequence 5′-GTGGACTATGACATCCTATTT -3′; clone 83, TRCN0000244083 containing target sequence 5′-TGTAGGAGCAGGATGGAAATA-3′). The shRNA lentiviral clones were obtained from the National RNAi Core Facility at the Academia Sinica in Taiwan. Endothelial HMEC-1 cells were seeded in 6-well plates at a density of 2×10^5 cells/well and infected with lentivirus at a M.O.I of 3 in M200 medium containing 8 μg/mL of polybrene at 37°C in 5% CO_2 for 24 h. After transduction, cells were treated with 5 μg/mL of puromycin (Sigma-Aldrich) for three generations to select the stable clone. The efficiency of ATG9A knockdown was evaluated by Western blot analysis.

Transmission Electron Microscopy (TEM)

Endothelial HMEC-1 cells were infected with GAS at M.O.I. of 1 for 30 min, and treated with gentamicin to kill extracellular bacteria. After treatment, cells were maintained in NAM-supplemented medium for additional 2 h. Cells were detached by trypsinization and gently resuspended in phenol red-free medium containing 5% FBS and 40% Dextran T2000. Cells were kept on ice and cryo-fixed in a high-pressure freezer (Leica EM HPM100) according to manufacturer's instructions. After fixation, samples were dehydrated by freeze substitution and embedded in plastic (Epon812, TAAB Laboratories Equipment, Aldermaston, United Kingdom). The embedded samples were cut into 70 nm ultrathin sections, and stained with saturated uranyl acetate and Reynolds lead citrate solution. TEM images were acquired with a JEOL JEM-1011 transmission electron microscope (JEOL, JEM-1011, Tokyo, Japan).

Western Blotting

Cells were infected with GAS as described above. At indicated times post infection, cells were lysed with RIPA lysis buffer

containing protease inhibitor (Promega). Cell lysates were quantified using Bradford protein assay (Bio-Rad, Hercules, CA, United States) and boiled in SDS sample buffer for 10 min. All samples were then separated by 12% SDS-polyacrylamide gel and transferred onto the polyvinylidene difluoride (PVDF) membranes (Millipore, Boston, MA, United States). After blocking with 5% skim milk, the membranes were incubated with primary antibody against ATG9A (Abcam Technology, Cambridge, MA, United States), LC3 (MBL, Nagoya, Japan), or β-actin (novusbio, Littleton, CO) in blocking buffer at 4°C for overnight. After incubation, horseradish peroxidase (HRP)-conjugated secondary antibody (Jackson Immunoresearch Laboratories, West Grove, PA, United States) were used to visualize the Western blot. All images were acquired with the ImageQuant LAS-4000 imaging system (GE Healthcare Life Sciences, Pittsburgh, PA, United States).

Confocal Microscopy

The confocal microscopy was performed as previously described (Hsieh et al., 2018). All images were acquired and analyzed with the confocal microscope LSM 880 (Carl Zeiss, Jena, Germany). The percentage of colocalization between GAS, autophagosomal LC3 (MBL), lysosomal marker LAMP1 (Cell Signaling Technology, Danvers, MA, United States), and acidotropic probe Lysotracker (Invitrogen Molecular Probes, Eugene, OR, United States) was quantified from five independent fields of each experiment.

Statistical Analysis

Data are representative of three independent experiments and expressed as the mean ± standard error of the mean (SEM). GraphPad Prism 5.0 was used to plot all graphs and evaluate statistical significance by 1-way or 2-way ANOVA with *post hoc* Tukey's or Bonferroni multiple comparison test, respectively. Results were considered as statistically significant when the *p*-value was less than 0.05, and marked by an asterisk (*) in figures.

RESULTS

Nicotinamide Inhibits Intracellular GAS Growth in Endothelial Cells

NAD-glycohydrolase is a potent glycohydrolase that cleaves β-NAD to produce nicotinamide and ADP-ribose, leading to disrupted energy homeostasis and promoting bacterial survival in host cells (Bastiat-Sempe et al., 2014; Hsieh et al., 2018; Pajuelo et al., 2018). However, the underlying mechanisms associating NAD^+ metabolism with intracellular bacterial survival are complicated and poorly understood. To clarify whether NAD^+ depletion or accumulated metabolic products contributed to GAS survival in endothelial cells, exogenous β-NAD, NADH and nicotinamide (NAM) were supplemented in culture medium, and the GAS survival in endothelial HMEC-1 cells was evaluated by a colony-forming assay. The results revealed no significant difference in the bacterial count of intracellular GAS within endothelial HMEC-1 cells under different concentrations of β-NAD^+ and NADH after 5 h of infection (**Figures 1A,B**). In contrast, supplementation with 2.5 to 40 mM of NAM was able to inhibit the intracellular GAS growth within endothelial HMEC-1 cells in a time and dose-dependent manner (**Figure 1C**).

Nicotinamide has been extensively used to treat various clinical disorders and skin infections (Rolfe, 2014). To investigate whether NAM directly inhibits bacterial survival, the TSBY broth was supplemented with NAM to measure the growth of GAS *in vitro*. The results showed that the bacterial growth was not affected in the presence of 2.5 to 40 mM of NAM, when compared with the control medium (**Supplementary Figure S1A**). Previous reports showed that the enzymatic activity of NADase is required for GAS survival in host cells (O'Seaghdha and Wessels, 2013; Sharma et al., 2016). To explore whether NAM could mediate intracellular GAS growth through repressing NADase activity, the GAS was cultured to mid-logarithmic phase in TSBY broth supplemented with different concentrations of NAM. The results showed that deletion of *nga* (SW957) caused loss of the NADase activity, compared to wild-type NZ131. Moreover, NADase activity of wild-type NZ131 in GAS culture supernatant was decreased to 45 and 60.8% in the presence of 40 mM NAM at 4 and 6 h of incubation, respectively (**Supplementary Figure S1B**), when compared with untreated group. In order to clarify whether NAM reduced NADase activity to mediate intracellular GAS survival, the NADase-encoding gene (*nga*) was deleted by homologous recombination (Hsieh et al., 2018). The endothelial HMEC-1 cells were then infected with *nga* mutant under NAM treatment. The results showed that the intracellular growth of the NADase mutant was significant decreased when compared with the wild-type NZ131 in non-NAM-treated HMEC-1 cells. Moreover, supplementation with exogenous NAM further reduced both wild-type NZ131 and NADase mutant growth in HMEC-1 cells during infection (**Figure 1D**). These results imply that NADase activity contributes to intracellular GAS survival, but NAM may trigger other mechanisms to inhibit intracellular GAS growth in endothelial cells.

Nicotinamide Elevates Intracellular NAD^+ Content and the NAD^+/NADH Ratio of Endothelial Cells

Nicotinamide can be used as a precursor for the biosynthesis of NAD^+ through the salvage pathway in mammalian cells (Canto et al., 2015). We were interested to understand whether NAM could increase the intracellular NAD^+ content in endothelial cells. NAM was supplemented in culture medium, and then intracellular NAD^+ and NAD^+/NADH ratio were evaluated at indicated time points post infection. As expected, the results revealed that NZ131-infected cells have lower NAD^+ content and NAD^+/NADH ratios than non-infected cells at 1 and 5 h of infection. In contrast, supplementation with exogenous NAM could significantly increase intracellular NAD^+ content and the NAD^+/NADH ratio in endothelial cells at 1 and 5 h of infection, compared with untreated cells (**Figures 2A,B**). In order to illustrate the correlation between exogenous NAM supplementation and intracellular NAD^+ content in endothelial

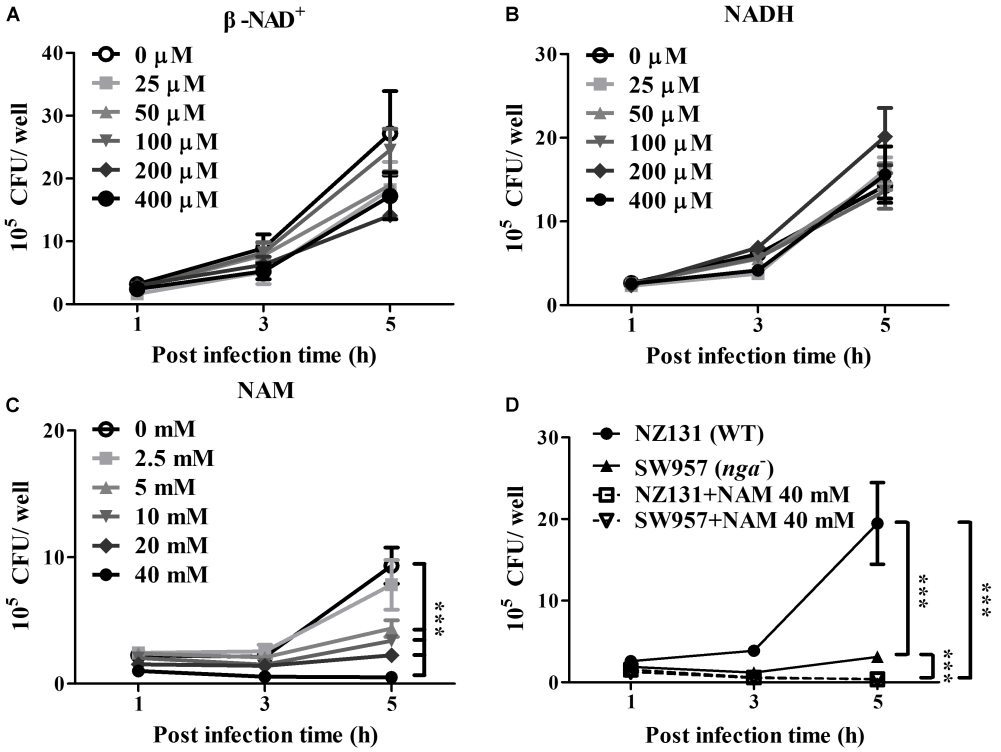


FIGURE 1 | Nicotinamide inhibits intracellular GAS growth in endothelial cells. The intracellular growth of GAS was analyzed in endothelial HMEC-1 cells under exogenous NAD^+ substrate supplementation. HMEC-1 cells were infected with wild-type NZ131 at M.O.I. of 1 and gentamicin was used to kill extracellular bacteria. Cells subsequently were maintained in the fresh M200 medium with/without different concentrations of **(A)** β-NAD, **(B)** NADH, or **(C)** nicotinamide (NAM) for an additional 2 and 4 h. The intracellular bacterial load was counted by CFU-based assay. **(D)** The intracellular growth of GAS was analyzed in the NAM-treated endothelial HMEC-1 cells. HMEC-1 cells were infected with wild-type NZ131 or NADase-knockout strain SW957, treated with gentamicin, and maintained in 40 mM of NAM. The intracellular bacterial load was counted by CFU-based assay. The data represent the means ± SEM of at least three independent experiments. ***$p < 0.001$ (2-way ANOVA).

cells during GAS infection, the intracellular NAD^+ content and NAD^+/NADH ratio of HMEC-1 cells were analyzed by treatment with different concentrations of NAM (2.5 to 40 mM). The results showed that supplementation with exogenous NAM can gradually elevate intracellular NAD^+ content and the NAD^+/NADH ratio in NZ131-infected HMEC-1 cells (**Figures 2C,D**), suggesting that both intracellular NAD^+ level and NAD^+/NADH ratio are correlated with NAM treatment in endothelial cells during GAS infection. In addition to NAM, the intracellular NAD^+ content of NZ131-infected cells was also measured after exogenous β-NAD and NADH supplementation. The results showed that the intracellular NAD^+ content was not increased after β-NAD and NADH treatment, when compared with untreated and NAM-treated cells at 5 h of infection (**Supplementary Figure S2**).

Nicotinamide adenine dinucleotide (NAD^+) is considered as a critical molecule involved in diverse physiological functions including energy production and cell viability (Filomeni et al., 2015). To illustrate whether increased NAD^+ contributed to cell survival, the protease-based cytotoxicity assay was used to measure the cell viability of GAS-infected cells under NAM treatment. The results showed that the GAS infection induced higher cytotoxicity than non-infected cells, but NAM treatment can dramatically decrease cytotoxicity in both non-infected and GAS-infected cells at 1 and 5 h of infection (**Figure 2E**).

Nicotinamide Modulates Autophagic Activation to Inhibit GAS Growth in Endothelial Cells

Several studies have indicated that NAD^+ homeostasis plays a prominent role in the regulation of autophagy responses (Preyat and Leo, 2016; Zhang et al., 2016). Accordingly, the autophagy-related ATG9A was knocked down by lentiviral short hairpin RNA (shRNA) to explore whether the increasing NAD^+ was associated with autophagy-mediated intracellular GAS clearance in endothelial cells. The results of Western blot analysis showed that the clone 82 had the highest knockdown efficiency of ATG9A, compared to wild-type and negative control shLuc-infeced HMEC-1 cells (**Figure 3A**). In addition, the autophagy-related LC3-II form was decreased in ATG9A-knockdown cells when compared to wild-type cells (**Supplementary Figure S3A**). The clone 82 was subsequently chosen to infect with GAS for analyzing bacterial intracellular growth under NAM treatment. The results showed that the intracellular bacterial load was significantly increased in ATG9A-knockdown HMEC-1 cells

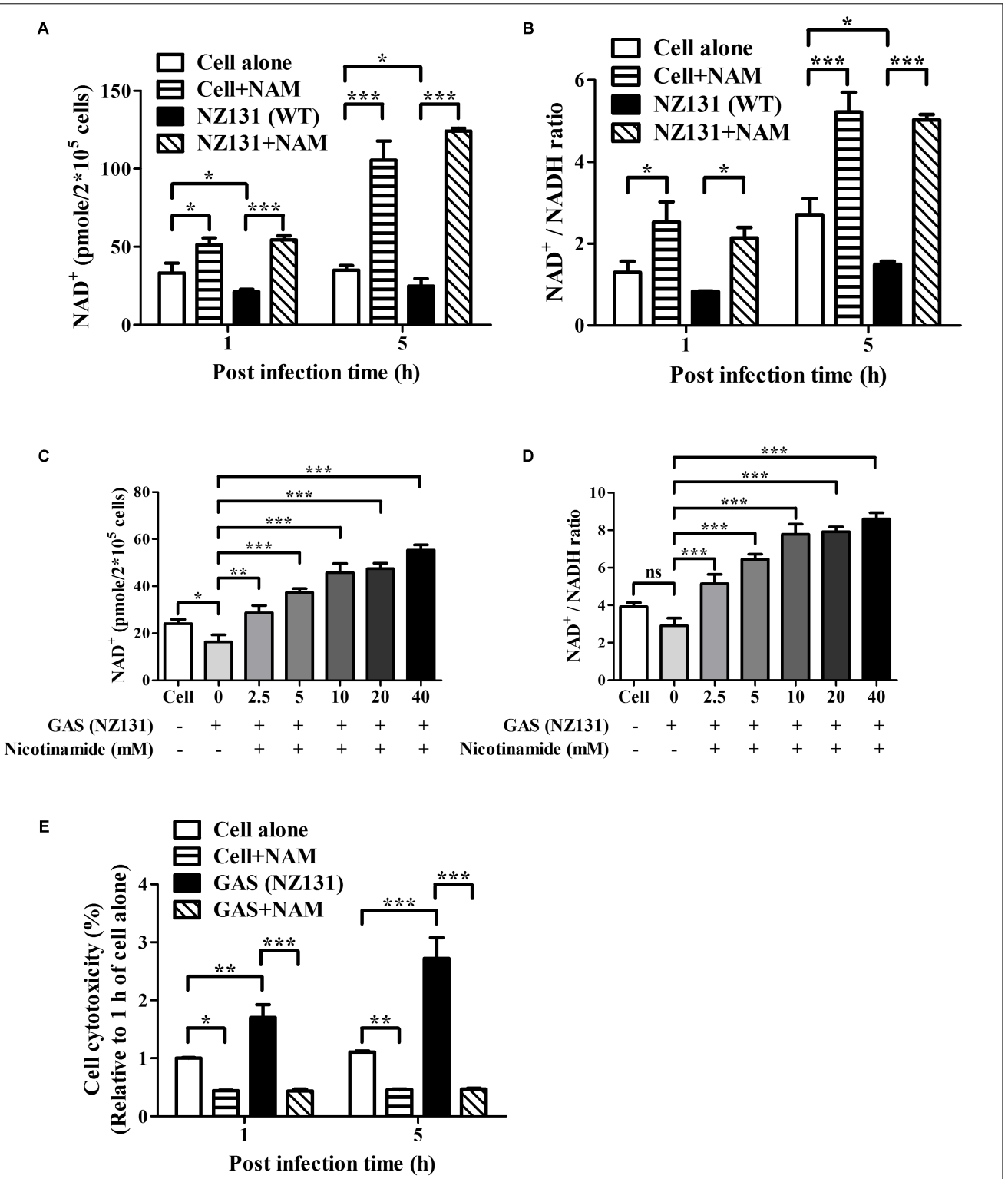


FIGURE 2 | Nicotinamide increases intracellular NAD+ content and NAD+/NADH ratio of endothelial cells. **(A,C)** The intracellular NAD+ content of endothelial HMEC-1 cells was measured after nicotinamide (NAM) treatment. The cell lysates were extracted from GAS-infected HMEC-1 cells with/without NAM treatment and analyzed by a NAD+/NADH quantification kit. **(B,D)** The NAD+/NADH ratio of GAS-infected HMEC-1 cells was calculated following a formula: [($NADH_{total}$ – NADH)/NADH]. **(E)** The cell viability of GAS-infected cells after NAM treatment were measured by CytoTox-Glo cytotoxicity assay kit. The cell culture supernatants were harvested after GAS infection with/without NAM treatment at indicated time point, and then co-incubated with luminogenic substrate. The cell cytotoxicity was expressed as relative percentage compared to 1 h of cell alone. The data represent the means ± SEM of at least three independent experiments. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ (1- or 2-way ANOVA).

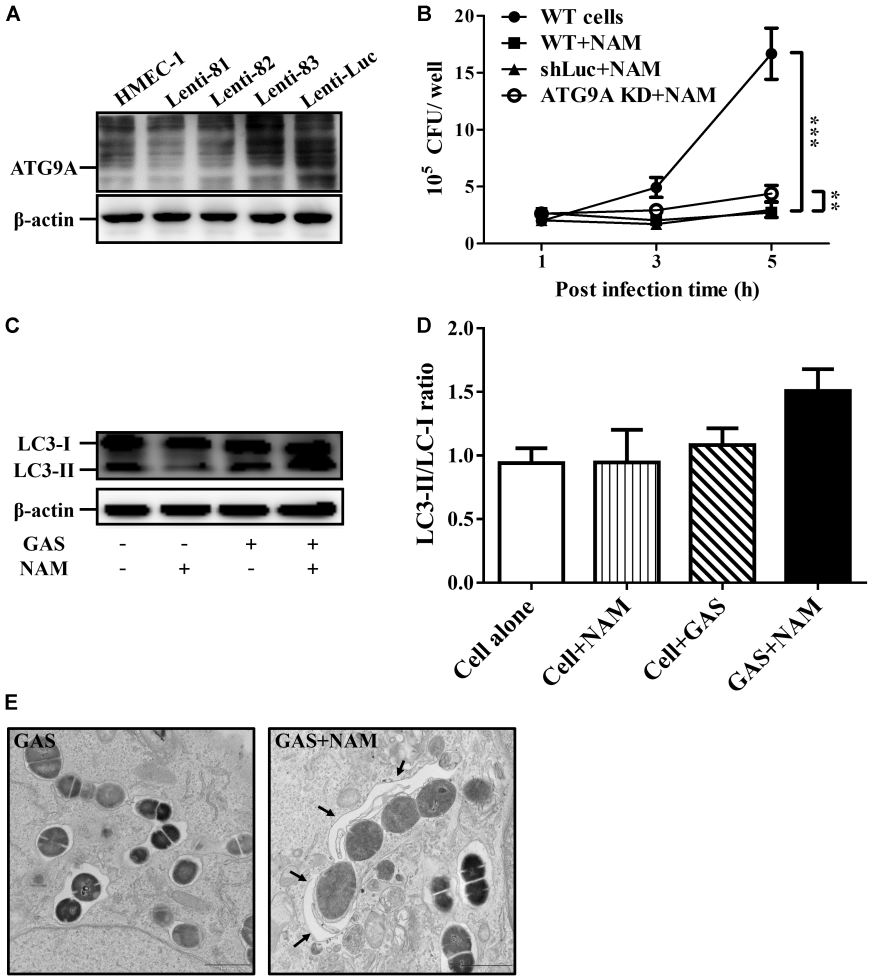


FIGURE 3 | Nicotinamide activates autophagy to inhibit intracellular GAS growth in endothelial cells. **(A)** The expression of ATG9A in endothelial HMEC-1 cells. HMEC-1 cells were transfected with three lentivirus-based shRNAs (shATG9A 81, 82, and 83) to silence the ATG9A expression or Luciferase shRNA (shLuc) for a negative control. The expression of ATG9A was examined by Western blot analysis. **(B)** The intracellular growth of GAS was evaluated in ATG9A-knockdown (KD) HMEC-1 cells under nicotinamide (NAM) treatment. The wild-type and lentivirus-infected HMEC-1 cells (shLuc and clone 82) were infected with GAS at M.O.I. of 1, and gentamicin was used to kill extracellular bacteria. The intracellular bacterial load was counted by CFU-based assays. The data represent the means ± SEM of at least three independent experiments. $^{**}p < 0.01$; $^{***}p < 0.001$ (2-way ANOVA). **(C)** The LC3 conversion of GAS-infected HMEC-1 cells under NAM treatment were analyzed by Western blotting. **(D)** The relative band intensities of LC3-I and LC3-II were measured by densitometry analysis using ImageJ software and expressed as the ratio of LC3-II/LC3-I. **(E)** Cells were infected with GAS and treated with/without NAM at 1 h of infection. The membrane structure of GAS-containing vacuoles were observed by conventional transmission electron microscopy (TEM). Black arrow indicate the double membrane compartment. Scale bar = 1 μm.

under 40 mM of NAM supplementation, when compared to wild-type and shLuc-infected cells (**Figure 3B**). In order to clarify whether autophagy could be activated by NAM treatment, the LC3 conversion was analyzed in endothelial HMEC-1 cells treated with NAM during GAS infection. The results of the Western blotting showed that a trend of increasing level of LC3-II form was observed in NAM-treated cells at 5 h of infection, however, it has no significance when compared to non-treated cells (lane 3 versus lane 4, p-value = 0.1, **Figures 3C,D**). In contrast, the LC3-II form was not induced in ATG9A-knockdown HMEC-1 cells after NAM treatment (**Supplementary Figure S3B**). Next, transmission electron microscopy (TEM) was further used to visualize the autophagosome membrane structure in GAS-infected HMEC-1 cells. The images revealed that intracellular GAS was surrounded by the single membrane structure, but double membrane-engulfed GAS was markedly observed in NAM-treated HMEC-1 cells (**Figure 3E**). Theses evidences suggest that intracellular GAS can be sequestered by autophagy in endothelial cells after NAM treatment.

Nicotinamide Enhances Intracellular Trafficking of GAS-Containing Vacuoles to Lysosome and Acidification in Endothelial Cells

Previous studies have shown that GAS can survive in host cells through avoidance of lysosomal degradation with bacteria-containing vacuoles (O'Seaghdha and Wessels, 2013; Sharma

et al., 2016; Hsieh et al., 2018). Therefore, whether NAM could enhance lysosomal recruitment with GAS-containing vacuoles to inhibit GAS growth in endothelial cells was further analyzed. The intracellular localization of bacteria, LC3 decorated vesicles, and lysosomes in endothelial HMEC-1 cells were observed by confocal microscopy after NAM treatment. Confocal images showed that the intracellular bacteria were colocalized with LC3-positive vacuoles, but the colocalization of intracellular bacteria with LC3 puncta was reduced in HMEC-1 cells without NAM treatment after 5 h of infection. In contrast, more intracellular bacteria were colocalized with LC3-positive vacuoles in NAM-treated HMEC-1 cells compared to without NAM treatment after 5 h of infection (**Figure 4A**). Furthermore, the membrane glycoprotein LAMP-1 was used as a marker for visualizing the colocalization of lysosomes with intracellular GAS in endothelial HMEC-1 cells. The results showed that only 25 and 17% of intracellular GAS was associated with LAMP-1 in untreated cells at 3 and 5 h of infection, respectively. Under the NAM treatment, 61 and 74% of intracellular GAS was colocalized with LAMP-1 at 3 and 5 h of infection, respectively (**Figures 4A,B**).

Sufficient acidification is required for lysosomal enzyme activity to eliminate intracellular bacteria in endothelial cells and macrophages (Bastiat-Sempe et al., 2014; Valderrama and Nizet, 2018). In order to clarify whether the GAS-containing vacuoles were successfully acidified after NAM treatment, the fluorescent indicator Lysotracker was used to observe the acidification within NAM-treated endothelial HMEC-1 cells during GAS infection. The results revealed that about 40% of intracellular bacteria were colocalized with Lysotracker under NAM-treated and untreated conditions at 3 h of infection. However, the colocalization of intracellular GAS with Lysotracker was higher in NAM-treated endothelial HMEC-1 cells compared to cells without NAM treatments at 5 h of infection (51 vs. 13%, respectively) (**Figures 5A,B**). The activity of lysosomal H^+-ATPase was inhibited by bafilomycin A1 to evaluate the roles of acidification for intracellular survival of GAS in NAM-treated endothelial HMEC-1 cells. The results showed that the intracellular bacterial load was significantly increased following treatment with bafilomycin A1 at 5 h of infection compared to untreated cells under NAM-supplemented conditions (**Figure 5C**). These results indicate that treatment with exogenous NAM markedly enhances lysosome-mediated acidification to inhibit intracellular GAS growth in endothelial cells.

Nicotinamide Specifically Inhibits Multiplication of Intracellular GAS in Endothelial Cells

Nicotinamide and its derivatives have been extensively used to treat bacterial and viral infections (Murray, 2003; Singhal and Cheng, 2019). Based on these findings, we sought to examine whether NAM could inhibit other pathogens' survival in endothelial cells. The serotype M1 and M49 GAS, MRSA, and *S.* Typhimurium were utilized to infect endothelial HMEC-1 cells and the intracellular bacterial growth was measured by a colony-forming assay. The results showed that the intracellular growth of both serotype M1 (strain A20) and M49 (strain NZ131) was significantly reduced in NAM-treated endothelial HMEC-1 cells at 5 h of infection, compared to non-treated cells (**Figure 6A**). However, supplementation with exogenous NAM did not significantly inhibit the growth of MRSA and *S.* Typhimurium in endothelial HMEC-1 cells (**Figure 6B**), implying that these pathogens may use different mechanisms to escape NAM-mediated clearance in endothelial cells.

DISCUSSION

Nicotinamide adenine dinucleotide (NAD^+) is one of the most important molecules involved in energy generation and signal transduction to control cellular metabolism in live cells (Canto et al., 2015). However, several studies found that many pathogens express multiple toxins to influence intracellular NAD^+ homeostasis of infected cells, which results in energy collapse, impaired host defense, immunopathology, and increased pathogen survival (Bastiat-Sempe et al., 2014; Sun et al., 2015; Bhat et al., 2016; Mesquita et al., 2016; Pajuelo et al., 2018). In this study, we found that supplementation with exogenous NAM could increase intracellular NAD^+ levels and the NAD^+/NADH ratio that contributed to enhance intracellular GAS clearance by autophagy in endothelial cells (**Figure 7**).

In order to clarify the correlation between the intracellular NAD^+ content and bacterial survival of infected cells, the exogenous NAD^+-associated substrates were supplemented in GAS-infected endothelial cells and bacterial survival was analyzed by a colony-formation assay. Our results demonstrated that supplementation with exogenous NAM not only dramatically increased intracellular NAD^+ level and NAD^+/NADH ratio, but also inhibited GAS growth in endothelial cells in a dose- and time-dependent manner (**Figures 1**, **2**). In contrast to NAM, the intracellular NAD^+ content and bacterial load were not affected by the exogenous β-NAD and NADH treatments (**Supplementary Figure S2**), which might be due to the membrane impermeability (Canto et al., 2015). In addition to NAM, intracellular NAD^+ is also generated from tryptophan through a *de novo* biosynthetic pathway in the cells (Moffett and Namboodiri, 2003). Recent studies have shown that tryptophan catabolism plays a crucial role against bacterial and viral infections, including those caused by *Clostridium difficile*, *Mycobacterium tuberculosis*, and hepatitis B and C viruses (Larrea et al., 2007; El-Zaatari et al., 2014; Schmidt and Schultze, 2014). However, whether tryptophan is involved in NAD^+ fluctuation influencing GAS survival in endothelial cells needs to be evaluated. These results indicate that the importance of intracellular NAD^+ homeostasis on pathogens' survival in the host.

Our results have shown that the given exogenous NAM can increase the conversion of LC3-I to LC3-II and induce autophagic double membrane formation to inhibit GAS survival in endothelial cells (**Figures 3–5**). The LC3-II level and ratio

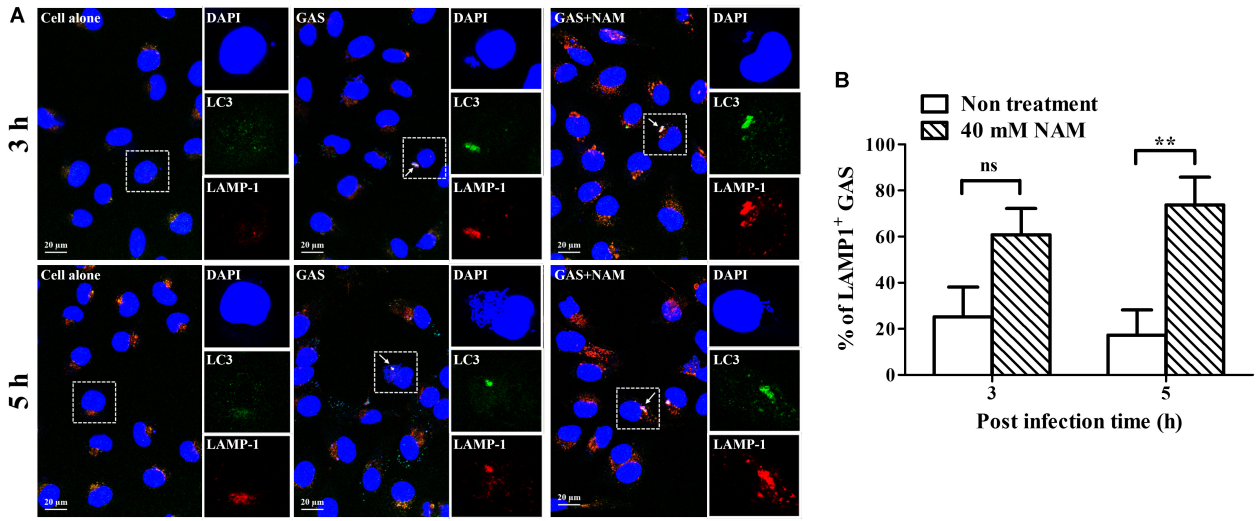


FIGURE 4 | Nicotinamide enhances GAS-containing vacuole trafficking to lysosomes in endothelial cells. **(A)** The colocalization of intracellular GAS with autophagosomes and lysosomes was analyzed in nicotinamide (NAM)-treated endothelial HMEC-1 cells. Confocal microscopy was used to observe the colocalization of GAS (DAPI, blue) with LC3 (Alexa 488, green) and LAMP-1 (Alexa 594, red) in HMEC-1 cells under NAM treatment at 3 and 5 h of infection. Representative images are shown from three independent experiments. Scale bar = 20 μm. **(B)** The intracellular bacteria associated with LAMP-1 at 3 and 5 h of infection were quantified from confocal microscopy images. At least 100 intracellular GAS were quantified for each time point in at least three independent experiments. The data represent the means ± SEM of at least three independent experiments. ns, not significant; $^{**}p < 0.01$ (2-way ANOVA).

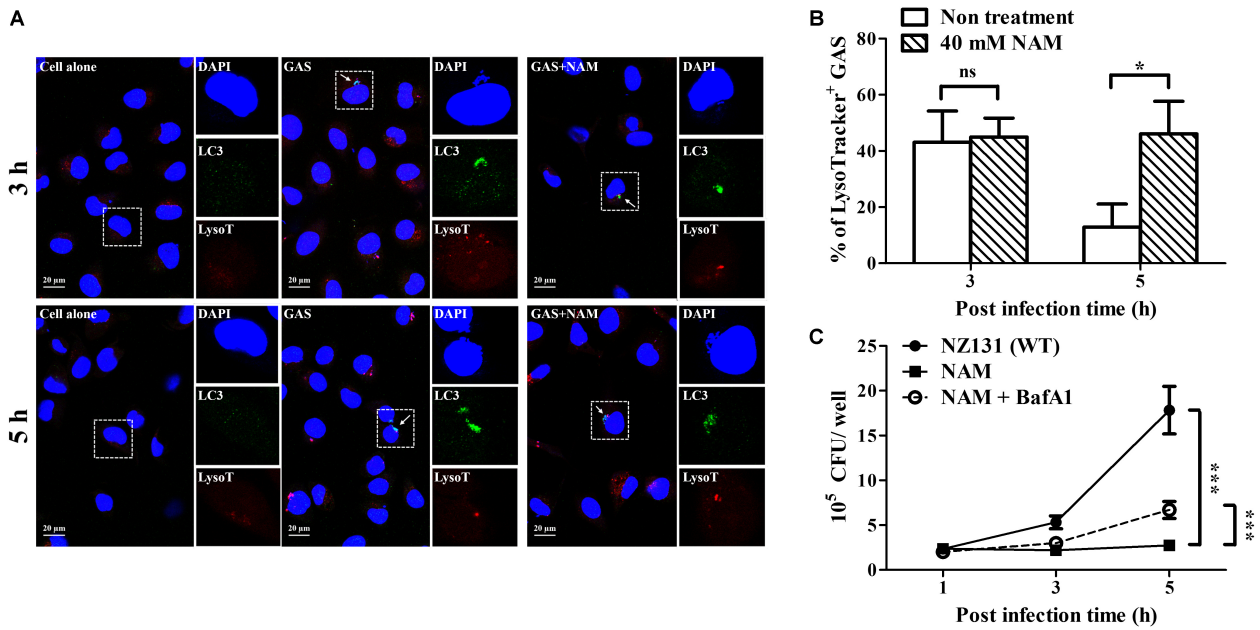


FIGURE 5 | Nicotinamide increases acidification of GAS-containing vacuole in endothelial cells. **(A)** The colocalization of intracellular GAS with acidified autophagosomes was analyzed in nicotinamide (NAM)-treated HMEC-1 cells. The intracellular localization of GAS (DAPI, blue), LC3 (Alexa 488, green), and acidification (acidotropic indicator LysoTracker Red DND-99, red) in endothelial cells was observed by confocal microscopy at 3 and 5 h of infection. Representative images are shown from three independent experiments. Scale bar = 20 μm. **(B)** The percentage of intracellular bacteria associated with LysoTracker was quantified at 3 and 5 h of infection. At least 100 intracellular GAS were quantified for each time point in at least three independent experiments. **(C)** The intracellular growth of GAS was analyzed in endothelial HMEC-1 cells under bafilomycin A1 (BafA1) and NAM cotreatment. HMEC-1 cells with/without pretreatment with 200 nM BafA1, were then infected with GAS at M.O.I. of 1 in presence or absence of 40 mM NAM. The intracellular bacterial load was counted by CFU-based assays. The data represent the means ± SEM of at least three independent experiments. ns, not significant; $^{*}p < 0.05$; $^{***}p < 0.001$ (2-way ANOVA).

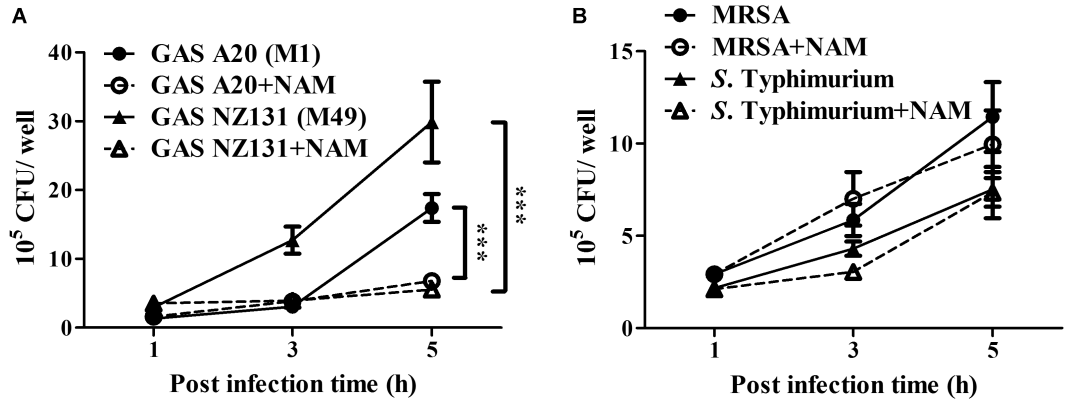


FIGURE 6 | Nicotinamide specifically inhibits intracellular multiplication of GAS in endothelial cells. The intracellular growth of GAS, MRSA, and *S.* Typhimurium was analyzed in nicotinamide (NAM)-treated endothelial HMEC-1 cells. HMEC-1 cells were infected with **(A)** serotype M1 A20 at M.O.I. of 10, M49 NZ131 at M.O.I of 1 or **(B)** MRSA and *S.* Typhimurium at M.O.I of 1. The extracellular bacteria were killed by gentamicin and intracellular bacterial load was counted by CFU-based assays. The data represent the means ± SEM of at least three independent experiments. ***$p < 0.001$ (2-way ANOVA).

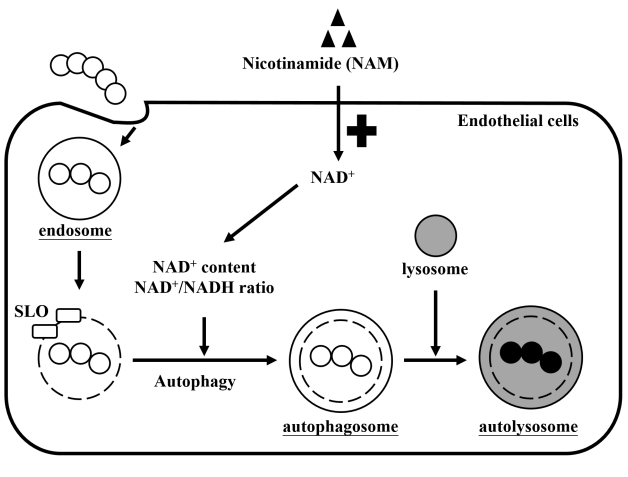


FIGURE 7 | Effects of nicotinamide on GAS survival in endothelial cells. GAS is internalized into cells through endocytosis to form the endosomes. However, SLO is expressed to damage the endosomal membrane and further induces autophagy responses. GAS depletes intracellular NAD^+ promoting their intracellular survival, but cells supplemented with exogenous nicotinamide (NAM) can restore the intracellular NAD^+ content and NAD^+/NADH ratio, resulting in GAS clearance by autophagy in endothelial cells.

of LC3-II/LC3-I have been considered as a reliable marker for reflecting the activation of autophagy (Kadowaki and Karim, 2009). Recently, several studies have shown that supplementation with NAM can induce the conversion of LC3-I to LC3-II in chronic hypoxic myocardial cells, which support our results that the LC3-II was increased after NAM treatment in GAS-infected endothelial cells (Kang and Hwang, 2009; Hwang and Song, 2017; Li et al., 2019). These evidences indicate that NAM can increase autophagic activation under different stress conditions. However, the LC3-II-decorated single-membrane phagosomes or vacuoles (LAP) recently has been found in microbial survival within phagocytic and non-phagocytotic cells (Schille et al., 2018; Upadhyay and Philips, 2019). Therefore, the signaling pathways and membrane structure can be utilized to define the difference between the conventional autophagosome and LC3-associated phagosome (LAPosome) (Martinez et al., 2015; Schille et al., 2018). Autophagy is a tightly regulated process influenced by many NAD^+-consuming enzymes including ADP-ribose polymerase-1 (PARP-1) and sirtuins (Zhang et al., 2016). Several studies have demonstrated that the NAD^+-dependent sirtuins are involved in bacterial survival by regulating autophagy via activation of the AMPK/mTOR signaling pathway in intestinal epithelial cells and macrophages (Ganesan et al., 2017; Al Azzaz et al., 2018; Li et al., 2019; Yang et al., 2019). Moreover, exogenous NAM treatment can enhance Sirtuin-1 activity to control autophagy-mediated mitochondrial quality in human fibroblasts (Kang and Hwang, 2009; Jang et al., 2012; Shen et al., 2017). On the other hand, reactive oxygen species (ROS) are considered as another critical mediators of NAD^+ in autophagy responses (Filomeni et al., 2015; Preyat and Leo, 2016). We recently found that GAS infection can induce ROS production to stimulate single membrane LAPosome formation for bacterial survival and activate the inhibition of mTOR for GAS escaping from autophagic clearance in endothelial cells. In contrast, suppression of ROS level significantly enhance GAS clearance through converting defective LAPosome to conventional double membrane autophagosome in endothelial cells (Lu et al., 2017; Cheng et al., 2019). Moreover, autophagy could be upregulated by supplementation with exogenous NAD^+, which can suppress PARP-1-mediated necrotic death of retinal pigment epithelium cells under oxidative stress (Zhu et al., 2016). These evidences suggest that NAM may modulate sirtuins activity or ROS production to regulate signaling transduction of conventional autophagy for intracellular GAS clearance in endothelial cells, but the underlying mechanisms require further investigation.

Nicotinamide has been considered as a safe and readily available therapeutic agent to improve the metabolic diseases, neurodegeneration, and skin infection through raising

intracellular NAD^+ level (Knip et al., 2000; Fricker et al., 2018). In this study, we showed that supplementation with exogenous NAM greatly reduced the intracellular growth of both serotype M1 and M49 GAS in endothelial cells, but did not dramatically inhibit the growth of MRSA and *S.* Typhimurium (**Figure 6**). Recent studies have shown that the *M. tuberculosis* depletes intracellular NAD^+ content to enhance their intracellular survival, but treatment with NAM can restore intracellular NAD^+ level and reduce the bacterial load of *M. tuberculosis* in macrophages (Sun et al., 2015; Pajuelo et al., 2018). Interestingly, NADase depleting intracellular NAD^+ is required for the intracellular survival of GAS and *M. tuberculosis* in host cells, but the NADase activity of MRSA and *S.* Typhimurium cannot be determined in the culture supernatants (data not shown). These results imply that GAS and *M. tuberculosis* may share similar NAD^+-associated mechanisms for intracellular survival in host cells.

In summary, we conclude that intracellular GAS survival within endothelial cells is linked to NAD^+ homeostasis, indicating that NAM and its analogs may provide novel therapeutic strategies against GAS infection.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

Y-SL, CC-N, P-JT, S-YW, C-CL, TN, and J-JW conceived and designed the study. C-LH, S-YH, H-MH, Y-NH, P-XZ, and J-JW analyzed the intracellular bacterial survival. C-LH, S-YH, Y-LC, and J-JW analyzed the cell response. C-LH, S-YH, S-LL, HO, TN, and J-JW did the images acquisition. C-LH, CC-N, P-XZ, and J-JW wrote the manuscript.

FUNDING

This work was supported by grants MOST105-2320-B006-009, MOST106-2320-B010-039, MOST 107-2320-B010-021, and MOST 108-2320-B010-004 from Ministry of Science and Technology, Taiwan.

ACKNOWLEDGMENTS

We are very grateful to Prof. Robert M. Jonas for their helpful comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00117/full#supplementary-material

REFERENCES

Al Azzaz, J., Rieu, A., Aires, V., Delmas, D., Chluba, J., Winckler, P., et al. (2018). Resveratrol-induced xenophagy promotes intracellular bacteria clearance in intestinal epithelial cells and macrophages. *Front. Immunol.* 9:3149. doi: 10.3389/fimmu.2018.03149

Bah, A., and Vergne, I. (2017). Macrophage autophagy and bacterial infections. *Front. Immunol.* 8:1483. doi: 10.3389/fimmu.2017.01483

Barnett, T. C., Bowen, A. C., and Carapetis, J. R. (2018). The fall and rise of group A *Streptococcus* diseases. *Epidemiol. Infect.* 15, 1–6. doi: 10.1017/S0950268818002285

Barnett, T. C., Liebl, D., Seymour, L. M., Gillen, C. M., Lim, J. Y., Larock, C. N., et al. (2013). The globally disseminated M1T1 clone of group A *Streptococcus* evades autophagy for intracellular replication. *Cell Host Microbe* 14, 675–682. doi: 10.1016/j.chom.2013.11.003

Bastiat-Sempe, B., Love, J. F., Lomayesva, N., and Wessels, M. R. (2014). Streptolysin O and NAD-glycohydrolase prevent phagolysosome acidification and promote group A *Streptococcus* survival in macrophages. *mBio* 5:e1690-14. doi: 10.1128/mBio.01690-14

Bhat, S. A., Iqbal, I. K., and Kumar, A. (2016). Imaging the NADH:NAD+ homeostasis for understanding the metabolic response of *Mycobacterium* to physiologically relevant stresses. *Front. Cell. Infect. Microbiol.* 6:145. doi: 10.3389/fcimb.2016.00145

Bricker, A. L., Cywes, C., Ashbaugh, C. D., and Wessels, M. R. (2002). NAD+-glycohydrolase acts as an intracellular toxin to enhance the extracellular survival of group A streptococci. *Mol. Microbiol.* 44, 257–269. doi: 10.1046/j.1365-2958.2002.02876.x

Canto, C., Menzies, K. J., and Auwerx, J. (2015). NAD+ metabolism and the control of energy homeostasis: a balancing act between mitochondria and the nucleus. *Cell Metab.* 22, 31–53. doi: 10.1016/j.cmet.2015.05.023

Castrejón-Jiménez, N. S., Leyva-Paredes, K., Hernandez-González, J. C., Luna-Herrera, J., and García-Pérez, B. E. (2015). The role of autophagy in bacterial infections. *Biosci. Trends* 9, 149–159. doi: 10.5582/bst.2015.01035

Chandrasekaran, S., and Caparon, M. G. (2015). The *Streptococcus pyogenes* NAD+ glycohydrolase modulates epithelial cell PARylation and HMGB1 release. *Cell. Microbiol.* 17, 1376–1390. doi: 10.1111/cmi.12442

Chandrasekaran, S., and Caparon, M. G. (2016). The NADase-negative variant of the *Streptococcus pyogenes* toxin NAD+ glycohydrolase induces JNK1-mediated programmed cellular necrosis. *mBio* 7:e2215. doi: 10.1128/mBio.02215-15

Cheng, Y. L., Kuo, C. F., Lu, S. L., Hiroko, O., Wu, Y. N., Hsieh, C. L., et al. (2019). Group A *Streptococcus* induces LAPosomes via SLO/β1 integrin/NOX2/ROS pathway in endothelial cells that are ineffective in bacterial killing and suppress xenophagy. *mBio* 10:e2148-19. doi: 10.1128/mBio.02148-19

Cunningham, M. W. (2008). Pathogenesis of group A streptococcal infections and their sequelae. *Adv. Exp. Med. Biol.* 609, 29–42. doi: 10.1007/978-0-387-73960-1_3

El-Zaatari, M., Chang, Y. M., Zhang, M., Franz, M., Shreiner, A., McDermott, A. J., et al. (2014). Tryptophan catabolism restricts IFN-γ-expressing neutrophils and *Clostridium* difficile immunopathology. *J. Immunol.* 193, 807–816. doi: 10.4049/jimmunol.1302913

Filomeni, G., De Zio, D., and Cecconi, F. (2015). Oxidative stress and autophagy: the clash between damage and metabolic needs. *Cell Death Differ.* 22, 377–388. doi: 10.1038/cdd.2014.150

Fricker, R. A., Green, E. L., Jenkins, S. I., and Griffin, S. M. (2018). The influence of nicotinamide on health and disease in the central nervous system. *Int. J. Tryptophan Res.* 11:1178646918776658. doi: 10.1177/1178646918776658

Ganesan, R., Hos, N. J., Gutierrez, S., Fischer, J., Stepek, J. M., Daglidu, E., et al. (2017). *Salmonella typhimurium* disrupts Sirt1/AMPK checkpoint control of mTOR to impair autophagy. *PLoS Pathog.* 13:e1006227. doi: 10.1371/journal.ppat.1006227

Hsieh, C. L., Huang, H. M., Hsieh, S. Y., Zheng, P. X., Lin, Y. S., Chiang-Ni, C., et al. (2018). NAD-glycohydrolase depletes intracellular NAD+ and inhibits acidification of autophagosomes to enhance multiplication of group A *Streptococcus* in endothelial cells. *Front. Microbiol.* 9:1733. doi: 10.3389/fmicb.2018.01733

Hwang, E. S., and Song, S. B. (2017). Nicotinamide is an inhibitor of SIRT1 *in vitro*, but can be a stimulator in cells. *Cell. Mol. Life Sci.* 74, 3347–3362. doi: 10.1007/s00018-017-2527-8

Jang, S. Y., Kang, H. T., and Hwang, E. S. (2012). Nicotinamide-induced mitophagy: event mediated by high NAD+/NADH ratio and SIRT1 protein activation. *J. Biol. Chem.* 287, 19304–19314. doi: 10.1074/jbc.M112.363747

Kadowaki, M., and Karim, M. R. (2009). Cytosolic LC3 ratio as a quantitative index of macroautophagy. *Methods Enzymol.* 452, 199–213. doi: 10.1016/S0076-6879(08)03613-6

Kang, H. T., and Hwang, E. S. (2009). Nicotinamide enhances mitochondria quality through autophagy activation in human cells. *Aging Cell.* 8, 426–438. doi: 10.1111/j.1474-9726.2009.00487.x

Kaplan, E. L., Chhatwal, G. S., and Rohde, M. (2006). Reduced ability of penicillin to eradicate ingested group A streptococci from epithelial cells: clinical and pathogenetic implications. *Clin. Infect. Dis.* 43, 1398–1406. doi: 10.1086/508773

Kimmey, J. M., and Stallings, C. L. (2016). Bacterial pathogens versus autophagy: implications for therapeutic interventions. *Trends Mol. Med.* 22, 1060–1076. doi: 10.1016/j.molmed.2016.10.008

Klionsky, D. J., Abdelmohsen, K., Abe, A., Abedin, M. J., Abeliovich, H., Acevedo Arozena, A., et al. (2016). Guidelines for the use and interpretation of assays for monitoring autophagy (3rd edition). *Autophagy* 12, 1–222. doi: 10.1080/15548627.2015.1100356

Knip, M., Douek, I. F., Moore, W. P., Gillmor, H. A., McLean, A. E., Bingley, P. J., et al. (2000). Safety of high-dose nicotinamide: a review. *Diabetologia* 43, 1337–1345. doi: 10.1007/s001250051536

Kocaturk, N. M., and Gozuacik, D. (2018). Crosstalk between mammalian autophagy and the ubiquitin-proteasome system. *Front. Cell Dev. Biol.* 6:128. doi: 10.3389/fcell.2018.00128

Larrea, E., Riezu-Boj, J. I., Gil-Guerrero, L., Casares, N., Aldabe, R., Sarobe, P., et al. (2007). Upregulation of indoleamine 2,3-dioxygenase in hepatitis C virus infection. *J. Virol.* 81, 3662–3666. doi: 10.1128/JVI.02248-06

Li, W., Zhu, L., Ruan, Z. B., Wang, M. X., Ren, Y., and Lu, W. (2019). Nicotinamide protects chronic hypoxic myocardial cells through regulating mTOR pathway and inducing autophagy. *Eur. Rev. Med. Pharmacol. Sci.* 23, 5503–5511. doi: 10.26355/eurrev_201906_18220

Lu, S. L., Kawabata, T., Cheng, Y. L., Omori, H., Hamasaki, M., Kusaba, T., et al. (2017). Endothelial cells are intrinsically defective in xenophagy of *Streptococcus pyogenes*. *PLoS Pathog.* 13:e1006444. doi: 10.1371/journal.ppat.1006444

Lu, S. L., Kuo, C. F., Chen, H. W., Yang, Y. S., Liu, C. C., Anderson, R., et al. (2015). Insufficient acidification of autophagosomes facilitates group A *Streptococcus* survival and growth in endothelial cells. *mBio* 6:e1435-15. doi: 10.1128/mBio.01435-15

Marouni, M. J., Barzilai, A., Keller, N., Rubinstein, E., and Sela, S. (2004). Intracellular survival of persistent group A streptococci in cultured epithelial cells. *Int. J. Med. Microbiol.* 294, 27–33. doi: 10.1016/j.ijmm.2004.01.001

Martinez, J., Malireddi, R. K., Lu, Q., Cunha, L. D., Pelletier, S., Gingras, S., et al. (2015). Molecular characterization of LC3-associated phagocytosis reveals distinct roles for Rubicon, NOX2 and autophagy proteins. *Nat. Cell Biol.* 17, 893–906. doi: 10.1038/ncb3192

Mesquita, I., Varela, P., Belinha, A., Gaifem, J., Laforge, M., Vergnes, B., et al. (2016). Exploring NAD+ metabolism in host-pathogen interactions. *Cell. Mol. Life Sci.* 73, 1225–1236. doi: 10.1007/s00018-015-2119-4

Mestre, M. B., and Colombo, M. I. (2013). Autophagy and toxins: a matter of life or death. *Curr. Mol. Med.* 13, 241–251. doi: 10.2174/1566524011313020002

Michos, A., Gryllos, I., Hakansson, A., Srivastava, A., Kokkotou, E., and Wessels, M. R. (2006). Enhancement of streptolysin O activity and intrinsic cytotoxic effects of the group A streptococcal toxin, NAD-glycohydrolase. *J. Biol. Chem.* 281, 8216–8223. doi: 10.1074/jbc.M511674200

Moffett, J. R., and Namboodiri, M. A. (2003). Tryptophan and the immune response. *Immunol. Cell Biol.* 81, 247–265. doi: 10.1046/j.1440-1711.2003.t01-1-01177.x

Murray, M. F. (2003). Nicotinamide: an oral antimicrobial agent with activity against both *Mycobacterium tuberculosis* and human immunodeficiency virus. *Clin. Infect. Dis.* 36, 453–460. doi: 10.1086/367544

O'Neill, A. M., Thurston, T. L. M., and Holden, D. W. (2016). Cytosolic replication of group A *Streptococcus* in human macrophages. *mBio* 7:e00020-16. doi: 10.1128/mBio.00020-16

O'Seaghdha, M., and Wessels, M. R. (2013). Streptolysin O and its co-toxin NAD-glycohydrolase protect group A *Streptococcus* from xenophagic killing. *PLoS Pathog.* 9:e1003394. doi: 10.1371/journal.ppat.1003394

Osterlund, A., Popa, R., Nikkila, T., Scheynius, A., and Engstrand, L. (1997). Intracellular reservoir of *Streptococcus pyogenes in vivo*: a possible explanation for recurrent pharyngotonsillitis. *Laryngoscope* 107, 640–647. doi: 10.1097/00005537-199705000-00016

Pajuelo, D., Gonzalez-Juarbe, N., Tak, U., Sun, J., Orihuela, C. J., and Niederweis, M. (2018). NAD+ depletion triggers macrophage necroptosis, a cell death pathway exploited by *Mycobacterium tuberculosis*. *Cell Rep.* 24, 429–440. doi: 10.1016/j.celrep.2018.06.042

Preyat, N., and Leo, O. (2016). Complex role of nicotinamide adenine dinucleotide in the regulation of programmed cell death pathways. *Biochem. Pharmacol.* 101, 13–26. doi: 10.1016/j.bcp.2015.08.110

Rolfe, H. M. (2014). A review of nicotinamide: treatment of skin diseases and potential side effects. *J. Cosmet. Dermatol.* 13, 324–328. doi: 10.1111/jocd.12119

Sakurai, A., Maruyama, F., Funao, J., Nozawa, T., Aikawa, C., Okahashi, N., et al. (2010). Specific behavior of intracellular *Streptococcus pyogenes* that has undergone autophagic degradation is associated with bacterial streptolysin O and host small G proteins Rab5 and Rab7. *J. Biol. Chem.* 285, 22666–22675. doi: 10.1074/jbc.M109.100131

Schille, S., Crauwels, P., Bohn, R., Bagola, K., Walther, P., and van Zandbergen, G. (2018). LC3-associated phagocytosis in microbial pathogenesis. *Int. J. Med. Microbiol.* 308, 228–236. doi: 10.1016/j.ijmm.2017.10.014

Schmidt, S. V., and Schultze, J. L. (2014). New insights into IDO biology in bacterial and viral infections. *Front. Immunol.* 5:384. doi: 10.3389/fimmu.2014.00384

Sharma, O., O'Seaghdha, M., Velarde, J. J., and Wessels, M. R. (2016). NAD+-glycohydrolase promotes intracellular survival of group A *Streptococcus*. *PLoS Pathog.* 12:e1005468. doi: 10.1371/journal.ppat.1005468

Shen, C., Dou, X., Ma, Y., Ma, W., Li, S., and Song, Z. (2017). Nicotinamide protects hepatocytes against palmitate-induced lipotoxicity via SIRT1-dependent autophagy induction. *Nutr. Res.* 40, 40–47. doi: 10.1016/j.nutres.2017.03.005

Singhal, A., and Cheng, C. Y. (2019). Host NAD+ metabolism and infections: therapeutic implications. *Int. Immunol.* 31, 59–67. doi: 10.1093/intimm/dxy068

Sun, J., Siroy, A., Lokareddy, R. K., Speer, A., Doornbos, K. S., Cingolani, G., et al. (2015). The tuberculosis necrotizing toxin kills macrophages by hydrolyzing NAD. *Nat. Struct. Mol. Biol.* 22, 672–678. doi: 10.1038/nsmb.3064

Turner, C. E., Abbott, J., Lamagni, T., Holden, M. T., David, S., Jones, M. D., et al. (2015). Emergence of a new highly successful acapsular group A *Streptococcus* clade of genotype *emm89* in the United Kingdom. *mBio* 6:e00622. doi: 10.1128/mBio.00622-15

Upadhyay, S., and Philips, J. A. (2019). LC3-associated phagocytosis: host defense and microbial response. *Curr. Opin. Immunol.* 60, 81–90. doi: 10.1016/j.coi.2019.04.012

Valderrama, J. A., and Nizet, V. (2018). Group A *Streptococcus* encounters with host macrophages. *Future Microbiol.* 13, 119–134. doi: 10.2217/fmb-2017-0142

Velarde, J. J., O'Seaghdha, M., Baddal, B., Bastiat-Sempe, B., and Wessels, M. R. (2017). Binding of NAD+-glycohydrolase to streptolysin O stabilizes both toxins and promotes virulence of group A *Streptococcus*. *mBio* 8:e1382-17. doi: 10.1128/mBio.01382-17

Walker, M. J., Barnett, T. C., McArthur, J. D., Cole, J. N., Gillen, C. M., Henningham, A., et al. (2014). Disease manifestations and pathogenic mechanisms of group A *Streptococcus*. *Clin. Microbiol. Rev.* 27, 264–301. doi: 10.1128/CMR.00101-13

Yang, H., Hu, J., Chen, Y. J., and Ge, B. (2019). Role of Sirt1 in innate immune mechanisms against *Mycobacterium tuberculosis* via the inhibition of TAK1 activation. *Arch. Biochem. Biophys.* 667, 49–58. doi: 10.1016/j.abb.2019.04.006

Zhang, D. X., Zhang, J. P., Hu, J. Y., and Huang, Y. S. (2016). The potential regulatory roles of NAD+ and its metabolism in autophagy. *Metabolism* 65, 454–462. doi: 10.1016/j.metabol.2015.11.010

Zheng, P. X., Chung, K. T., Chiang-Ni, C., Wang, S. Y., Tsai, P. J., Chuang, W. J., et al. (2013). Complete genome sequence of *emm1 Streptococcus pyogenes* A20,

a strain with an intact two-component system, CovRS, isolated from a patient with necrotizing fasciitis. *Genome Announc.* 1:e149-12. doi: 10.1128/genomeA.00149-12

Zhu, L., Olsen, R. J., Nasser, W., Beres, S. B., Vuopio, J., Kristinsson, K. G., et al. (2015). A molecular trigger for intercontinental epidemics of group A *Streptococcus*. *J. Clin. Invest.* 125, 3545–3559. doi: 10.1172/JCI82478

Zhu, Y., Zhao, K. K., Tong, Y., Zhou, Y. L., Wang, Y. X., Zhao, P. Q., et al. (2016). Exogenous NAD+ decreases oxidative stress and protects H2O2-treated RPE cells against necrotic death through the up-regulation of autophagy. *Sci. Rep.* 6:26322. doi: 10.1038/srep26322

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 13 March 2020
doi: 10.3389/fgene.2020.00216

Exploring Phenotypes for Disease Resilience in Pigs Using Complete Blood Count Data From a Natural Disease Challenge Model

Xuechun Bai[1], Austin M. Putz[2], Zhiquan Wang[1], Frédéric Fortin[3], John C. S. Harding[4], Michael K. Dyck[1], Jack C. M. Dekkers[2], Catherine J. Field[1], Graham S. Plastow[1] and PigGen Canada[†]*

[1] Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada, [2] Department of Animal Science, Iowa State University, Ames, IA, United States, [3] Centre de Développement du Porc du Québec, Inc., Quebec City, QC, Canada, [4] Department of Large Animal Clinical Sciences, University of Saskatchewan, Saskatoon, SK, Canada

OPEN ACCESS

Edited by:
Androniki Psifidi,
Royal Veterinary College (RVC),
United Kingdom

Reviewed by:
Marcos Lopes,
Topigs Norsvin, Brazil
Shogo Tsuruta,
University of Georgia, United States

***Correspondence:**
Graham S. Plastow
plastow@ualberta.ca

[†]PigGen Canada authors are listed in the Acknowledgments of the article

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: 25 October 2019
Accepted: 24 February 2020
Published: 13 March 2020

Citation:
Bai X, Putz AM, Wang Z, Fortin F, Harding JCS, Dyck MK, Dekkers JCM, Field CJ, Plastow GS and PigGen Canada (2020) Exploring Phenotypes for Disease Resilience in Pigs Using Complete Blood Count Data From a Natural Disease Challenge Model. Front. Genet. 11:216. doi: 10.3389/fgene.2020.00216

Disease resilience is a valuable trait to help manage infectious diseases in livestock. It is anticipated that improved disease resilience will sustainably increase production efficiency, as resilient animals maintain their performance in the face of infection. The objective of this study was to identify phenotypes related to disease resilience using complete blood count (CBC) data from a wean-to-finish natural disease challenge model, established to mimic the disease pressure caused by many common pathogens at the commercial level of pig production. In total, 2433 F1 crossbred (Landrace × Yorkshire) barrows that went through the natural disease challenge model were classified into four groups (resilient, average, susceptible, and dead) based on their divergent responses in terms of growth and individual treatment. Three sets of blood samples for CBC analysis were drawn at 2-weeks before, and at 2- and 6-weeks after the challenge: Blood 1, Blood 3, and Blood 4 respectively. CBC of Blood 1 taken from healthy pigs before challenge did not show differences between groups. However, resilient animals were found to be primed to initiate a faster adaptive immune response and recover earlier following infection, with greater increases of lymphocyte concentration from Blood 1 to Blood 3 and for hemoglobin concentration and hematocrit from Blood 3 to Blood 4, but a lower neutrophil concentration from Blood 3 to Blood 4 than in susceptible and dead animals ($FDR < 0.05$). The CBC traits in response to the challenge were found to be heritable and genetically correlated with growth and treatment, which may indicate the potential for developing CBC under disease or commercial conditions as a phenotype in commercial systems as part of developing predictions for disease resilience.

Keywords: natural disease challenge model, disease resilience, complete blood count, genetic parameters, pigs

INTRODUCTION

Disease resilience is defined as an animal's ability to maintain a relatively undepressed performance in the face of infection (Albers et al., 1987; Mulder and Rashidi, 2017). In pig breeding, disease resistance, which is defined as the ability to suppress establishment and subsequent development of infection, has been generally discussed in terms of making genetic improvement of herd health (Albers et al., 1987; Bishop and Stear, 2003; Guy et al., 2012). For example, the discovery of a polymorphism at bp 307 (G/A) in the fucosyltransferase gene (*FUT1*) associated with susceptibility/resistance to infection with F18 fimbriated *Escherichia coli* (ECF18) made it possible to select for ECF18 resistant pigs (Meijerink et al., 1997, 2000). Pigs that are homozygous for the resistant allele are resistant to ECF18 due to the non-adhesion of ECF18 in the small intestine (Meijerink et al., 1997; Bao et al., 2012). However, such complete resistance to a pathogen is not common, and selection for resistance to a specific pathogen may have unfavorable consequences for other production traits (Wilkie and Mallard, 1999; Guy et al., 2012). Currently, the challenge of infectious diseases in the pig industry is that a multitude of pathogens exists around the world (Zimmerman et al., 2012). Some pathogens, including porcine reproductive and respiratory syndrome virus (PRRSV), can also modulate the immune system to increase susceptibility to other pathogens while suppressing the immunologic memory of the host for the same pathogen (Zhu et al., 2010). Therefore, selective breeding for resilient animals that can maintain a relatively undepressed performance in a commercial system that typically harbors abundant infectious agents could be a pragmatic way to help maintain or even improve the productivity of the swine industry.

Direct selection for disease resilience is generally not feasible, because it is impractical to obtain heritable measures of resilience in the high health nucleus herds where the selection of elite breeding animals takes place (Wilkie and Mallard, 1999). Moreover, it is also challenging to appropriately characterize resilience because it is a complex trait composed of multiple biological functions, such as production, health, nutrient status, and other dynamic elements, including the efficiency of immune response and the rate of recovery from infection (Friggens et al., 2017). Many studies have explored the relationship of immune traits with performance. These include the use of white blood cell traits (**Figure 1**), which are reported to be moderately to highly heritable and genetically correlated with an animal's performance (Henryon et al., 2006; Clapperton et al., 2008, 2009; Flori et al., 2011; Mpetile et al., 2015). In addition to white blood cells, red blood cells and platelets have also been shown to play multiple roles in the immune system to help defend against pathogens, and these also have the potential to be genetically correlated with an animal's performance (Gershon, 1997; Liepke et al., 2003; Jiang et al., 2007; Rondina and Garraud, 2014; Hottz et al., 2018). Complete blood count (CBC) is a clinical measure used to evaluate the concentration and relative proportion of circulating blood cells and may be a practical measure of immune response and, therefore, could be a candidate phenotype for disease resilience. Moreover, CBC also evaluates the volume and concentration of red blood cells and hemoglobin to provide information about oxygen-carrying capacity and anemia, which are of concern during the disease process, with further impacts on animal performance (George-Gay and Parker, 2003).

Therefore, the objectives of this study were: (1) to assess CBC profiles of pigs that exhibited divergent performance in terms of growth and individual treatment in response to a polymicrobial infectious challenge; and (2) to estimate heritabilities of CBC traits and genetic correlations of CBC with growth and treatment rates following the disease challenge.

MATERIALS AND METHODS

This study was carried out in accordance with the Canadian Council on Animal Care guidelines (CCAC[1]). The protocol was approved by the Animal Protection Committee of the Centre de Recherche en Sciences Animales de Deschambault (15PO283) and the Animal Care and Use Committee at the University of Alberta (AUP00002227). The project was fully overseen by the Centre de Développement du Porc du Québec (CDPQ) and the herd veterinarian together with project veterinarians.

Natural Disease Challenge Model and Data Collection

A natural disease challenge model was established for wean-to-finish pigs at Deschambault, in the province of Québec, Canada. There were two main facilities in the model: (1) a healthy quarantine unit providing a 3-week nursery after weaning, and (2) a test station that consisted of a 4-week late nursery stage (40 to 68 days of age on average) and a grow-to-finish stage for approximately 16 weeks (69 to 181 days of age on average). The number of pigs per pen was approximately 4, 7, and 13 for the healthy quarantine unit, the test station late nursery, and the test station grow-to-finisher, respectively. Pigs were first exposed to the challenge in the test station in the late nursery, which aimed to represent and simulate a severe disease pressure caused by multiple pathogens found at the commercial level of production to maximize the expression of phenotypic and genetic differences associated with resilience. The test station barn was operated as a high health status facility prior to the introduction of the disease agents. Common disease-causing pathogens found in commercial farms were established by co-introducing commercial seeder pigs with known diseases with the first four batches of healthy pigs, including two viruses (three different strains of PRRSV and two strains of swine influenza A virus), five bacterial pathogens (*Mycoplasma hyopneumoniae*, *Haemophilus parasuis*, *Brachyspira hampsonii*, *Salmonella enterica* serovar *typhimurium*, and *Streptococcus suis*), and two parasites (*Cystoisospora suis* and *Ascaris suum*). For the data used in this study, every batch was confirmed to have been exposed to PRRSV in the test station based on randomly sampling of blood from a subset of individuals for RT-PCR 4 weeks post-challenge and enzyme-linked immunosorbent assay (ELISA) 6 weeks post-challenge. In addition to the introduced

[1]https://www.ccac.ca/en/certification/about-certification

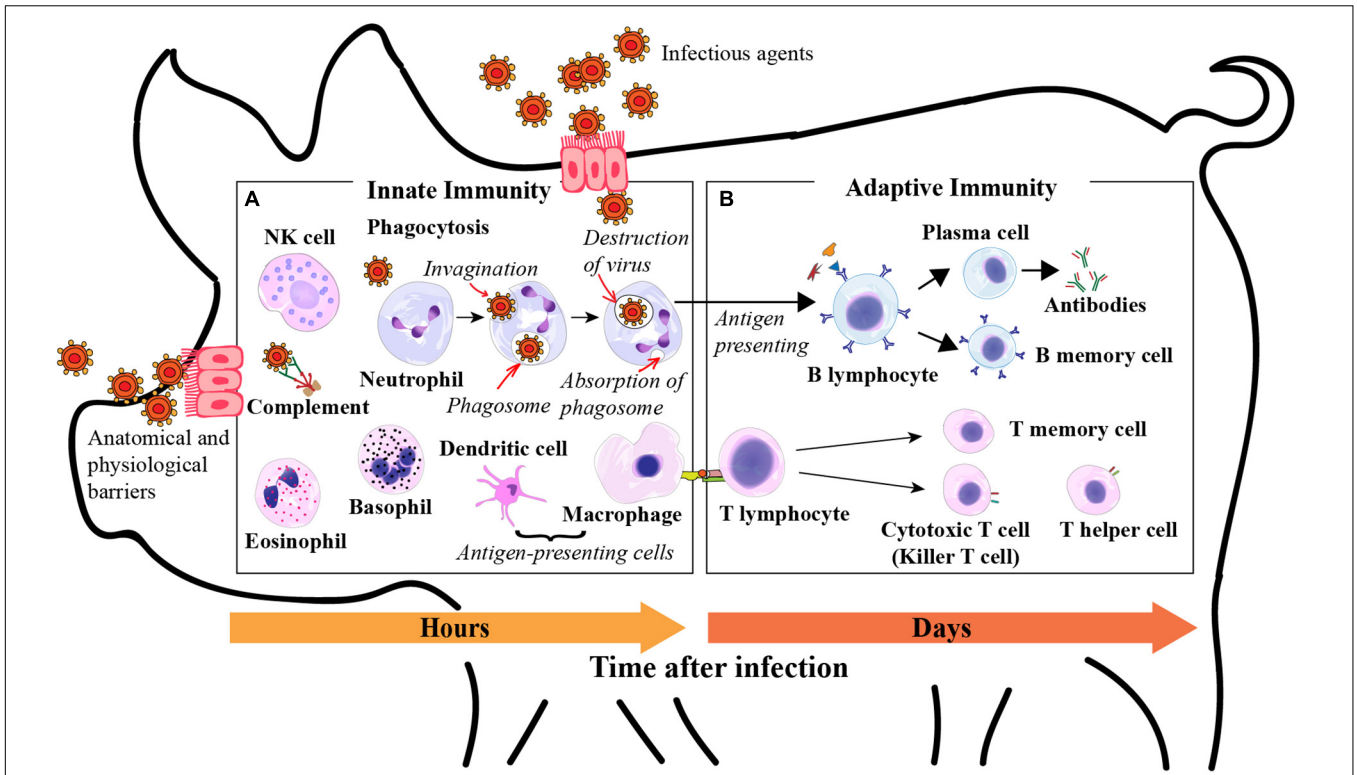


FIGURE 1 | Roles of white blood cells in innate immunity. **(A)** Phagocytosis is the process by which phagocytic cells recognize and ingest microbes for intracellular killing. Phagocytes include neutrophils, monocytes, dendritic cells, and eosinophils; Neutrophils, eosinophils, and basophils are granulocytes, the granules present in their cytoplasm contain biochemical mediators that serve inflammatory and immune functions; Eosinophils and basophils combat parasites through production of toxic proteins and histamine respectively; Dendritic cells produce cytokines that recruit white blood cells and initiate adaptive immune responses, and also present antigens to the adaptive immune system; Natural killer (NK) cells are a class of lymphocytes that recognize and kill infected cells to stop the spread of an infection; The complement system consists of a set of plasma proteins that act together to defend against extracellular pathogens. Roles of white blood cells in adaptive immunity. **(B)** B lymphocytes mediate humoral immunity by secreting antibodies into the circulation and mucosal fluid to neutralize and eliminate extracellular infectious agents; T lymphocytes characterize cell-mediated immunity and kill host cells that are harboring infectious agents in the cytoplasm. Derived from Janeway et al. (2001), Abbas et al. (2015), and Elsevier Health Sciences and Khan Academy (2019).

pathogens, other multiple disease-causing pathogens were also identified in the challenge facility, including porcine circovirus type-2 (PCV2), porcine rotavirus A, *Erysipelothrix rhusiopathiae*, *Staphylococcus hyicus*, and some undefined minor pathogens. However, these were not necessarily identified in all batches, as disease pressure varied by batch and on a seasonal basis. Not all pigs were exposed to all pathogens, as would be the case on a commercial farm. The level of mortality in a batch was carefully monitored and adjustments made to address the situation for reasons of animal ethics. For example, if the average mortality rate of each batch was more than 8% during the nursery stage, then group medication was applied through water and feed on a batch-level. If the challenge based on previous data was deemed to be too severe then direct nose-to-nose contact between batches in the challenge nursery was stopped.

Healthy F1 crossbred (Landrace × Yorkshire) castrated male weaned pigs were provided in rotation by seven genetic suppliers, all members of PigGen Canada. A total of 2743 pigs were introduced in 42 batches at 3-week intervals. Each batch consisted of approximately 65 or 75 pigs from one of the genetic suppliers. Every seven batches constituted a cycle. All weaned pigs arrived at an average age of 21 days old and were housed in a healthy quarantine unit, representing a 3-week nursery stage. For the first cycle, the quarantine unit and test station were in the same building connected by a hallway, but strict biosecurity protocols were practiced between them. Since the biosecurity practices were insufficient to stop the spread of pathogens from the test station to the clean quarantine, a separate quarantine unit located approximately 1 km south of the test station was set up for cycles 2 to 6 and kept free of disease by adhering to strict biosecurity protocols. Every 3 weeks, a new batch of approximately 40-day-old pigs was transferred from the quarantine nursery to the test station late nursery and exposed to the disease challenge by direct nose-to-nose contact with the preceding batch for 1 week (**Figure 2A**). The challenge was set up as a continuous flow system in order to maintain a steady disease challenge without repeatedly introducing commercial pigs and pathogens. During periods of very high challenge pressure, as identified by rates of morbidity and mortality, batches (n = 12) were not challenged by direct nose-to-nose contact. Pigs in these batches were allocated to nursery pens physically separated from the preceding batch (**Figure 2B**) to help maintain the

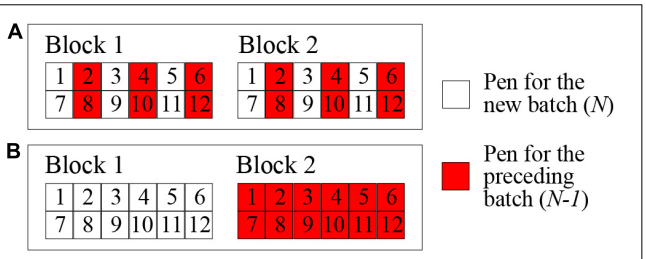


FIGURE 2 | Pen arrangements in the test station late nursery for the nose-to-nose direct challenge **(A)** and for the indirect contact during the period of excessively high pressure of challenge **(B).**

mortality rate below the target level established by the Animal Protection Committee.

The first blood sample for CBC (Blood 1) was collected on all pigs in the quarantine nursery at an average age of 26 days, 5 days post-arrival from their farm of origin. Two weeks after the first sampling, pigs were transferred to the test station and naturally exposed to multiple pathogens from 40 to approximately 181 days of age, when they reached the target slaughter weight of 130 kg. Blood 2 was collected immediately before transferring to the test station nursery at 40 days of age to measure antibody-mediated responses for a separate study of the immune response. No CBC was obtained on these Blood 2 samples. The second CBC blood sample (Blood 3) was collected on all pigs 2-weeks after transferring to the test station, at an average age of 54 days. The third CBC blood sample (Blood 4) was collected at approximately 82 days of age, 4 weeks after the collection of Blood 3, and 6 weeks after the transfer to the test station.

All blood samples were taken from the jugular vein, and the samples for CBC were collected into K2 ethylenediaminetetraacetic acid (EDTA) tubes (BD Vacutainer® blood collection tubes, New Jersey, United States). Samples were shipped overnight with ice packs and received by the University of Alberta for the CBC analysis using the ADVIA® 2120i Hematology System (Siemens Healthineers, Erlangen, Germany) within 24 to 48 h.

Body weights of each pig were measured every 3 weeks. Mortality and morbidity of each batch, as well as the reasons for death were also recorded on an individual pig basis. All medical treatments were recorded, including individual medication given on a case-by-case basis throughout the lifetime of the pig, as well as group treatments that were given on a batch-level. Of note, due to significant problems in managing the associated impact caused by PCV2 in cycle 1, Ingelvac CircoFLEX® PCV2 vaccination (Boehringer Ingelheim, Ingelheim am Rhein, Germany) was administered intramuscularly as per the label instructions to pigs before entering the test station from the second cycle onwards.

Genotyping

The genotyping of animals was performed at Delta Genomics (Edmonton, AB, Canada) using the 650K Affymetrix Axiom® Porcine Genotyping Array. In total, 658,692 single nucleotide polymorphisms (SNPs) were included on the chip. Raw Affymetrix SNP data for each cycle were processed separately at Delta Genomics with the Axiom Analysis Suite, using all defaults. Missing genotypes were imputed using FImpute (Sargolzaei et al., 2014). Sscrofa 11.1 was used as the reference genome. Quality control was performed using the preGSf90 software from the BLUPF90 family of programs to remove SNPs with a minor allele frequency lower than 0.01 and call rates lower than 0.90. Overall, genotypes for 2593 animals from all six cycles were used, with 475,839 SNPs remaining after processing and quality control.

Traits

The CBC traits used for this study were grouped into three categories: (1) six white blood cell traits, including total white blood cell concentration (WBC, $10^3/\mu L$), neutrophil concentration (NEU, $10^3/\mu L$), lymphocyte concentration (LYM, $10^3/\mu L$), monocyte concentration (MONO, $10^3/\mu L$), eosinophil concentration (EOS, $10^3/\mu L$), and basophil concentration (BASO, $10^3/\mu L$); (2) seven red blood cell traits, consisting of red blood cell concentration (RBC, $10^6/\mu L$), hemoglobin concentration (HGB, g/L), hematocrit (HCT, %), which measures the volume percentage of packed red blood cells in blood, mean corpuscular volume (MCV, fL), mean corpuscular hemoglobin (MCH, pg), mean corpuscular hemoglobin concentration (MCHC, g/L), and red blood cell distribution width (RDW, %), which evaluates the variability in size of red blood cells; and (3) two platelet traits, including platelet concentration (PLT, $10^3/\mu L$) and mean platelet volume (MPV, fL). In addition to these measurable traits for each blood sample, changes of CBC traits between blood samples collected at different time points were also calculated for each animal, which will be referred to as $\Delta 13$ for the change from Blood 1 to Blood 3 (Blood3 – Blood1), $\Delta 34$ for the change from Blood 3 to Blood 4 (Blood4 – Blood3), and $\Delta 14$ for the change from Blood 1 to Blood 4 (Blood4 – Blood1).

The growth rate of each animal in the grow-to-finish phase (GFGR) was estimated using linear regression of body weights collected from an average of 69 days old to the endpoint, i.e., when the pig died or when it reached the target slaughter weight at approximately 181 days old. The GFGR for animals that died before reaching the grow-to-finish stage was set to missing in the analyses. Treatment rate (TR) for each animal was the number of treatment events in the natural challenge barn, standardized by the number of days spent in the natural challenge barn (TR = number of treatment events/days × 100%). Group treatments given on the batch-level were not included because these would be accounted for in the model by fitting the fixed effect of batch. The TR for animals that died before receiving any treatment was set to missing.

Classification of Pigs Based on Resilience

Based on resilience indicated by phenotypes of GFGR and TR, pigs were classified into four groups as "resilient (RES)," "average (MID)," "susceptible (SUS)," and "dead (DEAD)" by batch. Within each batch, slaughtered pigs that had equal or higher GFGR than the third quartile (Q3, 75% quartile), and equal or lower TR than the first quartile (Q1, 25% quartile)

of all slaughtered pigs in the batch were classified as RES; slaughtered pigs that had equal or lower GFGR than the Q1 and equal or higher TR than the Q3 of all slaughtered pigs in the batch were regarded as SUS; the rest of the slaughtered animals, which had moderate TR and GFGR, were classified as MID (**Figure 3**). The influence caused by the environmental changes and differences among batches were controlled and minimized by classifying animals within each batch. Among 2593 genotyped pigs, mortalities (n = 160) caused by hernia, fighting, fracture, sampling, or sudden death due to unclear reasons were excluded from the analysis. Of the remaining 2433 pigs, 505 (21%) pigs that died as a result of infectious disease were classified as DEAD. For the 1928 pigs that were slaughtered at market body weight in the six cycles, 213 (9%) pigs were in the RES group, 1505 (61%) pigs were in the MID group, and 210 (9%) pigs were in the SUS group.

Statistical Analyses

Removal of Outliers

Due to the relative complexity of the sample handling, shipping conditions, and laboratory analysis, outliers for the measures of CBC traits could be the result of damaged samples with hematological issues including hemolysis and clotting, or mechanical problems of the Hematology System used to measure CBC from blood samples. Such outliers were detected and removed using the Adjusted Boxplot in R (R Core Team, 2017, Package 'robustbase'). It is a robust measure of skewness in the determination of thresholds for the removal of outliers and can avoid erroneously declaring points as outliers in a skewed distribution (Hubert and Vandervieren, 2008). The skewness of a CBC trait was measured using Medcouple (Brys et al., 2004). Thresholds for removing outliers for CBC measures were determined by several parameters, including Medcouple (MC), first quartile (Q1), third quartile (Q3), and interquartile range between Q1 and Q3 (IQR). The lower and upper bounds for a right-skewed distribution (MC > 0) were Q1 − $1.5^{(-4MC)}$ × IQR and Q3 + $1.5^{(3MC)}$ × IQR; for a left-skewed distribution (MC < 0), the lower and upper bounds were Q1 − $1.5^{(-3MC)}$ × IQR and Q3 + $1.5^{(4MC)}$ × IQR; and for a symmetric distribution (MC = 0), the outliers were removed using Tukey's boxplot (lower bound Q1 − 1.5 × IQR, upper bound: Q3 + 1.5 × IQR) (Seo, 2006; Hubert and Vandervieren, 2008). All CBC measures outside of the upper and lower bounds were removed as outliers.

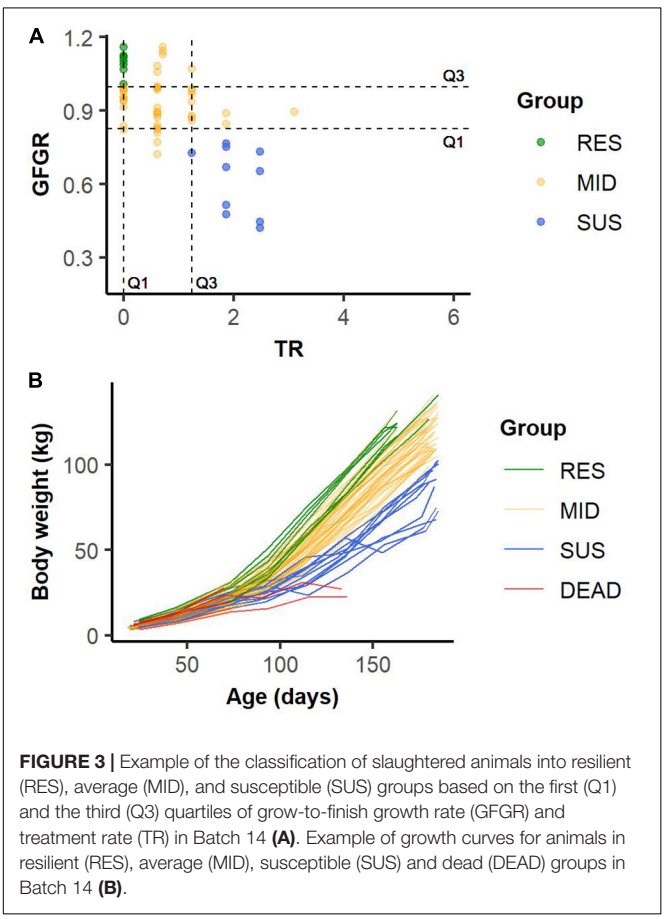


FIGURE 3 | Example of the classification of slaughtered animals into resilient (RES), average (MID), and susceptible (SUS) groups based on the first (Q1) and the third (Q3) quartiles of grow-to-finish growth rate (GFGR) and treatment rate (TR) in Batch 14 **(A)**. Example of growth curves for animals in resilient (RES), average (MID), susceptible (SUS) and dead (DEAD) groups in Batch 14 **(B)**.

Models

The likelihood ratio test in ASReml 4.1 was used to determine the significance of different environmental random terms for litter and pen effects by comparing the full model, including batch, bleed age, litter, pen, and genetic effects to reduced models without each litter or pen effect (Hagger, 1998; Gilmour et al., 2015).

The CBC phenotype data were analyzed using linear mixed effects models to estimate the least-squares means for CBC traits by group (RES, MID, SUS, and DEAD), and the Tukey–Kramer test was applied for pairwise comparisons of the difference between groups in R (R Core Team, 2017, packages 'lme4' and 'lsmeans'). White blood cell traits were $\log_{10}$-transformed because of residual heterogeneity. In the mixed model, batch was fitted as a fixed effect to control and minimize the influence of the environmental changes among batches, group was also fitted as a fixed effect, and bleeding age was fitted as a covariate. Of note, for the changes of CBC between time points, bleeding age of Blood 1 was fitted for Δ13 and Δ14, and Blood 3 bleeding age was fitted for Δ34 since the 4-week interval between each blood sampling was the same for all animals. Random terms, including the litter and pen effects were fitted if significant ($p < 0.05$).

Heritabilities and genetic correlations of CBC traits with resilience traits were estimated in ASReml4.1 using pairwise bivariate models, with batch, bleed age, litter, and pen effects as described above for estimating the difference between resilience groups. Analyses for GFGR and TR included the fixed effect of batch for both traits, and random effects of litter and pen if significant ($p < 0.05$). Animal genetic effects were fitted using the genomic relationship matrix for 2593 animals, rather than the pedigree-based relationship matrix because the complete pedigree was unavailable due to the use of pooled semen in some batches. The genomic relationship matrix was constructed using $ZZ'/2\sum p_i(1-p_i)$, where Z contains centered genotypes codes and p_i is the minor allele frequency for locus i (VanRaden, 2008). The average estimate of corresponding pairwise bivariate analyses was reported as the heritability for each trait. In the bivariate models, batch was fitted as a fixed effect for both traits.

The likelihood ratio test was applied to test the significance of estimates for heritabilities and genetic correlations in ASReml 4.1, where the log-likelihood of full models were compared to restricted models that constrained the genetic variance and the genetic covariance to zero, respectively (Gilmour et al., 2015).

The model used in ASReml 4.1 can be written as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} Z_3 & 0 \\ 0 & Z_4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

where $\mathbf{y}_1$ and $\mathbf{y}_2$ denote vectors of observations for traits 1 and 2; $\mathbf{X}_1$ and $\mathbf{X}_2$ are incidence matrices relating fixed effects to $\mathbf{y}_1$ and $\mathbf{y}_2$, $\mathbf{b}_1$ and $\mathbf{b}_2$ are vectors of fixed effects for traits 1 and 2; $\mathbf{Z}_1$ and $\mathbf{Z}_2$ represent design matrices that associate observations of traits 1 and 2 to vectors of animal genetic effects $\mathbf{g}_1$ and $\mathbf{g}_2$; $\mathbf{c}_1$ and $\mathbf{c}_2$ are vectors of random effects, including litter and pen effects when they were significant ($p < 0.05$); $\mathbf{Z}_3$ and $\mathbf{Z}_4$ are incidence matrices relating $\mathbf{y}_1$ and $\mathbf{y}_2$ to random effects $\mathbf{c}_1$ and $\mathbf{c}_2$; $\mathbf{e}_1$ and $\mathbf{e}_2$ are vectors of unknown and random residuals for traits 1 and 2 (Miar et al., 2014a,b; Gilmour et al., 2015).

When random effects **c** and residuals errors **e** are uncorrelated, and identically distributed following a normal distribution, the (co-)variances of random effects are assumed to be

$$Var\begin{bmatrix} g_1 \\ g_2 \\ c_1 \\ c_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} G\sigma^2_{g_1} & G\sigma_{g_1g_2} & 0 & 0 & 0 & 0 \\ G\sigma_{g_1g_2} & G\sigma^2_{g_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & I\sigma^2_{c_1} & I\sigma_{c_1c_2} & 0 & 0 \\ 0 & 0 & I\sigma_{c_1c_2} & I\sigma^2_{c_2} & 0 & 0 \\ 0 & 0 & 0 & 0 & I\sigma^2_{e_1} & I\sigma_{e_1e_2} \\ 0 & 0 & 0 & 0 & I\sigma_{e_1e_2} & I\sigma^2_{e_2} \end{bmatrix}$$

where $\boldsymbol{G}$ is the genomic relationship matrix, $\boldsymbol{I}$ is the identity matrix, σ^2_g is the additive genetic variance, σ^2_c is the random effect variance, and σ^2_e is the residual variance. σ_{g1g2}, $\sigma_{c_1c_2}$, and $\sigma_{e_1e_2}$ are covariances between two traits due to the additive genetic effects, common random effects, and residual effects, respectively. Heritability (h^2) of a trait was estimated using variance components obtained from the bivariate analyses, and the average estimates of corresponding pairwise bivariate analyses were reported as the heritabilities:

$$h^2 = \sigma^2_g/(\sigma^2_g + \sigma^2_c + \sigma^2_e)$$

and the genetic correlation (r_g) between two traits was estimated as:

$$r_g = \sigma_{g1g2}/\sigma_{g1}\sigma_{g2}$$

RESULTS

Descriptive Statistics for CBC Traits

Table 1 summarizes the descriptive statistics for the CBC data of 2593 genotyped animals after removing outliers. Most traits were recorded on all animals in Blood 1, but some samples for Blood 3 and Blood 4 were unavailable for animals that died prior to the sampling. Relevant random effects fitted in the models for CBC traits are presented in **Table 2**. The random effect of litter was fitted for GFGR, and pen effects in the test station late nursery and the grow-to-finish stage were fitted for TR.

Group Differences in CBC Traits

White Blood Cell Traits

Results comparing the least-squares means of white blood cell traits in groups with different responses to the natural disease challenge are shown in **Table 3**. In Blood 1, no significant difference was found between groups for any of the white blood cell traits. However, in Blood 3, the RES group had a significantly higher LYM, and the LYM for the MID group was also significantly higher than for the DEAD group (*FDR* = 0.0003). In Blood 4, the RES and MID groups had significantly lower NEU levels than both the SUS and DEAD groups (*FDR* = 0.0002). For the count of LYM in Blood 4, the DEAD group was significantly lower than both the RES and MID groups (*FDR* = 0.0012).

Results comparing the least-squares means of changes in white blood cell traits between groups are summarized in **Table 4**. All white blood cell traits increased from Blood 1 to Blood 3 shown as positive Δ13. The increase of LYM was significantly higher for the RES group than for the other groups (*FDR* = 0.0002), but no significant difference was found among the MID, SUS and DEAD groups. Changes of white blood cell traits from Blood 3 to Blood 4 were not as dramatic as those from Blood 1 to Blood 3, except for LYM, which had a higher increase from Blood 3 to Blood 4 for all groups. The WBC, LYM, and MONO levels increased continuously for all groups based on positive Δ13 and Δ34, but EOS and BASO decreased from Blood 3 to Blood 4 based on negative Δ34. NEU showed a tendency to decrease in the RES and MID groups, which was opposite to the positive NEU in the SUS and DEAD groups for Δ34 (*FDR* < 0.0024). Additionally, a significant difference in NEU among groups was also identified for Δ14, which represents the overall change of NEU from Blood 1 to Blood 4. Δ14 for NEU were positive for all groups, but the SUS and DEAD groups had significantly higher increases in NEU than the RES and MID groups (*FDR* = 0.0002). Compared with Blood 1, which was collected in the quarantine unit, the other white blood cell traits, including WBC, LYM, MONO, EOS, and BASO, also increased significantly in Blood 4, although no significant differences based on Δ14 were found between groups.

Red Blood Cell and Platelet Traits

Results of comparing red blood cell and platelet traits in the RES, SUS, MID, and DEAD groups are summarized in **Table 5**. No significant differences were identified between groups for either red blood cell or platelet traits in Blood 1. However, for Blood 3, RDW and MPV were significantly higher in the DEAD group than in the RES and MID groups (*FDR* < 0.002). For Blood 4, several red blood cell traits showed significant differences between groups. Notably, HGB, HCT, and MCH were found to be significantly lower in the SUS and DEAD groups than in the RES and MID groups (*FDR* < 0.0005). Moreover, RBC was significantly higher in

TABLE 1 | Descriptive statistics for complete blood count (CBC) traits in Blood 1, Blood 3, and Blood 4 after removing outliers, including the number of animals per trait (n), mean, standard deviation (SD), minimum (Min), and maximum (Max) values.

Traits[1]	Blood 1					Blood 3					Blood 4					Reference intervals[2]	
	n	Mean	*SD*	Min	Max	*n*	Mean	*SD*	Min	Max	*n*	Mean	*SD*	Min	Max	0 to 42 days	42 day to 2 years
WBC, $10^3/\mu L$	2222	11.47	3.67	5.64	28.21	2284	19.11	5.09	8.28	36.53	1802	21.92	6.15	9.23	43.01	9.62–25.20	11.35–28.90
NEU, $10^3/\mu L$	2375	4.76	2.38	1.33	14.71	2322	10.34	4.01	1.64	23.61	1808	9.95	4.65	2.48	28.37	2.35–11.90	2.00–10.40
LYM, $10^3/\mu L$	2425	5.61	1.85	2.39	12.65	2326	6.47	2.21	2.06	13.57	1840	9.82	3.11	3.67	21.09	4.02–12.50	5.30–17.90
MONO, $10^3/\mu L$	2440	0.32	0.21	0.04	1.23	2364	0.82	0.59	0.05	3.70	1890	1.01	0.74	0	4.06	0.05–2.30	0–3.70
EOS, $10^3/\mu L$	2474	0.47	0.40	0	2.61	2213	0.71	0.75	0.12	4.35	1807	0.60	0.48	0.12	3.01	0–0.50	0–1.30
BASO, $10^3/\mu L$	2096	0.13	0.23	0.02	1.69	2264	0.84	1.36	0.06	8.51	1798	0.33	0.32	0.05	2.09	NA[3]	NA
RBC, $10^6/\mu L$	2373	6.15	0.60	4.27	7.52	2242	5.79	0.67	3.82	7.55	1767	6.28	0.57	4.51	7.67	4.87–7.88	5.88–8.19
HGB, g/L	2434	116.45	13.46	73	148	2239	100.59	10.35	68	126	1730	104.95	9.71	69	125	80.8–119	112–147
HCT,%	2310	37.12	4.10	24	44	2228	32.81	3.63	22.10	41.80	1723	35.25	3.14	28	43	28.22–39.80	32.30–42.60
MCV, fL	2444	61.25	5.45	44.5	73.40	2339	57.02	3.59	49.60	69.50	1879	55.78	3.42	46.80	65.40	43.40–64.50	47.50–59.20
MCH, pg	2318	18.73	2.03	12.50	23.60	2153	17.52	1.26	14.70	21.80	1719	16.72	1.19	13.40	20.10	12.40–19.30	16.30–20.60
MCHC, g/L	2245	305.88	12.06	274	340	2150	307.40	15.77	268	366	1708	300.22	13.31	264	345	273–314	333–358
RDW,%	2473	21.97	4.02	15.80	39.90	2321	18.45	1.61	15.90	25.10	1873	18.61	1.40	15.60	23.10	NA	NA
PLT, $10^3/\mu L$	2457	285.13	177.18	0	949	2351	365.46	182.69	35	1062	1872	337.08	150.87	47	784	374.3–1080.8	118.9–522.9
MPV, fL	2435	14.63	3.35	8.30	26.20	2180	15.33	3.72	10.10	30.80	1849	13.57	2.01	9.30	20.50	NA	NA

[1]WBC: total white blood cell concentration; NEU: neutrophil concentration; LYM: lymphocyte concentration; MONO: monocyte concentration; EOS: eosinophil concentration; BASO: basophil concentration; RBC: red blood cell concentration; HGB: hemoglobin concentration; HCT: hematocrit; MCV: mean corpuscular volume; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; RDW: red blood cell distribution width; PLT: platelet concentration; MPV: mean platelet volume. [2]Suggested reference intervals for CBC traits of 0 to 42 days-old pigs and 42 days-old to 2 years-old pigs (Iowa State University's Clinical Pathology Laboratory, 2011). [3]Not applicable.

TABLE 2 | Random effects included in the models for the analyses of complete blood count (CBC) traits.

Traits[1]	Blood 1		Blood 3			Blood 4				Δ13[2]			Δ34[3]				Δ14[4]			
	Litter	Pen1[5]	Litter	Pen1	Pen2[6]	Litter	Pen1	Pen2	Pen3[7]	Litter	Pen1	Pen2	Litter	Pen1	Pen2	Pen3	Litter	Pen1	Pen2	Pen3
WBC	√[8]	NS[9]	√	NS	√	NS	NS	NS	√	√	NS	√	NS	NS	NS	√	NS	NS	NS	√
NEU	√	NS	√	NS	√	NS	NS	NS	√	√	NS	√	NS	NS	NS	√	NS	NS	NS	√
LYM	√	NS	√	NS	√	NS	NS	NS	√	√	NS	NS	NS	NS	NS	√	NS	NS	NS	√
MONO	√	NS	NS	NS	√	√	√	NS	NS	NS	NS	NS	√	NS	NS	√	√	NS	NS	NS
EOS	√	√	NS	√	NS	NS	NS	NS	NS	NS	NS	NS	NS	√	NS	NS	√	NS	NS	√
BASO	NS	√	NS	√	√	√	√	NS	√	NS	√	√	NS	√	√	NS	NS	NS	NS	√
RBC	√	√	√	√	√	NS	NS	NS	√	√	√	NS	NS	NS	NS	√	√	√	NS	NS
HGB	√	√	√	√	√	NS	NS	NS	√	√	√	√	NS	NS	NS	√	√	√	NS	√
HCT	√	√	√	√	√	NS	NS	NS	√	√	√	NS	NS	NS	NS	√	√	√	NS	√
MCV	√	√	√	NS	NS	√	NS	NS	√	√	√	NS	√	NS	NS	√	√	NS	NS	NS
MCH	√	√	√	NS	√	√	NS	NS	NS	√	√	NS	√	NS	NS	√	√	NS	NS	√
MCHC	√	√	√	√	NS	NS	NS	NS	√	√	√	NS	NS	NS	NS	√	√	√	NS	√
RDW	√	√	√	√	√	√	√	NS	√	√	√	NS	√	NS	NS	√	√	NS	NS	NS
PLT	√	√	NS	NS	√	√	√	NS	NS	√	√	NS	NS	NS	NS	√	√	√	NS	NS
MPV	√	√	NS	NS	√	NS	NS	NS	√	NS	√	NS	NS	√	NS	NS	NS	√	NS	NS

[1]WBC: total white blood cell concentration; NEU: neutrophil concentration; LYM: lymphocyte concentration; MONO: monocyte concentration; EOS: eosinophil concentration; BASO: basophil concentration; RBC: red blood cell concentration; HGB: hemoglobin concentration; HCT: hematocrit; MCV: mean corpuscular volume; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; RDW: red blood cell distribution width; PLT: platelet concentration; MPV: mean platelet volume. [2]The change of CBC traits from Blood 1 to Blood 3; [3]The change of CBC traits from Blood 3 to Blood 4; [4]The change of CBC traits from Blood 1 to Blood 4. [5]The pen arrangement in the healthy quarantine unit; [6]The pen arrangement in the test station late nursery; [7]The pen arrangement in the test station grow-to-finish stage. [8]Significant random effect that was included in the model. [9]Not significant.

TABLE 3 | Least-squares means ± standard errors for white blood cell traits[1] in Blood 1, 3, and 4 of animals from the resilient (RES), average (MID), susceptible (SUS), and dead (DEAD) groups.

Blood 1, $10^3/\mu L$	**RES**	**MID**	**SUS**	**DEAD**	**FDR-group[2]**
log_{10} (WBC)	1.03 ± 0.01^{a3}	1.04 ± 0.00^{a}	1.04 ± 0.01^{a}	1.03 ± 0.01^{a}	0.55
log_{10} (NEU)	0.62 ± 0.01^{a}	0.64 ± 0.01^{a}	0.63 ± 0.01^{a}	0.62 ± 0.01^{a}	0.55
log_{10} (LYM)	0.71 ± 0.01^{a}	0.73 ± 0.00^{a}	0.73 ± 0.01^{a}	0.72 ± 0.01^{a}	0.29
log_{10} (MONO)	-0.61 ± 0.02^{a}	-0.59 ± 0.01^{a}	-0.58 ± 0.02^{a}	-0.58 ± 0.01^{a}	0.42
log_{10} (EOS)	-0.47 ± 0.02^{a}	-0.49 ± 0.01^{a}	-0.50 ± 0.02^{a}	-0.50 ± 0.01^{a}	0.84
log_{10} (BASO)	-1.15 ± 0.02^{a}	-1.15 ± 0.01^{a}	-1.14 ± 0.02^{a}	-1.16 ± 0.01^{a}	0.88
Blood 3, $10^3/\mu L$	**RES**	**MID**	**SUS**	**DEAD**	**FDR-group**
log_{10} (WBC)	1.27 ± 0.01^{a}	1.27 ± 0.00^{a}	1.25 ± 0.01^{a}	1.26 ± 0.01^{a}	0.18
log_{10} (NEU)	0.97 ± 0.01^{a}	0.98 ± 0.00^{a}	0.97 ± 0.01^{a}	0.98 ± 0.01^{a}	0.56
log_{10} (LYM)	**0.82 ± 0.01^{c4}**	**0.79 ± 0.00^{b}**	**0.77 ± 0.01^{ab}**	**0.75 ± 0.01^{a}**	**<0.0001**
log_{10} (MONO)	-0.18 ± 0.02^{a}	-0.21 ± 0.01^{a}	-0.23 ± 0.02^{a}	-0.22 ± 0.01^{a}	0.27
log_{10} (EOS)	-0.32 ± 0.02^{a}	-0.30 ± 0.01^{a}	-0.33 ± 0.02^{a}	-0.33 ± 0.01^{a}	0.15
log_{10} (BASO)	-0.51 ± 0.02^{a}	-0.51 ± 0.01^{a}	-0.55 ± 0.02^{a}	-0.49 ± 0.01^{a}	0.15
Blood 4, $10^3/\mu L$	**RES**	**MID**	**SUS**	**DEAD**	**FDR-group**
log_{10} (WBC)	1.31 ± 0.01^{a}	1.32 ± 0.00^{a}	1.34 ± 0.01^{a}	1.34 ± 0.01^{a}	0.21
log_{10} (NEU)	**0.93 ± 0.01^{a}**	**0.95 ± 0.01^{a}**	**1.02 ± 0.01^{b}**	**1.03 ± 0.02^{b}**	**<0.0001**
log_{10} (LYM)	**0.98 ± 0.01^{b}**	**0.98 ± 0.00^{b}**	**0.95 ± 0.01^{ab}**	**0.92 ± 0.01^{a}**	**0.0009**
log_{10} (MONO)	-0.16 ± 0.02^{a}	-0.15 ± 0.01^{a}	-0.14 ± 0.02^{a}	-0.18 ± 0.03^{a}	0.67
log_{10} (EOS)	-0.33 ± 0.01^{a}	-0.33 ± 0.01^{a}	-0.29 ± 0.02^{a}	-0.30 ± 0.02^{a}	0.46
log_{10} (BASO)	-0.61 ± 0.02^{a}	-0.59 ± 0.01^{a}	-0.57 ± 0.02^{a}	-0.57 ± 0.02^{a}	0.40

[1]WBC: total white blood cell concentration; NEU: neutrophil concentration; LYM: lymphocyte concentration; MONO: monocyte concentration; EOS: eosinophil concentration; BASO: basophil concentration. [2]FDR-group: adjusted p-values for the significant level of group effect using the Benjamini and Hochberg correction (FDR) in R to control false positives from multiple comparisons (R Core Team, 2017 Package 'stats'). [3]Values in a column suffixed with different letters are significantly different from each other at FDR < 0.05. [4]Significant differences among RES, MID, SUS and DEAD groups are highlighted in bold (FDR < 0.05).

the RES and MID groups than in the SUS and DEAD groups (*FDR* = 0.0036), and MCV was significantly lower in the DEAD group than in the others. In contrast, RDW and MPV were found to be significantly higher in the DEAD group than in the RES in Blood 4.

Table 6 summarizes the results of comparing the least-squares means of changes in red blood cell and platelet traits between groups. In contrast to the increase in white blood cell traits, all red blood cell traits decreased from Blood 1 to Blood 3, except for MCHC, which increased significantly in the DEAD group. Apart from MCHC, the drop for the other red blood cell traits from Blood 1 to Blood 3 did not show a tendency of being different between groups. The MPV in the SUS group was the only platelet trait that did not show a significantly positive Δ13 due to a relatively large standard error. Changes of platelet traits based on Δ13 did not show significant differences between groups. In contrast to the decreasing trend of red blood cell traits from Blood 1 to Blood 3, RBC and HCT increased significantly from Blood 3 to Blood 4 for all groups based on positive Δ34. Moreover, HGB also increased for both the RES and MID groups from Blood 3 to Blood 4, and Δ34 for HGB of these groups was significantly different from Δ34 for the SUS and DEAD groups (*FDR* = 0.0002), which were not found to be significantly different from zero. The MCV decreased continuously based on negative Δ34, and the DEAD group showed a more dramatic drop in MCV than the RES and MID groups (*FDR* = 0.0003). MCH and MCHC also kept decreasing based on negative Δ34, and the decrease of MCH for the DEAD group was significantly higher than for the RES and MID groups. Platelet traits also reduced from Blood 3 to Blood 4 for all groups, except for PLT in the SUS group, which did not show a significantly negative Δ34 due to a relatively large standard error.

Although several traits increased slightly from Blood 3 to Blood 4, for the overall changes from Blood 1 to Blood 4, all traits decreased significantly based on negative Δ14, except for RBC and PLT. Comparing Blood 4 to Blood 1, RBC increased slightly for the RES and MID groups, but it showed a tendency to return to the same level as in Blood 1 for the SUS and DEAD groups. PLT increased significantly from Blood 1 to Blood 4 for the RES, MID, and SUS groups, with no significant change identified for the DEAD group. MCHC was the only trait that showed a significant difference between groups for Δ14, which was lower in the SUS group than in the RES group (*FDR* = 0.04).

Estimates of Heritability

The GFGR was estimated to be moderately heritable (0.15 ± 0.04), but the heritability estimate of TR was low (0.04 ± 0.01). Heritability estimates for CBC traits with standard errors are in **Table 7**. Most CBC traits were moderately heritable, with estimates ranging from 0.11 ± 0.03 to 0.27 ± 0.04. A few red blood cell traits showed moderate to high heritability estimates, ranging from 0.30 ± 0.04 to 0.53 ± 0.05, including RBC, MCV, and MCH in Blood 3 and 4. Estimates of heritability were low for some CBC traits,

TABLE 4 | Least-squares means ± standard errors for changes of white blood cell traits[1] between Blood 1, 3, and 4 of animals in the resilient (RES), average (MID), susceptible (SUS), and dead (DEAD) groups.

Δ13[2], 10[3]/μL	**RES**	**MID**	**SUS**	**DEAD**	**FDR-group[5]**
WBC	8.39 ± 0.39[a6]	7.68 ± 0.17[a]	7.10 ± 0.40[a]	7.74 ± 0.29[a]	0.24
NEU	5.52 ± 0.27[a]	5.52 ± 0.11[a]	5.35 ± 0.29[a]	5.93 ± 0.20[a]	0.38
LYM	**1.69 ± 0.17[b7]**	**0.85 ± 0.07[a]**	**0.67 ± 0.17[a]**	**0.51 ± 0.12[a]**	**<0.0001**
MONO	0.56 ± 0.03[a]	0.49 ± 0.01[a]	0.42 ± 0.03[a]	0.47 ± 0.02[a]	0.08
EOS	0.23 ± 0.04[a]	0.26 ± 0.01[a]	0.18 ± 0.04[a]	0.24 ± 0.03[a]	0.38
BASO	0.85 ± 0.08[a]	0.66 ± 0.04[a]	0.63 ± 0.07[a]	0.79 ± 0.06[a]	0.08
Δ34[3], 10[3]/μL	**RES**	**MID**	**SUS**	**DEAD**	**FDR-group**
WBC	2.13 ± 0.51[a]	2.89 ± 0.23[a]	4.06 ± 0.54[a]	3.91 ± 0.68[a]	0.08
NEU	**−0.60 ± 0.41[a]**	**−0.33 ± 0.17[a]**	**1.36 ± 0.40[b]**	**1.77 ± 0.51[b]**	**<0.0001**
LYM	3.08 ± 0.25[a]	3.41 ± 0.11[a]	3.24 ± 0.26[a]	2.78 ± 0.32[a]	0.32
MONO	0.16 ± 0.05[a]	0.20 ± 0.02[a]	0.24 ± 0.05[a]	0.13 ± 0.06[a]	0.65
EOS	−0.08 ± 0.04[a]	−0.13 ± 0.02[a]	−0.04 ± 0.04[a]	−0.03 ± 0.06[a]	0.16
BASO	−0.54 ± 0.06[a]	−0.45 ± 0.03[a]	−0.38 ± 0.06[a]	−0.40 ± 0.08[a]	0.42
Δ14[4], 10[3]/μL	**RES**	**MID**	**SUS**	**DEAD**	**FDR-group**
WBC	10.39 ± 0.48[a]	10.75 ± 0.19[a]	11.60 ± 0.51[a]	11.73 ± 0.63[a]	0.27
NEU	**4.75 ± 0.35[a]**	**5.11 ± 0.14[a]**	**6.83 ± 0.36[b]**	**7.32 ± 0.45[b]**	**<0.0001**
LYM	4.63 ± 0.23[a]	4.37 ± 0.09[a]	4.06 ± 0.24[a]	3.56 ± 0.31[a]	0.08
MONO	0.72 ± 0.04[a]	0.71 ± 0.02[a]	0.66 ± 0.04[a]	0.61 ± 0.05[a]	0.37
EOS	0.12 ± 0.03[a]	0.13 ± 0.01[a]	0.22 ± 0.03[a]	0.19 ± 0.04[a]	0.12
BASO	0.15 ± 0.03[a]	0.20 ± 0.01[a]	0.24 ± 0.02[a]	0.23 ± 0.03[a]	0.12

[1]WBC: total white blood cell concentration; NEU: neutrophil concentration; LYM: lymphocyte concentration; MONO: monocyte concentration; EOS: eosinophil concentration; BASO: basophil concentration. [2]The change of complete blood count (CBC) traits from Blood 1 to Blood 3; [3]The change of CBC traits from Blood 3 to Blood 4; [4]The change of CBC traits from Blood 1 to Blood 4. [5]FDR-group: adjusted p-values for the significant level of group effect using the Benjamini and Hochberg correction (FDR) in R to control false positives from multiple comparisons (R Core Team, 2017 Package 'stats'). [6]Values in a column suffixed with different letters are significantly different from each other at FDR < 0.05. [7]Significant differences among RES, MID, SUS and DEAD groups are highlighted in bold (FDR < 0.05).

including BASO, HGB, and HCT in Blood 1, PLT in Blood 3 and Blood 4, RDW in Blood 4, and also for the changes of many CBC traits based on Δ13, Δ34, and Δ14. Genetic variances of several traits, especially MONO, and some changes of EOS, BASO, HCT, PLT, and MPV were not found to be significantly different from zero based on likelihood ratio tests, which compared full models to restricted models that constrained the genetic variance to zero in ASReml 4.1 ($p > 0.05$) (Gilmour et al., 2015).

Estimates of Genetic Correlations

GFGR and TR were estimated to be negatively correlated, with a genetic correlation of −0.50 ± 0.16. Estimates of genetic correlations for CBC traits that showed significant differences among groups (RES, MID, SUS, and DEAD) and the resilience traits of GFGR and TR are summarized in **Table 8**. LYM in Blood 3 and its change based on Δ13, which had the highest levels in the RES group, showed significantly negative genetic correlations with TR of −0.38 ± 0.18 and −0.46 ± 0.24, respectively. HCT based on Δ34, which was significantly higher in the RES and MID groups, showed a high negative genetic correlation with TR (−0.82 ± 0.47). NEU in Blood 4, RDW in Blood 4, and the change of NEU based on Δ14, which all had higher counts in the SUS and DEAD groups, showed significantly positive genetic correlations with TR. Genetic correlations between these CBC traits and GFGR showed a tendency of being opposite to the positive genetic correlations with TR but had relatively large standard errors. NEU based on Δ34, which was significantly positive in the SUS and DEAD groups but not significantly different from zero in the RES and MID groups, was estimated to have a negative genetic correlation with GFGR (−0.45 ± 0.21). TR showed a tendency to have a positive genetic correlation with the NEU based on Δ34 but had a large standard error (0.44 ± 0.26). For CBC traits from Blood 1, RDW was the only trait that showed a significantly positive genetic correlation with TR (0.41 ± 0.20), while none of the other CBC traits from Blood 1 showed significant correlations with TR or GFGR due to having low estimates and relatively high standard errors (**Supplementary Table 1**). Estimates of genetic correlations for CBC traits within Blood 1, Blood 3, and Blood 4 are summarized in **Supplementary Table 2**, while estimates of genetic correlations for each CBC trait between Blood 1, Blood 3, and Blood 4 are shown in **Supplementary Table 3**. Genetic correlations between Δ13, Δ34, and Δ14 were also estimated for each CBC trait and are summarized in **Supplementary Table 4**.

DISCUSSION

CBC Traits and Disease Resilience

Hematopoiesis, including the establishment and maintenance of all circulating cellular blood components, relies on the proliferation and differentiation of hematopoietic stem cells

TABLE 5 | Least-squares means ± standard errors for red blood cell and platelet traits[1] in Blood 1, 3, and 4 of animals in in the resilient (RES), average (MID), susceptible (SUS), and dead (DEAD) groups.

Blood 1	**RES**	**MID**	**SUS**	**DEAD**	**FDR-group[2]**
RBC, $10^6/\mu L$	6.18 ± 0.03[a3]	6.18 ± 0.01[a]	6.13 ± 0.03[a]	6.15 ± 0.02[a]	0.58
HGB, g/L	117.15 ± 0.71[a]	116.84 ± 0.31[a]	116.89 ± 0.72[a]	116.50 ± 0.50[a]	0.92
HCT,%	37.45 ± 0.23[a]	37.50 ± 0.10[a]	37.63 ± 0.23[a]	37.22 ± 0.16[a]	0.51
MCV, fL	61.31 ± 0.28[a]	61.36 ± 0.13[a]	62.07 ± 0.28[a]	61.39 ± 0.20[a]	0.20
MCH, pg	18.67 ± 0.11[a]	18.67 ± 0.05[a]	18.83 ± 0.11[a]	18.69 ± 0.08[a]	0.67
MCHC, g/L	306.86 ± 0.68[a]	305.74 ± 0.29[a]	305.70 ± 0.68[a]	305.11 ± 0.47[a]	0.32
RDW,%	21.94 ± 0.22[a]	21.77 ± 0.10[a]	21.91 ± 0.22[a]	22.12 ± 0.16[a]	0.32
PLT, $10^3/\mu L$	281.02 ± 10.26[a]	283.85 ± 4.29[a]	290.49 ± 10.42[a]	286.12 ± 7.10[a]	0.93
MPV, fL	14.57 ± 0.16[a]	14.70 ± 0.07[a]	14.98 ± 0.16[a]	14.79 ± 0.11[a]	0.40
Blood 3	**RES**	**MID**	**SUS**	**DEAD**	**FDR-group**
RBC, $10^6/\mu L$	5.81 ± 0.04[a]	5.77 ± 0.02[a]	5.75 ± 0.04[a]	5.79 ± 0.03[a]	0.74
HGB, g/L	101.66 ± 0.60[a]	101.03 ± 0.25[a]	100.63 ± 0.61[a]	101.32 ± 0.44[a]	0.68
HCT,%	32.92 ± 0.20[a]	32.87 ± 0.09[a]	32.85 ± 0.21[a]	32.71 ± 0.15[a]	0.87
MCV, fL	57.15 ± 0.20[a]	57.15 ± 0.08[a]	57.22 ± 0.20[a]	56.80 ± 0.14[a]	0.29
MCH, pg	17.51 ± 0.07[a]	17.54 ± 0.03[a]	17.52 ± 0.08[a]	17.49 ± 0.06[a]	0.93
MCHC, g/L	306.66 ± 0.95[a]	306.68 ± 0.60[a]	304.65 ± 0.97[a]	307.79 ± 0.79[b]	0.14
RDW,%	**18.26 ± 0.10[a4]**	**18.40 ± 0.0[a]**	**18.58 ± 0.10[ab]**	**18.72 ± 0.07[b]**	**0.0004**
PLT, $10^3/\mu L$	390.20 ± 11.04[a]	362.14 ± 11.10[a]	305.14 ± 0.81[a]	363.44 ± 8.12[a]	0.26
MPV, fL	**14.59 ± 0.16[a]**	**14.92 ± 0.06[a]**	**15.01 ± 0.16[ab]**	**15.51 ± 0.12[b]**	**<0.0001**
Blood 4	**RES**	**MID**	**SUS**	**DEAD**	**FDR-group**
RBC, $10^6/\mu L$	**6.36 ± 0.04[b]**	**6.32 ± 0.01[b]**	**6.16 ± 0.04[a]**	**6.22 ± 0.05[ab]**	**0.0009**
HGB, g/L	**106.52 ± 0.60[b]**	**105.16 ± 0.25[b]**	**100.78 ± 0.62[a]**	**100.94 ± 0.81[a]**	**<0.0001**
HCT,%	**35.54 ± 0.20[b]**	**35.21 ± 0.08[b]**	**34.15 ± 0.22[a]**	**34.05 ± 0.28[a]**	**<0.0001**
MCV, fL	**56.03 ± 0.20[b]**	**55.74 ± 0.08[b]**	**55.44 ± 0.20[b]**	**54.39 ± 0.26[a]**	**<0.0001**
MCH, pg	**16.89 ± 0.07[b]**	**16.78 ± 0.03[b]**	**16.57 ± 0.07[a]**	**16.37 ± 0.09[a]**	**<0.0001**
MCHC, g/L	301.73 ± 0.72[a]	301.17 ± 0.30[a]	299.45 ± 0.73[a]	299.97 ± 0.94[a]	0.15
RDW,%	**18.31 ± 0.09[a]**	**18.57 ± 0.04[b]**	**18.84 ± 0.09[c]**	**18.89 ± 0.12[bc]**	**0.0001**
PLT, $10^3/\mu L$	352.11 ± 9.84[a]	337.37 ± 4.02[a]	354.67 ± 10.25[a]	339.54 ± 13.13[a]	0.38
MPV, fL	**13.31 ± 0.11[a]**	**13.41 ± 0.04[a]**	**13.63 ± 0.11[ab]**	**14.12 ± 0.13[b]**	**<0.0001**

[1]RBC: red blood cell concentration; HGB: hemoglobin concentration; HCT: hematocrit; MCV: mean corpuscular volume; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; RDW: red blood cell distribution width; PLT: platelet concentration; MPV: mean platelet volume. [2]FDR-group: adjusted p-values for the significant level of group effect using the Benjamini and Hochberg correction (FDR) in R to control false positives from multiple comparisons (R Core Team, 2017 Package 'stats'). [3]Values in a column suffixed with different letters are significantly different from each other at FDR < 0.05. [4]Significant differences among RES, MID, SUS and DEAD groups are highlighted in bold (FDR < 0.05).

(HSCs) (Orkin and Zon, 2008; Zaretsky et al., 2014). In response to disturbances of the hematopoietic equilibrium, such as infection, extensive proliferation and increased differentiation of HSCs are required to meet the higher demand of immune effector cells (Shahbazian et al., 2004; Singh et al., 2008; Johns et al., 2009; Yáñez et al., 2009; Baldridge et al., 2011; Boettcher and Manz, 2017). In the natural challenge model, our results showed that all white blood cell traits increased significantly from Blood 1 to Blood 4, although some traits, including NEU, EOS, and BASO, decreased from Blood 3 to Blood 4 (**Tables 4**, **5**). According to the reference intervals, white blood cell traits have the tendency to increase slightly with age, except for NEU, which tends to decrease with age (**Table 1**) (Iowa State University's Clinical Pathology Laboratory, 2011). Eze et al. (2011) indicated that white blood cell traits did not vary significantly between clinically healthy piglets and adults raised under an intensive management system. Therefore, the significant increases of all white blood cell traits observed here are likely to result from recruiting phagocytes (monocytes, neutrophils), immunocytes (lymphocytes), and granulocytes (neutrophils, eosinophils, and basophils) to drive immune responses at the early stage of infection (George-Gay and Parker, 2003; Mitre and Nutman, 2006; Rothenberg and Hogan, 2006; Porwit et al., 2011).

Notably, resilient pigs had significantly higher LYM for Blood 3 and based on Δ13 compared to the other three groups. Lymphocytes are mainly indicative of initiation and execution of the adaptive immune responses due to their essential and multiple roles in adaptive immunity (**Figure 1B**). Higher LYM in the blood of resilient pigs may indicate earlier and greater adaptive immune responses and increase the transport of lymphocytes to the infected tissues. Resilient pigs may be primed to orchestrate immune responses against a wide variety of pathogens more efficiently together with the higher concentrations of lymphocytes in infected tissues at the early stage of infection and, therefore, limiting the adverse effect caused

TABLE 6 | Least-squares means ± standard errors for changes of red blood cell and platelet traits[1] between Blood1, Blood 3, and Blood 4 of animals in the resilient (RES), average (MID), susceptible (SUS), and dead (DEAD) groups.

Δ13[2]	**RES**	**MID**	**SUS**	**DEAD**	**FDR-group[5]**
RBC, $10^6/\mu L$	−0.43 ± 0.05[a6]	−0.44 ± 0.02[a]	−0.43 ± 0.05[a]	−0.33 ± 0.03[a]	0.16
HGB, g/L	−15.57 ± 0.93[a]	−15.77 ± 0.38[a]	−16.25 ± 0.96[a]	−14.72 ± 0.68[a]	0.59
HCT,%	−4.58 ± 0.31[a]	−4.68 ± 0.13[a]	−4.83 ± 0.32[a]	−4.59 ± 0.23[a]	0.93
MCV, fL	−4.11 ± 0.26[a]	−4.31 ± 0.10[a]	−4.74 ± 0.26[a]	−4.44 ± 0.18[a]	0.43
MCH, pg	−1.22 ± 0.10[a]	−1.20 ± 0.04[a]	−1.30 ± 0.10[a]	−1.17 ± 0.07[a]	0.81
MCHC, g/L	**−0.63 ± 1.00[a7]**	**0.91 ± 0.42[ab]**	**−0.92 ± 1.04[a]**	**3.13 ± 0.74[b]**	**0.01**
RDW,%	−3.58 ± 0.20[a]	−3.37 ± 0.08[a]	−3.32 ± 0.20[a]	−3.63 ± 0.14[a]	0.44
PLT, $10^3/\mu L$	105.93 ± 14.20[a]	76.20 ± 5.73[a]	67.59 ± 10.26[a]	67.59 ± 10.26[a]	0.29
MPV, fL	0.29 ± 0.22[a]	0.40 ± 0.09[a]	0.10 ± 0.22[a]	0.79 ± 0.16[a]	0.14
Δ34[3]	**RES**	**MID**	**SUS**	**DEAD**	**FDR-group**
RBC, $10^6/\mu L$	0.56 ± 0.05[a]	0.54 ± 0.02[a]	0.40 ± 0.05[a]	0.36 ± 0.06[a]	0.01
HGB, g/L	**6.23 ± 0.79[b]**	**4.32 ± 0.34[b]**	**0.59 ± 0.81[a]**	**−0.64 ± 1.03[a]**	**<0.0001**
HCT,%	**2.61 ± 0.27[b]**	**2.22 ± 0.12[b]**	**1.15 ± 0.29[a]**	**1.04 ± 0.38[a]**	**0.0002**
MCV, fL	**−1.01 ± 0.20[b]**	**−1.40 ± 0.08[b]**	**−1.84 ± 0.20[ab]**	**−2.55 ± 0.25[a]**	**<0.0001**
MCH, pg	**−0.59 ± 0.07[c]**	**−0.73 ± 0.03[bc]**	**−0.94 ± 0.08[ab]**	**−1.01 ± 0.10[a]**	**0.0007**
MCHC, g/L	−4.22 ± 1.06[a]	−4.57 ± 0.45[a]	−4.61 ± 1.10[a]	−3.74 ± 1.43[a]	0.94
RDW,%	0.06 ± 0.10[a]	0.22 ± 0.04[a]	0.30 ± 0.10[a]	0.15 ± 0.13[a]	0.46
PLT, $10^3/\mu L$	−43.81 ± 14.00[a]	−30.55 ± 5.90[a]	−2.77 ± 14.53[a]	−52.74 ± 18.46[a]	0.21
MPV, fL	**−1.34 ± 0.17[ab]**	**−1.55 ± 0.07[a]**	**−1.24 ± 0.17[ab]**	**−0.76 ± 0.22[b]**	**0.02**
Δ14[4]	**RES**	**MID**	**SUS**	**DEAD**	**FDR-group**
RBC, $10^6/\mu L$	0.18 ± 0.05[a]	0.16 ± 0.02[a]	0.05 ± 0.05[a]	0.09 ± 0.06[a]	0.22
HGB, g/L	−10.75 ± 0.98[a]	−11.76 ± 0.39[a]	−14.96 ± 1.04[a]	−13.42 ± 1.35[a]	0.06
HCT,%	−1.93 ± 0.32[a]	−2.22 ± 0.12[a]	−3.15 ± 0.35[a]	−2.96 ± 0.43[a]	0.06
MCV, fL	−5.38 ± 0.29[a]	−5.84 ± 0.11[a]	−6.70 ± 0.30[a]	−6.74 ± 0.39[a]	0.01
MCH, pg	**−1.81 ± 0.11[b]**	**−1.99 ± 0.05[ab]**	**−2.33 ± 0.12[a]**	**−2.20 ± 0.15[ab]**	**0.02**
MCHC, g/L	−5.53 ± 1.02[a]	−5.40 ± 0.44[a]	−7.01 ± 1.04[a]	−5.05 ± 1.27[a]	0.58
RDW,%	−3.36 ± 0.22[a]	−2.93 ± 0.09[a]	−2.84 ± 0.23[a]	−3.68 ± 0.29[a]	0.08
PLT, $10^3/\mu L$	69.91 ± 13.48[a]	56.42 ± 5.35[a]	60.74 ± 14.30[a]	22.48 ± 18.13[a]	0.32
MPV, fL	−1.12 ± 0.18[a]	−1.24 ± 0.07[a]	−1.15 ± 0.19[a]	−0.40 ± 0.25[a]	0.05

[1]RBC: red blood cell concentration; HGB: hemoglobin concentration; HCT: hematocrit; MCV: mean corpuscular volume; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; RDW: red blood cell distribution width; PLT: platelet concentration; MPV: mean platelet volume. [2]The change of complete blood count (CBC) traits from Blood 1 to Blood 3; [3]The change of CBC traits from Blood 3 to Blood 4; [4]The change of CBC traits from Blood 1 to Blood 4. [5]FDR-group: adjusted p-values for the significant level of group effect using the Benjamini and Hochberg correction (FDR) in R to control false positives from multiple comparisons (R Core Team, 2017 Package 'stats'). [6]Values in a column suffixed with different letters are significantly different from each other at FDR < 0.05. [7]Significant differences among RES, MID, SUS and DEAD groups are highlighted in bold (FDR < 0.05).

by infectious challenges (Wilkie and Mallard, 1999; Badri and Wood, 2003; Zabriskie, 2009; Zhu et al., 2010; Luckheeram et al., 2012). This was also indicated by the negative genetic relationships of TR with LYM in Blood 3 and its change based on Δ13. A higher increase of LYM from Blood 1 to Blood 3 should favor resilience, which is related to a lower TR. Neutrophils, which increased significantly from Blood 1 to Blood 3 for all groups are both present as phagocytes and granulocytes in the innate immune response to defend against bacterial pathogens (**Figure 1A**) (Pham, 2006; Kolaczkowska and Kubes, 2013; Boettcher and Manz, 2017). However, after moving animals into the grow-to-finish stage, between Blood 3 and Blood 4, NEU showed the tendency to decrease in the RES and MID groups, which was opposite to the significant rise observed for the SUS and DEAD groups. Thus, NEU in Blood 4, and its changes based on Δ34 and Δ14 were also significantly lower for the RES and MID groups compared to the SUS and DEAD groups. Sustained high levels of NEU for the SUS and DEAD groups may be related to ongoing bacterial infection. The decrease of NEU in the blood of the RES and MID groups may indicate the recovery and resolution of inflammation when pathogens were brought under control by early initiation and efficient adaptive immune responses in resilient animals with higher increase of LYM from Blood 1 to Blood 3 (Savill, 1997; Nathan, 2006). Alternatively, it may reflect that neutrophils were already transported to the infected tissues to defend against pathogens in the RES and MID groups. These suggested processes need to be further explored for example, by monitoring the pathogen load in animals and identifying signs of the resolution of inflammation, such as the exodus of neutrophils in infected tissues and "stop signals" or checkpoints of inflammation, including lipoxins, Resolvins, and D-series prostaglandins (Serhan et al., 2007). Positive genetic correlations of TR with NEU in Blood 4 and its change based on Δ14, and the

TABLE 7 | Estimates of heritability ± standard error for complete blood count (CBC) traits.

Traits[1]	Blood 1	Blood 3	Blood 4	Δ13[2]	Δ34[3]	Δ14[4]
WBC	**0.16 ± 0.04**[5]	**0.22 ± 0.04**	**0.19 ± 0.04**	**0.09 ± 0.04**	**0.14 ± 0.04**	**0.15 ± 0.04**
NEU	**0.18 ± 0.04**	**0.18 ± 0.04**	**0.13 ± 0.04**	**0.11 ± 0.04**	**0.11 ± 0.04**	**0.07 ± 0.04**
LYM	**0.21 ± 0.04**	**0.21 ± 0.04**	**0.30 ± 0.04**	**0.11 ± 0.04**	**0.20 ± 0.04**	**0.24 ± 0.04**
MONO	0.05 ± 0.03	**0.12 ± 0.03**	0.02 ± 0.03	**0.08 ± 0.03**	0.00 ± 0.00	0.05 ± 0.04
EOS	**0.22 ± 0.04**	**0.19 ± 0.04**	**0.27 ± 0.04**	**0.07 ± 0.03**	0.00 ± 0.03	**0.08 ± 0.04**
BASO	**0.08 ± 0.04**	**0.10 ± 0.03**	**0.13 ± 0.04**	**0.06 ± 0.04**	**0.06 ± 0.04**	0.06 ± 0.05
RBC	**0.27 ± 0.04**	**0.30 ± 0.04**	**0.34 ± 0.05**	**0.08 ± 0.04**	0.04 ± 0.04	**0.08 ± 0.05**
HGB	**0.08 ± 0.03**	**0.16 ± 0.04**	**0.28 ± 0.05**	**0.16 ± 0.04**	**0.11 ± 0.04**	**0.09 ± 0.05**
HCT	**0.09 ± 0.03**	**0.23 ± 0.04**	**0.23 ± 0.04**	0.04 ± 0.03	0.04 ± 0.04	**0.10 ± 0.05**
MCV	**0.19 ± 0.04**	**0.38 ± 0.04**	**0.46 ± 0.05**	**0.08 ± 0.03**	**0.22 ± 0.05**	0.06 ± 0.04
MCH	**0.18 ± 0.04**	**0.39 ± 0.04**	**0.53 ± 0.05**	**0.15 ± 0.04**	**0.13 ± 0.05**	0.06 ± 0.05
MCHC	**0.13 ± 0.04**	**0.25 ± 0.04**	**0.26 ± 0.05**	**0.17 ± 0.04**	**0.20 ± 0.05**	**0.07 ± 0.05**
RDW	**0.13 ± 0.03**	**0.14 ± 0.04**	**0.08 ± 0.04**	**0.14 ± 0.04**	**0.09 ± 0.04**	**0.18 ± 0.05**
PLT	**0.15 ± 0.03**	**0.07 ± 0.03**	**0.08 ± 0.04**	0.01 ± 0.03	0.00 ± 0.03	0.04 ± 0.03
MPV	**0.11 ± 0.03**	**0.19 ± 0.04**	**0.23 ± 0.04**	0.02 ± 0.03	**0.10 ± 0.04**	**0.08 ± 0.04**

[1]WBC: total white blood cell concentration; NEU: neutrophil concentration; LYM: lymphocyte concentration; MONO: monocyte concentration; EOS: eosinophil concentration; BASO: basophil concentration; RBC: red blood cell concentration; HGB: hemoglobin concentration; HCT: hematocrit; MCV: mean corpuscular volume; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; RDW: red blood cell distribution width; PLT: platelet concentration; MPV: mean platelet volume. [2]The change of CBC traits from Blood 1 to Blood 3; [3]The change of CBC traits from Blood 3 to Blood 4; [4]The change of CBC traits from Blood 1 to Blood 4. [5]Significant estimates of genetic variances are highlighted in bold based on the likelihood ratio test by comparing full models to restricted models that constrained genetic variances to zero in ASReml 4.1 ($p < 0.05$).

negative genetic correlation of GFGR with NEU based on Δ34 together may indicate that higher NEU in the grow-to-finish stage has a negative relationship with resilience, which is associated with increased TR and decreased GFGR.

Unlike the situation of white blood cells, red blood cell traits declined from Blood 1 to Blood 3 to the same degree for all groups, except for MCHC, which did not show a significant decrease (**Table 6**). By comparing clinically healthy grower to finisher pigs, Ježek et al. (2018) suggested that red blood cell traits, including RBC, HGB, HCT, MCV, and MCH, increased with age. The reference intervals from Iowa State University's Clinical Pathology Laboratory (2011) also indicated a tendency for red blood cell traits to increase with age in pigs. Therefore, significant decreases in red blood cell traits from Blood 1 to Blood 3 are likely caused by the challenge of bacterial pathogens, which could damage circulating blood cells and accelerate hemolysis for iron to support bacterial cellular processes of respiration and replication (Barrett-Connor, 1972; Kent, 1994; Viana, 2011; Cassat and Skaar, 2013). This, however, changed during the late stage of infection for the RES and MID groups, for which HGB and HCT increased significantly from Blood 3 to Blood 4. Although red blood cell traits may increase with age, the significantly higher increase of HGB and HCT from Blood 3 to Blood 4 of more resilient animals may also suggest a better performance and faster recovery from infection by providing a higher level of iron and oxygen to the host (Morera and MacKenzie, 2011). Moreover, hemoglobin has been found to directly participate in immune responses as a source of bioactive peptides that exhibit antimicrobial activity against bacteria (El Bishlawy, 1999; Liepke et al., 2003). The higher increase of HGB from Blood 3 to Blood 4 of resilient animals are expected to enhance immune responses and work together with the other immune cells to defend against pathogens. Although relatively large standard errors are reported, highly negative genetic correlations of TR with HGB and HCT based on Δ34 and in Blood 4 may indicate that higher HGB and HCT during the late stage of infection favors resilience, which is related to lower TR. In addition, the significant increase in RDW has been identified to be a valuable index for assessing various pathological conditions, including inflammation and respiratory diseases in humans (Goyal et al., 2017). Our results also showed higher levels of RDW in Blood 3 and Blood 4 for less resilient animals. According to the highly positive genetic correlation of TR with RDW in Blood 4 (0.89 ± 0.26), higher RDW after challenge may have adverse effects associated with increasing the TR.

Significant genetic correlations of CBC traits with resilience traits suggest that a well-functioning immune system plays an essential role in resilient animals to maintain performance and prevent death from infection. An adequate nutritional status is necessary for the normal functioning of various components of the immune system because the immune system is energetically expensive (Coop and Kyriazakis, 2001; McDade, 2005; Nelson and Williams, 2007; Calder, 2013). Any changes in resource demands by the immune system can create significant differences in the level of fitness and performance that are related to resilience (Stearns, 1976). When nutrient resources are limited by decreased feed intake in response to disease challenge, a trade-off is expected to occur between the immune system and other nutrient-demands, such as growth (Lochmiller and Deerenberg, 2000; Doeschl-Wilson et al., 2009; Rauw, 2012; Putz et al., 2018). Although the negative genetic correlation between GFGR and TR could be the result of decreasing feed

TABLE 8 | Estimates of genetic correlations ± standard errors for complete blood count (CBC) traits that showed significant differences among groups with the resilience traits of grow-to-finish growth rate (GFGR) and treatment rate (TR).

Traits[1]	GFGR	TR
Blood3		
LYM	0.10 ± 0.18	**−0.38 ± 0.18**[2]
RDW	−0.07 ± 0.21	0.39 ± 0.22
MPV	0.09 ± 0.18	0.26 ± 0.18
Blood4		
NEU	−0.31 ± 0.20	**0.50 ± 0.23**
LYM	0.16 ± 0.15	−0.28 ± 0.16
RBC	0.15 ± 0.15	−0.08 ± 0.17
HGB	0.04 ± 0.16	−0.25 ± 0.18
HCT	0.10 ± 0.17	−0.33 ± 0.19
MCV	−0.08 ± 0.15	−0.16 ± 0.16
MCH	−0.03 ± 0.14	−0.21 ± 0.15
RDW	−0.12 ± 0.28	**0.89 ± 0.26**
MPV	0.09 ± 0.17	0.11 ± 0.19
Δ13[3]		
LYM	0.15 ± 0.23	**−0.46 ± 0.24**
MCHC	−0.25 ± 0.20	0.26 ± 0.21
Δ34[4]		
NEU	**−0.45 ± 0.21**	0.44 ± 0.26
RBC	−0.33 ± 0.45	−0.35 ± 0.43
HGB	0.01 ± 0.25	−0.32 ± 0.28
HCT	−0.29 ± 0.44	**−0.82 ± 0.47**
MCV	0.03 ± 0.19	0.02 ± 0.20
MCH	0.25 ± 0.25	0.14 ± 0.28
MPV	−0.15 ± 0.26	−0.27 ± 0.28
Δ14[5]		
NEU	−0.32 ± 0.26	**0.76 ± 0.29**
MCV	0.02 ± 0.33	−0.02 ± 0.35
MCH	0.00 ± 0.35	0.26 ± 0.36

[1]NEU: neutrophil concentration; LYM: lymphocyte concentration; MONO: monocyte concentration; RBC: red blood cell concentration; HGB: hemoglobin concentration; HCT: hematocrit; MCV: mean corpuscular volume; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; RDW: red blood cell distribution width. [2]Significant estimates of genetic correlations are highlighted in bold based on the likelihood ratio test by comparing full models to restricted models that constrained the genetic covariance to zero in ASReml 4.1 ($p < 0.05$). [3]The change of CBC traits from Blood 1 to Blood 3; [4]The change of CBC traits from Blood 3 to Blood 4; [5]The change of CBC traits from Blood 1 to Blood 4.

intake in challenged pigs, it might further indicate the trade-off and competing demands for the investment of nutrients in growth and immune function. In susceptible and dead animals, the infection may not be eliminated effectively as a result of a weak immune response. Therefore, decreased feed intake, along with prolonged infection, may further compromise the immune system, leading to a more severe disease state, and increased susceptibility to other pathogens (Keusch, 2003; Nelson and Williams, 2007; Hine et al., 2014). Conversely, the significant changes of CBC traits over time in RES animals, including higher LYM based on Δ13, higher HGB and HCT based on Δ34, and lower NEU based on Δ34 together, are expected to indicate the allocation of more resources toward immunity during the infection stage to help limit infection in resilient animals. Once the infection is brought under control by an efficient immune response, resilient animals may recover earlier from the infection, which could allow them to allocate more resources to maintain a higher growth rate in the grow-to-finish stage (McDade, 2005; Calder, 2013).

Estimates of Heritabilities

Estimates of heritabilities for CBC traits have been reported in many studies (**Table 9**). Some of these were conducted under a controlled environment with limited disease challenges and types of pathogens (Clapperton et al., 2008, 2009). Others were conducted under a lower health status condition with multiple pathogens (Henryon et al., 2006; Flori et al., 2011; Mpetile et al., 2015). Heritability estimates for CBC traits in the natural challenge model in this study were within the range of estimates reported in these studies. Additionally, we were able to provide heritability estimates for novel CBC traits that capture changes of CBC in response to the challenge of infection. Heritability estimates for many CBC traits, especially red blood cells, were observed to be higher in Blood 3 and Blood 4 than in Blood 1, possibly because genetic variances of these traits may be more fully expressed in a lower health environment when there is the challenge of infection (Clapperton et al., 2008, 2009).

Heritability estimates for GFGR and TR in this study were 0.15 ± 0.04 and 0.04 ± 0.01, respectively. Guy et al. (2018) estimated the heritability of treatments for a relatively high-health herd to be between 0.04 ± 0.03 and 0.06 ± 0.04. Putz et al. (2018) estimated the heritability of finishing average daily gain (FinADG) to be 0.25 ± 0.07 based on the phenotypes of the first three cycles of this natural challenge model. Moreover, the heritability for treatment rate adjusted to 180 days for animals that reached 65 days of age (TRT180) was estimated to 0.29 ± 0.07 by Putz et al. (2018). Our use of phenotypes and genotypes on a larger population with 2593 animals of six cycles resulted in relatively lower estimates of heritabilities and lower standard errors for both growth and treatment traits. Moreover, heritability estimates for the treatment rate were different since the definitions of this trait were not the same. In Putz et al. (2018), animals that died before the age of 65 days were excluded, but we included all animals unless they died without receiving any treatment. Moreover, we used additional batches of animals that were introduced into the natural challenge. As disease pressure varied by batch and on a seasonal basis, treatment rates could change accordingly. Moreover, treatment rates may also change with many other non-infectious factors, such as the level of stress caused by weather and transport in these batches (Bishop and Woolliams, 2014). Therefore, the heritability estimates for treatment rates are expected to change correspondingly.

CONCLUSION

Resilience is a valuable attribute in livestock to manage infectious diseases and sustainably increase production efficiency,

TABLE 9 | Heritability estimates of complete blood count traits in related studies reported in the literature.

Traits[1]	Henryon et al., 2006	Clapperton et al., 2008				Clapperton et al., 2009		Flori et al., 2011	Mpetile et al., 2015
		SPF[2]	Non-SPF[3]	Start-test[4]	End-test[5]	SPF	Non-SPF		
WBC	0.25 ± 0.05	0.06 ± 0.11	0.37 ± 0.16	0.24 ± 0.15	0.18 ± 0.11	0.29 ± 0.13	0.28 ± 0.11	0.73 ± 0.20	0.23 ± 0.19
NEU	0.22 ± 0.04	–	–	–[6]	–	–	–	0.61 ± 0.20	0.31 ± 0.21
LYM	0.24 ± 0.05	–	–	–	–	–	–	0.72 ± 0.21	0.15 ± 0.19
MONO	0.22 ± 0.04	0.58 ± 0.18	0.58 ± 0.18	0.52 ± 0.17	0.59 ± 0.14	0.26 ± 0.11	0.16 ± 0.13	0.38 ± 0.20	0.36 ± 0.20
EOS	0.30 ± 0.05	–	–	–	–	–	–	0.80 ± 0.21	0.58 ± 0.12
BASO	–	–	–	–	–	–	–	–	0.12 ± 0.19
RBC	–	–	–	–	–	–	–	0.43 ± 0.20	0.62 ± 0.25
HGB	–	–	–	–	–	–	–	–	0.56 ± 0.13
HCT	–	–	–	–	–	–	–	–	0.06 ± 0.14
MCV	–	–	–	–	–	–	–	–	0.47 ± 0.24
RDW	–	–	–	–	–	–	–	0.70 ± 0.20	0.34 ± 0.25
MCH	–	–	–	–	–	–	–	–	0.37 ± 0.24
MCHC	–	–	–	–	–	–	–	–	0.04 ± 0.16
PLT	–	–	–	–	–	–	–	0.56 ± 0.19	0.11 ± 0.23
MPV	–	–	–	–	–	–	–	–	0.38 ± 0.25

[1]WBC: total white blood cell concentration; NEU: neutrophil concentration, LYM: lymphocyte concentration; MONO: monocyte concentration; EOS: eosinophil concentration; BASO: basophil concentration; RBC: red blood cell concentration; HGB: hemoglobin concentration; HCT: hematocrit; MCV: mean corpuscular volume; RDW: red blood cell distribution width; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; PLT: platelet concentration; MPV: mean platelet volume. [2]Specific pathogen-free (SPF), free of all major swine pathogens; [3]Non-specific pathogen-free (Non-SPF), lower health status condition with the challenge of enzootic pneumonia, Pasteurella multocida, Actinobacillus pleuropneumoniae, Leptospira Bratislava, Salmonella typhimurium, and porcine multi-wasting syndrome; [4]Blood samples collected from animals in both SPF and non-SPF farms at the average of 89 days old; [5]Blood samples collected from animals in both SPF and non-SPF farms at the average of 148 days old; [6]Heritability estimate of the trait was not reported in the study.

as resilient animals can maintain their performance without the need for intensive treatment. Consequently, there is an increasing focus on exploring the potential to select for resilience. Although CBC in Blood 1 is attractive as a potential predictor trait for resilience, as it is a cost-effective phenotype that can be collected from nucleus breeding herds with high health, no significant differences in CBC traits between resilience groups were identified for Blood 1 and estimates of genetic correlations of Blood 1 CBC traits with resilience were not significantly different from zero. Alternatively, for CBC under disease, resilient animals were found to have a greater increase of lymphocyte levels in the blood collected at 2-weeks after challenge, higher levels of hemoglobin and hematocrit, but a significantly lower level of the neutrophil concentration based on the changes from 2- to 6-weeks. Therefore, these changes of CBC traits in response to a disease challenge could provide a measure of resilience. Several of the latter CBC traits were found to be heritable and genetically correlated with resilience. Thus, these CBC traits may have the potential to be further developed as a phenotype for prediction of resilience by collecting data from commercial systems.

DATA AVAILABILITY STATEMENT

Because the data were generated on samples from commercially owned animals, the data analyzed in the current study are not publicly available, but they can be made available for non-commercial use by the corresponding author on reasonable request.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Protection Committee of the Centre de Recherche en Sciences Animales de Deschambault (15PO283) and the Animal Care and Use Committee at the University of Alberta (AUP00002227).

AUTHOR CONTRIBUTIONS

XB analyzed the data and wrote the manuscript with help from GP and ZW. FF, JH, PC, MD, JD, and GP designed the project and developed protocols for the natural disease challenge model. FF oversaw the sample collection and scheduling. JH was in charge of veterinary oversight on the project. CF provided support on CBC data measurement. GP was in charge of the database and genotyping for the project. AP and JD further processed the genotype data and provided the genomic relationship matrix for the project. All authors helped with the interpretation of results and reviewed and approved the final manuscript.

FUNDING

This project was funded by the Genome Canada, Genome Alberta, PigGen Canada, Swine Innovation Porc, and Alberta Agriculture and Forestry. This research is also part of the AMR – One Health Consortium, funded by the Major Innovation Fund program of the Alberta Ministry of Economic Development, Trade and Tourism.

ACKNOWLEDGMENTS

The following individuals served as collaborators and representatives for member companies of the PigGen Canada Consortium: Mr. D. Vandenbroek and Mr. B. DeVries, Alliance Genetics Canada, St. Thomas, ON, Canada; Dr. N. Dion and Ms. S. Blanchette, AlphaGene, St.-Hyacinthe, QU, Canada; Dr. T. Rathje, DNA Genetics, Columbus, NE, United States; Mr. M. Duggan, FastGenetics, Saskatoon, SK, Canada; Dr. R. Kemp, Genesus, London, ON, Canada; Dr. P. Charagu, Hypor, Regina, SK, Canada; Dr. P. Mathur, Topigs Norsvin, Helvoirt, Netherlands; JH, University of Saskatchewan, Saskatoon, SK, Canada. They participated in the development and implementation of the project, coordinated the sources of piglets and the collection of associated data, and discussed with the research team regularly during the execution of the natural challenge model to contribute to the project. Thanks to Mr. Michael Lowing at Delta Genomics (Edmonton, AB, Canada) for processing the raw genotypes from the Affymetrix SNP chips. Thanks to Dr. Jason Grant at Livestock Gentec, University of Alberta for managing the project database and helping with questions concerning genotyping and SNP map file questions. Thanks to Dr. Tianfu Yang and Dr. Chunyan Zhang at Livestock Gentec, University of Alberta for helping with questions on modeling and the use of ASReml. The help of Yan Meng, Janelle Jiminez, Ziqi Yang, and Jiehan Lim at the University of Alberta is appreciated for the processing of blood samples and managing the database of samples and CBC measures. Thanks to Ms. Susan Goruk and Ms. Marnie Newell at the University of Alberta for helping with the processing of blood samples and problem solving of technical problems with the CBC instrument.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00216/full#supplementary-material

REFERENCES

Abbas, A. K., Lichtman, A. H., and Pillai, S. (2015). *Basic Immunology: Functions and Disorders of the Immune System*. Amsterdam: Elsevier.

Albers, G. A., Gray, G. D., Piper, L. R., Barker, J. S., Le Jambre, L. F., and Barger, I. A. (1987). The genetics of resistance and resilience to haemonchus contortus infection in young merino sheep. *Int. J. Parasitol.* 17, 1355–1363. doi: 10.1016/0020-7519(87)90103-90102

Badri, M., and Wood, R. (2003). Usefulness of total lymphocyte count in monitoring highly active antiretroviral therapy in resource-limited settings. *AIDS* 17, 541–545. doi: 10.1097/01.aids.0000050811.06065.7f

Baldridge, M. T., King, K. Y., and Goodell, M. A. (2011). Inflammatory signals regulate hematopoietic stem cells. *Trends Immunol.* 32, 57–65. doi: 10.1016/j.it.2010.12.003

Bao, W.-B., Ye, L., Zi, C., Su, X.-M., Pan, Z.-Y., Zhu, J., et al. (2012). Study on the age-dependent tissue expression of FUT1 gene in porcine and its relationship to E. *coli F*18 receptor. *Gene* 497, 336–339. doi: 10.1016/j.gene.2012.01.035

Barrett-Connor, E. (1972). Anemia and infection. *Am. J. Med.* 52, 242–253. doi: 10.1016/0002-9343(72)90073-3

Bishop, S. C., and Stear, M. J. (2003). Modeling of host genetics and resistance to infectious diseases: understanding and controlling nematode infections. *Vet. Parasitol.* 115, 147–166. doi: 10.1016/s0304-4017(03)00204-8

Bishop, S. C., and Woolliams, J. A. (2014). Genomics and disease resistance studies in livestock. *Livest. Sci.* 166, 190–198. doi: 10.1016/j.livsci.2014.04.034

Boettcher, S., and Manz, M. G. (2017). Regulation of inflammation-and-infection-driven hematopoiesis. *Trends Immunol.* 38, 345–357. doi: 10.1016/j.it.2017.01.004

Brys, G., Hubert, M., and Struyf, A. (2004). A robust measure of skewness. *J. Comput. Graph. Stat.* 13, 996–1017. doi: 10.1198/106186004x12632

Calder, P. C. (2013). Feeding the immune system. *Proc. Nutr. Soc.* 72, 299–309. doi: 10.1017/S0029665113001286

Cassat, J. E., and Skaar, E. P. (2013). Iron in infection and immunity. *Cell Host Microbe* 13, 509–519. doi: 10.1016/j.chom.2013.04.010

Clapperton, M., Diack, A. B., Matika, O., Glass, E. J., Gladney, C. D., Mellencamp, M. A., et al. (2009). Traits associated with innate and adaptive immunity in pigs: heritability and associations with performance under different health status conditions. *Genet. Sel. Evol.* 41:54. doi: 10.1186/1297-9686-41-54

Clapperton, M., Glass, E. J., and Bishop, S. C. (2008). Pig peripheral blood mononuclear leukocyte subsets are heritable and genetically correlated with performance. *Animal* 2, 1575–1584. doi: 10.1017/s1751731108002929

Coop, R. L., and Kyriazakis, I. (2001). Influence of host nutrition on the development and consequences of nematode parasitism in ruminants. *Trends Parasitol.* 17, 325–330. doi: 10.1016/S1471-4922(01)01900-1906

Doeschl-Wilson, A. B., Brindle, W., Emmans, G., and Kyriazakis, I. (2009). Unravelling the relationship between animal growth and immune response during micro-parasitic infections. *PLoS One* 4:e7508. doi: 10.1371/journal.pone.0007508

El Bishlawy, I. M. (1999). Red blood cells, hemoglobin and the immune system. *Med. Hypotheses* 53, 345–346. doi: 10.1054/mehy.1997.0778

Elsevier Health Sciences and Khan Academy (2019). *Immune system*. Available online at: https://www.khanacademy.org/test-prep/mcat/organ-systems#the-immune-system (accessed September 3, 2019).

Eze, J. I., Onunkwo, J. I., Shoyinka, S. V. O., Chah, F. K., Ngene, A. A., Okolinta, N., et al. (2011). Haematological profiles of pigs raised under intensive management system in South-Eastern Nigeria. *Niger. Vet. J.* 31, 115–123. doi: 10.4314/nvj.v31i2.68958

Flori, L., Gao, Y., Laloë, D., Lemonnier, G., Leplat, J.-J., Teillaud, A., et al. (2011). Immunity Traits in pigs: substantial genetic variation and limited covariation. *PLoS One* 6:e22717. doi: 10.1371/journal.pone.0022717

Friggens, N. C., Blanc, F., Berry, D. P., and Puillet, L. (2017). Review: deciphering animal robustness. A synthesis to facilitate its use in livestock breeding and management. *Animal* 11, 2237–2251. doi: 10.1017/S175173111700088X

George-Gay, B., and Parker, K. (2003). Understanding the complete blood count with differential. *J. Perianesth. Nurs.* 18, 96–117. doi: 10.1053/jpan.2003.50013

Gershon, H. (1997). The anti-inflammatory role of the erythrocyte: impairment in the elderly. *Arch. Gerontol. Geriatr.* 24, 157–165. doi: 10.1016/S0167-4943(96)00748-740

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., and Thompson, R. (2015). *ASReml User Guide Release 4.1 Functional Specification*. Hemel Hempstead: VSN International Ltd.

Goyal, H., Lippi, G., Gjymishka, A., John, B., Chhabra, R., and May, E. (2017). Prognostic significance of red blood cell distribution width in gastrointestinal disorders. *World J. Gastroenterol.* 23, 4879–4891. doi: 10.3748/wjg.v23.i27.4879

Guy, S. Z. Y., Li, L., Thomson, P. C., and Hermesch, S. (2018). "Genetic parameters for health of the growing pig using medication records," in *Proceeding 11th World Congress of Genetics Applied to Livestock Production*, Auckland.

Guy, S. Z. Y., Thomson, P. C., and Hermesch, S. (2012). Selection of pigs for improved coping with health and environmental challenges: breeding for resistance or tolerance? *Front. Genet.* 3:281. doi: 10.3389/fgene.2012.00281

Hagger, C. (1998). Litter, permanent environmental, ram-flock, and genetic effects on early weight gain of lambs. *J. Anim. Sci.* 76, 452–457. doi: 10.2527/1998.762452x

Henryon, M., Heegaard, P. M. H., Nielsen, J., Berg, P., and Juul-Madsen, H. R. (2006). Immunological traits have the potential to improve selection of pigs for resistance to clinical and subclinical disease. *Anim. Sci.* 82:597. doi: 10.1079/asc200671

Hine, C. B., Mallard, B. A., Ingham, A. B., and Colditz, I. G. (2014). "Immune competence in livestock,"in *Breeding Focus 2014 - Improving Resilience*, ed S. Hermesch, (Armidale: University of New England), 49–64.

Hottz, E. D., Bozza, F. A., and Bozza, P. T. (2018). Platelets in immune response to virus and immunopathology of viral infections. *Front. Med.* 5:121. doi: 10.3389/fmed.2018.00121

Hubert, M., and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Comput. Stat. Data Anal.* 52, 5186–5201. doi: 10.1016/j.csda.2007.11.008

Iowa State University's Clinical Pathology Laboratory (2011). *Reference Intervals*. Available online at: https://vetmed.iastate.edu/vpath/services/diagnostic-services/clinical-pathology/testing-and-fees/reference-intervals (accessed September 3, 2019).

Janeway, C. A., Traver, P., Walport, M., and Schlomchik, M. J. (2001). *Immunobiology: The Immune System in Health and Disease*, 5th Edn. New York: Taylor & Francis, Inc.

Ježek, J., Starič, J., Nemec, M., Plut, J., Oven, I. G., Klinkon, M., et al. (2018). The influence of age, farm, and physiological status on pig hematological profiles. *J. Swine Health Prod.* 26, 72–78.

Jiang, N., Tan, N. S., Ho, B., and Ding, J. L. (2007). Respiratory protein-generated reactive oxygen species as an antimicrobial strategy. *Nat. Immunol.* 8, 1114–1122. doi: 10.1038/ni1501

Johns, J. L., Macnamara, K. C., Walker, N. J., Winslow, G. M., and Borjesson, D. L. (2009). Infection with *Anaplasma phagocytophilum* induces multilineage alterations in hematopoietic progenitor cells and peripheral blood cells. *Infect. Immun.* 77, 4070–4080. doi: 10.1128/IAI.00570-579

Kent, S. (1994). The etiology of the anemia of chronic disease and infection. *J. Clin. Epidemiol.* 47, 23–33. doi: 10.1016/0895-4356(94)90030-90032

Keusch, G. T. (2003). The history of nutrition: malnutrition, infection and immunity. *J. Nutr.* 133, 336S–340S. doi: 10.1093/jn/133.1.336S

Kolaczkowska, E., and Kubes, P. (2013). Neutrophil recruitment and function in health and inflammation. *Nat. Rev. Immunol.* 13, 159–175. doi: 10.1038/nri3399

Liepke, C., Baxmann, S., Heine, C., Breithaupt, N., Ständker, L., and Forssmann, W.-G. (2003). Human hemoglobin-derived peptides exhibit antimicrobial activity: a class of host defense peptides. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 791, 345–356. doi: 10.1016/S1570-0232(03)00245-9

Lochmiller, R. L., and Deerenberg, C. (2000). Trade-offs in evolutionary immunology: just what is the cost of immunity? *Oikos* 88, 87–98. doi: 10.1034/j.1600-0706.2000.880110.x

Luckheeram, R. V., Zhou, R., Verma, A. D., and Xia, B. (2012). CD4 T Cells: differentiation and Functions. *Clin. Dev. Immunol.* 2012, 1–12. doi: 10.1155/2012/925135

McDade, T. W. (2005). Life history, maintenance, and the early origins of immune function. *Am. J. Hum. Biol.* 17, 81–94. doi: 10.1002/ajhb.20095

Meijerink, E., Fries, R., Vögeli, P., Masabanda, J., Wigger, G., Stricker, C., et al. (1997). Two α (1,2) fucosyltransferase genes on porcine Chromosome 6q11 are closely linked to the blood group inhibitor (S) and *Escherichia coli* F18 receptor (ECF18R) loci. *Mamm. Genome* 8, 736–741. doi: 10.1007/s003359900556

Meijerink, E., Neuenschwander, S., Fries, R., Dinter, A., Bertschinger, H. U., Stranzinger, G., et al. (2000). A DNA polymorphism influencing α (1,2) fucosyltransferase activity of the pig FUT1 enzyme determines susceptibility of small intestinal epithelium to *Escherichia coli* F18 adhesion. *Immunogenetics* 52, 129–136. doi: 10.1007/s002510000263

Miar, Y., Plastow, G. S., Bruce, H. L., Moore, S. S., Durunna, O. N., Nkrumah, J. D., et al. (2014a). Estimation of genetic and phenotypic parameters for ultrasound and carcass merit traits in crossbred beef cattle. *Can. J. Anim. Sci.* 94, 273–280. doi: 10.4141/cjas2013-115

Miar, Y., Plastow, G. S., Moore, S. S., Manafiazar, G., Charagu, P., Kemp, R. A., et al. (2014b). Genetic and phenotypic parameters for carcass and meat quality traits in commercial crossbred pigs. *J. Anim. Sci.* 92, 2869–2884. doi: 10.2527/jas.2014-7685

Mitre, E., and Nutman, T. B. (2006). Basophils, basophilia and helminth infections. *Chem. Immunol. Allergy* 90, 141–156. doi: 10.1159/000088886

Morera, D., and MacKenzie, S. A. (2011). Is there a direct role for erythrocytes in the immune response? *Vet. Res.* 42:89. doi: 10.1186/1297-9716-42-89

Mpetile, Z., Young, J. M., Gabler, N. K., Dekkers, J. C. M., and Tuggle, C. K. (2015). Assessing peripheral blood cell profile of Yorkshire pigs divergently selected for residual feed intake. *J. Anim. Sci.* 93:892. doi: 10.2527/jas.2014-8132

Mulder, H. A., and Rashidi, H. (2017). Selection on resilience improves disease resistance and tolerance to infections. *J. Anim. Sci.* 95, 3346–3358. doi: 10.2527/jas.2017.1479

Nathan, C. (2006). Neutrophils and immunity: challenges and opportunities. *Nat. Rev. Immunol.* 6, 173–182. doi: 10.1038/nri1785

Nelson, K. E., and Williams, C. (2007). *Infectious Disease Epidemiology: Theory and Practice*, 3ed Edn. Burlington, MA: Jones & Bartlett Learning.

Orkin, S. H., and Zon, L. I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* 132, 631–644. doi: 10.1016/j.cell.2008.01.025

Pham, C. T. N. (2006). Neutrophil serine proteases: specific regulators of inflammation. *Nat. Rev. Immunol.* 6, 541–550. doi: 10.1038/nri1841

Porwit, A., McCullough, J., and Erber, W. N. (2011). *Blood and Bone Marrow Pathology*. London: Churchill Livingstone.

Putz, A. M., Harding, J. C. S., Dyck, M. K., Fortin, F., Plastow, G. S., Dekkers, J. C. M., et al. (2018). Novel resilience phenotypes using feed intake data from a natural disease challenge model in wean-to-finish Pigs. *Front. Genet.* 9:660. doi: 10.3389/fgene.2018.00660

R Core Team (2017). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Vienna: R Core Team.

Rauw, W. M. (2012). Immune response from a resource allocation perspective. *Front. Genet.* 3:267. doi: 10.3389/fgene.2012.00267

Rondina, M. T., and Garraud, O. (2014). Emerging evidence for platelets as immune and inflammatory effector cells. *Front. Immunol.* 5:653. doi: 10.3389/fimmu.2014.00653

Rothenberg, M. E., and Hogan, S. P. (2006). The eosinophil. *Annu. Rev. Immunol.* 24, 147–174. doi: 10.1146/annurev.immunol.24.021605.090720

Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15:478. doi: 10.1186/1471-2164-15-478

Savill, J. (1997). Apoptosis in resolution of inflammation. *J. Leukoc. Biol.* 61, 375–380. doi: 10.1002/jlb.61.4.375

Seo, S. (2006). *A review and Comparison of Methods for Detecting Outliers in Univariate Data Sets*. Master's Thesis, University of Pittsburgh, Pittsburgh.

Serhan, C. N., Brain, S. D., Buckley, C. D., Gilroy, D. W., Haslett, C., O'Neill, L. A. J., et al. (2007). Resolution of inflammation: state of the art, definitions and terms. *FASEB J.* 21, 325–332. doi: 10.1096/fj.06-7227rev

Shahbazian, L. M., Quinton, L. J., Bagby, G. J., Nelson, S., Wang, G., and Zhang, P. (2004). *Escherichia coli* pneumonia enhances granulopoiesis and the mobilization of myeloid progenitor cells into the systemic circulation. *Crit. Care Med.* 32, 1740–1746. doi: 10.1097/01.CCM.0000132900.84627.90

Singh, P., Yao, Y., Weliver, A., Broxmeyer, H. E., Hong, S.-C., and Chang, C. H. (2008). Vaccinia virus infection modulates the hematopoietic cell compartments in the bone marrow. *Stem Cells* 26, 1009–1016. doi: 10.1634/stemcells.2007-2461

Stearns, S. C. (1976). Life-history tactics: a review of the ideas. *Q. Rev. Biol.* 51, 3–47. doi: 10.1086/409052

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-2980

Viana, M. B. (2011). Anemia and infection: a complex relationship. *Rev. Bras. Hematol. Hemoter.* 33, 90–92. doi: 10.5581/1516-8484.20110024

Wilkie, B., and Mallard, B. (1999). Selection for high immune response: an alternative approach to animal health maintenance? *Vet. Immunol. Immunopathol.* 72, 231–235. doi: 10.1016/S0165-2427(99)00136-131

Yáñez, A., Murciano, C., O'Connor, J.-E., Gozalbo, D., and Gil, M. L. (2009). Candida albicans triggers proliferation and differentiation of hematopoietic stem and progenitor cells by a MyD88-dependent signaling. *Microbes Infect.* 11, 531–535. doi: 10.1016/j.micinf.2009.01.011

Zabriskie, J. B. (2009). *Essential Clinical Immunology*. Cambridge: Cambridge University Press.

Zaretsky, A. G., Engiles, J. B., and Hunter, C. A. (2014). Infection-induced changes in hematopoiesis. *J. Immunol.* 192, 27–33. doi: 10.4049/jimmunol.1302061

Zhu, J., Yamane, H., and Paul, W. E. (2010). Differentiation of effector CD4 T cell populations (*). *Annu. Rev. Immunol.* 28, 445–489. doi: 10.1146/annurev-immunol-030409-101212

Zimmerman, J. J., Karriker, L. A., Ramirez, A., Schwartz, K. J., and Stevenson, G. W. (2012). *Diseases of Swine*. Hoboken, NJ: John Wiley & Sons.

Conflict of Interest: FF was employed by company Centre de Développement du Porc du Québec, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 April 2020
doi: 10.3389/fmicb.2020.00576

Integrated Transcriptomic and Proteomic Analyses of the Interaction Between Chicken Synovial Fibroblasts and *Mycoplasma synoviae*

*Rui Liu[1,2], Bin Xu[1,2], Shengqing Yu[3], Jingfeng Zhang[1,2], Huawei Sun[1,2], Chuanmin Liu[1,2], Fengying Lu[1,2], Qunxing Pan[1,2] and Xiaofei Zhang[1,2]**

[1] Key Laboratory of Veterinary Biological Engineering and Technology, Ministry of Agriculture, Institute of Veterinary Medicine, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [2] National Center for Engineering Research of Veterinary Bio-products, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [3] Shanghai Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Shanghai, China

OPEN ACCESS

Edited by:
Dong Xia,
Royal Veterinary College (RVC), United Kingdom

Reviewed by:
Pablo Smircich,
Instituto de Investigaciones Biológicas Clemente Estable (IIBCE), Uruguay
Guilherme Campos Tavares,
Federal University of Minas Gerais, Brazil

***Correspondence:**
Xiaofei Zhang
xiaofei0804@sina.com

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: 05 December 2019
Accepted: 16 March 2020
Published: 03 April 2020

Citation:
Liu R, Xu B, Yu S, Zhang J, Sun H, Liu C, Lu F, Pan Q and Zhang X (2020) Integrated Transcriptomic and Proteomic Analyses of the Interaction Between Chicken Synovial Fibroblasts and Mycoplasma synoviae. Front. Microbiol. 11:576. doi: 10.3389/fmicb.2020.00576

Mycoplasma synoviae (MS), which causes respiratory disease, eggshell apex abnormalities, infectious synovitis, and arthritis in avian species, has become an economically detrimental poultry pathogen in recent years. In China, the disease is characterized by infectious synovitis and arthritis. However, the mechanism by which MS causes infectious synovitis and arthritis remains unknown. Increasing evidence suggests that synovial fibroblasts (SF) play a key role in the pathogenesis of arthritis. Here, both RNA sequencing and tandem mass tag analyses are utilized to compare the response of primary chicken SF (CSF) following infection with and without MS. The host response between non-infected and infected cells was remarkably different at both the mRNA and protein levels. In total, 2,347 differentially expressed genes (DEGs) (upregulated, n = 1,137; downregulated, n = 1,210) and 221 differentially expressed proteins (DEPs) (upregulated, n = 129; downregulated, n = 92) were detected in the infected group. A correlation analysis indicated a moderate positive correlation between the mRNA and protein level changes in MS-infected CSF. At both the transcriptomic and proteomic levels, 149 DEGs were identified; 88 genes were upregulated and 61 genes were downregulated in CSF. Additionally, part of these regulated genes and their protein products were grouped into seven categories: proliferation-related and apoptosis-related factors, inflammatory mediators, proangiogenic factors, antiangiogenic factors, matrix metalloproteinases, and other arthritis-related proteins. These proteins may be

involved in the pathogenesis of MS-induced arthritis in chickens. To our knowledge, this is the first integrated analysis on the mechanism of CSF-MS interactions that combined transcriptomic and proteomic technologies. In this study, many key candidate genes and their protein products related to MS-induced infectious synovitis and arthritis were identified.

Keywords: correlation analysis, chicken synovial fibroblasts, *Mycoplasma synoviae*, RNA-Seq, TMT

INTRODUCTION

Mycoplasma synoviae (MS) is a common poultry and extracellular pathogen that leads to acute or chronic respiratory diseases, infectious synovitis, and arthritis in avian species (Fletcher et al., 1976; Catania et al., 2016b; Michiels et al., 2016; Sun et al., 2017), and eggshell apex abnormalities in chickens (Feberwee et al., 2009; Catania et al., 2010). Control of MS infections generally involves eradication of the pathogen from breeder flocks, antibiotic usage, improvements in housing conditions, and vaccination with MS-H, a temperature-sensitive MS strain that is widely used in many countries (Gerchman et al., 2008). However, subacute and chronic infections make the control and elimination of this pathogen particularly difficult. For example, in China, the disease has resulted in the loss of millions of chickens in many regions and has negatively affected the economy of the poultry industry from 2010 to 2015 (Sun et al., 2017).

The clinical characteristics of MS infection are well known (Catania et al., 2016a; Sun et al., 2017; Ball et al., 2018; Kordafshari et al., 2019; Lorenc et al., 2019; Wu et al., 2019), and a number of genomic, proteomic, phenotype microarrays, and other analyses have been conducted. However, only a few MS proteins, including variable lipoprotein hemagglutinin (Narat et al., 1998; Lavric et al., 2007), cysteine protease (Cizelj et al., 2011), neuraminidase (Bercic et al., 2011), the putative nuclease MS53_0284 (Vasconcelos et al., 2005; Cizelj et al., 2016), and matrix metalloproteinase-2 (MMP-2) (Cizelj et al., 2016), have been identified as virulence determinants. In terms of host-pathogen interactions, Goret et al. (2017) showed that *M. hominis* lipoproteins played a role in the activation of human dendritic cells and their immune responses. Addis et al. (2011) utilized a proteomic approach to investigate the impact of *M. agalactiae* infection on mammary epithelium. Another study employed phenotype microarrays to investigate the influence of MS on the global metabolic activity of chicken chondrocytes (CCH) (Dusanic et al., 2014). A separate study performed a quantitative real-time polymerase chain reaction (qRT-PCR) assay of gene expression in an *in vitro* co-culture of MS and CCH (Cizelj et al., 2016). Nevertheless, the molecular pathogenesis of the disease is not well understood, and the nature of the host-pathogen interaction during MS infection has not been clarified. This lack of knowledge can be attributed to the absence of data on the effects of MS infection in host cell responses.

The aim of this study was to investigate the interaction between primary chicken synovial fibroblasts (CSF) and MS strain, HN01. RNA sequencing (RNA-Seq)-based transcriptomics and tandem mass tag (TMT)-based proteomics analyses were applied to identify differentially expressed genes (DEGs) and differentially expressed proteins (DEPs) in response to MS strain HN01 infection in CSF. Using obtained comprehensive quantitative datasets, key genes and their protein products involved in MS-induced arthritis were extracted.

MATERIALS AND METHODS

MS Culture, Cell Culture, and Experimental Design

A wild-type MS strain, HN01, was isolated from broiler breeder flocks that exhibited clinical symptoms of MS from Henan Province. HN01 was grown in Frey mycoplasma broth, as previously described (Dusanic et al., 2009). The number of color-changing units (CCUs) was determined according to previously reported methods (Stemke and Robertson, 1982). Bacteria in the logarithmic phase of growth were used for cell infection.

Primary CSF were prepared from synovial tissues obtained from specific pathogen-free chickens. Synovial tissues were rinsed with sterile phosphate-buffered saline, minced into small pieces, dried with sterilization filter paper, and spread on the bottom of six-well cell plates in Dulbecco's modified Eagle's medium, supplemented with 7.5% fetal bovine serum, 2.5% chicken serum, 100 μg/mL streptomycin, and 100 U/mL penicillin at 37°C for 4–5 days (**Supplementary Figure S1A**). The culture medium was replaced every 3 days until primary cultures reached 90% confluence and subsequently recultivated. Cells at passages 3–6 were used in this study. To characterize the cytological phenotype of synovial cultures, cells were stained with anti-vimentin antibody (Boster Biological Technology Co. Ltd., Wuhan, China) (Hirohata et al., 2009). The majority of cells (98%) were positive for vimentin staining, which was measured by a Nikon N-STORM microscope (Nikon, Tokyo, Japan) (**Supplementary Figure S1B**).

CSF were plated on six-well culture dishes at a density of 3×10^5 cells/well, and incubated overnight. After 24 h of serum starvation, cells were infected with MS at a MOI of 100 (5×10^7 CCU/well) for 48 h. Control cells were mock stimulated with serum free medium for 48 h. Mock- and MS-infected CSF were collected from non-exposed and exposed CSF 48 h post-infection

(hpi), respectively. Samples included three independent biological replicates.

Total RNA Isolation and cDNA Library Construction

Total RNA was extracted from mock- and MS-infected CSF using Trizol (Invitrogen, CA, United States) following the manufacturer's instructions. The integrity of total RNA was measured using a Nanodrop 2000 spectrophotometer and Agilent 2100 BioAnalyzer (Thermo Fisher Scientific, MA, United States). In a subsequent procedure, mRNA molecules were purified from total RNA using Oligo(dT)-attached magnetic beads. For quality control, the cDNA libraries were validated using an Agilent Technologies 2100 BioAnalyzer. Libraries were generated using the BGISEQ-500 platform (BGI, Shenzhen, China).

RNA-Seq Data Analysis

Raw data were filtered using SOAPnuke v1.4.0 (Chen et al., 2018) and Trimmomatic v0.36 (Bolger et al., 2014) to obtain clean reads and remove low-quality reads with >5% unknown nucleotides, reads containing >20% low-quality bases with quality values < 10, and reads containing adaptor sequences. Clean reads were stored in FASTQ format and used for quantitative analysis.

All high-quality clean reads were mapped onto the reference genes that were available in the *Gallus gallus* reference genome GRCg6a[1] using Bowtie2 v2.2.5 (Langmead and Salzberg, 2012). Matched reads were calculated at the gene expression level using RSEM software (Dewey and Li, 2011). For the analysis of DEGs, raw counts were utilized and compared across treatments using the DESeq2 method with a fold change (FC) ≥ 2 and adjusted p-value (q-value) ≤ 0.001 as the significance threshold (Wang et al., 2010).

Protein Preparation, TMT Labeling, and Peptide Fractionation

Three biological replicates for each group were used for proteomic analyses. Samples were sonicated on ice in lysis buffer (8M urea, 1% protease inhibitors). After centrifugation at 12,000 *g* for 10 min at 4°C, the supernatant was transferred to a new centrifuge tube. Protein concentrations and qualities were determined using a Pierce™ BCA Protein Assay kit (Thermo Fisher Scientific, MA, United States) and confirmed by SDS-PAGE. For digestion, the protein solution was treated with 5 mM dithiothreitol for 30 min at 56°C and alkylated with 11 mM iodoacetamide for 15 min at room temperature in the dark. Then, the sample was diluted to a urea concentration < 2 M by adding 100 mM triethyl ammonium bicarbonate (TEAB) (Sigma, St. Louis, MO, United States). For trypsin digestion, trypsin was added in a 50:1 protein-to-trypsin ratio overnight followed by a 100:1 ratio for 4 h. After trypsin digestion, the peptide was desalted using a Strata X C18 SPE column (Phenomenex, Torrance, United States) and vacuum dried. The peptide was reconstituted in 0.5 M TEAB and labeled by TMTs with a TMT labeling kit (Thermo Fisher Scientific, MA, United States) following the manufacturer's instructions. In brief, one unit of TMT reagent was defrosted and restructured in acetonitrile. Then, the peptide mixture was incubated for 2 h at room temperature, pooled, desalted, and vacuum dried by centrifugation. Tryptic peptides were separated by high pH reversed-phase HPLC using an Agilent 300Extend C18 column (5 μm particles, 4.6 mm × 250 mm) (Agilent Technologies, CA, United States). Peptides were first separated with a gradient of 8–32% acetonitrile (pH 9.0) for 1 h into 60 fractions. Subsequently, peptides were merged into nine fractions and vacuum dried.

Liquid Chromatography-Mass Spectrometry (LC-MS/MS) Analysis

Each fraction was resuspended in mobile phase A containing 0.1% formic acid and 2% acetonitrile. A linear gradient was formed as follows: an increase in solvent B (0.1% formic acid in 90% acetonitrile) from 9% to 25% for 30 min, an increase from 25% to 35% for 22 min, an increase to 80% for 4 min, and maintenance at 80% for the last 4 min at a constant flow rate of 350 nL/min on an EASY-nLC 1200 UPLC system (Thermo Fisher Scientific, MA, United States). Eluted peptides were subjected to a nanospray ionization source and analyzed by Oribitrap Fusion Lumos Mass Spectrometer (Thermo Fisher Scientific, MA, United States). A data-dependent manner that alternated between one mass spectrometry scan (60,000 resolving power) and 20 MS/MS (15,000 resolving power) scans with 30 s dynamic exclusion was performed.

Protein Identification and Quantification Based on TMT Data

Mass spectrometry proteome data were deposited to the ProteomeXchange Consortium via the iProX partner repository (Ma et al., 2019). The resulting MS/MS data were analyzed using the MaxQuant search engine v.1.5.2.8 (Tyanova et al., 2016). Tandem mass spectra were searched against the *G. gallus* reference transcriptome (GRCg6a). Trypsin/P was specified as the cleavage enzyme, and two missed cleavages were permissible. Mass tolerance was set to 20 ppm for precursor ions of the first search, 5 ppm for precursor ions of the main search, and 0.02 Da for fragment ions. Carbamidomethyl on Cys was specified as fixed modifications, and oxidation on Met was specified as variable modifications. For protein quantification, TMT-6plex was employed and the false discovery rate was adjusted to <1%. Proteins with a FC > 1.3 and a p-value < 0.05 were considered differentially expressed.

Statistical Analysis

To analyze biological replicates in the experimental groups and observe differences in the gene and protein expression changes between mock and infected groups, a principal component analysis (PCA) of the transcript and protein data was performed using the "princomp" function in R software package.

[1]https://www.ncbi.nlm.nih.gov/genome/111?genome_assembly_id=451987

Bioinformatics Analysis

For transcriptome data, DEGs were utilized for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses, which were further classified according to official classifications. Both GO terms and KEGG pathways with a q-value ≤ 0.05 were significantly enriched in DEGs according to the phyper function in the R software package.

For proteome data, DEPs were annotated with the GO and KEGG databases for functional analysis. For GO proteome annotation, the UniProt-GOA database[2] was searched. For proteins not annotated in the UniProt-GOA database, InterProScan software[3] was used to annotate the proteins' GO function based on the protein sequence alignment method. Additionally, the KEGG online service tool, KAAS[4], was used to annotate the proteins' KEGG database description. To test the enriched DEPs against all identified proteins in the GO and KEGG pathways analyses, Fisher's exact test (two-tailed) was conducted. The GO terms and KEGG pathways with a p-value < 0.05 were considered significant.

STRING database v.10.5[5] was used to analyze the protein-protein interactions (PPIs) of selected proteins. An interaction network with a confidence score ≥ 0.7 (high confidence) was constructed, and the data were visualized using the R package, "networkD3."

Association and Co-expression Analyses

To investigate the potential relevance of quantitative information between mRNA and proteins, subsets of mRNA and proteins were screened using the following cut-off values: FC ≥ 2.0 and q-value ≤ 0.001 (DEGs); FC > 1.3 and p-value < 0.05 (DEPs). Then, RNA-Seq data were used as a searchable database. All identified protein sequences were analyzed and queried with the RNA-Seq data. Concordance between the transcriptome and proteome data was determined by Pearson's correlation analysis with $R > 0.80$ denoting a significantly positive correlation and $0.50 < R < 0.80$ denoting a moderate positive correlation.

qRT-PCR

To confirm the RNA-Seq and TMT results, 19 DEGs were assessed using qRT-PCR. Reaction parameters were identical for all genes. Briefly, purified RNA was transformed into first-strand cDNA using PrimeScript^{TM} RT Master Mix (Takara, Japan) following the manufacturer's instructions. qRT-PCR with cDNA and the appropriate minus RT controls was performed on the ABI StepOnePlus Real-Time PCR system (Applied Biosystems, Foster City, CA, United States) using Power SYBR Green PCR Master mix (Applied Biosystems, Foster City, CA, United States). In brief, after the initial reaction conditions at 50°C for 2 min and denaturation at 95°C, cDNA was amplified by 40 cycles of PCR (95°C for 15 s, 55°C for 15 s, 72°C for 1 min) on the ABI StepOnePlus Real-Time PCR system. Relative RNA expression levels were normalized using *actin* (Wang et al., 2018) and calculated using the $2^{-\Delta\Delta Ct}$ method (Schmittgen and Livak, 2008). All primers used in this study are provided (**Supplementary Table S1**).

[2]http://www.ebi.ac.uk/GOA/

[3]http://www.ebi.ac.uk/interpro/

[4]http://www.genome.jp/kaas-bin/kaas_main

[5]http://string-db.org/

RESULTS

Transcriptomics Analysis of DEGs

To explore the molecular mechanisms involved in the response to MS infection and changes in gene expression, cDNA samples of CSF not exposed to MS infection and CSF 48 hpi were sequenced using the BGISEQ-500 platform. After sequencing and filtering for low-quality reads, clean reads were obtained. In total, 49.19–52.3 million raw sequence reads were generated from the mock and infected groups with percentages of total mapped reads and unique mapped reads ranging from 74.92 to 77.02 and 66.62 to 69.1, respectively (**Supplementary Table S2**). After mapping clean reads onto *G. gallus* genome, 15,400 genes were identified in the transcriptomics analysis (**Supplementary Table S3**). In response to infection, the expression levels of various host genes were remarkably altered. Compared to the untreated group and according to the high-throughput analysis, 2,347 DEGs (upregulated, n = 1,137; downregulated, n = 1,210) were identified in MS-infected CSF (**Figure 1A** and **Supplementary Table S4**).

PCA of Transcriptome and Proteome Data

PCA is a multivariate statistical analysis method that reduces multiple dimensions to a few independent variables while preserving as much of the original data as possible. PCA allows for grouping of samples with overall similar expression characteristics. And the higher the degree of aggregation between samples, the better the biological repeatability. Based on the PCA analysis, two experimental groups included six samples were clustered into 2 groups of a clear separation which indicated that there was a difference in the expression of genes and proteins between the experimental groups and the biological replicates were consistent in the transcriptomics and proteomics analyses (**Figures 2A,B**).

GO and KEGG Pathway Enrichment Analyses of DEGs

To elucidate the distribution of DEGs at the macro-level, each set of DEGs from the mock and treatment groups was mapped in accordance with the GO terms and KEGG pathways. The GO enrichment analysis revealed that upregulated genes in the infected group were strongly related to stimuli responses (i.e., defense, external stimulus/stimulus/chemical/organic substance/external biotic stimulus/biotic stimulus/oxygen-containing compound, immune response, and inflammatory response), multi-organism processes (i.e., response to bacteria or other organisms, including defense responses), cellular processes (i.e., cellular response to a chemical stimulus, including chemotaxis), membrane parts (i.e., intrinsic/integral components of the membrane or membrane parts), and

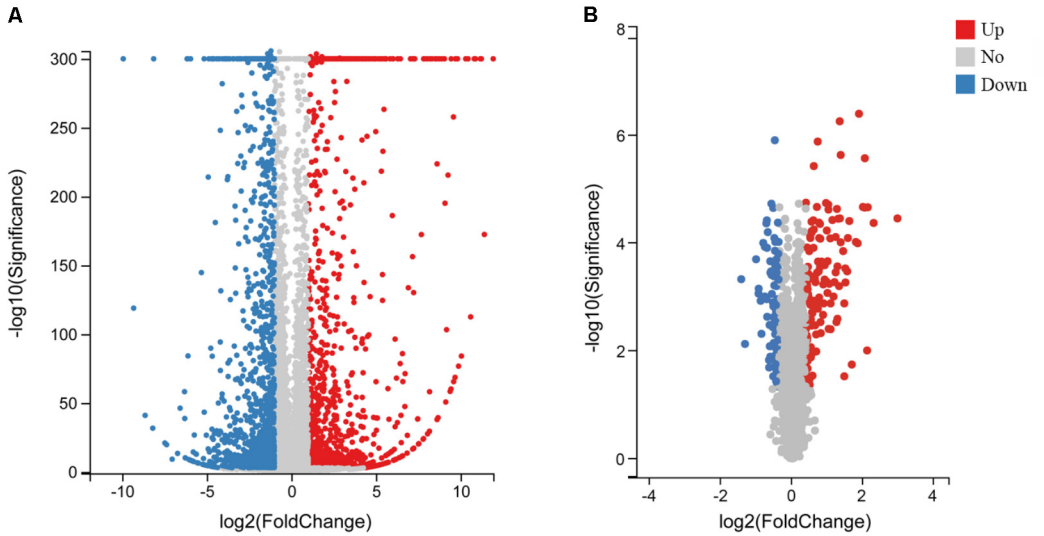


FIGURE 1 | Volcano plot of DEGs and DEPs in MS-infected CSF. Volcano plot of DEGs **(A)** and DEPs **(B)**. Gray dots indicate genes or proteins without significant differential expression, red dots indicate significantly upregulated genes or proteins, and blue dots indicate significantly downregulated genes or proteins. Horizontal and vertical coordinates indicate the fold change differences and *q*-values of DEGs and DEPs, respectively.

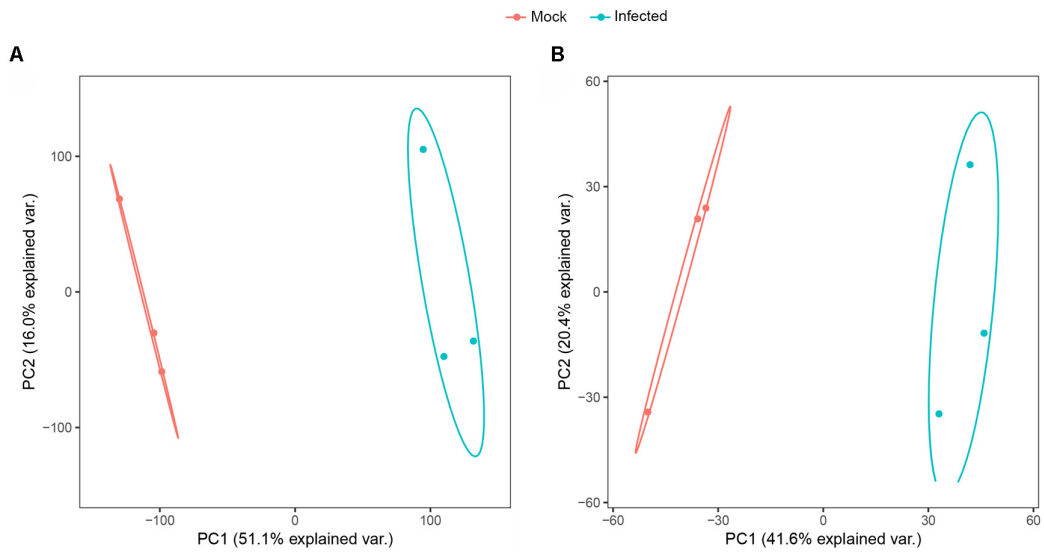


FIGURE 2 | Principal component analysis score plot of transcriptome **(A)** and proteome **(B)** profiles.

immune system processes (**Figure 3A**). Most downregulated genes in MS-incubated CSF were associated with biological adhesion (i.e., cell/biological adhesion), cellular processes (i.e., cell communication/differentiation/development, cellular developmental processes, and cell), extracellular region (i.e., the extracellular matrix or proteinaceous extracellular matrix), biological regulation (i.e., signal transduction, biological regulation, or the regulation of cell migration), and cells (i.e., cell periphery, plasma membrane, plasma membrane part, and intrinsic/integral components of the plasma membrane) (**Figure 3B**). The KEGG pathway enrichment analysis demonstrated that most upregulated genes in the infected group were associated with environmental information processing (i.e., cytokine-cytokine receptor interactions, the TNF/NF-kappa B/cAMP signaling pathway, and neuroactive ligand-receptor interaction), metabolism (i.e., arachidonic acid and caffeine metabolism), and organismal systems (i.e., Toll-like/IL-17/NOD-like receptor signaling pathway, complement and coagulation cascades, and mineral absorption) (**Figure 4A**). Additionally, environmental information processing (i.e., cell adhesion molecules, neuroactive ligand-receptor interactions, and the Ras signaling pathway) and organismal systems (i.e., axon guidance,

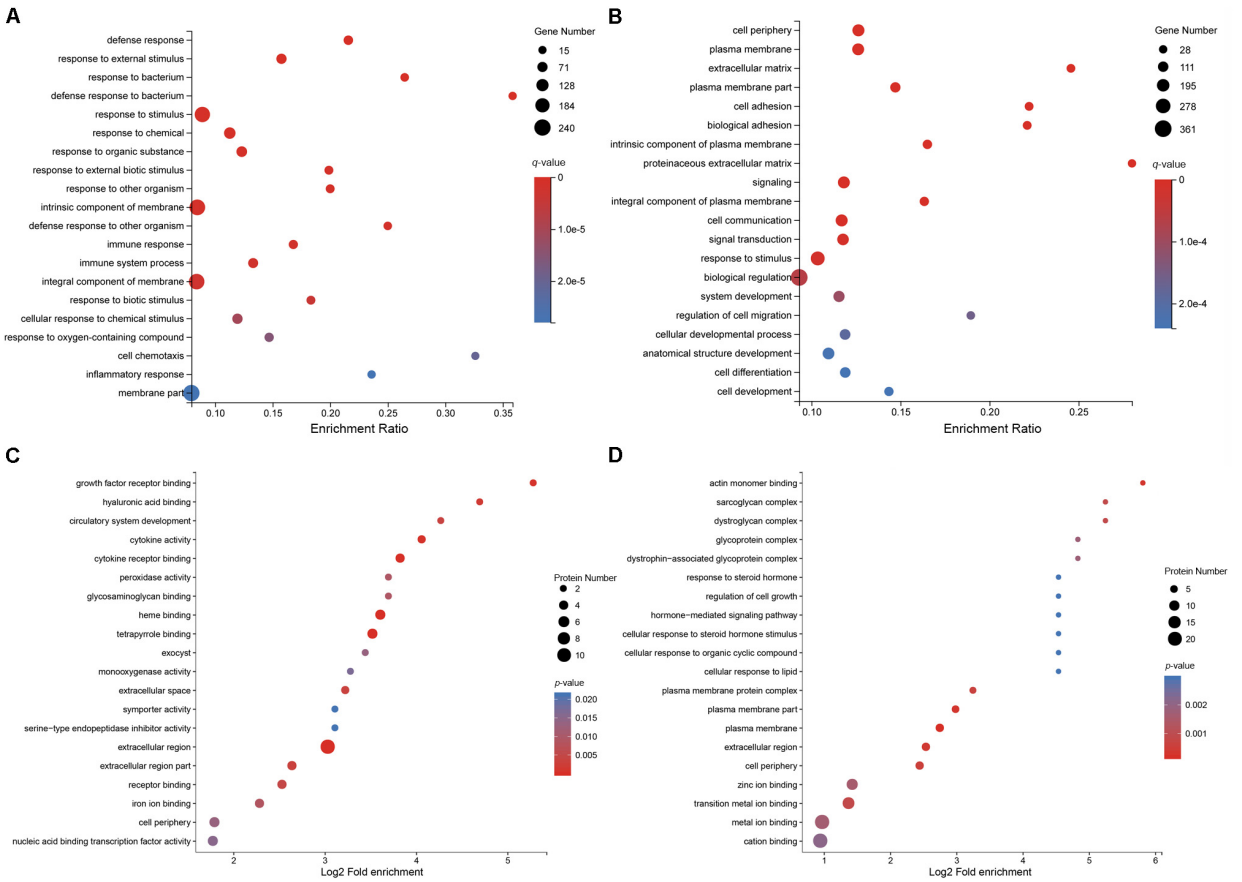


FIGURE 3 | GO enrichment analysis of DEGs and DEPs in MS-treated CSF. Statistics of the top 20 enriched GO term for upregulated DEGs **(A)**, the top 20 enriched GO term for downregulated DEGs **(B)**, the top 20 enriched GO term for upregulated DEPs **(C)**, and the top 20 enriched GO term for downregulated DEPs **(D)**.

leukocyte transendothelial migration, and the neurotrophin signaling pathway) were the most enriched pathways among downregulated genes (**Figure 4B**).

Proteomics Analysis of DEPs

To uncover the effects of MS infection on host gene expression at the translational level, TMT and LC-MS/MS analyses were performed. After quality validation, a total of 320,457 (82,753 matched) spectra were obtained. Of these spectra, 41,982 peptides (39,018 unique peptides) and 5,537 identified proteins (4,609 quantifiable proteins) were detected. Details on the protein mass spectrometry data are provided (**Supplementary Table S5**). In total, 221 DEPs were identified, of which, 129 proteins were upregulated and 92 proteins were downregulated with a FC $>$ 1.3 and p-value $<$ 0.05 (**Figure 1B** and **Supplementary Table S6**).

GO and KEGG Pathway Enrichment Analyses of DEPs

To further characterize the functions of these DEPs, a GO enrichment analysis was conducted. The GO enrichment analysis revealed that the majority of upregulated DEPs in the MS-infected group were related to the extracellular region (i.e., extracellular region, extracellular region part, and extracellular space), cells (i.e., periphery, exocysts, cortex, cortex and its components, and the cytoplasmic region), and binding (i.e., heme, tetrapyrrole, cytokine receptor, and growth factor receptor binding) (**Figure 3C**). Most downregulated DEPs were preferentially associated with cells (i.e., the plasma membrane and its components, periphery, the plasma membrane protein complex, and the dystroglycan complex), binding (i.e., actin monomers, transition metal ions, zinc ions, metal ions, cations, and polysaccharide/pattern binding), cellular processes (i.e., cellular response to steroid hormones, organic cyclic compounds, and lipids), and cellular component organization or biogenesis (i.e., actin cytoskeleton and actin filament organization) (**Figure 3D**).

Functional classification of the DEPs was performed by mapping to the KEGG pathways using KAAS. Results revealed that upregulated proteins in MS-infected CSF were largely involved in metabolism (i.e., steroid hormone biosynthesis, pentose and glucuronate interconversions, ascorbate and aldarate metabolism, retinol metabolism, and amino sugar and nucleotide sugar metabolism), organismal systems (i.e., the PPAR signaling pathway, intestinal immune network for IgA production, NOD-like receptor signaling pathway, and C-type lectin receptor

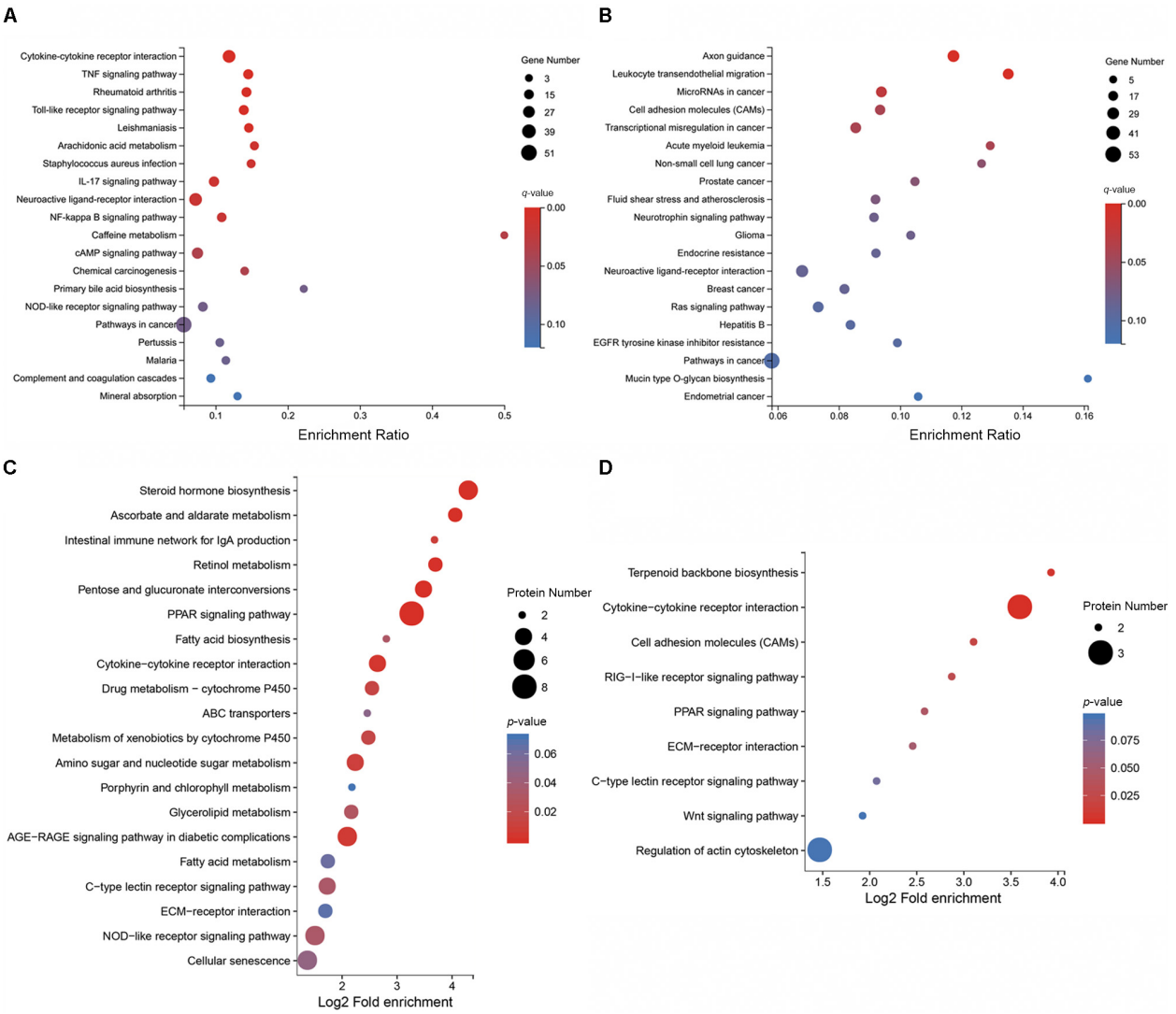


FIGURE 4 | KEGG pathway enrichment analysis of DEGs and DEPs in CSF exposed to MS. KEGG pathway enrichment analysis of the top 20 enriched pathways for upregulated DEGs **(A)**, the top 20 enriched pathways for downregulated DEGs **(B)**, the top 20 enriched pathways for upregulated DEPs **(C)**, and the top 9 enriched pathways for downregulated DEPs **(D)**.

signaling pathway), and environment information processing (i.e., cytokine-cytokine receptor interactions) (**Figure 4C**). Among the downregulated proteins in MS-infected CSF, enriched pathways included cytokine–cytokine receptor interactions, terpenoid backbone biosynthesis, cell adhesion molecules, the RIG-I-like receptor signaling pathway, and the PPAR signaling pathway (**Figure 4D**).

Correlation Analysis of the Transcriptome and Proteome Data in Mock and MS-Infected CSF

As the samples in transcriptomics and proteomics analyses were acquired simultaneously, the correlation between mRNA and protein expression profiles was comprehensively investigated. A total of 5,426 genes were detected at both the transcriptional and translational levels (**Figure 5A** and **Supplementary Table S7**). The Pearson's correlation coefficient was 0.6117 ($p < 0.01$) (**Figure 5B**), indicating a moderate positive correlation between the mRNA and protein level changes. Quantified genes and proteins were categorized into nine groups based on the pattern of changes at the mRNA and protein levels (**Figure 5B**). In the first quadrant (blue dots), the mRNA level was upregulated, but the protein level was downregulated, including three genes. In the second quadrant (purple dots), the mRNA level was upregulated, but the protein level did not change significantly, including 70 genes. In the third quadrant (red dots), both the mRNA and protein levels of 88 genes were upregulated. In the fourth quadrant (pink dots), the mRNA level did not show a clear change, while the

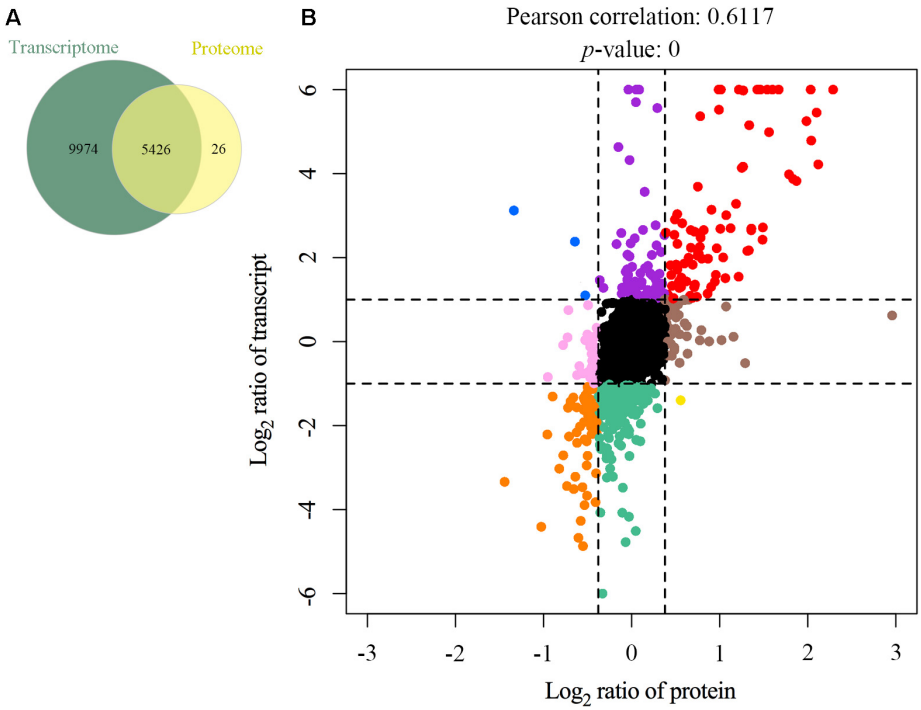


FIGURE 5 | Relationship patterns of all quantitative mRNA and proteins. Comparison of the number of transcribed genes and expressed proteins **(A)**. Nine-quadrant diagram **(B)**. The abscissa represents $\log_2$ expression ratio from proteomic profiling and the ordinate represents $\log_2$ expression ratio from transcriptomic profiling. Black indicates genes and proteins with no significant differences, red indicates upregulated genes and proteins, orange indicates downregulated genes and proteins, blue indicates upregulated genes but downregulated proteins, purple indicates upregulated genes but proteins with no significant differences, pink indicates genes with no significant differences but downregulated proteins, brown indicates genes with no significant differences but upregulated proteins, green indicates downregulated genes but proteins with no significant differences, and yellow indicates downregulated genes but upregulated proteins.

protein level was downregulated, including 33 genes. In the fifth quadrant (black dots), both the mRNA and protein levels of 4,028 genes did not change appreciably. In the sixth quadrant (brown dots), the mRNA level did not show a clear change, while the protein level was upregulated, including 46 genes. In the seventh quadrant (orange dots), both the mRNA and protein levels were synchronously downregulated, including 61 genes. In the eighth quadrant (green dots), the mRNA level was downregulated, but the protein level did not change significantly, including 207 genes. In the ninth quadrant (yellow dots), the mRNA level was downregulated, but the protein level was upregulated, including one gene (**Supplementary Table S8**). Furthermore, the majority of genes congregated in the center of the coordinate, indicating that these genes' expression levels were constant at both the transcriptional and translational levels. The data also revealed significantly discordant regulation at the transcriptional and translational levels (i.e., the first, second, fourth, eight, and ninth quadrants). Interestingly, the directions of mRNA and protein changes of four genes were opposing [i.e., apolipoprotein C3 (*Apoc3*), TNF receptor superfamily member 6b (*Tnfrsf6b*), lysyl oxidase like 1 (*Loxl1*), and apolipoprotein A-I (*Apoa1*)]. The latter could be caused by regulation at several levels, such as post-transcriptional processing, degradation of the transcript, translation, post-translational processing, and modification.

PPI Network of DEPs

To establish the interactions of DEPs, the STRING database was used to construct a PPI network (**Figure 6**). The PPI network consisted of 69 proteins, and the top 5 proteins with higher degree scores were selected as central proteins [i.e., IL-6 (NP_989959.1), MMP-9 (NP_989998.1), cystatin C (CST3, NP_990831.2), APOA1 (NP_990856.1), and cysteine-rich angiogenic inducer 61 (CYR61, NP_001026734.1)] (**Supplementary Table S9**). Among these proteins, IL-6 had the highest degree of connection and interaction with 14 proteins. These 5 proteins, especially IL-6, may play important roles in the development of acute infectious MS-induced arthritis.

DEGs and DEPs Involved in MS-Induced Arthritis

To further understand the mechanism of MS-induced arthritis, the top DEGs and DEPs were further filtered. Based on a review of the literature and the GO and KEGG enrichment analyses, the following key proteins and genes related to MS-induced arthritis were identified: extracellular matrix protein 1 (ECM1), transforming growth factor-β2 (TGF-β2), platelet-derived growth factor receptor alpha (PDGFRA), serum amyloid A (SAA), IL-6, IL-8, IL-1β, neuropilin-1 (NRP1), glutamate-ammonia ligase (GLUL), thrombospondin 2 (THBS2), regulator

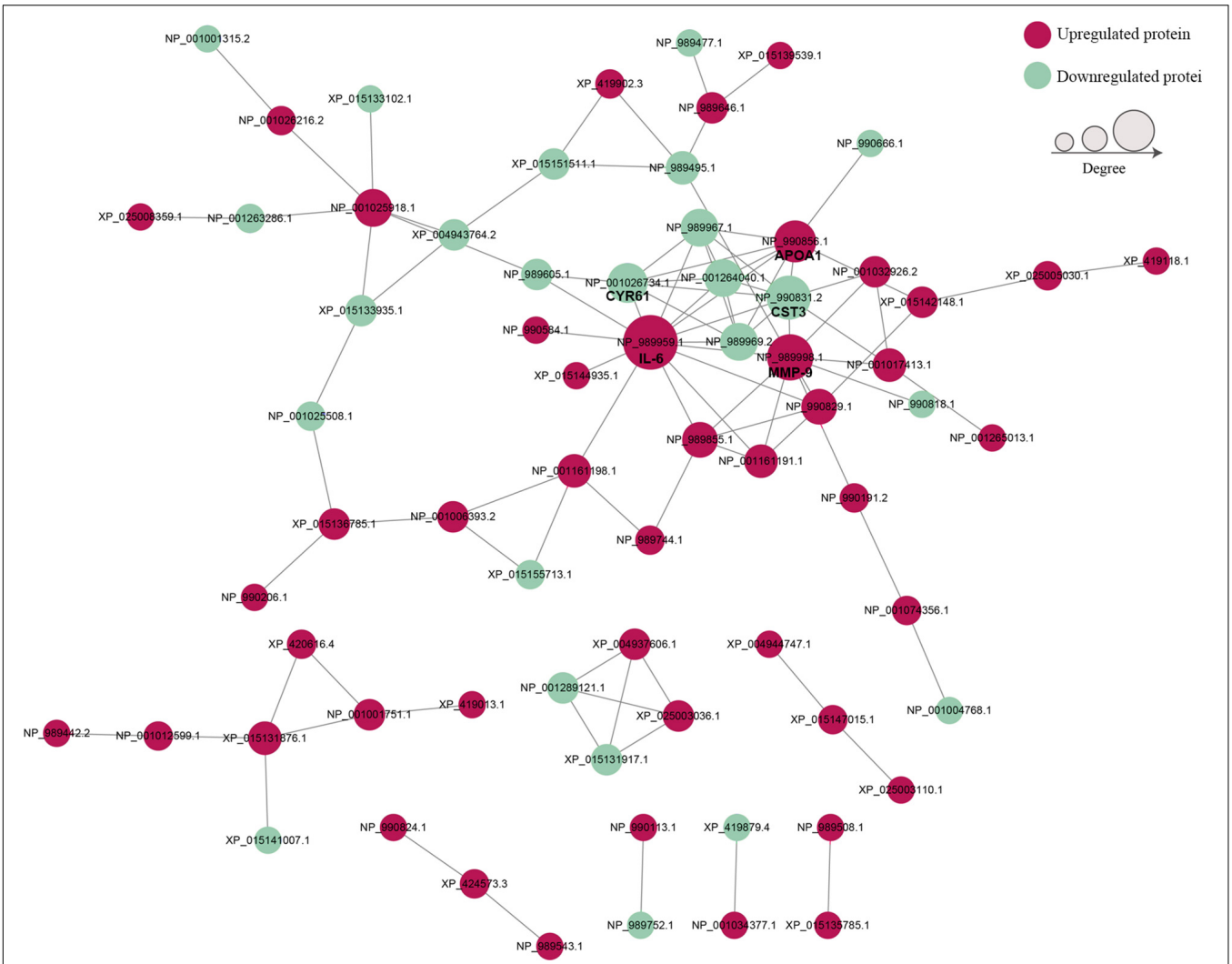


FIGURE 6 | The protein–protein interaction networks of DEPs. Red nodes indicate upregulated proteins and green nodes indicate downregulated proteins. Node size indicates high (large) or low (small) interaction degrees. Network edges represent protein–protein associations.

of cell cycle (RGCC), metalloproteinase inhibitor 3 (TIMP3), matrix metalloproteinases (MMPs), nitric oxide synthase 2 (*Nos2*), mitogen-activated protein kinase 11 (*Mapk11*), caspase 8 (*Casp8*), caspase 3 (*Casp3*), apoptosis-inducing factor (*Aifm1*), nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (*Nfκb1*), htrA serine peptidase 3 (*htrA3*), B-cell CLL/lymphoma 2 (*Bcl2*), *Apoa1*, Complement C1s *(C1s)*, metallothionein 4 (*Mt4*), and *Tlr15* (**Table 1**).

Validation of Gene Expression Levels

To further confirm the transcriptomic and proteomic results, 19 genes were selected for validation by qRT-PCR. Among these selected genes, 14 genes (i.e., *C1s*, *Capn6*, *Ecm1*, *Il-6*, *Il-8*, *Mmp-9*, *Mt4*, *Pdgfra*, *Ptgs2*, *Saa*, *Tgf-β2*, *Thbs2*, *Tlr15*, and *Vnn*) were upregulated, while the other 5 genes (i.e., *Apoa1*, *Cdh11*, *Cst3*, *Ctgf*, and *Timp3*) were downregulated. qRT-PCR results exhibited a similar expression tendency as the transcriptomics analysis, confirming the reliability of the transcriptome sequencing results (**Figure 7**). The trend in the expression levels of *Apoa1* revealed by qRT-PCR was inconsistent with the changes in abundance of its encoded protein detected in the TMT analysis. However, this trend will require further investigation.

DISCUSSION

MS is a major bacterial pathogen affecting the poultry industry worldwide with infections incurring enormous economic losses (Kreizinger et al., 2017; Sun et al., 2017; Xue et al., 2017; Kursa et al., 2019). Clinical signs of MS infection include infectious synovitis, primarily of the hock joints and foot pads (Kleven et al., 1975). Acute infectious arthritis caused by MS is characterized by damage to the joint cavity. In humans, observations have prompted speculation that synovial fibroblasts (SF) in arthritis

TABLE 1 | Critical genes and proteins involved in MS-induced arthritis.

	Gene name	Description	Gene ID	mRNA_log$_2$FC	pep_log$_2$FC	References
Proliferation-regulated factors	*Ecm1*	Extracellular matrix protein 1	107049123	2.13844959	0.754459974	Mongiat et al., 2003
	Tgf-β2	Transforming growth factor beta 2	421352	1.069379238	0.740711733	Allen et al., 1990; Elshabrawy et al., 2015
	Pdgfra	Platelet-derived growth factor receptor alpha	395509	2.757529736	1.358396262	Watanabe et al., 2002; Madarampalli et al., 2019
	Saa	Serum amyloid A	423079	11.92724632	1.599793852	Lee et al., 2006; Hatanaka et al., 2011; Connolly et al., 2016
Apoptosis-related factors	*Nos2*	Nitric oxide synthase 2	395807	3.393184476	NA	Lavric et al., 2007; Lavric et al., 2008; Dusanic et al., 2012
	Mapk11	Mitogen-activated protein kinase 11	417739	−1.955189293	NA	Dusanic et al., 2012
	Casp8	Caspase 8	395284	0.230119267	0.09085343	
	Casp3	Caspase 3	395476	0.062583414	−0.112474729	
	Aifm1	Apoptosis-inducing factor	428688	−0.062502318	0.041242982	
	Nfκb1	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	396033	0.581189669	0.051024003	
	htrA3	htrA serine peptidase 3	422868	0.582529237	0.260627908	
	Bcl2	B-cell CLL/lymphoma 2	396282	−1.128205565	NA	
Inflammatory mediators	*Il-1β*	Interleukin 1β	395196	10.44244078	0.988412026	Narat et al., 1998; Lavric et al., 2007; Dusanic et al., 2014
	Il-6	Interleukin 6	395337	8.876096951	2.033158667	
	Il-8	Interleukin 8	396495	9.695048495	2.287767778	Narat et al., 1998; Bodolay et al., 2002; Lavric et al., 2007; Dusanic et al., 2014
Pro-angiogenic factors	*Nrp1*	Neuropilin-1	395560	1.583486767	0.449957484	Kong et al., 2010
	Glul	Glutamate-ammonia ligase	396489	4.223126274	2.119024103	Eelen et al., 2018
	Thbs2	Thrombospondin 2	414837	1.537056154	0.564622052	Park et al., 2004
Anti-angiogenic factors	*Rgcc*	Regulator of cell cycle	418833	−3.326714915	−1.442222329	An et al., 2009
	Timp3	Metallopeptidase inhibitor 3	396483	−4.398041844	−1.02620507	Chen et al., 2014
MMPs	*Mmp-1*	Matrix metallopeptidase 1	418982	7.037356709	NA	Elshabrawy et al., 2015
	Mmp-2	Matrix metallopeptidase 2	386583	−1.07696931	−0.279283757	Elshabrawy et al., 2015; Cizelj et al., 2016
	Mmp-9	Matrix metallopeptidase 9	395387	5.365580167	0.779049553	Burrage et al., 2006; Elshabrawy et al., 2015
	Mmp-10	Matrix metallopeptidase 10	418981	5.831697437	NA	Elshabrawy et al., 2015
	Mmp-13	Matrix metallopeptidase 13	395683	−1.282777798	NA	
	Mmp-27	Matrix metallopeptidase 27	395850	3.830092934	NA	
Others	*Apoa1*	Apolipoprotein A1	396536	−1.382497329	0.556797247	de Seny et al., 2015
	C1s	Complement C1s	418294	1.975691423	0.867896464	Nakagawa et al., 1999
	Mt4	Metallothionein 4	396212	3.178141448	0.906890596	Youn et al., 2002
	Tlr15	Toll like receptor 15	421219	2.199816288	NA	Oven et al., 2013

NA, not available.

are key cells that drive pathological processes (Ospelt et al., 2004; Bartok and Firestein, 2010; Carmona-Rivera et al., 2017; Falconer et al., 2018). SF can release proinflammatory mediators and chemotactic factors that activate complement factors, resulting in inflammation (Bodolay et al., 2002; Mathew and Ravindran, 2014; Yoshitomi, 2019). SF may be proliferative, migratory, and invasive (Falconer et al., 2018). This ability of SF, along with the impairment of host-defense mechanisms, damages the joint cavity and articular cartilage (Mathew and Ravindran, 2014; Yoshitomi, 2019). According to a previous study, one mechanism of acute infectious arthritis may be direct microbial invasion (Mathew and Ravindran, 2014). In the acute phase of infection, live MS is detected in the SF fluids of infected birds, implying that MS may come into direct contact with CCH and CSF (Kerr and Olson, 1970; Dusanic et al., 2012).

A better understanding of host-bacteria interactions would elucidate the molecular mechanism of MS-induced arthritis. The application of high-throughput methods on MS-induced arthritis could lead to the discovery of pathogenesis-related genes and proteins. However, high-throughput methods have not been applied in most studies of host-MS interactions, with the exception of two studies that utilized an avian macrophage

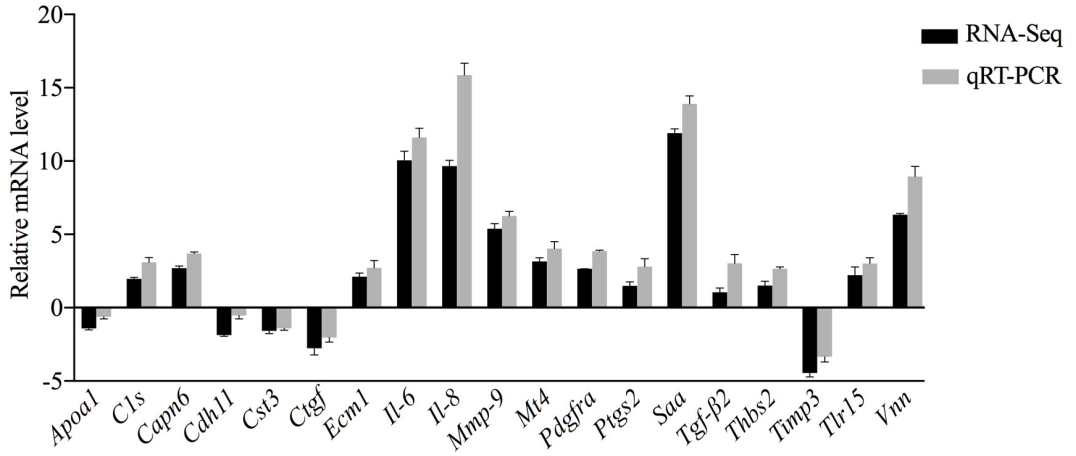


FIGURE 7 | Comparative analysis of the qRT-PCR and transcriptome sequencing results of DEGs in CSF exposed to MS.

microarray/avian innate immunity microarray (Lavric et al., 2008) and a phenotype microarray (Dusanic et al., 2014). To uncover the molecular mechanisms of MS-induced arthritis, primary CSF were prepared by the explanted tissue culture method in this study (**Supplementary Figure S1**). Then, the DEGs and DEPs in MS-infected CSF were investigated, and critical genes involved in arthritis were identified by transcriptome and proteome analyses.

Recent studies demonstrated that MS-invaded host cells incubated and altered the gene expression of these cells after 24, 48, or 72 h (Dusanic et al., 2009; Buim et al., 2011; Dusanic et al., 2012, 2014; Cizelj et al., 2016). Combining these studies and comprehensive analysis, we employed high-throughput methods to analyze the gene expression of MS-infected CSF 48 hpi. By utilizing DEGs and DEPs and drawing on previous research on arthritis, this study aimed to identify critical genes or proteins involved in arthritis caused by MS infection in chickens. In total, 15,400 genes and 5,537 proteins in MS-treated CSF were identified (**Supplementary Tables S3, S5**). Of which, 1,137 genes and 129 proteins were upregulated, while the remaining 1,210 genes and 92 proteins were downregulated (**Figure 1** and **Supplementary Tables S4, S6**). For the most part, the correlation analysis of the transcriptome and proteome data revealed that the changes in proteins were consistent with the transcriptome results. Although no difference was detected between the transcriptional and translational regulation of most of the identified genes, four genes opposed this trend (i.e., *Apoc3*, *Tnfrsf6b*, *Loxl1*, and *Apoa1*), which may be affected by post-transcriptional regulation. However, further research is needed to investigate this issue. Additionally, 149 genes (upregulated, $n = 88$; downregulated, $n = 61$) were significantly differentially expressed at both the transcriptional and translational levels in the MS-infected group (**Figure 4B** and **Supplementary Table S7**). To further extract related information on MS-induced infectious synovitis and arthritis, 149 genes were analyzed. Although a direct answer does not exist, the relevant literature and GO/KEGG enrichment analyses provide some clues.

A previous study reported hypertrophy and hyperplasia of synovial cells in the joints of MS-infected chickens with fibroplasia and acute inflammatory cells causing the synovial membrane to thicken (Kerr and Olson, 1970). This pathological phenomenon is possibly due to the hyperproliferation and defective apoptosis of SF (Zhang et al., 2009; Wilkinson and Henley, 2010). In the present study, several proteins, including ECM1 (Mongiat et al., 2003), TGF-β2 (Allen et al., 1990), PDGFRA (Watanabe et al., 2002; Madarampalli et al., 2019), and SAA (Lee et al., 2006; Hatanaka et al., 2011), that regulate the proliferation of human arthritis were upregulated at both the mRNA and protein levels. As these genes may also increase the proliferative capacity of CSF, their overexpression may explain CSF hyperplasia, as well as the altered resistance to apoptosis. A previous study demonstrated that apoptosis-induced death of MS-infected CCH was a result of the upregulation of several pro-apoptotic genes, including *Nos2*, *Mapk1*, *Casp8*, *Casp3*, *Aifm1*, *Nfκb1*, *htrA3*, and *Bcl2* (Dusanic et al., 2012). However, with the exception of *Nos2*, the transcription of these genes in CSF was unchanged or downregulated in this study. Dysregulation of apoptosis and the proliferation of CSF due to multiple genes may play roles in this hyperplasia.

Previous studies reported human SF-mediated inflammation and autoimmunity via the production of inflammatory mediators in rheumatoid arthritis (RA) (McInnes and Schett, 2011; Bustamante et al., 2017). In the present study, the expression of various inflammatory mediators, including *Il-6*, *Il-8*, *Il-1*β, and *Nos2*, increased in MS-infected CSF, which matches the findings of previous studies on MS. Lavric et al. (2007, 2008) found that viable MS and MS proteins stimulated chicken monocyte-derived macrophages and HD11, and CCH secreted IL-6, IL-8, and IL-1β. In a previous study that utilized phenotype microarrays to investigate the effects of MS on the global metabolic activity of CCH, MS-infected CCH were found to be sensitive to IL-6, IL-1β, IL-8, TNF-α, and IFN-γ (Dusanic et al., 2014). Previous studies also found that human SF secreted various cytokines and that cytokine networks at inflammatory sites largely contributed to the

pathogenesis of arthritis and the perpetuation of inflammation (Bartok and Firestein, 2010; Emery, 2012; Kihara et al., 2017). Exposure to these cytokines led to chrondrocyte deprivation in cartilage over time (McInnes and Schett, 2011). Another study found that through the secretion of cytokines, human SF played a role in the persistence of inflammation in the synovium via the recruitment and retention of effector cells in the immune system (Ritchlin, 2000). Thus, it is likely that proinflammatory cytokines upregulated in CSF play a role in the pathogenesis of MS-induced arthritis.

In addition to the production of inflammatory cytokines, another characteristic of synovium inflammation in RA is angiogenesis, which appears to be necessary for pannus development (MacDonald et al., 2018). The expansive synovial tissue, "pannus," which is found at the cartilage-bone interface, cloaks the cartilage and erodes it into bone (Bartok and Firestein, 2010). The ability of RASF to secrete a number of proangiogenic factors suggests that RASF also contributes to blood vessel growth, which is necessary for sustaining pannus formation and the arthritis process (Mor et al., 2005). An investigation conducted by Kerr and Olson (1970) observed the formation of granulomatous pannus in the medullary spaces of chickens inoculated experimentally or contact-infected with MS. The present study demonstrated that several proangiogenic factors were upregulated, including ECM1 (Mongiat et al., 2003), IL-8 (Bodolay et al., 2002), TGF-β2 (Elshabrawy et al., 2015), SAA (Connolly et al., 2016), MMP-9 (Burrage et al., 2006), NRP1 (Kong et al., 2010), GLUL (Eelen et al., 2018), and THBS2 (Park et al., 2004), while antiangiogenic factors, including RGCC (An et al., 2009), as well as TIMP3 (Chen et al., 2014), were downregulated at both the mRNA and protein levels. These findings suggest that MS-infected CSF may contribute to blood vessel growth, which is necessary for sustaining pannus formation and the arthritis process.

The growth of new blood vessels requires the breakdown of the surrounding ECM. Principal enzymes involved in this process include MMPs (Chen et al., 2014), as well as a family of zinc-containing, calcium-dependent proteinases that degrade major components of ECM (Murphy et al., 2002). Although collagenases (i.e., MMP-1, -8, and -13), gelatinases A and B (i.e., MMP-2 and -9), stromelysins (i.e., MMP-3, -10, and -11), and matrilysins (i.e., MMP-7 and -26) are expressed at low levels in normal human joint tissues, they are highly expressed in arthritic joints (Elshabrawy et al., 2015). In RA, cartilage destruction depends on the balance between MMPs and TIMPs. When the balance favors MMPs, cartilage degradation proceeds (Yoshitomi, 2019). In this study, *Mmp-1*, *-9*, *-10*, and *-27* were upregulated, while *Mmp-2* and *-13* were downregulated at the mRNA level. The MMP-9 protein was upregulated and TIMP was downregulated at the protein level. These findings suggest that CSF may contribute to MS-induced arthritis via MMPs.

Additionally, the expression of certain genes changed at the mRNA and/or protein levels, including *Apoa1* (de Seny et al., 2015), *C1s* (Nakagawa et al., 1999), *Mt4* (Youn et al., 2002), and *Tlr15* (Oven et al., 2013). A review of the literature revealed that these genes may participate in the pathogenesis of MS-induced infectious synovitis in chickens.

CONCLUSION

In this study, the gene and protein expression changes caused by MS strain HN01 infection in CSF were investigated. Several critical DEGs and DEPs involved in MS-induced arthritis were identified. These genes and their protein products in infected CSF consisted of proliferation- and apoptosis-related factors, inflammatory mediators, proangiogenic factors, antiangiogenic factors, and MMPs, as well as other arthritis-related proteins, indicating that these genes may play important roles in MS-induced arthritis. The integrated analysis of transcriptome and proteome data advances our understanding of the interplay between CSF and MS. However, future work is needed to elucidate the interactions observed therein.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI Sequence Read Archive: SRR10479009, SRR10479010, SRR10479011, SRR10479012, SRR10479013, and SRR10479014. ProteomeXchange Consortium: PXD016513.

ETHICS STATEMENT

The animal experiment was carried out in Jiangsu Academy of Agricultural Sciences with the approval of the Committee on the Ethics of Animal Experiments of (JAAS No. 20141107). The protocol conformed to the guidelines of Jiangsu Province Animal Regulations (Government Decree No. 45) in accordance with international law.

AUTHOR CONTRIBUTIONS

RL conceived and designed the experiments, carried out the experiments, analyzed the data, and wrote the manuscript. BX participated in performing experiments. SY participated in revising the manuscript. JZ, HS, CL, FL, and QP participated in analyzing and interpreting the data. XZ supervised and guided this work. All the authors have read and approved the final version of the manuscript.

FUNDING

This work was supported by the National Key Research and Development Program of China (2017YFD0500706), China Postdoctoral Science Foundation (2018M642194), and Jiangsu Planned Projects for Postdoctoral Research Funds (2018K204C).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00576/full#supplementary-material

REFERENCES

Addis, M. F., Pisanu, S., Ghisaura, S., Pagnozzi, D., Marogna, G., Tanca, A., et al. (2011). Proteomics and pathway analyses of the milk fat globule in sheep naturally infected by *Mycoplasma agalactiae* provide indications of the *in vivo* response of the mammary epithelium to bacterial infection. *Infect. Immun.* 79, 3833–3845. doi: 10.1128/IAI.00040-11

Allen, J. B., Manthey, C. L., Hand, A. R., Ohura, K., Ellingsworth, L., and Wahl, S. M. (1990). Rapid onset synovial inflammation and hyperplasia induced by transforming growth factor beta. *J. Exp. Med.* 171, 231–247. doi: 10.1084/jem.171.1.231

An, X., Jin, Y., Guo, H., Foo, S. Y., Cully, B. L., Wu, J., et al. (2009). Response gene to complement 32, a novel hypoxia-regulated angiogenic inhibitor. *Circulation* 120, 617–627. doi: 10.1161/CIRCULATIONAHA.108.841502

Ball, C., Forrester, A., and Ganapathy, K. (2018). Co-circulation of genetically diverse population of vaccine related and unrelated respiratory mycoplasmas and viruses in UK poultry flocks with health or production problems. *Vet. Microbiol/* 225, 132–138. doi: 10.1016/j.vetmic.2018.09.009

Bartok, B., and Firestein, G. S. (2010). Fibroblast-like synoviocytes: key effector cells in rheumatoid arthritis. *Immunol. Rev.* 233, 233–255. doi: 10.1111/j.0105-2896.2009.00859.x

Bercic, R. L., Cizelj, I., Dusanic, D., Narat, M., Zorman-Rojs, O., Dovc, P., et al. (2011). Neuraminidase of *Mycoplasma synoviae* desialylates heavy chain of the chicken immunoglobulin G and glycoproteins of chicken tracheal mucus. *Avian Pathol.* 40, 299–308. doi: 10.1080/03079457.2011.565311

Bodolay, E., Koch, A. E., Kim, J., Szegedi, G., and Szekanecz, Z. (2002). Angiogenesis and chemokines in rheumatoid arthritis and other systemic inflammatory rheumatic diseases. *J. Cell Mol. Med.* 6, 357–376. doi: 10.1111/j.1582-4934.2002.tb00514.x

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Buim, M. R., Buzinhani, M., Yamaguti, M., Oliveira, R. C., Mettifogo, E., Ueno, P. M., et al. (2011). *Mycoplasma synoviae* cell invasion: elucidation of the *Mycoplasma* pathogenesis in chicken. *Comp. Immunol. Microbiol. Infect. Dis.* 34, 41–47. doi: 10.1016/j.cimid.2009.11.001

Burrage, P. S., Mix, K. S., and Brinckerhoff, C. E. (2006). Matrix metalloproteinases: role in arthritis. *Front. Biosci.* 11:529–543. doi: 10.2741/1817

Bustamante, M. F., Garcia-Carbonell, R., Whisenant, K. D., and Guma, M. (2017). Fibroblast-like synoviocyte metabolism in the pathogenesis of rheumatoid arthritis. *Arthritis Res. Ther.* 19:110. doi: 10.1186/s13075-017-1303-3

Carmona-Rivera, C., Carlucci, P. M., Moore, E., Lingampalli, N., Uchtenhagen, H., James, E., et al. (2017). Synovial fibroblast-neutrophil interactions promote pathogenic adaptive immunity in rheumatoid arthritis. *Sci. Immunol.* 2:eaag3358. doi: 10.1126/sciimmunol.aag3358

Catania, S., Bilato, D., Gobbo, F., Granato, A., Terregino, C., Iob, L., et al. (2010). Treatment of eggshell abnormalities and reduced egg production caused by *Mycoplasma synoviae* infection. *Avian Dis.* 54, 961–964. doi: 10.1637/9121-110309-Case.1

Catania, S., Gobbo, F., Bilato, D., Gagliazzo, L., Moronato, M. L., Terregino, C., et al. (2016a). Two strains of *Mycoplasma synoviae* from chicken flocks on the same layer farm differ in their ability to produce eggshell apex abnormality. *Vet. Microbiol.* 193, 60–66. doi: 10.1016/j.vetmic.2016.08.007

Catania, S., Gobbo, F., Ramirez, A. S., Guadagnini, D., Baldasso, E., Moronato, M. L., et al. (2016b). Laboratory investigations into the origin of *Mycoplasma synoviae* isolated from a lesser flamingo (*Phoeniconaias minor*). *BMC Vet. Res.* 12:52. doi: 10.1186/s12917-016-0680-1

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., et al. (2018). SOAPnuke: a mapreduce acceleration supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* 7:gix120. doi: 10.1093/gigascience/gix120

Chen, Y. Y., Brown, N. J., Jones, R., Lewis, C. E., Mujamammi, A. H., Muthana, M., et al. (2014). A peptide derived from TIMP-3 inhibits multiple angiogenic growth factor receptors and tumour growth and inflammatory arthritis in mice. *Angiogenesis* 17, 207–219. doi: 10.1007/s10456-013-9389-y

Cizelj, I., Bercic, R. L., Dusanic, D., Narat, M., Kos, J., Dovc, P., et al. (2011). *Mycoplasma gallisepticum* and *Mycoplasma synoviae* express a cysteine protease CysP, which can cleave chicken IgG into Fab and Fc. *Microbiology* 157(Pt 2), 362–372. doi: 10.1099/mic.0.045641-0

Cizelj, I., Dusanic, D., Bencina, D., and Narat, M. (2016). Mycoplasma and host interaction: *In vitro* gene expression modulation in *Mycoplasma synoviae* and infected chicken chondrocytes. *Acta Vet. Hung.* 64, 26–37. doi: 10.1556/004.2016.003

Connolly, M., Rooney, P. R., McGarry, T., Maratha, A. X., McCormick, J., Miggin, S. M., et al. (2016). Acute serum amyloid A is an endogenous TLR2 ligand that mediates inflammatory and angiogenic mechanisms. *Ann. Rheum Dis.* 75, 1392–1398. doi: 10.1136/annrheumdis-2015-207655

de Seny, D., Cobraiville, G., Charlier, E., Neuville, S., Lutteri, L., Le Goff, C., et al. (2015). Apolipoprotein-A1 as a damage-associated molecular patterns protein in osteoarthritis: ex vivo and in vitro pro-inflammatory properties. *PLoS One* 10:e0122904. doi: 10.1371/journal.pone.0122904

Dewey, C. N., and Li, B. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *Bmc Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323

Dusanic, D., Bencina, D., Narat, M., and Oven, I. (2014). Phenotypic characterization of *Mycoplasma synoviae* induced changes in the metabolic and sensitivity profile of in vitro infected chicken chondrocytes. *Biomed Res. Int.* 2014:613730. doi: 10.1155/2014/613730

Dusanic, D., Bencina, D., Oven, I., Cizelj, I., Bencina, M., and Narat, M. (2012). *Mycoplasma synoviae* induces upregulation of apoptotic genes, secretion of nitric oxide and appearance of an apoptotic phenotype in infected chicken chondrocytes. *Vet. Res.* 43:7. doi: 10.1186/1297-9716-43-7

Dusanic, D., Bercic, R. L., Cizelj, I., Salmic, S., Narat, M., and Bencina, D. (2009). *Mycoplasma synoviae* invades non-phagocytic chicken cells in vitro. *Vet. Microbiol.* 138, 114–119. doi: 10.1016/j.vetmic.2009.02.014

Eelen, G., Dubois, C., Cantelmo, A. R., Goveia, J., Bruning, U., DeRan, M., et al. (2018). Role of glutamine synthetase in angiogenesis beyond glutamine synthesis. *Nature* 561, 63–69. doi: 10.1038/s41586-018-0466-7

Elshabrawy, H. A., Chen, Z., Volin, M. V., Ravella, S., Virupannavar, S., and Shahrara, S. (2015). The pathogenic role of angiogenesis in rheumatoid arthritis. *Angiogenesis* 18, 433–448. doi: 10.1007/s10456-015-9477-2

Emery, P. (2012). Optimizing outcomes in patients with rheumatoid arthritis and an inadequate response to anti-TNF treatment. *Rheumatology* 51(Suppl. 5), v22–v30. doi: 10.1093/rheumatology/kes115

Falconer, J., Murphy, A. N., Young, S. P., Clark, A. R., Tiziani, S., Guma, M., et al. (2018). Review: synovial cell metabolism and chronic inflammation in rheumatoid arthritis. *Arthritis Rheumatol.* 70, 984–999. doi: 10.1002/art.40504

Feberwee, A., de Wit, J. J., and Landman, W. J. (2009). Induction of eggshell apex abnormalities by *Mycoplasma synoviae*: field and experimental studies. *Avian Pathol.* 38, 77–85. doi: 10.1080/03079450802662772

Fletcher, O. J., Anderson, D. P., and Kleven, S. H. (1976). Histology of air sac lesions induced in chickens by contact exposure to *Mycoplasma synoviae*. *Vet. Pathol.* 13, 303–314. doi: 10.1177/030098587601300407

Gerchman, I., Lysnyansky, I., Perk, S., and Levisohn, S. (2008). In vitro susceptibilities to fluoroquinolones in current and archived *Mycoplasma gallisepticum* and *Mycoplasma synoviae* isolates from meat-type turkeys. *Vet. Microbiol.* 131, 266–276. doi: 10.1016/j.vetmic.2008.04.006

Goret, J., Beven, L., Faustin, B., Contin-Bordes, C., Le Roy, C., Claverol, S., et al. (2017). Interaction of Mycoplasma hominis PG21 with Human Dendritic Cells: Interleukin-23-Inducing Mycoplasmal Lipoproteins and Inflammasome Activation of the Cell. *J. Bacteriol.* 199:e00213-17. doi: 10.1128/JB.00213-17

Hatanaka, E., Dermargos, A., Armelin, H. A., Curi, R., and Campa, A. (2011). Serum amyloid A induces reactive oxygen species (ROS) production and proliferation of fibroblast. *Clin. Exp. Immunol.* 163, 362–367. doi: 10.1111/j.1365-2249.2010.04300.x

Hirohata, S., Yanagida, T., Tomita, T., and Yoshikawa, H. (2009). Differential influences of bucillamine and methotrexate on the generation of fibroblast-like cells from bone marrow CD34+ cells of rheumatoid arthritis patients. *Int. Immunopharmacol.* 9, 86–90. doi: 10.1016/j.intimp.2008.10.007

Kerr, K. M., and Olson, N. O. (1970). Pathology of chickens inoculated experimentally or contact-infected with *Mycoplasma synoviae*. *Avian Dis.* 14, 290–320.

Kihara, M., Davies, R., Kearsley-Fleet, L., Watson, K. D., Lunt, M., Symmons, D. P., et al. (2017). Use and effectiveness of tocilizumab among patients with rheumatoid arthritis: an observational study from the British society for rheumatology biologics register for rheumatoid arthritis. *Clin. Rheumatol.* 36, 241–250. doi: 10.1007/s10067-016-3485-5

Kleven, S. H., Fletcher, O. J., and Davis, R. B. (1975). Influence of strain of *Mycoplasma synoviae* and route of infection on development of synovitis or airsacculitis in broilers. *Avian Dis.* 19, 126–135.

Kong, J. S., Yoo, S. A., Kim, J. W., Yang, S. P., Chae, C. B., Tarallo, V., et al. (2010). Anti-neuropilin-1 peptide inhibition of synoviocyte survival, angiogenesis, and experimental arthritis. *Arthritis Rheum.* 62, 179–190. doi: 10.1002/art.27243

Kordafshari, S., Marenda, M. S., O'Rourke, D., Shil, P., and Noormohammadi, A. H. (2019). Mutation of oppF gene in the *Mycoplasma synoviae* MS-H vaccine strain and its implication for differential serological responses to vaccination versus field challenge. *Vet. Microbiol.* 231, 48–55. doi: 10.1016/j.vetmic.2019.02.029

Kreizinger, Z., Grozner, D., Sulyok, K. M., Nilsson, K., Hrivnak, V., Bencina, D., et al. (2017). Antibiotic susceptibility profiles of *Mycoplasma synoviae* strains originating from Central and Eastern Europe. *BMC Vet. Res.* 13:342. doi: 10.1186/s12917-017-1266-2

Kursa, O., Pakula, A., Tomczyk, G., Pasko, S., and Sawicka, A. (2019). Eggshell apex abnormalities caused by two different *Mycoplasma synoviae* genotypes and evaluation of eggshell anomalies by full-field optical coherence tomography. *BMC Vet. Res.* 15:1. doi: 10.1186/s12917-018-1758-8

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lavric, M., Bencina, D., Kothlow, S., Kaspers, B., and Narat, M. (2007). *Mycoplasma synoviae* lipoprotein MSPB, the N-terminal part of VlhA haemagglutinin, induces secretion of nitric oxide, IL-6 and IL-1beta in chicken macrophages. *Vet. Microbiol.* 121, 278–287. doi: 10.1016/j.vetmic.2006.12.005

Lavric, M., Maughan, M. N., Bliss, T. W., Dohms, J. E., Bencina, D., Keeler, C. L., et al. (2008). Gene expression modulation in chicken macrophages exposed to *Mycoplasma synoviae* or *Escherichia coli*. *Vet. Microbiol.* 126, 111–121. doi: 10.1016/j.vetmic.2007.06.011

Lee, M. S., Yoo, S. A., Cho, C. S., Suh, P. G., Kim, W. U., and Ryu, S. H. (2006). Serum amyloid A binding to formyl peptide receptor-like 1 induces synovial hyperplasia and angiogenesis. *J. Immunol.* 177, 5585–5594. doi: 10.4049/jimmunol.177.8.5585

Lorenc, Z., Pasko, S., Kursa, O., Pakula, A., and Salbut, L. (2019). Spectral technique for detection of changes in eggshells caused by *Mycoplasma synoviae*. *Poult. Sci.* 98, 3481–3487. doi: 10.3382/ps/pez150

Ma, J., Chen, T., Wu, S., Yang, C., Bai, M., Shu, K., et al. (2019). iProX: an integrated proteome resource. *Nucleic Acids Res.* 47, D1211–D1217. doi: 10.1093/nar/gky869

MacDonald, I. J., Liu, S. C., Su, C. M., Wang, Y. H., Tsai, C. H., and Tang, C. H. (2018). Implications of angiogenesis involvement in arthritis. *Int. J. Mol. Sci.* 19:2012. doi: 10.3390/ijms19072012

Madarampalli, B., Watts, G. F. M., Panipinto, P. M., Nguygen, H. N., Brenner, M. B., and Noss, E. H. (2019). Interactions between cadherin-11 and platelet-derived growth factor receptor-alpha signaling link cell adhesion and proliferation. *Biochim. Biophys. Acta Mol. Basis Dis.* 1865, 1516–1524. doi: 10.1016/j.bbadis.2019.03.001

Mathew, A. J., and Ravindran, V. (2014). Infections and arthritis. *Best Pract. Res. Clin. Rheumatol.* 28, 935–959. doi: 10.1016/j.berh.2015.04.009

McInnes, I. B., and Schett, G. (2011). The pathogenesis of rheumatoid arthritis. *N. Engl. J. Med.* 365, 2205–2219. doi: 10.1056/NEJMra1004965

Michiels, T., Welby, S., Vanrobaeys, M., Quinet, C., Rouffaer, L., Lens, L., et al. (2016). Prevalence of *Mycoplasma gallisepticum* and *Mycoplasma synoviae* in commercial poultry, racing pigeons and wild birds in Belgium. *Avian Pathol.* 45, 244–252. doi: 10.1080/03079457.2016.1145354

Mongiat, M., Fu, J., Oldershaw, R., Greenhalgh, R., Gown, A. M., and Iozzo, R. V. (2003). Perlecan protein core interacts with extracellular matrix protein 1 (ECM1), a glycoprotein involved in bone formation and angiogenesis. *J. Biol. Chem.* 278, 17491–17499. doi: 10.1074/jbc.M210529200

Mor, A., Abramson, S. B., and Pillinger, M. H. (2005). The fibroblast-like synovial cell in rheumatoid arthritis: a key player in inflammation and joint destruction. *Clin. Immunol.* 115, 118–128. doi: 10.1016/j.clim.2004.12.009

Murphy, G., Knauper, V., Atkinson, S., Butler, G., English, W., Hutton, M., et al. (2002). Matrix metalloproteinases in arthritic disease. *Arthritis Res.* 4(Suppl. 3), S39–S49. doi: 10.1186/ar572

Nakagawa, K., Sakiyama, H., Tsuchida, T., Yamaguchi, K., Toyoguchi, T., Masuda, R., et al. (1999). Complement C1s activation in degenerating articular cartilage of rheumatoid arthritis patients: immunohistochemical studies with an active form specific antibody. *Ann. Rheum. Dis.* 58, 175–181. doi: 10.1136/ard.58.3.175

Narat, M., Bencina, D., Kleven, S. H., and Habe, F. (1998). The hemagglutination-positive phenotype of *Mycoplasma synoviae* induces experimental infectious synovitis in chickens more frequently than does the hemagglutination-negative phenotype. *Infect. Immun.* 66, 6004–6009. doi: 10.1128/iai.66.12.6004-6009.1998

Ospelt, C., Neidhart, M., Gay, R. E., and Gay, S. (2004). Synovial activation in rheumatoid arthritis. *Front. Biosci.* 9:2323–2334. doi: 10.2741/1399

Oven, I., Resman, Rus K, Dušanić, D., Benčina, D., Keeler, CL Jr, and Narat, M. (2013). Diacylated lipopeptide from *Mycoplasma synoviae* mediates TLR15 induced innate immune responses. *Vet. Res.* 44:99. doi: 10.1186/1297-9716-44-99

Park, Y. W., Kang, Y. M., Butterfield, J., Detmar, M., Goronzy, J. J., and Weyand, C. M. (2004). Thrombospondin 2 functions as an endogenous regulator of angiogenesis and inflammation in rheumatoid arthritis. *Am. J. Pathol.* 165, 2087–2098. doi: 10.1016/S0002-9440(10)63259-2

Ritchlin, C. (2000). Fibroblast biology: effector signals released by the synovial fibroblast in arthritis. *Arthritis Res.* 2, 356–360.

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73

Stemke, G. W., and Robertson, J. A. (1982). Comparison of two methods for enumeration of mycoplasmas. *J. Clin. Microbiol.* 16, 959–961. doi: 10.1128/jcm.16.5.959-961.1982

Sun, S. K., Lin, X., Chen, F., Wang, D. A., Lu, J. P., Qin, J. P., et al. (2017). Epidemiological investigation of *Mycoplasma Synoviae* in native chicken breeds in China. *BMC Vet. Res.* 13:115. doi: 10.1186/s12917-017-1029-0

Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* 11, 2301–2319. doi: 10.1038/nprot.2016.136

Vasconcelos, A. T., Ferreira, H. B., Bizarro, C. V., Bonatto, S. L., Carvalho, M. O., Pinto, P. M., et al. (2005). Swine and poultry pathogens: the complete genome sequences of two strains of *Mycoplasma hyopneumoniae* and a strain of *Mycoplasma synoviae*. *J. Bacteriol.* 187, 5568–5577. doi: 10.1128/JB.187.16.5568-5577.2005

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138. doi: 10.1093/bioinformatics/btp612

Wang, Q., Yuan, X., Chen, Y., Zheng, Q., Xu, L., and Wu, Y. (2018). Endoplasmic reticulum stress mediated MDRV p10.8 *Protein-Induced Cell Cycle Arrest and Apoptosis Through the PERK/eIF2* alpha Pathway. *Front. Microbiol.* 9:e01327. doi: 10.3389/fmicb.2018.01327

Watanabe, N., Ando, K., Yoshida, S., Inuzuka, S., Kobayashi, M., Matsui, N., et al. (2002). Gene expression profile analysis of rheumatoid synovial fibroblast cultures revealing the overexpression of genes responsible for tumor-like growth of rheumatoid synovium. *Biochem. Biophys. Res. Commun.* 294, 1121–1129. doi: 10.1016/S0006-291X(02)00608-3

Wilkinson, K. A., and Henley, J. M. (2010). Mechanisms, regulation and consequences of protein SUMOylation. *Biochem. J.* 428, 133–145. doi: 10.1042/bj20100158

Wu, Q., Xu, X., Chen, Q., Zuo, K., Zhou, Y., Zhang, Z., et al. (2019). Rapid and visible detection of *Mycoplasma synoviae* using a novel polymerase spiral reaction assay. *Poult. Sci.* 98, 5355–5360. doi: 10.3382/ps/pez356

Xue, J., Xu, M. Y., Ma, Z. J., Zhao, J., Jin, N., and Zhang, G. Z. (2017). Serological investigation of *Mycoplasma synoviae* infection in China from 2010 to 2015. *Poult. Sci.* 96, 3109–3112. doi: 10.3382/ps/pex134

Yoshitomi, H. (2019). Regulation of immune responses and chronic inflammation by fibroblast-like synoviocytes. *Front. Immunol.* 10:1395. doi: 10.3389/fimmu.2019.01395

Youn, J., Hwang, S. H., Ryoo, Z. Y., Lynes, M. A., Paik, D. J., Chung, H. S., et al. (2002). Metallothionein suppresses collagen-induced arthritis via induction of TGF-beta and down-regulation of proinflammatory mediators. *Clin. Exp. Immunol.* 129, 232–239. doi: 10.1046/j.1365-2249.2002.01922.x

Zhang, Q., Wu, J., Cao, Q., Xiao, L., and Li, N. (2009). A critical role of Cyr61 in interleukin-17-Dependent proliferation of fibroblast-like synoviocytes in rheumatoid arthritis. *Arthritis Rheum.* 60, 3602–3612. doi: 10.1002/art.24999

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 09 April 2020
doi: 10.3389/fmicb.2020.00654

Dynamic Changes in the Global Transcriptome and MicroRNAome Reveal Complex miRNA-mRNA Regulation in Early Stages of the Bi-Directional Development of *Echinococcus granulosus* Protoscoleces

Yun Bai[1,2], Zhuangzhi Zhang[3], Lei Jin[2], Yongqiang Zhu[2], Li Zhao[3], Baoxin Shi[3], Jun Li[4], Gang Guo[4], Baoping Guo[4], Donald P. McManus[5], Shengyue Wang[2,6] and Wenbao Zhang[4]**

[1] Central Laboratory, Shanghai Skin Disease Hospital, Tongji University School of Medicine, Shanghai, China, [2] Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, China, [3] Veterinary Research Institute, Xinjiang Academy of Animal Sciences, Urumqi, China, [4] State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, Clinical Medical Research Institute, The First Affiliated Hospital of Xinjiang Medical University, Urumqi, China, [5] Molecular Parasitology Laboratory, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia, [6] National Research Center for Translational Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Background: Cystic echinococcosis is a life-threatening disease caused by the larval stages of the dog tapeworm *Echinococcus granulosus*. Protoscoleces (PSCs) of this worm have the ability of bi-directional development to either larval cysts or strobilar adult worms. However, the molecular mechanisms underlying this development process are unknown.

Results: RNA and small RNAs sequencing was employed to characterize the gene and miRNA expression at 0–24 h and 7–14 days in the bi-directional development of PSCs. A total of 963 genes and 31 miRNAs were differentially expressed in the early development of PSCs to adult worms whereas 972 genes and 27 miRNAs were differentially expressed in the early development of PSCs to cysts. Pairwise comparison between the two developmental patterns showed that 172 genes and 15 miRNAs were differentially expressed at three time-points. Most of these genes were temporally changed at 24 h or 7 days. GO enrichment analysis revealed that the differentially expressed genes in early adult worm development are associated with nervous system development and carbohydrate metabolic process; whereas, the differentially expressed genes in early cystic development are associated with transmembrane transporter activity and nucleoside triphosphatase activity. In addition, miR-71 and miR-219 regulated genes are likely involved in oxidation reduction in adult worm development.

OPEN ACCESS

Edited by:
Diego Robledo, The University of Edinburgh, United Kingdom

Reviewed by:
Majid Fasihi Harandi, Kerman University of Medical Sciences, Iran
Chafia Touil-Boukoffa, University of Sciences and Technology Houari Boumediene (USTHB), Algeria

***Correspondence:**
Wenbao Zhang
wenbaozhang2013@163.com
Shengyue Wang
wsy12115@rjh.con.cn

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *17 September 2019*
Accepted: *23 March 2020*
Published: *09 April 2020*

Citation:
Bai Y, Zhang Z, Jin L, Zhu Y, Zhao L, Shi B, Li J, Guo G, Guo B, McManus DP, Wang S and Zhang W (2020) Dynamic Changes in the Global Transcriptome and MicroRNAome Reveal Complex miRNA-mRNA Regulation in Early Stages of the Bi-Directional Development of Echinococcus granulosus Protoscoleces. Front. Microbiol. 11:654. doi: 10.3389/fmicb.2020.00654

Conclusion: The early stages of bi-directional development in *E. granulosus* PSCs are controlled by miRNAs and genes likely associated with nervous system development and carbohydrate metabolic process. ATP-dependent transporter genes are associated with cystic development. These results may be important for exploring the mechanisms underlying early development in *E. granulosus* providing novel information that can be used to discover new therapeutics for controlling cystic echinococcosis.

Keywords: *Echinococcus granulosus*, protoscoleces, microRNA, transcriptome, differential expression, bi-directional development

INTRODUCTION

Echinococcus granulosus sensus tricto is the causative agent of CE, which is an important zoonosis with an almost cosmopolitan global distribution (McManus et al., 2003; Moro and Schantz, 2009; Wen et al., 2019). The life cycle of *E. granulosus* is complex and involves two mammalian hosts: a definitive host in which mature adult worms are produced and an intermediate host where larval cysts occur. The mature adult worm resides in the small intestine of the definitive carnivore host (dog, wolf, fox) and releases eggs which contaminate vegetation and water sources. Following ingestion by an intermediate host such as a human or sheep, the eggs hatch to release larval oncospheres which penetrate through the intestinal wall and migrate via the circulatory system to various organs (mainly the liver and lungs). In these organs, the oncospheres develop into hydatid cysts over many months and then produce PSCs which may remain in an inactive state for years. However, when a definitive host swallows the infected organs, the PSCs may be released from the cysts and develop into adult worms in the dog intestine. In contrast, if a hydatid cyst ruptures within an intermediate or human host, each released PSC is capable of developing into a secondary hydatid cyst (Zhang et al., 2005).

The bi-directional development of PSCs is a remarkable feature of *E. granulosus*. However, the mechanisms underlying this developmental process remain largely undefined. Several transcriptional and proteomic studies have identified some DEGs in the different life cycle stages of *E. granulosus* (Monteiro et al., 2010; Parkinson et al., 2012; Cui et al., 2013; Zheng H. et al., 2013). However, these genes may only be responsible for stage-specific biological processes or structural changes and may not include the genes involved in early differentiation of the parasite into either a strobilated adult tapeworm or hydatid cyst. The successful *in vitro* culture of PSCs provides an important model for studying the early development and differentiation of *E. granulosus* (Smyth et al., 1966; Smyth, 1990b; Zhang et al., 2005) to either the cystic form or adult worms triggered by bile acids.

A recent study of cultured parasites identified 75 differentially expressed proteins at 24 h by the azidohomoalanine (AHA)-specific labeling method (Debarba et al., 2015), but the study provided very limited data of the associated transcriptional events occurring.

MicroRNAs are usually highly conserved throughout the animal kingdom; they are involved in embryonic development and cell differentiation of specific developmental stages. For example, lin-4 is highly expressed in the first larval stage of *Caenorhabditis elegans* and controls diverse postembryonic developmental events (Lee et al., 1993). In addition, some miRNAs are highly tissue-specific expressed and exhibit distinct temporal expression patterns in *Schmidtea mediterranea* (Palakodeti et al., 2006; Friedlander et al., 2009), *Schistosoma japonicum* (Huang et al., 2009; Hao et al., 2010), *Schistosoma mansoni* (de Souza Gomes et al., 2011), *Echinococcus multilocularis* (Cucher et al., 2015), and *Echinococcus canadensis* (Macchiaroli et al., 2015), suggesting that miRNAs play a key role in the differentiation and development of helminth worms. Although our previous study identified 42 mature miRNAs and 23 miRNAs* differently expressed in three life-stages of *E. granulosus* (Bai et al., 2015), the roles of these miRNAs in the early stages of the bi-directional development process have not been defined.

In the present study, we used NGS to profile global miRNAs and mRNAs expression during the *in vitro* culture of PSCs at multiple time points (0–24 h, 7–14 days). Bioinformatics analysis showed that nervous system development, carbohydrate metabolism, and ion transmembrane transporter may be critical in the early differentiation of PSCs into either adult worms or larval cysts. The results increase knowledge on the process of *E. granulosus* differentiation at the molecular level.

MATERIALS AND METHODS

Collection of PSCs

Sheep livers containing hydatid cysts of *E. granulosus* (G1 genotype) were collected from a slaughterhousein Urumqi, Xinjiang, China. PSCs were collected from the cysts by aspiration, and then centrifuged at 1000 × *g* for 10 min. After washing 5 times with PBS, pH 7.4 (Zheng H. et al., 2013), the viability of the PSCs was determined by trypan blue staining (Zhang et al., 2013). Only samples of PSCs with a viability greater than 95% were utilized for further study.

Abbreviations: 3′-UTRs, 3′-untranslational regions; AST, aspartate aminotransferase; CE, cystic echinococcosis; DEGs, differentially expressed genes; *E. granulosus*, *Echinococcus granulosus*; FDR, false discovery rate; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; miRNAs, microRNAs; NGS, next generation sequencing technology; NSD, PSC without strobilisation stimulus; PAGE, polyacrylamide gel electrophoresis; PSCs, protoscoleces; SSD, PSC with strobilisation stimulus; TPM, transcripts per million.

Culture of PSCs and Sample Preparation

Protoscoleces were cultured *in vitro* as described (Smyth, 1990a), with minor modifications (Zhang et al., 2005). Briefly, PSCs were digested at 37°C for 30 min with 1% (w/v) pepsin (Sigma, St. Louis, MO, United States) prepared in 0.85% (w/v) sodium chloride, pH 2.0. After the digestion procedure, the PSCs were washed three times with PBS containing antibiotics (100 IU/mL penicillin and 100 mg/mL streptomycin, Sigma, St. Louis, MO, United States) and incubated in 37°C RPMI 1640 medium (Invitrogen, San Diego, CA, United States) containing 20% (v/v) fetal calf serum (Gibco, Auckland, NZ, United States), 0.45% (w/v) yeast extract, 0.4% (w/v) glucose, 100 IU/ml of penicillin, and 100 ug/ml of streptomycin at 37°C for 30 min. A solid nutritive base was made by heating newborn calf serum at 76°C for 45 min. To obtain adult worms (strobilar development, SSD), 2000 PSCs/mL were cultured in the above medium containing 0.02% (w/v) sodium taurocholate at 37°C. To obtain cysts, PCS were cultured in the same system as SSD but without the addition of sodium taurocholate (NSD). In both cases, medium was changed every 3 days. Cultured PSCs were collected at day 1, 7, and 14 days after *on vitro* cultivation for transcriptome and microRNAome analysis. We also collected cultured PSC treated only with pepsinand collected immediately, representing the baseline of gene expression. All parasite materials were subjected to a final wash in PBS and then they were suspended in 10 volumes of RNAlater (Ambion, Austin, TX, United States), and stored at −80°C.

RNA Isolation

Total RNA was extracted from echinococcal samples collected at the different time points using mirVana^{TM} miRNA isolation kits (Ambion, Austin, TX, United States), according to the manufacturer's instructions. RNA concentration and purity were measured using a NanoDrop ND-1000 UV spectrophotometer (Nanodrop Technologies, Wilmington, DE, United States). RNA integrity was evaluated by an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, United States).

Library Preparation and Sequencing of mRNA

About 10 μg total RNA was used for the purification of RNAs containing poly(A) using the MicroPoly(A) Purist Kit (Ambion/Applied Biosystems) according to the manufacturer's protocols. cDNA was synthesized using the SuperScript Reverse Transcriptase II (Life Technologies, Gaithersburg, MD, United States) kit with random primers. Paired-end cDNA libraries were generated by TruSeq RNA kit (RS-122-2001, Illumina) and sequenced on the Illumina Hiseq 2500 next generation sequencing platform.

Small RNA Sequencing

About 5 μg total RNA was used to isolate small RNAs (size 18 to 30 nt) on denaturing PAGE gels. Then, the isolated small RNAs were prepared according to Illumina's Small RNA v1.5 Sample Preparation guide (Illumina, San Diego, CA, United States). Briefly, the small RNAs were ligated to Illumina's small RNA 3′ and 5′ adaptors. Subsequently cDNA was synthesized by reverse transcription and amplified by PCR (12 cycles). The amplified fragments were purified by 6% (w/v) PAGE. Sequencing of the small RNA library was undertaken using the Illumina Genome Analyzer II (Illumina, San Diego, CA, United States).

RNA-SEQ Analysis

Adaptors were trimmed from raw reads using Trimmomatic v0.35 (Bolger et al., 2014) (with the setting conditions: sliding window: 4 bp, minimum average PHRED quality: 10; minimum read length: 40bp). To obtain an accurate assembly of transcripts, we applied Trinity v20140413p1 (Grabherr et al., 2011) and PASA v2.0.2 (Haas et al., 2003) to reconstruct transcripts for every gene. For the Trinity *de novo* and genome-guided assembly (assembly number: ASM52419v1), we used the default parameters. Tophat 2.0.8 (Kim et al., 2013) was implemented to map the trimmed reads to the *E. granulosus* genome with a maximum of two mismatches. Minimum and maximum fragment sizes were set at 300 and 500 bp, respectively. The remainder of the parameters were left as default. Differential expression analysis between time points was performed using DESeq (version 1.24.0) after counting the number of reads per transcript using featureCount v1.4.6 (Liao et al., 2014). Significance was assessed as having an experiment-wide FDR < 0.05 (calculated using the Benjamini–Hochberg method) and a fold-change value >2 or <0.5.

miRNA Analysis

Small RNA reads were first trimmed from the 5′ adapter sequence through FASTX toolkits[1]. Then, we removed low quality sequence reads and collected small RNAs ranging from 18–30 nucleotides. After removing duplicated sequences from the initial dataset, we compared all sequence tags against a database of known miRNA[2], and profiled every annotated miRNA in each library. We discarded all sequences matching with known *E. granulosus* rRNA, tRNA, snRNAs, mRNAs, and repeat sequences in the *E. granulosus* genome. The remaining unique sequences were predicted by miRDeep 2.0.7 (Friedlander et al., 2012) and miRNA candidates were selected based on the criteria described previously (Bai et al., 2015). All known or novel mature miRNAs were counted by the quantifier module of miRDeep2 and relative expression levels were calculated by counting the numbers of respective miRNA reads normalized to the total number of annotated miRNAs (TPM) in each library. After log2 transformation of each expression value (TPM + 1), differential expression between the two developing directions was evaluated using DEGSeq. The significance value was adjusted using FDR (Benjamini–Hochberg) to correct for multiple testing.

miRNA Target Prediction

3′-untranslational regions were determined using the pipeline described in a previous study (Macchiaroli et al., 2017). We used complete gene structures reconstructed by PASA analysis. Only sequences located at the 3′ untranslated region of predicted

[1]http://hannonlab.cshl.edu/fastx_toolkit/index.html

[2]http://www.mirbase.org/, release 22.0

genes were regarded as potential 3′-UTR sequences. The 3′-UTR sequences were extracted from the *E. granulosus* genome and length distribution analysis was performed using R custom scripts. Then, two algorithms, miRanda (Enright et al., 2003) and TargetScan, were used to predict the target genes of all mature miRNAs. The miRanda Energy thresholds were set at ≤-20 kcal/mol and other thresholds used a default value (score threshold, 140; gap-opening penalty, -9; gap-extend penalty, -4).

Functional Annotation

Novel transcripts, including alter splicing transcripts of known genes and new identified unigenes, were determined by PASA. Each alter splicing transcript was supported by at least two uniquely mapped reads. Unigene sequences were aligned to the non-redundant protein database (nr) using BLASTX (*E*-value $< 10^{-5}$) for obtaining both protein and functional annotation information. Based on the annotations in the protein database, Blast2GO program[3] was used to obtain GO annotations for the DEGs. Then, the hypergeometric test was used to classify the GO category, and the FDR was calculated to correct the *P*-value. KEGG pathway analysis was carried out using the KEGG Automatic Annotation Server for ortholog assignment and pathway mapping[4]. The hypergeometric test was used to assess significant pathways of enrichment.

Validation of Differentially Expressed Genes and miRNAs by Quantitative RT-PCR

To validate the Illumina sequencing results, we randomly selected 10 DEGs and 10 miRNAs and used quantitative RT-PCR (qPCR) to determine their expression levels at the different time points. Total RNA was extracted using mirVana^{TM} miRNA isolation kits (Ambion, Austin, TX, United States) according to the manufacturer's instructions. The RNA was reverse transcribed by PrimeScript RT Enzyme Mix I (Takara Bio, Otsu, Japan). The primers used in the qPCR analysis are listed in **Supplementary Table S16**. We respectively evaluated 5.8S rRNA and U6 as house-keeping small RNAs and *GAPDH*, *EIF3* and ß-actin as reference genes. Normfinder was used to find the most stable expressed house-keeping gene under the test conditions. As a result, 5.8S rRNA (M = 0.08) and GAPDH (M = 0.23) were selected as reference genes for quantitative (q)-PCR analysis. The RT reaction mixture was incubated at 37°C for 30 minutes, then at 85°C for 5 s. A control was set up at the same time with no RNA input. With the cDNA products as a template, qPCR was carried out using SYBR® Premix Ex Taq^{TM} (Takara Bio, Otsu, Japan) in the StepOne Plus real-time system (Applied Biosystems, Carlsbad, CA, United States). For each qPCR, dissociation curve analysis was carried out to discriminate the specific products from the primer dimers. The CT-values were the average of three technical and three biological replicates and fold changes of miRNAs in different samples were calculated by the $2^{-\Delta\Delta Ct}$ method.

[3]http://www.blast2go.com/

[4]http://www.genome.jp/tools/kaas

RESULTS

We successfully used an *in vitro* culture method to promote the early development of PSCs into either adult worms of 3−5 proglottids or cystic metacestode cysts (up to 2 cm in size) (data not shown). The bi-directional development of PSCs in the culture system (with or without dog bile) is shown in **Figure 1**. Increased scolex evagination and the production of numerous calcareous corpuscles was observed within the first 3−7 days of cultivation in culture medium containing bile salt (sodium taurocholate). However, after 14 days culture, the calcareous corpuscles had disappeared; lateral excretory canals were conspicuous and a genital rudiment was present denoting the formation of the first proglottid. In culture medium without the bile salt, the PSCs developed into cysts; these developing PSCs contained calcareous corpuscles in the first 3−7 days as was evident during adult worm development. Cysts were then subsequently formed (**Figure 1**).

Summary of RNA and Small RNA Sequencing and the Alignment of Reads

We compared the transcriptomes and microRNAomes of the cultured *E. granulosus* PSCs harvested at 1, 7, and 14 days after culture in medium with (early adult worm development) or without (early cystic development) dog bile salt. A total of 394.92 million and 210.72 million Illumina raw reads were respectively generated from the transcriptomic and microRNAomic libraries. The raw reads were deposited in the National Center for Biotechnology Information Sequence Read Archive under the accession number PRJNA610943. After removing low-quality sequences, adaptor contaminants and small reads (less than 40 bp in RNA-seq and 18 nts in small RNA-seq), we obtained 352.59 million and 163.95 million clean reads, respectively. Of these reads, an average of 65.63% and 51.12% could be mapped to the *E. granulosus* genome (**Table 1**).

Due to the absence of sufficient information about alternative splicing isoforms and 5′ and 3′ untranslated regions (UTR) in current *E. granulosus* annotations, we performed both a *de novo* as well as a genome-guided assembly (ASM52419v1) to reconstruct the gene transcripts and then combined the outputs using the PASA pipeline. There were respectively 113,141 and 93,584 transcripts identified using these two assembly methods (**Supplementary Table S1**). We further combined them with known genome annotations, and identified 185 of 11,329 predicted genes that merged with neighboring genes. Furthermore, a total of 5,460 genes were re-annotated gene constructions. Among these, 3,219 genes were determined to have alternative splicing events, and 4,795 and 4,859 genes were found containing 5′-UTRs and 3′-UTRs, respectively. **Supplementary Figure S1** shows the length distribution of the annotated *E. granulosus* 3′-UTR sequences. A major proportion (90.44%) of the 9,839 3′-UTRs was shorter than 4,000 nucleotides. The median length of the total set of 3′-UTRs was 1,027 nucleotides and the average GC content was 42.45%. We then mapped the high quality sequence reads to the *E. granulosus* transcriptome. The alignment summary metrics are shown in

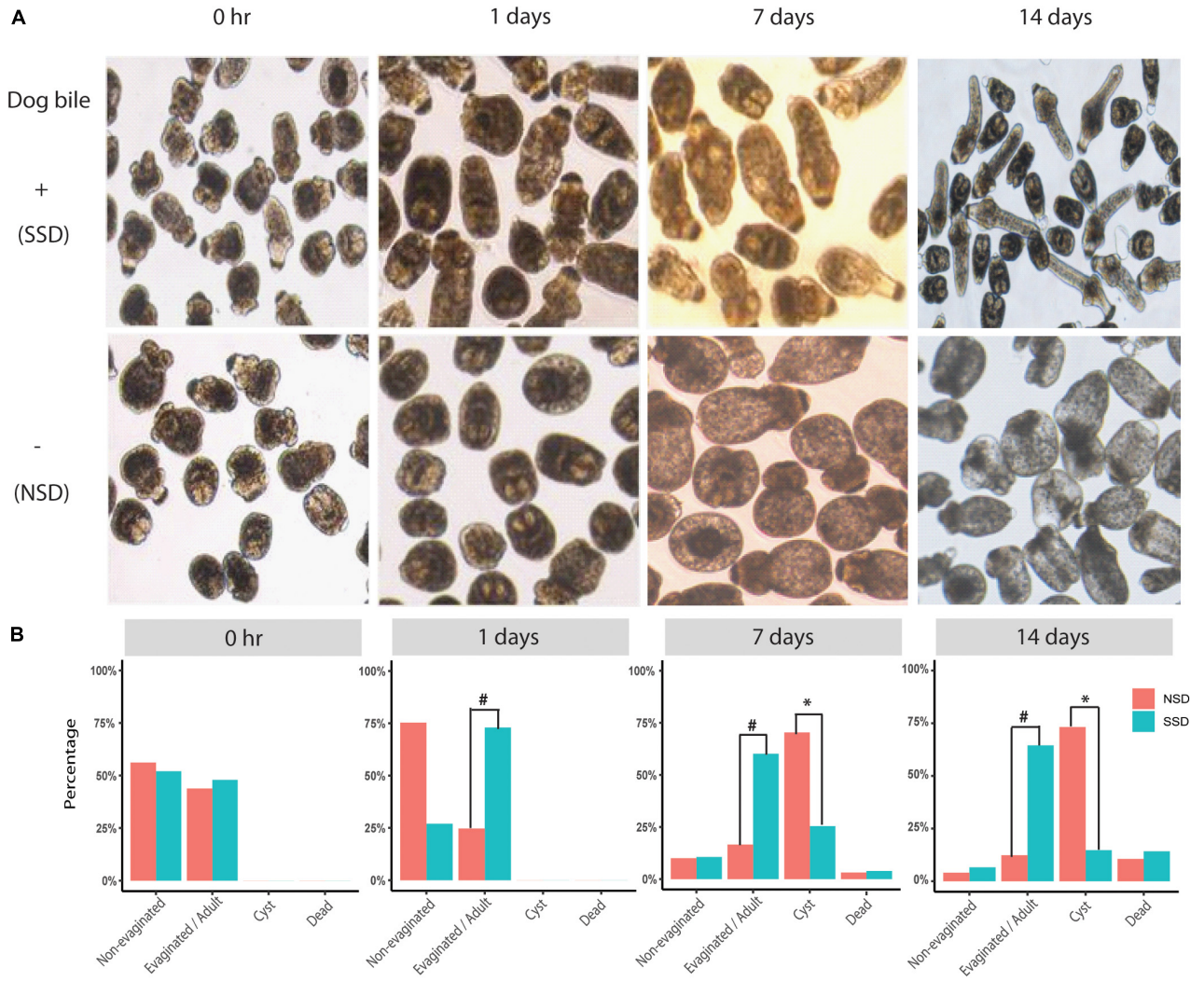


FIGURE 1 | Bi-directional development of *E. granulosus* protoscoleces *in vitro*. **(A)** The structural changes in *in vitro* cultured PSC in the presence or absence of bile salt as a strobilization stimulus. **(B)** A scale bar to summarize the percentages of the different development directions, including adult worms (Evagination/Adult), hydatid cyst (Cyst), dead (Dead), and no changes (Non-Evagination). #A significant increase in scolex evagination occurred at 1–7 days after stimulation with the dog bile salt ($P < 0.001$). When the PSC were cultured for 14 days, the calcareous corpuscles were much reduced and suckers and the first proglottid had formed in adult worms. *A significant increase of PSC developing to cysts at 7–14 days compared with SSD ($P < 0.001$).

Table 1. After elimination of ambiguous sequence matches, more than 18 million reads were mapped in each library. The number of sequence reads mapped per gene varied from one to >740,000.

For miRNA sequencing, we discarded all known non-coding RNAs, such as rRNA, tRNA, snoRNA, repeat-associated RNA, and degraded fragments of mRNAs. The remaining 70.82 million high quality sequences were used to search for both known and novel miRNAs. To date, there are 111 miRNA precursors of *E. granulosus* identified in the miRBase database. They could encode 218 mature miRNAs or miRNA stars sequences. By deep sequencing, we found that more than 40% high quality sequences were matched to the known miRNA precursors in each library (**Supplementary Table S2**). A total of 167 known mature miRNAs or miRNA stars were identified in the present study. miR-71-5p and miR-1-3p were the most abundant in all libraries. After using miRDeep2 to predict novel miRNA precursors, we identified 7 miRNA candidates (**Supplementary Table S3**). Among these, the mature sequence miR-10450b sequence was highly consistent with miR-10450a in the 5′ arm, and these were classified into the same miRNA family. None of the others showed homology with any other metazoan miRNAs and were thus defined as *E. granulosus* specific miRNAs.

Transcriptomic Changes in Early Development of PSC

To identify DEGs in PSCs during early bi-directional development, we compared the transcriptomes of PSCs treated with pepsin and cultured for 24 h and 7–14 days in the presence or absence of dog bile salt with the baseline which

TABLE 1 | Data summary of the *E. granulosus* transcriptome and microRNAome.

Samples ID	RNA Sequencing			Small RNA Sequencing		
	Total Clean Reads	High Quality Reads Genome Mapped (%)*	High Quality reads Tr anscripts mapped (%)#	Total Clean Reads (18–30 nt)	High Quality Reads Genome Mapped (%)*	High Quality Reads miRNAs Mapped (%)#
0 h	56,805,518	41,477,149 (73.02%)	27,483,646 (66.26%)	24,185,579	13,168,238 (54.45%)	7,249,946 (55.05%)
NSD						
24 h	45,865,714	27,534,887 (60.03%)	20,807,190 (75.57%)	22,664,542	13,374,352 (59.01%)	8,162,254 (61.03%)
7 days	47,387,288	30,641,662 (64.66%)	21,343,889 (69.66%)	22,053,979	12,312,969 (55.83%)	6,996,764 (56.82%)
14 days	55,597,998	38,266,289 (68.83%)	28,976,806 (75.72%)	22,113,695	11,027,824 (49.87%)	5,184,590 (47.01%)
SSD						
24 h	55,037,642	33,752,124 (61.33%)	19,137,393 (56.70%)	21,173,443	10,502,453 (49.60%)	3,546,522 (33.77%)
7 days	48,910,972	28,462,253 (58.19%)	18,073,569 (63.50%)	26,331,190	9,608,894 (36.49%)	5,773,972 (60.09%)
14 days	42,987,048	31,534,216 (73.36%)	23,182,955 (73.52%)	25,430,586	13,366,873 (52.56%)	6,859,834 (51.32%)

*NSD, PSC without strobilization stimulus (no bile salt added); SSD, PSC with strobilization stimuli (bile salt added). *The percentages were calculated by dividing the counts of total clean reads. #The percentages were calculated by dividing the counts of high quality reads genome mapped.*

TABLE 2 | Summary of the counts of differentially expressed genes over time in the two developmental (i.e., cystic or strobilar) directions.

	24 h vs. 0 h	7 days vs. 0 h	14 days vs. 0 h	7 days vs. 24 h	14 days vs. 7 days	Sum
PSC without strobilization stimulus (NSD)						
Total DEG	276	328	276	352	344	972
DEG (log_2 fold-change >1.0)	110	188	161	250	151	612
DEG (log_2 fold-change <−1.0)	166	140	115	102	193	558
PSC with strobilization stimulus (SSD)						
Total DEG	320	361	349	272	334	963
DEG (log_2 fold-change >1.0)	95	246	156	193	88	579
DEG (log_2 fold-change <−1.0)	225	115	193	79	246	622

DEG, differentially expressed genes. The statistical comparison was determined by DESeq. Log_2 fold-change cutoffs of +1.0 and −1.0.

is gene expression in the PSCs treated only with pepsin and collected immediately. As shown in **Table 2** and **Supplementary Table S4**, 963 DEGs were differentially expressed in the PSCs culture with bile salt strobilisation stimulus (SSD), including 579 genes that were up-regulated and that were 622 genes down-regulated. In PSCs developing cysts [i.e., without bile salt strobilisation stimulus in the culture medium (NSD)], 972 DEGs were identified, including 612 up-regulated genes and 558 down-regulated genes (**Supplementary Table S5**).

When comparing each time point (24 h and 7–14 days) to the baseline (0 h), we identified 52 DEGs (17 up-regulated and 35 down-regulated) in SSD and 28 DEGs (28 up-regulated and 0 down-regulated) in NSD which are differentially expressed at all three time points (**Figure 2A**). Based on the functional annotations of the up-regulated genes, most of the DEGs in SSD were characterized as hypothetical protein with those in NSD belonging to transporter, aminotransferase and protease (**Figure 2B**).

We also horizontally compared the expression profiles between the two developmental patterns at the different time-points. A total of 537 DEGs were found in SSD or NSD. After removing the non-differentially expressed genes by comparison with the baseline (0 h), we identified 172 DEGs in the two developmental patterns (118 genes were up-regulated in SSD and 63 genes were over-expressed in NSD; **Figure 2C** and **Supplementary Table S6**). Most of the DEGs were temporally changed, being mainly differentially expressed at 24 h or 7 days. There were 22 genes with increased expression at 24 h after the stimulation with bile acid and 97 genes were over-expressed at 7 days. A similar trend was also observed in NSD; there were 29 and 34 genes up-regulated at 24h and 7 days, respectively. Finally, only 6 DEGs were changed at 14d in both SSD and NSD. These results suggest that the early stage is an important time point guiding the early developmental direction of PSCs. Moreover, as shown in **Figure 2C**, we observed that 9 DEGs were changed in both SSD and NSD with expression changes at the different time-points, indicating that even with the same gene, its temporal and spatial pattern could determine the early developmental direction of PSCs. We used qPCR to validate the expression of 10 randomly selected DEGs of which 8 were shown to be in accordance with the expression profiles detected by the RNA sequencing (**Supplementary Table S7**).

Gene Function Analysis of DEGs During the Early Development of PSCs

We conducted GO functional analysis of the up-regulated and down-regulated DEGs in SSD and NSD, with 52.0 and 64.6%, respectively, of DEGs assigned a GO term. Further enrichment analysis revealed significant over-representation of 25 terms

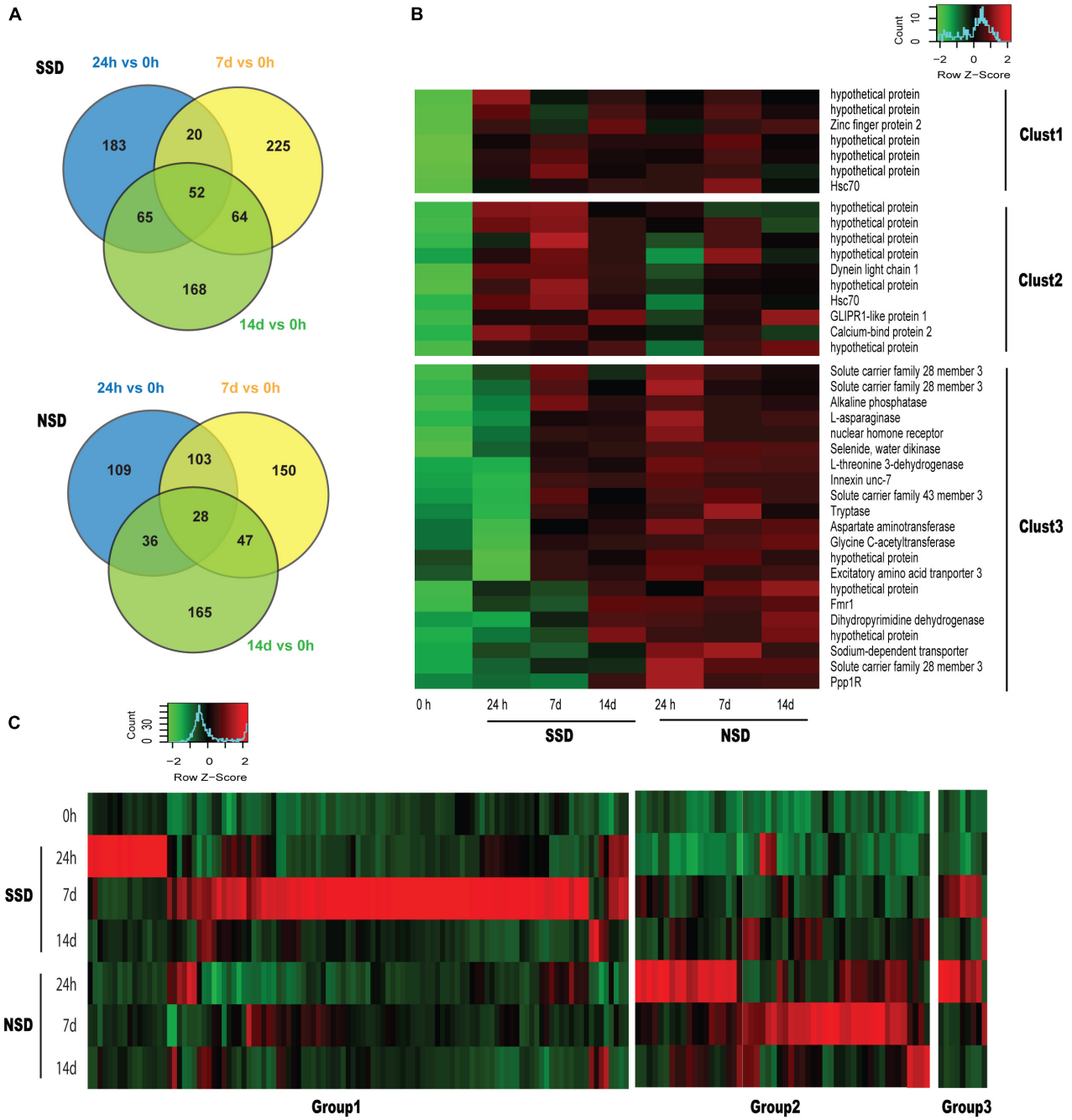


FIGURE 2 | Dynamic transcriptomic changes over time. SSD: *in vitro* culture of PSC with bile salt strobilization stimulus; NSD: *in vitro* culture of PSC without bile salt strobilization stimulus. **(A)** Venn diagram showing the number of overlapping differentially expressed genes (DEGs) across pairwise comparisons at each time point compared to the baseline (0 h). **(B)** Heat maps of sequencing data from 38 overlapping DEGs identified in **Figure 1A**. Clust1 represents 7 DEGs which were significantly increased in the both SSD and NSD. Clust2 and Clust3 represent 10 and 21 DEGs, respectively, which were specifically and consistantly over-expressed in SSD and NSD. A *P*-value < 0.05 and a fold-change >2.0 were used as the threshold criteria to define significant differences. **(C)** Heat maps showing 172 DEGs which were specifically over-expressed at SSD or NSD (118 in the SSD and 63 in NSD). Groups1 and 2 respectively represent 109 and 54 DEGs which were only up-regulated in SSD or NSD. Group 3 includes 9 DEGs which were simultaneously over-expressed in SSD and NSD.

in SSD (10 from up-regulated genes and 15 from down-regulated genes) and 25 in NSD (23 from up-regulated genes and 2 from down-regulated genes) (**Supplementary Table S8**). After comparing these between SSD and NSD, we identified 8 GO terms simultaneously existing in the two development directions, namely cellular homeostasis (GO: 0019725), cellular chemical homeostasis (GO: 0055082), cellular ion homeostasis (GO: 0006873), microtubule-based process (GO: 0007017),

microtubule associated complex (GO: 0005875), cytoskeleton (GO: 0005856), cytoskeleton part (GO:0044430) and microtubule cytoskeleton (GO: 0015630) (**Figure 3** and **Supplementary Table S9**); these were more prominent in NSD.

Additionally, we observed some GO terms were specifically enriched in SSD or NSD. For example, comparing the NSD, more up-regulated genes in SSD were associated with nervous system development (GO: 0007399), carbohydrate metabolic process (GO: 0005975) and cellular carbohydrate metabolic process (GO: 0044262). Meanwhile, there were 15 "Molecular Function" terms more prominent in the up-regulated genes of NSD than in SSD (**Supplementary Table S9**). Most of these are related to ion transmembrane transporter activity (GO: 0015075), substrate-specific transporter activity (GO: 0022857) and nucleoside-triphosphatase activity (GO: 0017111). Of note, we also observed that some biological processes had obvious stage-specificities. As shown in **Figure 3**, there were 18 over-expressed genes related to carbohydrate metabolic process (GO: 0005975) in SSD at 7d. This number was significantly higher than in NSD (8 genes) at the same time-point. In the GO terms of ion transmembrane transporter activity (GO: 0015075) and substrate-specific transporter activity (GO: 0022857), clear increases were evident in the amounts of up-regulated genes in NSD at 14d.

Expression Changes in miRNAs Over Time

We compared the expression profiles of miRNAs across the different stages of early bidirectional development. A total of 31 and 27 mature miRNAs, respectively, exhibited statistically significant changes (a threshold of relative expression >50, correct P-value < 0.01 and fold-change >2.0) in at least one of the three time-points during SSD and NSD (**Supplementary Table S10**). Further clustering of these miRNAs, based on similar expression patterns, identified 5 clusters in each development direction (**Supplementary Figure S2** and **Supplementary Table S11**). 7 days was the stage with the most significant expression change among the three time-points. There were, respectively, 13 up-regulated miRNAs (cluster 1) and 8 down-regulated miRNAs (cluster 2) identified at this time-point of SSD. A similar pattern was evident in NSD, where cluster 1 and cluster 2 respectively represented 9 and 7 specifically up- and down-regulated miRNAs. Additionally, the 24 h culture time point is also an important regulatory period in PSC development. There were respectively 5 miRNAs (cluster 3) and 3 miRNAs (cluster 4) that were specifically over expressed in SSD and NSD. In addition, we found that some miRNAs were consecutively differentially expressed at multiple stages, such as cluster 4 and cluster 5, respectively, which contained 2 and 3 miRNAs for which expression increased and decreased throughout early development in SSD. Meanwhile, cluster 5 consisted of miR-9 and miR-4989 that were strongly up-regulated in NSD at 7–14 days.

After further comparing the expression patterns of each miRNA, we identified 12 differentially expressed miRNAs between SSD and NSD (**Figure 4A**). Additionally, there were 3 miRNAs (miR-190-5p, miR-745-3p, and miR-71-5p) having the same trend throughout the three time-points of development, although these differences did not reach the 2-fold threshold (**Figure 4B**). Among these miRNAs, 10miRNAs (let-7-5p, miR-133-3p, miR-153-3p miR-219-5p, miR-3479a-3p, miR-2b-3p, miR-2c-5p, new-1-3p, miR-190-5p, and miR-745-3p) were

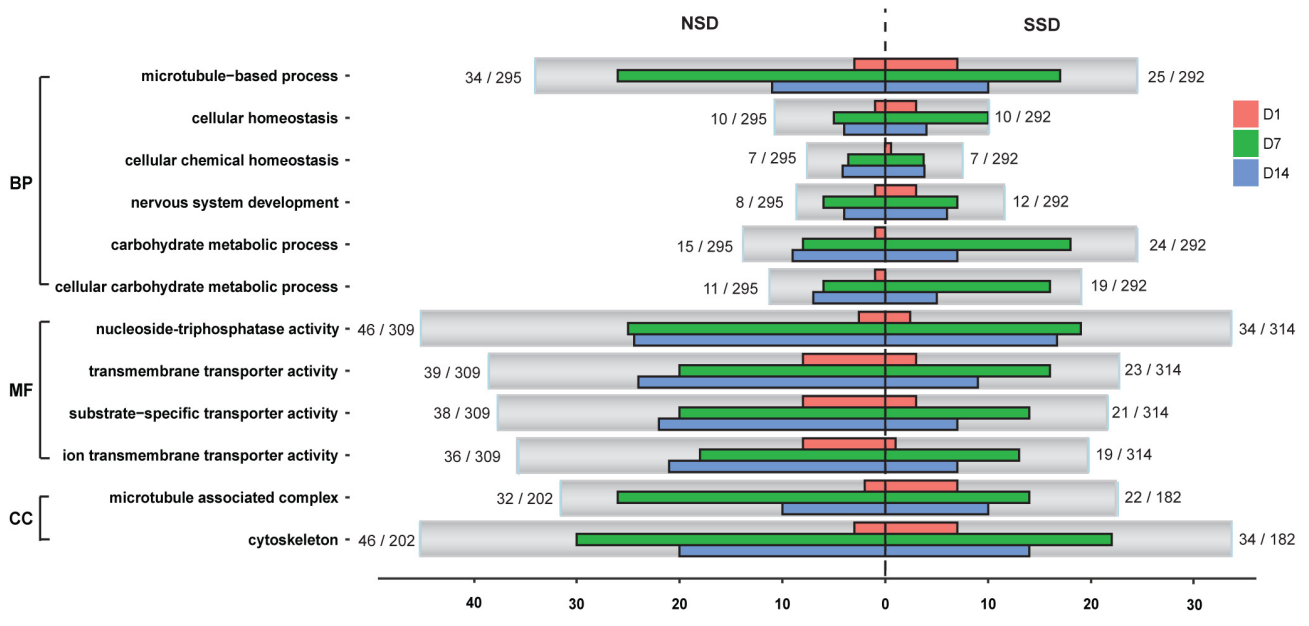


FIGURE 3 | Classification of GO annotations. The *X*-axis indicates the number of genes in a sub-category and the *Y*-axis indicates the GO sub-categories. D1, *in vitro* culture for 24 h; D7, *in vitro* culture for 7 days; D14, *in vitro* culture for 14 days. SSD, *in vitro* culture of PSC with bile salt strobilization stimulus; NSD, the *in vitro* culture of PSC without bile salt strobilization stimulus; BP, biological processes; MF, molecular functions; CC, cellular components.

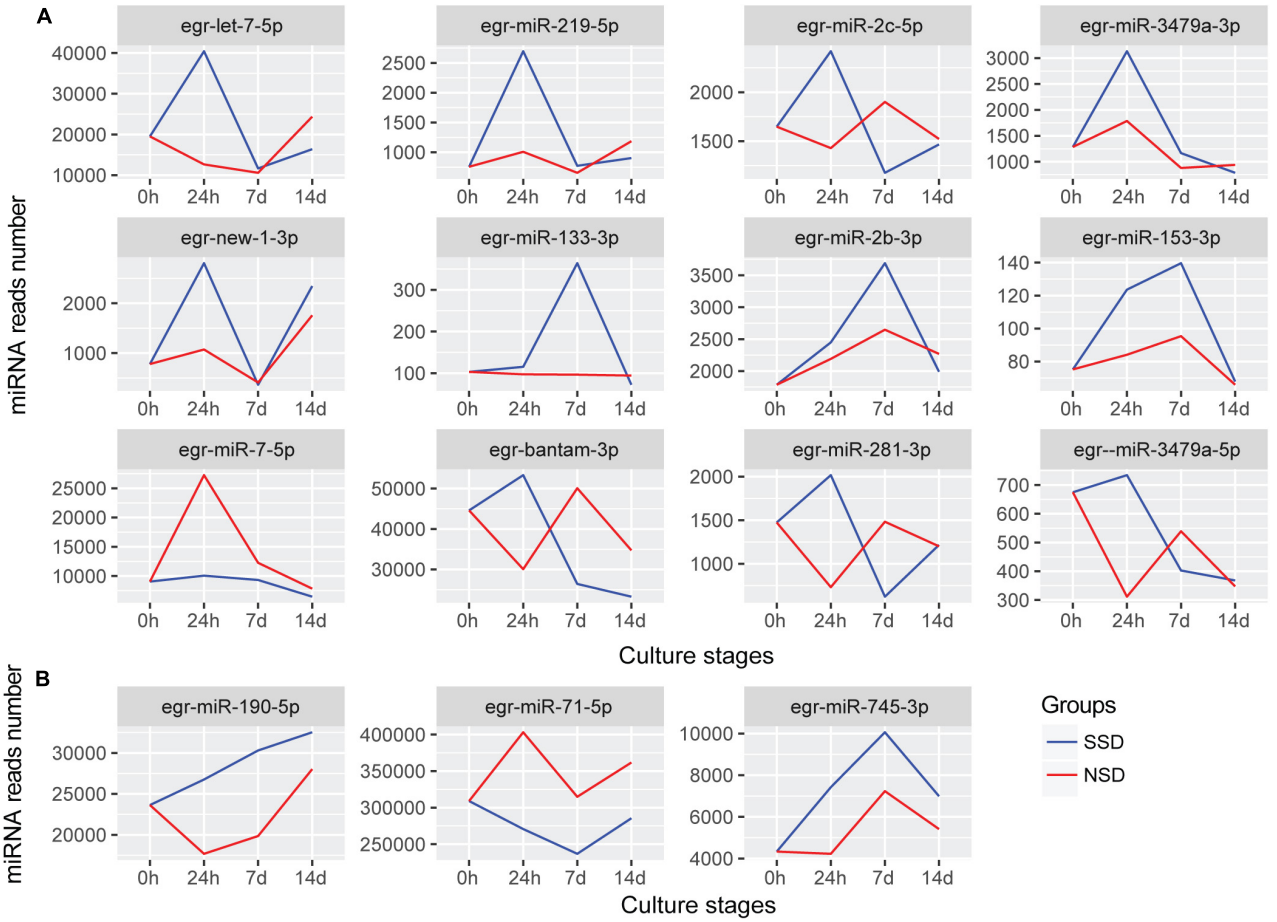


FIGURE 4 | miRNAs with different expression patterns between SSD and NSD. **(A)** The differentially expressed miRNAs with different expression patterns. **(B)** The miRNAs were differential expressed throughout the three time-points. The blue line indicates the expression levels of candidate miRNAs during *in vitro* culture in the presence of the bile salt as strobilization stimulus (SSD). The red line indicates the expression levels of candidate miRNAs during the *in vitro* culture without the bile salt strobilization stimulus (NSD). The *Y*-axis indicates the relative expression levels of each miRNA.

significantly up-regulated in SSD, and two miRNA (miR-7-5p and miR-71-5p) were specifically over-expressed in NSD. A clear characteristic of time-dependent expressions was observed in these miRNAs. For example, let-7-5p, miR-219-5p, miR-3479a-3p, miR-2c-5p, and new-1-3p were specifically over expressed at 24 h, while miR-2b-3p and miR-133-3p were only increased at 7 days. These consistent differences also imply that these miRNAs may be associated with the direction of early development of a PSC into either an adult worm or a secondary hydatid cyst. Finally, we used stem-loop real-time quantification RT-PCR to examine the expression of 10 randomly selected miRNAs of which 9 were shown to be in accordance with the expression profiles detected by the Illumina sequencing (**Supplementary Table S12**).

Target Gene Prediction

To evaluate the biological functions of the 15 differentially expressed miRNAs, we predicted putative target genes using the Miranda program. Based on our 3′-UTR results in the transcriptome, we found 3,601 genes possibly targeted by 242 mature miRNAs and star sequences. Among these, a total of 971 genes may be regulated by the differentially expressed miRNAs (**Supplementary Table S13**). We also used the worm TargetScan to validate our predictions. As shown in **Supplementary Table S13**, all targets of the differentially expressed miRNAs were confirmed except for one novel miRNA which was not found in the same miRNA family in the TargetScan database. Further enrichment analyses on the predicted targets, indicated that these target genes are more abundant in the GO term of nucleotide binding (adjust $P = 0.020$, $n = 162$, **Supplementary Table S14**); these include cytoplasmic polyadenylation element-binding protein 1 (EG_00812), Titin (EG_02049), Casein kinase I alpha (EG_04711), Heat shock cognate 70 kDa protein (EG_08863), Tyrosine-protein kinase Srms (EG_10535), and Fyn (EG_09408), some of which have been shown to be associated with germline development (Jin et al., 2001), cell asymmetric division (Banerjee et al., 2010), and adaptation processes of parasites in the host environment (Yang et al., 2012). Moreover, we also determined that some target genes, such as let-7, bantam and miR-71, regulated by the differentially expressed miRNAs

may be involved in the MAPK signaling (PATH: ko04010) and Wnt signaling pathways (PATH: ko04310) (**Supplementary Figure S3**). These results are consistent with a previous study (Macchiaroli et al., 2017), suggesting that these miRNAs maybe involved in *E. granulosus* development.

Finally, we observed a significant enrichment of DEGs associated with oxidation reduction process. As shown in **Supplementary Table S15**, there were 10 DEGs differentially expressed in this process, including Glyceraldehyde-3-phosphate dehydrogenase (EG_02307), NADH dehydrogenase 1 beta subunit 9 (EG_03558), Ferritin 1 (EG_08965), Protein yfeX (EG_07721) and Heat shock cognate 70 kDa protein (EG_09244, EG_09649). All of these genes were clearly up-regulated at 7 days in the SSD. Notably, some differentially expressed miRNAs between the SSD and the NSD may also regulate these targets. For example, a higher expression level of miR-71-5p was observed at baseline than at 7 days, yet its target gene heat shock cognate 70 kDa protein *HSC70* (EG_09649), was up-regulated at 7 days. Moreover, at 24 h and 7 days of SSD, we also observed a significant negative correlation between the expression of miR-219-5p and its putative target gene yfeX (EG_07721).

DISCUSSION

The bi-directional development of PSCs is a striking feature of the biology of *E. granulosus* (Zheng H. et al., 2013). In the presence of bile salt, a PSC can differentiate in a sexual direction to form a mature adult worm in the dog gastrointestinal tract. However, if a hydatid cyst ruptures within the intermediate or human host, each released PSC is capable of differentiating asexually into a new hydatid cyst. These *in vivo* developmental patterns can be replicated *in vitro* and the culture of PSCs provides a unique opportunity to determine how the gene pathways are modulated in these diverse environments. In the present study, we used NGS to study the transcriptional dynamics of PSCs cultured in the presence or absence of dog bile acid at three *in vitro* culture time points (24 h and 7–14 days). A total of 5,460 genes were re-annotated using the gene constructions based on the ESTs of genome-guided assembly. Among these, 3,219 of 10,044 detected genes underwent alternative splicing (32.0%). This ratio is very close to the proportions recorded for *Dictyocaulus viviparus* (34.6%) (Cantacessi et al., 2010) and *Cooperia oncophora* (33.1%) (Heizer et al., 2013), but is significantly higher than that reported for *C. elegans* (17%) in WormBase (Yook et al., 2012). This result is consistent with observations of parasitic nematode transcriptomes (Abubucker et al., 2014), which may be due to a reduction in the number of functional genes in parasite genomes (Zheng H. et al., 2013) or the increased genomic complexity that may be required to interact with multiple hosts/vectors (Abubucker et al., 2014).

Further expression pattern analysis of the *E. granulosus* transcriptome revealed that highly temporal gene expression was evident in both SSD and NSD. Only 52 and 28 DEGs, respectively, were found to have consistent expression changes at all three time-points. Although there are no reports of the dynamic changes in gene expression during PSC development, some stage-specific changes in morphology (Thompson, 2017) indicated that the study of PSC development needs to simultaneously consider temporal cues and environmental conditions. Through the comparison of DEGs between SSD and NSD, we found that most of the DEGs were detectable at the 24 h and 7 day culture points. This phenomenon suggests that the early developmental stage is critical for triggering the PSCs to differentiate either into an adult worm or a secondary cyst. Key regulatory genes that determine the developmental direction may be differentially expressed at this time. For example, neurogenic locus protein delta, an important ligand for the Notch receptor, is expressed at the early embryonic stage of *Drosophila* to regulate the correct separation of neural and epidermal cell lineages (Haenlin et al., 1994). In the present study, its gene homolog in *E. granulosus* was specifically over-expressed in SSD at 7d culture, suggesting that it may promote the differentiation and growth of the early nervous system in the PSC as it develops to an adult worm.

Nutrition and Energy Metabolism in the Early Stages of the Bi-Directional Development of PSCs

Nutritional metabolism is critical for cellular and organismal development. Usually, parasites need to take up considerable amounts of nutrients from the host to maintain their own growth and reproduction. *E. granulosus*, because of multiple hosts in its life-cycle, is confronted by a range of complex host environments, such as the liver or lungs in intermediate hosts or the small intestine in definitive hosts. This diversity has led these worms to modify their energy metabolism to adapt to the different hosts and environments. We showed previously that *E. granulosus* has complete pathways for glycolysis, the tricarboxylic acid cycle and the pentose phosphate pathway, but lacks the capability for *de novo* synthesis of pyrimidines, purines and most amino acids (except for alanine, aspartic acid and glutamic acid) (Zheng H. et al., 2013); indicating that *E. granulosus* has to rely on obtaining some essential nutrients directly from its mammalian hosts. In addition, more genes associated with aerobic metabolism were observed to be up-regulated in SSD, and these included phosphoglucomutase, glucokinase, fructose-bisphosphate aldolase, pyruvate dehydrogenase and citrate synthase. This increased level of expression can be interpreted in two ways. Firstly, all nutrients taken in by the adults of *E. granulosus* are absorbed through the tegument. However, due to the limitations of its small volume and surface area, the adult worms have to use aerobic glucose metabolism to yield a higher level of energy than anaerobic glycolysis. It has been reported that adults of *E. granulosus* worms tend to produce more acetate and succinate and less ethanol in the small intestine of the definitive host (Constantine et al., 1998). Secondly, compared with the metacestode cyst, the adult worm of *E. granulosus* has a remarkably complex nervous system (Koziol et al., 2013). This current study has revealed that many neuro-developmental related genes were over-expressed in SSD. Although there are no reports that have considered the relationship between nervous development and carbohydrate

metabolism in *E. granulosus*, studies on another cyclophyllidean cestode, *Mesocestoidescorti*, showed that the distribution of glycogen changed according to the distribution of nerve codes during its development from a larva to an adult worm (Cabrera et al., 2010); this implies that nervous system development may utilize carbohydrate metabolism to provide the necessary energy requirements.

Glycans are important components of a wide variety of microbial pathogens, including viruses, bacteria, and parasites, including *E. granulosus*. They are widely distributed at hydatid cyst membranes and in PSCs (Khoo et al., 1997), and have been shown to control host complement activation (Diaz et al., 1999). In the present study, we found that the levels of some enzymes involved in the synthesis of polysaccharides, including beta-1,3-N-acetyl-glucosaminyltransferase, UDP-glucose and 6-dehydrogenase, were increased in SSD, suggesting that glycans may not only protect the cystic stage from immune attack, but may also assist the adult worms in attaching to its definitive host. Furthermore, lectin fluorescence assays have shown that the glycans are widespread in the tegument and parenchyma components, including the reproductive system in adult *E. granulosus* (Casaravilla et al., 2003).

Ion Channel and Nutrient Transporters in the Cyst

Host-parasite crosstalk at the cellular and molecular levels is essential for *E. granulosus* survival and development. Our previous genomic study identified two different mechanisms of material exchange occur simultaneously in this tapeworm, including microtubule-dependent passive diffusion and ATP-dependent active transport (Zheng H. et al., 2013). In the present study, we observed that the up-regulated genes in the microtubule-based process were significantly enriched in the both SSD and NSD, a result suggesting that passive diffusion may be involved mainly in the provision of some critical metabolites for worm survival. Furthermore, we also noted a significant increase in the expression levels of ATP-dependent transporter genes in NSD, suggesting the more actively growing cyst requires more essential nutrients from its host.

Additionally, we also found that many calcium (Ca^{2+}) transporters were over-expressed in NSD. In general, Ca^{2+} gains access into cells across the plasma membrane through voltage-gated Ca^{2+} channels. In order to maintain a very large transmembrane electrochemical gradient, the cell has to remove Ca^{2+} through ATP-dependent Ca^{2+} transporters to maintain low concentrations of Ca^{2+} intracellularly (Carafoli, 1991). A number of studies have shown that praziquantel can act directly or indirectly on Ca^{2+} channels and effectively kill adult *Schistosoma* and *Echinococcus* worms (Andersen et al., 1981; Richards et al., 1988; Xiao et al., 1994; Greenberg, 2005; Wei et al., 2005). These findings not only suggest that Ca^{2+} homeostasis is critical to the survival and development of *E. granulosus*, but also imply that interfering with Ca^{2+} transport may be a novel alternative approach for humancystic echinococcosis treatment. A recent study has reported that low doses of tamoxifen reduced the survival of *E. granulosus* both *in vitro* and *in vivo* (Nicolao et al., 2014). The drug may induce increased release of ATP-dependent Ca^{2+} from the endoplasmic reticulum and disrupt Ca^{2+} homeostasis (Bollig et al., 2007).

Albendazole and mebendazole are proven effective drugs for the treatment of CE. These compounds not only inhibit the development of PSCs and germinal cells in a dose-dependent manner *in vitro*, but they also reduce cyst weight after oral application in infected mice (Liu et al., 2015). In the past, it was believed that benzimidazole treatment efficacy resulted mainly from selectively inhibiting the synthesis of microtubules in parasitic worms, and destroying extant cytoplasmic microtubes in their intestinal cells, thereby blocking the uptake of glucose and other nutrients (Schmidt, 1998). However, a recent study found that mebendazole can significantly inhibit the AST enzyme activity of hydatid cysts (Naderloo et al., 2016). Our present study also showed that the expression levels of AST and asparaginase were significantly increased during the process of NSD, suggesting that albendazole or mebendazole not only reduce the uptake of glucose and other nutrients, but also interfere with the synthesis of some essential amino acids.

miRNAs of *E. granulosus* as Novel Therapeutic Targets for Parasite Control

In the present study, we compared the expression profiles of miRNAs in the early stages of the bi-directional development of *E. granulosus*. We identified 15 miRNAs with differentially expressed patterns between SSD and NSD. These miRNAs were differentially expressed at either some specific time points or during the whole development process and may directly determine the developmental direction of PSC. Accordingly, effective inhibition of these miRNAs as novel targets may represent an effective new avenue targets for the treatment of echinococcosis.

One of the most abundant miRNAs in all the early PSC development libraries was miR-71-5p. In *C. elegans*, miR-71may enhance longevity through the negative regulation of PHA-4/FOXA and DAF-16/FOXO (Boulias and Horvitz, 2012; Smith-Vikos et al., 2014). Furthermore, miR-71 is involved in neuronal development in nematodes (Hsieh et al., 2012) and in the biological function of neoblasts, which are the only known cell type in the Platyhelminthes with the ability to undergo mitosis (Zheng Y. et al., 2013). We showed that miR-71 was significantly highly expressed in NSD and down-regulated in SSD, and suggest that its targets may be involved in the regulation of the MAPK and Wnt signaling pathways in SSD. A recent study proposed that MAPK may be a therapeutic target in *E. granulosus* through the suppression of MKK3/6 and MEK1/2 (Zhang et al., 2019). Furthermore, some Wnt-pathway related proteins, such as the frizzled receptor eg-fz4, have been shown to be highly expressed in PSC cultured *in vitro* toward adult maturation, and that they may be involved in early development of the nervous system and gut-derived organs (Pan et al., 2006; Dezaki et al., 2016; Zidek et al., 2018). These results are consistent with a previous study (Macchiaroli et al., 2017), and suggest that the down-regulation of miR-71 in SSD might trigger the MAPK and Wnt pathways in growth and development.

In addition, many components, including proteins, small molecular compounds and miRNAs, have been shown to be involved in the interactions between the mammalian host and *E. granulosus*; they can not only promote the growth of parasites, but also affect the immune response of the host (Zheng Y. et al., 2013). A study of comparative proteome profiling in *E. granulosus* identified many host proteins in human hydatid cyst fluid, indicating that the hydatid cyst is permeable allowing constant protein exchange between the metacestode and affected host tissue (Zeghir-Bouteldja et al., 2017). It has been shown that nematode miR-71 can be released into the host via exosomes (Buck et al., 2014). The transfection of *E. multilocularis* miR-71 mimics on murine macrophage cells significantly repressed the production of NO after treatment with lipopolysaccharide (LPS) (Zheng et al., 2016). The production has been shown to play an essential role in limiting the extent of infection and effecting immunosuppressive anti-Echinococcus immune responses (Zheng, 2013). These results suggest that the up-regulation of miR-71 in NSD may result in the interference of host innate immunity following its the secretion from the parasite to the host.

let-7 is an the important member of the miRNA family first identified in *C. elegan* (Sokol, 2012). Some studies have suggested that it can act as a developmental switch to control the transition from larva to adult (Reinhart et al., 2000). We previously showed that *E. granulosus* let-7 could complementary bind to the VDR 3′-UTR sequences, which may mediate a negative feedback loop controlling the bile acid signaling pathway in the early development of PSCs (Bai et al., 2015). In the present study, we observed a significant up-regulation at 1 days of SSD, further emphasizing that let-7 may regulate the expression of VDR and promote adult development in *E. granulosus*. Recently, an *in vitro* study showed that the exposure of albendazole to cultured PSCs significantly reduced let-7 expression and inhibited the differentiation of microcyst formation (Mortezaei et al., 2019); in contrast, higher let-7 expression and a lower growth inhibitory rate were observed in segmented worms, suggesting that let-7 may be an essential developmental regulator in the bi-directional development of *E. granulosus*.

Finally, although the synthesis of many *E. granulosus* proteins followed the expression of their corresponding transcripts, we observed a weak correlation between the transcriptome and newly synthesized proteins. This inconsistence may be ascribed to the following reasons: (1) the half-lives of the various proteins are very different. The level of a given protein is not only governed by protein synthesis, but is also dependent on the rates of degradation in different conditions; (2) some minor differences in RNA abundance may result in significant changes in the level of protein synthesis for a number of proteins. (3) post-transcriptional editing of miRNAs may also contribute to the differences evident between the transcriptome and the proteome.

In summary, the successful *in vitro* culture of PSCs represents an important model for studying the early development of *E. granulosus*. Using this model, we first characterized the dynamics of the mRNA transcriptome of cultured PSCs in the presence or absence of bile salt. We observed the bile salt stimulated significant effects on gene expression. The major gene expression changes occurring during strobilar development were related to carbohydrate metabolism and nervous system development. Additionally, we showed that the transcription of ion channel and ion transmembrane transporters was significantly increased during early cystic development. This feature supports the current view that most of the drugs for the effective treatment of CE target ion transmembrane transport in *E. granulosus*.

Furthermore, miRNA analysis revealed that 15 miRNAs had differentially expressed patterns between the two developmental (strobilar or cystic) directions. Among of these, miR-71 was highly expressed in *E. granulosus*, but not expressed in the mammalian host, and may thus be assessed in future as a candidate target for diagnositic marker for CE. Recently, a primary cell cultures of *E. multilocularis* found that 2'-O-methyl modified anti-miR-71 led to a significantly reduced level of miR-71 and abnormal development of PSCs (Perez et al., 2019). This outcome suggests that antisense oligonucleotides of miRNAs may be considered as a future clinical intervention strategy against echinococcosis. Therefore, understanding the transcriptional and regulatory characteristic of PSCs under the different *in vitro* culture conditions may be helpful in exploring the mechanisms governing cystic or adult worm development in *E. granulosus*. This information provides a new avenue to develop new interventions and therapeutics for the control of CE.

DATA AVAILABILITY STATEMENT

High throughput sequencing data have been submitted to the NCBI Sequence Read Archive (SRA) under accession number PRJNA610943.

AUTHOR CONTRIBUTIONS

WZ and SW coordinated the project. ZZ, LZ, BS, JL, and WZ participated in the *E. granulosus* sample preparation and RNA extraction. SW, YB, LJ, and YZ directed and performed the mRNA and small RNA sequencing. YB performed the sequence alignment and the novel miRNAs prediction, annotation, and expression analysis. YB, GG, BG, and LJ carried out the quantitative RT-PCR. SW, WZ, and DM commented on the mRNA and small RNA sequencing and analysis. YB, DM, SW, and WZ wrote the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (81271868, 31260272, and 81830066), the National Basic Research Program (973) of China (2010CB534906), the Shanghai Municipal Commission for Science and Technology (10XD1403200, 11DZ2292600, and 13DZ2291800), and the Shanghai Natural Science Foundation (13ZR1429500).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00654/full#supplementary-material

FIGURE S1 | Length distribution analysis of *E. granulosus* 3′-UTRs. The figure summarizes the length distribution of *E. granulosus* 3′-UTR. A major proportion of the 9,839 3′-UTRs (90.44%) were shorter than 4,000 nucleotides. The median length of the total set of 3′-UTRs was 1,027 nucleotides and the average GC content was 42.45%.

FIGURE S2 | Clustering of differentially expressed miRNAs during the *in vitro* development of PSCs. The differentially expressed miRNAs were clustered using the K-means method. Expression values were normalized and scaled between −1.0 and 1.0 (*Y*-axis). The *X*-axis indicates the time-points when RNA was subjected to transcriptome sequencing. A description of the pattern of expression belonging to each cluster is shown for each panel.

FIGURE S3 | Predicted targets of the differentially expressed miRNAs within the context of selected signaling pathways. **(A)** Wnt signaling pathway; **(B)** MAPK signaling pathway. Pink boxes represent gene orthologs present in *E. granulosus*. miRNAs that target genes in these pathways are shown in red.

TABLE S1 | Assembly of RNA-seq and alternatively spliced genes estimates. The table summarizes the counts of genes detected in RNA-seq. The percentages were calculated by dividing the counts of detected genes in all samples; AS, Alternatively splicing.

TABLE S2 | Raw data filtration and distribution of sequenced small RNAs across different categories. All small RNA reads were aligned to the reference *E. granulosus* genome and related RNA sequences using bowtie. One mismatch was permitted. The percentages were calculated by dividing the counts of reads matched to the genome. SSD: PSC with bile salt strobilization stimulus; NSD: PSC without bile salt strobilization stimulus.

TABLE S3 | All predicted novel miRNA precursors in *E. granulosus*. The table lists the sequences and MFE of candidate novel miRNAs in *E. granulosus*. If a miRNA candidate has multiple loci in the *E. granulosus* genome, we distinguish this candidate using a different number at the end of the miRNA names. Score, score of mirDeep; MFE, minimum free energy; Pre, miRNA precursors.

TABLE S4 | The differentially expressed genes in the developmental process of PSC with bile salt strobilization stimulus. The table lists all differentially expressed genes in the PSCs culture with bile salt strobilization stimulus. The statistical comparisons were determined by DESeq.

TABLE S5 | The differentially expressed genes in the developmental process of PSC without bile salt strobilization stimulus. The table lists all differentially expressed genes in the PSCs culture without bile salt strobilization stimulus. The statistical comparisons were determined by DESeq.

TABLE S6 | The differentially expressed genes between SSD and NSD. The table lists all differentially expressed genes between SSD and NSD. SSD: PSCs culture with bile salt strobilization stimulus; NSD, PSCs culture without bile salt strobilization stimulus.

TABLE S7 | Validation of the differentially expressed genes by real-time quantitative PCR. The table lists relative expression levels of 10 differentially expressed gene detected by RNAseq and qPCR. The qPCR fold changes were calculated by the $2^{-\Delta\Delta Ct}$ method.

TABLE S8 | Gene Ontology (GO) abundance analysis of the differentially expressed genes in SSD or NSD. The differentially expressed genes and ref represent the number of GO terms in the subgroup or whole gene group, respectively. The hypergeometric distribution test was used to find the significantly different GO term (*P*-value), and the false discovery rate (FDR) was calculated to correct the *P*-value (*Q*-value). SSD: PSCs culture with bile salt strobilization stimulus; NSD: PSCs culture without bile salt strobilization stimulus.

TABLE S9 | Comparison of Gene Ontology (GO) terms between SSD and NSD. The table lists the raw reads counts and relative expression levels of all detected mature miRNAs and miRNA stars in the three different life-stages. A statistical comparison was determined by DEGseq (Wang et al., 2010), and the q-value was the correction of the *P*-value with the Benjamini–Hochberg multiple test. PSC, protoscoleces.

TABLE S10 | Profiles of differentially expressed miRNAs from the different developmental stages of PSC with or without bile salt strobilization stimulus. The table lists the raw reads counts and relative expression levels of the differentially expressed mature miRNAs and miRNA stars in the different developmental stages of PSC with or without bile salt strobilization stimulus. A statistical comparison was determined by DEGseq (Wang et al., 2010), and the *Q*-value was the correction of the *P*-value with the Benjamini–Hochberg multiple test. The correct *P*-value < 0.001 and fold-change > 2.0 were used as the threshold criteria to define significant differences in miRNA expression. PSC, protoscoleces.

TABLE S11 | Clustering of the differentially expressed miRNAs from *in vitro* development of PSC. The table lists all differentially expressed miRNAs which have different patterns clustered by the K-means method.

TABLE S12 | Validation of the miRNA expression patterns by stem-loop real-time quantitative PCR. The table lists relative expression levels of 10 differentially expressed mature miRNAs and miRNA stars detected by NGS and stem-loop Q-PCR. The fold changes from stem-loop Q-PCR were calculated by the $2^{-\Delta\Delta Ct}$ method.

TABLE S13 | Target genes of the differentially expressed miRNAs. The table lists all potential target genes of the differentially expressed miRNAs predicted by Miranda (Enright et al., 2003). The score and energy represent the miRanda score and binding minimum free energy, respectively.

TABLE S14 | Nucleotide binding genes which targeted by the differentially expressed miRNAs. The table lists all nucleotide binding genes which regulated by the differentially expressed miRNAs. The score and energy represent the miRanda score and binding minimum free energy, respectively.

TABLE S15 | The differentially expressed miRNAs and their targets related with oxidation reduction process The table lists all differentially expressed gene which is related with oxidation reduction, some of them may regulated by the differentially expressed miRNAs.

TABLE S16 | Sequences of stem-loop RT primers, forward primers and reverse primer. The table lists all primer sequences used in the qPCR analysis. Normfinder was used to find the most stable expressed house-keeping genes under the test conditions. According to the results obtained, 5.8S rRNA ($M = 0.08$) and GAPDH ($M = 0.23$) were selected as reference genes for qPCR analysis. Furthermore, the PCR amplification efficiencies were calculated using the formula $[10^{(-1/S)}-1] \times 100$. Most of these were close to 100% (110%–90%).

REFERENCES

Abubucker, S., McNulty, S. N., Rosa, B. A., and Mitreva, M. (2014). Identification and characterization of alternative splicing in parasitic nematode transcriptomes. *Parasit Vectors* 7:151. doi: 10.1186/1756-3305-7-151

Andersen, F. L., Crellin, J. R., and Cox, D. D. (1981). Efficacy of praziquantel against immature *Echinococcus multilocularis* in dogs and cats. *Am. J. Vet. Res.* 42, 1978–1979.

Bai, Y., Zhang, Z., Jin, L., Kang, H., Zhu, Y., Zhang, L., et al. (2015). Genome-wide sequencing of small RNAs reveals a tissue-specific loss of conserved microRNA families in *Echinococcus granulosus*. *BMC Genomics* 15:736. doi: 10.1186/1471-2164-15-736

Banerjee, D., Chen, X., Lin, S. Y., and Slack, F. J. (2010). kin-19/casein kinase Ialpha has dual functions in regulating asymmetric division and terminal differentiation in *C. elegans* epidermal stem cells. *Cell Cycle* 9, 4748–4765. doi: 10.4161/cc.9.23.14092

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bollig, A., Xu, L., Thakur, A., Wu, J., Kuo, T. H., and Liao, J. D. (2007). Regulation of intracellular calcium release and PP1alpha in a mechanism for 4-hydroxytamoxifen-induced cytotoxicity. *Mol. Cell Biochem.* 305, 45–54. doi: 10.1007/s11010-007-9526-2

Boulias, K., and Horvitz, H. R. (2012). The *C. elegans* microRNA mir-71 acts in neurons to promote germline-mediated longevity through regulation of DAF-16/FOXO. *Cell Metab.* 15, 439–450. doi: 10.1016/j.cmet.2012.02.014

Buck, A. H., Coakley, G., Simbari, F., McSorley, H. J., Quintana, J. F., Le Bihan, T., et al. (2014). Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nat. Commun.* 5:5488. doi: 10.1038/ncomms6488

Cabrera, G., Espinoza, I., Kemmerling, U., and Galanti, N. (2010). Mesocestoides corti: morphological features and glycogen mobilization during in vitro differentiation from larva to adult worm. *Parasitology* 137, 373–384. doi: 10.1017/S0031182009991454

Cantacessi, C., Gasser, R. B., Strube, C., Schnieder, T., Jex, A. R., Hall, R. S., et al. (2010). Deep insights into *Dictyocaulus viviparus* transcriptomes provides unique prospects for new drug targets and disease intervention. *Biotechnol. Adv.* 29, 261–271. doi: 10.1016/j.biotechadv.2010.11.005

Carafoli, E. (1991). Calcium pump of the plasma membrane. *Physiol. Rev.* 71, 129–153.

Casaravilla, C., Malgor, R., and Carmona, C. (2003). Characterization of carbohydrates of adult *Echinococcus granulosus* by lectin-binding analysis. *J. Parasitol.* 89, 57–61.

Constantine, C. C., Bennet-Jenkins, E. M., Lymbery, A. J., Jenkins, D. J., Behm, C. A., and Thompson, R. C. (1998). Factors influencing the development and carbohydrate metabolism of *Echinococcus granulosus* in dogs. *J. Parasitol.* 84, 873–881.

Cucher, M., Macchiaroli, N., Kamenetzky, L., Maldonado, L., Brehm, K., and Rosenzvit, M. C. (2015). High-throughput characterization of *Echinococcus* spp. metacestode miRNomes. *Int. J. Parasitol.* 45, 253–267. doi: 10.1016/j.ijpara.2014.12.003

Cui, S. J., Xu, L. L., Zhang, T., Xu, M., Yao, J., Fang, C. Y., et al. (2013). Proteomic characterization of larval and adult developmental stages in *Echinococcus granulosus* reveals novel insight into host-parasite interactions. *J. Proteomics* 84, 158–175. doi: 10.1016/j.jprot.2013.04.013

de Souza Gomes, M., Muniyappa, M. K., Carvalho, S. G., Guerra-Sa, R., and Spillane, C. (2011). Genome-wide identification of novel microRNAs and their target genes in the human parasite *Schistosoma mansoni*. *Genomics* 98, 96–111. doi: 10.1016/j.ygeno.2011.05.007

Debarba, J. A., Monteiro, K. M., Moura, H., Barr, J. R., Ferreira, H. B., and Zaha, A. (2015). Identification of newly synthesized proteins by *Echinococcus granulosus* protoscoleces upon induction of strobilation. *PLoS Negl. Trop. Dis.* 9:e0004085. doi: 10.1371/journal.pntd.0004085

Dezaki, E. S., Yaghoubi, M. M., Spiliotis, M., Boubaker, G., Taheri, E., Almani, P. G., et al. (2016). Comparison of ex vivo harvested and in vitro cultured materials from *Echinococcus granulosus* by measuring expression levels of five genes putatively involved in the development and maturation of adult worms. *Parasitol. Res.* 115, 4405–4416. doi: 10.1007/s00436-016-5228-6

Diaz, A., Irigoin, F., Ferreira, F., and Sim, R. B. (1999). Control of host complement activation by the *Echinococcus granulosus* hydatid cyst. *Immunopharmacology* 42, 91–98.

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2003). MicroRNA targets in *Drosophila*. *Genome Biol.* 5:R1. doi: 10.1186/gb-2003-5-1-r1

Friedlander, M. R., Adamidi, C., Han, T., Lebedeva, S., Isenbarger, T. A., Hirst, M., et al. (2009). High-resolution profiling and discovery of planarian small RNAs. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11546–11551. doi: 10.1073/pnas.0905222106

Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52. doi: 10.1093/nar/gkr688

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Greenberg, R. M. (2005). Are Ca2+ channels targets of praziquantel action? *Int. J. Parasitol.* 35, 1–9. doi: 10.1016/j.ijpara.2004.09.004

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., and Smith, R. K. Jr. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31, 5654–5666.

Haenlin, M., Kunisch, M., Kramatschek, B., and Campos-Ortega, J. A. (1994). Genomic regions regulating early embryonic expression of the *Drosophila* neurogenic gene Delta. *Mech. Dev.* 47, 99–110.

Hao, L., Cai, P., Jiang, N., Wang, H., and Chen, Q. (2010). Identification and characterization of microRNAs and endogenous siRNAs in *Schistosoma japonicum*. *BMC Genomics* 11:55. doi: 10.1186/1471-2164-11-55

Heizer, E., Zarlenga, D. S., Rosa, B., Gao, X., Gasser, R. B., De Graef, J., et al. (2013). Transcriptome analyses reveal protein and domain families that delineate stage-related development in the economically important parasitic nematodes, *Ostertagia ostertagi* and *Cooperia oncophora*. *BMC Genomics* 14:118. doi: 10.1186/1471-2164-14-118

Hsieh, Y. W., Chang, C., and Chuang, C. F. (2012). The microRNA mir-71 inhibits calcium signaling by targeting the TIR-1/Sarm1 adaptor protein to control stochastic L/R neuronal asymmetry in *C. elegans*. *PLoS Genet.* 8:e1002864. doi: 10.1371/journal.pgen.1002864

Huang, J., Hao, P., Chen, H., Hu, W., Yan, Q., Liu, F., et al. (2009). Genome-wide identification of *Schistosoma japonicum* microRNAs using a deep-sequencing approach. *PLoS One* 4:e8206. doi: 10.1371/journal.pone.0008206

Jin, S. W., Arno, N., Cohen, A., Shah, A., Xu, Q., Chen, N., et al. (2001). In *Caenorhabditis elegans*, the RNA-binding domains of the cytoplasmic polyadenylation element binding protein FOG-1 are needed to regulate germ cell fates. *Genetics* 159, 1617–1630.

Khoo, K. H., Nieto, A., Morris, H. R., and Dell, A. (1997). Structural characterization of the N-glycans from *Echinococcus granulosus* hydatid cyst membrane and protoscoleces. *Mol. Biochem. Parasitol.* 86, 237–248. doi: 10.1016/s0166-6851(97)00036-4

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36. doi: 10.1186/gb-2013-14-4-r36

Koziol, U., Krohne, G., and Brehm, K. (2013). Anatomy and development of the larval nervous system in *Echinococcus multilocularis*. *Front. Zool.* 10:24. doi: 10.1186/1742-9994-10-24

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854. doi: 0092-8674(93)90529-Y

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Liu, C., Zhang, H., Yin, J., and Hu, W. (2015). *In vivo* and *in vitro* efficacies of mebendazole, mefloquine and nitazoxanide against cyst echinococcosis. *Parasitol. Res.* 114, 2213–2222. doi: 10.1007/s00436-015-4412-4

Macchiaroli, N., Cucher, M., Zarowiecki, M., Maldonado, L., Kamenetzky, L., and Rosenzvit, M. C. (2015). microRNA profiling in the zoonotic parasite *Echinococcus canadensis* using a high-throughput approach. *Parasit Vectors* 8:83. doi: 10.1186/s13071-015-0686-8

Macchiaroli, N., Maldonado, L. L., Zarowiecki, M., Cucher, M., Gismondi, M. I., Kamenetzky, L., et al. (2017). Genome-wide identification of microRNA targets in the neglected disease pathogens of the genus *Echinococcus*. *Mol. Biochem. Parasitol.* 214, 91–100. doi: 10.1016/j.molbiopara.2017.04.001

McManus, D. P., Zhang, W., Li, J., and Bartley, P. B. (2003). *Echinococcosis*. *Lancet* 362, 1295–1304. doi: 10.1016/S0140-6736(03)14573-4

Monteiro, K. M., de Carvalho, M. O., Zaha, A., and Ferreira, H. B. (2010). Proteomic analysis of the *Echinococcus granulosus* metacestode during infection of its intermediate host. *Proteomics* 10, 1985–1999. doi: 10.1002/pmic.200900506

Moro, P., and Schantz, P. M. (2009). *Echinococcosis*: a review. *Int. J. Infect. Dis.* 13, 125–133. doi: 10.1016/j.ijid.2008.03.037

Mortezaei, S., Afgar, A., Mohammadi, M. A., Mousavi, S. M., Sadeghi, B., and Harandi, M. F. (2019). The effect of albendazole sulfoxide on the expression of miR-61 and let-7 in different in vitro developmental stages of *Echinococcus granulosus*. *Acta Trop.* 195, 97–102. doi: 10.1016/j.actatropica.2019.04.031

Naderloo, Z., Farahnak, A., Golestani, A., Eshraghian, M. R., and Rad, M. B. M. (2016). *In vitro* study of mebendazole (Anthelmintic drug) Effects on the Aspartate Aminotransferase (AST) enzyme activity of hydatid cyst parasite. *J. Paramed. Sci.* 7, 41–47.

Nicolao, M. C., Elissondo, M. C., Denegri, G. M., Goya, A. B., and Cumino, A. C. (2014). In vitro and in vivo effects of tamoxifen against larval stage *Echinococcus granulosus*. *Antimicrob. Agents Chemother.* 58, 5146–5154. doi: 10.1128/AAC.02113-13

Palakodeti, D., Smielewska, M., and Graveley, B. R. (2006). MicroRNAs from the Planarian Schmidtea mediterranea: a model system for stem cell biology. *RNA* 12, 1640–1649. doi: 10.1261/rna.117206

Pan, C. L., Howell, J. E., Clark, S. G., Hilliard, M., Cordes, S., Bargmann, C. I., et al. (2006). Multiple Wnts and frizzled receptors regulate anteriorly directed cell and growth cone migrations in *Caenorhabditis elegans*. *Dev. Cell.* 10, 367–377. doi: 10.1016/j.devcel.2006.02.010

Parkinson, J., Wasmuth, J. D., Salinas, G., Bizarro, C. V., Sanford, C., Berriman, M., et al. (2012). A transcriptomic analysis of *Echinococcus granulosus* larval stages: implications for parasite biology and host adaptation. *PLoS Negl. Trop. Dis.* 6:e1897. doi: 10.1371/journal.pntd.0001897

Perez, M. G., Spiliotis, M., Rego, N., Macchiaroli, N., Kamenetzky, L., Holroyd, N., et al. (2019). Deciphering the role of miR-71 in *Echinococcus multilocularis* early development in vitro. *PLoS Negl. Trop. Dis.* 13:e0007932. doi: 10.1371/journal.pntd.0007932

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., et al. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403, 901–906. doi: 10.1038/35002607

Richards, K. S., Morris, D. L., Daniels, D., and Riley, E. M. (1988). *Echinococcus granulosus*: the effects of praziquantel, *in vivo* and *in vitro*, on the ultrastructure of equine strain murine cysts. *Parasitology* 96(Pt 2), 323–336.

Schmidt, J. (1998). Effects of benzimidazole anthelmintics as microtubule-active drugs on the synthesis and transport of surface glycoconjugates in *Hymenolepis microstoma*, *Echinostoma caproni*, and *Schistosoma mansoni*. *Parasitol. Res.* 84, 362–368.

Smith-Vikos, T., de Lencastre, A., Inukai, S., Shlomchik, M., Holtrup, B., and Slack, F. J. (2014). MicroRNAs mediate dietary-restriction-induced longevity through PHA-4/FOXA and SKN-1/Nrf transcription factors. *Curr. Biol.* 24, 2238–2246. doi: 10.1016/j.cub.2014.08.013

Smyth, J. D. (1990a). *In Vitro Cultivation of Parasitic Helminths*. Boca Raton, FL: CRC press.

Smyth, J. D. (1990b). Parasitological serendipity: from *Schistocephalus* to *Echinococcus*. *Int. J. Parasitol.* 20, 411–423. doi: 10.1016/0020-7519(90)90190-x

Smyth, J. D., Howkins, A. B., and Barton, M. (1966). Factors controlling the differentiation of the hydatid organism, *Echinococcus granulosus*, into cystic or strobilar stages *in vitro*. *Nature* 211, 1374–1377. doi: 10.1038/2111374a0

Sokol, N. S. (2012). Small temporal RNAs in animal development. *Curr. Opin. Genet. Dev.* 22, 368–373. doi: 10.1016/j.gde.2012.04.001

Thompson, R. C. (2017). Biology and systematics of *Echinococcus*. *Adv. Parasitol.* 95, 65–109. doi: 10.1016/bs.apar.2016.07.001

Wang, L., Feng, Z., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138. doi: 10.1093/bioinformatics/btp612

Wei, J., Cheng, F., Qun, Q., Nurbek, Xu, S. D., Sun, L. F., et al. (2005). Epidemiological evaluations of the efficacy of slow-released praziquantel-medicated bars for dogs in the prevention and control of cystic echinococcosis in man and animals. *Parasitol. Int.* 54, 231–236.

Wen, H., Vuitton, L., Tuxun, T., Li, J., Vuitton, D. A., Zhang, W., et al. (2019). Echinococcosis: advances in the 21st century. *Clin. Microbiol. Rev.* 32:e00075-18. doi: 10.1128/CMR.00075-18

Xiao, S., You, J., Mei, J., Jiao, P., Guo, H., Feng, J., et al. (1994). Early treatment with artemether and praziquantel in rabbits repeatedly infected with *Schistosoma japonicum* cercariae. *Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi* 12, 252–256.

Yang, J., Yang, L., Lv, Z., Wang, J., Zhang, Q., Zheng, H., et al. (2012). Molecular cloning and characterization of a HSP70 gene from *Schistosoma japonicum*. *Parasitol. Res.* 110, 1785–1793. doi: 10.1007/s00436-011-2700-1

Yook, K., Harris, T. W., Bieri, T., Cabunoc, A., Chan, J., Chen, W. J., et al. (2012). WormBase 2012: more genomes, more data, new website. *Nucleic Acids Res.* 40, D735–D741. doi: 10.1093/nar/gkr954

Zeghir-Bouteldja, R., Polome, A., Bousbata, S., and Touil-Boukoffa, C. (2017). Comparative proteome profiling of hydatid fluid from Algerian patients reveals cyst location-related variation in *Echinococcus granulosus*. *Acta Trop.* 171, 199–206. doi: 10.1016/j.actatropica.2017.03.034

Zhang, C., Li, J., Aji, T., Li, L., Bi, X., Yang, N., et al. (2019). Identification of functional MKK3/6 and MEK1/2 homologs from *Echinococcus granulosus* and investigation of protoscolecidal activity of mitogen-activated protein kinase signaling pathway inhibitors *in vitro* and *in vivo*. *Antimicrob. Agents Chemother.* 63:e01043-18. doi: 10.1128/AAC.01043-18

Zhang, J., Ye, B., Kong, J., Cai, H., Zhao, Y., Han, X., et al. (2013). In vitro protoscolicidal effects of high-intensity focused ultrasound enhanced by a superabsorbent polymer. *Parasitol. Res.* 112, 385–391. doi: 10.1007/s00436-012-3176-3

Zhang, W. B., Jones, M. K., Li, J., and McManus, D. P. (2005). *Echinococcus granulosus*: pre-culture of protoscoleces in vitro significantly increases development and viability of secondary hydatid cysts in mice. *Exp. Parasitol.* 110, 88–90. doi: 10.1016/j.exppara.2005.02.003

Zheng, H., Zhang, W., Zhang, L., Zhang, Z., Li, J., Lu, G., et al. (2013). The genome of the hydatid tapeworm *Echinococcus granulosus*. *Nat. Genet.* 45, 1168–1175. doi: 10.1038/ng.2757

Zheng, Y., Cai, X., and Bradley, J. E. (2013). microRNAs in parasites and parasite infection. *RNA Biol.* 10, 371–379. doi: 10.4161/rna.23716

Zheng, Y. (2013). Strategies of *Echinococcus* species responses to immune attacks: implications for therapeutic tool development. *Int. Immunopharmacol.* 17, 495–501. doi: 10.1016/j.intimp.2013.07.022

Zheng, Y., Guo, X., He, W., Shao, Z., Zhang, X., Yang, J., et al. (2016). Effects of *Echinococcus* multilocularis miR-71 mimics on murine macrophage RAW264.7. *Cells Int. Immunopharmacol.* 34, 259–262. doi: 10.1016/j.intimp.2016.03.015

Zidek, R., Machon, O., and Kozmik, Z. (2018). Wnt/beta-catenin signalling is necessary for gut differentiation in a marine annelid, *Platynereis dumerilii*. *Evodevo* 9:14. doi: 10.1186/s13227-018-0100-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 16 April 2020
doi: 10.3389/fmicb.2020.00703

TRIM21 Aggravates Herpes Simplex Virus Epithelial Keratitis by Attenuating STING-IRF3-Mediated Type I Interferon Signaling

*Tianchang Tan and Likun Xia**

Department of Ophthalmology, Shengjing Hospital of China Medical University, Shenyang, China

Herpes simplex virus-1 (HSV-1) is the leading cause of infectious blindness in the developed world. HSV-1 infection can occur anywhere in the eye, and the most common presentation is epithelial keratitis. In the HSV epithelial keratitis mice model, we detected the expression of TRIM21 and then investigated the clinical relationship between TRIM21 and HSV epithelial keratitis by silencing TRIM21. Through the clinical scores and histopathology examination, we found that TRIM21 can effectively reduce the severity of HSV epithelial keratitis. Furthermore, silencing TRIM21 significantly controlled the virus particle release at 1, 3, and 5 days post-HSV-1 infection. Notably, the production of IFN-β was enhanced, and the secretion of pro-inflammatory cytokines (IL-6 and TNF-a) was inhibited. Next, human corneal epithelial cells were pretreated with lentivirus or siRNA, respectively, so that TRIM21 expression was overexpressed or silenced. We focused on the regulation of STING-IRF3 and type I interferon signaling after infected with HSV-1. In conclusion, our results have identified that TRIM21 is abnormally high expressed in HSV epithelial keratitis. TRIM21 enhances the replication of HSV-1 in corneal epithelial cells via suppressing the production of type I IFN by inhibiting STING/IRF3 signaling. It also promotes the secretion of IL-6 and TNF-a, thereby aggravating the severity of HSV epithelial keratitis.

Keywords: herpes simplex virus-1, keratitis, TRIM21, type I interferon, STING

OPEN ACCESS

Edited by:
Benjamin Makepeace, University of Liverpool, United Kingdom

Reviewed by:
Sharvan Sehrawat, Indian Institute of Science Education and Research Mohali, India
Chao Qiu, Fudan University, China

***Correspondence:**
Likun Xia
xialk@sj-hospital.org

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *10 January 2020*
Accepted: *26 March 2020*
Published: *16 April 2020*

Citation:
Tan T and Xia L (2020) TRIM21 Aggravates Herpes Simplex Virus Epithelial Keratitis by Attenuating STING-IRF3-Mediated Type I Interferon Signaling. Front. Microbiol. 11:703. doi: 10.3389/fmicb.2020.00703

INTRODUCTION

Herpes simplex virus-1 (HSV-1) is a neurotropic double-stranded DNA virus, which is the leading cause of infectious blindness in the developed world (Rechenchoski et al., 2017). Epidemiology shows that HSV-1 has been recognized as a ubiquitous human pathogen, infecting 50–90% of the world population (Smith and Robinson, 2002). Eye disease caused by HSV-1 infection usually presents as epithelial keratitis, which accounts for 50–80% of ocular herpes (Labetoulle et al., 2005). Although HSV epithelial keratitis is self-limiting within one week, without adequate treatment, it may progress to stromal keratitis, leading to progressive corneal opacity (Wilhelmus, 2015).

Primary infection is usually caused by direct infection of the cornea by HSV-1. After infection, HSV-1 replicates in the corneal epithelial cells and then triggers innate immune signaling through the production of cytokines and chemokines. Subsequently, immune cells are recruited to the

site of primary infection, predominantly neutrophils and dendritic cells, which at this stage function to limit viral spread within corneal epithelial cells (Lobo et al., 2019). At present, antiviral medicines, interferon drops and superficial wiping have been used to cure HSV epithelial keratitis. However, controversies persist about the side effects of antiviral drugs, the effectiveness of topical interferons and the safety of debridement methods (Wilhelmus, 2015). The treatment objective of HSV epithelial keratitis is to inhibit viral replication and to prevent the active viral infection of corneal cells.

Type I interferon is one of the first cytokines secreted upon HSV-1 infection of the cornea, which can limit virus transmission by inhibiting virus replication and increasing the resistance of neighboring cells (Ellermann-Eriksen, 2005). Several studies have shown that type I interferon is necessary for controlling the replication of HSV-1 in the cornea, and it is also required for immune cells to recruit to the site of infection (Hendricks et al., 1991; Leib et al., 1999; Conrady et al., 2011b). Therefore, enhancing the ability of host cells to produce type I interferon may be the best treatment for HSV epithelial keratitis.

The tripartite motif (TRIM) protein family consists of up to 100 members and is the largest group of E3 ubiquitin ligases in mammals (Su et al., 2016). Many TRIMs proteins have been reported to have direct or indirect antiviral activity (Khan et al., 2019). TRIM21, also known as Ro52/SS-A or RNF81, has a B30.2 domain encoded in the C-terminal region, which comprises a combination of a PRY motif followed by a SPRY motif (Rhodes et al., 2005). B30.2 domain can be recruited to incoming viral cores and determines antiviral specificity (Towers, 2007). Recent studies have shown that TRIM21 can recruit proteasomes and savagely break down capsids so that the exogenous viral genomes are exposed prematurely to promote the activity of PRRs (Watkinson et al., 2015). HSV-1 belongs to an enveloped virus. Thus we suspect that TRIM21 may affect HSV-1 infection. In this study, we have investigated the role of TRIM21 in the host defense against HSV-1 infection in a murine model of HSV epithelial keratitis and explored its underlying mechanism in human corneal epithelial (HCE) cells.

MATERIALS AND METHODS

Mice, Cell, and Virus

Female 6-week old C57BL/6J mice were purchased from Liaoning Changsheng Biotechnology Company (Benxi, China) and maintained in the specific pathogen-free environment. All investigations followed guidelines of the Institutional Animal Care and Use Committee of China Medical University and adhered to the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research. During experimental procedures, suffering minimized by isoflurane anesthesia. The HCE cells (BNCC341100) harvested in Dulbecco's modified Eagle's medium (DMEM, Hyclone, China) with 10% fetal bovine serum (FBS, Hyclone). The Vero cells (ATCCCCL-81) maintained in Minimal Essential Media (MEM, Hyclone) supplemented with 10% FBS. The cells were maintained at 37°C with 5% CO_2 in a humidified atmosphere. HSV-1 Mckrae strain was propagated and titrated in Vero cell monolayers, followed by stored in aliquots at −80°C until used as previously described (Stremlau et al., 2004).

Mouse HSV Epithelial Keratitis Model and Pre-treatment

Development of HSV epithelial keratitis in C57BL/6J mice was studied with HSV-1 McKrae strain and $10^{6.6}$. $TCID_{50}$/ml of the virus was applied to the right cornea in 5 μl MEM after abrading in a cross-shaped pattern with a sterile 27-gage needle. While the control group was instead received 5 μl MEM, leaving them uninfected. For the pre-treatment, 3 μmol of siRNA-TRIM21 or siRNA-control were subconjunctivally injected into the right eye once a day for three times before establishing HSV epithelial keratitis models. The dosages used in the experiment were evaluated in the pre-experiment.

Clinical Observation and Scoring

Corneas were infected with HSV-1 and scored the severity of HSV epithelial keratitis by slit-lamp biomicroscopy (Kawa Co., Japan) in a blinded manner every day. The clinical scoring system was as follows (Komoto et al., 2015): 0, entire epithelial area intact; +1, diffuse punctate lesion; +2, dendritic lesion occupying less than 1/4 of the entire epithelial area; +3, severe dendritic lesion extending more than 1/4 of the entire epithelial area; +4, geographic lesion on the epithelial area.

Virus Titers Assay

Swabs harvested tear samples on the 1, 3, and 5 days after infection. Then all the swabs with tear samples were stocked in 1ml of MEM with 2% FBS and stored at −80°C until used. Titers were determined by $TCID_{50}$ assay on Vero cell monolayers with standard methodology. Briefly, cell culture lysis supernatants were diluted serially using 10-fold dilutions and titered on Vero cell monolayers by the $TCID_{50}$ assay.

Western Blotting Assay

The total protein from mouse corneas (3 corneas/sample/group) and HCE cells was extracted and prepared in a standardized manner for Western blotting. After SDS-PAGE, transfer and blocking with 5% non-fat milk at room temperature for an hour on a shaker, membranes were incubated overnight at 4°C with anti-TRIM21 antibody (Abcam, United Kingdom, used at 1:1000), STING (D2P2F) Rabbit mAb (Cell Signaling Technology, China, 1:1000), IRF-3 Rabbit mAb (Cell Signaling Technology, 1:1000), phospho-IRF3 Rabbit mAb (Cell Signaling Technology, 1:1000) and GAPDH antibody (Wanlei, China, 1:1000) as primary antibodies. Followed by the membranes were washed three times in Tris-buffered saline. Then, the membranes incubated with goat horseradish peroxidase-conjugated anti-mouse or anti-rabbit antibodies (Beyotime, China, 1:5000) as secondary reagents for 2 h at room temperature. Finally, an enhanced chemiluminescence kit (Wanlei, China) was utilized to visualize the membrane. The densitometry analysis was performed using ImageJ 6.0 software.

ELISA

Corneas (3 corneas/sample/group) were cut in small pieces and homogenized in 0.5 mL of PBS with 0.1% Tween 20. All samples were centrifuged at 13000 × *g* for 15 min to remove tissue debris. Then, levels of IL-6 and TNF-a in supernatants were detected by ELISA kits according to the manufacturer's instructions. ELISA kits were all purchased from R&D Systems.

Quantitative PCR Analysis

Total mRNA from infected corneas and cells were extracted using the RNAiso Plus kit (Takara, Japan), and then stored at -80°C until used. Total mRNA was reverse-transcribed using PrimeScript^{TM} RT reagent Kit with gDNA Eraser to generate cDNA (Takara, Japan). Quantitative PCR (qPCR) was performed using an SYBR Green (Takara, Japan) format with 7500 Fast Real-Time PCR Detection Systems (Life Technology). The sequences of the qPCR primers used in the analyses are summarized in **Table 1**. The expression levels of different molecules were normalized to GAPDH using ΔCt calculation.

Histopathology and Immunofluorescence Staining

Mice were sacrificed by cervical dislocation at the indicated time point, and then corneas were collected into 1.5 ml of 4% paraformaldehyde for fixation. Paraffin-embedded corneas were then sectioned at 5 μm thickness. The sections were incubated with H&E staining to determine the pathologic changes of corneas. For immunofluorescence staining, the sections were incubated with diluted TRIM21 primary antibody (Santa, United Kingdom, 1:150) overnight at 4°C. After washing thrice with PBS (10 min each time), fluorescence-conjugated secondary antibody (Santa, United Kingdom, 1:50) working solution were added and incubated at 37°C for 2 h. Subsequently, DAPI (Invitrogen, United Kingdom) was added for 5 min to visualize the nuclei. Finally, the fluorescence signal was observed under the fluorescence microscope (Olympus, Japan).

Statistical Analysis

Three independent experiments were performed for each assay. All results were expressed as mean ± standard deviation and analyzed by Student's *t*-test. Rank data comparison using the rank-sum test. The statistics are presented using GraphPad Prism five to evaluate the differences. A p-value < 0.05 was considered to be statistically significant. All methods were performed following the relevant guidelines and regulations.

TABLE 1 | List of primers.

Gene	Forward primer (5′–3′)	Reverse primer (5′–3′)
TRIM21	AGAGAGACTTCACCTGTTCTGT	TCAGTTCCCCTAATGCCACCT
IFN-β	CCCTATGGAGATGACGGAGA	CTGTCTGCTGGTGGAGTTCA
GAPDH	TGGATTTGGACGCATTGGTC	TTTGCACTGGTACGTGTTGAT

RESULTS

Expression of TRIM21 in Murine Corneas in Response to HSV-1 Infection

To explore the potential role of TRIM21 in HSV epithelial keratitis, its expression was measured in corneas of epithelial keratitis mice at 0, 2, and 4 days post-HSV-1 infection (dpi). As shown in **Figures 1A,B**, compared with uninfected corneas, HSV-1 infected corneas displayed abnormally elevated TRIM21 level. Overall, as the course of HSV epithelial keratitis progressed, the expression of TRIM21 continued to increase. To further confirm whether TRIM21 was involved in the pathological process of HSV epithelial keratitis, we detected the localization of TRIM21 in corneas by immunofluorescence, which labeled TRIM21 with green fluorescence (**Figure 1C**). Results showed that TRIM21 was predominantly expressed in the corneal epithelium at 0 dpi compared with the expression of TRIM21 was increased significantly at 3 dpi. TRIM21 mainly expressed in the cytoplasm of corneal epithelial cells.

The Contribution of TRIM21 in HSV Epithelial Keratitis

To further ascertain the role of TRIM21 in HSV epithelial keratitis, siRNA transfection was used to limit TRIM21 expression in the corneas before establishing the HSV epithelial

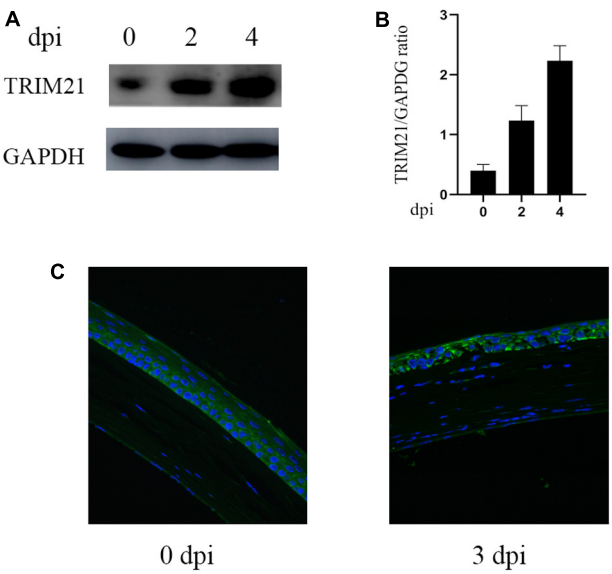


FIGURE 1 | Abnormal expression of TRIM21 in C57BL/6 mouse cornea after HSV-1 infection. **(A)** The protein expression level of TRIM21 was examined by western blotting in corneas at 0, 2, and 4 days post HSV-1 infection (dpi). Three corneas were used per experiment. **(B)** Relative integrated density values quantitated TRIM21 protein levels after normalization to GAPDH. Data were representative of three individual experiments and represented as the mean ± SD. **(C)** At 0 and 3 dpi, the expression of corneal TRIM21 was evaluated by immunofluorescent staining. TRIM21 staining showed green fluorescence, and nuclear staining showed blue fluorescence. Magnification: 400×.

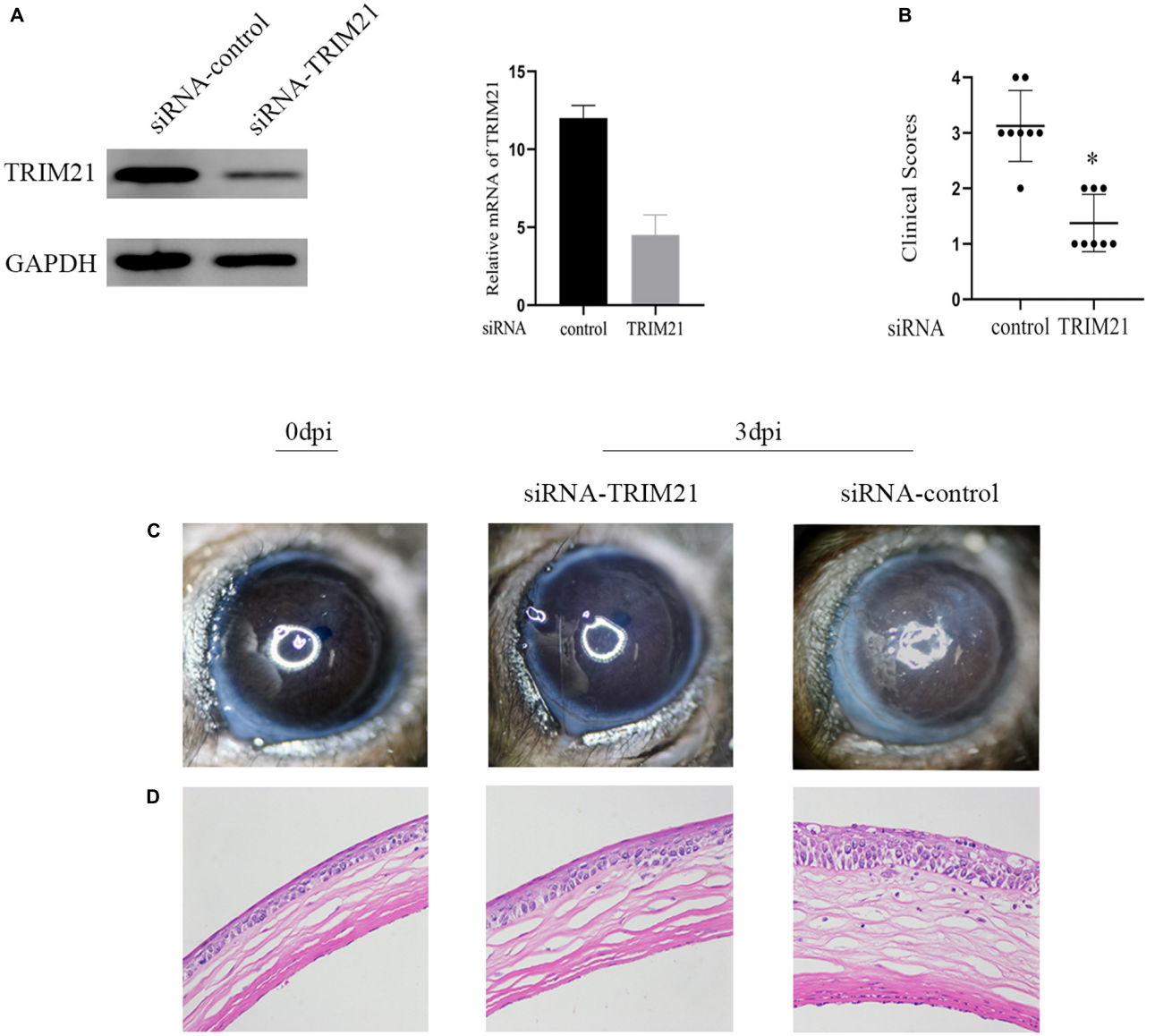


FIGURE 2 | Silencing TRIM21 alleviates the severity of HSV epithelial keratitis. **(A)** The protein (left) and mRNA (right) levels of TRIM21 in corneas were assessed to confirm the transfection efficiency of siRNA before establishing HSV epithelial keratitis models. **(B)** HSV epithelial keratitis clinical scores at 3 dpi. $n = 8$ per time point. **(C)** Representative photographs of corneas from each group of mice at 3 dpi corresponding to **(D)** histopathology results. Magnification: 400×.

keratitis mice model. Knockdown efficiency (**Figure 2A**) was assessed by the protein and mRNA levels of TRIM21 in corneas. After that, corneas were infected with HSV-1 and then examined to score the severity of HSV epithelial keratitis by hand-held slit lamp microscope in a blinded manner every day. We found that the clinical scores and the degree of corneal opacity in the siRNA-TRIM21 group were significantly lower than that in the siRNA-control group ($p < 0.05$, **Figures 2B,C**). **Figure 2B** showed that the number of corneal lesion score of ≤2 in the siRNA-TRIM21 group (8/8) was more than the number in the siRNA-control group (1/8). As shown in **Figure 2D**, compared with histopathology in the siRNA-TRIM21 treated corneas, corneal epithelial cells in the siRNA-control treated corneas showed obvious proliferation and growth disorders. Corneal epithelial layer showed vacuole-like changes, and a small number of inflammatory cells infiltrated into the superficial layer of the corneal stroma. These results indicated that silencing TRIM21 can reduce the severity of HSV epithelial keratitis.

Silencing TRIM21 Controls HSV-1 Replication in HSV-1 Infected Corneas

Tear samples from the siRNA-TRIM21 and siRNA-control treatment groups were subjected to $TCID_{50}$ assay to investigate the effect of TRIM21 on virus replication. As shown in **Figure 3A**,

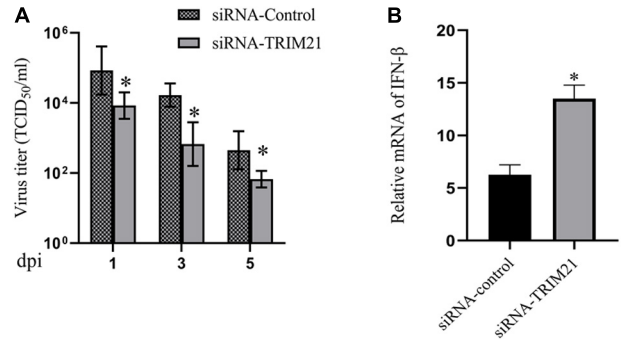


FIGURE 3 | Silencing TRIM21 controls HSV-1 replication in HSV epithelial keratitis. **(A)** Viral titers in tear films from siRNA-TRIM21 and siRNA-control treatment mice were determined by $TCID_{50}$ assay on days 1, 3, and 5 post-infection. **(B)** The relative mRNA level of IFN-β in corneas from each group measured by qPCR. $n = 3$ per group. Data are shown as mean ± SD of three independent experiments. $^*P < 0.05$.

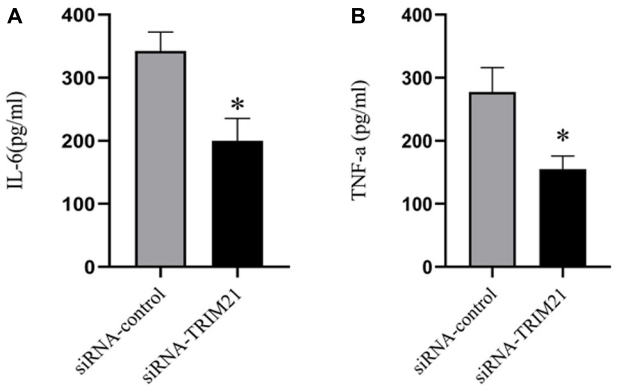


FIGURE 4 | Silencing TRIM21 reduces pro-inflammatory cytokines in HSV epithelial keratitis. ELISA measured the protein expression levels of corneal IL-6 **(A)** and TNF-a **(B)**. $n = 3$ per group. Data are shown as mean ± SD of three independent experiments. $^*P < 0.05$.

silencing TRIM21 controlled the virus particle release at 1, 3, and 5 dpi significantly. Type I interferon signaling is identified as a critical step for controlling viral replication. Therefore, we further examined the transcript level of IFN-β in corneas at 3 dpi. The qPCR results showed that the IFN-β expression in the siRNA-TRIM21 treated corneas was approximately twice that of the siRNA-control treated corneas (**Figure 3B**).

Silencing Trim21 Reduces Pro-inflammatory Cytokines in Hsv-1 Infected Corneas

To determine the effects of TRIM21 on pro-inflammatory cytokines in HSV epithelial keratitis, three corneas from the siRNA-TRIM21 and siRNA-control treatment groups were pooled to obtain one sample for analysis of IL-6 and TNF-a expression levels by ELISA. As shown in **Figure 4**, the expression of IL-6 and TNF-a were both increased in HSV epithelial keratitis at 3 dpi. Compared with the siRNA-control treated corneas, the expression of these two pro-inflammatory cytokines in the siRNA-TRIM21 treated corneas was significantly suppressed.

TRIM21 Inhibited STING-IRF3 Signaling in HSV-1 Infected HCE Cells

Next, we explored the mechanism of TRIM21 affecting HSV-1 replication in human corneal epithelial cells (HCE cells). To interfere with TRIM21 expression, lentivirus-TRIM21 (LV-TRIM21), LV-control, siRNA-TRIM21, and siRNA-control were used to transfect HCE cells, respectively. First of all, transfection efficiency (**Figure 5A**) was assessed by the protein and mRNA levels of TRIM21 in HCE cells. Subsequently, HSV-1 infected the pretreated cells at MOI of 10. qPCR results showed that the expression of IFN-β was enhanced in siRNA-TRIM21 pretreated cells and suppressed in LV-TRIM21 pretreated cells (**Figure 5B**). In order to explore the mechanism of TRIM21 on the production of IFN-β in HSV-1 infected HCE cells, Western Blotting was used to detect the STING/IRF3 signaling pathway. As shown in **Figure 5C**, the expression of STING and phosphorylated IRF3 was enhanced in siRNA-TRIM21 pretreated cells after HSV-1 infection 12 h. Conversely, the expression of STING and phosphorylated IRF3 was reduced in LV-TRIM21 pretreated cells. When we pretreated HCE cells with both siRNA-TRIM21 and siRNA-STING, qPCR results showed that the level of IFN-β production decreased significantly compared to siRNA-TRIM21 pretreated cells alone (**Figure 5D**). The above results showed that in HSV-1 infected HCE cells, TRIM21 regulates the production of type I interferon by regulating the STING/IRF3 signaling pathway.

DISCUSSION

B30.2 domain in the TRIM21 protein structure can identify the invading virus so that TRIM21 plays a vital role in the process of virus infection. HSV-1 infection can occur anywhere in the eye, and the most common presentation is epithelial keratitis, which usually self-healing within a week after infection. However, the underlying mechanism of TRIM21 in HSV epithelial keratitis remains unclear. Our study demonstrated that TRIM21 is abnormally high expressed in HSV epithelial keratitis and aggravates the severity of epithelial keratitis by promoting type I interferon signaling.

Previous studies have shown that TRIMs protein can be expressed in most normal tissues (Hatakeyama, 2017). This study is the first to confirm that TRIM21 is constitutively expressed in the cornea of C57BL/6J mice without pathogen infection. The initial manifestation of HSV epithelial keratitis is punctate keratopathy. As the virus spreads uncontrollably, the disease progresses from punctate keratitis to dendritic keratitis or a geographic ulcer. HSV-1 McKrae strain ($10^{6.6}$ $TCID_{50}$/ml) was applied to the right cornea of C57BL/6J mice. We found that the expression of TRIM21 in corneas was abnormally increased at 2 and 4 days post HSV-1 infection (dpi).

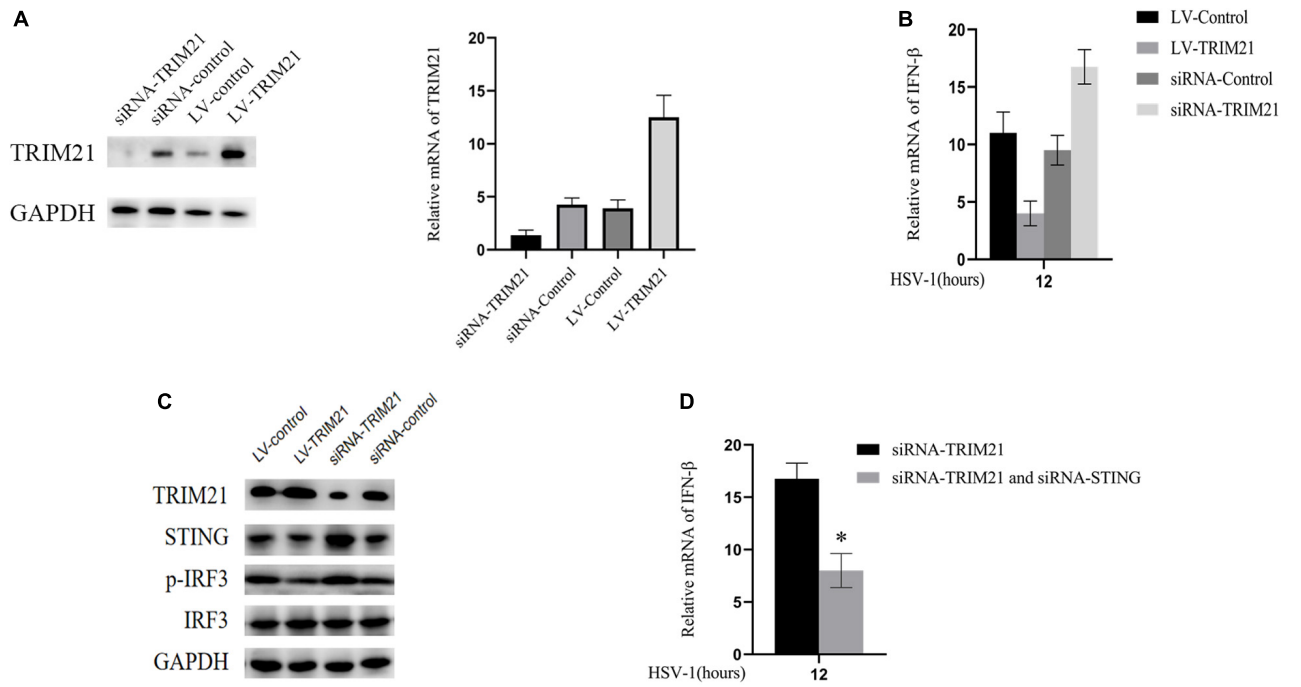


FIGURE 5 | TRIM21 inhibited STING-IRF3 signaling in HSV-1-infected HCE cells. **(A)** The protein (left) and mRNA levels (right) of TRIM21 in HCE cells were assessed to confirm the transfection efficiency of lentivirus and siRNA before HSV-1 infection. **(B)** The mRNA level of IFN-β in each group was examined after HSV-1 infection 12 h. **(C)** The protein expression levels of TRIM21, STING, phosphorylation of IRF3, and IRF3 were examined after HSV-1 infection 12 h. **(D)** HCE cells pretreated with both siRNA-TRIM21 and siRNA-STING. After HSV-1 infection 12 h, qPCR detected the production of IFN-β. Data are shown as mean ± SD of three independent experiments. $^*P < 0.05$.

Immunofluorescence results revealed that, at 3 dpi, TRIM21 was located in the cytoplasm of corneal epithelial cells and compared with the normal cornea, its expression was enhanced with strong fluorescence. These results suggest that TRIM21 is involved in the development of HSV epithelial keratitis.

To further ascertain the role of TRIM21 in HSV epithelial keratitis, siRNA was used to limit TRIM21 expression in the corneas before establishing the HSV epithelial keratitis mice model. Through the clinical scores and histopathology examination, we found that TRIM21 can effectively reduce the severity of HSV epithelial keratitis. Previous research has shown that the severity of HSV epithelial keratitis is related to the virus clearance rate (Babu et al., 1996; Twardy et al., 2011). In this study, silencing TRIM21 controlled the virus particle release at 1, 3, and 5 dpi significantly, indicating that silencing TRIM21 increased the virus clearance rate. Once HSV-1 infects corneal epithelial cells, recognition of viral ligands by pattern recognition receptors (PRRs) in host cells will elevate the transcription of various interleukins and tissue necrosis factor to clear the virus (Akira et al., 2001; Akira and Takeda, 2004). At the same time, the inflammatory cascade induces an excessive immune response leading to progressive corneal opacification. At 3 dpi, we found that the expression of pro-inflammatory cytokines (IL-6 and TNF-a) in the siRNA-TRIM21 treated corneas was significantly repressed. Consistent with previous reports, IL-6 and TNF-a are the major drivers of corneal inflammation, and various strategies to block these cytokines have proven effective in decreasing the severity of HSV epithelial keratitis in mice (Rajasagi et al., 2017).

Cytosolic DNA sensors are the most recently described class of PRRs, which include cGAS (Li et al., 2013), AIM2 (Fernandes-Alnemri et al., 2009), DAI (Takaoka et al., 2007), DDX41 (Zhang et al., 2011) and IFI16 (Unterholzner et al., 2010). Viral nucleic acids of HSV-1, recognized by various PRRs, can act as activators of type I interferons signaling pathways so that promote antiviral immune responses. Several studies have identified that the antiviral response mediated by type I interferon is a key way to limit viral replication (Gresser et al., 1979; Zawatzky et al., 1982; Ellermann-Eriksen et al., 1986). In the absence of type I interferon signaling (interferon receptor A1 knocked out and CD118 −/−), the virus replication ability was enhanced, and the mortality of mice was increased, which highlighted the crucial defensive role of type I interferon in the control of HSV-1 infection (Leib et al., 1999; Luker et al., 2003; Conrady et al., 2011a). In the research of CVB3 and HBV, TRIM21 functions as a positive regulator of type I interferon, promoting a comprehensive antiviral response (Liu et al., 2018; Zhang et al., 2018). However, in the research of JEV infection, TRIM21 functions as a negative regulator of type I interferon, promoting JEV virus replication (Manocha et al., 2014). In our study, qPCR results showed that silencing TRIM21 enhanced type I interferon signaling at 3 days post HSV-1 infection.

The stimulator of interferon genes (STING) acts as a converging point of cytosolic DNA receptors to promote signal

transmission to downstream effector IRF3, leading to the expression of type I interferon (Jiang et al., 2017). In order to survive in host cells, HSV-1 encodes a variety of proteins that inhibit STING-IRF3 signaling to prevent host cells from producing type I interferon (Christensen et al., 2016). It means that STING-IRF3 signaling is critical in host defense against the virus (Abe et al., 2013; Barber, 2014). In the experiment of HCE cells infected with HSV-1, we found that up-regulating TRIM21 inhibited STING-IRF3 signaling and reduced IFN-β expression, while silencing TRIM21 enhanced STING-IRF3 signaling and increased IFN-β expression. HCE cells were afterward pretreated with siRNA to silence both TRIM21 and STING at the same time. We found that the expression of IFN-β relatively decreased compared to the siRNA-TRIM21 treated alone. Thus, these results show that the effect of TRIM21 on type I interferon is mediated by STING-IRF3 signaling.

Collectively, our results have identified conclusions as follows. TRIM21 is constitutively expressed in normal corneas and is abnormally high expressed in HSV epithelial keratitis. Highly expressed TRIM21 enhances the replication of HSV-1 in corneal epithelial cells via suppressing the production of type I IFN by inhibiting STING/IRF3 signaling. It also promotes the production of pro-inflammatory cytokines IL-6 and TNF-a to the point of synergistically exacerbating the severity of HSV epithelial keratitis.

DATA AVAILABILITY STATEMENT

The data used to support the findings of this study are available from the corresponding author upon request.

ETHICS STATEMENT

The animal study was reviewed and approved by the Ethics Committee of Shengjing Hospital, China Medical University (Ethics Number: 2018PS128K).

AUTHOR CONTRIBUTIONS

LX and TT contributed to the design of the experiments. TT contributed to the conduction of the experiments and editing of the manuscript.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (No: 30772394) and the Provincial Natural Science Foundation of Liaoning (No: 20180530083).

REFERENCES

Abe, T., Harashima, A., Xia, T., Konno, H., Konno, K., Morales, A., et al. (2013). STING recognition of cytoplasmic DNA instigates cellular defense. *Mol. Cell* 50, 5–15. doi: 10.1016/j.molcel.2013.01.039

Akira, S., and Takeda, K. (2004). Toll-like receptor signalling. *Nat. Rev. Immunol.* 4, 499–511. doi: 10.1038/nri1391

Akira, S., Takeda, K., and Kaisho, T. (2001). Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat. Immunol.* 2, 675–680. doi: 10.1038/90609

Babu, J. S., Thomas, J., Kanangat, S., Morrison, L. A., Knipe, D. M., and Rouse, B. T. (1996). Viral replication is required for induction of ocular immunopathology by herpes simplex virus. *J. Virol.* 70, 101–107. doi: 10.1128/jvi.70.1.101-107.1996

Barber, G. N. (2014). STING-dependent cytosolic DNA sensing pathways. *Trends Immunol.* 35, 88–93. doi: 10.1016/j.it.2013.10.010

Christensen, M. H., Jensen, S. B., Miettinen, J. J., Luecke, S., Prabakaran, T., Reinert, L. S., et al. (2016). HSV -1 ICP 27 targets the TBK 1-activated STING signalsome to inhibit virus-induced type I IFN expression. *EMBO J.* 35, 1385–1399. doi: 10.15252/embj.201593458

Conrady, C. D., Halford, W. P., and Carr, D. J. J. (2011a). Loss of the type i interferon pathway increases vulnerability of mice to genital herpes simplex virus 2 infection. *J. Virol.* 85, 1625–1633. doi: 10.1128/jvi.01715-1710

Conrady, C. D., Jones, H., Zheng, M., and Carr, D. J. J. (2011b). A functional type I interferon pathway drives resistance to cornea herpes simplex virus type 1 infection by recruitment of leukocytes. *J. Biomed. Res.* 25, 111–119. doi: 10.1016/S1674-8301(11)60014-6

Ellermann-Eriksen, S. (2005). Macrophages and cytokines in the early defence against herpes simplex virus. *Virol. J.* 2:59. doi: 10.1186/1743-422X-2-59

Ellermann-Eriksen, S., Liberto, M. C., Iannello, D., and Mogensen, S. C. (1986). X-linkage of the early in vitro alpha/beta interferon response of mouse peritoneal macrophages to herpes simplex virus type 2. *J. Gen. Virol.* 67(Pt 6), 1025–1033. doi: 10.1099/0022-1317-67-6-1025

Fernandes-Alnemri, T., Yu, J. W., Datta, P., Wu, J., and Alnemri, E. S. (2009). AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature* 458, 509–513. doi: 10.1038/nature07710

Gresser, I., Maury, C., Kress, C., Blangy, D., and Maunoury, M. T. (1979). Role of interferon in the pathogenesis of virus diseases in mice as demonstrated by the use of anti-interferon serum. VI. Polyoma virus infection. *Int. J. Cancer* 24, 178–183. doi: 10.1002/ijc.2910240209

Hatakeyama, S. (2017). TRIM family proteins: roles in autophagy, immunity, and carcinogenesis. *Trends Biochem. Sci.* 42, 297–311. doi: 10.1016/j.tibs.2017.01.002

Hendricks, R. L., Weber, P. C., Taylor, J. L., Koumbis, A., Tumpey, T. M., and Glorioso, J. C. (1991). Endogenously produced interferon α protects mice from herpes simplex virus type 1 corneal disease. *J. Gen. Virol.* 72, 1601–1610. doi: 10.1099/0022-1317-72-7-1601

Jiang, Y., Zhu, Y., Liu, Z. J., and Ouyang, S. (2017). The emerging roles of the DDX41 protein in immunity and diseases. *Protein Cell* 8, 83–89. doi: 10.1007/s13238-016-0303-304

Khan, R., Khan, A., Ali, A., and Idrees, M. (2019). The interplay between viruses and TRIM family proteins. *Rev. Med. Virol.* 29:e2028. doi: 10.1002/rmv.2028

Komoto, S., Higaki, S., Fukuda, M., and Shimomura, Y. (2015). Effects of antiviral medications on herpetic epithelial keratitis in mice. *Jpn. J. Ophthalmol.* 59, 194–200. doi: 10.1007/s10384-015-0375-373

Labetoulle, M., Auquier, P., Conrad, H., Crochard, A., Daniloski, M., Bouée, S., et al. (2005). Incidence of herpes simplex virus keratitis in France. *Ophthalmology* 112, 888–895. doi: 10.1016/j.ophtha.2004.11.052

Leib, D. A., Harrison, T. E., Laslo, K. M., Machalek, M. A., Moorman, N. J., and Virgin, H. W. (1999). Interferons regulate the phenotype of wild-type and mutant herpes simplex viruses in vivo. *J. Exp. Med.* 189, 663–672. doi: 10.1084/jem.189.4.663

Li, X. D., Wu, J., Gao, D., Wang, H., Sun, L., and Chen, Z. J. (2013). Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. *Science* 341, 1390–1394. doi: 10.1126/science.1244040

Liu, H., Li, M., Song, Y., and Xu, W. (2018). TRIM21 restricts Coxsackievirus B3 replication, cardiac and pancreatic injury via interacting with MAVS and positively regulating IRF3-mediated type-I interferon production. *Front. Immunol.* 9:2479. doi: 10.3389/fimmu.2018.02479

Lobo, A. M., Agelidis, A. M., and Shukla, D. (2019). Pathogenesis of herpes simplex keratitis: the host cell response and ocular surface sequelae to infection and inflammation. *Ocul. Surf.* 17, 40–49. doi: 10.1016/j.jtos.2018.10.002

Luker, G. D., Prior, J. L., Song, J., Pica, C. M., and Leib, D. A. (2003). Bioluminescence imaging reveals systemic dissemination of herpes simplex virus type 1 in the absence of interferon receptors. *J. Virol.* 77, 11082–11093. doi: 10.1128/jvi.77.20.11082-11093.2003

Manocha, G. D., Mishra, R., Sharma, N., Kumawat, K. L., Basu, A., and Singh, S. K. (2014). Regulatory role of TRIM21 in the type-I interferon pathway in Japanese encephalitis virus-infected human microglial cells. *J. Neuroinflamm.* 11:24. doi: 10.1186/1742-2094-11-24

Rajasagi, N. K., Bhela, S., Varanasi, S. K., and Rouse, B. T. (2017). Frontline science: aspirin-triggered resolvin D1 controls herpes simplex virus-induced corneal immunopathology. *J. Leukoc. Biol.* 102, 1159–1171. doi: 10.1189/jlb.3hi1216-511rr

Rechenchoski, D. Z., Faccin-Galhardi, L. C., Linhares, R. E. C., and Nozawa, C. (2017). Herpesvirus: an underestimated virus. *Folia Microbiol.* 62, 151–156. doi: 10.1007/s12223-016-0482-487

Rhodes, D. A., De Bono, B., and Trowsdale, J. (2005). Relationship between SPRY and B30.2 protein domains. Evolution of a component of immune defence. *Immunology* 116, 411–417. doi: 10.1111/j.1365-2567.2005.02248.x

Smith, J. S., and Robinson, N. J. (2002). Age-specific prevalence of infection with herpes simplex virus types 2 and 1: a global review. *J. Infect. Dis.* 186, S3–S28. doi: 10.1086/343739

Stremlau, M., Owens, C. M., Perron, M. J., Kiessling, M., Autissier, P., and Sodroski, J. (2004). The cytoplasmic body component TRIM5α restricts HIV-1 infection in old world monkeys. *Nature* 427, 848–853. doi: 10.1038/nature02343

Su, C., Zhan, G., and Zheng, C. (2016). Evasion of host antiviral innate immunity by HSV-1, an update. *Virol. J.* 13:38. doi: 10.1186/s12985-016-0495-495

Takaoka, A., Wang, Z., Choi, M. K., Yanai, H., Negishi, H., Ban, T., et al. (2007). DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* 448, 501–505. doi: 10.1038/nature06013

Towers, G. J. (2007). The control of viral infection by tripartite motif proteins and cyclophilin A. *Retrovirology* 4:40. doi: 10.1186/1742-4690-4-40

Twardy, B. S., Channappanavar, R., and Suvas, S. (2011). Substance P in the corneal stroma regulates the severity of herpetic stromal keratitis lesions. *Investig. Ophthalmol. Vis. Sci.* 52, 8604–8613. doi: 10.1167/iovs.11-8089

Unterholzner, L., Keating, S. E., Baran, M., Horan, K. A., Jensen, S. B., Sharma, S., et al. (2010). IFI16 is an innate immune sensor for intracellular DNA. *Nat. Immunol.* 11, 997–1004. doi: 10.1038/ni.1932

Watkinson, R. E., McEwan, W. A., Tam, J. C. H., Vaysburd, M., and James, L. C. (2015). TRIM21 promotes cGAS and RIG-I sensing of viral genomes during infection by antibody-opsonized virus. *PLoS Pathog.* 11:e05253. doi: 10.1371/journal.ppat.1005253

Wilhelmus, K. R. (2015). Antiviral treatment and other therapeutic interventions for herpes simplex virus epithelial keratitis. *Cochrane Database Syst. Rev.* 2015:CD002898. doi: 10.1002/14651858.CD002898.pub5

Zawatzky, R., Gresser, I., DeMaeyer, E., and Kirchner, H. (1982). The role of interferon in the resistance of C57BL/6 mice to various doses of herpes simplex virus type 1. *J. Infect. Dis.* 146, 405–410. doi: 10.1093/infdis/146.3.405

Zhang, J. F., Xiong, H. L., Cao, J. L., Wang, S. J., Guo, X. R., Lin, B. Y., et al. (2018). A cell-penetrating whole molecule antibody targeting intracellular HBx suppresses hepatitis B virus via TRIM21-dependent pathway. *Theranostics* 8, 549–562. doi: 10.7150/thno.20047

Zhang, Z., Yuan, B., Bao, M., Lu, N., Kim, T., and Liu, Y. J. (2011). The helicase DDX41 senses intracellular DNA mediated by the adaptor STING in dendritic cells. *Nat. Immunol.* 12, 959–965. doi: 10.1038/ni.2091

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 29 May 2020
doi: 10.3389/fmicb.2020.01098

A Novel miRNA—hlo-miR-2—Serves as a Regulatory Factor That Controls Molting Events by Targeting CPR1 in *Haemaphysalis longicornis* Nymphs

*Wen-Ge Liu[1,2], Jin Luo[1], Qiao-Yun Ren[1], Zhi-Qiang Qu[1], Han-Liang Lin[3], Xiao-Feng Xu[1], Jun Ni[1], Rong-Hai Xiao[4], Rong-Gui Chen[5], Muhammad Rashid[1], Ze-Gong Wu[1], Yang-Chun Tan[1], Xiao-Fei Qiu[1], Jian-Xun Luo[1], Hong Yin[1,6], Hui Wang[1,7], Zeng-Qi Yang[2], Sa Xiao[2] and Guang-Yuan Liu[1]**

[1] State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, China, [2] College of Veterinary Medicine, Northwest A&F University, Yangling, China, [3] Xinjiang Animal Health Supervision Station, Ürümqi, China, [4] Ruili Entry-Exit Inspection and Quarantine Bureau Inspection and Quarantine Comprehensive Technology Center, Yunnan, China, [5] Ili Center of Animal Disease Control and Diagnosis, Ili, China, [6] Jiangsu Co-innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonoses, Yangzhou, China, [7] Department of Engineering, Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom

OPEN ACCESS

Edited by:
Dong Xia,
Royal Veterinary College (RVC), United Kingdom

Reviewed by:
Longxian Zhang,
Henan Agricultural University, China
Tetsuya Tanaka,
Kagoshima University, Japan

***Correspondence:**
Guang-Yuan Liu
liuguangyuan@caas.cn;
particks@126.com

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *19 January 2020*
Accepted: *04 May 2020*
Published: *29 May 2020*

Citation:
Liu W-G, Luo J, Ren Q-Y, Qu Z-Q, Lin H-L, Xu X-F, Ni J, Xiao R-H, Chen R-G, Rashid M, Wu Z-G, Tan Y-C, Qiu X-F, Luo J-X, Yin H, Wang H, Yang Z-Q, Xiao S and Liu G-Y (2020) A Novel miRNA—hlo-miR-2—Serves as a Regulatory Factor That Controls Molting Events by Targeting CPR1 in Haemaphysalis longicornis Nymphs. Front. Microbiol. 11:1098. doi: 10.3389/fmicb.2020.01098

Successful completion of the molting process requires new epidermal growth and ecdysis of the old cuticle in *Haemaphysalis longicornis* (*H. longicornis*). MicroRNAs (miRNAs) participate in the development of organisms by inhibiting the expression of their target mRNAs. In this study, a novel tick-specific miRNA was identified and denoted hlo-miR-2 that serves as a novel regulator of molting events in *H. longicornis* nymphs by targeting a cuticular protein. The full length of this cuticular protein was first obtained and named it CPR1. A qRT-PCR analysis showed that hlo-miR-2 and CPR1 exhibit significant tissue and temporal specificity and that their transcription levels are negatively correlated during the molting process. CPR1, as a direct target of hlo-miR-2, was identified by a luciferase reporter assay *in vitro*. Agomir treatment indicated that the overexpression of hlo-miR-2 significantly reduced the protein expression level of CPR1, decreased the molting rate and delayed the molting time point in *H. longicornis* nymphs. RNA interference (RNAi) experiments demonstrated that CPR1 was significantly associated with the molting process in *H. longicornis* nymphs. Phenotypic rescue experiments convincingly showed that hlo-miR-2 participated in molting events by targeting CPR1 in *H. longicornis* nymphs. In summary, we present evidence demonstrating that miRNAs constitute a novel important regulator of molting events in addition to hormones. The described functional evidence implicating CPR1 in molting events contributes to an improved understanding of the distinct functions of the CPR family in ticks and will aid the development of a promising application of cuticular protein RNAi in tick control.

Keywords: *H. longicornis*, hlo-miR-2, cuticular protein, molting event, RNA interference

INTRODUCTION

As a three-host tick, *H. longicornis* requires three hosts and undergoes two molting processes during its lifetime (Yatsenko and Shcherbata, 2014). The molting process, which can cause the generation of new epidermal growth and ecdysis of the old cuticle, is a prerequisite for the growth and development of *H. longicornis* ticks (Freitak et al., 2012; Zhou et al., 2013). Numerous external and internal stimuli, which might include seasonal changes, food intake, hormones and other factors, participate in the beginning of the molting process. The sharp increase in ecdysteroid (20-E) titers in the *H. longicornis* body during molting is consistent with the deposition of the new epicuticle and is an important factor in initiating the formation of the exuvial space (apolysis). A histological study of the tick cuticle ultrastructure has shown that the tick cuticle consists of two parts, the epicuticle and the procuticle, and the procuticle consists of two layers, namely, the inner and the outer endocuticle (Boulias and Horvitz, 2012; Asgari, 2013). Because cuticular proteins (CPs) are the main components of the arthropod cuticle, variations in the amounts and types of these proteins greatly affect the structure and properties of the arthropod cuticle (Choi and Hyun, 2012). Hence, CPs and CP-encoding genes can be used as relevant models for studying molting in ticks and the potential underlying molecular mechanisms. Furthermore, the molting process is very important in tick development, and interfering with the molting process of ticks will notably prevent or delay the development of ticks, resulting in the blockage of tick development and even death. Therefore, investigations of the effects of interference with the molting process will provide new ideas for the development of tick vaccines and the prevention and control of ticks.

MicroRNAs (miRNAs) are non-coding RNAs with a length of approximately 22 nt that posttranscriptionally regulate their target genes by binding to the 3′ untranslated regions (UTRs) of the corresponding mRNAs (Cenik et al., 2011; Fullaondo and Lee, 2012). Since the discovery of lin-4, this regulatory biological factor has been found in many organisms (Bejarano et al., 2010), such as mosquitoes, Drosophila, mites, ticks and other arthropods, and has been confirmed to be involved in the blood sucking, spawning, molting, and larval development processes (Caygill and Johnston, 2008; Huang et al., 2014; Hussain and Asgari, 2014; Kozomara and Griffiths-Jones, 2014; Zhang et al., 2015). A recent study on ticks revealed that miRNA-275 is involved in the physiological processes of blood digestion and ovary development and that miR-375 impacts oviposition and egg hatching in *H. longicornis* (Hao et al., 2017; Malik et al., 2019).

In our previous study, various miRNAs, including conserved and potentially novel miRNAs, were identified by deep sequencing of *H. longicornis* at different developmental stages. Among these miRNAs, a potential novel miRNA, named novel miRNA-2, exhibited high expression at the unengorged larval, unengorged nymph and unengorged adult developmental stages but not at the egg stage. Based on this finding, we speculated that novel miRNA-2 might play an important role in the maintenance of the developmental state of *H. longicornis*. Therefore, this study aimed to reveal the biological role of novel miRNA-2 in the development of *H. longicornis*.

MATERIALS AND METHODS

Tick and Tissue Samples

Colonies of *H. longicornis* were established by the Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, China (LVRI). The ticks were reared at 25°C in cages with 90% relative humidity and allowed to feed on the blood of New Zealand white rabbits. Under a stereoscopic microscope (Olympus, Japan), semi-saturated *H. longicornis* was dissected, and the tissues in the body cavity (salivary gland, midgut, ovary, and Malpighian tubules) were isolated and collected. Epidermal tissue was obtained by flushing with RNase-free water to remove the contents of the residual body cavity. A study conducted by Liu and Jiang (1998) found that the molting period of *H. longicornis* nymphs lasts 16.9 days, whereas our study revealed a duration of 15.47 days (Liu and Jiang, 1998). Therefore, we set the duration of the experimental observation in this study to 16 days. To more comprehensively and conveniently observe the dynamic changes during the molting period, we selected an experimental scheme involving the collection of samples every other day. The collected tissues were placed in the TRIzol reagent (Takara, Dalian, China) for the extraction of total RNA. The use of vertebrates in this study was approved by the Animal Ethics Committee of LVRI (Approval No. LVRIAEC2012-011).

RNA Isolation and Quantitative Real-Time PCR (qRT-PCR)

Total RNA was extracted using the TRIzol reagent (Takara, Dalian, China). cDNAs of the mature sequence of novel miRNA-2 and the CP CPR1 (accession number MT274605) were produced using a PrimeScript™ RT Reagent Kit (Takara) according to the manufacturer's recommended protocol. All primer designs were performed using Primer Premier v5.0, and the primer details are shown in **Table 1**. qRT-PCR of miRNA and mRNA was performed using a SYBR® *Premix EX Taq*™ II Kit (Takara, Dalian, China) according to the manufacturer's instructions. Each sample was measured in triplicate, and the relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method and normalized to the expression of the housekeeping gene β-actin.

Abbreviations: Ag, agomir; Ant, antagomir; cDNA, complementary deoxyribonucleic acid; CDS, coding sequence; CPR, cuticular protein with Rebers and Riddiford consensus; CPs, cuticular proteins; Cy3, sulfo-cyanine 3; DAPI, 2-(4-amidinophenyl)-6-indolecarbamidine dihydrochloride; DMEM, Dulbecco's modified Eagle medium; DNA, deoxyribonucleic acid; FAM, carboxyfluorescein; FBS, fetal bovine serum; HRP, horseradish peroxidase; iCut, interference RNA for CPR1; miRNA, micro-ribonucleic acid; NC, negative control; PBS, phosphate buffered saline; PVDE, polyvinylidene fluoride; qRT-PCR, quantitative real-time PCR; qRT-PCR, quantitative reverse transcription polymerase chain reaction; RNA, ribonucleic acid; RNAi, RNA interference; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TBST, Tris-buffered saline Tween-20; USA, United States of America; UTRs, untranslated regions.

TABLE 1 | Primers used in this study.

Purpose	Primer name	Primer sequence (5′–3′)
cDNA synthesis of miRNAs	Novel miRNA-2 SL primer	GTCGTATCCAGTGCAGGGTCCGAGG TATTCGCACTGGATACGACACTGTC
Real-time PCR analysis	Novel miRNA-2 F	CGGGCAAGAGAGCAATCCGT
	Novel miRNA-2 R	CAGTGCAGGGTCCGAGGTAT
	cuticular F1	ACAAGATCGCAATCCTCCTG
	Cuticular R1	TCCTGGTTGGCGTAGTTGAA
	β-actin F	CGTTCCTGGGTATGGAATCG
	β-actin R	TCCACGTCGCACTTCATGAT
Amplify the full length of the cuticular	GSP-F1	CAGCTTTCCTCTAAGGTCCGCTTCC
	GSP-R1	CATCTACAGGGGGGTGGATATGGAT
	GSP-F2	CGGACCCGAGGCCGCCATCGCCAAC
	GSP-R2	GGAAAGACGCTGTACGAGATCCCGC
Amplify the 3′ UTR of the cuticular	Cuticular F2	AGCTTTGTTTAAACCCTGCGTGTGATT CCGTGTA
	Cuticular R2	GCTCTAGAAAGAGGGAAGGAACGGG TAG
In situ hybridization	Cuticular probes	FAM-GGCAACGCAGGCGCAGGCCAGGAG GATT-FAM
	Novel miRNA-2 probes	CY3-ACTGTCCACGGATTGCTCTCTT

SL means Stem loop. FAM and CY3 means green and red dye, respectively. Restriction sites are underlined.

Simulation of the Secondary Structure of the Novel miRNA-2 Precursor

RNAfold 2.4.13[1] was used to simulate the secondary structure of the novel miRNA-2 precursor. The following running parameters of RNAfold were used for the simulation: -p, calculates the partition function and base pairing probability matrix in addition to the MFE structure; -d2, this previous check is ignored, and dangling energies are added to the bases adjacent to a helix on both sides in any case; and -noLP, produces structures without lonely pairs (helices of length 1). For partition function folding, the algorithm disallows pairs that can only occur in isolation, and other pairs might still occasionally occur as helices of length 1 (Mathews et al., 2004).

Prediction and Amplification of the Full Length of the Target Gene

Three miRNA target prediction programs, namely, RNAhybrid[2], PITA[3] and miRanda[4], were used to predict the target gene of novel miRNA-2. The rapid amplification of cDNA ends (RACE) technique was used to amplify the full length of the target gene. The two pairs of GSP primers were used to amplify the 3′ UTR and 5′ UTR of the CPR1 according to the RACE kit instructions (Takara, Dalian, China). All the primers were designed using Primer Premier v5.0, and the primer details are shown in **Table 1**.

Luciferase Reporter Assay

293-T cells (cultured by LVRI) were grown in DMEM (Gibco, Waltham, MA, United States) supplemented with 10% fetal bovine serum (FBS) (Gibco, Waltham, MA, United States) and 1% 1x antibiotic-antimycotic (Gibco). Luciferase constructs were generated by inserting the 3′ UTR of the CPR1 gene into the pmirGLO vector (Promega, Madison, WI, United States). A pair of primers was used to amplify the 3′ UTR of CPR1 and were designed using Primer Premier v5.0. The primer details are shown in **Table 1**. One hundred micrograms of reconstructed pmirGLO vector was cotransfected with 50 nM synthetic novel miRNA-2 mimic (RiboBio, Guangzhou, China) or negative control (NC) (RiboBio, Guangzhou, China) into 293-T cells using Lipofectamine 3000 (Invitrogen, United States). Transfection with only the reconstructed pmirGLO vector was performed as a blank control treatment. Forty-eight hours after transfection, a dual luciferase reporter assay was performed using the Dual-Glo® Luciferase Assay System (Promega) according to the manufacturer's instructions. The firefly luciferase activity of the pmirGLO vector was measured for the normalization of Renilla luciferase activity. The treatments were performed in triplicate, and the transfections were repeated three times (Hao et al., 2017).

Colocalization of Novel miRNA-2 and CPR1 by Fluorescence *in situ* Hybridization (FISH)

For the assessment of novel miRNA-2 and CPR1 colocalization, antisense RNA detection probes targeting novel miRNA-2 and the CPR1 gene were designed and labeled with the dual fluorophores Cy3 and FAM, respectively, at Servicebio Biotechnology Company (Wuhan, China). A scrambled sequence and the sense probe for the target gene were used as NCs. Two days after microinjection, the tick nymphs were dissected in cold PBS buffer, fixed overnight in 4% paraformaldehyde, and incubated overnight at 37°C for probe hybridization (Nuovo et al., 2009; Yang et al., 2016). The tick samples were washed in PBS containing 5% Triton X-100 (v/v) for 10 min and stained with DAPI G1012 (Servicebio, Wuhan, China) at room temperature for 8 min. The signals of CPR1 and novel miRNA-2 were detected using an Eclipse CI upright fluorescence microscope (Nikon, Japan). The probes for the miRNA and target genes are listed in **Table 1**.

Agomir and Antagomir Treatment

Agomir (Ag), micrON^{TM} agomir negative control #22 (NC), antagomir (Ant), and micrOFF^{TM} Ant negative control #22 (NC) were synthesized by RiboBio Company (RiboBio, Guangzhou, China).

Five days after blood meal ingestion, *H. longicornis* nymphs were dosed at a site between the third and fourth legs. The

[1]http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi

[2]https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid/

[3]https://genie.weizmann.ac.il/pubs/mir07/mir07_data.html

[4]http://www.microrna.org/microrna/home.do

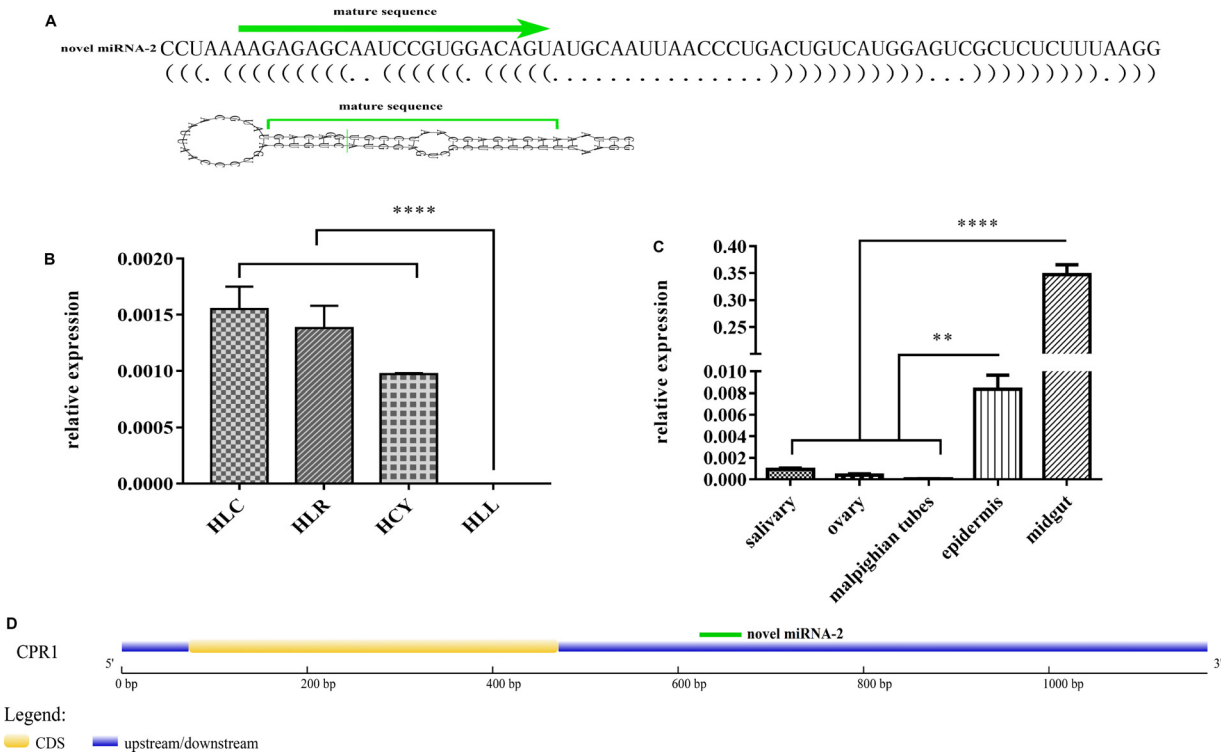


FIGURE 1 | Characterization, expression patterns and target gene prediction of novel miRNA-2. **(A)** The novel miRNA-2 precursor is 69 nt in length with a stem-loop structure and the mature miRNA is 22 nt in length. **(B)** The relative expression levels of novel miRNA-2 in eggs, unengorged larvae, unengorged nymphs and unengorged adults were analyzed by qRT-PCR. HLL, HLY, HLR, and HLC are representative of eggs, unengorged larvae, unengorged nymphs, and unengorged adults of *H. longicornis*, respectively. **(C)** The relative expression levels of novel miRNA-2 in the salivary gland, midgut, ovarian, Malpighian tubules, and epidermal tissues were analyzed by qRT-PCR. **(D)** Structure of the CPR1 gene and location of the novel miRNA-2-binding site. All the data represent the results from three biological replicates, each of which included three technical replicates, and are presented as the means ± SEMs. $^{**}P < 0.01$ and $^{****}P < 0.0001$.

experimental group was administered 800 μM novel miRNA-2 Ag or Ant in 0.2 μL of RNase-free water, the NC group was administered 800 μM micr*ON*™ Ag or micr*OFF*™ Ant in 0.2 μL of RNase-free water, and no injection was administered to the blank control group. Each group included 50 *H. longicornis* nymphs collected 5 days after blood meal ingestion.

Thirty ticks were used for observation of the molting timepoint and molting rate, and the rest were used to detect the expression of related genes and proteins. Two days after injection, total RNA and protein were extracted from the ticks using the TRIzol™ reagent (Takara, Dalian, China) in accordance with the manufacturer's instructions. Moreover, the number of successfully completed molts and the molting timepoints after microinjection were observed and recorded.

RNAi and Rescue Experiments

The double-stranded siRNA targeting CPR1 (iCut) and the NC for iCut, which exhibited no significant sequence similarity to the *H. longicornis* sequence, was synthesized by RiboBio (Guangzhou, China). Five days after blood meal ingestion, *H. longicornis* nymphs were dosed at a site between the third and fourth legs: the experimental group was administered 800 μM iCut in 0.2 μL of RNase-free water, and NC group was administered 800 μM NC iCut in 0.2 μL of RNase-free water, and no injection was administered to the blank control group.

In the rescue experiments, the tick nymphs were divided into four groups and treated as follows: coinjection with 0.2 μL of an Ant/iCut mixture (800 μM Ant and 800 μM iCut), coinjection with 0.2 μL of an Ant/NC iCut mixture (800 μM Ant and 800 μM NC iCut), coinjection with 0.2 μL of an NC Ant/iCut mixture (800 μM NC Ant and 800 μM iCut), and coinjection with 0.2 μL of an NC Ant/NC iCut mixture (800 μM NC Ant and 800 μM NC iCut).

Each group included 50 *H. longicornis* nymphs obtained 5 days after blood meal ingestion. Thirty ticks were used for observation of the molting timepoint and molting rate, and the rest were used for the detection of related genes and proteins. Two days after injection, total RNA and protein from the ticks were extracted using the TRIzol™ reagent (Takara, Dalian, China) in accordance with the manufacturer's instructions. Moreover, the number of successfully completed molts and the molting timepoints after microinjection were observed and recorded.

Western Blot Analysis

Total protein from the ticks was extracted using the TRIzol™ reagent (Takara, Dalian, China) according to

the manufacturer's instructions, separated on 12% SDS–PAGE gels and transferred to PVDF membranes (Millipore, United States). The membranes were then blocked in TBST containing 5% skim milk for 1 h at room temperature. For the detection of CPR1, the membranes were incubated overnight at 4°C with monoclonal antibodies against CPR1 at a 1:300 dilution and then incubated with a 1:5000 dilution of an anti-rabbit HRP-conjugated secondary antibody (Solarbio, Beijing, China) at room temperature for 1 h. CPR1 expression was detected by chemiluminescence (Thermo Fisher Scientific, United States).

Scanning Electron Microscopy

The samples were fixed in 3% glutaraldehyde at room temperature for 2 h and transferred to 4°C for storage. The samples were fixed in 1% osmium tetroxide for 2 h, dehydrated in an alcohol gradient (30–50–70–80–90–95–100–100% alcohol), and incubated with isoamyl acetate for 15 min. The samples were then dried in a K850 critical point dryer (Quorum, United Kingdom) and subjected to gold sputtering for approximately 30 s with an MSP-2S instrument (IXRF, United States). Observations and imaging were performed under an SU8100 scanning electron microscope (HITACHI, Japan) (Liu et al., 2001).

Statistical Analysis

All the data were analyzed using GraphPad Prism 7, and the results are presented as the means ± SEMs. The F test was used for equal variance detection, and the means were compared using Student's *t*-test or the Wilcoxon test based on the following significance levels: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

RESULTS

Characterization and Validation of Novel miRNA-2

The potential novel miRNA-2 needs to be further characterized and validated. One of the most notable characteristics defining a miRNA gene is a precursor with a typical stem-loop structure, and at most four symmetrical mismatches within the duplex and at most two asymmetric bulges of up to 2 bases are allowed (Ambros et al., 2003). Close scrutiny of the secondary structure of novel miRNA-2 precursor indicated that its stem-loop structure conformed to the established standards (**Figure 1A**).

Tissue-Specific and Developmental Expression Patterns of Novel miRNA-2

Novel miRNA-2 was highly expressed in unengorged larvae, unengorged nymphs and unengorged adults but was not found at the egg stage by qRT-PCR analysis (**Figure 1B**). To determine the tissue expression patterns of novel miRNA-2, we performed a RT-PCR analysis of the five tissues of *H. longicornis*. The results demonstrated that novel miRNA-2 was abundantly expressed in both the midgut and epidermis but was expressed at low levels in the other tissues (salivary gland, ovary, and Malpighian tubules) (**Figure 1C**).

Prediction of a Novel miRNA-2 Target Gene and Luciferase Reporter Assay-Based Identification of the Binding Site of Novel miRNA-2 on the Target Gene *in vitro*

Three publicly available bioinformatic tools for miRNA target gene prediction were used to predict the target gene of novel miRNA-2. The same prediction results obtained from all of the tools showed that the CPR1 gene contains a binding site for novel miRNA-2. We obtained the full-length sequence of the CPR1 gene using RACE technology. The full length of the CPR1 gene contained 1171 bp, and the coding sequence (CDS) was 339 bp in length (Supplementary Data 1). The putative novel miRNA-2-binding site in CPR1 is located at position 661, which is located at the 3′ UTR of CPR1, and this binding is a classical binding mode (**Figure 1D**).

The 3′ UTR of the CPR1 gene was inserted into the pmirGLO vector and cotransfected into 293-T cells along with a novel miRNA-2 mimic, whereas NC mimic/novel miRNA-2 and novel miRNA-2 were transfected alone as negative and blank controls, respectively. The fluorescence results showed that the luciferase activity of the pmirGLO- CPR1/novel miRNA-2 group was 72.12% of the values found for the NC and blank control groups, which indicated that CPR1 was a target gene of novel miRNA-2 *in vitro* (**Figure 2A**).

Tissue-Specific Expression Patterns of CPR1

To assess the expression of CPR1 in different tissues, we performed a RT-PCR analysis of the five tissues of *H. longicornis*. As shown in **Figure 2B**, CPR1 was abundantly expressed in the epidermis but exhibited low expression in the other tissues (midgut, salivary gland, ovary, and Malpighian tubules).

Detection of the Dynamic Expression of Novel miRNA-2 and CPR1 During the Molting Process of *H. longicornis* Nymphs by qRT-PCR

To compare the expression levels of novel miRNA-2 and CPR1, we performed a thorough time-course assay of these miRNAs in *H. longicornis* nymphs during the molting process by qRT-PCR analysis. Eight time points were selected for the collection of RNA samples from *H. longicornis* nymphs: 2, 4, 6, 8, 10, 12, 14, and 16 days post blood meal ingestion. The expression level of novel miRNA-2 in *H. longicornis* nymphs gradually increased from day 0 to 6 post blood meal ingestion, reached a peak on day 6, decreased sharply starting on day 8 and remained at a low level until the end of the molting process. Interestingly, CPR1 exhibited a lower expression level from day 0 to 8 post blood meal ingestion, and its level then increased rapidly beginning

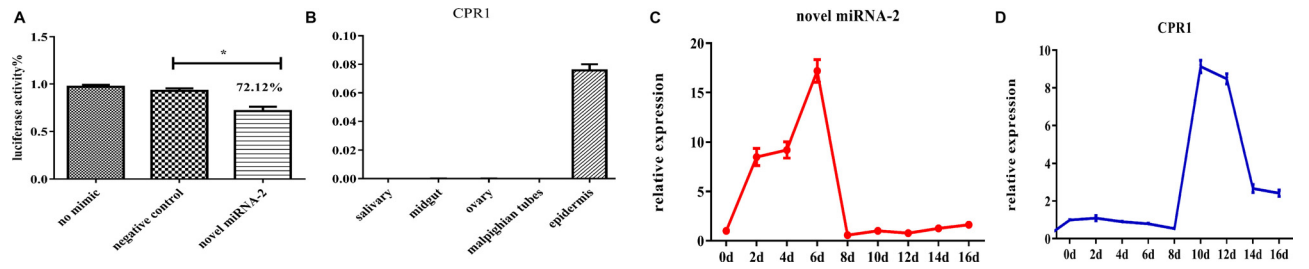


FIGURE 2 | The target gene identification and expression patterns analysis of novel miRNA-2 and CPR1. **(A)** Dual luciferase reporter assay of CPR1. The data indicate the percent activity and are presented as the means ± SEMs from triplicate samples. The percentages are shown. *$P < 0.05$. **(B)** The relative expression levels of CPR1 in salivary gland, midgut, ovarian, Malpighian tubules, and epidermal tissues were analyzed by qRT-PCR. **(C)** The relative expression level of novel miRNA-2 in the epidermis was analyzed every two days (D: day) during the molting process of nymphs by qRT-PCR. **(D)** The relative expression level of CPR1 in the epidermis was analyzed every two days (D: day) during the molting process of nymphs by qRT-PCR. All the data are presented as the results from three biological replicates, each of which included three technical replicates, and are presented as the means ± SEMs.

on day 10 and remained high until the end of the molting process (**Figures 2C,D**).

Colocalization of the Novel miRNA-2 and Its Target Gene CPR1

To determine whether novel miRNA-2 and CPR1 were colocalized in the epidermal tissue of tick nymphs, we performed *in situ* analyses of novel miRNA-2 and its target CPR1 via miRNA/mRNA fluorescence *in situ* hybridization. Both novel miRNA-2 and CPR1 were widely detected in the epidermal cells of ticks (**Figure 3**), which suggested that in the tick integument, CPR1 interacts directly with novel miRNA-2 in a spatially dependent manner.

Novel miRNA-2 Overexpression Results in Severe Molting Defects

To determine the function of novel miRNA-2 in molting events, we overexpressed novel miRNA-2 by microinjecting Ag into *H. longicornis* nymphs at 5 days post blood meal ingestion. The transcript and protein levels of CPR1 were assessed on day 2 after microinjection. The qRT-PCR results showed that the CPR1 mRNA level was not significantly reduced in the Ag-treated group (0.956 ± 0.03445 fold), and as expected, the expression in the NC and blank control groups was unaffected. Moreover, the Western blot results showed that the expression of CPR1 in the Ag-treated group was lower than that in the NC (0.989 ± 0.05615 fold) and blank control (0.979 ± 0.06109 fold) groups, which suggested that overexpression of novel miRNA-2 can downregulate CPR1 expression.

Furthermore, the molting timepoint and molting rate were recorded. The results showed that the molting timepoint was significantly delayed to 19.23 ± 0.3944 days in the Ag-treated group compared with the NC group (15.47 ± 0.1417 days) (Student's *t*-test, $P < 0.0001$). Moreover, the molting rate of the Ag-treated group was markedly decreased to 73.33% of the value obtained for the NC and blank control groups. These results strongly indicated that novel miRNA-2 overexpression exerts a negative effect on the molting process in *H. longicornis* nymphs (**Figure 4**).

Novel miRNA-2 Depletion Results in Severe Molting Defects

To decipher the role of novel miRNA-2 in molting events, we blocked the expression of novel miRNA-2 by microinjecting Ant into *H. longicornis* nymphs at 5 days post blood meal ingestion and assessed the transcript and protein levels of CPR1 on day 2 after the microinjection. The qRT-PCR results showed that the mRNA expression level of CPR1 was not increased in the Ant-treated group (0.976 ± 0.05658) compared with the other groups, and no significant difference was observed between the NC (0.985 ± 0.05381) and blank control (0.973 ± 0.0771) groups. However, the Western blot analysis showed that the protein expression level of CPR1 was increased by Ant treatment compared with that found in the NC and blank control groups, which suggested that the blockage of novel miRNA-2 expression can upregulate CPR1 expression.

The molting timepoint of the Ant-treated group was reduced to 14.39 ± 0.1654 days compared with that of the NC group (15.4 ± 0.1633 days) (Student's *t*-test, $P < 0.0001$). Moreover, the molting rate did not differ significantly between the Ant-treated and NC control groups (**Figure 5**).

The epidermal ultrastructure of the tick samples on day 2 after the microinjection was observed by scanning electron microscopy. The analysis of the NC and blank control groups showed that the tip of the microvilli-protruding portion was precipitated in the form of a cap, which was dispersed but tended to increase in number and aggregate. In the Ant group, these cap-like structures aggregated and formed a continuous layer, and clear cortical fusion occurred, which resulted in a clear exuvial space. These results showed that the blockade of novel miRNA-2 expression accelerated the molting process of *H. longicornis* nymphs (**Figure 6**).

Effects of CPR1 RNA Interference (RNAi) on *H. longicornis* Nymph Molting Events

Because CPR1 was expressed in the epidermis and its expression changed during the normal molting process of *H. longicornis*

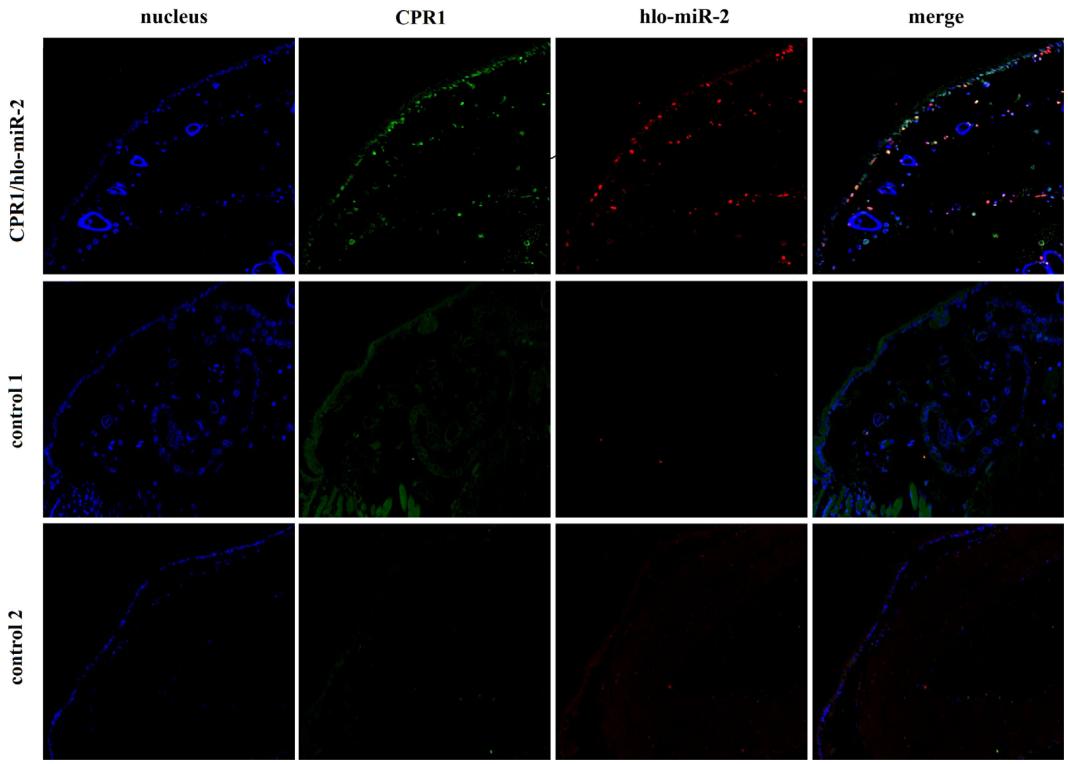


FIGURE 3 | Colocalization of novel miRNA-2 with the target gene CPR1 in epidermal cells of ticks. The nucleic acid probe targeting novel miRNA-2 was conjugated to the dual fluorophore Cy3 (red), whereas the nucleic acid probe for the target gene CPR1 was conjugated to the dual fluorophore FAM (green). Probes targeting a scrambled nucleotide sequence were used as the control for miRNA, and probes targeting the sense nucleotide sequence were used as the control for the target gene. The overlapping of green (CPR1) and red signals (novel miRNA-2) is indicated by a yellow signal, which indicated the colocalization of novel miRNA-2 and its target gene CPR1. Control 1, scrambled miRNA and CPR1 antisense probes; Control 2, novel miRNA-2 and CPR1 sense probes. The images were acquired using an Eclipse CI upright fluorescence microscope (Nikon).

nymphs, we examined the protein and mRNA expression of CPR1 and the molting timepoint and rate after CPR1 RNAi to further evaluate the potential roles of this gene. The mRNA expression of CPR1 were assessed on day 2 after iCut injection, and the results showed that the mRNA levels of CPR1 were significantly reduced in the iCut-treated group (0.725 ± 0.02172) compared with other group and not appreciably differ between the NC (0.996 ± 0.03407) and blank control (0.984 ± 0.02575) groups, Moreover, the Western blot analysis showed that the protein expression level of CPR1 was decreased by iCut treatment compared with that found in the NC and blank control groups, which suggested that a siRNA designed to target CPR1 successfully decreased the CPR1 transcript and protein levels in *H. longicornis* nymphs.

We also found that the molting rate of the iCut group was significantly decreased to 46.67% of the value obtained for the NC and blank control groups, whereas no significant difference was observed between the NC and blank control groups. Additionally, the molting timepoint was extended to 21.71 ± 0.3387 days in the iCut group compared with that in the NC group (15.5 ± 0.1216 days) (Student's *t*-test, $P < 0.0001$), whereas no significant difference in the molting timepoint was observed between the NC and blank control groups (**Figure 7**).

Phenotype Rescue Confirms That CPR1 Is a Direct Target of Novel miRNA-2 *in vivo*

To further verify whether CPR1 is an authentic target of novel miRNA-2 during *H. longicornis* nymph molting events, behavior phenotype rescue experiments were performed by microinjecting Ant for the blockage of novel miRNA-2 expression and using designed siRNA for the RNAi-mediated depletion of the CPR1 gene in *H. longicornis* nymphs at 5 days post blood meal ingestion. This method has been successfully used for validating miRNA mutant phenotypes in mosquitoes and ticks (Lucas et al., 2015; Hao et al., 2017).

The molting timepoint of the Ant/iCut coinjection group was significantly delayed to 17.95 ± 0.3104 days compared with that of the Ant/NC iCut group (14.32 ± 0.0171 days) (Student's *t*-test, $P < 0.0001$), but this delay was shorter than that obtained for the iCut/NC Ant group (21.4 ± 0.3625 days) (Student's *t*-test, $P < 0.0001$). However, the molting rate of the Ant/iCut group was notably declined to 76% of that

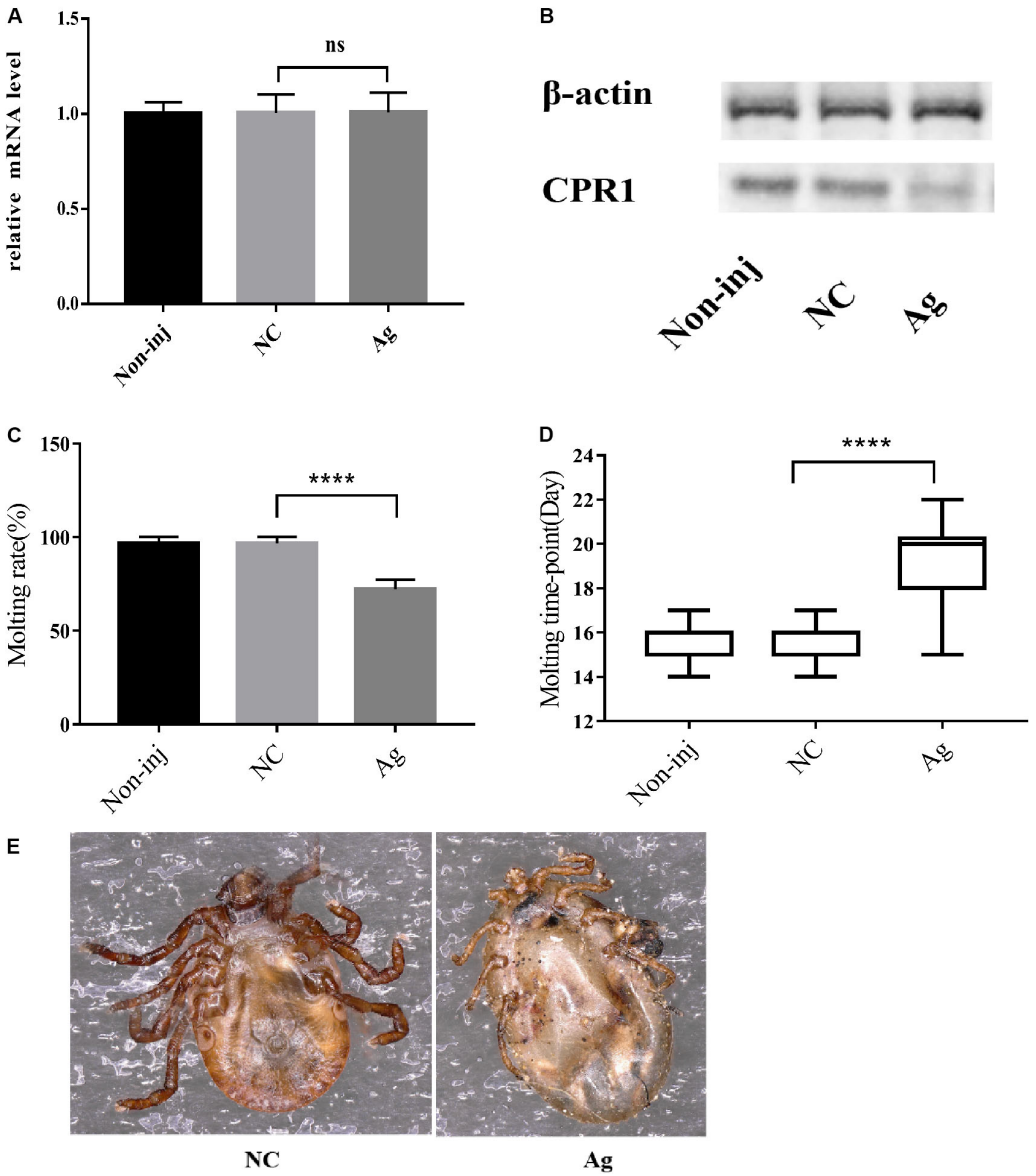


FIGURE 4 | The overexpression of novel miRNA-2 significantly disrupted protein accumulation and delayed the molting time with increasing variation. **(A)** Transcription level of the target gene CPR1 after novel miRNA-2 overexpression. **(B)** Protein level of the target gene CPR1 after novel miRNA-2 overexpression. **(C)** Molting rate of nymphs injected with a novel miRNA-2 agomir. **(D)** Molting timepoint of nymphs injected with a novel miRNA-2 agomir. The data are shown as the means ± SEMs and were evaluated by Student's *t*-test. ****$P < 0.0001$. **(E)** Representative phenotypes after feeding with novel miRNA-2 agomir. The injection of novel miRNA-2 agomir prevented the molting process, which led to high mortality. Ag, novel miRNA-2 agomir; NC, negative control.

obtained for the Ant group but was higher than that of the iCut group (46.67%) (**Figure 8**). These results suggested that the RNAi-mediated silencing of CPR1 resulted in severe phenotypes resulting from the inability of *H. longicornis* nymphs to undergo normal molting and, as expected, ameliorated the adverse phenotypes caused by the silencing of novel miRNA-2. Hence, the phenotypic rescue achieved by the administration of CPR1 RNAi to novel miRNA-2-depleted *H. longicornis* nymphs confirmed that CPR1 is an authentic target of novel miRNA-2 *in vivo*.

DISCUSSION

In this study, we first discovered and verified the presence of novel miRNA-2 in *H. longicornis*. The full-length novel miRNA-2 precursor has a length of 62 nt and exhibits a typical stem-loop structure, and its mature sequence is 22 nt in length. We named this miRNA hlo-miR-2.

hlo-miR-2 exhibits substantial stage and tissue specificity in *H. longicornis*. Specifically, it is highly abundant in unengorged larvae, unengorged nymphs and unengorged adults but not in

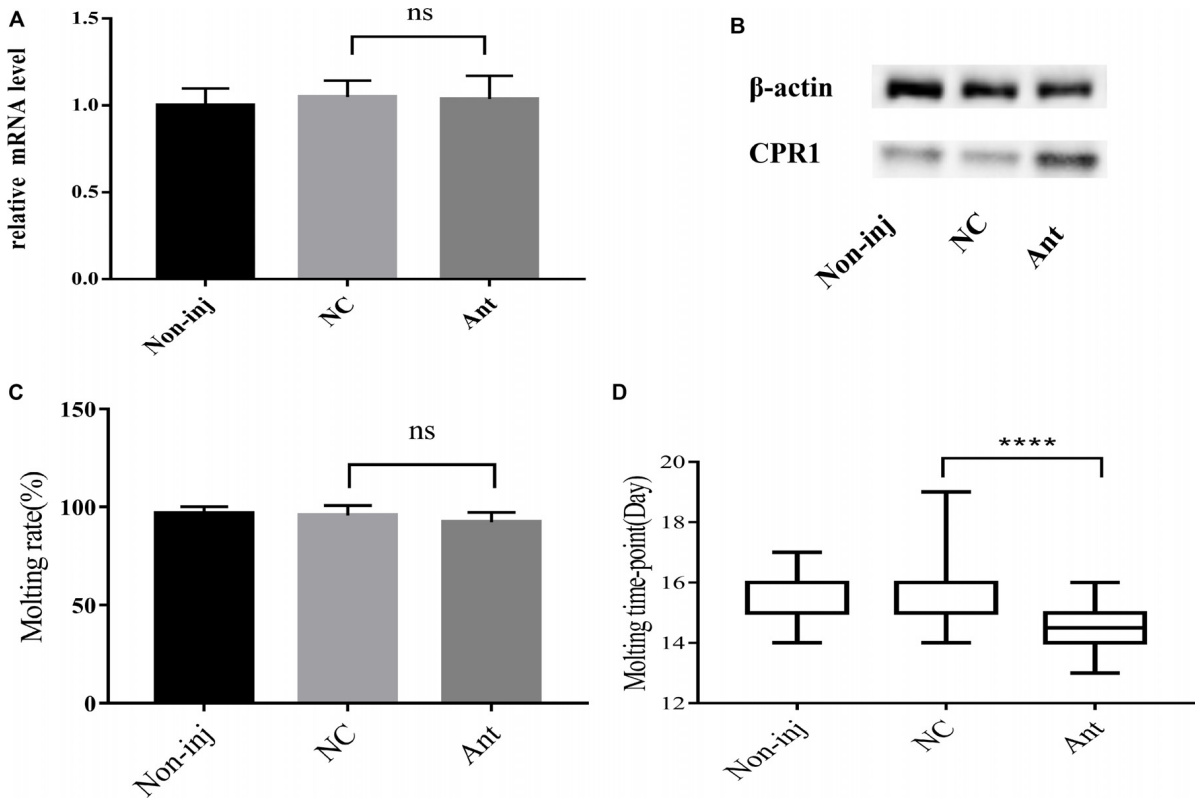


FIGURE 5 | The blockade of novel miRNA-2 expression disrupted protein accumulation and impaired molt synchrony. **(A)** Transcription level of the target gene CPR1 after the inhibition of novel miRNA-2 expression. **(B)** Protein level of the target gene CPR1 after the inhibition of novel miRNA-2 expression. **(C)** Molting rate of nymphs injected with novel miRNA-2 Ant. **(D)** Molting timepoint of nymphs injected with novel miRNA-2 Ant. The data are shown as the means ± SEMs and were evaluated by Student's *t*-test. $^{****}P < 0.0001$.

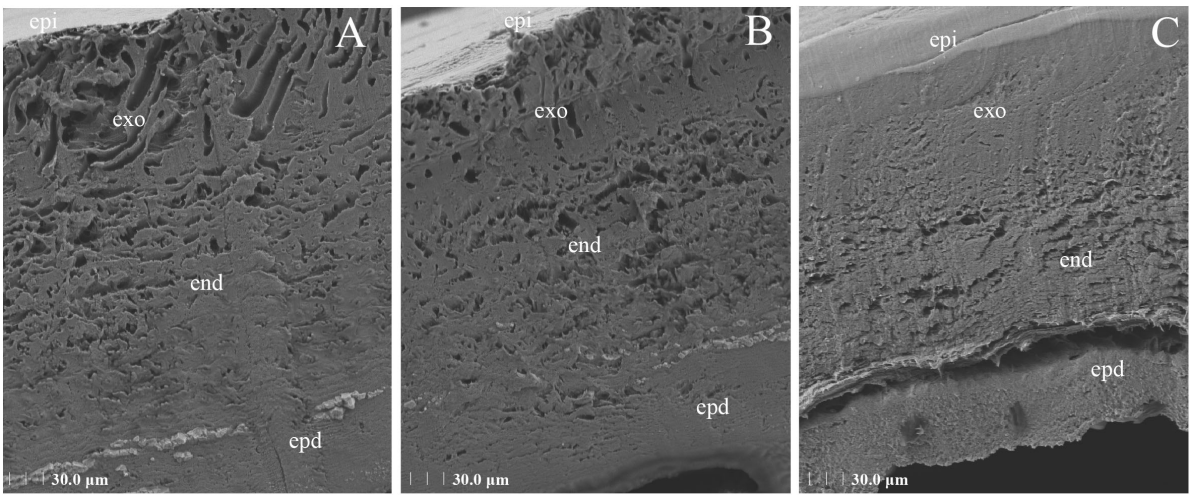


FIGURE 6 | Ultrastructural architecture of the cuticle after injection of novel miRNA-2 antagomir. The tick nymph cuticle has three horizontal layers: the outermost layer is the epicuticle (epi), the second layer is the procuticle (pro), which contains two layers, namely, the exocuticle (exo) and the endocuticle (end), and the third layer is the epidermis (epd). In the NC and the blank control groups **(A,B)**, the tip of the microvilli-protruding portion was precipitated in the form of a cap, which was dispersed but tended to accumulate and aggregate. In the antagomir group **(C)**, these cap-like structures aggregated and formed a continuous layer, and clear cortical fusion occurred, resulting in a clear exuvial space.

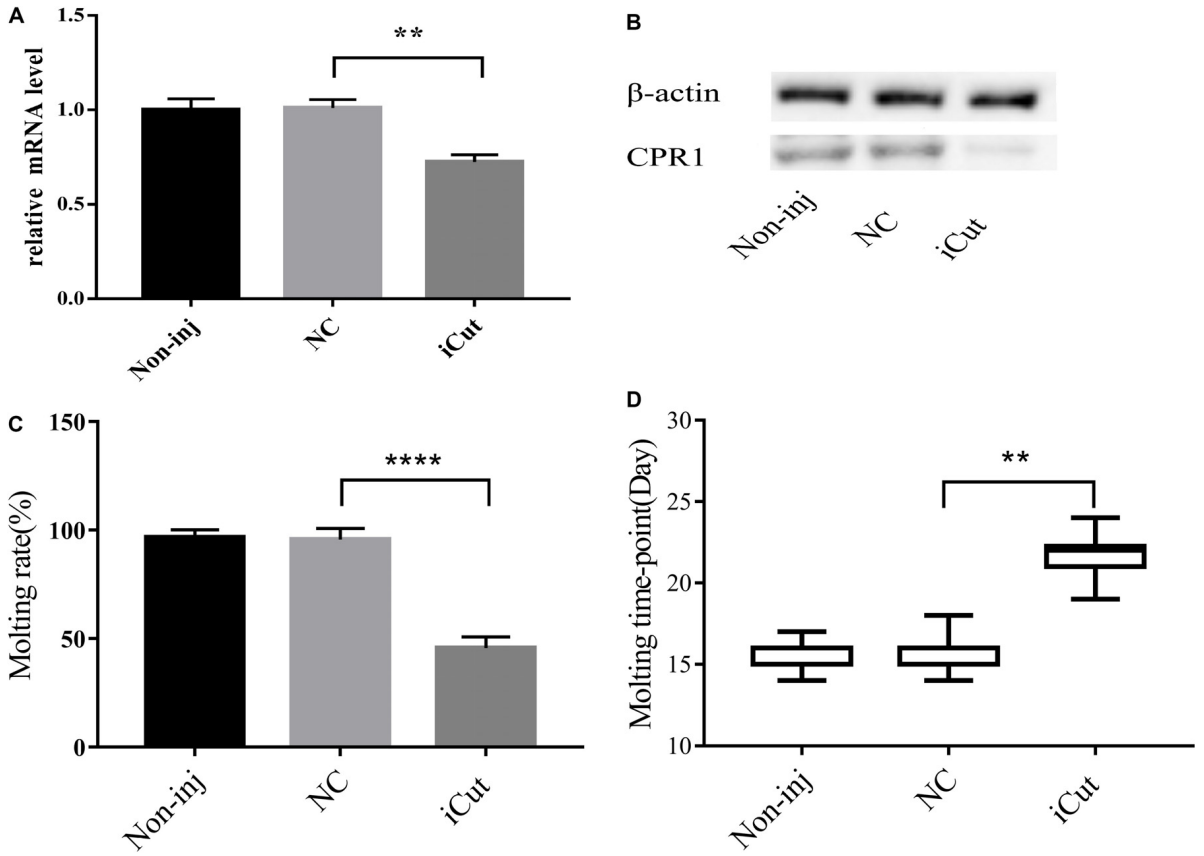


FIGURE 7 | CPR1 RNAi significantly disrupted protein accumulation and delayed the molting timepoint. **(A)** Transcription level of the target gene CPR1 after the inhibition of CPR1. **(B)** Protein level of the target gene CPR1 after the inhibition of CPR1. **(C)** Molting rate of nymphs injected with CPR1-targeting siRNA. **(D)** Molting timepoint of nymphs injected with CPR1-targeting siRNA. The data are shown as the means ± SEMs and were evaluated by Student's *t*-test. $^{**}P < 0.01$ and $^{****}P < 0.0001$.

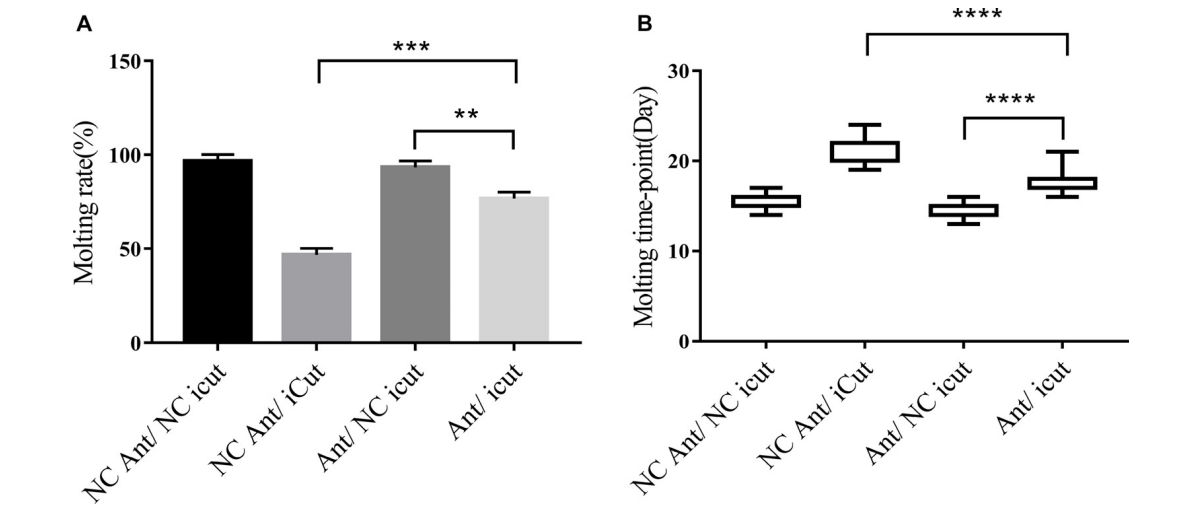


FIGURE 8 | CPR1 RNAi partially rescued the novel miRNA-2-depletion phenotypes. **(A)** CPR1 RNAi improved the molting rate of nymphs treated with novel miRNA-2 antagomir. **(B)** CPR1 RNAi improved the molting timepoint of nymphs treated with novel miRNA-2 antagomir. The data are shown as the means ± SEMs and were evaluated by Student's *t*-test. $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

eggs. This miRNA is also highly expressed in epidermal and midgut tissues compared with other tissues.

Our results identified CPR1 as a target gene of hlo-miR-2 and showed that this gene was also highly abundant in epidermal tissue. Moreover, the qRT-PCR analysis showed that the levels of hlo-miR-2 and CPR1 were negatively correlated during the molting process of *H. longicornis*.

This study provides the first determination of the full-length sequence of CPR1, which was 1171 bp in length. The CDS was 339 bp in length, and the binding site for hlo-miR-2 was located in the 3′ UTR of CPR1. The miRNA/mRNA fluorescence *in situ* hybridization results showed that CPR1 interacted directly with novel miRNA-2 in a spatially dependent manner.

The epidermis provides a barrier for the defense of ticks against adverse environmental changes and mechanical stimuli, and ticks must periodically shed and resynthesize their epidermis during development. Members of the CPR family, which is the largest family of CPs, are expressed in the procuticle and play a role in binding to chitin (Willis, 2010; Dittmer et al., 2015; Vannini et al., 2015; Arasteh et al., 2016). A simulation of the secondary structure of the protein indicated that this cuticular gene sequence contained a conserved chitin-binding 4 region, which is within a Rebers & Riddiford (R&R) motif and belongs to the CPR family (Tetreau et al., 2015).

As determined through dual luciferase reporter assay, the cotransfection of a luciferase reporter vector containing the *H. longicornis* CPR1 3′ UTR with the hlo-miR-2 mimic resulted in decreased *Renilla* luciferase activity *in vitro*, which suggested that CPR1 is a direct target gene of hlo-miR-2 *in vitro*.

The overexpression of hlo-miR-2 decreased CPR1 expression, significantly delayed the molting timepoint of *H. longicornis* nymphs, and markedly decreased the molting rate. The depletion of hlo-miR-2 in *H. longicornis* nymphs increased CP expression and accelerated the molting process in *H. longicornis* nymphs. These results suggested that hlo-miR-2 plays an important role in *H. longicornis* nymph molting events. The data obtained from the siRNA-mediated silencing of CPR1 during the *H. longicornis* nymph molting process suggested that CPR1 is necessary and sufficient for molting events.

The phenotypic rescue experiment showed that the siRNA/Ant group exhibited a delayed molting timepoint and a decreased molting rate compared with the Ant group, which indicated that the Ant can bind to hlo-miR-2 *in vivo* and counteract the inhibition of hlo-miR-2; thus, only the inhibitory effect of the CPR1 siRNA is reflected in the phenotype. Additionally, these results suggested that CPR1 significantly affects the formation of the original epidermis, whereas hlo-miR-2 has a significant inhibitory effect on the posttranslational level of CPR1, which indicated that CPR1 is a target gene of hlo-miR-2 *in vivo*.

A histological study showed that the *H. longicornis* 20-E titers increased sharply on day 6 after nymph engorgement, and this event was accompanied by the initiation of apolysis and the appearance of the exuvial space (Liu et al., 2001). Additionally, the same results were observed in *Ornithodoros moubata* and *Amblyomma hebraeum* (Diehl et al., 1982; Dotson et al., 1991). However, in this study, we found that hlo-miR-2 was maintained at a high level during the first 6 days after engorgement, and after this time, its level suddenly decreased and remained low until the end of the molting period. This trend is exactly opposite to those found for CPR1 transcription and the 20-E titer, which indicated that high hlo-miR-2 expression might inhibit molt initiation after *H. longicornis* nymph engorgement, but this possibility needs further study.

Previous studies of *Bombyx mori*, *Manduca sexta* and locusts have suggested that the level of *LmTwdl1*, a member of the Tweedle family of CPs, gradually increased beginning on day 5 after nymph engorgement, which is a vital period for cuticle renewal, and that this gene has conserved essential functions in molting events (Soares et al., 2011; Dittmer et al., 2015; Song et al., 2016). A study conducted by Liu et al. (2001) showed that deposition of the new epicuticle began on day 8 after nymph engorgement, which coincides with the time at which the 20-E titer peaked. However, in this study, the transcriptional level of CPR1 continued to increase from day 10 after engorgement until the molting process was completed. Furthermore, the silencing of CPR1 resulted in high mortality during the molting process before the next stage, which indicated that CPR1 plays a critical role during the late molting stage in *H. longicornis* nymphs.

CONCLUSION

In conclusion, our study provides the first demonstration that hlo-miR-2 is an authentic miRNA in *H. longicornis* that exhibits significant tissue and temporal specificity. The results from RNAi experiments showed that CPR1 is highly associated with the molting process in *H. longicornis* nymphs. More importantly, through dual luciferase reporter assays and phenotypic rescue experiments, we uncovered the role of hlo-miR-2 in molting events in *H. longicornis* nymphs. Furthermore, we identified CPR1 as a direct target of hlo-miR-2, and the obtained functional evidence verified that the hlo-miR-2-mediated regulation of CPR1 is essential for molting development in *H. longicornis* nymphs. In summary, our research indicates that miRNAs constitute a novel important regulator of the molting process in addition to hormones. The functional evidence implicating CPR1 in molting events contributes to a better understanding of the distinct functions of the CPR family in ticks and will aid the development of the promising application of cuticular RNAi in tick control.

DATA AVAILABILITY STATEMENT

The datasets generated or analyzed during this study are included in this published article and its additional files.

ETHICS STATEMENT

The animal study was reviewed and approved by Approval No. LVRIAEC2012-011.

AUTHOR CONTRIBUTIONS

G-YL and W-GL designed the experiments. W-GL, Z-QQ, JL, Q-YR, Z-GW, JN, Y-CT, X-FQ, X-FX, H-LL, R-HX, and R-GC performed the experiments. W-GL analyzed the data and wrote the manuscript. G-YL, Z-QY, J-XL, SX, MR, HW, and HY critically revised the manuscript. All the authors read and approved the final version of the manuscript.

FUNDING

This study was financially supported by the NSFC (No. 31572511), the National Key Research and Development Program of China (2019YFC1200502), NPRC-2019-194-30, ASTIP, NBCIS CARS-38 and the State Key Laboratory of Veterinary Etiological Biology Project.

REFERENCES

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., et al. (2003). A uniform system for microRNA annotation. *RNA* 9, 277–279. doi: 10.1261/rna.2183803

Arasteh, S., Kazemnejad, S., Khanjani, S., Heidari-Vala, H., Akhondi, M. M., and Mobini, S. (2016). Fabrication and characterization of nano-fibrous bilayer composite for skin regeneration application. *Methods* 99, 3–12. doi: 10.1016/j.ymeth.2015.08.017

Asgari, S. (2013). MicroRNA functions in insects. *Insect Biochem. Mol. Biol.* 43, 388–397.

Bejarano, F., Smibert, P., and Lai, E. C. (2010). miR-9a prevents apoptosis during wing development by repressing *Drosophila* LIM-only. *Dev. Biol.* 338, 363–373. doi: 10.1016/j.ydbio.2009.11.025

Boulias, K., and Horvitz, H. R. (2012). The *C. elegans* microRNA mir-71 acts in neurons to promote germline-mediated longevity through regulation of DAF-16/FOXO. *Cell Metab.* 15, 439–450. doi: 10.1016/j.cmet.2012.02.014

Caygill, E. E., and Johnston, L. A. (2008). Temporal regulation of metamorphic processes in *Drosophila* by the let-7 and miR-125 heterochronic microRNAs. *Curr. Biol.* 18, 943–950. doi: 10.1016/j.cub.2008.06.020

Cenik, E. S., Fukunaga, R., Lu, G., Dutcher, R., Wang, Y., Tanaka, T. M., et al. (2011). Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol. Cell* 42, 172–184. doi: 10.1016/j.molcel.2011.03.002

Choi, I. K., and Hyun, S. (2012). Conserved microRNA miR-8 in fat body regulates innate immune homeostasis in *Drosophila*. *Dev. Comp. Immunol.* 37, 50–54. doi: 10.1016/j.dci.2011.12.008

Diehl, P. A., Germond, J., and Morici, M. (1982). Correlations between ecdysteroid titers and integument structure in nymphs of the tick, *Amblyomma hebraeum* Koch (Acarina: Ixodidae). *Rev. Suisse Zool.* 89, 859–868. doi: 10.1007/BF00383769

Dittmer, N. T., Tetreau, G., Cao, X., Jiang, H., Wang, P., and Kanost, M. R. (2015). Annotation and expression analysis of cuticular proteins from the tobacco hornworm, *Manduca sexta*. *Insect Biochem. Mol. Biol.* 62, 100–113. doi: 10.1016/j.ibmb.2014.12.010

Dotson, E. M., Connat, J. L., and Diehl, P. (1991). Cuticle Deposition and Ecdysteroid Titers during Embryonic and Larval Development of the Argasid Tick *Ornithodoros moubat* (Murray, 1877, sensu Walton, 1962)(Ixodoidea: Argasidae). *Gen. Comp. Endocrinol.* 82, 386–400. doi: 10.1016/0016-6480(91)90314-v

Freitak, D., Knorr, E., Vogel, H., and Vilcinskas, A. (2012). Gender- and stressor-specific microRNA expression in *Tribolium castaneum*. *Biol. Lett.* 8, 860–863. doi: 10.1098/rsbl.2012.0273

Fullaondo, A., and Lee, S. Y. (2012). Identification of putative miRNA involved in *Drosophila melanogaster* immune response. *Dev. Comp. Immunol.* 36, 267–273. doi: 10.1016/j.dci.2011.03.034

Hao, J., Luo, J., Chen, Z., Ren, Q., Guo, J., Liu, X., et al. (2017). MicroRNA-275 and its target Vitellogenin-2 are crucial in ovary development and blood digestion of *Haemaphysalis longicornis*. *Parasites Vect.* 10:253. doi: 10.1186/s13071-017-2153-1

Huang, Y., Dou, W., Liu, B., Wei, D., Liao, C. Y., Smagghe, G., et al. (2014). Deep sequencing of small RNA libraries reveals dynamic expression patterns of microRNAs in multiple developmental stages of *Bactrocera dorsalis*. *Insect Mol. Biol.* 23, 656–667. doi: 10.1111/imb.12111

Hussain, M., and Asgari, S. (2014). MicroRNAs as mediators of insect host-pathogen interactions and immunity. *J. Insect Physiol.* 70, 151–158. doi: 10.1016/j.jinsphys.2015.01.010

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181

Liu, J., Li, Y., and Jiang, Z. (2001). Correlation between 20-hydroxyecdysone levels and cuticulogenesis in nymphs of *Haeymaphysalis longicornis*. *Acta Entomol. Sin.* 44, 46–51.

Liu, J. Z., and Jiang, Z. J. (1998). Studies on the bionomics of haemaphysalis longicornis neumann (Acari:ixodidae) under laboratory conditions. Kun Chong xue bao. *Acta Entomol. Sin.* 41, 280–283.

Lucas, K. J., Zhao, B., Roy, S., Gervaise, A. L., and Raikhel, A. S. (2015). Mosquito-specific microRNA-1890 targets the juvenile hormone-regulated serine protease JHA15 in the female mosquito gut. *RNA Biol.* 12, 1383–1390. doi: 10.1080/15476286.2015.1101525

Malik, M. I., Nawaz, M., Hassan, I. A., Zhang, H., Gong, H., Cao, J., et al. (2019). A microRNA profile of saliva and role of miR-375 in *Haemaphysalis longicornis* (Ixodida: Ixodidae). *Parasites Vect.* 12:68. doi: 10.1186/s13071-019-3318-x

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7287–7292. doi: 10.1073/pnas.0401799101

Nuovo, G. J., Elton, T. S., Nana-Sinkam, P., Volinia, S., Croce, C. M., and Schmittgen, T. D. (2009). A methodology for the combined in situ analyses of the precursor and mature forms of microRNAs and correlation with their putative targets. *Nat. Protoc.* 4, 107–115. doi: 10.1038/nprot.2008.215

Soares, M. P., Silva-Torres, F. A., Elias-Neto, M., Nunes, F. M., Simões, Z. L., and Bitondi, M. (2011). Ecdysteroid-dependent expression of the tweedle and peroxidase genes during adult cuticle formation in the honey bee, *Apis mellifera*. *PLoS One* 6:e20513. doi: 10.1371/journal.pone.0020513

Song, T. Q., Yang, M. L., Wang, Y. L., Liu, Q., Wang, H. M., Zhang, J., et al. (2016). Cuticular protein LmTwdl1 is involved in molt development of the migratory locust. *Insect Sci.* 23, 520–530. doi: 10.1111/1744-7917.12342

Tetreau, G., Dittmer, N. T., Cao, X., Agrawal, S., Chen, Y. R., Muthukrishnan, S., et al. (2015). Analysis of chitin-binding proteins from *Manduca sexta* provides new insights into evolution of peritrophin A-type chitin-binding domains in insects. *Insect Biochem. Mol. Biol.* 62, 127–141. doi: 10.1016/j.ibmb.2014.12.002

Vannini, L., Bowen, J. H., Reed, T. W., and Willis, J. H. (2015). The CPCFC cuticular protein family: anatomical and cuticular locations in Anopheles gambiae and distribution throughout *Pancrustacea*. *Insect Biochem. Mol. Biol.* 65, 57–67. doi: 10.1016/j.ibmb.2015.07.002

Willis, J. H. (2010). Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochem. Mol. Biol.* 40, 189–204. doi: 10.1016/j.ibmb.2010.02.001

Yang, M., Wang, Y., Jiang, F., Song, T., Wang, H., Liu, Q., et al. (2016). miR-71 and miR-263 jointly regulate target genes chitin synthase and chitinase to control *Locust* molting. *PLoS Genet.* 12:e1006257. doi: 10.1371/journal.pgen.1006257

Yatsenko, A. S., and Shcherbata, H. R. (2014). *Drosophila* miR-9a targets the ECM receptor Dystroglycan to canalize myotendinous junction formation. *Dev. Cell* 28, 335–348. doi: 10.1016/j.devcel.2014.01.004

Zhang, X., Zheng, Y., Cao, X., Ren, R., Yu, X. Q., and Jiang, H. (2015). Identification and profiling of *Manduca sexta* microRNAs and their possible roles in regulating specific transcripts in fat body, hemocytes, and midgut. *Insect Biochem. Mol. Biol.* 62, 11–22. doi: 10.1016/j.ibmb.2014.08.006

Zhou, J., Zhou, Y., Cao, J., Zhang, H., and Yu, Y. (2013). Distinctive microRNA profiles in the salivary glands of *Haemaphysalis longicornis* related to tick blood-feeding. *Exp. Appl. Acarol.* 59, 339–349. doi: 10.1007/s10493-012-9604-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 09 June 2020
doi: 10.3389/fgene.2020.00604

Transcriptome Profiling of Pacu (*Piaractus mesopotamicus*) Challenged With Pathogenic *Aeromonas hydrophila*: Inference on Immune Gene Response

*Vito Antonio Mastrochirico-Filho[1], Milene Elissa Hata[1], Rafael Yutaka Kuradomi[1], Milena Vieira de Freitas[1], Raquel Belini Ariede[1], Daniel Guariz Pinheiro[2], Diego Robledo[3], Ross Houston[3] and Diogo Teruo Hashimoto[1]**

[1] Aquaculture Center, São Paulo State University (Unesp), Jaboticabal, Brazil, [2] Faculty of Agricultural and Veterinary Sciences, São Paulo State University (Unesp), Jaboticabal, Brazil, [3] The Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Edinburgh, United Kingdom

OPEN ACCESS

Edited by:
Argyris Papantonis, University Medical Center Göttingen, Germany

Reviewed by:
Bhanwar Lal Puniya, University of Nebraska-Lincoln, United States
Leda Torres, National Institute of Pediatrics of Mexico, Mexico

***Correspondence:**
Diogo Teruo Hashimoto
diogo.hashimoto@unesp.br

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: 28 February 2020
Accepted: 18 May 2020
Published: 09 June 2020

Citation:
Mastrochirico-Filho VA, Hata ME, Kuradomi RY, de Freitas MV, Ariede RB, Pinheiro DG, Robledo D, Houston R and Hashimoto DT (2020) Transcriptome Profiling of Pacu (Piaractus mesopotamicus) Challenged With Pathogenic Aeromonas hydrophila: Inference on Immune Gene Response. Front. Genet. 11:604. doi: 10.3389/fgene.2020.00604

Pacu (*Piaractus mesopotamicus*) is a Neotropical fish of major importance for South American aquaculture. Septicemia caused by *Aeromonas hydrophila* bacteria is currently considered a substantial threat for pacu aquaculture that have provoked infectious disease outbreaks with high economic losses. The understanding of molecular aspects on progress of *A. hydrophila* infection and pacu immune response is scarce, which have limited the development of genomic selection for resistance to this infection. The present study aimed to generate information on transcriptome of pacu in face of *A. hydrophila* infection, and compare the transcriptomic responses between two groups of time-series belonging to a disease resistance challenge, peak mortality (HM) and mortality plateau (PM) groups of individuals. Nine RNA sequencing (RNA-Seq) libraries were prepared from liver tissue of challenged individuals, generating ∼160 million 150 bp pair-end reads. After quality trimming/cleanup, these reads were assembled *de novo* generating 211,259 contigs. When the expression of genes from individuals of HM group were compared to individuals from control group, a total of 4,413 differentially expressed transcripts were found (2,000 upregulated and 2,413 downregulated candidate genes). Additionally, 433 transcripts were differentially expressed when individuals from MP group were compared with those in the control group (155 upregulated and 278 downregulated candidate genes). The resulting differentially expressed transcripts were clustered into the following functional categories: cytokines and signaling, epithelial protection, antigen processing and presentation, apoptosis, phagocytosis, complement system cascades and pattern recognition receptors. The proposed results revealing relevant differential gene expression on HM and PM groups which will contribute to a better understanding of the molecular defense mechanisms during *A. hydrophila* infection.

Keywords: disease resistance, complement system, *de novo* assembly, RNA-Seq, aquaculture

INTRODUCTION

Disease outbreaks are considered a primary constraint and a continuous challenge to the growth of aquaculture sector (FAO, 2018). Detrimental effects due improper management practices, high-density culture systems and fluctuating water parameters have facilitated the spreading of bacteria or other pathogens, resulting in frequent and often severe disease outbreaks. This scenario has led to the billions of dollars in losses to aquaculture every year, that may exceed 40% of global capacity of production (Lafferty et al., 2015; Stentiford et al., 2017; Owens, 2019). While economic impact of disease outbreaks are under-reported for fish production in South America, a quarter of Brazilian large producers have reported moderate to high production losses due to disease outbreaks in the major aquaculture regions (Benavides and Maciel, 2015).

The incidence of sanitary problems linked to intensification of fish production from South America have resulted in large mortalities of Pacu (*Piaractus mesopotamicus*), one of the most important native fish species produced in Brazil (IBGE, 2016; Woynárovich and Van Anrooy, 2019). Pacu is considered a promising species for Brazilian aquaculture due to its large size, high quality meat and low-cost diet (Woynárovich and Van Anrooy, 2019). However, one of the main causes of disease associated with large mortalities for this species is the infection by *Aeromonas hydrophila* (Carraschi et al., 2012; Farias et al., 2016; Mastrochirico-Filho et al., 2019).

Aeromonas hydrophila is considered the etiological agent of Motile Aeromonas Septicemia (*MAS*), and is generally associated with symptoms such as reddened or rotten fins, small ulcerative lesions on the skin and hemorrhage, often followed by mortality of individuals within hours after manifestation of disease (Silva et al., 2012; Stratev and Odeyemi, 2017). Despite the lack of official data on the economic losses in fish production attributed to *A. hydrophila* infection, Brazilian fish farmers have reported that *MAS* prevalence is responsible for 20–30% of annual mortalities on pacu farming. However, insufficient knowledge on preventive measures and immune mechanisms of susceptibility and resistance are available to *MAS* (Belém-Costa and Cyrino, 2006; Farias et al., 2016; Zanuzzo et al., 2017; Mastrochirico-Filho et al., 2019).

RNA-Seq has been considered an effective technique for understanding the transcriptomic changes that occur during infection, exploring the activation of defense mechanisms involved in fast responses of infected hosts, even in organisms without a reference genome (Sudhagar et al., 2018). Over the last years, an increasing effort has been accomplished to characterize host-pathogen interaction by differentially expressed genes (DEGs) of fish species submitted to *A. hydrophila* resistance challenges, such as blunt snout bream *Megalobrama amblycephala* (Tran et al., 2015); grass carp *Ctenopharyngodon idella* (Dang et al., 2016); darkbarbel catfish *Pelteobagrus vachellii* (Qin et al., 2017); hybrid sturgeon (*Huso dauricus* × *Acipenser schrenckii*) (Jiang et al., 2018); and other species (Maekawa et al., 2019). These expression profiling have been considered useful to obtain information about the genomic makeup of immune responses at a molecular level, contributing to a better understanding of genetic resistance to disease, and guiding selective breeding strategies to produce more resistant strains. Additionally, in order to develop knowledge to suppress the disease, a better understanding of host-pathogen interaction between fish and bacteria, during critical periods of a disease outbreak is needed.

Therefore, the aim of the study was to characterize the transcriptome of pacu after *A. hydrophila* infection. A *de novo* assembly was necessary, since pacu is considered a non-model species and does not have a reference genome until the present moment. Further, the identification of differentially expressed genes involved in the response to this bacteria was performed aiming immune system genes, comparing differences between the actions of genomic mechanisms under two progressive time trends of a disease resistance challenge, peak mortality (HM) and mortality plateau (PM).

MATERIALS AND METHODS

Ethics Statement

This study was conducted in strict accordance with the recommendations of the National Council for Control of Animal Experimentation (CONCEA) (Brazilian Ministry for Science, Technology and Innovation) and was approved by the Ethics Committee on Animal Use (CEUA number 19.005/17) of Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus Jaboticabal, SP, Brazil. All biosafety procedures related to the management of bacteria and infected animals were carried out. In addition, all measures related to the safety of the participants of the present study in handling the strains of bacteria were ensured.

Generation of Pacu Families

Individuals used for RNA sequencing were obtained from an experimental challenge carried out in 14 full-sibling families of pacu, generated by a hierarchical mating scheme using 4 dams and 14 sires (3 or 4 sires per dam). Induced spawning was performed using carp pituitary extract dissolved in saline solution (0.9% NaCl) and applied in two dosages, with a 12 h interval (first and second dosage of 0.6 and 5.4 mg/kg, respectively). For sires, a single dosage was used, at the same time of the second dosage for dams, equivalent to 1.5 mg/kg of carp pituitary extract.

After hatching in 20 l conical fiberglass incubators, the larvae were fed with artemia nauplii for 20 days. Gradually, the feed was replaced by 50% of crude protein. In the fingerling stage, 1.2 mm pelleted feeds were used (40% of crude protein), being gradually replaced by 2 to 3 mm pelleted feeds (36% of crude protein) provided twice daily in 60 l tanks. Animals used in the experiment were pit-tagged to maintain the pedigree information during the challenge experiments. Laterally, fish were kept in 800 l fiberglass tanks at the Laboratory of Genetics in Aquaculture and Conservation (LaGeAC), at

the São Paulo State University (UNESP), Jaboticabal city (São Paulo, Brazil).

Experimental Challenge by *Aeromonas hydrophila*

Bacterial challenge was performed using a strain of *A. hydrophila* isolated from an outbreak of pacu aeromoniasis in a commercial facility from the São Paulo State, by the Laboratory of Microbiology and Parasitology of Aquatic Organisms, at UNESP, Jaboticabal city (São Paulo, Brazil). The strain of *A. hydrophila* was the same of that used by Mastrochirico-Filho et al. (2019). The strain was cultured in Agar Trypticase Soy (ATS), Vegitone (Sigma-Aldrich) for 24 h (28°C). The colony was then transferred to a nutrient tryptic soy broth (TSB) (Sigma-Aldrich) and cultured for 24 h (28°C). After bacteria growth, the culture was centrifuged at 5,000 × *g* for 10 min (4°C, in the Eppendorf Centrifuge 5810), forming a bacterial pellet suspended in a saline solution (PBS) and washed twice.

The challenge test performed in this study was carried out by intraperitoneal inoculation, according to protocols of Mastrochirico-Filho et al. (2019). The LD_{50} (lethal dose in 50% of individuals) was previously tested in 60 randomly chosen individuals from the same pacu families using concentrations adjusted by optical density of the solution at 0.400, 0.600, and 0.800 at 625 nm in spectrophotometer (2100 Unico, Japan). A sample of 100 μl of the LD50 was removed from the inoculum to perform serial dilutions and plate counts in duplicate on Trypticase Soy Agar (TSA).

Prior to challenge, the presence of *A. hydrophila* and other pathogens, such as *Flavobacterium columnare* and *Streptococcus agalactiae* in the populations was discarded checking subsamples of the populations by routine microbiological tests. In the experimental design of the challenge test, the individuals were distributed into three communal tanks of 2 m^3 (length = 2 m, width = 1 m, depth = 1 m), where approximately ten individuals from each family were randomly distributed into each treatment tank. Prior to the inoculation of bacteria, fish were anesthetized with benzocaine (0.1 mg/l) and weighed. The mean weight of animals prior to the bacterial inoculation time was 23.0 ± 9.06 g. Individual fish were injected by intraperitoneal inoculation of the predefined LD_{50} dose of live cells of *A. hydrophila* (8×10^5 CFU/g body weight), according to protocols carried out by Mastrochirico-Filho et al. (2019). Moreover, approximately ten fish from each family were also used as control and kept in a separated tank (called as control tank) of 2 m^3 (length = 2 m, width = 1 m, depth = 1 m). Individuals of the control tank were injected by intraperitoneal inoculation of saline solution (PBS).

Each treatment and control tank were maintained with an independent water recirculation system, fitted with mechanical and biological filters, external aeration system, and controlled temperature using thermal controller connected to heaters (2 × 500 w). Water quality parameters were determined daily during 14 days of the challenge. Temperature, dissolved oxygen and pH were measured with a Multiparameter Water Quality Checker U-50 (Horiba, Kyoto, Japan). Water temperature was maintained at 30°C (*SD* = 0.5), with dissolved oxygen at 5.8 mg/L (*SD* = 1.0) and pH at 7.0 (*SD* = 1.1) during the challenge period. No water was exchanged during the challenge, but the tanks were topped off to compensate for evaporation. The control tank had the same condition when compared with treatment tanks, except for an antibacterial treatment through a UV filter.

Fish mortality was observed throughout the entire day (24 h) in the initial 3 days of challenge, and in intervals of 8 h in the remaining days of the challenge experiment. Fish that showed mortality by clinical signs of *A. hydrophila* infection (e.g., disequilibrium, hemorrhage, isolation from the group) were recorded and removed immediately from the tanks. Necropsy examination and microbiological tests were performed in a subsample of dead fish in order to confirm mortality by *A. hydrophila* and discard other pathogens. The isolation of bacteria was performed in a specific growth medium (phenol red agar and ampicillin) incubated at 28°C for 48 h. All surviving fish were examined externally for clinical signs of disease.

Tissue Collection, RNA Isolation and Sequencing

According to experimental challenges by *A. hydrophila* in different fish species, the response of the immune system is very fast to this bacteria infection, occurring when mortality rates are generally detected in the first 24 h after exposure, whereas the mortality plateau generally occurs 72 h after infection (Mahapatra et al., 2008; Ødegård et al., 2010; Xiong et al., 2017; Mastrochirico-Filho et al., 2019). Therefore, fish presenting classical symptoms of *A. hydrophila* infection at 24 h after bacteria inoculation were collected and considered as belonging to HM (period of high mortality) group. When the mortality plateau occurred at 72 h after bacteria inoculation, fish were also collected and considered as belonging to PM (period of mortality plateau) group.

Laterally, three randomly selected families were sampled for RNA sequencing. Therefore, liver tissue from three individuals belonging to the HM group and three individuals belonging to the PM group were sampled from each family. Three control fish at each time trends of mortality were collected from the same families. Liver was collected from all individuals and stored in RNAlater®. This organ was chosen because it plays a critical role in the coordination of innate and adaptive immune responses (Kumar et al., 2017; Rebl and Goldammer, 2018), and it has been used in several studies involving transcriptome analysis in fish for a better understanding of the behavior of the immune system in face of bacterial infections (Tran et al., 2015; Maekawa et al., 2019; Jiang et al., 2020).

RNA was extracted using RNeasy Plus Universal Tissue Kit (Qiagen, Valencia, CA, United States) following the manufacturer's specifications. After extraction, total RNA was treated with DNAse I to remove any DNA contamination. Total RNA was quantified by Qubit RNA Assay Kit in a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, United States) and RNA integrity verified using a 2100 Bioanalyzer (Agilent Technologies). After mRNA extraction by poly-A tail selection, cDNA libraries were constructed using the TruSeq RNA Library Prep Kit protocol (Illumina). Six cDNA libraries were created from the combination of three pooled samples per family (family

A, V, and R) diluted to the same concentration, for each time-series of mortality (HM and PM). At the same time, three cDNA libraries (one per family) were created for control samples, pooling three samples belonging to time-series (HM + PM), totalizing 9 cDNA libraries (**Table 1**). Sequencing was performed on the Illumina Hiseq^{TM} 2500 platform.

Data Processing and Analysis

All RNA-Seq data generated for this study were deposited in the National Center for Biotechnology Information Sequence Read Archive under BioProject ID PRJNA632934. Bioinformatics pipelines are available in **Supplementary Table S1**. The quality of raw data was initially checked using FASTQC v. 0.11.9 (Andrews, 2010). Quality filtering and removal of residual Illumina adaptors were conducted on read pairs using TRIMMOMATIC v.0.36 (Bolger et al., 2014). Leading and trailing bases with a Phred score less than 20 were removed. Additionally, the read was trimmed if a sliding window average Phred score over four bases was less than 20. Following other trimming strategies adopted on studies approaching transcriptomic changes in fish submitted to disease infections, only pair reads greater than 36 bp were maintained (Robledo et al., 2014, 2018).

After filtering out low quality reads, the remained reads were used to assemble a *de novo* transcriptome using TRINITY v.2.9.1, adopting default parameters (Grabherr et al., 2011) as kmer-length value of 25 and contiguous sequences (contigs) longer than 200 bp. Read representation was evaluated by mapping the cleaned reads back to generated assembly with BOWTIE2 v.2.3.4.3 (Langmead and Salzberg, 2012). *A. hydrophila* genome (Yang et al., 2016), zebrafish ribosomal RNA (rRNA) sequences and pacu mitochondrial genome (da Pimentel et al., 2014) were then downloaded from the NCBI (National Center for Biotechnology Information) database and mapped against the assembly using BOWTIE2 v.2.3.4.3 (Langmead and Salzberg, 2012) to remove any contamination. Redundancy was reduced using CD-HIT-EST v.4.6.8 software by clustering sequences at 95% identity (Li and Godzik, 2006).

TABLE 1 | Preparation of cDNA pooled libraries of *P. mesopotamicus* families (A, V, R) challenged on *A. hydrophila* infection.

Treatment	Treatment samples			Control samples		
	Family A	Family V	Family R	Family A	Family V	Family R
HM	3	3	3	3	3	3
PM	3	3	3	3	3	3

Three pools per family:

Pool 1	HM treatment
Pool 2	PM treatment
Pool 3 (control)	HM + PM

The pools were constructed containing three individuals per family. Three cDNA libraries were constructed for control group (one per family) adding individuals belonging HM + PM time-trends of mortality.

Transcriptome statistics were generated using TRANSRATE v1.0.2 (Smith-Unna et al., 2016).

Annotation and homology searches of assembled transcripts were conducted using the TRINOTATE pipeline (Bryant et al., 2017). The assembled transcripts were translated into long open reading frames by TRANSDECODER (Haas et al., 2013). Homology searches were performed against the UniprotKB/Swiss-Prot database (BLASTX - E-value $1e^{-3}$) to identify known protein sequences (Altschul et al., 1990). Gene Ontology (GO) (Ashburner et al., 2000), eggNOG/COG (Non-supervised orthologous groups) (Powell et al., 2012), and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2012) databases were used to assign functions and biology pathways to transcripts.

KALLISTO v0.43.1 software was used to compute expression levels for each pooled samples (HM, PM and control) by pseudoalignments (Bray et al., 2016). Initially, an index was built with default k-mer size based on the contigs assembled by TRINITY. After, in order to quantify abundances of transcripts, TPM (transcripts per million reads) expression levels of genes for each of the nine pooled samples were computed. DESeq2 (Love et al., 2014) used negative binomial distribution, data normalization and FDR (false recovery rate) correction through Benjamini-Hochberg (BH) method, to identify genes differentially expressed comparing the gene expression from pooled samples belonging to HM and PM groups against pooled samples from control groups. To control false discovery rates of multiple-hypothesis testing, the *p*-values calculated by DESeq2 were processes to *q*-values. Then, the list of differentially expressed transcripts were generated with cutoff *q*-values less than 0.05 and $\log_2$FC (fold change (condition/control) for a gene) > 1 or $\log_2\text{FC} < -1$) as described in similar transcriptome studies (Marancik et al., 2015; Polinski et al., 2016; Jiang et al., 2020).

RESULTS

Disease Challenge Experiment

During the *A. hydrophila* resistance challenge, moribund individuals demonstrated lethargic behavior, erratic swimming and red spotted skin lesions on the operculum, head, fins and gills. No symptoms related to infection were reported in individuals in the unchallenged control group. High mortality rates (93.1%) were observed mainly at the first day after inoculation (HM period). Resistance rates of sampled families for RNA sequencing ranged from 16 to 32%. *A. hydrophila* was successfully isolated from infected individuals, confirming the bacterial infection. The last mortality was recorded at approximately 65 h post inoculation (PM period).

Sequencing and *de novo* Assembly

RNA sequencing, generated from liver samples, yielded between 15 and 20 million paired-end reads per pooled library. Approximately 12% of low-quality reads (quality score < 20) were removed. This resulted in approximately 143 million (87.8%) high quality paired-end reads, with a total length of

TABLE 2 | RNA-seq statistics in *P. mesopotamicus* families (A, V, R) challenged on *A. hydrophila* infection.

Library	Condition	Raw reads	Filtered reads (%)
C-A	Control	16,636,300	14,741,119 (88.6)
C-V	Control	18,560,164	16,282,254 (87.7)
C-R	Control	17,224,277	15,222,441 (88.4)
HM-A	Peak mortality	20,012,222	17,463,031 (87.3)
HM-V	Peak mortality	19,385,555	17,086,377 (88.1)
HM-R	Peak mortality	15.571.531	13,619,104 (87.5)
PM-A	Mortality plateau	16,760,301	14,861,247 (88.7)
PM-V	Mortality plateau	19,393,490	17,046,959 (87.9)
PM-R	Mortality plateau	19,998,031	17,244,459 (86.2)
Total		163.541.871	143.566.991 (87.8)

TABLE 3 | Summary of *de novo* assembly statistics of *P. mesopotamicus* liver transcriptome in *A. hydrophila* disease challenge.

Features	Results
Total bases of filtered reads	39,806,523,666
Total assembled bases	169,652,373
Number of transcripts	238,054
Average transcript length	713
Maximum transcript length	31,450
N50	1296
% reads mapped to the assembly	∼92
% paired-reads adequately mapped	76

39.81×10^9 bp to be explored in the analysis of the pacu transcriptome (**Table 2**). *De novo* assembly produced 238,054 contigs longer than 200 bp, totalizing 169,652,373 assembled bases. Additionally, the average contig length was 712.6 bp with an N50 value of 1,296 bp. Approximately 92% of the filtered reads mapped back to the assembly, of which 76% mapped adequately (both pair reads aligned in the correct orientation and in the same contig) (**Table 3**).

Posteriorly, 115 sequences mapped to the *A. hydrophila* genome, 20 sequences mapped to pacu mitochondrial genome and 90 transcripts identified as part of zebrafish ribosomal RNA (rRNA) were identified as contaminants, and therefore were excluded of the analysis. In addition, CD-HIT-EST clustered around 11.2% (26,570) of the transcripts resulting in a non-redundant assembly containing 211,259 sequences.

Gene Annotation

A total of 25,807 (12.2%) transcripts were annotated based on currently known proteins in the UniprotKB/Swiss-Prot database (**Supplementary Table S2**). Most of the annotate transcripts (15,547; 60%) showed significant homology with *Astyanax mexicanus* proteins, followed by *Danio rerio*, *Cyprinus carpio*, *Clupea harengus*, and *Salmo salar* proteins with 2,340 (9%); 1,368 (5.3%); 574 (2.2%); and 422 (1,6%) matched transcripts, respectively. Laterally, 23,420 transcripts were assigned into one or more GO terms classifying 20,851 transcripts into biological processes, 20,604 transcripts into molecular function and 19,884 transcripts into cellular component category (**Figure 1**). The most frequent GO term subcategories were cellular process, metabolic process and biological regulation for biological process; binding and catalytic activity for molecular function; and cell, cell part and organelle for cellular component. Furthermore, GO annotation revealed pertinent information on 2,726 transcripts assigned to immune system processes.

Cluster of Orthologous Groups (COG) using the EggNOG database assigned 4,931 transcripts to 25 COG functional categories (**Supplementary Table S3** and **Figure 2**). The largest clusters were composed by transcripts of unknown function (1,075 transcripts, 21.8%), followed by clusters related to signal transduction mechanisms (884 transcripts, 17.9%), post-translational modification and protein turnover (chaperones systems) (510 transcripts, 10.3%). Other clusters related to transcripts involved in intracellular trafficking secretion and vesicular transport (296 transcripts, 6.0%), RNA processing and modification (187 transcripts, 3.8%) and defense mechanisms (61 transcripts, 1.2%) were well represented. Further, 8,422 transcripts were mapped to 405 KEGG pathways (**Supplementary Table S4**). The most represented KEGG pathways are shown in **Figure 3**. "Signal Transduction" pathway had the largest number of matched transcripts (1,540, 18.3%), followed by "Global and Overview Maps" (1,374, 16.3%), "Infectious Disease (viral)" (814, 9.6%), "Endocrine System" (732, 8.7%), and "Immune System" (703, 8.3%) pathways.

Differential Gene Expression Analysis

Differential expression analyses between pooled samples belonging to control and HM group showed 4,413 differentially expressed transcripts, where 2,000 transcripts were up-regulated (45.3%) and 2,413 transcripts were down-regulated (54.7%) in HM group (q-value < 0.05, $\log_2 FC > |1|$) (**Figure 4A**, **Table 4**, and **Supplementary Table S5**). Additionally, 433 transcripts were differentially expressed between PM group and control fish, with 155 up-regulated (35.8%) and 278 down-regulated (64.2%) transcripts considering the PM group (q-value < 0.05, $\log_2 FC > |1|$) (**Figure 4B**, **Table 4**, and **Supplementary Table S6**).

A significant fraction of the differentially expressed transcripts showed $\log_2 FC > |5|$, specifically 400 genes (9%) differentially expressed for HM group, and 5 genes (1.2%) differentially expressed in PM group. Further, a total of 4,271 genes (96.8%) were differentially expressed only in HM group, whereas 291 genes (67.2%) were differentially expressed exclusively in PM group (**Figures 4A,B** and **Table 4**).

In this study, only genes specifically related to the immune responses against *A. hydrophila* and differentially expressed in HM (**Table 5**) and PM (**Table 6**) groups were considered. Therefore, 57 transcripts identified as immune transcripts were considered differentially expressed in the HM group (31 up-regulated and 26 down-regulated transcripts) (q-value < 0.09) (**Table 5**), whereas 23 immune transcripts were considered differentially expressed in the PM group (15 up-regulated and 8 down-regulated transcripts) (q-value < 0.07) (**Table 6**). For didactic purposes, these transcripts were grouped and

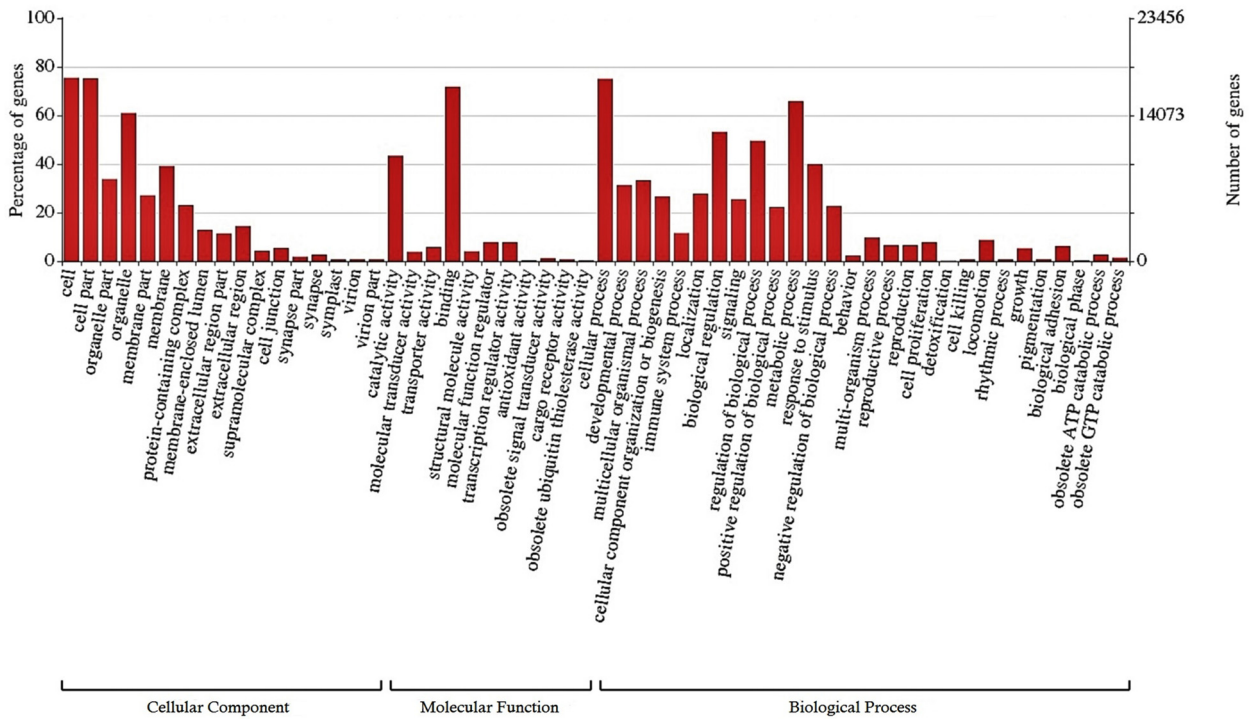


FIGURE 1 | Functional classification of the liver transcriptome of pacu (*P. mesopotamicus*) exposed to *A. hydrophila* infection. The transcripts were classified in the *x*-axis, according to Gene Ontology (GO) categories at level 2: cellular component, molecular function and biological process (E-value < 1e-3). The primary *y*-axis represents the percentage of genes classified in each subcategory, whereas the secondary *y*-axis represents the number of classified genes.

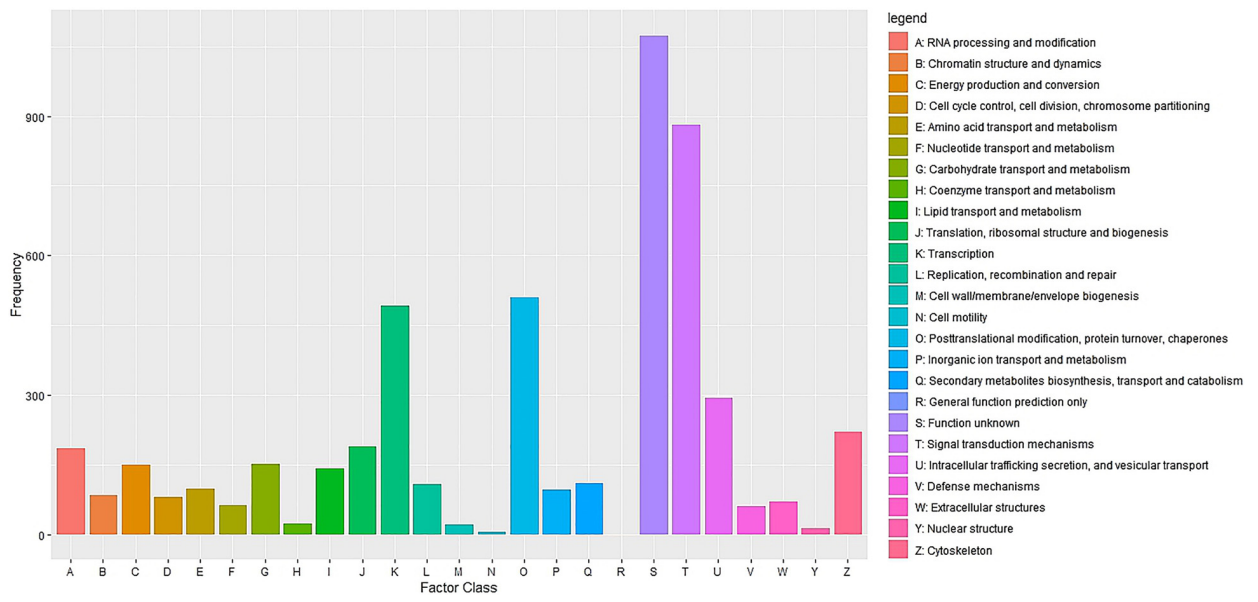


FIGURE 2 | Functional classification of the pacu liver transcriptome exposed to *A. hydrophila* infection based on eggnog/COG orthologous groups (E-value < 1e-3). 4,931 pacu transcripts were classified in 25 functional categories arranged in the *x*-axis. The *y*-axis presents the frequency of each represented category.

classified according to the important immune mechanisms that may be related to fighting infections such as those caused by *A. hydrophila*, such as: cytokines and signaling, epithelial protection, antigen processing and presentation, apoptosis, phagocytosis, complement system cascades and pattern recognition receptors (**Tables 5**, **6**).

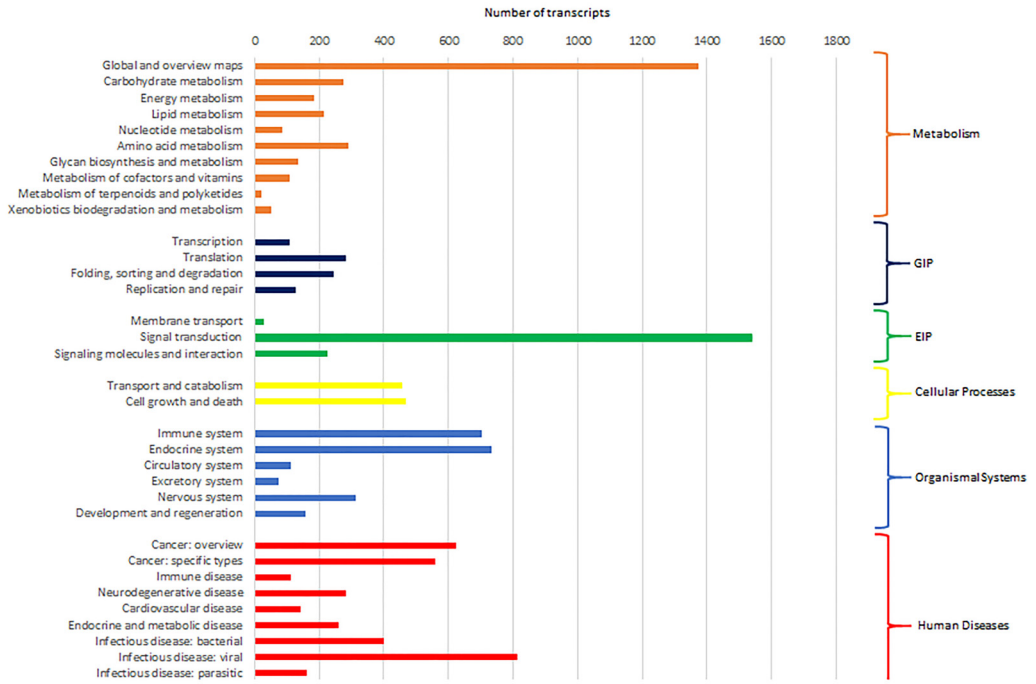


FIGURE 3 | Number of pacu transcripts (*x*-axis) assigned to each KEGG pathway arranged in the *y*-axis. The categories GIP and EIP stand for "Genetic Information Processing" and "Environmental Information Processing."

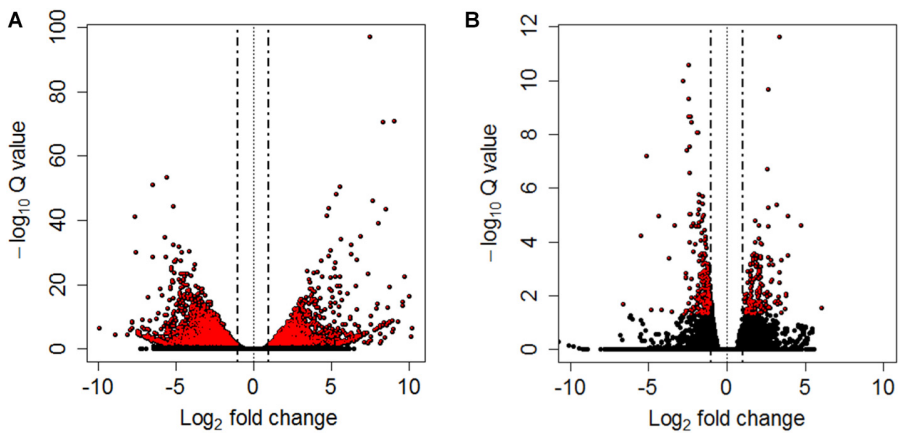


FIGURE 4 | Volcano plots between controls and HM **(A)** and PM **(B)** pacu (*P. mesopotamicus*) challenged with *A. hydrophila*. Each dot represents the expression (*x*-axis, Log_2 fold change), and significance (*y*-axis, $-\text{Log}_{10}$ *Q*-value) of each transcript. Differentially expressed transcripts (*q*-value < 0.05) are red.

DISCUSSION

De novo assembly from pacu liver RNA sequencing generated a high-quality transcriptome containing 211,259 transcripts, presenting an N50 value of 1,296 bp and average length of 713 bp. Moreover, default settings (kmer-length fixed to 25 and minimum contig length of 200 bp) used by the Trinity software provided an optimal assembly, as they subsequently allowed 92% of the reads to be mapped back to the transcriptome, having few transcriptomic information missing. The evident quality of the transcriptome was similar to other RNA-seq studies involving infections in fish (Wang et al., 2013; Kumar et al., 2017;

TABLE 4 | Number of differentially expressed transcripts of pacu (*P. mesopotamicus*) after infection with *A. hydrophila*.

Genes differentially expressed	Total	Upregulated	Downregulated
HM	4,413	2,000	2,413
PM	433	155	278
Only in HM fish	4,271	1,959	2,312
Only in PM fish	291	104	187

Comparisons of HM and PM fish vs. control. Upregulations means more expression in the HM/PM animals when compared to controls (transcript q-value < 0.05). LFC > 1 were considered upregulated genes and LFC < −1 were considered downregulated genes.

Song et al., 2017). Therefore, this immune-enriched *de novo* pacu liver transcriptome can be used as framework to investigate transcriptomic changes upon infection by *A. hydrophila* and other pathogens.

This is the first study on gene expression of pacu infected by *A. hydrophila*. The choice to focus only on genes related to the immune system was due to the generation of subsidies for future studies on application of immunostimulating treatments that could increase the pacu resistance when exposed to infectious outbreaks of *A. hydrophila*. Additionally, this results will contribute to develop effective strategies to protect pacu production from this threatening bacterial disease, and could be exploited through future development of breeding programs to increase the genetic resistance of pacu to *A. hydrophila* infection. Our results were similar to previous transcriptomic analyses in different species of fish infected by *A. hydrophila* (Tran et al., 2015; Dang et al., 2016; Qin et al., 2017; Jiang et al., 2018; Maekawa et al., 2019) showing transcripts related to innate immune system, which consist in a complex set of pathways that mediate the activation of phagocytosis, inflammation and destruction of the pathogen by antimicrobial substances, stimulating adaptive immune responses (Tort et al., 2003). In order to facilitate the discussion we have grouped these differentially expressed genes in categories based on their putative molecular mechanisms.

Cytokines and Signaling

Cytokines are small proteins released by immune cells acting as signaling molecules that regulate immunity, inflammation and hematopoiesis. We found the *Interleukin-1 receptor type 1* (*IL1R1*) and *Interleukin-1 receptor type 2* (*IL1R2*) transcripts upregulated in HM and PM groups. These transcripts are part of the Interleukin-1 receptor complex, involved in many cytokine-induced inflammatory responses.

We also found four chemokines upregulated and four downregulated in HM group. Chemokines are cytokines that regulate the leukocyte migration and the differentiation of recruited cells under inflammatory conditions, in order to orchestrate the first steps of both innate and adaptive immune responses (Bird and Tafalla, 2015). Further, we found downregulated genes involved in chemokine activation in HM group, and one transcript related to the negative regulation of cytokines (*Suppressor of cytokine signaling, SOCS1*), downregulated in PM group. Excess production of anti-inflammatory cytokines can compromise the ability to clear microorganisms through suppression of immune cell function. If a balance is not maintained, the result is either an excessive proinflammatory response or an immunosuppression (Chaudhry et al., 2015).

Phagocytosis

Phagocytosis is an essential cellular process that captures and destroys foreign particles, such as small unicellular organisms. One of the major effects of *A. hydrophila* infection in fish is characterized by the production of toxins that can cause severe tissue destruction, affecting the epithelial cells due to the immune phagocytic activity in response to the bacteria (Cipriano et al., 2001; Abdelhamed et al., 2017). According to the results, *Natural resistance-associated macrophage protein 2* (*NRAMP2*) transcript was upregulated only in PM group. This protein has been previously considered an important regulator of macrophages in the response to *A. hydrophila* (Jiang et al., 2018). Further, a previous study on *cytochrome B-245 beta chain* (*CYBB*) transcription, upregulated in HM group, revealed that deficiency in the production of the protein impairs phagocytosis-mediated defense against *A. hydrophila*. This means that even if the phagocytes are able to phagocytize the bacteria, *A. hydrophila* can survive into the phagocytic vacuoles (Jiang et al., 2018).

Pattern Recognition Receptors

Interleukin-1 receptor-associated kinase 1 (*IRAK1*) transcript is directly involved in the recognition of pathogen-derived products and the consequent initiation of immune and inflammatory responses. This protein is considered part of Toll-like receptor complex, and it is usually expressed on sentinel cells, such as macrophages and dendritic cells, which recognize structurally conserved molecules derived from pathogens. *Interleukin 1 receptor-associated kinase 3* (*IRAK3*), which inhibits the dissociation of *IRAK1* from the Toll-like receptor signaling complex (Shan et al., 2015), is upregulated in HM group. Therefore, these results can suggest that these genes are involved in the induction of the inflammatory response in response to *A. hydrophila* infection in pacu.

Complement System

The complement system is a key component of the innate immune system of fish. The activation of the complement pathway leads to opsonization of pathogens, recruitment of inflammatory and immune competent cells, and direct killing of pathogens through lysis (Merle et al., 2015). In the present study, practically all complement transcripts identified were downregulated in HM group (especially C1, C3, and C4), suggesting that complement pathways is not activated following *Aeromonas hydrophila* infection, which is different than reported in previous studies considering other fish species (Tran et al., 2015; Dang et al., 2016; Jiang et al., 2018). Complement system inactivation in HM group could be an important factor leading to the failure of the defense system, and consequent host death during the PM time-series that we found in the present study. However, further studies need to be performed in order to

TABLE 5 | List of the most significant differentially expressed immune system transcripts in HM group of *P. mesopotamicus* exposed to *A. hydrophila* infection vs. the controls.

Category	Gene name	Log_2 Fold	Diff	Q-value
Epithelial protection	Guanylin	10.11	UP	0.0002
	NADPH oxidase organizer 1	9.29	UP	3.33e-09
	Protein-glutamine gamma-glutamyltransferase K	5.27	UP	2.49e-23
	Microfibril-associated glycoprotein 4	3.18	UP	2.61e-10
	Transmembrane serine protease 4	−5.37	DOWN	4.72e-20
Cytokines and signaling	Interleukin-1 receptor type 2	7.86	UP	5.02e-07
	Prostaglandin E synthase	7.45	UP	6.25e-98
	Relaxin-3	6.85	UP	7.69e-36
	C-C motif chemokine 20	6.50	UP	1.78e-08
	C-C motif chemokine 24	4.63	UP	1.78e-06
	Interleukin-1 beta	4.39	UP	0.001
	C-C motif chemokine 13	4.19	UP	0.004
	Interleukin-1 receptor type 1	3.56	UP	1.82e-06
	C-X-C motif chemokine 10	3.36	UP	0.01
	Carcinoembryonic antigen-related cell adhesion molecule 20	2.50	UP	1.87e-09
	C-X-C motif chemokine 14	−4.26	DOWN	6.85e-19
	Cytokine-dependent hematopoietic cell linker	−4.21	DOWN	0.08
	C-C chemokine receptor type 5	−3.49	DOWN	0.001
	Chemokine-like receptor 1	−3.46	DOWN	0.06
	C-X-C motif chemokine 11	−3.17	DOWN	0.06
	Atypical chemokine receptor 4	−2.70	DOWN	0.03
Apoptosis	Lactate dehydrogenase A	5.58	UP	7.92e-35
	Protein KIBRA	3.37	UP	8.96e-17
	Apoptosis-associated speck-like protein containing a CARD	2.73	UP	3.20e-06
	Perforin-1	−4.95	DOWN	0.0004
Phagocytosis	Monocarboxylate transporter 4	6.57	UP	2.13e-20
	C-type lectin domain family 4 member M	3.76	UP	1.50e-15
	Cytochrome B-245 beta chain	3.64	UP	1.73e-21
	Carnitine O-acetyltransferase	−4.38	DOWN	1.65e-20
	Macrophage migration inhibitory factor	−2.30	DOWN	6.23e-05
Pattern recognition receptor	Interleukin 1 receptor associated kinase 3	4.40	UP	2.31e-23
	Interleukin-1 receptor-associated kinase 1	2.72	UP	0.02
	Toll-interacting protein	1.75	UP	6.50e-05
	Protein NLRC3	−3.21	DOWN	0.02
Complement system	Tissue factor pathway inhibitor 2	3.93	UP	1.87e-18
	Complement component C6	2.70	UP	1.87e-06
	Coagulation factor XI	−4.99	DOWN	2.21e-28
	Complement C1q subcomponent subunit B	−4.04	DOWN	0.0001
	Coagulation factor VII	−4.00	DOWN	1.35e-17
	Complement C1q subcomponent subunit A	−3.65	DOWN	0.06
	Coagulation Factor IX	−3.53	DOWN	7.52e-20
	Complement C1r-A subcomponent	−2.49	DOWN	0.02
	Complement C3	−1.78	DOWN	4.48e-05
	Complement factor I	−1.64	DOWN	0.002
	Complement C4	−1.36	DOWN	0.005
	Complement C4-B	−1.35	DOWN	0.003
Antigen processing and presentation	Polyhomeotic-like protein 2	7.63	UP	5.51e-47
	Heat shock 70 kDa protein	6.26	UP	2.60e-30
	S-adenosylmethionine synthase 1	5.16	UP	7.61e-26
	Purine nucleoside phosphorylase	4.82	UP	1.05e-29
	Adhesion G protein-coupled receptor E1	4.15	UP	0.0006
	CD97 antigen	2.28	UP	1.30e-05
	Cathepsin-O	−5.19	DOWN	3.74e-45
	Retinol-binding protein 4	−4.12	DOWN	6.11e-06
	MHC class I-related gene protein	−3.28	DOWN	0.01
	CD48 antigen	−3.12	DOWN	1.77e-09
	B-cell linker protein	−2.11	DOWN	0.03

TABLE 6 | List of most significant differentially expressed immune system transcripts in PM group of *P. mesopotamicus* exposed to *A. hydrophila* infection vs. the controls.

	Category and Gene name	Log_2 Fold	Diff	Q-value
Cytokines and signaling	TAB2 - TGF-beta-activated kinase 1 and MAP3K7-binding protein 2	1.91	UP	0.0007
	Innate immunity activator protein	2.60	UP	0.01
	Suppressor of cytokine signaling	−2.38	DOWN	0.0001
Apoptosis	E3 ubiquitin-protein ligase	2.69	UP	0.0007
	Ankyrin repeat and KH domain-containing protein 1	1.46	UP	0.06
	Krueppel-like factor 15	−2.13	DOWN	0.0002
	Glyceraldehyde 3-phosphate dehydrogenase	−1.43	DOWN	0.0003
Phagocytosis	Rho GTPase-activating protein 35	3.16	UP	4.15e-06
	Natural resistance-associated macrophage protein 2	2.00	UP	2.47e-05
	25-hydroxyvitamin D-1 alpha hydroxylase	−5.12	DOWN	6.34e-08
	F173B - Protein FAM173B	−2.00	DOWN	0.0001
	Macrophage migration inhibitory factor	−1.58	DOWN	0.001
Pattern recognition receptor	Ladderlectin	5.97	UP	0
	DENN domain-containing protein 5B	1.91	UP	0.0003
	Proto-oncogene c-Fos	1.68	UP	0.0004
Complement system	Coagulation factor XI	1.62	UP	0.0003
	Plasma kallikrein	1.82	UP	1.66e-05
Antigen processing and presentation	Early growth response protein 1	3.35	UP	2.30e-12
	Rho GTPase-activating protein 5	2.14	UP	0.0003
	B-cell lymphoma 6 protein homolog	2.02	UP	0.02
	High affinity immunoglobulin gamma Fc receptor I	1.33	UP	0.04
	Hemopexin	−2.58	DOWN	3.92e-08
	Adenosylhomocysteinase B	−1.51	DOWN	1.00e-05

validate these genes and corroborate this hypothesis. In the same way, several coagulation factors were also downregulated in HM group, such as *Coagulation factor VII* (*F7*), *Coagulation Factor IX* (*F9*), and *Coagulation Factor XI* (*F11*), indicating the absence of blood coagulation, and favoring the occurrence of hemorrhages, as observed in HM group with individuals presenting clinical signs of *A. hydrophila* infection.

CONCLUSION

Transcriptome analysis by RNA-seq allowed the identification and comparative quantification of gene expression in pacu infected by *A. hydrophila*. *De novo* liver transcriptome assembled in this study enables capturing expression levels of specific transcripts associated with important immune pathways, such as the coagulation factor and complement system cascades. Using this transcriptome, we have described for the first time the response of pacu infected by *A. hydrophila* in two critical periods of infection (PM and HM) comparing individuals belonging two distinct groups based on different phases of bacterial infection (acute mortality and plateau of mortality). To our knowledge this is the first descriptive study on transcriptome analysis of infected pacu that may serve as a reference to understand the genetic mechanisms underlying the host-pathogen interaction. The differences in the transcriptomic responses to a standardized *A. hydrophila* challenge was able to provide a initial data set to begin to understand the impact of selective breeding on genetic basis of disease resistance. Therefore, further studies trying to connect the genomic and genetic basis of resistance to *A. hydrophila* are fundamental to enable the implementation of *A. hydrophila* resistance into pacu breeding programs. Despite these limitations, these results may suggest possible suppression of genes belonging to the immune system that may contribute to susceptibility to infection, as for example genes related with complement system and coagulation factor cascades. Such supression of important immune system genes will serve as subsidies for future studies aiming at the use of prophylactic and immunostimulating treatments that could increase the pacu resistance when exposed to infectious outbreaks of *A. hydrophila*.

DATA AVAILABILITY STATEMENT

All RNA-Seq data generated for this study were deposited in the National Center for Biotechnology Information Sequence Read Archive under BioProject ID PRJNA632934. Bioinformatics pipelines are available in **Supplementary Table S1**.

ETHICS STATEMENT

This study was conducted in strict accordance with the recommendations of the National Council for Control of Animal Experimentation (CONCEA) (Brazilian Ministry for Science, Technology and Innovation) and was approved by the Ethics Committee on Animal Use (CEUA number 19.005/17) of Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus Jaboticabal, SP, Brazil.

AUTHOR CONTRIBUTIONS

VM-F and DH: conceptualization, funding acquisition, and roles/writing – original draft. VM-F, DR, RH, and DH: formal analysis. VM-F, MH, RK, MF, RA, and DP: investigation and methodology. DH: project administration. DH, DR, and RH: supervision. All authors: writing – review and editing.

FUNDING

This work was supported by the São Paulo Research Foundation (FAPESP grants 2016/21011-9, 2017/26900-9, and 2016/18294-9), the Comisión Nacional de Investigación Científica y Tecnológica (International Call FAPESP-CONICYT 2018/08416-5), the National Council for Scientific and Technological Development (CNPq grant 311559/2018-2), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES – Finance Code 001), and the BBSRC Institute Strategic Program Grants (BB/P013759/1 and BB/P013740/1).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00604/full#supplementary-material

REFERENCES

Abdelhamed, H., Ibrahim, I., Baumgartner, W., Lawrence, M. L., and Karsi, A. (2017). Characterization of histopathological and ultrastructural changes in channel catfish experimentally infected with virulent *Aeromonas hydrophila*. *Front. Microbiol.* 8:1519. doi: 10.3389/fmicb.2017.01519

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed April 15, 2018).

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556.Gene

Belém-Costa, A., and Cyrino, J. E. P. (2006). Antibiotic resistence of *Aeromonas hydrophila* isolated from *Piaractus mesopotamicus* (Holmberg, 1887) and *Oreochromis niloticus* (Linnaeus, 1758). *Sci. Agric.* 63, 281–284. doi: 10.1590/s0103-90162006000300011

Benavides, M. V., and Maciel, P. O. (2015). *Melhoramento de Espécies Aquícolas Com Foco na Resistência a Doenças*. Palmas: Embrapa Pesca e Aquicultura.

Bird, S., and Tafalla, C. (2015). Teleost chemokines and their receptors. *Biology* 4, 756–784. doi: 10.3390/biology4040756

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., et al. (2017). A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18, 762–776. doi: 10.1016/j.celrep.2016.12.063

Carraschi, S. P., Cruz, C., Neto, J. G. M., Moraes, F., Rossi, O. D., Neto, A. N., et al. (2012). Evaluation of experimental infection with *Aeromonas hydrophila* in pacu (*Piaractus mesopotamicus*) (Holmberg, 1887). *Int. J. Fish. Aquac.* 4, 81–84. doi: 10.5897/IJFA11.008

Chaudhry, H., Zhou, J., Zhong, Y., Ali, M. M., McGuire, F., Nagarkatti, P. S., et al. (2015). Role of cytokines as a double-edged sword in sepsis. *In Vivo* 27, 669–684.

Cipriano, R. C., Bullock, G. L., and Pyle, S. W. (2001). *Aeromonas hydrophila and motile aeromonad septicemias of fish. Fish disease leaflet*. Washington: U.S. Fish and Wildlife Service.

da Pimentel, J. S. M., do Carmo, A. O., de Maciel, D. C. L., de Faria Siqueira, F., and Kalapothakis, E. (2014). Complete mitochondrial genome sequence of *Piaractus mesopotamicus* (Holmberg, 1887). *Mitochondrial DNA* 1736, 1–2. doi: 10.3109/19401736.2014.971297

Dang, Y., Xu, X., Shen, Y., Hu, M., Zhang, M., Li, L., et al. (2016). Transcriptome analysis of the innate immunity-related complement system in spleen tissue of *Ctenopharyngodon idella* infected with *Aeromonas hydrophila*. *PLoS One* 11:e0157413. doi: 10.1371/journal.pone.0157413

FAO, (2018). *The State of World Fisheries and Aquaculture 2018 – Meeting the Sustainable Development Goals*. Rome: FAO.

Farias, T. H. V., Levy-Pereira, N., de Oliveira Alves, L., de Carla Dias, D., Tachibana, L., Pilarski, F., et al. (2016). Probiotic feeding improves the immunity of pacus, *Piaractus mesopotamicus*, during *Aeromonas hydrophila* infection. *Anim. Feed Sci. Tech.* 211, 137–144. doi: 10.1016/j.anifeedsci.2015.11.004

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

IBGE (2016). *Produção da pecuária municipal v. 44*. Rio de Janeiro: IBGE.

Jiang, H., Wang, M., Fu, L., Zhong, L., Liu, G., Zheng, Y., et al. (2020). Liver transcriptome analysis and cortisol immune-response modulation in lipopolysaccharide-stimulated in channel catfish (*Ictalurus punctatus*). *Fish Shellfish Immun.* 101, 19–50. doi: 10.1016/j.fsi.2020.03.024

Jiang, N., Fan, Y., Zhou, Y., Wang, W., Ma, J., and Zeng, L. (2018). Transcriptome analysis of *Aeromonas hydrophila* infected hybrid sturgeon (*Huso dauricus* × *Acipenser schrenckii*). *Sci. Rep.* 8, 1–18. doi: 10.1038/s41598-018-36376-2

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40, 109–114. doi: 10.1093/nar/gkr988

Kumar, R., Sahoo, P. K., and Barat, A. (2017). Transcriptome profiling and expression analysis of immune responsive genes in the liver of Golden mahseer (*Tor putitora*) challenged with *Aeromonas hydrophila*. *Fish Shellfish Immun.* 67, 655–666. doi: 10.1016/j.fsi.2017.06.053

Lafferty, K. D., Harvell, C. D., Conrad, J. M., Friedman, C. S., Kent, M. L., Kuris, A. M., et al. (2015). Infectious diseases affect marine fisheries and aquaculture economics. *Ann. Rev. Mar. Sci.* 7, 471–496. doi: 10.1146/annurev-marine-010814-015646

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21. doi: 10.1186/s13059-014-0550-8

Maekawa, S., Wang, P. C., and Chen, S. C. (2019). Comparative study of immune reaction against bacterial infection from transcriptome analysis. *Front. Immunol.* 10:153. doi: 10.3389/fimmu.2019.00153

Mahapatra, K., Das Gjerde, B., Sahoo, P. K., Saha, J. N., Barat, A., Sahoo, M., et al. (2008). Genetic variations in survival of rohu carp (*Labeo rohita*, Hamilton) after *Aeromonas hydrophila* infection in challenge tests. *Aquaculture* 279, 29–34. doi: 10.1016/j.aquaculture.2008.03.054

Marancik, D., Gao, G., Paneru, B., Ma, H., Hernandez, A. G., Salem, M., et al. (2015). Whole-body transcriptome of selectively bred, resistant-, control-, and susceptible-line rainbow trout following experimental challenge with *Flavobacterium psychrophilum*. *Front. Genet.* 5:453. doi: 10.3389/fgene.2014.00453

Mastrochirico-Filho, V. A., Ariede, R. B., Freitas, M. V., Lira, L. V. G., Agudelo, J. F. G., Pilarski, F., et al. (2019). Genetic parameters for resistance to *Aeromonas hydrophila* in the Neotropical fish pacu (*Piaractus mesopotamicus*). *Aquaculture* 513:734442. doi: 10.1016/j.aquaculture.2019.734442

Merle, N. S., Noe, R., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V., and Roumenina, L. T. (2015). Complement system part II: role in immunity. *Front. Immunol.* 6:257. doi: 10.3389/fimmu.2015.00257

Ødegård, J., Olesen, I., Dixon, P., Jeney, Z., Nielsen, H. M., Way, K., et al. (2010). Genetic analysis of common carp (*Cyprinus carpio*) strains. II: resistance to koi herpesvirus and *Aeromonas hydrophila* and their relationship with pond survival. *Aquaculture* 304, 7–13. doi: 10.1016/j.aquaculture.2010.03.017

Owens, L. (2019). "Disease principles," in *Aquaculture: Farming Aquatic Animals and Plants*, 3rd Edn, eds J. S. Lucas, P. C. Southgate, and C. S. Tucker, (Chichester: Wiley-Blackwell), 203–216.

Polinski, M. P., Bradshaw, J. C., Inkpen, S. M., Richard, J., Fritsvold, C., Poppe, T. T., et al. (2016). De novo assembly of Sockeye salmon kidney transcriptomes reveal a limited early response to piscine reovirus with or without infectious hematopoietic necrosis virus superinfection. *BMC Genomics* 17:848. doi: 10.1186/s12864-016-3196-y

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., et al. (2012). eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* 40, 284–289. doi: 10.1093/nar/gkr1060

Qin, C., Gong, Q., Wen, Z., Yuan, D., Shao, T., Wang, J., et al. (2017). Transcriptome analysis of the spleen of the darkbarbel catfish *Pelteobagrus vachellii* in response to *Aeromonas hydrophila* infection. *Fish Shellfish Immun.* 70, 498–506. doi: 10.1016/j.fsi.2017.09.042

Rebl, A., and Goldammer, T. (2018). Under control: the innate immunity of fish from the inhibitors' perspective. *Fish Shellfish Immun.* 77, 328–349. doi: 10.1016/j.fsi.2018.04.016

Robledo, D., Gutiérrez, A. P., Barría, A., Yáñez, J. M., and Houston, R. D. (2018). Gene expression response to sea lice in Atlantic salmon skin: RNA sequencing comparison between resistant and susceptible animals. *Front. Genet.* 9:287. doi: 10.3389/fgene.2018.00287

Robledo, D., Ronza, P., Harrison, P. W., Losada, P. P., Bermúdez, R., Pardo, B. G., et al. (2014). RNA-seq analysis reveals significant transcriptome changes in turbot (*Scophthalmus maximus*) suffering severe enteromyxosis. *BMC Genomics* 15:1149. doi: 10.1186/1471-2164-15-1149

Shan, S. J., Liu, D. Z., Wang, L., Zhu, Y. Y., Zhang, F. M., Li, T., et al. (2015). Identification and expression analysis of irak1 gene in common carp *Cyprinus carpio* L .: indications for a role of antibacterial and antiviral immunity. *J. Fish Biol.* 87, 241–255. doi: 10.1111/jfb.12714

Silva, B. C., Mouriño, J. L. P., Vieira, F. N., Jatobá, A., Seiffert, W. Q., and Martins, M. L. (2012). Haemorrhagic septicaemia in the hybrid surubim (*Pseudoplatystoma corruscans*×*Pseudoplatystoma fasciatum*) caused by *Aeromonas hydrophila*. *Aquac. Res.* 43, 908–916. doi: 10.1111/j.1365-2109.2011.02905.x

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., and Kelly, S. (2016). TransRate: reference free quality assessment of de novo transcriptome assemblies. *Genome Res.* 26:021626. doi: 10.1101/gr.196469.115

Song, X., Hu, X., Sun, B., Bo, Y., Wu, K., Xiao, L., et al. (2017). A transcriptome analysis focusing on inflammation-related genes of grass carp intestines following infection with *Aeromonas hydrophila*. *Sci. Rep.* 7, 1–12. doi: 10.1038/srep40777

Stentiford, G. D., Sritunyalucksana, K., Flegel, T. W., Williams, B. A. P., Withyachumnarnkul, B., Itsathitphaisarn, O., et al. (2017). New paradigms to help solve the global aquaculture disease crisis. *PLoS Pathog.* 13:e1006160. doi: 10.1371/journal.ppat.1006160

Stratev, D., and Odeyemi, O. A. (2017). An overview of motile *Aeromonas septicaemia* management. *Aquac. Int.* 25, 1095–1105. doi: 10.1007/s10499-016-0100-3

Sudhagar, A., Kumar, G., and El-Matbouli, M. (2018). Transcriptome analysis based on RNA-Seq in understanding pathogenic mechanisms of diseases and the immune system of fish: a comprehensive review. *Int. J. Mol. Sci.* 19, 245. doi: 10.3390/ijms19010245

Tort, L., Balasch, J. C., and Mackenzie, S. (2003). Fish immune system. A crossroads between innate and adaptative responses. *Inmunologia* 22, 277–286.

Tran, N. T., Gao, Z., Zhao, H., Yi, S., Chen, B., Zhao, Y., et al. (2015). Transcriptome analysis and microsatellite discovery in the blunt snout bream (*Megalobrama amblycephala*) after challenge with *Aeromonas hydrophila*. *Fish Shellfish Immun.* 45, 72–82. doi: 10.1016/j.fsi.2015.01.034

Wang, R., Sun, L., Bao, L., Zhang, J., Jiang, Y., Yao, J., et al. (2013). Bulk segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish. *BMC Genomics* 14:929. doi: 10.1186/1471-2164-14-929

Woynárovich, A., and Van Anrooy, R. (2019). *Field Guide to the Culture of Tambaqui (Colossoma macropomum, Cuvier, 1816)*. FAO Fisheries and Aquaculture Technical Paper No. 624. Rome: FAO.

Xiong, X., Chen, Y., Liu, L., Wang, W., Robinson, N. A., and Gao, Z. (2017). Estimation of genetic parameters for resistance to *Aeromonas hydrophila* in blunt snout bream (*Megalobrama amblycephala*). *Aquaculture* 479, 768–773. doi: 10.1016/j.aquaculture.2017.07.011

Yang, W., Li, N., Li, M., Zhang, D., and An, G. (2016). Complete genome sequence of fish pathogen *Aeromonas hydrophila* JBN2301. *Genome Announc.* 4, 16–17. doi: 10.1128/genomeA.01615-15.Copyright

Zanuzzo, F. S., Sabioni, R. E., Montoya, L. N. F., Favero, G., and Urbinati, E. C. (2017). *Aloe vera* enhances the innate immune response of pacu (*Piaractus mesopotamicus*) after transport stress and combined heat killed *Aeromonas hydrophila* infection. *Fish Shellfish Immun.* 65, 198–205. doi: 10.1016/j.fsi.2017.04.013

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 June 2020
doi: 10.3389/fmicb.2020.01301

NF-κB/TWIST1 Mediates Migration and Phagocytosis of Macrophages in the Mice Model of Implant-Associated *Staphylococcus aureus* Osteomyelitis

Yutian Wang[1,2†], Yihuang Lin[1,2†], Caiyu Cheng[1,2], Pengyu Chen[1,2], Ping Zhang[1], Hangtian Wu[1,2], Kaiqun Li[1,2], Ye Deng[1,2], Jikun Qian[1,2], Xianrong Zhang[1,2] and Bin Yu[1,2]**

[1] Department of Orthopaedics, Nanfang Hospital, Southern Medical University, Guangzhou, China, [2] Guangdong Provincial Key Laboratory of Bone and Cartilage Regenerative Medicine, Nanfang Hospital, Southern Medical University, Guangzhou, China

Staphylococcus aureus (*S. aureus*) infection-induced osteomyelitis is a great challenge in clinic treatment. Identification of the essential genes and biological processes that are specifically changed in mononuclear cells at an early stage of *S. aureus* osteomyelitis is of great clinical significance. Based on transcriptional dataset GSE16129 available publicly, a bioinformatic analysis was performed to identify the differentially expressed genes of osteomyelitis caused by *S. aureus* infection. ERBB2, TWIST1, and NANOG were screened out as the most valuable osteomyelitis-related genes (OMRGs). A mice model of implant-associated *S. aureus* osteomyelitis was used to verify the above genes. We found significantly up-regulated expression of TWIST1 in macrophages and accumulation of macrophages around the infected implant. Meanwhile, *S. aureus* infection increased the expression of TWIST1, MMP9, and MMP13, and stimulated the migration and phagocytosis function of Raw 264.7 cells. Additionally, knock-down of the expression of TWIST1 by siRNA could significantly down-regulate MMP9 and MMP13 and suppress the migration and phagocytosis ability of macrophages in response to *S. aureus* infection. Furthermore, we found that NF-κB signaling was activated in Raw 264.7 cells by *S. aureus* and that inhibition of NF-κB signaling by Bay11–7082 blocked the expression of TWIST1, MMP9, and MMP13 as well as cell migration and phagocytosis evoked by *S. aureus*. Our findings demonstrate that NF-κB/TWIST1 is necessary for migration and phagocytosis of macrophages in response to *S. aureus* infection. Our study highlights the essential role of NF-κB/TWIST1 in early innate immune response to *S. aureus* infection in bone.

Keywords: ***Staphylococcus aureus*, osteomyelitis, macrophage, bioinformatics, bone remodeling**

OPEN ACCESS

Edited by:
Diego Robledo,
The University of Edinburgh,
United Kingdom

Reviewed by:
Elisabeth Seebach,
Heidelberg University Hospital,
Germany
Haider Abdul-Lateef Mousa,
University of Basrah, Iraq
Tao Jin,
University of Gothenburg, Sweden

***Correspondence:**
Xianrong Zhang
xianrongzh@smu.edu.cn
Bin Yu
yubin@smu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: *17 January 2020*
Accepted: *22 May 2020*
Published: *12 June 2020*

Citation:
Wang Y, Lin Y, Cheng C, Chen P, Zhang P, Wu H, Li K, Deng Y, Qian J, Zhang X and Yu B (2020) NF-κB/TWIST1 Mediates Migration and Phagocytosis of Macrophages in the Mice Model of Implant-Associated Staphylococcus aureus Osteomyelitis. Front. Microbiol. 11:1301. doi: 10.3389/fmicb.2020.01301

INTRODUCTION

Osteomyelitis is an inflammatory process in bone resulting from infection by such micro-organisms as bacteria, fungi, or mycobacteria (Xiong and Pamer, 2015; Momodu and Savaliya, 2019). Its incidence in the pediatric population is approximately 5–8/10,000 each year (Riise et al., 2008; Peltola and Paakkonen, 2014). Its major etiology can be hematogenous, tracking from adjacent foci of infection, and direct inoculation from trauma or surgery (Kavanagh et al., 2018). An acute process is considered when the duration of symptoms is less than 2 weeks (Yeo and Ramachandran, 2014). Acute hematogenous osteomyelitis usually affects the skeleton in children as the metaphysis of a growing long bone contains abundant blood vessels with leaky endothelium and sluggish blood flow, which is commonly the primary site of infection (Stephen et al., 2012; Whyte and Bielski, 2016). Delay in diagnosis and inappropriate treatment may lead to septicaemia and intense inflammatory reaction, which destroys bone structures with longitudinal growth arrest or bone defects, and may even bring about deteriorated outcomes of multiorgan failure, and death (Fink et al., 1977).

As the most common causative pathogen of osteomyelitis (Peltola and Paakkonen, 2014), *Staphylococcus aureus* (*S. aureus*) infection can trigger strong inflammatory response through its virulence factors and structural components (Bogoslowski et al., 2018; Putnam et al., 2019), leading to a substantial bone loss during osteomyelitis (Mbalaviele et al., 2017). Peripheral blood mononuclear cells (PBMCs), the precursors of dendritic cells, osteoclasts and macrophages, are key actors recruited which exert an immediate and potent immune response against *S. aureus* infection, playing an important role in inflammatory bone loss (Xiong and Pamer, 2015; de Vries et al., 2019). Subjects with suppressed monocytes/macrophages may have an increased susceptibility to *S. aureus* infection (Knobloch et al., 2019), whereas early accumulation of inflammatory monocytes/macrophages may serve as a reservoir for intracellular *S. aureus* survival, thereby promoting bacterial resistance to antibiotic treatment (Fischer et al., 2019). In a previous study, we demonstrated that the G-CSF-mediated bone loss might be due to aggregation of $F4/80^+$ macrophages (Hou et al., 2019). However, the events of migration and immune response by macrophages at an acute stage of *S. aureus* infection in bone have been poorly understood. Clarification of macrophage migration in response to *S. aureus* infection in bone is essential for identification of targets in treatment of osteomyelitis induced by *S. aureus* infection.

High-throughput transcriptional analysis is a useful technique to investigate the multidimensional networks of molecules and cells in response to stimulus (Kulasingam and Diamandis, 2008). In the present study, we downloaded the transcriptome profiles of GSE16129 (Ardura et al., 2009) and analyzed the differentially expressed genes (DEGs) in mononuclear/macrophage cells between *S. aureus* osteomyelitis patients (median age 7.5 years) and healthy controls (median age 6 years) using bioinformatics methods. Finally, TWIST1, NANOG, and ERBB2 were screened out as the DEGs most likely related to immunity of marrow and bone metabolism. Our study revealed that *S. aureus* infection might stimulate NF-κB/TWIST1 signaling, thereby promoting migration and phagocytosis of macrophages.

MATERIALS AND METHODS

Microarray Data Preprocessing

The genes expression profile GSE16129 from the study by Ardura et al. (2009) was downloaded from the publicly available gene expression omnibus database[1] based on three platforms (GPL96, GPL97, and GPL6106). A bioinformatic analysis was performed. As there was only one osteomyelitis sample with *S. aureus* infection in the GPL6106 platform, we did not include it in the present study. In these platforms, the total RNA extracted from PBMCs of healthy controls and patients was utilized for gene expression microarrays. R statistical software (version 3.5.2, R Project for Statistical Computing[2]) was used to perform the analysis process.

Analysis and Screen of DEGs

To further analyze the genes related to *S. aureus* osteomyelitis, three groups comparing healthy control (Ctrl) vs osteomyelitis-free infection (OFI), Ctrl vs osteomyelitis infection (OMI) and Ctrl vs *S. aureus* infection (SI; **Figures 1A,B**) were analyzed by limma package[3] (Diboun et al., 2006) and R statistical software (version 3.5.2). In order to reduce the false positive rate, the *P*-value was adjusted using Benjamini and Hochberg false discovery rate method. Only genes with $|\log2\ FC\ (\text{fold change})| > 1$ and adjusted *P*-value < 0.05 were identified as the DEGs. To screen the overlapping and unique genes in three comparison groups, Venn diagram[4] was used to analyze different DEGs (**Figures 1C,D**). The DEGs observed from Ctrl vs. OMI but not from Ctrl vs. OFI were considered as osteomyelitis-related genes (OMRGs).

Gene Ontology (GO) Functional Annotation, Kyoto Encyclopedia of Genes and Genomes (KEGG) Enrichment Analysis, Protein-Protein Interaction (PPI) Network Analysis and Module Identification

The database for annotation, visualization and integrated discovery (DAVID)[5] was used to perform GO functional annotation (Ashburner et al., 2000), and KEGG pathway enrichment analysis (Kanehisa et al., 2008) of the OMRGs. The significantly enriched GO terms and KEGG pathways with the thresholds of *P*-value <0.05 were selected for analyzing *S. aureus* osteomyelitis-associated biological process

[1]http://www.ncbi.nlm.nih.gov/geo/
[2]https://www.r-project.org/
[3]www.bioconductor.org
[4]http://bioinfogp.cnb.csic.es/tools/venny/index.html
[5]http://david.abcc.ncifcrf.gov

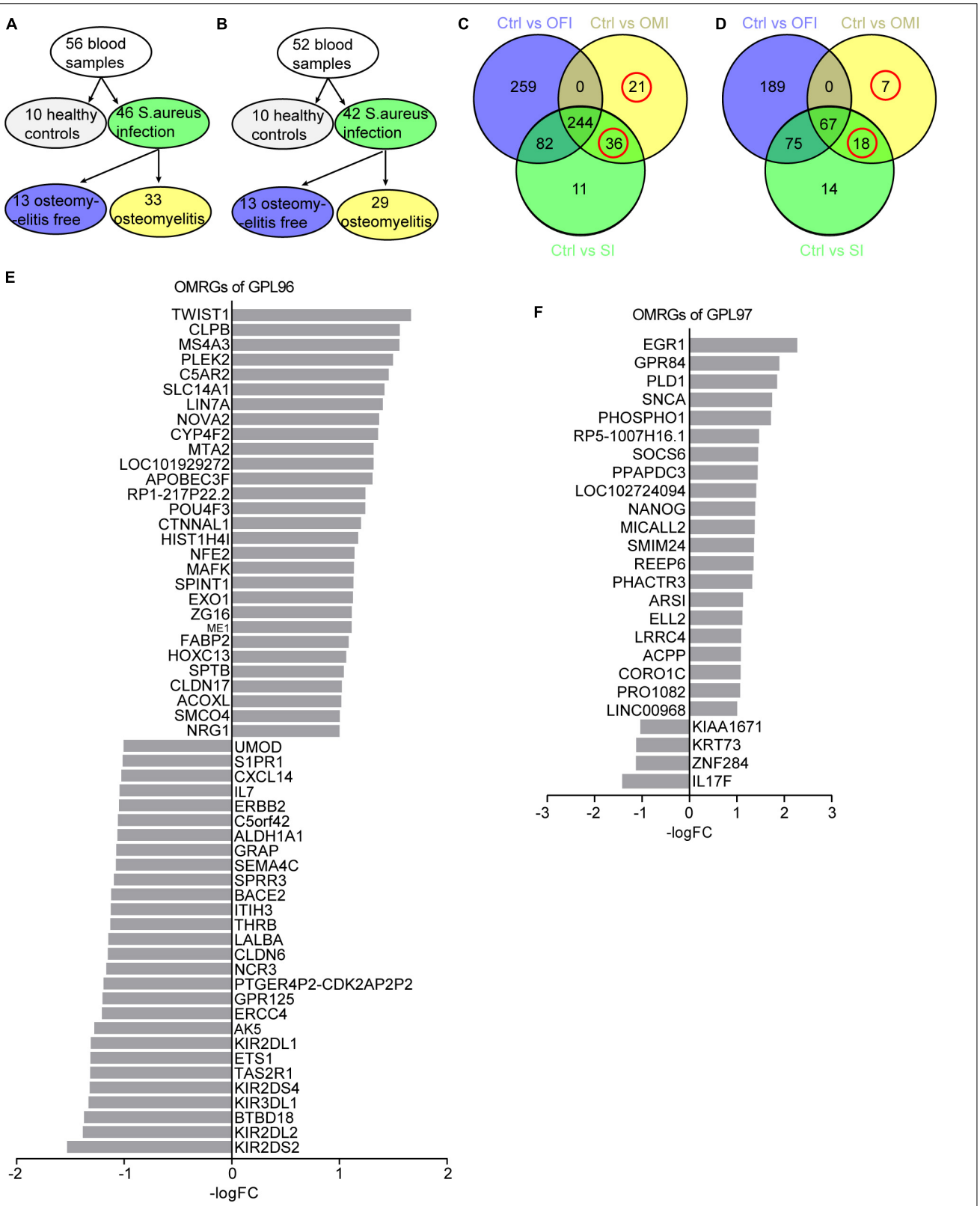


FIGURE 1 | Sample grouping and DEGs from OFI, OMI, and SI groups. **(A,B)** Grouping plots of transcriptional file from platform GPL96 and GPL97. **(C,D)** Venn diagram of DEGs and OMRGs of GPL96 and GPL97. Red circle represents the OMRGs. **(E,F)** All the up-regulated and downregulated OMRGs of GPL 96 and GPL97. The horizontal axis shows the -log (fold change), and the vertical axis represents gene names.

and pathways. The online tool STRING database[6] was used to establish the PPI network and identify the key proteins and important protein modules (Szklarczyk and Jensen, 2015). The Cytoscape software (version 3.6.1) was used to analyze degree of distribution, closeness centrality and betweenness centrality. The nodes that owned a relatively large number of edges (interactions) were identified as core proteins that play a major role in the PPI network. The top 15 core proteins with a high degree of relatedness were selected. Then, the module analysis was conducted with the MCODE plugin of Cytoscape software with thresholds of degree cutoff = 2, node score cutoff = 0.2, k-core $\geq$ 2, and max. depth = 100 (Bader and Hogue, 2003).

Bacterial Strain and Preparation

Staphylococcus aureus was isolated from the osteomyelitis patients in Nanfang Hospital and identified using PHOENIX 100 (Becton Dickinson Microbiology System, United States). For the infection experiments, *S. aureus* was inoculated in 10 ml fresh tryptic soy broth (TSB) and incubated overnight at 37°C with shaking at 200 rpm/min. Bacteria were collected by centrifugation, washed three times with phosphate-buffered saline (PBS), and re-suspended in PBS. The concentration of *S. aureus* was adjusted to an OD of 0.5 at 600 nm, approximately equal to 1×10^8 CFU/ml, and then diluted to different concentrations for infection in mice osteomyelitis model or in Raw264.7 macrophage cell line.

Animals

Eight-week old male C57BL/6 mice were purchased from Laboratory Animal Center of Southern Medical University (Guangzhou, China). Mice were housed under specific pathogen-free conditions at 24–26°C with a 12-h light/dark cycle and access to food and water *ad libitum*. All interventions and animal care procedures were performed according to the Guidelines and Policies for Animal Surgery provided by Nanfang Hospital. The osteomyelitis model used in this study was modified from that reported by Bernthal et al. (2010). Briefly, after anesthesia by intraperitoneal injection of tribromoethanol and sterilization, a 5 mm incision was made along the dorsal side of right femur to expose the bone surface. A hole on the cortical bone of mid-diaphysis of the femur was drilled with a 25-gage syringe needle, without penetrating the cortical bone on the other side. Next, a 2 mm-long sterile stainless steel needle (0.3 mm in diameter) was inserted into the bone marrow cavity through the hole. Mice were divided randomly into *S. aureus* osteomyelitis group (n = 19) and control group (n = 19). In the *S. aureus* osteomyelitis group, 1×10^6 CFU/ml *S. aureus* in 2 μl of PBS was injected into the canal of the bone marrow cavity using a micro syringe (Sangon, China, Wuhan) while the control mice were injected with the same volume of sterile PBS. The hole was sealed with bone wax, and the incisions were closed with 5–0 sutures. Analgesics was given for perioperative analgesia to minimize pain. On day 3 after surgery and infection, animals were killed by cervical vertebra dislocation, and the right femurs were collected for further analysis.

[6]http://www.string-db.org/

Histochemistry, Immunohistochemistry, and Immunofluorescence

Femurs were fixed with 4% paraformaldehyde overnight, decalcified in 10% EDTA (pH 7.4) for 4 weeks, and finally embedded in paraffin. Longitudinally oriented 4 μm-thick sagittal sections were cut and processed for staining. From every consecutive ten sections, two were chosen for hematoxylin and eosin (H&E) staining.

Immunohistochemistry staining was conducted according to the standard protocol. After being deparaffinized and rehydrated, sections were incubated with primary antibodies to TWIST1 (Cat. 25465-1-AP, Proteintech, Wuhan, China), and F4/80 (Cat.71299, Cell Signal Technology, MA, United States) overnight at 4°C. After washing in PBS, sections were incubated with a biotinylated secondary antibody and then with an avidin-biotinylated horseradish peroxidase complex (Vectastain ABC Kit, Vector Laboratories, United States) according to the manufacturer's protocol. Peroxidase activity was revealed by DAB (PW017, Sangon Biotech, Shanghai, China).

For immunofluorescence staining, frozen sections were incubated with primary antibody to TWIST1 (Cat.25465-1-AP, Proteintech, Wuhan, China), and F4/80 (Cat.71299, Cell Signal Technology, MA, United States) overnight at 4°C, followed by incubation with Alexa Fluor 594-conjugated goat anti-rabbit IgG (H + L; Cat.SA00006-4; Proteintech, Rosemont, IL, United States), and Alexa Fluor 488-conjugated goat anti-mouse IgG (H + L; Cat. SA00006-1; Proteintech) for 1 h at room temperature. Nuclei were counterstained with DAPI (E607303–0002; BBI Life Science, Shanghai, China). The sections were observed through a BX63 microscope (Olympus, Tokyo, Japan). To quantify the positive-stained cells adjacent to implant, we randomly chose 2 non-overlapping area in three sections of each mice sample.

Total RNA Extraction and Real-Time Quantitative PCR (qPCR)

To obtain tissue RNA, bone was homogenized in TRIzol reagent (TaKaRa, Dalian China), followed by total RNA isolation. RNA was then reverse transcribed to cDNA with PrimeScript^{TM} RT reagent Kit (TaKaRa, Dalian, China). qPCR was conducted using the iQ5 (Bio-Rad, Hercules, CA, United States) with SYBRP remix Ex Taq^{TM} (TaKaRa, Dalian, China). The relative expression of genes was normalized to GAPDH and processed by the $2^{-\Delta\Delta Ct}$ method. All primers are listed in **Supplementary Table S1**.

Western Blotting

Bone medullary cavity contents were collected and then lysed by RIPA buffer with proteinase inhibitors (KGP250; KeyGen BioTech, Jiangning, China). Equal amounts of protein (30 μg) were resolved by SDS-PAGE on 10% polyacrylamide gels and then transferred to PVDF membrane. After being blocked with

5% BSA solution at room temperature for 1 h and washed with 1 × tris-buffered saline tween (TBST) three times, membranes were then incubated with the primary antibodies to TWIST1 (Cat. 25465-1-AP, Proteintech, Wuhan, China), P65 (Cat. AF5006, Affinity Cincinnati, OH, United States), phospho-P65 (Cat. AF2006, Affinity Cincinnati, OH, United States), MMP9 (Cat. 13667, Cell Signal Technology, MA, United States), and MMP13 (Cat. ab39012, Abcam, Cambridge, MA, United States). The membranes were further incubated with a HRP-conjugated anti-rabbit (ab6721, Abcam, Cambridge, MA, United States), anti-mouse (ab6728, Abcam, Cambridge, MA, United States) secondary antibody for 1 h at room temperature. All blots were developed using ECL (MilliporeSigma, Burlington, MA, United States). Images were captured with Tanon chemiluminescence apparatus (Tanon-5200Multi, Tanon, China, Shanghai).

Cell Culture and Treatments

Mouse Raw264.7 cell line was purchased from Cell Bank, Shanghai Institute of Biochemistry and Cell Biology at the Chinese Academy of Sciences (Shanghai, China). Cells were seeded at a density of 2×10^5 cells/well in 6-well plate and cultured in high-glucose medium DMEM (HyClone, United States) containing 10% fetal bovine serum (FBS; Gibco, United States). To evaluate the expression of genes in macrophage in response to *S. aureus* infection, cells were serum starved overnight, then infected by *S. aureus* with the multiplicity of infection (MOI) range from 0.01 to 10. After 1 h infection, extracellular *S. aureus* were killed with 20 μg/ml gentamicin (Cat. B20192, Sigma Aldrich, St. Louis, Missouri, United States) for 30 min. After washing with PBS three times, cells were allowed to grow in fresh 10% FBS medium for additional 24 h, and then RNA was harvested for mRNA expression analysis.

To knockdown TWIST1 expression, Raw 264.7 cells (2×10^5 cells/well) were seeded in 6-well plate and cultured for 24 h, and then transient RNA interference was performed following the manufacturer's instructions. Cells were transfected with small interference RNAs for TWIST1 (siRNAs; siBDM1999A, RIBOBIO, GUANGZHOU, China) using Lipofectamine 3000 (Invitrogen, Carlsbad, California, United States). The sequences of designed siRNAs targeting TWIST1 were as follows: si-TWIST1-1 (CAAGATTCAGACCCTCAAA), si-TWIST1-2 (GATGGCAAGCTGCAGCTAT), and si-Twist1-3 (GACTCCAAGATGGCAAGCT). 24 h after transfection, cells were infected with *S. aureus* at MOI of 0.01 for 1 h. Next, after killing the extracellular *S. aureus* with 20 μg/ml gentamicin for 30 min, cells were allowed to grow in fresh 10% FBS medium for additional 24 h, or 48 h. Then RNA and protein were harvested for further analysis.

To evaluate the role of NF-κB signaling in microphage during *S. aureus* infection, cells were pretreated with 10 μM BAY11-7082 (BAY; Cat. S7352; Selleck, Houston, United States) for 2 h, and then infected with *S. aureus* at a MOI of 0.01 for 1 h. Next, *S. aureus* was killed by adding 20 μg/ml gentamicin for 30 min. Next, cells were washed with PBS three times and cultured in fresh 10% FBS medium for additional 48 h, and then proteins were harvested for western blotting.

Trypan Blue Staining

Trypan blue staining was used to determine the optimal MOI and observe the viability of Raw 264.7 cells with *S. aureus* infection in different MOI. 5×10^5 cells were plated in 6-well plates. 24 h later, cells were infected with *S. aureus* at MOI of 0.01 to 10 for 1 h. Next, after the extracellular *S. aureus* eliminated by treatment with 20 μg/ml gentamicin for 30 min, cells were cultured with fresh 10% FBS medium. After culture for 24 h, cells were washed with PBS three times, and stained with 0.04% trypan blue (C0040, Solarbio, Beijing, China) for 3 min. The plates were observed through a BX63 inverted microscope (Olympus, Tokyo, Japan) and cells were counted using Image Lab Software (Bio-Rad, CA, United States).

Macrophage Migration Assay

Macrophage migration was evaluated using wound-healing assay and transwell-based migration assay. For wound-healing assay, Raw264.7 were seeded at 1×10^6 cells/well in 6-well plate and cultured for 12 h, cells were then infected with *S. aureus* at 0.01 MOI. After 1 h infection, cells were treated with 20 μg/ml gentamicin for 30 min to kill the extracellular *S. aureus*. After being washed with PBS for three times, cells were allowed to grow in fresh 5% FBS medium, and a 200-μl micro pipette tip was used to scraped the macrophage monolayer. Wound-healing gap closure was analyzed and captured by Image Lab Software (Bio-Rad, CA, United States) at four time points (0, 24, 48, and 72 h). For transwell-based migration assays, 1×10^5 cells/well infected with *S. aureus* at 0.01 MOI for 1 h were seeded in the 8 μm upper chamber with serum-free medium, the lower chamber were supplied with 500 μl 10% FBS medium. After 24 h of culture, membranes were fixed in 4% paraformaldehyde and stained with the 1% crystal violet solution. The migration cells were counted using Image Lab Software.

Macrophage Phagocytosis

2×10^5 Raw 264.7 cells were infected with *S. aureus* at a MOI of 0.01 for 1 h, following by treatment with 20 μg/ml gentamicin for 30 min to kill extracellular bacteria. Cells were washed with PBS for three times, followed by lysis with 0.2% Triton. The cell lysis mixture was cultured on TSB agar plates overnight at 37°C. Bacteria colonies were counted and set as N0. To evaluate the phagocytosis of macrophage, after extracelluar bacteria eliminated, cells were allowed to grow in fresh 10% FBS medium for an additional 1 h. Then cells were lysed and cell lysis mixture was grown on TSB agar plates, and bacteria colonies were counted and set as N1. The rate of phagocytosis was calculated as $N0/(2 \times 10^5)$ (%), and the rate of bacterial killing was calculated as (N0-N1)/N0 (%).

Statistical Analysis

The results were expressed as mean ± S.E.M. The significance of variability was analyzed by two-tailed Student's *t* test, one-way analysis of variance (ANOVA) followed by Dunnett's test, or Mann–Whitney *U*-test. Quantification was performed from at least three independent experimental groups. $P < 0.05$ was

considered to be significant in all tests. All data were analyzed using SPSS 20 (IBM, NY, United States).

RESULTS

Analysis of DEGs From Microarray Data

GPL96 platform consists of 56 samples of PBMCs, with 10 from healthy people, 33 from osteomyelitis patients and 13 from osteomyelitis-free patients after *S. aureus* infection for 4–6 days (**Figure 1A**). GPL97 platform consists of 52 PBMCs samples with 10 from healthy people, 29 from osteomyelitis patients and 13 from osteomyelitis-free patients after *S. aureus* infection for 4–6 days. (**Figure 1B**). The GPL96 and GPL97 raw data were normalized (**Supplementary Figures S1A,B**) and converted into expression values by the robust multi-array average (RMA) algorithm (Irizarry et al., 2003).

As shown in **Figure 1C** and **Supplementary Figure S1C**, 57 OMRGs were considered to be differentially expressed among three comparison groups in GPL96 platform, and 25 OMRGs in GPL97 platform (**Figure 1D** and **Supplementary Figure S1D**). Specifically, there were 29 up-regulated, and 28 down-regulated OMRGs in GPL96 platform (**Figure 1E**) and 21 up-regulated and 4 down-regulated OMRGs in GPL97 platform (**Figure 1F**). Together, we got 82 unique OMRGs in GPL96 and GPL97 platforms, which were selected for further GO and KEGG pathways analysis.

GO Functional Annotation, KEGG Pathway and PPI Network Analysis of the DEGs

The biological function of the DEGs can be predicted by GO functional annotation analysis which depicts three biological functions including biological process, cellular component, and molecular function (Ashburner et al., 2000). GO analysis was performed using the online tool DAVID. The results showed biological process of the OMRGs enriched in positive regulation of transcription from RNA polymerase II promoter, bone mineralization, and regulation of immune response. The genes enriched at bone mineralization were S1PR1, TWIST1, and PHOSPHO1. For cellular component, OMRGs were assembled at membrane structure, such as bicellular tight junction, apical plasma membrane, and basolateral plasma membrane (**Figure 2A**). With regard to molecular function analysis, OMRGs were enriched in specific DNA binding, actin filament binding, and Erbb-3 class receptor binding (**Figure 2A**).

We performed DAVID to analyze the KEGG pathways of the OMRGs. Results showed that OMRGs were enriched in natural killer cell mediated cytotoxicity and antigen processing and presentation, especially pathways associated with the antigen processing and presentation (**Figures 2B,C**).

Protein-protein interaction networks were constructed using DEGs and protein interaction information obtained from STRING database. Results showed that there were 30 nodes and 30 edges in PPI network of OMRGs. Among this PPI network of OMRGs, erbb2 receptor tyrosine kinase 2 (ERBB2), twist family basic helix-loop-helix (bHLH) transcription factor 1 (TWIST1), Nanog homeobox (NANOG) were the top 3 core proteins according to the degree of distribution, closeness centrality, and betweenness centrality (**Figure 2D** and **Supplementary Table S2**).

TWIST1 Up-Regulation Might Be Associated With Macrophages Accumulation at Infection Site

As TWIST1 was the most up-regulated gene in OMRGs of GPL96 platform, NANOG and ERBB2 were closely associated with TWIST1, and they were top 3 core proteins as well in PPI network, we established *S. aureus* osteomyelitis model, and evaluated the mRNA expression of these three genes in peripheral blood cells and bone of *S. aureus*-infected mice. Results showed that TWIST1 and NANOG were significantly up-regulated in bone marrow cells (**Figure 3A**), but not in peripheral blood cells (**Figure 3B**), of *S. aureus* osteomyelitis mice.

TWIST1 was reported to be involved in tumor cells migration (Xu et al., 2017; Gou et al., 2018). As macrophages and neutrophils are the main phagocytes to defend *S. aureus* infection in early infection stage (Miller and Cho, 2011), we speculate that *S. aureus* infection may stimulate cells migration by up-regulating the expression of TWIST1. As shown in H&E staining results, there were a lot of macrophages accumulated around the infected implant on day 3 after infection (**Figure 3C**). Consistent with the histochemistry results, immunohistochemistry staining for F4/80, a marker for macrophage, also showed significantly increased positive staining (**Figures 3D,E**).

To further evaluate the changes of macrophage at infection site in bone marrow, the mRNA expression levels of CD86, IL-1β, IL-6, and COX2, marker genes of the classically activated macrophages (M1), and ARG1, IDO1, YM1, and CCR7, marker genes of the alternatively activated macrophages (M2; Parisi et al., 2018), were evaluated in the bone marrow from *S. aureus* osteomyelitis mice. Results showed that the mRNA expression levels of IL-1β, IL-6, and COX2 were dramatically up-regulated, while YM1 expression levels were down-regulated (**Figures 3F,G**). All these data indicate that the up-regulated expression of TWIST1 might be associated with macrophages accumulation at infection site.

TWIST1 Is Activated in Macrophages Around the Infection Site

To confirm the activated expression of TWIST1 in *S. aureus* osteomyelitis, protein was harvested from mice bone marrow at day 3 after infection. Consistent with the up-regulated expression level in mRNA, the protein level of TWIST1 was significantly increased in bone marrow infected by *S. aureus* (**Figures 4A,B**). To further verify this finding, we performed immunohistochemistry staining. Results showed a much higher amount of TWIST1-positive staining around the infection site than that in the control group (**Figures 4C,D**).

As macrophages were recruited around the infection site and TWIST1 was highly expressed around the infection, we speculated that the expression of TWIST1 might be up-regulated

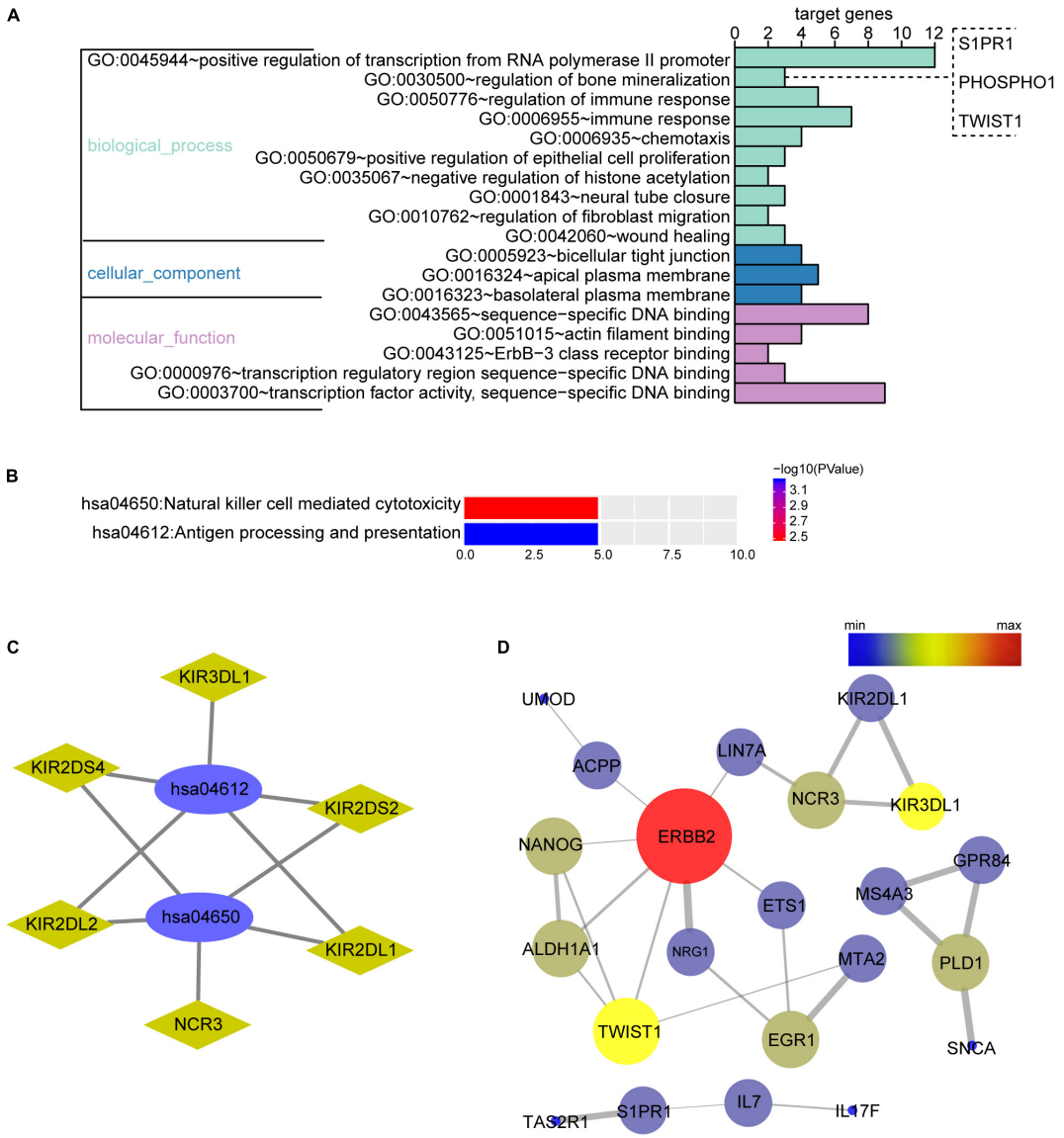


FIGURE 2 | Biological function enrichment analysis of the OMRGs. **(A)** Bar chart of gene ontology (GO) analysis. Genes associated with biological process, cellular component, and molecular function were analyzed for OMRGs. The horizontal axis shows the number of the target genes, and the vertical axis represents GO ID and annotation. **(B)** Bar chart of the KEGG analysis for OMRGs. The horizontal axis shows the number of the target genes, the right vertical axis -log10 (*P* value), and the left vertical axis KEGG pathway ID and annotation. **(C)** Plots show the relationship between the OMRGs and the KEGG pathways. The ellipse plots show the KEGG pathway ID, the rhombus plots the target gene, and the green plots down-regulated genes. **(D)** PPI network analysis of the OMRGs. The nodes represent the OMRGs in the PPI network, and the lines show the interaction among the OMRGs. The size and color of nodes are associated with the significance degree of proteins, while the thickness of lines is proportional to the strength of interactions among proteins.

in macrophages around infection. Double-immunofluorescence of F4/80 and TWIST1 results demonstrated dramatically increased intensity of F4/80^{+}TWIST1^{+} staining (**Figures 4E,F**).

TWIST1 May Be Associated With Macrophage Polarization in Response to *S. aureus* Infection

To verify the effect of *S. aureus* infection on the expression of TWIST1 in macrophages, Raw 264.7 cells were infected with *S. aureus* at different MOIs [the concentration of *S. aureus* (CFU) to Raw 264.7 cells, MOI = 10, 1, 0.1, 0.01]. In accordance with the result *in vivo*, we found significantly up-regulated expression of TWIST1. The mRNA level and protein level of TWIST1 decreased gradually with the severity of infection (**Figure 5A** and **Supplementary Figures S2A,B**). We believed that this phenomenon might be associated with increased cell mortality due to high MOI of infection (**Supplementary Figures S2C,D**). Additionally, western blotting results confirmed the up-regulated expression of

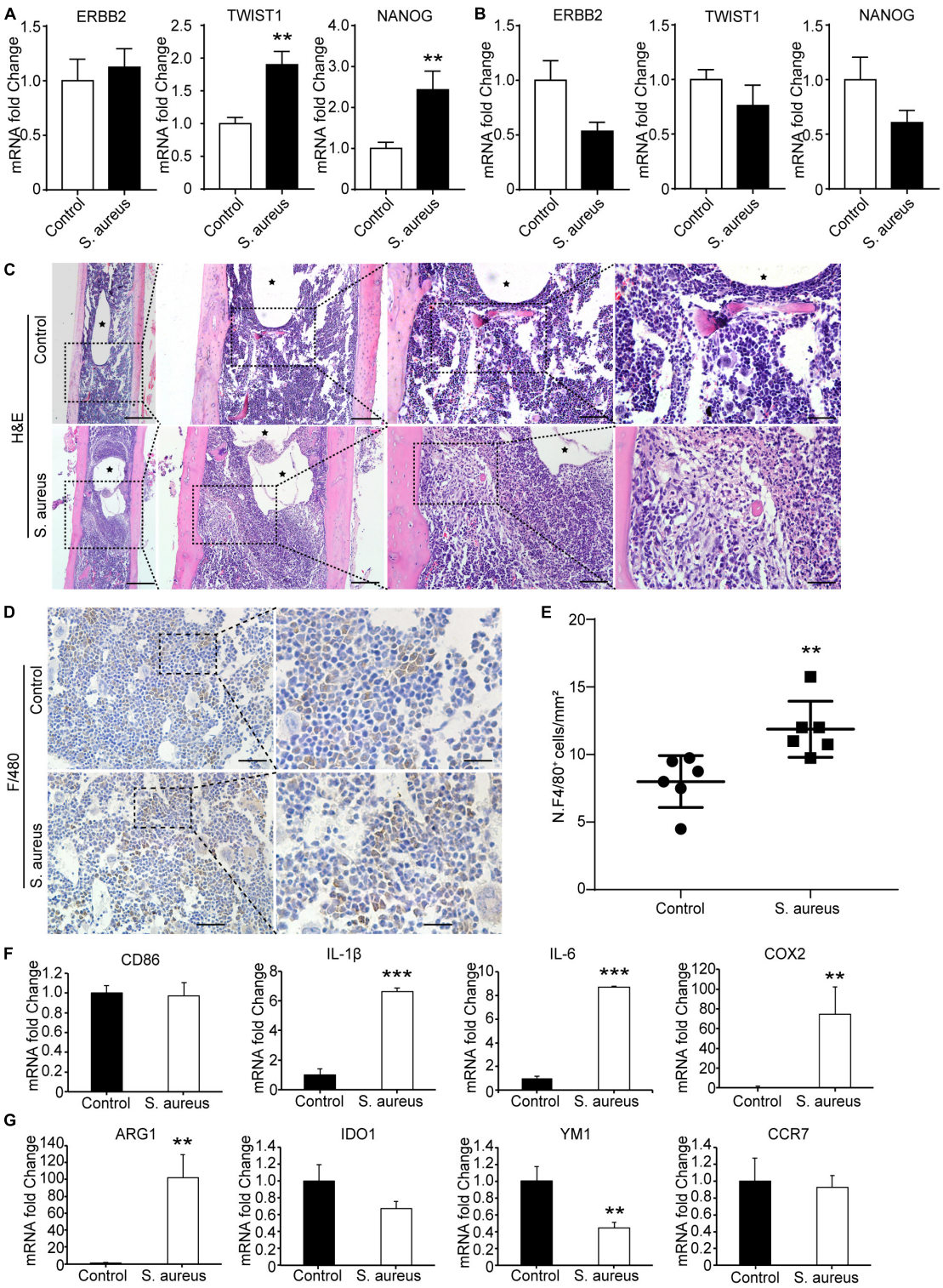


FIGURE 3 | Macrophages accumulate around the infected implant in bone marrow of osteomyelitis mice. **(A,B)** qPCR analysis of the mRNA expression of TWSIT1, NONAG, and ERBB2 in bone marrow cells **(A)** and in peripheral blood cells **(B)** from *S. aureus* osteomyelitis mice and control group. $n = 6$/group. **(C)** H&E staining for the bone from mice with *S. aureus* infected implant and control mice. Scale bars form left to right represent 500, 200, 100, and 20 μm, respectively. Black stars represent implant track. $n = 6$/group. **(D)** Representative images of immunohistochemistry staining for F4/80 in the bone of *S. aureus* infected mice and control mice. **(E)** Quantitative analysis of the number of $F4/80^+$ cells around an implant (*N.* $F4/80^+/mm^2$). Scale bars form left to right represent 100 and 20 μm, respectively. $N = 5$/group, $^{**}p < 0.01$. **(F,G)** qPCR analysis of the mRNA expression of marker genes associated with macrophage polarization in bone marrow cells. $n = 6$/group. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.0001$, and Mann–Whitney *U*-test.

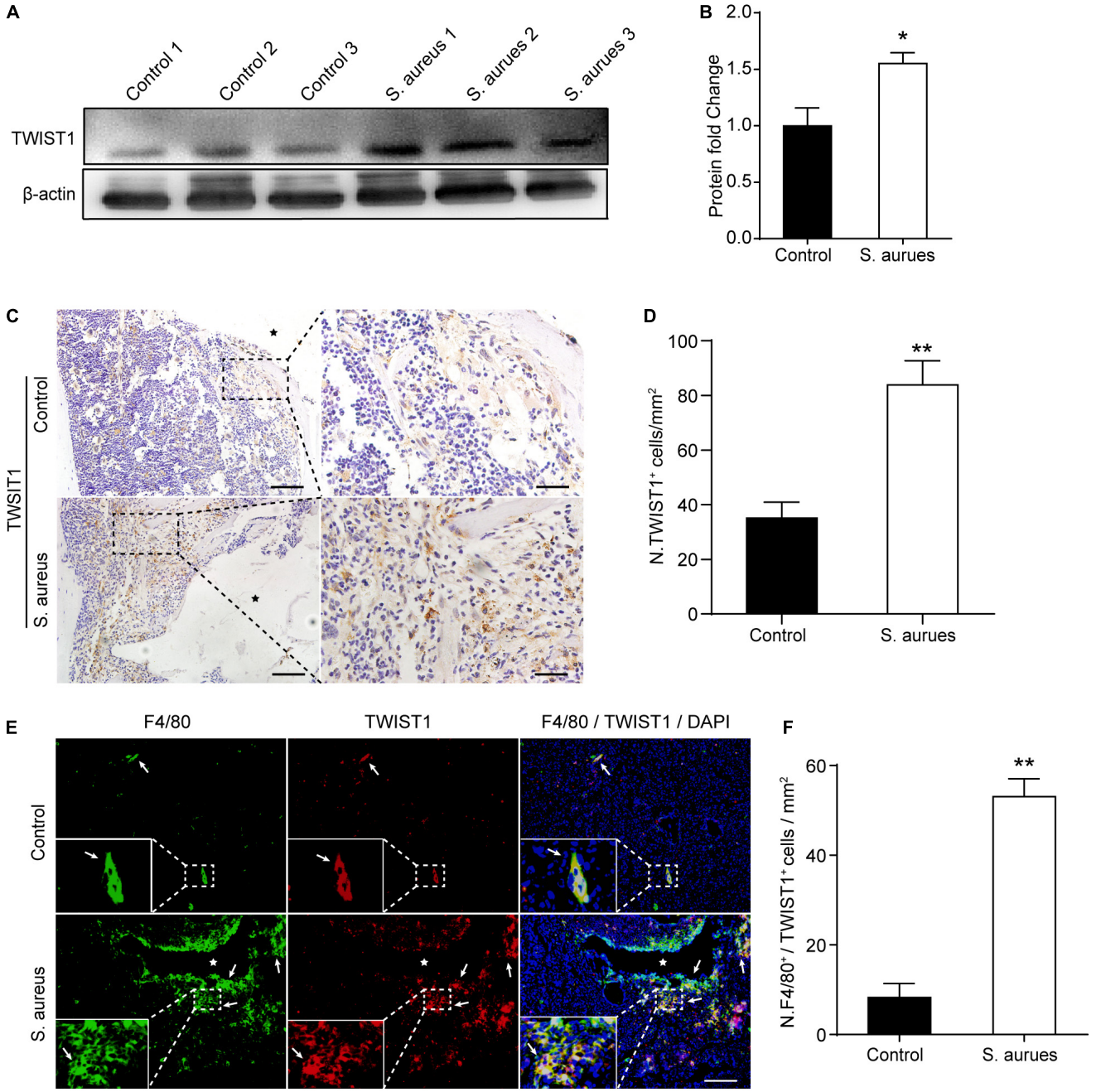


FIGURE 4 | *S. aureus* infection activates the expression of TWIST1 in macrophages. **(A,B)** Western blotting analysis **(A)** and quantification **(B)** of TWIST1 expression in the bone marrow from *S. aureus* infected bone and control bone. n = 6/group. **(C)** Representative images of immunohistochemistry staining for TWIST1 in bone from *S. aureus* infected mice and control mice. Scale bars form left to right represents 100 μm and 20 μm, respectively, and stars represent the implant track. **(D)** Quantitative analysis of the number of TWIST1$^+$ cells per tissue area around implant (*N*. TWIST1$^+$/mm^2). n = 6/group. **(E)** Representative images of double-immunofluorescence staining for TWIST1 and F4/80 in bone from *S. aureus* infected mice and control mice. Scale bar represents 100 μm. Stars locate the implant track white arrows point to positive stained cells. **(F)** Quantitative analysis of the number of F4/80$^+$TWIST1$^+$ cells per tissue area around implant (N. F4/80$^+$TWIST1$^+$/mm^2). n = 4/group. $^*p < 0.05$, $^{**}p < 0.01$, and Mann–Whitney *U*-test.

TWIST1 in macrophages after infection with *S. aureus* at 0.01 MOI (**Figures 5B,C**). To explore the role of TWIST1 in Raw 264.7 cells after infection, transient RNA interference (Si-TWIST1) was performed to screen out the best interference effect in mRNA level (**Figure 5D**) and protein level (**Figures 5E,F**). The marker genes for M1 macrophage were dramatically up-regulated in Raw 264.7 macrophage cells *S. aureus* infection (**Figure 5G**). Further,

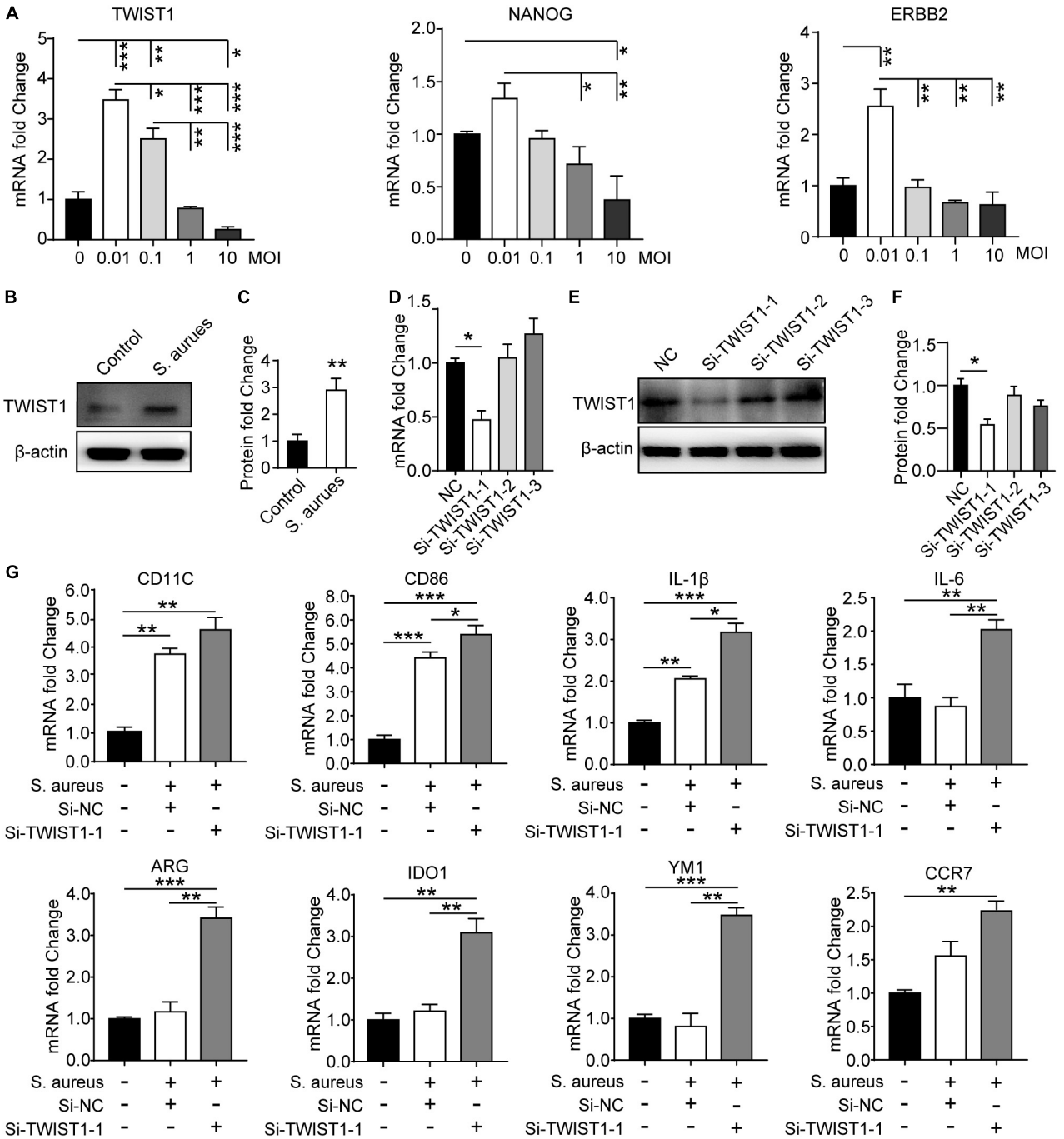


FIGURE 5 | *S. aureus* infection activates the expression of TWIST1 in Raw 264.7 cells. **(A)** qPCR analysis of the mRNA expression of TWIAT1, NANOG, and ERBB2 in Raw 264.7 cells in response to *S. aureus* infection with MOI [the ratio of the *S. aureus* colonies number (CFU) to the Raw 264.7 cells number] at 0.01, 0.1, 1.0, and 10. ANOVA followed by Dunnett's test, $n = 3$/group. **(B)** Immunoblot and **(C)** quantitative analysis of TWIST1 in Raw 264.7 cells after infection. Two-tailed Student's t test, $n = 3$/group. **(D)** qPCR analysis of the effect of siRNA fragments on the mRNA expression of TWIST1 in Raw 264.7 cells. ANOVA followed by Dunnett's test, $n = 3$/group. **(E)** Immunoblot and **(F)** quantitative analysis of the effect of siRNA fragments on the protein expression of TWIST1 in Raw 264.7 cells. ANOVA followed by Dunnett's test, $n = 3$/group. **(G)** qPCR analysis of the effect of knocking down TWIST1 on the mRNA expression of marker genes for macrophage polarization in Raw 264.7 cells after *S. aureus* infection. ANOVA followed by Dunnett's test, $n = 3$/group. $^*p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

the mRNA expression levels of marker genes associated with macrophage polarization were considerably changed after knocking down TWIST1. Results showed that the mRNA expression levels of ARG, IDO1 and YM1, marker genes for M2 macrophages, were dramatically up-regulated (**Figure 5G**). The data indicate that TWIST1 might be

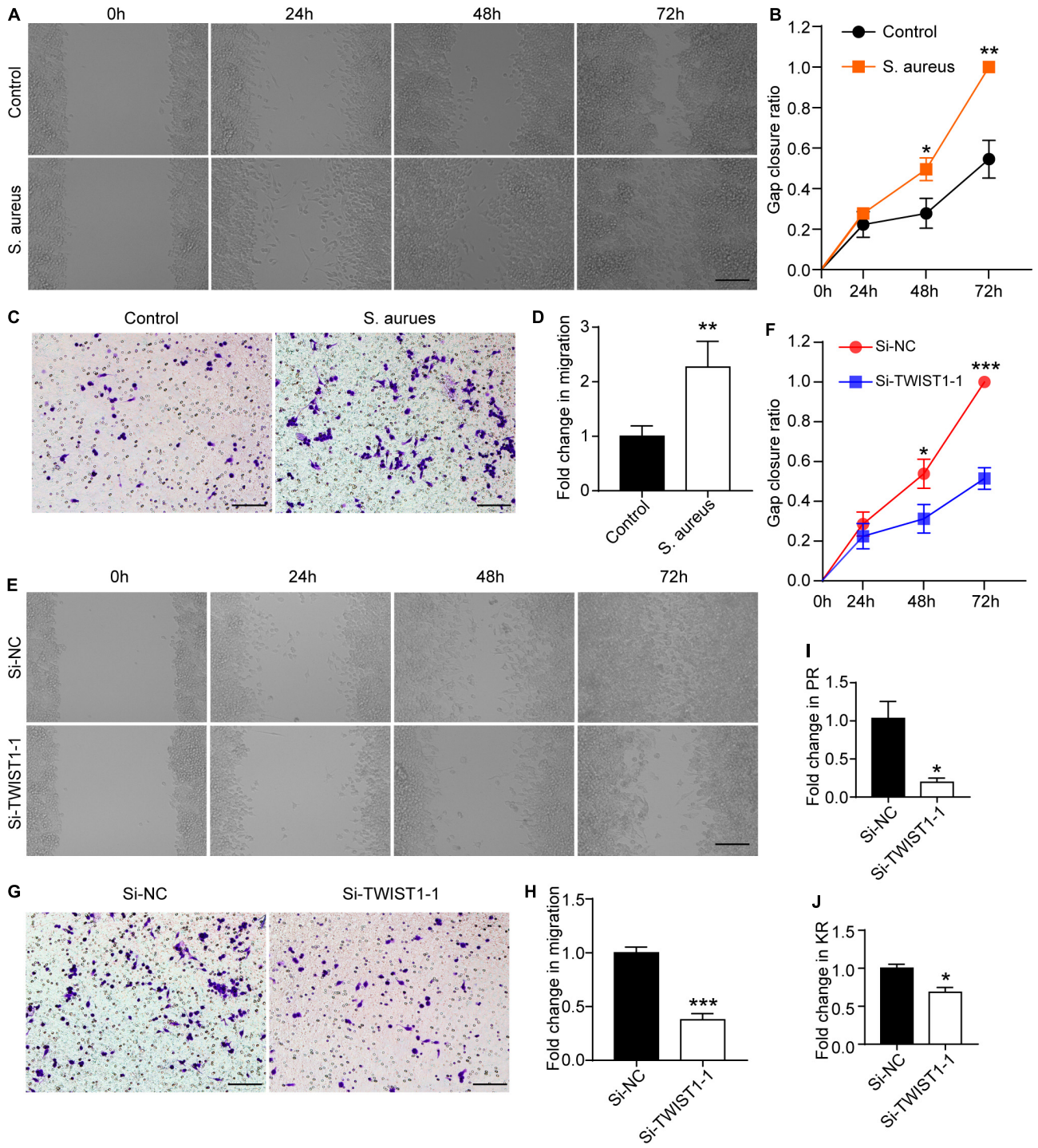


FIGURE 6 | TWIST1 mediates the migration and phagocytosis of macrophage in response to *S. aureus* infection. Wound-healing assay was used to evaluate the effect of *S. aureus* on macrophage migration. After 1×10^6 cells were seeded in 6-well plate and cultured for 12 h, they were infected with *S. aureus* at 0.01 MOI. **(A)** Representative images for wound-healing assay and **(B)** quantitative analysis of gap closure ratio of Raw 264.7 cells at 0, 24, 48, and 72 h after *S. aureus* infection. Scale bar represents 100 μm. **(C)** Transwell-based migration assays and **(D)** quantitative analysis of the migrated cells of Raw 264.7 cells at 24 h after *S. aureus* infection at 0.01 MOI. Scale bar represents 100 μm. **(E)** Wound-healing assay and **(F)** quantitative analysis of gap closure ratio of Raw 264.7 cells at 0, 24, 48, and 72 h after *S. aureus* infection. Scale bar represents 100 μm. NC represents negative control. Cells were transfected with negative random fragments. **(G)** Transwell-based migration assays and **(H)** quantitative analysis of the migrated cells of Raw 264.7 cells at 24 h after *S. aureus* infection. Scale bar represents 100 μm. Quantitative analysis of the effect of knocking down TWIST1 on Phagocytosis rate (PR; **I**) and killing rate (KR; **J**) of Raw 264.7 cells in response to *S. aureus* infection. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and Two-tailed Student's *t* test, $n = 3$/group.

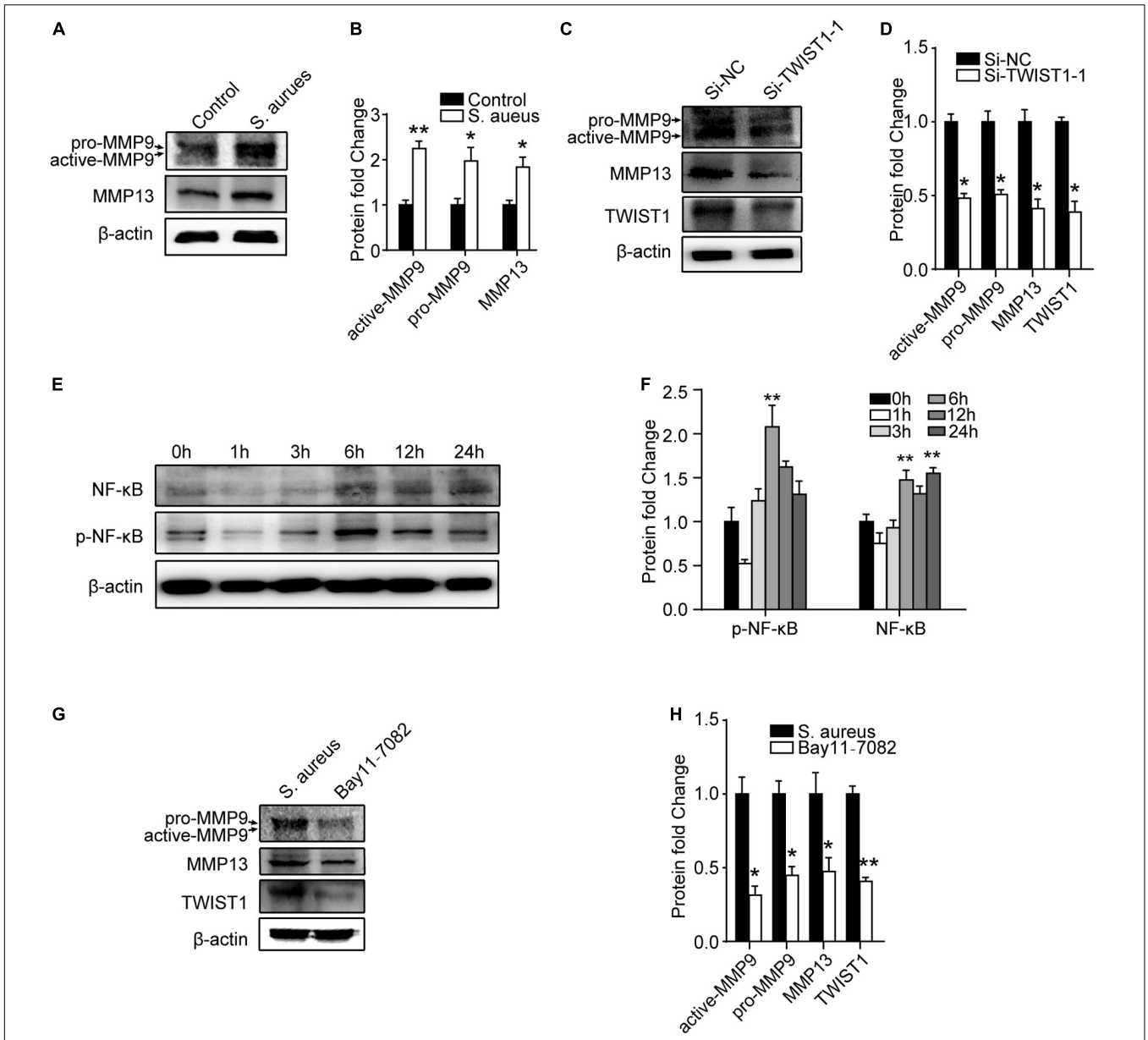


FIGURE 7 | NF-κB/TWIST1 mediates the effect of *S. aureus* on macrophages. **(A)** Immunoblot and **(B)** quantitative analysis of MMP9 and MMP13 expression in Raw 264.7 cells in response to *S. aureus* infection. Two-tailed Student's *t* test, $n = 3$/group. **(C)** Immunoblot and **(D)** quantitative analysis of the effect of knocking down TWIST1 on the expression of MMP9 and MMP13 in Raw 264.7 cells after *S. aureus* infection. Two-tailed Student's *t* test, $n = 3$/group. **(E)** Immunoblot and quantitative analysis **(F)** of NF-κB in Raw 264.7 cells at 0, 1, 3, 6, 12, and 24 h after *S. aureus* infection. ANOVA followed by Dunnett's test, $n = 3$/group. **(G)** Immunoblot and quantitative analysis **(H)** of the effect of NF-κB inhibitor (Bay11-7082) on the expression of MMP9 and MMP13 in Raw 264.7 cells after *S. aureus* infection. Two-tailed Student's *t* test, $n = 3$/group. $^{*}p < 0.05$, and $^{**}p < 0.01$.

associated with macrophage polarization in response to *S. aureus* infection.

TWIST1 May Regulate the Migration and Phagocytosis of Macrophage

As there were more macrophages at the infection site, we speculated that TWIST1 might be associated with the migration and recruitment of macrophages to the infection sites in marrow. To prove this speculation, the wound-healing assay and transwell-based migration assays were performed. Interestingly, the Raw 264.7 cells infected by *S. aureus* migrated faster than control cells treated with vehicle (**Figures 6A–D**), and the down-regulation of TWIST1 inhibited the migration of Raw 264.7 cells in response to *S. aureus* (**Figures 6E,H**). In addition, knocking-down TWIST1 significantly reduced the phagocytosis rate and killing rate of Raw 264.7 in response to *S. aureus* infection (**Figures 6I,J**).

S. aureus Promotes Macrophage Migration Through P65/TWIST1/MMP9/MMP13 Axis

It was reported that TWIST1 promoted extracellular matrix degradation and cells migration through upregulating matrix metallopeptidases especially MMP9 and MMP13 (Ding et al., 2019). We speculated that *S. aureus* infection might stimulate TWIST1 expression, thereby up-regulating MMP9 and MMP13 expression. Interestingly, we found that MMP9 (active-MMP9 and pro-MMP9), and MMP13 in Raw 264.7 cells were significantly up-regulated in response to *S. aureus* infection (**Figures 7A,B**). Furthermore, knocking-down the TWIST1 dramatically suppressed the expression of MMP9 and MMP13 (**Figures 7C,D**).

As NF-κB (P65) was reported to be associated with the expression of TWIST1 (Li C. W. et al., 2012; Li S. et al., 2012; Liu et al., 2015), we evaluated the expression of NF-kB after *S. aureus* infection. Activation of NF-kB was observed in a time-dependent manner 6 h after *S. aureus* infection (**Figures 7E,F**). To further evaluate the role of NF-kB in *S. aureus* induced expression of TWIST1/MMP9/13, P65 signaling inhibitor, Bay 11–7082 (10 μM, pretreatment 6 h) was used to block NF-κB when infected by *S. aureus*. Results showed that Bay11-7082 significantly suppressed the protein levels of TWIST1, MMP9, and MMP13 after *S. aureus* infection (**Figures 7G,H**).

DISCUSSION

Staphylococcus aureus is a leading cause for bloodstream infection-induced osteomyelitis in children (Street et al., 2015). Understanding the pathogenesis of early host response is important in prevention and elimination of *S. aureus* osteomyelitis. In the present study, bioinformatics was used to investigate the system-level responses of PBMCs to *S. aureus* osteomyelitis. ERBB2, TWIST1, and NANOG were screened out as the most important OMRGs. We found dramatically up-regulated expression of TWIST1 and macrophage recruitment at the infection site of implant-related osteomyelitis mice model. Further, we demonstrated that *S. aureus* stimulated polarization, migration and phagocytosis of macrophage by up-regulating NF-κB/TWIST1 signaling.

Gene ontology analysis revealed that biological process of the OMRGs was enriched in the genes enriched at bone mineralization, like down-regulated S1PR1, and up-regulated TWIST1 and PHOSPHO1. Intriguingly, these genes have been shown to target bone formation and bone matrix mineralization (Isenmann et al., 2009; Dillon et al., 2019; Lewis et al., 2019; Xiao et al., 2019). It is likely that there are multiple genes involved in infection-induced changes in bone remodeling, but the relative importance of these distinct genes in bone remodeling during infection remains to be determined.

Protein-protein interaction network analysis of OMRGs revealed changes of 3 core proteins, down-regulated ERBB2, up-regulated TWIST1, and NANOG. TWIST1 is a bHLH transcription factor, over-expression of which stimulates proliferation of bone marrow mononuclear cells, aggravates the inflammatory balance (Wang et al., 2015; Mahadik et al., 2018), and is related to cells migration (Xu et al., 2017; Gou et al., 2018; Zhou et al., 2019). NANOG is a key transcription factor for pluripotency of $CD14^{+}$ monocytes, involved in immune response, repair and regeneration of damaged tissue (Seta and Kuwana, 2010). ERBB2, the second ErbB-family member, is ubiquitously expressed and commonly altered in metastatic breast, prostate and ovarian cancers (Arteaga and Engelman, 2014). A recent study reveals a critical function of ERBB2 in innate immune modulation. ERBB2 inhibitor or ablation of ERBB2 expression leads to a dramatic suppression of HSV-1 infection (Wu et al., 2019). Therefore, up-regulation of TWIST1 and NANOG and down-regulation of ERBB2 might indicate activation of immune cells in response to *S. aureus* osteomyelitis.

It is somewhat unexpected that the up-regulated expression of TWIST1 and NANOG was only found in the cells from bone marrow but not in those from blood and the expression of ERBB2 was only down-regulated in the cells from blood in implant-associated osteomyelitis mice models. This might have been because the implant-associated infection model we used to observe the genes expression was only at an early stage of *S. aureus* infection when the infection had not spread to elsewhere. It will be interesting to determine in a future study the changes in expression of these genes at different infection stages in mice models of hematogenous and implant-associated osteomyelitis. In addition, the healthy controls and osteomyelitis patients were at a median age of 6–7.5 years old in the PBMCs samples for the transcriptional data whereas the implant-associated infection mice were 8 weeks old in the OMRGs verification, which is equivalent to about 20 years of age in human (Dutta and Sengupta, 2016). Therefore, age difference might have been another major cause leading to the above discrepancy.

Macrophages are actively involved in clearance of bacteria, thereby providing the first line of defense against invading pathogens (Motwani and Gilroy, 2015). Our data provided evidence that the amount of macrophage was significantly increased around *S. aureus*-infected implant. Additionally, we found up-regulated expression of TWIST1 in macrophage around *S. aureus*-infected implant. TWIST1 has been found to be involved in promoting cell migration through up-regulating the expression of MMPs in cancer cells (Xu et al., 2017; Gou et al., 2018; Zhou et al., 2019). However, little is known about the role of TWIST1 in regulating macrophage function under *S. aureus* infection. Our data reveal that *S. aureus* infection may stimulate the expression of TWIST1 which up-regulates the expression of MMP9 and MMP 13, thereby promoting the migration of macrophages toward the infection sites for scavenging *S. aureus*.

It is known that the balance between classically activated (M1) macrophages and alternatively activated (M2) macrophages governs the fate of an organ in infection. M1 macrophages release cytokines which can cause tissue damage while M2 macrophages secret factors to suppress the inflammation and promote tissue remodeling (Shapouri-Moghaddam et al., 2018).

NF-κB is considered to be a member of pro-inflammatory family of main transcription factors in inflammatory processes. It participates in M1 macrophage polarization induced by various stimulations (Gao et al., 2018; Genard et al., 2018). Consistent with their reports, our data demonstrate that NF-κB pathway activation is associated with TWIST1 expression and M1 polarization in response to *S. aureus* infection. Our data suggest that *S. aureus* may activate TWIST1 expression through NF-κB, thereby promoting macrophage migration and M1 polarization. Of note, TWIST1 may limit inflammation through a negative feedback loop that represses NF-κB activity (Šošić et al., 2003; Merindol et al., 2014). It is possible that TWIST1 and NF-κB may form a negative regulatory loop in modulating macrophage reaction during *S. aureus* infection.

This study has 2 limitations. First, we did not further investigate the role of up-regulated NANOG and down-regulated ERBB2 in *S. aureus* infection. Additional further studies are warranted to determine in which cells NANOG and ERBB2 are regulated and their roles in *S. aureus* infection. Secondly, we only evaluated the mRNA expression of 3 core genes obtained from PPI network analysis of OMRGs. It will be important in the future studies to verify more OMRGs and their roles in animal osteomyelitis models.

In conclusion, we identified the DEGs of osteomyelitis caused by *S. aureus* infection based on publicly available transcriptional dataset GSE16129 using bioinformatic analysis. Our *in vivo* and *in vitro* data demonstrate that NF-κB/TWIST1 is necessary for *S. aureus* infection-induced macrophage migration and phagocytosis. Our study highlights a previously unknown fundamental function of TWIST1 in the early stages of the innate immune response to *S. aureus* infection in bone.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/gds/?term=gse16129), GSE16129 (Ardura et al., 2009).

ETHICS STATEMENT

The publicly available GSE16129 dataset involving human participants were reviewed and approved by Institutional Review Boards of the University of Texas Southwestern Medical Center and Children's Medical Center of Dallas (IRB #0802-447) and Baylor Institute of Immunology Research (BIIR, IRB #002-141). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin (Ardura et al., 2009). The animal study was reviewed and approved by Animal experiment ethics committee of Nanfang hospital.

AUTHOR CONTRIBUTIONS

XZ and BY conceived and designed the study. YW performed the bioinformatics analyses and drafted the manuscript and carried out the experiment. YL and CC established a model of osteomyelitis. YW and PZ performed *in vitro* experiments. PC and HW drafted sections of the manuscript. PZ, KL, YD, and JQ composed the figures in this manuscript. XZ and BY revised and approved the manuscript.

FUNDING

This work was supported by National Natural Science Foundation of China (No. 81772366), Guangdong Provincial Science and Technology Plan Projects (No. 2016B090913004), and Academy and Research Foundation of Guangdong Province (No. 2013B090600140), and Science and Technology Project of Zhanjiang (grant number: 2018A01033).

ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr. Ardura, Banchereau and their team for uploading the datasets, related with our study, to the GEO public database. The authors thank Professor Allen P. Liang (Nanfang Hospital, Southern Medical University) for English proofreading of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01301/full#supplementary-material

FIGURE S1 | Sample data preprocessing. **(A,B)** Box plots of the expression value before and after normalization of GPL 96 and GPL97. Blue plots represent the unnormalized expression value, the red pots the normalized expression value, the horizontal axis the samples, and the vertical axis the expression value. **(C,D)** Volcano plots for DEGs of platform GPL96 and GPL97. Plots of the relationship between -log10 (adjusting *p* value) and log2 (fold change) for Ctrl vs OFI, Ctrl vs OMI, and Ctrl vs SI. Red plots represent the up-regulated or down-regulated genes that log2 (fold change) >1 or <-1, also -log10 (adjusted *P*-value) >1.301.

FIGURE S2 | To determine the optimal MOI for *S. aureus* infection of Raw264.7 macrophages **(A)** Immunoblot and quantitative analysis **(B)** of TWIST1 in Raw 264.7 cells after *S. aureus* infection in different MOIs. **(C)** Trypan blue staining of Raw 264.7 cells after *S. aureus* infection for 24 h in different MOIs. **(D)** Quantitative analysis of viability (%Ctrl) of Raw 264.7 cells after *S. aureus* infection for 24 h in different MOIs. ANOVA followed by Dunnett's test, $n = 3$/group. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

REFERENCES

Ardura, M. I., Banchereau, R., Mejias, A., Di Pucchio, T., Glaser, C., Allantaz, F., et al. (2009). Enhanced monocyte response and decreased central memory T cells in children with invasive *Staphylococcus aureus* infections. *PLoS One* 4:e5446. doi: 10.1371/journal.pone.000544

Arteaga, C. L., and Engelman, J. A. (2014). ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* 25, 282–303. doi: 10.1016/j.ccr.2014.02.025

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bader, G. D., and Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4:2. doi: 10.1186/1471-2105-4-2

Bernthal, N. M., Stavrakis, A. I., Billi, F., Cho, J. S., Kremen, T. J., Simon, S. I., et al. (2010). A mouse model of post-arthroplasty *Staphylococcus aureus* joint infection to evaluate in vivo the efficacy of antimicrobial implant coatings. *PLoS One* 5:e12580. doi: 10.1371/journal.pone.0012580

Bogoslowski, A., Butcher, E. C., and Kubes, P. (2018). Neutrophils recruited through high endothelial venules of the lymph nodes via PNAd intercept disseminating *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U.S.A.* 115, 2449–2454. doi: 10.1073/pnas.1715756115

de Vries, T. J., El Bakkali, I., Kamradt, T., Schett, G., Jansen, I. D. C., and D'Amelio, P. (2019). What are the peripheral blood determinants for increased osteoclast formation in the various inflammatory diseases associated with bone loss? *Front. Immunol.* 10:505. doi: 10.3389/fimmu.2019.00505

Diboun, I., Wernisch, L., Orengo, C. A., and Koltzenburg, M. (2006). Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. *BMC Genomics* 7:252. doi: 10.1186/1471-2164-7-252

Dillon, S., Staines, K. A., Millán, J. L., and Farquharson, C. (2019). How to build a bone: PHOSPHO1, biomineralization, and beyond. *JBMR Plus* 3:e10202. doi: 10.1002/jbm4.10202

Ding, X., Li, F., and Zhang, L. (2019). Knockdown of Delta-like 3 restricts lipopolysaccharide-induced inflammation, migration and invasion of A2058 melanoma cells via blocking Twist1-mediated epithelial-mesenchymal transition. *Life Sci.* 226, 149–155. doi: 10.1016/j.lfs.2019.04.024

Dutta, S., and Sengupta, P. (2016). Men and mice: relating their ages. *Life Sci.* 152, 244–248. doi: 10.1016/j.lfs.2015.10.025

Fink, C. W., Dich, V. Q., Howard, J. Jr., and Nelson, J. D. (1977). Infections of bones and joints in children. *Arthritis Rheum.* 20(2 Suppl.), 578–583.

Fischer, K. J., Yajjala, V. K., Bansal, S., Bauer, C., Chen, R., and Sun, K. (2019). Monocytes represent one source of bacterial shielding from antibiotics following influenza virus infection. *J. Immunol.* 202, 2027–2034. doi: 10.4049/jimmunol.1801471

Gao, X. R., Ge, J., Li, W. Y., Zhou, W. C., Xu, L., and Geng, D. Q. (2018). NF-kappaB/let-7f-5p/IL-10 pathway involves in wear particle-induced osteolysis by inducing M1 macrophage polarization. *Cell Cycle* 17, 2134–2145. doi: 10.1080/15384101.2018.1515549

Genard, G., Wera, A. C., Huart, C., Le Calve, B., Penninckx, S., Fattaccioli, A., et al. (2018). Proton irradiation orchestrates macrophage reprogramming through NFkappaB signaling. *Cell Death Dis.* 9:728. doi: 10.1038/s41419-018-0757-9

Gou, W., Zhou, X., Liu, Z., Wang, L., Shen, J., Xu, X., et al. (2018). CD74-ROS1 G2032R mutation transcriptionally up-regulates Twist1 in non-small cell lung cancer cells leading to increased migration, invasion, and resistance to crizotinib. *Cancer Lett.* 422, 19–28. doi: 10.1016/j.canlet.2018.02.032

Hou, Y., Qin, H., Jiang, N., Liu, G., Wu, H., Bai, L., et al. (2019). G-CSF partially mediates bone loss induced by *Staphylococcus aureus* infection in mice. *Clin. Sci. (Lond.)* 133, 1297–1308. doi: 10.1042/CS20181001

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249–264. doi: 10.1093/biostatistics/4.2.249

Isenmann, S., Arthur, A., Zannettino, A. C., Turner, J. L., Shi, S., Glackin, C. A., et al. (2009). TWIST family of basic helix-loop-helix transcription factors mediate human mesenchymal stem cell growth and commitment. *Stem Cells* 27, 2457–2468.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36(Database issue), D480–D484. doi: 10.1093/nar/gkm882

Kavanagh, N., Ryan, E. J., Widaa, A., Sexton, G., Fennell, J., O'Rourke, S., et al. (2018). Staphylococcal osteomyelitis: disease progression, treatment challenges, and future directions. *Clin. Microbiol. Rev.* 31:e00084-17.

Knobloch, J., Panek, S., Yanik, S. D., Jamal Jameel, K., Bendella, Z., Jungck, D., et al. (2019). The monocyte-dependent immune response to bacteria is suppressed in smoking-induced COPD. *J. Mol. Med. (Berl.)* 97, 817–828. doi: 10.1007/s00109-019-01778-w

Kulasingam, V., and Diamandis, E. P. (2008). Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat. Clin. Pract. Oncol.* 5, 588–599. doi: 10.1038/ncponc1187

Lewis, K. J., Choi, R. B., Pemberton, E. Z., Bullock, W. A., Firulli, A. B., and Robling, A. G. (2019). Twist1 inactivation in dmp1-expressing cells increases bone mass but does not affect the Anabolic response to sclerostin neutralization. *Int. J. Mol. Sci.* 20:4427.

Li, C. W., Xia, W., Huo, L., Lim, S. O., Wu, Y., Hsu, J. L., et al. (2012). Epithelial-mesenchymal transition induced by TNF-alpha requires NF-kappaB-mediated transcriptional upregulation of Twist1. *Cancer Res.* 72, 1290–1300. doi: 10.1158/0008-5472.CAN-11-3123

Li, S., Kendall, S. E., Raices, R., Finlay, J., Covarrubias, M., Liu, Z., et al. (2012). TWIST1 associates with NF-kappaB subunit RELA via carboxyl-terminal WR domain to promote cell autonomous invasion through IL8 production. *BMC Biol.* 10:73. doi: 10.1186/1741-7007-10-73

Liu, Y., Mayo, M. W., Xiao, A., Hall, E. H., Amin, E. B., Kadota, K., et al. (2015). Loss of BRMS1 promotes a mesenchymal phenotype through NF-kappaB-dependent regulation of Twist1. *Mol. Cell. Biol.* 35, 303–317. doi: 10.1128/MCB.00869-14

Mahadik, K., Prakhar, P., Rajmani, R. S., Singh, A., and Balaji, K. N. (2018). c-Abl-TWIST1 epigenetically dysregulate inflammatory responses during mycobacterial infection by co-regulating bone morphogenesis protein and miR27a. *Front. Immunol.* 9:85. doi: 10.3389/fimmu.2018.00085

Mbalaviele, G., Novack, D. V., Schett, G., and Teitelbaum, S. L. (2017). Inflammatory osteolysis: a conspiracy against bone. *J. Clin. Invest.* 127, 2030–2039.

Merindol, N., Riquet, A., Szablewski, V., Eliaou, J. F., Puisieux, A., and Bonnefoy, N. (2014). The emerging role of Twist proteins in hematopoietic cells and hematological malignancies. *Blood Cancer J.* 4:e206. doi: 10.1038/bcj.2014.22

Miller, L. S., and Cho, J. S. (2011). Immunity against *Staphylococcus aureus* cutaneous infections. *Nat. Rev. Immunol.* 11, 505–518. doi: 10.1038/nri3010

Momodu, I. I., and Savaliya, V. (2019). *Osteomyelitis*. Treasure Island, FL: StatPearls Publishing.

Motwani, M. P., and Gilroy, D. W. (2015). Macrophage development and polarization in chronic inflammation. *Semin. Immunol.* 27, 257–266. doi: 10.1016/j.smim.2015.07.002

Parisi, L., Gini, E., Baci, D., Tremolati, M., Fanuli, M., Bassani, B., et al. (2018). Macrophage polarization in chronic inflammatory diseases: killers or builders? *J. Immunol. Res.* 2018:8917804. doi: 10.1155/2018/8917804

Peltola, H., and Paakkonen, M. (2014). Acute osteomyelitis in children. *N. Engl. J. Med.* 370, 352–360. doi: 10.1056/NEJMra1213956

Putnam, N. E., Fulbright, L. E., Curry, J. M., Ford, C. A., Petronglo, J. R., Hendrix, A. S., et al. (2019). MyD88 and IL-1R signaling drive antibacterial immunity and osteoclast-driven bone loss during *Staphylococcus aureus* osteomyelitis. *PLoS Pathog.* 15:e1007744. doi: 10.1371/journal.ppat.1007744

Riise, O. R., Kirkhus, E., Handeland, K. S., Flato, B., Reiseter, T., Cvancarova, M., et al. (2008). Childhood osteomyelitis-incidence and differentiation from other acute onset musculoskeletal features in a population-based study. *BMC Pediatr.* 8:45. doi: 10.1186/1471-2431-8-45

Seta, N., and Kuwana, M. (2010). Derivation of multipotent progenitors from human circulating CD14+ monocytes. *Exp. Hematol.* 38, 557–563. doi: 10.1016/j.exphem.2010.03.015

Shapouri-Moghaddam, A., Mohammadian, S., Vazini, H., Taghadosi, M., Esmaeili, S. A., Mardani, F., et al. (2018). Macrophage plasticity, polarization, and function in health and disease. *J. Cell. Physiol.* 233, 6425–6440. doi: 10.1002/jcp.26429

Šošić, D., Richardson, J. A., Yu, K., Ornitz, D. M., and Olson, E. N. (2003). Twist regulates cytokine gene expression through a negative feedback loop that represses NF-kappaB activity. *Cell* 112, 169–180. doi: 10.1016/s0092-8674(03)00002-3

Stephen, R. F., Benson, M. K., and Nade, S. (2012). Misconceptions about childhood acute osteomyelitis. *J. Child. Orthop.* 6, 353–356. doi: 10.1007/s11832-012-0435-x

Street, M., Puna, R., Huang, M., and Crawford, H. (2015). Pediatric acute hematogenous osteomyelitis. *J. Pediatr. Orthop.* 35, 634–639. doi: 10.1097/BPO.0000000000000332

Szklarczyk, D., and Jensen, L. J. (2015). Protein-protein interaction databases. *Methods Mol. Biol.* 1278, 39–56. doi: 10.1007/978-1-4939-2425-7_3

Wang, N., Guo, D., Zhao, Y. Y., Dong, C. Y., Liu, X. Y., Yang, B. X., et al. (2015). TWIST-1 promotes cell growth, drug resistance and progenitor clonogenic capacities in myeloid leukemia and is a novel poor prognostic factor in acute myeloid leukemia. *Oncotarget* 6, 20977–20992. doi: 10.18632/oncotarget.4007

Whyte, N. S., and Bielski, R. J. (2016). Acute hematogenous osteomyelitis in children. *Pediatr. Ann.* 45, e204–e208. doi: 10.3928/00904481-20160428-01

Wu, S., Zhang, Q., Zhang, F., Meng, F., Liu, S., Zhou, R., et al. (2019). HER2 recruits AKT1 to disrupt STING signalling and suppress antiviral defence and antitumour immunity. *Nat. Cell Biol.* 21, 1027–1040. doi: 10.1038/s41556-019-0352-z

Xiao, L., Zhou, Y., Friis, T., Beagley, K., and Xiao, Y. (2019). S1P-S1PR1 signaling: the "Sphinx" in osteoimmunology. *Front. Immunol.* 10:1409. doi: 10.3389/fimmu.2019.01409

Xiong, H., and Pamer, E. G. (2015). Monocytes and infection: modulator, messenger and effector. *Immunobiology* 220, 210–214. doi: 10.1016/j.imbio.2014.08.007

Xu, Y., Qin, L., Sun, T., Wu, H., He, T., Yang, Z., et al. (2017). Twist1 promotes breast cancer invasion and metastasis by silencing Foxa1 expression. *Oncogene* 36, 1157–1166. doi: 10.1038/onc.2016.286

Yeo, A., and Ramachandran, M. (2014). Acute haematogenous osteomyelitis in children. *BMJ* 348:g66. doi: 10.1136/bmj.g66

Zhou, J., Zhang, J., Xu, M., Ke, Z., Zhang, W., and Mai, J. (2019). High SRC-1 and Twist1 expression predicts poor prognosis and promotes migration and invasion by inducing epithelial-mesenchymal transition in human nasopharyngeal carcinoma. *PLoS One* 14:e0215299. doi: 10.1371/journal.pone.0215299

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 19 June 2020
doi: 10.3389/fgene.2020.00610

Integration of Transcriptome, Gross Morphology and Histopathology in the Gill of Sea Farmed Atlantic Salmon (*Salmo salar*): Lessons From Multi-Site Sampling

Elżbieta Król[1], Patricia Noguera[2], Sophie Shaw[3], Eoin Costelloe[1], Karina Gajardo[4], Victoria Valdenegro[4], Ralph Bickerdike[5], Alex Douglas[1] and Samuel A. M. Martin[1]**

[1] School of Biological Sciences, Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen, United Kingdom, [2] Fish Health and Welfare, Marine Scotland Science, Aberdeen, United Kingdom, [3] Centre for Genome-Enabled Biology and Medicine, University of Aberdeen, Aberdeen, United Kingdom, [4] BioMar AS, Trondheim, Norway, [5] Scottish Sea Farms, Stirling, United Kingdom

OPEN ACCESS

Edited by:
Dong Xia,
Royal Veterinary College (RVC), United Kingdom

Reviewed by:
Aleksei Krasnov, Nofima, Norway
Mark D. Fast, University of Prince Edward Island, Canada

***Correspondence:**
Alex Douglas
a.douglas@abdn.ac.uk
Samuel A. M. Martin
sam.martin@abdn.ac.uk

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *29 February 2020*
Accepted: *19 May 2020*
Published: *19 June 2020*

Citation:
Król E, Noguera P, Shaw S, Costelloe E, Gajardo K, Valdenegro V, Bickerdike R, Douglas A and Martin SAM (2020) Integration of Transcriptome, Gross Morphology and Histopathology in the Gill of Sea Farmed Atlantic Salmon (Salmo salar): Lessons From Multi-Site Sampling. Front. Genet. 11:610. doi: 10.3389/fgene.2020.00610

The gill of teleost fish is a multifunctional organ involved in many physiological processes such as gas exchange, osmotic and ionic regulation, acid-base balance and excretion of nitrogenous waste. Due to its extensive interface with the environment, the gill plays a key role as a primary mucosal defense tissue against pathogens, as manifested by the presence of the gill-associated lymphoid tissue (GIALT). In recent years, the prevalence of multifactorial gill pathologies has increased significantly, causing substantial losses in Atlantic salmon aquaculture. The transition from healthy to unhealthy gill phenotypes and the progression of multifactorial gill pathologies, such as proliferative gill disease (PGD), proliferative gill inflammation (PGI) and complex gill disorder (CGD), are commonly characterized by epithelial hyperplasia, lamellar fusion and inflammation. Routine monitoring for PGD relies on visual inspection and non-invasive scoring of the gill tissue (gross morphology), coupled with histopathological examination of gill sections. To explore the underlying molecular events that are associated with the progression of PGD, we sampled Atlantic salmon from three different marine production sites in Scotland and examined the gill tissue at three different levels of organization: gross morphology with the use of PGD scores (macroscopic examination), whole transcriptome (gene expression by RNA-seq) and histopathology (microscopic examination). Our results strongly suggested that the changes in PGD scores of the gill tissue were not associated with the changes in gene expression or histopathology. In contrast, integration of the gill RNA-seq data with the gill histopathology enabled us to identify common gene expression patterns associated with multifactorial gill disease, independently from the origin of samples. We demonstrated that the gene expression patterns associated with multifactorial gill disease were dominated by two processes: a range of immune responses driven by pro-inflammatory cytokines and the events associated with tissue damage and repair, driven by caspases and angiogenin.

Keywords: proliferative gill disease, gene expression, RNA-seq, immune response, gill inflammation, aquaculture, climate change

INTRODUCTION

The gill of teleost fish is a multifunctional organ involved in many physiological processes such as gas exchange, osmotic and ionic regulation, acid-base balance and excretion of nitrogenous waste (Evans et al., 2005). To facilitate these functions, the gill tissue has evolved into a highly complex system of branching vascular structures (lamellae), separated from the external milieu only by a thin layer of gill epithelium and mucosa (Koppang et al., 2015; Salinas, 2015). The densely packed lamellar structure of the gill is highly advantageous because it provides a large surface area for oxygen transfer, amounting to approximately 0.1–0.4 m^2 of lamellar surface per kg of body mass (Maina, 2011; Park et al., 2014). However, having such an extensive interface with the environment comes at a high price (reviewed in Nilsson, 2007). Firstly, the large respiratory surface area may contribute to the increased water and ion fluxes that need to be counteracted by energetically expensive ion pumping. That so-called osmorespiratory compromise, which is a trade-off between high gill permeability (to promote respiratory gas exchange) and low gill permeability (to limit unfavorable water and ion fluxes), has been demonstrated in freshwater fish during exercise and in hypoxia-intolerant species (e.g., salmonids) during the periods of exposure to low-oxygen water (Onukwufor and Wood, 2018; Morgenroth et al., 2019). Secondly, the highly complex vasculature of the gill is prone to mechanical injuries, increasing the risk of hemorrhage. Because the gill is the only organ in fish that receives the entire cardiac output, large or frequent gill hemorrhages can be fatal. Thirdly, the large lamellar surface area may facilitate the uptake of toxic substances, both those that occur naturally (such as ammonia, algal toxins and metal ions) as well as man-made pollutants (e.g., pesticides, herbicides, fertilizers, detergents, industrial chemicals and microplastics) (McIntyre et al., 2018; Kintner and Brierley, 2019). Finally, the close proximity of blood and environmental water across the large respiratory surface area provides (1) major ports of entry for pathogens via transepithelial transport and (2) attachment sites for water-born parasites. To combat pathogens and parasites, the gill is equipped with the gill-associated lymphoid tissue (GIALT), adding another dimension to the importance of this organ to fish functioning and survival (Koppang et al., 2015; Salinas, 2015; Xu et al., 2020).

To avoid 'the large interface' penalties, fish constantly adjust the functional size of the gill tissue to match their current oxygen demands (Nilsson, 2007). When faced with changes in their need for oxygen uptake (due to changes in metabolic rate or water oxygen content), fish are known to (1) regulate the water flow over the gill by adjusting the volume and frequency of buccal pumping and/or (2) regulate the blood flow inside the gill to alter the perfusion levels of lamellae. New evidence suggests that many teleost species may also undergo extensive gill tissue remodeling, during which lamellae are either completely embedded in a cell mass (to minimize the respiratory surface area) or fully protruded (to maximize the functional size of the tissue), with many potential intermediate stages between these extremes (Sollid et al., 2003; Anttila et al., 2015). Whether these regulatory and compensatory mechanisms aiming to optimize gill performance are sufficient for fish to cope with climate change, especially in the areas of high human impacts, remains unknown.

Fish are predominantly ectothermic and any increases in water temperature resulting from climate change will increase their metabolic rate and thus demands for oxygen. The Q_{10} for metabolism in fish (the fold increase in metabolic rate with a 10% increase in water temperature) varies from 1.5 to 2 (White et al., 2006). But the elevated oxygen demand is not the only challenge that the gill tissue is facing. Warmer water holds less oxygen and rising global temperatures alter oceanic circulation, which altogether may decrease the oxygen supply available for fish to support their increased metabolism (McBryan et al., 2013). It has been estimated that the open ocean lost ∼2% (77 billion metric tons) of its oxygen over the past 50 years (reviewed in Breitburg et al., 2018). The extent of deoxygenation of coastal waters may even be greater, due to eutrophication from agriculture and sewage runoff (Diaz and Rosenberg, 2008). Furthermore, environmental changes, including temperature increases, have been linked to enhanced expression of marine infectious diseases (Burge et al., 2014). Many marine organisms, including marine pathogens, are shifting their distributions poleward or to deeper waters as ocean temperatures warm (Nye et al., 2009; Pinsky et al., 2013). As a result, many species of fish are at risk of being exposed to novel pathogens and parasites that were not part of their evolutionary history (LeBlanc et al., 2019). Climate change may also affect the fish-pathogen interactions that are already well established, for example by altering pathogen virulence, overall pathogenicity and/or host immune response (Benedicenti et al., 2019; Laurin et al., 2019). Given the fact that in fish most environmental stressors converge at the level of the gill, performance of this tissue and its capacity to accommodate environmental change are central to understand short- and long-term impacts of warming and hypoxia on fish species (Akbarzadeh et al., 2018; McCormick and Regish, 2018; Houde et al., 2019). Monitoring gill health is especially important in farmed marine fish, which are restricted to coastal waters and have limited microhabitat choice.

In recent years, the prevalence of gill damage and disease in farmed marine fish has significantly increased, leading to substantial losses in the Scottish and global Atlantic salmon (*Salmo salar*) aquaculture industry (Young et al., 2007; Shinn et al., 2015; Nowak and Archibald, 2018). Although gill-related mortality varies from site to site and between years, July to December in Northern Hemisphere is considered the higher risk period because the sea during these months is warmer. With the projected increase in sea water temperature associated with climate change, gill pathologies are now regarded as the greatest challenge to the farmed salmon sector due to direct and indirect impacts on productivity. Losses may occur through direct mortalities, poor growth rates linked to decreased feed efficiency and because of the increased risk of co-infection (Rodger, 2007; Downes et al., 2018). The transition from healthy to unhealthy gill tissue happens through the progression of various gill pathologies such as amoebic gill disease (AGD), proliferative gill disease (PGD), proliferative gill inflammation (PGI) and complex gill disorder (CGD), which are commonly characterized by epithelial hyperplasia, lamellar fusion and inflammation

(Steinum et al., 2010; Matthews et al., 2013; Mitchell et al., 2013; Herrero et al., 2018; Gjessing et al., 2019). The diagnostic of gill pathologies is often difficult because single-cause diseases (such as AGD) are compounded by multi-cause (multifactorial) disorders such as PGD, PGI or CGD, with complex etiology and potentially synergistic effects of co-infections (Gjessing et al., 2017; Gunnarsson et al., 2017; English et al., 2019). Due to overlapping symptoms of multifactorial gill inflammation, the terms PGD, PGI and CGD are often used interchangeably, with the first two historically referring to salmon farmed in Scotland (PGD) and Norway (PGI), while the last one highlighting the growing complexity of gill diseases and encompassing the syndromes referred to as PGD and PGI (Rozas-Serri, 2019). The multifactorial gill diseases are associated with a number of causative agents, including phytoplankton, zooplankton, viruses, fungi, bacteria and larger parasites (Gjessing et al., 2017; Herrero et al., 2018), with climate change contributing to emerging new aquaculture pathogens (Bayliss et al., 2017). Furthermore, gill diseases not only directly affect fish health and performance, but also creates a challenge for farmed salmon producers as they are required to treat compromised fish, increasing the risk of mechanical injuries to the gill vasculature (Powell et al., 2005). To meet the challenge of salmon gill health, the sector needs non-invasive monitoring tools for (1) early detection of gill pathologies, (2) diagnostics that is sufficient to establish appropriate treatment and (3) evaluation of effectiveness of treatment and recovery of gill function. It is important that such initiatives target multifactorial diseases, as they are becoming progressively more prevalent (Rahel and Olden, 2008).

The current practice is to regularly check sea farmed salmon for any signs of gill inflammation by *in situ* gross morphology examination of gill arches, using the PGD scoring system from 0 (no macroscopic pathology) to 5 (severe macroscopic pathology) (Gill Health Initiative, 2015; Bloecher et al., 2018). This is a non-invasive procedure that requires only light anesthesia and can be performed on a weekly basis in fish from across a facility. However, the applicability of the PGD scoring system to grade inflammation has never been tested in a systematic way. Because PGD is a multifactorial disease, progression from low to high PGD scores is likely to have a large component of case specificity, depending on the combination of pathogens involved, the overall health status of the fish and the environmental conditions. Despite complex etiology, the phenotypes generated by PGD are macroscopically similar (inflammation and hyperplasia of respiratory epithelium), which suggests that at least some of the underlying mechanisms associated with this pathology may be common (Mitchell and Rodger, 2011; Bruno et al., 2013). Unlocking these shared mechanisms is essential for early detection of gill disease, developing strategies to improve gill health and for fine-tuning salmon husbandry practices to the challenging conditions of a rapidly changing marine environment.

To identify the common gene expression patterns associated with gill inflammation, we examined Atlantic salmon from three marine production sites in Scotland. The sites and time of sampling were chosen to ensure (1) differences in fish cohort (hatchery origin, on-site treatments and overall health status), (2) diversity of pathogen and (3) differences in local environment. For each site, we selected fish with low and high PGD scores, analyzed their gill transcriptome (RNA-seq) and gill histopathology (microscopic examination), and then integrated these data to explore gene expression patterns associated with multifactorial gill disease.

MATERIALS AND METHODS

Sampling and Gross Morphology

Fish (Atlantic salmon, *Salmo salar*) were sampled at three marine production sites belonging to Scottish Sea Farms (A on Isle of Mull and B and C in Shetland) between October 2017 and March 2018 (for details see **Table 1** and **Supplementary Figure 1**). All fish were of strain Fanad and originated from the same egg fertilization batch. They were reared in different hatcheries (Couldoran, Pettigo-Damph and Knock-Frisa for sites A, B and C, respectively) for one year and entered the sea in spring 2017.

During sampling, randomly selected fish (76 in total) were placed in an anesthetic bath (∼20 g of MS-222/150 L) for 5–10 min, followed by macroscopic examination (gross morphology) of gill tissue as per routine procedure to monitor fish welfare indicators. This procedure is performed by fish health professionals, using semi-quantitative 6-grade scoring systems for PGD, from 0 (no visual pathology) to 5 (severe visual pathology) (Gill Health Initiative, 2015; Bloecher et al., 2018). Each of the 8 gill arches (4 on left and 4 on right side) was scored separately, generating in total 8 PGD scores per fish. Although scoring for PGD is typically non-destructive and requires only light anesthesia (e.g., 10 g of MS-222/150 L), fish in our study were subjected to terminal anesthesia for subsequent harvesting of tissue samples. Immediately after PGD scoring, fish were bled and the gill arch with the highest PGD score was excised for both transcriptome profiling and histopathological examination. For transcriptomic profiling, 3 transverse sections (gill arch ∼3 mm long with 3–4 pairs of filaments) from dorsal, medial and ventral regions of the gill (**Figure 1**) were transferred to RNAlater® (Sigma-Aldrich, St. Louis, MO, United States), kept at

TABLE 1 | Sampling details, fish metrics and background.

Site	Date	SWT (°C)[1]	Number of fish	Body mass (kg)[2]	Body length (cm)[2]	$K^{2,3}$
A (Isle of Mull)	19 Oct 2017	13.4	31	2.4 ± 0.5	55 ± 3	1.4 ± 0.1
B (Shetland)	24 Nov 2017	13.9	20	2.4 ± 0.4	55 ± 3	1.4 ± 0.2
C (Shetland)	22 March 2018	5.7	25	2.6 ± 0.7	55 ± 4	1.5 ± 0.2

[1]Sea water temperature at 5 m depth; [2]Values are means ± standard deviations; [3]Fulton's condition factor (K) calculated as 100 × body mass in g/body length in cm^3.

4°C overnight for equilibration and then stored at −80°C prior to RNA extraction. All remaining tissue from the same gill arch was transferred to freshly prepared seawater Davidson's fixative (Cadoret et al., 2013) for 24 h, followed by a short-term exposure to 10% neutral buffered formalin before tissue processing for histopathological examination.

Selection of Fish With Low and High PGD Scores

Grouping all 76 sampled fish by their highest PGD score and site demonstrated that each site had fish with PGD scores 1, 2 or 3, but not 0 or 5, and only rarely 4 (**Table 2**). Thus, the comparison between fish with low and high PGD scores was restricted to PGD scores 1 and 3, respectively. In total, 45 gill arches (27 with PGD score 1 and 18 with PGD score 3) were subjected to total RNA extraction, but the final RNA-seq sample size dropped to 43 gill arches (from 43 fish), due to losses at the both pre- and post-sequencing stage (**Table 2**).

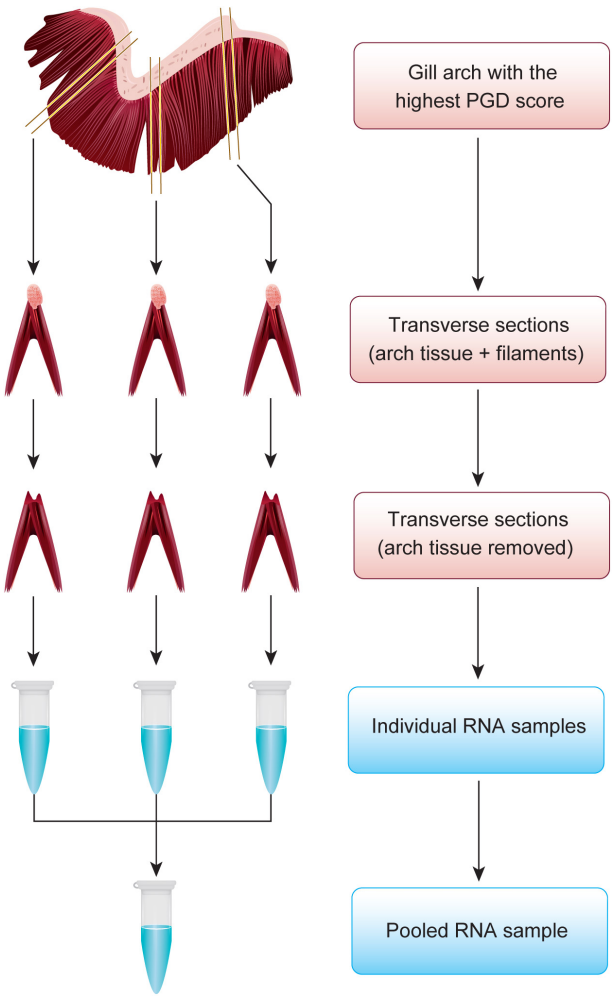


FIGURE 1 | Approach to representatively divide the gill arch with the highest PGD score between transcriptome profiling and histopathological examination. For transcriptomic profiling, 3 transverse sections from dorsal, medial and ventral regions of the gill were subjected to total RNA extraction to generate individual RNA samples, which were subsequently pooled. All remaining tissue was used for histopathology.

RNA Extraction

Total RNA extraction was performed on individual gill transverse sections (45 gills × 3 sections) after removing the arch tissue and leaving only full-length filaments for further processing (**Figure 1**). Briefly, the RNA was isolated by homogenization of ∼100 mg of gill tissue in TRIzol® Reagent (Ambion by Life Technologies, Carlsbad, CA, United States), using 3 mm tungsten carbide beads and a TissueLyser II Disruption System (Qiagen GmbH, Hilden, Germany). Following isolation, the RNA was quantified by spectrophotometry (NanoDrop Technologies, Wilmington, DE, United States) and its integrity was confirmed by electrophoresis (Agilent Technologies, Santa Clara, CA, United States). The three individual RNA samples that originated from the same gill were then pooled to generate a single RNA sample per gill tissue (n = 45 RNA pools in total), with an equimolar contribution of RNA from dorsal, medial and ventral regions of the gill to each pool. All but one pooled gill RNA samples had a 260/280 ratio > 1.8 and RIN number > 9.3, thus meeting the criteria for RNA-sequencing. The sample with degraded RNA was eliminated from further processing.

RNA-seq Library Preparation and Sequencing

RNA-seq library preparation and sequencing were carried out by Edinburgh Genomics at the University of Edinburgh (United Kingdom). The libraries for each of the 44 samples were constructed using the TruSeq Stranded mRNA Sample Preparation Kit (Illumina, San Diego, CA, United States), according to the manufacturer's instructions. The paired-end sequencing (50 bp from each end) was performed on the NovaSeq 6000 system with S2 flow cell (Illumina, San Diego, CA, United States) at a sequencing depth of ∼50 million read pairs per library. The raw reads in BCL format were converted to FastQ format with bcl2fastq2 Conversion Software v2.19.1 (Illumina, San Diego, CA, United States). All raw sequences have been deposited in the ArrayExpress repository[1] under accession number E-MTAB-8855.

Read Mapping

To assess the quality of the sequencing data, reads were analyzed with FastQC v0.11.8 (Andrews, 2010). Sequencing adaptors and sequences shorter than 20 bp were removed using Flexbar v3.4.0 (Dodt et al., 2012). Filtered reads were then mapped to the Atlantic salmon reference genome ICSASG_v2 (GenBank: GCF_000233375.1, Lien et al., 2016) using HISAT2 v2.1.0. (Kim et al., 2015) with the stranded library preparation parameter. Overall, alignment rates ranged from 93.2 to 95.7%. Aligned reads were counted at gene locations using featureCounts v1.6.4 (Liao et al., 2014). For multi-mapping reads, a fractional count (1/n) was generated for each reported alignment of the

[1]http://www.ebi.ac.uk/arrayexpress/

TABLE 2 | Results of gross gill scoring for proliferative gill disease (PGD), performed on all 8 gill arches of individual fish to identify the gill arch with the highest PGD score for each fish.

		Number of fish grouped by their highest PGD score[2]					
Site	**Median PGD score[1]**	**Score 0**	**Score 1**	**Score 2**	**Score 3**	**Score 4**	**Score 5**
A (Isle of Mull)	1 (range 1–3.5, n = 31)	-	**10**	10	**10**	1	-
B (Shetland)	1 (range 1–2, n = 20)	-	**8**[3]	9	**3**	-	-
C (Shetland)	1 (range 0–3, n = 25)	-	**9**	10	**5**[4]	1	-

The gill arches (one from each fish) subjected to transcriptome and histopathological examination (PGD scores 1 and 3) are presented in bold. [1]Values are median PGD scores per fish (calculated from all 8 gill arches) per site (calculated for n fish), with range referring to fish with minimal and maximal median PGD scores at each site; [2]- denotes no fish; [3]Final n = 7, because one total RNA sample was degraded; [4]Final n = 4, because one RNA-seq sample was identified as an outlier and removed from analysis.

multi-mapping read, with n reflecting the total number of alignments reported for that read.

Differential Gene Expression

Differential expression analysis was performed using the Bioconductor package edgeR v3.22.5 (Robinson et al., 2010) in R v3.5.1 (R Core Team, 2018). Genes with a CPM (count per million) < 1 in three or more samples were removed, resulting in 35996 genes for analysis. Filtered counts were subsequently normalized using a trimmed mean of *M*-values (TMM) between each pair of samples. Based on exploratory data analysis, one library was identified as an outlier and removed from subsequent analysis (for details on sample size see **Table 2**). Data for the remaining 43 fish were modeled using a negative binomial generalized log-linear mixed model that included both group (PGD scores 1 and 3) and site (A, B and C) as fixed effects. In total, four contrasts were generated from the same model: PGD 3 vs PGD 1 (with the site effect blocked) and then sites A vs C, B vs C and A vs B (with the group effect blocked). Differentially expressed genes (DEGs) were identified at false discovery rate (FDR) < 0.01 and absolute Log_2 fold change (FC) > 1.

Functional Analysis of Gene Expression

Salmon DEGs were mapped to human orthologs to generate HGNC (HUGO Gene Nomenclature Committee) gene identifiers for functional analysis of the RNA-seq results. This approach has been demonstrated to improve biological interpretation of the salmon gene expression profiles by providing access to well-annotated databases and tools for mammalian model organisms, despite limitations of the mapping due to the extra genome duplication events in teleost fish and species-specific differences in gene function and molecular pathways (Song et al., 2014; Król et al., 2016). Mapping was done by aligning salmon transcript sequences to the protein sequences from the human genome (release 88, downloaded from Ensembl at[2]) using BLASTX (version 2.2.31) (Camacho et al., 2009) with an E-value cut off of 0.00001 and a maximum of one target sequence for every transcript. As one transcript can have multiple hits against one target sequence, custom Python scripts were used to filter the blast results to contain only the hit with the highest identity for each transcript. Although the majority of the salmon genes mapped to a unique human ortholog, some salmon genes mapped to the same human ortholog (for details see Results). To obtain a single expression value (mean Log_2 FC) per HGNC gene identifier, the expression of the salmon genes mapped to the same human ortholog was either averaged (if contributing salmon genes had similar expression profiles) or based on the expression of the salmon gene that was more abundant (if contributing salmon genes had contrasting expression profiles). For biological interpretation of the RNA-seq results, human orthologs of the salmon DEGs along with their Log_2 FC values were analyzed using Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, www.qiagen.com/ingenuity) to explore (1) enrichment of canonical pathways, (2) upstream regulators and (3) downstream effects associated with these genes. The same set of genes was also submitted to PANTHER Classification System (Mi et al., 2013) to perform Gene Ontology (GO) enrichment analysis.

Gill Histopathology

Gill tissue was routinely dehydrated in ethanol, equilibrated in xylene and embedded in paraffin wax according to standard histological techniques (Bancroft and Gamble, 2007). Sagittal sections (3 μm) of the gill arch were cut with a microtome and mounted onto microscope slides. These sections were then subjected to haematoxylin and eosin (H&E) staining. All sections were digitized at 40 × magnification, using the Olympus dotSlide 2.1 Virtual Slide System (Olympus Corporation, Tokyo, Japan). The resultant images were randomized to ensure blinded examination and then scored using a system developed to assess gill histopathology in sea farmed Atlantic salmon (Mitchell et al., 2012), with minor amendments. The details of the scoring system are presented in **Table 3**.

Prior to the analysis, three variables were excluded (bE, bT and ob) as scores were invariant across fish (see Results). To compare the gill histopathology scores between PGD groups (1 and 3) and also between fish from different sites (A, B and C), non-metric multidimensional scaling (NMDS) was performed using the metaMDS function from the vegan package (Oksanen et al., 2019) in R. For the NMDS analysis, a Jaccards dissimilarity index was used to calculate the dissimilarity matrix, two dimensions were specified and 100 random starts used. Similarities between fish histopathology scores were visualized using a biplot with 95% confidence ellipses around the group centroids and variable vectors included.

[2]https://www.ensembl.org/Homo_sapiens/Info/Index

TABLE 3 | Semi-quantitative scoring system used for histopathological examination of gill tissue in sea farmed Atlantic salmon (adapted and modified from Mitchell et al., 2012).

Parameter	Description
Index criteria with scores 0–3[1]	
LH	Lamellar hyperplasia
LF	Lamellar fusion
LO	Lamellar oedema
CA	Cellular anomalies
Ancillary criteria with scores 0–3[1]	
In	Inflammation (presence of inflammatory cells outside blood vessels)
Eg	Eosinophilic granular cells (numbers higher than normal in gill filaments)
Cc	Chloride cells (numbers higher than normal and/or abnormal location)
Cd	Circulatory disturbances (thrombi, telangiectasis, congestion)
Ib	Interlamellar blood (hemorrhage)
Ch	Cellular hypertrophy (hydropic degeneration of lamellar cells)
bE	*Epitheliocystis*-like bacteria
bT	*Tenacibaculum*-like bacteria (mats of filamentous bacteria on lamellar surfaces)
Ob	Other bacteria
Pp	Protist parasites, *Neoparamoeba*-like
Op	Other parasites or agents

[1]*Scoring system 0–3, with 0 = no pathological changes, 1 = mild changes affecting < 10% of gill tissue, 2 = moderate changes affecting 10–50% of gill tissue, and 3 = severe changes affecting > 50% of gill tissue.*

RESULTS

Phenotypic Characteristics of Fish

Body mass of 43 fish subjected to the RNA-seq experiment varied from 1.5 to 4.0 kg, with their body length ranging from 47 to 62 cm (**Figures 2A,B**). The effects of group (PGD 1 and PGD 3) and site (A, B and C) on body mass and length were not significant ($P > 0.05$, 2-way ANOVA). As a result, the fish did not differ in their Fulton's condition factor K ($P > 0.05$, 2-way ANOVA), which varied from 1.1 to 1.7 (**Figure 2C**). Given the small sample size (**Table 2**), these results should be treated with caution.

Gill Transcriptome

The NMDS analysis of 43 gill transcriptomes revealed no differences between 26 fish with PGD score 1 and 17 fish with PGD score 3, as indicated by their overlapping 95% confidence intervals (**Figure 3A**). This finding was reinforced by differential gene expression analysis, which showed 0 DEGs for comparison PGD 3 vs PGD 1 (**Table 4** and **Supplementary Table 1**). In contrast, gill transcriptomes were clearly separated by site, with no overlap in 95% confidence intervals for 20 fish from site A, 10 fish from site B and 13 fish from site C (**Figure 3B**). The site-specific differences in the gill transcriptomes are reflected in the number of DEGs identified between the sites: 1360 for A vs C, 708 for B vs C and 240 for A vs B (**Table 4**, for the lists of DEGs see **Supplementary Tables 2–4**).

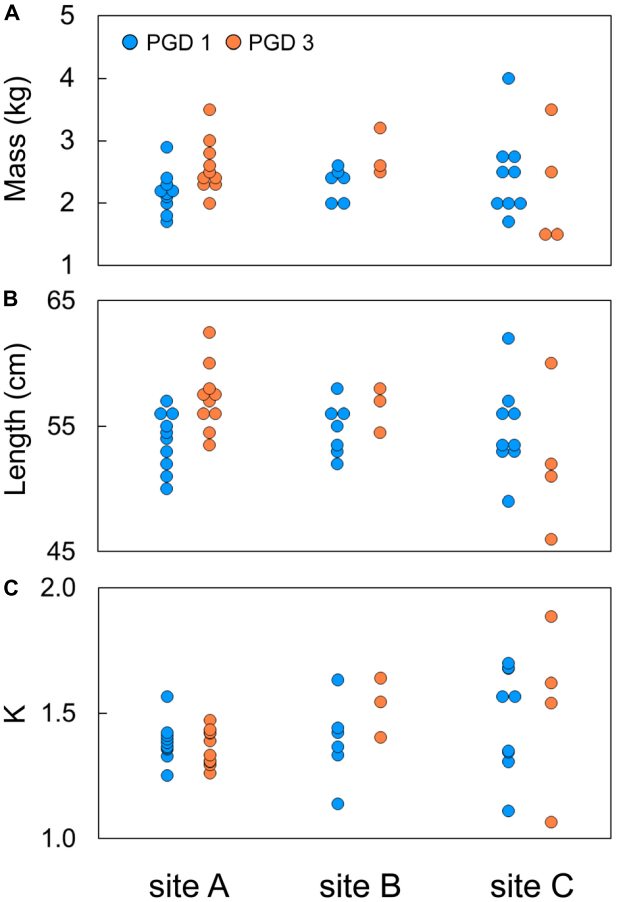


FIGURE 2 | Individual body mass **(A)**, body length **(B)** and Fulton's condition factor K **(C)** of 43 fish subjected to RNA-seq experiment, plotted by group (PGD 1 and PGD 3) and site (A, B and C). For formula to calculate K, see **Table 1**.

Gill Histopathology

Histopathological examination of the gill tissue generated 15 scores (ranging from 0 to 3) per fish, focusing on 4 index and 11 ancillary criteria (**Supplementary Table 5**). Distribution of the histopathological scores by PGD score (1 and 3) and site (A, B and C) for each of 15 criteria is presented in **Supplementary Figure 2**. The NMDS analysis of the gill histopathology results for 43 fish showed no differences between 26 fish with PGD score 1 and 17 fish with PGD score 3, as indicated by their overlapping 95% confidence intervals (**Figure 4A**). In contrast, gill histopathology differed between sites, with 1) fish from site C being clearly separated from other fish and 2) fish from site A overlapping with fish from site B (**Figure 4B**). The position of the site-associated clusters along the NMDS1 axis indicates that fish from site C (with lower scores) show relatively low level of gill histopathology, while fish from sites A and B (with higher scores) have gill tissue with relatively moderate changes in histopathology. The main drivers of these differences are lamellar hyperplasia (LH), lamellar fusion (LF), cellular anomalies (CA)

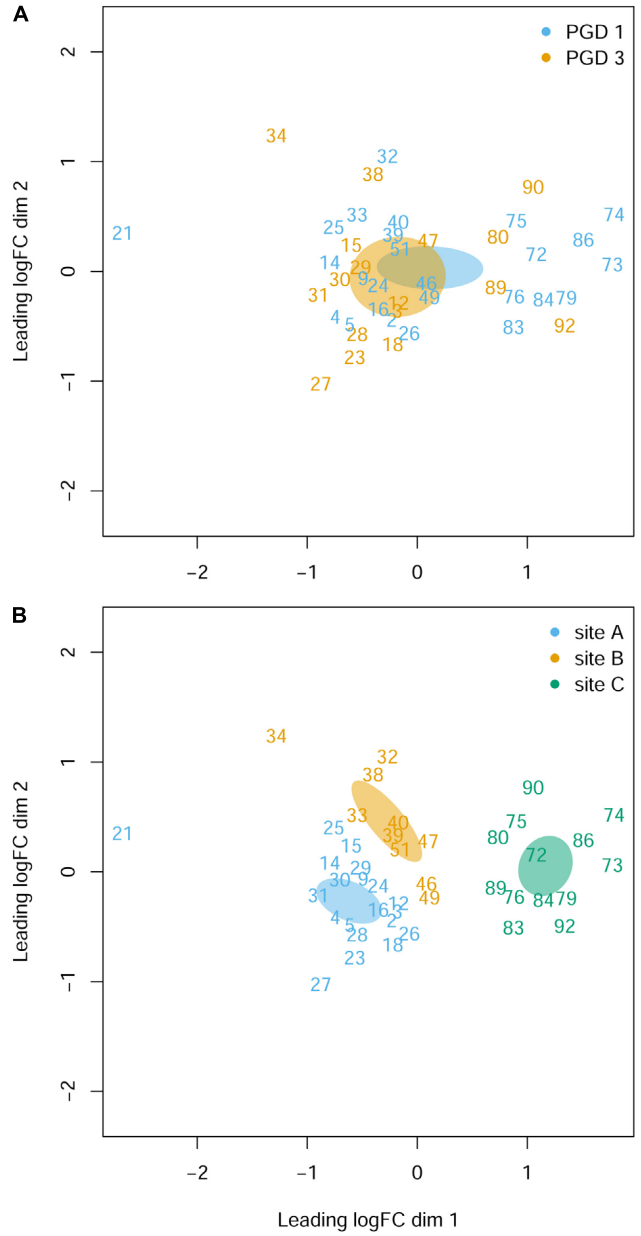


FIGURE 3 | Gill transcriptome of 43 fish grouped by PGD score **(A)** and site **(B)**. Each panel shows a NMDS plot of gene expression profiles between different fish (numbers are fish IDs). The distances on the plots correspond to the leading fold change (FC), which is the average (root-mean-square) Log_2 FC for the 500 genes most divergent between each pair of fish. Ellipses indicate 95% confidence intervals, overlapping for fish grouped by PGD score **(A)** but not for fish grouped by site **(B)**. The stress value of the NMDS ordination is 0.162.

and *Neoparamoeba*-like protist parasites (pp), as evidenced by the vectors paralleled to the NMDS1 axis in **Figure 4B**. Overall, fish from site A and independently from site B developed significantly higher grade inflammation than fish from site C, thus facilitating the search for the molecular bases of multifactorial gill disease.

TABLE 4 | Results of differential gene expression analysis performed on gill transcriptome of 43 fish to elucidate the differences between groups (PGD 3 vs PGD 1) and sites (A vs C, B vs C and A vs B).

		Number of differentially expressed genes (DEGs)		
Comparison	**N[1]**	**Total**	**Upregulated**	**Downregulated**
PGD 3 vs PGD 1	17 vs 26	0	0	0
sites A vs C	20 vs 13	1360	858	502
sites B vs C	10 vs 13	708	492	216
sites A vs B	20 vs 10	240	124	116

Genes were considered differentially expressed at FDR < 0.01 and absolute Log_2 FC > 1. [1]Number of fish used for each comparison.

Identification of Gene Expression Patterns Associated With Multifactorial Gill Disease

Our approach to identify the common gene expression patterns of non-specific gill inflammation is explained in **Figure 5**. Specifically, all gill transcriptomes from sites A and B (with moderate gill histopathology) were independently compared to all gill transcriptomes from site C (with low gill histopathology), which resulted in 1360 and 708 DEGs (at FDR < 0.01 and absolute Log_2 FC > 1) for A vs C and B vs C contrasts, respectively (**Table 4**, for the lists of DEGs see **Supplementary Tables 2**, **3**). A relatively large number of genes were common between these two comparisons (462 DEGs in total), including 354 protein-coding genes, 35 immunoglobulin gene segments, 15 pseudogenes and 58 non-coding RNAs (**Table 5**, for the list of common DEGs see **Supplementary Table 6**). The Log_2 FC values for these genes were averaged between A vs C and B vs C contrasts to provide one expression value per common gene. Importantly, all common DEGs show the same direction of change in both the moderate-to-low histopathology contrasts, i.e., they are either upregulated or downregulated in relation to site C, as indicated in **Figure 5**. That commonality of the response suggests that the identified DEGs are part of the gene expression profile indicative of progression from lower to higher grade inflammation of gill tissue. To understand the underlying mechanisms, we focused on the 354 protein-coding DEGs and predicted their functionality through mapping to the human orthologs.

Mapping Salmon Genes to Human Orthologs

The results of blasting all 462 salmon DEGs associated with multifactorial gill disease against the protein sequences from the human genome (BLASTX, *E*-value < 0.00001, top hit) are presented in **Table 5** and **Supplementary Table 6**. Among them, 311 of 354 protein-coding salmon genes were matched to HGNC gene identifiers, but not all of them were unique. Specifically, 191 protein-coding salmon transcripts mapped uniquely to one human ortholog, while the remaining 120 protein-coding salmon transcripts mapped to 44 human orthologs, with > 1 salmon gene mapping to the same HGNC gene identifier (**Supplementary Table 7**). The expression of salmon genes mapped to the same

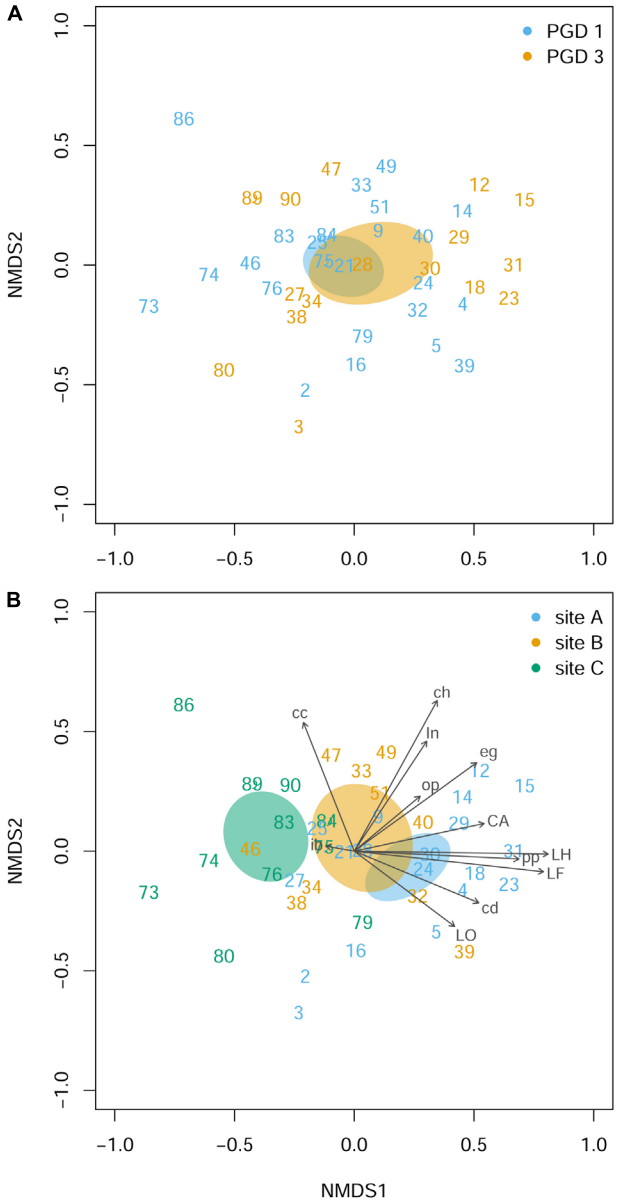


FIGURE 4 | Gill histopathology of 43 fish grouped by PGD score **(A)** and site **(B)**. Each panel shows a NMDS plot of histopathological profiles between different fish (numbers are fish IDs). The distances on the plots are calculated from the scores of 12 histopathological parameters (LH, LF, LO, CA, in, eg, cc, cd, ib, ch, pp and op), with 3 parameters (bE, bT and ob) removed from the analysis because all scores were 0 (for parameter abbreviations and details see **Table 3** and **Supplementary Figure 2**). Ellipses indicate 95% confidence intervals, overlapping for fish grouped by PGD score **(A)** and for fish from sites A and B but not from C **(B)**. Vectors are fitted to visualize the contribution of histopathological parameters to the ordination. The stress value of the NMDS ordination 0.190.

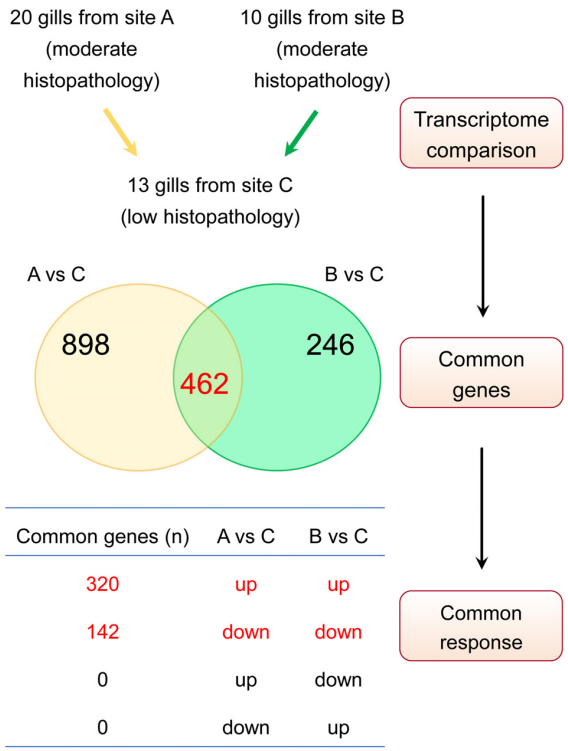


FIGURE 5 | Approach to identify gene expression patterns associated with multifactorial gill disease. Based on the histopathological examination, A and B were classified as two independent sites with relatively moderate gill histopathology, while C was classified as a site with relatively low gill histopathology. Firstly, the lists of differentially expressed genes (DEGs) for A vs C and B vs C contrasts were generated (FDR < 0.01 and absolute Log_2 FC > 1), yielding 1360 and 708 transcripts, respectively. Secondly, both lists of DEGs were checked for common genes (462 in total). Finally, the list of common genes was screened for a commonality of response to ensure that only genes either upregulated or downregulated in both sets of DEGs were considered to constitute a common gene expression profile of non-specific gill inflammation.

TABLE 5 | Characterisation of 462 potential gene expression markers associated with gill disease.

	Number of genes			
Gene type	**Total**	**HGNC +**	**HGNC –**	**IPA/GO**
Protein-coding	354	311	43	311 → 235
Immunoglobulin gene segments	35	0	35	0
Pseudogenes	15	1	14	0
Non-coding RNA	58	23	35	0

Mapping to human orthologs (BLASTX, E-value < 0.00001, top hit) generated in total 334 salmon genes with HGNC gene identifiers (HGNC +) and 128 salmon genes with no human orthologs (HGNC –). Only 311 protein-coding salmon genes with 235 unique HGNC gene identifiers were subjected to functional analysis of gene expression (IPA/GO). Abbreviations: HGNC, HUGO Gene Nomenclature Committee; IPA, Ingenuity Pathway Analysis; GO, Gene Ontology.

human ortholog was averaged to provide one expression value (mean Log_2 FC) per HGNC gene identifier, if the response of the salmon genes was consistent (i.e., all contributing salmon genes were either upregulated or downregulated, but not both). In two cases with the contrasting expression of salmon genes (PRF1 and CTSV, for details see **Supplementary Table 7**), the mean Log_2 FC was based on the expression of the salmon gene that was more abundant (higher CPM). As a result,

the expression patterns of 311 protein-coding salmon DEGs were represented by 235 human orthologs and their corresponding 235 expression values (Log_2 FC for 191 unique mapping and mean Log_2 FC for 44 multiple mapping), which were then used to predict functionality.

Functional Analysis of Gene Expression Patterns Associated With Multifactorial Gill Disease

Biological interpretation of the transcriptomic changes in gills with moderate histopathology (sites A and B) vs low histopathology (site C) was performed on the expression profiles of 311 protein-coding salmon DEGs converted into 235 human orthologs (**Supplementary Table 7**), using (1) IPA for enrichment of canonical pathways, upstream regulators and downstream effects and (2) PANTHER Classification System for enrichment of GO terms (Biological Process). The predictions made by IPA also require the expression values, while GO enrichment analysis is based on the list of DEGs.

IPA identified 13 canonical pathways that were altered in gills with moderate histopathology (sites A and B) vs low histopathology (site C) at *P*-value < 0.01, with contribution of 35 DEGs in total (**Figure 6** and **Supplementary Table 8**). Most of these pathways are associated with 1) cellular immune response (IL-17 Signaling, IL-6 Signaling, Granzyme A Signaling, Crosstalk between Dendritic Cells and Natural Killer Cells, Agranulocyte Adhesion and Diapedesis and HMGB1 Signaling), 2) cytokine signaling (IL-17 Signaling, IL-6 Signaling, Acute Phase Response Signaling, Role of JAK family kinases in IL-6-type Cytokine Signaling, TNFR2 Signaling and HMGB1 Signaling) and (3) tissue damage and repair. The tissue damage and repair were evidenced by alterations of pathways related to cellular stress and injury (Autophagy and HMGB1 Signaling), apoptosis (TNFR2 Signaling and JAK/Stat Signaling), cellular growth (STAT3 Pathway and JAK/Stat Signaling) and proliferation and development (STAT3 Pathway and JAK/Stat Signaling).

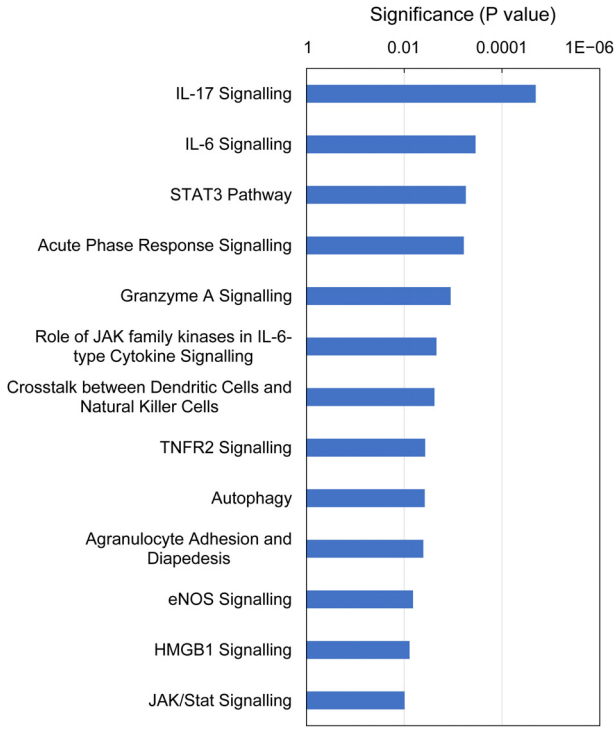


FIGURE 6 | Top canonical pathways altered in Atlantic salmon gills with moderate histopathology (sites A and B) vs low histopathology (site C). The analysis was performed on 311 salmon genes mapped to 235 human orthologs, using Ingenuity Pathway Analysis (IPA) and *P*-value < 0.01. For details and list of corresponding genes see **Supplementary Table 8**.

IPA analysis of upstream regulators is based on prior knowledge of predictable effects between transcriptional regulators (e.g., transcription factors, cytokines, microRNAs, receptors, kinases, chemicals and drugs) and their target genes, stored in the Ingenuity Knowledge Base. When such analysis was performed on the gene expression profiles indicative of multifactorial gill disease, IPA identified 15 top upstream regulators associated with 101 of the 235 submitted DEGs, all highly significant (overlap *P*-value < 0.0001) and activated (*z*-score > 2) (**Figure 7** and **Supplementary Table 9**). Among them were one endotoxin (LPS), one pattern recognition receptor (NOD2), 10 cytokines (IL1B, IFNG, TNF, IL2, LIF, IL12B, IL12A, IL1, IFNA2 and IL6), two growth factors (PDGF and AGT) and one transcription factor (NFKB).

IPA analysis of downstream effects predicts potential outcomes from gene expression data, using the Ingenuity Knowledge Base of differential gene expression in varying disease and functional states. The predictions made for gills with moderate histopathology (sites A and B) vs low histopathology (site C) included 13 downstream disease/functions at *P*-value < 9.4E-08, with nearly all DEGs (230 of 235) contributing to the effects (**Figure 8** and **Supplementary Table 10**). The dominant features of these predictions are (1) immune and inflammatory diseases (Inflammatory Response, Infectious Diseases, Inflammatory Disease, Immune Cell Trafficking, Immunological Disease and Humoral Immune Response), (2) tissue damage and repair (Organismal Injury and Abnormalities, Tissue Morphology, Connective Tissue Disorders, Cellular Movement, Cell Death and Survival, Cellular Function and Maintenance) and (3) intra-tissue communication (Cell-To-Cell Signaling and Interaction).

GO enrichment analysis identified 15 GO terms (Biological Process) that were overrepresented with the majority of DEGs (209 of 235) associated with multifactorial gill inflammation, generating fold enrichments from 1.2 to 2.6 at Bonferroni-corrected *P*-value < 0.05 (**Figure 9** and **Supplementary Table 11**). The enriched GO terms pointed towards increased (1) intra-tissue communication (Cellular response to cytokine stimulus and Cell communication), (2) presence of external stimuli (Response to external stimulus and Response to oxygen-containing compound), (3) activated immune response (Immune system process) and 4) ongoing tissue remodeling (Animal organ development and Regulation of biological quality).

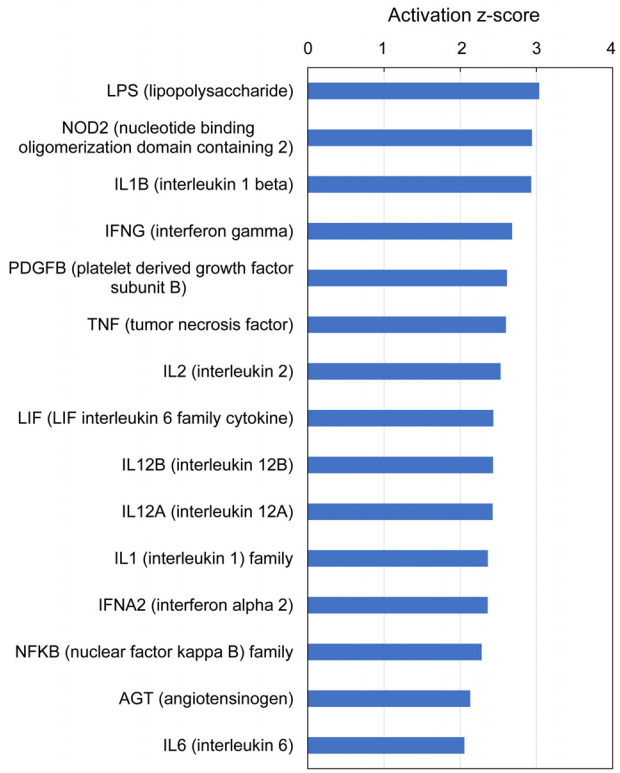


FIGURE 7 | Top upstream regulators predicted from gene expression in Atlantic salmon gills with moderate histopathology (sites A and B) vs low histopathology (site C). The analysis was performed on 311 salmon genes mapped to 235 human orthologs, using Ingenuity Pathway Analysis (IPA). For details and list of corresponding genes see **Supplementary Table 9**.

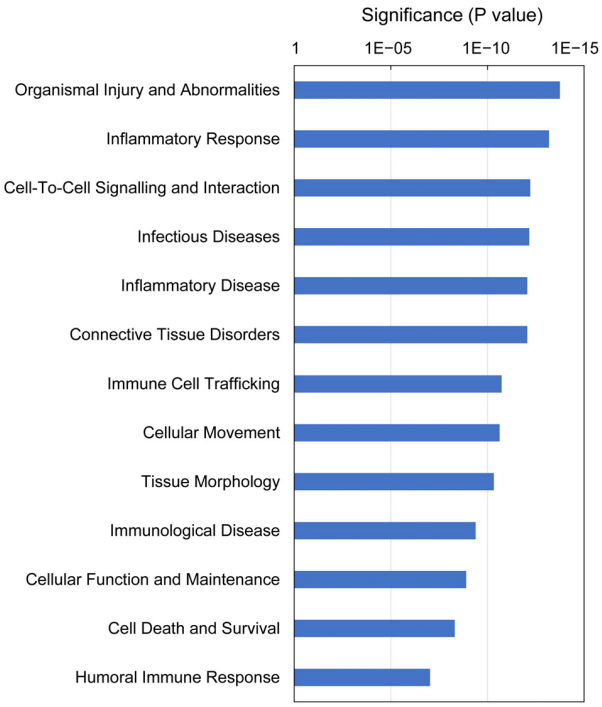


FIGURE 8 | Top downstream effects predicted from gene expression in Atlantic salmon gills with moderate histopathology (sites A and B) vs low histopathology (site C). The analysis was performed on 311 salmon genes mapped to 235 human orthologs, using Ingenuity Pathway Analysis (IPA). For details and list of corresponding genes see **Supplementary Table 10**.

Top Genes Associated With Multifactorial Gill Disease

Top genes were defined here as the DEGs with the largest magnitude of change in expression (absolute Log_2 FC $>$ 2), thus including the protein-coding genes that were at least 4 times higher or 4 times lower expressed in gills with moderate histopathology (sites A and B) vs low histopathology (site C). The list of these top genes is presented in **Table 6**, with 43 upregulated salmon genes mapped to 25 unique human orthologs and 14 downregulated salmon genes mapped to 12 unique human orthologs.

DISCUSSION

Our study is the first to focus on the gill transcriptome of Atlantic salmon in a highly variable and complex environment of the marine production sites. We specifically chose to work on fish farmed in the open sea-based system (such as net pens), because (1) the use of net pens is currently the most common form of salmon production (Philis et al., 2019) and (2) in the open pens, only nets separate the fish from the environment, making them prone to developing gill pathologies but also assisting with the tissue recovery through high water exchange and oxygenation (Shinn et al., 2015; Nowak and Archibald, 2018). Farming in net pens is typically associated with the multifactorial gill diseases (PGD, PGI or CGD) rather than the single-cause gill pathologies, such as AGD (Herrero et al., 2018; Gjessing et al., 2019; Laurin et al., 2019). So far, the extensive transcriptomic profiling of the gill tissue has been performed only in salmon exposed to the controlled environment of the closed-system tanks, following a single-pathogen challenge to induce either AGD (Morrison et al., 2006; Wynne et al., 2008a; Bloecher et al., 2018; Boison et al., 2019; Robledo et al., 2019) or infectious salmon anemia (Valenzuela-Miranda et al., 2015). In most cases, fish were subjected to a single-dose challenge of the infectious organisms, with only few studies evaluating the effects of re-infection (Bloecher et al., 2018; Boison et al., 2019) or co-exposure to other infectious agents, such as hydroids (Bloecher et al., 2018) and *Yersinia ruckeri* (Valdenegro-Vega et al., 2015). To closely resemble the dynamics of multifactorial gill diseases observed in the production systems, fish need to be continuously exposed to a variety of infectious agents and environmental stressors that impact their health.

To evaluate the robustness and reliability of the PGD scores (gross morphology) in conveying information about gill health, we sampled salmon at three geographically distant locations in autumn and spring, without prior knowledge of the biotic and abiotic conditions of these sites. By doing this, we included in our data set the gill samples that were diverse in terms of origin,

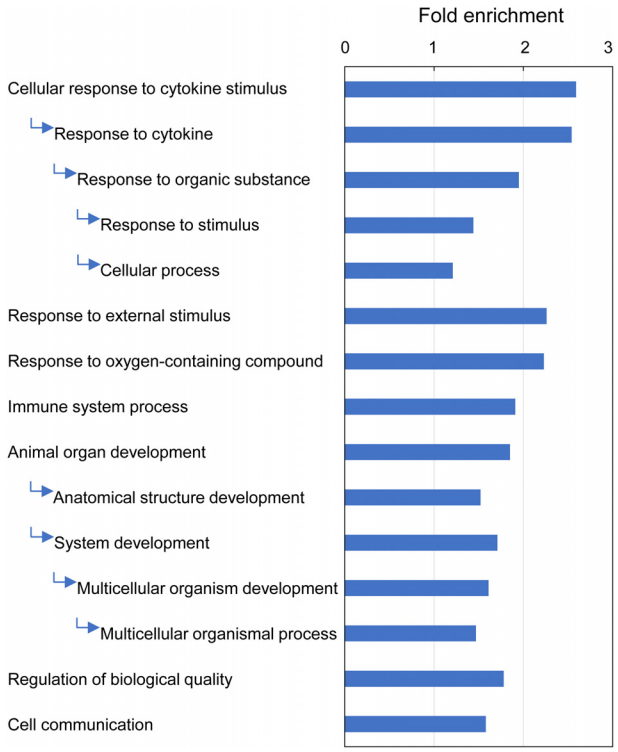


FIGURE 9 | Top GO terms (Biological Process) associated with differentially expressed genes (DEGs) in Atlantic salmon gills with moderate histopathology (sites A and B) vs low histopathology (site C). The analysis was performed on 311 salmon genes mapped to 235 human orthologs, using PANTHER overrepresentation test and Bonferroni-corrected *P*-value < 0.05. The GO terms are sorted by the fold enrichment of the most specific categories, with their parent terms indented directly below. For details and list of corresponding genes see **Supplementary Table 11**.

time of sampling, fish background and the overall environmental milieu (**Table 1**). The multi-site sampling approach to identify the common aspects of the multifactorial gill inflammation has been used before (Kvellestad et al., 2005; Steinum et al., 2010; Gjessing et al., 2019), but not in the context of the transcriptome analysis or underlying molecular mechanisms. Our study is the first to investigate the gill tissue from salmon at three different levels of organization: gross morphology (macroscopic examination), histopathology (microscopic examination) and whole transcriptome (gene expression) (**Figure 1**). We have focused on the gill arch with the highest PGD score per fish, as this arch may provide more accurate information about the general state of the gill health than the arch sampled randomly or the arch sampled because of its position, as practiced elsewhere (Wise et al., 2008; Steinum et al., 2010; Benedicenti et al., 2019; Gjessing et al., 2019). For better integration of the transcriptomic data with the gill gross morphology and histopathology, the tissue fragments used for the total RNA extraction aimed to represent the whole surface area of the gill rather than some specific areas of the tissue (**Figure 1**). This is in contrast to many gill gene expression (qPCR) studies, which tend to focus on either interbranchial lymphoid tissue or visible pathologies of the gill tissue, such as mucoid patches, hyperplastic lesions and lamellar fusions (Pennacchi et al., 2016; Marcos-López et al., 2018; Benedicenti et al., 2019).

PGD Scores as Measures of Gill Health

Effective treatment of salmon gill diseases requires development of diagnostic and prognostic tools that are non-invasive and suitable for a frequent use in the rapidly changing marine environment (Herrero et al., 2018; Gjessing et al., 2019; Rozas-Serri, 2019). We have specifically focused on the applicability of the PGD scores (gross morphology) to reflect gill health and underlying pathology, as this scoring system is already in use as part of routine fish welfare indicator assessment of salmon health in seawater farms worldwide (Gill Health Initiative, 2015; Bloecher et al., 2018). By comparing 43 gill samples with low (1) and high (3) PGD scores across three marine production sites in Scotland, we found that these two groups of gill tissue classified as different at the macroscopic level (PGD 1 and PGD 3) were in fact indistinguishable at the level of whole-transcriptome gene expression (**Figure 3A**) and also indistinguishable at the level of microscopic histopathology (**Figure 4A**). Furthermore, we could not identify any single gene that was expressed differently between the two groups (**Table 4**). Our results strongly suggest that the changes in gross morphology were not supported by the changes in gene expression or histopathology. The lack of detectable transcriptomic and/or histopathological changes associated with the progression of the PGD scores questions the suitability of the gross scoring system as diagnostic and prognostic tools to monitor and control both existing and emerging gill diseases. Instead, the PGD scores are good proxies for monitoring changes in macroscopic gill surface area available for respiration, the knowledge of which may be essential for fish health management and husbandry practices in aquaculture settings.

The PGD scoring system of salmon gill is based on the 6 macroscopic grades, from 0 (no visual pathology) to 5 (severe visual pathology) (Gill Health Initiative, 2015; Bloecher et al., 2018). It is important to realize that our comparison was done on the groups of tissue that differed only by two grades (PGD 1 and PGD 3) and were in the middle range of the spectrum, with a bias towards less damaged gills. Although our intention was to compare the transcriptome and histopathology of gill tissues with the full spectrum of PGD scores (e.g., PGD 0 and PGD 5) from the same farm location, fish with such a large range in gill gross morphology were not found during the sampling events, not only within the three production sites described in the current study (**Table 2**), but also at the four other locations (Scottish Sea Farms) visited by us in years 2017–2019 (R. Bickerdike and S. A. M. Martin, unpublished data). It is unusual for the salmon farmed in the sea to have all 8 gill arches scored as PGD 0 (apart from the first few weeks following the transfer from the freshwater facility), while the macroscopic gill damage classified as PGD 5 is very rare and would not be expected to be found until later in the production cycle, and as a result of a specific acute event or from cumulative significant environmental insults over a period of time (R. Bickerdike, personal communication).

TABLE 6 | Top genes altered in Atlantic salmon gills with moderate histopathology (sites A and B) vs low histopathology (site C), based on FDR < 0.01 and absolute Log_2 FC > 2.

Atlantic salmon genes				Human gene orthologs (HGNC identifiers)		
Gene ID	**Transcript ID**	**Gene name**	**Log_2 FC[1]**	**Symbol**	**Name**	**Log_2 FC[2]**
LOC106604507	XM_014199172.1	ribonuclease-like 3	4.6	ANG	angiogenin	4.6
LOC106598253	XM_014189304.1	ribonuclease-like 3	4.6	ANG	angiogenin	
LOC100196060	NM_001141089.1	angiogenin-1	4.5	ANG	angiogenin	
LOC100196525	NM_001141554.1	chymotrypsin-like	4.2	CTRL	chymotrypsin like	4.2
LOC106577309	XM_014155258.1	glutathione peroxidase 6-like	3.9	GPX2	glutathione peroxidase 2	3.9
LOC106561635	XM_014125761.1	ladderlectin-like	3.8	REG1B	regenerating family member 1 beta	3.8
LOC106577833	XM_014156195.1	interleukin-8-like	3.4	CXCL9	C-X-C motif chemokine ligand 9	3.4
LOC106601490	XM_014193713.1	RING finger protein 208-like	3.4	RNF152	ring finger protein 152	3.3
LOC106601491	XM_014193714.1	RING finger protein 186-like	3.2	RNF152	ring finger protein 152	
LOC106566533	XM_014134632.1	complement C1q tumor necrosis factor-related protein 3-like	3.2	C1QTNF3	C1q and TNF related 3	3.1
LOC106566537	XM_014134640.1	complement C1q tumor necrosis factor-related protein 3-like	3.2	C1QTNF3	C1q and TNF related 3	
LOC106566534	XM_014134633.1	complement C1q tumor necrosis factor-related protein 3-like	3.0	C1QTNF3	C1q and TNF related 3	
LOC106567034	XM_014135842.1	complement C1q-like protein 4	2.8	C1QL4	complement C1q like 4	2.8
LOC106573018	XM_014147615.1	carboxypeptidase A1-like	2.9	CPA1	carboxypeptidase A1	2.7
LOC106598577	XM_014189609.1	carboxypeptidase A1-like	2.8	CPA1	carboxypeptidase A1	
LOC100195857	NM_001140886.1	carboxypeptidase A1	2.6	CPA1	carboxypeptidase A1	
LOC106561558	XM_014125655.1	carboxypeptidase A1-like	2.3	CPA1	carboxypeptidase A1	
LOC100136358	XM_014214975.1	nitric oxide synthase 2	2.4	NOS2	nitric oxide synthase 2	2.4
LOC106594149	XM_014185514.1	C-type lectin domain family 4 member E-like	2.4	CLEC4E	C-type lectin domain family 4 member E	2.4
LOC106591222	XM_014182433.1	zymogen granule membrane protein 16-like	2.8	ZG16	zymogen granule protein 16	2.4
LOC106562680	XM_014127661.1	zymogen granule membrane protein 16-like	2.5	ZG16	zymogen granule protein 16	
LOC106562681	XM_014127663.1	zymogen granule membrane protein 16-like	2.4	ZG16	zymogen granule protein 16	
LOC106584756	XM_014170303.1	zymogen granule membrane protein 16-like	1.9	ZG16	zymogen granule protein 16	
LOC106567571	XM_014137044.1	H-2 class II histocompatibility antigen, A-K alpha chain-like	2.4	HLA-DPA1	major histocompatibility complex, class II, DP alpha 1	2.4
LOC106585685	XM_014172160.1	neurotrophic receptor tyrosine kinase 2	2.3	NTRK2	neurotrophic receptor tyrosine kinase 2	2.3
LOC100286614	NM_001146553.1	high choriolytic enzyme 2	2.7	ASTL	astacin like metalloendopeptidase	2.2
LOC106586984	XM_014174772.1	high choriolytic enzyme 1-like	2.7	ASTL	astacin like metalloendopeptidase	
LOC100195775	NM_001140804.1	high choriolytic enzyme 1	1.3	ASTL	astacin like metalloendopeptidase	
LOC106591797	XM_014183059.1	protein disulfide-isomerase-like	2.2	PDIA2	protein disulfide isomerase family A member 2	2.2
LOC106595494	XM_014186866.1	protein disulfide-isomerase A2-like	2.1	PDIA2	protein disulfide isomerase family A member 2	
LOC100196492	NM_001141521.2	cysteine dioxygenase type 1	2.1	CDO1	cysteine dioxygenase type 1	2.1

(Continued)

TABLE 6 | Continued

Atlantic salmon genes				Human gene orthologs (HGNC identifiers)		
Gene ID	**Transcript ID**	**Gene name**	**Log_2 FC[1]**	**Symbol**	**Name**	**Log_2 FC[2]**
LOC100136458	NM_001123590.1	tumor necrosis factor alpha-2 precursor	2.1	LTA	lymphotoxin alpha	2.1
LOC106599048	XM_014190103.1	retinol-binding protein 1-like	2.1	RBP1	retinol binding protein 1	2.1
LOC106573692	XM_014148970.1	thyrotropin-releasing hormone receptor-like	2.1	TRHR	thyrotropin releasing hormone receptor	2.1
LOC106589989	XM_014180468.1	myosin-7-like	2.1	MYH7	myosin heavy chain 7	2.1
LOC106581492	XM_014163567.1	P2Y purinoceptor 13-like	2.1	P2RY12	purinergic receptor P2Y12	2.1
LOC106581219	XM_014163130.1	collagenase 3-like	2.5	MMP13	matrix metallopeptidase 1	2.1
LOC106613110	XM_014215041.1	collagenase 3-like	1.7	MMP13	matrix metallopeptidase 1	
LOC106581616	XM_014163747.1	aconitate decarboxylase 1	2.0	ACOD1	aconitate decarboxylase 1	2.0
LOC106609709	XM_014208680.1	complement C1q-like protein 2	2.3	C1QL2	complement C1q like 2	2.0
LOC106596487	XM_014187765.1	complement C1q-like protein 2	2.3	C1QL2	complement C1q like 2	
LOC106592632	XM_014183971.1	complement C1q-like protein 2	2.3	C1QL2	complement C1q like 2	
LOC106601034	XM_014192902.1	complement C1q-like protein 2	1.1	C1QL2	complement C1q like 2	
LOC106609915	XM_014208958.1	sialic acid synthase-like	−2.0	NANS	N-acetylneuraminate synthase	−2.1
LOC106594767	XM_014186155.1	sialic acid synthase-like	−2.1	NANS	N-acetylneuraminate synthase	
LOC106609889	XM_014208923.1	sialic acid synthase-like	−2.2	NANS	N-acetylneuraminate synthase	
LOC106562772	XM_014127771.1	neuronal acetylcholine receptor subunit alpha-7-like	−2.1	CHRNA7	cholinergic receptor nicotinic alpha 7 subunit	−2.1
LOC106566856	XM_014135351.1	succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial-like	−2.3	SDHB	succinate dehydrogenase complex iron sulfur subunit B	−2.3
LOC106603364	XM_014196939.1	sodium channel subunit beta-4-like	−2.4	SCN4B	sodium voltage-gated channel beta subunit 4	−2.4
LOC106594819	XM_014186204.1	corticosteroid 11-beta-dehydrogenase isozyme 2-like	−2.4	HSD11B2	hydroxysteroid 11-beta dehydrogenase 2	−2.4
LOC100194860	NM_001139889.1	Gamma-aminobutyric acid receptor subunit delta	−2.4	GABRD	gamma-aminobutyric acid type A receptor delta subunit	−2.4
LOC106607156	XM_014203794.1	heat-stable enterotoxin receptor-like	−2.4	GUCY2C	guanylate cyclase 2C	−2.4
LOC106597204	XM_014188425.1	pinopsin-like	−2.5	OPN3	opsin 3	−2.5
LOC106585409	XM_014171569.1	spermatid perinuclear RNA-binding protein-like	−2.6	STRBP	spermatid perinuclear RNA binding protein	−2.6
LOC106596671	XM_014187940.1	protein APCDD1-like	−2.9	APCDD1	APC down-regulated 1	−2.9
LOC106586762	XM_014174368.1	growth hormone-regulated TBC protein 1-A-like	−2.9	GRTP1	growth hormone regulated TBC protein 1	−2.9
LOC106606833	XM_014203242.1	uncharacterized LOC106606833	−3.1	MSLNL	mesothelin like	−3.1

The Atlantic salmon genes are grouped by the common human ortholog (BLASTX, E-value < 0.00001, top hit) and sorted from the highest to the lowest Log_2 FC associated with the HGNC identifier. [1]Mean Log_2 FC for A vs C and B vs C comparisons, as detailed in ***Supplementary Table 7****; [2]mean expression of salmon genes mapped to the same human ortholog.*

Discrepancies between macroscopic and microscopic examination of the gill tissue have been reported by previous studies, mainly in the context of AGD (Pennacchi et al., 2016). For example, it has been shown that the AGD scoring system (gross morphology) may not be a reliable method of confirming the disease in cases of light severity of AGD (Clark and Nowak, 1999; Zilberg et al., 2001). This is because small AGD-associated lesions, which affect < 10 lamellae and are easy to detect under the microscope, are typically overlooked during the visual inspection of the gill (Adams et al., 2004). As a result, fish exposed to *Neoparamoeba perurans* and then classified as clear of the AGD symptoms during the macroscopic gill scoring for AGD may in fact need to be re-classified at the level of histopathological examination (Wynne et al., 2008b). Further complexity to the diagnostic problems is added when the gill disease is multifactorial (Wise et al., 2008; Gjessing et al., 2019; Noguera et al., 2019). More broadly, the poor diagnostic and prognostic value of the PGD scores demonstrated in our study is consistent with the limited applicability of the gross morphology to diagnose complex diseases in livestock and humans, which typically need extensive histopathology and molecular profiling to confirm and prognosticate (Hoffmann et al., 2009; Ahmed and Abedalthagafi, 2016; Mobadersany et al., 2018).

Histopathology-Directed Analysis of Gill Transcriptome Between Sites

By sampling sea farmed Atlantic salmon at three production sites (A, B and C) in Scotland, we demonstrated that the gill samples from different sites had different histopathology (**Figure 4B**) and different transcriptomic profiles (**Figure 3B**), with 240 to 1360 genes identified as differentially expressed between the sites (**Table 4**). The drivers of this pronounced site-to-site variability in gill histopathology and transcriptome are unknown and may reflect the differences in (1) fish cohort, including their hatchery origin, past and present husbandry practices and overall health status, (2) local biotic conditions (e.g., diversity of pathogen), (3) local abiotic conditions (e.g., sea water temperature) and (4) interplay and interactions between all the above factors. Identification of these drivers requires more rigorous spatial and temporal sampling regime (Jokinen et al., 2012; Maestrini and Basso, 2018) and is beyond the scope of the current study.

The NMDS plots revealed that the gill samples from site C were different (i.e., clearly separated from sites A and B) not only at the level of their transcriptome (**Figure 3B**), but also in terms of their histopathology (**Figure 4B**). Because the NMDS ordination of the histopathological data is based on the same set of parameters (i.e., LH, LF, LO, CA, in, eg, cc, cd, ib, ch, pp and op) across all fish, lower NMDS1 values of gill samples from site C reflect lower scores of the contributing parameters (lower grade inflammation) than higher NMDS1 values of gill samples from sites A and B (**Figure 4B**, **Supplementary Table 5**, and **Supplementary Figure 2**). In contrast, the NMDS ordination of the transcriptomic data (**Figure 3B**) is based on the expression of 500 genes that are most divergent between each pair of fish (amounting to 903 sets of genes for all combinations of 43 fish in total) rather than being performed on the same set of genes across all fish (Robinson et al., 2010). Thus, the higher leading Log_2 FC dim 1 values of gill samples from site C refer to their generally higher magnitude of transcriptomic changes rather than to the specific set of genes that separates them from sites A and B, with latter requiring more detailed analysis of the existing data to perform.

Because gill samples were more different between sites than within sites (**Figures 3B**, **4B**), we performed the histopathology-directed analysis of gill transcriptome on all fish belonging to each site rather than on the individual fish within or between sites (**Figure 5**). Specifically, we classified all fish from site C as having gills with low histopathology, while all fish from site A and independently all fish from site B as having gills with moderate histopathology. Comparing the two independent sets of the moderately inflamed gill transcriptomes (A and B) with the low inflammation gill transcriptome (C), in the A vs C and B vs C contrasts, led to the identification of the common DEGs that are likely to constitute the gene expression patterns of multifactorial gill inflammation. Similar approaches utilizing independent sampling and site-to-site comparisons to establish common transcriptomic responses to a range of environmental factors have been successfully applied in studies ranging from aquatic toxicology (Oleksiak, 2008; Defo et al., 2018) to human medicine (Mueller et al., 2009; Rohart et al., 2017).

Gene Expression Patterns of Multifactorial Gill Disease

Transcriptome profiling by RNA-seq has become a powerful tool for identification of genes and molecular pathways involved in the progression from health to disease in farmed fish (Sun et al., 2012; Martin and Król, 2017; Houston and Macqueen, 2019; Ronza et al., 2019). Combining RNA-seq assays with an independent sampling regime is especially important for understanding multifactorial diseases, because it allows for the potential identification of the common gene expression patterns of the disease that is essential for developing treatment and prevention strategies (Cookson et al., 2009; Bhuju et al., 2012). This approach has for example been used in the context of non-specific inflammatory diseases in fish intestine (reviewed in Martin et al., 2016). The present study is the first to focus on the transcriptomic patterns of multifactorial gill disease, a condition increasingly prevalent in salmon sea farms.

Transcriptomic comparison of the two independent sets of moderate histopathology gill samples (sites A and B) with the low histopathology gill samples (site C) led to the identification of 462 common DEGs in total, 311 of which were protein-coding transcripts mapped to 235 human orthologs (**Table 5**). The identification of these genes as part of the gene expression patterns of multifactorial gill disease was further supported by their uniform direction of changes in the expression levels (**Figure 5**). Subsequent functional analysis of these genes by IPA and GO pointed towards a range of immune responses as the most dominant feature of the moderate histopathology gills.

At the gene level, the immune responses were driven by pro-inflammatory cytokines (IL17F, CXCL9, CXCL10, CCL4L1 and TNF superfamily members: LTA and TNFSF14), cytokine receptors (TNFRSF1B, TNFRSF6B and CXCR1) and regulators of cytokine expression and signaling (C1QTNF3, SOCS1 and SOCS3), all of which were upregulated apart from TNFSF14 (**Supplementary Table 7**). The proteins encoded by IL17 family genes have been shown to stimulate the production of several other cytokines, including IL6 and IL8 in splenocytes of rainbow trout using recombinant IL17A/F2a (Monte et al., 2013) as well as IL1β, IL6, IL8 and TNF-α in head kidney leukocytes of grass carp using recombinant IL17A/F1 (Du et al., 2015). Furthermore, the emerging role of IL17 cytokines in gill mucosal immunity is supported by reports showing differential expression of these genes following the exposure to *Aeromonas hydrophila* in common carp (Dong et al., 2019) and after challenge with *Ichthyophthirius multifiliis* in rainbow trout (Syahputra et al., 2019). According to our BLASTX results, CXCL9 represents salmon IL8, whose high expression in our data set (Log_2 FC = 3.4, **Supplementary Table 7**) is consistent with upregulation of IL17F (Zou and Secombes, 2016). The implication of IL8 in gill immune responses has been previously demonstrated in Atlantic cod (Caipang et al., 2010) and rainbow trout (Olsen et al., 2011; Santana et al., 2016). The protein encoded by LTA (salmon TNF-α2) mediates a large variety of inflammatory, immunostimulatory and antiviral responses in mammals (Upadhyay and Fu, 2013), showing induction at the level of gene expression in gills of Atlantic bluefin tuna during natural infection with Digenea (Pleić et al., 2015), but not in Atlantic salmon following the infection with *Neoparamoeba* spp. to induce AGD (Morrison et al., 2007). Cytokines were also overwhelmingly present in our functional analysis of gene expression patterns in the moderately inflamed gills. Specifically, 6 of the top 13 IPA canonical pathways were related to cytokine signaling (**Figure 6** and **Supplementary Table 8**) and 10 of the top 15 IPA upstream regulators were predicted to be cytokines (**Figure 7** and **Supplementary Table 9**). The presence of cytokines among DEGs and also as upstream regulators is consistent with a cytokine signaling cascade, during which one cytokine stimulates its target cells to produce another cytokines (Tisoncik et al., 2012). As a result of cytokine presence among DEGs, 6 of the top 13 IPA downstream effects were predicted to be immunological and inflammatory diseases (**Figure 8** and **Supplementary Table 10**). Altogether, our results identified cytokine-driven immune response as a hallmark of multifactorial gill disease.

Besides the immune response, the multifactorial gill disease from sites A and B had a common transcriptomic profile indicative of tissue damage and repair. Part of the tissue damage could be inflicted by high levels of nitric oxide (NO), suggested by upregulation of NOS2 (Log_2 FC = 2.4, **Supplementary Table 7**). This gene encodes a NO synthase that is inducible by a combination of lipopolysaccharide (LPS) and certain cytokines (Okamoto et al., 2004). Although NO is produced as a first-line defense against invading pathogens, its strong cytotoxic effects may also damage the tissue of the host (Abramson et al., 2001). Furthermore, high levels of NO may contribute to oxidative stress through production of reactive oxygen species (ROS) (Girouard et al., 2009), which could potentially explain the upregulation of GPX2 in our data set (Log_2 FC = 3.9, **Supplementary Table 7**). The protein encoded by this gene belongs to the glutathione peroxidase family, members of which catalyze the reduction of organic hydroperoxides and hydrogen peroxide (H_2O_2) by glutathione, and thereby protect cells against oxidative damage (Brigelius-Flohé and Maiorino, 2013). The potential role of GPX2 in the antioxidant defense against increased ROS at gill mucosal surfaces has been recently discussed in the context of exposing Atlantic salmon to peracetic acid (Soleng et al., 2019). It is well established that the repair of inflamed tissue requires efficient removal of damaged cells through controlled cell death (apoptosis) and concurrent cell proliferation to regenerate damaged structures and build up lost tissue, with both processes closely linked to the activity of cysteine-dependent aspartate-directed proteases (caspases) (Fogarty and Bergmann, 2017). In our study, both CASP14 (salmon caspase-14-like) and other genes associated with apoptosis and autophagy (RNF152, CTSL, CTSG, CTSV, RAB32 and TGM2) were upregulated, pointing towards ongoing tissue repair in the moderately inflamed gills (**Supplementary Table 7**). The increased activity of caspases during gill inflammation has been previously documented in rainbow trout challenged with *Aeromonas salmonicida* (Rojas et al., 2015) and grass carp after infection with *Flavobacterium columnare* (Chen et al., 2019). Because of the complexity and size of gill vasculature, the repair of the gill tissue requires extensive vasculogenesis (formation of new blood vessels from vascular precursor cells), angiogenesis (process of outgrowing vessels from the existing vasculature) and arteriogenesis (remodeling of arteries where collateral arterial anastomoses undergo abluminal expansion) (Shi et al., 2017). One of the most potent mediators of new blood vessel formation is angiogenin (Hoang and Raines, 2017), represented in our data set by ANG (salmon angiogenin-1 and ribonuclease-like 3), which was the gene with the highest level of upregulation (Log_2 FC = 4.6, fold change ∼24) in the moderate histopathology vs low histopathology gills (**Table 6** and **Supplementary Table 7**). This result is consistent with the earlier study conducted by Valdenegro-Vega et al. (2014), showing increased protein levels of angiogenin (fold change ∼12) in the gills of Atlantic salmon following four successive infections with *Neoparamoeba perurans*. Among other genes involved in angiogenesis were ADM, BMP10 and VCAN, all upregulated in the moderately inflamed gills (**Supplementary Table 7**). Overall, 5 of top 13 IPA canonical pathways (**Figure 6** and **Supplementary Table 8**) and 5 of top 13 IPA downstream effects (**Figure 8** and **Supplementary Table 10**) were associated with cell death and proliferation related to tissue damage and repair, highlighting their importance in multifactorial gill disease.

Conclusion and Future Directions

The gill is central for understanding the impacts of climate change on fish health. Recent increases in gill pathologies in sea farmed Atlantic salmon highlight the need to establish the molecular basis of multifactorial gill disease (frequently referred to as PGD, PGI or CGD) to improve diagnosis and preventive

measures of this condition. To ensure the multifactorial etiology of gill disease, we sampled Atlantic salmon from three different production sites in Scotland and then examined the gill tissue at three different levels of organization: gross morphology with the use of PGD scores (macroscopic examination), whole transcriptome (gene expression by RNA-seq) and histopathology (microscopic examination). By exploring the association between gill transcriptome and gill gross morphology (PGD scores), we found that the PGD scores were less effective in providing a graded assessment of gill health status than expected, and they did not convey any information about the underlying pathology and/or tissue deterioration. In contrast, integration of the gill RNA-seq data with the gill histopathology enabled us to identify common gene expression patterns associated with multifactorial gill disease. We demonstrated that the gene expression patterns associated with multifactorial gill disease were dominated by two processes: a range of immune responses driven by pro-inflammatory cytokines and the events associated with tissue damage and repair, driven by caspases and angiogenin.

Previous studies of the gill inflammation have typically focused on single-cause pathologies (such as AGD), in the experiments performed on fish exposed to the controlled environment of the closed-system tanks. Although these studies are very important from a mechanistic point of view of specific pathologies, they may have limited relevance for the fish facing a changing environment at farm locations, e.g., rising sea water temperature, deoxygenation and surges of existing and emerging pathogens with the complexity of co-infections, coupled with husbandry practices. Performing more transcriptomic studies in the field rather than in the lab would benefit both academia and industry. The multi-site approach is also important because in this study, we demonstrated large and significant effects of sampling site (sensu lato) on gill transcriptome and histopathology. The drivers of this site-to-site variability are presently unknown and require more specific sampling regime.

Our histopathology-directed analysis of gill transcriptomes identified in total 462 genes that we claim to constitute the gene expression profile of multifactorial gill inflammation, including 354 protein-coding genes, 35 immunoglobulin gene segments, 15 pseudogenes and 58 non-coding RNAs. It is, however, important to remember that our analysis was based on the gill samples from three production sites in one specific part of the world (coastal waters of Scotland), with sampling events covering October, November and March. Substantial amount of work is therefore needed to test the association of these genes with multifactorial gill diseases at different times of year in Scotland and worldwide. The diagnostic and therapeutic value of these transcripts is currently unknown and require further studies.

DATA AVAILABILITY STATEMENT

The RNA-seq data have been deposited in the ArrayExpress repository (http://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-8855. The R and Python scripts are available on request to AD and SS, respectively.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because the sampling of fish was carried out under established protocols for routine health assessments in accordance with RSPCA Assured Welfare Standards for farmed Atlantic salmon and under supervision of the company veterinarian.

AUTHOR CONTRIBUTIONS

SM, AD, EK, VV, KG, and RB conceived and designed the study, interpreted results, and gave final approval of the manuscript. All authors (apart from SS and AD) were involved in sampling. EK extracted RNA, performed IPA analysis, and drafted the manuscript with AD, who oversaw bioinformatics and statistical analysis. SS mapped salmon genes to human orthologs to enable functional analysis of gene expression. PN performed histopathological examination of samples, generated scores, and interpreted results. EC provided support in data exploration. All authors discussed and commented on the manuscript.

FUNDING

The study was supported by the Scottish Aquaculture Innovation Centre (SAIC grant SL 2017 08, 'Nutritional Aspects of Gill Disease in Atlantic Salmon'). The authors declare that this study received funding from BioMar AS and Scottish Sea Farms (SSF). The funders had the following involvement with the study: contribution to the experimental design, selection of sampling sites, participation in sampling, contribution to the interpretation of the results, and approval of the manuscript.

ACKNOWLEDGMENTS

We thank SSF farm personnel for accommodating our research, performing gross morphology scoring and helping with sampling. Edinburgh Genomics is partly supported through core grants from NERC (R8/H10/56), MRC (MR/K001744/1), and BBSRC (BB/J004243/1). We also thank two reviewers for their comments on the earlier draft of the article. EK dedicates this paper to her late mother, Irena Król.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00610/full#supplementary-material

REFERENCES

Abramson, S. B., Amin, A. R., Clancy, R. M., and Attur, M. (2001). The role of nitric oxide in tissue destruction. *Best Pract. Res. Clin. Rheumatol.* 15, 831–845. doi: 10.1053/berh.2001.0196

Adams, M. B., Ellard, K., and Nowak, B. F. (2004). Gross pathology and its relationship with histopathology of amoebic gill disease (AGD) in farmed Atlantic salmon, *Salmo salar* L. *J. Fish Dis.* 27, 151–161. doi: 10.1111/j.1365-2761.2004.00526.x

Ahmed, A. A., and Abedalthagafi, M. (2016). Cancer diagnostics: the journey from histomorphology to molecular profiling. *Oncotarget* 7, 58696–58708. doi: 10.18632/oncotarget.11061

Akbarzadeh, A., Günther, O. P., Houde, A. L., Li, S., Ming, T. J., Jeffries, K. M., et al. (2018). Developing specific molecular biomarkers for thermal stress in salmonids. *BMC Genomics* 19:749. doi: 10.1186/s12864-018-5108-9

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed February 11, 2020).

Anttila, K., Lewis, M., Prokkola, J. M., Kanerva, M., Seppänen, E., Kolari, I., et al. (2015). Warm acclimation and oxygen depletion induce species-specific responses in salmonids. *J. Exp. Biol.* 218, 1471–1477. doi: 10.1242/jeb.119115

Bancroft, J. D., and Gamble, M. (2007). *Theory and Practice of Histological Techniques*, 6th Edn. London: Churchill Livingstone.

Bayliss, S. C., Verner-Jeffreys, D. W., Bartie, K. L., Aanensen, D. M., Sheppard, S. K., Adams, A., et al. (2017). The promise of whole genome pathogen sequencing for the molecular epidemiology of emerging aquaculture pathogens. *Front. Microbiol.* 8:121. doi: 10.3389/fmicb.2017.00121

Benedicenti, O., Pottinger, T. G., Collins, C., and Secombes, C. J. (2019). Effects of temperature on amoebic gill disease development: does it play a role? *J. Fish Dis.* 42, 1241–1258. doi: 10.1111/jfd.13047

Bhuju, S., Aranday-Cortes, E., Villarreal-Ramos, B., Xing, Z., Singh, M., and Vordermeier, H. M. (2012). Global gene transcriptome analysis in vaccinated cattle revealed a dominant role of IL-22 for protection against bovine tuberculosis. *PLoS Pathog.* 8:e1003077. doi: 10.1371/journal.ppat.1003077

Bloecher, N., Powell, M., Hytterød, S., Gjessing, M., Wiik-Nielsen, J., Mohammad, S. N., et al. (2018). Effects of cnidarian biofouling on salmon gill health and development of amoebic gill disease. *PLoS One* 13:e0199842. doi: 10.1371/journal.pone.0199842

Boison, S. A., Gjerde, B., Hillestad, B., Makvandi-Nejad, S., and Moghadam, H. K. (2019). Genomic and transcriptomic analysis of amoebic gill disease resistance in Atlantic salmon (*Salmo salar* L.). *Front. Genet.* 10:68. doi: 10.3389/fgene.2019.00068

Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., et al. (2018). Declining oxygen in the global ocean and coastal waters. *Science* 359:eaam7240. doi: 10.1126/science.aam7240

Brigelius-Flohé, R., and Maiorino, M. (2013). Glutathione peroxidases. *Biochim. Biophys. Acta* 1830, 3289–3303. doi: 10.1016/j.bbagen.2012.11.020

Bruno, D. W., Noguera, P. A., and Poppe, T. T. (2013). *A Colour Atlas of Salmonid Diseases*, 2nd Edn. Dordrecht: Springer.

Burge, C. A., Eakin, C. M., Friedman, C. S., Froelich, B., Hershberger, P. K., Hofmann, E. E., et al. (2014). Climate change influences on marine infectious diseases: implications for management and society. *Annu. Rev. Mar. Sci.* 6, 249–277. doi: 10.1146/annurev-marine-010213-135029

Cadoret, K., Bridle, A. R., Leef, M. J., and Nowak, B. F. (2013). Evaluation of fixation methods for demonstration of *Neoparamoeba perurans* infection in Atlantic salmon, *Salmo salar* L., gills. *J. Fish Dis.* 36, 831–839.

Caipang, C. M., Lazado, C. C., Brinchmann, M. F., and Kiron, V. (2010). Infection-induced changes in expression of antibacterial and cytokine genes in the gill epithelial cells of Atlantic cod, *Gadus morhua* during incubation with bacterial pathogens. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 156, 319–325. doi: 10.1016/j.cbpb.2010.04.009

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Chen, K., Zhou, X. Q., Jiang, W. D., Wu, P., Liu, Y., Jiang, J., et al. (2019). Dietary phosphorus deficiency caused alteration of gill immune and physical barrier function in the grass carp (*Ctenopharyngodon idella*) after infection with *Flavobacterium columnare*. *Aquaculture* 506, 1–13. doi: 10.1016/j.aquaculture.2019.03.018

Clark, A., and Nowak, B. F. (1999). Field investigations of amoebic gill disease in Atlantic salmon, *Salmo salar* L., in Tasmania. *J. Fish Dis.* 22, 433–443. doi: 10.1046/j.1365-2761.1999.00175.x

Cookson, W., Liang, L., Abecasis, G., Moffatt, M., and Lathrop, M. (2009). Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* 10, 184–194. doi: 10.1038/nrg2537

Defo, M. A., Douville, M., Giraudo, M., Brodeur, P., Boily, M., and Houde, M. (2018). RNA-sequencing to assess the health of wild yellow perch (*Perca flavescens*) populations from the St. Lawrence River, Canada. *Environ. Pollut.* 243, 1657–1668. doi: 10.1016/j.envpol.2018.09.133

Diaz, R. J., and Rosenberg, R. (2008). Spreading dead zones and consequences for marine ecosystems. *Science* 321, 926–929. doi: 10.1126/science.1156401

Dodt, M., Roehr, J. T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR—flexible barcode and adapter processing for next-generation sequencing platforms. *Biology* 1, 895–905. doi: 10.3390/biology1030895

Dong, C., Kong, S., Zheng, X., Zhang, J., Nie, G., Li, X., et al. (2019). Genome-wide identification of interleukin-17 (IL17) in common carp (*Cyprinus carpio*) and its expression following *Aeromonas hydrophila* infection. *Gene* 686, 68–75. doi: 10.1016/j.gene.2018.10.038

Downes, J. K., Yatabe, T., Marcos-Lopez, M., Rodger, H. D., MacCarthy, E., O'Connor, I., et al. (2018). Investigation of co-infections with pathogens associated with gill disease in Atlantic salmon during an amoebic gill disease outbreak. *J. Fish Dis.* 41, 1217–1227. doi: 10.1111/jfd.12814

Du, L., Feng, S., Yin, L., Wang, X., Zhang, A., Yang, K., et al. (2015). Identification and functional characterization of grass carp IL-17A/F1: an evaluation of the immunoregulatory role of teleost IL-17A/F1. *Dev. Comp. Immunol.* 51, 202–211. doi: 10.1016/j.dci.2015.03.014

English, C. J., Tyml, T., Botwright, N. A., Barnes, A. C., Wynne, J. W., Lima, P. C., et al. (2019). A diversity of amoebae colonise the gills of farmed Atlantic salmon (*Salmo salar*) with amoebic gill disease (AGD). *Eur. J. Protistol.* 67, 27–45. doi: 10.1016/j.ejop.2018.10.003

Evans, D. H., Piermarini, P. M., and Choe, K. P. (2005). The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. *Physiol Rev.* 85, 97–177. doi: 10.1152/physrev.00050.2003

Fogarty, E., and Bergmann, A. (2017). Killers creating new life: caspases drive apoptosis-induced proliferation in tissue repair and disease. *Cell Death Differ.* 24, 1390–1400. doi: 10.1038/cdd.2017.47

Gill Health Initiative (2015). *Gill Diseases in Scotland by Iain Berrill (SSPO)*. Available online at: https://www.fishhealth.ie/fhu/health-surveillance/aquaplan-fish-health-management-ireland/gill-health-initiative (accessed February 11, 2020).

Girouard, H., Wang, G., Gallo, E. F., Anrather, J., Zhou, P., Pickel, V. M., et al. (2009). NMDA receptor activation increases free radical production through nitric oxide and NOX2. *J. Neurosci.* 29, 2545–2552. doi: 10.1523/jneurosci.0133-09.2009

Gjessing, M. C., Steinum, T., Olsen, A. B., Lie, K. I., Tavornpanich, S., Colquhoun, D. J., et al. (2019). Histopathological investigation of complex gill disease in sea farmed Atlantic salmon. *PLoS One* 14:e0222926. doi: 10.1371/journal.pone.0222926

Gjessing, M. C., Thoen, E., Tengs, T., Skotheim, S. A., and Dale, O. B. (2017). Salmon gill poxvirus, a recently characterized infectious agent of multifactorial gill disease in freshwater- and seawater-reared Atlantic salmon. *J. Fish Dis.* 40, 1253–1265. doi: 10.1111/jfd.12608

Gunnarsson, G. S., Karlsbakk, E., Blindheim, S., Plarre, H., Imsland, A. K., Handeland, S., et al. (2017). Temporal changes in infections with some pathogens associated with gill disease in farmed Atlantic salmon (*Salmo salar* L). *Aquaculture* 468, 126–134. doi: 10.1016/j.aquaculture.2016.10.011

Herrero, A., Thompson, K. D., Ashby, A., Rodger, H. D., and Dagleish, M. P. (2018). Complex gill disease: an emerging syndrome in farmed Atlantic salmon (*Salmo salar* L.). *J. Comp. Pathol.* 163, 23–28. doi: 10.1016/j.jcpa.2018.07.004

Hoang, T. T., and Raines, R. T. (2017). Molecular basis for the autonomous promotion of cell proliferation by angiogenin. *Nucleic Acids Res.* 45, 818–831. doi: 10.1093/nar/gkw1192

Hoffmann, B., Beer, M., Reid, S. M., Mertens, P., Oura, C. A. L., van Rijn, P. A., et al. (2009). A review of RT-PCR technologies used in veterinary virology

and disease control: sensitive and specific diagnosis of five livestock diseases notifiable to the World Organisation for Animal Health. *Vet. Microbiol.* 139, 1–23. doi: 10.1016/j.vetmic.2009.04.034

Houde, A. L. S., Akbarzadeh, A., Günther, O. P., Li, S., Patterson, D. A., Farrell, A. P., et al. (2019). Salmonid gene expression biomarkers indicative of physiological responses to changes in salinity and temperature, but not dissolved oxygen. *J. Exp. Biol.* 222:jeb198036. doi: 10.1242/jeb.198036

Houston, R. D., and Macqueen, D. J. (2019). Atlantic salmon (*Salmo salar* L.) genetics in the 21st century: taking leaps forward in aquaculture and biological understanding. *Anim. Genet.* 50, 3–14. doi: 10.1111/age.12748

Jokinen, C. C., Edge, T. A., Koning, W., Laing, C. R., Lapen, D. R., Miller, J., et al. (2012). Spatial and temporal drivers of zoonotic pathogen contamination of an agricultural watershed. *J. Environ. Qual.* 41, 242–252. doi: 10.2134/jeq2011.0203

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kintner, A., and Brierley, A. S. (2019). Cryptic hydrozoan blooms pose risks to gill health in farmed North Atlantic salmon (*Salmo salar*). *J. Mar. Biol. Assoc. U.K.* 99, 539–550. doi: 10.1017/s002531541800022x

Koppang, E. O., Kvellestad, A., and Fischer, U. (2015). "Fish mucosal immunity: gill," in *Mucosal Health in Aquaculture*, eds B. H. Beck and E. Peatman (Amsterdam: Academic Press), 93–133. doi: 10.1016/b978-0-12-417186-2.00005-4

Król, E., Douglas, A., Tocher, D. R., Crampton, V. O., Speakman, J. R., Secombes, C. J., et al. (2016). Differential responses of the gut transcriptome to plant protein diets in farmed Atlantic salmon. *BMC Genomics* 17:156. doi: 10.1186/s12864-016-2473-0

Kvellestad, A., Falk, K., Nygaard, S. M., Flesja, K., and Holm, J. A. (2005). Atlantic salmon paramyxovirus (ASPV) infection contributes to proliferative gill inflammation (PGI) in seawater-reared *Salmo salar*. *Dis. Aquat. Organ.* 67, 47–54. doi: 10.3354/dao067047

Laurin, E., Jaramillo, D., Vanderstichel, R., Ferguson, H., Kaukinen, K. H., Schulze, A. D., et al. (2019). Histopathological and novel high-throughput molecular monitoring data from farmed salmon (*Salmo salar* and *Oncorhynchus* spp.) in British Columbia, Canada, from 2011–2013. *Aquaculture* 499, 220–234. doi: 10.1016/j.aquaculture.2018.08.072

LeBlanc, F., Ditlecadet, D., Arseneau, J. R., Steeves, R., Boston, L., Boudreau, P., et al. (2019). Isolation and identification of a novel salmon gill poxvirus variant from Atlantic salmon in Eastern Canada. *J. Fish Dis.* 42, 315–318. doi: 10.1111/jfd.12922

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., et al. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* 533, 200–205. doi: 10.1038/nature17164

Maestrini, B., and Basso, B. (2018). Drivers of within-field spatial and temporal variability of crop yield across the US Midwest. *Sci. Rep.* 8:14833. doi: 10.1038/s41598-018-32779-3

Maina, J. N. (2011). *Bioengineering Aspects in the Design of Gas Exchangers: Comparative Evolutionary, Morphological, Functional, and Molecular Perspectives*. Heidelberg: Springer-Verlag.

Marcos-López, M., Calduch-Giner, J. A., Mirimin, L., MacCarthy, E., Rodger, H. D., O'Connor, I., et al. (2018). Gene expression analysis of Atlantic salmon gills reveals mucin 5 and interleukin 4/13 as key molecules during amoebic gill disease. *Sci. Rep.* 8:13689. doi: 10.1038/s41598-018-32019-8

Martin, S. A. M., Dehler, C. E., and Król, E. (2016). Transcriptomic responses in the fish intestine. *Dev. Comp. Immunol.* 64, 103–117. doi: 10.1016/j.dci.2016.03.014

Martin, S. A. M., and Król, E. (2017). Nutrigenomics and immune function in fish: new insights from omics technologies. *Dev. Comp. Immunol.* 75, 86–98. doi: 10.1016/j.dci.2017.02.024

Matthews, C. G. G., Richards, R. H., Shinn, A. P., and Cox, D. I. (2013). Gill pathology in Scottish farmed Atlantic salmon, *Salmo salar* L., associated with the microsporidian Desmozoon lepeophtherii Freeman et Sommerville, 2009. *J. Fish Dis.* 36, 861–869.

McBryan, T. L., Anttila, K., Healy, T. M., and Schulte, P. M. (2013). Responses to temperature and hypoxia as interacting stressors in fish: implications for adaptation to environmental change. *Integr. Comp. Biol.* 53, 648–659. doi: 10.1093/icb/ict066

McCormick, S. D., and Regish, A. M. (2018). Effects of ocean acidification on salinity tolerance and seawater growth of Atlantic salmon *Salmo salar* smolts. *J. Fish Biol.* 93, 560–566. doi: 10.1111/jfb.13656

McIntyre, J. K., Lundin, J. I., Cameron, J. R., Chow, M. I., Davis, J. W., Incardona, J. P., et al. (2018). Interspecies variation in the susceptibility of adult Pacific salmon to toxic urban stormwater runoff. *Environ. Pollut.* 238, 196–203. doi: 10.1016/j.envpol.2018.03.012

Mi, H., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* 8, 1551–1566. doi: 10.1038/nprot.2013.092

Mitchell, S. O., Baxter, E. J., Holland, C., and Rodger, H. D. (2012). Development of a novel histopathological gill scoring protocol for assessment of gill health during a longitudinal study in marine-farmed Atlantic salmon (*Salmo salar*). *Aquacult. Int.* 20, 813–825. doi: 10.1007/s10499-012-9504-x

Mitchell, S. O., and Rodger, H. D. (2011). A review of infectious gill disease in marine salmonid fish. *J. Fish Dis.* 34, 411–432. doi: 10.1111/j.1365-2761.2011.01251.x

Mitchell, S. O., Steinum, T. M., Toenshoff, E. R., Kvellestad, A., Falk, K., Horn, M., et al. (2013). 'Candidatus Branchiomonas cysticola' is a common agent of epitheliocysts in seawater-farmed Atlantic salmon *Salmo salar* in Norway and Ireland. *Dis. Aquat. Organ.* 103, 35–43. doi: 10.3354/dao02563

Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D. A., Barnholtz-Sloan, J. S., Velázquez Vega, J. E., et al. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci. U.S.A.* 15, E2970–E2979. doi: 10.1073/pnas.1717139115

Monte, M. M., Wang, T., Holland, J. W., Zou, J., and Secombes, C. J. (2013). Cloning and characterisation of rainbow trout interleukin-17A/F2 (IL-17A/F2) and IL-17 receptor A: expression during infection and bioactivity of recombinant IL-17A/F2. *Infect. Immun.* 81, 340–353. doi: 10.1128/iai.00599-12

Morgenroth, D., Ekström, A., Hjelmstedt, P., Gräns, A., Axelsson, M., and Sandblom, E. (2019). Hemodynamic responses to warming in euryhaline rainbow trout: implications of the osmo-respiratory compromise. *J. Exp. Biol.* 222:jeb207522. doi: 10.1242/jeb.207522

Morrison, R. N., Cooper, G. A., Koop, B. F., Rise, M. L., Bridle, A. R., Adams, M. B., et al. (2006). Transcriptome profiling the gills of amoebic gill disease (AGD)-affected Atlantic salmon (*Salmo salar* L.): a role for tumor suppressor p53 in AGD pathogenesis? *Physiol. Genomics* 26, 15–34. doi: 10.1152/physiolgenomics.00320.2005

Morrison, R. N., Zou, J., Secombes, C. J., Scapigliati, G., Adams, M. B., and Nowak, B. F. (2007). Molecular cloning and expression analysis of tumour necrosis factor-α in amoebic gill disease (AGD)-affected Atlantic salmon (*Salmo salar* L.). *Fish Shellfish Immunol.* 23, 1015–1031. doi: 10.1016/j.fsi.2007.04.003

Mueller, P. P., Drynda, A., Goltz, D., Hoehn, R., Hauser, H., and Peuster, M. (2009). Common signatures for gene expression in postnatal patients with patent arterial ducts and stented arteries. *Cardiol. Young* 19, 352–359. doi: 10.1017/S1047951109004260

Nilsson, G. E. (2007). Gill remodeling in fish – a new fashion or an ancient secret? *J. Exp. Biol.* 210, 2403–2409. doi: 10.1242/jeb.000281

Noguera, P., Olsen, A. B., Hoare, J., Lie, K. I., Marcos-López, M., Poppe, T. T., et al. (2019). Complex gill disorder (CGD): a histopathology workshop report. *Bull. Eur. Assoc. Fish Pathol.* 39, 172–176.

Nowak, B. F., and Archibald, J. M. (2018). Opportunistic but lethal: the mystery of Paramoebae. *Trends Parasitol.* 34, 404–419. doi: 10.1016/j.pt.2018.01.004

Nye, J. A., Link, J. S., Hare, J. A., and Overholtz, W. J. (2009). Changing spatial distribution of fish stocks in relation to climate and population size on the Northeast United States continental shelf. *Mar. Ecol. Prog. Ser.* 393, 111–129. doi: 10.3354/meps08220

Okamoto, T., Gohil, K., Finkelstein, E. I., Bove, P., Akaike, T., and van der Vliet, A. (2004). Multiple contributing roles for NOS2 in LPS-induced acute airway inflammation in mice. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 286, L198–L209.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2019). *Vegan: Community Ecology Package. R package version 2.5-6*. Available online at: https://CRAN.R-project.org/package=vegan (accessed February 11, 2020).

Oleksiak, M. F. (2008). Changes in gene expression due to chronic exposure to environmental pollutants. *Aquat. Toxicol.* 90, 161–171. doi: 10.1016/j.aquatox.2008.08.010

Olsen, M. M., Kania, P. W., Heinecke, R. D., Skjoedt, K., Rasmussen, K. J., and Buchmann, K. (2011). Cellular and humoral factors involved in the response of rainbow trout gills to *Ichthyophthirius multifiliis* infections: molecular and immunohistochemical studies. *Fish Shellfish Immunol.* 30, 859–869. doi: 10.1016/j.fsi.2011.01.010

Onukwufor, J. O., and Wood, C. M. (2018). The osmorespiratory compromise in rainbow trout (*Oncorhynchus mykiss*): the effects of fish size, hypoxia, temperature and strenuous exercise on gill diffusive water ?uxes and sodium net loss rates. *Comp. Biochem. Physiol. A* 219–202, 10–18. doi: 10.1016/j.cbpa.2018.02.002

Park, K., Kim, W., and Kim, H. Y. (2014). Optimal lamellar arrangement in fish gills. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8067–8070. doi: 10.1073/pnas.1403621111

Pennacchi, Y., Adams, M. B., Nowak, B. F., and Bridle, A. R. (2016). Immune gene expression in the gills of Atlantic salmon (*Salmo salar* L.) following experimental reinfection with Neoparamoeba perurans. *Aquaculture* 464, 410–419. doi: 10.1016/j.aquaculture.2016.07.025

Philis, G., Ziegler, F., Gansel, L. C., Jansen, M. D., Gracey, E. O., and Stene, A. (2019). Comparing life cycle assessment (LCA) of salmonid aquaculture production systems: status and perspectives. *Sustainability* 11:2517. doi: 10.3390/su11092517

Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L., and Levin, S. A. (2013). Marine taxa track local climate velocities. *Science* 341, 1239–1242. doi: 10.1126/science.1239352

Pleić, I. L., Bušelić, I., Trumbić, Ž., Bočina, I., Šprung, M., and Mladineo, I. (2015). Expression analysis of the Atlantic bluefin tuna (*Thunnus thynnus*) pro-inflammatory cytokines, IL-1β, TNFα1 and TNFα2 in response to parasites *Pseudocycnus appendiculatus* (Copepoda) and *Didymosulcus katsuwonicola* (Digenea). *Fish Shellfish Immunol.* 45, 946–954. doi: 10.1016/j.fsi.2015.06.008

Powell, M. D., Harris, J. O., Carson, J., and Hill, J. V. (2005). Effects of gill abrasion and experimental infection with *Tenacibaculum maritimum* on the respiratory physiology of Atlantic salmon *Salmo salar* affected by amoebic gill disease. *Dis. Aquat. Org.* 63, 169–174. doi: 10.3354/dao063169

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rahel, F. J., and Olden, J. D. (2008). Assessing the effects of climate change on aquatic invasive species. *Conserv. Biol.* 22, 521–533. doi: 10.1111/j.1523-1739.2008.00950.x

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Robledo, D., Hamilton, A., Gutiérrez, A. P., Bron, J. E., and Houston, R. D. (2019). Characterising the mechanisms underlying genetic resistance to amoebic gill disease in Atlantic salmon using RNA sequencing. *bioRxiv* [Preprint]. doi: 10.1101/699561

Rodger, H. D. (2007). Gill disorders: An emerging problem for farmed Atlantic salmon (*Salmo salar*) in the marine environment? *Fish Vet. J.* 9, 38–48.

Rohart, F., Eslami, A., Matigian, N., Bougeard, S., and Lê Cao, K. A. (2017). MINT: a multivariate integrative method to identify reproducible molecular signatures across independent experiments and platforms. *BMC Bioinformatics* 18:128. doi: 10.1186/s12859-017-1553-8

Rojas, V., Camus-Guerra, H., Guzman, F., and Mercado, L. (2015). Pro-inflammatory caspase-1 activation during the immune response in cells from rainbow trout *Oncorhynchus mykiss* (Walbaum 1792) challenged with pathogen-associated molecular patterns. *J. Fish Dis.* 38, 993–1003. doi: 10.1111/jfd.12315

Ronza, P., Robledo, D., Bermúdez, R., Losada, A. P., Pardo, B. G., Martínez, P., et al. (2019). Integrating genomic and morphological approaches in fish pathology research: the case of turbot (*Scophthalmus maximus*) enteromyxosis. *Front. Genet.* 10:26. doi: 10.3389/fgene.2019.00026

Rozas-Serri, M. (2019). Gill diseases in marine salmon aquaculture with an emphasis on amoebic gill disease. *CAB Rev.* 14, 1–15.

Salinas, I. (2015). The mucosal immune system of teleost fish. *Biology* 4, 525–539. doi: 10.3390/biology4030525

Santana, P. A., Guzmán, F., Forero, J. C., Luna, O. F., and Mercado, L. (2016). Hepcidin, cathelicidin-1 and IL-8 as immunological markers of responsiveness in early developmental stages of rainbow trout. *Dev. Comp. Immunol.* 62, 48–57. doi: 10.1016/j.dci.2016.04.014

Shi, X., Zhang, W., Yin, L., Chilian, W. M., Krieger, J., and Zhang, P. (2017). Vascular precursor cells in tissue injury repair. *Transl. Res.* 184, 77–100. doi: 10.1016/j.trsl.2017.02.002

Shinn, A. P., Pratoomyot, J., Bron, J. E., Paladini, G., Brooker, E. E., and Brooker, A. J. (2015). Economic costs of protistan and metazoan parasites to global mariculture. *Parasitology* 142, 196–270. doi: 10.1017/s0031182014001437

Soleng, M., Johansen, L. H., Johnsen, H., Johansson, G. S., Breiland, M. W., Rormark, L., et al. (2019). Atlantic salmon (*Salmo salar*) mounts systemic and mucosal stress responses to peracetic acid. *Fish Shellfish Immunol.* 93, 895–903. doi: 10.1016/j.fsi.2019.08.048

Sollid, J., De Angelis, P., Gundersen, K., and Nilsson, G. (2003). Hypoxia induced adaptive and reversible gross morphological changes in crucian carp gills. *J. Exp. Biol.* 206, 3667–3673. doi: 10.1242/jeb.00594

Song, Y., Salbu, B., Teien, H. C., Sørlie Heier, L., Rosseland, B. O., Høgåsen, T., et al. (2014). Hepatic transcriptomic profiling reveals early toxicological mechanisms of uranium in Atlantic salmon (*Salmo salar*). *BMC Genomics* 15:694. doi: 10.1186/1471-2164-15-694

Steinum, T., Kvellestad, A., Colguhoun, D. J., Heum, M., Mohammad, S., Grontvedt, R. N., et al. (2010). Microbial and pathological findings in farmed Atlantic salmon *Salmo salar* with proliferative gill inflammation. *Dis. Aquat. Org.* 91, 201–211. doi: 10.3354/dao02266

Sun, F. Y., Peatman, E., Li, C., Liu, S. K., Jiang, Y. L., Zhou, Z. C., et al. (2012). Transcriptomic signatures of attachment, NF-kappa B suppression and IFN stimulation in the catfish gill following columnaris bacterial infection. *Dev. Comp. Immunol.* 38, 169–180. doi: 10.1016/j.dci.2012.05.006

Syahputra, K., Kania, P. W., Al-Jubury, A., Marnis, H., Setyawan, A. C., and Buchmann, K. (2019). Differential immune gene response in gills, skin, and spleen of rainbow trout *Oncorhynchus mykiss* infected by *Ichthyophthirius multifiliis*. *PLoS One* 14:e0218630. doi: 10.1371/journal.pone.0218630

Tisoncik, J. R., Korth, M. J., Simmons, C. P., Farrar, J., Martin, T. R., and Katze, M. G. (2012). Molecular innate immunity in teleost fish: review and future perspectives. Into the eye of the cytokine storm. *Microbiol. Mol. Biol. Rev.* 76, 16–32. doi: 10.1128/MMBR.05015-11

Upadhyay, V., and Fu, Y. X. (2013). Lymphotoxin signalling in immune homeostasis and the control of microorganisms. *Nat. Rev. Immunol.* 13, 270–279. doi: 10.1038/nri3406

Valdenegro-Vega, V. A., Cook, M., Crosbie, P., Bridle, A. R., and Nowak, B. F. (2015). Vaccination with recombinant protein (*r* 22C03), a putative attachment factor of *Neoparamoeba perurans* , against AGD in Atlantic salmon (*Salmo salar*) and implications of a co-infection with *Yersinia ruckeri*. *Fish Shellfish Immunol.* 44, 592–602. doi: 10.1016/j.fsi.2015.03.016

Valdenegro-Vega, V. A., Crosbie, P., Bridle, A., Leef, M., Wilson, R., and Nowak, B. F. (2014). Differentially expressed proteins in gill and skin mucus of Atlantic salmon (*Salmo salar*) affected by amoebic gill disease. *Fish Shellfish Immunol.* 40, 69–77. doi: 10.1016/j.fsi.2014.06.025

Valenzuela-Miranda, D., Boltaña, S., Cabrejos, M. E., Yáñez, J. M., and Gallardo-Escárate, C. (2015). High-throughput transcriptome analysis of ISAV-infected Atlantic salmon Salmo salar unravels divergent immune responses associated to head-kidney, liver and gills tissues. *Fish Shellfish Immunol.* 45, 367–377. doi: 10.1016/j.fsi.2015.04.003

White, C. R., Phillips, N. F., and Seymour, R. S. (2006). The scaling and temperature dependence of vertebrate metabolism. *Biol. Lett.* 2, 125–127. doi: 10.1098/rsbl.2005.0378

Wise, D. J., Griffin, M. J., Terhune, J. S., Pote, L. M., and Khoo, L. H. (2008). Induction and evaluation of proliferative gill disease in channel catfish fingerlings. *J. Aquat. Anim. Health* 20, 236–244. doi: 10.1577/H08-023.1

Wynne, J. W., O'Sullivan, M. G., Cook, M. T., Stone, G., Nowak, B. F., Lovell, D. R., et al. (2008a). Transcriptome analyses of amoebic gill disease-affected Atlantic salmon (*Salmo salar*) tissues reveal localized host gene suppression. *Mar. Biotechnol.* 10, 388–403. doi: 10.1007/s10126-007-9075-4

Wynne, J. W., O'Sullivan, M. G., Stone, G., Cook, M. T., Nowak, B. F., Lovell, D. R., et al. (2008b). Resistance to amoebic gill disease (AGD) is characterised by the transcriptional dysregulation of immune and cell cycle pathways. *Dev. Comp. Immunol.* 32, 1539–1560. doi: 10.1016/j.dci.2008.05.013

Xu, Z., Takizawa, F., Casadei, E., Shibasaki, Y., Ding, Y., Sauters, T. J. C., et al. (2020). Specialization of mucosal immunoglobulins in pathogen control and microbiota homeostasis occurred early in vertebrate evolution. *Sci. Immunol.* 5:eaay3254. doi: 10.1126/sciimmunol.aay3254

Young, N. D., Crosbie, P. B., Adams, M. B., Nowak, B. F., and Morrison, R. N. (2007). *Neoparamoeba perurans* n. sp., an agent of amoebic gill disease of Atlantic salmon (*Salmo salar*). *Int. J. Parasitol.* 37, 1469–1481. doi: 10.1016/j.ijpara.2007.04.018

Zilberg, D., Gross, A., and Munday, B. L. (2001). Production of salmonid amoebic gill disease by exposure to *Paramoeba* sp. harvested from the gills of infected fish. *J. Fish Dis.* 24, 79–82. doi: 10.1046/j.1365-2761.2001.00271.x

Zou, J., and Secombes, C. J. (2016). The function of fish cytokines. *Biology* 5:23. doi: 10.3390/biology5020023

Conflict of Interest: KG and VV were employed by the company BioMar AS. RB was employed by the company Scottish Sea Farms (SSF).

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 June 2020
doi: 10.3389/fmicb.2020.01420

Analysis of Genetic Variation in the Bovine *SLC11A1* Gene, Its Influence on the Expression of NRAMP1 and Potential Association With Resistance to Bovine Tuberculosis

*Angela Holder[1], Rachel Garty[1], Charlotte Elder[1], Paula Mesnard[1,2], Celine Laquerbe[1,2], Marie-Christine Bartens[1], Mazdak Salavati[1,3], Muhammad Zubair Shabbir[4], Thomas Tzelos[3], Timothy Connelly[3], Bernardo Villarreal-Ramos[5,6] and Dirk Werling[1]**

[1] Department of Pathobiology and Population Sciences, Royal Veterinary College, Hertfordshire, United Kingdom, [2] EPLEFPA Agricampus La Roque, Rodez, France, [3] The Roslin Institute, The University of Edinburgh, Midlothian, United Kingdom, [4] University of Veterinary and Animal Sciences, Lahore, Pakistan, [5] Institute of Biological Environmental and Rural Sciences (IBERS), Aberystwyth University, Aberystwyth, United Kingdom, [6] APHA, Webybridge, United Kingdom

OPEN ACCESS

Edited by:
Miklos Fuzi,
Semmelweis University, Hungary

Reviewed by:
Ana Cláudia Coelho,
University of Trás-os-Montes and Alto Douro, Portugal
David E. MacHugh,
University College Dublin, Ireland

***Correspondence:**
Dirk Werling
dwerling@rvc.ac.uk

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: 10 March 2020
Accepted: 02 June 2020
Published: 30 June 2020

Citation:
Holder A, Garty R, Elder C, Mesnard P, Laquerbe C, Bartens M-C, Salavati M, Shabbir MZ, Tzelos T, Connelly T, Villarreal-Ramos B and Werling D (2020) Analysis of Genetic Variation in the Bovine SLC11A1 Gene, Its Influence on the Expression of NRAMP1 and Potential Association With Resistance to Bovine Tuberculosis. Front. Microbiol. 11:1420. doi: 10.3389/fmicb.2020.01420

Bovine tuberculosis (bTB), caused by *Mycobacterium bovis*, is a chronic zoonotic disease where host genetics is thought to contribute to susceptibility or resistance. One of the genes implicated is the *SLC11A1* gene, that encodes for the natural resistance-associated macrophage protein 1 (NRAMP1). The aim of this study was to identify *SLC11A1* polymorphisms and to investigate any resulting functional differences in NRAMP1 expression that might be correlated with resistance/susceptibility to *M. bovis* infection. Sequencing of the *SLC11A1* gene in cDNA isolated from Brown Swiss, Holstein Friesian, and Sahiwal cattle identified five single nucleotide polymorphisms (SNPs) in the coding region, but only one of these (SNP4, c.1066C>G, rs109453173) was present in all three cattle breeds and therefore warranted further investigation. Additionally, variations of 10, 11, and 12 GT repeats were identified in a microsatellite (MS1) in the *SLC11A1* 3′UTR. Measurement of NRAMP1 expression in bovine macrophages by ELISA showed no differences between cells generated from the different breeds. Furthermore, variations in the length of the MS1 microsatellite did not impact on NRAMP1 protein expression as analyzed by luciferase reporter assay. However, further analysis of the ELISA data identified that the presence of the alternative G allele at SNP4 was associated with increased expression of NRAMP1 in bovine macrophages. Since NRAMP1 has been shown to influence the survival of intracellular pathogens such as *M. bovis* through the sequestering of iron, it is possible that cattle expressing the alternative G allele might have an increased resistance to bTB through increased NRAMP1 expression in their macrophages.

Keywords: SLC11A1, NRAMP1, bovine tuberculosis, genetics, cattle breeds

INTRODUCTION

Bovine tuberculosis (bTB), caused by *Mycobacterium bovis*, is a chronic disease of worldwide importance due to the major economic and zoonotic implications associated with it. In the United Kingdom, approximately 30,000 cattle are slaughtered each year due to bTB at a cost of around £100 million (UK Government Bovine TB, 2020), while the annual cost of bTB to worldwide agricultural is estimated to be around three billion US dollars (Allen et al., 2010). However, bTB does not only affect cattle, it also represents a risk to human health, with 10–15% of human TB cases worldwide being attributed to an *M. bovis* infection (Michel et al., 2010).

Recent research into controlling the spread of bTB has focused on improving diagnostic testing (Schiller et al., 2010) as well as the development of suitable vaccines for livestock and/or wildlife reservoirs with the ability to differentiate between vaccinated and infected animals (Chambers et al., 2014; Buddle et al., 2018). However, it has been observed that different cattle breeds vary in their ability to clear bacterial pathogens, suggesting that the host's genetics play an important role in generating a potentially protective immune response (Rupp and Boichard, 2003; Allen et al., 2010; le Roex et al., 2013). Indeed, a comparison between breeds showed that Simmental and Brown Swiss (BS) cattle demonstrated a lower clinical mastitis frequency compared to Holstein-Friesian (HF) cattle (Rupp and Boichard, 2003), while studies in Ethiopia identified a lower incidence rate as well as reduced severity of *M. bovis* infections in zebu cattle (*Bos indicus*) compared to HF cattle (*Bos taurus*) (Ameni et al., 2007; Vordermeier et al., 2012). Additionally, monocyte-derived macrophages (MDM) from BS cattle produced more reactive oxygen species and showed increased phagocytic and microbicidal capabilities compared to MDM generated from HF cattle, suggesting that BS cattle might be more effective in dealing with bacterial pathogens compared HF cattle (Gibson et al., 2016). Therefore, the identification of genetic traits impacting on susceptibility or resistance to bTB might offer alternative ways of controlling the disease through selective breeding (Warner et al., 1987; Allen et al., 2010; le Roex et al., 2013). Indeed, genomic regions and candidate genes were recently identified in HF cattle which may provide an opportunity to further understand pathways critical to cattle susceptibility to bTB (Raphaka et al., 2017).

The genome-wide association study (GWAS) approach has provided a tool to identify chromosomal regions in the bovine genome that are associated with susceptibility to bTB in order to try and identify specific candidate genes for further analysis (Finlay et al., 2012; Bermingham et al., 2014; Raphaka et al., 2017; Ring et al., 2019). However, one of the potentially most promising candidate gene was identified in murine studies performed over 30 years ago. Here, a locus (originally designated Ity/Lsh/Bcg) was identified that was responsible for innate resistance to the intracellular pathogens *Salmonella typhimurium*, *Leishmania donovani* and *M. bovis* BCG (Plant et al., 1982; Skamene et al., 1982). This locus has since been identified as *SLC11A1*, a member of the solute carrier family 11, that encodes for the natural resistance-associated macrophage protein 1 (NRAMP1). NRAMP1 is a multi-pass membrane protein that regulates macrophage activation in infectious and autoimmune diseases by acting as a transporter protein for protons, iron and other divalent cations (Blackwell et al., 2000).

A previous study has identified single nucleotide polymorphisms (SNPs) in the bovine *SLC11A1* gene that are associated with susceptibility to bTB (Cheng et al., 2015). A study in Taiwanese HF cattle identified SNPs in exon 4 and intron 4 that were associated with enhanced susceptibility to bTB (Cheng et al., 2015), while in a Chinese HF cattle population, associations with bTB susceptibility/resistance were found with SNPs in exon 11 and introns 5 and 9 (Liu et al., 2017). Additionally polymorphisms in the 3′UTR of the bovine *SLC11A1* gene, that have been shown to cause changes to the lengths of two microsatellites (MS1 and MS2), have been identified in various breeds of cattle (Paixao et al., 2006; Vazquez-Flores et al., 2006; Hasenauer et al., 2013). Interestingly, an increase in the length of the MS1 microsatellite has been described to reduce traits indicative of bTB infection in African zebu cattle (Kadarmideen et al., 2011). In contrast, no association between the MS1 microsatellite and response to the tuberculin skin test was found in a study performed using HF and Jersey cattle in Argentina (Hasenauer et al., 2018). Furthermore, no associations between the MS2 microsatellite and susceptibility/resistance to bTB have currently been identified (Barthel et al., 2000; Hasenauer et al., 2018). Therefore, in the present study we characterized polymorphisms in the bovine *SLC11A1* gene in two *B. taurus* breeds, BS and HF, both present in the United Kingdom, and compared these to those found in Sahiwal cattle, a *B. indicus* breed. Furthermore, functional differences in NRAMP1 protein expression resulting from the genetic variation identified were subsequently investigated in order to potentially explain some of the previously identified breed differences seen regarding susceptibility to bTB.

MATERIALS AND METHODS

Samples

Peripheral blood mononuclear cells (PBMCs) were purified by density gradient centrifugation (Lymphoprep, STEMCELL technologies) using whole blood from 15 BS and 15 HF cattle reared at the Royal Veterinary College (Hatfield, United Kingdom) and Cancourt Farm (Swindon, United Kingdom), and sampled under Home Office license (PPL7009059). RNA was extracted from the PBMCs using the RNeasy Mini Kit (Qiagen) and cDNA was synthesized using the iScript cDNA Synthesis Kit (Bio-Rad). Complementary DNA samples from 15 Sahiwal cattle, reared on farms in Pakistan, were a provided by The Roslin Institute (University of Edinburgh, Edinburgh, United Kingdom). The sample size used in this study was limited by aforementioned Home Office license.

Sequencing and Analysis of the Bovine *SLC11A1* Gene

All 45 samples were used for sequencing. Specific primers to amplify the coding region of the bovine *SLC11A1* gene

were designed using publicly available sequence information[1] and Primer Blast[2] (**Supplementary Table S1**). Primers to amplify a specific region of the *SLC11A1* 3′UTR containing a previously described microsatellite [MS1, c.1647+61GT(10_13)] were taken from the current literature (Hasenauer et al., 2013; **Supplementary Table 1**).

PCR reactions were carried out using Phusion High-Fidelity Master Mix (Thermo Fisher Scientific). PCRs were performed in 50 μl reactions containing 25 μl 2× Phusion Master Mix, 3 μl cDNA and 2.5 μl of each primer (forward and reverse, 10 μM). Due to the GC rich nature of the *SLC11A1* gene 3% DMSO was included in PCR reactions used to amplify the coding region. Thermocycling conditions for the coding region consisted of an initial denaturation step at 98°C for 30 s, followed by 36 cycles of 98°C for 10 s (denaturation), 65.5°C for 30 s (annealing) and 72°C for 30 s (extension), with a final extension step at 72°C for 10 min (Mastercycler Pro S, Eppendorf). For amplification of the MS1 microsatellite the annealing and elongation steps were changed to 62°C for 20 s and 72°C for 15 s, respectively. The PCR products generated were separated by agarose gel electrophoresis, purified using the QIAquick Gel Extraction Kit (Qiagen), and then submitted for sequencing (MRC PPU DNA Sequencing Service, Dundee University).

Single nucleotide polymorphisms were identified using CLC Main Workbench version 6.0.2 (CLCbio, Aarhus, Denmark). Allele and genotype frequencies were calculated for each SNP and compared between cattle breeds using a Fisher Exact Test[3]. Phylogenetic analysis was performed in CLC Main Workbench using the coding region genotypes identified in the study and the bovine reference sequence[4]. Heterozygous loci were labeled using IUPAC ambiguity codes. A maximum likelihood phylogenetic tree was constructed using a Jukes Cantor model and 100 bootstrap iterations.

Measurement of NRAMP1 Protein in Bovine Monocyte-Derived Macrophages by ELISA

Monocytes were isolated from PBMCs (11 BS and 12 HF) by magnetic activated cell sorting (MACS). Briefly, PBMCs were incubated with paramagnetic beads labeled with an anti-CD14 antibody (MicroBeads, Miltenyi Biotec) and $CD14^+$ monocytes were isolated by positive selection on LS columns in a MidiMACS magnetic separator (Miltenyi Biotech) as described before (Gibson et al., 2016). Monocytes were cultured at approximately 1×10^6 cells/ml in RPMI 1640 medium containing GlutaMAX-1 (Gibco) supplemented with 10% Fetal Bovine Serum (FBS, Sigma-Aldrich) and 1% Penicillin-Streptomycin (PenStrep, Gibco). Cells were differentiated into MDM by culturing them for 6 days in the presence of 20 ng/ml recombinant bovine (rbo)M-CSF (Kingfisher Biotech). After 6 days in culture, MDM were harvested using Accutase (Gibco), washed twice in PBS and counted using an automated cell counter (TC20, Bio-Rad). MDM were lysed subsequently by freeze-thawing three times, before being centrifuged at 1000 × *g* for 15 min to remove any debris. Cell lysates were assayed for NRAMP1 concentrations using a bovine specific NRAMP1 ELISA kit (BlueGene) according to the manufacturer's instructions. Standards and samples were assayed in duplicate. A standard curve was generated using optical density (OD) 450 nm values obtained for the standards provided, and the concentration of NRAMP1 protein in the cell lysates was interpolated from this. Subsequently, NRAMP1 concentrations were normalized for cell number.

Dual-Glo miRNA Target Luciferase Reporter Assay

A 117 bp region of the *SLC11A1* 3′UTR containing the MS1 microsatellite was amplified by PCR (see **Supplementary Table 1** for primers) using cDNA samples from cows representing the three different lengths of MS1 identified in the study (10, 11, and 12 repeats of GTs). PCR products were purified by gel extraction (QIAquick Gel Extraction Kit, Qiagen) and ligated into a dual-luciferase miRNA target expression vector (pmirGLO, Promega) using the *PmeI* and *XbaI* restriction sites located in the multiple cloning site of the vector and T4 DNA ligase (Thermo Fisher Scientific).

Two different cell lines, chinese hamster ovary (CHO) and murine macrophage-like RAW264.7 cells were transfected with recombinant pmirGLO constructs. CHO cells were plated at 1×10^5 cells/well in 24 well plates in Minimal Essential Medium (MEM) supplemented with 10% FBS, 1% non-essential amino acid solution (Sigma-Aldrich), 1% GlutaMAX-1, 1% PenStrep (Gibco), and cultured overnight until they reached 70–90% confluency. At this point, CHO cells were transfected with 800 ng plasmid DNA using Turbofect (Thermo Fisher Scientific) according to the manufacturer's instructions. RAW264.7 cells were plated at the same density but cultured in RPMI-1640 medium containing GlutaMAX-1 (Gibco), supplemented 10% FBS (Sigma-Aldrich), 1 mM Sodium Pyruvate and 1% PenStrep (Gibco). They were transfected with 500 ng plasmid DNA using Lipofectamine 3000 (Invitrogen) according to the manufacturer's instructions. Untransfected cells were used as a negative control.

Twenty-four hours after transfection, cells were assayed for both firefly and renilla luciferase activity, using the Dual-Glo Luciferase Assay System (Promega). The cells were lysed in 200 μl 1% Triton ×100 (Sigma-Aldrich) and 50 μl was transferred to triplicate wells on a 96 well white walled plate (Corning). Fifty microliters Dual-Glo Luciferase Reagent was added to each well as a substrate for firefly luciferase. After 15 min incubation on a rotating platform luciferase activity was measured using a luminometer (SpectraMax L, Molecular Devices Ltd.). Next, 50 μl Dual-Glo Stop & Glo Reagent was added, and after another 15 min incubation luminescence was measured for a second time to obtain a reading for renilla luciferase activity. Firefly luciferase activity was then normalized against renilla luciferase activity.

Statistical Analysis

Groups were compared using either an unpaired *t*-test or one-way ANOVA following a Shapiro–Wilk test to confirm normality.

[1] https://www.ncbi.nlm.nih.gov/nuccore/NM_174652.2

[2] https://www.ncbi.nlm.nih.gov/tools/primer-blast/

[3] http://vassarstats.net/index.html

[4] https://www.ncbi.nlm.nih.gov/nuccore/NM_174652.2

Significant differences were donated as followed in the figures: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

RESULTS

Analysis of Polymorphisms in the Bovine *SLC11A1* Gene

Sequencing of the *SLC11A1* coding region in cDNA generated from BS, HF, and Sahiwal cattle identified five SNPs. One of these was synonymous while the remaining four resulted in amino acid substitutions (**Supplementary Table 2**).

The *SLC11A1* coding region in BS and HF cattle appeared to contain less genetic variation compared to Sahiwal cattle (**Table 1** and **Supplementary Table 3**). Four out of the five SNPs identified occurred at significantly higher frequencies in Sahiwal cattle compared to BS and HF cattle. In fact, SNP1 and SNP5 (c.87A>G, rs109614179 and c.1592G>C, rs110347562) were found to be present only in the Sahiwal cattle. SNP2 (c.650C>T, rs109915208) was only found in HF cattle, although it was present at quite a low frequency (10%). SNP4 (c.1066C>G, rs109453173) was the only one identified as having an alternative allele frequency of >10% in all three cattle breeds, suggesting it has the potential to be a significantly polymorphic locus. None of the SNPs were found to occur at significantly different frequencies between the BS and HF cattle.

Analysis of the MS1 *SLC11A1* 3'UTR microsatellite [c.1647+61GT(10_13)] in BS, HF, and Sahiwal cattle found very little difference in the frequency of the different lengths of this GTn polymorphism between the three cattle breeds (**Table 2** and **Supplementary Table 3**).

Phylogenetic analysis was used to assess evolutionary similarities within the coding region of the *SLC11A1* gene between BS, HF and Sahiwal cattle (**Supplementary Figure 1**). The bovine reference sequence, which originates from a Hereford bull, was included in the analysis for comparative reason. The analysis shows that the sequences cluster into three main groups, with clusters in two of these groups being independent of cattle breed, and only Sahiwal cattle genotypes being present in the third group. This is unusual as sequences from different sub-species, such as the *B. taurus* and *B. indicus* cattle, would be expected to diverge into separate clades on a molecular phylogenetic tree.

NRAMP1 Protein Expression in Bovine Macrophages Is Associated With a Specific Polymorphism in the Coding Region of the *SLC11A1* Gene

As NRAMP1 protein concentration in MDMs generated from BS and HF cattle did not show any significant difference as analyzed by ELISA (**Figure 1A**), we assessed the impact of *SLC11A1* SNP4 (c.1066C>G, rs109453173) instead. Analysis of NRAMP1 expression in MDMs from cows possessing each of the three different *SLC11A1* SNP 4 genotypes (CC, CG, GG) resulted in significantly higher NRAMP1 expression in cells from the alternative GG genotype cows compared to the reference CC genotype ($p = 0.025$) (**Figure 1B**). Similarly, MDMs from cows with the heterozygous cows CG genotype also had increased NRAMP1 expression compared to the reference CC genotype, although this difference did not reach significance. This suggests the G allele has an additive effect on NRAMP1 expression. Combining the results for the CG and GG genotype revealed a significantly increased NRAMP1 expression in MDMs generated from cows carrying the alternative G allele ($p = 0.049$; **Figure 1C**). The influence of the alternative G allele at SNP4 was then analyzed separately by breed (**Figure 1D**). Although NRAMP1 expression is clearly higher in MDMs with the alternative G allele in both, BS and HF, the differences were not significant. This is likely to be the result of reduced numbers in each of the groups compared to when the breeds were pooled together for the previous analysis.

The *SLC11A1* 3′UTR MS1 Microsatellite Does Not Appear to Influence Post-transcriptional Regulation in Macrophages

A dual luciferase reporter assay was employed to assess the possible functional effect of the MS1 microsatellite in the *SLC11A1* 3′UTR. Recombinant plasmid DNA constructs were designed that contained a 117 bp region of the *SLC11A1* 3′UTR, representing the three lengths of MS1 microsatellite identified in this study (10, 11, or 12 GTs), located downstream of the firefly luciferase coding sequence and within its 3′UTR.

Chinese hamster ovary and RAW264.7 cells transfected with recombinant pmirGLO constructs were assayed for both renilla luciferase activity, to estimate transfection efficiency, and firefly luciferase activity, which is the primary reporter gene (**Supplementary Figure 2**). Firefly luciferase activity was then normalized against renilla luciferase activity and the constructs containing the different lengths of microsatellite compared to the empty pmirGLO plasmid (**Figure 2** and **Supplementary Figure 2**).

In CHO cells, expression of the reporter protein appeared to be reduced in cells transfected with the pmirGLO/MS1 11×GTs construct, compared to both the empty pmirGLO plasmid and the constructs containing lengths of 10× and 12×GTs (**Figure 2A**). However, this difference was not found to be significant. In contrast, in murine RAW264.7 cells transfection with the pmirGLO/MS1 11×GTs construct gave higher expression of the reporter protein compared to the others, but again this difference was not significant (**Figure 2B**). This suggests that differences in the number of GT repeats in the *SLC11A1* MS1 microsatellite do not appear to significantly impact post-transcriptional regulation and that this may vary depending on the cell line used.

DISCUSSION

This study was designed to identify potential polymorphisms in the bovine *SLC11A1* gene and examine the influence these might have on the expression of NRAMP1, a protein that plays

TABLE 1 | Analysis of *SLC11A1* coding region SNPs in Brown Swiss, Holstein-Friesian and Sahiwal cattle.

SNP ID	Cattle breed	Allele frequency (%)		*P*-value	Genotype frequency (%)			*P*-value
		Reference	Alternative		Reference	Heterozygous	Alternative	
SNP1	BS	30 (100)	0		15 (100)	0	0	
c.87A>G	HF	30 (100)	0	1	15 (100)	0	0	1
rs109614179	SW	22 (73.3)	8 (26.7)	**0.0002**	9 (60)	4 (26.7)	2 (13.3)	**0.0018**
SNP2	BS	30 (100)	0		15 (100)	0	0	
c.650C>T	HF	27 (90)	3 (10)	0.237	13 (86.6)	1 (6.7)	1 (6.7)	0.483
rs109915208	SW	30 (100)	0	0.104	15 (100)	0	0	0.318
SNP3	BS	29 (96.7)	1 (3.3)		14 (93.3)	1 (6.7)	0	
c.961G>A	HF	30 (100)	0	1	15 (100)	0	0	1
rs109551090	SW	21 (70)	9 (30)	**0.0002**	8 (53.4)	5 (33.3)	2 (13.3)	**0.0038**
SNP4	BS	23 (76.7)	7 (23.3)		9 (60)	5 (33.3)	1 (6.7)	
c.1066C>G	HF	26 (86.7)	4 (13.3)	0.506	12 (80)	2 (13.3)	1 (6.7)	0.682
rs109453173	SW	17 (56.7)	13 (43.3)	**0.0421**	6 (40)	5 (33.3)	4 (26.7)	0.188
SNP5	BS	30 (100)	0		15 (100)	0	0	
c.1592G>C	HF	30 (100)	0	1	15 (100)	0	0	1
rs110347562	SW	25 (83.3)	5 (16.7)	**0.0097**	11 (73.3)	3 (10)	1 (6.7)	**0.018**

Genotype and allele frequencies were compared between cattle breeds for each SNP. P-values were generated using Fisher's exact test. Comparisons were made between Brown Swiss and Holstein-Friesian cattle (upper P-values) and between all three cattle breeds (lower P-values). Significant P-values are shown in bold. BS, Brown Swiss; HF, Holstein-Friesian; SW, Sahiwal.

TABLE 2 | Analysis of the MS1 *SLC11A1* 3′UTR microsatellite in Brown Swiss, Holstein-Friesian, and Sahiwal cattle.

Cattle breed	Allele frequency (%)			*P*-value	Genotype frequency (%)						*P*-value
	10	11	12		10/10	10/11	10/12	11/11	11/12	12/12	
BS	3 (10)	12 (40)	15 (50)		1 (6.7)	0	1 (6.7)	0	12 (80)	1 (6.7)	
HF	2 (6.7)	12 (40)	16 (53.3)	1	1 (6.7)	0	0	0	12 (80)	2 (13.3)	0.999
SW	2 (6.6)	14 (46.7)	14 (46.7)	0.966	0	1 (6.7)	1 (6.7)	0	13 (86.6)	0	0.889

Genotype and allele frequencies were compared between cattle breeds for the different numbers of GT repeats in the microsatellite. P-values were generated using Fisher's exact test. Comparisons were made between Brown Swiss and Holstein-Friesian cattle (upper P-values) and between all three cattle breeds (lower P-values). Significant P-values are shown in bold. BS, Brown Swiss; HF, Holstein-Friesian; SW, Sahiwal.

an important role in dealing with intracellular pathogens, and in particular *M. bovis* (Allen et al., 2010). Sequencing of the *SLC11A1* coding region in cDNA from BS, HF and Sahiwal cattle identified five SNPs. All these SNPs have been identified previously and their details cataloged in the relevant databases[5]. For the present study, the non-synonymous polymorphism located in exon 11 (SNP4, c.1066C>G, rs109453173) was of special interest, as this was the only SNP present in all three cattle breeds. The presence of the alternative G allele for this SNP was associated with higher expression of NRAMP1 in MDM. Sequencing of the *SLC11A1* 3′UTR confirmed the presence of a microsatellite [MS1, c.1647+61GT(10_13)], with variable numbers of GT repeats in all three cattle breeds, but none of these seemed to significantly impact on protein expression.

Four of the five SNPs identified in the bovine SLC11A1 coding region were found to be non-synonymous and therefore have the potential to alter protein function. Sequence alignments, hydropathy profiling and epitope mapping studies have predicted the structure of mammalian NRAMP1 to consist of 12 hydrophobic trans-membrane domains (TMD) joined by alternating extracellular and cytoplasmic loops (Nevo and Nelson, 2006; Cellier et al., 2007; Montalbetti et al., 2013), and analysis performed using the structure of bovine NRAMP1 suggests that this is the same in cattle (Feng et al., 1996; Triantaphyllopoulos et al., 2019). SNP2 and SNP3 (c.650C>T, rs109915208 and c.961G>A, rs109551090), identified in this study resulted in amino acid changes (alanine to valine, aspartic acid to asparagine) in the third and fourth extracellular loops, respectively. SNP4 (c.1066C>G, rs109453173) resulted in proline to alanine substitution in TDM8, and SNP5 (c.1592G>C, rs110347562) resulted in an arginine to proline change in the c-terminus of the protein. As there is limited information regarding the function of NRAMP1 available, the effects these changes might have are difficult to assess. Mutation studies on another NRAMP family member (DMT1/NRAMP2) have shown that the first extracellular loop is involved in metal ion binding (Cohen et al., 2003), while TMD 2, 4, and 6 are essential for metal transport (Lam-Yuk-Tseung et al., 2003; Nevo and Nelson, 2004; Nevo, 2008). Since none of the polymorphisms

[5]https://www.ensembl.org/Bos_taurus/Gene/Variation_Gene/Table?db=core;g=ENSBTAG00000015520;r=2:106392590-106403650

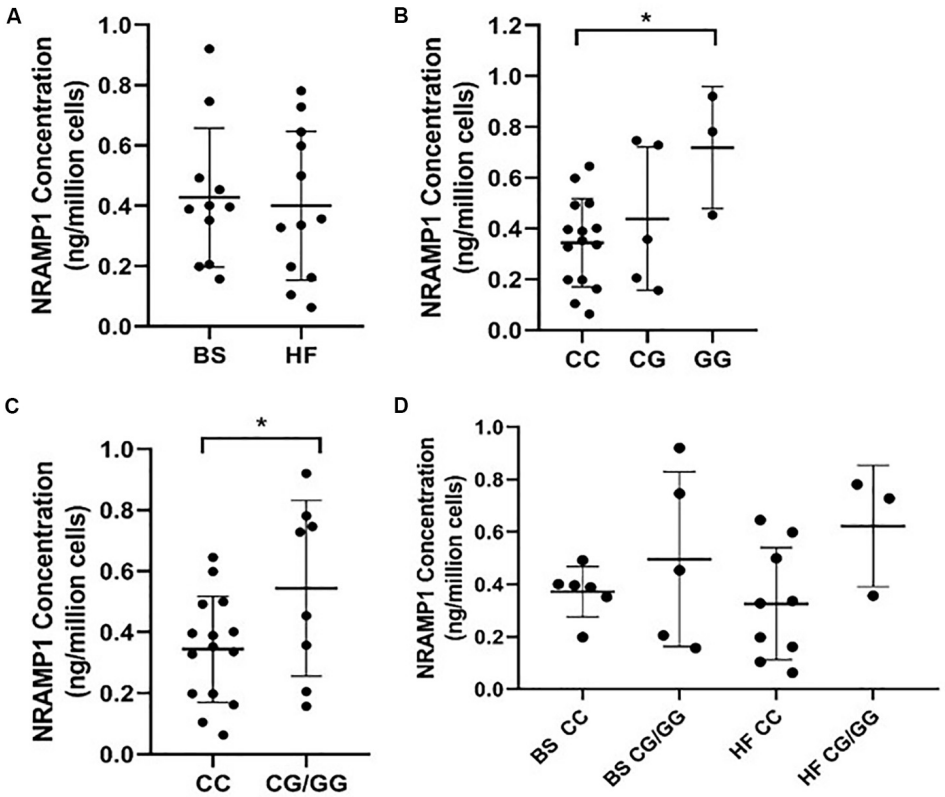


FIGURE 1 | NRAMP1 expression in bovine MDMs is associated with *SLC11A1* SNP4 (c.1066C>G). NRAMP1 protein was measured in cell lysates by ELISA and concentrations were normalized for cell number. Each sample is represented by a filled circle on the dot plots. Samples are grouped according to **(A)** cattle breed **(B)** *SLC11A1* SNP4 genotype, **(C)** the presence or absence of the minor G allele, **(D)** cattle breed and the presence or absence of the minor G allele. The mean and SD of each group are represented by horizontal lines. *P*-values were calculated using **(B)** a one-way ANOVA with *post hoc* Tukey HSD test and **(C)** an unpaired *t*-test. **P*-value < 0.05. BS, Brown Swiss; HF, Holstein-Friesian.

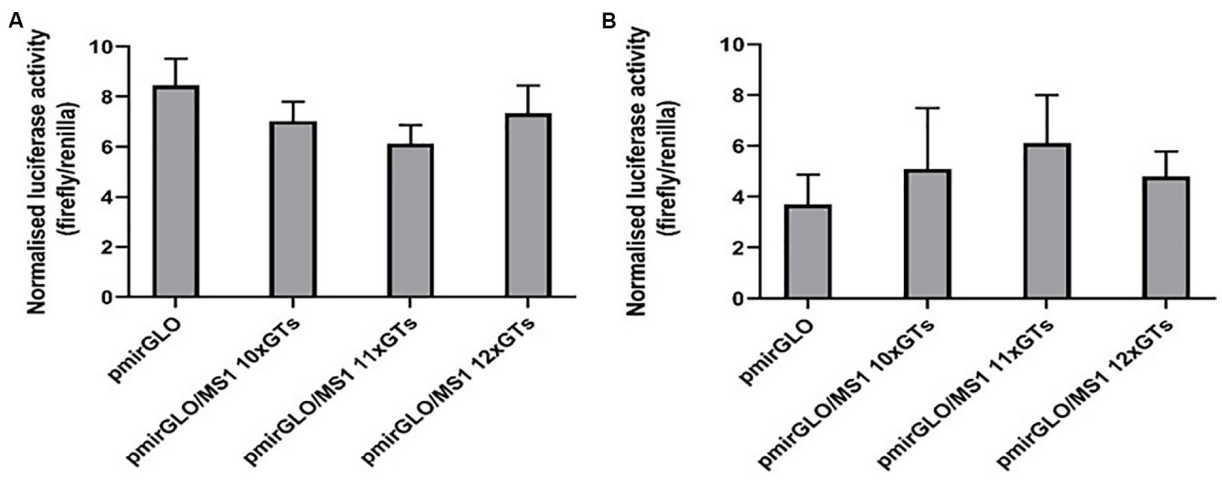


FIGURE 2 | Different lengths of the *SLC11A1* 3′UTR MS1 microsatellite do not appear to influence post-transcriptional regulation. **(A)** CHO and **(B)** RAW 264.7 cells were transfected with pmirGLO constructs containing different lengths of the MS1 microsatellite. After 24 h renilla and firefly luciferase activity was measured using the Dual-GLO luciferase assay system. Normalized luciferase activity (firefly/renilla) was then calculated for each construct. Results are the mean ± SEM of three experiments.

identified in this study are located in these critical regions, it is difficult to assume how severe their effects might be. However, molecular modeling of the structure of bovine NRAMP1 (Triantaphyllopoulos et al., 2019) using the crystal structures of three different prokaryotic NRAMP orthologs (ScaDMT, DraNramp, and EcoDMT) (Ehrnstorfer et al., 2014; Bozzi et al., 2016; Ehrnstorfer et al., 2017) as comparison suggested that TMD 1, 2, 6, and 7 form a bundle of α-helices involved in ion transportation, while TMD 3, 4, 5, 8, 9, and 10 form a scaffold that supports this bundle. Therefore, it is possible that the change in TDM 8 caused by SNP4 potentially alters the structure of the scaffold and therefore the function of bovine NRAMP1.

A comparison of the *SLC11A1* coding region between the three cattle breeds used in this study revealed that the Sahiwal cattle (*B. indicus*) showed a higher genetic variation compared to BS and HF cattle (*B. taurus*), with four out of the five SNP alleles occurring at significantly higher frequencies in the Sahiwal cattle. Despite this observed genetic variation, the molecular phylogenetic analysis unexpectedly demonstrated no divergence between the two cattle sub-species. This might indicate that certain polymorphisms in the bovine *SLC11A1* gene are being maintained in cattle populations through balancing selection, a suggestion that has also been described for the human *SLC11A1* gene (Blackwell and Searle, 1999; Searle and Blackwell, 1999; Sanjeevi et al., 2000); however, a larger study across global cattle populations would be required to confirm this. Therefore, differences in the frequency of alleles at maintained polymorphic sites within the *SLC11A1* gene could be contributing to previously described differences in resistance to *M. bovis* between *B. indicus* and *B. taurus* breeds (Ameni et al., 2007; Vordermeier et al., 2012). Previous studies also suggested that BS cattle are more resistant to bacterial pathogens compared to HF cattle (Rupp and Boichard, 2003; Gibson et al., 2016). However, our current work does not allow for an assumption to be made whether this is in part due to SNPs in NRAMP1, as no significant differences in *SLC11A1* SNP allele frequencies were observed between these two cattle breeds, potentially due to the small sample size. SNP4 (c.1066C>G, rs109453173) was the only polymorphism present in all three cattle breeds and therefore warranted further investigation. The frequency of the alternative allele at this SNP was found to be highest in the Sahiwal cattle, thought to be more resistant to bTB (Ameni et al., 2007; Vordermeier et al., 2012), and lowest in HF cattle, which seem to be more susceptible to bTB (Ameni et al., 2007; Vordermeier et al., 2012). Interestingly, this SNP has previously been associated with resistance/susceptibility to bTB in Chinese Holstein cattle, where the presence of the alternative allele in the CG and GG genotypes appeared to confer resistance to *M. bovis* infection (Liu et al., 2017).

Cellular lysates from MDMs did not show any differences in the NRAMP1 concentration between MDM generated from BS or HF cattle. This result is consistent with our analysis regarding the *SLC11A1* gene polymorphisms, that also showed no significant difference between these two cattle breeds. Unfortunately, due to import restriction based on the occurrence of Foot-and-Mouth Disease in Pakistan, PBMCs from Sahiwal cattle were not available for this study. Since SNP4 was the only polymorphism present in all three cattle breeds, we compared the impact between three different genotypes (CC, CG, and GG) identified for this polymorphism on MDM NRAMP1 concentration. Combining NRAMP1 values obtained for carriers of the CG and GG genotypes showed that the presence of the G allele was significantly associated with increased NRAMP1 expression, compared to the CC genotype. Previous studies showed that NRAMP1 has the ability to lower the phagosomal iron concentrations, thus impacting on survival and growth of intracellular pathogens, such as *M. bovis*, by depriving them of this essential nutrient (Forbes and Gros, 2001; Cellier et al., 2007). Therefore, it is tempting to speculate that cattle carrying the alternative G allele have an increased resistance to bTB through increased NRAMP1 expression in their macrophages, which would be supported by the results published recently in Chinese HF cattle (Liu et al., 2017). In this study, the authors indeed describe an increased resistance to *M. bovis* in cattle carrying the alternative G allele. In addition, we found the frequency of this G allele to be highest in DNA of *B. indicus* Sahiwal cattle, which also seem to have an increased bTB resistance (Ameni et al., 2007; Vordermeier et al., 2012).

A comparison of the number of GT repeats in the MS1 microsatellite present in the 3′UTR of *SLC11A1* [c.1647+61GT(10_13)] found no significant difference between the three cattle breeds included in this study. Our result differs from that of a previous study describing an increased number of genotypes for the MS1 microsatellite in *B. indicus* cattle breeds compared to *B. taurus* breeds (Hasenauer et al., 2013). Conflicting data also exists regarding the role this microsatellite plays in resistance to bTB. Indeed, whereas a study performed in African zebu cattle found that higher numbers of GT repeats in the MS1 microsatellite were associated with a reduction in traits indicative of bTB (Kadarmideen et al., 2011), a study using Argentine cattle found no association between different lengths of this microsatellite and response to the tuberculin skin test (Hasenauer et al., 2018). If the MS1 microsatellite in the 3′UTR of the bovine *SLC11A1* gene indeed plays a biological role in NRAMP1 expression, it is likely to be due to differences in post-transcriptional regulation of the mRNA, possibly by influencing mRNA stability, similarly as described for other molecules (Citores et al., 2004; Chen et al., 2007). Our attempt to investigate this using a luciferase reporter construct in an *in vitro* approach did not identify a significant difference in luciferase expression in either CHO or RAW264.7 cells, suggesting that different numbers of GT repeats in the *SLC11A1* MS1 microsatellite did not impact on NRAMP1 protein expression in this system. Whether the results would differ if an appropriate bovine macrophage cell line is employed remains to be seen.

CONCLUSION

In conclusion, our study demonstrated that variations in the bovine *SLC11A1* gene are associated with a functional difference

in the expression of NRAMP1 in bovine MDM. Since NRAMP1 is known to regulate the level of iron within phagosomes, this provides further evidence that *SLC11A1*/NRAMP1 might indeed influence the susceptibility to intracellular pathogens, and in particular *M. bovis*. However, bTB is a complex disease which is likely to be influenced by polymorphisms in multiple genes, particularly those related to the immune system and host-pathogen interactions.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The use of animals for blood collection was approved by the Royal Veterinary College (RVC) Ethical Committee, which subsequently resulted in granting Home Office License Number PPL7009059.

AUTHOR CONTRIBUTIONS

AH, RG, CE, PM, CL, and M-CB performed the experiments. AH, RG, CE, and MS analyzed the data. TT, MZS, and TC provided the valuable reagents. AH wrote the manuscript. BV-R and DW designed the study and revised the manuscript. All the authors read and approved the final manuscript.

FUNDING

The work described in this manuscript was funded through internal grants to DW. Genomic regions and candidate genes identified in the present study provide an opportunity to further understand pathways critical to cattle susceptibility to bTB.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01420/full#supplementary-material

REFERENCES

Allen, A. R., Minozzi, G., Glass, E. J., Skuce, R. A., McDowell, S. W., Woolliams, J. A., et al. (2010). Bovine tuberculosis: the genetic basis of host susceptibility. *Proc. Biol. Sci.* 277, 2737–2745. doi: 10.1098/rspb.2010.0830

Ameni, G., Aseffa, A., Engers, H., Young, D., Gordon, S., Hewinson, G., et al. (2007). High prevalence and increased severity of pathology of bovine tuberculosis in Holsteins compared to zebu breeds under field cattle husbandry in central Ethiopia. *Clin. Vaccine Immunol.* 14, 1356–1361. doi: 10.1128/CVI.00205-07

Barthel, R., Piedrahita, J. A., McMurray, D. N., Payeur, J., Baca, D., Suarez Guemes, F., et al. (2000). Pathologic findings and association of *Mycobacterium bovis* infection with the bovine NRAMP1 gene in cattle from herds with naturally occurring tuberculosis. *Am. J. Vet. Res.* 61, 1140–1144.

Bermingham, M. L., Bishop, S. C., Woolliams, J. A., Pong-Wong, R., Allen, A. R., McBride, S. H., et al. (2014). Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. *Heredity (Edinb)* 112, 543–551. doi: 10.1038/hdy.2013.137

Blackwell, J. M., and Searle, S. (1999). Genetic regulation of macrophage activation: understanding the function of Nramp1 (=Ity/Lsh/Bcg). *Immunol. Lett.* 65, 73–80. doi: 10.1016/s0165-2478(98)00127-8

Blackwell, J. M., Searle, S., Goswami, T., and Miller, E. N. (2000). Understanding the multiple functions of Nramp1. *Microbes Infect.* 2, 317–321.

Bozzi, A. T., Bane, L. B., Weihofen, W. A., Singharoy, A., Guillen, E. R., Ploegh, H. L., et al. (2016). crystal structure and conformational change mechanism of a bacterial nramp-family divalent metal transporter. *Structure* 24, 2102–2114. doi: 10.1016/j.str.2016.09.017

Buddle, B. M., Vordermeier, H. M., Chambers, M. A., and de Klerk-Lorist, L. M. (2018). Efficacy and safety of BCG vaccine for control of tuberculosis in domestic livestock and wildlife. *Front. Vet. Sci.* 5:259. doi: 10.3389/fvets.2018.00259

Cellier, M. F., Courville, P., and Campion, C. (2007). Nramp1 phagocyte intracellular metal withdrawal defense. *Microbes Infect.* 9, 1662–1670. doi: 10.1016/j.micinf.2007.09.006

Chambers, M. A., Carter, S. P., Wilson, G. J., Jones, G., Brown, E., Hewinson, R. G., et al. (2014). Vaccination against tuberculosis in badgers and cattle: an overview of the challenges, developments and current research priorities in Great Britain. *Vet. Rec.* 175, 90–96. doi: 10.1136/vr.102581

Chen, T. M., Kuo, P. L., Hsu, C. H., Tsai, S. J., Chen, M. J., Lin, C. W., et al. (2007). Microsatellite in the 3' untranslated region of human fibroblast growth factor 9 (FGF9) gene exhibits pleiotropic effect on modulating FGF9 protein expression. *Hum. Mutat.* 28, 98. doi: 10.1002/humu.9471

Cheng, Y., Huang, C., and Tsai, H. J. (2015). Relationship of Bovine SLC11A1 (Formerly NRAMP1) Polymorphisms to the risk of Bovine tuberculosis in Holstein cattle. *J. Veterinar. Sci. Technol.* 6:5. doi: 10.4172/2157-7579.1000247

Citores, M. J., Rua-Figueroa, I., Rodriguez-Gallego, C., Durantez, A., Garcia-Laorden, M. I., Rodriguez-Lozano, C., et al. (2004). The dinucleotide repeat polymorphism in the 3'UTR of the CD154 gene has a functional role on protein expression and is associated with systemic lupus erythematosus. *Ann. Rheum. Dis.* 63, 310–317. doi: 10.1136/ard.2003.006148

Cohen, A., Nevo, Y., and Nelson, N. (2003). The first external loop of the metal ion transporter DCT1 is involved in metal ion binding and specificity. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10694–10699. doi: 10.1073/pnas.1934572100

Ehrnstorfer, I. A., Geertsma, E. R., Pardon, E., Steyaert, J., and Dutzler, R. (2014). Crystal structure of a SLC11 (NRAMP) transporter reveals the basis for transition-metal ion transport. *Nat. Struct. Mol. Biol.* 21, 990–996. doi: 10.1038/nsmb.2904

Ehrnstorfer, I. A., Manatschal, C., Arnold, F. M., Laederach, J., and Dutzler, R. (2017). Structural and mechanistic basis of proton-coupled metal ion transport in the SLC11/NRAMP family. *Nat. Commun.* 8:14033. doi: 10.1038/ncomms14033

Feng, J., Li, Y., Hashad, M., Schurr, E., Gros, P., Adams, L. G., et al. (1996). Bovine natural resistance associated macrophage protein 1 (Nramp1) gene. *Genome Res.* 6, 956–964. doi: 10.1101/gr.6.10.956

Finlay, E. K., Berry, D. P., Wickham, B., Gormley, E. P., and Bradley, D. G. (2012). A genome wide association scan of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle. *PLoS One* 7:e30545. doi: 10.1371/journal.pone.0030545

Forbes, J. R., and Gros, P. (2001). Divalent-metal transport by NRAMP proteins at the interface of host-pathogen interactions. *Trends Microbiol.* 9, 397–403. doi: 10.1016/s0966-842x(01)02098-4

Gibson, A. J., Woodman, S., Pennelegion, C., Patterson, R., Stuart, E., Hosker, N., et al. (2016). Differential macrophage function in Brown Swiss and Holstein Friesian cattle. *Veterinar. Immunol. Immunopathol.* 181, 15–23. doi: 10.1016/j.vetimm.2016.02.018

Hasenauer, F. C., Caffaro, M. E., Czibener, C., Comerci, D., Poli, M. A., and Rossetti, C. A. (2013). Genetic analysis of the 3' untranslated region of the bovine SLC11A1 gene reveals novel polymorphisms. *Mol. Biol. Rep.* 40, 545–552. doi: 10.1007/s11033-012-2091-0

Hasenauer, F. C., Garbaccio, S. G., Caffaro, M. E., Garro, C., Huertas, P., Poli, M. A., et al. (2018). Exploring the association between polymorphisms at 3'UTR SLC11A1 gene microsatellites and resistance to tuberculosis: a case-control study in *Bos taurus* dairy cattle. *Livestock Sci.* 210, 1–7. doi: 10.1016/j.livsci.2018.01.012

Kadarmideen, H. N., Ali, A. A., Thomson, P. C., Muller, B., and Zinsstag, J. (2011). Polymorphisms of the SLC11A1 gene and resistance to bovine tuberculosis in African Zebu cattle. *Anim. Genet.* 42, 656–658. doi: 10.1111/j.1365-2052.2011.02203.x

Lam-Yuk-Tseung, S., Govoni, G., Forbes, J., and Gros, P. (2003). Iron transport by Nramp2/DMT1: pH regulation of transport by 2 histidines in transmembrane domain 6. *Blood* 101, 3699–3707. doi: 10.1182/blood-2002-07-2108

le Roex, N., van Helden, P. D., Koets, A. P., and Hoal, E. G. (2013). Bovine TB in livestock and wildlife: what's in the genes? *Physiol. Genomics* 45, 631–637. doi: 10.1152/physiolgenomics.00061.2013

Liu, K., Zhang, B., Teng, Z., Wang, Y., Dong, G., Xu, C., et al. (2017). Association between SLC11A1 (NRAMP1) polymorphisms and susceptibility to tuberculosis in Chinese Holstein cattle. *Tuberculosis (Edinb)* 103, 10–15. doi: 10.1016/j.tube.2016.11.003

Michel, A. L., Muller, B., and van Helden, P. D. (2010). Mycobacterium bovis at the animal-human interface: a problem, or not? *Vet. Microbiol.* 140, 371–381. doi: 10.1016/j.vetmic.2009.08.029

Montalbetti, N., Simonin, A., Kovacs, G., and Hediger, M. A. (2013). Mammalian iron transporters: families SLC11 and SLC40. *Mol. Aspects Med.* 34, 270–287. doi: 10.1016/j.mam.2013.01.002

Nevo, Y. (2008). Site-directed mutagenesis investigation of coupling properties of metal ion transport by DCT1. *Biochim. Biophys. Acta* 1778, 334–341. doi: 10.1016/j.bbamem.2007.10.007

Nevo, Y., and Nelson, N. (2004). The mutation F227I increases the coupling of metal ion transport in DCT1. *J. Biol. Chem.* 279, 53056–53061. doi: 10.1074/jbc.M408398200

Nevo, Y., and Nelson, N. (2006). The NRAMP family of metal-ion transporters. *Biochim. Biophys. Acta* 1763, 609–620. doi: 10.1016/j.bbamcr.2006.05.007

Paixao, T. A., Ferreira, C., Borges, A. M., Oliveira, D. A., Lage, A. P., and Santos, R. L. (2006). Frequency of bovine Nramp1 (Slc11a1) alleles in Holstein and Zebu breeds. *Vet. Immunol. Immunopathol.* 109, 37–42. doi: 10.1016/j.vetimm.2005.07.018

Plant, J. E., Blackwell, J. M., O'Brien, A. D., Bradley, D. J., and Glynn, A. A. (1982). Are the Lsh and Ity disease resistance genes at one locus on mouse chromosome 1? *Nature* 297, 510–511. doi: 10.1038/297510a0

Raphaka, K., Matika, O., Sanchez-Molano, E., Mrode, R., Coffey, M. P., Riggio, V., et al. (2017). Genomic regions underlying susceptibility to bovine tuberculosis in Holstein-Friesian cattle. *BMC Genet.* 18:27. doi: 10.1186/s12863-017-0493-7

Ring, S. C., Purfield, D. C., Good, M., Breslin, P., Ryan, E., Blom, A., et al. (2019). Variance components for bovine tuberculosis infection and multi-breed genome-wide association analysis using imputed whole genome sequence data. *PLoS One* 14:e0212067. doi: 10.1371/journal.pone.0212067

Rupp, R., and Boichard, D. (2003). Genetics of resistance to mastitis in dairy cattle. *Vet. Res.* 34, 671–688. doi: 10.1051/vetres:2003020

Sanjeevi, C. B., Miller, E. N., Dabadghao, P., Rumba, I., Shtauvere, A., Denisova, A., et al. (2000). Polymorphism at NRAMP1 and D2S1471 loci associated with juvenile rheumatoid arthritis. *Arthritis Rheum.* 43, 1397–1404. doi: 10.1002/1529-0131(200006)43:6<1397::AID-ANR25<3.0.CO;2-6

Schiller, I., Oesch, B., Vordermeier, H. M., Palmer, M. V., Harris, B. N., Orloski, K. A., et al. (2010). Bovine tuberculosis: a review of current and emerging diagnostic techniques in view of their relevance for disease control and eradication. *Transbound Emerg. Dis.* 57, 205–220. doi: 10.1111/j.1865-1682.2010.01148.x

Searle, S., and Blackwell, J. M. (1999). Evidence for a functional repeat polymorphism in the promoter of the human NRAMP1 gene that correlates with autoimmune versus infectious disease susceptibility. *J. Med. Genet.* 36, 295–299.

Skamene, E., Gros, P., Forget, A., Kongshavn, P. A., St Charles, C., and Taylor, B. A. (1982). Genetic regulation of resistance to intracellular pathogens. *Nature* 297, 506–509. doi: 10.1038/297506a0

Triantaphyllopoulos, K. A., Baltoumas, F. A., and Hamodrakas, S. J. (2019). Structural characterization and molecular dynamics simulations of the caprine and bovine solute carrier family 11 A1 (SLC11A1). *J. Comput. Aided Mol. Des.* 33, 265–285. doi: 10.1007/s10822-018-0179-x

UK Government Bovine TB (2020). Available online at: https://www.gov.uk/environment/bovine-tuberculosis (accessed May 21, 2020).

Vazquez-Flores, F., Alonso, R., Villegas-Sepulveda, N., Arriaga, C., Pereira-Suarez, A. L., Mancilla, R., et al. (2006). A microsatellite study of bovine solute carrier family 11 a1 (Slc11a1) gene diversity in Mexico in relation to bovine tuberculosis. *Genetics Mol. Biol.* 29, 503–507. doi: 10.1590/S1415-47572006000300019

Vordermeier, M., Ameni, G., Berg, S., Bishop, R., Robertson, B. D., Aseffa, A., et al. (2012). The influence of cattle breed on susceptibility to bovine tuberculosis in Ethiopia. *Comp. Immunol. Microbiol. Infect. Dis.* 35, 227–232. doi: 10.1016/j.cimid.2012.01.003

Warner, C. M., Meeker, D. L., and Rothschild, M. F. (1987). Genetic control of immune responsiveness: a review of its use as a tool for selection for disease resistance. *J. Anim. Sci.* 64, 394–406. doi: 10.2527/jas1987.642394x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

REVIEW
published: 10 July 2020
doi: 10.3389/fgene.2020.00677

Combining Multiple Approaches and Models to Dissect the Genetic Architecture of Resistance to Infections in Fish

Clémence Fraslin[1†], Edwige Quillet[1], Tatiana Rochat[2], Nicolas Dechamp[1], Jean-Francois Bernardet[2], Bertrand Collet[2], Delphine Lallias[1] and Pierre Boudinot[2]**

[1] INRAE, AgroParisTech, GABI, Université Paris-Saclay, Jouy-en-Josas, France, [2] INRAE, UVSQ, VIM, Université Paris-Saclay, Jouy-en-Josas, France

OPEN ACCESS

Edited by:
Androniki Psifidi, Royal Veterinary College (RVC), United Kingdom

Reviewed by:
Timothy D. Leeds, United States Department of Agriculture (USDA), United States
Elisa Domínguez-Hüttinger, National Autonomous University of Mexico, Mexico

***Correspondence:**
Delphine Lallias
Delphine.Lallias@inrae.fr
Pierre Boudinot
Pierre.Boudinot@inrae.fr

†Present address:
Clémence Fraslin, The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, United Kingdom

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *06 March 2020*
Accepted: *02 June 2020*
Published: *10 July 2020*

Citation:
Fraslin C, Quillet E, Rochat T, Dechamp N, Bernardet J-F, Collet B, Lallias D and Boudinot P (2020) Combining Multiple Approaches and Models to Dissect the Genetic Architecture of Resistance to Infections in Fish. Front. Genet. 11:677. doi: 10.3389/fgene.2020.00677

Infectious diseases represent a major threat for the sustainable development of fish farming. Efficient vaccines are not available against all diseases, and growing antibiotics resistance limits the use of antimicrobial drugs in aquaculture. It is therefore important to understand the basis of fish natural resistance to infections to help genetic selection and to develop new approaches against infectious diseases. However, the identification of the main mechanisms determining the resistance or susceptibility of a host to a pathogenic microbe is challenging, integrating the complexity of the variation of host genetics, the variability of pathogens, and their capacity of fast evolution and adaptation. Multiple approaches have been used for this purpose: (i) genetic approaches, QTL (quantitative trait loci) mapping or GWAS (genome-wide association study) analysis, to dissect the genetic architecture of disease resistance, and (ii) transcriptomics and functional assays to link the genetic constitution of a fish to the molecular mechanisms involved in its interactions with pathogens. To date, many studies in a wide range of fish species have investigated the genetic determinism of resistance to many diseases using QTL mapping or GWAS analyses. A few of these studies pointed mainly toward adaptive mechanisms of resistance/susceptibility to infections; others pointed toward innate or intrinsic mechanisms. However, in the majority of studies, underlying mechanisms remain unknown. By comparing gene expression profiles between resistant and susceptible genetic backgrounds, transcriptomics studies have contributed to build a framework of gene pathways determining fish responsiveness to a number of pathogens. Adding functional assays to expression and genetic approaches has led to a better understanding of resistance mechanisms in some cases. The development of knock-out approaches will complement these analyses and help to validate putative candidate genes critical for resistance to infections. In this review, we highlight fish isogenic lines as a unique biological material to unravel the complexity of

host response to different pathogens. In the future, combining multiple approaches will lead to a better understanding of the dynamics of interaction between the pathogen and the host immune response, and contribute to the identification of potential targets of selection for improved resistance.

Keywords: fish, disease resistance, QTL mapping, transcriptomics, host–pathogen interactions, immunity, functional assays, fish isogenic lines

INTRODUCTION

The outcome of infections is determined by multiple factors including the direct tissue damages caused by the pathogenic microbe (or parasite), the capacity of the immune system to prevent or limit its multiplication, and the adverse effects of the inflammation induced by their confrontation. Infectious diseases exert powerful selection pressures on host populations, affecting survival, growth, and fertility of the host. Host–pathogen interactions are therefore shaped by a dynamic co-evolution of the virulence factors of the pathogen and the defense mechanisms of the host. Within the same host species, this co-evolution contributes to marked differences in susceptibility between host populations from disease-free areas or from areas where the pathogen is present (Hedrick et al., 2003). Emergent diseases caused by pathogens that have recently crossed a host species barrier are often devastating because their virulence has not been attenuated by these evolutionary processes. At the population level, it is important to note that commensalism is not always the ideal final equilibrium, and that fast transmission or high durability of the pathogen can promote an evolution of host–parasite relationships toward a state of severe disease (Ewald, 1983).

Multiple anti-infectious host defense mechanisms control the pathogenesis. They can act at every step of host–pathogen interactions, from the entry into host organism/cell to the production and dissemination of the pathogen. All mechanisms of immunity can potentially affect the resistance or susceptibility to a given pathogen. Furthermore, many other (i.e., non-immune) mechanisms like virus receptors or tissue tropism may affect the efficiency of the pathogen biology or the beneficial/harmful result of the host response.

The identification of the main mechanisms determining the resistance or susceptibility of a host to a pathogenic microbe is therefore extremely challenging, integrating the complexity of the variation of host genetics and the variability of pathogens and their capacity of fast evolution and adaptation. Genetic approaches such as QTL (quantitative trait loci) mapping make it possible to disentangle these complex interactions by providing information on the genetic determinism of host resistance and, ultimately, on the underlying genetic variations (and thus mechanisms) that make a host resistant or susceptible (Houston et al., 2020). Functional assays (comparative transcriptome analysis, *in vitro* culture models) also provide insights into mechanisms of interaction between the pathogen and its host and can help in identifying genes that play a key role in host response to infection. Combining such positional and functional approaches is very promising, as exemplified by the identification of genes involved in intrinsic restriction of retroviruses: the gene Fv1 (Friend-virus susceptibility gene-1) responsible for the susceptibility of mice to Murine Leukemia Virus was identified by Stoye and colleagues using a positional cloning strategy (Best et al., 1996), while the gene *trim5* responsible for the resistance of rhesus cells to HIV-1 was cloned in parallel using a cDNA expression library by direct selection of virus-resistant transfected cells (Stremlau et al., 2004). In this work, we did not address the interactions between fish resistance or susceptibility, and the variation of virulence within pathogen species. It is certainly an important—and understudied—issue, but the discussion of these mechanisms is beyond the scope of the present work.

Infectious diseases remain a major threat for the development and the economic and environmental efficiency of fish farming. Bacterial diseases can be treated by antibiotics, but such treatments lead to the development of resistant microbes, which reduce treatment efficiency and represent a significant issue for animal and human health. Vaccines can efficiently protect fish against infectious diseases, and indeed allowed a drastic reduction of antibiotic treatments in Nordic salmon aquaculture. However, vaccines are not available against all diseases. Moreover, they are generally efficient when administered by injection, which is not possible for small individuals. There are no vaccines against many viral diseases and no vaccine protecting against fish parasites (Collins et al., 2019; Ma et al., 2019). Recent discoveries on probiotics raise hope for beneficial adjustment of gut microbiota, but no such treatment has been fully validated to date (Conti et al., 2014). Hence, genetic selection of fish with improved resistance to the main infectious diseases in a given environment remains a highly sought-after objective in aquaculture (Houston et al., 2020).

Fish lifestyle in aquaculture conditions has an important impact on the interactions between farmed fish and their pathogens. The concentrations of animals in cages or small water bodies allow major outbreaks (Lafferty et al., 2015). Also, this enhances the transmission efficiency, hence allowing pathogens to evolve higher virulence and pathogenicity. Importantly, as for other farmed species, "domestication" and selection for positive

Abbreviations: Ag, antigen; AGD, amoebic gill disease; BCWD, bacterial cold water disease; BS, binary survival; CaHV, Carp herpes virus; CMS, cardiomyopathy syndrome; CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; eQTL, expression QTL; ESC, enteric septicemia in catfish; GWAS, genome-wide association studies; IFN, interferon; IPNV, infectious pancreatic necrosis virus; ISAV, infectious salmon anemia virus; INHV, infectious hematopoietic necrosis virus; KO, knock-out; LA, linkage analysis; lncRNA, long non-coding RNA; MHC, major histocompatibility complex; NNV, nervous necrosis virus; PMCV, piscine myocarditis virus; QTL, quantitative trait loci; reQTL, RNA-eQTL; SNP, single-nucleotide polymorphism; TCR, T-cell receptor; TH, T helper; TLR, Toll like receptor; TTD, time to death; VHS, viral hemorrhagic septicemia; VHSV, viral hemorrhagic septicemia virus; VNN, viral nervous necrosis.

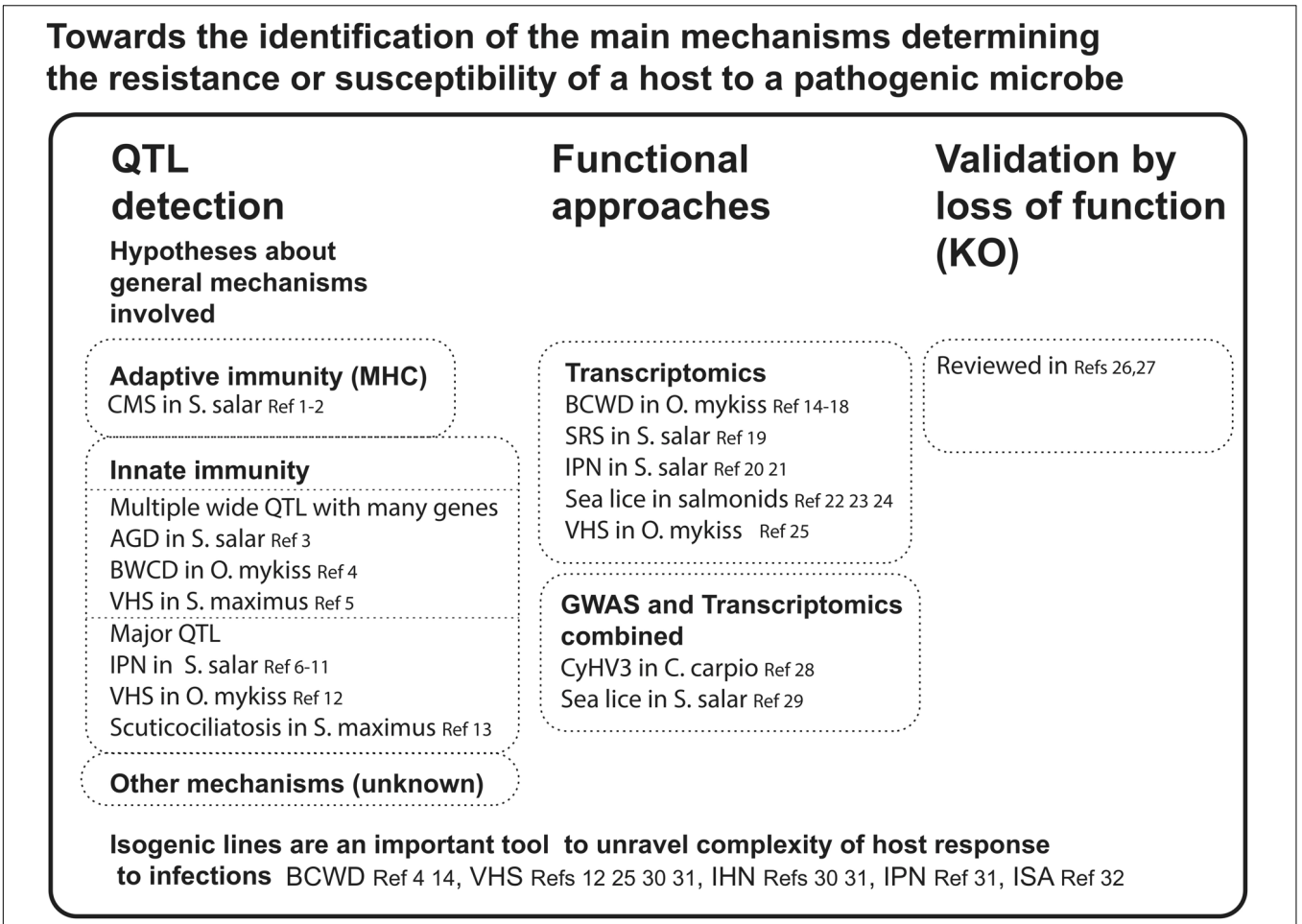


FIGURE 1 | Toward the identification of the main mechanisms determining the resistance or susceptibility of a host to a pathogenic microbe. CMS, cardiomyopathy syndrome; AGD, amoebic gill disease; BCWD, bacterial cold water disease; VHS, viral hemorrhagic septicemia; IPN, infectious pancreatic necrosis; INH, infectious hematopoietic necrosis; ISA, infectious salmon anemia; SRS, salmonid rickettsial septicemia; CyHV-3, cyprinid herpesvirus; QTL, quantitative trait loci; GWAS, genome-wide association study. References: [1]Boison et al., 2019a; [2]Hillestad and Moghadam, 2019; [3]Boison et al., 2019b; [4]Fraslin et al., 2018; [5]Rodríguez-Ramilo et al., 2014; [6]Houston et al., 2008; [7]Houston et al., 2010; [8]Houston et al., 2012; [9]Gheyas et al., 2010; [10]Moen et al., 2009; [11]Moen et al., 2015; [12]Verrier et al., 2013a; [13]Saura et al., 2019; [14]Langevin et al., 2012; [15]Marancik et al., 2015; [16]Paneru et al., 2016; [17]Zwollo et al., 2017; [18]Moore et al., 2018; [19]Dettleff et al., 2015; [20]Reyes-López et al., 2015; [21]Robledo et al., 2016; [22]Sutherland et al., 2014; [23]Braden et al., 2015; [24]Holm et al., 2015; [25]Verrier et al., 2018; [26]Dehler et al., 2019; [27]Gratacap et al., 2019; [28]Palaiokostas et al., 2018b; [29]Robledo et al., 2019; [30]Verrier et al., 2012; [31]Verrier et al., 2013b; [32]Biacchesi et al., 2007[32].

traits such as fast growth, food efficiency etc. at the industrial scale of modern aquaculture has led to a genetic homogenization of fish stocks and may have been detrimental for resistance to (at least some) pathogens.

In a context of globalization leading to severe problems due to invasive species (including pathogens), selection of resistance to diseases that are important locally may not be sufficient on the long term. The production of robust fish constituting interesting compromises between specific resistances and a general capacity to deal with multiple aggressors might be the ultimate aim.

Tolerance, i.e., the ability to limit pathogenesis of a given pathogen burden, is another important parameter of fish/pathogen interactions and survival, which shows genetic variability within animal populations (Råberg et al., 2007). Only a few reports have been published on such mechanisms in fish. For example, increased tolerance to viral infection by chikungunya virus has been shown in zebrafish in which the *inos* gene had been knocked down (Levraud et al., 2014). Since both mechanisms and genetic variability of tolerance remains poorly known in fish, this review does not cover this subject.

In this review, we focus on the different approaches developed to better understand and take benefit from resistance mechanisms against pathogens in fish (**Figure 1**). We first discuss examples of identification of QTL of resistance to pathogenic microbes. We then focus on transcriptomics studies and functional assays that describe host–pathogen interactions. The cross information from these two approaches allows a better understanding of the dynamics of interaction between the pathogen and the host immune response and contributes to the identification of potential targets of selection for improved resistance. Finally, we emphasize the importance of combining several approaches (QTL mapping and resequencing or gene

expression) and provide a few perspectives on new approaches (such as eQTL, reQTL, and other combination of QTL with expression studies) that should help to identify the genes responsible for resistance to pathogens in fish.

DESCRIBING THE GENETIC ARCHITECTURE OF RESISTANCE TRAITS IN FISH TO IDENTIFY UNDERLYING GENETIC MECHANISMS

Mapping of QTL Associated With Disease Resistance in Fish

One approach to better understand the host–pathogen interaction in fish is to describe the genetic architecture of traits related to host response to diseases. The detection and mapping of QTL associated with resistance traits after infections provides information along the genome on the position and effect of genetic variants associated with a trait of interest. QTL mapping relies on preferential association between known molecular genetic markers (microsatellites or SNPs) and an unknown causative mutation that affect the trait. The two main types of analysis used to detect QTL are (1) linkage analysis (LA) in full or half-sib families with low-density markers spread throughout the genome and (2) genome-wide association studies (GWAS) performed in large populations using thousands of SNPs. In fish, numerous QTL studies aiming at describing the genetic architecture of disease-related traits have been published in the past 20 years with an intensification since 2013: we found more than 60 studies reporting QTL associated with different resistance traits in 16 fish species infected with 30 pathogens (see **Tables 1**, **2** and **Supplementary Table 1**).

In those studies, fish were either experimentally infected by immersion (29 studies) or by intraperitoneal, intramuscular, or intracelomical injections (27 studies). In six studies (Grimholt et al., 2003; Moen et al., 2009; Palaiokostas et al., 2018b; Robledo et al., 2018; Saura et al., 2019; Tadmor-Levi et al., 2019), infection was performed through cohabitation with previously infected fish. Finally, resistance phenotypes after a natural outbreak in farms were analyzed in two studies (Houston et al., 2008; Fraslin et al., 2019). Four studies combined resistance traits measured after two types of infection: natural outbreak and injection (Boison et al., 2019a), cohabitation and injection (Gonen et al., 2015), or immersion and injection (Fraslin et al., 2018; Rodríguez et al., 2019). While cohabitation or immersion routes are expected to be more representative of the natural route of infection and challenge the barriers to pathogen entry in the host, injection warrants the administration of a given dose of pathogen and allows addressing other (later) resistance mechanisms.

The main resistance trait investigated was the binary survival (BS, dead *vs.* alive status of the fish at the end of the challenge, in 24 studies), which is the principal outcome of infection and a relatively easy trait to measure. In many cases, this phenotype was refined with the analysis of time to death (TTD), measured as the interval between infection and death (analyzed alone in 5 studies and with BS in 23 studies). In three studies, a survival analysis using a Cox model that combined information of TTD and BS was performed (Verrier et al., 2013a; Fraslin et al., 2018; Saura et al., 2019). When possible, for all diseases caused by parasite and some viral diseases, the record of pathogen load was used as a resistance trait (in nine studies). Disease resistance was also assessed using a lesion score on gills for the amoebic gill disease (AGD) (Robledo et al., 2018; Boison et al., 2019b) or on heart for the cardiomyopathy syndrome (CMS) (Boison et al., 2019a; Hillestad and Moghadam, 2019). Some studies were able to refine resistance traits based on the appearance of disease signs and on detailed information about fish fitness. Endurance was measured as the time interval between the infection (Fraslin et al., 2018) or the first disease sign (Saura et al., 2019) and the death of the fish. In Saura et al. (2019), they also distinguished between resistance (time interval between infection and first disease sign) and resilience (measured as time between infection and death). Reliable and quantifiable resistance phenotypes are key for QTL analysis. In practice, detailed measurements of resistance (as survival, pathogen load, lesion score, or TTD) require a lot of effort (monitoring of fish several times a day) and are costly to perform on a high number of fish, thus limiting the detection power of QTL mapping studies. For an extensive comment on how to measure different resistance traits and the different types of experimental challenges or natural outbreak used in aquatic species, one could refer to Robinson et al. (2017).

The fish used in the majority of the studies presented in this review are half- or full-sibs from commercial populations produced by factorial mating designs (48 studies). In some studies, fish were produced using parents from lines with divergent phenotypes, by simple cross (F1 or F2 in 13 studies), backcross (BC in 9 studies) or gynogenesis, producing doubled haploid fish (DH in 2 studies). Those fish with divergent phenotypes can come from either lines that have been divergently selected for generations (e.g., Ozaki et al., 2001; Vallejo et al., 2014a,b) or from isogenic lines. When available, isogenic lines are useful tools to investigate molecular bases of complex traits as the genetic homogeneity within a line allows the repetition of measurement on the same genetic background through time and in different environments. When multiple isogenic lines are available, the genomic variability between lines is important [theoretically twice the variability of the population they were produced from Falconer and Mackay (1996)] and they can exhibit a wide range of contrasted phenotypes useful for QTL detection studies. However, in practice, only a few lines are available, reducing the variability range available for such studies. The use of doubled haploid fish (fully homozygous) increases the power of genetic analysis for a given number of fish (Martinez et al., 2002) but limits the fine mapping of QTL using LA approaches, detecting wide QTL. However, the production of DH fish is complicated to achieve with low survival rates (Komen and Thorgaard, 2007) and thus was only used in a limited number of studies on rainbow trout (Verrier et al., 2013a, 2018; Fraslin et al., 2018). Selection of fish from families with extreme phenotypes is a well-known approach to increase QTL detection power (Lander and Botstein, 1989) that has been used in many studies (e.g., in Houston et al., 2008, 2010; Geng et al., 2015; Palti et al., 2015; Wang et al., 2019).

TABLE 1 | Summary of the relevant QTL studies pointing toward the implication of adaptive mechanisms in disease resistance.

Fish species	Disease	Challenge type	Experimental design (genotyped fish in the analysis)	Genetic molecular markers	Resistance traits	Number of detected QTL	Size of detected QTL	% of variance explained by QTL	Publication
Atlantic salmon (*S. salar*)	CMS	Injection (ip) and natural outbreak	571 and 4312 and 901 (FS and HF in three populations)	55,735 SNPs	BS and lesion score	2 QTL	0.2–31 Mb	8–31% of Vg	Boison et al., 2019a
Atlantic salmon (*S. salar*)	CMS	Injection (ip)	1144 fish (60 FS)	48K SNPs	Lesion score and viral load	2 QTL	320 kb (Ssa12) and 19 Mb (Ssa27)	Up to 57% of Vg (Ssa27)	Hillestad and Moghadam, 2019
Atlantic salmon (*S. salar*)	Sea lice	Immersion	94 fish (3 FS)	SSR markers located within 3′ of MHC I (in two chromosomes)	Parasite count	2 QTL	28 cM	13% of Vp	Gharbi et al., 2009
Atlantic salmon (*S. salar*)	ISA	Injection (ip)	1031 fish (25 FS)	17 polymorphic repeats located within 3′ of MHC I and MHC II	BS	5 MHC I and II alleles			Grimholt et al., 2003
Atlantic salmon (*S. salar*)	Furunculosis	Cohabitation	1182 fish (33 FS)	17 polymorphic repeats located within 3′ of MHC I and MHC II	BS	7 MHC I and II alleles			Grimholt et al., 2003
Rainbow trout (*O. mykiss*)	IPN	Immersion	199 fish (BC and F1)	226 markers (microsatellites AFPLs, MHC I and II genes)	BS	9 QTL		8–15% of Vp	Ozaki et al., 2007
Turbot (*S. maximus*)	Scuticociliatosis	Injection (ic)	758 fish (4 FS)	211 microsatellites	BS and TTD	5–10 QTL	7–30 cM	7–22% of Vp	Rodríguez-Ramilo et al., 2013
Red sea bream (*P. major*)	RSI disease	Injection	120 and 400 fish (2 F1)	458 microsatellites	BS	1 QTL		12–31% of Vp	Sawayama et al., 2017

QTL, quantitative trait loci; CMS, cardiomyopathy syndrome; ISA, infectious salmon anemia; IPN, infectious pancreatic necrosis; RSI, red seabream iridovirus; ip, intraperitoneal; ic, intracelomical; BS, binary survival that takes into account the outcomes of the challenge as dead or alive; TTD, time to death, measured as the time (in days or hours) between the infection and the day of death of a fish; Vp, phenotypic variance of the resistance trait; Vg, genetic or genomic variance of the resistance trait; FS; full-sib; HS, half-Sib; BC, back-cross; SNP, single-nucleotide polymorphism; MHC, major histocompatibility complex; cM, centi-Morgan. Size of detected QTL corresponds to either confidence intervals when calculated or cumulated windows with significant SNPs.

TABLE 2 | Summary of the relevant QTL studies pointing toward the implication of innate or intrinsic mechanisms in disease resistance.

Fish species	Disease	Challenge type	Experimental design (genotyped fish in the analysis)	Genetic molecular markers	Resistance traits	Number of detected QTL	Size of detected QTL	% variance explained by QTL	Publication
Atlantic salmon (*S. salar*)	AGD	Immersion	1333 fish (136 FS)	53,109 SNPs	Lesion score	3 QTL + smaller QTL	6 kb–1.2 Mb	4.6–5.3% of Vg	Boison et al., 2019b
Rainbow trout (*O. mykiss*)	BCWD	Immersion, injection (im)	DH fish 195 for immersion, 115 for injection	2130 SNPs	BS, TTD, COX endurance	15 QTL	20–80 cM	7–18% of Vp	Fraslin et al., 2018
Atlantic salmon (*S. salar*)	IPN	Natural outbreak	584 fish (10 FS)	Less than 100 microsatellites	BS	4 QTL	10–69 cM	8.9–24.6% of Vp	Houston et al., 2008
Atlantic salmon (*S. salar*)	IPN	Immersion	1321 fish (10 FS)	7 microsatellites	BS and TTD	3 QTL	10 cM	50.9% of Vp for Ssa26	Houston et al., 2010
Atlantic salmon (*S. salar*)	IPN	Immersion	28 fish (2 FS) and 9,000 fish (400 FS)	6712 SNPs	BS	1 QTL	2 cM		Houston et al., 2012
Atlantic salmon (*S. salar*)	IPN	Immersion	634 fry (20 FS) over 2 mph	8 microsatellites	BS	1 major QTL on Ssa26	7 cM	0–43.9% of Vp	Gheyas et al., 2010
Atlantic salmon (*S. salar*)	IPN	Cohabitation	1896 fish (10 FS) at 10 mph	136 microsatellites	BS	3 QTL with 1 major QTL	4 cM	29% of Vp 83% of Vg	Moen et al., 2009
Atlantic salmon (*S. salar*)	IPN	Immersion	4140 fish (207 FS)	3 microsatellites	BS	1 major QTL on Ssa26			Moen et al., 2015
Turbot (*S. maximus*)	Scuticociliatosis	Cohabitation	1394 fish (36 FS)	18,125 SNPs	Resistance Endurance Resilience	4 QTL: 1 major and 3 minors	Up to 9.3 Mb	Up to 33% Vg	Saura et al., 2019
Turbot (*S. maximus*)	VHS	Injection (ic)	280 fish (3 FS)	93 microsatellites	BS and TTD	7 QTL	4–30 cM	3–14% Vp	Rodríguez-Ramilo et al., 2014
Rainbow trout (*O. mykiss*)	VHS	Immersion	1325 DH fish + 450 DH fish (VREFT)	131–142 microsatellites	VREFT and COX	7 QTL (1 major)	1 cM	Up to 65% of Vp	Verrier et al., 2013a

AGD, amoebic gill disease; BCWD, bacterial cold water disease; IPN, infectious pancreatic necrosis; VHS, viral hemorrhagic septicemia; BS, binary survival that takes into account the outcomes of the challenge as dead or alive; TTD, time to death, measured as the time (in days or hours) between the infection and the day of death of a fish; COX, survival analysis with a Cox model (Cox, 1972); VREFT, viral replication in excised tissue; mph, month post hatching; Vp, phenotypic variance of the resistance trait; Vg, genetic or genomic variance of the resistance trait; FS, full-sib families; DH, doubled homozygous fish; BC, back-cross; SNP, single-nucleotide polymorphism; QTL, quantitative trait loci; cM, centi-Morgan; im, intramuscular; ic, intracelomical. Size of detected QTL correspond to either confidence intervals when calculated or cumulated windows with significant SNPs.

Until 2014, the majority of QTL mapping studies were performed using LA approaches with a few hundred microsatellites markers and detected QTL locating within large confidence intervals on chromosomes. Such QTL contained hundreds of positional genes being potential candidate genes (Moen et al., 2007; Houston et al., 2008; Massault et al., 2011; Rodríguez-Ramilo et al., 2011; Vallejo et al., 2014a). From 2015, with the development of genotyping by sequencing techniques (Robledo et al., 2017) and of medium- to high-density SNP arrays in aquaculture species, tens or even hundreds of thousands of SNPs could be produced. Their use allowed an important narrowing of the confidence intervals of QTL detected by GWAS to less than hundreds of bp (Wang et al., 2017; Shi et al., 2018; Kong et al., 2019). This constitutes an important step regarding the use of QTL in selection and, together with the continuous improvement of the quality of genome sequence assembly in a number of species, in the identification of genetic variants involved in individual resistance variation.

Resistance Mechanisms Hypothesized From QTL Analyses in Fish

The identification of underlying mechanisms from QTL analysis is notoriously difficult as QTL mapping identifies polymorphisms associated with a trait that can be distant from the causal mutation. Immune mechanisms are obvious targets when looking for resistance mechanisms, but the complexity of immune response and the high number of immune genes greatly complicate the task. Additionally, pathogens (especially intracellular ones such as mycobacteria or viruses) interact with many cell proteins that are not necessarily connected to immunity, but may affect the susceptibility of the host. In the following subsection, we will comment on a few QTL studies that pointed toward adaptive (see also **Table 1**) or innate mechanisms (see also **Table 2**). However, in most studies, even when QTL have been identified, the precise underlying mechanisms remain unknown (see also **Supplementary Table 1**).

QTL Pointing Toward Adaptive Mechanisms of Resistance/Susceptibility

Specific recognition of antigens expressed by pathogens by lymphocyte receptors diversified through VDJ somatic rearrangements of their genes is the archetype of adaptive response. During lymphocyte differentiation, immunoglobulin or T-cell receptor (TCR) loci are subjected to random genomic rearrangements of V, D, and J genes, resulting in the expression of a unique antigen (Ag) receptor by each lymphocyte. During the differentiation of lymphocytes, selective antigen-specific processes lead to elimination of most autoreactive cells and to proliferation of mature T and B cells. Clones specific to pathogen epitopes are expanded during antigen-driven responses. While immunoglobulins can recognize free native Ag, TCRs are specific for peptides that are processed by the antigen processing pathways in specialized cells and that are presented at the surface by MHC (major histocompatibility complex) molecules. The diversity (i.e., polymorphism) of the peptide binding groove of MHC molecules determines the repertoire of peptides that can be made available to T cell responses. Hence, variations in resistance to pathogens depending on adaptive responses may correspond to variations (1) of Ag receptor genes, (2) of the machinery mediating the recombination, (3) of genes affecting lymphocyte biology, or (4) of genes affecting Ag processing or presentation, including *mhc* genes. The large variation of *mhc* genes and the specificity of the effects of such variations make them targets of choice for pathogen-specific resistance mechanisms. **Table 1** summarizes the relevant QTL studies pointing toward the implication of adaptive mechanisms in disease resistance (also, reviewed in Yamaguchi and Dijkstra, 2019).

Resistance associated to MHC

By genotyping polymorphic repeats located within the 3′UTRS of MHC class I and MHC class II genes, Grimholt et al. (2003) found highly significant associations between MHC class I (UBA) and class II (DAA-DAB) gene polymorphism and resistance to both a virus [infectious salmon anemia virus (ISAV)] and a bacterium (*Aeromonas salmonicida*) in Atlantic salmon. This was one of the first studies to establish the functional role of MHC molecules in disease resistance in fish, pointing toward adaptive mechanisms. A few low-resolution QTL mapping studies followed in Salmonids, specifically targeting MHC. Detailed information on the experimental design is available in **Table 1** and **Supplementary Table 1**. In rainbow trout, Ozaki et al. (2007) performed a linkage analysis (LA) using 226 marker loci that included MCH class Ia, Ib, and II genes, genotyped in 199 individuals. They detected nine QTL associated with IPNV (infectious pancreatic necrosis virus) resistance and showed that only MHC class Ib gene mapped into a QTL, but finer mapping was not performed. Investigating Atlantic salmon resistance to sea lice, Gharbi et al. (2009) restricted their QTL analysis to the two linkage groups containing MHC class I and MHC class II genes but failed to show a proximal effect of MHC regions on salmon lice abundance. In turbot, Rodríguez-Ramilo et al. (2013) detected several QTL associated with resistance and survival time to the parasite *Philasterides dicentrarchi* (causative agent of scuticociliatosis). This study was based on 211 microsatellites and included the mapping of the MHC II B gene in the reference turbot map. Interestingly, MHC II B mapped within the confidence interval of one of the QTL, supporting this gene as being a candidate gene requiring further investigation. However, due to the low number of markers (average of 3.6–8.1 markers per linkage group), the detected QTL were associated with wide confidence intervals. In red sea bream, Sawayama et al. (2017) constructed low-density sex-specific genetic maps and mapped a candidate gene (MHC class II B). The QTL analysis for resistance to RSIV (red sea bream iridovirus) detected one major QTL, closely located to MHC class II B candidate gene. Although, in those studies, further mapping/association studies would be often required to definitively confirm that it is indeed the *mhc* genes that are involved.

Adaptive immunity is likely involved in the resistance to piscine myocarditis virus in Atlantic salmon

CMS, caused by piscine myocarditis virus (PMCV), is a severe inflammatory cardiac disease that affects farmed Atlantic salmon

during the seawater production stage. With no commercial vaccine efficient against the virus, this disease causes major losses to the Norway salmon aquaculture industry (Hillestad and Moghadam, 2019). Two independent studies pointed toward the same genomic regions associated with resistance to the disease in different salmon populations from Norway, with several candidate genes shared and identified as involved in adaptive immunity (Boison et al., 2019a; Hillestad and Moghadam, 2019; see **Table 1** and **Supplementary Table 1** for details). Boison et al. (2019a) analyzed three populations from the Mowi breeding program. Resistance was assessed as BS for two populations infected during a natural outbreak, whereas in the absence of mortality in a third population (experimentally challenged by intraperitoneal injection of PMCV), resistance was assessed using a histology score of heart tissue. Hillestad and Moghadam (2019) analyzed a single population from SalmoBreed, challenged by intraperitoneal injection. As no mortality occurred during the challenge, viral loads were estimated in heart tissue by qRT-PCR and used as the resistance phenotypes. In the two studies, more than 500 fish (and up to 4300) per population were genotyped with a 55K SNP array. Despite being based on different genetic backgrounds and investigating resistance measured as different traits (BS, histopathological score, or viral load), both studies detected QTL located on the same two chromosomes, Ssa12 and Ssa27, suggesting that resistance to PMCV is controlled by similar regions within different populations. Interestingly, in both studies, GWAS performed on a finer phenotype detected QTL on SSa27 that were more significant and explained a higher proportion of genetic variance: 31.7% (based on histology scores of heart tissue) compared to 7.9–14% (based on BS) in Boison et al. (2019a); 57% (based on viral load) in Hillestad and Moghadam (2019). A fine phenotype might better reflect the true impact of the virus than BS and thus better reflect the true resistance of fish. Those studies highlight the importance of fine phenotyping for QTL mapping. Further examination to identify potential candidate genes involved in resistance to PMCV focused on 50 kb regions in Boison et al. (2019a) and on a 2 Mb-block in Hillestad and Moghadam (2019). On Ssa12, positional candidate genes involved in adaptive immunity included the *magi1* gene, known to impact the pathogenicity of coxsackievirus and adenovirus by regulating the amount of *car* receptors, in Boison et al. (2019a); two putative *MHC II antigen* genes and the *T cell transcription factor EB-like* gene (*tfeb*) in Hillestad and Moghadam (2019). On Ssa27, the regions investigated contained several genes with functions in antigen processing and presentation (both studies), as well as genes involved in the inhibition of viral cell replication (Boison et al., 2019a). Finally, Hillestad and Moghadam (2019) corroborated their findings with a previously published transcriptomic study (Timmerhaus et al., 2011) and showed that some of the genes identified by GWAS were differentially expressed between PMCV-infected and non-infected animals, reinforcing the potential role of adaptive immune response in Atlantic salmon resistance to PMCV.

QTL Pointing Toward Innate or Intrinsic Candidate Mechanisms of Resistance or Susceptibility to Infections

Beside adaptive responses, pathogens are involved in a large number of interactions with their host. The so-called "innate" immune response comprises a wide collection of defense mechanisms based on factors encoded in the genome of the host (i.e., which are not subjected to somatic diversification like B or T cell receptors). Genes directly involved in innate immunity typically play a role in pathogen sensing (such as Toll-like receptors and RIGI-like receptors), in signaling pathways responsible for transduction from sensors to effectors mechanisms (such as Myd88, NFkB), or in the effector phase of the responses that kills or inhibits pathogens (such as lysozymes, cathepsin, chitinase, complement, agglutinins, precipitins, etc.). Such genes can significantly affect host susceptibility to a pathogen. However, host resistance or susceptibility is determined by a much wider repertoire of genes: any host protein involved in an important interaction during the pathogen cycle may influence the efficiency of the infection. Hence, innate or intrinsic (i.e., expressed at steady state) mechanisms potentially underlying QTL of resistance to infections are extremely diverse, which makes understanding of underlying mechanisms very difficult. Of course, specialized defense pathways such as the type I IFN system for viral infections constitute *a priori* relevant targets for the corresponding pathogens. **Table 2** summarizes the relevant QTL studies pointing toward the implication of innate or intrinsic mechanisms in disease resistance.

Many QTL studies typically detect multiple, quite wide QTL containing a number of genes potentially relevant for innate immunity

We present hereafter three examples: resistance to *Neoparamoeba perurans* (causative agent of AGD) in Atlantic salmon (Boison et al., 2019b), to *Flavobacterium psychrophilum* (causative agent of bacterial cold water disease, BCWD) in rainbow trout (Fraslin et al., 2018), and to viral hemorrhagic septicemia virus (VHSV) in turbot (Rodríguez-Ramilo et al., 2014). Details of experimental design and results are presented in **Table 2** and **Supplementary Table 1**. QTL detected in those three studies were quite wide, with one QTL spanning up to 3 Mb in Boison et al. (2019b) and QTL with confidence intervals of up to 30 cM in Rodríguez-Ramilo et al. (2014) or even 100 cM in Fraslin et al. (2018), due to the relatively low number of fish and markers. Therefore, the number of genes within those QTL was important, and in order to detect candidate genes among them, the authors combined QTL detection with transcriptomic or comparative genetics approaches. Eventually, those studies evidenced the role of 20, 14, and 6 genes potentially relevant for innate response to AGD, BCWD, and VHS, respectively. Boison et al. (2019b) performed a GWAS to investigate Atlantic salmon resistance to *N. perurans*, using 53K SNPs genotyped in 1333 smolts challenged by immersion. Resistance was assessed using a qualitative scoring of gill lesions, and a transcriptomic analysis of naïve and infected fish was used to narrow the candidate genes' list. In total, they detected three significant QTL as well as other suggestive QTL. Within the wide confidence intervals

of the three significant QTL, they localized more than 20 genes including genes from the cadherin family, as the *protocadherin Fat 4* gene, and genes involved in the proinflammatory response of cytokine, as the *interleukine-18-binding protein* gene. In Fraslin et al. (2018), we detected QTL associated with rainbow trout resistance to *F. psychrophilum* after both immersion and injection experimental challenges using 310 doubled haploid fish from isogenic lines with contrasted resistance to the disease and genotyped for 2K RADseq SNPs. In this study we detected 15 QTL associated with resistance traits with wide confidence intervals (from 20 up to 100 cM). Interestingly, only three QTL were common between the two infectious routes, which suggests a more or less important role of certain mechanisms according to the mode of infection, or the existence of resistance mechanisms that are only triggered by one infection route. Using a previous study on the transcriptome response to infection in two isogenic lines of trout (Langevin et al., 2012), we identified 14 immune-related genes located within or close to the confidence interval of the QTL. The list included several genes involved in proinflammatory cytokine pathways, several interferon-stimulated genes, a TLR, an antimicrobial peptide, and the *c3* gene from the complement cascade. Interestingly, even if the study was performed on 5-month-old fish that already have a mature immune system with B and T cells, we did not detect genes belonging to the adaptive immunity. In Rodríguez-Ramilo et al. (2014), turbot resistance to VHSV was investigated by a low-density LA approach performed in 270 fish. Eleven suggestive QTL associated with TTD and/or BS were detected. Interestingly, one QTL was located within the LG20 that was syntenic with Omy3, previously detected as a major QTL associated with VHSV resistance in trout (Verrier et al., 2013a) and also pointed to the *tlr7* gene. Further, to identify other candidate genes associated with resistance, they used a 0.5- to 1-Mb window around markers within syntenic blocks between *Gasterosteus aculeatus* and turbot. They detected some genes related to innate immune response such as *sugt1* coding for a protein with an essential role in the activation of Nod-like receptor that recognize invasive bacteria and *zc3hav1* coding for a protein related to induce innate immunity to viral infection. They also detected genes involved in apoptosis and in macrophage differentiation.

A few studies detected a major resistance QTL, explaining a large proportion of the genetic variance associated with positional candidate genes pointing toward the involvement of innate or intrinsic mechanisms

We present hereafter two studies as examples of major QTL detected with no causative mutation evidenced: resistance to VHSV in rainbow trout (Verrier et al., 2013a) and to *P. dicentrarchi* in turbot (Saura et al., 2019). We also develop the case of resistance to infectious pancreatic necrosis virus (IPNV) in Atlantic salmon (Houston et al., 2008, 2010, 2012; Moen et al., 2009; Gheyas et al., 2010), the only example of disease resistance study in fish in which a candidate causative mutation was evidenced (Moen et al., 2015). In Verrier et al. (2013a), a low-density linkage analysis approach performed on two families of doubled haploids trout allowed the detection of seven QTL associated with resistance to VHSV, including a major QTL on chromosome Omy3 also associated with viral replication in excised fin tissue infected *in vitro* (VREFT). This major QTL on Omy3 explained 33–49% of VREFT and 44–65% of TTD phenotypic variances. Several genes involved in innate immunity were located in the region of this major QTL, including Toll-like receptors 7 and 8 (Palti et al., 2010), cytokine receptors, and a trim (Verrier et al., 2013a). Functional studies performed in parallel demonstrated the involvement of the type I IFN pathway, as described below (see section "Isogenic doubled haploid lines, a tool of choice to unravel the complexity of host response to different pathogens and their interactions"). Saura et al. (2019), performed a QTL detection on 1800 turbot infected by cohabitation with *P. dicentrarchi*, a causative agent of scuticociliatosis. Using 18K SNPs, they detected one QTL associated with endurance (defined as the time between the first disease sign and the death of the fish) and four QTL associated with TTD (named resilience in the study) including a major QTL explaining 33% of the genetic variance of the trait. They performed a functional enrichment of candidate genes in this major QTL using the turbot transcriptome as a reference and detected 32 immune-related genes involved in tissue regeneration, response to wounding, inflammatory response, activation of NF-kappa-B pathway, and several genes involved in the activation of the innate immune response (activation of the TLR21 or the MAPK signaling pathways for example).

The case of IPN (infectious pancreatic necrosis) caused by IPNV in Atlantic salmon (*Salmo salar*) is one of the most famous example of identification of a major QTL associated with disease resistance in a fish species, potentially explained by the variations of an epithelial cadherin that prevents the binding and entry of the virus in the host cell. A major QTL was detected on the chromosome Ssa26 (linkage group LG21) in pre-smolt (2 mph) and post-smolt (10–13 mph) Scottish (Houston et al., 2008, 2010, 2012; Gheyas et al., 2010) and Norwegian (Moen et al., 2009, 2015) salmon populations. The markers associated with the resistant genotype are now routinely used in marker assisted selection by salmon breeding companies in Norway (AquaGen) and Scotland (Hendrix genetics, former Landcatch Natural Selection Ltd) for which IPN is no longer a concern. In Houston et al. (2008), this major QTL was first detected on post-smolts and explained up to 79% of the phenotypic variance in four segregating families. As the marker density was low (two to eight markers per linkage group), the confidence interval of this QTL was quite wide (10 cM). Later studies conducted on different Scottish sub-populations and including more fish and more markers (Houston et al., 2010, 2012; Gheyas et al., 2010) detected the same major QTL at a younger stage (2 mph) and were able to narrow the confidence interval down to 2 cM. The same QTL was detected at two stages (pre- and post-smolt) in a Norwegian population (Moen et al., 2009). This major QTL explained 83% of the genetic variance and 29% of the phenotypic variance of BS and had a confidence interval of 4 cM. However, in all these studies, they identified neither the causal mutation nor the underlying genetic resistance mechanisms but hypothesized the role of innate immunity as this major QTL was detected at young stage (2 mph), potentially before the development of an adaptive response (Zapata et al., 2006). By sequencing 22

homozygous resistant fish and 23 homozygous susceptible fish, Moen et al. (2015) identified a putative functional mutation within the cadherin domain of the epithelial cadherin gene (*cdh1-1*), causing a serine-to-proline amino acid shift. They confirmed the interest of this *cdh1-1* gene by targeting it in a functional approach using an immunofluorescence analysis in liver tissue showing IPNV binding to Cdh1-1 to enter the cell and infect the host. Thus, they hypothesized that resistance to IPNV in Atlantic salmon seemed to be explained by the virus failing to enter host cells due to *cdh1-1* mutation. However, while this putative functional mutation was in high linkage disequilibrium with three other SNPs strongly associated with resistance, it did not entirely correspond to the genotypes at the QTL. They hypothesized the existence of a second causative polymorphism involved in a two-locus model controlling resistance to IPNV. The fact that the *cdh1-1* mutation is the causative variant associated with salmon resistance to IPNV is still being debated as two later studies (Reyes-López et al., 2015; Robledo et al., 2016) reported that IPNV could replicate within resistant fish and thus could infect host cells, underlying the existence of a different causative mutation driving resistance to IPNV, still undiscovered.

Lack of Identification of Candidate Genes for Underlying Mechanisms of Resistance Can Be Due to Multiple Factors

In many cases, QTL mapping studies established a long list of positional candidate genes (i.e., genes located within QTL regions) but were not able to narrow down the list of immune-related genes to point toward innate or adaptive immunity, either because QTL were too wide or because they detected too many QTL. In other studies, no candidate genes were identified because no or poor quality genome assembly and genome annotation were available at the time of the publication. **Supplementary Table 1** presents detailed experimental design and results of studies that identified QTL associated with resistance in various fish species but in which no immune mechanisms could be investigated or hypothesized.

In many studies, no functional candidate genes or underlying immune mechanisms could be proposed to explain the fish resistance because the detected QTL were too wide (sometimes corresponding to the whole linkage group). Firstly, the important size of the detected QTL might be due to a small number of markers used to produce a genetic map not dense enough to refine the QTL position, that could correspond to the whole linkage group as in Moen et al. (2004), Massault et al. (2011), Wiens et al. (2013), and Liu et al. (2016a). An example of refinement of QTL position by increasing the number of markers is the case of Asian sea bass resistance to nervous necrosis virus (NNV). Liu et al. (2016a) used 149 microsatellite markers to detect wide QTL (up to 13 cM) associated with both BS and TTD. In a following study performed on the same fish population genotyped with 3K SNPs, Liu et al. (2016b) were able to narrow some confidence interval to 1 cM and thus identified 62 positional candidate genes. With the development of genotyping by sequencing techniques as well as medium- to high-density SNP arrays, the number of markers used will no longer be the limiting factor. Secondly, increasing the number of genotyped individuals remains critical to refine the distances between markers on the genetic maps (dependent on the recombination rate) and to increase the analysis power for QTL detection. The production of an important number of fish is easier than in terrestrial species (as each fish can produce thousands of offspring); however, the recording of fine resistance phenotypes as well as the genotyping of an important number of fish is expensive and thus is becoming the main limiting factor for a powerful and fine mapping of QTL. Four studies performed GWAS with medium- to high-density SNPs arrays (31–690K) but using a relatively small number of fish (between 192 and 720), resulting in the detection of QTL with large confidence intervals: from 3 to 8 Mb for BCWD resistance in rainbow trout (Fraslin et al., 2019); from 27 kb to 3.7 Mb for resistance to enteric septicemia in catfish (ESC) (Zhou et al., 2017; Shi et al., 2018; Tan et al., 2018). No candidate genes' list was established in Fraslin et al. (2019) because too many immune-related genes were positioned within those wide QTL. In the catfish studies, the list of potential genes involved in ESC resistance still contained many (more than 50) immune-related genes; thus, no clear mechanism of disease resistance could be inferred. In all those four studies, mapping of QTL along with gene expression study would have been needed to distinguish functional candidate genes among positional immune genes.

In other studies, no gene was identified because of the lack or poor quality of genome assembly or genome annotation available at the time QTL studies were published (Ozaki et al., 2013; Campbell et al., 2014; Vallejo et al., 2014b; Liu et al., 2015). High-quality genomic data and annotations are critical for identification of candidate causative mutations among significant SNPs detected by QTL approaches. Unfortunately, genome assembly is particularly difficult in many farmed fish species including salmonids and a number of cyprinids, due to recent whole genome duplications. Additionally, both the number of species of interest and the limited size of the fish biologists' communities has hampered the progress of detailed functional annotation. Overall, the quality of the genomic information available is currently improving at a fast pace, thanks to the development of new sequencing technologies producing very long reads, and of new analysis tools. Also, the development of functional annotation initiatives, such as FAASG (Functional Annotation of All Salmonid Genomes), will be extremely important to speed up the progress of QTL approaches (Macqueen et al., 2017).

Finally, in other studies, because of the polygenic architecture of disease resistance traits, the authors did not investigate the positional genes located within the numerous medium or minor effect QTL detected. For example, Palaiokostas et al. (2018a) showed that sea bass (*Dicentrarchus labrax*) resistance to NNV is a trait controlled by at least seven QTL with minor to moderate effects on the trait (from 1.5 to 4% of genetic variance explained by each QTL). In Atlantic salmon, different studies also evidenced a polygenic architecture of the traits with an important number of QTL detected for resistance to *Gyrodactylus salaris* (Gilbey et al., 2006), *N. perurans* (AGD, Robledo et al., 2018), or sea lice (Tsai et al., 2016). Despite being a drawback when investigating the underlying mechanisms of a trait, polygenic architecture

of disease resistance will not prevent the improvement of fish resistance using selective breeding. If the traditional pedigree-based selective breeding can be used to improve disease resistance in the absence of a major QTL (Sonesson et al., 2011; Gjedrem and Rye, 2018), genomic selection will usually perform better as it better takes into account the within-family variation (Ødegård et al., 2011). In those studies on VNN, AGD, and sea lice, the authors estimated that genomic prediction would result in a 10–27% increase in the accuracy of estimated breeding values of fish, more efficiently sorting fish between resistant and susceptible than with only pedigree information.

In summary, many studies in diverse fish species have investigated the genetic determinism of resistance to bacterial, viral, or parasitic diseases using QTL mapping or GWAS analyses. The large majority evidenced a clear genetic control of resistance traits, but to date, even when a major QTL controlling resistance was detected, only one study (Moen et al., 2015) could pinpoint a putative causal mutation that is still being debated. Yet, in a few cases, substantiated information about underlying mechanisms (i.e., role of receptor for IPNV in salmon, role of innate immunity for VHSV in trout) were obtained. In the next section, we will focus on transcriptomics and functional approaches aiming at deciphering molecular mechanisms involved in host–pathogen interactions.

HOST–PATHOGEN INTERACTION MECHANISMS: SELECTION OF FUNCTIONAL CANDIDATE GENES BASED ON TRANSCRIPTOMICS AND FUNCTIONAL ASSAYS

Transcriptomics and the Main Innate Responses to Pathogens: Building a Framework of Fish Responsiveness to Pathogens

Gene expression profiling is one of the possible approaches to get insight into the mechanisms explaining differential resistance to pathogens. Since high-throughput technologies have made possible comprehensive transcriptome description, one would expect that such analyses would, by comparing resistant (R) vs. susceptible (S) individuals, point toward defense mechanisms or pathways required for immunity. When performed on cohorts of significant size, such approaches can identify genetic variations potentially associated with severe or mild forms of the disease (Yamagishi et al., 2014). Furthermore, it is also crucial to understand how the host and the pathogen influence one another, particularly for complex pathogens such as bacteria or protozoans; transcriptome analyses can prove critical to shed light on such interactions.

In fish, transcriptome (and proteome) responses induced by viral or bacterial infections have been extensively studied in multiple tissues and species (reviewed in Verrier et al., 2011; Salinas, 2015; Benard et al., 2016; Martin et al., 2016; Ye et al., 2018), providing a rich resource for comparison. Still, transcriptome studies comparing resistant and susceptible genetic backgrounds have been performed only in a limited number of models. It is also important to note that one should keep in mind that the key factor for resistance may not be directly associated with large differential gene expression (i.e., below the detection threshold). The transcriptional response of rainbow trout to *F. psychrophilum* was compared between isogenic lines of rainbow trout with contrasted susceptibility to the infection (Langevin et al., 2012). The pronephros transcriptome was analyzed using micro-arrays 5 days after infection and showed a typical immune response involving antimicrobial peptides, complement, cytokines, and matrix enzymes. Key genes of the inflammatory response were more induced in susceptible animals where the bacterial load was also much higher. Although a complement C3 gene showed stronger induction in the resistant fish, it did not show an obvious contribution of the complement cascade to the variation of susceptibility to the infection. Analyses also revealed an extensive divergence between the transcriptomes of non-infected resistant and susceptible fish. A RNAseq whole-body transcriptome study was performed on resistant and susceptible fish produced by three generations of selection, comparing naïve and infected fish 1 or 5 days post-infection (Marancik et al., 2015). At both time points, the transcriptional response of resistant and susceptible fish was very different, with a large majority of genes up- or down-regulated only in one line. The expression of long non-coding RNAs (lncRNAs) was also analyzed in the same context (Paneru et al., 2016). Pairwise analyses between relevant conditions identified 556 differentially expressed lncRNAs. The pattern of correlation between differentially regulated lncRNAs and protein-coding genes suggested that some lncRNAs might control the expression of protein-coding genes involved in immunity and susceptibility to the pathogen. Differences between immune cell subsets from resistant and susceptible lines of rainbow trout have also been observed (Zwollo et al., 2017; Moore et al., 2018). For example, neutrophil-like cells expressing high-level Q4E (a trout neutrophil marker), MPO, Pu1, EBF, and IL1 were twice as abundant in the spleen of susceptible line compared to the resistant line. Also, fish from the resistant line have more IgT^+ B cells in their spleen, pronephros, and blood than fish from the susceptible line. Further studies are required to clarify whether correlations between fish resistance to *F. psychrophilum* and the frequency of particular cell subsets mean that these cells play a role in the response to the infection.

In the case of *Piscirickettsia salmonis* in Atlantic salmon, deep sequencing of pronephros transcriptome revealed an exacerbated innate response in susceptible fish, while a few antibacterial genes like C-type lysozyme were more induced in resistant fish and may be involved in the resistance (Dettleff et al., 2015).

Two studies investigated the response of Atlantic salmon to the birnavirus IPNV, in the pronephros (Reyes-López et al., 2015) and in fry (Robledo et al., 2016). Surprisingly, in the first study, in susceptible families, many differentially expressed genes at day 1 returned to basal values at day 5 post-infection, while in resistant families, unlike susceptible families, most induced genes remained highly expressed at day 5. In the second study, while significant viral titer was observed in both resistant and

susceptible fish at the analyzed time points, the inflammatory and cytokine response was stronger in susceptible fish, while resistant fish responded less with a noticeable induction of genes of the M2 macrophage pathway.

Transcriptome responses to the infection were also analyzed in resistant and susceptible fish to a few other viruses: VHSV in rainbow trout (Verrier et al., 2018, see below), herpesvirus CaHV in gibel carp (Gao et al., 2017; Mou et al., 2018), ISAV in Atlantic salmon (Dettleff et al., 2017), and Singapore grouper iridovirus in orange-spotted grouper (Yang et al., 2019).

Similar approaches were also followed with fish showing contrasted levels of resistance to parasites. For example, transcriptomic responses of the anterior kidney and skin of salmon with different levels of susceptibility to salmon louse (*Lepeophtheirus salmonis*) were compared (Sutherland et al., 2014; Braden et al., 2015; Holm et al., 2015). Studies comparing different salmon species (Sutherland et al., 2014; Braden et al., 2015) showed that species-specific pathways can be associated with resistance or susceptibility. Early proinflammatory TH1-type pathway appears to be generally important, but a later regulatory TH2-type response was observed in the skin of resistant coho salmon. Comparing low and high responding families of Atlantic salmon, Holm et al. (2015) observed higher expression of immune genes in resistant fish, suggesting that immunosuppression could explain susceptibility.

Overall, these studies have contributed to build a framework of fish responsiveness to a number of pathogens. It is very important to understand what pathways are differentially expressed in resistant and susceptible fish, to point to putative causal mechanisms and genes. Since several resistance mechanisms can act in combination, and pathogens can be highly variable, such global profiling is invaluable even when a main causal gene has been identified. However, such strategies also have significant pitfalls. It is often very difficult to distinguish variations that explain the extent and severity of infection from those that simply relate to its extent. Thus, the type I IFN response is a key factor to contain virus infection, but it is also a good indicator of the degree of the viral infection itself. The kinetics of transcriptome changes is often a critical parameter to identify resistance factors, but it is complicated to produce large temporal series of comprehensive datasets. Finally, the choice of tissues/organs analyzed is also difficult and can orientate toward genes relating consequences rather than causes of the resistance or susceptibility.

Isogenic Doubled Haploid Lines, a Tool of Choice to Unravel the Complexity of Host Response to Different Pathogens and Their Interactions

Doubled haploid isogenic lines have been established in several fish species (Franěk et al., 2019). They offer unparalleled opportunities for a number of in-depth studies of complex traits. Indeed, for a given genetic background, it is possible to work with the whole organism, tissues, or derived cells at the same time, and with different pathogens or route of infection. These levels of investigation at different scales, and the comparison of different lineages, are very useful in dissecting the underlying mechanisms.

Isogenic Doubled Haploid Lines of Rainbow Trout: Susceptibility and Resistance to VHSV and the Type I IFN Response

A wide range of susceptibility to VHSV was observed among nine homozygous isogenic lines of rainbow trout produced by gynogenesis from a domestic population (INRA Synthetic strain, Quillet et al., 2007). While all these isogenic lines were susceptible to VHSV infection after injection, they exhibited a surprising diversity of susceptibility levels leading to final survival from 100 to 0% after waterborne infection, with some lines having intermediate positions. The virus was seldom found in the spleen, blood, or pronephros of resistant fish, confirming a resistant (not a tolerant) phenotype. To get insight into the mechanism of resistance to the virus, two functional assays were developed. First, *in vitro* cultures of fin explants (Quillet et al., 2001) from resistant and susceptible fish were infected by the virus to assess their capacity to support viral production. The virus production was well correlated to the susceptibility level of the fish from which fins had been sampled. This correlation was further supported by the detection, in two different trout families, of a single major QTL governing both survival of fish after waterborne infection and virus production in *in vitro* cultures of fin explants (Verrier et al., 2013a). In parallel, fibroblast cell lines were developed from several isogenic trout lines (Verrier et al., 2012). Their susceptibility to VHSV infection was remarkably consistent with the susceptibility of the parental fish, indicating that resistance mechanisms associate with innate or intrinsic factors. Interestingly, comparison of the type I IFN response of the cell lines derived from the two fully resistant fish lines revealed two distinct mechanisms: the high resistance of one cell line (named B57) was largely due to an early interferon IFN induction, which was not observed in susceptible cells, while the other cell line (A02) was more refractory to infection—as the corresponding fish—due to different mechanisms.

Deep sequencing of the transcriptome of one resistant (the same B57 line) and one susceptible cell line after stimulation with inactivated VHSV revealed a stronger and earlier response in the resistant background (Verrier et al., 2018). Many typical interferon-stimulated genes were induced, such as *irf1*, *rsad2*, and members of the *ifi44*, *gig*, *parp*, *ifit*, and *cd9* families, as well as many fintrim genes. Also, several major factors of the antiviral jak/stat pathway were much more expressed in non-stimulated B57 cells compared to the susceptible cells, possibly contributing to faster antiviral response.

Altogether, these data illustrate that functional assays provide a better understanding of the pathways involved in resistance. The multiplicity of phenotypes underscores the complexity of the mechanisms, likely based on a similarly complex genetic architecture. Even if the functional assays and transcriptome analyses did not identify the causal gene(s), they have built a framework of the mechanisms and pathways involved in the response of resistant and of susceptible fish to VHSV. This knowledge is fundamental to fully understand the structure of the

resistance as well as the possible subversive mechanisms evolved by the pathogen.

Isogenic Doubled Haploid Lines of Rainbow Trout: Specificity of Susceptibility and Resistance to Multiple Pathogens

The work of VHSV resistance pointed toward innate mechanisms, probably connected to the strength and kinetics of type I IFN response. One may expect that such factors or pathways may be generic and would mediate a significant level of resistance against other viruses. The fish line B57, which was highly resistant to VHSV, showed high to very high resistance levels to all tested viruses (**Table 3**). Consistent results were observed *in vitro* using the corresponding cell line. Interestingly, while our functional assays indicated that the line A2 was resistant due to other mechanisms, these fish were also highly resistant to IHNV. However, testing 10 different isogenic lines, no overall correlation between resistance to those two rhabdoviruses was observed (Verrier et al., 2013b). This seems to be the general picture, as further experiments with a number of isogenic lines indicated that resistance levels to different viruses are not generally correlated. For example, A36 was highly susceptible to VHSV and IPNV, while being rather resistant to IHNV and ISAV; in contrast, B3 was susceptible to IHNV and IPNV and more resistant to VHSV and ISAV (**Table 3**).

This collection of isogenic lines was also challenged with two bacterial pathogens, *F. psychrophilum* (BCWD) and *A. salmonicida* (furunculosis), and contrasted resistance and susceptibility were observed between isogenic lines (**Table 3**, unpublished results). This character was largely influenced by the bacterial species, as well as partly by the route of bacterial infection and by the genetic background of *F. psychrophilum*. Two isogenic lines showed high susceptibility (A36, A22) or strong resistance (AP2, B3) toward *F. psychrophilum* infection,

TABLE 3 | Resistance and susceptibility of a collection of rainbow trout and cell lines against multiple pathogens.

Pathogen, strain, and route of administration	Genetic background							
	A2		B57		A36		B45	
	Fish	Fibr	Fish	Fibr.	Fish	Fibr.	Fish	Fibr.
VHSV, bath	R*		R		S		S	
VHSV, *in vitro*		R		R		S		I
IHNV, bath	R		R		R		S	
IHNV, *in vitro*		R		R				
IPNV, bath			R		S			
ISAV, bath			R		R		S	
[a]*F. psychrophilum* FRDGSA 1882/11, bath	R		R		S		I	
[b]*F. psychrophilum* FRDGSA 1882/11, IM	I		S		S		S	
[b]*F. psychrophilum* OSU THCO2-90, IM	R		S		S		R	
[c]*A. salmonicida* 36-75R, IM	n.d.		S		R		R	

Pathogen, strain, and route of administration	Genetic background								
	A22		A3		AP2	B3	A32	Syn	RTG
	Fish	Fibr.	Fish	Fibr.	Fish	Fish	Fish	Fish	Fibr.
VHSV, bath	S (0%)		S		S	R	R	S	
VHSV		S		S					S
IHNV, bath	S (17%)		R		S	S	S		
IHNV		S		S					
IPNV, bath	I (55%)		S		S	S	I	I	
ISAV	S (0%)		I			R			
[a]*F. psychrophilum* FRDGSA 1882/11, bath	S		R		R	R	R	R	
[b] *F. psychrophilum* FRDGSA 1882/11, IM	S		I		R	R	I	n.d.	
[b]*F. psychrophilum* OSU THCO2-90, IM	S		I		R	R	n.d.	R	
[c]*A. salmonicida* 36-75R, IM	R		I		n.d.	S	S	n.d.	

**S, susceptible; R, resistant.* [a]*Resistance/susceptibility phenotype estimation using the percentage of survival 28 days after immersion challenge. S: susceptible (0–30%), R: resistant (>70%), I: intermediate (30–70%).* [b]*Isogenic lines were classified into three groups (S, susceptible; R, resistant; I, intermediate) based on the time to death (TD) values or % of survival after intramuscular injection challenge. TD of group S < TD of group I < TD of group R.* [c]*Resistance/susceptibility phenotype estimation using the percentage of survival 13 days following the IM challenge. S: susceptible (0–30%), R: resistant (>70%), I: intermediate (30–70%). n.d., not done; Fibr, fibroblasts; VHS, viral hemorrhagic septicemia; ISAV, infectious salmon anemia virus; IHNV, infectious hematopoietic necrosis virus; IPNV, infectious pancreatic necrosis virus. Data are compiled from Quillet et al. (2007) [VHSV], Biacchesi et al. (2007) [ISAV], and Verrier et al. (2013b) [IHNV], or unpublished (Flavobacterium psychrophilum, IPNV).*

whatever the route or strain used for the challenge. However, for several isogenic lines, this trait was influenced by the route of infection, which may be indicative of different mechanisms of resistance in the different lines. For instance, B57 was highly susceptible by injection but gained in resistance by bath, while A2 and A3 showed intermediate resistance by intramuscular injection compared to bath. The existence of favorable genetic factors controlling defense mechanisms triggered after waterborne infection in line B57 was further supported by the results of a QTL analysis (Fraslin et al., 2018). The resistance/susceptibility status was also compared between fish challenged with two *F. psychrophilum* strains from different genetic lineages as defined by Multi Locus Sequence Analysis, namely, FRDGSA 1882/11 (clonal complex ST90) isolated from rainbow trout and OSU THCO2-90 (clonal complex ST9) isolated from Coho salmon (Duchaud et al., 2018), leading to opposite ranking for one isogenic line (B45) only. Some interactions between the host genetics and the bacterial genotype may occur and should be considered when selecting for BCWD resistance. This study also pointed to a negative relationship of the resistance character between *F. psychrophilum* and *A. salmonicida*, highlighting the importance of screening for resistance to various pathogens in breeding schemes.

Table 3 also illustrates that there is no straightforward correlation (or reverse correlation) between the resistance to viral and bacterial infections. These data illustrate the strong interest of comparing the resistance levels to multiple pathogens across stable genetic backgrounds. Within a given species, it confirms the complexity and diversity of genetic architectures of resistance to different viruses or bacteria. The strong impact of the route of infection provides further insight into the type of mechanisms involved and might deserve further investigation. Combined with QTL data, genome sequencing of isogenic lines will help determine the best candidate genes/sequence polymorphisms for mechanisms responsible for resistance and susceptibility. These models may also constitute an interesting context to test the interactions between susceptibility factors to different pathogens.

Importance of KO Approaches *in vivo* and *in vitro*

In order to validate the role of a locus on a particular phenotype of resistance or sensitivity to a pathogen, loss-of-function methodologies are very valuable (Vidalin et al., 2009). For many years, the development of knock-out approaches was limited to animal models relevant to the medical research community, mainly mice (Hall et al., 2009). Today, with the advent of CRISPR/Cas9-based genome editing combined with an increasing number of well assembled genomes, it can be extended to farmed fish species (Yáñez et al., 2015). Proofs of concepts of efficient gene KO by CRIPSR/Cas9 have been established *in vitro* for chinook salmon cells (Dehler et al., 2016) or *in vivo* in Atlantic salmon (Edvardsen et al., 2014). However, the latter is faced with the difficulty to produce non-mosaic animals (Mehravar et al., 2019).

Cell lines can be valuable to verify the relevance of one or several candidate genes. In the case of certain traits such as innate immunity to viral pathogens (Langevin et al., 2019), phenotypes obtained in a KO cell line can be extrapolated to the whole animal with some degree of accuracy. This is the case of the type I interferon (IFN1) system, an important arm of the immunity to viruses. Dehler et al. (2019) generated a chinook salmon cell line (named GS2) with the gene encoding for Signal transducer and activator of transcription (*stat2*), an important molecule involved in the signaling of IFN1, disrupted. The transcriptomic characterization of the GS2 cells demonstrated that they had lost their ability to respond to IFN1.

In addition, the production of additional KO cell lines in the same species suggested that the CRISPR/Cas9 approach allows the generation of homozygous null mutations at one or several paralogous loci. This is particularly well suited to the European commercial fish species, salmonids, and carp, with a duplicated genome.

Genome-wide KO *in vitro* screen is also possible in well-established mammalian–lentivirus systems (Shalem et al., 2014, 2015) and may be applicable in fish cells in the near future (Gratacap et al., 2020). Genome editing can also target non-coding sequences and contribute to validation of mutation in, for instance, regulatory sequences (Banerjee and Sherwood, 2017; Kang et al., 2019).

Thus, *in vitro* knock-out could help validate phenotypes in a certain number of biological functions that can be accurately replicated in cell lines (such as antiviral innate mechanisms). In a second stage, *in vitro*-validated single guide RNA can be used to generate KO animals within the same species. Nuclear transfer from edited somatic fish cells to oocytes is a strategy that merits some future attention (Ju et al., 2003) in order to circumvent the current problems of mosaicism.

Perspectives

QTL analyses draw a first picture of the genetic architecture of disease resistance. With the continuous development of new technologies allowing the genotyping of tens to hundreds of thousands of SNPs in larger datasets, the precision of the QTL position has dramatically increased in recent studies (confidence intervals reduced to sometimes tens of kb), with promising opportunities for the identification of underlying acting factors.

With a few exceptions, disease resistance in fish appears to be mainly controlled by several genomic regions having a minor to medium impact on the trait, underlying a complex determinism. In many studies, resistance was controlled by tens of minor or medium effect QTL, revealing the polygenic architecture of the trait. In others, a few QTL—sometimes major ones—that had a bigger impact were detected alongside with smaller-sized QTL, revealing an oligogenic architecture of resistance. Only a few examples of simple determinism were reported with one or two major QTL associated with Atlantic salmon resistance to IPNV (Houston et al., 2008, 2010, 2012; Moen et al., 2009, 2015; Gheyas et al., 2010) and PMCV (Boison et al., 2019a; Hillestad and Moghadam, 2019), red sea bream resistance to RSIV (Sawayama et al., 2017), and rainbow trout resistance to whirling disease (Baerwald et al., 2011).

Besides being often polygenic, disease resistance is also a complex trait for which detected QTL might often be in

interactions with each other. Epistasis has been shown to play an important role in complex traits but is often neglected as it is complex to measure and take into account (Flint and Mott, 2001; Carlborg and Haley, 2004). Two studies have investigated epistatic interactions between QTL associated with disease resistance in fish. Fraslin et al. (2018) detected 15 QTL associated with resistance to *F. psychrophilum* in doubled haploid rainbow trout. In this study, by fixing the alleles at the five main QTL detected previously as co-factors in a new model, they were able to reveal five new QTL, underlying the polygenic and complex nature of rainbow trout resistance to *F. psychrophilum*. In Tadmor-Levi et al. (2019), they tested pairwise interactions between their QTL by including QTL effect as a fixed effect in a new model and detected multiple epistatic QTL associated with resistance to cyprinid herpes virus-3 (CyHV-3) in five common carp families. Interestingly, once again, some QTL that were not significant in families in a simple model became significant when epistasis was accounted for. In total, three types of interactions were evidenced, "enhancing-like or synergistic" interaction when the allele at one QTL enhances the effect of the allele at the second QTL, "compensation-like" interaction when each QTL alternatively contributes to the resistance and "nullifying" interaction when both QTL have opposite effects that cancel each other's main effect. In Fraslin et al. (2018), they also evidenced "counter-acting interactions," which might correspond to negative feedback loops, when the favorable or deleterious effects of the genotype at one QTL was reversed according to the genotype at the interacting QTL. While complicated to evidence, epistatic interactions between disease resistance might be common and underline the complexity of resistance mechanisms pathways.

When the QTL detected were not too wide and when an annotated genome assembly was available, those QTL studies also identified putative candidate genes located within the QTL, underlying complex immune pathways that pointed toward adaptive or innate immunity (see **Tables 1**, **2**, respectively). However, even when a narrow list of positional candidate genes was established, only one study (Moen et al., 2015) identified the putative causative mutation.

GWAS approaches to map QTL and detect positional candidate genes can be completed with different approaches to refine both QTL location and candidate genes' list by integrating genetic diversity, comparative genomics, or functional studies. GWAS, genetic diversity (F_{st}), and nucleotide diversity filtration approaches can be combined to select the most promising positional candidate genes, as in Zhou et al. (2019) for the resistance to *Vibrio harveyi* in the Chinese tongue sole (*Cynoglossus semilaevis*): in this way, nine significant SNPs located in six candidate regions (23 putative candidate genes) were identified from an initial list of 79 regions. In a different perspective, Yáñez et al. (2019) combined GWAS approaches and comparative genomics to detect QTL associated with resistance to *P. salmonis*, an intracellular bacterium responsible for salmon rickettsial syndrome (SRS), in three salmonids species (Atlantic salmon, coho salmon, and rainbow trout). Using between 9K and 42K SNPs, they described an oligogenic architecture of the trait in rainbow trout and coho salmon with the top 200 SNPs explaining between 70 and 90% of genetic variance of BS or TTD in the two species. However, in Atlantic salmon, resistance to *P. salmonis* was a polygenic trait with no SNPs explaining over 5% of the genetic variance and the top 200 SNPs explaining 30% of both BS and TTD. To narrow the long list of positional candidate genes, they classified SNPs and corresponding genes into four categories according to (1) function of protein domain or (2) orthology among species, (3) proximity to the top SNP explaining the highest percentage of genetic variance, and finally (4) presence in more than one genomic region explaining more than 1% of the genetic variance within species. The list of best positional candidates involved in salmonids resistance to *P. salmonis* could be reduced to 21 genes belonging to at least two of those four categories. This study shows that comparative genomics can be an interesting approach to detect highly conserved resistance mechanisms to one pathogen among various fish species (here three salmonids); however, by doing so, species-specific mechanisms are not investigated. In three previous publications performed by the same group in Chile on rainbow trout (Barria et al., 2019), Atlantic salmon (Correa et al., 2015), and coho salmon (Barría et al., 2018), species-specific QTL have been identified, pointing to involvement of both innate and adaptive immunity in Atlantic salmon resistance to *P. salmonis*, and of innate immunity in rainbow trout and coho salmon (see **Supplementary Table 1**).

Recent studies combined QTL mapping with whole-genome sequencing (WGS) of pooled or single DNA samples, and expression analysis (RNAseq or qPCR) to search for candidates genes underlying disease susceptibility/resistance. Palaiokostas et al. (2018b) combined QTL mapping and WGS to refine the position of QTL of carp resistance to koi herpesvirus, and detected a promising positional and functional candidate gene. After performing WGS on two pooled DNA libraries of 30 susceptible and 30 resistant fish, they annotated new SNPs in the main QTL region, and identified a potential causal mutation responsible for a premature stop codon in the E3 ubiquitin ligase *trim25*, a major Interferon Stimulated Gene involved in RIG I and IFN signaling. The mutation responsible for the premature stop codon was rare, but more common in the susceptible fish than in the resistant fish pools. *trim25* gene therefore appeared as a promising functional candidate to explain resistance to koi herpesvirus in carp. In another study on the genetic architecture of Atlantic salmon resistance to sea lice (*Caligus rogercressey*), Robledo et al. (2019) succeeded to identify a few top positional and functional candidate genes by combining GWAS and transcriptomics study in the fish skin. Those genes included *tob1*, a transcription factor that negatively regulates cell proliferation, and the serine/threonine-protein kinase 17B (*stk17b*) showing also the highest fold change between healthy and diseased skin.

These examples illustrate that future projects to understand the genetic bases of fish resistance to diseases will combine a number of approaches taking advantage of novel genomic and sequencing technologies.

eQTL (expression quantitative trait loci) and reQTL (RNA-eQTL) will certainly constitute another important approach to identify relevant causal genetic variants in fish as it is the case in mammals (Gat-Viks et al., 2013). Importantly, fast functional tests such as *in vitro* models and the powerful loss-of-function screens based on CRISPR/Cas systems will likely complement the arsenal used by fish immuno-geneticists to foster the development of integrated pipelines for fast identification of key genes regulating resistance to infections. When available, isogenic fish lines with contrasted phenotypes will provide interesting models to explore the mechanistic diversity of resistance, as illustrated by studies on VHSV in rainbow trout (Verrier et al., 2013a, 2018).

A largely unresolved question is how genetic resistance and susceptibility to pathogens may have an effect on, or may be correlated to, the efficiency of vaccination against these pathogens or others. There are multiple possible links between vaccine development and natural resistance to pathogens. First, resistance based on adaptive immune responses, in particular those determined by MHC haplotypes expressed by an individual, often depend on the structure of selected immune repertoires of lymphocytes, which favor efficient and protective responses. In such cases, it is likely that genetic resistance to a pathogen, and good responses to vaccination against the same pathogen, would be correlated. Robust vaccines would therefore have to induce protective responses even in susceptible animals naturally expressing repertoires devoid of the most relevant specificities for this response. Second, when the resistance to a pathogen is mediated by components of innate immunity-for example by pathogen sensing—it would likely impact the strength of immune responses induced by vaccines or even, more generally, by many adjuvants. Finally, the resistant or susceptible status of experimental fish is an essential parameter to take into account when testing the protective effects of a vaccine. Robust tests of vaccine efficiency require control groups with high levels of mortality upon challenge, but it might be also interesting to test the response induced in naturally resistant individuals. While it remains difficult to integrate such aspects in the vaccine development for the time being, a better understanding of the mechanisms of genetic resistance to pathogens should contribute to a better design of vaccines and adjuvants in the future.

CONCLUSION

In this review, we focused on resistance and susceptibility to infections in fish. Upcoming progress in the field will certainly lead to a better understanding of the importance of endurance and tolerance/resilience (Yáñez et al., 2010; Levraud et al., 2014; Fraslin et al., 2018; Fraslin, 2018; Saura et al., 2019). Finally, a dynamic vision of robustness should emerge, in the context of a given aquatic farming system and its pathobiome.

AUTHOR CONTRIBUTIONS

CF, EQ, DL, and PB planned the review topic, reviewed the manuscript, and wrote and edited the manuscript. TR, ND, J-FB, and BC wrote specific sections of the manuscript, synthesized the information for specific tables, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE) and by the European Commission under the Work Programme of the 7th Framework Programme for Research and Technological Development of the European Union [Grant Agreements 311993 (TARGETFISH), 613611 (FISHBOOST), and 262336 (AQUAEXCEL)].

ACKNOWLEDGMENTS

We would like to thank Dr. Christian Michel for his advices and for sharing his data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00677/full#supplementary-material

REFERENCES

Baerwald, M. R., Petersen, J. L., Hedrick, R. P., Schisler, G. J., and May, B. (2011). A major effect quantitative trait locus for whirling disease resistance identified in rainbow trout (*Oncorhynchus mykiss*). *Heredity* 106, 920–926. doi: 10.1038/hdy.2010.137

Banerjee, B., and Sherwood, R. I. (2017). A CRISPR view of gene regulation. *Curr. Opin. Syst. Biol.* 1, 1–8. doi: 10.1016/j.coisb.2016.12.016

Barría, A., Christensen, K. A., Yoshida, G. M., Correa, K., Jedlicki, A., Lhorente, J. P., et al. (2018). Genomic predictions and genome-wide association study of resistance against *Piscirickettsia salmonis* in Coho Salmon (*Oncorhynchus kisutch*) using ddRAD sequencing. *G3* 8, 1183–1194. doi: 10.1534/g3.118.200053

Barria, A., Marín-Nahuelpi, R., Cáceres, P., López, M. E., Bassini, L. N., Lhorente, J. P., et al. (2019). Single-step genome-wide association study for resistance to *Piscirickettsia salmonis* in Rainbow Trout (*Oncorhynchus mykiss*). *G3* 9, 3833–3841. doi: 10.1534/g3.119.400204

Benard, E. L., Rougeot, J., Racz, P. I., Spaink, H. P., and Meijer, A. H. (2016). Transcriptomic approaches in the zebrafish model for tuberculosis-insights into host- and pathogen-specific determinants of the innate immune response. *Adv. Genet.* 95, 217–251. doi: 10.1016/bs.adgen.2016.04.004

Best, S., Tissier, P. L., Towers, G., and Stoye, J. P. (1996). Positional cloning of the mouse retrovirus restriction gene Fvl. *Nature* 382, 826–829. doi: 10.1038/382826a0

Biacchesi, S., Le Berre, M., Le Guillou, S., Benmansour, A., Brémont, M., Quillet, E., et al. (2007). Fish genotype significantly influences susceptibility of juvenile rainbow trout, *Oncorhynchus mykiss* (Walbaum), to waterborne infection with

infectious salmon anaemia virus. *J. Fish Dis.* 30, 631–636. doi: 10.1111/j.1365-2761.2007.00854.x

Boison, S. A., Ding, J., Leder, E., Gjerde, B., Bergtun, P. H., Norris, A., et al. (2019a). QTLs associated with resistance to cardiomyopathy syndrome in Atlantic salmon. *J. Hered.* 110, 727–737. doi: 10.1093/jhered/esz042

Boison, S. A., Gjerde, B., Hillestad, B., Makvandi-Nejad, S., and Moghadam, H. K. (2019b). Genomic and transcriptomic analysis of amoebic gill disease resistance in Atlantic salmon (*Salmo salar* L.). *Front. Genet.* 10:68. doi: 10.3389/fgene.2019.00068

Braden, L. M., Koop, B. F., and Jones, S. R. M. (2015). Signatures of resistance to *Lepeophtheirus salmonis* include a TH2-type response at the louse-salmon interface. *Dev. Comp. Immunol.* 48, 178–191. doi: 10.1016/j.dci.2014.09.015

Campbell, N. R., LaPatra, S. E., Overturf, K., Towner, R., and Narum, S. R. (2014). Association mapping of disease resistance traits in rainbow trout using restriction site associated DNA sequencing. *G3* 4, 2473–2481. doi: 10.1534/g3.114.014621

Carlborg, Ö, and Haley, C. S. (2004). Epistasis: too often neglected in complex trait studies? *Nat. Rev. Genet.* 5, 618–625. doi: 10.1038/nrg1407

Collins, C., Lorenzen, N., and Collet, B. (2019). DNA vaccination for finfish aquaculture. *Fish Shellfish Immunol.* 85, 106–125. doi: 10.1016/j.fsi.2018.07.012

Conti, F., Abnave, P., and Ghigo, E. (2014). Unconventional animal models: a booster for new advances in host-pathogen interactions. *Front. Cell. Infect. Microbiol.* 4:142. doi: 10.3389/fcimb.2014.00142

Correa, K., Lhorente, J. P., López, M. E., Bassini, L., Naswa, S., Deeb, N., et al. (2015). Genome-wide association analysis reveals loci associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar* L.) chromosomes. *BMC Genomics* 16:854. doi: 10.1186/s12864-015-2038-7

Cox, D. R. (1972). Regression models and life-tables. *J. R. Stat. Soc. Ser. B Methodol.* 34, 187–220.

Dehler, C. E., Boudinot, P., Martin, S. A. M., and Collet, B. (2016). Development of an efficient genome editing method by CRISPR/Cas9 in a fish cell line. *Mar. Biotechnol.* 18, 449–452. doi: 10.1007/s10126-016-9708-6

Dehler, C. E., Lester, K., Pelle, G. D., Jouneau, L., Houel, A., Collins, C., et al. (2019). Viral resistance and IFN signaling in STAT2 knockout fish cells. *J. Immunol.* 203, 465–475. doi: 10.4049/jimmunol.1801376

Dettleff, P., Bravo, C., Patel, A., and Martinez, V. (2015). Patterns of *Piscirickettsia salmonis* load in susceptible and resistant families of *Salmo salar*. *Fish Shellfish Immunol.* 45, 67–71. doi: 10.1016/j.fsi.2015.03.039

Dettleff, P., Moen, T., Santi, N., and Martinez, V. (2017). Transcriptomic analysis of spleen infected with infectious salmon anemia virus reveals distinct pattern of viral replication on resistant and susceptible Atlantic salmon (*Salmo salar*). *Fish Shellfish Immunol.* 61, 187–193. doi: 10.1016/j.fsi.2017.01.005

Duchaud, E., Rochat, T., Habib, C., Barbier, P., Loux, V., Guérin, C., et al. (2018). Genomic diversity and evolution of the fish pathogen *Flavobacterium psychrophilum*. *Front. Microbiol.* 9:138. doi: 10.3389/fmicb.2018.00138

Edvardsen, R. B., Leininger, S., Kleppe, L., Skaftnesmo, K. O., and Wargelius, A. (2014). Targeted mutagenesis in Atlantic salmon (*Salmo salar* L.) using the CRISPR/Cas9 system induces complete knockout individuals in the F0 generation. *PLoS One* 9:e108622. doi: 10.1371/journal.pone.0108622

Ewald, P. W. (1983). Host-parasite relations, vectors, and the evolution of disease severity. *Annu. Rev. Ecol. Syst.* 14, 465–485. doi: 10.1146/annurev.es.14.110183.002341

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4 Edn. Harlow: Pearson.

Flint, J., and Mott, R. (2001). Finding the molecular basis of quantitative traits: successes and pitfalls. *Nat. Rev. Genet.* 2, 437–445. doi: 10.1038/35076585

Franěk, R., Baloch, A. R., Kašpar, V., Saito, T., Fujimoto, T., Arai, K., et al. (2019). Isogenic lines in fish – a critical review. *Rev. Aquac.* doi: 10.1111/raq.12389

Fraslin, C. (2018). *Bases Génétiques de la Réponse à L'infection par Flavobacterium psychrophilum chez la Truite arc-en-ciel: Approche Expérimentale et Perspectives en Sélection*. Ph.D. thesis, Université Paris-Saclay, Saint-Aubin.

Fraslin, C., Brard-Fudulea, S., D'Ambrosio, J., Bestin, A., Charles, M., Haffray, P., et al. (2019). Rainbow trout resistance to bacterial cold water disease: two new quantitative trait loci identified after a natural disease outbreak on a French farm. *Anim. Genet.* 50, 293–297. doi: 10.1111/age.12777

Fraslin, C., Dechamp, N., Bernard, M., Krieg, F., Hervet, C., Guyomard, R., et al. (2018). Quantitative trait loci for resistance to *Flavobacterium psychrophilum* in rainbow trout: effect of the mode of infection and evidence of epistatic interactions. *Genet. Sel. Evol.* 50:60. doi: 10.1186/s12711-018-0431-9

Gao, F.-X., Wang, Y., Zhang, Q.-Y., Mou, C.-Y., Li, Z., Deng, Y.-S., et al. (2017). Distinct herpesvirus resistances and immune responses of three gynogenetic clones of gibel carp revealed by comprehensive transcriptomes. *BMC Genomics* 18:561. doi: 10.1186/s12864-017-3945-6

Gat-Viks, I., Chevrier, N., Wilentzik, R., Eisenhaure, T., Raychowdhury, R., Steuerman, Y., et al. (2013). Deciphering molecular circuits from genetic variation underlying transcriptional responsiveness to stimuli. *Nat. Biotechnol.* 31, 342–349. doi: 10.1038/nbt.2519

Geng, X., Sha, J., Liu, S., Bao, L., Zhang, J., Wang, R., et al. (2015). A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance. *BMC Genomics* 16:196. doi: 10.1186/s12864-015-1409-4

Gharbi, K., Glover, K. A., Stone, L. C., MacDonald, E. S., Matthews, L., Grimholt, U., et al. (2009). Genetic dissection of MHC-associated susceptibility to *Lepeophtheirus salmonis* in Atlantic salmon. *BMC Genet.* 10:20. doi: 10.1186/1471-2156-10-20

Gheyas, A. A., Houston, R. D., Mota-Velasco, J. C., Guy, D. R., Tinch, A. E., Haley, C. S., et al. (2010). Segregation of infectious pancreatic necrosis resistance QTL in the early life cycle of Atlantic Salmon (*Salmo salar*). *Anim. Genet.* 41, 531–536. doi: 10.1111/j.1365-2052.2010.02032.x

Gilbey, J., Verspoor, E., Mo, T. A., Sterud, E., Olstad, K., Hytterød, S., et al. (2006). Identification of genetic markers associated with Gyrodactylus salaris resistance in Atlantic salmon *Salmo salar*. *Dis. Aquat. Organ.* 71, 119–129. doi: 10.3354/dao071119

Gjedrem, T., and Rye, M. (2018). Selection response in fish and shellfish: a review. *Rev. Aquac.* 10, 168–179. doi: 10.1111/raq.12154

Gonen, S., Baranski, M., Thorland, I., Norris, A., Grove, H., Arnesen, P., et al. (2015). Mapping and validation of a major QTL affecting resistance to pancreas disease (salmonid alphavirus) in Atlantic salmon (*Salmo salar*). *Heredity* 115, 405–414. doi: 10.1038/hdy.2015.37

Gratacap, R. L., Regan, T., Dehler, C. E., Martin, S. A. M., Boudinot, P., Collet, B., et al. (2020). Efficient CRISPR/Cas9 genome editing in a salmonid fish cell line using a lentivirus delivery system. *BMC Biotechnol.* 20:35. doi: 10.1186/s12896-020-00626-x

Gratacap, R. L., Wargelius, A., Edvardsen, R. B., and Houston, R. D. (2019). Potential of genome editing to improve aquaculture breeding and production. *Trends Genet.* 35, 672–684. doi: 10.1016/j.tig.2019.06.006

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., et al. (2003). MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics* 55, 210–219. doi: 10.1007/s00251-003-0567-8

Hall, B., Limaye, A., and Kulkarni, A. B. (2009). Overview: generation of gene knockout mice. *Curr. Protoc. Cell Biol.* 44, 19.12.1–19.12.17. doi: 10.1002/0471143030.cb1912s44

Hedrick, R. P., McDowell, T. S., Marty, G. D., Fosgate, G. T., Mukkatira, K., Myklebust, K., et al. (2003). Susceptibility of two strains of rainbow trout (one with suspected resistance to whirling disease) to *Myxobolus cerebralis* infection. *Dis. Aquat. Organ.* 55, 37–44. doi: 10.3354/dao055037

Hillestad, B., and Moghadam, H. K. (2019). Genome-wide association study of Piscine Myocarditis Virus (PMCV) resistance in Atlantic salmon (*Salmo salar*). *J. Hered.* 110, 720–726. doi: 10.1093/jhered/esz040

Holm, H., Santi, N., Kjøglum, S., Perisic, N., Skugor, S., and Evensen, Ø (2015). Difference in skin immune responses to infection with salmon louse (*Lepeophtheirus salmonis*) in Atlantic salmon (Salmo salar L.) of families selected for resistance and susceptibility. *Fish Shellfish Immunol.* 42, 384–394. doi: 10.1016/j.fsi.2014.10.038

Houston, R. D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., Jenkins, T. L., et al. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nat. Rev. Genet.* 21, 389–409. doi: 10.1038/s41576-020-0227-y

Houston, R. D., Davey, J. W., Bishop, S. C., Lowe, N. R., Mota-Velasco, J. C., Hamilton, A., et al. (2012). Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* 13:244. doi: 10.1186/1471-2164-13-244

Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Mota-Velasco, J. C., Gheyas, A. A., et al. (2010). The susceptibility of Atlantic salmon fry to freshwater

infectious pancreatic necrosis is largely explained by a major QTL. *Heredity* 105, 318–327. doi: 10.1038/hdy.2009.171

Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J. B., et al. (2008). Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics* 178, 1109–1115. doi: 10.1534/genetics.107.082974

Ju, B., Pristyazhnyuk, I., Ladygina, T., Kinoshita, M., Ozato, K., and Wakamatsu, Y. (2003). Development and gene expression of nuclear transplants generated by transplantation of cultured cell nuclei into non-enucleated eggs in the medaka *Oryzias latipes*. *Dev. Growth Differ.* 45, 167–174. doi: 10.1034/j.1600-0854.2004.00687.x

Kang, J. G., Park, J. S., Ko, J.-H., and Kim, Y.-S. (2019). Regulation of gene expression by altered promoter methylation using a CRISPR/Cas9-mediated epigenetic editing system. *Sci. Rep.* 9:11960. doi: 10.1038/s41598-019-48130-3

Komen, H., and Thorgaard, G. H. (2007). Androgenesis, gynogenesis and the production of clones in fishes: a review. *Aquaculture* 269, 150–173. doi: 10.1016/j.aquaculture.2007.05.009

Kong, S., Ke, Q., Chen, L., Zhou, Z., Pu, F., Zhao, J., et al. (2019). Constructing a high-density genetic linkage map for large yellow croaker (*Larimichthys crocea*) and mapping resistance trait against ciliate parasite *Cryptocaryon irritans*. *Mar. Biotechnol.* 21, 262–275. doi: 10.1007/s10126-019-09878-x

Lafferty, K. D., Harvell, C. D., Conrad, J. M., Friedman, C. S., Kent, M. L., Kuris, A. M., et al. (2015). Infectious diseases affect marine fisheries and aquaculture economics. *Annu. Rev. Mar. Sci.* 7, 471–496. doi: 10.1146/annurev-marine-010814-015646

Lander, E. S., and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185–199.

Langevin, C., Blanco, M., Martin, S. A. M., Jouneau, L., Bernardet, J.-F., Houel, A., et al. (2012). Transcriptional responses of resistant and susceptible fish clones to the bacterial pathogen *Flavobacterium psychrophilum*. *PLoS One* 7:e39126. doi: 10.1371/journal.pone.0039126

Langevin, C., Boudinot, P., and Collet, B. (2019). IFN signaling in inflammation and viral infections: new insights from fish models. *Viruses* 11:302. doi: 10.3390/v11030302

Levraud, J.-P., Palha, N., Langevin, C., and Boudinot, P. (2014). Through the looking glass: witnessing host–virus interplay in zebrafish. *Trends Microbiol.* 22, 490–497. doi: 10.1016/j.tim.2014.04.014

Liu, P., Wang, L., Wan, Z. Y., Ye, B. Q., Huang, S., Wong, S.-M., et al. (2016a). Mapping QTL for resistance against viral nervous necrosis disease in Asian seabass. *Mar. Biotechnol.* 18, 107–116. doi: 10.1007/s10126-015-9672-6

Liu, P., Wang, L., Wong, S.-M., and Yue, G. H. (2016b). Fine mapping QTL for resistance to VNN disease using a high-density linkage map in Asian seabass. *Sci. Rep.* 6:32122. doi: 10.1038/srep32122

Liu, S., Vallejo, R. L., Palti, Y., Gao, G., Marancik, D. P., Hernandez, A. G., et al. (2015). Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Front. Genet.* 6:298. doi: 10.3389/fgene.2015.00298

Ma, J., Bruce, T. J., Jones, E. M., and Cain, K. D. (2019). A review of fish vaccine development strategies: conventional methods and modern biotechnological approaches. *Microorganisms* 7:569. doi: 10.3390/microorganisms7110569

Macqueen, D. J., Primmer, C. R., Houston, R. D., Nowak, B. F., Bernatchez, L., Bergseth, S., et al. (2017). Functional annotation of all salmonid genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genomics* 18:484. doi: 10.1186/s12864-017-3862-8

Marancik, D., Gao, G., Paneru, B., Ma, H., Hernandez, A. G., Salem, M., et al. (2015). Whole-body transcriptome of selectively bred, resistant-, control-, and susceptible-line rainbow trout following experimental challenge with *Flavobacterium psychrophilum*. *Front. Genet.* 5:453. doi: 10.3389/fgene.2014.00453

Martin, S. A. M., Dehler, C. E., and Król, E. (2016). Transcriptomic responses in the fish intestine. *Dev. Comp. Immunol.* 64, 103–117. doi: 10.1016/j.dci.2016.03.014

Martinez, V. A., Hill, W. G., and Knott, S. A. (2002). On the use of double haploids for detecting QTL in outbred populations. *Heredity* 88, 423–431.

Massault, C., Franch, R., Haley, C., de Koning, D. J., Bovenhuis, H., Pellizzari, C., et al. (2011). Quantitative trait loci for resistance to fish pasteurellosis in gilthead sea bream (*Sparus aurata*). *Anim. Genet.* 42, 191–203. doi: 10.1111/j.1365-2052.2010.02110.x

Mehravar, M., Shirazi, A., Nazari, M., and Banan, M. (2019). Mosaicism in CRISPR/Cas9-mediated genome editing. *Dev. Biol.* 445, 156–162. doi: 10.1016/j.ydbio.2018.10.008

Moen, T., Baranski, M., Sonesson, A. K., and Kjøglum, S. (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics* 10:368. doi: 10.1186/1471-2164-10-368

Moen, T., Fjalestad, K. T., Munck, H., and Gomez-Raya, L. (2004). A multistage testing strategy for detection of quantitative trait Loci affecting disease resistance in Atlantic salmon. *Genetics* 167, 851–858. doi: 10.1534/genetics.103.013227

Moen, T., Sonesson, A. K., Hayes, B., Lien, S., Munck, H., and Meuwissen, T. H. E. (2007). Mapping of a quantitative trait locus for resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*): comparing survival analysis with analysis on affected/resistant data. *BMC Genet.* 8:53. doi: 10.1186/1471-2156-8-53

Moen, T., Torgersen, J., Santi, N., Davidson, W. S., Baranski, M., Ødegård, J., et al. (2015). Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. *Genetics* 200, 1313–1326. doi: 10.1534/genetics.115.175406

Moore, C., Hennessey, E., Smith, M., Epp, L., and Zwollo, P. (2018). Innate immune cell signatures in a BCWD-Resistant line of rainbow trout before and after in vivo challenge with *Flavobacterium psychrophilum*. *Dev. Comp. Immunol.* 90, 47–54. doi: 10.1016/j.dci.2018.08.018

Mou, C.-Y., Wang, Y., Zhang, Q.-Y., Gao, F.-X., Li, Z., Tong, J.-F., et al. (2018). Differential interferon system gene expression profiles in susceptible and resistant gynogenetic clones of gibel carp challenged with herpesvirus CaHV. *Dev. Comp. Immunol.* 86, 52–64. doi: 10.1016/j.dci.2018.04.024

Ødegård, J., Baranski, M., Gjerde, B., and Gjedrem, T. (2011). Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects. *Aquac. Res.* 42, 103–114. doi: 10.1111/j.1365-2109.2010.02669.x

Ozaki, A., Khoo, S. K., Yoshiura, Y., Ototake, M., Sakamoto, T., Dijkstra, J. M., et al. (2007). Identification of additional quantitative trait loci (QTL) responsible for susceptibility to infectious pancreatic necrosis virus in rainbow trout. *Fish Pathol.* 42, 131–140. doi: 10.3147/jsfp.42.131

Ozaki, A., Sakamoto, T., Khoo, S., Nakamura, K., Coimbra, M. R., Akutsu, T., et al. (2001). Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). *Mol. Genet. Genomics* 265, 23–31. doi: 10.1007/s004380000392

Ozaki, A., Yoshida, K., Fuji, K., Kubota, S., Kai, W., Aoki, J., et al. (2013). Quantitative Trait Loci (QTL) Associated with resistance to a monogenean parasite (*Benedenia seriolae*) in yellowtail (*Seriola quinqueradiata*) through genome wide analysis. *PLoS One* 8:e64987. doi: 10.1371/journal.pone.0064987

Palaiokostas, C., Cariou, S., Bestin, A., Bruant, J.-S., Haffray, P., Morin, T., et al. (2018a). Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. *Genet. Sel. Evol.* 50:30. doi: 10.1186/s12711-018-0401-2

Palaiokostas, C., Robledo, D., Vesely, T., Prchal, M., Pokorova, D., Piackova, V., et al. (2018b). Mapping and sequencing of a significant quantitative trait locus affecting resistance to koi herpesvirus in common carp. *G3* 8, 3507–3513. doi: 10.1534/g3.118.200593

Palti, Y., Gahr, S. A., Purcell, M. K., Hadidi, S., Rexroad, C. E., and Wiens, G. D. (2010). Identification, characterization and genetic mapping of TLR7, TLR8a1 and TLR8a2 genes in rainbow trout (*Oncorhynchus mykiss*). *Dev. Comp. Immunol.* 34, 219–233. doi: 10.1016/j.dci.2009.10.002

Palti, Y., Vallejo, R. L., Gao, G., Liu, S., Hernandez, A. G., Rexroad, C. E., et al. (2015). Detection and validation of QTL affecting bacterial cold water disease resistance in rainbow trout using restriction-site associated DNA sequencing. *PLoS One* 10:e0138435. doi: 10.1371/journal.pone.0138435

Paneru, B., Al-Tobasei, R., Palti, Y., Wiens, G. D., and Salem, M. (2016). Differential expression of long non-coding RNAs in three genetic lines of rainbow trout in response to infection with *Flavobacterium psychrophilum*. *Sci. Rep.* 6:36032. doi: 10.1038/srep36032

Quillet, E., Dorson, M., Aubard, G., and Torhy, C. (2001). *In vitro* viral haemorrhagic septicaemia virus replication in excised fins of rainbow trout: correlation with resistance to waterborne challenge and genetic variation. *Dis. Aquat. Organ.* 45, 171–182. doi: 10.3354/dao045171

Quillet, E., Dorson, M., Le Guillou, S., Benmansour, A., and Boudinot, P. (2007). Wide range of susceptibility to rhabdoviruses in homozygous clones of rainbow trout. *Fish Shellfish Immunol.* 22, 510–519. doi: 10.1016/j.fsi.2006.07.002

Råberg, L., Sim, D., and Read, A. F. (2007). Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. *Science* 318, 812–814. doi: 10.1126/science.1148526

Reyes-López, F. E., Romeo, J. S., Vallejos-Vidal, E., Reyes-Cerpa, S., Sandino, A. M., Tort, L., et al. (2015). Differential immune gene expression profiles in susceptible and resistant full-sibling families of Atlantic salmon (*Salmo salar*) challenged with infectious pancreatic necrosis virus (IPNV). *Dev. Comp. Immunol.* 53, 210–221. doi: 10.1016/j.dci.2015.06.017

Robinson, N. R., Gjedrem, T., and Quillet, E. (2017). "Improvement of disease resistance by genetic methods," in *Fish Diseases, Prevention and Control Strategies*, ed. G. Jeney (Cambridge, MA: Academic Press), 21–50. doi: 10.1016/B978-0-12-804564-0.00002-8

Robledo, D., Gutiérrez, A. P., Barría, A., Lhorente, J. P., Houston, R. D., and Yáñez, J. M. (2019). Discovery and functional annotation of quantitative trait loci affecting resistance to sea lice in Atlantic salmon. *Front. Genet.* 10:56. doi: 10.3389/fgene.2019.00056

Robledo, D., Matika, O., Hamilton, A., and Houston, R. D. (2018). Genome-wide association and genomic selection for resistance to Amoebic Gill Disease in Atlantic salmon. *G3* 8, 1195–1203. doi: 10.1534/g3.118.200075

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., and Houston, R. (2017). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquac.* 10, 670–682. doi: 10.1111/raq.12193

Robledo, D., Taggart, J. B., Ireland, J. H., McAndrew, B. J., Starkey, W. G., Haley, C. S., et al. (2016). Gene expression comparison of resistant and susceptible Atlantic salmon fry challenged with Infectious Pancreatic Necrosis virus reveals a marked contrast in immune response. *BMC Genomics* 17:279. doi: 10.1186/s12864-016-2600-y

Rodríguez, F. H., Flores-Mara, R., Yoshida, G. M., Barría, A., Jedlicki, A. M., Lhorente, J. P., et al. (2019). Genome-wide association analysis for resistance to infectious pancreatic necrosis virus identifies candidate genes involved in viral replication and immune response in rainbow trout (*Oncorhynchus mykiss*). *G3* 9, 2897–2904. doi: 10.1534/g3.119.400463

Rodríguez-Ramilo, S. T., De La Herrán, R., Ruiz-Rejón, C., Hermida, M., Fernández, C., Pereiro, P., et al. (2014). Identification of quantitative trait loci associated with resistance to viral haemorrhagic septicaemia (VHS) in turbot (*Scophthalmus maximus*): a comparison between bacterium, parasite and virus diseases. *Mar. Biotechnol.* 16, 265–276. doi: 10.1007/s10126-013-9544-x

Rodríguez-Ramilo, S. T., Fernández, J., Toro, M. A., Bouza, C., Hermida, M., Fernández, C., et al. (2013). Uncovering QTL for resistance and survival time to *Philasterides dicentrarchi* in turbot (*Scophthalmus maximus*). *Anim. Genet.* 44, 149–157. doi: 10.1111/j.1365-2052.2012.02385.x

Rodríguez-Ramilo, S. T., Toro, M. A., Bouza, C., Hermida, M., Pardo, B. G., Cabaleiro, S., et al. (2011). QTL detection for *Aeromonas salmonicida* resistance related traits in turbot (*Scophthalmus maximus*). *BMC Genomics* 12:541. doi: 10.1186/1471-2164-12-541

Salinas, I. (2015). The mucosal immune system of teleost fish. *Biology* 4, 525–539. doi: 10.3390/biology4030525

Saura, M., Carabaño, M., Fernández, A., Cabaleiro, S., Doeschl-Wilson, A., Anacleto, O., et al. (2019). Disentangling genetic variation for resistance and endurance to scuticociliatosis in turbot using pedigree and genomic information. *Front. Genet* 7:539. doi: 10.3389/fgene.2019.00539

Sawayama, E., Tanizawa, S., Kitamura, S.-I., Nakayama, K., Ohta, K., Ozaki, A., et al. (2017). Identification of quantitative trait loci for resistance to RSIVD in red sea bream (*Pagrus major*). *Mar. Biotechnol.* 19, 601–613. doi: 10.1007/s10126-017-9779-z

Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelson, T., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87. doi: 10.1126/science.1247005

Shalem, O., Sanjana, N. E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR–Cas9. *Nat. Rev. Genet.* 16, 299–311. doi: 10.1038/nrg3899

Shi, H., Zhou, T., Wang, X., Yang, Y., Wu, C., Liu, S., et al. (2018). Genome-wide association analysis of intra-specific QTL associated with the resistance for enteric septicemia of catfish. *Mol. Genet. Genomics* 293, 1365–1378. doi: 10.1007/s00438-018-1463-0

Sonesson, A. K., Gjerde, B., and Robinson, N. (2011). A simple selection scheme to improve disease resistance and growth. *Aquaculture* 319, 337–341. doi: 10.1016/j.aquaculture.2011.07.009

Stremlau, M., Owens, C. M., Perron, M. J., Kiessling, M., Autissier, P., and Sodroski, J. (2004). The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* 427, 848–853. doi: 10.1038/nature02343

Sutherland, B. J. G., Koczka, K. W., Yasuike, M., Jantzen, S. G., Yazawa, R., Koop, B. F., et al. (2014). Comparative transcriptomics of Atlantic *Salmo salar*, chum *Oncorhynchus keta* and pink salmon *O. gorbuscha* during infections with salmon lice *Lepeophtheirus salmonis*. *BMC Genomics* 15:200. doi: 10.1186/1471-2164-15-200

Tadmor-Levi, R., Hulata, G., and David, L. (2019). Multiple interacting QTLs affect disease challenge survival in common carp (*Cyprinus carpio*). *Heredity* 123, 565–578. doi: 10.1038/s41437-019-0224-0

Tan, S., Zhou, T., Wang, W., Jin, Y., Wang, X., Geng, X., et al. (2018). GWAS analysis using interspecific backcross progenies reveals superior blue catfish alleles responsible for strong resistance against enteric septicemia of catfish. *Mol. Genet. Genomics* 293, 1107–1120. doi: 10.1007/s00438-018-1443-4

Timmerhaus, G., Krasnov, A., Nilsen, P., Alarcon, M., Afanasyev, S., Rode, M., et al. (2011). Transcriptome profiling of immune responses to cardiomyopathy syndrome (CMS) in Atlantic salmon. *BMC Genomics* 12:459. doi: 10.1186/1471-2164-12-459

Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Bron, J. E., Taggart, J. B., et al. (2016). Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet. Sel. Evol.* 48:47. doi: 10.1186/s12711-016-0226-9

Vallejo, R. L., Palti, Y., Liu, S., Evenhuis, J. P., Gao, G., Rexroad, C. E., et al. (2014a). Detection of QTL in rainbow trout affecting survival when challenged with *Flavobacterium psychrophilum*. *Mar. Biotechnol.* 16, 349–360. doi: 10.1007/s10126-013-9553-9

Vallejo, R. L., Palti, Y., Liu, S., Marancik, D. P., and Wiens, G. D. (2014b). Validation of linked QTL for bacterial cold water disease resistance and spleen size on rainbow trout chromosome Omy19. *Aquaculture* 432, 139–143. doi: 10.1016/j.aquaculture.2014.05.003

Verrier, E. R., Dorson, M., Mauger, S., Torhy, C., Ciobotaru, C., Hervet, C., et al. (2013a). Resistance to a Rhabdovirus (VHSV) in rainbow trout: identification of a major QTL related to innate mechanisms. *PLoS One* 8:e55302. doi: 10.1371/journal.pone.0055302

Verrier, E. R., Ehanno, A., Biacchesi, S., Le Guillou, S., Dechamp, N., Boudinot, P., et al. (2013b). Lack of correlation between the resistances to two rhabdovirus infections in rainbow trout. *Fish Shellfish Immunol.* 35, 9–17. doi: 10.1016/j.fsi.2013.03.369

Verrier, E. R., Genet, C., Laloë, D., Jaffrezic, F., Rau, A., Esquerre, D., et al. (2018). Genetic and transcriptomic analyses provide new insights on the early antiviral response to VHSV in resistant and susceptible rainbow trout. *BMC Genomics* 19:482. doi: 10.1186/s12864-018-4860-1

Verrier, E. R., Langevin, C., Benmansour, A., and Boudinot, P. (2011). Early antiviral response and virus-induced genes in fish. *Dev. Comp. Immunol.* 35, 1204–1214. doi: 10.1016/j.dci.2011.03.012

Verrier, E. R., Langevin, C., Tohry, C., Houel, A., Ducrocq, V., Benmansour, A., et al. (2012). Genetic resistance to rhabdovirus infection in teleost fish is paralleled to the derived cell resistance status. *PLoS One* 7:e33935. doi: 10.1371/journal.pone.0033935

Vidalin, O., Muslmani, M., Estienne, C., Echchakir, H., and Abina, A. M. (2009). In vivo target validation using gene invalidation, RNA interference and protein functional knockout models: it is the time to combine. *Curr. Opin. Pharmacol.* 9, 669–676. doi: 10.1016/j.coph.2009.06.017

Wang, L., Liu, P., Huang, S., Ye, B., Chua, E., Wan, Z. Y., et al. (2017). Genome-wide association study identifies loci associated with resistance to viral nervous necrosis disease in Asian seabass. *Mar. Biotechnol.* 19, 255–265. doi: 10.1007/s10126-017-9747-7

Wang, W., Tan, S., Luo, J., Shi, H., Zhou, T., Yang, Y., et al. (2019). GWAS analysis indicated importance of NF-κB signaling pathway in host resistance against motile aeromonas septicemia disease in catfish. *Mar. Biotechnol.* 21, 335–347. doi: 10.1007/s10126-019-09883-0

Wiens, G. D., Vallejo, R. L., Leeds, T. D., Palti, Y., Hadidi, S., Liu, S., et al. (2013). Assessment of genetic correlation between bacterial cold water disease resistance and spleen index in a domesticated population of rainbow trout:

identification of QTL on chromosome Omy19. *PLoS One* 8:e75749. doi: 10.1371/journal.pone.0075749

Yamagishi, J., Natori, A., Tolba, M. E. M., Mongan, A. E., Sugimoto, C., Katayama, T., et al. (2014). Interactive transcriptome analysis of malaria patients and infecting *Plasmodium falciparum*. *Genome Res.* 24, 1433–1444. doi: 10.1101/gr.158980.113

Yamaguchi, T., and Dijkstra, J. M. (2019). Major histocompatibility complex (MHC) genes and disease resistance in fish. *Cells* 8:378. doi: 10.3390/cells8040378

Yáñez, J. M., Newman, S., and Houston, R. D. (2015). Genomics in aquaculture to better understand species biology and accelerate genetic progress. *Front. Genet.* 6:128. doi: 10.3389/fgene.2015.00128

Yáñez, J. M., Smith, P. A., Manneschi, G., Guajardo, A., Rojas, M. E., Díaz, S., et al. (2010). "Unravelling Genetic Co-variation For Resistance And Tolerance Against *Piscirickettsia salmonis* Infection In Atlantic Salmon (Salmo salar)," in *Proceedings of the Ninth World Congress Genetics Applied Livestock Production. Presented at the 9th World Congress Genetics Applied LIvestock Productions*, (Leipzig), 4.

Yáñez, J. M., Yoshida, G. M., Parra, Á, Correa, K., Barría, A., Bassini, L. N., et al. (2019). Comparative genomic analysis of three salmonid species identifies functional candidate genes involved in resistance to the intracellular bacterium *Piscirickettsia salmonis*. *Front. Genet.* 10:665. doi: 10.3389/fgene.2019.00665

Yang, M., Wang, Q., Wang, S., Wang, Y., Zeng, Q., and Qin, Q. (2019). Transcriptomics analysis reveals candidate genes and pathways for susceptibility or resistance to Singapore grouper iridovirus in orange-spotted grouper (*Epinephelus coioides*). *Dev. Comp. Immunol.* 90, 70–79. doi: 10.1016/j.dci.2018.09.003

Ye, H., Lin, Q., and Luo, H. (2018). Applications of transcriptomics and proteomics in understanding fish immunity. *Fish Shellfish Immunol.* 77, 319–327. doi: 10.1016/j.fsi.2018.03.046

Zapata, A., Diez, B., Cejalvo, T., Gutiérrez-de Frías, C., and Cortés, A. (2006). Ontogeny of the immune system of fish. *Fish Shellfish Immunol.* 20, 126–136. doi: 10.1016/j.fsi.2004.09.005

Zhou, Q., Su, Z., Li, Y., Liu, Y., Wang, L., Lu, S., et al. (2019). Genome-wide association mapping and gene expression analyses reveal genetic mechanisms of disease resistance variations in *Cynoglossus semilaevis*. *Front. Genet.* 10:1167. doi: 10.3389/fgene.2019.01167

Zhou, T., Liu, S., Geng, X., Jin, Y., Jiang, C., Bao, L., et al. (2017). GWAS analysis of QTL for enteric septicemia of catfish and their involved genes suggest evolutionary conservation of a molecular mechanism of disease resistance. *Mol. Genet. Genomics* 292, 231–242. doi: 10.1007/s00438-016-1269-x

Zwollo, P., Hennessey, E., Moore, C., Marancik, D. P., Wiens, G. D., and Epp, L. (2017). A BCWD-resistant line of rainbow trout exhibits higher abundance of IgT+ B cells and heavy chain tau transcripts compared to a susceptible line following challenge with *Flavobacterium psychrophilum*. *Dev. Comp. Immunol.* 74, 190–199. doi: 10.1016/j.dci.2017.04.019

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 18 August 2020
doi: 10.3389/fgene.2020.00913

High Cysteine Membrane Proteins (HCMPs) Are Up-Regulated During *Giardia*-Host Cell Interactions

*Dimitra Peirasmaki[1], Showgy Y. Ma'ayeh[1], Feifei Xu[1], Marcela Ferella[2], Sara Campos[3], Jingyi Liu[1] and Staffan G. Svärd[1,4]**

[1] Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden, [2] Eukaryotic Single Cell Genomics Platform, Karolinska Institute, Science for Life Laboratory (SciLifeLab), Solna, Sweden, [3] Department of Molecular Biology, Max Planck Institute for Infection Biology, Berlin, Germany, [4] Science for Life Laboratory (SciLifeLab), Uppsala University, Uppsala, Sweden

Giardia intestinalis colonizes the upper small intestine of humans and animals, causing the diarrheal disease giardiasis. This unicellular eukaryotic parasite is not invasive but it attaches to the surface of small intestinal epithelial cells (IECs), disrupting the epithelial barrier. Here, we used an *in vitro* model of the parasite's interaction with host IECs (differentiated Caco-2 cells) and RNA sequencing (RNAseq) to identify differentially expressed genes (DEGs) in *Giardia*, which might relate to the establishment of infection and disease induction. *Giardia* trophozoites interacted with differentiated Caco-2 cells for 1.5, 3, and 4.5 h and at each time point, 61, 89, and 148 parasite genes were up-regulated more than twofold, whereas 209, 265, and 313 parasite genes were down-regulated more than twofold. The most abundant DEGs encode hypothetical proteins and members of the High Cysteine Membrane Protein (HCMP) family. Among the up-regulated genes we also observed proteins associated with proteolysis, cellular redox balance, as well as lipid and nucleic acid metabolic pathways. In contrast, genes encoding kinases, regulators of the cell cycle and arginine metabolism and cytoskeletal proteins were down-regulated. Immunofluorescence imaging of selected, up-regulated HCMPs, using C-terminal HA-tagging, showed localization to the plasma membrane and peripheral vesicles (PVs). The expression of the HCMPs was affected by histone acetylation and free iron-levels. In fact, the latter was shown to regulate the expression of many putative giardial virulence factors in subsequent RNAseq experiments. We suggest that the plasma membrane localized and differentially expressed HCMPs play important roles during *Giardia*-host cell interactions.

Keywords: diarrhea, chromatin, RNAseq, small intestine, protozoa

OPEN ACCESS

Edited by:
Benjamin Makepeace, University of Liverpool, United Kingdom

Reviewed by:
Steven Singer, Georgetown University, United States
Andrew Paul Jackson, University of Liverpool, United Kingdom

***Correspondence:**
Staffan G. Svärd
Staffan.svard@icm.uu.se

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: *19 February 2020*
Accepted: *22 July 2020*
Published: *18 August 2020*

Citation:
Peirasmaki D, Ma'ayeh SY, Xu F, Ferella M, Campos S, Liu J and Svärd SG (2020) High Cysteine Membrane Proteins (HCMPs) Are Up-Regulated During Giardia-Host Cell Interactions. Front. Genet. 11:913. doi: 10.3389/fgene.2020.00913

INTRODUCTION

Diarrheal disease is still a major cause of death in children under 5 years of age, although it is both preventable and treatable (Kotloff et al., 2013). It is also the leading cause of malnutrition in young children (Kotloff et al., 2013). *Giardia intestinalis*, also known as *Giardia lamblia* or *Giardia duodenalis*, is a unicellular protozoan parasite that infects humans and other mammals, causing the diarrheal disease giardiasis (Ankarklev et al., 2010). Giardiasis is a major cause of

diarrhea worldwide with around 180 million symptomatic infections reported in humans every year (reviewed in Ryan et al., 2019). In developing countries, point prevalences range between 1.5 and 73.4%, reflecting a large number of asymptomatic infections (Farthing, 1997; Cacciò and Sprong, 2011). Elderly, malnourished or immunodeficient people are at risk of acquiring giardiasis but children are the most affected population by the disease, with reported effects on their growth, nutrition and cognitive function (Simsek et al., 2004; Prado et al., 2005; Nematian et al., 2008). The symptoms of acute giardiasis include watery diarrhea, abdominal pain, vomiting and weight loss. If the disease becomes chronic, symptoms of malabsorption (e.g., malaise and weight loss) become more prominent (Farthing, 1997; Ortega and Adam, 1997; Gascón, 2006) and children can become stunted (Rogawski et al., 2017).

Little is known about how *Giardia* causes disease; it is not invasive, does not secrete any known toxins and it causes little inflammation in the intestine (Buret, 2007, 2008; Ankarklev et al., 2010). Nevertheless, *Giardia* trophozoites can damage the intestinal epithelial cells (IECs) and the damage results in a decrease of the total absorptive area in the intestine and villus atrophy, leading to the malabsorption of water, glucose and electrolytes and maldigestion due to the loss of digestive enzymes on the IEC brush border (Scott et al., 2000; Troeger et al., 2007; Cotton et al., 2011; Humen et al., 2011; Solaymani-Mohammadi and Singer, 2011; Bénéré et al., 2012). Several reports have shown that the structural damage of the IECs is induced by *Giardia* trophozoite attachment (Magne et al., 1991; Céu Sousa et al., 2001) and release of excretory-secretory products (ESPs) like metabolic enzymes (Ringqvist et al., 2008), cysteine proteases (CPs, Liu et al., 2018), high cysteine membrane proteins (HCMPs), tenascins (Dubourg et al., 2018) and variable surface proteins (VSPs) (Ringqvist et al., 2011; Emery et al., 2016; Ma'ayeh et al., 2017; Dubourg et al., 2018).

Simple host–parasite interaction models have been established using axenic *Giardia* trophozoites from different assemblages [WB, P1 and NF (assemblage A) and GS (assemblage B)] and different intestinal cell-lines (Caco-2, HT-29, and IEC6, Roxström-Lindquist et al., 2005; Ringqvist et al., 2011; Ma'ayeh and Brook-Carter, 2012; Ferella et al., 2014). This has identified differentially expressed genes (DGEs) in both cell types and generated information about putative virulence factors (Einarsson et al., 2016a). However, these analyses have suffered from poor sensitivity of the methods used for DGE analyses, fragmented and incompletely annotated parasite genomes and problematic medium effects (Emery-Corbin et al., 2020; Jex et al., 2020).

In this study, we used our well-established model system of *Giardia* trophozoites (isolate WB) and human IECs (differentiated Caco-2 cells), combined with medium pre-incubation and medium controls, a new version of the *Giardia* WB genome (Xu et al., 2020) and an in-house RNA sequencing (RNAseq) pipeline to generate a more complete differentially expressed genes (DGE) data set from trophozoites interacting with IECs. We used the new data set together with earlier data to pin down the most commonly differentially expressed *Giardia* trophozoite genes during interactions with human IECs *in vitro*. The combined data suggest that the stress-regulated HCMPs and other cysteine-rich proteins are important during host–parasite interactions. Follow-up studies showed that up-regulated HCMPs localize to the trophozoite plasma membrane and that they are regulated by histone acetylation and levels of free-iron in the medium.

MATERIALS AND METHODS

Cell Culture

The human colon adenocarcinoma cell line (Caco-2, clone TC7) (Roxström-Lindquist et al., 2005) was used in the experiments at a differentiated state (Lievin-Le Moal, 2013). Caco-2 cells (Passage no. 6–8) were cultured in 25 cm^2 tissue culture flasks (T25) filled with 10 ml of complete Dulbecco's Modified Eagle Medium (DMEM) containing 10% fetal bovine serum (FBS), 2 mM GlutaMAX (Gibco, Thermo Fisher Scientific, MA, United States), 1× MEM non-essential amino acid solution (Sigma-Aldrich, MO, United States) and 1× of penicillin-streptomycin 100× solution (10,000 units penicillin and 10 mg streptomycin/ml) (Sigma-Aldrich, MO, United States). Culture flasks were incubated in a humidified incubator (10% CO_2 and 37°C) for 21 days until differentiation during which the media were changed twice weekly. For host–parasite interactions, heat inactivated (HI)-FBS (Gibco, Thermo Fisher Scientific) was used in the DMEM to avoid adverse effects of active serum components on trophozoites.

Giardia intestinalis isolate WB, clone C6 (ATCC 30957) trophozoites were cultured in 50 or 10 ml tubes filled with TYDK medium (Keister, 1983), supplemented with 10% heat-inactivated bovine serum (Gibco, Thermo Fisher MA, United States). All tubes were incubated at 37°C until reaching peak density (80% confluence) upon which they were used in the experiments. All materials used in the TYDK medium were purchased from Sigma-Aldrich unless otherwise stated.

Cell–Cell Interactions

At the beginning of the experiment, both trophozoites and human cell cultures (T25) were washed twice with warm PBS (37°C) and replenished with fresh DMEM with 10% HI-FBS (50 ml for trophozoite culture and 9 ml for human cell culture). Incubating the trophozoites in complete DMEM prior to the experiment was deemed necessary to reduce shifts in gene transcription due to the change of media (i.e., from TYDK to DMEM). Initially, all cultures were incubated in a humidified tissue culture incubator for 2 h at 37°C after which trophozoites were processed for addition to the differentiated Caco-2 cells. This involved incubating culture tubes on ice (10 min) for detachment, counting (Neubauer Chamber slide), pelleting (centrifugation; 7 min, 750 × *g* and 4°C), resuspension in 1 ml of DMEM (3×10^7 cells) and finally addition to the differentiated Caco-2 cells (T25 flasks). Trophozoites were incubated with the differentiated Caco-2 cells for 1.5, 3, and 4.5 h (10% CO_2 at 37°C). At the end of each time point, the interaction medium was removed from the flask and the co-culture was lysed directly in 1.5 ml of lysis buffer for RNA

extraction. The lysis buffer was included in the PureLink RNA Mini Kit (Ambion, Thermo Fisher Scientific) together with beta-mercaptoethanol, added directly prior to cell lysis. For the starting control, trophozoites, pre-incubated for 2 h in DMEM as above, were pelleted and lysed directly in the RNA lysis buffer. All samples were collected in RNase-free Eppendorf tubes and frozen immediately in dry ice. Samples were kept at −80°C until RNA extraction. The experiment was repeated 3 times.

Iron Regulation of HCMPs

To minimize the heterogeneity of expressed VSPs and HCMPs in the culture, we started off by creating a clonal trophozoite population from the original WB-C6 clone using serial dilution. Briefly, a trophozoite culture was diluted in TYDK and seeded as single cells into the wells of a 96-well plate. Upon reaching 70–80% confluence, a clonal culture was selected used to seed 10 ml tubes containing TYDK, TYDK without added ferric (i.e., ferric ammonium citrate), and TYDK without ferric and supplemented with 50 μM of the metal ion chelator 2,2′-Bipyridyl (Sigma-Aldrich, MO, United States). The concentration of 2,2′-Bipyridyl was the highest that did not affect growth of *Giardia* trophozoites in TYDK. Iron analyses at ALS (Umeå, Sweden) showed that the standard TYDK contained 5.5 mg/l iron, TYDK without added iron and with added chelator 1.6 mg/ml. It should be noted that the chelator reduces the available, free iron but the total level is not decreased. Upon reaching 80% confluence, trophozoites were detached on ice (10 min), collected by centrifugation (7 min, 750 × *g*, and 4°C), washed with ice-cold PBS, pelleted (7 min, 750 × *g*, and 4°C), lysed in 1.5 ml of Trizol and collected in Eppendorf tubes. All collected samples were frozen immediately in dry ice and kept at −80°C until RNA extraction. The experiment was repeated 3 times.

RNA Extraction, Library Preparation, and RNA Sequencing

The samples collected above were processed according to the instructions provided in the PureLink RNA Mini Kit (Ambion, Thermo Fisher scientific). A DNaseI treatment step (PureLink DNase Set, Ambion, Thermo Fisher Scientific) was performed during RNA extraction to remove genomic DNA before eluting the RNA. RNA quality was assessed by evaluating the 260/280 and 260/230 ratios (NanoDrop 1000 Spectrometer, Thermo Fisher Scientific) and electrophoresing the samples (500 ng) on a 1.5% Tris-Borate-EDTA (TBE) agarose gel containing 20 mM of guanidium isothiocyanate (GITC). Sequencing libraries were prepared from 500 ng of total RNA using the TruSeq stranded mRNA library preparation kit (Cat no. RS-122-2101/2102, Illumina Inc.), which included a polyA selection step. Libraries preparation was performed following the manufacturers' protocol (no. 15031047). The quality of prepared libraries was evaluated using a Fragment Analyzer from Advanced Analytical (AATI) using the DNF-910 kit and they were quantified by qPCR using the Library quantification kit for Illumina (KAPA Biosystems) on a CFX384 Touch instrument (Bio-Rad) prior to cluster generation and sequencing. Sequencing was carried out on an Illumina NovaSeq6000 instrument (NVCS v 1.3.0/RTA v3.3.3) according to the manufacturer's instructions. De-multiplexing and conversion to FASTQ format was performed using the bcl2fastq2 (2.20.0.422) software, provided by Illumina[1]. Additional statistics on sequencing quality were compiled with an in-house script from the FASTQ-files, RTA and BCL2FASTQ2 output files. RNA sequencing was performed at the SciLifeLab NovaSeq Sequencing Platform, Uppsala University, Sweden.

Bioinformatics Analyses of RNA Sequencing Data

We have produced our own pipeline for RNAseq analyses in *Giardia* and all scripts used in the bioinformatics analyses are available upon request. Essentially STAR v020201 (Dobin et al., 2013) was used to map the RNA-Seq reads to the new, more complete *Giardia* WB reference genome (Xu et al., 2020) and the raw counts per gene were generated with the parameter "–quantMode GeneCounts." Downstream gene differential expression analysis was carried out in R using edgeR v3.6.8 (Robinson et al., 2010) workflow. Quasi-likelihood (QL) *F*-test (glmQLFTest) was used to determine significant differential gene expression. Genes with adjusted p-value ≤ 0.05 were considered significant and reported here. Raw reads and the processed raw counts per gene were deposited at gene Expression Omnibus (GEO), available as accession ID GSE144004 for the interaction experiment and GSE136820 for the iron depletion experiment.

Construction of Plasmids With Epitope-Tagged HCMPs and Transfection of Trophozoites

Three HCMPs were selected for further immunofluorescence studies (ORFs 7715, 91707, and 115066). The selection was based on their differential gene expression in all earlier published *Giardia*-host cell interaction reports (Ringqvist et al., 2011; Ma'ayeh and Brook-Carter, 2012; Ferella et al., 2014). The selected genes were PCR-amplified from genomic DNA of the WB isolate. For C-terminal HA-tagging primers were designed to amplify the gene of interest, including 100 bp upstream of the start codon (i.e., putative promoter region) and the complete codon region, except the stop codon. Adaptor sequences containing unique restriction sites were added to the 5′ end of primers and the whole sequence was checked for inframe translation. All primers were analyzed using the online tool OligoAnalyzer 3.1 – Integrated DNA Technologies (IDT) and are shown in **Supplementary Table S1**. Each PCR reaction contained 20 ng of DNA, 0.6 μL of 10 mM dNTPs, 1.2 μL of each forward and reverse primers (200 nM), 6 μL of 5× High Fidelity PCR buffer with $MgCl_2$ (Thermo Fisher Scientific), 0.5 μL of Phusion Hot Start II High-Fidelity DNA Polymerase (proofreading – Thermo Fisher Scientific) and up to 30 μL of sterile water. PCR amplification was performed following the standard protocols, including initial denaturation and enzyme

[1]http://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html

activation (98°C, 5 min), denaturation (15 s), annealing (variable annealing temperature based on primers, see **Supplementary Table S1**) and primer extension (35 cycles, 72°C for 80 s), and a final extension step at 72°C for 5 min. Amplicons were gel-purified using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The purified products were double-digested with the appropriate restriction enzymes (HindIII and EcoRI for ORFs 7715 and 91707 and XbaI and EcoRI for ORF 115066) and ligated into the episomal pPAC-3xHA-C-terminus plasmid vector cleaved with the same restriction enzymes as previously described (Jerlström-Hultqvist et al., 2012). Plasmids (∼20 μg) were transfected into *Giardia* trophozoites (Jerlström-Hultqvist et al., 2012) and the transfectants were grown at 37°C under puromycin selection (50 μg/ml) to establish a stable transfectant line.

The same plasmids were used to produce integrated C-terminal constructs. The C-terminal constructs were cleaved with restriction enzymes with unique sites in each HCMP gene (7715-Sph1, 91707-MfeI, and 115066-PmaC1), generating linearized constructs that were gel purified and 20 μg linearized plasmid was used in transfection as above. Genomic PCRs using primers outside of the construct and in the plasmids were used to verify that the plasmids were integrated into the correct genomic sites.

The three HCMPs were also cloned with N-terminal HA tags downstream of the signal sequence of each HCMP, followed by the complete HCMP gene sequences, including the stop codons. The cloning was done in two steps, generating two PCR products that were fused and cloned into the pPAC plasmid without the HA tag (cut at the EcoRI and SmaI site for ORFs 7715 and 91707 and XbaI and SmaI sites for ORF 115066). Plasmids were transfected into *Giardia* trophozoites as previously described and positive clones were selected under puromycin pressure (50 μ g/ml).

Immunofluorescence of HCMP Transfectants

Transfectants were examined by immunofluorescence for the localization of the HCMPs. Fifteen microliter of trophozoite culture were pipetted onto a poly-L-lysine coated microscopy slide and allowed to attach for 5 min at 37°C in a humidified chamber. Attached trophozoites were fixed for 20 min using 15 μl of 4% paraformaldehyde (PFA) in PBS and incubated at 37°C. The fixative was then removed using vacuum suction and neutralized with 15 μl of 0.1 M glycine dissolved in PBS. Cells were rinsed in PBS and permeabilized with 0.1% Triton-X 100 in PBS for 30 min at 37°C, washed with PBS and blocked overnight (4°C) in a blocking solution (2% bovine serum albumin dissolved in PBS and 0.05% Triton-X 100). Next day, the blocking solution was removed and each well of the slide was incubated with 20 μl of an Anti–HA mouse monoclonal antibody (1:500 dilution, Sigma-Aldrich) for 1 h at room temperature. Next, the antibody was removed using vacuum suction and the cells were washed with PBS followed by 1-h incubation at room temperature with 20 μl of Alexa Fluor 488-conjugated Goat anti-mouse antibody (1:250 dilution, Thermo Fisher). Finally, antibodies were removed and the wells were washed with PBS and mounted using Vectashield containing the DNA stain 4′, 6′-diamidino-2-phenyldone (DAPI) (Vector Laboratories, place and country). The slides were stored at −20°C in darkness. The slides were examined using a Zeiss Axioplan2 fluorescence microscope and the images were processed using the software Axiovision Rel. 4.8.

Transfectants were also followed by IFA for 37 generations after initial transfection. Two hundred cells were counted each time and valuated according to the cellular localization of the expressed protein.

Oxidative Stress Regulation of HCMPs

To study the medium effects on HCMP expression we made use of our C-terminally HA-tagged episomal constructs of HCMP 7715, 91707, and 115066. Transfected parasites (1×10^7 cells) were incubated for 3 h in 6-well plates filled with medium at 5 different conditions: TYDK, DMEM, DMEM + Caco-2 cells, DMEM + 1 mM cysteine, and DMEM + 1 mM cysteine + Caco-2 cells. Trophozoites were detached on ice (10 min), collected by centrifugation (7 min, 750 × *g*, and 4°C), washed with ice-cold PBS, pelleted (7 min, 750 × *g* and 4°C), lysed in 1.5 ml of Trizol and collected in Eppendorf tubes. All collected samples were frozen immediately in dry ice and kept at −80°C until RNA extraction. The experiment was repeated 3 times.

Chromatin Regulation of HCMPs

To determine whether the HCMPs are epigenetically regulated, we used inhibitors of histone deacetylases (HDACs) to examine their effect on RNA levels of selected HCMPs showing differential gene expression in this and earlier gene expression studies; GL50803_7715, GL50803_9620, GL50803_11309, GL50803_91707, and GL50803_115066. The inhibitors of NAD^+-independent HDAC inhibitors [trichostatin A (TSA) and sodium butyrate (NaB)] and the inhibitor of NAD^+-dependent HDACs [nicotinamide (Nt)] have been previously shown to interfere with epigenetic regulation in *Giardia* (Carranza et al., 2016) and thus, they were selected for this study. The earlier reported conditions for treatment were used and trophozoites (10^4 starting trophozoites) were incubated in TYDK for 1 or 3 days with each inhibitor at a final concentrations of 200 nM for TSA, 20 nM for NaB and 10 nM for Nt (Carranza et al., 2016). At the end of all incubations, treated trophozoites were detached from the tubes wall (10 min on ice), pelleted (750 × *g*, 4°C and 10 min) and lysed in 1.5 ml of Trizol (Ambion, Thermo Fisher Scientific) and transferred into Eppendorf tubes. For the control, trophozoites incubated for the same duration without the inhibitor were processed as stated previously and lysed directly in Trizol. All collected samples were frozen immediately in dry ice and kept at −80°C until RNA extraction. RNA extraction was performed according to the instructions provided with the Trizol reagent. Extracted RNA was checked for quality as stated in a previous section. The experiment was repeated four times.

Quantitative Polymerase Chain Reaction (qPCR)

High quality intact RNA samples were processed for cDNA synthesis. First, 1 μg of each sample was DNase I (Fermentas, Thermo Fisher) treated and the DNA-free RNA was reverse transcribed according to the instructions in the Revert Aid H Minus cDNA Synthesis Kit (Thermo Fisher Scientific). Oligo-$(dT)_{18}$ primers were used for cDNA synthesis. cDNA reaction mixtures were diluted (1.5–5 ng template per reaction) and used in the qPCR reactions together with 250 nM of each primer and the Maxima SYBR Green qPCR Master Mixes (Thermo Fisher Scientific). Reactions were set up in 20 μl volume and run following the manufacturer's instructions with the inclusion of melt curve analysis in the end of the run. qPCRs were performed in a StepOnePlus thermal cycler (Applied Biosystems, Carlsbad, CA, United States). All primers (Sigma-Aldrich) used in the qPCRs are listed in **Supplementary Table S1**. The gene encoding tryptophanyl-tRNA-synthetase (TtRNA) (GL50803_3032) was used as an endogenous control in the qPCR reactions (Einarsson et al., 2016b). The fold change in gene expression was calculated using the $\Delta\Delta Ct$ method. Significant changes in RNA levels between treatments and controls were assessed using one-way analysis of variance (ANOVA) at $\alpha < 0.05$ followed by the Dunnett's multiple comparison test at $P < 0.05$.

RESULTS

RNAseq Analyses of *Giardia*-Host Cell Interactions

RNA sequencing was performed on *Giardia intestinalis* trophozoites (isolate WB) incubated with and without differentiated Caco-2 cells, in three biological replicates. A major difference in this experiment compared to earlier studies was a 2 h pre-incubation of the trophozoites in DMEM before addition to the Caco-2 cells in order to reduce the medium effects that have been seen to dominate gene expression changes in earlier experiments (Ringqvist et al., 2011; Ma'ayeh and Brook-Carter, 2012; Ferella et al., 2014). We also used the new version of the *Giardia* WB genome, which is more complete and better annotated (Xu et al., 2020). EdgeR was used to analyze the RNA-Seq reads to obtain the differentially expressed genes (DEGs). Overall, an average of 40.4 million paired-end reads were obtained for the starting trophozoite population (DMEM pre-incubated trophozoites), amongst which 79.4% mapped to the *Giardia* WB reference genome (Xu et al., 2020). For trophozoites co-incubated with IECs, we obtained 42.3, 40.4, and 44.7 million paired-end reads on average at the 1.5, 3, and 4.5 h time points, with 53.1, 51.3, and 47.4% of the respective reads mapped to the *Giardia* WB genome. We did comparisons of gene expression between starting trophozoites and trophozoites after different times of interaction with Caco-2 cells. EdgeR identified 2,840, 2,968, and 3,103 DEGs (adjusted p-value ≤ 0.05) for each timepoint compared to the starting trophozoites, and 2,306 of the DEGs were shared between all timepoints (**Figure 1A**). The overlap of DEGs between the different timepoints increases with time and there are less unique DEGs at 3 h (**Figure 1A**). Comparisons between each time-point generated only 2DEGs from the 1.5–3 h comparison (HCMP114089 and FixW6289) and none from the 3–4.5 h comparison. Detailed information about all the DEGs can be found in **Supplementary Table S2** (1.5 h), **Supplementary Table S3** (3 h), and **Supplementary Table S4** (4.5 h). For each time point, 61, 89, and 148 DEGs were up-regulated more than twofold ($\log_2 FC > 1$) whereas 209, 265, and 316 DEGs were down-regulated more than twofold ($\log_2 FC < -1$), respectively (**Supplementary Tables S2–S4**). A total of 35 genes were up-regulated threefold or more (**Table 1**). Only five of these DEGs were found among the 30 most up-regulated genes in a microarray-based study of trophozoites co-incubated with differentiated Caco-2 cells (ORFs 5800, 10659, 114210, 115066, and 137727) (Ringqvist et al., 2011), showing the effects of the different approaches. It was also a partial overlap with genes up-regulated in three different *Giardia* isolates (WB, P-1, and NF) incubated with murine IEC6 cells (ORFs 5800, 114210, 115066, and 137727) (Ma'ayeh and Brook-Carter, 2012). Three of the commonly up-regulated DEGs (ORFs 10659, 115066, and 137727) are HCMPs and it is the most represented gene family (8 of 35 genes, 23%) in **Table 1**. The majority of the 32 DEGs down-regulated fourfold or more were down-regulated already at 1.5 h (**Table 2**). Several of these DEGs are involved in the regulation of cell cycle such as MAD2 (ORF100955) and cyclin (ORF 17400) as well arginine metabolizing enzymes (e.g., arginine deiminase, and carbamate kinase), suggesting a reduced arginine consumption and replication of the parasites already after 1.5 h co-incubation.

A GO term analysis for the enrichment of biological terms within the up-regulated DEG population at the three time points showed overlapping functions, involving cellular processes, metabolic processes, gene expression, ribosome biogenesis and translation (**Figure 1B**). The commonality in these biological functions can be explained by the high number of overlapping DEGs among the three time points. A further curation of DEGs for the enrichment of molecular functions produced groups for catalytic activity, hydrolase activity, RNA binding, nucleic acid binding and ribosome biogenesis (**Supplementary Figure S1**). Exclusive to the 1.5 h time point was the emergence of the functional groups, peptidase activity (GO:0008233, 32 DEGs) and oxidoreductase activity (GO:0016491, 25 DEGs). The peptidase activity group contained genes encoding cysteine proteases (e.g., ORFs 10217, 15564, 114915, and 137680), metalloproteases (e.g., ORFs 9508, 16823, and 93551), proteasome subunits (e.g., 3209, 13127, 33166, and 91643) and HCMPs with putative leishmanolysin-like peptidase activity (ORFs, 9620, 112432, and 137715) (**Supplementary Table S2**). The oxidoreductase group contained genes encoding proteins with peroxidase activity including the peroxiredoxins, nitroreductases and NADH oxidase (e.g., ORFs 3042, 6289, 9355, 9719 14521, 15009, 15307, 16076, 22677, and 23888). We further analyzed the down-regulated DEGs at the three time points for GO terms and this showed an enrichment of the biological functions, cellular processes, metabolic process and phosphorylation including phosphate-containing compounds metabolism (**Figure 1C**). A molecular functions

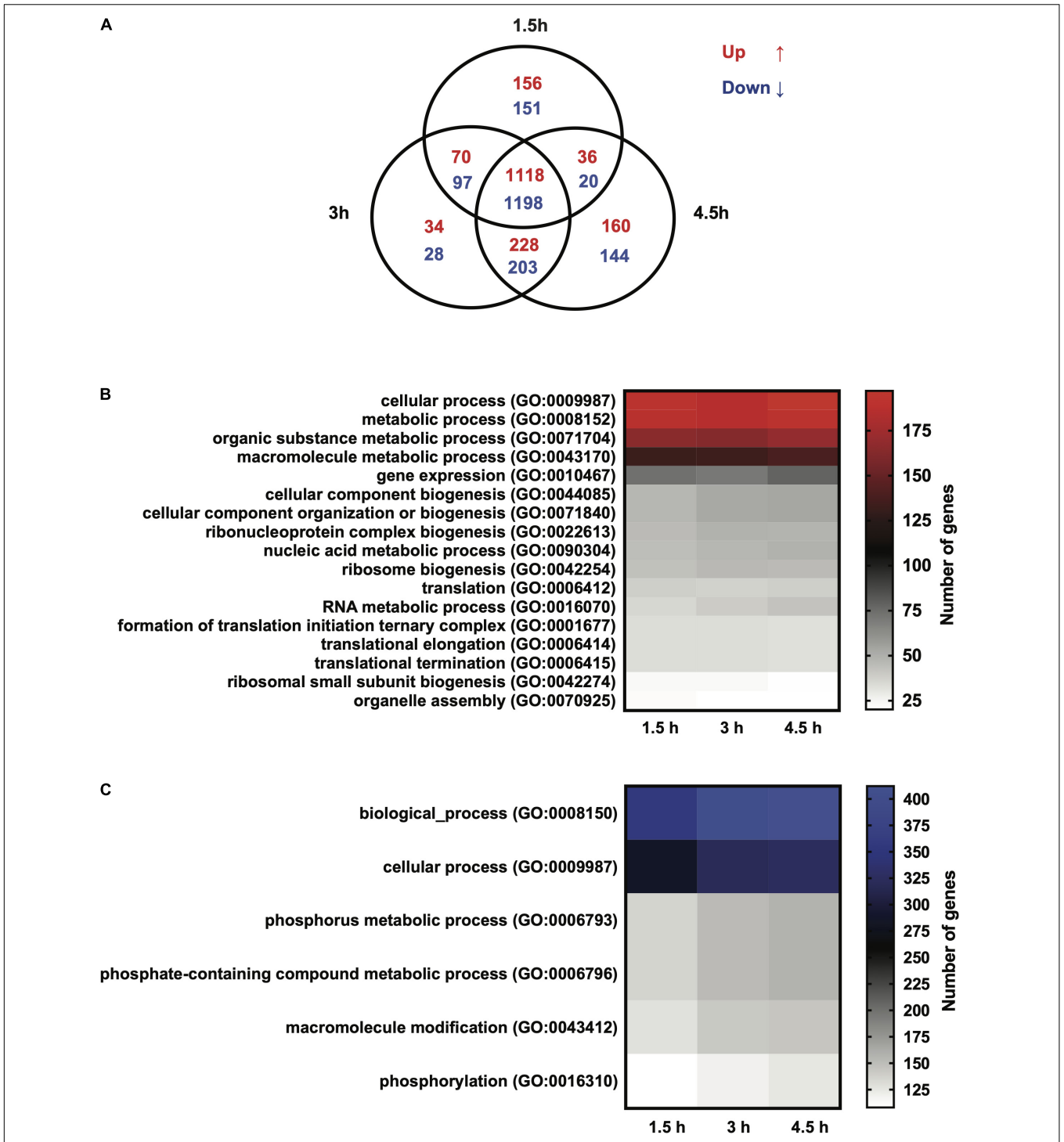


FIGURE 1 | (A) Venn diagram showing the number of differentially expressed genes (DEGs) in *Giardia intestinalis* WB trophozoites incubated with differentiated Caco-2 cells *in vitro* for 1.5, 3, and 4.5 h. **(B)** Gene Ontology (GO) analysis of the DEGs at the three time points represented by a heat map. The map shows the up-regulated DEGs enriched for biological functions. **(C)** Gene Ontology (GO) analysis of DEGs at the three time points represented by a heat map. The map shows the down-regulated DEGs enriched for biological functions. Note that up-regulated (red) and down-regulated (blue) heat maps are color-coded.

query, on the other hand, showed many functions associated with binding such as organic cyclic compounds, nucleotides ribonucleotide, nucleoside phosphate, ATP, carbohydrate derivatives and ions (**Supplementary Figure S1**). Analyses of the 3 and 4.5 h time points for GO terms produced more molecular groups with the down-regulated DEGs, including

TABLE 1 | Genes up-regulated more than threefold during co-incubation of *Giardia intestinalis* WB isolate trophozoites with differentiated Caco-2 cells *in vitro* for 1.5, 3, and 4.5 h.

Gene ID	Description	1.5 h	3 h	4.5 h
GL50803_3470	ABC transporter family protein	4.6	6.1	6.1
GL50803_137727	High cysteine membrane protein EGF-like	4.0	4.3	4.6
GL50803_5800	Lipid binding protein	3.2	4.0	4.6
GL50803_26679	Hypothetical protein	2.1	3.7	4.3
GL50803_d15250	High cysteine membrane protein Group 6	3.2	4.3	4.0
GL50803_8227	ATP-binding cassette protein 5	3.5	4.3	3.7
GL50803_27713	Hypothetical protein	1.4	2.3	4.0
GL50803_102110	Chromosome segregation protein SMC	NA	2.6	4.0
GL50803_17121	Bip	4.0	2.1	2.0
GL50803_114089	High cysteine membrane protein Group 4	NA	2.3	3.7
GL50803_8883	Hypothetical protein	2.0	3.0	3.7
GL50803_114210	Hypothetical protein	1.6	2.5	3.7
GL50803_7195	Glutamate synthase	3.5	3.7	3.5
GL50803_115066	High cysteine membrane protein VSP-like	2.5	3.0	3.5
GL50803_32419	Hypothetical protein	2.3	3.2	3.5
GL50803_113415	Hypothetical protein	2.6	3.0	3.5
GL50803_16568	Transcription factor, putative	3.5	3.0	3.5
GL50803_16424	Disk-associated protein	2.0	2.5	3.2
GL50803_9620	High cysteine membrane protein Group 2	2.3	3.2	3.2
GL50803_11309	High cysteine membrane protein Group 1	NA	2.5	3.2
GL50803_4191	Kinase, CMGC CDK	NA	1.6	3.1
GL50803_61026	Hypothetical protein	2.2	2.9	3.1
GL50803_14164	Ankyrin repeat protein 1	2.1	2.3	3.0
GL50803_10659	High cysteine membrane protein Group 1	1.9	2.5	3.0
GL50803_8377	Putative TM nitroreductase	2.1	2.5	3.0
GL50803_112885	Adenine phosphoribosyltransferase	1.9	2.5	3.0
GL50803_22547	High cysteine membrane protein Group 2	2.6	2.8	3.0
GL50803_26835	PIG-X/PBN1 family protein	1.9	2.5	3.0
GL50803_22504	Hypothetical protein	2.3	2.6	3.0
GL50803_34093	Ribosomal L38e	2.0	2.3	3.0
GL50803_14670	Protein disulfide isomerase PDI3	3.0	2.1	2.3
GL50803_114636	Hypothetical protein	2.6	3.0	2.8
GL50803_15450	Hypothetical protein	2.8	3.0	3.0
GL50803_29692	Hypothetical protein	NA	2.7	3.0

Expression values in fold change.

transferase activity (transferring phosphorus-containing groups), phosphotransferase activity, kinase activity and hydrolase activity (**Supplementary Figure S1**). Overall, the above changes show the complexity of cellular and metabolic responses in *Giardia* trophozoites incubated with IECs.

In total 53 DEGs encode HCMPs that are expressed in at least two timepoints (**Supplementary Table S5**). 34 of the differentially expressed HCMPs were up-regulated with 13 being up-regulated more than twofold (**Supplementary Table S5**). On the other hand, 19 HCMPs were down-regulated at least two timepoints with four that were down-regulated to less than 0.5-fold (**Supplementary Table S5**). In addition, there were 17 HCMPs that were only identified at specific interaction time points; 6 (7 up- and 3 down-regulated) at 1.5 h, 1 (down-regulated) at 3 h and 10 (5 up- and 5-down-regulated) at 4.5 h (**Supplementary Table S5**). Overall, these results show that 70 of a total of 116 (60%) HCMP genes in the WB genome are differentially expressed during trophozoite-IEC interactions.

Localization of Up-Regulated HCMP Proteins During *Giardia*-Host Cell Interactions

Due to the lack of expression and localization data of most HCMPs, three HCMP genes (HCMP 7715, 91707, and 115066) that were differentially expressed during IEC interactions in our and in earlier studies (Ringqvist et al., 2011; Ma'ayeh and Brook-Carter, 2012; Ferella et al., 2014) were selected for further characterization and cloned in episomal *Giardia* expression-vectors tagged with a 3XHA epitope at the C-terminus (Jerlström-Hultqvist et al., 2012). Microscopic examination of the transfected parasites revealed a plasma membrane localization for all 3 HCMPs (**Figures 2A–C**), combined with

TABLE 2 | Genes down-regulated more than fourfold during co-incubation of *Giardia intestinalis* WB isolate trophozoites with differentiated Caco-2 cells *in vitro* for 1.5, 3, and 4.5 h.

Gene ID	Description	1.5 h	3 h	4.5 h
GL50803_21116	Kinase, CMGC CMGC-GL1	0.1	0.1	0.1
GL50803_100955	Mitotic spindle checkpoint protein MAD2	0.2	0.1	0.1
GL50803_112103	Arginine deiminase	NA	0.2	0.1
GL50803_17090	Giardia trophozoite antigen GTA-1	0.3	0.2	0.1
GL50803_87577	Hypothetical protein	0.2	0.2	0.1
GL50803_16367	Hypothetical protein	0.2	0.1	0.1
GL50803_102322	Glycosyltransferase, putative	0.2	0.1	0.1
GL50803_10527	Hypothetical protein	0.2	0.2	0.1
GL50803_17400	Cyclin	0.2	0.2	0.1
GL50803_16534	Ankyrin repeat protein 1	0.2	0.2	0.2
GL50803_7538	Hypothetical protein	0.2	0.2	0.2
GL50803_8044	Seven transmembrane protein 1	0.2	0.2	0.2
GL50803_7137	Hypothetical protein	0.2	0.2	0.2
GL50803_16802	Kinase, CMGC CDK	0.2	0.2	0.2
GL50803_8263	Hypothetical protein	0.2	0.2	0.2
GL50803_5188	Ankyrin repeat protein 1	0.3	0.2	0.2
GL50803_86933	Hypothetical protein	0.2	0.2	0.3
GL50803_11940	Hypothetical protein	0.3	0.2	0.3
GL50803_95593	Kinase, NEK	0.3	0.2	0.2
GL50803_13651	Hypothetical protein	0.3	0.2	0.2
GL50803_16453	Carbamate kinase	0.5	0.3	0.2
GL50803_14971	SMC family protein	0.3	0.2	0.2
GL50803_8726	Hypothetical protein	0.3	0.2	0.2
GL50803_8980	Hypothetical protein	0.3	0.2	0.2
GL50803_7390	tRNA-dihydrouridine synthase 2	0.4	0.2	0.2
GL50803_16343	Median body protein	0.3	0.3	0.2
GL50803_11151	Hypothetical protein	0.3	0.3	0.2
GL50803_113610	GlcNAc-PI synthesis protein	0.3	0.3	0.2
GL50803_2622	Phosphatidylinositol-4-phosphate 5-kinase	0.4	0.3	0.25
GL50803_17558	Kinase, CMGC DYRK	0.3	0.3	0.25
GL50803_13133	Hypothetical protein	0.3	0.3	0.25
GL50803_5183	Hypothetical protein	0.3	0.25	0.25

Expression values in fold change.

spotty cytoplasmic fluorescence for HCMPs 7715 and 91707 (**Figures 2A,B**). The initial plasma membrane localization of the recently transfected, episomal 91707-HA construct (**Figure 2B**) gradually changed to a more internal localization around the pheripheral vesicles (PVs) during later generations (**Supplementary Figure S2**), whereas the other two constructs were stable for at least 30 generations. In order to verify that the localization from the episomal constructs is not an artifact due to over-expression we also produced integrated versions of the C-terminally tagged HCMPs. The episomal vectors were cut in the coding sequences of the HCMPs and the constructs were integrated into their cognate chromosomal loci as in Gourguechon and Cande (2011). The integrated constructs showed a spotty plasma membrane localization for all three HCMPs (**Supplementary Figure S3**) and more than 80% of the cells were positive, compared to 50% or less in the episomal constructs (**Supplementary Figure S2**). However, the immunofluorescence signal per cell was weaker, suggesting lower expression. Co-staining with Cellmask, a plasma membrane stain, supported the localization to the plasma membrane of HCMP115066 (**Supplementary Figure S3**). For HCMPs with N-terminal HA-fusions, the expressed proteins stayed in the ER, suggesting a disturbed protein transport (**Supplementary Figure S4**).

To gain further insights into the localization of the selected HCMPs in co-culture with IECs, the three C-terminal episomal HA fusion transfectants were added to differentiated Caco-2 cells and the localization of the HCMPs was examined after 0 and 3 h of co-incubation (**Figures 2D–I**). The interactions were performed in DMEM as opposed to the localization above, which were performed after growth in *Giardia* growth medium (TYDK). All examined HCMPs localized to the plasma membrane initially (**Figures 2D,F,H**). After 3 h co-incubation with IECs, the signal for HCMP 7715 was lost on the parasites (**Figure 2E**). It is possible that this protein had been secreted since we could previously detect HCMP 7715 in the culture supernatant during co-incubation with IECs (Ma'ayeh et al., 2017). For the HCMPs 91707 and 115066, a stronger signal could

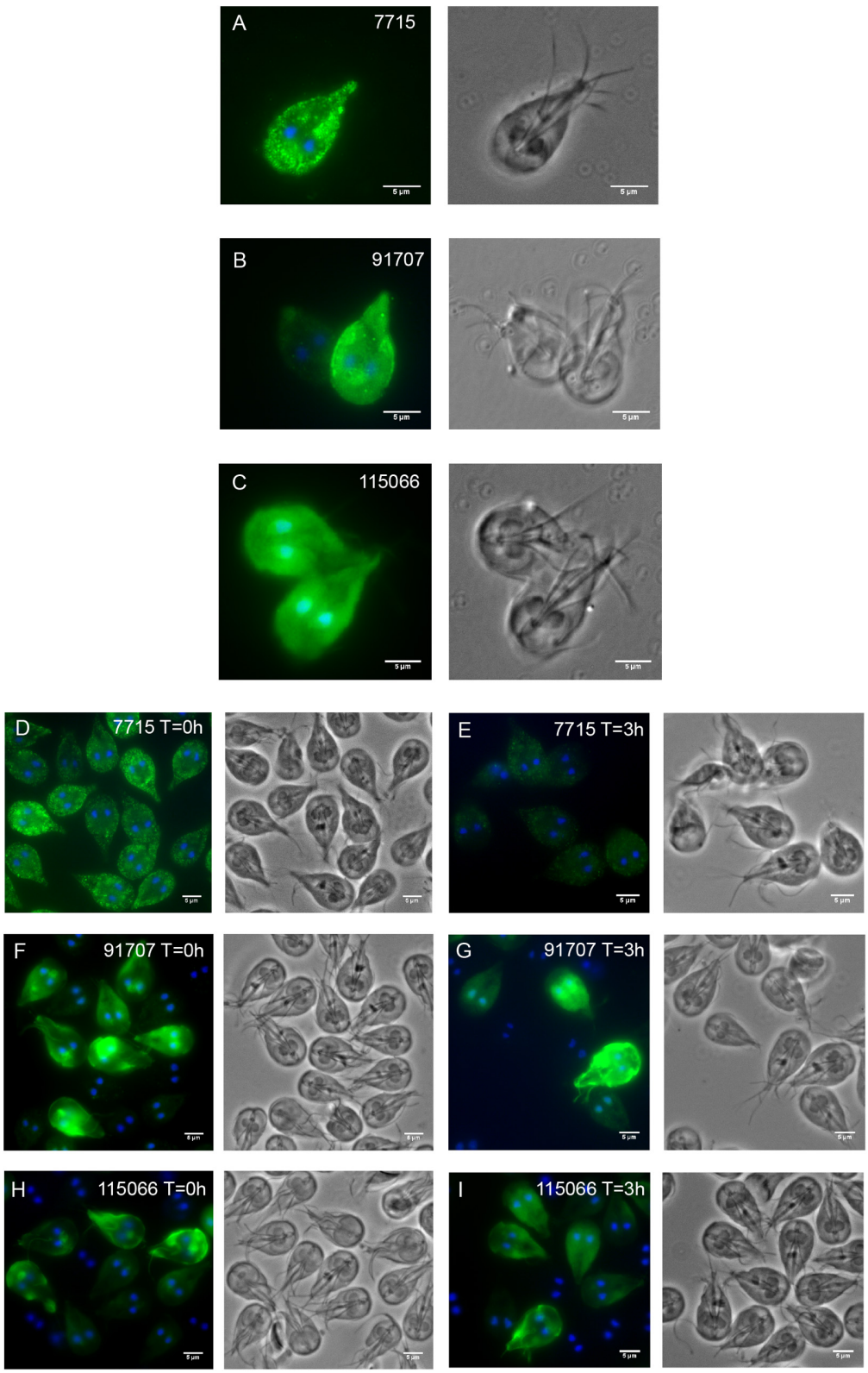


FIGURE 2 | Fluorescent microscopy analyses of three C-terminally HA-tagged, episomally expressed HCMPs. Panels **(A–C)** (**A**-7715, **B**-91707, and **C**-115066) show *Giardia* trophozoites after axenic growth to 80% confluence in TYDK. The 3 transfectants were also co-incubated with human IECs **(D–I)** and epitope tagged HCMP localization was studied at the start of interaction (T = 0 h, **D,F,H**) and after 3 h interaction with differentiated Caco-2 cells (T = 3 h, **E,G,I**).

be detected at the plasma membrane and within the cytoplasm (**Figures 2G,I**).

Regulation of HCMP Genes During *Giardia*-Host Cell Interactions

Histone deacetylases (HDACs) regulate the expression of VSPs (Carranza et al., 2016), whose gene organization and structure are similar to that of HCMPs (Xu et al., 2020). Earlier studies of the effect of an inhibitor specific for NAD^+-independent HDACs (FR235222) on *Giardia* showed that it affected the expression of both VSPs and HCMPs (Sonda et al., 2010). We, therefore, decided to test whether the HCMPs up-regulated upon IEC contact can be epigenetically regulated using three different HDAC inhibitors that have been shown to interfere with NAD^+-dependent (Nicotinamide, Nt) and NAD^+-independent [Trichostatin A (TSA) and sodium Butyrate (NaB)] HDACs. Trophozoites were treated with TSA, NaB, or Nt at the same concentrations (TSA 200 nM, NaB 20 nM, and Nt 10 mM) and time (1 day) as used in earlier studies of VSP expression (Carranza et al., 2016) but with an extension to 3 days to get 80% confluent tubes and see long-term effects. Changes in the RNA levels of 5 differentially expressed HCMPs during interaction with differentiated Caco-2 cells (7715, 9620, 11309, 91707, and 115066) were studied using qPCR (**Figure 3A**). Neither NaB nor Nt had any effects on the RNA levels of the selected HCMPs. Nevertheless, trophozoites treated with TSA showed a significant increase in the RNA levels ranging between 4- and 16-fold for the HCMPs 7715, 9620, 91707, and 115066. Therefore, our results indicate that some of the differentially expressed HCMPS during *Giardia*-IEC interactions are regulated by NAD^+-independent HDACs like VSPs.

Stress with metronidazole, hydrogen peroxide or growth at 39°C affect expression of certain HCMPs (Ansell et al., 2016).

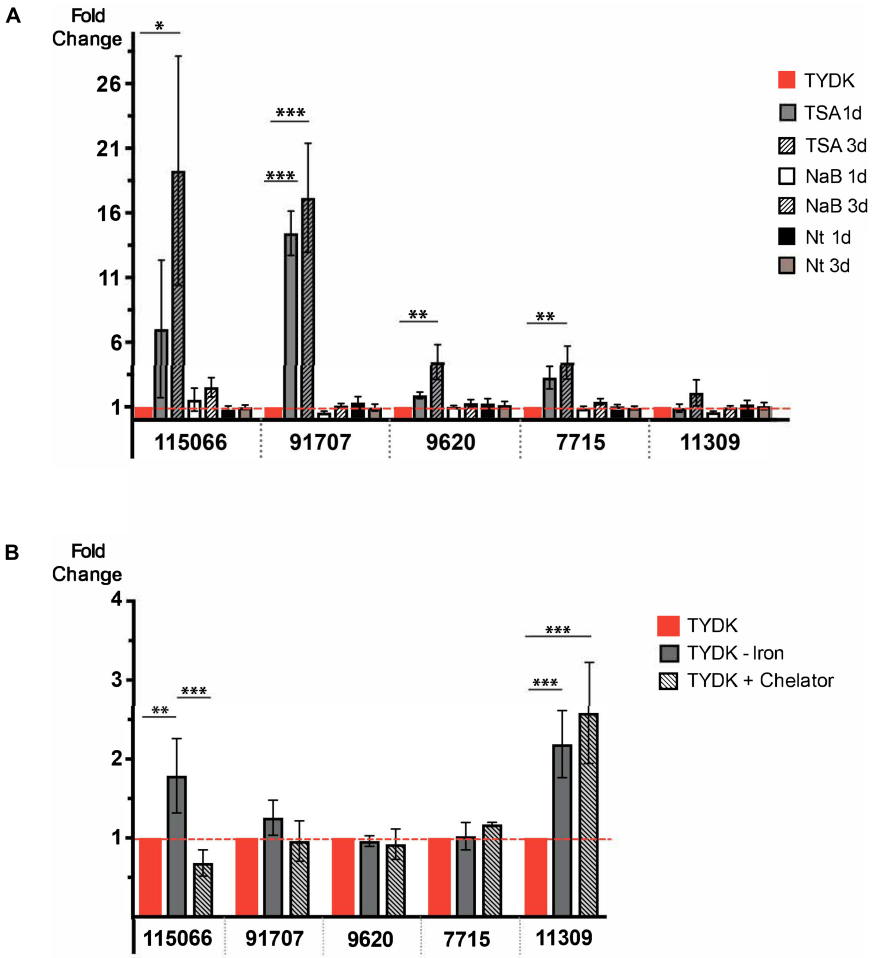


FIGURE 3 | (A) Gene expression analyses of five HCMP genes (HCMP 9620, 7715, 11309, 91707, and 115066) during growth of trophozoites (10^4 starting trophozoites) for 1 and 3 days in the presence of three HDAC inhibitors [Trichostatin A (TSA), Sodium butyrate (NaB) and Nicotinamide (Nt)]. The experiment was repeated four times. **(B)** Expression of HCMPs 9620, 7715, 11309, 91707, and 115066 in medium containing different levels of iron; TYDK medium with extra iron (TYDK+ Iron), TYDK medium without extra iron (TYDK– Iron) and TYDK without extra iron and addition of 50 μM of the metal ion chelator 2,2′-Bipyridyl (TYDK-Iron + Chelator). The experiment was repeated three times. Levels of expression were measured by qPCR in both panels. Significant results are indicated by $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

It is possible that oxidative stress in the DMEM medium, due to a much lower level of cysteine compared to in TYDK, is inducing the expression of the HCMPs. We tested this by growing our episomal transfectants for 3 h in different media and in the absence or presence of Caco-2 cells and the HCMP transcript levels were analyzed with qPCR (**Supplementary Figure S5**). It is clear that the change of medium from TYDK to DMEM has a large effect on HCMP expression but inclusion of 1 mM cysteine in the DMEM medium rather increased the HCMP expression (twofold, **Supplementary Figure S5**). The presence of IECs significantly increased the expression of the episomal HCMPs but inclusion of 1 mM cysteine reduced the level of up-regulation (**Supplementary Figure S5**). Thus, oxidative stress can affect the expression of the HCMPs but this is not the only factor leading to up-regulation during trophozoite-Caco-2 interactions.

Oxidative stress in cells is known to be associated with the cellular levels of free iron ions and DMEM contain much less free iron (0.1 mg/l) compared to TYDK (5.5 mg/l). We therefore tested whether HCMPs can be regulated by iron. To do this, we grew *Giardia* trophozoites in complete TYDK medium and in TYDK media without addition of extra iron (1.6 mg/l) with or without the addition of 50 μM of the metal ion chelator 2,2′-Bipyridyl (see section "Materials and Methods"). RNA was extracted and qPCR was used to study changes in RNA levels of the three HCMP genes. No significant differences in the RNA levels were seen for the HCMPs 91707, 9620, and 7715 in response to growth in medium with threefold less iron (**Figure 3B**). However, the lower levels of iron significantly increased the RNA levels of HCMP 115066 (1.7-fold) and HCMP 11309 (2.3-fold, **Figure 3B**). A further increase in the RNA levels of HCMP 11309 was seen when iron was chelated in the medium, whereas this reduced expression of HCMP115066 (**Figure 3B**). These findings suggest that iron might play a role in the regulation of HCMPs.

Giardia Responses to Iron Depletion

The observed effects of iron on the expression of a small set of HCMPs and the fact the iron is a well-known regulator of virulence in intestinal bacteria (Begg, 2019) and protozoa (Gastelum-Martínez et al., 2018) inspired us to study the effects of iron on the total *Giardia* transcriptome. RNA Sequencing was performed on trophozoites grown in media with different levels of free iron to assess global changes in gene expression. We used a recently established clone of the WB-C6 isolate (see section "Materials and Methods") in order to be able to better study the expression of variable gene families like HCMPs and VSPs. Overall, reducing the iron in the *Giardia* growth medium threefold (TYDK-Fe) resulted in 225 DEGs; 205 were up-regulated and 20 were down-regulated (**Supplementary Table S6**). Six genes were up-regulated twofold or more, including the HCMPs 24880 and 11309 and the iron containing enzyme pyruvate-flavodoxin oxidoreductase (PFOR, ORF17063). Among the most up-regulated genes we noted several other HCMPs (ORFs 6372, 103454, 113987, and 115066), another PFOR (ORF 114609) and several putative cation transporters (ORFs 32658, 92246, and 96670). We also detected higher expression of the secreted, putative giardial virulence factors cathepsin B 14019 and 16779 (Liu et al., 2018, 2019) and two tenascin-like proteins (ORFs 95162 and 114815) (Dubourg et al., 2018).

Chelation of free iron in the *Giardia* medium using 50 μM of 2,2′-Bipyridyl (TYDK-Fe + Chelator) resulted in 613 DEGs with 185 being up-regulated and 428 down-regulated (**Supplementary Table S7**). Four of the genes were up-regulated twofold or more, including HCMPs 11309 and 25816. Among the up-regulated genes we also noticed several encystation-specific genes (Cyst-wall proteins -1 and -2 and glucosamine-6-phosphate), oxidative stress related genes (Peroxoredoxin-1, superoxide reductase, Ferredoxin, Thioredoxin reductase and NADPH oxidoreductase) and histones (H2A, H3, and H4). Only two VSPs were found among the DEGs (ORFs 33279 and 111933) in TYDK-Fe whereas 15 other VSP genes were up-regulated in the presence of the chelator (**Supplementary Table S7**). Totally 15 HCMPs were up-regulated when iron was not added to the TYDK medium, whereas 10 of the HCMPs were differentially expressed (5 up- and 5 down-regulated) in response to chelation of metal ions (**Table 3**). Only 11 DEGs (9 up- and 2 down-regulated) were common between the TYDK-Fe and TYDK-Fe + chelator conditions and the HCMPs 11309, 16716, 25816, and 115066 were found to be DEGs in both treatments (**Table 3**). The level of chelator used in the experiments was set at the highest concentration that did not affect the growth rate of *Giardia* WB trophozoites but 2,2′-Bipyridyl also chelate other metal ions, giving broader effects than the reduction of iron in the medium. This can explain the differences between the data sets and the large number of down-regulated genes in the presence of the chelator. The effect of metal ion chelation produced an enrichment of general GO terms such as nucleic acid binding, RNA binding, transporter activity or structural constituent of the ribosome and therefore we analyzed the DEGs further using the complete biological and molecular functions. This, in turn, enriched for many biological and molecular GO terms such as translation, gene expression, ribosome biogenesis and others involving the metabolism of nucleosides, pyrimidine nucleosides, glycosyl- and organonitrogen compounds, macromolecules synthesis and amino acid activation. Cellular oxidant detoxification (GO: 0098869) also emerged within the enriched group, indicating a relation between iron and oxidative stress.

DISCUSSION

Several mechanisms have been proposed to be important for induction of symptoms during a *Giardia* infection but there is yet no consensus. However, data generated during the last years have shed light on the disease mechanism, which seems to be multi-factorial (Einarsson et al., 2016a). Most of this data has been generated using different *in vitro* models of giardiasis using human or rat intestinal epithelial cell-lines combined with trophozoites or parasite extracts and it has been complemented by experimental infections in mice, gerbils and data from human giardiasis patients (Emery-Corbin et al., 2020; Jex et al., 2020). Various strategies have been developed to study the interplay

TABLE 3 | Differentially expressed high cysteine membrane proteins (HCMPs) in *Giardia intestinalis* WB isolate trophozoites in response to different levels of iron.

Gene ID	Description	TYDK-Fe	TYDK-Fe + Chelator
GL50803_24880	High cysteine membrane protein Group 2	2.4	NA
GL50803_103454	High cysteine membrane protein Group 1	1.8	NA
GL50803_113987	High cysteine membrane protein Group 3	1.8	NA
GL50803_6372	High cysteine membrane protein	1.7	NA
GL50803_16936	High cysteine membrane protein EGF-like	1.7	NA
GL50803_15317	High cysteine membrane protein Group 1	1.6	NA
GL50803_112135	High cysteine membrane protein VSP-like	1.6	NA
GL50803_114991	High cysteine membrane protein EGF-like	1.6	NA
GL50803_103943	High cysteine membrane protein Group 3	1.5	NA
GL50803_27717	High cysteine membrane protein Group 3	1.4	NA
GL50803_21321	High cysteine membrane protein Group 5	1.3	NA
GL50803_11309	High cysteine membrane protein Group 1	2.3	2.9
GL50803_25816	High cysteine membrane protein Group 1	1.7	2.8
GL50803_115066	High cysteine membrane protein VSP-like	1.7	0.7
GL50803_16716	High cysteine membrane protein Group 5	1.3	0.8
GL50803_114617	High cysteine membrane protein Group 4	NA	1.2
GL50803_101589	High cysteine membrane protein	NA	1.2
GL50803_112432	High cysteine membrane protein Group 5	NA	1.2
GL50803_104087	High cysteine membrane protein	NA	0.8
GL50803_91707	High cysteine membrane protein Group 1	NA	0.7
GL50803_14017	High cysteine membrane protein Group 2	NA	0.6

and the outcome of interaction between *Giardia* and host cell *in vitro* and *in vivo* (Roxström-Lindquist et al., 2005; Ringqvist et al., 2011; Ferella et al., 2014; Kraft et al., 2017; Ma'ayeh et al., 2017). Here, we have used our well-characterized *in vitro* model of interaction between differentiated Caco-2 cells and *G. intestinalis* WB trophozoites (Roxström-Lindquist et al., 2005) combined with RNA sequencing to study transcriptional changes in trophozoites during the early hours of co-incubation with the human cells. We reasoned that early hours on interaction are important for establishing infection, counteracting intestinal epithelial cell defenses and the expression of virulence factors. One problem with studies of changed gene expression in *Giardia* trophozoites during interactions with intestinal epithelial cells *in vitro* are the control parasites; *Giardia* trophozoites do not do well in the growth mediums adapted for the intestinal epithelial cells. In this model, we pre-incubated the trophozoites in DMEM for 2 h prior to the addition to the Caco-2 cells to reduce the background effects on gene expression due to the change in medium. The levels of up- and down-regulation of gene expression is smaller in this study compared to what was earlier observed (Ringqvist et al., 2011), suggesting that a big part of the transcriptional changes in earlier studies can be attributed to the change of growth medium, but still 4-times more DEGs were identified in this study compared to the earlier studies (Emery-Corbin et al., 2020). Nevertheless, there are common transcriptional changes detected in all studies, even if different intestinal cell lines and methods for RNA analyses were used (Ringqvist et al., 2011; Ferella et al., 2014; Ma'ayeh et al., 2017). This makes it possible to summarize differential gene expression changes in *Giardia* trophozoites during host cell interactions and make some general comments. First, the largest changes in gene expression are induced by medium changes and not direct interaction with IECs. Second, the level of up- and down-regulation of RNA expression in *Giardia* is relatively small compared to other organisms, suggesting either a very tight level of regulation or post-transcriptional regulation. Third, up-regulation of metabolic, encystation and oxidative stress related genes is seen in all studies and this is most likely due to the stress induced during the interaction experiments. However, this type of stressors, oxidative and nutrient stress and induction of encystation, is also seen *in vivo* when mice are infected by *Giardia* WB trophozoites (Pham et al., 2017).

All earlier transcriptomal analyses of *Giardia*-host cell interactions have suggested an important role for proteases during *Giardia*-host interactions (Emery-Corbin et al., 2020). In this study, up-regulation was seen of a large number of cysteine proteases (CPs, encoded by ORFs 3169, 10217, 11209, 15564, 16160, 16779, 17516, 29304, 112831, 113303, 113656, 114165, 114773, 114915, and 137680). Most of these CPs have been shown to be up-regulated during host–parasite interactions *in vivo* (Pham et al., 2017) and to be more expressed in parasites recently axenized from human patients (4 generations) compared to parasites that have been grown for 50 generations *in vitro* (Ankarklev et al., 2015). Several recent studies have highlighted the importance of *Giardia* CPs in infections, specifically, the degradation of chemokines, antibodies, antimicrobial peptides, tight junction proteins and in encystation (Williams and Coombs, 1995; Jiménez et al., 2000; Coradi and Guimarães, 2006; Rodríguez-Fuentes et al., 2006; Carvalho et al., 2008; Ringqvist et al., 2011; Ma'ayeh and Brook-Carter, 2012; Ferella et al., 2014; Bhargava et al., 2015; Emery et al., 2016; Dubourg et al., 2018; Liu et al., 2018). It will be important to further characterize the

CPs encoded by ORFs 10217, 16160, 16779, 17516, and 137680 in order to understand the virulence of *Giardia*.

The production of reactive oxygen species (ROS) represents one of the main epithelial cell defenses to fight off trophozoites early during infection (Ma'ayeh et al., 2015). Nevertheless, the ability of microaerophilic trophozoites to counteract host oxidative defenses has been investigated previously *in vitro*, highlighting the roles of the thioredoxin system, NADH oxidases, peroxiredoxins oxidoreductases, nitroreductases and nicotinamide co-factors in neutralizing host ROS (Ma'ayeh et al., 2015; Mastronicola et al., 2016; Pham et al., 2017). In fact, the genes encoding the above enzymes were differentially expressed at 1.5 h and thereafter (**Supplementary Tables S2–S4**), corroborating the above findings. Furthermore, the anti-oxidative stress responses have been also reported in trophozoites during mice infections, suggesting that *in vitro* results also can be seen *in vivo* (Pham et al., 2017).

Genes encoding hypothetical proteins have been found to be the main group of up-regulated genes in all gene expression studies of *Giardia*-host cell interactions (Ringqvist et al., 2011; Ma'ayeh and Brook-Carter, 2012; Pham et al., 2017) and also in this study hypothetical proteins dominated the DEGs. However, the hypothetical proteins are often *Giardia*-specific and have unknown function, which makes it difficult to make any conclusions about their function. A recent study modeled the structure of proteins encoded by 5,000 ORFs, including most hypothetical proteins, in *Giardia* using I-TASSER (Ansell et al., 2019). This changed the annotation of 212 hypothetical proteins and the new annotations have been used in this study. The large membrane protein encoded by ORF 114210 is glycosylated (Ratner et al., 2008) and the protein structure modeling using I-TASSER shows that it is similar to complement factor H (Ansell et al., 2019). Complement factor H is a large, soluble glycoprotein found in serum that suppresses reactions of the alternative pathway of the complement system to host cells (Parente et al., 2017). The complement system has been shown to be important in the immune defense against of *Giardia* in mice (Li et al., 2016; Coelho and Singer, 2018) and it is possible that this protein is used by the parasite to prevent complement-mediated immune mechanisms. Another gene, ORF5800, that earlier was annotated to encode a hypothetical protein, has been shown to be similar to lipid transporters in the StART family (Ansell et al., 2019). *Giardia* cannot produce most of its lipids and instead needs to scavenge them from the host (Jarroll et al., 1981; Mendez et al., 2015). Expression of genes involved in lipid metabolism in *Giardia* have been shown to be up-regulated when WB parasites infect mice (Pham et al., 2017) and this is also true for ORF5800. The role of the protein encoded by ORF5800 during *Giardia* infections will be interesting to study in more detail.

The *Giardia* gene family that shows the largest levels of differential gene expression during *Giardia*-host interactions *in vitro* and *in vivo* are the High-Cysteine Membrane Proteins (HCMPs). It is a yet uncharacterized family of proteins with high resemblance to the cysteine-rich Variant Surface Proteins (VSPs) (Davids et al., 2006). HCMPs often co-localize with VSPs in variable parts of the genome (Xu et al., 2020) and there are isolate-specific differences in the HCMP repertoires (Franzén et al., 2009; Ankarklev et al., 2015). Structurally, the VSPs have one conserved transmembrane domain, multiple extracellular CXXC motifs and a short cytoplasmic tail with the sequence CRGKA (Ankarklev et al., 2010). HCMPs also have multiple, extracellular CXXC motifs but often also multiple CXC motifs and the cytoplasmic domain is variable in size (Davids et al., 2006). Originally 61 HCMP genes were identified in the *Giardia* WB genome (Morrison et al., 2007) but in our new, more complete assembly of the WB genome we identified 116 HCMP genes (Xu et al., 2020). VSPs are involved in antigenic variation in order for the parasite to avoid being eliminated by the adaptive immune system and one main VSP is expressed on the cell surface of each trophozoite (Prucca and Lujan, 2009). The VSPs have been shown to be regulated epigenetically (Kulakova et al., 2006; Carranza et al., 2016) and post-transcriptionally by small RNAs (Prucca et al., 2008; Saraiya et al., 2014) but the understanding of VSP regulation is still limited. The HCMPs were first studied in the association with encystation (Davids et al., 2006) but we know now that they are also induced by other types of stressors (e.g., type of medium, temperature, drugs or oxidative stress) (Ansell et al., 2016; Einarsson et al., 2016b). One major limitation of earlier studies of HCMP gene expression has been the poor genome annotation but also limitations in the microarray, subtractive hybridization and serial-analyses of gene expression (SAGE) in differentiating HCMPs. Here we used the newly assembled WB genome (Xu et al., 2020), which is more complete and better annotated, in combination with RNAseq to show that 70 of the 116 HCMP genes in the WB isolate are differentially expressed during co-incubation with differentiated Caco-2 cells (**Supplementary Table S5**). In fact, certain HCMPs have earlier been shown to be up-regulated during interaction with other intestinal cell lines (HT29 and IEC-6) (Ferella et al., 2014; Ma'ayeh et al., 2017) and to be more highly expressed in the mouse intestine compared to growth *in vitro* in the standard medium TYDK (Pham et al., 2017) but this is the most complete study so far of HCMP gene expression. A summary of the data generated here together with earlier published data show that the HCMPs have different expression patterns in different conditions but we did not see any correlation between protein sequence similarity and expression profile (**Supplementary Table S8**). The expression patterns of the VSPs compared to the HCMPs suggest that the two groups display different types of regulation, even if there is some overlap between the families (**Supplementary Tables S2–S4**). Only 27 of the 341 VSP genes in the *Giardia* WB genome (Xu et al., 2020) are up-regulated during co-incubation with differentiated Caco-2 cells and the level of up-regulation is lower than in the HCMP group (no VSPs among the 200 most up-regulated genes during the interaction compared to 13 HCMPs). Further studies are needed in order to define the relationship between VSPs and HCMPs and to see if there are sub-classes within each group with different expression profiles.

We used epitope-tagging to localize a few of the HCMPs which were up-regulated during interaction with IECs (e.g., ORFs 7715, 91707, and 115066). These HCMPs were plasma membrane-associated but they were also partially localized around the peripheral vesicles, ER and nuclear membrane. The localization

HCMP91707 varied with generation numbers in culture, which suggests that certain HCMPs can display a dynamic localization between internal and external membranes. Some HCMPs are also released into the surrounding environment during host cell interactions (Ma'ayeh et al., 2017). The presence of epidermal growth factor (EGF)-like domains within many of the HCMPs might relate to possible functions associated with this dynamic localization inside and outside the giardial cell. EGF-like domains are usually present in secreted proteins but they can function intracellularly upon ligand binding to mediate the release of transcriptional factors or modulate transcription within the nucleus (e.g., NOTCH signaling) (Kiyota and Kinoshita, 2004; Lai, 2004). This domain is also associated with diverse functions including calcium binding (Handford et al., 1991), adhesion (Stenflo, 1991), protein–protein interactions and intracellular and extracellular signaling (Wouters et al., 2005). It will be interesting to identify which exact role HCMPs and EGF-like domain have in *Giardia*.

This and earlier studies (Sonda et al., 2010) suggest a putative role of NAD^+-independent HDACs in the regulation of certain HCMPs, implying that HCMPs are regulated at chromatin structure level, similar to VSPs. HCMPs are structurally similar to VSP and they have a proximal chromosomal location to VSPs throughout the genome (Adam, 2000; Morrison et al., 2007). Interestingly, a unique NAD^+-independent histone deacetylase has been found to modulate VSP switching (Carranza et al., 2016). Blocking this HDAC with a specific drug induced the expression of encystation-specific genes, VSPs and certain HCMPs (Sonda et al., 2010). Epigenetic regulation is a very complicated process and in-depth analysis of histone modifications, DNA methylation and chromatin-modifying enzymes is warranted to draw a comprehensive conclusion on HCMP regulation during different stages of the parasite life-cycle and during infection.

Iron is an essential trace-metal for all organisms and it is a well-known regulator of virulence genes in many pathogenic bacteria (Begg, 2019) but also in pathogenic protozoa, including *Entamoeba histolytica*, *Trichomonas vaginalis*, *Trypanosoma*, and *Leishmania* (Niu et al., 2016; Gastelum-Martínez et al., 2018; Paiva et al., 2018). This is because iron is an essential constituent of many proteins in the pathogens, including metabolic and antioxidant enzymes and virulence factors (Arroyo et al., 2015). In the duodenum, were ferric iron (Fe^{2+}) absorption occurs in humans, the level of accessible iron is limited for invading pathogens, resulting in a competition for free iron with the host cells, and iron has been shown to be a major regulator of virulence genes in intestinal bacteria (Begg, 2019). Here, we studied the effect of reducing the iron concentration in the growth medium or its fixation by a chelation agent on *Giardia* trophozoites. The level of total iron in the standard *Giardia* growth medium TYDK is very high (5 mg/ml) but the levels of iron available for *Giardia in vivo* is much lower (Deschemin et al., 2016). Thus, it will be important to determine what *Giardia* genes that are differentially expressed at lower levels of free-iron compared to in the standard TYDK medium. In the first data set we did not add extra iron to the *Giardia* growth medium and this reduced the total levels of iron threefold. This resulted in 225 DEGs, including up-regulation of several genes that earlier have been suggested to be virulence genes (**Supplementary Table S6**). The level of up- and down-regulation of RNA expression in *Giardia* is relatively small compared to other organisms as noted above and this can also be seen in the analysis of iron-regulated genes. We detected 613 DEGs in the chelator treated parasites compared to cells grown in TYDK but the effect was mainly a down-regulation of a broad repertoire of genes. The addition of a metal ion chelator in to the medium should reduce the free iron but it is also binding other metal ions, increasing the risk of non-iron related responses. This calls for alternative methods for specific reduction of iron to lower levels than in the standard TYDK medium. We noticed the induction of many HCMPs and VSPs in both treatments, indicating that HCMP and VSP genes transcription can also be regulated by iron. It will be interesting to see whether there are iron response elements within the sequences of these induced genes similar to that in *Entamoeba* (Soto-Castro et al., 2017). Furthermore, the withdrawal of iron resulted in the enrichment of functional groups relating to iron functions, such as the oxidoreductase activity, ATPase activity, nucleoside triphosphatase activity, and pyrophosphatase activity and nucleoside metabolism. It was interesting to see the pyrophosphatase activity group enriched since the use of pyrophosphate in *Giardia* produces more ATP in energy limiting conditions (Ma'ayeh and Brook-Carter, 2012). This overall indicate a compensatory response for energy, antioxidant functions and the acquisition of nucleosides to counteract iron withdrawal. The role of iron during *Giardia* infections is not well-characterized but our expression datasets can be start for further studies of this important factor.

To conclude, gene expression analyses of *Giardia* trophozoites during interaction with IECs *in vitro* have identified a group of genes that are up-regulated during interactions. This group contains genes encoding cysteine proteinases, oxidative stress response proteins, metabolic proteins and certain HCMP genes. We have also identified genes that are regulated by the level of iron in the growth medium. The recent development of CRISPR repression and mutagenesis systems (Lin et al., 2019; McInally et al., 2019) will make it possible to define the role of these genes in *Giardia's* virulence. Mutant parasites can be tested *in vitro* in enteroid systems or *in vivo* in mice or gerbils and this can lead to a more detailed molecular understanding of giardiasis.

DATA AVAILABILITY STATEMENT

The data sets generated here can be found at Gene Expression Omnibus (GEO); accession ID GSE144004 for the interaction experiment and GSE136820 for the iron depletion experiment.

AUTHOR CONTRIBUTIONS

DP, SM, FX, MF, SC, and JL performed the experiments, analyzed the data, and wrote parts of the manuscript.

SS conceived and designed the experiments and wrote the first draft of the manuscript. All authors revised the manuscript.

FUNDING

This study was supported by a grant from Vetenskapsrådet-M (2017-02918) to SS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

The sequencing facility at SciLifeLab, Uppsala University is acknowledged for support with the RNASeq data generation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00913/full#supplementary-material

REFERENCES

Adam, R. D. (2000). The *Giardia lamblia* genome. *Int. J. Parasitol.* 30, 475–484. doi: 10.1016/s0020-7519(99)00191-5

Ankarklev, J., Franzén, O., Peirasmaki, D., Jerlström-Hultqvist, J., Lebbad, M., Andersson, J., et al. (2015). Comparative genomic analyses of freshly isolated *Giardia intestinalis* assemblage A isolates. *BMC Genomics* 16:697. doi: 10.1186/s12864-015-1893-6

Ankarklev, J., Jerlström-Hultqvist, J., Ringqvist, E., Troell, K., and Svärd, S. G. (2010). Behind the smile: cell biology and disease mechanisms of *Giardia* species. *Nat. Rev. Microbiol.* 8, 413–422. doi: 10.1038/nrmicro2317

Ansell, B. R. E., Mcconville, M. J., Baker, L., Korhonen, P. K., Emery, S. J., Svärd, S. G., et al. (2016). Divergent transcriptional responses to physiological and *Xenobiotic* stress in *Giardia duodenalis*. *Antimicrob. Agents Chemother.* 60, 6034–6045. doi: 10.1128/aac.00977-16

Ansell, B. R. E., Pope, B. J., Georgeson, P., Emery-Corbin, S. J., and Jex, A. R. (2019). Annotation of the Giardia proteome through structure-based homology and machine learning. *Gigascience* 8:giy150.

Arroyo, R., Ochoa, T., Tai, J.-H., and De La Garza, M. (2015). Iron and parasites. *Biomed. Res. Int.* 2015:291672. doi: 10.1155/2015/291672

Begg, S. L. (2019). The role of metal ions in the virulence and viability of bacterial pathogens. *Biochem. Soc. Trans.* 47, 77–87. doi: 10.1042/bst20180275

Bénéré, E., Van Assche, T., Van Ginneken, C., Peulen, O., Cos, P., and Maes, L. (2012). Intestinal growth and pathology of *Giardia duodenalis* assemblage subtype AI, AII, B and E in the gerbil model. *Parasitology* 139, 424–433. doi: 10.1017/s0031182011002137

Bhargava, A., Cotton, J. A., Dixon, B. R., Gedamu, L., Yates, R. M., and Buret, A. G. (2015). *Giardia duodenalis* surface cysteine proteases induce cleavage of the intestinal epithelial cytoskeletal protein villin via myosin light chain kinase. *PLoS One* 10:e0136102. doi: 10.1371/journal.pone.0136102

Buret, A. G. (2007). Mechanisms of epithelial dysfunction in giardiasis. *Gut* 56, 316–317. doi: 10.1136/gut.2006.107771

Buret, A. G. (2008). Pathophysiology of enteric infections with *Giardia duodenalius*. *Parasite* 15, 261–265. doi: 10.1051/parasite/2008153261

Cacciò, S. M., and Sprong, H. (2011). "Epidemiology of giardiasis in humans," in *Giardia: A Model Organism*, eds H. D. Luján, and S. Svärd, (Vienna: Springer Vienna).

Carranza, P. G., Gargantini, P. R., Prucca, C. G., Torri, A., Saura, A., Svärd, S., et al. (2016). Specific histone modifications play critical roles in the control of encystation and antigenic variation in the early-branching eukaryote *Giardia lamblia*. *Intern. J. Biochem. Cell Biol.* 81, 32–43. doi: 10.1016/j.biocel.2016.10.010

Carvalho, T. B. D., David, É. B., Coradi, S. T., and Guimarães, S. (2008). Protease activity in extracellular products secreted in vitro by trophozoites of *Giardia duodenalis*. *Parasitol. Res.* 104:185. doi: 10.1007/s00436-008-1185-z

Céu Sousa, M., Gonçalves, C. A., Bairos, V. A., and Poiares-da-Silva, J. (2001). Adherence of *Giardia lamblia* trophozoites to int-407 human intestinal cells. *Clin. Diagn. Lab. Immunol.* 8, 258–265. doi: 10.1128/CDLI.8.2.258-265.2001

Coelho, C. H., and Singer, S. M. (2018). Recent advances in the *Giardia*-host relationship reveal danger lurking behind the smile. *PLoS Negl. Trop. Dis.* 12:e0006625. doi: 10.1371/journal.pone.000006625

Coradi, S. T., and Guimarães, S. (2006). *Giardia duodenalis*: protein substrates degradation by trophozoite proteases. *Parasitol. Res.* 99, 131–136. doi: 10.1007/s00436-005-0124-5

Cotton, J. A., Beatty, J. K., and Buret, A. G. (2011). Host parasite interactions and pathophysiology in *Giardia* infections. *Intern. J. Parasitol.* 41, 925–933. doi: 10.1016/j.ijpara.2011.05.002

Davids, B. J., Reiner, D. S., Birkeland, S. R., Preheim, S. P., Cipriano, M. J., Mcarthur, A. G., et al. (2006). A new family of giardial cysteine-rich non-Vsp protein genes and a novel cyst protein. *PLoS One* 1:e44. doi: 10.1371/journal.pone.0000044

Deschemin, J. C., Noordine, M. L., Remot, A., Willemetz, A., Afif, C., Canonne-Hergaux, F., et al. (2016). The microbiota shifts the iron sensing of intestinal cells. *FASEB J.* 30, 252–261. doi: 10.1096/fj.15-276840

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Dubourg, A., Xia, D., Winpenny, J. P., Al Naimi, S., Bouzid, M., Sexton, D. W., et al. (2018). Giardia secretome highlights secreted tenascins as a key component of pathogenesis. *Gigascience* 7, 1–13.

Einarsson, E., Ma'ayeh, S., and Svärd, S. G. (2016a). An up-date on Giardia and giardiasis. *Curr. Opin. Microbiol.* 34, 47–52. doi: 10.1016/j.mib.2016.07.019

Einarsson, E., Troell, K., Hoeppner, M. P., Grabherr, M., Ribacke, U., and Svärd, S. G. (2016b). Coordinated changes in gene expression throughout encystation of *Giardia intestinalis*. *PLoS Negl. Trop. Dis.* 10:e0004571. doi: 10.1371/journal.pone.004571

Emery, S. J., Lacey, E., and Haynes, P. A. (2016). Quantitative proteomics in *Giardia duodenalis*—Achievements and challenges. *Mol. Biochem. Parasit.* 208, 96–112. doi: 10.1016/j.molbiopara.2016.07.002

Emery-Corbin, S. J., Gruttner, J., and Svard, S. (2020). Transcriptomic and proteomic analyses of *Giardia intestinalis*: intestinal epithelial cell interactions. *Adv. Parasitol.* 107, 139–171. doi: 10.1016/bs.apar.2019.11.002

Farthing, M. J. (1997). The molecular pathogenesis of giardiasis. *J. Pediatr. Gastroenterol. Nutr.* 24, 79–88. doi: 10.1097/00005176-199701000-00018

Ferella, M., Davids, B. J., Cipriano, M. J., Birkeland, S. R., Palm, D., Gillin, F. D., et al. (2014). Gene expression changes during *Giardia*Ű host cell interactions in serum-free medium. *Mol. Biochem. Parasitol.* 197, 21–23. doi: 10.1016/j.molbiopara.2014.09.007

Franzén, O., Jerlström-Hultqvist, J., Castro, E., Sherwood, E., Ankarklev, J., Reiner, D. S., et al. (2009). Draft genome sequencing of *Giardia intestinalis* assemblage b isolate gs: is human giardiasis caused by two different species? *PLoS Pathog.* 5:e1000560. doi: 10.1371/journal.pone.1000560

Gascón, J. (2006). Epidemiology, etiology and pathophysiology of traveler's diarrhea. *Digestion* 73(Suppl. 1), 102–108. doi: 10.1159/000089785

Gastelum-Martínez, A., León-Sicairos, C., Plata-Guzmán, L., Soto-Castro, L., León-Sicairos, N., and De La Garza, M. (2018). Iron-modulated virulence factors of *Entamoeba histolytica*. *Future Microbiol.* 13, 1329–1341. doi: 10.2217/fmb-2018-0066

Gourguechon, S., and Cande, W. Z. (2011). Rapid tagging and integration of genes in *Giardia intestinalis*. *Eukaryot. Cell* 10, 142–145. doi: 10.1128/ec.00190-10

Handford, P. A., Mayhew, M., Baron, M., Winship, P. R., Campbell, I. D., and Brownlee, G. G. (1991). Key residues involved in calcium-binding motifs in EGF-like domains. *Nature* 351, 164–167. doi: 10.1038/351164a0

Humen, M. A., Perez, P. F., and Lievin-Le Moal, V. (2011). Lipid raft-dependent adhesion of *Giardia intestinalis* trophozoites to a cultured human enterocyte-like Caco-2/TC7 cell monolayer leads to cytoskeleton-dependent functional injuries. *Cell Microbiol.* 13, 1683–1702. doi: 10.1111/j.1462-5822.2011.01647.x

Jarroll, E. L., Muller, P. J., Meyer, E. A., and Morse, S. A. (1981). Lipid and carbohydrate metabolism of *Giardia lamblia*. *Mol. Biochem. Parasitol.* 2, 187–196. doi: 10.1016/0166-6851(81)90099-2

Jerlström-Hultqvist, J., Stadelmann, B., Birkestedt, S., Hellman, U., and Svärd, S. G. (2012). Plasmid vectors for proteomic analyses in giardia: purification of virulence factors and analysis of the proteasome. *Eukaryot. Cell* 11, 864–873. doi: 10.1128/ec.00092-12

Jex, A. R., Svard, S., Hagen, K. D., Starcevich, H., Emery-Corbin, S. J., Balan, B., et al. (2020). Recent advances in functional research in *Giardia intestinalis*. *Adv. Parasitol.* 107, 97–137. doi: 10.1016/bs.apar.2019.12.002

Jiménez, J. C., Uzcanga, G., Zambrano, A., Di Prisco, M. C., and Lynch, N. R. (2000). Identification and partial characterization of excretory/secretory products with proteolytic activity in *Giardia intestinalis*. *J. Parasitol.* 86, 859–862.

Keister, D. B. (1983). Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Trans. R. Soc. Trop. Med. Hygiene* 77, 487–488.

Kiyota, T., and Kinoshita, T. (2004). The intracellular domain of X-Serrate-1 is cleaved and suppresses primary neurogenesis in *Xenopus laevis*. *Mech. Dev.* 121, 573–585. doi: 10.1016/j.mod.2004.03.034

Kotloff, K. L., Nataro, J. P., Blackwelder, W. C., Nasrin, D., Farag, T. H., Panchalingam, S., et al. (2013). Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* 382, 209–222. doi: 10.1016/s0140-6736(13)60844-2

Kraft, M. R., Klotz, C., Bücker, R., Schulzke, J.-D., and Aebischer, T. (2017). Giardia's epithelial cell interaction in vitro: mimicking asymptomatic infection? *Front. Cell. Infect. Microbiol.* 7:421. doi: 10.3389/fcimb.2017.00421

Kulakova, L., Singer, S. M., Conrad, J., and Nash, T. E. (2006). Epigenetic mechanisms are involved in the control of *Giardia lamblia* antigenic variation. *Mol. Microbiol.* 61, 1533–1542. doi: 10.1111/j.1365-2958.2006.05345.x

Lai, E. C. (2004). Notch signaling: control of cell communication and cell fate. *Development* 131, 965–973. doi: 10.1242/dev.01074

Li, E., Tako, E. A., and Singer, S. M. (2016). Complement activation by *Giardia duodenalis* parasites through the lectin pathway contributes to mast cell responses and parasite control. *Infect. Immun.* 84, 1092–1099. doi: 10.1128/iai.00074-16

Lin, Z.-Q., Gan, S.-W., Tung, S.-Y., Ho, C.-C., Su, L.-H., and Sun, C.-H. (2019). Development of CRISPR/Cas9-mediated gene disruption systems in *Giardia lamblia*. *PLoS One* 14:e0213594. doi: 10.1371/journal.pone.0213594

Lievin-Le Moal, V. (2013). Dysfunctions at human intestinal barrier by water-borne protozoan parasites: lessons from cultured human fully differentiated colon cancer cell lines. *Cell Microbiol.* 15, 860–869. doi: 10.1111/cmi.12126

Liu, J., Fu, Z., Hellman, L., and Svärd, S. G. (2019). Cleavage specificity of recombinant *Giardia intestinalis* cysteine proteases: degradation of immunoglobulins and defensins. *Mol. Biochem. Parasitol.* 227, 29–38. doi: 10.1016/j.molbiopara.2018.10.004

Liu, J., Ma'ayeh, S., Peirasmaki, D., Lundström-Stadelmann, B., Hellman, L., and Svärd, S. G. (2018). Secreted *Giardia intestinalis* cysteine proteases disrupt intestinal epithelial cell junctional complexes and degrade chemokines. *Virulence* 9, 879–894. doi: 10.1080/21505594.2018.1451284

Ma'ayeh, S. Y., and Brook-Carter, P. T. (2012). Representational difference analysis identifies specific genes in the interaction of *Giardia duodenalis* with the murine intestinal epithelial cell line. IEC-6. *Int. J. Parasitol.* 42, 501–509. doi: 10.1016/j.ijpara.2012.04.004

Ma'ayeh, S. Y., Knörr, L., and Svärd, S. G. (2015). Transcriptional profiling of *Giardia intestinalis* in response to oxidative stress. *Int. J. Parasitol.* 45, 925–938. doi: 10.1016/j.ijpara.2015.07.005

Ma'ayeh, S. Y., Liu, J., Peirasmaki, D., Hörnaeus, K., Bergström Lind, S., Grabherr, M., et al. (2017). Characterization of the *Giardia intestinalis* secretome during interaction with human intestinal epithelial cells: the impact on host cells. *PLoS Negl. Trop. Dis.* 11:e0006120. doi: 10.1371/journal.pone.0006120

Magne, D., Favennec, L., Chochillon, C., Gorenflot, A., Meillet, D., Kapel, N., et al. (1991). Role of cytoskeleton and surface lectins in *Giardia duodenalis* attachment to Caco2 cells. *Parasitol. Res.* 77, 659–662. doi: 10.1007/BF00928679

Mastronicola, D., Falabella, M., Forte, E., Testa, F., Sarti, P., and Giuffre, A. (2016). Antioxidant defence systems in the protozoan pathogen *Giardia intestinalis*. *Mol. Biochem. Parasitol.* 206, 56–66. doi: 10.1016/j.molbiopara.2015.12.002

McInally, S. G., Hagen, K. D., Nosala, C., Williams, J., Nguyen, K., Booker, J., et al. (2019). Robust and stable transcriptional repression in Giardia using CRISPRi. *Mol. Biol. Cell* 30, 119–130. doi: 10.1091/mbc.e18-09-0605

Mendez, T. L., De Chatterjee, A., Duarte, T., De Leon, J., Robles-Martinez, L., and Das, S. (2015). Sphingolipids, lipid rafts, and *Giardial* encystation: the show must go on. *Curr. Trop. Med. Rep.* 2, 136–143. doi: 10.1007/s40475-015-0052-0

Morrison, H. G., Mcarthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., et al. (2007). Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* 317, 1921–1926. doi: 10.1126/science.1143837

Nematian, J., Gholamrezanezhad, A., and Nematian, E. (2008). Giardiasis and other intestinal parasitic infections in relation to anthropometric indicators of malnutrition: a large, population-based survey of schoolchildren in Tehran. *Ann. Trop. Med. Parasitol.* 102, 209–214. doi: 10.1179/136485908X267876

Niu, Q., Li, S., Chen, D., Chen, Q., and Chen, J. (2016). Iron acquisition in *Leishmania* and its crucial role in infection. *Parasitology* 143, 1347–1357. doi: 10.1017/s0031182016000858

Ortega, Y. R., and Adam, R. D. (1997). Giardia: overview and update. *Clin. Infect. Dis.* 25, 545–550.

Paiva, C. N., Medei, E., and Bozza, M. T. (2018). ROS and *Trypanosoma cruzi*: Fuel to infection, poison to the heart. *PLoS Pathog.* 14:e1006928. doi: 10.1371/journal.ppat.1006928

Parente, R., Clark, S. J., Inforzato, A., and Day, A. J. (2017). Complement factor H in host defense and immune evasion. *Cell Mol. Life Sci.* 74, 1605–1624. doi: 10.1007/s00018-016-2418-4

Pham, J. K., Nosala, C., Scott, E. Y., Nguyen, K. F., Hagen, K. D., Starcevich, H. N., et al. (2017). Transcriptomic profiling of high-density giardia foci encysting in the murine proximal intestine. *Front. Cell. Infect. Microbiol.* 7:227. doi: 10.3389/fcimb.2017.00227

Prado, M. S., Cairncross, S., Strina, A., Barreto, M. L., Oliveira-Assis, A. M., and Rego, S. (2005). Asymptomatic giardiasis and growth in young children; a longitudinal study in Salvador, Brazil. *Parasitology* 131, 51–56. doi: 10.1017/s0031182005007353

Prucca, C. G., and Lujan, H. D. (2009). Antigenic variation in *Giardia lamblia*. *Cell. Microbiol.* 11, 1706–1715. doi: 10.1111/j.1462-5822.2009.01367.x

Prucca, C. G., Slavin, I., Quiroga, R., Elias, E. V., Rivero, F. D., Saura, A., et al. (2008). Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature* 456, 750–754. doi: 10.1038/nature07585

Ratner, D. M., Cui, J., Steffen, M., Moore, L. L., Robbins, P. W., and Samuelson, J. (2008). Changes in the *N*-Glycome, Glycoproteins with Asn-Linked Glycans, of *Giardia lamblia* with Differentiation from Trophozoites to Cysts. *Eukaryotic Cell* 7, 1930–1940. doi: 10.1128/EC.00268-08

Ringqvist, E., Avesson, L., Söderbom, F., and Svärd, S. G. (2011). Transcriptional changes in Giardia during host–parasite interactions. *Int. J. Parasitol.* 41, 277–285. doi: 10.1016/j.ijpara.2010.09.011

Ringqvist, E., Palm, J. E. D., Skarin, H., Hehl, A. B., Weiland, M., Davids, B. J., et al. (2008). Release of metabolic enzymes by *Giardia* in response to interaction with intestinal epithelial cells. *Mol. Biochem. Parasitol.* 159, 85–91. doi: 10.1016/j.molbiopara.2008.02.005

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Rogawski, E. T., Bartelt, L. A., Platts-Mills, J. A., Seidman, J. C., Samie, A., Havt, A., et al. (2017). Determinants and impact of giardia infection in the first 2 years of Life in the MAL-ED birth cohort. *J. Pediatric. Infect. Dis. Soc.* 6, 153–160. doi: 10.1093/jpids/piw082

Rodríguez-Fuentes, G. B., Cedillo-Rivera, R., Fonseca-Liñán, R., Argüello-García, R., Muñoz, O., Ortega-Pierres, G., et al. (2006). *Giardia duodenalis*: analysis of secreted proteases upon trophozoite-epithelial cell interaction in vitro. *Mem. Inst. Oswaldo Cruz* 101, 693–696. doi: 10.1590/s0074-02762006000600020

Roxström-Lindquist, K., Ringqvist, E., Palm, D., and Svärd, S. (2005). *Giardia lamblia*-induced changes in gene expression in differentiated Caco-2 human intestinal epithelial cells. *Infect. Immun.* 73, 8204–8208. doi: 10.1128/iai.73.12.8204-8208.2005

Ryan, U., Hijjawi, N., Feng, Y., and Xiao, L. (2019). *Giardia*: an under-reported foodborne parasite. *Int. J. Parasitol.* 49, 1–11. doi: 10.1016/j.ijpara.2018.07.003

Saraiya, A. A., Li, W., Wu, J., Chang, C. H., and Wang, C. C. (2014). The microRNAs in an ancient protist repress the variant-specific surface protein expression by targeting the entire coding sequence. *PLoS Pathog.* 10:e1003791. doi: 10.1371/journal.pone.1003791

Scott, K. G. E., Logan, M. R., Klammer, G. M., Teoh, D. A., and Buret, A. G. (2000). Jejunal brush border microvillous alterations ingiardia muris-infected mice: role of T lymphocytes and Interleukin-6. *Infect. Immun.* 68, 3412–3418. doi: 10.1128/iai.68.6.3412-3418.2000

Simsek, Z., Zeyrek, F. Y., and Kurcer, M. A. (2004). Effect of *Giardia* infection on growth and psychomotor development of children aged 0–5 years. *J. Trop. Pediatrics.* 50, 90–93. doi: 10.1093/tropej/50.2.90

Solaymani-Mohammadi, S., and Singer, S. M. (2011). Host immunity and pathogen strain contribute to intestinal disaccharidase impairment following gut infection. *J. Immunol.* 187, 3769–3775. doi: 10.4049/jimmunol.1100606

Sonda, S., Morf, L., Bottova, I., Baetschmann, H., Rehrauer, H., Caflisch, A., et al. (2010). Epigenetic mechanisms regulate stage differentiation in the minimized protozoan *Giardia lamblia*. *Mol. Microbiol.* 76, 48–67. doi: 10.1111/j.1365-2958.2010.07062.x

Soto-Castro, L., Plata-Guzmán, L. Y., Figueroa-Angulo, E. E., Calla-Choque, J. S., Reyes-López, M., and De La Garza, M. (2017). Iron responsive-like elements in the parasite *Entamoeba histolytica*. *Microbiology (Reading, England)* 163, 1329–1342. doi: 10.1099/mic.0.000431

Stenflo, J. (1991). Structure-function relationships of epidermal growth factor modules in vitamin K-dependent clotting factors. *Blood* 78, 1637–1651.

Troeger, H., Epple, H. J., Schneider, T., Wahnschaffe, U., Ullrich, R., Burchard, G. D., et al. (2007). Effect of chronic *Giardia lamblia* infection on epithelial transport and barrier function in human duodenum. *Gut* 56, 328–335. doi: 10.1136/gut.2006.100198

Williams, A. G., and Coombs, G. H. (1995). Multiple protease activities in *Giardia intestinalis* trophozoites. *Int. J. Parasitol.* 25, 771–778. doi: 10.1016/0020-7519(94)00201-x

Wouters, M. A., Rigoutsos, I., Chu, C. K., Feng, L. L., Sparrow, D. B., and Dunwoodie, S. L. (2005). Evolution of distinct EGF domains with specific functions. *Protein Science* 14, 1091–1103. doi: 10.1110/ps.041207005

Xu, F., Jex, A., and Svard, S. G. (2020). A chromosome-scale reference genome for *Giardia intestinalis* WB. *Sci. Data* 7:38.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 September 2020
doi: 10.3389/fgene.2020.01011

Genetic Analysis of Antibody Response to Porcine Reproductive and Respiratory Syndrome Vaccination as an Indicator Trait for Reproductive Performance in Commercial Sows

*Leticia P. Sanglard[1], Rohan L. Fernando[1], Kent A. Gray[2], Daniel C. L. Linhares[3], Jack C. M. Dekkers[1], Megan C. Niederwerder[4] and Nick V. L. Serão[1]**

[1] Department of Animal Science, Iowa State University, Ames, IA, United States, [2] Smithfield Premium Genetics, Rose Hill, NC, United States, [3] Department of Veterinary Diagnostic and Production Animal Medicine, Iowa State University, Ames, IA, United States, [4] Department of Diagnostic Medicine/Pathobiology, Kansas State University, Manhattan, KS, United States

OPEN ACCESS

Edited by:
Androniki Psifidi,
Royal Veterinary College (RVC),
United Kingdom

Reviewed by:
Katrina Morris,
University of Edinburgh,
United Kingdom
Xiaoxia Dai,
United States Department
of Agriculture, United States

***Correspondence:**
Nick V. L. Serão
serao@iastate.edu

Specialty section:
This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 30 January 2020
Accepted: 07 August 2020
Published: 11 September 2020

Citation:
Sanglard LP, Fernando RL,
Gray KA, Linhares DCL,
Dekkers JCM, Niederwerder MC and
Serão NVL (2020) Genetic Analysis
of Antibody Response to Porcine
Reproductive and Respiratory
Syndrome Vaccination as an Indicator
Trait for Reproductive Performance
in Commercial Sows.
Front. Genet. 11:1011.
doi: 10.3389/fgene.2020.01011

We proposed to investigate the genomic basis of antibody response to porcine reproductive and respiratory syndrome (PRRS) virus (PRRSV) vaccination and its relationship to reproductive performance in non-PRRSV-infected commercial sows. Nine hundred and six F1 replacement gilts (139 ± 17 days old) from two commercial farms were vaccinated with a commercial modified live PRRSV vaccine. Blood samples were collected about 52 days after vaccination to measure antibody response to PRRSV as sample-to-positive (S/P) ratio and for single-nucleotide polymorphism (SNP) genotyping. Reproductive performance was recorded for up to 807 sows for number born alive (NBA), number of piglets weaned, number born mummified (MUM), number of stillborn (NSB), and number of pre-weaning mortality (PWM) at parities (P) 1–3 and per sow per year (PSY). Fertility traits such as farrowing rate and age at first service were also analyzed. BayesC0 was used to estimate heritability and genetic correlations of S/P ratio with reproductive performance. Genome-wide association study (GWAS) and genomic prediction were performed using BayesB. The heritability estimate of S/P ratio was 0.34 ± 0.05. High genetic correlations (r_g) of S/P ratio with farrowing performance were identified for NBA P1 (0.61), PWM P2 (-0.70), NSB P3 (-0.83), MUM P3 (-0.84), and NSB PSY (-0.90), indicating that genetic selection for increased S/P ratio would result in improved performance of these traits. A quantitative trait locus was identified on chromosome 7 (∼25 Mb), at the major histocompatibility complex (MHC) region, explaining ∼30% of the genetic variance for S/P ratio, mainly by SNPs ASGA0032113, H3GA0020505, and M1GA0009777. This same region was identified in the bivariate GWAS of S/P ratio and reproductive traits, with SNP H3GA0020505 explaining up to 10% (for NBA P1) of the genetic variance of reproductive performance. The heterozygote genotype at H3GA0020505 was associated with greater S/P ratio and NBA P1 (P = 0.06), and lower MUM P3 and NSB P3 (P = 0.07). Genomic prediction

accuracy for S/P ratio was high when using all SNPs (0.67) and when using only those in the MHC region (0.59) and moderate to low when using all SNPs excluding those in the MHC region (0.39). These results suggest that there is great potential to use antibody response to PRRSV vaccination as an indicator trait to improve reproductive performance in commercial pigs.

Keywords: antibody response, porcine reproductive and respiratory syndrome vaccination, heritability, reproductive performance, genetic correlation, bivariate genome-wide association study

INTRODUCTION

Antibody response, measured as sample-to-positive (S/P) ratio, to porcine reproductive and respiratory syndrome (PRRS) in PRRS virus (PRRSV)-infected sows has been proposed as an indicator trait for selection for improved reproductive performance in sows during a PRRS outbreak (Serão et al., 2014). These authors reported that S/P ratio measured at approximately 46 days after the outbreak had high heritability (h^2), with an estimate of 0.45. In addition, these authors showed that S/P ratio was highly genetically favorably correlated with litter size traits during PRRSV infection, such as number born alive (NBA), with an estimate of 0.73 and number of stillborn (NSB) of -0.72. Putz et al. (2019) also reported a high and negative r_g estimate between S/P ratio and NSB (-0.73). Nonetheless, these authors obtained a weak r_g estimate between S/P and NBA (0.05) and a low h^2 estimate for S/P ratio (0.17). However, waiting for PRRS outbreaks to occur to collect data might limit the use of S/P ratio as a selection tool. Ideally, this strong relationship between S/P ratio and performance in PRRSV-infected sows would also be favorable in non-infected pigs.

A more feasible practice would be to use antibody response to PRRSV vaccination, which is a commonly used tool to control PRRS in commercial herds. PRRSV vaccination with a modified live virus (MLV) vaccine stimulates the same mechanisms of evasion as are developed during natural infection (Lopez and Osorio, 2004). Initially, T-cell-mediated response is stimulated with the production of interferon-gamma, and later non-neutralizing antibodies also play a role against the virus (Lopez and Osorio, 2004). Part of this response is controlled by genes located in the major histocompatibility complex (MHC), and this region may play a role in the relationship between immune response to vaccination and reproductive performance. Indeed, haplotypes in the MHC class I and II regions have been previously associated with reproductive traits, such as ovulation rate, embryo development, and litter size in non-infected pigs (Vaiman et al., 1998). Genetic variation in this region has been associated with S/P ratio in naturally PRRSV-infected sows (Serão et al., 2014) and in F1 replacement gilts (Serão et al., 2016). In addition, Serão et al. (2016) reported h^2 estimates ranging from 0.28 to 0.47, as the proportion of seroconverted animals increased in the dataset. Although there was no confirmation on whether the replacement gilts in Serão et al. (2016) were PRRSV-vaccinated or naturally PRRSV-infected, or even both, these authors hypothesized that PRRSV vaccination would yield similar results at the genetic level for S/P ratio as in Serão et al. (2014). More directly investigating the relationship between antibody response to PRRSV vaccination and subsequent reproductive performance, Abella et al. (2019) evaluated the impact of PRRS vaccination on growing pigs at 6–7 weeks of age at the time of vaccination. These authors measured S/P ratio at 42 days after vaccination and obtained an h^2 estimate of 0.69. Also, they developed a phenotyping criterion to discriminate susceptible and resilient sows based on viral load at 7 and 21 days post vaccination (Abella et al., 2019). At the phenotypic level, they reported that susceptible sows had greater antibody response to vaccination and higher NSB than those classified as resilient. However, in this study, they measured antibody in nursery pigs, when the energy being used for growth needs to be channeled for production of antibody after the vaccination, which may affect the relationship between antibody response and future performance.

Thus, in our study, we proposed to investigate the genomic basis of S/P ratio to PRRSV vaccination and its relationship with reproductive performance in non-infected commercial sows in an independent dataset (from a different breeding company) than the animals from Serão et al. (2016). For that purpose, we aimed to (1) estimate genetic parameters and genomic prediction accuracy (GPA) for antibody response to PRRSV vaccination, (2) assess the phenotypic and genetic relationships of antibody response to PRRSV vaccination with reproductive performance in commercial sows, and (3) identify regions of the genome [quantitative trait loci (QTLs)] that are associated with these traits.

MATERIALS AND METHODS

All methods described in this study were approved by the Institutional Animal Care and Use Committee at Iowa State University (IACUC# 6-17-8551-S).

Phenotypic and Genotypic Data

Nine hundred and six naïve F1 (Landrace × Large White) replacement gilts from two commercial farms in North Carolina, United States, were vaccinated (139 ± 17 days old) intramuscularly with a commercial PRRS MLV vaccine (Ingelvac PRRS MLV, Boehringer Ingelheim Animal Health, Ames, IA, United States), following the manufacturer's guidelines. Throughout the study, the farm performed periodic diagnostic tests for PRRSV and all animals (included or not in the study) did not show PCR-positive tests or signs of natural PRRSV infection. Blood samples were collected using Lavender Top Vacutainer tubes (Becton, Dickinson and Company, Franklin Lakes, NJ,

United States) at approximately 50 days after vaccination in three contemporary groups (CGs), in which each CG represented 1 day of blood collection (52 and 53 days post vaccination for one farm, and 46 days post vaccination on the other farm). After collection, a drop of blood from each sample was used on blood cards (Neogen Genomics, Lincoln, NE, United States) for subsequent genotyping. The remaining blood samples were shipped to the Veterinary Diagnostic Laboratory at Iowa State University (Ames, IA, United States), where samples were processed using the laboratory's standard procedures, for measurement of Immunoglobulin G against PRRSV, as S/P ratio, using a commercial ELISA test (IDEXX PRRS X3, IDEXX Laboratories Inc., Westbrook, ME, United States).

A subset of 807 animals from the animals with information of S/P ratio, located on PRRS-negative commercial farm, had farrowing performance recorded for up to three parities from January 2018 (∼150 days after blood collection) to December 2018 for the following litter size traits: NBA, number of stillborn (NSB), number born mummified (MUM), number of piglets weaned (NW), and number of pre-weaning mortality from the total number being weaned (PWM). Number born dead (NBD) was calculated as the sum of MUM and NSB, and total number born (TNB) was calculated as the sum of NBA and NBD. For animals with three parities, the number of piglets per sow per year (PSY) was calculated for each trait as the sum of the phenotype across the three parities divided by the difference in days between the third and first farrowing, multiplied by 365 days. We also analyzed fertility traits such as farrowing rate (FR), which was defined as the probability of an inseminated sow to farrow, age at first service (AFS), and farrowing interval (FI), which was the difference in days between the farrow date of parities 1 and 2, and between parities 2 and 3. The summary statistics of the data are presented in **Table 1**.

Blood cards were shipped to Neogen Genomics (Lincoln, NE, United States) for DNA extraction according to their standard procedures. Then, DNA from each animal was used for genotyping using the GGP Porcine HD (Neogen GeneSeek) for a total of 50,697 single-nucleotide polymorphisms (SNPs). Genotypes were set to missing if GC score < 0.50, SNPs with call rate < 0.90 were removed, and animals with genotype call rate < 0.90 were removed. After quality control, 45,536 SNPs and 906 animals were used for subsequent analyses. Positions of SNPs on the genome were based on the *Sus scrofa* 11.1 assembly.

TABLE 1 | Descriptive statistics of the data.

Traits	*N*	Mean	*SD*	Min	Max
S/P ratio	906	1.41	0.45	0.06	2.55
AFS	807	247.1	27.9	164.0	392.0
FI (P1 vs. P2)	616	161.8	28.5	129.0	343.0
FI (P2 vs. P3)	460	152.6	13.8	131.0	262.0
Parity 1					
FR	807	0.94	0.25	0.00	1.00
NBA	744	11.63	2.99	0.00	19.00
NSB	744	0.50	0.92	0.00	10.00
MUM	714	0.36	1.02	0.00	13.00
NBD	744	0.85	1.46	0.00	13.00
TNB	744	12.49	2.87	4.00	20.00
NW	735	11.03	2.47	0.00	15.00
PWM	735	2.27	2.07	0.00	11.00
Parity 2					
FR	755	0.81	0.40	0.00	1.00
NBA	608	12.67	3.35	0.00	22.00
NSB	608	0.45	0.86	0.00	7.00
MUM	608	0.28	0.88	0.00	12.00
NBD	608	0.73	1.31	0.00	12.00
TNB	608	13.40	3.37	4.00	24.00
NW	605	10.74	2.43	0.00	15.00
PWM	605	2.03	1.90	0.00	10.00
Parity 3					
FR	608	0.75	0.44	0.00	1.00
NBA	458	12.88	2.99	0.00	20.00
NSB	458	0.49	0.86	0.00	6.00
MUM	458	0.20	0.49	0.00	3.00
NBD	458	0.69	1.07	0.00	6.00
TNB	458	13.57	3.15	4.00	21.00
NW	458	9.01	4.08	0.00	15.00
PWM	458	2.05	2.03	0.00	12.00
Per sow per year					
NBA	448	44.45	7.43	21.00	66.36
NSB	448	0.63	0.56	0.00	2.36
MUM	448	0.41	0.53	0.00	2.90
NBD	448	1.00	0.70	0.00	3.06
TNB	448	47.15	7.61	21.00	73.48
NW	426	37.04	7.24	9.54	54.22
PWM	426	6.37	0.86	0.00	20.58

AFS, age at first service (days); FI, farrow interval (days); P1, parity 1; P2, parity 2; P3, parity 3; FR, farrowing rate (1 if the sow farrowed or 0 if not after first service); NBA, number born alive; NSB, number of stillborn; MUM, number born mummified; NBD, number born dead; TNB, total number born; NW, number weaned; PWM, pre-weaning mortality; N, number of records.

Statistical Analyses

Heritability and Genetic Correlations

Bayesian analysis (BayesC0; Habier et al., 2011) was used to estimate (co)variance parameters using the following model for each parity separately:

$$y = \mu + Xb + Wu + \sum_{i=1}^{k} z_i a_i \delta_i + e$$

where y_{ij} is a vector of phenotypic response (S/P ratio or reproductive performance); μ is the intercept; X is the incidence matrix relating the fixed effects to the response; $\boldsymbol{b}$ is a vector of fixed effects: CG for S/P ratio, farm for farrowing performance, and number of piglets cross-fostered and NBA as covariate (for NW and PWM); $\boldsymbol{W}$ is the incidence matrix relating the random effects to the response; $\boldsymbol{u}$ is the vector of random effects: combination of month/year of farrowing for NBA, NSB, MUM, NBD, and TNB, month/year of weaning for NW and PWM, and month/year of birth for AFS, PSY, FR, and FI traits; z_i is the vector of genotypes for SNP i (coded as 0, 1, and 2); α_i is the allele substitution effect of SNP i; δ_i is an indicator whether SNP i was included ($\delta_i = 1$) or excluded ($\delta_i = 0$) in the model for a

given iteration of the Monte Carlo Markov chain (MCMC) (for BayesC0, $\delta_i = 1$); and e is the vector of residuals. FR was a binary trait and was analyzed with a threshold model.

Bayesian analyses consisted of 50,000 MCMCs, with the first 5,000 discarded as burn-in. At every 100th iteration of the chain, the breeding value of each individual used in the analysis was calculated as the sum of its genotypes multiplied by the sampled marker effects. The variance of the sampled breeding values was used as the sampled additive genetic variance in that iteration. The sampled additive genetic variance was divided by the sampled phenotypic variance (sum of sampled additive and residual variances) at each iteration to obtain the sampled heritability (h^2). Then, the estimate of h^2 was calculated as the posterior mean of the sampled h^2. The posterior standard deviation of the h^2 samples was used as the standard error of the estimate.

Bivariate analyses were performed between S/P ratio and reproductive performance using the same fixed and random effects as used for the univariate analyses, fitting the following model (Cheng et al., 2018a):

$$y_j = \mu + Xb + Wu + \sum_{i=1}^{m} z_{ij} D_i \beta_i + e_j$$

where y_j is a vector of phenotypes of t ($t = 2$) traits for individual j; μ is a vector of overall means for t traits; $\boldsymbol{X}$ is $\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}$, where X_1 and X_2 are the incidence matrices relating the fixed effects to the response for S/P ratio ($k = 1$) and reproductive performance ($k = 2$), respectively; $\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$, where b_1 and b_2 are the vectors of fixed effects for S/P ratio and reproductive performance, respectively; W is the incidence matrix relating the random effects to the response (reproductive traits); u is the vector of random effects (month/year of farrow); z_{ij} is the genotype covariate at locus i for individual j (coded as 0, 1, and 2); m is the number of genotyped loci; D_i is a diagonal matrix with elements $\text{diag}(D_i) = \delta_i = (\delta_{i1}, \delta_{i2})$, where δ_{ik} is an indicator variable indicating if the marker effect of locus i for trait k is zero or not and, in this case, there are $2^t = 4$ combinations for δ_i: (0, 0), (1, 0), (0, 1), and (1, 1), with (0, 0) representing the proportion of markers not being fitted in both traits and so on (note that for BayesC0, all markers are being fitted for both traits); β_i is the vector of marker effect for loci β_i, where $\sim$ MVN (0, $\boldsymbol{G}$), where $G = \begin{bmatrix} \sigma^2_\beta i_1 & \sigma_\beta i_{1,2} \\ \sigma_\beta i_{1,2} & \sigma^2_\beta i_2 \end{bmatrix}$ and is assumed to have an inverse Wishart prior distribution, $W_t^{-1}(S_\beta, \nu_\beta)$; and e_j is the vector of residuals of t traits for individual j, where $e_j \sim$ MVN (0, $\boldsymbol{R}$), where $R = \begin{bmatrix} \sigma^2_{e_1} & \sigma_{e_{1,2}} \\ \sigma_{e_{1,2}} & \sigma^2_{e_2} \end{bmatrix}$ and is assumed to have an inverse Wishart prior distribution, $W_t^{-1}(S_e, \nu_e)$.

The genetic correlation (r_g) was estimated as the posterior mean of the correlation between the sampled genomic breeding values for each animal for each trait at each iteration and its standard deviation across iterations as the standard error. The proportion of the covariance between S/P ratio and reproductive performance that was explained by a 1-Mb window was calculated as the covariance between sampled window breeding for each animal obtained based on the SNPs in the 1-Mb window divided by the covariance between sampled breeding values for each animal obtained based on all SNPs across the genome. All analyses were performed in the *JWAS* package (Cheng et al., 2018b), written in the Julia programing language (Bezanson et al., 2017).

Genome-Wide Association Studies

Univariate and bivariate GWASs were performed for S/P ratio and for S/P ratio with reproductive performance, respectively, using the same models as before. First, BayesCπ (Habier et al., 2011) was used to estimate the proportion of markers to be fitted in the model. Then, BayesB (Habier et al., 2011) was used with the estimated π to identify QTL within 1-Mb SNP windows that explained most of the genetic variance accounted for by the SNPs and that had a posterior probability of inclusion (PPI) greater than 0.7 (Garrick and Fernando, 2013). A bivariate GWAS was performed when the estimate of r_g between S/P ratio, and the reproductive trait analyzed was larger than 0.50 to investigate QTL associated with the genetic covariance between the two traits. All analyses were performed using the *JWAS* package. Candidate genes in the QTL regions were identified using Ensembl BioMart (Kinsella et al., 2011). Linkage disequilibrium (LD) between SNPs within QTL regions was estimated as r^2 using Plink (Purcell et al., 2007) and plotted using Haploview (Barrett et al., 2005).

Effect of Major Single-Nucleotide Polymorphisms on Antibody Response and Reproductive Traits

Results from the bivariate GWAS were used to perform additional analyses to evaluate the impact of SNPs on traits evaluated in this study. For this, SNPs with PPI > 0.70 or that explained more than 1% of the genetic variance explained by markers (TGVM) for each trait were simultaneously fitted as categorical fixed effects in the model used for estimation of genetic parameters. Orthogonal contrasts were used to test the additive and dominance effects for each marker. Significant associations and tendency were considered at the significance levels of $P \leq 0.05$ and $P \leq 0.10$, respectively. Analyses were performed in ASReml v4.0 (Gilmour et al., 2015).

Genomic Prediction

Genomic prediction was performed for S/P ratio using the same statistical model as used for GWAS, using BayesB ($\pi = 0.995$) and BayesC0. We used BayesB based on our results and on the literature, which showed that S/P ratio has a major QTL accounting for ~25% of the genetic variance. The BayesB method assumes different genetic variance per locus, allowing for the QTL to be well captured while reducing the effect of the small-effect QTL. Due to this major QTL for S/P ratio, we also analyzed the data removing SNPs located in this major QTL. With many more markers to be estimated and assumption equal variance, we chose to use BayesC0, as this method is more suitable under these circumstances. Nonetheless, this is the first study showing

the genomic prediction for S/P ratio to vaccination, and thus, we investigated both methods to identify the one that would be more accurate.

A three-fold cross-validation was used for genomic prediction analyses. For this, data from two CGs were used for training, and the data from the remaining CG were used as validation. This was repeated until all CGs were used as validation. The number of individuals in each validation dataset was 378, 257, and 252 for CGs 1, 2, and 3, respectively. Marker effects were estimated (i.e., trained) based on the training population in three scenarios: (1) using the whole genome (ALL), using markers for the MHC QTL on SSC 7 for each training population (MHC), and using the rest of markers not included in the MHC scenario (REST). For the MHC scenario, a GWAS was performed for each training population to preselect the markers. The regions defined for each training population were SSC 7 25–26 Mb (29 SNPs) for two training populations (CGs 1 with 3, and CGs 2 with 3) and SSC 7 23–26 Mb (81 SNPs) for the other training population (CGs 1 with 2). GPA was calculated as follows:

$$\text{GPA} = \frac{\frac{\sum_{i=1}^{3} n_i r_{i(GEBV,\, y^*)}}{\sum_{i=1}^{3} n_i}}{\sqrt{h^2}}$$

where n_i is the number of individuals in the ith validation dataset (i = 1, 2, 3), $r_{i(GEBV,\, y^*)}$ is the correlation between the genomic estimated breeding values (GEBVs) and phenotypes adjusted for estimates of fixed-effects (y^*) for the ith validation dataset, and h^2 is the heritability estimate using the whole dataset. All analyses were performed using the *JWAS* package.

RESULTS

Genetic Parameters

Estimates of h^2 are presented in **Table 2**. h^2 estimates for S/P ratio and AFS were moderate, with 0.34 and 0.29, respectively, and higher than those for reproductive performance. This indicates that a faster genetic progress could be obtained for S/P ratio compared with reproductive performance. All other fertility and litter size traits showed low h^2 (<0.20) and were similar across parities and sows per year. The highest and lowest h^2 estimates for these traits were 0.16 (FR P3) and <0.001 (NW and PWM P1, NSB and NBD P2, and MUM and NBD PSY), respectively. In general, the average h^2 for FR (0.16) across parities was higher than for litter size traits (0.05). As expected, fertility and litter size traits had low h^2 estimates, indicating slow genetic progress for these traits.

The r_g estimates between S/P ratio with reproductive traits are presented in **Table 2**. There were few moderate-to-high r_g estimates with favorable direction, such as between S/P ratio with NBA P1 (0.61), PWM P2 (-0.70), NSB P3 (-0.84), MUM P3 (-0.83), and NSB PSY (-0.90). These estimates indicate that selection for increased S/P ratio would result in improved farrowing performance for these traits. A moderate non-favorable r_g estimate was observed between S/P ratio and NW P1 (-0.35). For other traits, r_g estimates were overall low and favorable.

TABLE 2 | Estimates of heritability (h^2) and correlations (genetic and phenotypic; r_g and r_p, respectively) between antibody response [sample-to-positive (S/P) ratio] to porcine reproductive and respiratory syndrome virus vaccination and reproductive traits.

Traits	h^2	r_g	r_p
S/P ratio	0.34(0.05)	–	–
AFS	0.29(0.05)	−0.25(0.13)	−0.05(0.03)
FI (P1 vs. P2)	0.07(0.03)	−0.16(0.25)	−0.02(0.04)
FI (P2 vs. P3)	<0.001(0.01)	0.07(0.26)	0.07(0.04)
Parity 1			
FR	0.16(0.04)	−0.04(0.13)	−0.04(0.03)
NBA	0.06(0.02)	0.61(0.16)	0.02(0.03)
NSB	0.03(0.02)	−0.02(0.20)	−0.001(0.03)
MUM	0.01(0.005)	−0.17(0.23)	−0.002(0.03)
NBD	0.05(0.02)	−0.05(0.29)	0.02(0.03)
TNB	0.08(0.04)	0.30(0.19)	0.01(0.03)
NW	<0.001(0.01)	−0.35(0.26)	−0.08(0.03)
PWM	<0.001(0.01)	−0.14(0.27)	0.09(0.03)
Parity 2			
FR	0.15(0.03)	−0.05(0.15)	−0.06(0.03)
NBA	0.10(0.04)	−0.15(0.21)	−0.09(0.03)
NSB	<0.001(0.01)	−0.03(0.19)	−0.06(0.04)
MUM	0.02(0.01)	0.19(0.33)	0.05(0.03)
NBD	<0.001(0.01)	−0.17(0.18)	−0.007(0.04)
TNB	0.15(0.05)	−0.19(0.63)	−0.10(0.04)
NW	0.11(0.06)	0.01(0.21)	0.04(0.04)
PWM	0.07(0.07)	−0.70(0.10)	−0.02(0.03)
Parity 3			
FR	0.16(0.03)	0.08(0.14)	−0.005(0.03)
NBA	0.04(0.01)	0.02(0.20)	0.001(0.04)
NSB	0.001(0.0007)	−0.84(0.05)	−0.01(0.04)
MUM	0.003(0.003)	−0.83(0.11)	0.01(0.04)
NBD	0.01(0.009)	−0.19(0.23)	0.002(0.04)
TNB	0.12(0.03)	0.01(0.22)	0.006(0.04)
NW	0.08(0.09)	−0.11(0.38)	−0.04(0.04)
PWM	0.21(0.14)	−0.18(0.44)	−0.02(0.03)
Per sow per year			
NBA	0.07(0.04)	0.20(0.16)	−0.07(0.04)
NSB	0.01(0.005)	−0.90(0.05)	−0.07(0.04)
MUM	<0.001(0.01)	−0.03(0.30)	0.01(0.04)
NBD	<0.001(0.01)	−0.24(0.36)	−0.06(0.05)
TNB	0.09(0.05)	−0.02(0.26)	−0.08(0.04)
NW	0.04(0.09)	−0.17(0.37)	−0.06(0.04)
PWM	0.10(0.10)	−0.13(0.27)	0.06(0.04)

AFS, age at first service; FI, farrow interval; P1, parity 1; P2, parity 2; P3, parity 3; FR, farrowing rate; NBA, number born alive; NSB, number of stillborn; MUM, number of mummies; NBD, number born dead (NSB + MUM); TNB, total number born (NBA + NBD); NW, number weaned; PWM, pre-weaning mortality.

Genome-Wide Association Studies

The GWAS results are presented in **Figure 1** for the univariate and bivariate analyses. For the univariate analysis of S/P ratio (**Table 3**), we identified a region on *Sus scrofa* chromosome (SSC) 7 (23–26 Mb), the MHC region, explaining 30% of the TGVM (PPI = 1). This variance was mainly explained by SNPs ASGA0032113, H3GA0020505, and M1GA0009777,

TABLE 3 | Results for the genome-wide association study (GWAS) analysis for sample-to-positive (S/P) ratio (single trait) to porcine reproductive and respiratory syndrome virus vaccination measure as and bi-trait for S/P ratio and reproductive performance.

Traits	SSC	Window		#SNPs	S/P ratio		Reproductive trait	
		Start (Mb)	End (Mb)		TGVM (%)	PPI	TGVM (%)	PPI
Single trait	7	25,003,013	25,967,157	29	30.00	1.00	–	–
	7	23,037,875	23,985,825	18	3.50	0.73	–	–
Bi-trait								
NBA P1	7	25,003,013	25,967,157	29	20.35	0.99	10.36	0.84
PWM P2	7	25,003,013	25,967,157	29	19.67	0.84	7.96	0.76
NSB P3	7	25,003,013	25,967,157	29	22.38	1.00	4.74	0.42
MUM P3	7	25,003,013	25,967,157	29	22.12	0.99	3.61	0.58
NSB PSY	7	25,003,013	25,967,157	29	21.80	0.99	4.98	0.51

NBA, number born alive; PWM, pre-weaning mortality; NSB, number of stillborn; MUM, number of piglets mummified; P1, parity 1; P2, parity 2; and P3, parity 3; PSY, pigs per sow per year; SSC, Sus scrofa chromosome; #SNPs, number of SNPs within the window; TGVM, total genetic variance explained by the markers in the window; PPI, posterior probability of inclusion.

which explained 21% (PPI = 1), 10.5% (PPI = 0.93), and 3.5% (PPI = 0.78) of the TGVM, respectively.

For the bivariate analysis between S/P ratio and reproductive traits showing $r_g > 0.50$ with S/P ratio (**Table 2**), results are presented in **Table 3**. For all analyses, a similar region identified for the univariate analysis of S/P ratio on SSC 7 was found for all the traits, on the MHC class II region (SSC 7, 25–26 Mb; **Figure 1B**). This region explained 10.4% (PPI = 0.84), 7.9% (PPI = 0.53), 4.7% (PPI = 0.42), 3.6% (PPI = 0.58), and 4.9% (PPI = 0.51) of the TGVM for NBA P1, PWM P2, NSB P3, MUM P3, and NSB PSY, respectively; and an average of 21.4% (*SD* = 1.1%; PPI $\geq$ 0.99) for S/P ratio across all bivariate GWASs. Additionally, this region explained 34, 25, 26, 35, and 90% of the genetic covariance explained for by the markers (TGCoVM) between S/P ratio with NBA P1, PWM P2, NSB P3, MUM P3, and NSB PSY, respectively. These results indicate that this major region on the MHC for S/P ratio is also associated with farrowing traits.

The effect of the main SNPs explaining most of the %TGVM in the QTL identified in the univariate (S/P ratio; ASGA0032113, H3GA0020505, and M1GA0009777) and bivariate analysis (reproductive traits; H3GA0020505) is shown in **Figure 2**. These three SNPs were significantly ($P < 0.001$) associated with S/P ratio to PRRSV vaccination. The additive effect was significant ($P \leq 0.06$) for all of them, and the dominance effect was significant for H3GA0020505 ($P = 0.002$). For H3GA0020505 and M1GA0009777, genotypes AA (1.7 ± 0.11 and 1.8 ± 0.17, respectively) and AB (1.8 ± 0.06 and 1.7 ± 0.04, respectively) had greater S/P ratio than BB (1.5 ± 0.06 and 1.5 ± 0.04, respectively), while for ASGA0032113, S/P ratio increased from AA to BB, with 1.5 ± 0.06, 1.7 ± 0.06, and 1.8 ± 0.08, for AA, AB, and BB, respectively (**Figure 2A**).

The main SNP associated with farrowing performance in the bivariate analysis was H3GA0020505. This marker had significant ($P \leq 0.07$) effect on NBA P1, MUM P3, and NSB P3. For these traits, there was a dominance effect ($P \leq 0.06$), whereas there was an additive effect ($P = 0.01$) also for MUM P3 (**Figure 2B**). For NBA, AB (12.1 ± 0.22) had greater NBA than BB (11.5 ± 0.14), but these two genotypes did not differ from AA (11.8 ± 0.72). For MUM P3, AA (0.06 ± 0.02) had worse performance than AB (0.03 ± 0.001) and BB (0.03 ± 0.001).

Effect of Major Histocompatibility Complex Single-Nucleotide Polymorphisms on Antibody Response and Reproductive Traits

The effect of the main SNPs explaining most of the %TGVM in the QTL identified in the univariate (S/P ratio; ASGA0032113, H3GA0020505, and M1GA0009777) and bivariate analyses (reproductive traits; H3GA0020505) is shown in **Figure 2**. These three SNPs were significantly ($P < 0.001$) associated with S/P ratio to PRRSV vaccination. There was a tendency ($P \leq 0.06$) for the additive effect for all of them, and the dominance effect was significant for H3GA0020505 ($P = 0.002$). For H3GA0020505 and M1GA0009777, genotypes AA (1.7 ± 0.11 and 1.8 ± 0.17, respectively) and AB (1.8 ± 0.06 and 1.7 ± 0.04, respectively) had greater S/P ratio than BB (1.5 ± 0.06 and 1.5 ± 0.04, respectively), while for ASGA0032113, S/P ratio increased from AA to BB, with 1.5 ± 0.06, 1.7 ± 0.06, and 1.8 ± 0.08, for AA, AB, and BB, respectively (**Figure 2A**).

The main SNP associated with farrowing performance in the bivariate analysis was H3GA0020505. This marker shows a tendency ($P \leq 0.07$) for NBA P1, MUM P3, and NSB P3. For NBA P1, there was a tendency for the dominance effect ($P = 0.06$), whereas for MUM P3 and NSB P3, the additive and dominance effects were significant ($P \leq 0.04$; **Figure 2B**). For NBA, AB (12.1 ± 0.22) had greater NBA than BB (11.5 ± 0.14), but these two genotypes did not differ from AA (11.8 ± 0.72). For MUM P3, AA (0.06 ± 0.02) had worse performance than AB (0.03 ± 0.001) and BB (0.03 ± 0.001).

Genomic Prediction

GPAs for S/P ratio are presented in **Figure 3** for three scenarios: (1) using all the SNPs (ALL), (2) using SNPs on the MHC region, and (3) using SNPs only outside the MHC region (REST). The MHC region defined for the training populations without CG 1 or 2 was the same between them and narrower than the

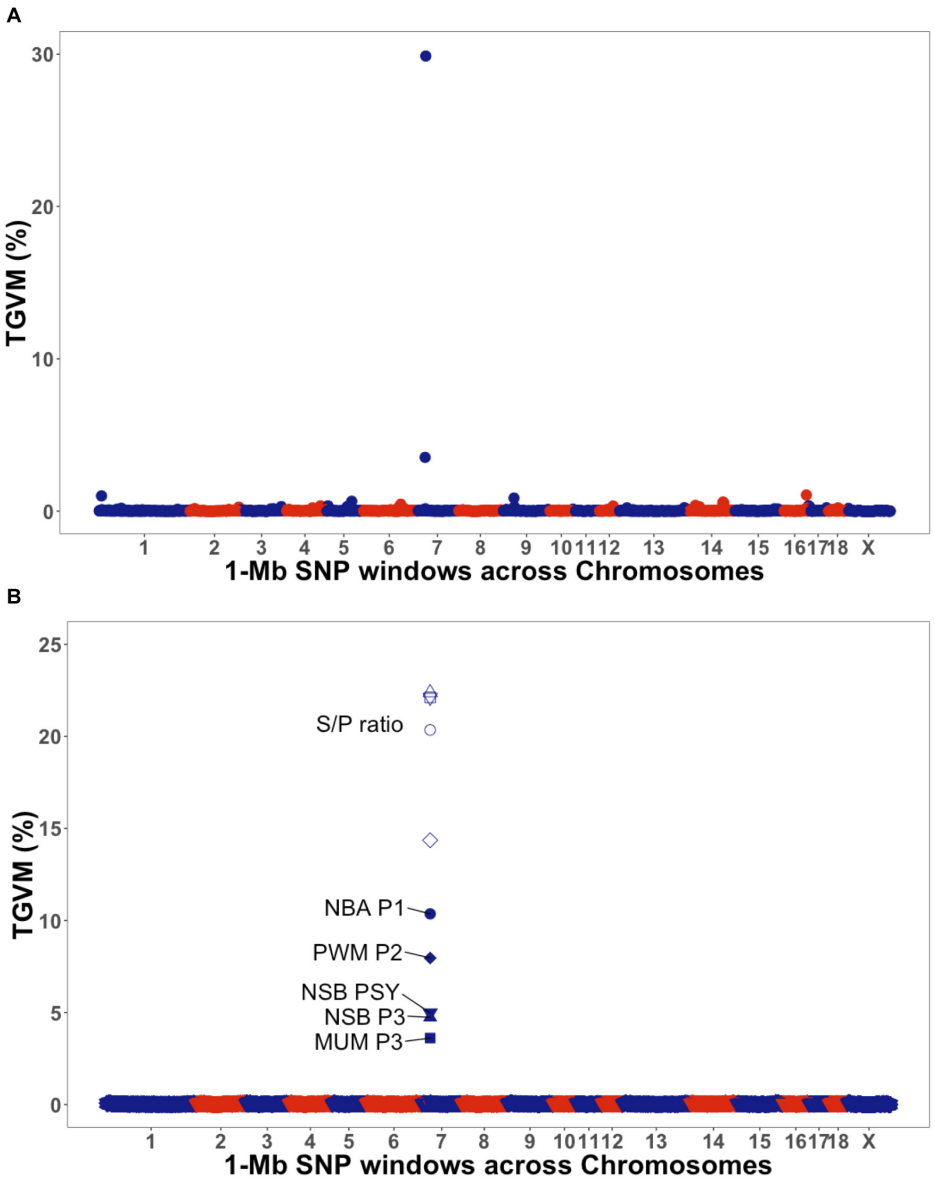


FIGURE 1 | Manhattan plot for univariate and bivariate genome-wide association studies. The *y*-axis represents the percentage of total genetic variance explained for by markers (TGVM). The *x*-axis represents the position of 1-Mb single-nucleotide polymorphism (SNP) window across the genome. **(A)** Manhattan plot for univariate of antibody response [sample-to-positive (S/P) ratio] to porcine reproductive and respiratory syndrome virus vaccination (S/P ratio). **(B)** Manhattan plot for bivariate of S/P ratio (open symbols) and farrowing performance (solid symbols): number born alive at first parity (NBA P1; circle), pre-weaning mortality at second parity (PWM P2; diamond), number of stillborn at third parity (NSB P3; triangle), pre-weaning mortality at third parity (PWM P3; circle crossed), and number of piglets mummifies at third parity (NSB PSY; inverted triangle).

region defined by training population without CG 3. The TGVM explained by the MHC region for region SSC 7 25–26 Mb was 26.06% (PPI = 1.00; CGs 2 and 3 in the training dataset), 23.94% (PPI = 1.00; CGs 1 and 3 in the training dataset), and 30.22% (PPI = 1; CGs 1 and 2 in the training dataset), whereas for the region SSC 7, 23–26 Mb was 10.0% (PPI = 0.95; CGs 1 and 2 in the training dataset).

GPAs were moderate to high, ranging from 0.39 for REST (BayesB) to 0.59 for ALL (BayesB). GPAs were higher for BayesB than for BayesC0 for scenarios ALL and MHC, but not for REST. This may be explained by the major QTL on the MHC region identified for all training population and was excluded in the scenario REST. We observed a large standard deviation for the MHC scenario when using BayesC0, much larger than when using BayesB. The GPAs for training population without CGs 1, 2, and 3 were 0.71, 0.68, and 0.17, respectively, for BayesC0, and were 0.58, 0.68, and 0.51, respectively, for BayesB. Thus, selection and training of markers using CGs 1 and 2 and validating CG 3 did not work well using BayesC0. Among scenarios, ALL showed better accuracy than did MHC and REST for BayesB, whereas ALL and MHC were comparable for BayesC0. Although there was substantial

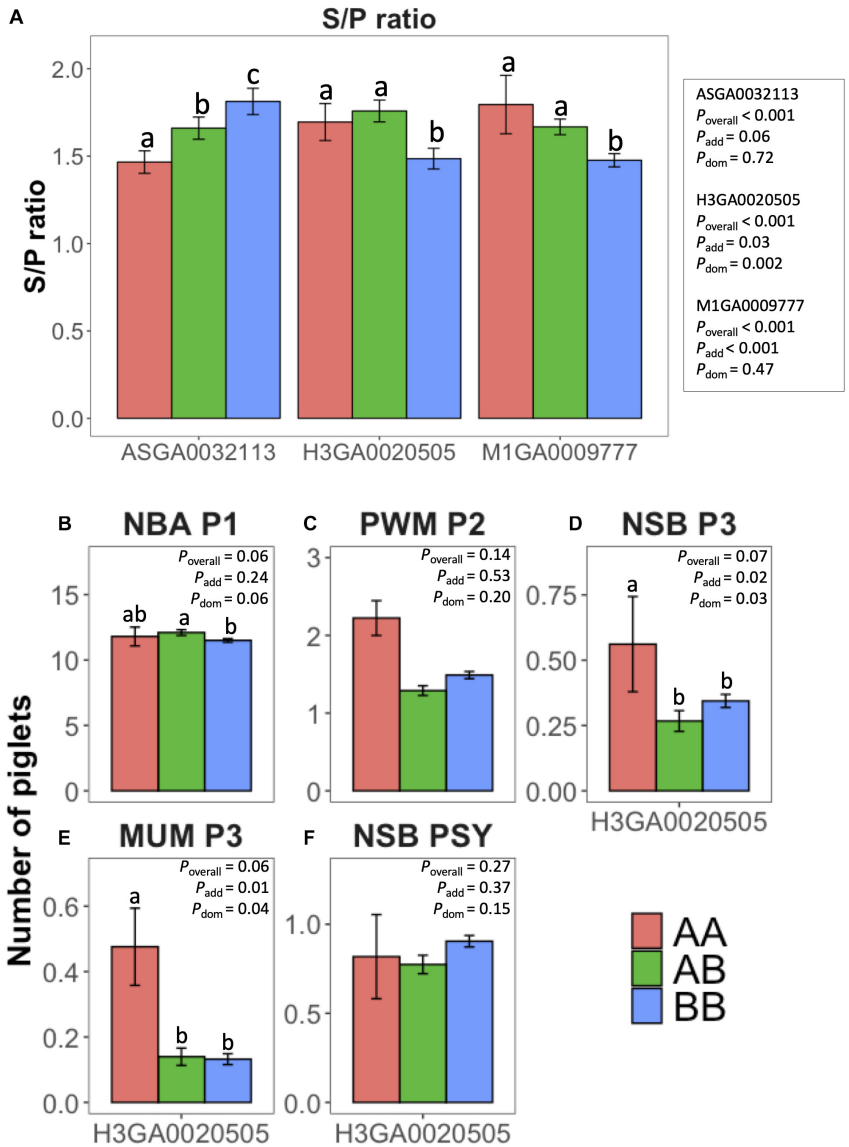


FIGURE 2 | Effect of selected single-nucleotide polymorphisms (SNPs) from univariate [antibody response (S/P ratio) to porcine reproductive and respiratory syndrome virus vaccination] and bivariate (farrowing performance) genome-wide association studies. *y*-axis represents least squares means for **(A)** S/P ratio, **(B)** number born alive at first parity (NBA P1), **(C)** pre-weaning mortality at second parity (PWM P2), **(D)** number of stillborn (NSB) at third parity (NSB P3), **(E)** number of piglets mummified at third parity (MUM P3), and **(F)** NSB per sow per year (NSB PSY). Colors represent the genotypes AA (coral), AB (green), and BB (blue) of the SNP evaluated: ASGA0032113, M1GA0009777, and H3GA0020505. Error bars represent standard errors. Different letters over the bars represent significant difference between the genotypes within a SNP. *P*-values for the overall *F*-test, additive effect, and dominance effect of each SNP are represented by $P_{overall}$, P_{add}, and P_{dom}, respectively.

decrease in GPA from MHC to REST for BayesB, this was not the case for BayesC0. In fact, overall, BayesC0 had similar GPAs across all scenarios.

DISCUSSION

Genetic Parameters

Heritabilities

The h^2 estimates reported for litter size traits in this study were slightly lower than estimates from literature using commercial F1 gilts (Ehlers et al., 2005; Strange et al., 2013). Nonetheless, these were low as expected, and within those previously reported in the literature when accounting for the SE of the estimates. The h^2 estimate for AFS was moderate but higher than previously shown in F1 (Large White × Yorkshire) gilts (0.16; Kumari and Rao, 2010). Our estimate was more similar to what has been reported for Landrace (0.31; Holm et al., 2005). The h^2 estimates for PSY traits were also low. Noppibool et al. (2017) reported h^2 of 0.07 ± 0.03 for NBA PSY and 0.13 ± 0.03 for NW PSY in purebred sows (Landrace and Yorkshire). Abell et al. (2013) reported an h^2 of 0.11 ± 0.01 in a study including Landrace,

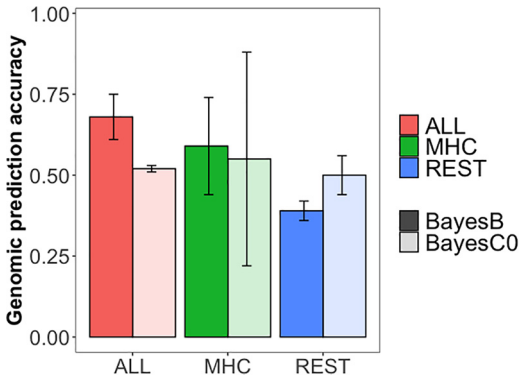


FIGURE 3 | Genomic prediction accuracies for antibody response [sample-to-positive (S/P) ratio] to porcine reproductive and respiratory syndrome virus vaccination. The colors indicate the set of single-nucleotide polymorphisms (SNPs) used in the analysis: whole dataset (ALL, coral), only the SNPs in the major histocompatibility complex region (chromosome 7, 23, to 26 Mb; MHC, green), and only SNPs outside the MHC region (REST, blue). The transparency indicates the method of estimation used: darker colors represent BayesB analysis (Pi = 0.995), and lighter colors represent BayesC0. Error bars represent the standard deviation across three cross-validation folds.

Large White, and F1 sows. To the best of our knowledge, there are no reports in the literature regarding the mortality traits PSY. In our study, the h^2 for FR was low but slightly higher to what has been reported by Bloemhof et al. (2015) of 0.10 ± 0.01 in a crossbred (Yorkshire × Landrace) population. For FI, the estimates of heritability in the literature are low (<0.10) in pure (Cavalcante Neto et al., 2009) and crossbred pig populations (Nagyné-kiszlinger et al., 2013), similar to what we have found. These results reaffirm that selection over litter size traits is challenging due to its low heritability.

In this study, F1 replacement gilts followed the standard acclimation procedures for the farms enrolled in our study. Moreover, these animals were sourced from high health multiplier herds and represent standard high-producing sows in the US swine industry. Therefore, our results are well representative of current genetics used for commercial production.

The number of studies investigating the genetic variation for antibody response to PRRSV has increased in the literature in the past few years. Studies vary on the age of animals (young and adult), type of exposure to PRRSV (vaccination, infection or both), and the assay used to measure antibodies. We measured S/P ratio ∼52 days after vaccination using a commercial ELISA test, when animals were ∼34 weeks old, and we obtained a moderate h^2 estimate for S/P ratio, with 0.34 ± 0.05. The first h^2 estimate for S/P ratio reported in the literature was by Serão et al. (2014), with 0.45 ± 0.13 in multiparous purebred sows during a PRRS outbreak. These authors measured S/P ratio at about 46 days after the PRRS outbreak using the same ELISA test utilized in our study. Also using purebred pregnant sows under a PRRS outbreak, Putz et al. (2019) reported a much lower h^2 estimate, of 0.17 ± 0.05 during the PRRS outbreak. In their study, antibody response to PRRSV was measured at about 60 days after the PRRS outbreak. There were three main differences between Serão et al. (2014) and Putz et al. (2019) that could be associated with different estimates of h^2 observed. First, the time of sample collection in Putz et al. (2019) was of about 2 weeks after the PRRS outbreak than in Serão et al. (2014), which could indicate a different immunological response of the animals (further discussion below). Second, Putz et al. (2019) used a different method of antibody measurement, based on microsphere (or microbead) assay, in contrast to using a standard commercial ELISA test. Finally, all animals in Putz et al. (2019) were re-exposed to the PRRSV before (∼30 days) serum samples were collected to measure antibody. However, Hickmann et al. (2019) reported moderate h^2 estimates of S/P in Duroc (0.33 ± 0.06) and Landrace (0.28 ± 0.07) sow populations during a PRRS outbreak, measured about 54 days after the outbreak using the same ELISA test used in our study. As in Putz et al. (2019), these animals were re-exposed to the PRRSV prior to collection of samples to measure antibody response. Therefore, low h^2 estimate found in Putz et al. (2019) for S/P ratio is likely associated with the differences in diagnostic assays.

Differently than in the studies described above, in our study, we investigated the relationship between S/P ratio and vaccination with reproductive performance in non-infected sows. Serão et al. (2016), using the same PRRS ELISA test as in our study, reported h^2 estimates ranging from 0.28 ± 0.04 to 0.47 ± 0.06, as the proportion of seroconverted animals increased in the dataset; however, there was no confirmation on whether the replacement gilts in were PRRSV-vaccinated or PRRSV-infected or even both. More similar to our study, Abella et al. (2019) estimated h^2 for S/P ratio at 42 days after PRRSV vaccination and obtained a higher h^2 estimate than we observed in our study (0.69 ± 0.10). Although they used the same ELISA test as us, the animals in their study were growing pigs at 6–7 weeks of age, and it is possible that the redirect of energy being used to growth for antibody production may have affected future performance. Another important point to be highlighted is that in the study by Abella et al. (2019), the herds were endemic for PRRS, and, therefore, the animals were in constant health challenge during vaccination and farrowing. Therefore, our study is the first one, to the best of our knowledge, to estimate h^2 for S/P ratio in PRRS-vaccinated F1 replacement gilts.

Several environmental factors may be involved in the different h^2 estimates observed across the different studies, such as the time of sample collection, the assay used for antibody measurement, the strains affecting the population, breeds, and genetic background, among others. The difference observed between our studies and PRRS outbreak studies may be the higher viral load during an outbreak compared with vaccination, although the use of MLV stimulates a similar immune response when compared with natural infection (Ellingson et al., 2010). This may result in a stronger immune response, which exacerbates the genetic variability between the individuals. The humoral response to PRRSV is well known to be delayed (Montaner-Tarbes et al., 2019). Therefore, measuring antibody response of about 6–8 weeks after infection, as used in all these studies, seems reasonable to capture the genetic variability for antibody response to PRRSV (Lopez and Osorio, 2004).

Although slightly lower than expected, the h^2 estimate for S/P ratio in this study was still substantially high, indicating that genetic selection to change response to PRRSV vaccine is possible. Nonetheless, assessment of h^2 at different time points after PRRSV vaccination is needed to identify how S/P ratio response changes genetically across time.

Genetic Correlations of Sample-to-Positive Ratio With Reproductive Traits

In this study, we evaluated the relationship of S/P ratio in PRRSV-vaccinated gilts with their subsequent reproductive performance up to three parities. In general, results depended on the parity being analyzed. Phenotypic correlations were consistently close to 0, indicating that relationships between S/P ratio and performance may exist at the genetic and environmental levels depending on the parity and trait analyzed.

In our study, however, S/P ratio had high and moderate positive genetic correlations with NBA P1 (0.60) and TNB P1 (0.30), respectively. The low negative r_g between S/P ratio with MUM P1 (-0.17) and NSB P1 (-0.05) may explain the lower r_g between S/P ratio with TNB P1 compared with the r_g with NBA P1, which indicates that selection for increased S/P ratio would increase NBA and decrease mortality traits at parity 1. Also, S/P ratio was highly negatively correlated with MUM P3 (-0.83) and NSB P3 (-0.84), indicating that the use of S/P ratio as a selection tool would have positive long-term effects. For the r_g estimates between S/P ratio with NSB P3 (-0.84) and MUM P3 (-0.83), although the standard errors for these estimates were low, these estimates were inconsistent with the one for NBD (-0.19), since NBD is calculated as a function of NSB and MUM. The h^2 estimates for NSB P3 (0.001) and MUM P3 (0.003) were very low in our study and, thus, have very low genetic variances for these traits in comparison with their respective phenotypic variances. Therefore, although highly genetically correlated with S/P ratio, the very low h^2 estimates for these traits indicate that there is limited genetic improvement of these traits. On the other hand, the posterior probabilities of these r_g estimates being smaller than zero were of 1 (data not shown), indicating that there is a negative genetic correlation between S/P ratio with NSB P3 and MUM P3.

Although not directly comparable, our r_g estimates of S/P ratio with NBA and NSB are similar to those reported by Serão et al. (2014), at 0.73 and -0.72, respectively, suggesting that the proposition of using S/P ratio as an indicator trait for reproductive performance in PRRSV-infected sows should be more general, as this relationship seems to also hold between S/P ratio in vaccinated gilts and reproductive performance in the absence of PRRSV infection. Putz et al. (2019) also reported a similar estimate of r_g between S/P with NSB, of -0.73 during a PRRS outbreak but not for NBA, which was positive but very low (0.05). Abella et al. (2019) also investigated the relationship between S/P ratio in vaccinated young gilts and subsequent reproductive performance in endemic herds. In contrast to our results, they found a small but positive phenotypic correlation between antibody response to PRRSV vaccination and NSB when looking at extreme phenotypes for response to PRRSV vaccination (resistant and susceptible gilts). Nonetheless, it is important to highlight that, among other differences between the two studies, their study was based on a phenotypic relationship and not genetic. We also tried to investigate the extreme phenotypes based on S/P ratio and associate them with reproductive performance, but it was not significant for any of the traits investigated (data not shown).

In addition to these results, we also investigated the relationship of S/P ratio with AFS, FR, and PSY, which had not yet been reported in the literature. The r_g between S/P ratio and AFS was moderate and negative (-0.25) while null with FR, indicating that selection for increased S/P ratio would result in younger gilts being serviced. Analysis of PSY traits has the advantage of taking in consideration the interval between farrow events, which also reflects the overall farrowing efficiency of the sows. In general, the r_g estimates between S/P ratio with PSY trait were low to moderate, with exception of NSB, which was high and negative (-0.90), in accordance with the r_g estimate between S/P ratio and NSB P3 (-0.84).

In general, results in our study indicate that genetic selection for increased antibody response to PRRSV vaccination would indirectly increase the reproductive performance in commercial sows. In order to evaluate the impact of direct selection for increased S/P ratio, we calculated the indirect response to selection on reproductive performance using the genetic parameters estimated in this study. For all deterministic simulations, intensity of selection was set to 5% for traits with $r_g > 0.50$ with S/P ratio (**Figure 4**). The correlated response to selection shown to be more efficient than direct response to those traits, with about 8% for NBA P1 (**Figure 4A**) and 72, 53, 49, and 77% more efficient for PWM P2, NSB P3, MUM P3, and NSB PSY (**Figure 4B**). Thus, selection for increased S/P ratio would result in substantial favorable gains for several reproductive traits. It is important to note that these calculations of efficiency of indirect response to selection when selecting for S/P ratio are simplistic, as in reality, animals are selected using an index with different economic weights. Therefore, a more comprehensive simulation is needed in follow up studies including costs associated with vaccination, measurements of S/P ratio, and genotyping. Nonetheless, the use of genomic selection allows breeding companies to use data on antibody response collected at the commercial levels to estimate breeding values for sires in the breeding herd with better antibody response to PRRS vaccination, which makes the proposed use of S/P ratio as an indicator trait for reproductive performance a feasible strategy.

Genome-Wide Association Studies

The major QTL identified in the MHC region explaining 30% of the TGVM for S/P ratio in our study is the same previously associated with S/P ratio in sows naturally infected with wild-type PRRSV by Serão et al. (2014) and validated by Serão et al. (2016) in an independent dataset in F1 replacement gilts and by Hickmann et al. (2019) in two outbreak herds. Haplotypes in this region have also been previously significantly associated with production traits, such as average daily gain and backfat thickness in non-infected pigs (Jung et al., 1989; Vaiman et al., 1998). The MHC region spans from around 23–26 Mb in the SSC 7 (Hammer et al., 2020). Two of the main SNPs associated with S/P ratio, ASGA0032113 and H3GA0020505, were located at

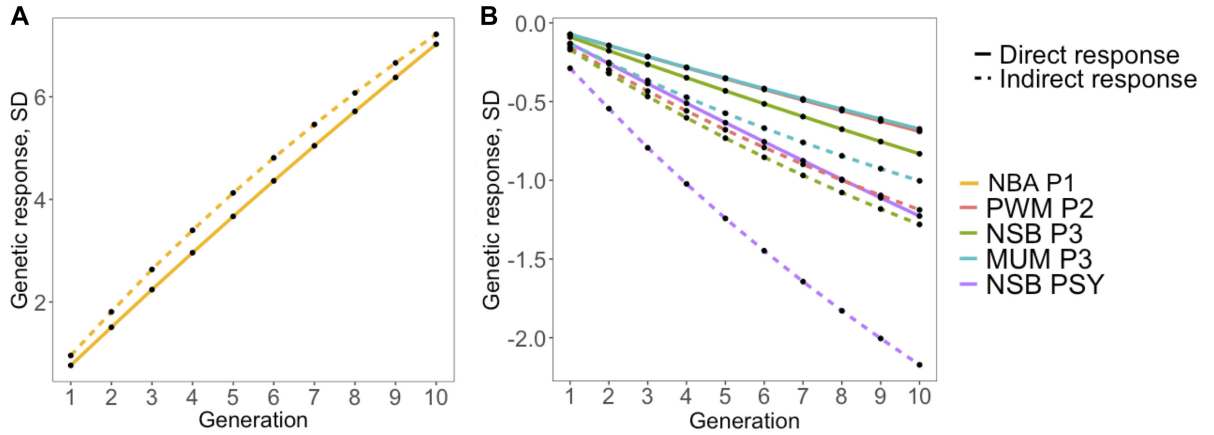


FIGURE 4 | Simulated response to selection for reproductive performance after 10 generations based (indirect) or not (direct) on antibody response, measured as sample-to-positive (S/P) ratio, to porcine reproductive and respiratory syndrome virus vaccination. The *y*- and *x*-axes represent response to selection in genetic standard deviations and generations, respectively. Direct and indirect response to selection is represented by solid and dashed lines, respectively, assuming 5% selection intensity. Colors represent **(A)** number born alive at first parity at first parity (NBA P1; golden), **(B)** pre-weaning mortality at second parity (PWM P2; coral), number of stillborn (NSB) at third parity (NSB P3; green), number of piglets mummified at third parity (MUM P3; blue), and NSB per sow per year (NSB PSY; purple).

25 Mb (MHC class II), while M1GA0009777 was located at 24 Mb (MHC class III).

The MHC region is well known for the presence of several immune-related genes with potential to be candidate causal genes associated with Ab response to PRRSV vaccination. The MHC class II harbors genes relating to peptide presentation of the adaptive immune system, such as swine leukocyte antigen (*SLA*)-*DR*, *SLA-DQ*, *SLA-DM*, and *SLA-DO* proteins; transporter-associated with antigen processing genes, such as ATP binding cassette subfamily B member transporter 1 (*TAP1*) and 2 (*TAP2*); and proteasomes, such as Proteasome 20S subunit beta 8 (*PSMB8*) and 9 (*PSMB9*) (Hammer et al., 2020). Interestingly, *SLA-DRA* has been shown to affect production and immune traits in swine (Lunney et al., 2009). In addition, Jiang et al. (2013) compared the lung transcriptome of PRRSV-infected gilts (at 6–8 weeks of age) and non-infected gilts and observed that *SLA-DRA*, *SLA-DRB1*, *SLA-DQA1*, *SLA-DQA2*, *SLA-DMB*, and *SLA-DOA* had significantly lower expression in the infected group compared with the non-infected group. *SLA-DRA*, *SLA-DRB1*, and *SLA-DQA1* are antigen binding genes, while *SLA-DMB* and *SLA-DOA* are involved in epitope loading of MHC class II molecules. Thus, all these genes are strong potential candidate genes since their lower expression may contribute to the attenuation of the immune response by the virus, which prolongs the PRRSV infection.

Similarly, the MHC class III region includes important genes for immune defense mechanisms and inflammation, including the tumor necrosis factor gene families, such as the tumor necrosis factor (*TNF*), lymphotoxin alpha (*LTA*), and lymphotoxin beta (*LTB*), components of the complement cascade, such as complement C2 (*C2*), complement C4A (*C4A*), and complement factor beta (*CFB*); heat shock proteins, such as *HSP1A*, *HSP1B*, and *HSP1L*; and genes with complex functions, such as tenascin XB (*TNXB*) and notch receptor 4 (*NOTCH4*). Some of them are potential candidate causal genes given their previous association with PRRSV. Among them, *C2* has been associated with PRRSV susceptibility and was downregulated in pregnant sows with higher antibody response at 35 days post PRRSV vaccination (Yang et al., 2018). In addition, *TNF* and heat shock proteins had significantly lower expression in the PRRSV-infected group compared with the non-infected group (Jiang et al., 2013). *C2* and *C4A* are involved in the activation of non-specific immune response (Vaiman et al., 1998), and their activation is associated with the mode of action of non-neutralizing antibody during PRRSV infection (Montaner-Tarbes et al., 2019). Heat shock proteins interact with viral protein and enhance the development of innate and adaptative system (Jiang et al., 2013). Therefore, the low expression of such genes may be associated with the weakened immune response observed during PRRSV infection.

Therefore, several genes located in this region are very likely to be associated with the strong relationship between the humoral immune response to PRRSV exposure (vaccination or natural infection) and the MHC region. However, in our study, we used SNP data to perform associations between them and variation in S/P ratio. Although we obtain strong and somewhat narrow associations, including the identification of specific SNPs explaining the majority of the variation, our analyses do not allow us to make any sort of cause and effect. Therefore, additional studies are needed to further investigate the causal variants that could be explained by the SNPs identified in our studies.

Interestingly, in the bivariate GWAS, the major region associated with the covariance between S/P ratio and all reproductive performance was the MHC class II. Recently, a study with Landrace and Large White during a PRRS outbreak observed that sows carrying specific genotypes of the *TAP1* gene, located in this region, were more PRRS resilient than others (Laplana et al., 2020). In this study, female piglets at 6–7 weeks of age were PRRSV-vaccinated and, after a period, underwent a PRRS outbreak (Laplana et al., 2020). The resilience was

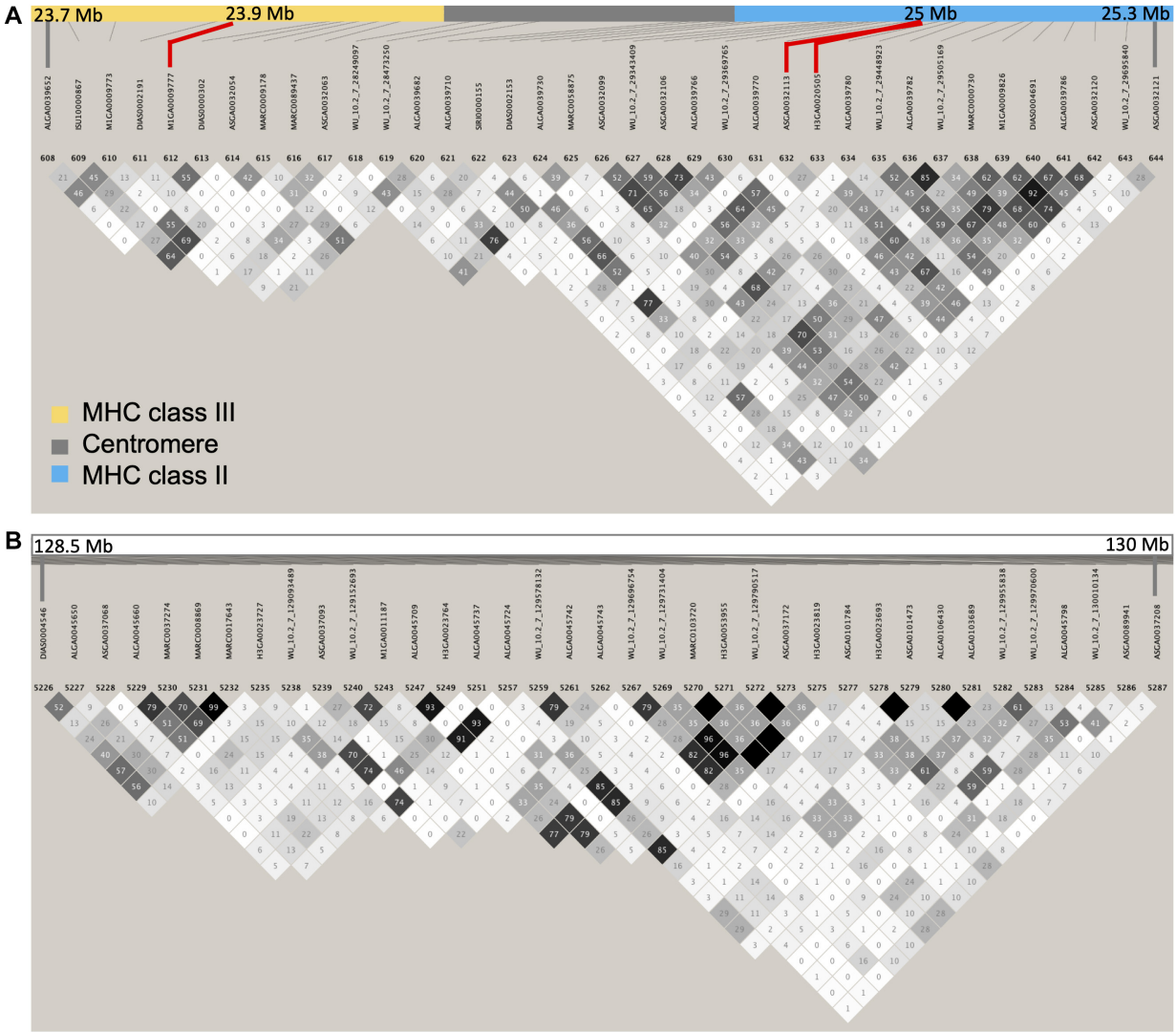


FIGURE 5 | Linkage disequilibrium (LD) map. Plots represent LD map of the data used in our study for the **(A)** major histocompatibility complex (MHC) region (chromosome 7: 23–26 Mb) and **(B)** the quantitative trait locus (QTL) previously identified for sample-to-positive (S/P) ratio by Serão et al. (2014, 2016). The SNPs explaining greater proportion of the genetic variance for S/P ratio to porcine reproductive and respiratory syndrome virus vaccination (ASGA0032113, M1GA0009777, and H3GA0020505) and reproductive performance (H3GA0020505) are shown in red in panel **(A)**. The MHC regions class II (∼24.7–25.3 Mb) and class III (∼23.6–24.3 Mb) are in blue and yellow, respectively, separated by the centromere (∼24.3–24.7 Mb) being shown in gray. Differently than in Serão et al. (2014), our region on SSC 7 between 128.5 and 130 Mb (B) for their proposed QTL showed overall low LD.

measured as the capacity to maintain reproductive performance, such as NBA, during the outbreak (Laplana et al., 2020). Additionally, a polymorphism in the Hydroxysteroid 17-Beta Dehydrogenase 8 (*HSD17B8*) gene, located at 25.2 Mb, has been associated with reproductive traits in pigs (Ma et al., 2015). Haplotypes in the MHC class I and II regions have been previously associated with reproductive traits, such as ovulation rate, embryo development, and litter size in non-infected pigs (Vaiman et al., 1998). These results suggest that genes in the MHC region seem to have a direct effect on reproductive performance, and it is possible that these genes are in LD with genes controlling the immune response to PRRSV (Gautschi and Gaillard, 1990). Therefore, some of those candidate genes involved in the immune response to PRRSV and reproductive performance might be physically close to each other, which generates the high r_g between these two traits in our study and in Serão et al. (2014). On the other hand, the average LD within the MHC region (SSC 7: 23–26 Mb) was of r^2 = 0.44, and the LD map shows very little overall LD in this region (**Figure 5A**), similarly to the one depicted by Serão et al. (2014), suggesting that the genetic correlation between S/P ratio and reproductive performance could be due to pleiotropy instead. The LD within a region may vary in different populations, which may explain, along with other environmental variations, the different results being obtained across the different studies. Finally, the MHC region was not identified in the univariate GWAS for reproductive

performance (data not shown), nor any other major regions. Farrowing performance traits are explained by several loci with low effect along the genome; therefore, the use of bivariate analysis may be advantageous since measurements of one trait can be informative for other traits (Cheng et al., 2018a).

Interestingly, the other major QTL found by Serão et al. (2014, 2016) on SSC 7, called Mb 130, was not identified in our study. According to Serão et al. (2014), this region is in high LD and harbors genes associated with immune response, such as TNF receptor associated factor 3 (*TRAF3*), which is involved in the innate immune response and induces NF-kappa-B activation. We also constructed an LD map of this region (**Figure 5B**), but contrarily to Serão et al. (2014), we observed a low LD in this region. Hence, one possible explanation for the fact that we did not identify the region Mb 130 in our study may be due to SNPs on that region not being in LD with the QTL. In addition, although this region was identified in the overall GWAS in Serão et al. (2016), this region was not identified for all analyses when data on each one of the seven breeding companies were analyzed separately (NVL Serão, *personal communication*). Similarly, Hickmann et al. (2019) also did not find this region in multiparous Duroc and Landrace sows during a PRRS outbreak, using population of animals sourced from the same breeding company used in our study. Thus, another possible reason for not identifying this region may be that this region does not segregate in all commercial swine populations.

Summarizing, the major QTL identified in this study for vaccination was the same as the one previously identified in PRRSV-infected animals, supporting that the genetic control to vaccination may be similar to infection. This is reasonable, as we used an MLV vaccine, which has an attenuated form of the PRRSV. Also, the high genetic correlation between S/P ratio to PRRSV vaccination and reproductive performance, which was partially explained by this QTL, could be due to LD between genes controlling immune response and reproductive performance, or pleiotropy in the MHC region.

Effect of Major Histocompatibility Complex Single-Nucleotide Polymorphisms on Antibody Response and Reproductive Traits

The main SNPs associated with S/P ratio were responsible for most of the variance accounted by the window, and after fitting them simultaneously as fixed effects in the model, the percentage of the variance explained by that window dropped to about null, supporting that these SNPs are playing a major role explaining the variability in S/P ratio. These three SNPs were not the same as the ones identified in Serão et al. (2014). Six out of 10 SNPs in the MHC region and in the Mb 130 region (unmapped in the current version of the SNP map) identified in Serão et al. (2014) were present in our dataset (positions based on the 11.1 assembly): ASGA003186 (SSC 7–22 Mb), MARC0058875 (SSC 7–24.8 Mb), ASGA0032151 (SSC 7–25.9 Mb), BGIS000074 (unmapped), MARC0037274 (unmapped), and ASGA0037093 (unmapped). We also evaluated their impact on S/P ratio by fitting all of them simultaneously as fixed effects in the model, along with the three SNPs identified in our analysis, but none of these six SNPs from Serão et al. (2014) were significantly ($P \geq 0.14$) associated with S/P ratio in our dataset (data not shown).

The SNP H3GA0020505 was associated with S/P ratio and farrowing performance in the bivariate analysis. Although this was the SNP explained most of variance for all traits analyzed, after fitting it as a fixed effect in the statistical model, this SNP tended to be associated with NBA, MUM, and NSB. The AA genotype was favorable for S/P ratio, NBA, and NSB but not for MUM, being the heterozygous genotype (i.e., AB) the only simultaneously favorable for all these three traits. Assuming that greater S/P ratio is favorable, this SNP seems to have a complete dominance mode-of-action for the A allele, and the highest values of S/P ratio could be obtained with AA and AB genotypes. However, the AA genotype was unfavorable for MUM P3, with the other two genotypes (AB and BB) having the highest performance. For this trait, the dominance seems to be the opposite than for S/P ratio, with the B allele dominating the A allele. The biological explanation for this is beyond what can be concluded using SNP data, but in case of pleiotropy, the transcripts for of AB genotypes could be the only ones participating on complex pathways that favor both traits simultaneously. This same rationale could be used for PMW P2 and NSB P3, where AB and BB genotypes had similar and favorable performances, although no significant association between H3GA0020505 and these two traits were identified. The trend for these two traits was clear (i.e., AA > AB = BB) and similar to the that for MUM P3, but the lack of significant associations for PWM P2 and NSB P3 could be due to the high standard error of the AA genotype, which can be explained by its very low frequency in this population. The frequency of the AA, AB, and BB genotypes were 0.074, 0.464, and 0.462, respectively. The consistent superiority of the heterozygote across all of these traits facilitates selection for improved performance in this situation. Since these F1 gilts were generated from two maternal lines (Landrace and Large White), the selection for opposite alleles could be performed for each breed in order to result in 100% heterozygote F1 animals for this locus. Investigation of genotypic frequencies in the purebred populations used to create these F1 gilts indicates that this locus is in Hardy–Weinberg equilibrium (data not shown), suggesting that this novel locus has not yet been selected for.

We further explored the impact of H3GA0020505 on the relationship between S/P ratio and these traits discussed in this section. After fitting this SNP as fixed effect in the model, the covariance explained by the window containing this SNP dropped from 34 to 26% and 90 to 74% for NBA P1 and NSB PSY, respectively, but not for the other traits. The genetic correlation between these two traits and S/P ratio also decreased for NBA, from 0.61 to 0.50 for NBA P1, supporting the results that this SNP plays a major role in the relationship between these two traits. In addition, we observed that the genetic variances for S/P ratio and reproductive performance, and the genetic covariances between these traits decreased as well. Although this happened for all the traits, the genetic correlations between S/P ratio with PWM P2, NSB P3, and MUM P3 did not decrease (or had just a slight decrease; data not shown), showing that the remaining

SNPs were capable of maintaining the relationship between these traits. Altogether, our results provided potential genetic markers (i.e., SNP) that could be explored in marker-assisted selection schemes to increase S/P ratio and farrowing performance in commercial sows.

Genomic Prediction

The GPAs were moderate to high, showing that genomic information is able to accurately predict S/P ratio in commercial gilts vaccinated for PRRS. GPAs for ALL and MHC were similar and greater than for REST. Interestingly, results for REST using BayesC0 were greater than using BayesB. These results are in accordance with the GWAS results, in which the MHC region plays an important role in the prediction of S/P ratio and using only SNPs in the region promotes a substantial GPA. In contrast, no regions outside the MHC were identified in the GWAS, supporting the result using BayesC0 for REST, since this method assumes that no major QTLs control the trait being analyzed.

The use of the MHC region for selection should be taken with caution, since lower genetic variability in this region may be associated with limited immune response and, consequently, impaired productive/reproductive performance (Vaiman et al., 1998). If that is the case, it is valid to note that even after removing the MHC region for genomic prediction, the GPA was still moderate, indicating the possibility of not using the MHC region to promote genetic improvement for S/P ratio. Nonetheless, studies in the literature evaluating the impact of direct selection for changes in the swine MHC are scarce, and thus, additional studies are needed to better understand the negative impact that this selection may cause. Mallard et al. (1998) created lines of pigs with high and low immune response based on estimated breeding values for antibody and cellular immune response. Selection for high immune response improved the resistance to specific infectious pathogens and increased the weight gain without altering the *SLA-DRA* gene expression (Mallard et al., 1998). These results indicate that it is possible to select for lines based on immune response, and it would have an effect on the immune response to infectious diseases and productive performance. Interestingly, the heterozygote genotype of the SNP associated with antibody response and farrowing performance (i.e., H3GA0020505) had the most favorable performance for these traits, suggesting that selection of parents with opposite alleles may also be an alternative to improve these traits. Indeed, it has been previously shown that in humans and pigs, MHC compatibility in parents was associated with impaired pregnancy (Gautschi and Gaillard, 1990). This brings great possibilities for selection within line in the nucleus to obtain opposite homozygotes in maternal lines to create heterozygote F1 replacement gilts for improved reproduction at the commercial level. However, the impact of this selection on the performance of commercial three-cross hogs should also be evaluated to verify whether this locus has any impact on economically important traits in hogs.

Our results are slightly different from those reported by Serão et al. (2016). These authors performed analyses using phenotype and genotype data on replacement gilts to train markers and used two populations for validation: the purebred population under PRRS outbreak in Serão et al. (2014) and the same F1 population used to train markers in a seven-fold cross-validation. In Serão et al. (2016), GPAs using ALL were generally similar to using only markers in the MHC region, such as results found in our study. However, in our study, we found a much lower reduction in GPA for REST (0.50 ± 0.06) compared with ALL (0.52 ± 0.01) and MHC (0.55 ± 0.33), in contrast to in Serão et al. (2016), who reported a GPA of 0.12 for REST and 0.31 and 0.28 for ALL and MHC, respectively, when validating on crossbred animals. This must be due to the fact that in our study we used animals from the same breeding company, whereas in Serão et al. (2016), they used animals from different breeding companies. The greater genetic relatedness in our study should also be the main reason for overall greater GPAs compared with those in Serão et al. (2016). In fact, our results are more applicable to the industry than those of Serão et al. (2016), as selection is performed using animals from the same breeding company. Nonetheless, the use of vaccination versus infection could also help in explaining these differences. In general, these results indicate that the use of genomic information would be efficient to estimate genomic breeding values for S/P ratio.

CONCLUSION

In this study, we showed that S/P ratio to PRRSV vaccination is moderately heritable and has favorable genetic correlation with reproductive performance in commercial pigs, especially with NBA at first parity, NSB and MUM in the third parity, and NSB per sow per year. A major QTL on the MHC classes II and III region explained most of the genetic variance of S/P ratio; and, more specifically, the region on MHC class II was associated with S/P ratio and reproductive performance, simultaneously. Three SNPs (ASGA0032113, H3GA0020505, and M1GA0009777) in these regions explained the vast majority of the genetic variance for S/P ratio within the MHC. In addition, SNP H3GA0020505 was associated with S/P ratio and reproductive performance, with the heterozygote genotype yielding the most favorable performance across these traits. Finally, the accuracy of genomic prediction was fairly high when using all SNPs available and only those located in the MHC region. Altogether, these results indicate that genetic selection for increased S/P ratio after PRRS MLV vaccination would result in indirect response for genetic improvement of farrowing performance in commercial sows. However, the improvement of reproductive performance may not be observed in all parities. Nonetheless, the majority of the genetic correlations that were not high were close to zero and/or in the right direction. Thus, genetic selection for S/P ratio might not generate unfavorable response in the reproductive performance in other parities. Future work is needed to validate our results for S/P ratio to PRRSV vaccination and its relationship with farrowing performance. In addition, it is necessary to evaluate if this genetic relationship exists in different populations

(i.e., different breeding source), while evaluating additional time points for collection of S/P ratio and using data on sows with greater number of parities (i.e., >3).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are not publicly available. Data may be available from authors upon request and authorization from the company that generated the data.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee at Iowa State University (IACUC# 6-17-8551-S).

AUTHOR CONTRIBUTIONS

LS performed the data analyses, interpreted the results, and drafted the manuscript. KG, JD, MN, and NS developed the research project. LS, RF, JD, and NS conceived the statistical analyses. LS and NS prepared the first draft of the manuscript. KG and DL coordinated the collection of materials. MN and DL coordinated the antibody measurement analysis. All the authors contributed to the final manuscript, read and approved the final manuscript.

FUNDING

This study was funded by the Smithfield Premium Genetics, the Iowa Pork Producers Association, and the ISU Graduate College.

REFERENCES

Abell, C. E., Mabry, J. W., Dekkers, J. C. M., and Stalder, K. J. (2013). Relationship between litters per sow per year sire breeding values and sire progeny means for farrowing rate, removal parity and lifetime born alive. *J. Anim. Breed. Genet.* 130, 64–71. doi: 10.1111/j.1439-0388.2012.01012.x

Abella, G., Novell, E., Tarancon, V., Varona, L., Pena, R. N., Estany, J., et al. (2019). Identification of resilient sows in porcine reproductive and respiratory syndrome virus–infected farms1. *J. Anim. Sci.* 97, 3228–3236. doi: 10.1093/jas/skz192

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: a fresh approach to numerical computing. *SIAM Review*, 59, 65–98.

Bloemhof, S., Kause, A., Knol, E. F., Van Arendonk, J. A. M., and Misztal, I. (2015). Heat stress effects on farrowing rate in sows: genetic parameter estimation using within-line and crossbred models 1. *J. Anim. Sci.* 90, 2109–2119. doi: 10.2527/jas2011-4650

Cavalcante Neto, A., Lui, J. F., Sarmento, J. L. R., Ribeiro, M. N., Monteiro, J. M. C., Fonseca, C., et al. (2009). Estimation models of variance components for farrowing interval in swine. *Brazillian Arch. Biol. Technol.* 52, 69–76. doi: 10.1590/s1516-89132009000100009

Cheng, H., Fernando, R., and Garrick, D. (2018a). *JWAS: Julia Implementation of Whole-Genome Analyses Software. Proceedings of the World Congress on Genetics Applied to Livestock Production, 11.859*. Auckland, New Zealand.

Cheng, H., Kizilkaya, K., Zeng, J., Garrick, D., and Fernando, R. (2018b). Genomic prediction from multiple-trait bayesian. *Genetics* 209, 89–103. doi: 10.1534/genetics.118.300650/-/DC1.1

Ehlers, M. J., Mabry, J. W., Bertrand, J. K., and Stalder, K. J. (2005). Variance components and heritabilities for sow productivity traits estimated from purebred versus crossbred sows. *J. Anim. Breed. Genet.* 122, 318–324. doi: 10.1111/j.1439-0388.2005.00533.x

Ellingson, J. S., Wang, Y., Layton, S., Ciacci-Zanella, J., Roof, M. B., and Faaberg, K. S. (2010). Vaccine efficacy of porcine reproductive and respiratory syndrome virus chimeras. *Vaccine* 28, 2679–2686. doi: 10.1016/j.vaccine.2009.12.073

Garrick, D. J., and Fernando, R. L. (2013). Implementing a QTL detection study (GWAS) using genomic prediction methodology. *Methods Mol. Biol.* 1019, 275–298. doi: 10.1007/978-1-62703-447-0-11

Gautschi, C., and Gaillard, C. (1990). Influence of major histocompatibility complex on reproduction and production traits in swine. *Anim. Genet.* 21, 161–170. doi: 10.1111/j.1365-2052.1990.tb03221.x

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., and Thompson, R. (2015). *ASReml User Guide Release 4.1 Functional Specification*, VSN International Ltd., Hemel Hempstead, United Kingdom.

Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186. doi: 10.1186/1471-2105-12-186

Hammer, S. E., Ho, C., Ando, A., Rogel-gaillard, C., Charles, M., Tector, M., et al. (2020). Importance of the MAJOR HISTOCOMPATIBILITY COMPLEX (Swine Leukocyte Antigen) in swine health and biomedical research. *Annu. Rev. Anim. Biosci.* 8, 1–14.

Hickmann, F. M. W., Braccini, J., Kramer, L., Sanglard, L. P., Gray, K. A., Huang, Y., et al. (2019). "Host-genomic scan for total antibody response during a PRRSV outbreak in purebred sows," in *PRRS Meeting*, Chicago, IL.

Holm, B., Bakken, M., Vangen, O., and Rekaya, R. (2005). Genetic analysis of age at first service, return rate, litter size, and weaning-to-first service interval of gilts and sows. *J. Anim. Sci.* 83, 41–48. doi: 10.2527/2005.83141x

Jiang, Z., Zhou, X., Michal, J. J., Wu, X. L., Zhang, L., Zhang, M., et al. (2013). Reactomes of porcine alveolar macrophages infected with porcine reproductive and respiratory syndrome virus. *PLoS One* 8:e0059229. doi: 10.1371/journal.pone.0059229

Jung, Y. C., Rothschild, M. F., Flanagan, M. P., Lauber, C. L., and Warner, C. M. (1989). Association of restriction fragment length polymorphisms of swine leucocyte antigen class I genes with production traits of Duroc and Hampshire boars. *Anim. Genet.* 20, 79–91. doi: 10.1111/j.1365-2052.1989.tb00845.x

Kinsella, R. J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., et al. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* 2011:bar030. doi: 10.1093/database/bar030

Kumari, B. P., and Rao, D. S. (2010). Effect of non-genetic factors on the reproductive traits in crossbred pigs. *J. Vet. Anim. Sci.* 6, 1–4. doi: 10.3923/rjnasci.2012.1.3

Laplana, M., Estany, J., Fraile, L. J., and Pena, R. N. (2020). Resilience effects of sgk1 and tap1 dna markers during prrsv outbreaks in reproductive sows. *Animals* 10, 1–11. doi: 10.3390/ani10050902

Lopez, O. J., and Osorio, F. A. (2004). Role of neutralizing antibodies in PRRSV protective immunity. *Vet. Immunol. Immunopathol.* 102, 155–163. doi: 10.1016/j.vetimm.2004.09.005

Lunney, J. K., Ho, C. S., Wysocki, M., and Smith, D. M. (2009). Molecular genetics of the swine major histocompatibility complex, the SLA complex. *Dev. Comp. Immunol.* 33, 362–374. doi: 10.1016/j.dci.2008.07.002

Ma, Y., Chen, N., Li, F., Fu, W., Han, Y., Chang, Y., et al. (2015). Bovine HSD17B8 gene and its relationship with growth and meat quality traits. *Sci. Bull.* 60, 1617–1621. doi: 10.1007/s11434-015-0882-0

Mallard, B. A., Wilkie, B. N., Kennedy, B. W., Gibson J., and Quinton, M. (1998). "Immune responsiveness in swine: eight generations of selection for high and low immune response in Yorkshire pigs," in *6th World Congress of Genetics Applied to Livestock Production*, Armidale, Australia, 257–264.

Montaner-Tarbes, S., del Portillo, H. A., Montoya, M., and Fraile, L. (2019). Key Gaps in the knowledge of the porcine respiratory reproductive syndrome virus (PRRSV). *Front. Vet. Sci.* 6:38. doi: 10.3389/fvets.2019.00038

Nagyné-kiszlinger, H., Farkas, J., Kövér, G., and Nagy, I. (2013). Selection for reproduction traits in hungarian pig breeding in a two-way cross. *Anim. Sci. Pap. Rep.* 31, 315–322.

Noppibool, U., Elzo, M. A., Koonawootrittriron, S., and Suwanasopee, T. (2017). Genetic relationships between length of productive life and lifetime production ef fi ciency in a commercial swine herd in Northern Thailand. *Anim. Sci. J.* 88, 213–221. doi: 10.1111/asj.12647

Purcell, S., Neale, B., Todd-brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Putz, A. M., Schwab, C. R., Sewell, A. D., Holtkamp, D. J., Zimmerman, J. J., Baker, K., et al. (2019). The effect of a porcine reproductive and respiratory syndrome outbreak on genetic parameters and reaction norms for reproductive performance in pigs. *J. Anim. Sci.* 97, 1101–1116. doi: 10.1093/jas/sky485

Serão, N. V. L., Kemp, R. A., Mote, B. E., Willson, P., Harding, J. C. S., Bishop, S. C., et al. (2016). Genetic and genomic basis of antibody response to porcine reproductive and respiratory syndrome (PRRS) in gilts and sows. *Genet. Sel. Evol.* 48:51. doi: 10.1186/s12711-016-0230-0

Serão, N. V. L., Matika, O., Kemp, R. A., Harding, J. C. S., Bishop, S. C., Plastow, G. S., et al. (2014). Genetic analysis of reproductive traits and antibody response in a PRRS outbreak herd. *J. Anim. Sci.* 92, 2905–2921. doi: 10.2527/jas.2014-7821

Strange, T., Ask, B., and Nielsen, B. (2013). Genetic parameters of the piglet mortality traits stillborn, weak at birth, starvation, crushing, and miscellaneous in crossbred pigs 1. *J. Anim. Sci.* 91, 1562–1569. doi: 10.2527/jas2012-5584

Vaiman, M., Chardon, P., and Rothschild, M. F. (1998). Porcine major histocompatibility complex. *Rev. Sci. Tech.* 17, 95–107.

Yang, T., Zhang, F., Zhai, L., He, W., Tan, Z., Sun, Y., et al. (2018). Transcriptome of porcine PBMCs over two generations reveals key genes and pathways associated with variable antibody responses post PRRSV vaccination. *Sci. Rep.* 8, 1–12. doi: 10.1038/s41598-018-20701-w

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 23 September 2020
doi: 10.3389/fmicb.2020.555730

Identification and Analysis of the Tegument Protein and Excretory-Secretory Products of the Carcinogenic Liver Fluke *Clonorchis sinensis*

Yunliang Shi[1,2], Kai Yu[3], Anli Liang[4], Yan Huang[1], Fangqi Ou[1], Haiyan Wei[1], Xiaoling Wan[1], Yichao Yang[1], Weiyu Zhang[3] and Zhihua Jiang[1]**

[1] Institute of Parasitic Disease Control and Prevention, Guangxi Zhuang Autonomous Region Center for Disease Control and Prevention, Nanning, China, [2] Guangxi Key Laboratory for the Prevention and Control of Viral Hepatitis, Guangxi Zhuang Autonomous Region Center for Disease Control and Prevention, Nanning, China, [3] College of Animal Science and Technology, Guangxi University, Nanning, China, [4] Xiangsihu College of Guangxi University for Nationalities, Nanning, China

OPEN ACCESS

Edited by:
Diego Robledo,
The University of Edinburgh,
United Kingdom

Reviewed by:
Adam Dowle,
University of York, United Kingdom
David Timson,
University of Brighton,
United Kingdom
Russell M. Morphew,
Aberystwyth University,
United Kingdom

***Correspondence:**
Yunliang Shi
syunliang2008@126.com
Zhihua Jiang
gxcdcjzh@163.com

Specialty section:
This article was submitted to Infectious Diseases, a section of the journal Frontiers in Microbiology

Received: 26 April 2020
Accepted: 26 August 2020
Published: 23 September 2020

Citation:
Shi Y, Yu K, Liang A, Huang Y, Ou F, Wei H, Wan X, Yang Y, Zhang W and Jiang Z (2020) Identification and Analysis of the Tegument Protein and Excretory-Secretory Products of the Carcinogenic Liver Fluke Clonorchis sinensis. Front. Microbiol. 11:555730. doi: 10.3389/fmicb.2020.555730

Liver fluke proteins, including excretory-secretory products (ESPs) and tegument proteins, are critical for the pathogenesis, nutrient metabolism, etiology and immune response of liver cancer. To understand the functions of various proteins in *Clonorchis sinensis* physiology and human clonorchiasis, the ESPs and tegument proteins of *C. sinensis* were identified. Supernatants containing ESPs from adult *C. sinensis* after culture for 6 h were harvested and concentrated. The tegument was detached using a freeze/thaw method and successively extracted using various extraction buffers. The outer surface proteins of *C. sinensis* were labeled with biotin, and the biotinylated proteins were purified. The ESP, tegument and labeled outer surface proteins were identified and analyzed by high-resolution LC-MS/MS. The identified proteins were compared with those of other flukes, and the protein functions associated with pathogenesis, carcinogenesis and potential vaccine antigens and drug targets were predicted and analyzed. A total of 175 proteins were identified after the 6-h culture of *C. sinensis* ESPs. A total of 352 tegument proteins were identified through sequential solubilization of the isolated teguments, and a subset of these proteins were localized to the surface membrane of the tegument by labeling with biotin. Thirty identified proteins, including annexins, actin and tetraspanins, were identified as potential immunomodulators and promising vaccine antigens. Interestingly, among the 352 tegument proteins, as many as 155 were enzymes, and most were oxidoreductases, hydrolases or transferases. A comparison of the outer surface proteins of *C. sinensis* with those of other flukes indicated that flukes have some common outer surface proteins, such as actin, tetraspanin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and annexin. Granulin, thioredoxin peroxiredoxin, carbonyl reductase 1 and cystatin were identified in the *C. sinensis* proteome and predicted to be related to liver

disease and cancer. The analysis of the *C. sinensis* proteome could contribute to a more in-depth understanding of complex parasite-host relationships, improve the diagnosis of clonorchiasis and benefit research on the pathogenesis and development of novel interventions, drugs and vaccines to control *C. sinensis* infection.

Keywords: *C. sinensis*, proteomics, tegument, biotinylation, cholangiocarcinoma

INTRODUCTION

Clonorchis sinensis is a food-borne zoonotic parasite, and humans infected with the liver fluke *C. sinensis* constitute a serious public health problem in many parts of southeast Asia, including China, Korea, and Vietnam (Lun et al., 2005). Approximately 35 million people are infected worldwide. People infected with *C. sinensis* mainly digest raw fish contaminated with metacercaria. *C. sinensis* adults, which dwell in bile ducts, cause subclinical or clinical hepatic and biliary disease with biliary epithelial hyperplasia, periductal fibrosis, and cystic changes in the ducts and even facilitate the development of cholangiocarcinoma (CCA) (Lee et al., 1993; Olnes and Erlich, 2004; Fried et al., 2011). Both clinical observations and epidemiological evidence strongly implicate *C. sinensis* infection in the pathogenesis of CCA (Choi et al., 2006, 2010; Shi et al., 2017).

Excretory-secretory products (ESPs) are antigens that include a complex mixture of secreted proteins and some extracellular vesicles (EVs) secreted through the parasite's tegument and products from oral openings or the gut of the parasite. ESPs are involved in parasite–host interactions and play key roles in the migration of *C. sinensis* adults to intrahepatic bile ducts and in the development of liver damage and even CCA (Kim Y. J. et al., 2009; Pak et al., 2009, 2019). Knowledge of the components of ESPs has value in rendering a diagnosis (Kang et al., 2019; Kim et al., 2019). Identification of the composition of the ESPs could provide attractive materials for the identification of causative agent candidates and new drug targets. However, the proteomic profiles of the ESPs in *C. sinensis* remain obscure.

The tegument constitutes the outermost surface of the parasite and is composed of a syncytial layer. The outer surface proteins of the tegument are enriched in enzymes and glycans and thus enable the parasite to directly contact immune cells, antibodies and various cytokines. Tegument proteins are thought to play key roles in parasite reproduction, signal transduction, nutrition intake, pathogenesis and the modulation of host immune responses and are therefore promising new drug targets and vaccine candidates (Dalton et al., 2004; Jones et al., 2004; Loukas et al., 2007). Previous studies have identified tegument and outer surface proteins (labeled with biotin) of *Opisthorchis viverrini* (Mulvenna et al., 2010b), *Schistosoma mansoni* (van Balkom et al., 2005; Braschi and Wilson, 2006; Braschi et al., 2006; Sotillo et al., 2015), *Schistosoma japonicum* (Mulvenna et al., 2010a), *Streptococcus bovis* (de la Torre-Escudero et al., 2013), and *Fasciola hepatica* (Wilson et al., 2011; Ravidà et al., 2016), and some important proteins have been revealed to be useful for immunologic diagnosis, drug targeting and vaccine development (Mossallam et al., 2015; Chen et al., 2016). Here, we characterize both the ESPs and the tegument proteins of *C. sinensis* through tryptic digestion coupled with high-resolution LC-MS/MS and describe the proteins that have potential roles in pathogenesis and carcinogenesis and that can serve as potential vaccine antigens and drug targets.

MATERIALS AND METHODS

Preparation of *C. sinensis* ESP Proteins

Clonorchis sinensis ESPs were obtained using established methods (Kim Y. J. et al., 2009). In brief, *C. sinensis* metacercariae were obtained from naturally infected cyprinoid fish in Hengxian County, Guangxi Province, China. The fish were digested with pepsin-HCl and washed, and the metacercariae were collected and used to infect rats (*Mesocricetus auratus*) via stomach incubation. The rats were maintained at the animal research facility of Guangxi Disease Prevention and Control Center using protocols approved by the Guangxi Disease Prevention and Control Center Animal Ethics Committee. Fresh adult flukes were recovered from the bile ducts of euthanized rats infected for 40 days and washed several times with normal saline containing penicillin (200 U/ml) and streptomycin (200 U/ml). Dead or dying flukes were discarded, and the ESPs were prepared by incubating viable flukes in modified PBS (Invitrogen) containing penicillin (100 U/ml) and streptomycin (100 U/ml) at 37°C in 5% CO_2. Supernatants containing the ESPs were collected after 6 h, centrifuged at 12,000 × *g* and 4°C for 30 min to remove the parasites' eggs from the media, concentrated to 100–250 μg/ml using a 2-kDa spin concentrator (Sartorius, Germany) and aliquoted for storage at −80°C.

Preparation of Whole Tegument Proteins

The whole tegument (WT) proteins of *C. sinensis* were prepared as previously described for *O. viverrini* (Mulvenna et al., 2010b). Briefly, fresh adult flukes were obtained from the infected rats as described above. Healthy and undamaged flukes were used for tegument detachment using a freeze/thaw method, successively extracted with various extraction buffers and pooled together (Mulvenna et al., 2010b). The extractions were precipitated with three volumes of acetone at −20°C.

Preparation of Biotinylated Tegument Proteins

The biotinylated tegument proteins (BTs) of *C. sinensis* were prepared as previously described for *O. viverrini* (Mulvenna et al., 2010b). Briefly, undamaged flukes were labeled with 1 mM

sulfosuccinimidyl-6-[biotinamido]hexanoate (EZ-Link™ sulfo-NHS-LC-biotin, Pierce) in Hank's balanced salt solution (HBSS), and teguments were removed and solubilized. The extractions were combined and incubated with 240 μl of streptavidin-agarose beads (GE Healthcare, United Kingdom). The proteins bound to the streptavidin were then eluted with 2% SDS, the supernatants were combined, and the proteins were precipitated using three volumes of acetone at −20°C.

Fluorescence Microscopy-Based Observation of Surface Biotinylation

To observe the surface biotinylation of *C. sinensis*, the flukes were biotinylated with 1 mM EZ-Link sulfo-NHS-SS-biotin, as described previously, washed three times in HBSS, incubated with streptavidin-FITC for 30 min at room temperature and washed three times with HBSS. Unlabeled (no biotin) flukes incubated with streptavidin-FITC alone and biotin-labeled flukes not incubated with streptavidin-FITC were used to measure the autofluorescence and as negative controls. The samples were visualized using a Zeiss Axio Imager M2 ApoTome fluorescence microscope (Zeiss, Germany) equipped with an AxioCam MRN at 40 × magnification.

Protein Purification Prior to Digestion

The crude protein precipitate of *C. sinensis* was added to 1:50 (W/V) lysis buffer (2 mM EDTA, 8 M urea, 10 mM DTT and 1% protease inhibitor cocktail), sonicated for 1–2 min, and centrifuged at 13,000 × *g* and 4°C for 10 min to remove debris. The supernatant was collected and precipitated using three volumes of acetone for 3 h at −20°C. After centrifugation at 4°C and 12,000 × *g* for 10 min, the protein deposit was redissolved in urea buffer [8 M urea and 100 mM triethylammonium bicarbonate (TEA)]. The concentration of the protein was determined using a modified Bradford protein assay kit (Sangon Biotech, China) according to the manufacturer's instructions.

Trypsin Digestion and Peptide Desalting

For trypsin digestion, 10 μg of the sample proteins was reduced with 10 mM DTT at 37°C for 60 min and alkylated with 25 mM iodoacetamide (IAM) at room temperature for 30 min in darkness. The urea concentration of the protein sample was diluted to less than 2 M by adding 100 mM TEAB. The protein pool of each sample was digested overnight at 37°C with sequencing-grade modified trypsin at a protein:trypsin mass ratio of 50:1 and then subjected to further digestion for 4 h with trypsin at a ratio of 100:1. After trypsin digestion, the peptides were desalted on a Strata X SPE column and vacuum-dried.

High-Resolution LC-MS/MS Analysis

The experiment was performed with a NanoLC 1000 LC-MS/MS system using a Proxeon EASY-nLC 1000 coupled to a Thermo Fisher LTQ-Orbitrap Elite mass spectrometer. There is only one biological replicate of each fraction WT, BT, and ESP, and proteomic run of each fraction was performed only one. The trypsin-digested fractions were reconstituted in 0.1% FA, and 2 μg of each sample was directly loaded onto a reversed-phase precolumn (Acclaim PepMap®100 C18, 3 μm, 100 Å, 75 μm × 2 cm) with 100% solvent A (0.1 M acetic acid in water) at 5 μl/min. The peptides eluted from the trap column were loaded onto a reversed-phase analytical column (Acclaim PepMap® RSLC C18, 2 μm, 100 Å, 50 μm × 15 cm). The gradient program was as follows: 15 to 35% solvent B (0.1% FA in 98% ACN) over 45 min, 35 to 98% solvent B over 5 min and 98% solvent B for 5 min; a constant flow rate of 250 nl/min was maintained throughout the program. The eluent was sprayed with an NSI source with a 1.8-kV electrospray voltage and then analyzed by tandem mass spectrometry (MS/MS) using an LTQ-Orbitrap Elite instrument. The mass spectrometer was operated in the data-dependent mode, which involves automatically switching between MS and MS/MS. Full-scan MS spectra (from m/z 350 to 1800) were acquired with the Orbitrap at a resolution of 60,000. Ion fragments were detected at a resolution of 15,000, and the 20 most intense precursors were selected for HCD fragmentation at the collision energy of 38% in the MS survey scan with 45.0-s dynamic exclusion.

Database Searching and Data Processing

Using the Sequest software integration tool in Proteome Discoverer (version 1.3, Thermo Scientific), the MS/MS raw data were searched against a *C. sinensis* database downloaded from the UniProt database with additional contaminant and host proteins. Trypsin was chosen as the enzyme, and two missed cleavages were allowed. Carbamidomethylation (C) was set as a fixed modification, and oxidation (M) and acetylation in the N-Term were set as variable modifications. The searches were performed using a peptide mass tolerance of 10 ppm, a product ion tolerance of 0.02 Da, and a false discovery rate (FDR) of 5% against a decoy database, and single peptide matches were included in the lists of accepted protein identifications. Comparisons were made with the results obtained using two additional database searching software programs (Mascot and MaxQuant) with the same parameter settings.

Bioinformatics Analysis

The subcellular localization of the identified proteins was predicted using WoLF PSORT software, and the sequences were predicted using an updated version of PSORT/PSORT II in WoLF PSORT (Paul et al., 2007) and literature searches. The enriched proteins identified in a Gene Ontology analysis were classified by GO annotation. The identified proteins (UniProt ID) were annotated with the UniProt-GOA database[1] and then classified via Gene Ontology annotations based on three categories: biological processes, cellular components and molecular functions. For each category, a two-tailed Fisher's exact test was employed to test the enrichment of the differentially expressed proteins compared with all the identified proteins. Correction for multiple hypothesis testing was performed using

[1]http://www.ebi.ac.uk/GOA/

standard FDR correction methods such as the Benjamini–Hochberg procedure. Differences in the GO-enriched proteins with a corrected p-value < 0.05 were considered significant.

A pathway enrichment analysis was also conducted. The Encyclopedia of Genes and Genomes (KEGG)[2] was used to identify the enriched pathways based on a two-tailed Fisher's exact test to determine the enrichment of the identified proteins in each fraction compared with all the identified proteins. First, the KEGG online service tool KAAS[3] was used to annotate the proteins according to the description in the KEGG database. The annotation results were mapped to the KEGG pathway database using the KEGG online KEGG mapper service tool. Correction for multiple hypothesis testing was performed using standard FDR control methods and the Benjamini–Hochberg multiple correction procedure. Differences in pathway proteins with a corrected p-value < 0.05 were considered significant. The proteins were classified into hierarchical pathway categories according to terms in the KEGG website.

RESULTS

Proteins Identified Among the *C. sinensis* ESPs

Approximately 120 μg of *C. sinensis* ESPs was obtained by culturing worms for a period of 6 h in PBS. The proteomic characterization of this mixture yielded 175 unique proteins to be identified (**Supplementary Table 1**). Using peptide-spectrum matches (PSMs) as a guide for determining the relative abundance of the identified proteins (James et al., 2015), the top five most abundant proteins were found to be basement membrane-specific heparan sulfate proteoglycan core protein, glutathione transferase, dynein, retinal dehydrogenase 1 and myoglobin. Some proteases, such as glutathione transferase, dehydrogenase and cysteine protease, were also among the top 10 most common proteins in the samples. The prediction of the subcellular localization of the proteins showed that most of the proteins were located in the cytosol (68 proteins), nucleus (23 proteins), mitochondria (22 proteins), and extracellular space (19 proteins) (**Figure 1A**). Seventeen proteins were uncharacterized proteins with no known homology.

Identification of *C. sinensis* Tegument Proteins

The tegument of approximately 200 adult *C. sinensis* was stripped and sequentially extracted, and a proteomic analysis of these fractions yielded 352 unique protein read identifies (**Supplementary Table 2**). The values of the PSMs indicated that the top ten most abundant proteins were nesprin-1, dynein, myosin, basement membrane-specific heparan sulfate proteoglycan core protein, and voltage-dependent anion channel protein 2. The majority of the tegument proteins were predicted to be localized in the cytosol (36.9%), mitochondria (17.9%), extracellular space (11.9%), nucleus (11.3%), and plasma membrane (10.5%). Forty-two extracellular and 37 plasma membranes were identified (**Figure 1A**), and 17 proteins were uncharacterized proteins with no homology to other proteins in the GenBank database.

Identification of BTs After Labeling the Surface of Live Worms

The localization of biotin on the labeled worms, which was observed by fluorescence microscopy, showed a clear zone of labeling around the exterior of the worm, and the biotin did not penetrate into the tegument membrane or the gut membrane (**Figure 2**). Approximately 200 live and undamaged adult *C. sinensis* worms presented biotin labeling on their outer surface. Thirty proteins were identified among the biotin-labeled proteins, and according to the PSM values, the most abundant proteins were actin, 70-kDa heat shock protein, prostaglandin-H2, glutamate dehydrogenase, thioredoxin peroxidase, succinate dehydrogenase and the Ras-related protein Rab-27B. Fourteen cytoskeletal proteins were identified in the biotinylated tegument fraction, and these included glyceraldehyde-3-phosphate dehydrogenase (GAPDH), elongation factor 1-alpha, and prostaglandin-H2. Seven mitochondrial proteins, including glutamate dehydrogenase (NAD(P)+), voltage-dependent anion channel protein 2, succinate dehydrogenase and uridine phosphorylase, and five plasma membrane-associated proteins, including tetraspanin, solute carrier family 25, glycerol kinase, ADP/ATP carrier and sodium/glucose cotransporter 4, were also found in this fraction (**Table 1**).

The outer surface proteins of *C. sinensis* were compared with the identified outer surface proteins of *O. viverrini* (Mulvenna et al., 2010b), *S. mansoni* (van Balkom et al., 2005; Braschi and Wilson, 2006; Braschi et al., 2006; Sotillo et al., 2015), *S. japonicum* (Mulvenna et al., 2010a), *S. bovis* (de la Torre-Escudero et al., 2013), and *F. hepatica* (Wilson et al., 2011; Ravidà et al., 2016; **Table 1**), and actin, tetraspanin, GAPDH and annexin, among other proteins, were found in all of these flukes. The comparison of *C. sinensis* and *O. viverrini*, which have similar lifecycles, revealed that these worms have a number of the same outer surface proteins, such as actin, ADP/ATP carrier, NAD(P)+, GAPDH, pyruvate carboxylase, tetraspanin, glycerol kinase, thioredoxin peroxidase, succinate dehydrogenase and annexin (**Table 1**). However, some proteins, such as sodium/glucose cotransporter 4 and fatty-acid amide hydrolase 1, appeared to be specific to *C. sinensis* and were not found in *O. viverrini*, *S. mansoni*, *S. japonicum*, *S. bovis*, and *F. hepatica* (**Table 1**).

Comparison of the Identified ESPs, WT Proteins and BTs

The mass spectrometry data were submitted to ProteomeXchange and were given the accession number PXD019116. The subcellular localization of the identified proteins was predicted using WoLF PSORT software, and the sequences were predicted with an updated version of PSORT/PSORT II in WoLF PSORT. A more in-depth analysis based on the subcellular localization showed some differences

[2]https://www.kegg.jp/

[3]https://www.genome.jp/tools/kaas/

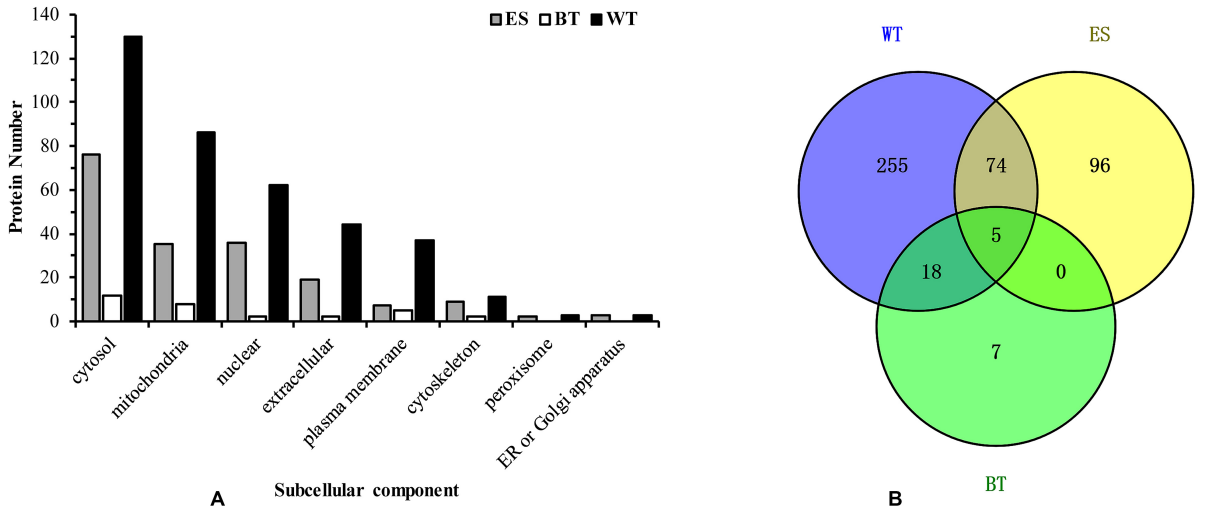


FIGURE 1 | Subcellular location of the *Clonorchis sinensis* proteins. Bar graph showing the abundance of excretory-secretory products, tegument proteins, and outer surface proteins at different subcellular locations **(A)**. Venn diagrams of common excretory-secretory product proteins, tegument proteins and outer surface proteins **(B)**.

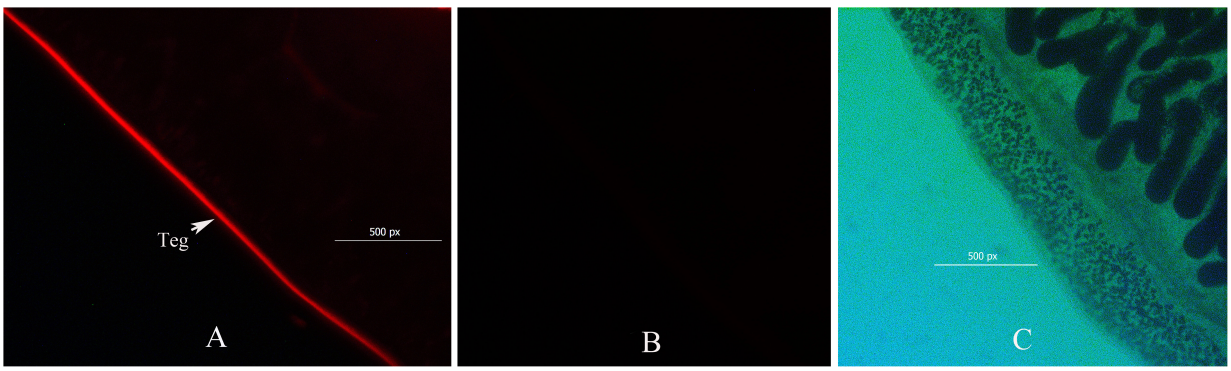


FIGURE 2 | Immunofluorescence micrograph showing the surface biotinylation of adult *Clonorchis sinensis*. Transformed worms were incubated with biotin and probed with streptavidin-FITC. The short-chain thiol-cleavable biotin was incorporated only into the outer surface of the tegument **(A)**. Streptavidin-FITC did not bind to the tegument of non-biotinylated worms **(B)**. Bright field showing worm biotin probed with streptavidin-FITC **(C)**.

in the expression profiles of the proteins in the ESP, WT, and BT fractions. Most of the ESP proteins are cytoplasmic, nuclear or extracellular proteins, whereas the WT and BT fractions mostly contain cytoplasmic, nuclear, mitochondrial and plasma membrane proteins (**Figure 1A**). A Venn diagram of the identified proteins in the ESP, WT and BT fractions was constructed. Among the identified proteins, 30, 352, and 175 were only found in the BT, WT, and ESP fractions, respectively, five proteins were identified in all three fractions, 23 proteins were found in both the BT and WT fractions, five proteins were found in both the BT and ESP fractions, and 79 proteins were found in both the WT and ESP fractions (**Figure 1B**).

Analyses of the Biological Processes and Pathways of the Identified Proteins

The GO annotation results showed that the ESPs, WT proteins and outer surface proteins were associated with different biological processes. The ESPs were found to be associated with complicated biological processes, whereas the WT and outer surface proteins were associated with substantially fewer biological processes, mainly cellular, metabolic and transmembrane transport processes (**Figure 3**). The identified proteins in the ESP fraction were predicted to be mainly associated with the following pathways: chemical carcinogenesis, metabolism of xenobiotics by cytochrome 450, drug metabolism-cytochrome P450, insulin signaling, glutathione metabolism and fructose and mannose metabolism. The analysis also revealed that the identified WT proteins were mainly associated with oxidative phosphorylation, valine/leucine and isoleucine degradation, biosynthesis of antibiotics, Huntington's disease and Alzheimer's disease, whereas the BTs were mainly associated with calcium, cGMP-PKG, Hippo signaling and the HTLV-I infection pathway (**Figure 4**).

TABLE 1 | Proteins identified from the *Clonorchis sinensis* biotinylated tegument preparation fraction.

GI	SC	CO	UP	Description	Ov	Sm	Sj	Sb	Fh
H2KRT8	27.89	40.69	1	Actin beta/gamma 1	+	+	+	+	+
G7YCA9	16.99	35.48	1	Actin beta/gamma 1					
H2KVT7	9.20	34.27	3	Prostaglandin-H2 D-isomerase	–	–	–	–	+
H2KPM6	0.00	9.65	1	Annexin A7	+	+	+	+	+
H2KPI5	1.84	5.28	1	Voltage-dependent anion channel protein	+	+	+	–	+
H2KPE2	3.97	21.07	1	Glyceraldehyde-3-phosphate dehydrogenase	+	+	+	+	–
H2KNU1	7.01	18.97	2	Tegument antigen	–	–	–	–	–
G7YTI6	0.00	15.60	1	Ras-related protein Rab-27B	–	–	–	–	–
G7YIE7	0.00	23.39	1	Heat shock 70-kDa protein	–	+	+	+	+
G7YD93	4.27	12.98	2	Fructose-bisphosphate aldolase	–	+	–	+	–
G7YBA9	0.00	12.77	1	Beta-2-syntrophin	–	–	–	–	–
G7YA24	1.90	8.97	1	Uridine phosphorylase	–	–	–	–	–
E2JF28	0.00	13.85	1	Thioredoxin peroxidase	+	–	+	+	–
1 WA65	6.19	14.10	2	Elongation factor 1-alpha	–	–	+	–	–
G7YK67	2.16	6.77	1	Fatty-acid amide hydrolase	–	–	–	–	–
G7YGH4	3.00	17.31	1	Uncharacterized protein	–	–	–	–	–
G7Y7T6	5.53	19.74	1	Glutathione S-transferase	–	–	+	+	–
G7YW97	73.79	32.46	1	Propionyl-CoA carboxylase alpha chain	–	–	–	–	+
G7YT72	4.74	12.92	1	Endophilin-B1	–	+	+	–	+
G7YS59	2.49	15.73	1	Pyruvate carboxylase	+	–	–	–	+
G7YQG7	0.00	14.29	2	Succinate dehydrogenase	+	–	–	–	+
G7YLC0	6.86	13.61	1	Glutamate dehydrogenase	+	–	–	+	+
G7YHC3	0.00	15.35	1	Uncharacterized protein	–	–	–	–	–
G7YFK3	10.36	12.52	2	Propionyl-CoA carboxylase beta chain	–	–	–	–	+
G7YFT8	2.06	21.43	1	Fatty acid-binding protein type 3	–	+	–	+	–
A7XWR4	14.68	21.19	2	ADP/ATP carrier	+	+	+	+	–
G7YXC3	3.26	6.52	1	Solute carrier family 25 member	–	–	–	–	–
G7YB65	2.21	8.21	1	Tetraspanin	+	+	+	+	+
H2KSG0	2.11	8.32	1	Glycerol kinase	+	–	–	–	–
H2KVD0	1.99	3.03	1	Sodium/glucose cotransporter 4	–	–	–	–	–

The abbreviations are defined as follows: GI, gene accession; UP, unique peptide; SC, spectral count; CO, percent coverage. The presence and presence of the indicated protein in the outer surface of O. viverrini (Mulvenna et al., 2010b), S. mansoni (van Balkom et al., 2005; Braschi and Wilson, 2006; Braschi et al., 2006; Sotillo et al., 2015), S. japonicum (Mulvenna et al., 2010a), S. bovis (de la Torre-Escudero et al., 2013), and F. hepatica (Wilson et al., 2011; Ravidà et al., 2016) is denoted with a "+" and "–," respectively.

DISCUSSION

The genome and transcriptome of *C. sinensis* have been revealed (Young et al., 2010; Wang et al., 2011b; Huang et al., 2013), and this knowledge greatly contributes to our understanding of different processes, such as parasite development, nutrition, immune evasion, and modulation. The completed *C. sinensis* genome was sequenced and annotated, and approximately 16,000 protein-coding gene models have been predicted (Wang et al., 2011b); in addition, a large number of secreted and tegument proteins have been identified and are available in GenBank, which makes the search for *C. sinensis* proteins accessible and reliable. Several reports have indicated that the ESPs of *C. sinensis* might play key pathogenic roles in the development of hepatic and bile duct diseases (Kim H. G. et al., 2009, Kim et al., 2019). However, few studies have examined the identities of *C. sinensis* ESPs. Zheng et al. (2011, 2013) identified 110 and 267 proteins in *C. sinensis* ESPs by LC–MS/MS after culture for 0–48 h (Zheng et al., 2011) and 0–6 h (Zheng et al., 2013), respectively. In the current study, we identified 175 ESPs, and the differences in the number of identified proteins might be associated with the protocol used for protein preparation, the protein search software program and the parameter settings. In our research, we used liquid enzyme hydrolysis, not gel electrophoresis, to effectively reduce the loss of protein. In the protein search, we set a rigorous quality error for the secondary spectrum. Because not all details of the proteins identified by Zheng et al. (2011) and Zheng et al. (2013) are included in the papers and published databases, we are unable to perform an in-depth comparison. However, some of the proteins, such as enolase, actin, heat shock protein, and Ras-related protein, were found in both the previous investigations and the current study.

The ESPs of parasites include molecules that might be useful in the development of diagnostics, vaccines, and drug therapies. In the current study, a total of 175 proteins were identified after 6 h of the *in vitro* culture of *C. sinensis*, and these included some important proteins, such as enolase, which is a cytosolic glycolytic enzyme that localizes to the cell surface and

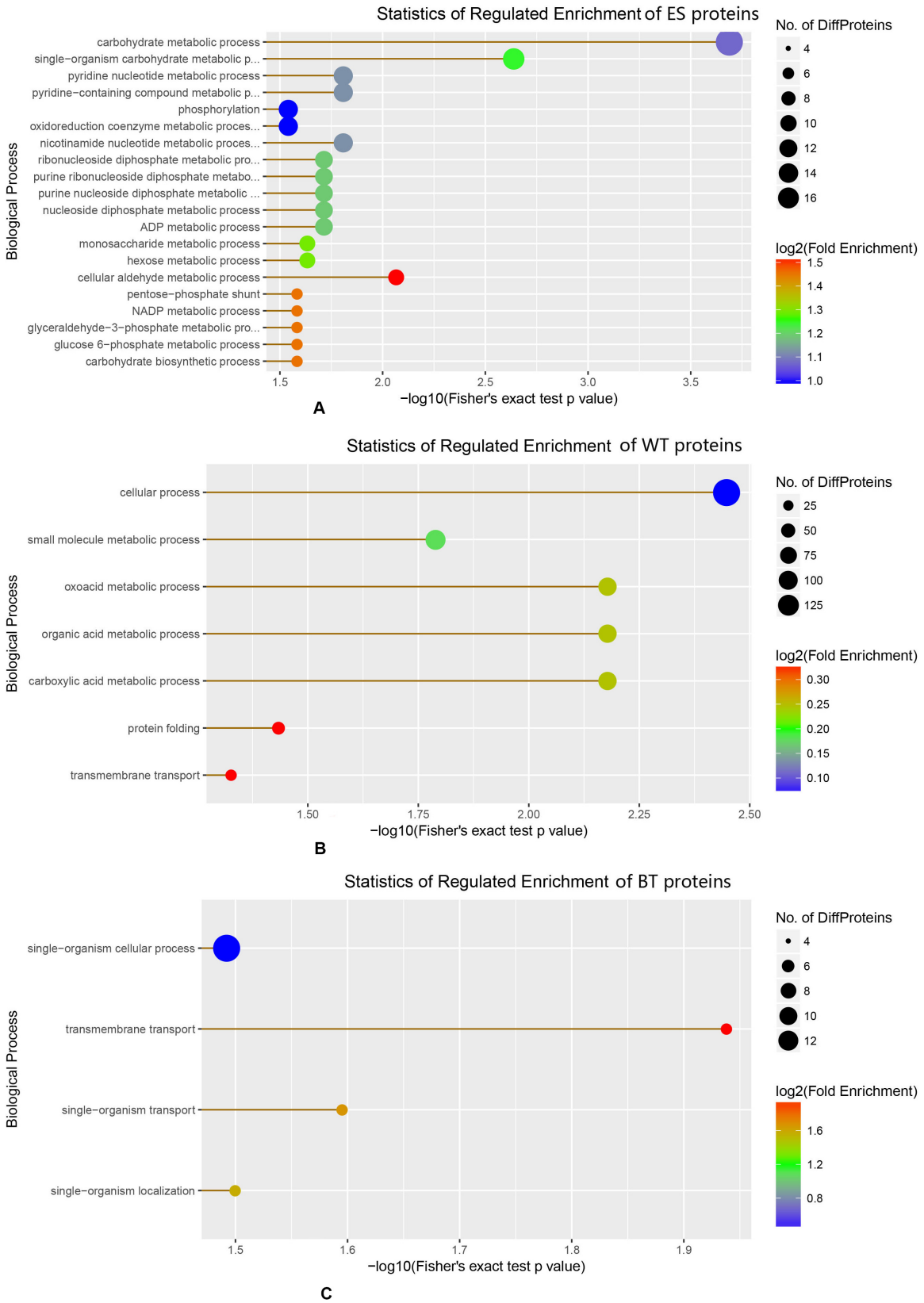


FIGURE 3 | Enrichment of different biological proteins among the excretory-secretory product, tegument protein, and outer surface protein fractions. **(A)** Statistics of regulated enrichment of excretory-secretory protein. **(B)** Statistics of regulated enrichment of tegument protein. **(C)** Statistics of regulated enrichment of biotinylated tegument protein.

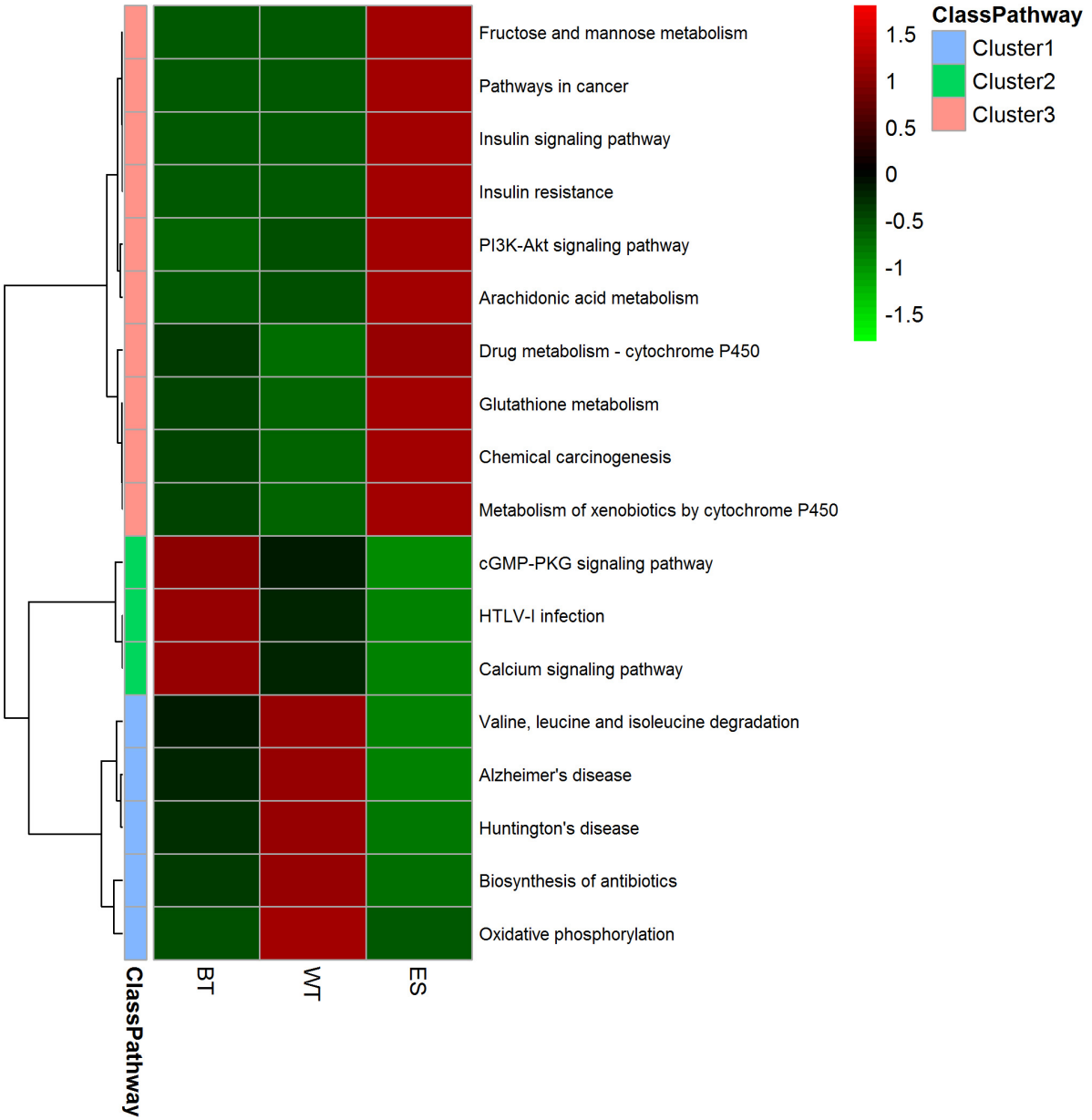


FIGURE 4 | Heatmap showing the pathway clusters of excretory-secretory products, tegument proteins and outer surface proteins. The abundance of the proteins is represented by different colors and compared among the three protein fractions. The proteins were grouped into 10 different Gene Ontology (GO) categories based on GO annotations and six cellular location categories predicted with UniLoc and literature searches. The clustering was performed using Euclidean distances.

tegument in helminths. In some schistosome species (*S. bovis* and *S. japonicum*), enolase has been identified at the host-parasite surface and serves as a PLMG activator (de la Torre-Escudero et al., 2010; Yang et al., 2010; Figueiredo et al., 2015). In *C. sinensis*, enolase is an important glycolytic enzyme required for parasite growth (Wang et al., 2011a), and knockdown of the enolase gene in *C. sinensis* increases worm mortality (Wang et al., 2014a). Cathepsins have been identified as good diagnostic antigens, in addition, it has been associated with nutrient uptake (Na et al., 2008; Ma et al., 2019), and several cysteine protease subtypes, such as cathepsin A, B, D, E, and F, have been found in *C. sinensis* (Wang et al., 2011b; Kang et al., 2019); After 6 h of *C. sinensis* culture, we identified cathepsin C in the ESP fraction, and cathepsin C was specifically distributed in the suckers (oral and ventral suckers), eggs, vitellarium, intestines, and testis of *C. sinensis*, and it plays roles in the digestion of host proteins, nutrition assimilation, and immune invasion. Furthermore, cathepsin C might be a potential diagnostic antigen and drug target against *C. sinensis* infection (Liang et al., 2014).

In the ESPs of flukes that should contain some EVs, EVs can be purified from the ESPs by differential ultracentrifugation, and the EVs of some trematodes have been identified from the secretory products (Marcilla et al., 2012; Chaiyadet et al., 2015; Davis et al., 2019). Previous studies have shown that EVs of trematodes contain some of the proteins previously identified as components of ESPs; specifically, more than half of the molecules identified in the secretomes of *Echinostoma caproni* and *F. hepatica* are found in these exosome-like vesicles (Marcilla et al., 2012). At present, the EVs of *C. sinensis* have not been studies, and we could not identify overlapping proteins between ESPs and EVs.

Previous studies have identified the tegument proteins of *O. viverrini* (Mulvenna et al., 2010b), *F. hepatica* (Wilson et al., 2011), *S. japonicum* (Mulvenna et al., 2010a), *S. mansoni* (van Balkom et al., 2005; Braschi and Wilson, 2006; Braschi et al., 2006; Sotillo et al., 2015), and *S. bovis* (de la Torre-Escudero et al., 2013). Using the techniques described, the WT of *C. sinensis* was removed by freeze-thawing and successively extracted with buffers of increasing solubilization strength, which enabled the enrichment of membrane-associated proteins, and proteins within and exposed at the tegument surface were identified. Of the 352 identified *C. sinensis* tegument proteins identified, many were proteases, and these proteases might be closely related to nutrient uptake, immune escape and migration. A cathepsin B-like cysteine proteinase was identified in *C. sinensis*, and a novel tegument cathepsin B was identified in *S. mansoni* (Caffrey et al., 2002). Cathepsin B-like cysteine proteinase can decompose the skin tissue and help the cercaria penetrate into skin (Wilson and Coulson, 2006). The cathepsin B-like cysteine proteinase in *C. sinensis* might facilitate larval migration. The identified enolases have been previously studied and are expressed during the adult worm, metacercaria, cercaria and egg life stages of *C. sinensis*. Enolases exhibit a significantly higher expression level at the adult stage and are deposited on the tegument of the adult worm and cyst wall of the metacercaria. Enolases might play key roles in the growth of parasites and might be promising oral vaccine candidates (Wang et al., 2011a, 2014b). Most of the identified proteases were oxidoreductases, hydrolases and transferases, which indicated that the tegument is key for nutrient uptake and energetic metabolism.

Whole fluke biotinylation is a technique that has been used to label and identify the outer surface of tegument proteins. The reagent used in this study (Sulfo-NHS-LC-biotin) can react with exposed primary amines and preferentially labels surface-exposed proteins. Previous studies used biotin labeling to identify the outer surface proteins of *O. viverrini* and *S. mansoni* (Mulvenna et al., 2010b; Sotillo et al., 2015). The outer surfaces of *O. viverrini* (Mulvenna et al., 2010b), *S. mansoni* (van Balkom et al., 2005; Braschi and Wilson, 2006; Braschi et al., 2006; Sotillo et al., 2015), *S. japonicum* (Mulvenna et al., 2010a), *S. bovis* (de la Torre-Escudero et al., 2013), and *F. hepatica* (Wilson et al., 2011; Ravidà et al., 2016) have been identified, and the number of identified outer surface proteins ranges from 22 to 45. Thirty proteins were identified from the biotin-labeled outer surface of *C. sinensis*, and various proteins, such as actin, tetraspanin, GAPDH and annexin, were found in *C. sinensis*, *O. viverrini*, *S. mansoni*, *S. japonicum*, *S. bovis*, and *F. hepatica* (**Table 1**), which indicates that these outer surface proteins of flukes might have similar functions. Two tetraspanins (TSP1 and TSP2) were identified in the outer surface of *S. mansoni* and *S. japonicum* and function in tegument renewal (Tran et al., 2010), and these have been found to be promising vaccine candidates to protect against schistosomiasis (Tran et al., 2010; Chen et al., 2016). Tetraspanin 2 was previously found in *C. sinensis* (Kim et al., 2011), but the function of tetraspanins in *C. sinensis* remains unknown. Annexin is another important protein identified in the *C. sinensis* tegument. In *S. mansoni*, annexin binds to the surface membranes of the tegument in a calcium-dependent manner and is considered a potential vaccine candidate (Tararam et al., 2010). Annexin B30, which is composed of four annexin repeats, was previously found in *C. sinensis* and is distributed in the tegument, intestine, and eggs of adult worms and in the tegument and vitellarium of metacercaria. This protein plays a role in adjusting the host immune response during *C. sinensis* infection (He et al., 2014).

Considering the different biological fluids in which each parasite resides, such as the blood for schistosomes and the bile ducts for *O. viverrini* and *C. sinensis*, the tegument proteins of these two different flukes are expected to show some differences, particularly in the composition of the proteins exposed to hosts, which might be relevant to the host immune response and to nutrient uptake and metabolism. Pyruvate carboxylase, succinate dehydrogenase and glycerol kinase were found only on the outer surface of *C. sinensis* and *O. viverrini* but not in *Schistosoma*. These three enzymes are associated with nutrition absorption and energy metabolism. Pyruvate carboxylase is an anaplerotic enzyme that plays important roles in various cellular metabolic pathways, including fatty acid synthesis, amino acid synthesis, gluconeogenesis and glucose-induced insulin secretion. This enzyme is involved in tumorigenesis in several cancers, including breast cancer, non-small cell lung cancer, glioblastoma, renal carcinoma, and gall bladder cancer (Lao-On et al., 2018). Glycerol kinase is a highly conserved enzyme that functions at the junction of lipid synthesis and carbohydrate metabolism. This transferase catalyzes the transfer of a phosphorus group from ATP to glycerol and is a critical enzyme at the junction of fat and carbohydrate metabolism. Previous studies have shown that *Plasmodium falciparum* glycerol kinase catalyzes the ATP-dependent phosphorylation of glycerol to glycerol-3-phosphate (Naidoo and Coetzer, 2013). *Trypanosome* glycerol kinase catalyzes not only the forward reaction (ATP-dependent glycerol phosphorylation) but also the reverse reaction. Glycerol kinase is important for trypanosome survival and might be a promising drug target (Balogun et al., 2017). Succinate dehydrogenase is a functional member of both the Krebs cycle and the aerobic respiratory chain. Previous studies have shown that *C. sinensis* succinate dehydrogenase has varying activities during treatments with pyquiton, bithionol, and menichlopholan, and these activities are associated with morphological alterations of the tegument of the flukes (Hamajima et al., 1979; Pang et al., 1990). The determination of whether these bile duct fluke-specific enzymes play a key role in *C. sinensis* or *O. viverrini* requires further study. Some of the identified surface

proteins of *C. sinensis*, such as Ras-related protein, beta-2-syntrophin, fatty-acid amide hydrolase, solute carrier family 25 and sodium/glucose cotransporter 4, were not found in other flukes and seem to be specific to *C. sinensis*. The functions of these outer surface proteins in *C. sinensis* have not been reported and are worth further research.

In this study, the proteins that might contribute to carcinogenesis were analyzed. Granulin, which is reportedly associated with carcinogenesis caused by *O. viverrini* (Smout et al., 2009), was identified in the *C. sinensis* tegument. In mammals, granulin is able to modulate cell growth (Shoyab et al., 1990; Culouscou et al., 1993; Zhou et al., 1993), and granulin is also a helminth-derived growth factor that causes the proliferation of mammalian cells (Smout et al., 2009, 2011). In addition, granulin was previously identified in the ESP supernatant fraction of *O. viverrini* after 24 h of culture (Mulvenna et al., 2010b). In *C. sinensis*, granulin was previously identified among the ESPs and is localized to the tegument and testes of the adult worm (Wang et al., 2017). In the current study, the absence of granulin in the supernatant obtained after 6 h of culture indicated that the protein is found at a low abundance among the ESPs. In addition to granulin, thioredoxin peroxiredoxin, carbonyl reductase 1 (CBR1) and cystatin might contribute to the carcinogenic process. Thioredoxin is a growth factor (Nguyen et al., 2006) that is overexpressed in many aggressive forms of cancer (Yoo et al., 2006). In *O. viverrini*, thioredoxin inhibits oxidative stress-induced apoptosis of bile duct epithelial cells, which indicates that liver fluke oxidoreductase might induce cholangiocarcinogenesis (Matchimakul et al., 2015). In *C. sinensis*, thioredoxin is localized to the tegument, vitelline gland, intestine, and intrauterine eggs of adult worms and is involved in the immunoregulation of the host immune response (Zhou et al., 2013), but the function of thioredoxin related to carcinogenesis remains unknown. Oxidoreductase-peroxiredoxin and carbonyl reductase 1 (CBR1) have been identified among the ESPs of *C. sinensis*, and both are overexpressed in liver cancer. The inhibition of CBR1 activity can enhance the effectiveness and decrease the cardiotoxicity of the anticancer drug anthracycline daunorubicin (DNR) in hepatocellular carcinoma (Huang et al., 2010). Cystatin is a superfamily of cysteine protease inhibitors that participates in various physiological and pathological processes. Cystatin stimulates interferon gamma-dependent nitric oxide production by macrophages (Vray et al., 2002), and two cystatins (Stifin-1 and Stifin-2) have been identified in *C. sinensis* (Kang et al., 2011, 2014). Cystatin might contribute to carcinogenesis at sites of inflammation via DNA damage and subsequent malignant transformation. Other proteins, such as FKBP12-rapamycin complex-associated protein, the molecular chaperone HtpG, phosphoenolpyruvate carboxykinase and growth factor receptor-binding protein 2, are predicted to play roles in the cancer pathway, but whether these proteins contribute to liver disease and liver cancer in *C. sinensis* infection needs further research. Liver fluke-induced CCA is a multifactorial pathological process associated with long-term infection and live fluke material-induced inflammation, and the release of carcinogenic substances by parasites is an important factor. The identified proteins will help elucidate the potentially important molecules and mechanisms of liver fluke-induced CCA.

CONCLUSION

In the current study, we identified the ESPs and tegument proteome of *C. sinensis*, and these data could contribute to a more in-depth understanding of the complex parasite-host relationship, improve the diagnosis of clonorchiasis and benefit research on the pathogenesis and development of novel interventions, drugs and vaccines to control *C. sinensis* infection.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by The Institutional Review Board (IRB) approval by Guangxi Institutional Review Board (GXIRB) (IRB00001594).

AUTHOR CONTRIBUTIONS

YS, WZ, and ZJ contributed to the study conception and design. YS, KY, AL, YY, YH, WZ, ZJ, FO, and HW contributed to the sample collection and the acquisition, analysis and interpretation of the data. YS and KY contributed to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by self-financed projects of the Health and Health Committee of Guangxi Zhuang Autonomous Region (Z20170136), the Guangxi Zhuang Autonomous Region Center for Disease Prevention and Control Young Science Foundation (201501), and Development and Application of Medical and Health Appropriate Technology of Guangxi (S2018090).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.555730/full#supplementary-material

Supplementary Table 1 | The information of the identified excretory-secretory proteins.

Supplementary Table 2 | The information of the identified tegument proteins.

REFERENCES

Balogun, E. O., Inaoka, D. K., Shiba, T., Tokuoka, S. M., Tokumasu, F., Sakamoto, K., et al. (2017). Glycerol kinase of African trypanosomes possesses an intrinsic phosphatase activity. *Biochim. Biophys. Acta* 1861, 2830–2842. doi: 10.1016/j.bbagen.2017.07.028

Braschi, S., Borges, W. C., and Wilson, R. A. (2006). Proteomic analysis of the schistosome tegument and its surface membranes. *Mem. Inst. Oswaldo Cruz* 101, 205–212. doi: 10.1590/s0074-02762006000900032

Braschi, S., and Wilson, R. A. (2006). Proteins exposed at the adult schistosome surface revealed by biotinylation. *Mol. Cell. Proteomics* 5, 347–356. doi: 10.1074/mcp.m500287-mcp200

Caffrey, C. R., Salter, J. P., Lucas, K. D., Khiem, D., Hsieh, I., Lim, K. C., et al. (2002). SmCB2, a novel tegumental cathepsin B from adult *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* 121, 49–61. doi: 10.1016/s0166-6851(02)00022-1

Chaiyadet, S., Sotillo, J., Smout, M., Cantacessi, C., Jones, M. K., Johnson, M. S., et al. (2015). Carcinogenic liver fluke secretes extracellular vesicles that promote cholangiocytes to adopt a tumorigenic phenotype. *J. Infect. Dis.* 212, 1636–1645. doi: 10.1093/infdis/jiv291

Chen, L., Chen, Y., Zhang, D., Hou, M., Yang, B., Zhang, F., et al. (2016). Protection and immunological study on two tetraspanin-derived vaccine candidates against schistosomiasis japonicum. *Parasite Immunol.* 38, 589–598. doi: 10.1111/pim.12338

Choi, D., Lim, J. H., Lee, K. T., Lee, J. K., Choi, S. H., Heo, J. S., et al. (2006). Cholangiocarcinoma and *Clonorchis sinensis* infection: a case–control study in Korea. *J. Hepatol.* 44, 1066–1073. doi: 10.1016/j.jhep.2005.11.040

Choi, D., Lim, J. H., Lee, K. T., Lee, J. K., Choi, S. H., Heo, J. S., et al. (2010). Gallstones and *Clonorchis sinensis* infection: a hospital-based case-control study in Korea. *J. Gastroenterol. Hepatol.* 23, e399–e404. doi: 10.1111/j.1440-1746.2007.05242.x

Culouscou, J. M., Carlton, G. W., and Shoyab, M. (1993). Biochemical analysis of the epithelin receptor. *J. Biol. Chem.* 268, 10458–10462.

Dalton, J. P., Skelly, P., and Halton, D. W. (2004). Role of the tegument and gut in nutrient uptake by parasitic platyhelminths. *Can. J. Zool.* 82, 211–232. doi: 10.1139/z03-213

Davis, C. N., Phillips, H., Tomes, J. J., Swain, M. T., Wilkinson, T. J., Brophy, P. M., et al. (2019). The importance of extracellular vesicle purification for downstream analysis: a comparison of differential centrifugation and size exclusion chromatography for helminth pathogens. *PLoS Negl. Trop. Dis.* 13:e0007191. doi: 10.1371/journal.pntd.0007191

de la Torre-Escudero, E., Manzano-Román, R., Pérez-Sánchez, R., Siles-Lucas, M., and Oleaga, A. (2010). Cloning and characterization of a plasminogen-binding surface-associated enolase from *Schistosoma bovis*. *Vet. Parasitol.* 173, 76–84. doi: 10.1016/j.vetpar.2010.06.011

de la Torre-Escudero, E., Pérez-Sánchez, R., Manzano-Román, R., and Oleaga, A. (2013). In vivo intravascular biotinylation of *Schistosoma bovis* adult worms and proteomic analysis of tegumental surface proteins. *J. Proteom.* 94, 513–526. doi: 10.1016/j.jprot.2013.09.020

Figueiredo, B. C., Da'dara, A. A., Oliveira, S. C., and Skelly, P. J. (2015). Schistosomes enhance plasminogen activation: the role of tegumental enolase. *PLoS Pathog.* 11:e1005335. doi: 10.1371/journal.ppat.1005335

Fried, B., Reddy, A., and Mayer, D. (2011). Helminths in human carcinogenesis. *Cancer Lett.* 305, 239–249. doi: 10.1016/j.canlet.2010.07.008

Hamajima, F., Fujino, T., Yamagami, K., and Eriguchi, N. (1979). Studies on the in vitro effects of bithionol and menichlopholan on flukes of *Clonorchis sinensis*, *Metagonimus takahashii* and *Paragonimus miyazakii*. *Int. J. Parasitol.* 9, 241–249. doi: 10.1016/0020-7519(79)90035-3

He, L., Ren, M., Chen, X., Wang, X., Li, S., Lin, J., et al. (2014). Biochemical and immunological characterization of annexin B30 from *Clonorchis sinensis* excretory/secretory products. *Parasitol. Res.* 113, 2743–2755. doi: 10.1007/s00436-014-3935-4

Huang, W., Ding, L., Huang, Q., Hu, H., Liu, S., Yang, X., et al. (2010). Carbonyl reductase 1 as a novel target of (-)-epigallocatechin gallate against hepatocellular carcinoma. *Hepatology* 52, 703–714. doi: 10.1002/hep.23723

Huang, Y., Chen, W., Wang, X., Liu, H., Chen, Y., Guo, L., et al. (2013). The carcinogenic liver fluke, *Clonorchis sinensis*: new assembly, reannotation and analysis of the genome and characterization of tissue transcriptomes. *PLoS One* 8:e54732. doi: 10.1371/journal.pone.0054732

James, A. M., Victor, F., Theresa, C., Steve, W., Yan, Y., Stephen, S., et al. (2015). Toward the complete characterization of host cell proteins in biotherapeutics via affinity depletions, LC-MS/MS, and multivariate analysis. *MAbs* 7, 1128–1137. doi: 10.1080/19420862.2015.1082017

Jones, M. K., Gobert, G. N., Zhang, L., Sunderland, P., and McManus, D. P. (2004). The cytoskeleton and motor proteins of human schistosomes and their roles in surface maintenance and host-parasite interactions. *BioEssays* 26, 752–765. doi: 10.1002/bies.20058

Kang, J. M., Ju, H. L., Lee, K. H., Kim, T. S., Pak, J. H., Sohn, W. M., et al. (2014). Identification and characterization of the second cysteine protease inhibitor of *Clonorchis sinensis* (CsStefin-2). *Parasitol. Res.* 113, 47–58. doi: 10.1007/s00436-013-3624-8

Kang, J. M., Lee, K. H., Sohn, W. M., and Na, B. K. (2011). Identification and functional characterization of CsStefin-1, a cysteine protease inhibitor of *Clonorchis sinensis*. *Mol. Biochem. Parasitol.* 177, 126–134. doi: 10.1016/j.molbiopara.2011.02.010

Kang, J. M., Yoo, W. G., Lê, H. G., Thái, T. L., Hong, S. J., Sohn, W. M., et al. (2019). Partial characterization of two cathepsin D family aspartic peptidases of *Clonorchis sinensis*. *Korean J. Parasitol.* 57, 671–680. doi: 10.3347/kjp.2019.57.6.671

Kim, H. G., Han, J., Kim, M. H., Cho, K. H., Shin, I. H., Kim, G. H., et al. (2009). Prevalence of clonorchiasis in patients with gastrointestinal disease: a Korean nationwide multicenter survey. *World J. Gastroenterol.* 15, 86–94. doi: 10.3748/wjg.15.86

Kim, J. G., Ahn, C. S., Sripa, B., Eom, K. S., Kang, I., Sohn, W. M., et al. (2019). *Clonorchis sinensis* omega-class glutathione transferases are reliable biomarkers for serodiagnosis of clonorchiasis and opisthorchiasis. *Clin. Microbiol. Infect.* 25, 109.e1–109.e6. doi: 10.1016/j.cmi.2018.03.042

Kim, T. Y., Chung, E. J., Sohn, W. M., Hong, S. H., and Yong, T. S. (2011). Molecular characterization of *Clonorchis sinensis* tetraspanin 2 extracellular loop 2. *Parasitol. Res.* 110, 707–711. doi: 10.1007/s00436-011-2546-6

Kim, Y. J., Choi, M. H., Hong, S. T., and Bae, Y. M. (2009). Resistance of cholangiocarcinoma cells to parthenolide-induced apoptosis by the excretory–secretory products of *Clonorchis sinensis*. *Parasitol. Res.* 104, 1011–1016. doi: 10.1007/s00436-008-1283-y

Lao-On, U., Attwood, P. V., and Jitrapakdee, S. (2018). Roles of pyruvate carboxylase in human diseases: from diabetes to cancers and infection. *J. Mol. Med.* 96, 237–247. doi: 10.1007/s00109-018-1622-0

Lee, J. H., Rim, H. J., and Bak, U. B. (1993). Effect of *Clonorchis sinensis* infection and dimethylnitrosamine administration on the induction of cholangiocarcinoma in Syrian golden hamsters. *Korean J. Parasitol.* 31, 21–30. doi: 10.3347/kjp.1993.31.1.21

Liang, P., He, L., Xu, Y., Chen, X., Huang, Y., Ren, M., et al. (2014). Identification, immunolocalization, and characterization analyses of an exopeptidase of papain superfamily, (cathepsin C) from *Clonorchis sinensis*. *Parasitol. Res.* 113, 3621–3629. doi: 10.1007/s00436-014-4027-1

Loukas, A., Tran, M., and Pearson, M. S. (2007). Schistosome membrane proteins as vaccines. *Int. J. Parasitol.* 37, 257–263. doi: 10.1016/j.ijpara.2006.12.001

Lun, Z. R., Gasser, R. B., Lai, D. H., Li, A. X., Zhu, X. Q., Yu, X. B., et al. (2005). Clonorchiasis: a key foodborne zoonosis in China. *Lancet Infect. Dis.* 5, 31–41. doi: 10.1016/s1473-3099(04)01252-6

Ma, C., Liang, K., Tang, L., He, S., Liu, X., He, M., et al. (2019). Identification and characteristics of a cathepsin L-like cysteine protease from *Clonorchis sinensis*. *Parasitol. Res.* 118, 829–835. doi: 10.1007/s00436-019-06223-y

Marcilla, A., Trelis, M., Cortés, A., Sotillo, J., Cantalapiedra, F., Teresa, M., et al. (2012). Extracellular vesicles from parasitic helminths contain specific excretory/secretory proteins and are internalized in intestinal host cells. *PLoS One* 7:e45974. doi: 10.1371/journal.pone.0045974

Matchimakul, P., Rinaldi, G., Suttiprapa, S., Mann, V. H., Popratiloff, A., Laha, T., et al. (2015). Apoptosis of cholangiocytes modulated by thioredoxin of carcinogenic liver fluke. *Int. J. Biochem. Cell Biol.* 65, 72–80. doi: 10.1016/j.biocel.2015.05.014

Mossallam, S. F., Amer, E. I., Ewaisha, R. E., Khalil, A. M., Aboushleib, H. M., and Bahey-El-Din, M. (2015). Fusion protein comprised of the two schistosomal antigens, Sm14 and Sm29, provides significant protection against *Schistosoma mansoni* in murine infection model. *BMC Infect. Dis.* 15:147. doi: 10.1186/s12879-015-0906-z

Mulvenna, J., Moertel, L., Jones, M. K., Nawaratna, S., Lovas, E. M., Gobert, G. N., et al. (2010a). Exposed proteins of the *Schistosoma japonicum* tegument. *Int. J. Parasitol.* 40, 543–554. doi: 10.1016/j.ijpara.2009.10.002

Mulvenna, J., Sripa, B., Brindley, P. J., Gorman, J., Jones, M. K., Colgrave, M. L., et al. (2010b). The secreted and surface proteomes of the adult stage of the carcinogenic human liver fluke *Opisthorchis viverrini*. *Proteomics* 10, 1063–1078. doi: 10.1002/pmic.200900393

Na, B. K., Kang, J. M., and Sohn, W. M. (2008). CsCF-6, a novel cathepsin F-like cysteine protease for nutrient uptake of *Clonorchis sinensis*. *Int. J. Parasitol.* 38, 493–502. doi: 10.1016/j.ijpara.2007.09.001

Naidoo, K., and Coetzer, T. L. (2013). Reduced glycerol incorporation into phospholipids contributes to impaired intra-erythrocytic growth of glycerol kinase knockout plasmodium falciparum parasites. *Biochim. Biophys. Acta Gen. Subj.* 1830, 5326–5334. doi: 10.1016/j.bbagen.2013.08.006

Nguyen, P., Awwad, R. T., Smart, D. D. K., Spitz, D. R., and Gius, D. (2006). Thioredoxin reductase as a novel molecular target for cancer therapy. *Cancer Lett.* 236, 164–174. doi: 10.1016/j.canlet.2005.04.028

Olnes, M. J., and Erlich, R. (2004). A review and update on cholangiocarcinoma. *Oncology* 66, 167–179. doi: 10.1159/000077991

Pak, J. H., Lee, J. Y., Jeon, B. Y., Dai, F., Yoo, W. G., and Hong, S. J. (2019). Cytokine production in cholangiocarcinoma cells in response to *Clonorchis sinensis* excretory-secretory products and their putative protein components. *Korean J. Parasitol.* 57, 379–387. doi: 10.3347/kjp.2019.57.4.379

Pak, J. H., Moon, J. H., Hwang, S. J., Cho, S. H., Seo, S. B., and Kim, T. S. (2009). Proteomic analysis of differentially expressed proteins in human cholangiocarcinoma cells treated with *Clonorchis sinensis* excretory-secretory products. *J. Cell. Biochem.* 108, 1376–1388. doi: 10.1002/jcb.22368

Pang, X., Li, B., and Yu, X. (1990). Histochemical changes in *Clonorchis sinensis* after pyquiton treatment. *Chin. J. Parasitol. Parasit. Dis.* 8, 32–34.

Paul, H., Keun-Joon, P., Takeshi, O., Naoya, F., Hajime, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259

Ravidà, A., Cwiklinski, K., Aldridge, A. M., Clarke, P., Thompson, R., Gerlach, J. Q., et al. (2016). *Fasciola hepatica* surface tegument: glycoproteins at the interface of parasite and host. *Mol. Cell. Proteomics* 15, 3139–3153. doi: 10.1074/mcp.m116.059774

Shi, Y., Jiang, Z., Yang, Y., Zheng, P., Wei, H., Lin, Y., et al. (2017). *Clonorchis sinensis* infection and co-infection with the hepatitis B virus are important factors associated with cholangiocarcinoma and hepatocellular carcinoma. *Parasitol. Res.* 116, 2645–2649. doi: 10.1007/s00436-017-5572-1

Shoyab, M., McDonald, V. L., Byles, C., Todaro, G. J., and Plowman, G. D. (1990). Epithelins 1 and 2: isolation and characterization of two cysteine-rich growth-modulating proteins. *Proc. Natl. Acad. Sci. U.S.A.* 87, 7912–7916. doi: 10.1073/pnas.87.20.7912

Smout, M. J., Laha, T., Mulvenna, J., Sripa, B., Suttiprapa, S., Jones, A., et al. (2009). A granulin-like growth factor secreted by the carcinogenic liver fluke, *Opisthorchis viverrini*, promotes proliferation of host cells. *PLoS Pathog.* 5:e1000611. doi: 10.1371/journal.ppat.1000611

Smout, M. J., Mulvenna, J. P., Jones, M. K., and Loukas, A. (2011). Expression, refolding and purification of Ov-GRN-1, a granulin-like growth factor from the carcinogenic liver fluke, causes proliferation of mammalian host cells. *Protein Expr. Purif.* 79, 263–270. doi: 10.1016/j.pep.2011.06.018

Sotillo, J., Pearson, M., Becker, L., Mulvenna, J., and Loukas, A. (2015). A quantitative proteomic analysis of the tegumental proteins from *Schistosoma mansoni* schistosomula reveals novel potential therapeutic targets. *Int. J. Parasitol.* 45, 505–516. doi: 10.1016/j.ijpara.2015.03.004

Tararam, C. A., Farias, L. P., Wilson, R. A., and Leite, L. C. C. (2010). *Schistosoma mansoni* annexin 2: molecular characterization and immunolocalization. *Exp. Parasitol.* 126, 146–155. doi: 10.1016/j.exppara.2010.04.008

Tran, M. H., Freitas, T. C., Cooper, L., Gaze, S., Gatton, M. L., Jones, M. K., et al. (2010). Suppression of mRNAs encoding tegument tetraspanins from *Schistosoma mansoni* results in impaired tegument turnover. *PLoS Pathog.* 6:e1000840. doi: 10.1371/journal.ppat.1000840

van Balkom, B. W. M., van Gestel, R. A., Brouwers, J. F. H. M., Krijgsveld, J., Tielens, A. G. M., Heck, A. J. R., et al. (2005). Mass spectrometric analysis of the *Schistosoma mansoni* tegumental sub-proteome. *J. Proteome Res.* 4, 958–966. doi: 10.1021/pr050036w

Vray, B., Hartmann, S., and Hoebeke, J. (2002). Immunomodulatory properties of cystatins. *Cell. Mol. Life Sci.* 59, 1503–1512. doi: 10.1007/s00018-002-8525-4

Wang, C., Lei, H., Tian, Y., Shang, M., Wu, Y., Li, Y., et al. (2017). *Clonorchis sinensis* granulin: identification, immunolocalization, and function in promoting the metastasis of cholangiocarcinoma and hepatocellular carcinoma. *Parasit. Vectors* 10:262. doi: 10.1186/s13071-017-2179-4

Wang, X., Chen, W., Hu, F., Deng, C., Zhou, C., Lv, X., et al. (2011a). *Clonorchis sinensis* enolase: identification and biochemical characterization of a glycolytic enzyme from excretory/secretory products. *Mol. Biochem. Parasitol.* 177, 135–142. doi: 10.1016/j.molbiopara.2011.02.011

Wang, X., Chen, W., Huang, Y., Sun, J., Men, J., Liu, H., et al. (2011b). The draft genome of the carcinogenic human liver fluke *Clonorchis sinensis*. *Genome Biol.* 12:R107. doi: 10.1186/gb-2011-12-10-r107

Wang, X., Chen, W., Tian, Y., Huang, Y., Li, X., and Yu, X. (2014a). RNAi-mediated silencing of enolase confirms its biological importance in *Clonorchis sinensis*. *Parasitol. Res.* 113, 1451–1458. doi: 10.1007/s00436-014-3785-0

Wang, X., Chen, W., Tian, Y., Mao, Q., Lv, X., Shang, M., et al. (2014b). Surface display of *Clonorchis sinensis* enolase on *Bacillus subtilis* spores potentializes an oral vaccine candidate. *Vaccine* 32, 1338–1345. doi: 10.1016/j.vaccine.2014.01.039

Wilson, R. A., and Coulson, P. S. (2006). Schistosome vaccines: a critical appraisal. *Mem. Inst. Oswaldo Cruz* 101(Suppl. 1), S13–S20. doi: 10.1590/s0074-02762006000900004

Wilson, R. A., Wright, J. M., de Castro-Borges, W., Parker-Manuel, S. J., Dowle, A. A., Ashton, P. D., et al. (2011). Exploring the *Fasciola hepatica* tegument proteome. *Int. J. Parasitol.* 41, 1347–1359. doi: 10.1016/j.ijpara.2011.08.003

Yang, J., Qiu, C., Xia, Y., Yao, L., Fu, Z., Yuan, C., et al. (2010). Molecular cloning and functional characterization of *Schistosoma japonicum* enolase which is highly expressed at the schistosomulum stage. *Parasitol. Res.* 107, 667–677. doi: 10.1007/s00436-010-1913-z

Yoo, M. H., Xu, X. M., Carlson, B. A., Gladyshev, V. N., and Hatfield, D. L. (2006). Thioredoxin reductase 1 deficiency reverses tumor phenotype and tumorigenicity of lung carcinoma cells. *J. Biol. Chem.* 281, 13005–13008. doi: 10.1074/jbc.c600012200

Young, N. D., Campbell, B. E., Hall, R. S., Jex, A. R., Cantacessi, C., Laha, T., et al. (2010). Unlocking the transcriptomes of two carcinogenic parasites, *Clonorchis sinensis* and *Opisthorchis viverrini*. *PLoS Negl. Trop. Dis.* 4:e719. doi: 10.1371/journal.pntd.0000719

Zheng, M., Hu, K., Liu, W., Hu, X., Hu, F., Huang, L., et al. (2011). Proteomic analysis of excretory secretory products from *Clonorchis sinensis* adult worms: molecular characterization and serological reactivity of a excretory–secretory antigen-fructose-1,6-bisphosphatase. *Parasitol. Res.* 109, 737–744. doi: 10.1007/s00436-011-2316-5

Zheng, M., Hu, K., Liu, W., Li, H., Chen, J., and Yu, X. (2013). Proteomic analysis of different period excretory secretory products from *Clonorchis sinensis* adult worms: molecular characterization, immunolocalization, and serological reactivity of two excretory secretory antigens—methionine aminopeptidase 2 and acid phosphatase. *Parasitol. Res.* 112, 1287–1297. doi: 10.1007/s00436-012-3264-4

Zhou, C., Bian, M., Liao, H., Mao, Q., Li, R., Zhou, J., et al. (2013). Identification and immunological characterization of thioredoxin transmembrane-related protein from *Clonorchis sinensis*. *Parasitol. Res.* 112, 1729–1736. doi: 10.1007/s00436-013-3331-5

Zhou, J., Gao, G., Crabb, J. W., and Serrero, G. (1993). Purification of an autocrine growth factor homologous with mouse epithelin precursor from a highly tumorigenic cell line. *J. Biol. Chem.* 268, 10863–10869.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 09 October 2020
doi: 10.3389/fgene.2020.543890

Integrating Genetic and Genomic Analyses of Combined Health Data Across Ecotypes to Improve Disease Resistance in Indigenous African Chickens

Georgios Banos[1,2,3†], Victoria Lindsay[4†], Takele T. Desta[5], Judy Bettridge[6,7,8], Enrique Sanchez-Molano[1], Adriana Vallejo-Trujillo[5], Oswald Matika[1], Tadelle Dessie[7], Paul Wigley[6], Robert M. Christley[6], Peter Kaiser[1], Olivier Hanotte[3,5,7] and Androniki Psifidi[1,3,4]*

[1] The Roslin Institute, The University of Edinburgh, Edinburgh, United Kingdom, [2] Scotland's Rural College, Edinburgh, United Kingdom, [3] Centre for Tropical Livestock Genetics and Health, Edinburgh, United Kingdom, [4] Royal Veterinary College, University of London, London, United Kingdom, [5] School of Life Sciences, University of Nottingham, Nottingham, United Kingdom, [6] Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom, [7] LiveGene – Centre for Tropical Livestock Genetics and Health, International Livestock Research Institute, Addis Ababa, Ethiopia, [8] Natural Resources Institute, University of Greenwich, London, United Kingdom

OPEN ACCESS

Edited by:
Theodore J. Perkins,
University of Ottawa, Canada

Reviewed by:
Priyanka Baloni,
Institute for Systems Biology (ISB), United States
Lujiang Qu,
China Agricultural University, China

***Correspondence:**
Androniki Psifidi
apsifidi@rvc.ac.uk

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Genetics

Received: 18 March 2020
Accepted: 04 September 2020
Published: 09 October 2020

Citation:
Banos G, Lindsay V, Desta TT, Bettridge J, Sanchez-Molano E, Vallejo-Trujillo A, Matika O, Dessie T, Wigley P, Christley RM, Kaiser P, Hanotte O and Psifidi A (2020) Integrating Genetic and Genomic Analyses of Combined Health Data Across Ecotypes to Improve Disease Resistance in Indigenous African Chickens. Front. Genet. 11:543890. doi: 10.3389/fgene.2020.543890

Poultry play an important role in the agriculture of many African countries. The majority of chickens in sub-Saharan Africa are indigenous, raised in villages under semi-scavenging conditions. Vaccinations and biosecurity measures rarely apply, and infectious diseases remain a major cause of mortality and reduced productivity. Genomic selection for disease resistance offers a potentially sustainable solution but this requires sufficient numbers of individual birds with genomic and phenotypic data, which is often a challenge to collect in the small populations of indigenous chicken ecotypes. The use of information across-ecotypes presents an attractive possibility to increase the relevant numbers and the accuracy of genomic selection. In this study, we performed a joint analysis of two distinct Ethiopian indigenous chicken ecotypes to investigate the genomic architecture of important health and productivity traits and explore the feasibility of conducting genomic selection across-ecotype. Phenotypic traits considered were antibody response to Infectious Bursal Disease (IBDV), Marek's Disease (MDV), Fowl Cholera (PM) and Fowl Typhoid (SG), resistance to *Eimeria* and cestode parasitism, and productivity [body weight and body condition score (BCS)]. Combined data from the two chicken ecotypes, Horro (n = 384) and Jarso (n = 376), were jointly analyzed for genetic parameter estimation, genome-wide association studies (GWAS), genomic breeding value (GEBVs) calculation, genomic predictions, whole-genome sequencing (WGS), and pathways analyses. Estimates of across-ecotype heritability were significant and moderate in magnitude (0.22–0.47) for all traits except for SG and BCS. GWAS identified several significant genomic associations with health and productivity traits. The WGS analysis revealed putative candidate genes and mutations for IBDV (*TOLLIP, ANGPTL5, BCL9, THEMIS2*), MDV (*GRM7*), SG (*MAP3K21*), *Eimeria* (*TOM1L1*) and

cestodes (*TNFAIP1, ATG9A, NOS2*) parasitism, which warrant further investigation. Reliability of GEBVs increased compared to within-ecotype calculations but accuracy of genomic prediction did not, probably because the genetic distance between the two ecotypes offset the benefit from increased sample size. However, for some traits genomic prediction was only feasible in across-ecotype analysis. Our results generally underpin the potential of genomic selection to enhance health and productivity across-ecotypes. Future studies should establish the required minimum sample size and genetic similarity between ecotypes to ensure accurate joint genomic selection.

Keywords: GWAS, GEBV, WGS, indigenous chickens, body weight, infectious diseases, antibody responses, Ethiopia

INTRODUCTION

Village poultry are important in low- and middle-income countries around the world (Hassan et al., 2004; Khobondo et al., 2015) and, therefore, are the focus of many development programs (Dana et al., 2010, 2011; Dinka et al., 2010; Ngeno, 2011; Magothe et al., 2012; Khobondo et al., 2015). However, such programs have often been unsustainable (Dinka et al., 2010; Dana et al., 2011). For example, in Ethiopia, previous poultry development programs concentrated on ways to increase the productivity of chickens by introducing commercial (exotic) breeds that perform highly under (semi) intensive management conditions. Whilst this has been a relatively successful approach in peri-urban areas (FAO, 2008), it has not translated well into rural areas and smallholder farmers. In the latter, system productivity is usually low and constrained by disease, predation and scarcity of feed (Pica-Ciamarra and Dhawan, 2009), among other factors. However, village chickens in these settings possess the advantages of being widely accessible and well-adapted to the local environmental conditions, while requiring lower inputs compared to commercial chickens (Psifidi et al., 2016; Bettridge et al., 2018). Moreover, consumers exhibit a strong preference for indigenous chicken meat and eggs compared to commercial types (Dana et al., 2010), which suggests that there is a viable market to be supplied. Besides the often unsustainable attempts to introduce exotic commercial birds (Dinka et al., 2010; Dana et al., 2011) and/or implement cross-breeding programs (Magothe et al., 2012; Khobondo et al., 2015), genetic improvement of the indigenous African village chickens presents a promising avenue. In fact, a breeding program is already in place aiming to improve both bird growth and egg production of local Ethiopian chickens (Wondmeneh et al., 2014). Moreover, in addition to focusing on increasing productivity, a focus toward raising the efficiency of the production system has also been considered beneficial (Bettridge et al., 2018). This is especially true for rural smallholders who may be able to make significant gains in income and nutrition, but of necessity can afford little in the way of economic or opportunity costs that some of the previous chicken development programs in Ethiopia have required.

We previously used a high-density genome-wide array to genotype two distinct unselected Ethiopian indigenous chicken ecotypes (village chickens from the Jarso and Horro geographic regions) and performed genomic studies on each ecotype separately to investigate the genetic architecture of six major infectious diseases (Marek's Disease, Infectious Bursal Disease, Fowl Cholera, Fowl Typhoid, and *Eimeria* and cestode parasitism) and two production traits (live body weight and body condition score) (Psifidi et al., 2016). The outcomes of that study suggested that concomitant genetic improvement for enhanced disease resistance and productivity in each indigenous chicken ecotype is feasible, although small population size would challenge the accuracy of selection. Indeed, successful genomic selection programs require sufficient numbers of individual birds with genotypic and phenotypic data, which may be a challenge within most indigenous chicken ecotypes. Therefore, the use of information across multiple ecotype populations presents, in principle, an attractive alternative to increase the relevant numbers and the accuracy of genomic selection. This has not been investigated in chickens before although there is evidence of benefit in across-breed genomic selection in other species such as cattle (Hozé et al., 2014; Iheshiulor et al., 2016).

In the present study, we extended the previous work of Psifidi et al. (2016) described above and jointly analyzed the same individual bird data from the Horro and Jarso indigenous ecotypes in order to increase the sample size and identify common genomic regions controlling the traits of interest. Moreover, we generated and analyzed whole-genome sequencing data of the two ecotypes to identify candidate genes and mutations within the relevant genomic regions, and performed pathway and network analysis in order to increase our understanding of the genomic architecture of the traits under investigation. We also examined the feasibility of joint across-ecotype genomic selection aiming to enhance antibody response, resistance to parasitic infectious diseases and productivity.

MATERIALS AND METHODS

Ethics Statement

All work was conducted with the approval of the University of Liverpool Research Ethics Committee (reference RETH000410).

Animals, Sampling, and Phenotyping

Details of the bird populations used in the present study, sampling strategy implemented and phenotyping of individual birds have been described in detail in Psifidi et al. (2016). Briefly,

the two indigenous ecotypes were located in the Jarso geographic region in arid eastern Ethiopia and in the Horro region in sub-humid western Ethiopia (Desta et al., 2013). These two regions are about 900 km far away from each other. Multistage cross-sectional sampling was performed considering two market sheds per geographic region, each represented by two villages; within each village, two chickens, over 6 months of age, were randomly sampled from each of 25 households. A total of 760 individual bird samples, 376 from Jarso and 384 from Horro, were collected in four rounds over 2 years at 6 month intervals, covering the pre- and post-rainy seasons of data collection (Desta et al., 2013; Bettridge et al., 2014).

Fresh feces were collected for parasite egg measurements, and brachial blood collected into tubes with sodium citrate for serology and on FTA cards for DNA extraction from each of the birds. *Eimeria* spp. oocysts and cestode spp. eggs were counted with a modified version of the concentration McMaster technique as described in Permin and Hansen (1998). Antibody titers for Infectious Bursal Disease (IBDV) were measured using a Flockscreen antibody ELISA kit (x-OvO, Dunfermline, United Kingdom). Antibody titers for Marek's Disease (MDV), Fowl Cholera (*Pasteurella multocida*, henceforth PM) and Fowl Typhoid (*Salmonella enterica serovar Gallinarum*, henceforth SG) were measured using in-house developed ELISAs, as described in Bettridge et al. (2014). Live bodyweight (BW, g) and body condition score (BCS) on a 0–3 scale (Gregory and Robins, 1998) were also recorded on each bird. All data except for BCS were $\log_{10}$-transformed in order to normalize the respective distributions.

All birds were genotyped using the high density single nucleotide polymorphism (SNP) genome-wide DNA array (Affymetrix® Axiom® HD) consisting of 580,954 SNPs uniformly distributed across the genome (Kranis et al., 2013). These data were subjected to the following quality control thresholds using PLINK v1.09 (Purcell et al., 2007): minor allele frequency < 0.05, call rate $< 95\%$, and Hardy-Weinberg equilibrium ($P < 10^{-6}$). After quality control, 359,470 SNP markers were kept for further analyses, distributed across all chromosomes.

Genetic Parameter Estimation

Variance components and genetic parameters were estimated for the combined population of the two ecotypes as described previously in detail in Psifidi et al. (2016) for all traits, using a mixed linear univariate model that included the fixed effects of geographic region, village, calendar season, sex and age of bird, and ELISA plate (for antibody titers only), and the random additive genetic effect of the individual bird. Genomic relationships among birds were calculated using a combined kinship matrix obtained with the Genome-wide Efficient Mixed Model Association (GEMMA) algorithm (Zhou and Stephens, 2014) and were included in the analyses. Estimates of variance components were obtained with the method of Restricted Maximum Likelihood and used to calculate the heritability of each trait as the ratio of the additive genetic to the total phenotypic variance. Bivariate analyses were subsequently conducted with the same model to estimate phenotypic and genetic correlations among the studied traits. All analyses were performed using the ASReml 3.0 software (Gilmour et al., 2009).

Genome-Wide Association Studies

For each studied trait, a genome-wide association analysis (GWAS) was performed using combined data across the two ecotypes. The GEMMA algorithm (Zhou and Stephens, 2014) and the same linear mixed model as in the genetic parameter estimation step described above were used. After Bonferroni correction for multiple testing, significance thresholds were $P \leq$ 1.39E-07 and 2.78E-06 for genome-wide significant ($P \leq 0.05$) and suggestive (one false positive per genome scan) levels, respectively, corresponding to $-\log_{10}$(P) of 6.85 and 5.55.

All significant and suggestive significant SNP markers identified in the GWAS were annotated using the galGal4 assembly according to the Affymetrix map file and their locations were then re-mapped from galGal4 to galGal6 using UCSC liftOver tool (Hinrichs et al., 2006). In addition, the genes located within 100 kb upstream and 100 kb downstream of these markers were also annotated using the BioMart (Smedley et al., 2015) data mining tool[1] within the Ensembl database and the galGal6 assembly. We chose these 200-kb windows based on the average Linkage Disequilibrium (LD) calculated previously for the Horro and Jarso chicken populations (Desta, 2015); mild LD ($r^2 \sim 0.2$) rarely exceeds 100 kb. This allowed us to catalog and create lists of all genes located around the identified significant markers that are potentially associated with antibody response, resistance to parasitic infection and production traits.

Whole Genome Sequencing Analysis

Six Horro and 14 Jarso birds were whole-genome sequenced (WGS; paired-end with a read length of 150 bp and average coverage of 40 X) on an Illumina HiSeqX platform. The sequencing reads were mapped to galGal6 using BWA-MEM (Li, 2013) and variant calling was performed using GATK tools according to Best Practices (DePristo et al., 2011; Poplin et al., 2018). Variant filtration was performed using GATK VQSR with 1 million validated SNPs and more than 20 million known chicken SNPs (from dbSNP).

Based on the GWAS results, the candidate regions (200 kb windows around each of the significant SNP) for each of the studied traits were extracted from the WGS data for both the Horro and Jarso chickens. The identified variants in these regions were annotated and their predicted effects on the encoding protein were assessed using the Ensembl Variant Effect Predictor tool (McLaren et al., 2016). The variants with a predicted high (stop gained/lost, start lost, frameshift variant, or splice acceptor/donor variants) and moderate (in-frame insertion/deletion, missense, or protein altering variants) impact were filtered out and further analyzed to highlight putative candidate genes and variants.

Pathways and Network Analysis

Pathways and network analyses were performed in the combined population based on the GWAS results. We reasoned that

[1] http://www.ensembl.org/biomart/martview/

the corresponding quantitative trait loci (QTL) regions would contain genes contributing to common pathways associated with each of the studied traits. Therefore, initially the lists of annotated genes located within the QTL regions for each bird phenotype were analyzed using the Ingenuity Pathway Analysis (IPA) program (Krämer et al., 2014). We sought evidence of gene set enrichment in order to identify potential underlying canonical pathways and networks associated with the studied traits. IPA constructs multiple possible upstream regulators, pathways and networks serving as hypotheses for the biological mechanism underlying the data based on a large-scale causal network derived from the Ingenuity Knowledge Base. Subsequently, IPA infers the most suitable pathways and networks based on their statistical significance (*P*-values obtained using Fisher's exact test), after correcting for a baseline threshold (Krämer et al., 2014).

Estimation of Genomic Breeding Values and Genomic Predictions

Genomic breeding values (GEBVs) were simultaneously estimated for all birds of the two ecotypes using a genomic best linear unbiased prediction model including the same effects as the genetic parameter estimation model [fixed effects of geographic region, village, calendar season, sex and age of bird, and ELISA plate (for antibody titers only], and the random additive genetic effect of the individual bird). Reliability of the GEBV of each bird and trait was calculated as:

$$Rel = 1 - \frac{PEV}{\sigma_A^2}$$

where PEV is the prediction error variance of the GEBV and $\sigma^2{}_A$ is the additive genetic variance of the trait.

A threefold cross-validation was then performed to assess the accuracy of genomic predictions for the combined population of Jarso and Horro ecotypes. Briefly, we divided the dataset into three subsets, each consisting of approximately equal proportions of birds from each ecotype, and predicted the GEBVs in each subset (validation subset) based on the analysis of genotypic and phenotypic data from the other two (reference subsets). This was repeated three times for each studied trait, with different random sample division in each repeat to reduce the effect produced by specific random combinations of animals within subsets on prediction. Accuracy of genomic prediction was defined as the Pearson correlation coefficient between the GEBVs in the validation subset and the corresponding adjusted phenotypic values. The latter were the residuals from a fixed-effect model analysis including all fixed effects described in the genetic parameter estimation step. A second accuracy measurement was derived by dividing the above-mentioned correlation by the square root of the trait heritability.

Derivation of GEBVs and cross-validation of genomic predictions were also performed separately within each ecotype for comparison with the across-ecotype results.

RESULTS

Descriptive Statistics and Genetic Parameters of Studied Traits

Means and standard deviations of the studied traits across the two ecotypes are presented in **Table 1**. Heritability estimates of these traits for the combined population of the two ecotypes are summarized in **Table 2**. Moderately high estimates were derived for MDV and IBDV antibody responses, and for BW. A moderately low heritability was estimated for PM antibody responses, and cestodes and *Eimeria* parasitism, while a low heritability was estimated for SG antibody responses and BCS. All trait heritability estimates were significantly ($P < 0.05$) greater than zero with the exception of SG and BCS.

Significantly ($P < 0.05$) different from zero phenotypic and genetic correlations were identified pairwise between PM and SG antibody responses, IBDV and SG antibody responses, and BW and BCS (**Table 3**). Moreover, a significant negative phenotypic and genetic correlation was estimated between *Eimeria* and cestodes parasitism. There were no significant genetic correlations of either production trait with any of the antibody response or parasitic infection traits.

Genome-Wide Association Studies

Genome-wide association studies outcomes revealed several new significant and suggestive genome-wide SNP associations for all the traits with the exception of PM antibody response (**Figure 1** and **Supplementary Figure S1**, **Supplementary Table S1**). We identified novel significant SNPs for IBDV (*P*-values 1.73×10^{-8} to 1.63×10^{-8}) and SG (4.68×10^{-8} to 2.23×10^{-8}) antibody titers as well as cestodes parasitism (1.35×10^{-7} to 5.88×10^{-8}). Moreover, suggestive genome-wide associations were identified for MDV antibody response and *Eimeria* parasitic resistance. A very strong genome-wide significant association was identified for BW (5.32×10^{-8} to 9.87×10^{-14}) with SNP markers

TABLE 1 | Means and standard deviations (STD) of antibody responses, resistance to parasitic infections and production traits across Horro and Jarso ecotypes.

Trait	Mean	*STD*
IBDV	0.04	0.15
MDV	0.21	0.58
SG	1.31	1.40
PM	0.96	0.77
Eimeria	38.83	202.16
Cestode	2.91	21.57
Body weight	1.33	0.31
BCS	1.55	0.60

IBDV log transformed antibody titers measured for Infectious Bursal Disease virus, MDV log-transformed antibody titers measured for Marek's Disease virus, SG log-transformed antibody titers measured for Salmonella enterica serovar Gallinarum, PM log-transformed antibody titers measured for Pasteurella multocida; Eimeria resistance to Eimeria parasitism (log-transformed egg counts/gr of fecal), Cestodes resistance to cestodes parasitism (log-transformed egg counts/gr of fecal), BW live body weight (log-transformed g), BCS body condition score (scale 0 to 3).

TABLE 2 | Heritability estimates (h^2) for antibody responses, resistance to parasitic infections and production traits across Horro and Jarso ecotypes.

	IBDV	MDV	SG	PM	*Eimeria*	Cestodes	BW	BCS
h^2	**0.47**	**0.42**	0.07	**0.29**	**0.22**	**0.28**	**0.43**	0.16
SE	**0.10**	**0.11**	0.10	**0.10**	**0.10**	**0.11**	**0.09**	0.10

Heritability estimates in bold attained statistical significance ($P < 0.05$). IBDV log transformed antibody titers measured for Infectious Bursal Disease virus, MDV log-transformed antibody titers measured for Marek's Disease virus, SG log-transformed antibody titers measured for Salmonella enterica serovar Gallinarum, PM log-transformed antibody titers measured for Pasteurella multocida; Eimeria resistance to Eimeria parasitism (log-transformed egg counts/gr of fecal), Cestodes resistance to cestodes parasitism (log-transformed egg counts/gr of fecal), BW live body weight (log-transformed g), BCS body condition score (scale 0 to 3); SE, standard error.

on chromosome 4. The same genomic association with BW had been previously reported separately within the Horro and Jarso ecotypes (Psifidi et al., 2016). Moreover, 22% of the SNPs identified in the present across-ecotype analyses had previously been identified in within-ecotype analyses of the same traits (Psifidi et al., 2016). Manhattan plots displaying the significant association results are shown in **Figure 1**; the corresponding Q-Q plots, which corroborate the genomic association results, are shown in **Supplementary Figure S1**. The list of all SNP markers with a significant or suggestive genome-wide association with the studied traits is presented in **Supplementary Table S1**.

The candidate regions identified by the GWAS for each of these traits contained a limited number of protein-coding genes (n = 180) and non-coding RNAs (n = 14), summarized in **Supplementary Table S2**.

Whole Genome Sequencing Analysis

Whole-genome sequencing analysis revealed the presence of high and/or moderate impact genetic variants located within the candidate regions for all studied traits. The majority of the identified genetic variants were common to both Horro and Jarso ecotypes (**Table 4**). Nevertheless, some unique genetic variants for each of the two ecotypes were also identified (**Table 4**). Genetic variants of interest identified in several immune related genes for all the antibody response and parasitic infection traits: IBDV (*TOLLIP, ANGPTL5, BCL9, THEMIS2*), MDV (*GRM7*), SG (*MAP3K21*), *Eimeria* (*TOM1L1*), and cestodes (*TNFAIP1, ATG9A, NOS2*). Details of the genetic variants with a high and moderate predicted impact on the encoded protein identified in the candidate regions of the studied traits are presented in the **Supplementary Table S3**.

Pathway and Network Analysis

The pathways and networks constructed from the gene products located in the candidate regions (based on the across-ecotype GWAS analyses) for each of the studied traits are presented in **Figure 2** and **Supplementary Figure S2**, respectively. Enriched pathways for IBDV and MDV antibody responses, and for cestodes parasitic resistance were mainly related to innate and adaptive immune responses (**Figure 2**). Examples include NF-Kb, toll-like receptor, interferon, B-cell response, T-cell, and Wnt/β-catenin signaling (**Figure 2**). In accordance with the previous within-ecotype analysis (Psifidi et al., 2016), the enriched pathways for BW were related to androgen signaling and pentose phosphate (**Figure 2**), which are associated with gluconeogenesis processes.

Moreover, several networks of molecular interactions were constructed using the list of genes in the candidate regions for antibody response to IBDV, MDV and SG, as well as resistance to parasitic infections (**Supplementary Figure S2**). The networks associated with resistance to parasitic infection were mostly related to cell death and survival, cell to cell signaling and interaction, immune trafficking, and hematological and immunological diseases (**Supplementary Figure S2**). For BCS, a network related to cellular function and maintenance as well as skeletal and muscular system development and function was constructed (**Supplementary Figure S2**).

Estimation of Genomic Breeding Values and Genomic Predictions

Reliability estimates of GEBVs and accuracy of genomic predictions from the across- and within-ecotype analyses are summarized in **Table 5**. Across-ecotype analysis resulted in a moderate to high GEBV reliability (0.37–0.80) for all traits with a statistically significant heritability (**Table 1**). The joint analysis of the two ecotypes improved the GEBV reliability of all traits considerably, compared to within-ecotype analysis (**Table 5**). On the other hand, genomic prediction accuracies from the across-ecotype analyses were not higher compared to within-ecotype. In the latter case, predictions were more accurate within the Horro ecotype, probably due to the higher heritability of the traits compared to Jarso. Interestingly, for some traits, such as IBDV antibody titer and parasitic infection resistance, within-ecotype predictions were not attainable and an across-ecotype analysis was the only option.

DISCUSSION

The present study set out to investigate the possibility of combining across-ecotype chicken data to study the genomic architecture of health and productivity traits, and examine the possibility of genomic selection in Ethiopian indigenous chickens. The GWAS analyses of the combined bird ecotype data revealed the presence of multiple novel genomic regions significantly associated with the traits of interest. This may be an indication that the two populations may have shared a common genetic profile in the recent history and confirmed that a larger sample size increases the power of genomic detection. Genomic and WGS information were integrated in order to detect and prioritize candidate genes, genetic variants and underlying pathways and networks for antibody response to bacterial and viral infections, resistance to parasitic infections and

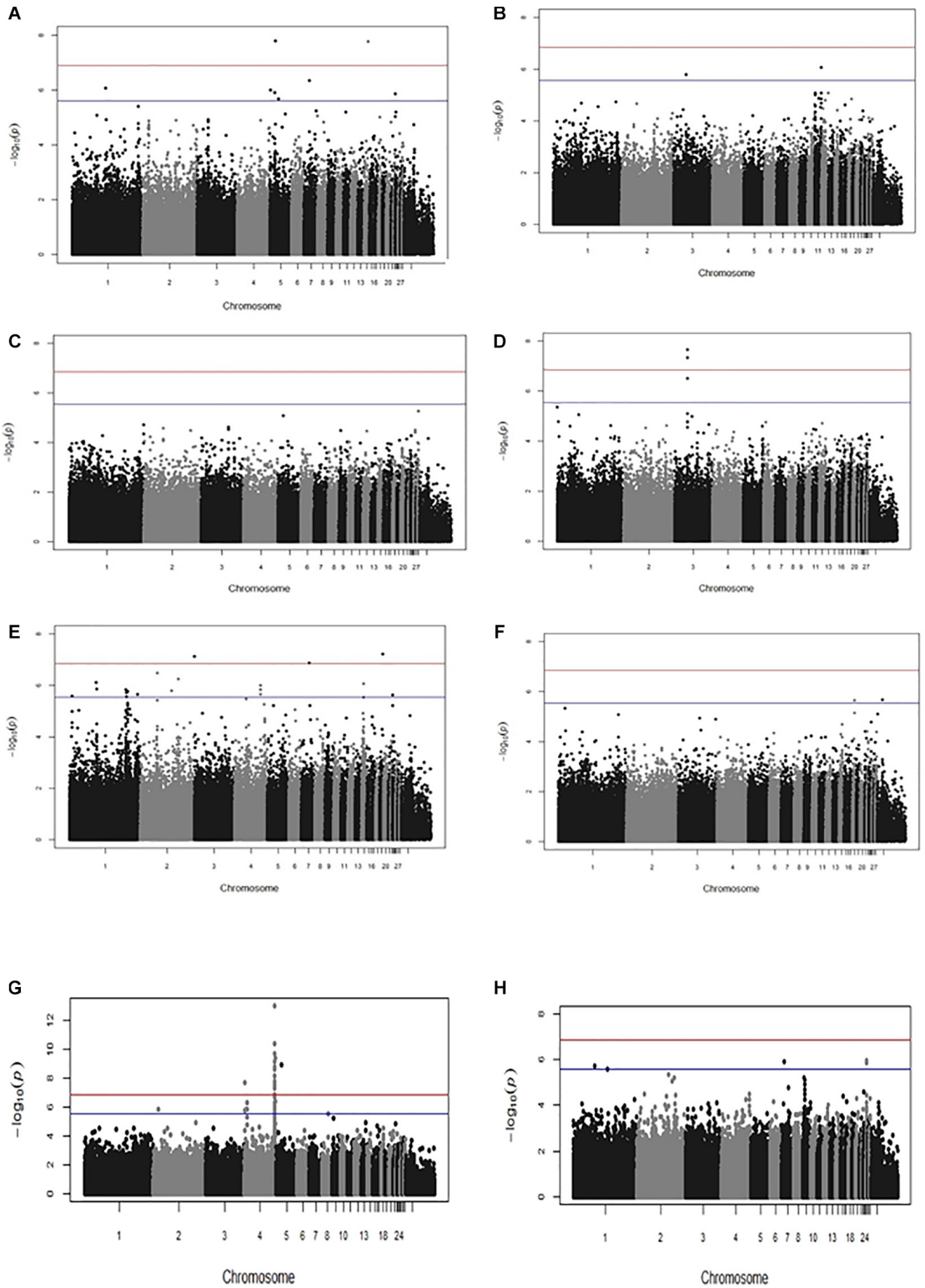


FIGURE 1 | Manhattan plots displaying the results of the genome-wide association analyses performed across Horro and Jarso chicken ecotypes. Genomic location (horizontal axis) is plotted against $-\log_{10}$(*P*-value); significant and suggestive significant genome-wide thresholds are shown as red and blue lines, respectively. Manhattan plots are for: **(A)** Infectious Bursal Disease virus (IBDV) antibody titer, **(B)** Marek's Disease virus (MDV) antibody titer, **(C)** *Pasteurella multocida* (PM) antibody titer, **(D)** *Salmonella enterica* serovar *Gallinarum* (SG) antibody titer, **(E)** Cestode parasitism, **(F)** *Eimeria* parasitism, **(G)** Live body weight (BW), **(H)** Body condition score (BCS).

TABLE 3 | Genetic (above diagonal) and phenotypic (below diagonal) correlations between traits studied across Horro and Jarso ecotypes (standard errors in parentheses).

	IBDV	MDV	PM	SG	Cestodes	*Eimeria*	BW	BCS
IBDV		0.03 (0.18)	0.16 (0.23)	**0.81 (0.41)**	0.13 (0.23)	0.05 (0.25)	−0.07 (0.16)	−0.30 (0.30)
MDV	0.05 (0.04)		0.14 (0.25)	0.19 (0.45)	−0.26 (0.25)	−0.05 (0.27)	0.01 (0.18)	−0.12 (0.31)
PM	0.05 (0.04)	−0.00 (0.04)		**1.36 (0.43)**	−0.47 (0.32)	−0.55 (0.34)	0.25 (0.23)	0.23 (0.39)
SG	**0.13 (0.04)**	−0.03 (0.04)	**0.40 (0.02)**		−0.67 (0.75)	−0.06 (0.60)	0.41 (0.51)	0.47 (0.72)
Cestodes	−0.02 (0.04)	0.02 (0.04)	0.01 (0.04)	0.01 (0.04)		**−0.63 (0.27)**	0.14 (0.21)	0.32 (0.38)
Eimeria	−0.03 (0.04)	0.04 (0.04)	**−0.11 (0.04)**	**−0.10 (0.04)**	**−0.17 (0.04)**		−0.03 (0.24)	1.23 (0.79)
BW	−0.01 (0.04)	0.01 (0.04)	**0.11 (0.04)**	**0.09 (0.04)**	0.01 (0.04)	−0.02 (0.04)		**0.54 (0.21)**
BCS	0.06 (0.04)	0.02 (0.04)	0.04 (0.04)	−0.01 (0.04)	−0.01 (0.04)	0.05 (0.04)	**0.30 (0.04)**	

Estimates in bold attained statistical significance ($P < 0.05$). IBDV antibody titers measured for Infectious Bursal Disease virus, MDV antibody titers measured for Marek's Disease virus, SG antibody titers measured for Salmonella enterica serovar Gallinarum, PM antibody titers measured for Pasteurella multocida; Eimeria resistance to Eimeria parasitism, Cestodes resistance to cestodes parasitism, BW live body weight, BCS body condition score.

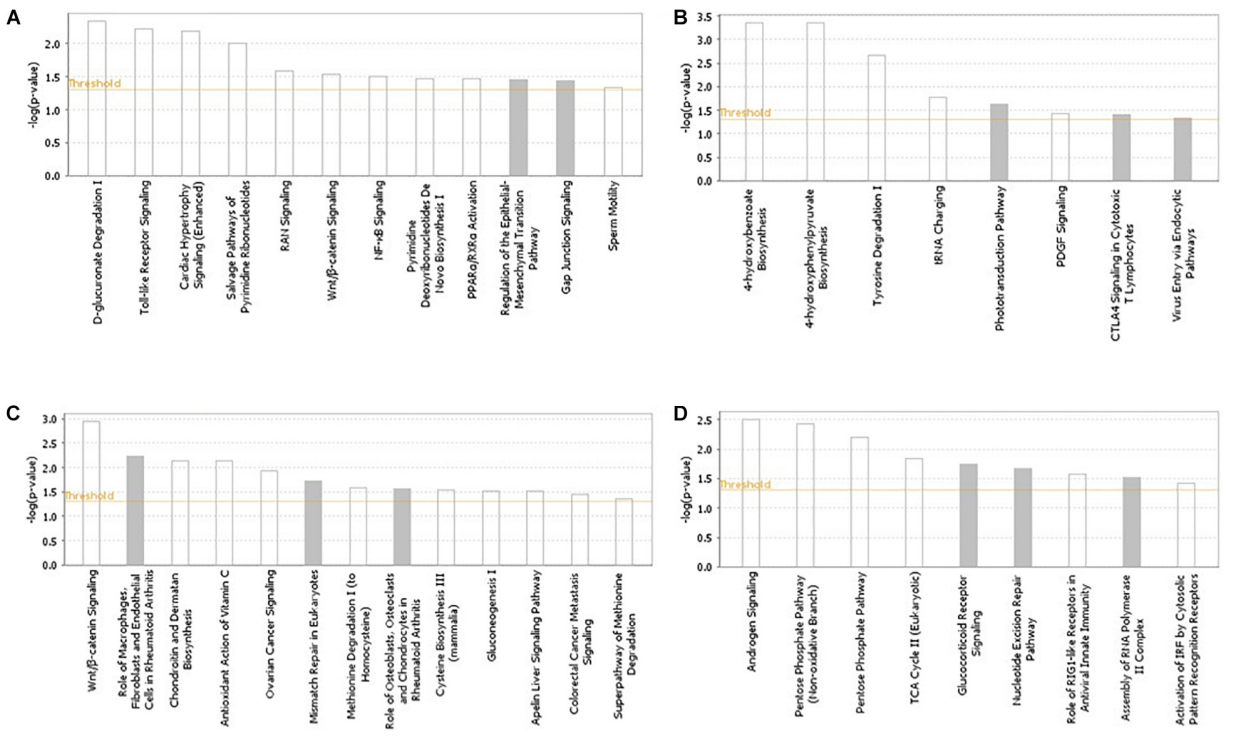


FIGURE 2 | Pathways analysis results using the IPA software across Horro and Jarso chicken ecotypes. The most highly represented canonical pathways of genes located in the candidate genomic regions for **(A)** Infectious Bursal Disease virus (IBDV) antibody titer, **(B)** Marek's Disease virus (MDV) antibody titer, **(C)** Cestodes parasitism resistance, **(D)** Live body weight. The solid yellow line represents the significance threshold. The line with squares represents the ratio of the genes within each pathway to the total number of genes in the pathway.

production traits. Moreover, the current study explored, for the first time to our knowledge, the possibility of performing genomic selection across distinct indigenous African chicken ecotypes and demonstrated the scope for merging such data in genetic improvement programs.

The across-ecotype population heritability estimates fell within the range of those previously calculated separately for the Horro and Jarso ecotypes (Psifidi et al., 2016). Most studied traits, except for BCS and SG, were heritable and amenable to improvement with genetic selection. Importantly, BCS and SG were found to be genetically correlated with BW and PM, respectively, which are highly heritable traits and, therefore, the former would indirectly benefit from genetic selection and improvement in the latter. Moreover, the *Eimeria* resistance heritability in Horro birds, which was not previously estimable within-ecotype (Psifidi et al., 2016), was estimable and significant when calculated in the across-ecotype analysis in the present study. This suggests that, despite the genetically distinct profiles of the two ecotypes (Psifidi et al., 2016), there is sufficient shared genetic material to allow for an across-ecotype analysis of Horro and Jarso indigenous birds. These findings are in accordance with a previous study comparing WGS data from

TABLE 4 | WGS analysis results.

Genetic variants	Traits							
Population	**IBDV**	**MDV**	**PM**	**SG**	**Cestodes**	***Eimeria***	**BW**	**BCS**
Unique to Horro birds								
High impact variants	0	0	0	0	0	0	0	0
Moderate impact variants	46	21	13	5	25	0	0	8
Affected genes	18	6	4	2	14	0	0	4
Unique to Jarso birds								
High impact variants	0	0	0	0	1	0	12	3
Moderate impact variants	99	1	16	4	38	16	0	28
Affected genes	40	1	7	2	23	2	2	11
Present in both Horro and Jarso birds								
High impact variants	5	0	1	0	2	0	0	1
Moderate impact variants	290	2	87	0	169	31	107	116
Affected genes	57	2	9	0	30	3	16	22

Genetic variants with a predicted high and moderate impact identified in the candidate regions for each of the studied traits in Horro and Jarso chickens. IBDV, titres measured for Infectious Bursal Disease virus; MDV, antibody titres measured for Marek's Disease virus; SG, antibody titres measured for Salmonella enterica serovar Gallinarum; PM, antibody titres measured for Pasteurella multocida; Eimeria, resistance to Eimeria parasitism; Cestodes, resistance to cestodes parasitism; BW, live body weight; BCS, body condition score.High impact: predicted high impact variant by Ensembl Variant Effect Predictor tool; Moderate impact: predicted moderate impact variant by Ensembl Variant Effect Predictor tool; Affected genes: number of genes containing one or more predicted high or moderate impact variant within the 200 kb window around the significant SNP marker identified by the GWAS analyses.

TABLE 5 | Reliability of genomic breeding values (GEBV) and cross-validation accuracy of genomic predictions from within- and across-ecotype analyses.

	IBDV	**MDV**	**SG**	**PM**	***Eimeria***	**Cestodes**	**BW**	**BCS**
Within-ecotype Horro								
GEBV Reliability	Non-estimable	0.54	Non-estimable	0.39	Non-estimable	Non-estimable	Not estimable	Non-estimable
Accuracy	Non-estimable	0.40	Non-estimable	0.36	Non-estimable	Non-estimable	0.39	Non-estimable
Relative Accuracy	Non-estimable	0.66	Non-estimable	0.57	Non-estimable	Non-estimable	0.53	Non-estimable
Within-ecotype Jarso								
GEBV Reliability	Non-estimable	0.21	Non-estimable	Non-estimable	0.16	0.58	0.07	0.25
Accuracy	0.13	0.27	Non-estimable	Non-estimable	0.18	0.20	0.29	0.17
Relative Accuracy	0.24	0.39	Non-estimable	Non-estimable	0.26	0.31	0.44	0.26
Across-ecotype								
GEBV Reliability	0.37	0.55	Non-estimable	0.80	0.76	0.80	0.45	Non-estimable
Accuracy	0.17	0.20	Non-estimable	0.16	0.10	0.16	0.23	Non-estimable
Relative Accuracy	0.24	0.32	Non-estimable	0.34	0.22	0.37	0.34	Non-estimable

IBDV, antibody titers to Infectious Bursal Disease virus; MDV, antibody titers to Marek's Disease virus; SG, antibody titers to Salmonella enterica serovar Gallinarum; PM, antibody titers to Pasteurella multocida; Eimeria, resistance to Eimeria parasitism; Cestodes, resistance to cestodes parasitism; BW, live bodyweight; BCS, body condition score; Relative Accuracy, Accuracy corrected by the square root of heritability. Non-estimable results signified failure of the calculating algorithm to converge.

Horro and Jarso birds to Saudi Arabian (AlQurin, Goliggah, and Al Oyoun districts) and Sri Lankan (Puttalam district) indigenous chickens and red junglefowl (Yunnan and Hainan provinces, China) aiming to explore genetic differentiation between the populations (Lawal et al., 2018). According to the latter, both Ethiopian ecotypes clustered together distinctly from the other indigenous birds and the junglefowl, indicating closer genetic relationships between these ecotypes. Whilst there was some separation between Horro and Jarso within the Ethiopian cluster, these results supported analyses of an across-ecotypes population, particularly in combination with selective sweep results indicating different selection pressures in Ethiopian birds from the other indigenous birds and red junglefowl (Lawal et al., 2018).

Phenotypic and genetic correlations between traits were generally consistent with previous reports from within-ecotype analyses (Psifidi et al., 2016). However, the genetic correlation between resistance to the two parasitic infections (*Eimeria* and cestodes) was strongly negative (−0.63) in the combined population compared to a positive estimate in Horro (0.68) and a zero estimate in Jarso in previous within-ecotype analyses (Psifidi et al., 2016). *Eimeria* co-infection responses are complex and dependent on the nature of the co-infection. Co-infection of chickens with *Eimeria* and *Toxoplasma gondii* in a previous experimental study (Hiob et al., 2017) demonstrated no significant difference in pathology or immune response to *Eimeria* infection alone; however, *Eimeria* oocyst excretion was actually lower in the co-infected birds than mono-infected

(Hiob et al., 2017). This could potentially explain the negative phenotypic and genetic correlation between *Eimeria* and cestode oocysts burden estimated in the present study. However, new studies based on controlled challenge experiments are needed in order to confirm these findings. The other phenotypic and genetic correlation results support previous observations of PM and SG co-infections (Biswas et al., 2005) as well as IBDV and SG co-infections (Chaka et al., 2012) in village poultry. In accordance with previous studies (Psifidi et al., 2016), *Eimeria* parasitism displayed negative phenotypic correlations with PM and SG antibody responses that might be suggestive of parasitic immunosuppression.

Importantly, there were no significant genetic correlations of production traits with antibody response and resistance to parasitic infection, in agreement with the previous within-ecotype analyses (Psifidi et al., 2016). This result would greatly facilitate the development of concomitant breeding programs aiming to improve simultaneously production and disease resistance traits, which would have been hampered by potentially antagonistic correlations between traits.

Genomic analyses based on joined data of the two ecotypes combined with whole genome sequencing analysis identified several novel significant SNP markers, genes of interest and genetic variants, which are suitable candidates for further investigation to determine if they can be exploited in future breeding programs. About one fifth (22%) of the SNPs identified in the across-ecotype analysis here had previously been identified in the separate analyses of the two ecotypes (Psifidi et al., 2016), attesting to a partially common genetic basis of these traits in the two populations. Interestingly, some of the SNP associations previously identified in the within-ecotype analyses (Psifidi et al., 2016) now attained an even higher significance, suggesting that the bigger sample size in the joint analysis increased the GWAS power to not only identify novel associations but also confirm previously identified ones.

Similarly, pathway analysis not only confirmed the presence of previously identified pathways but also identified enrichment for several immune pathways, which had not been detectable in the previous within-ecotype analyses (Psifidi et al., 2016). For the IBDV antibody responses among the novel identified pathways was the toll-like receptor (TLR) signaling, which has been previously linked to IBDV. *TLR3* has been found to be up-regulated in IBDV-infected chickens (Ou et al., 2017), and the function of this TLR has been considered vital to innate responses to viral infection (Yang et al., 2016). We also identified two missense variants in the toll-interacting protein (*TOLLIP*) gene, which is a component of the TLR signaling pathway that acts to inhibit cell activation, and five missense variants in the single immunoglobulin IL-1R-related molecule (*SIGIRR*), which is a TLR/IL-1 receptor system antagonist. One novel pathway found here to be associated with MDV antibody titer was cytotoxic T lymphocyte (CTL) signaling. MDV is a lymphotropic virus and has previously been associated with lowered ratio of CTLs to $CD4^+$ T cells in infected chicken skin (Heidari et al., 2016) due to immunosuppression. A comparison of MDV resistant and susceptible chicken lines demonstrated that baseline levels of specifically CD8αα T cells were different between the two lines, indicating that these cells play an important role in effective immune response to MDV infection (Perumbakkam et al., 2016).

Furthermore, cestode resistance was associated with Wnt/β-catenin signaling and with networks linked to digestive system development and function, and with cellular morphology and abnormalities. Interestingly, cestode infections can be treated with anthelmintic drugs such as niclosamide (Katz, 1977), mebendazole or pyrvinium (Hamilton and Rath, 2017), which are also used to treat Wnt-dependent cancers (Lu et al., 2011; Osada et al., 2011; Zhang et al., 2017) due to their Wnt inhibitory action (Chen et al., 2009). These results suggest a potential involvement of the Wnt signaling pathway to cestode parasitism resistance.

Live body weight was unsurprisingly associated with androgen signaling, which has long been known to affect body growth (Li et al., 2020). In addition, two missense variants were identified in the adipose triglyceride lipase (*ATPL/PNPLA2*) gene that has been previously associated with chicken growth and fat deposition traits (Fennell and Scanes, 1992; Fennell et al., 1996; Nie et al., 2010). Moreover, previously identified polymorphisms in this gene have been associated with myopathic phenotypes (Akiyama et al., 2007; Fischer et al., 2007). The majority of high and moderate predicted impact variants were present in both ecotypes and all studied traits, with the notable exception of BW where 12 predicted high impact variants, all of them located in two novel long non-coding RNAs (lncRNAs), were found uniquely in Jarso chickens, with no shared high impact variants found in Horro chickens. There was a significant difference in BW between the two ecotypes previously (Psifidi et al., 2016) [unpaired two-tailed *t*-test; $t(758) = 5.747$; $P < 0.0001$], with Jarso animals being the smaller ones, which may explain the number of predicted high impact variants found only in Jarso birds in candidate regions linked to live body weight and growth. This suggestion is in accordance with previous studies of humans, mice and cattle, which have reported that many lncRNAs play crucial roles in the development of skeletal muscle and cardiac lineages (Sweta et al., 2019). Furthermore, the QTLs identified for BW in the Ethiopian ecotypes are quite unique. A search in the Chicken QTL database[2] identified 337 genomic associations across chicken chromosomes 2, 4 and 5 for BW, from a total of 69 different studies. However, none of these were within 1 Mb of genomic markers identified in the present study. All the above findings warrant further investigation in order to validate and fully dissect the underlying molecular mechanisms connecting genotypes with disease and productivity phenotypes, and determine how to best inform future breeding programs.

An advantage of using genomic predictions for selection to improve traits of interest compared to using more targeted approaches is that the effects of all genetic markers across the genome are considered simultaneously. This allows potentially modest additive effects of individual SNPs that may not meet stringent statistical tests to be modeled into phenotypic predictions of complex traits. Furthermore, the identification of causative genes and mutations could increase the accuracy of the estimated genomic predictions

[2]https://www.animalgenome.org/cgi-bin/QTLdb/GG/

(van Binsbergen et al., 2015; Edwards et al., 2016). In the present study, genomic analyses across-ecotype increased the reliability of individual bird GEBVs compared to within-ecotype analyses, likely due to the larger sample size. Moreover, the across-ecotype analyses allowed relatively accurate genomic predictions for traits that could not be attained within-ecotype for this dataset. However, for most traits the accuracy of predictions did not improve across-ecotype compared to within-ecotype suggesting that the linkage disequilibrium between SNPs and QTL does not always persist across the two populations. These findings corroborate the GWAS results showing a partial overlap of genomic associations within and across ecotypes. Moreover, the cross-validation method considered in the present study required approximately equal proportions of Jarso and Horro birds within each subset in order to improve the balance in prediction across ecotypes. Nevertheless, methods of genomic predictions should be consistent with the actual aims of genomic selection; for example, if the cross-region movement of birds is not an option, then genomic predictions should be weighed toward the ecotypes that are more common in the specific geographic region. An alternative approach to cross-validation in assessing the accuracy of genomic predictions would be forward validation, which makes use of data from earlier years to train the model and predict performance in the recent years. Since it uses all available data, forward validation might expedite genomic selection response. This approach has previously been successfully applied in studies of indigenous cattle breeds (Neves et al., 2014; Silva et al., 2016; Boison et al., 2017) but not of chicken.

Results obtained in the present study apply to the two Ethiopian ecotypes but the principle of across-ecotype genomic selection may be expanded to other chicken populations raised in sub-Saharan Africa provided that the phenotypes of interest and the associated markers are to some extend common in different populations. Size of available phenotypic and genomic data, and the genetic similarity of the ecotypes would be the key determinants of genomic selection success in each case. Therefore, further studies aiming to genetically and phenotypically characterize other indigenous chicken ecotypes are required to inform and underpin future breeding programs. In addition, independent studies should establish the required minimum sample size to derive reliable GEBVs and accurate genomic predictions in these settings.

DATA AVAILABILITY STATEMENT

The Horro and Jarso whole genome sequences used in the present study are deposited in the NCBI Sequence Read Archive (SRA) under accession number SRP142580.

ETHICS STATEMENT

All work was conducted with the approval of the University of Liverpool Research Ethics Committee (reference RETH000410).

AUTHOR CONTRIBUTIONS

RC, OH, PW, PK, and TD conceived the experimental design and secured funding. AP, OH, and GB designed the genetic studies. TTD and JB performed the collection of samples and the phenotyping. GB, VL, and AP with input from ESM and OM performed the genetic parameter and the genomic prediction analyses. GB and AP collated and edited the genotyping data and performed the GWAS analysis. AVT and OH generated the WGS data and VL with input from AP analyzed the genetic variants. AP performed the pathway analysis. AP, OH, RC, PW, TD, GB, VL, and ESM interpreted these results. GB and VL wrote the manuscript with input from AP. All other co-authors provided manuscript editing and feedback and read and approved the final manuscript.

FUNDING

We thank Biotechnology and Biological Sciences Research Council (BBSRC), the UK Foreign Commonwealth & Development Office (FCDO, previous DFID) and the Scottish Government for providing funding for the 'Reducing the impact of infectious disease on poultry production in Ethiopia' project under the Combating Infectious Diseases of Livestock for International Development (CIDLID) program (BB/H009396/1, BB/H009159/1, and BB/H009051/1). Part of this research was funded by the Bill & Melinda Gates Foundation and with UK aid from the UK Foreign, Commonwealth and Development Office (Grant Agreement OPP1127286) under the auspices of the Centre for Tropical Livestock Genetics and Health (CTLGH), established jointly by the University of Edinburgh, SRUC (Scotland's Rural College), and the International Livestock Research Institute. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation nor the UK Government.

ACKNOWLEDGMENTS

We thank the Chicken Health for Development project team members and the farmers and development agents in the Jarso and Horro districts for their assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.543890/full#supplementary-material

REFERENCES

Akiyama, M., Sakai, K., Ogawa, M., McMillan, J. R., Sawamura, D., and Shimizu, H. (2007). Novel duplication mutation in the patatin domain of adipose triglyceride lipase (PNPLA2) in neutral lipid storage disease with severe myopathy. *Muscle Nerve* 36, 856–859. doi: 10.1002/mus.20869

Bettridge, J. M., Lynch, S. E., Brena, M. C., Melese, K., Dessie, T., Terfa, Z. G., et al. (2014). Infection-interactions in Ethiopian village chickens. *Vet. Prevent. Med.* 117, 358–366. doi: 10.1016/j.prevetmed.2014.07.002

Bettridge, J. M., Psifidi, A., Terfa, Z. G., Desta, T. T., Lozano-Jaramillo, M., Dessie, T., et al. (2018). The role of local adaptation in sustainable production of village chickens. *Nat. Sustain.* 1, 574–582. doi: 10.1038/s41893-018-0150-9

Biswas, P. K., Biswas, D., Ahmed, S., Rahman, A., and Debnauth, N. C. (2005). A longitudinal study of the incidence of major endemic and epidemic diseases affecting semi-scavenging chickens reared under the participatory livestock development project areas in Bangladesh. *Avian Pathol.* 34, 303–312. doi: 10.1080/03079450500178972

Boison, S. A., Utsunomiya, A. T. H., Santos, D. J. A., Neves, H. H. R., Carvalheiro, R., Mészáros, G., et al. (2017). Accuracy of genomic predictions in Gyr (Bos indicus) dairy cattle. *J. Dairy Sci.* 100, 1–12. doi: 10.3168/jds.2016-11811

Chaka, H., Goutard, F., Bisschop, S. P. R., and Thompson, P. N. (2012). Seroprevalence of Newcastle disease and other infectious diseases in backyard chickens at markets in Eastern Shewa zone, Ethiopia. *Poult. Sci.* 91, 862–869. doi: 10.3382/ps.2011-01906

Chen, M., Wang, J., Lu, J., Bond, M. C., Ren, X.-R., Lyerly, H. K., et al. (2009). The anti-helminthic niclosamide inhibits Wnt/Frizzled1 signaling. *Biochemistry* 48, 10267–10274. doi: 10.1021/bi9009677

Dana, N., Van der Waaij, L. H., Dessie, T., and van Arendonk, J. A. (2010). Production objectives and trait preferences of village poultry producers of Ethiopia: implications for designing breeding schemes utilizing indigenous chicken genetic resources. *Trop. Anim. Health Product.* 42, 1519–1529. doi: 10.1007/s11250-010-9602-6

Dana, N., Vander Waaij, E., and van Arendonk, J. A. (2011). Genetic and phenotypic parameter estimates for body weights and egg production in Horro chicken of Ethiopia. *Trop. Anim. Health Product.* 43, 21–28. doi: 10.1007/s11250-010-9649-4

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.

Desta, T. T. (2015). *Phenomic and Genomic Landscape of Ethiopian Village Chickens*. Nottingham: University of Nottingham.

Desta, T. T., Dessie, T., Bettridge, J., Lynch, S. E., Melese, K., Collins, M., et al. (2013). Signature of artificial selection and ecological landscape on morphological structures of Ethiopian village chickens. *Anim. Genet. Resour.* 52, 17–29. doi: 10.1017/s2078633613000064

Dinka, H., Chala, R., Dawo, F., Bekana, E., and Leta, S. (2010). Major constraints and health management of village poultry production in Rift Valley of Oromia, Ethiopia. *Am. Euras. J. Agric. Environ. Sci.* 9, 529–533.

Edwards, S. M., Sørensen, I. F., Sarup, P., Mackay, T. F., and Sørensen, P. (2016). Genomic Prediction for Quantitative Traits Is Improved by Mapping Variants to Gene Ontology Categories in Drosophila melanogaster. *Genetics* 203, 1871–1883.

FAO (2008). *Poultry Sector Country Review*. Rome: Food and Agriculture Organization.

Fennell, M. J., Radecki, S. V., Proudman, J. A., and Scanes, C. G. (1996). The suppressive effects of testosterone on growth in young chickens appears to be mediated via a peripheral androgen receptor; studies of the anti-androgen ICI 176,334. *Poult. Sci.* 75, 763–766. doi: 10.3382/ps.0750763

Fennell, M. J., and Scanes, C. G. (1992). Inhibition of growth in chickens by testosterone, 5α-dihydrotestosterone, and 19-nortestosterone. *Poult. Sci.* 71, 357–366. doi: 10.3382/ps.0710357

Fischer, J., Lefèvre, C., Morava, E., Mussini, J.-M., Laforêt, P., Negre-Salvayre, A., et al. (2007). The gene encoding adipose triglyceride lipase (PNPLA2) is mutated in neutral lipid storage disease with myopathy. *Nat. Genet.* 39, 28–30. doi: 10.1038/ng1951

Gilmour, A., Gogel, B., Cullis, B., and Thompson, R. (2009). *ASReml User Guide Release 3.0*. Hemel Hempstead: VSN International Ltd.

Gregory, N. G., and Robins, J. K. (1998). A body condition scoring system for layer hens. *N. Z. J. Agric. Res.* 41, 555–559. doi: 10.1080/00288233.1998.9513338

Hamilton, G., and Rath, B. (2017). Repurposing of anthelminthics as anticancer drugs. *Oncomedicine* 2, 142–149.

Hassan, M., Afify, M., and Aly, M. (2004). Genetic resistance of Egyptian chickens to infectious bursal disease and Newcastle disease. *Trop. Anim. Health Product.* 36, 1–9. doi: 10.1023/b:trop.0000009524.47913.d4

Heidari, M., Wang, D., Delekta, P., and Sun, S. (2016). Marek's disease virus immunosuppression alters host cellular responses and immune gene expression in the skin of infected chickens. *Vet. Immunol. Immunopathol.* 180, 21–28. doi: 10.1016/j.vetimm.2016.08.013

Hiob, L., Koethe, M., Schares, G., Goroll, T., Daugshies, A., and Bangoura, B. (2017). Experimental Toxoplasma gondii and Eimeria tenella co-infection in chickens. *Parisitol. Res.* 116, 3189–3203. doi: 10.1007/s00436-017-5636-2

Hinrichs, A. S., Karolchik, D., Baertsch, R., Barber, G. P., Bejerano, G., Clawson, H., et al. (2006). The UCSC genome browser database: update 2006. *Nucleic Acids Res.* 34(suppl_1), D590–D598. doi: 10.1093/nar/gkj144

Hozé, C., Fritz, S., Phocas, F., Boichard, D., Ducrocq, V., and Croiseau, P. (2014). Efficiency of multi-breed genomic selection for dairy cattle breeds with different sizes of reference population. *J. Dairy Sci.* 97, 3918–3929. doi: 10.3168/jds.2013-7761

Iheshiulor, O. O., Woolliams, J. A., Yu, X., Wellmann, R., and Meuwissen, T. H. (2016). Within- and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. *Genet. Sel. Evol.* 48:15.

Katz, M. (1977). Anthelmintics. *Drugs* 13, 124–136. doi: 10.1016/s0140-6736(01)35813-0

Khobondo, J., Muasya, T., Miyumo, S., Okeno, T. O., Wasike, C., Mwakubambanya, R., et al. (2015). Genetic and nutrition development of indigenous chicken in Africa. *Livestock Res. Rural Dev.* 27:2015.

Krämer, A., Green, J., Pollard, J. Jr., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30, 523–530. doi: 10.1093/bioinformatics/btt703

Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S., et al. (2013). Development of a high density 600K SNP genotyping array for chicken. *BMC Genom.* 14:59. doi: 10.1186/1471-2164-14-59

Lawal, R. A., Al-Atiyat, R. M., Aljumaah, R. S., Silva, P., Mwacharo, J. M., and Hanotte, O. (2018). Whole-genome resequencing of red junglefowl and indigenous village chicken reveal new insights on the genome dynamics of the species. *Front. genet.* 9:264. doi: 10.3389/fgene.2018.00264

Li, D., Wang, Q., Shi, K., Lu, Y., Yu, D., Shi, X., et al. (2020). Testosterone promotes the proliferation of chicken embryonic myoblasts via androgen receptor mediated PI3K/Akt signaling pathway. *Intern. J. Mol. Sci.* 21:1152. doi: 10.3390/ijms21031152

Li, H. (2013). *Aligning Sequence Reads, Clone Sequences and Assembly Contigs With BWA-MEM*. arXiv, 1303.3997.

Lu, W., Lin, C., Roberts, M. J., Waud, W. R., Piazza, G. A., and Li, Y. (2011). Niclosamide suppresses cancer cell growth by inducing Wnt co-receptor LRP6 degradation and inhibiting the Wnt/β-Catenin pathway. *PLoS One* 6:e29290. doi: 10.1371/journal.pone.0029290

Magothe, T., Okeno, T., Muhuyi, W., and Kahi, A. (2012). Indigenous chicken production in Kenya: I. Current status. *World Poult. Sci. J.* 68, 119–132. doi: 10.1017/s0043933912000128

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., et al. (2016). The ensembl variant effect predictor. *Genome Biol.* 17:122.

Neves, H. H., Carvalheiro, R., O'Brien, A. M. P., Utsunomiya, Y. T., do Carmo, A. S., Schenkel, F. S., et al. (2014). Accuracy of genomic predictions in Bos indicus (Nellore) cattle. *Genet. Sel. Evol.* 46:17. doi: 10.1186/1297-9686-46-17

Ngeno, K. (2011). *Genetic Analysis of Growth Patterns in Different Ecotypes of Indigenous Chicken Populations in Kenya*. Njoro: Egerton University.

Nie, Q. H., Fang, M. X., Xie, L., Shen, X., Liu, J., Luo, Z. P., et al. (2010). Associations of ATGL gene polymorphisms with chicken growth and fat traits. *J. Appl. Genet.* 51, 185–191. doi: 10.1007/bf03195726

Osada, T., Chen, M., Yang, X. Y., Spasojevic, I., Vandeusen, J. B., Hsu, D., et al. (2011). Antihelminth compound niclosamide downregulates Wnt signaling and elicits antitumor responses in tumors with activating APC mutations. *Cancer Res.* 71, 4172–4182. doi: 10.1158/0008-5472.can-10-3978

Ou, C., Wang, Q., Zhang, Y., Kong, W., Zhang, S., Yu, Y., et al. (2017). Transcription profiles of the responses of chicken bursae of Fabricius to IBDV in different timing phases. *Virol. J.* 14:93.

Permin, A., and Hansen, J. W. (1998). *Epidemiology, Diagnosis and Control of Poultry Parasites*. Rome: FAO.

Perumbakkam, S., Hunt, H. D., and Cheng, H. H. (2016). Differences in CD8αα and cecal microbiome community during proliferation and late cytolytic phases of Marek's disease virus infection are associated with genetic resistance to Marek's disease. *FEMS Microbiol. Ecol.* 92:fiw188. doi: 10.1093/femsec/fiw188

Pica-Ciamarra, U., and Dhawan, M. (2009). *A Rapid Rural Appraisal of the Family-Based Poultry Distribution Scheme of West Bengal, India. Pro-Poor Livestock Policy Initiative (PPLPI) Research Report (FAO)*. Rome: FAO.

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., et al. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* [Preprint], doi: 10.1101/201178v3

Psifidi, A., Banos, G., Matika, O., Desta, T. T., Bettridge, J., Hume, D. A., et al. (2016). Genome-wide association studies of immune, disease and production traits in indigenous chicken ecotypes. *Genet. Select. Evol.* 48:74.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., et al. (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* 43, W589–W598. doi: 10.1093/nar/gkv350

Silva, R. M. O., Fragomeni, B. O., Lourenco, D. A. L., Magalhães, A. F. B., Irano, N., Carvalheiro, R., et al. (2016). Accuracies of genomic prediction of feed efficiency traits using different prediction and validation methods in an experimental Nelore cattle population. *J. Anim. Sci.* 94, 3613–3623. doi: 10.2527/jas.2016-0401

Sweta, S., Dudnakova, T., Sudheer, S., Baker, A. H., and Bhushan, R. (2019). Importance of long non-coding RNAs in the development and disease of skeletal muscle and cardiovascular lineages. *Front. Cell Dev. Biol.* 7:228. doi: 10.3389/fcell.2019.00228

van Binsbergen, R., Calus, M. P. L., Bink, M. C. A. M., van Eeuwijk, F. A., Schrooten, C., and Veerkamp, R. F. (2015). Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. *Genet. Select. Evol.* 47:71.

Wondmeneh, E., van Der Waaij, E., Tadelle, D., Udo, H., and Van Arendonk, J. (2014). Adoption of exotic chicken breeds by rural poultry keepers in Ethiopia. *Acta Agric. Scand.Sect. A Anim. Sci.* 64, 210–216. doi: 10.1080/09064702.2015.1005658

Yang, Y., Wang, S.-Y., Huang, Z.-F., Zou, H.-M., Yan, B.-R., Luo, W.-W., et al. (2016). The RNA-binding protein Mex3B is a coreceptor of Toll-like receptor 3 in innate antiviral response. *Cell Res.* 26, 288–303. doi: 10.1038/cr.2016.16

Zhang, C., Zhang, Z., Zhang, S., Wang, W., and Hu, P. (2017). Targeting of Wnt/β-catenin by anthelmintic drug pyrvinium enhances sensitivity of ovarian cancer cells to chemotherapy. *Med. Sci. Monit.* 23:266. doi: 10.12659/msm.901667

Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* 11:407. doi: 10.1038/nmeth.2848

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership