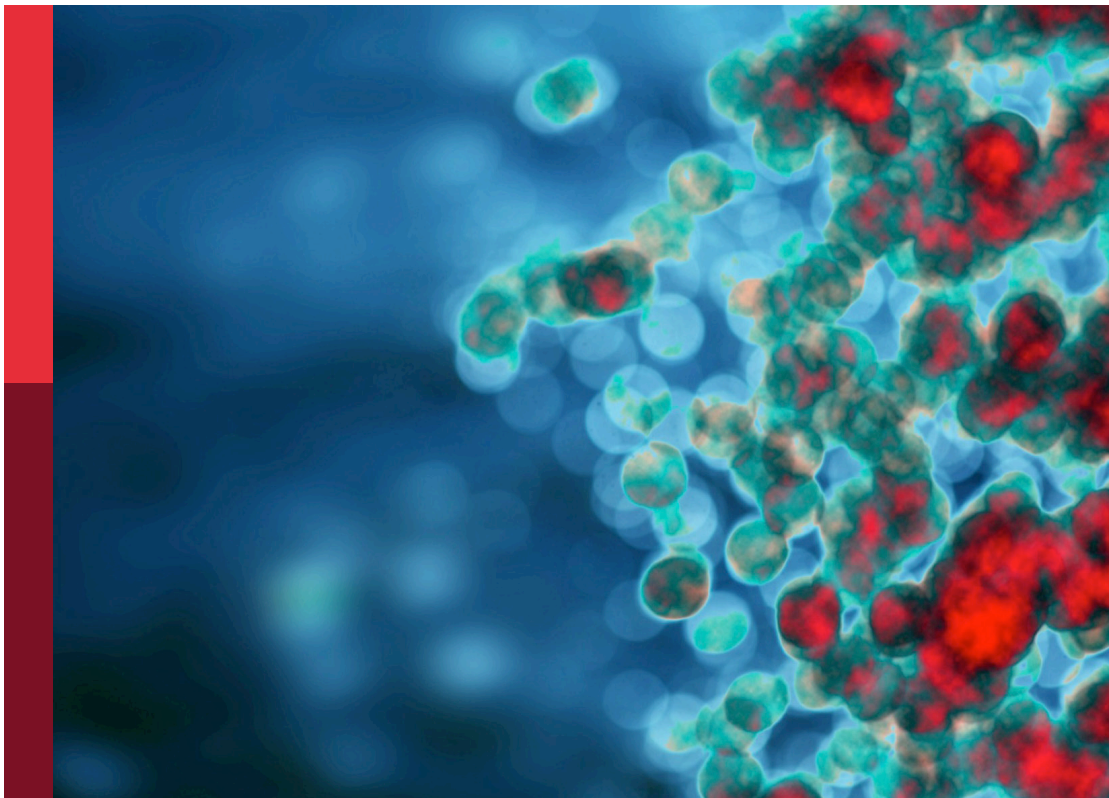
frontiers | Research Topics

Revolutionizing immunological disease understanding through single cell multi-omics technologies

Edited by
Jennie R. Lill, Veronica Anania, Ying Zhu and Zora Modrusan

Published in
Frontiers in Immunology



ISSN 1664-8714
ISBN 978-2-8325-6449-3
DOI 10.3389/978-2-8325-6449-3

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

Revolutionizing immunological disease understanding through single cell multi-omics technologies

Topic editors

Jennie R. Lill — Genentech Inc., United States
Veronica Anania — Genentech Inc., United States
Ying Zhu — Genentech Inc., United States
Zora Modrusan — Genentech Inc., United States

Citation

Lill, J. R., Anania, V., Zhu, Y., Modrusan, Z., eds. (2025). *Revolutionizing immunological disease understanding through single cell multi-omics technologies*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6449-3

Table of contents

frontiers | Frontiers in Immunology

TYPE Editorial
PUBLISHED 30 May 2025
DOI 10.3389/fimmu.2025.1628120

OPEN ACCESS

EDITED AND REVIEWED BY
Simon Mitchell,
Brighton and Sussex Medical School,
United Kingdom

*CORRESPONDENCE
Zora Modrusan
modrusan.zora@gene.com

RECEIVED 13 May 2025
ACCEPTED 15 May 2025
PUBLISHED 30 May 2025

CITATION
Zhu Y, Anania VG, Lill JR and Modrusan Z (2025) Editorial: Revolutionizing immunological disease understanding through single cell multi-omics technologies. *Front. Immunol.* 16:1628120. doi: 10.3389/fimmu.2025.1628120

Editorial: Revolutionizing immunological disease understanding through single cell multi-omics technologies

Ying Zhu, Veronica G. Anania, Jennie R. Lill and Zora Modrusan*

Genentech Inc., South San Francisco, CA, United States

KEYWORDS

single-cell omics, single cell RNA sequencing (scRNA-seq), cellular heterogeneity, tumor immunology, spatial technologies, targeted therapies, precision medicine

Editorial on the Research Topic
Revolutionizing immunological disease understanding through single cell multi-omics technologies

Introduction

Single cell approaches have revolutionized our understanding of many different immunological diseases and in this special issue, several new examples are highlighted that pave the way for new treatment options including better understanding of Th17 biology in fibrosis, HIV pathogenesis, IgA Nephropathy pathobiology, and identification of key immune subsets in rheumatoid arthritis and other autoimmune diseases. Spatial technologies have furthered our knowledge of gene expression and protein colocalization, opening new paths for studying tissue-specific cellular dynamics. Advancements like the SENSE method for cryopreserving whole blood have simplified single-cell analyses, making them more viable for clinical use. Innovative multiplexing strategies and developments in proteomics and Raman spectroscopy are broadening the capabilities of single-cell technologies, allowing for comprehensive profiling that can enhance precision medicine. Together, these advances highlight the potential of single-cell omics to lead the future of immunological research and clinical practice, facilitating the creation of novel therapeutic strategies and personalized medicine.

Understanding disease mechanisms through single-cell RNA sequencing

Deng et al. studied fibrotic skin diseases like keloids, hypertrophic scars, and scleroderma, characterized by excessive fibroblast growth and extracellular matrix buildup. They used fluorescence-activated cell sorting to isolate CD45+ immune cells from keloid and normal scar tissues, then applied scRNA-seq to map the immune cell

landscape. The study found a significant increase in Th17 cells, which promote fibroblast proliferation, collagen expression, and migration via IL-17A secretion. This increase in Th17 cells in other fibrotic conditions suggests a common mechanism in skin fibrosis, advancing the understanding of these diseases and identifying potential therapeutic targets.

Hong et al. used multiplex cytokine assays and scRNA-seq to explore immunological differences between ACPA-positive and ACPA-negative early rheumatoid arthritis (eRA) patients. They found that ACPA+ eRA patients had higher levels of interferon-gamma (IFN-γ) and interleukin-12 (IL-12), indicating a Th1 immune response. The study identified 17 distinct cell types, with notable expansions of IL1B+ proinflammatory monocytes, CD8+ CCL4+ T cells, and IL7R+ T cells in ACPA+ eRA. These cells showed upregulated IFN-γ response genes, suggesting enhanced IFN-driven monocyte-T cell interactions. IFITM2 and IFITM3 were identified as potential biomarkers for ACPA+ eRA. These findings indicate that ACPA+ eRA is characterized by a more active IFN-mediated immune response, potentially guiding personalized treatment strategies targeting type I and II interferon pathways.

scRNA-seq has been crucial in understanding immunity related to pathogenic invasion. Knoll et al. conducted a comprehensive study on immune cell reprogramming in people living with HIV (PLHIV), revealing persistent functional changes in monocytes even with long-term antiretroviral therapy (ART). Using various omics technologies, the study identified significant transcriptomic changes in monocytes, indicating an "anti-viral" state with upregulated IFN signaling pathways, like acute HIV infection. This suggests ongoing immune activation despite ART. The research also explored drug repurposing to reverse the pro-inflammatory monocyte phenotype in PLHIV, identifying sunitinib and doxycycline as promising candidates. These findings support the larger 2000-HIV multicenter cohort study, highlighting the need for transcriptomics to define disease endotypes and predict outcomes. The study emphasizes verifying chromatin-level differences via ATAC-seq and further exploring monocyte-mediated immune dysregulation. Despite a small sample size, it lays the groundwork for future research to refine therapeutic strategies and understand the immune landscape in PLHIV under ART.

Nephrotic syndrome (NS) is a severe form of IgA nephropathy (NS-IgAN) with unclear pathogenesis, marked by immune cell imbalances and kidney damage. Chen et al. used scRNA-seq on peripheral blood mononuclear cells and kidney cells from pediatric NS-IgAN patients to investigate this condition. They found increased intermediate monocytes (IMs) expressing VSIG4, MHC class II molecules, and genes related to oxidative phosphorylation. Classical and non-classical monocytes showed elevated CCR2, possibly linked to kidney injury. Two regulatory T cell subsets were identified, with Treg2 cells expressing high CCR4 and GATA3, potentially aiding kidney recovery. Podocyte injury was associated with increased CCL2, PRSS23, and epithelial-mesenchymal transition genes. PTGDS was suggested as a potential podocyte marker due to its decreased expression after injury. This study provides insights into NS-IgAN pathogenesis and could guide future targeted therapies.

Huang et al. describe the integration of single-cell and spatial transcriptomic analyses to unravel the cellular heterogeneity and molecular mechanisms underlying ulcerative colitis (UC), a chronic inflammatory bowel disease characterized by immune dysregulation. By identifying distinct monocyte subtypes and leveraging machine learning techniques, two key genes, GNG5 and TIMP1, were highlighted as central to UC pathogenesis. GNG5, downregulated in UC, is implicated in anti-inflammatory pathways such as PPAR signaling, while TIMP1, upregulated, exhibits pro-inflammatory effects and correlates with T cell exhaustion markers like TIGIT and CTLA4. Spatial transcriptomic data, immunohistochemical validation in human UC lesions, and experimental findings from a DSS-induced colitis mouse model confirmed these gene expression patterns. TIMP1 was further shown to co-localize with macrophages and promote Th17-driven inflammation, suggesting its dual role in macrophage activation and immune depletion. These findings provide a foundation for developing targeted therapeutic strategies aimed at mitigating chronic inflammation and immune dysfunction in UC.

Spatial technologies

Spatial technologies are essential for pinpointing specific cell types and gene expression locations. Moos et al. demonstrated how spatial-temporal single-cell transcriptomic sequencing can analyze genetic mutations in pulmonary epithelial nodes related to pulmonary fibrosis (PF) and interstitial lung diseases. Using a clinical PF dataset and a murine model with SP-C gene mutations, they investigated monocyte/macrophage changes in fibrotic lungs. The study found heterogeneous activation of CD68 + macrophages, especially near injury sites. Ingenuity Pathway Analysis showed asynchronous activation of extracellular matrix reorganization and ApoE signaling in alveolar macrophages. scRNA-seq identified pro-fibrogenic signaling from Trem2+ macrophages. Although genetic deletion of ApoE had limited impact on inflammation, the study suggests ApoE as a biomarker for active macrophages in tissue remodeling. These findings provide insights into macrophage heterogeneity and cell-cell interactions in fibrotic diseases.

Advances in spatial proteomics and protein colocalization are crucial for understanding cellular mechanisms and developing novel algorithms. Rhomberg-Kauert et al. introduced Molecular Pixelation (MPX), a method that provides spatial information on surface proteins in single cells, allowing for *in silico* graph representation of protein neighborhoods. To analyze this data modality, local assortative methods were adapted to assess spatial relationships, enabling evaluation of pairwise colocalization and similarity among multiple proteins. MPX was tested on datasets showing its ability to detect stimuli effects, such as T cells treated with a chemokine to study uropod formation, and cancerous B-cell lines treated with rituximab, providing insights into cell polarity.

This computational approach enhances understanding of immune responses and cell surface protein reorganization, potentially guiding new therapeutic designs. MPX offers high throughput, sensitivity, and three-dimensional analysis, surpassing traditional microscopy, and enabling deep phenotyping at single-cell resolution. The method can also analyze other biological spatial data represented as graphs, showcasing its broad applicability and potential to advance spatial proteomics.

Sample preparation is often a bottleneck in single-cell methodologies, especially with biofluids. Satpathy et al. developed the SENSE method for single-step cryopreservation of whole blood (WB), streamlining cell suspension preparation for scRNA-seq. This method overcomes the limitations of labor-intensive multistep processes unsuitable for clinical use. In a comparative analysis of six blood samples, the SENSE method produced highly viable single-cell suspensions, with 22,353 cells showing a viability rate of 86.3 ± 1.51%. It yielded high-quality transcriptomic profiles comparable to traditional PBMC methods and showed higher T-cell enrichment, allowing for detailed T-cell subtype characterization. Both methods captured transcriptional and cellular networks across cell types, with minimal batch effects, except in myeloid cells. The SENSE method's simplicity and effectiveness make it promising for widespread clinical and research adoption, facilitating single-cell assays and translational research.

Khoshbakht et al. introduced a label-free sample multiplexing strategy based on the souporcell algorithm, enabling cost-effective scRNA-seq and flow cytometry analyses of paired blood and skin samples. This protocol addresses the complexity and cost of current methods, applicable to both healthy and inflamed skin. It allows simultaneous RNA and protein analysis on the same lesion, reducing costs by 2–4 times. The strategy minimizes batch effects and examines the impact of different enzymatic incubation durations (1, 3, and 16 hours, with and without enzyme P) on flow cytometry results. It includes bioinformatic demultiplexing and a step-by-step guide, making it accessible for newcomers. This approach aims to enhance single-cell analysis accessibility, potentially extending to other dermatological disorders and aiding in understanding immune mechanisms and identifying new therapeutic targets.

Single-cell omics techniques for clinical samples have traditionally focused on genomic, transcriptomic, and more recently, proteomic methodologies. Raman spectroscopy has emerged as a complementary bioanalytical tool due to its ability to characterize the biophysical properties of biomolecules. Chadokiya et al. review how molecular tumor characterization is crucial for identifying predictive biomarkers to improve precision immunotherapy. However, challenges like tumor heterogeneity and limited biomarker efficacy hinder accurate treatment predictions. This study highlights label-free Raman spectroscopy as a non-invasive tool for profiling precision immunotherapy, capable of unifying various omics data. With its ability to distinguish immune cell types and detect molecular changes, Raman spectroscopy offers a promising approach for enhancing treatment prediction and monitoring in cancer care. As it evolves, Raman spectroscopy could become a cost-effective, patient-focused tool integrated into clinical practice for precise immunotherapy.

Future directions

Future developments in single-cell omics technologies are set to transform our understanding and treatment of immunological and tumor-related diseases. Integrating multi-omics data including genomic, transcriptomic, proteomic, and metabolomic at the single-cell level, along with spatial information, will provide a comprehensive view of cellular states and interactions, aiding in the discovery of new biomarkers and therapeutic targets. Advances in artificial intelligence and machine learning will be vital for analyzing the large datasets from these technologies, enabling predictive models for disease progression and treatment response. Enhancing spatial omics technologies will offer insights into cell and molecular organization within tissues, enriching our understanding of tissue architecture and function. Expanding single-cell techniques to less-studied cell types and rare diseases will uncover new areas of human health. Making single-cell omics cost-effective and user-friendly will be crucial for their integration into routine clinical practice, allowing personalized healthcare for more patients. These advancements will deepen our understanding of complex biological systems and lead to innovative therapies and precision medicine tailored to individual needs.

Author contributions

YZ: Writing – review & editing. VA: Writing – review & editing. JL: Conceptualization, Writing – original draft, Writing – review & editing. ZM: Conceptualization, Project administration, Writing – original draft, Writing – review & editing.

Conflict of interest

Authors YZ, VA, JL, and ZM were employed by Genentech Inc.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. For refining and polishing original draft manuscript.

Publisher's note

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 16 October 2023
DOI 10.3389/fimmu.2023.1231937

OPEN ACCESS

EDITED BY
David Perlman,
Health Outlook, United States

REVIEWED BY
Silvia Galván-Peña,
Harvard Medical School, United States
Kaushik Muralidharan,
Nationwide Children's Hospital,
United States

*CORRESPONDENCE
Qiu Li
liqiu809@hospital.cqmu.edu.cn
Haiping Yang
480847@cqmu.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 31 May 2023
ACCEPTED 20 September 2023
PUBLISHED 16 October 2023

CITATION
Chen Q, Jiang H, Ding R, Zhong J, Li L, Wan J, Feng X, Peng L, Yang X, Chen H, Wang A, Jiao J, Yang Q, Chen X, Li X, Shi L, Zhang G, Wang M, Yang H and Li Q (2023) Cell-type-specific molecular characterization of cells from circulation and kidney in IgA nephropathy with nephrotic syndrome.
Front. Immunol. 14:1231937.
doi: 10.3389/fimmu.2023.1231937

Cell-type-specific molecular characterization of cells from circulation and kidney in IgA nephropathy with nephrotic syndrome

Qilin Chen[1,2,3†], Huimin Jiang[1,2,3†], Rong Ding[4], Jinjie Zhong[1,2,3], Longfei Li[4], Junli Wan[1,2], Xiaoqian Feng[1,2,3], Liping Peng[1,2,3], Xia Yang[1,2,3], Han Chen[1,2], Anshuo Wang[1,2], Jia Jiao[1,2], Qin Yang[1,2], Xuelan Chen[1,2], Xiaoqin Li[1,2], Lin Shi[1,2], Gaofu Zhang[1,2], Mo Wang[1,2], Haiping Yang[1,2*] and Qiu Li[1,2,3*]

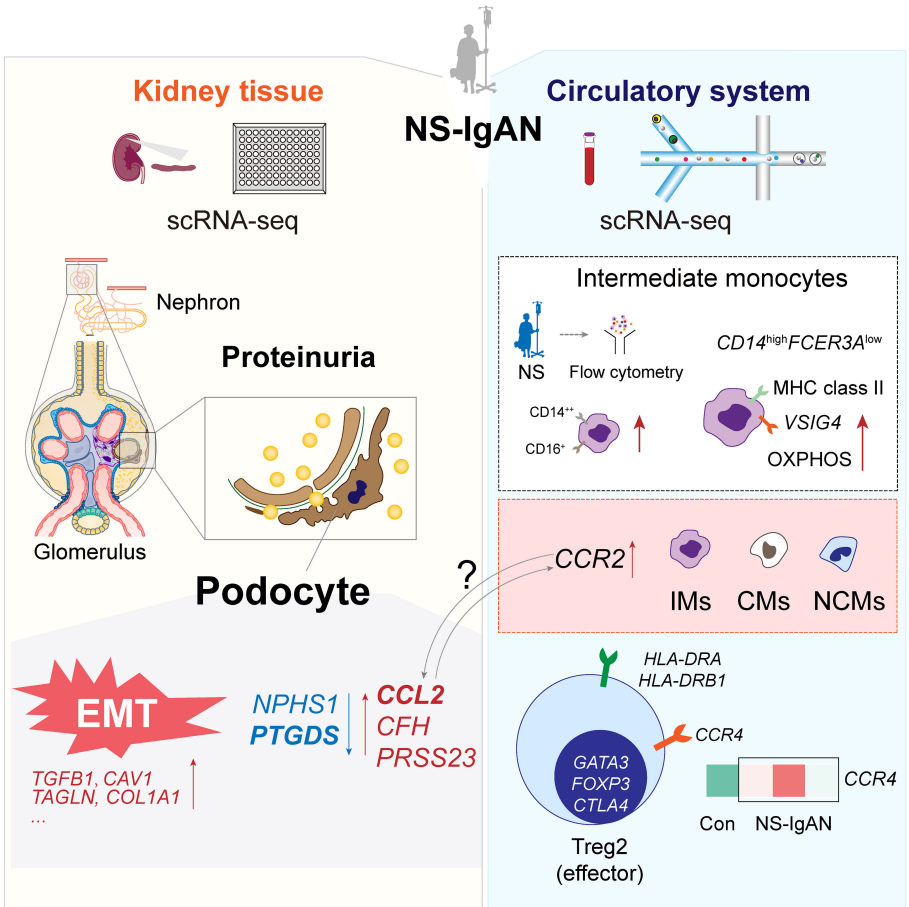
[1]Department of Nephrology, Children's Hospital of Chongqing Medical University, Chongqing, China, [2]National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing, China, [3]Chongqing Key Laboratory of Pediatrics, Chongqing, China, [4]Nanjing Jiangbei New Area Biopharmaceutical Public Service Platform Co. Ltd, Nanjing, Jiangsu, China

Nephrotic syndrome (NS) is a relatively rare and serious presentation of IgA nephropathy (IgAN) (NS-IgAN). Previous research has suggested that the pathogenesis of NS-IgAN may involve circulating immune imbalance and kidney injury; however, this has yet to be fully elucidated. To investigate the cellular and molecular status of NS-IgAN, we performed single-cell RNA sequencing (scRNA-seq) of peripheral blood mononuclear cells (PBMCs) and kidney cells from pediatric patients diagnosed with NS-IgAN by renal biopsy. Consistently, the proportion of intermediate monocytes (IMs) in NS-IgAN patients was higher than in healthy controls. Furthermore, flow cytometry confirmed that IMs were significantly increased in pediatric patients with NS. The characteristic expression of *VSIG4* and MHC class II molecules and an increase in oxidative phosphorylation may be important features of IMs in NS-IgAN. Notably, we found that the expression level of *CCR2* was significantly increased in the CMs, IMs, and NCMs of patients with NS-IgAN. This may be related to kidney injury. Regulatory T cells (Tregs) are classified into two subsets of cells: Treg1 (*CCR7*^{high}, *TCF7*^{high}, and *HLA-DR*^{low}) and Treg2 (*CCR7*^{low}, *TCF7*^{low}, and *HLA-DR*^{high}). We found that the levels of Treg2 cells expressed significant levels of *CCR4* and *GATA3*, which may be related to the recovery of kidney injury. The state of NS in patients was closely related to podocyte injury. The expression levels of *CCL2*, *PRSS23*, and genes related to epithelial-mesenchymal transition were significantly increased in podocytes from NS-IgAN patients. These represent key features of podocyte injury. Our analysis suggests that *PTGDS* is

significantly downregulated following injury and may represent a new marker for podocytes. In this study, we systematically analyzed molecular events in the circulatory system and kidney tissue of pediatric patients with NS-IgAN, which provides new insights for targeted therapy in the future.

KEYWORDS

IgA nephropathy, nephrotic syndrome, single-cell RNA sequencing, peripheral blood mononuclear cells, podocyte



GRAPHICAL ABSTRACT

1 Introduction

IgA nephropathy (IgAN) is the most common form of primary glomerulonephritis worldwide (1). With the increased use of kidney biopsy in children, there is a growing concern related to pediatric patients with IgAN (2). The diagnostic hallmark of IgAN is the predominance of IgA deposits in the glomerular mesangium. The pathogenesis of IgAN remains unclear but it is considered an autoimmune disease. A "multi-hit" hypothesis has been proposed to explain the pathogenesis of IgAN (1). This hypothesis outlines that increased levels of galactose-deficient IgA1 (Gd-IgA1), the production of autoantibodies, the deposition of immune complexes in the mesangial area of the glomeruli, and the secretion of cytokines, chemokines, and complements by mesangial cells ultimately lead to further kidney injury (3). In children, IgAN has long been considered to be a benign disease, with remission usually occurring after timely treatment (4). Nevertheless, patients with heavy proteinuria at biopsy often have a poor prognosis and there is a significant lack of therapeutic options for such patients (4). Nephrotic syndrome (NS) is a

relatively rare and serious presentation of IgAN (NS-IgAN). This condition has a low incidence (4), ranging from approximately 4% to 10% (4, 5). Therefore, there is limited information relating to the molecular mechanism of NS-IgAN, particularly in pediatric populations.

NS-IgAN patients not only exhibit typical manifestations of nephrotic syndrome (e.g., edema, heavy proteinuria, and hypoproteinemia), they also show glomerular mesangial IgA deposition and extensive foot process effacement on kidney biopsy (6). When considering NS-IgAN, the presentation of the typical histological characteristics of IgAN can indicate the existence of two glomerular diseases: IgAN and "podocytopathy" (6). The injury and loss of podocytes are key factors that contribute to progressive proteinuria and filtration dysfunction in IgAN (7). In addition, mesangial-derived humoral factors, such as tumor necrosis factor, complement components, and angiotensin II, may alter glomerular permeability in the presence of proteinuria (8). Existing literature suggests that abnormal crosstalk between circulating immune cells and kidney cells may contribute to the occurrence and progression of IgAN. Systematic dissection of the molecular characteristics of the circulating immune system and kidney tissue will help us to identify the mechanisms underlying proteinuria and provide potential strategies for improving the treatment and prognosis of pediatric patients with NS-IgAN.

Single-cell transcriptome sequencing (scRNA-seq) technology can achieve unbiased and high-throughput transcriptome sequencing at the single-cell level and has been applied in clinical research for various diseases, including kidney and autoimmune diseases (9). Previous scRNA-seq studies of IgAN have attempted to identify the molecular features of kidney cells or circulating immune cells in IgAN (10–12). However, there is still a significant lack of research on the systematic identification of the changes in circulating immune cells and kidney cells that occur in IgAN (9). To provide valuable insight into the molecular features of local tissues and the circulatory system, we simultaneously performed scRNA-seq on kidney biopsies and peripheral blood mononuclear cells (PBMCs) from pediatric patients with NS-IgAN. We found that the composition of monocyte subsets was altered in NS-IgAN and this was accompanied by the increased expression of *CCR2*. In addition, we found that *PTGDS* was significantly downregulated in podocyte injury. Collectively, our research provides a new understanding of the molecular characteristics of cell types in NS-IgAN and may facilitate the development of new targeted therapies.

2 Materials and methods

2.1 Ethical approval

This study for scRNA-seq analysis was reviewed and approved by the Institutional Review Board of the Children's Hospital of Chongqing Medical University (File Number: 2022 Research 124). The research for flow cytometric analysis was reviewed and approved by the Institutional Review Board of the Children's Hospital of Chongqing Medical University. The File Number is 2022 Research 35. Written informed consent was obtained from all participants and their guardians.

2.2 Sample collection and patient details

In this study, we collected PBMCs from the blood and kidney tissues from three pediatric patients diagnosed with IgAN by renal biopsy, all of whom had IgAN with nephrotic-level proteinuria (NS-IgAN). The PBMCs were harvested from samples that remained after routine tests had been performed. Patients were classified as NGC, SGC, or LGC according to their glucocorticoid treatment at the time of sampling. We also included two kidney tissues as control kidney samples: one sample was Wilms' tumor paracancerous tissue from a child (CTRL-1) while the other sample was from a resected duplex kidney (CTRL-2). All participants were enrolled from the Children's Hospital of Chongqing Medical University, and each sample was collected for scRNA-seq individually. For comparative analysis, we downloaded scRNA-seq datasets from the Gene Expression Omnibus database. The kidney tissue datasets included one healthy adult (CTRL-3), four adult patients with IgAN (merged as HU-IgAN) from GSE171314, and PBMC datasets from three healthy children (Con-1, Con-2, Con-3) from GSE206295. Detailed information relating to the participants is shown in Supplementary Table 1. To further investigate the changes in monocyte subsets in NS patients, we recruited a cohort of 13 healthy children, 26 NS patients with proteinuria, and 6 NS patients who were in remission after treatment. We collected blood from samples that remained after routine testing for flow cytometry.

2.3 Preparation of single-cell suspensions

To prepare PBMCs, we collected 2 mL of venous blood in EDTA collection vessels, which were then taken to the laboratory on ice. PBMCs were isolated using Ficoll medium (TBD, Tianjin, China) and cryopreserved according to the 10X genomics recommended protocol (CG00039). To dissociate the kidney tissue into single-cell suspensions, each fresh kidney sample was washed three times with Hanks' balanced salt solution and immediately stored in GEXSCOPE tissue preservation solution (Singleron Biotechnologies, Nanjing, China) at 2–8°C. Then, the tissue was cut into small pieces and digested in 1 mL of Tissue Dissociation Mix (Singleron Biotechnologies) at 37°C for 15 minutes before being passed through a 40 μm filter. After centrifugation at 3500 g for 5 minutes, cell pellets were resuspended in 1 mL of cold PBS. To remove red blood cells, 2 mL of GEXSCOPE Red Blood Cell Lysis Buffer (Singleron) was added into the cell suspension and incubated at 25°C for 10 minutes. Cells were then centrifuged at 300 g for 5 min and resuspended in cold PBS. Next, cells were stained with trypan blue (Beyotime, Shanghai, China) and counted with a TC20 automated cell counter (Bio-Rad, California, USA). Sample processing and analysis were permitted once cell viability exceeded 85%.

2.4 scRNA-seq library construction and sequencing

Several different sequencing methods were used for the samples of PBMCs and kidneys. For PBMCs, each sample was diluted to a final concentration of 700–1200 cells μl^{-1} and loaded onto a Chromium Single Cell Controller (10X Genomics, San Francisco, USA). The libraries for scRNA-seq were constructed using a Chromium Next GEM Single Cell 3′ GEM, Library and Gel Bead Kit v3.1 (10X Genomics) and then sequenced using an Illumina NovaSeq 6000 system. For each kidney sample, the single-cell suspension was adjusted to a concentration of approximately 300 cells μl^{-1}. A GEXSCOPE Single Cell RNA Library Kit (Singleron Biotechnologies) was then used to construct a single-cell RNA-seq library for kidney samples. The libraries were then sequenced with an Illumina HiSeq X 10 system. Each sample of PBMCs and kidney tissue was processed independently.

2.5 scRNA-seq data processing

Raw sequencing reads from PBMCs were processed using Cell Ranger (version 6.0.0), including demultiplexing, genome alignment (GRCh38), barcode counting, and unique molecular identifier (UMI) processing. Similarly, raw data from the kidney tissues were processed by Celescope (version 1.10.0). We then used the Seurat (version 4.1.0) package to perform downstream analysis. To exclude low-quality cells, the cells were filtered by gene counts and UMI counts; cells with a high mitochondrial content were removed. Detailed information relating to the quality control (QC) threshold settings is given in Supplementary Table 2. After cell filtering, 53,571 PBMCs and 47,602 renal cells were captured for downstream analysis. Next, we used DoubletFinder (version 2.0.3) to identify doublets and removed clusters with a high proportion of doublets. To remove batch-effects, we integrated sample datasets via the "integrate" function in Seurat. Principal component analysis (PCA) was then performed on the top 2000 highly variable genes; the top 20 PCs were used for subsequent analysis. We then used a graph-based clustering algorithm to identify clusters, thus allowing us to construct a K-Nearest Neighbor (KNN) graph by Euclidean distance. The Louvain algorithm was used to group cells and optimize modules. To display the distribution of cells by status, we used the uniform manifold approximation and projection (UMAP) algorithms to visualize clustered cells. First, cells were identified by SingleR (v1.6.1). This software compared the transcriptome of each cell cluster to various reference datasets (e.g., human primary cell atlas, Blueprint/ENCODE, Database of Immune Cell Expression, Novershtern hematopoietic data, and Monaco immune data). In order to exclude the influence of automatic assignment, we also manually adjusted the identities of clusters by combining data with the expression levels of canonical marker genes and existing annotated scRNA-Seq data (Supplementary Tables 3–7, 9).

2.6 Differentially expressed genes and enrichment analysis

Differentially expressed genes (DEGs) were computed using the FindMarkers function of Seurat. DEGs were defined as genes detected in at least 10% of cells, with a threshold of 0.25 log fold change using the Two-tailed Wilcoxon Rank-Sum Test with $P < 0.01$ following Bonferroni correction. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed with the Metascape web tool (www.metascape.org) and ClusterProfiler (version 4.2.2).

2.7 Calculation and Analysis of epithelial-mesenchymal transition scores

We used cell scores to evaluate the degree to which individual cells expressed a certain predefined expression gene set. This allowed us to define meaningful functions and states. The cell scores were calculated using the Seurat function "AddModuleScore", which calculated the average expression of genes from the predefined gene set in the respective cell. The control gene sets were randomly selected based on aggregate expression level bins. The final gene set score was obtained by subtracting the control score from the predefined gene set score. We then used several well-defined EMT markers (Supplementary Table 8) to define the EMT score.

To assess the statistical significance of scores, for each NS-IgAN patient or HU-IgAN group, the EMT scores were compared with that of the control group (CTRL) using the two-tailed Mann-Whitney *U*-test. Differences with a *P* value <0.05 were considered significant, $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$, and $^{****}P<0.0001$.

2.8 Ligand–receptor analysis

CellphoneDB was used with default parameters to reveal changes in interactions between different cell types.

2.9 Pseudo-time analysis

R package Monocle2 (version 2.18.0) was used to perform pseudo-time analysis. To clarify the role of genes in cell fate decisions, branched expression analysis modeling (BEAM) from Monocle2 was applied.

2.10 GWAS expression analysis

The defined IgAN susceptibility genes were obtained based on previous genome-wide association studies (GWASs) in IgAN combined with NephQTL and eQTL (cf) online analysis tools, and were provided in Supplementary Table 10. Cell-type specific

expression of IgAN susceptibility genes was calculated by the average relative logarithmic expression values.

2.11 Cell culture and treatment

Human immortalized podocytes were provided from the lab of Zhihong Liu, and the cells were cultured as previously described (13). Cells were grown at the permissive temperature of 33°C (in 5% CO_2) and differentiated at 37°C (in 5% CO_2). After differentiating for 7-14 days, podocytes were treated with 0.25 μg mL^{-1} doxorubicin (Sigma-Aldrich, Saint Louis, USA) and 0.5 μg mL^{-1} doxorubicin for 24 hours.

2.12 Immunofluorescence

The kidney tissue was dewaxed and then heat-mediated antigen repair was performed in sodium citrate solution (pH=6.0) for 15min, and blocked with goat serum. Cultured podocytes growing on a glass slide were fixed in 4% paraformaldehyde for 15 min. The sections and cells were incubated with the following primary antibodies: Prostaglandin D Synthase (PGDS) (ABclonal, Wuhan, China), anti-nephrin (PROGEN, Darmstadt, Germany) at 4°C overnight. Then incubated with the appropriate secondary antibody for 45 or 60 min at room temperature: anti-guinea pig IgG antibody conjugated with Alexa Fluor 568 (Invitrogen, California, USA), anti-rabbit IgG antibody conjugated with Alexa Fluor 633 (Invitrogen). Nuclei were counterstained with Hoechst (Thermo Fisher, Boston, USA). Fluorescence signals were viewed under a fluorescence microscope (Nikon A1R, Tokyo, Japan). NIS-Element (version 5.5) was used to quantify PGDS and nephrin staining intensity.

2.13 Flow cytometric analysis

200μL whole blood was incubated with CD14-PE (BioLegend, California, USA) and CD16-APC (BioLegend, California, USA) antibodies. After staining for 20 minutes at room temperature in the dark, erythrocytes (BD Pharmingen, New Jersey, USA) in the samples were lysed by incubation with lysing solution for 5 minutes. Following centrifugation (300g/5 minutes, 4°C) and washing with PBS, cells were then examined using BD FACSCanto™. The datasets were analyzed using FlowJo (version 10.4.2).

2.14 Statistical analysis

Data are expressed as means ± standard deviation (SD). The EMT scores and gene expression levels were analyzed by the two-tailed Mann-Whitney *U*-test in SPSS 26.0. Differences in staining intensity between the two groups were analyzed by the two-tailed Student's *t* test. All other analyses involved the two-tailed Wilcoxon Rank-Sum Test. Significance was defined as $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$. No specific indication is given if data were not significant. Graphs were generated by ggplot2 (version 3.3.5) and pheatmap (version 1.0.12) in R.

3 Results

3.1 Construction of a cell atlas of renal tissue and circulating immune cells in children with NS-IgAN

scRNA-seq was performed on PBMCs and kidney cells from three pediatric patients, all of whom had IgAN with nephrotic-level proteinuria, hypoalbuminemia, and hyperlipidemia (NS-IgAN) (Figure 1A; Supplementary Table 1). Kidney biopsy revealed the mesangial deposition of IgA and extensive or partial foot process effacement (Supplementary Figures 1A–C). It is worth noting that all three patients showed the pathological characteristics of mesangial hypercellularity. These patients had accepted different treatments at the time point of sampling, one patient was not on glucocorticoid (GC) therapy (NGC), one patient was on short-term GC therapy (SGC), and one patient was on long-term GC therapy (LGC) (Supplementary Figures 1D–F). Following the scRNA-seq of PBMCs, the raw data of three NS-IgAN patients and three healthy controls (Cons) from GSE206295 were merged and 53,571 PBMCs were captured following QC (Figure 1B; Supplementary Table 2). We annotated 30 cell types of PBMCs based on the expression of canonical markers, including $CD4^{+}$ T cells (*CD3D*, *CD4*), $CD8^{+}$ T cells (*CD3D*, *CD8A*), natural killer (NK) cells (*KLRB1*, *GNLY*), B cells (*CD19*, *CD79A*), myeloid cells (*CD68*, *LST1*), γδT cells (*CD3D*, *TRDC*), and megakaryocytes (*PPBP*, *PF4*) (Supplementary Figure 2A; Supplementary Table 3). For the kidney dataset, we downloaded data from GSE171314 as the HU-IgAN group, including four adult IgAN patients with hematuria and no nephrotic-level proteinuria (Supplementary Table 1). In addition, we performed scRNA-seq on the kidney tissues from two pediatric patients and included a single-cell public database (GSE171314) from one adult kidney as a control group (CTRL) (Figure 1A). The data from a total of ten kidney samples from three groups (CTRL, NS-IgAN, HU-IgAN) were integrated and 47,602 kidney cells were captured after QC (Supplementary Table 2). There were 26 clusters after dimension reduction (Supplementary Figure 2B). Then, 16 major cell types were annotated with the expression of canonical marker genes (Figure 1C; Supplementary Table 4). The expression levels of canonical marker genes for all cell types in the PBMCs and kidney cells are shown in Figures 1D, E.

3.2 The proportion of intermediate monocytes expressing MHC class II molecules is significantly increased in the disease

Our investigation of the composition ratio of each sample in the same cell type found that the composition of SGC and LGC samples in myeloid cell types was prominent (Figure 1B). In response, we re-

10.3389/fimmu.2023.1231937

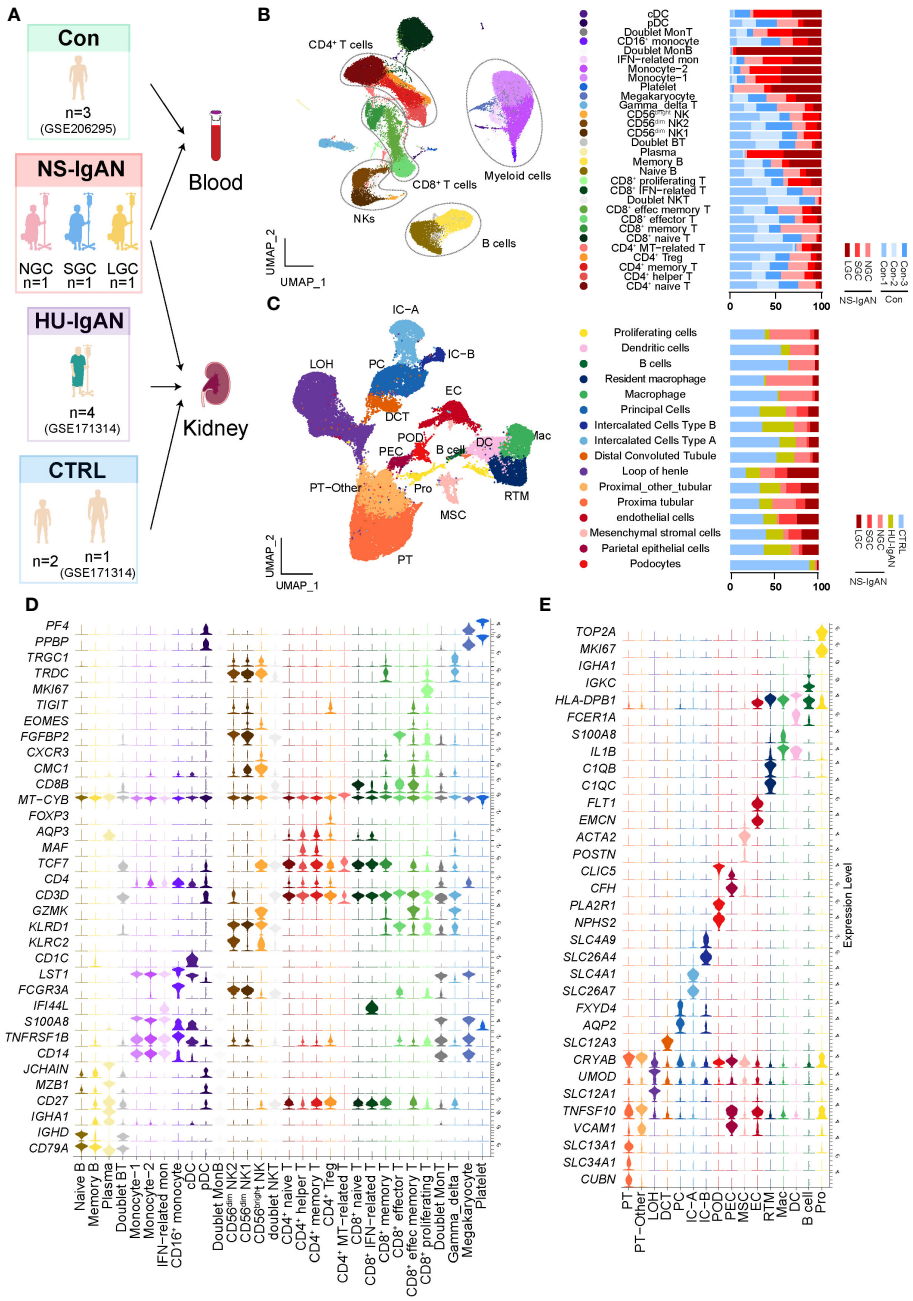


FIGURE 1
The landscape of PBMCs and kidney cells in NS-IgAN patients and healthy controls identified by single-cell transcriptomic analysis. **(A)** Schematic diagram of the study design for scRNA-seq. Con, n=3; NS-IgAN, n=3 (NGC, n=1; SGC, n=1; LGC, n=1); CTRL, n=3; HU-IgAN, n=4. **(B)** Distribution of 30 cell clusters in PBMCs. The figure on the left is a two-dimensional UMAP visualization of PBMCs. Different colors represent 30 clusters. The figure on the right is the percentage of each sample in each cell type. **(C)** Distribution of 16 cell types in kidney cells. The figure on the left is a two-dimensional UMAP visualization of kidney cells. Different colors represent 16 cell types. The figure on the right is the percentage of each sample in each cell type **(D)** Violin plot showing marker genes for each PBMC cell type. **(E)** Violin plot showing marker genes for each kidney cell type.

clustered the myeloid cell types into 14 cell clusters (Figure 2A). According to the expression of canonical marker genes for each cluster, we defined the cell types of clusters (Figure 2B, Supplementary Table 5). According to the expression levels of *CD14* and *FCER3A* (encoding CD16 protein), we respectively defined classical (CMs, *CD14*^{high}, *FCER3A*^{neg}), intermediate (IMs, *CD14*^{high}, *FCER3A*^{low}), and non-classical monocytes (NCMs, *CD14*^{low}, *FCER3A*^{high}) (Figure 2B) (14). Of these, the CMs were composed of four cell clusters. Cluster 5 expressed interferon-related genes (*ISG15*, *MX1*) and was defined as CD14^{+} IFN-related monocytes (IFN-Mon) (Figure 2B). Notably, we found that cluster 9 expressed *CCL4*, *IL1B*, *ICAM1*, and *CXCL2*. This cluster was defined as inflammatory monocytes (INF-M) (Figure 2B). Although the proportion of INF-M in myeloid cells did not differ significantly between disease and healthy samples (Figure 2C), we found that this group of cells was dominant in the disease group (Figure 2A).

10.3389/fimmu.2023.1231937

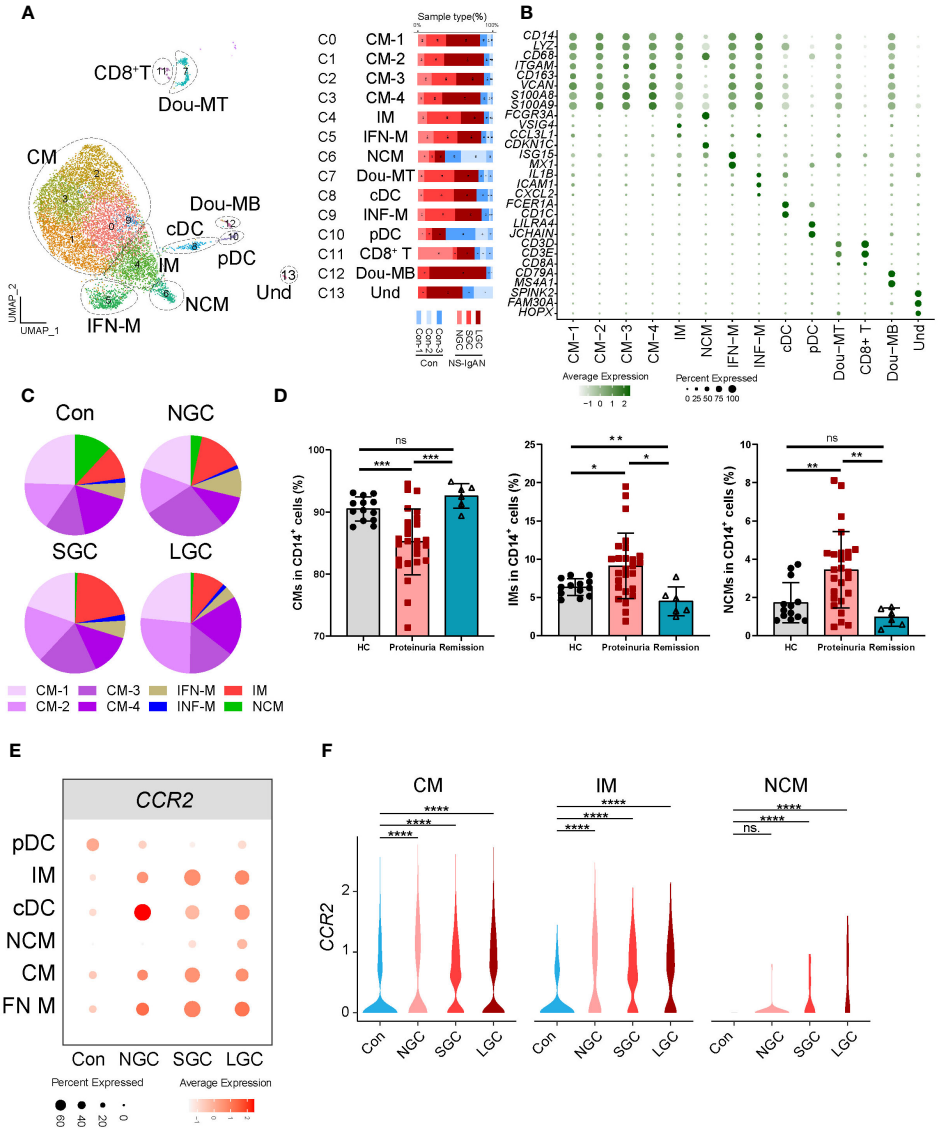


FIGURE 2
Molecular characterization of myeloid cells in NS-IgAN. **(A)** 13 clusters were visualized by UMAP plotting through re-clustering analysis of myeloid cells. **(B)** Violin plots showing expression of marker genes in 13 clusters. **(C)** The pie chart shows the proportion of different monocyte subsets in monocytes. **(D)** The proportion of monocyte subsets in NS patients, from left to right are CMs, IMs, and NCMs in flow cytometry. **(E, F)** Expression level of *CCR2* in monocyte subsets from different samples. All differences with $P < 0.05$ are indicated, **P*<0.05, ***P*<0.01, ****P*<0.001, and *****P*<0.0001, ns means no significant difference.

Figure 2C shows the composition ratio of each monocyte subset to $CD14^+$ cells in the controls and three cases. In NS-IgAN, the proportions of CM 1-4 and IMs tended to increase, and the proportion of NCMs tended to decrease; however, these trends were not statistically significant (Figure 2C, Supplementary Figure 3A). Previous evidence proved that monocyte subsets undergo changes in IgAN (15); our data suggested that similar changes may also exist in NS-IgAN. Since NS is the main clinical manifestation of NS-IgAN, we collected the peripheral blood of 26 NS patients with proteinuria, 6 NS patients with remission after treatment, and 13 healthy children to investigate changes in monocyte subsets in NS by flow cytometry. Monocytes were divided into classical monocytes (CMs, $CD14^{++}CD16^-$), intermediate monocytes (IMs, $CD14^{++}CD16^+$), and non-classical monocytes (NCMs, $CD14^+CD16^{++}$) through flow cytometry by the expression levels of CD14 and CD16 (Supplementary Figure 3B). The proportion of CMs in NS patients with proteinuria was significantly lower than that in healthy children, and the proportion of IMs and NCMs was significantly increased (Figure 2D). Following the glucocorticoid-induced remission of proteinuria, the proportion of CMs and NCMs returned to normal levels; the proportion of IMs was even lower than healthy controls (Figure 2D). Although the results of scRNA-seq and flow cytometry were different when compared between CMs and NCMs, the results consistently indicated a trend for an increased proportion of IMs. We found that IMs were characterized by high expression levels of *VSIG4*, *HLA-DPA1*, *HLA-DPB1*, and other MHC class II molecules (Supplementary Figure 3C). KEGG and GO enrichment analysis of the IMs showed

that their function was mainly related to cytokine-cytokine receptor interaction, the chemokine signaling pathway, antigen processing and presentation, and MHC class II protein complex assembly (Supplementary Figures 4A, B). DEG analysis of IMs in the Cons and NS-IgAN groups showed that the expression of *VSIG4*, *LYZ*, *HLA-C*, *FN1*, and *FCGRT* was significantly increased in NS-IgAN (Supplementary Figure 4C). The significant up-regulation of DEGs in the IMs of NS-IgAN patients were related to oxidative phosphorylation by KEGG enrichment analysis (Supplementary Figure 4D). These results suggest that increased oxidative phosphorylation may be an important feature of IMs in NS-IgAN.

3.3 *CCR2* was significantly expressed in the IMs of NS-IgAN patients

DEG analysis of monocytes in the Cons and NS-IgAN groups showed that the expression levels of *CCR2* were significantly higher in the CMs (CM1-4), IMs, and NCMs of NS-IgAN patients (Figures 2E, F). CCR2 mediates monocyte chemoattractant recruitment to inflammatory regions and promotes the production of inflammatory cytokines (16). It has been reported that *Ccr2*-deficient mice with adriamycin-induced nephropathy showed reduced levels of injury, along with reduced macrophage and fibrocyte infiltration and inflammation in the kidney (17). To investigate the effect of high expression levels of *CCR2* on monocytes in NS-IgAN, we divided monocytes into $CCR2^{+}$ monocytes and $CCR2^{-}$ monocytes for further analysis. We found that the proportion of $CCR2^{+}$ monocytes was slightly increased in NS-IgAN, although this was not statistically significant (Supplementary Figure 4E). We also found that compared with $CCR2^{-}$ monocytes, $CCR2^{+}$ monocytes expressed high levels of *LYZ* and *HLA-DQA2* (Supplementary Figure 4F). *HLA-DQA2* encodes HLA class II alpha-chain proteins to constitute MHC class II molecules. A recent GWAS study of a Chinese IgAN cohort revealed significant associations between three HLA polymorphisms, thus indicating the extensive involvement of HLA-mediated immunity in IgAN development (18). Collectively these results suggested that the proportion of $CCR2^{+}$ monocytes is increased in NS-IgAN and that this may participate in the pathogenetic process.

3.4 $GATA3^{+}$ effector Tregs expressed high levels of *CCR4* in NS-IgAN

The proportion of Tregs in the peripheral blood of IgAN was significantly lower than that in controls, thus an increase in the proportion of Treg cells could improve clinical prognosis (19). We performed a re-clustering analysis of $CD4^{+}$ T cells and identified 13 clusters (Supplementary Figure 5A); these were defined as eight cell types by the expression of canonical marker genes (Figures 3A, B; Supplementary Table 6). Of these, two clusters expressed *FOXP3* and *IL2RA*; we defined these as Treg1 and Treg2 cells (Figure 3C). According to the canonical marker genes expressed by different subsets of Tregs (20, 21), we found that Treg1 cells expressed high levels of *CCR7*, *SELL*, and *TCF7*; which may indicate that the cells are in a naive state; Treg2 cells expressed high levels of *CCR4* and HLA class II molecules, which may indicate that the cells are in an effector state (Figure 3D). The two groups of Treg cells consistently expressed some classical genes, including *TIGIT*, *IKZF2H*, and *RTKN2*, but also had molecular expression characteristics (Supplementary Figures 5B, C). We found that the proportion of Treg2 cells in NS-IgAN tended to be higher than that in the Cons group, while the proportion of Treg1 cells tended to be lower, although these differences were not statistically significant (Figure 3E). The results of GO enrichment analysis suggested that the functions of Treg2 cells were mainly related to peptide antigen binding, MHC protein complex, and antigen processing and presentation (Figure 3F). In Treg2 cells from NS-IgAN patients, we found that the expression levels of genes such as *FOS*, *JUN*, and *JUNB* were decreased, while those of *CCR4* were upregulated (Figure 3G). Recent research reported that a significant increase in the number of $GATA3^{+}$ Tregs in the kidney was closely related to disease remission (22). CCR4 is known as an important chemokine receptor that promotes the infiltration of $GATA3^{+}$ Tregs in the kidney during the later phases of injury (22). Our results suggest that there may be an increased proportion of $CCR4^{+}$ $GATA3^{+}$ Tregs in the circulation of NS-IgAN patients, thus indicating that dynamic changes of this special subset of Tregs may be involved in recovering from kidney injury.

Considering the important role of B cells in IgAN, we also performed a re-clustering analysis of B cells (Supplementary Figure 6A). We defined 12 clusters as 6 cell types by the expression of marker genes (Supplementary Figures 6B, C, Supplementary Table 7). Consistent with expectations, the *IGHA1* encoding IgA was mainly expressed in plasma cells. Pseudo-time analysis of B cells suggested that plasma cells at the end of differentiation were increasingly dominant in the disease, and expressed *IGHA1* and *IGHA2* (Supplementary Figures 6D–G). We compared the expression of key genes related to IgAN in diseased and healthy B cells; analysis suggested that *C1GALT1* expression was downregulated in IgAN; this may be related to the formation of Gd-IgA1 (Supplementary Figure 6H). In addition, the expression of IgG and IgA-related genes in the B cells of SGC and LGC patients was upregulated (Supplementary Figure 6H). Findings related to Treg cells and plasma cells are restrictive and limited by the number of captured cells, meaning that further research needs to be undertaken to validate these findings.

3.5 Podocytes in NS-IgAN expressed high levels of *CCL2* and EMT characteristics

Understanding alterations in the molecular characteristics of kidney cells may help us to understand the downstream mechanisms of kidney injury in NS-IgAN. Considering the clinical characteristics of patients with NS-IgAN who have "nephrotic-range" proteinuria, we focused on the podocyte cluster that significantly expressed *NPHS2* and *FGF1*. Podocytes specifically expressed NPHS2 and FGF1 at both protein and transcriptional levels (Figures 1E, 4A, B). Unfortunately, only two

10.3389/fimmu.2023.1231937

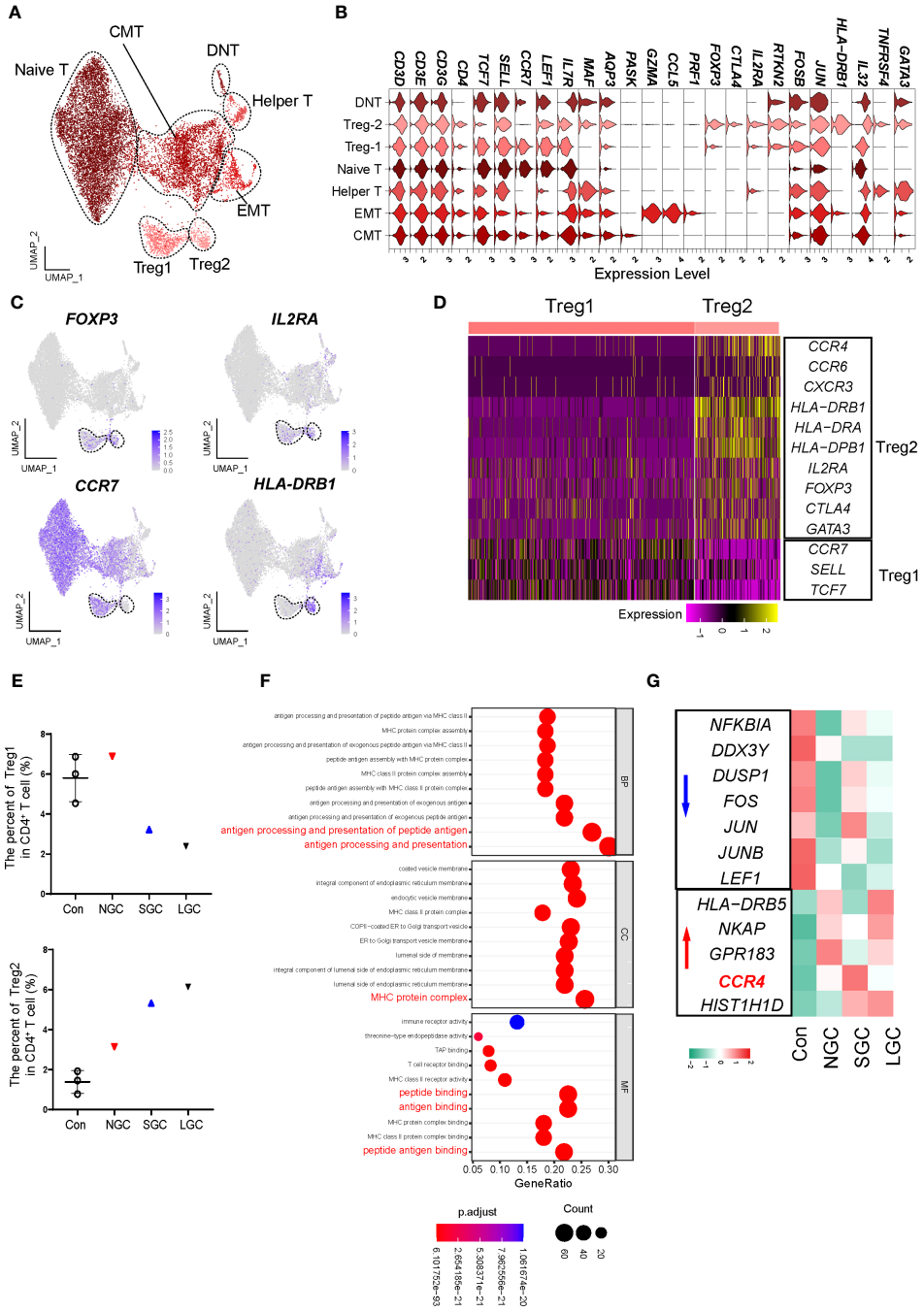


FIGURE 3
Molecular characterization of CD4$^+$ T cells in NS-IgAN. **(A)** 7 distinct subsets were visualized by UMAP plotting through re-clustering analysis of CD4$^+$ T cells. **(B)** Violin plots showing expression of marker genes in 7 subsets. **(C)** UMAP plots showing color-coded expression of four representative markers, *FOXP3*, *IL2RA*, *CCR7*, and *HLA-DRB1* in CD4$^+$ T cells. **(D)** Heatmap exhibiting the differential expressing genes of Treg1 and Treg2. **(E)** The percent of Treg1 and Treg2 in CD4$^+$ T cells. **(F)** GO terms identified by differential pathway enrichment and GO analysis via comparison of Treg2 cells vs. others. GO terms are labeled with name, and sorted by $-\log_{10}$(P) value. A darker color indicates a smaller P value. Interesting terms are labeled in red. **(G)** Heatmap exhibiting DEGs of Treg2 in Con, NGC, SGC, and LGC.

podocytes were captured in the SGC patient, and we did not include this group in the subsequent comparative analysis (Supplementary Table 4). We found that some genes showed a downward trend in terms of their expression by podocytes in the NS-IgAN group, including the canonical marker genes (*NPHS1* and *CLIC5*), the genes encoding membranous nephropathy-associated autoantigens (*PLA2R1* and *THSD7A*), and genes that have not been extensively investigated in podocytes, such as *PCOLCE2* and *PTGDS* (Figure 4C). Our data indicated that *CCL2* was highly expressed in the podocytes of NS-IgAN patients (Figure 4C). It has been reported that the expression of *CCL2* in podocytes is closely related to podocyte injury and proteinuria (24). In the adriamycin-induced

10.3389/fimmu.2023.1231937

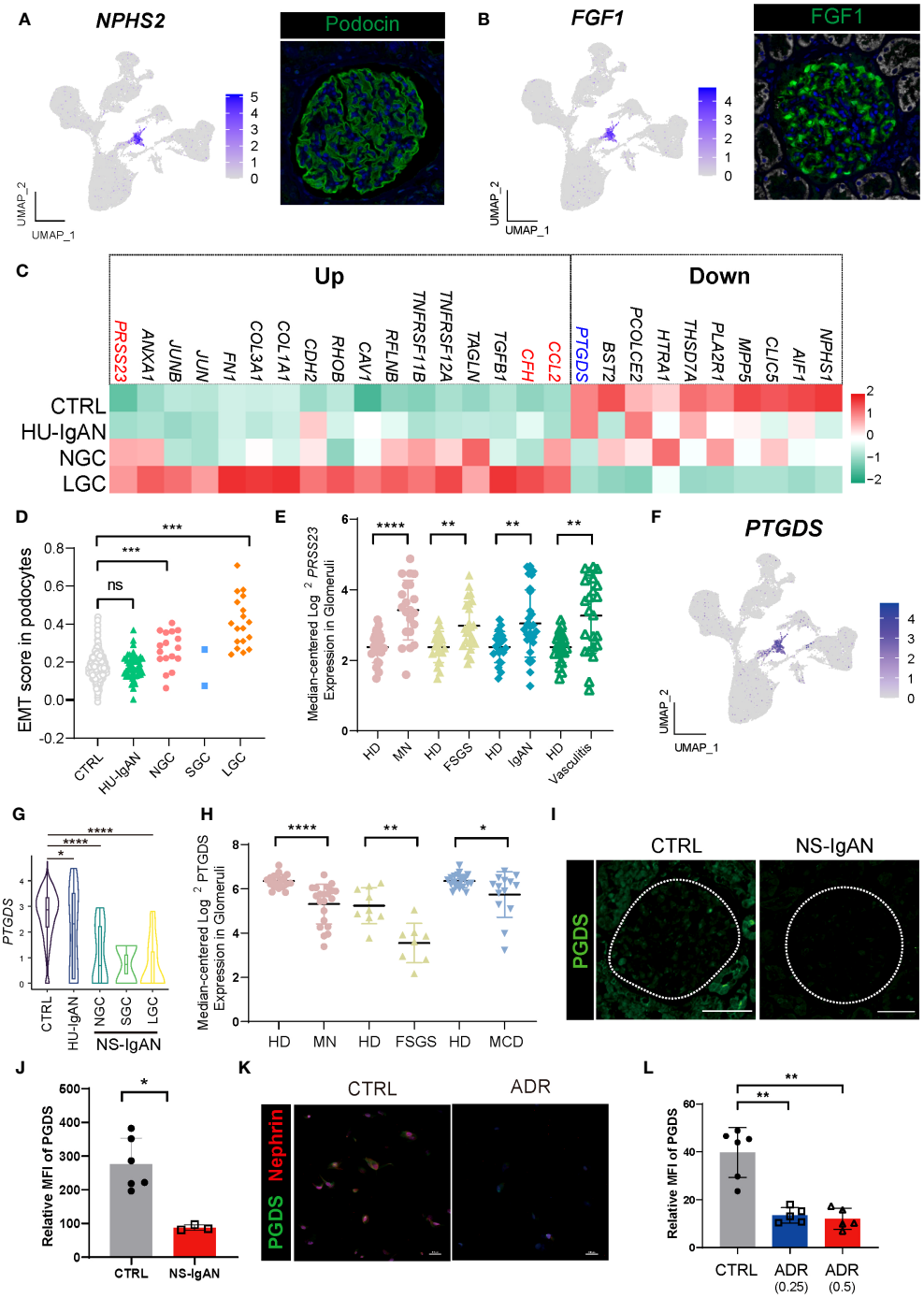


FIGURE 4
Molecular characterization of podocytes in NS-IgAN. **(A, B)** UMAP plots showing expression of *NPHS2* **(A)** and *FGF1* **(B)** in kidney and fluorescence staining of podocin and FGF1 in healthy kidney from The Human Protein Atlas database (https://www.proteinatlas.org/). **(C)** Heatmap showing up- and down-regulated DEGs of podocytes in the CTRL, HU-IgAN, NGC, and LGC groups. **(D)** EMT gene set scores for podocytes in CTRL, HU-IgAN, NGC, SGC, and LGC groups. **(E)** *PRSS23* mRNA expression in glomeruli of human biopsy specimens with pathological diagnosis of MN, FSGS, IgAN, or vasculitis disease compared with normal kidneys. Data are from previously published microarray studies by Ju et al. (2013) (23) and were subjected to further analysis using Nephroseq. A two-tailed Mann–Whitney *U* test was used for each comparison. All differences with $P < 0.05$ are indicated, *P<0.05, **P<0.01, ***P<0.001, and ****P<0.0001. **(F)** UMAP plots showing the expression of *PTGDS* in the kidney. **(G)** Statistical analysis of the expression level of *PTGDS* in podocytes in HU-IgAN, NGC, and LGC compared with CTRL. Since only 2 podocytes were captured in SGC patients, the comparison between SGC and CTRL was not performed. **(H)** *PTGDS* mRNA expression in glomeruli of human biopsy with pathological diagnosis of MN, FSGS, IgAN, or vasculitis disease compared with normal kidneys. Data are from previously published microarray studies by Ju et al. (2013) (23) and were subjected to further analysis using Nephroseq. **(I)** Immunofluorescence staining of PGDS in glomeruli of the NS-IgAN patient and the healthy control child. **(J)** Statistical analysis of PGDS glomerular mean fluorescence intensity quantified using NIS-Elements software. **(K)** Representative pictures of PGDS and nephrin staining in podocytes exposed to vehicle (CTRL) or Adriamycin (ADR). **(L)** Statistical analysis of the mean fluorescence intensity of PGDS between vehicle and ADR podocytes. A two-tailed Mann–Whitney *U* test was used for each comparison. All differences with $P < 0.05$ are indicated, *P<0.05, **P<0.01, ***P<0.001, and ****P<0.0001. ns means no significant difference.

mouse model of nephropathy, researchers found that CCL2 in the kidney may recruit the infiltration of inflammatory and pro-fibrotic bone marrow-derived cell populations through its receptor CCR2; furthermore, a deficiency of *Ccr2* in mice can ameliorate renal injury (17). We detected high expression levels of *CCL2* in NS-IgAN glomerular podocytes and significantly increased expression levels of *CCR2* in circulating monocytes (Figure 2F), thus indicating that crosstalk may occur in patients between podocytes and circulating monocytes through the CCL2-CCR2 ligand receptor. Furthermore, *CFH* was highly expressed in the podocytes of LGC (Figure 4C). Complement factor H (CFH) is one of the important circulating regulators of the alternative pathway, serves as an essential cofactor for complement factor I (CFI)-mediated C3b cleavage (25). Podocytes produce functionally active complement components, such as CFH; these could influence the local glomerular complement activation and regulation (26).

We also found that genes related to epithelial-to-mesenchymal transition (EMT) were significantly upregulated in NS-IgAN, including *TGFB1*, *CAV1*, *TAGLN*, and *COL1A1* (Figure 4C). We constructed an EMT gene set (Supplementary Table 8) and compared the gene set scores of podocytes between different groups. The EMT scores of podocytes in the NGC and LGC were significantly increased, while the EMT scores of podocytes in HU-IgAN did not change significantly (Figure 4D). EMT is an important feature of podocyte injury (27). Recent studies have found that podocytes in the urine of patients with NS also have EMT characteristics (28). Our results suggest that EMT is an important molecular feature of NS-IgAN podocytes. Interestingly, the expression of *PRSS23* was significantly elevated in NS-IgAN patients and to a greater degree than in HU-IgAN patients (Figure 4C). Bulk RNA-seq results also confirmed the reduced expression of *PRSS23* in various glomerular diseases (Figure 4E). These results suggest that a novel serine protease encoded by *PRSS23* may be related to podocyte injury.

3.6 The expression of *PTGDS* significantly decreased in damaged podocytes

By performing the scRNA-seq on kidney tissues, we found that *PTGDS* was only specifically expressed in the podocytes (Figure 4F). In addition, *PTGDS* was significantly downregulated in NS-IgAN podocytes ($P < 0.0001$) (Figure 4G). The glomerular transcriptome sequencing of various glomerular diseases, including membranous nephropathy (MN), focal segmental glomerulosclerosis (FSGS), and minimal change disease (MCD) confirmed that the transcriptional level of *PTGDS* was significantly decreased (Figure 4H). Hence, we reviewed the previously published scRNA-seq data of kidney tissues and found that *PTGDS* only expressed at high levels in human kidney podocytes (Supplementary Figures 7A, B). By using immunofluorescence and immunohistochemistry experiments, we were able to confirm the downregulation of PGDS (prostaglandin D2 synthase, encoded by *PTGDS*) in the glomeruli of patients with NS-IgAN (Figures 4I, J; Supplementary Figures 8A, B). The induction of injury in human immortalized podocytes *in vitro* (by applying doxorubicin) also led to a significant reduction in the expression levels of PGDS (Figures 4K, L; Supplementary Figures 8C, D). PGDS, also known as β-Trace protein (BTP), is an emerging novel marker for glomerular filtration rate (29). PGDS is produced *de novo* by both the glomeruli and LOHs in monkey kidneys (30), although the handling process of BTP in the kidney is uncertain. BTP is an eicosanoid that plays a role in a variety of important physiological processes, including vasodilation, inflammation, and adipogenesis (29). Our findings, and those of other researchers, indicate that *PTGDS* may represent a candidate marker gene for podocytes at the transcriptional level. *PTGDS* is expressed at high levels in normal podocytes and at far lower levels in injured podocytes.

3.7 Crosstalk between kidney cell types in NS-IgAN

We investigated cell-to-cell communication between kidney cells by applying CellphoneDB. In the CTRL group, the cell types present in the proximal nephrons and kidney immune cells exhibited obvious intercellular communication, whereas distal convoluted tubules, intercalated cells and principal cells located in the distal nephron did not (Figure 5A). Our results suggest that the intercellular communication of the cells associated with the proximal nephron may be stronger than that of the cells associated with the distal nephron. In the CTRL and HU-IgAN groups, PECs may have served as the center for intercellular crosstalk in kidney cells (Figure 5A). In the NGC and LGC patients, the center of intercellular crosstalk was replaced by podocytes and mesenchymal stromal cells (MSCs) (Figure 5A). In an injury state, podocytes were affected by collagen-related, extracellular matrix-related, and inflammation-related signals from PECs and MSCs, such as COL3A1, FN1, and CCL2 (Figure 5B). These findings were consistent with previous findings related to IgAN (7, 8).

3.8 The molecular characteristics of renal lymphocytes

Kidney immune cells were re-clustered and defined into eight cell subsets by the expression of canonical marker genes (31) (Figure 5C, Supplementary Table 9). *C1QA*, *C1QB*, and *C1QC* are characteristically expressed in two subsets of tissue-resident macrophages (RTMs). Of these, RTM-1 expressed *LYVE1*, *IGF1*, and *CD209*, and RTM-2 expressed *TREM2* and *MARCO* (Figure 5D). Our results suggest that *FCN1* and *VCAN* may be specific markers of monocytes in the human kidney. Classical monocytes (CMs) in the kidney represented the main subset of monocytes and expressed *CXCL3*. In contrast, NCMs expressed *CX3CR1* (Figure 5D). *XCR1*, *CLEC9A*, and *IDO1* were expressed in Classical DC-1, and *CD1C* and *CD1E* were expressed in Classical DC-2. Plasmacytoid DC expressed *IL3RA*, *CLEC4C*, *GZMB*, and *TCL1A*. It is worth noting that cluster 12 simultaneously expressed the marker genes of B cells, including *CD79A* and *MZB1*, and also expressed the marker genes of T cells such as *CD3D*, *CD3E*, and *TRAC* (Figure 5D). These results indicated that this cluster may be composed of B lymphocytes and T lymphocytes; hence, we defined this cluster as

10.3389/fimmu.2023.1231937

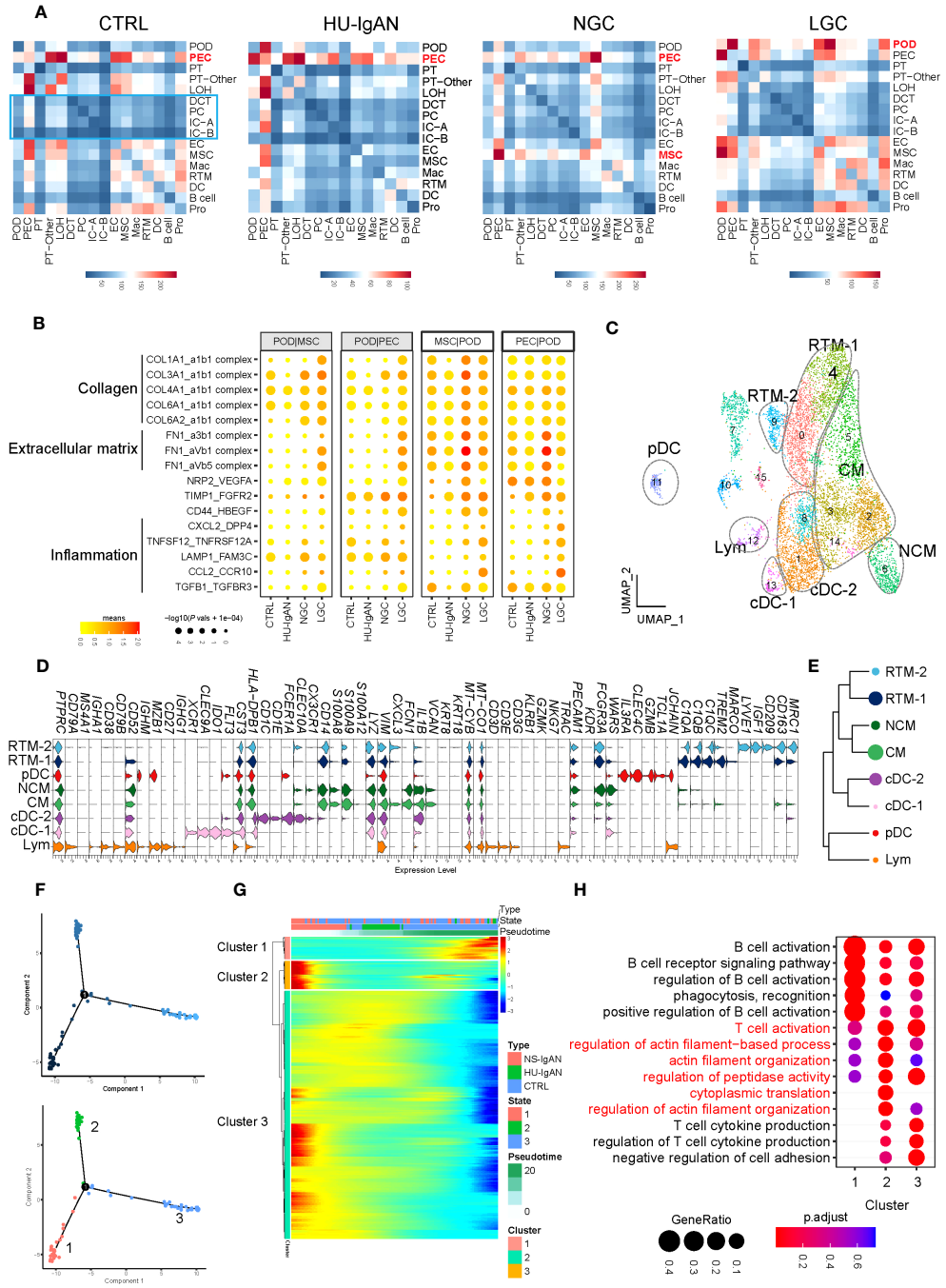


FIGURE 5
Molecular characterization of interactions between kidney cells and kidney immune cells. **(A)** Cell-to-cell crosstalk networks between kidney cells in CTRL, HU-IgAN, NGC, and LGC groups. **(B)** Bubble chart showing dysregulated cell-type specific interactions between mesenchymal stromal cells (MSCs) and podocytes (PODs) or parietal epithelial cells (PECs) and podocytes (PODs) in the CTRL, HU-IgAN, NGC, and LGC group. **(C)** 8 cell subtypes were visualized by UMAP plotting through re-clustering analysis of kidney immune cells. **(D)** Violin plots showing marker genes of 8 kidney immune cell subtypes. **(E)** Developmental tree analysis showed the relationship between different kidney immune cells. **(F)** Cell trajectory map of lymphocytes showing the pseudo-time (Top). Ordering single cells along a cell conversion trajectory using Monocle 2. Three states were identified based on their distribution in the cell trajectory map (Bottom). **(G)** Heatmap: each column represents one cell and each row represents the expression of one gene. Cells are ordered by Monocle-based pseudotime analysis and the color represents expression levels. **(H)** The GO enrichment function items of cluster1-3 in **(G)** are shown by the bubble diagram.

lymphocytes. We performed a developmental analysis of eight groups of kidney immune cells and confirmed the difference between tissue-resident macrophages and monocytes (Figure 5E). Previous studies comprehensively demonstrated the molecular characteristics of myeloid cell subsets in the kidney (31–34); however, there was a lack of understanding of lymphocyte subsets in the kidney. We further performed pseudo-time analysis on lymphocytes and found that they were in three different states (states 1–3) in the kidneys (Figure 5F),

with different molecular characteristics (clusters 1–3) (Figure 5G). Lymphocytes in state 3 significantly expressed genes related to cluster 1 (Figure 5G), which were related to B cell activation, phagocytosis, and recognition (Figure 5H). Lymphocytes in state 1 significantly expressed genes related to cluster 3 (Figure 5G). These results were related to the production of cytokines from T cells (Figure 5H). Surprisingly, state 2 lymphocytes significantly expressed genes related to cluster 2 (Figure 5G). These findings are related to T cell activation and actin filaments (Figure 5H). These results suggest that there may be a group of T cells with actin filament-related functions in the kidney. By applying a scRNA-seq technology platform, we were fortunate to capture immune cells in the kidney and provide a preliminary exploration of the molecular characteristics of lymphocytes. However, due to the limited number of immune cells captured from the kidney tissue involved in this study, we were not able to conduct further analysis.

3.9 DCs in circulation and kidney tissue expressed high levels of genes related to HLA

Based on the SNPs detected in a recent GWAS meta-analysis of IgAN (18), we combined NephQTL and eQTL (cf.) to predict susceptibility genes that may be affected in different regions of the kidney tissue (Supplementary Table 10) and mapped these genes to kidney cells (Figure 6A) and PBMCs (Figure 6B). In the kidney tissue, renal resident macrophages, macrophages, and DC cells expressed high levels of genes related to HLA (Figure 6A). In circulation, multiple B cell subsets (except plasma cells), cDCs, and pDCs expressed high levels of genes related to HLA, while monocytes did not (Figure 6B). Previous studies had identified HLA molecules as the main disease related susceptibility loci for IgAN (18). Research evidence also indicated that DCs play key roles in the pathogenesis of IgAN (35). Our present data showed that DCs express high levels of HLA-related genes in both the circulation and kidney tissue. This indicated that attention needs to be paid to the changes in DC cells in the circulation and kidney tissue of NS-IgAN patients in future research (Figures 6A, B). The CFH-related genes *CFHR1*, *CFHR3*, and *CFHR4* were exclusively expressed in kidney MSCs (Figure 6A). We performed a re-clustering of the MSCs that significantly expressed *POSTN* and *ACTA2* (Supplementary Table 4; Supplementary Figure 9A). This analysis suggested that MSCs were a mixed subset composed of fibroblasts, myofibroblasts, vascular smooth muscle cells, pericytes, and mesangial cells (Supplementary Figure 9B). Unfortunately, in our study, mesangial cells could not be defined by canonical marker genes such as *PDGFRB*, *PIEZO2*, *ITGA8*, and *GATA3* (Supplementary Figure 9B).

4 Discussion

Glomerular diseases are still classified based on histological descriptions; however, these do not help capture the systemic mechanisms that drive the disease, nor are they suitable for target identification and drug development (36). Transcriptome sequencing at single-cell resolution, as represented by scRNA-seq and single-nucleus RNA-seq (snRNA-seq), is a powerful and new approach to unbiased analysis (37). However, it is challenging to create suitable designs for scRNA-seq experiments because each option requires the user to make informed decisions to obtain interpretable results (37, 38). Considering its low requirement for cell viability, we chose plate-based scRNA-seq technology to study core needle biopsies from the kidneys of patients (37, 39). We performed strict QC on the data and selected threshold settings based on the specificity of kidney cells (Supplementary Table 2). To identify mesangial cells, we performed a re-clustering of the MSCs (40). Our data suggested that MSCs may be composed of fibroblasts, myofibroblasts, vascular smooth muscle cells, pericytes, and mesangial cells (Supplementary Figure 9). As we were not able to annotate subgroups of cells by canonical marker genes, we were not able to perform downstream analysis of mesangial cells. Different methodological and technological platforms can lead to bias in capturing glomerular cell types; our results provide evidence for future single-cell research on mesangial-related glomerular diseases. Considering that immune cells in the kidney may play an important role in disease, we did not select snRNA-seq for our research to avoid losing molecular information related to immune cells (37) (Figures 5C–H).

Monocytes are innate immune cells that can be divided into three subsets based on the expression of CD14 and CD16 on the cell surface: classical, intermediate, and non-classical monocytes (14). Under inflammatory conditions, monocytes in the blood may migrate to the tissues and differentiate into mononuclear phagocytes in local regions of tissue (41). Previous research in the field of chronic kidney disease (CKD) has detected significant expansion of IMs (42), thus suggesting that changes in monocyte subsets may play an important role in CKD. Recent studies have found that CKD is associated with an increase in the number of unique proinflammatory IMs (HLA-DR^{high} IMs), as well as the migration of monocytes and endothelial adhesion abnormalities (43). Due to the limitation of renal biopsy in children with NS and the low incidence rate of NS-IgAN, we recruited children with NS as a validation patient cohort. The consistency of scRNA-seq data with the flow cytometry results of the validation cohort confirmed that IMs may be more abundant in NS-IgAN patients. Future studies on glomerular diseases need to pay specific attention to this group of IMs, particularly in terms of MHC class II analysis and chemokine signaling pathway functionality. In several human and experimental studies of proteinuria nephropathy, the expression of CCL2 was significantly localized in glomerular podocytes (17). It has also been confirmed that CCL2 does not only directly affect the actin cytoskeleton of podocytes (44), thus affecting the permeability of the slit diaphragm to albumin (45), but also causes indirect damage to podocytes by attracting macrophages and promoting inflammation (46). The expression of CCL2 in the kidney will recruit monocytes/macrophages expressing CCR2 into the circulation to be transported to the injury site and promote the differentiation of these myeloid cells toward the proinflammatory "M1" phenotype (47). Previous research suggested that the persistent infiltration of M1 macrophages and related inflammation can play a crucial pathogenic role in the development of podocyte dysfunction (48). The inhibition of CCR2 has been found to improve outcomes in animal models of FSGS (17), and several trials are currently underway to evaluate the impact of CCR2 inhibitors on FSGS patients (NCT03649152, NCT03536754,

10.3389/fimmu.2023.1231937

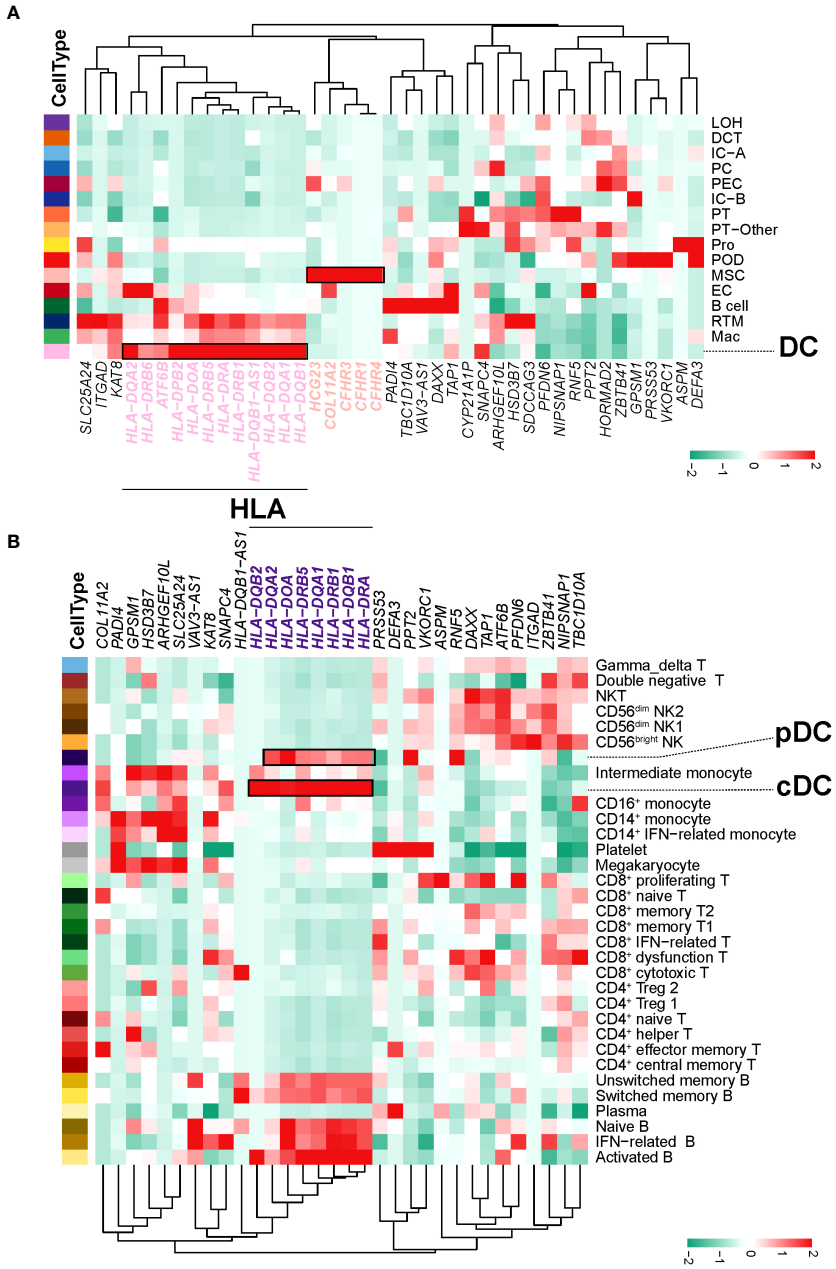


FIGURE 6
IgAN disease risk susceptibility genes expression. **(A)** The heatmap shows, for each IgAN susceptibility gene, the average expression over all cells in each kidney cell type. **(B)** The heatmap shows, for each IgAN susceptibility gene, the average expression over all cells in each PBMC cell subset.

NCT03703908) (49). In addition, urinary CCL2 can be used as a biomarker for kidney inflammation (50) and these specific levels may be related to the extent of proteinuria (51). These data indicated that we cannot only gain an understanding of inflammatory progression in the kidney by detecting the urinary levels of CCL2. We may also be able to delay the chronic progression of the disease through new drugs targeting the CCL2-CCR2 ligand receptor pathway.

Podocyte damage is the key to the formation of proteinuria. The expression levels of EMT-related genes and *PRSS23* were significantly increased in podocytes from patients with NS-IgAN (Figures 4C–E). EMT is a functional and morphological alteration in podocyte injury (27, 28). The serine protease encoded by *PRSS23* activates protease-activated receptor 2, which is known to be associated with TGFβ1-induced podocyte injury in the rat model of doxorubicin nephropathy (52). The expression levels of canonical marker genes in the podocytes were significantly downregulated in NS-IgAN conditions, including those of *NPHS1*, *CLIC5*, and *MPP5* (Figure 4C). Our scRNA-seq results indicated that *PTGDS* was specifically expressed in podocytes and were significantly downregulated in NS-IgAN (Figures 4G–J). Unlike the remarkable specificity at the transcriptional level, immunohistochemistry and immunofluorescence results suggested that the specificity of PGDS at the protein level was limited (Supplementary Figure 10). BTP is a heterogeneous monomeric glycoprotein that is the consequence of post-translational N-

glycosylation resulting in different glycoforms of varying molecular weight (53). The presence of multiple isomers certainly affects the measurement of BTP; the molecular structure of BTP produced by podocytes needs to be investigated further. Previous studies have suggested that in the early stages of diabetes nephropathy and CKD, the urinary levels of PGDS increased significantly (54). In the process of being excreted by the urine, some PGDS would be reabsorbed into the tubules and degraded by the lysosomes of tubule cells. We also noted that the staining intensity of PGDS was significantly reduced in the tubule cells of NS-IgAN patients (Figure 4I, Supplementary Figure 10). Whether this was due to the reduction of PGDS in podocytes or the reduced reabsorption capacity of renal tubules needs to be investigated in future research.

In our study, scRNA-seq of PBMCs and renal tissues were performed only on three pediatric patients with clinical manifestations of nephrotic syndrome. There was significant heterogeneity among the samples, and long-term cohort observations with larger sample sizes are required to truly elucidate the pathogenesis of NS-IgAN. Whether NS-IgAN is a combination of two diseases, is not discussed herein and needs to be further explored in future studies (4). Even though the findings of this study await future validation, they provide a rigorous framework for future research.

Data availability statement

The original contributions presented in the study are publicly available. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics statement

Children's Hospital of Chongqing Medical University approved this study (File Number: 2022 Research 124 and 2022 Research 35). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. Ethical approval was not required for the studies on animals in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

QC and HJ contributed equally to this work and share first authorship. QC conceptualized this project, interpreted the results of the analysis, prepared figures, and wrote the original paper. HJ enrolled and followed up with the patients, collected the samples, interpreted part of the results, and performed validation experiments. RD and JZ performed the single-cell transcriptome sequencing analyses and offered help with the bioinformatics analysis. JW, XF, and LP interpreted the results of the analysis, prepared figures, and wrote the manuscript. LL constructed the scRNA-seq library and performed the next-generation sequencing. XY, HC, AW, JJ, QY, XC, and XL prepared the single-cell suspensions of PBMCs and kidney biopsies. LS, GZ, and MW discussed the draft paper and critically reviewed the manuscript. HY and QL conceived and supervised the project and reviewed the manuscript. All authors read and commented on the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key R&D Program of China (Grant No. 2022YFC2705101), the Chongqing Science and Health Joint Medical Research Project (Grant No. 2023GGXM001), the Program for Youth Innovation in Future Medicine, Chongqing Medical University (Grant No. W0098), and through the Special Fund For Postdoctoral Research Projects of Chongqing, China (Grant No. 2011010006602511).

Acknowledgments

We thank all the participants and their guardians for this study.

Conflict of interest

Authors RD and LL are employed by the company Nanjing Jiangbei New Area Biopharmaceutical Public Service Platform.

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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1231937/full#supplementary-material

References

1. Lai KN, Tang SC, Schena FP, Novak J, Tomino Y, Fogo AB, et al. IgA nephropathy. *Nat Rev Dis Primers* (2016) 2:16001. doi: 10.1038/nrdp.2016.1

2. Nie S, He W, Huang T, Liu D, Wang G, Geng J, et al. The spectrum of biopsy-proven glomerular diseases among children in China: A national, cross-sectional survey. *Clin J Am Soc Nephrol* (2018) 13(7):1047–54. doi: 10.2215/cjn.11461017

3. Schena FP, Rossini M, Abbrescia DI, Zaza G. The molecular mechanisms of inflammation and scarring in the kidneys of immunoglobulin A nephropathy : Gene involvement in the mechanisms of inflammation and scarring in kidney biopsy of IgAN patients. *Semin Immunopathol* (2021) 43(5):691–705. doi: 10.1007/s00281-021-00891-8

4. Shima Y, Nakanishi K, Sato M, Hama T, Mukaiyama H, Togawa H, et al. IgA nephropathy with presentation of nephrotic syndrome at onset in children. *Pediatr Nephrol* (2017) 32(3):457–65. doi: 10.1007/s00467-016-3502-6

5. Kim JK, Kim JH, Lee SC, Kang EW, Chang TI, Moon SJ, et al. Clinical features and outcomes of IgA nephropathy with nephrotic syndrome. *Clin J Am Soc Nephrol* (2012) 7(3):427–36. doi: 10.2215/CJN.04820511

6. Pattrapornpisut P, Avila-Casado C, Reich HN. IgA nephropathy: core curriculum 2021. *Am J Kidney Dis* (2021) 78(3):429–41. doi: 10.1053/j.ajkd.2021.01.024

7. Lemley KV, Lafayette RA, Safai M, Derby G, Blouch K, Squarer A, et al. Podocytopenia and disease severity in IgA nephropathy. *Kidney Int* (2002) 61 (4):1475–85. doi: 10.1046/j.1523-1755.2002.00269.x

8. Lai KN, Tang SC, Guh JY, Chuang TD, Lam MF, Chan LY, et al. Polymeric IgA1 from patients with IgA nephropathy upregulates transforming growth factor-beta synthesis and signal transduction in human mesangial cells *via* the renin-angiotensin system. *J Am Soc Nephrol* (2003) 14(12):3127–37. doi: 10.1097/01.asn.0000095639.56212.bf

9. Stewart BJ, Ferdinand JR, Clatworthy MR. Using single-cell technologies to map the human immune system - implications for nephrology. *Nat Rev Nephrol* (2020) 16 (2):112–28. doi: 10.1038/s41581-019-0227-3

10. Zheng Y, Lu P, Deng Y, Wen L, Wang Y, Ma X, et al. Single-cell transcriptomics reveal immune mechanisms of the onset and progression of igA nephropathy. *Cell Rep* (2020) 33(12):108525. doi: 10.1016/j.celrep.2020.108525

11. Tang R, Meng T, Lin W, Shen C, Ooi JD, Eggenhuizen PJ, et al. A partial picture of the single-cell transcriptomics of human igA nephropathy. *Front Immunol* (2021) 12:645988. doi: 10.3389/fimmu.2021.645988

12. Zeng H, Wang L, Li J, Luo S, Han Q, Su F, et al. Single-cell RNA-sequencing reveals distinct immune cell subsets and signaling pathways in IgA nephropathy. *Cell Biosci* (2021) 11(1):203. doi: 10.1186/s13578-021-00706-1

13. Wu J, Zheng C, Wang X, Yun S, Zhao Y, Liu L, et al. MicroRNA-30 family members regulate calcium/calcineurin signaling in podocytes. *J Clin Invest* (2015) 125 (11):4091–106. doi: 10.1172/JCI81061

14. Ziegler-Heitbrock L, Ancuta P, Crowe S, Dalod M, Grau V, Hart DN, et al. Nomenclature of monocytes and dendritic cells in blood. *Blood* (2010) 116(16):e74–80. doi: 10.1182/blood-2010-02-258558

15. Cox SN, Serino G, Sallustio F, Blasi A, Rossini M, Pesce F, et al. Altered monocyte expression and expansion of non-classical monocyte subset in IgA nephropathy patients. *Nephrol Dial Transpl* (2015) 30(7):1122–232. doi: 10.1093/ndt/gfv017

16. Franca CN, Izar MCO, Hortencio MNS, do Amaral JB, Ferreira CES, Tuleta ID, et al. Monocyte subtypes and the CCR2 chemokine receptor in cardiovascular disease. *Clin Sci (Lond)* (2017) 131(12):1215–24. doi: 10.1042/CS20170009

17. Wilkening A, Krappe J, Muhe AM, Lindenmeyer MT, Eltrich N, Luckow B, et al. C-C chemokine receptor type 2 mediates glomerular injury and interstitial fibrosis in focal segmental glomerulosclerosis. *Nephrol Dial Transpl* (2020) 35(2):227–39. doi: 10.1093/ndt/gfy380

18. Li M, Wang L, Shi DC, Foo JN, Zhong Z, Khor CC, et al. Genome-wide meta-analysis identifies three novel susceptibility loci and reveals ethnic heterogeneity of genetic susceptibility for IgA nephropathy. *J Am Soc Nephrol* (2020) 31(12):2949–63. doi: 10.1681/ASN.2019080799

19. Huang H, Sun W, Liang Y, Peng Y, Long XD, Liu Z, et al. CD4 (+)CD 25 (+)Treg cells and IgA nephropathy patients with tonsillectomy: a clinical and pathological study. *Int Urol Nephrol* (2014) 46(12):2361–9. doi: 10.1007/s11255-014-0851-6

20. Wing JB, Tanaka A, Sakaguchi S. Human FOXP3(+) regulatory T cell heterogeneity and function in autoimmunity and cancer. *Immunity* (2019) 50 (2):302–16. doi: 10.1016/j.immuni.2019.01.020

21. Zemmour D, Charbonnier LM, Leon J, Six E, Keles S, Delville M, et al. Single-cell analysis of FOXP3 deficiencies in humans and mice unmasks intrinsic and extrinsic CD4(+) T cell perturbations. *Nat Immunol* (2021) 22(5):607–19. doi: 10.1038/s41590-021-00910-8

22. Sakai R, Ito M, Komai K, Iizuka-Koga M, Matsuo K, Nakayama T, et al. Kidney GATA3(+) regulatory T cells play roles in the convalescence stage after antibody-mediated renal injury. *Cell Mol Immunol* (2021) 18(5):1249–61. doi: 10.1038/s41423-020-00547-x

23. Ju W, Greene CS, Eichinger F, Nair V, Hodgin JB, Bitzer M, et al. Defining cell-type specificity at the transcriptional level in human disease. *Genome Res* (2013) 23 (11):1862–73. doi: 10.1101/gr.155697.113

24. Ren J, Xu Y, Lu X, Wang L, Ide S, Hall G, et al. Twist1 in podocytes ameliorates podocyte injury and proteinuria by limiting CCL2-dependent macrophage infiltration. *JCI Insight* (2021) 6(15):e148109. doi: 10.1172/jci.insight.148109

25. Pickering MC, Cook HT. Translational mini-review series on complement factor H: renal diseases associated with complement factor H: novel insights from humans and animals. *Clin Exp Immunol* (2008) 151(2):210–30. doi: 10.1111/j.1365-2249.2007.03574.x

26. Muhlig AK, Keir LS, Abt JC, Heidelbach HS, Horton R, Welsh GI, et al. Podocytes produce and secrete functional complement C3 and complement factor H. *Front Immunol* (2020) 11:1833. doi: 10.3389/fimmu.2020.01833

27. Li Y, Kang YS, Dai C, Kiss LP, Wen X, Liu Y. Epithelial-to-mesenchymal transition is a potential pathway leading to podocyte dysfunction and proteinuria. *Am J Pathol* (2008) 172(2):299–308. doi: 10.2353/ajpath.2008.070057

28. Latt KZ, Heymann J, Jessee JH, Rosenberg AZ, Berthier CC, Arazi A, et al. Urine single-cell RNA sequencing in focal segmental glomerulosclerosis reveals inflammatory signatures. *Kidney Int Rep* (2022) 7(2):289–304. doi: 10.1016/j.ekir.2021.11.005

29. White CA, Ghazan-Shahi S, Adams MA. beta-Trace protein: a marker of GFR and other biological pathways. *Am J Kidney Dis* (2015) 65(1):131–46. doi: 10.1053/j.ajkd.2014.06.038

30. Nagata N, Fujimori K, Okazaki I, Oda H, Eguchi N, Uehara Y, et al. *De novo* synthesis, uptake and proteolytic processing of lipocalin-type prostaglandin D synthase, beta-trace, in the kidneys. *FEBS J* (2009) 276(23):7146–58. doi: 10.1111/j.1742-4658.2009.07426.x

31. Stewart BJ, Ferdinand JR, Young MD, Mitchell TJ, Loudon KW, Riding AM, et al. Spatiotemporal immune zonation of the human kidney. *Science* (2019) 365 (6460):1461–6. doi: 10.1126/science.aat5031

32. Zimmerman KA, Bentley MR, Lever JM, Li Z, Crossman DK, Song CJ, et al. Single-cell RNA sequencing identifies candidate renal resident macrophage gene expression signatures across species. *J Am Soc Nephrol* (2019) 30(5):767–81. doi: 10.1681/ASN.2018090931

33. McEvoy CM, Murphy JM, Zhang L, Clotet-Freixas S, Mathews JA, An J, et al. Single-cell profiling of healthy human kidney reveals features of sex-based transcriptional programs and tissue-specific immunity. *Nat Commun* (2022) 13 (1):7634. doi: 10.1038/s41467-022-35297-z

34. Conway BR, O'Sullivan ED, Cairns C, O'Sullivan J, Simpson DJ, Salzano A, et al. Kidney single-cell atlas reveals myeloid heterogeneity in progression and regression of kidney disease. *J Am Soc Nephrol* (2020) 31(12):2833–54. doi: 10.1681/asn.2020060806

35. Takechi H, Oda T, Hotta O, Yamamoto K, Oshima N, Matsunobu T, et al. Clinical and immunological implications of increase in CD208+ dendritic cells in tonsils of patients with immunoglobulin A nephropathy. *Nephrol Dial Transpl* (2013) 28(12):3004–13. doi: 10.1093/ndt/gft399

36. Kalantar-Zadeh K, Jafar TH, Nitsch D, Neuen BL, Perkovic V. Chronic kidney disease. *Lancet* (2021) 398(10302):786–802. doi: 10.1016/s0140-6736(21)00519-5

37. Deleersnijder D, Callemeyn J, Arijs I, Naesens M, Van Craenenbroeck AH, Lambrechts D, et al. Current methodological challenges of single-cell and single-nucleus RNA-sequencing in glomerular diseases. *J Am Soc Nephrol* (2021) 32(8):1838–52. doi: 10.1681/ASN.2021020157

38. Lafzi A, Moutinho C, Picelli S, Heyn H. Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nat Protoc* (2018) 13(12):2742–57. doi: 10.1038/s41596-018-0073-y

39. Xu J, Shen C, Lin W, Meng T, Ooi JD, Eggenhuizen PJ, et al. Single-cell profiling reveals transcriptional signatures and cell-cell crosstalk in anti-PLA2R positive idiopathic membranous nephropathy patients. *Front Immunol* (2021) 12:683330. doi: 10.3389/fimmu.2021.683330

40. Avraham S, Korin B, Chung JJ, Oxburgh L, Shaw AS. The Mesangial cell - the glomerular stromal cell. *Nat Rev Nephrol* (2021) 17(12):855–64. doi: 10.1038/s41581-021-00474-8

41. Narasimhan PB, Marcovecchio P, Hamers AAJ, Hedrick CC. Nonclassical monocytes in health and disease. *Annu Rev Immunol* (2019) 37:439–56. doi: 10.1146/annurev-immunol-042617-053119

42. Naicker SD, Cormican S, Griffin TP, Maretto S, Martin WP, Ferguson JP, et al. Chronic kidney disease severity is associated with selective expansion of a distinctive intermediate monocyte subpopulation. *Front Immunol* (2018) 9:2845. doi: 10.3389/fimmu.2018.02845

43. Cormican S, Negi N, Naicker SD, Islam MN, Fazekas B, Power R, et al. Chronic kidney disease is characterized by expansion of a distinct proinflammatory intermediate monocyte subtype and by increased monocyte adhesion to endothelial cells. *J Am Soc Nephrol* (2023) 34(5):793–808. doi: 10.1681/ASN.0000000000000083

44. Lee EY, Chung CH, Khoury CC, Yeo TK, Pyagay PE, Wang A, et al. The monocyte chemoattractant protein-1/CCR2 loop, inducible by TGF-beta, increases podocyte motility and albumin permeability. *Am J Physiol Renal Physiol* (2009) 297(1): F85–94. doi: 10.1152/ajprenal.90642.2008

45. Burt D, Salvidio G, Tarabra E, Barutta F, Pinach S, Dentelli P, et al. The monocyte chemoattractant protein-1/cognate CC chemokine receptor 2 system affects

cell motility in cultured human podocytes. *Am J Pathol* (2007) 171(6):1789–99. doi: 10.2353/ajpath.2007.070398

46. Bruggeman LA, Drawz PE, Kahoud N, Lin K, Barisoni L, Nelson PJ. TNFR2 interposes the proliferative and NF-kappaB-mediated inflammatory response by podocytes to TNF-alpha. *Lab Invest* (2011) 91(3):413–25. doi: 10.1038/labinvest.2010.199

47. Fujinaka H, Yamamoto T, Takeya M, Feng L, Kawasaki K, Yaoita E, et al. Suppression of anti-glomerular basement membrane nephritis by administration of anti-monocyte chemoattractant protein-1 antibody in WKY rats. *J Am Soc Nephrol* (1997) 8(7):1174–8. doi: 10.1681/ASN.V871174

48. Bell RMB, Denby L. Myeloid heterogeneity in kidney disease as revealed through single-cell RNA sequencing. *Kidney360* (2021) 2(11):1844–51. doi: 10.34067/KID.0003682021

49. De Vriese AS, Wetzels JF, Glassock RJ, Sethi S, Fervenza FC. Therapeutic trials in adult FSGS: lessons learned and the road forward. *Nat Rev Nephrol* (2021) 17(9):619–30. doi: 10.1038/s41581-021-00427-1

50. Puthumana J, Thiessen-Philbrook H, Xu L, Coca SG, Garg AX, Himmelfarb J, et al. Biomarkers of inflammation and repair in kidney disease progression. *J Clin Invest* (2021) 131(3):e139927. doi: 10.1172/JCI139927

51. Eardley KS, Zehnder D, Quinkler M, Lepenies J, Bates RL, Savage CO, et al. The relationship between albuminuria, MCP-1/CCL2, and interstitial macrophages in chronic kidney disease. *Kidney Int* (2006) 69(7):1189–97. doi: 10.1038/sj.ki.5000212

52. Wang Y, He Y, Wang M, Lv P, Liu J, Wang J. Role of protease-activated receptor 2 in regulating focal segmental glomerulosclerosis. *Cell Physiol Biochem* (2017) 41(3):1147–55. doi: 10.1159/000464121

53. Urade Y, Hayaishi O. Biochemical, structural, genetic, physiological, and pathophysiological features of lipocalin-type prostaglandin D synthase. *Biochim Biophys Acta* (2000) 1482(1-2):259–71. doi: 10.1016/s0167-4838(00)00161-8

54. Oda H, Shiina Y, Seiki K, Sato N, Eguchi N, Urade Y. Development and evaluation of a practical ELISA for human urinary lipocalin-type prostaglandin D synthase. *Clin Chem* (2002) 48(9):1445–53.

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 20 November 2023
DOI 10.3389/fimmu.2023.1275136

OPEN ACCESS

EDITED BY
Ying Zhu,
Genentech Inc., United States

REVIEWED BY
Aurobind Vidyarthi,
Yale University, United States
Soumya Panigrahi,
Case Western Reserve University, United States

*CORRESPONDENCE
Anna C. Aschenbrenner
anna.aschenbrenner@dzne.de

†These authors share senior authorship

RECEIVED 09 August 2023
ACCEPTED 18 October 2023
PUBLISHED 20 November 2023

CITATION
Knoll R, Bonaguro L, dos Santos JC, Warnat-Herresthal S, Jacobs-Cleophas MCP, Blümel E, Reusch N, Horne A, Herbert M, Nuesch-Germano M, Otten T, van der Heijden WA, van de Wijer L, Shalek AK, Händler K, Becker M, Beyer MD, Netea MG, Joosten LAB, van der Ven AJAM, Schultze JL and Aschenbrenner AC (2023) Identification of drug candidates targeting monocyte reprogramming in people living with HIV. *Front. Immunol.* 14:1275136. doi: 10.3389/fimmu.2023.1275136

Identification of drug candidates targeting monocyte reprogramming in people living with HIV

Rainer Knoll[1,2], Lorenzo Bonaguro[1,2], Jéssica C. dos Santos[3,4], Stefanie Warnat-Herresthal[1,2], Maartje C. P. Jacobs-Cleophas[3,4], Edda Blümel[1], Nico Reusch[1], Arik Horne[1,5], Miriam Herbert[1,6], Melanie Nuesch-Germano[1], Twan Otten[3,4], Wouter A. van der Heijden[3,4], Lisa van de Wijer[3,4], Alex K. Shalek[7,8,9], Kristian Händler[10,11], Matthias Becker[1], Marc D. Beyer[1,10], Mihai G. Netea[3,4,12], Leo A. B. Joosten[3,4,13], Andre J. A. M. van der Ven[3,4], Joachim L. Schultze[1,2,10†] and Anna C. Aschenbrenner[1*†]

[1]Systems Medicine, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Bonn, Germany, [2]Genomics and Immunoregulation, Life and Medical Sciences Institute, University of Bonn, Bonn, Germany, [3]Department of Internal Medicine, Radboud University Medical Center, Nijmegen, Netherlands, [4]Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, Netherlands, [5]Systems Hematology, Stem Cells & Precision Medicine, Max Delbrück Center - Berlin Institute for Medical Systems Biology (MDCBIMSB), Berlin, Germany, [6]In Vivo Cell Biology of Infection, Max Planck Institute for Infection Biology (MPIIB), Berlin, Germany, [7]Broad Institute at Massachusetts Institute of Technology (MIT) and Harvard, Boston, MA, United States, [8]Ragon Institute of Mass General Hospital (MGH), MIT, and Harvard, Cambridge, MA, United States, [9]Department of Chemistry, Institute for Medical Engineering and Science, Koch Institute, Cambridge, MA, United States, [10]Platform for Single Cell Genomics and Epigenomics (PRECISE), DZNE and University of Bonn, Bonn, Germany, [11]Institute for Human Genetics, University Hospital Schleswig-Holstein, Lübeck, Germany, [12]Immunology and Metabolism, Life and Medical Sciences Institute, University of Bonn, Bonn, Germany, [13]Department of Medical Genetics, Iuliu Hatieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

Introduction: People living with HIV (PLHIV) are characterized by functional reprogramming of innate immune cells even after long-term antiretroviral therapy (ART). In order to assess technical feasibility of omics technologies for application to larger cohorts, we compared multiple omics data layers.

Methods: Bulk and single-cell transcriptomics, flow cytometry, proteomics, chromatin landscape analysis by ATAC-seq as well as *ex vivo* drug stimulation were performed in a small number of blood samples derived from PLHIV and healthy controls from the 200-HIV cohort study.

Results: Single-cell RNA-seq analysis revealed that most immune cells in peripheral blood of PLHIV are altered in their transcriptomes and that a specific functional monocyte state previously described in acute HIV infection is still existing in PLHIV while other monocyte cell states are only occurring acute infection. Further, a reverse transcriptome approach on a rather small number of PLHIV was sufficient to identify drug candidates for reversing the transcriptional phenotype of monocytes in PLHIV.

10.3389/fimmu.2023.1275136

Discussion: These scientific findings and technological advancements for clinical application of single-cell transcriptomics form the basis for the larger 2000-HIV multicenter cohort study on PLHIV, for which a combination of bulk and single-cell transcriptomics will be included as the leading technology to determine disease endotypes in PLHIV and to predict disease trajectories and outcomes.

KEYWORDS

systems immunology, transcriptomics, HIV, monocytes, inflammation, drug repurposing

Introduction

For people living with HIV (PLHIV), major risk factors for developing cardiovascular diseases (CVDs), neurocognitive impairment, frailty, and cancer are persistent low-grade inflammation and immune dysfunction even under long-term effective antiretroviral therapy (ART) (1–6). Although the adaptive immune system appears to play an important role (7), there is a growing body of evidence that suggests changes in the innate immune system as exemplified by elevated levels of circulating soluble CD163 and sCD14 derived from monocytes are critical (1, 8, 9). We and others have recently demonstrated that concentrations of pro-inflammatory monocyte-derived cytokines are elevated in serum from PLHIV, which was further validated when peripheral blood mononuclear cells were stimulated *ex vivo* with a number of pathogens or their derivatives resulting in increased levels of IL-1β (1, 10–14).

While CMV infection (15), the HIV reservoir itself (16), as well as microbial translocation (17) have been proposed as potential drivers of low-grade inflammation, the complex interplay between the different immune cell compartments in PLHIV is not fully understood. To study the role of different immune cells in the pathophysiology of persistent inflammation in PLHIV it will be necessary to apply higher-resolution single-cell technologies to larger cohorts of PLHIV (18–20). Based on our previous experience applying single-cell technologies to better understand the pathophysiology of COVID-19 (21–23) or chronic obstructive pulmonary disease (COPD) (24), we have recently suggested that large-scale studies should be preceded by smaller optimization studies for clinical application of omics technologies to a particular disease setting (25, 26).

Here, we describe a study using bulk and single-cell transcriptomics technologies as well as chromatin landscaping by ATAC-seq under clinically applicable conditions to assess the reprogramming of the peripheral immune cell compartment in PLHIV cohorts. Despite heterogeneity between individuals, scRNA-seq combined with bulk transcriptomics on a limited number of PLHIV included in this pilot study revealed important new information concerning the involvement of the monocyte compartment in persistent low-grade inflammation. Further, a reverse transcriptome approach in this setup allowed the identification of drug candidates reducing the inflammatory endophenotype, which we validated experimentally in an independent group of PLHIV.

Results

Bulk transcriptomes from PBMC of PLHIV are dominated by monocyte-related proinflammatory programs

We previously demonstrated in a cross-sectional study that PLHIV exhibits a proinflammatory profile in monocyte- but not lymphocyte-derived cytokines (1). We recalled five male PLHIV using long-term suppressive ART (mean 7.4 years) from the 200-HIV study with no overt clinical symptoms at the time of blood draw, determined as normal progressors, to investigate whether higher-resolution technologies down to the single-cell level would reveal further information about molecular and functional changes within the peripheral immune system in PLHIV. We generated a multi-layer dataset including selected soluble factors in plasma, multicolor flow cytometry (MCFC), bulk RNA-seq, Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) and microwell-based scRNA-seq comparing five age- and sex-matched healthy controls (Figure 1A; Supplementary Table S1).

The MCFC data generated here indicate that the five PLHIV chosen were representative of the 200-HIV cohort with similar alterations in the circulating immune cell compartment (e.g. higher CD8+ and lower CD4+ T as well as NK cell population frequencies in PLHIV versus healthy donors) (11) (Figure S1A). Principal component analysis (PCA) of bulk RNA-seq of PBMC revealed a disease-associated separation of the samples (Figure 1B). Exploration of these alterations by differential gene expression analysis resulted in 287 up- and 914 down-regulated genes in PLHIV compared to control ($|FC|>1.5$, adj p-value <0.05, with independent hypothesis weighting (IHW) correction) (Figure S1B). Inspection of those differentially expressed genes (DEGs) in more detail by hierarchical clustering revealed four transcript clusters similarly regulated across the donors (Figure 1C). One cluster revealed a group of commonly upregulated early innate immune response genes for PLHIV and a second cluster contained typical interferon response genes (Figure 1C), which was corroborated by functional enrichment analysis (Figure 1D; Supplementary Table S2). Upregulation of alarmins *S100A8* and *S100A9* (cluster 1), which have been previously associated with inflammation (27, 28) indicated a strong signal from the myeloid cell compartment. In cluster 2, *STAT1*, previously linked to enhanced inflammation in HIV (29, 30), was strongly expressed. Both heatmap visualization (Figure 1C) and

10.3389/fimmu.2023.1275136

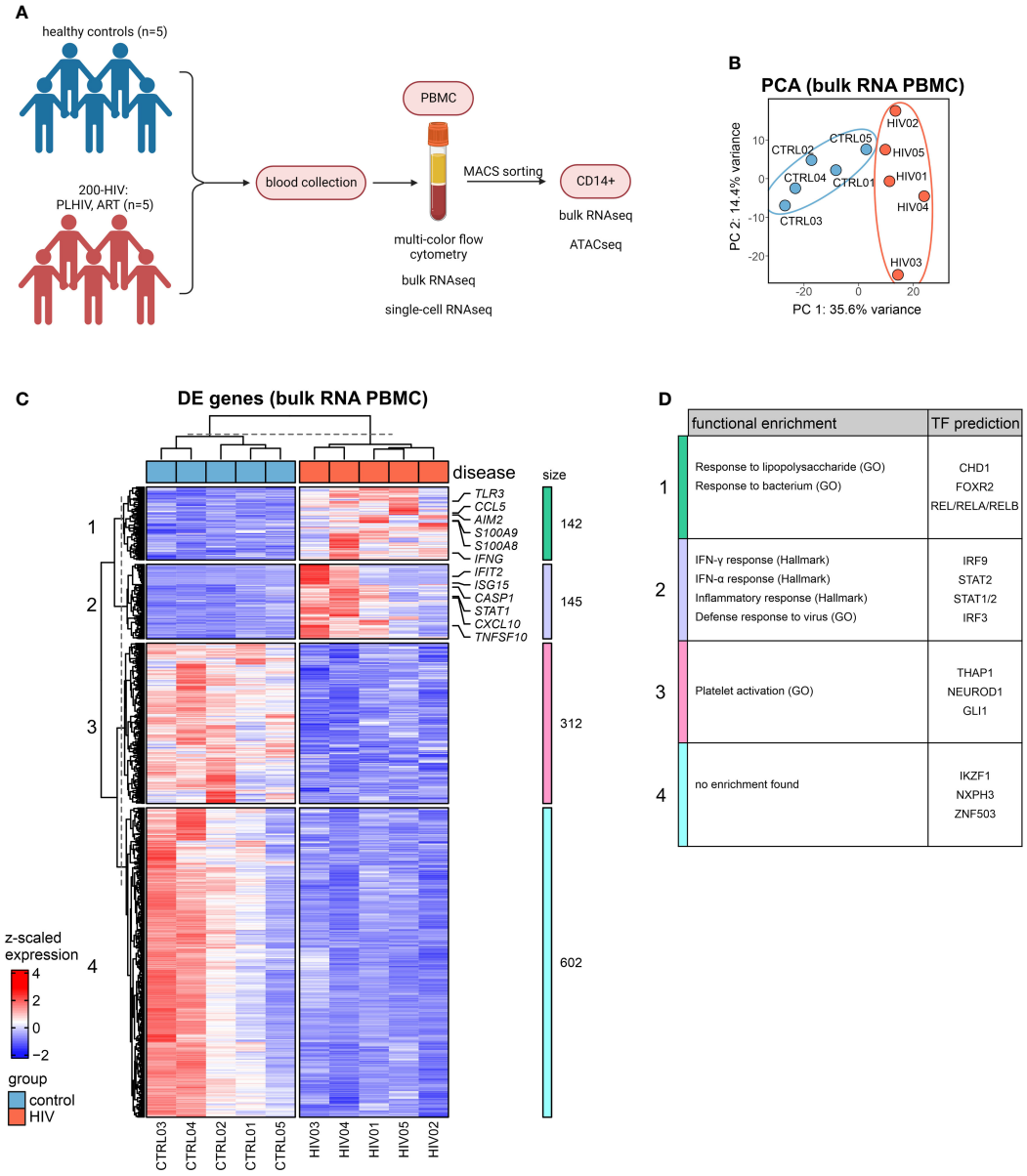


FIGURE 1
Bulk transcriptomes from PBMC in PLHIV are dominated by monocyte-related proinflammatory programs **(A)** Overview of the study design. **(B)** Principal component analysis (PCA) of bulk RNA-seq data from PBMCs. **(C)** Heatmap of DEG (adj. p.val<0.05, |FC|>1.5) from bulk PBMC transcriptomes based on HIV vs. control (see Figure S1C) and hierarchical clustering of genes into 4 clusters. **(D)** Functional enrichment using the GO and Hallmark databases and transcription factor (TF) prediction of gene clusters from **(C)** (full list see Supplementary Table S2).

gene set variation analysis (Figure S1C) showed the highest heterogeneity among the five patients in genes belonging to cluster 2.

Collectively, analysis of bulk transcriptomes from PBMCs of PLHIV revealed upregulation of innate and myeloid proinflammatory gene programs.

Bulk transcriptomics of monocytes in PLHIV reveals enriched IFN-signaling

The bulk transcriptomes of PBMCs pointed towards the involvement of myeloid cells in PLHIV, and indeed plasma concentrations indicated elevated monocyte-specific soluble factors in circulation such as sCD163 and sCD14, a classical marker of HIV disease progression and monocyte activation (8, 31, 32), while other markers such as liver-derived C-reactive protein (CRP) did not show a significant elevation in these PLHIV (Figure S2A). Consequently, we isolated $CD14^+$ monocytes from the same donors (Figure S2B) and analyzed their transcriptomes. DEGs were calculated for the comparison of PLHIV vs. control, resulting in 65 up- and 6 down-regulated genes (|FC|>1.5, p-value <0.05, IHW) (Figures 2A, S2C). Upregulated genes included several type I IFN-related genes such as *CXCL10*, *STAT2*, *MX2*, and *XAF1* (Figures 2B, S2D). Functional enrichment analysis of the upregulated DEGs

10.3389/fimmu.2023.1275136

supports these findings on the pathway level with IFN response and response to the virus being the most highly enriched terms (Figure 2C). The intersection of the CD14$^+$ DEG with those from the PBMC data revealed 3 shared downregulated (*HERC2P10, HSBP1L1, PHLDB3*) and 21 upregulated (e.g. *CXCL10, SERPING1, GBP1*) genes, most of which belong to cluster 2 of the PBMC DEGs (Figure 2D).

To investigate a possible epigenetic component of the disease-associated changes, we performed ATAC-seq of sorted CD14$^+$ monocytes. Using default analysis criteria (|FC|>1.5, adj. p-value < 0.05), we identified no differentially accessible regions (DARs) when comparing cells from PLHIV with control donors (Figure S2E).

Collectively, the CD14$^+$ monocytes in PLHIV show clear signs of transcriptional activation of IFN-mediated pathways which is not significantly impacted by chromatin packing.

"Anti-viral" monocyte state is persistent in PLHIV

To address whether changes in the transcriptomes of PBMCs (Figure 1), as well as isolated CD14+ monocytes (Figure 2), are due to general alterations in the transcriptional programs of the myeloid compartment or due to the presence of disease-specific cell states, scRNA-seq was performed on PBMCs of the same individuals (Figure 3). Transcriptomes from 31,566 single cells were produced representing all major immune cell types of the peripheral circulation according to cluster-specific markers known in literature, such as monocytes (*LYZ, S100A9, S100A8*), CD4$^+$ T cells (*IL7R, TRAT1*), CD8$^+$ T cells (*GZMH, CCL5, CD3G*) and NK cells (*GNLY, NKG7, KLRF1*) (Figures 3A; S3A). Density-based coloring of the UMAP for PLHIV and control groups disclosed a major transcriptional shift in the monocyte cluster, in the CD8$^+$ T cell cluster, but not in the CD4$^+$ T cell cluster (Figure 3B). These differences are also reflected in changes in the number of DEG (log2FC=0.25, adj. p-value<0.05, min.pct=0.1) (Figure 3C). Compared to other immune cell populations, monocytes showed the highest number of DEGs comparing PLHIV with controls, 90 up- and 25 down-regulated genes. Functional enrichment analysis on the HIV-specific up-regulated DEG of the monocyte compartment included terms such as 'IFN-γ response", "IFN-α response" and "response to virus" (Figure 3D), in line with the PBMC and CD14 bulk RNA-seq data (Figures 1B, 2C). Similar to the bulk data produced from CD14$^+$ monocytes, 19 genes were also upregulated in the monocyte cluster resulting from scRNA-seq, including *XAF1* and *GBP1* (Figures S3B, E; Supplementary Table S3). To confirm the upregulation of the genes from that intersection, we measured protein levels of SAMD9L, VAMP5, IFIT3, GBP1, SELL, and EIF2AK2, which are

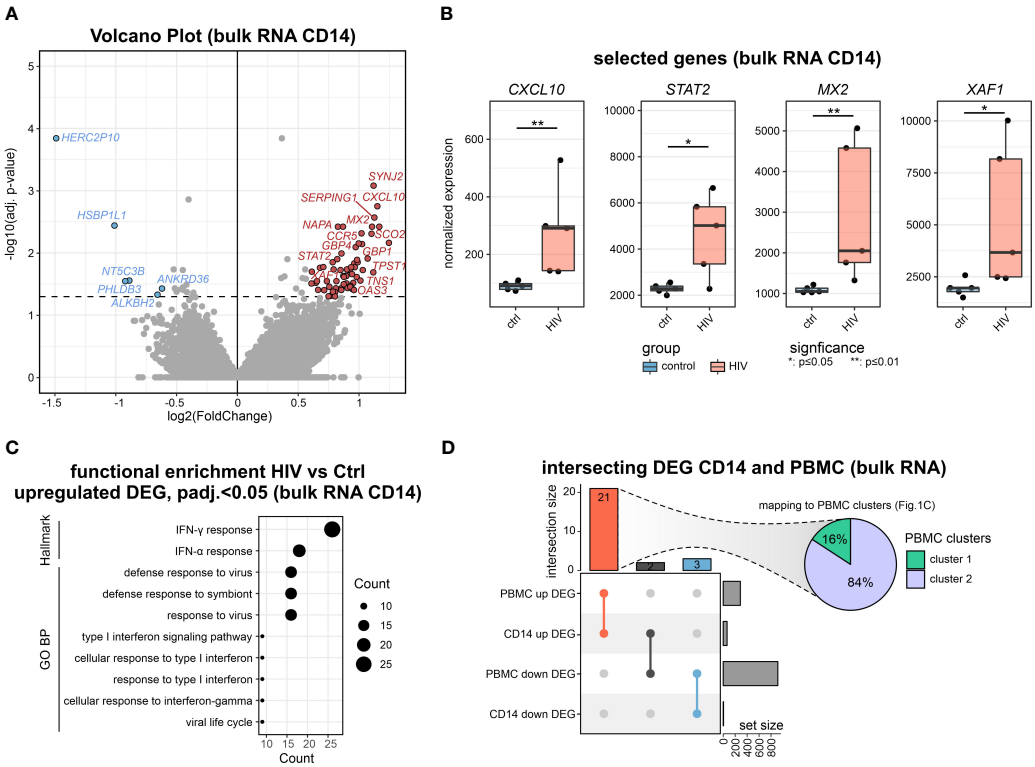


FIGURE 2
Bulk transcriptomics of monocytes in PLHIV mainly reveals IFN-signaling **(A)** Volcano plot showing the DEGs (adj. p.val<0.05, |FC|>1.5) in HIV vs. control of bulk CD14$^+$ monocyte transcriptomes. **(B)** Boxplot and whisker of selected HIV-specific genes. Wilcoxon rank-sum for statistical testing (*: p-value <0.05, **: p-value <0.01). **(C)** Functional enrichment using the GO and Hallmark databases upregulated DEG (HIV vs Ctrl). **(D)** Intersecting DEG for the comparison of HIV vs Ctrl in bulk transcriptomes from CD14$^+$ monocytes and PBMCs. Commonly upregulated DEG mapped to PBMC clusters from Figure 1C.

10.3389/fimmu.2023.1275136

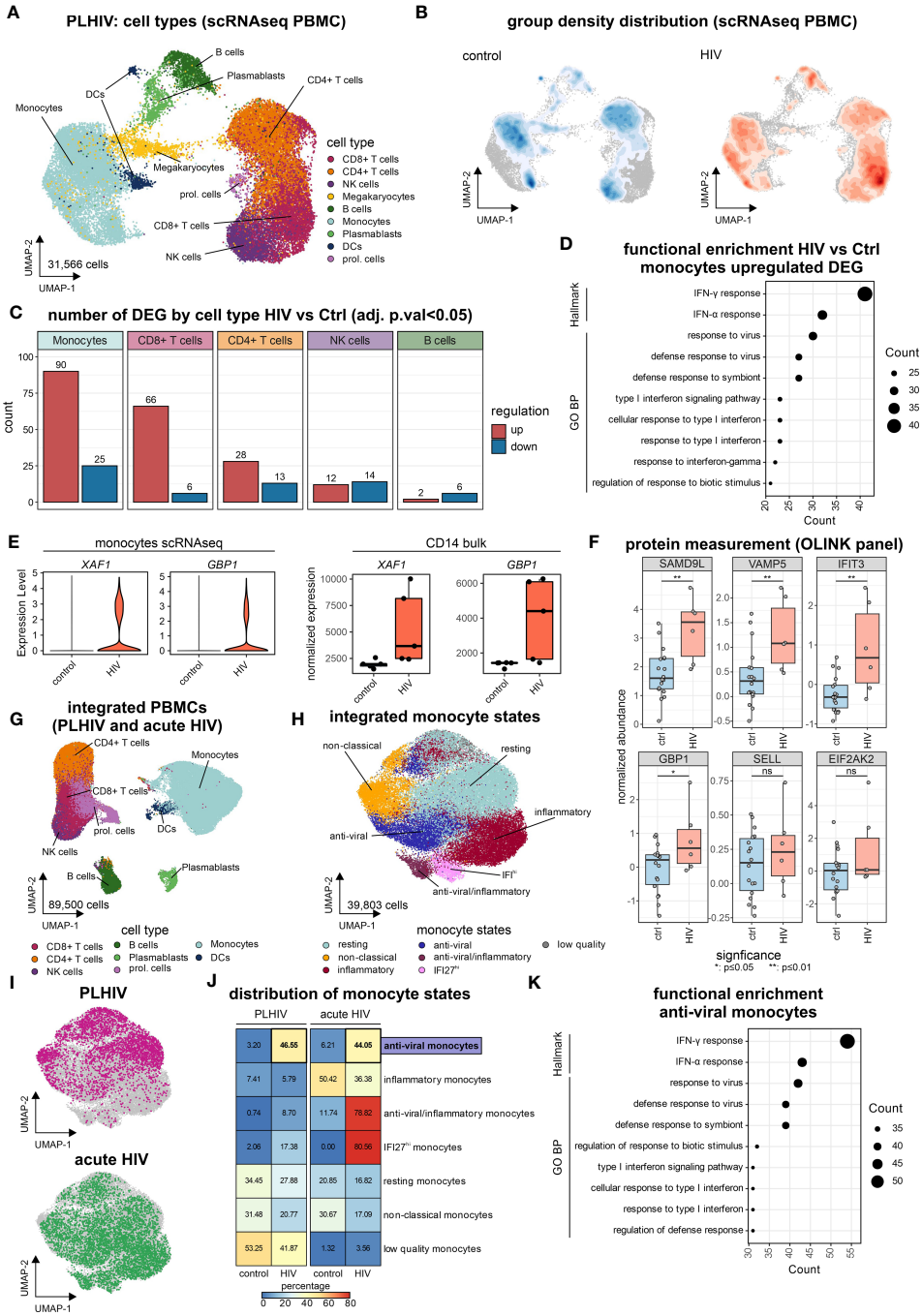


FIGURE 3
"Anti-viral" monocyte state is persistent in PLHIV **(A)** UMAP of PBMCs from PLHIV patients (n= 31,566 cells) indicating identified cell types. **(B)** UMAP from **(A)** colored by disease group density distribution. **(C)** Number of DEG (adj. p.val<0.05, |log2FC|>0.25, min.pct=0.1) by major cell types for the comparison HIV vs Ctrl. **(D)** Functional enrichment using the GO and Hallmark databases for HIV-specific (up-regulated) genes in monocytes. **(E)** Marker expression of *XAF1* and *GBP1* by disease group for monocytes extracted from scRNA-seq data (left panel) and bulk CD14^{+} monocytes (right panel). **(F)** Protein level quantification for SAMD9L, VAMP5, IFIT3, GBP1, SELL, and EIF2AK2 using the Olink system. Wilcoxon rank-sum for statistical testing (ns: not significant, *: p-value <0.05, **: p-value <0.01). **(G)** UMAP of integrated PBMCs from PLHIV **(A)** and acute HIV (Kazer et al., n= 59,286 cells) for commonly present cell types in both datasets, identified cell types are indicated (total dataset n= 89,500 cells). **(H)** UMAP of integrated monocyte subset (n= 39,803 cells) from PLHIV and acute HIV annotated by signatures from Kazer et al. and cluster marker expression. **(I)** UMAP of integrated monocytes colored by dataset origin (PLHIV and acute HIV), each n= 10,000 cells. **(J)** Confusion matrix heatmap showing the distribution of monocyte cell states for disease groups stratified by dataset. **(K)** Functional enrichment using the GO and Hallmark databases for markers (from Figure S3E) of the 'anti-viral' monocyte state.

all related to IFN responses (Figure 3F). In PLHIV, all six proteins showed elevated levels compared to healthy controls with SAMD9L, VAMP5, IFIT3, and GBP1 being significant.

To relate our findings from chronic HIV in PLHIV to acute HIV, in which patients did not yet receive ART and have high viremia, we integrated the newly produced data with our previously published results using the same microwell-based single-cell technology describing several inflammatory monocyte states in acute HIV infection (33) resulting in 89,500 single-cell transcriptomes (Figures 3G, S3C, D). To investigate the possible presence of chronic disease-specific cell states within the monocyte compartment, we subsetted the monocytes of the integrated scRNA-seq dataset (Figure 3H). Clustering of the monocyte compartment resulted in seven monocyte substrates, which could be annotated based on the previously reported acute monocyte states (33). These included several inflammatory monocyte states associated with acute HIV infection, e.g. anti-viral/inflammatory or *IFI27*^{hi} monocytes (Figures 3H, S3E). Monocytes from our new data predominantly exhibited resting and non-classical states, irrespective of HIV group (Figures 3I, J). Chronic HIV was characterized by an 'anti-viral' monocyte state that was also found during acute infection (Figure 3J). This 'anti-viral' monocyte state expresses interferon-related genes, e.g. *IFIT3* and *ISG15* (Figure S3E), and is strongly enriched for the hallmarks 'IFNγ response' and 'IFNα response' as well as the GO term 'response to virus' (Figure 3K), reminiscent of our results in PBMCs (Figures 1C, D) and CD14$^+$ monocytes (Figure 2C).

Even within the resting and non-classical monocyte substates that do not exhibit major changes in proportions between the clinical groups (Figure 3J), differentially expressed genes (log2FC=0.25, adj. p-value<0.05, min.pct=0.1) for PLHIV vs. controls (resting: 70 DEGs, non-classical: 36 DEGs) had a substantial overlay with the DE genes identified from bulk PBMC data, i.e. clusters 1 and 2 (Figures 1C, S3F; Supplementary Table S3). ScRNA-seq also revealed heterogeneity in cell state distribution in the group of the PLHIV, which was not apparent in the healthy individuals (Figure S3G).

Collectively, single-cell transcriptomics identified monocytes as the major cause of changes in PLHIV. Common alterations were evident across all identified cell states, including resting and non-classical monocytes, yet scRNA-seq uncovered elevated numbers of monocytes in the 'anti-viral' cell state in chronic HIV that had been described for acute HIV infection. Thus, pathology in PLHIV is a combination of molecular alterations and proportion changes that could only be revealed by analysis on the single-cell level.

Drug repurposing to reverse monocyte reprogramming in PLHIV

To illustrate how to identify potential drug targets for reversing a molecular phenotype, here the changes observed in monocytes, we performed a drug repurposing approach (Figure 4A) following a previously established methodology (34). In brief, genes up- and down-regulated in PLHIV who are under ART from scRNA-seq monocytes, bulk RNA-seq PBMCs, and bulk RNA-seq CD14$^+$ monocytes were subjected to the drug prediction databases iLINCS and CLUE (35, 36), resulting in 519 predicted drugs (Supplementary Table S4). From those drugs, 17,641 signatures were retrieved from iLINCS and used as input for GSEA on the bulk RNA-seq CD14$^+$ monocytes and PBMC datasets. Drug signatures were then clustered by their delta normalized enrichment score (ΔNES), resulting in 50 clusters (Figure 4B; Supplementary Table S4). The ΔNES indicates the efficiency of the respective drug signature to reverse the PLHIV-specific signature, with higher ΔNES indicating a more complete reversal. Cluster 43, consisting of 32 signatures, showed the highest ΔNES for CD14$^+$ monocytes and also a high ΔNES for PBMCs (Figure 4C). To decipher the commonalities of those drug responses, we investigated recurring target genes of all drug signatures in the cluster (Figure 4D). A majority of genes were interferon-associated such as *IFI27*, *OAS1*, *MX1*, and *IFI44L*, and the target genes were strongly enriched in the 'anti-viral' and 'anti-viral/inflammatory' monocyte states (Figure S4A).

Of the 32 drug signatures, we chose five among the top 20 drugs according to ΔNES for CD14$^+$ monocytes in PLHIV. Four of them had been studied in the context of HIV infection [trametinib (37), sunitinib (38, 39), sitagliptin (40, 41), clofarabine (42)], but had not been reported to alter transcriptional programs in monocytes. Additionally, the predicted antibiotic doxycycline, for which neither anti-viral nor immune-modulating function has been reported, was chosen as well. Instead of addressing the viral life cycle, this approach predicts a potential impact on the host's immune response to these drugs. To test this hypothesis and validate our predictions, we performed *in vitro* experiments stimulating PBMC from PLHIV with the respective drugs.

Six independent PLHIV were recruited, PBMCs were isolated and co-cultured in the presence of the selected drugs or with DMSO as control (Figure 4A, right panel). After overnight incubation, RNA was extracted and bulk transcriptomics was performed to measure transcriptional changes induced by the respective treatment (Figure S4B). The different *in vitro* treatments resulted in prominent transcriptional changes in the PBMCs, evident in the PCA with the strongest alterations induced by doxycycline followed by trametinib, sunitinib, and clofarabine (Figure 4E). Differential expression analysis reflected this finding in the number of DE genes (Figure S4C). Of note, doxycycline, trametinib, and sunitinib induced a greater number of downregulated DEGs.

Based on our previous findings, we tested the influence of the different treatments by analyzing the reduction of gene signature enrichment for 1) the recurring target genes of cluster 43 identified from the drug repurposing pipeline (n=35), 2) the 'anti-viral monocyte' markers from our integrated single-cell RNA-seq analysis (n=137), and the hallmark terms 3) 'IFNγ response' (n=200) and 4) 'inflammatory response' (n=200) (Figure 4F). Sunitinib and doxycycline showed the most significant impact, strongly reversing the four different HIV-specific gene signatures. Trametinib also showed strong, clofarabine a moderate, and sitagliptin no reductions of the four signatures in our *in vitro* verification experiment. These differential effects of the different

10.3389/fimmu.2023.1275136

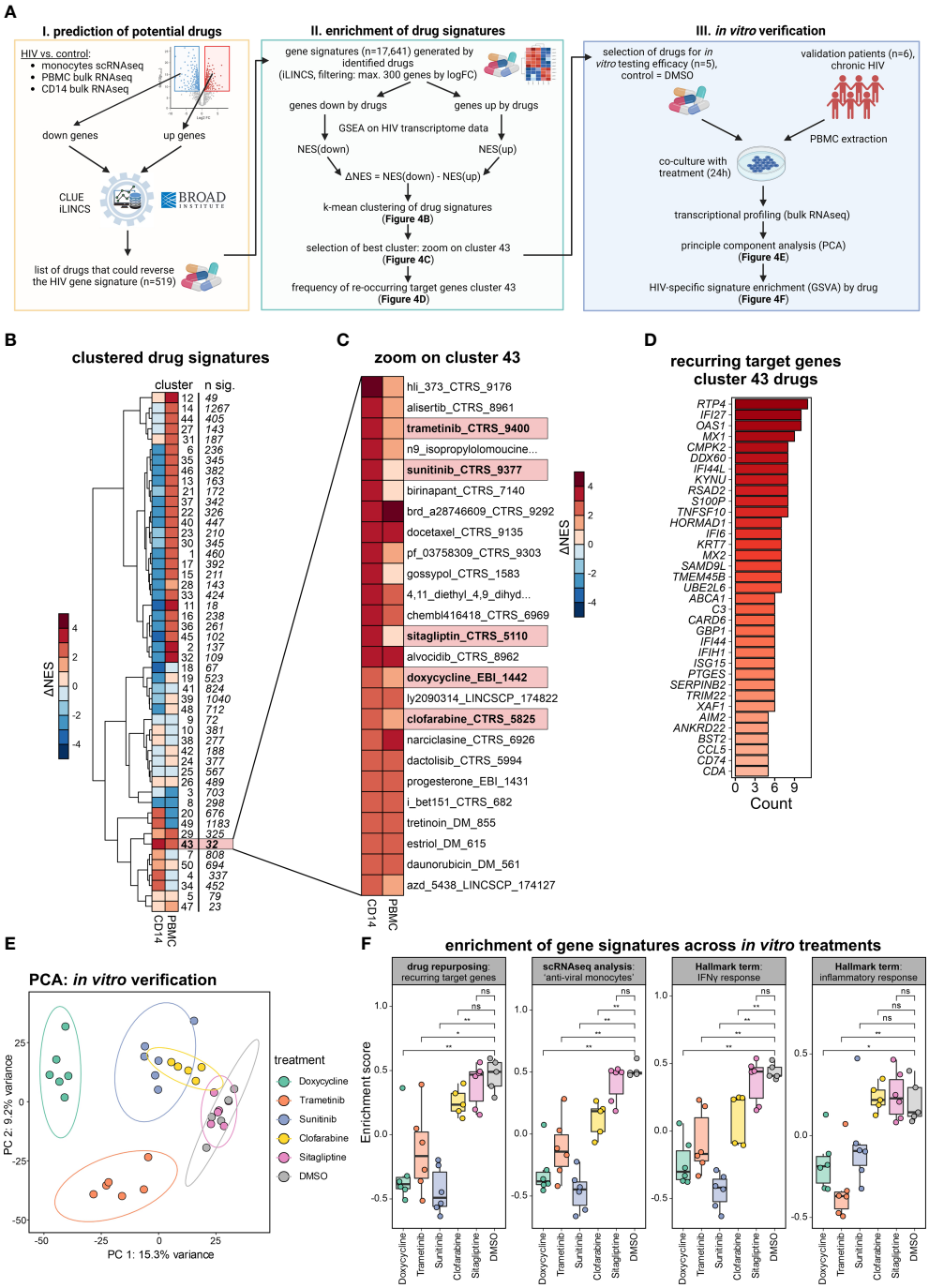


FIGURE 4
Drug repurposing to reverse monocyte reprogramming in PLHIV. **(A)** Drug prediction workflow and follow-up *in vitro* verification, NES=normalized enrichment score. **(B)** Heatmap showing hierarchical clustering (k-mean=50) of ΔNES from all drug signatures (n= 17,641) as groups enriched on transcriptomes from bulk CD14$^+$ monocytes and bulk PBMCs. **(C)** Zoom into cluster 43 from **(B)**, depicting all involved drug signatures. **(D)** Recurring target genes of drug signatures identified in cluster 43 from **(C)**. **(E)** Principal component analysis (PCA) of bulk PBMC transcriptomes of the *in vitro* verification experiment (five HIV donors with six conditions). Samples colored by treatment, DMSO as untreated control. **(F)** Enrichment of gene signatures across *in vitro* treatments, analyzed signatures include the recurring target genes from cluster 43 (drug repurposing), marker for the 'anti-viral' monocytes (integrated scRNA-seq analysis), and the hallmark terms 'IFNγ response' and 'inflammatory response'. Wilcoxon rank-sum for statistical testing (ns: not significant, *: p-value <0.05, **: p-value <0.01).

drugs are also seen on the gene level when investigating the top leading edge genes of the four signatures by each drug (Figure S4D).

Taken together, we predicted drugs that could reverse the altered monocyte-derived signatures and confirmed our repurposing approach *in vitro* with the drugs doxycycline and sunitinib strongly reversing the HIV-specific gene signatures, making them repurposed drug candidates of interest.

Discussion

In the present study, we illustrate in a small group of PLHIV derived from our previous cross-sectional 200-HIV cohort study (1) that single-cell and bulk transcriptomes of isolated immune cells revealed reprogramming in multiple cellular compartments in PLHIV, with innate immune cells, in particular monocytes, showing most profound changes. We further illustrate that a certain cellular state of monocytes, previously reported in acute HIV infection can be observed in PLHIV, while other cell states associated with acute inflammation are specific for acute HIV and absent in PLHIV. Long-term usage of ART in PLHIV results in undetectable viral loads and restores CD4 cell counts to normal levels, and therefore PLHIV patients differ from people with an acute HIV infection that have high-level viremia and reduced CD4 cell counts (33). Despite the small number of PLHIV studied, which clearly showed heterogeneity in their transcriptional profiles, we also illustrate that combined bulk and single-cell data of these PLHIV was already sufficient to predict drug candidates for reversing the observed transcriptional deviations in the monocyte compartment. While technically applicable to a cohort study setting, ATAC-seq of this small number of PLHIV did not reveal any significant differences, which clearly points towards the need for larger cohorts when assessing chromatin landscape differences. As such the study reported here provides the necessary information to include sophisticated transcriptome and epigenome data generation to be integrated into the larger 2000-HIV cohort study currently recruiting PLHIV including elite controllers.

The combined analysis of bulk transcriptomes from PBMC and purified $CD14^+$ monocytes together with single-cell transcriptomes from blood allowed us already in a rather small number of PLHIV to define major changes within the peripheral immune cell compartment, e.g. the identification of a gene cluster characterized by IFN signaling. The higher-resolution information from scRNA-seq revealed that some of the changes observed in the PBMC-derived transcriptomes was due to molecular changes in monocytes including cell-state differences, but not due to cell-type distribution differences, further supporting the use of higher-resolution technologies in larger cohort studies. While IFN-signaling related molecular changes (cluster 2, Figures 1C, D) were also captured in bulk transcriptomes from purified $CD14^+$ monocytes (Figure 2), the overall information content from purified $CD14^+$ monocytes was surprisingly low, indicating that many of the changes observed in PBMC are derived from other monocyte cell states ($CD14^{low/-}$) and other cell types. Single-cell transcriptomes clearly corroborated this hypothesis showing that basically all immune cell types exhibited transcriptional changes in PLHIV. With the lowest information content and highest technical effort, we concluded that cell-type isolation procedures are not suitable for larger cohort studies on PLHIV. Moreover, when assessing DEG in monocytes using both bulk and single-cell transcriptomes, we detected less DEG in bulk and only a small intersection with single-cell data (n=19, Figure S3B). Differences in experimental sample handling or sequencing resolution could explain this small intersection, however, even though certain genes were not tested to be significantly altered in both methods, the general pathway activation towards IFN responses was uncovered by both methods.

The systemic assessment of single-cell transcriptomes derived from PBMC of PLHIV revealed that major transcriptional reprogramming was mainly observed in monocytes and $CD8^+$ T cells with fewer changes in $CD4^+$ T cells, NK cells, and B cells. Focusing on the cell compartment with the major changes, we revealed a cell state composition in PLHIV including the well-described classical and non-classical monocyte states, but also a cell state we previously termed 'anti-viral monocytes' in acute HIV infection (33). Interestingly, this particular cell state showed high heterogeneity between PLHIV, which will have to be studied in larger cohorts to better define whether there is a pattern reminiscent of PLHIV endotypes or whether this might be explained by individual clinical incidents prior to blood sampling. Despite the heterogeneity of this monocyte state, the assessment of genes enriched in gene clusters derived from bulk transcriptomes indicated that even classical and non-classical monocyte states in PLHIV are characterized by elevated expression of cluster 2 genes, supporting the notion that despite the observed heterogeneity, persistent IFN signaling seems to be a major hallmark of persistent inflammation in PLHIV (43). Based on these informative and promising results we propose to integrate these levels of omics technologies into larger PLHIV studies.

As we identified a major theme for persistent inflammation in this small number of PLHIV, we addressed whether this information would already be sufficient to identify drug candidates by a reverse transcriptome approach (34). Interestingly, while most therapeutic strategies are currently addressing alternative antiviral drugs with less toxicity or treatment strategies aiming at minimizing ART toxicity, fewer drug regimens address immunomodulation itself including the use of purinergic P2X receptor inhibitors (44) or statins (45–49). In clinical studies testing the efficacy of these therapeutic approaches to lower inflammation in PLHIV, mainly soluble mediators (e.g. CRP, sCD14, IL-6, sCD163) measured in plasma or serum were used as readouts, while high-resolution technologies to address molecular changes in immune cells were not reported. We exemplified here, how such an approach could be applied to the identification of drug candidates lowering the inflammatory response observed in PLHIV. We focused on a cluster of drugs with a particularly high probability of reversing the transcriptional alterations observed in monocytes and experimentally validated a small number of drug candidates. A surprising finding was that the antibiotic doxycycline induced the strongest effect mainly reducing gene expression. Together with sunitinib, doxycycline was most effective in reversing gene expression alterations of 1) the major target genes used for drug prediction, 2) of the marker genes expressed in monocytes with the 'antiviral' cell state, 3) of the hallmark genes related to IFN signaling, and 4) hallmark genes related to the pro-inflammatory response. These findings strongly suggested that drugs such as doxycycline might not only function as antibiotics but also modulate host immune responses. This is similarly true for the drug candidates sunitinib and trametinib, which have been developed for completely different purposes (50,

51). Importantly, the modulation of the monocyte-related immune activation should not be considered yet as being unrestrictedly helpful for PLHIV, as it is not yet entirely clear whether these gene programs would play a clinically beneficial role or not. As these findings have to be considered as proof-of-concept, further investigations using more drug candidates, different drug concentrations, and further optimized computational and miniaturized experimental procedures in a larger group of PLHIV are certainly warranted to more quickly identify promising new drug candidates counteracting the inflammatory state in PLHIV under ART therapy.

Limitations of the study

The present study was conceptualized based on the previous cross-sectional 200-HIV cohort study (1) to determine whether the combination of high-resolution and high-content technologies such as bulk and scRNA-seq data would lead to additional insights into the pathophysiology of immune deviations in PLHI and therefore, only a limited number of individuals were included in this study. As the main purpose was to determine the best strategy to scale these technologies to larger clinical cohorts, we were surprised that despite a rather small number of individuals studied and obvious heterogeneity within the group of PLHIV, we could retrieve important information about major molecular changes on transcriptome level in all immune compartments. However, it became also clear that other layers, e.g. chromatin landscapes as assessed by ATAC-seq require a much larger number of individuals to determine whether immune cells in PLHIV are also altered on this epigenetic level. Based on these initial findings, we have now started to include these technologies in the much larger 2000-HIV cohort study of approx. 2000 PLHIV to study aspects such as disease heterogeneity, potential disease endotypes, and association of cellular changes with clinical trajectories, or to determine potential biomarkers predicting disease outcome. Certainly, the observation that innate immune cells such as monocytes show the most pronounced transcriptional reprogramming in PLHIV was unexpected and will be one major focus within the currently being assembled cohort of PLHIV. Moreover, the identification of these monocyte-derived programs also opens new avenues toward the identification of new mechanisms on how transcriptional alterations contribute to immune dysregulation in PLHIV.

Methods

Lead contact

Dr. Anna C. Aschenbrenner, anna.aschenbrenner@dzne.de.

Materials availability

This study did not generate unique reagents.

Data and code availability

Bulk RNA-seq datasets and single-cell RNA-seq data have been deposited at the European Genome-phenome Archive (EGA) and are publicly available under the accession numbers.

All original code is stored on FASTGenomics: https://beta.fastgenomics.org/p/HIV_Pilot

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Study cohort

Five PLHIV were recruited from the outpatient HIV clinic of the Radboud University Medical Center on March 26-28th 2019. Included patients were five males of Dutch/Western-European ethnicity who were receiving cART for more than 6 months and latest HIV-RNA levels ≤200 copies/ml. Ethical approval was granted by the Ethical Committee of the Radboud University Medical Center Nijmegen, the Netherlands under registration number NL42561.091.12). Additionally, five age-/sex-matched healthy volunteers were included as controls (age 43-61), and ethical approval was granted by the Ethical Committee of the Radboud University Medical Center Nijmegen, the Netherlands under registration number NL32357.091.10). For the *in vitro* verification experiments of drugs, six additional male PLHIV were recruited (age 26-43, with ethical approval granted by the Ethical Committee of the Radboud University Medical Center Nijmegen, the Netherlands under registration number NL68056.091.18). Written consent was obtained from all participants involved in this study and experiments were conducted according to the Declaration of Helsinki principles.

PBMC isolation

Human peripheral blood mononuclear cells (PBMCs) were isolated by dilution of blood in pyrogen-free PBS and differential density centrifugation over Ficoll-Paque (GE Healthcare, UK) as previously described by (52). Briefly, the interphase layer was collected, and cells were washed with cold PBS. Cells were resuspended in RPMI 1640 culture medium (Roswell Park Memorial Institute medium; Invitrogen, USA) supplemented with 50 g/mL gentamicin, 2 mM glutamax (Gibco, Life Technologies, USA), and 1 mM pyruvate (Gibco) and quantified. A fraction of PBMCs was viably frozen for later use. The cell suspension was spun down for 5 min at 300g, 4°C, after which all supernatant was removed. Cells were very gently resuspended in freezing medium (90% fetal calf serum, 10% DMSO) and aliquoted into cryovials. They were placed first at -80°C in a CoolCell freezing container (Corning), after which they were transported the next day on dry ice and moved to liquid nitrogen storage. For the thawing of PBMCs,

one vial of 5 million cells was thawed in 10ml RMPI medium supplemented with 10% FCS.

Preparation of Seq-Well arrays/libraries/sequencing

Seq-Well arrays and libraries were prepared from isolated PBMCs as described previously (24).

Measurements of plasma markers

Clinical plasma markers were measured using ELISA (Duoset or Quantikine, R&D Systems) for IL18-BP, IL-18, hsCRP, sCD14, sCD163 or using SimplePlex Cartridges (Protein Simple) for IL-6, all performed according to manufacturers' instructions. As a reference, the mean of healthy controls from van der Heijden et al. (1) were used.

Isolation of CD14+ monocytes

CD14+ monocytes were isolated from PBMC by magnetic-activated cell sorting (MACS) positive selection with CD14 microbeads (Miltenyi Biotec), according to the manufacturer's instructions. Depending on the available PBMC counts used as input, either MS or LS columns were used (Miltenyi Biotec). After isolation, cells were again resuspended in a Dutch modified RPMI culture medium (Invitrogen) supplemented with 50 µg/mL gentamycin, 2 mM glutamax and 1 mM pyruvate (Gibco, Life Technologies).

Flow cytometry

Frozen PBMCs were thawed then stained for surface markers (Supplementary Table S1) in DPBS with BD Horizon Brilliant Stain Buffer (Becton Dickinson) for 30min at 4°C. To distinguish live from dead cells, the cells were incubated with LIVE/DEAD Fixable Yellow Dead Cell Stain Kit (1:1000 – Thermo Scientific). Following staining and washing, the cell suspension was fixed with 4% PFA for 10 min at room temperature to prevent any possible risk of contamination due to aerosol formation during sample handling and acquisition. Flow cytometry analysis was performed on a BD Symphony instrument (Becton Dickinson) configured with 5 lasers (UV, violet, blue, yellow-green, red).

ATAC-seq

Frozen PBMCs were thawed and sorted on a BD FACSAria III (Blue, Yellow-Green, Red, and Violet lasers), and 20,000 live $CD14^+$ cells were sorted and spun down at 500×g for 5 min at 4°C. The cell pellet was washed with 50 µL of cold 1x PBS buffer and spun down at 500 ×g for 5 min at 4°C. The pellet was then resuspended in 50 µL of cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630) and spun down immediately at 500×g for 10 min at 4°C. The supernatant was then discarded, and the transposition reaction was immediately performed. To perform the transposition reaction, a mixture of transposase, 5x TAPS-DMF buffer (50mM TAPS (T5130 SIGMA), 25mM MgCl2, 50% DMF (N, N-Dimethylformamide)), and water was combined and added to the cell pellet. The transposition reaction was incubated at 37°C for 30 min. Following transposition, the DNA was purified using a Qiagen MinElute Kit. The transposed DNA was eluted in 10 µL of water, and purified DNA was stored at 4°C until the following day or at -20°C.

To amplify the transposed DNA fragments, a PCR mixture was prepared using the purified DNA, nuclease-free water, customized Nextera PCR primers, and NEBNext High-Fidelity 2x PCR Master Mix. The PCR mixture was cycled as follows: 72°C for 5 min, 98°C for 30 sec, 98°C for 10 sec, 63°C for 30 sec, and 72°C for 1 min. Steps 3-5 were then repeated 11 times for a total of 12 cycles. The PCR products were then purified using a Qiagen MinElute Kit and eluted in 12 µL of water. To validate the quality and concentration of the PCR products, gel electrophoresis was performed using the TapeStation and Agilent High Sensitivity D1000 kit.

Protein measurements

Proteomic profiling of selected markers was performed as described before (53). In brief, venous whole-blood samples were collected in EDTA tubes and centrifuged into plasma, and then stored at -80°C. Protein measurements were performed by Olink Proteomics AB using the Olink Explore platform. QC and normalization were performed by Olink services. For this study, protein markers of interest were selected.

In vitro verification of selected drugs

To verify the effectiveness of predicted drugs, six different PLHIV from the 200-HIV cohort were re-called, and the PBMCs were extracted and seeded in triplicates with 500,000 cells per replicate. The PBMCs were cultured for 24 hours in the presence of a selected subset of drugs from cluster 43, including trametinib (50 mM in 0.000002% DMSO), sunitinib (100 mM in 0.0001% DMSO), clofarabine (100 mM in 0.00001% DMSO), doxycycline (100 mM in H_2O) and sitagliptin (100 mM in 0.0001% DMSO) or DMSO (0.001%) as control. After incubation, replicates were collected in a total of 1 ml TRIzol reagent and processed for bulk RNA-seq.

Quantification and statistical analysis

RNA-sequencing analysis (bulk RNA PBMC, CD14)

Sequenced reads were aligned and quantified using STAR: ultrafast universal RNA-seq aligner (v2.7.3a) (54) and the human reference genome, GRCh38p13, from the Genome Reference Consortium. Raw counts were imported using the DESeqDataSetFromMatrix function

from DESeq2 (v1.32.0) (55) and rlog transformed according to the DESeq2 pipeline. DESeq2 was used for the calculation of normalized counts for each transcript using default parameters. All normalized transcripts with a maximum overall row mean lower than 10 were excluded resulting in 26,920 present transcripts. All present transcripts were used as input for principal component analysis (PCA). Differentially expressed genes were calculated for HIV vs. control using an independent hypothesis weighting (IHW) adjusted p-value cutoff of 0.05 and an absolute fold change (|FC|) of 1.5. DEGs were used as input for the k-mean clustered heatmap (k=4), generating four clusters.

RNA-sequencing analysis (drug verification analysis)

Sequenced reads were aligned and quantified using kallisto v0.44.0 (56) and the human reference genome, GRCh38p13, from the Genome Reference Consortium. Raw counts were imported using the DESeqDataSetFromTximport function from DESeq2 (v1.32.0) (55) and vst-transformed according to the DESeq2 pipeline. DESeq2 was used for the calculation of normalized counts for each transcript using default parameters. All normalized transcripts with a maximum overall row mean lower than 10 were excluded resulting in 37,952 present transcripts. Variation in the data was identified using the SVA package (v3.40) (57), and batch effects were removed with limma (v3.48.3) (58) using the first six surrogate variables (SVs), which were also added in the design of the dds object. All present transcripts were used as input for principal component analysis (PCA) of the batch-corrected data. Differentially expressed genes were calculated for HIV vs. control using a p-value cutoff of 0.05, an adjusted p-value (IHW) < 0.05 (independent hypothesis weighting), and a |FC|>2. DEGs were used as input for the clustered heatmap.

Transcription factor prediction analysis

The R package RcisTarget (version 1.12.0) (59) was used to predict the transcription factors potentially regulating heatmap cluster-specifically contained gene sets. The genomic regions of TF-motif search were limited to 10kb around the respective transcriptional start sites by using the RcisTarget-implemented "hg19-tss-centered-10kb-7species.mc9nr.feather" motifRanking file. Prediction was performed using the cisTarget function and the resulting top 3 predicted TF, according to their normalized enrichment scores (NES), were selected for each heatmap cluster.

Gene set ontology enrichment analysis

Gene set ontology enrichment analysis using the heatmap clusters as input was performed on the gene sets from the Gene Ontology (GO) biological process (BP) database (60, 61) and the Hallmark gene sets (62) using the R package clusterProfiler (version 4.0.5) (63). Ontologies with the highest and statistically significant enrichment were used for presentation.

Gene set variation analysis

For the enrichment of the genes included in the four different clusters of the DE heatmap (PBMC data) and for the enrichment of the four different transcriptional signatures for the *in vitro* verification of drugs, the GSVA package (version 1.40.1) (64) was applied.

Flow analysis

After pre-processing, compensated fluorescence intensities were exported from FlowJo (BD, v. 10.7.1). Exported.fcs files were imported in R with the flowCore package (v. 2.2.0). Fluorescence intensities were auto-logicle transformed, used for dimensionality reduction using the UMAP algorithm (umap package v. 0.2.7.0) (65) and clustered using the Phenograph package (v. 0.99.1) (66). Cell types were annotated for each cluster by respective marker expression. For visualization, the proportions of main cell types were calculated and stratified by disease group.

ATAC-seq analysis

Reads were aligned to human hg38 reference with bowtie2 (67). Samtools (68) was used to remove adapter offset and to create bam files. Open chromatin peaks were called using MACS2 (69), blacklisted regions (hg38-blacklist.v2.bed.gz, https://sites.google.com/site/anshulkundaje/projects/blacklists), the low covered peaks were excluded, and then the peaks were annotated with gene models from TxDb.Hsapiens.UCSC.hg38.knownGene using the ChIPseeker package (applying annotatePeaks function) (70). Downstream analysis was performed with the DESeq2 (v1.26.0) package (55). Differentially accessible regions (DAR) were detected with a |FC|>1.5 and a corrected p-value > 0.05. With these standard parameters, no DAR were identified.

ScRNA-seq data analysis

ScRNA-seq UMI count matrices were imported to R 4.1 and gene expression data analysis was performed using the Seurat package 4.0.4 (71, 72). Cells with more than 10% mitochondrial reads and less than 200 expressed genes were excluded from the analysis and only those genes present in more than 3 cells were considered for downstream analysis. Moreover, the genes *MT-RNR1* and *MT-RNR2* were excluded. Log-normalization, scaling, and dimensionality reduction steps were performed using the Seurat implemented functions. For scaling, the number of detected transcripts per cell was regressed out to correct for heterogeneity associated with differences in sequencing depth. For dimensionality reduction, PCA was performed on the top 2,000 variable genes identified using the vst method implemented in Seurat. Subsequently, UMAP was used for two-dimensional representation of the data structure using the first 30 PCs. Cell type annotation was based on the respective clustering results combined with the expression of known marker genes. DEG by celltype were calculated for the comparison of HIV vs control with a |log2FC|>0.25, adj. p-value<0.05 and min.pct=0.1.

Data integration

Data integration of the PLHIV PBMCs (this study) and the acute HIV PBMC dataset (33) were integrated using the harmony algorithm (73) based on the first 15 principal components. Prior to integration, the PLHIV dataset was subsetted for major cell types

present in acute HIV. Cell type annotation was based on the respective clustering results combined with the expression of known marker genes.

Integrated scRNA-seq monocyte analysis

The monocyte compartment was subsetted from the integrated PBMCs and subsequently normalized, scaled, and subjected to PCA calculation. For UMAP visualization, the first 10 harmony PCs were used. After clustering the integrated monocytes with the FindNeighbors and FindClusters function from Seurat, monocyte states were annotated according to the signatures described in acute HIV (33) and cluster-specific markers, separating the monocyte population into anti-viral (*TNFSF10, ISG15, IFIT2, IFIT3*), inflammatory (*IL8, IL1B, EREG*), anti-viral/inflammatory (*CCL2, CCL4*), IFI27/30^{hi} (*IFI27, IFI30*), HLA^{hi} (*HLA-DRB1, HLA-DQA1*), resting (*S100A8, S100A9, LYZ*) and non-classical (*FCGR3A, C1QA*) monocytes.

Confusion matrix

For each monocyte cell state, the relative proportion across the groups (HIV, control) was visualized as a fraction of samples from the respective condition contributing to the monocyte cell state stratified by dataset (PLHIV vs. acute HIV).

Drug prediction

To identify drugs that reverse the gene expression signature observed in the comparison HIV vs. control for bulk RNA-seq PBMCs, bulk RNA-seq CD14 monocytes, and scRNA-seq monocytes, the drug prediction databases iLINCS (http://www.ilincs.org/ilincs/), and CLUE (https://clue.io/) were accessed. As input for the drug prediction, the top 1000 (iLINCS) or the top 100 (CLUE) DEGs were used. Drugs reversing the HIV gene expression signature (defined by a negative score) comprised a total of 519 unique drugs. Using the iLINCS API (https://github.com/uc-bd2k/ilincsAPI/blob/master/usingIlincsApis.Rmd), every gene expression signature from each drug listed in the signature libraries iLINCS chemical perturbagens (LINCSCP), iLINCS targeted proteomics signatures (LINCSTP), Disease-related signatures (GDS), Connectivity Map signatures (CMAP), DrugMatrix signatures (DM), Transcriptional signatures from EBI Expression Atlas (EBI), Cancer therapeutics response signatures (CTRS), and Pharmacogenomics transcriptional signatures (PG) was downloaded. Labeling was performed in the following principle: "drug name"_"database"_"database ID". Signatures were ordered by fold change, and only the top 300 genes were used. This resulted in a total of 17,641 unique drug signatures each with an up- and downregulated set. Subsequently, GSEA was performed on the sequencing data for every up- and down-regulated set for each drug and each cluster comparison. The resulting normalized enrichment scores (NES) were used to calculate the delta NES for each drug, defined as ΔNES = NES (down) – NES (up), ergo the difference of the NES from the downregulated set and the NES from the upregulated set of each respective drug. These ΔNES values were then k-mean clustered (k = 40). The cluster with the highest ΔNES values for both CD14 and PBMCs was chosen and uniquely present drugs were shown. The leading edge genes of the downregulation signatures of these drugs (cluster 43) were examined, and the frequency was counted (recurring target genes).

Data visualization

For data visualization, the R packages Seurat, ggplot2 (version 3.3.5) (74), (https://ggplot2.tidyverse.org), pheatmap (version 1.0.12), and ComplexHeatmap (version 2.8.0) (75) were used.

Data availability statement

The datasets presented in this study can be found in online repositories. Bulk RNA-seq datasets and single-cell RNA-seq data have been deposited at the European Genome-phenome Archive (EGA) and are publicly available under the accession number EGAS00001007460.

All original code is stored on FASTGenomics: https://beta.fastgenomics.org/p/HIV_Pilot. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Ethics statement

Ethical approval was granted by the Ethical Committee of the Radboud University Medical Center Nijmegen, the Netherlands under registration number NL68056.091.18). Written consent was obtained from all participants involved in this study and experiments were conducted according to the Declaration of Helsinki principles. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

RK: Writing – original draft, Writing – review & editing, Data curation, Formal Analysis, Investigation, Visualization. LB: Data curation, Formal Analysis, Investigation, Writing – review & editing. JS: Data curation, Investigation, Validation, Writing – review & editing, Formal Analysis. SW-H: Data curation, Formal Analysis, Investigation, Writing – review & editing. MJ-C: Data curation, Investigation, Project administration, Resources, Writing – review & editing. EB: Writing – review & editing, Investigation. NR: Data curation, Formal Analysis, Investigation, Writing – review & editing. AH: Investigation, Writing – review & editing. MH: Investigation, Writing – review & editing, Formal Analysis. MN-G: Investigation, Formal Analysis, Writing – review & editing. TO: Writing – review & editing, Resources. WH: Resources, Writing – review & editing, Data curation, Investigation. LW: Resources,

Writing – review & editing. AS: Resources, Writing – review & editing, Methodology. KH: Writing – review & editing, Investigation. MB: Data curation, Writing – review & editing. MDB: Resources, Supervision, Writing – review & editing. MN: Resources, Supervision, Conceptualization, Funding acquisition, Writing – review & editing. LJ: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. AV: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. JS: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing, Writing – original draft, Investigation. AA: Conceptualization, Funding acquisition, Resources, Supervision, Writing – original draft, Writing – review & editing, Project administration, Data curation, Investigation.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the German Research Foundation (DFG) (INST 37/1049-1, INST 216/981-1, INST 257/605-1, INST 269/768-1, INST 217/988-1, INST 217/577-1, EXC2151 – 390873048; and SFB1454 – 432325352 to JS); Helmholtz-Gemeinschaft Deutscher Forschungszentren, Germany (sparse2big to JS and MN); EU projects SYSCID (733100 to JS); ERA CVD (00160389 to JS); ImmunoSep (847422 to JS); AA is supported by DFG under AS 637/1-1; AS 637/2-1; AS 637/3-1 and SFB1454 (432325352) and by the BMBF under IMMME/01EJ2204D. The 2000-HIV study is supported by an unrestricted grant from ViiV Healthcare (to AV, M.G.N., LJ). MN is supported by an ERC Advanced Grant (833247) and a Spinoza Grant of the Netherlands Organization for Scientific Research.

Acknowledgments

We are grateful to the patients and donors volunteering to participate in this study making this research possible.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1275136/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Blood transcriptomes of PLHIV are dominated by monocyte-related proinflammatory gene programs. **(A)** Overview of age and sex of the cohort by disease group. **(A)** Multicolor flow cytometry (MCFC) cell distribution for HIV and controls. **(B)** Number of DEG for the comparison HIV vs. Ctrl in bulk PBMCs transcriptomes; IHW multiple comparison adjustment and false discovery rate (FDR) cutoff of 5%, significant fold change of >|1.5|. **(C)** Gene set variation analysis (GSVA) of the genes from the four clusters of the DE heatmap (from).

SUPPLEMENTARY FIGURE 2
Bulk transcriptomics of monocytes in PLHIV mainly reveals IFN-signaling. **(A)** Boxplots of clinically relevant markers measured in the serum of PLHIV. References as blue bars (1). **(B)** Overview of MACS CD14 positive selection. **(C)** Number of DEG for the comparison HIV vs. Ctrl in bulk RNA-seq CD14; IHW multiple comparison adjustment and false discovery rate (FDR) cutoff of 5%, significant fold change of >|1.5|. **(D)** Boxplot and whisker of selected genes.

SUPPLEMENTARY FIGURE 3
'Anti-viral' monocyte state is persistent in PLHIV. **(A)** Cell type marker expression of the PLHIV dataset for all identified cell types. **(B)** Overlap of up-regulated DEG from monocytes extracted from scRNAseq and bulk $CD14^+$ transcriptomes (Supplementary Table S3). Genes contributing to the IFN-γ or IFN-α response pathways are indicated for the intersection as well as the uniquely identified DEG. **(C)** UMAP of integrated PBMCs from PLHIV and acute HIV split by dataset (total n= 89,500 cells, each 30,000 cells depicted). **(D)** Cell type marker expression of the integrated HIV dataset for all included cell types. **(E)** Monocyte cell state marker of the integrated monocytes from PLHIV and acute HIV. **(F)** Mapping of HIV-specific (upregulated) DEG of resting and non-classical monocyte states (for DEG see Supplementary Table S3) to PBMC clusters from . **(G)** Integrated monocyte UMAP subsetted for chronic HIV and stratified by donor.

SUPPLEMENTARY FIGURE 4
Drug repurposing to reverse monocyte reprogramming in PLHIV. **(A)** Enrichment of recurring target genes from cluster 43 in monocyte states of the integrated monocyte analysis (see). **(B)** Included samples by treatment condition after quality control (QC) for the *in vitro* verification experiment. **(C)** Number of DEG (adj. p-value<0.05, |FC|>2, IHW) for each treatment vs. control (DMSO). **(D)** Heatmap showing the union of top leading edge genes of each signature (from) for each treatment ranked by adj. p-value.

SUPPLEMENTARY TABLE 1
Donor overview.

SUPPLEMENTARY TABLE 2
Functional enrichment (GO and Hallmark) and transcription factor (TF) prediction of bulk RNA-seq PBMC heatmap clusters (related to Figure 1) and MCFC marker.

SUPPLEMENTARY TABLE 3
ScRNA-seq monocytes DEG (related to Figure 3).

SUPPLEMENTARY TABLE 4
Predicted drug clusters and signatures (related to Figure 4).

References

1. van der Heijden WA, Van de Wijer L, Keramati F, Trypsteen W, Rutsaert S, Horst RT, et al. Chronic HIV infection induces transcriptional and functional reprogramming of innate immune cells. *J Clin Invest Insight* (2021) 6(7). doi: 10.1172/jci.insight.145928

2. Deeks SG, Tracy R, Douek DC. Systemic effects of inflammation on health during chronic HIV infection. *Immunity* (2013) 39(4):633–45. doi: 10.1016/j.immuni.2013.10.001

3. INSIGHT START Study Group, Lundgren JD, Babiker AG, Gordin F, Emery S, Grund B, et al. Initiation of antiretroviral therapy in early asymptomatic HIV infection. *New Engl J Med* (2015) 373(9):795–807. doi: 10.1056/NEJMoa1506816

4. Hunt PW, Lee SA, Siedner MJ. Immunologic biomarkers, morbidity, and mortality in treated HIV infection. *J Infect Dis* (2016) 214 Suppl 2(Suppl 2):S44–50. doi: 10.1093/infdis/jiw275

5. Zicari S, Sessa L, Cotugno N, Ruggiero A, Morrocchi E, Concato C, et al. Immune activation, inflammation, and non-AIDS co-morbidities in HIV-infected patients under long-term ART. *Viruses* (2019) 11(3):200. doi: 10.3390/v11030200

6. Dirajlal-Fargo S, Funderburg N. HIV and cardiovascular disease: the role of inflammation. *Curr Opin HIV AIDS* (2022) 17(5):286–92. doi: 10.1097/COH.0000000000000755

7. Freeman ML, Shive CL, Nguyen TP, Younes S-A, Panigrahi S, Lederman MM. Cytokines and T-cell homeostasis in HIV infection. *J Infect Dis* (2016) 214 Suppl 2 (Suppl 2):S51–7. doi: 10.1093/infdis/jiw287

8. Sandler NG, Wand H, Roque A, Law M, Nason MC, Nixon DE, et al. Plasma levels of soluble CD14 independently predict mortality in HIV infection. *J Infect Dis* (2011) 203(6):780–90. doi: 10.1093/infdis/jiq118

9. Mensching L, Hoelzemer A. NK cells, monocytes and macrophages in HIV-1 control: impact of innate immune responses. *Front Immunol* (2022) 13:883728. doi: 10.3389/fimmu.2022.883728

10. McKibben RA, Margolick JB, Grinspoon S, Li X, Palella FJ, Kingsley LA, et al. Elevated levels of monocyte activation markers are associated with subclinical atherosclerosis in men with and those without HIV infection. *J Infect Dis* (2015) 211 (8):1219–28. doi: 10.1093/infdis/jiu594

11. Van de Wijer L, van der Heijden WA, Ter Horst R, Jaeger M, Trypsteen W, Rutsaert S, et al. The architecture of circulating immune cells is dysregulated in people living with HIV on long term antiretroviral treatment and relates with markers of the HIV-1 reservoir, cytomegalovirus, and microbial translocation. *Front Immunol* (2021) 12:661990. doi: 10.3389/fimmu.2021.661990

12. van der Heijden WA, Wan J, Van de Wijer L, Jaeger M, Netea MG, van der Ven AJ, et al. Plasmatic coagulation capacity correlates with inflammation and abacavir use during chronic HIV infection. *JAIDS J Acquired Immune Deficiency Syndromes* (2021) 87(1):711–9. doi: 10.1097/QAI.0000000000002633

13. van der Heijden WA, van de Wijer L, Jaeger M, Grintjes K, Netea MG, Urbanus RT, et al. Long-term treated HIV infection is associated with platelet mitochondrial dysfunction. *Sci Rep* (2021) 11(1):6246. doi: 10.1038/s41598-021-85775-5

14. Navas A, Van de Wijer L, Jacobs-Cleophas M, Schimmel-Naber AM, van Cranenbroek B, van der Heijden WA, et al. Comprehensive phenotyping of circulating immune cell subsets in people living with HIV. *J Immunol Methods* (2022) 507:113307. doi: 10.1016/j.jim.2022.113307

15. Gianella S, Letendre S. Cytomegalovirus and HIV: A dangerous pas de deux. *J Infect Dis* (2016) 214 Suppl 2(Suppl 2):S67–74. doi: 10.1093/infdis/jiw217

16. Hatano H. Immune activation and HIV persistence: considerations for novel therapeutic interventions. *Curr Opin HIV AIDS* (2013) 8(3):211–6. doi: 10.1097/COH.0b013e32835f9788

17. Brenchley JM, Price DA, Schacker TW, Asher TE, Silvestri G, Rao S, et al. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nat Med* (2006) 12(12):1365–71. doi: 10.1038/nm1511

18. Rajewsky N, Almouzni G, Gorski SA, Aerts S, Amit I, Bertero MG, et al. LifeTime and improving European healthcare through cell-based interceptive medicine. *Nature* (2020) 587(7834):377–86. doi: 10.1038/s41586-020-2715-9

19. Liu N, Jiang C, Cai P, Shen Z, Sun W, Xu H, et al. Single-cell analysis of COVID-19, sepsis, and HIV infection reveals hyperinflammatory and immunosuppressive signatures in monocytes. *Cell Rep* (2021) 37(1):109793. doi: 10.1016/j.celrep.2021.109793

20. Rood JE, Maartens A, Hupalowska A, Teichmann SA, Regev A. Impact of the human cell atlas on medicine. *Nat Med* (2022) 28(12):2486–96. doi: 10.1038/s41591-022-02104-7

21. Bernardes JP, Mishra N, Tran F, Bahmer T, Best L, Blase JI, et al. Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. *Immunity* (2020) 53(6):1296–1314.e9. doi: 10.1016/j.immuni.2020.11.017

22. Schulte-Schrepping J, Reusch N, Paclik D, Baßler K, Schlickeiser S, Zhang B, et al. Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell* (2020) 182(6):1419–1440.e23. doi: 10.1016/j.cell.2020.08.001

23. Krämer B, Knoll R, Bonaguro L, ToVinh M, Raabe J, Astaburuaga-García R, et al. Early IFN-α signatures and persistent dysfunction are distinguishing features of NK cells in severe COVID-19. *Immunity* (2021) 54(11):2650–2669.e14. doi: 10.1016/j.immuni.2021.09.002

24. Baßler K, Fujii W, Kapellos TS, Dudkin E, Reusch N, Horne A, et al. Alveolar macrophages in early stage COPD show functional deviations with properties of impaired immune activation. *Front Immunol* (2022) 13:917232. doi: 10.3389/fimmu.2022.917232

25. Bonaguro L, Schulte-Schrepping J, Ulas T, Aschenbrenner AC, Beyer M, Schultze JL. A guide to systems-level immunomics. *Nat Immunol* (2022) 23 (10):1412–23. doi: 10.1038/s41590-022-01309-9

26. Carraro C, Bonaguro L, Schulte-Schrepping J, Horne A, Oestreich M, Warnat-Herresthal S, et al. Decoding mechanism of action and sensitivity to drug candidates from integrated transcriptome and chromatin state. *eLife* (2022) 11. doi: 10.7554/eLife.78012

27. Wang S, Song R, Wang Z, Jing Z, Wang S, Ma J. S100A8/A9 in inflammation. *Front Immunol* (2018) 9:1298. doi: 10.3389/fimmu.2018.01298

28. Knoll R, Schultze JL, Schulte-Schrepping J. Monocytes and macrophages in COVID-19. *Front Immunol* (2021) 12:720109. doi: 10.3389/fimmu.2021.720109

29. Chaudhuri A, Yang B, Gendelman HE, Persidsky Y, Kanmogne GD. STAT1 signaling modulates HIV-1-induced inflammatory responses and leukocyte transmigration across the blood-brain barrier. *Blood* (2008) 111(4):2062–72. doi: 10.1182/blood-2007-05-091207

30. Appelberg KS, Wallet MA, Taylor JP, Cash MN, Sleasman JW, Goodenow MM. HIV-1 infection primes macrophages through STAT signaling to promote enhanced inflammation and viral replication. *AIDS Res Hum Retroviruses* (2017) 33(7):690–702. doi: 10.1089/AID.2016.0273

31. Lien E, Aukrust P, Sundan A, Müller F, Frøland SS, Espevik T. Elevated levels of serum-soluble CD14 in human immunodeficiency virus type 1 (HIV-1) infection: correlation to disease progression and clinical events. *Blood* (1998) 92(6):2084–92. doi: 10.1182/blood.V92.6.2084

32. Shive CL, Jiang W, Anthony DD, Lederman MM. Soluble CD14 is a nonspecific marker of monocyte activation. *AIDS* (2015) 29(10):1263–5. doi: 10.1097/QAD.0000000000000735

33. Kazer SW, Aicher TP, Muema DM, Carroll SL, Ordovas-Montanes J, Miao VN, et al. Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. *Nat Med* (2020) 26(4):511–8. doi: 10.1038/s41591-020-0799-2

34. Aschenbrenner AC, Mouktaroudi M, Krämer B, Oestreich M, Antonakos N, Nuesch-Germano M, et al. Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients. *Genome Med* (2021) 13(1):7. doi: 10.1186/s13073-020-00823-5

35. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* (2017) 171(6):1437–1452.e17. doi: 10.1016/j.cell.2017.10.049

36. Pilarczyk M, Fazel-Najafabadi M, Kouril M, Shamsaei B, Vasiliauskas J, Niu W, et al. Connecting omics signatures and revealing biological mechanisms with iLINCS. *Nat Commun* (2022/) 13:4678/. doi: 10.1038/s41467-022-32205-3

37. Dochi T, Akita A, Kishimoto N, Takamune N, Misumi S. Trametinib suppresses HIV-1 replication by interfering with the disassembly of human immunodeficiency virus type 1 capsid core. *Biochem Biophys Res Commun* (2018) 495(2):1846–50. doi: 10.1016/j.bbrc.2017.11.177

38. Guo J, Xu X, Rasheed TK, Yoder A, Yu D, Liang H, et al. Genistein interferes with SDF-1- and HIV-mediated actin dynamics and inhibits HIV infection of resting CD4 T cells. *Retrovirology* (2013) 10:62. doi: 10.1186/1742-4690-10-62

39. Fields JA, Metcalf J, Overk C, Adame A, Spencer B, Wrasidlo W, et al. The anticancer drug sunitinib promotes autophagyand protects from neurotoxicity in an HIV-1 Tat model of neurodegeneration. *J Neurovirology* (2017) 23(2):290–303. doi: 10.1007/s13365-016-0502-z

40. Best C, Struthers H, Laciny E, Royal M, Reeds DN, Yarasheski KE. Sitagliptin reduces inflammation and chronic immune cell activation in HIV+ adults with impaired glucose tolerance. *J Clin Endocrinol Metab* (2015) 100(7):2621–9. doi: 10.1210/jc.2015-1531

41. Dubé MP, Chan ES, Lake JE, Williams B, Kinslow J, Landay A, et al. A randomized, double-blinded, placebo-controlled trial of sitagliptin for reducing inflammation and immune activation in treated and suppressed human immunodeficiency virus infection. *Clin Infect Dis* (2019) 69(7):1165–72. doi: 10.1093/cid/ciy1051

42. Daly MB, Roth ME, Bonnac L, Maldonado JO, Xie J, Clouser CL, et al. Dual anti-HIV mechanism of clofarabine. *Retrovirology* (2016) 13:20. doi: 10.1186/s12977-016-0254-0

43. Soper A, Kimura I, Nagaoka S, Konno Y, Yamamoto K, Koyanagi Y, et al. Type I interferon responses by HIV-1 infection: association with disease progression and control. *Front Immunol* (2017) 8:1823. doi: 10.3389/fimmu.2017.01823

44. Soare AY, Durham ND, Gopal R, Tweel B, Hoffman KW, Brown JA, et al. P2X antagonists inhibit HIV-1 productive infection and inflammatory cytokines interleukin-10 (IL-10) and IL-1β in a human tonsil explant model. *J Virology* (2019) 93(1):e01186-18. doi: 10.1128/JVI.01186-18

45. Calza L, Trapani F, Bartoletti M, Manfredi R, Colangeli V, Borderi M, et al. Statin therapy decreases serum levels of high-sensitivity C-reactive protein and tumor necrosis factor-α in HIV-infected patients treated with ritonavir-boosted protease inhibitors. *HIV Clin Trials* (2012) 13(3):153–61. doi: 10.1310/hct1303-153

46. Eckard AR, Jiang Y, Debanne SM, Funderburg NT, McComsey GA. Effect of 24 weeks of statin therapy on systemic and vascular inflammation in HIV-infected subjects receiving antiretroviral therapy. *J Infect Diseases* (2014) 209(8):1156–64. doi: 10.1093/infdis/jiu012

47. Lo J, Lu MT, Ihenachor EJ, Wei J, Looby SE, Fitch KV, et al. Effects of statin therapy on coronary artery plaque volume and high-risk plaque morphology in HIV-infected patients with subclinical atherosclerosis: a randomised, double-blind, placebo-controlled trial. *Lancet HIV* (2015) 2(2):e52–63. doi: 10.1016/S2352-3018(14)00032-0

48. Eckard AR, Meissner EG, Singh I, McComsey GA. Cardiovascular disease, statins, and HIV. *J Infect Dis* (2016) 214 Suppl 2(Suppl 2):S83–92. doi: 10.1093/infdis/jiw288

49. Fitch KV, Fulda ES, Grinspoon SK. Statins for primary cardiovascular disease prevention among people with HIV: emergent directions. *Curr Opin HIV AIDS* (2022) 17(5):293–300. doi: 10.1097/COH.0000000000000752

50. Faivre S, Demetri G, Sargent W, Raymond E. Molecular basis for sunitinib efficacy and future clinical development. *Nat Rev Drug Discov* (2007) 6(9):734–45. doi: 10.1038/nrd2380

51. Lian T, Li C, Wang H. Trametinib in the treatment of multiple Malignancies harboring MEK1 mutations. *Cancer Treat Rev* (2019) 81:101907. doi: 10.1016/j.ctrv.2019.101907

52. Domínguez-Andrés J, Arts RJW, Bekkering S, Bahrar H, Blok BA, de Bree LCJ, et al. In vitro induction of trained immunity in adherent human monocytes. *STAR Protoc* (2021) 2(1):100365. doi: 10.1016/j.xpro.2021.100365

53. Vadaq N, Zhang Y, Vos WA, Groenendijk A, Blaauw M, Eekeren L, et al. High-throughput proteomic analysis reveals systemic dysregulation in virally suppressed people living with HIV. *J Clin Invest Insight* (2023) 8(11):e166166. doi: 10.1172/jci.insight.166166

54. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* (2013) 29(1):15–21. doi: 10.1093/bioinformatics/bts635

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

56. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* (2016) 34(5):525–7. doi: 10.1038/nbt.3519

57. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* (2012) 28(6):882–3. doi: 10.1093/bioinformatics/bts034

58. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007

59. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* (2017) 14(11):1083–6. doi: 10.1038/nmeth.4463

60. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. *Nat Genet* (2000) 25(1):25–9. doi: 10.1038/75556

61. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* (2019) 47(D1):D330–8. doi: 10.1093/nar/gky1055

62. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Systems* (2015) 1(6):417–25. doi: 10.1016/j.cels.2015.12.004

63. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: J Integr Biol* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118

64. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7

65. McInnes L, Healy J, Saul N, Großberger L. UMAP: uniform manifold approximation and projection for dimension reduction. *J Open Source Softw* (2018) 3(29):861. doi: 10.21105/joss.00861

66. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir ED, Tadmor MD, et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* (2015) 162(1):184–97. doi: 10.1016/j.cell.2015.05.047

67. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* (2012) 9(4):357–9. doi: 10.1038/nmeth.1923

68. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *GigaScience* (2021) 10(2). doi: 10.1093/gigascience/giab008

69. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of chIP-seq (MACS). *Genome Biol* (2008) 9(9):R137. doi: 10.1186/gb-2008-9-9-r137

70. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* (2015) 31(14):2382–3. doi: 10.1093/bioinformatics/btv145

71. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* (2018) 36(5):411–20. doi: 10.1038/nbt.4096

72. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single-cell data. *Cell* (2019) 177(7):1888–1902.e21. doi: 10.1016/j.cell.2019.05.031

73. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* (2019) 16(12):1289–96. doi: 10.1038/s41592-019-0619-0

74. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. (2016). Available at: https://ggplot2.tidyverse.org.

75. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* (2016) 32(18):2847–9. doi: 10.1093/bioinformatics/btw313

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 28 November 2023
DOI 10.3389/fimmu.2023.1271800

OPEN ACCESS

EDITED BY
Ying Zhu,
Genentech Inc., United States

REVIEWED BY
Suoqin Jin,
Wuhan University, China
Naveen Ramalingam,
Fluidigm Corporation, United States
Yang Xiao,
Columbia University, United States

*CORRESPONDENCE
Manoj K. Bhasin
manoj.bhasin@emory.edu

RECEIVED 02 August 2023
ACCEPTED 06 November 2023
PUBLISHED 28 November 2023

CITATION
Satpathy S, Thomas BE, Pilcher WJ, Bakhtiari M, Ponder LA, Pacholczyk R, Prahalad S, Bhasin SS, Munn DH and Bhasin MK (2023) The Simple prEservatioN of Single cElls method for cryopreservation enables the generation of single-cell immune profiles from whole blood. *Front. Immunol.* 14:1271800. doi: 10.3389/fimmu.2023.1271800

The Simple prEservatioN of Single cElls method for cryopreservation enables the generation of single-cell immune profiles from whole blood

Sarthak Satpathy[1,2], Beena E. Thomas[1,3], William J. Pilcher[1,2], Mojtaba Bakhtiari[1,3], Lori A. Ponder[4], Rafal Pacholczyk[5,6], Sampath Prahalad[3,4,7], Swati S. Bhasin[1,2,3], David H. Munn[5,8] and Manoj K. Bhasin[1,2,3]*

[1]Aflac Cancer and Blood Disorders Center, Children Healthcare of Atlanta, Atlanta, GA, United States, [2]Department of Biomedical Informatics, Emory University, Atlanta, GA, United States, [3]Department of Pediatrics, Emory University, Atlanta, GA, United States, [4]Division of Rheumatology, Children's Healthcare of Atlanta, Atlanta, GA, United States, [5]Georgia Cancer Center, Augusta University, Augusta, GA, United States, [6]Department of Biochemistry and Molecular Biology, Augusta University, Augusta, GA, United States, [7]Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, United States, [8]Department of Pediatrics, Augusta University, Augusta, GA, United States

Introduction: Current multistep methods utilized for preparing and cryopreserving single-cell suspensions from blood samples for single-cell RNA sequencing (scRNA-seq) are time-consuming, requiring trained personnel and special equipment, so limiting their clinical adoption. We developed a method, Simple prEservatioN of Single cElls (SENSE), for single-step cryopreservation of whole blood (WB) along with granulocyte depletion during single-cell assay, to generate high quality single-cell profiles (SCP).

Methods: WB was cryopreserved using the SENSE method and peripheral blood mononuclear cells (PBMCs) were isolated and cryopreserved using the traditional density-gradient method (PBMC method) from the same blood sample (n=6). The SCPs obtained from both methods were processed using a similar pipeline and quality control parameters. Further, entropy calculation, differential gene expression, and cellular communication analysis were performed to compare cell types and subtypes from both methods.

Results: Highly viable (86.3 $\pm$ 1.51%) single-cell suspensions (22,353 cells) were obtained from the six WB samples cryopreserved using the SENSE method. In-depth characterization of the scRNA-seq datasets from the samples processed with the SENSE method yielded high-quality profiles of lymphoid and myeloid cell types which were in concordance with the profiles obtained with classical multistep PBMC method processed samples. Additionally, the SENSE method cryopreserved samples exhibited significantly higher T-cell enrichment, enabling deeper characterization of T-cell subtypes. Overall, the SENSE and PBMC methods processed samples exhibited transcriptional, and cellular

communication network level similarities across cell types with no batch effect except in myeloid lineage cells.

Discussion: Comparative analysis of scRNA-seq datasets obtained with the two cryopreservation methods i.e., SENSE and PBMC methods, yielded similar cellular and molecular profiles, confirming the suitability of the former method's incorporation in clinics/labs for cryopreserving and obtaining high-quality single-cells for conducting critical translational research.

KEYWORDS

whole blood, cryopreservation, single cell profiling, density gradient, scRNA seq

1 Introduction

Recent advances in single-cell microfluidic technologies have resulted in a ubiquitous implementation of single-cell approaches to understand disease mechanisms and developmental biology (1–3). Single-cell assays provide high-resolution measurement of cell types/subtypes (4) and their molecular states associated with health/disease conditions (5). Single-cell assays have immense potential in the discovery of cell-specific biomarkers (6) and for gaining unprecedented insights into composite cell-to-cell interactions that drive therapeutic responses (7) for expanding disease diagnosis and therapeutic options (8). We are utilizing single cell assays for the development of single-cell atlases for multiple myeloma (MM) (9, 10), pediatric cancers (11) as well as chronic wounds (12–14), to identify next-generation prognostic biomarkers with high sensitivity and specificity. Recently, a comparative analysis by our group, of rapid and non-progressing MM patient samples using single-cell profiling (SCP), revealed a significant contribution of exhausted T-cells in the rapid progression of MM (9). The implementation of SCP in another study on diabetic foot ulcers (DFUs) resulted in the identification of a unique fibroblast population associated with the healing of chronic DFUs in diabetic patients (14). A major issue with the single-cell approach is that samples need to be immediately subjected to downstream processing for live cell capture or frozen viably, both of which require precious time and bench-work, often not feasible in a clinical setting. Therefore, developing and optimizing methodologies that enable stable cryopreservation of clinic/hospital-collected samples with minimal intervention is crucial for implementation of SCP assays as routine.

Sample preparation for bulk sequencing can be performed on samples collected in the clinic in tubes with RNA/DNA stabilizers (15) without the need for immediate pre-processing (16). Although this approach is easy and practical in a clinical setting, a major drawback is that bulk approaches only reveal the average behavior of all the different cell populations in a sample (17). On the contrary, single-cell assays measure individual cell profiles (18) and their transcriptional states in the complex tissue architecture (2). However, the inherent need for viable cells (19) for performing single-cell assays limits the adoption of single-cell assays in clinics (bench to bedside) as well as the collection of clinical samples for single-cell research (bedside to bench). The traditional method of isolating PBMCs involves multiple centrifugation steps, the addition of special density gradient reagents like Ficoll-Paque to facilitate the isolation of mononuclear cells (20) for downstream single-cell assays and significant time commitment (Figure 1A). These preprocessing steps to isolate PBMCs may inevitably delay sample cryopreservation, which can potentially introduce technical bias and artifacts during SCP (21). To overcome the limitations associated with traditional sample preservation for single-cell assays, we have developed and implemented the **S**imple pr**E**servatio**N** of **S**ingle c**E**lls **(SENSE)** method for one-step cryopreservation of whole blood (WB) by the direct addition of freezing solution. The SENSE method also incorporates a granulocyte removal step during single-cell assay steps, resulting in optimal capture of immune repertoire from WB samples. In this study for the first time, we have performed a deep characterization of the SENSE method-generated transcriptome profiles and compared it with the transcriptome profiles of the PBMCs isolated by the standard Ficoll-Paque gradient method. Comparative analysis was performed on the patient blood samples collected in clinic setting to pave the way for the clinical implementation of the SENSE method. Development and implementation of simplified cryopreservation of WB samples using methods like one-step SENSE method would result in a significant increase in the adoption of SCP in clinics and single/multi-center therapeutic trials and enable robust identification of next-generation diagnostics, prognostics, and therapeutic biomarkers.

Abbreviations: FC, fold change; SCP, Single Cell Profiling; SENSE, Simple prEservatioN of Single cElls; WB, whole blood; JIA, Juvenile Idiopathic Arthritis; PBMC, Peripheral Blood Mononuclear Cells; UMAP, Uniform Manifold Approximation and Projection.

2 Results

2.1 Whole blood cryopreservation by the SENSE method generated high-quality cells

We tested the feasibility and performance of the SENSE method for SCP on whole blood samples, collected in a clinic at Children's

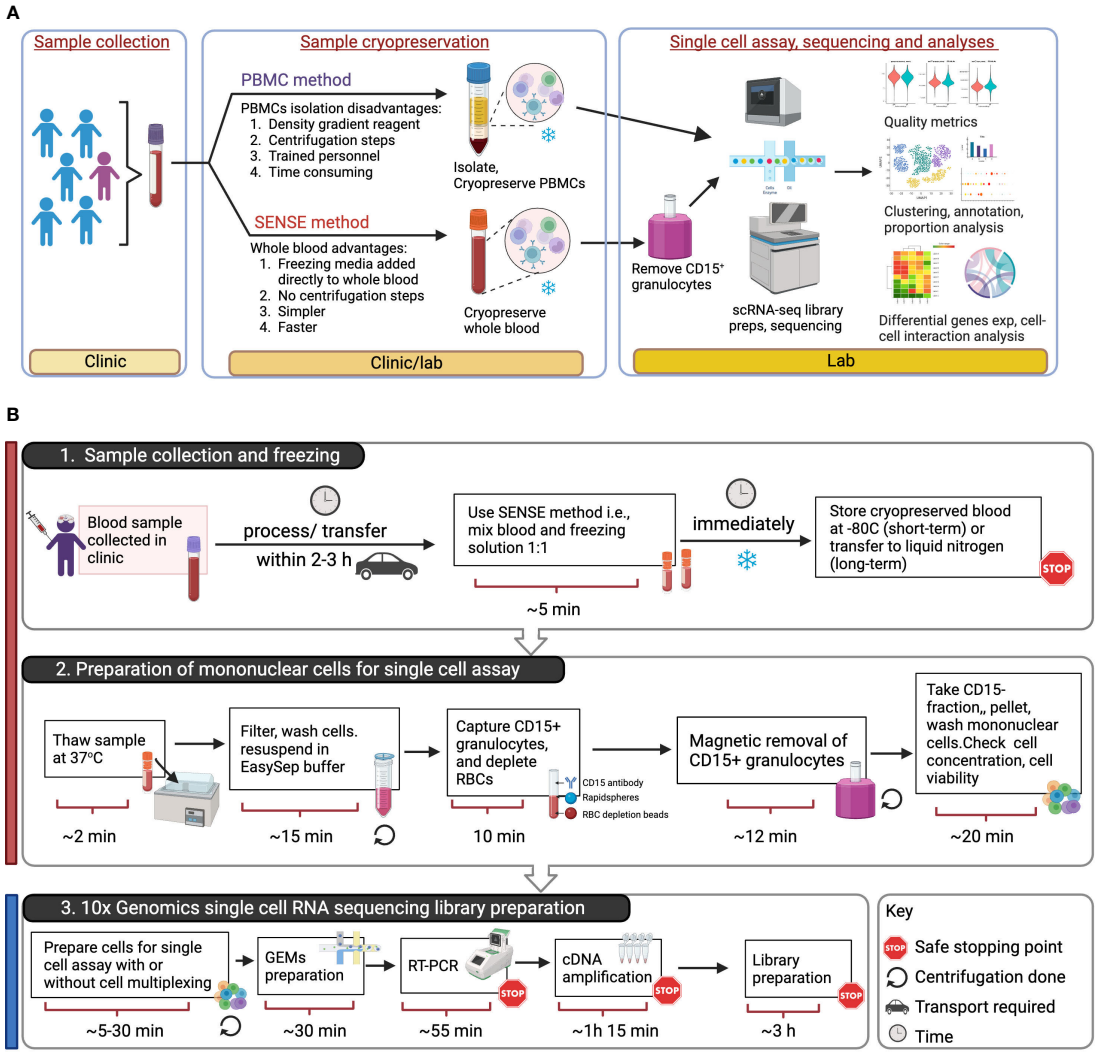


FIGURE 1
Overview of SENSE (Simple PrEservatioN of Single cElls) method for cryopreservation and single-cell immune profiles from whole blood. **(A)** Assay Overview: Blood samples were collected in EDTA tubes which were then split into two aliquots. One aliquot was processed using the traditional Ficoll-Paque density gradient method to isolate PBMCs, which were then cryopreserved. The other aliquot of blood was viably cryopreserved using the SENSE method, i.e., a cryoprotectant solution was added and the sample frozen. Post-thawing, the WB cells from the SENSE method were subjected to granulocyte depletion, and the $CD15^-$ fraction was collected, washed, and processed for single-cell profiling using the 10x Genomics method. The resulting single-cell data from the SENSE and PBMC methods were then compared to identify any differences in cell quality metrics and molecular profiles. **(B)** Schematic for single cell assays using sample stored and processed using the SENSE method involving: 1. simple cryopreservation of whole blood samples stably stored for short or long term in -80 °C or liquid nitrogen respectively, and 2. preparation of sample for single cell assay by removing granulocytes and RBCs. 3. Generation of single cell RNA sequencing libraries using appropriate 10x Genomics kits (5' or 3'with/without cell multiplexing). The color bars on the left-hand side serve as visual indicators, with the red bar denoting steps specific to SENSE method, while the blue bar represents steps common single cell profiling steps for both methods. The figure was prepared using BioRender.com.

Healthcare of Atlanta from Juvenile Idiopathic Arthritis (JIA) (n=5) and pediatric lupus (n=1) patients (Table 1). The blood samples (3.0 ml - 4.5 ml) were split into two equal aliquots and processed in parallel using the SENSE and PBMC methods. The Ficoll-Paque density gradient method was used to isolate PBMCs from one-half of the sample which were then frozen in freezing media (Fisher Scientific). These viably frozen PBMCs were thawed and profiled directly using 10x Genomics Next GEM single cell 3'v3.1 kits (as described in the methods section) (Figure 1A). The remaining half of the blood sample was processed using the SENSE method that involved freezing WB directly by adding freezing solution (80% FBS, 20% DMSO) at a ratio of 1:1 to obtain a final concentration of 40% FBS, 10% DMSO in the cryopreserved WB samples. In the SENSE method, frozen WB was thawed, and the mononuclear cells were collected and resuspended after the removal of $CD15^+$ and red blood cells as described in the detailed protocol (Supplementary Document 1; Figure 1B). The cells were then used for generating single-cell RNA sequencing (scRNA-seq) libraries. The data obtained from both methods were extensively studied by comparing various qualitative and quantitative parameters (Figure 1A). Simple single-step cryopreservation of WB made possible with incorporation of the SENSE method will promote the clinical implementation of SCP assays and expand single-cell research and discoveries (Supplementary Figure 1).

TABLE 1 Patient characteristics table.

Patient no.	Diagnosis age (years)	Sample collection age (years)	Sex	Race	Ethnicity	Diagnosis
1	6	7	Male	White	Non-hispanic	Oligoarticular JIA
2	2	16	Female	White	Non-hispanic	Polyarticular RF-
3	8	13	Female	White	Non-hispanic	Unifferntiated JIA
4	5	11	Male	AA/black	Non-hispanic	Systemic JIA
5	1	8	Female	Asian	Non-hispanic	Polyarticular RF+
6	15	16	Female	AA/black	Non-hispanic	Lupus

JIA, Juvenile idiopathic arthritis; RF, Rheumatoid factor. Patient 1-5: JIA, Patient 6: Pediatric lupus.

Comparative analysis of cell viability upon thawing of cryopreserved WB followed by granulocyte depletion (SENSE method) and thawing of cryopreserved PBMCs (density gradient method) revealed that the latter method yielded slightly higher viability (91 ± 1.64%) as compared to the former method (86.3 ± 1.51%), however, this difference was not significant (P=.10) (Figure 2A). A total of 20,024 and 23,502 cells were profiled from the PBMCs isolated using density gradient method and WB frozen using SENSE method respectively, hereby referred to as 'PBMC' and 'SENSE' for simplicity. The low-quality cell identification based on unique genes (<200), UMI count (<600), and mitochondrial transcripts (>20%) identified 793 and 1,149 low-quality cells with PBMC and SENSE methods respectively, that were filtered out from the subsequent analysis. This resulted in 19,231 and 22,353 high-quality cells from the PBMC and SENSE methods respectively (Figure 2B). The SENSE method was found to capture median gene counts and unique molecular identifiers (UMIs) comparable to the PBMC method, with a similar median representation of mitochondrial genes (Figure 2C). To check whether the SENSE method affected the integrity of the cells, the representation of genes in the membrane, extracellular, and ribosomal ontology categories were assessed. Cellular damage results in increasing the representation of the membrane genes and lowering the representation of extracellular genes (22). Comparative analysis showed profiling of similar proportions of cytoplasmic, membrane, extracellular, and ribosomal ontology categories with SENSE and PBMC methods, demonstrating that the former method is as robust as the latter method in obtaining high-quality cells from WB with no introduction of cellular damage artifacts (Figure 2D). SCP can be utilized to evaluate the cell cycle phases (i.e., G1, G2/M, and S) which significantly impact cellular gene expression and are vital in classifying cellular sub-populations in the single-cell assays. The comparative analysis revealed a broadly similar distribution of cells in various cell cycle phases between cells from all samples from the SENSE and PBMC methods (Figure 2E). Additionally, we also assessed the impact of sample processing by each method on the doublet rates as they are key confounders in the single cell data (23). The WB samples cryopreserved using the SENSE method had a lower percentage of doublets (2.41%, 539 cells) as compared to the PBMC method (4.76%, 916 cells) (Figure 2F), demonstrating that single cells cryopreserved using the SENSE method generated high-quality SCPs of clinical WB samples. Clusters with a high percentage of doublets were manually reviewed using canonical marker expression and excluded from downstream analysis.

Altogether, these single-cell quality assessment analyses demonstrate that the SENSE method is a reliable and effective method for WB single-cell profiling by preserving high-quality cells that yield comparable results to the traditional density gradient PBMCs isolation method.

2.2 Cellular profile and enrichment between SENSE and PBMCs methods

The high-quality cells obtained after filtration and normalization steps were clustered based on the gene expression profiles using Seurat (24). The initial 21 clusters obtained from the integrated scRNA-seq data of samples processed using PBMC and SENSE methods, were annotated to obtain 11 major cell types from various lineages using canonical marker genes: B-Cells (*MS4A1*$^+$, *CD79A*$^+$), Memory B-cells (*CD19*$^+$, *IGLC2*$^+$), NK cells (*NKG7*$^+$, *KLRD1*$^+$, *CD3D*$^-$), Myeloid cells (*CD14*$^+$, *MNDA*$^+$, *FCGR3A*$^+$, *FCN*$^+$), CD4+ Naïve T-cells (*CD3D*$^+$, *CD4*$^+$, *CCR7*$^+$, *LEF1*$^+$), CD4$^+$ Memory T-cells (*CD3D*$^+$, *CD4*$^+$, *TRAAD*$^+$, *TNFRSF4*$^+$), IFN T-cells (*CD3D*$^+$, *ISG15*$^+$, *STAT1*$^+$, *IFI6*$^+$), CD8$^+$ Effector T-cells (*GZMA*$^+$, *GZMB*$^+$, *CD8A*$^+$), CD8$^+$ Naive T-cells (*CD8A*$^+$, *CCR7*$^+$, *LEF1*$^+$, *TCF7*$^+$), CD8$^+$ Memory T-cells (*CCL5*$^+$, *GZMB*$^+$, *CD8A*$^+$), and platelets (*SNCA*$^+$) (Figures 3A, B). Using the doublet detection algorithm of the DoubletFinder package (25), we identified two outlier clusters exhibiting doublet proportions greatly exceeding other clusters: Db 1 (95% doublets) and Db 2 (44.1% doublets). The remaining clusters demonstrated notably lower doublet percentages (averaging at 1.7 ± 0.8%). We reviewed the canonical markers expression in these doublet-enriched clusters to explore if they express markers of cell types from different lineages and correctly flagged doublets. Cluster Db1 highly expressed both pDC and T-cell related markers (*CD4, JCHAIN, MZB1, IRF8, CLEC4C*), whereas cluster Db2 highly expressed both plasma cell and T cell markers (*CD8A, JCHAIN, MZB1, CD38*). We also observed a small number of cells in both PBMC (n=120 cells) and SENSE (n=182 cells) methods that were enriched with mitochondrial genes (Mt Enriched) (Figures 3A, B). On average, T-cells were the largest cluster among all patients, followed by myeloid cells, and B-cells (Figure 3A). All the identified cell types (except platelets) were detected in samples processed using either SENSE or PBMC method (split UMAP in Figure 3C). Regardless of the processing method used, cells of the same type consistently clustered together,

10.3389/fimmu.2023.1271800

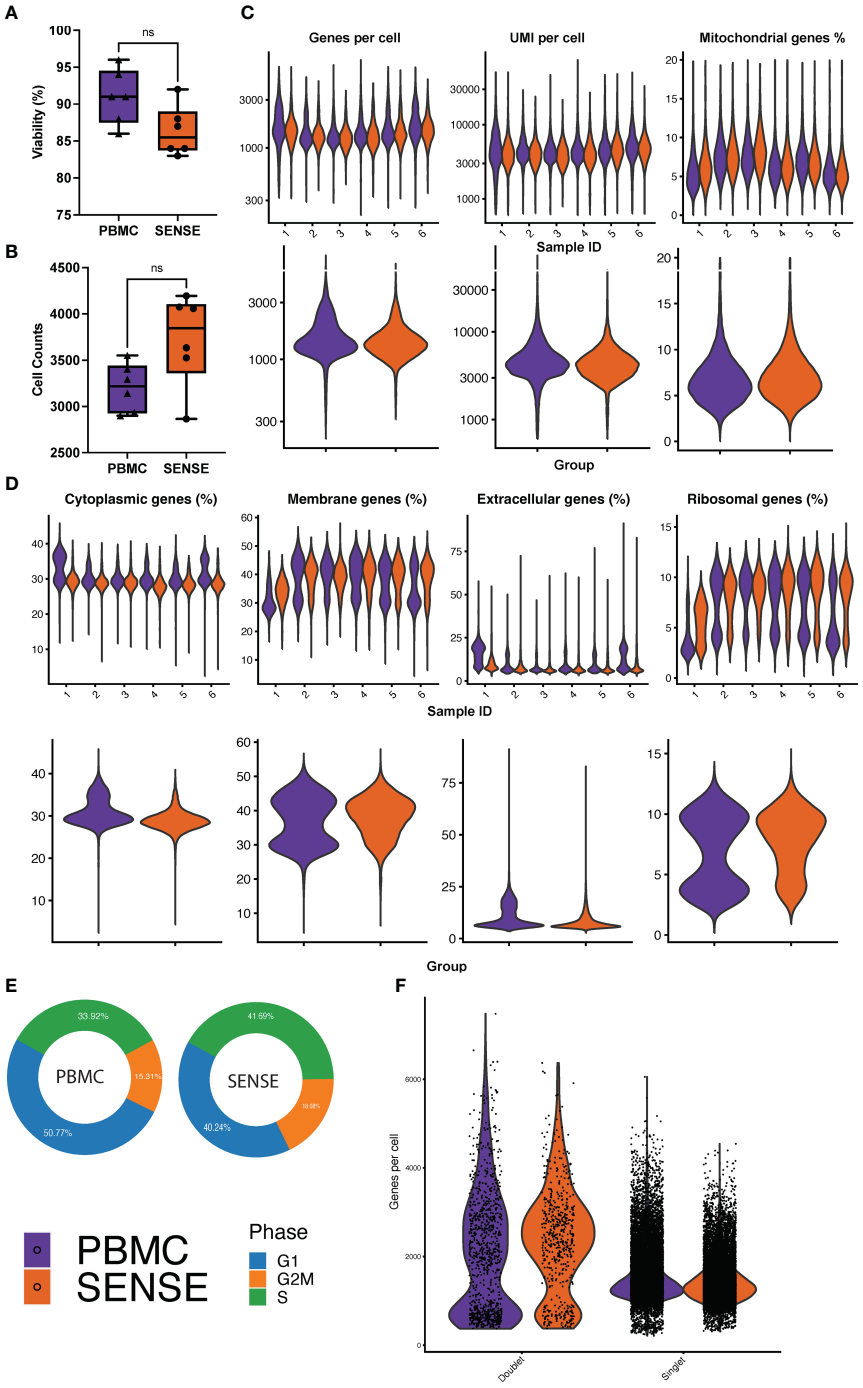


FIGURE 2
Comparative analysis of cell quality of SENSE and PBMCs methods. **(A)** Cell viability % boxplots, and **(B)** Single-cell counts boxplots with each dot representing an individual patient. The significance of the difference between the methods was tested using the paired Student's t-test. NS indicates non-significant differences with $P > .05$. **(C)** Count of Genes (log-scale), UMIs (log-scale), and proportion of mitochondrial genes per cell. The violin plots in the top panel show patient-wise information for the count of genes, UMIs, and proportion of mitochondrial genes per cell, while the violin plots in the bottom panel show the group-wise comparison of SENSE and PBMC methods. **(D)** Proportion of patient and group-wise genes in cytoplasmic, membrane, extracellular, and ribosomal gene ontology categories. **(E)** Proportion of cells from SENSE and PBMC methods in the G1, G2M, and S phases, and **(F)** Proportion of the singlets and doublets cells in the SENSE and PBMCs protocols.

highlighting their high transcriptome similarity. The split UMAP visualization and bar plots depicted lower enrichment of cells from the myeloid lineage (PBMC: n=6,085, SENSE: n=1,903) and higher enrichment of T cells (PBMC: n=8,558, SENSE: n=15,373) in the SENSE method as compared to the PBMC method (Figures 3C, D). Most of the major clusters had contributions of cells from each patient, implying a similar single-cell landscape across all patients (Figure 3D). Some clusters, such as IFNγ responsive T-cells (IFN T-cells), showed a disproportionately high contribution from a single patient (patient 1) with both processing methods (Figure 3D),

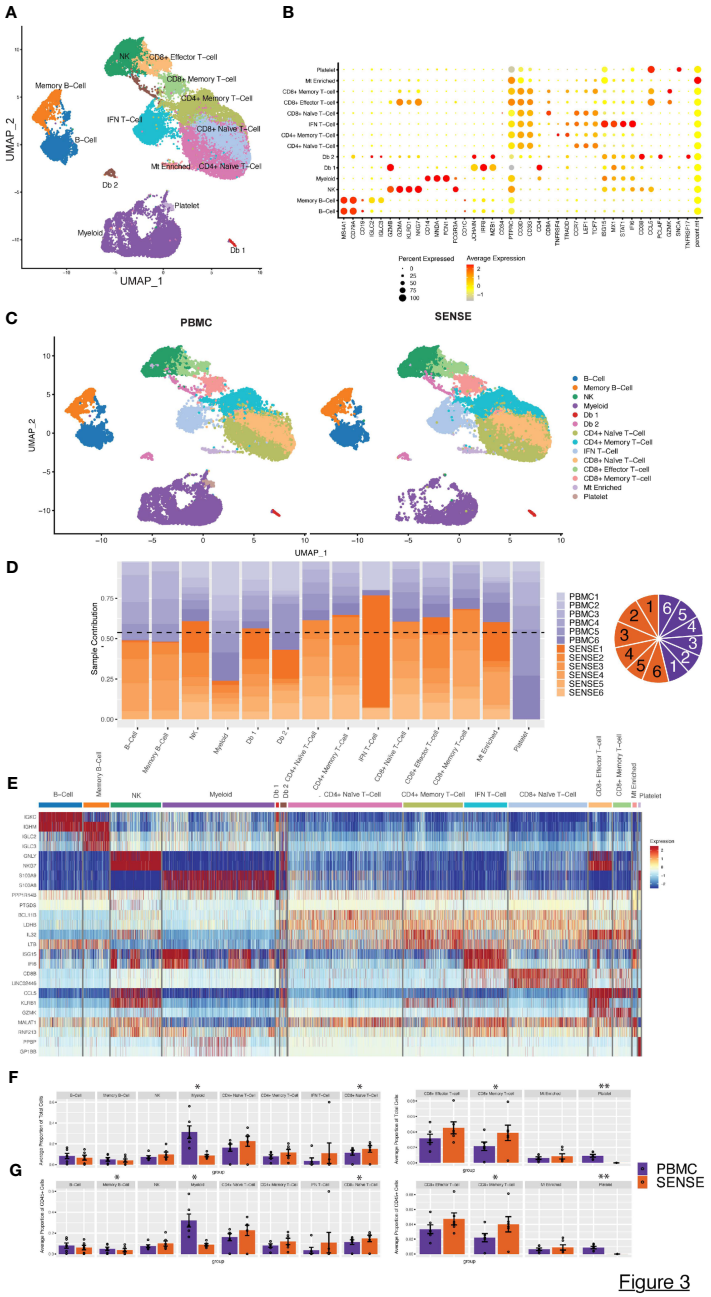


Figure 3

FIGURE 3
Characterization of blood single cell profiles obtained with SENSE and PBMCs methods. Blood samples for the analysis were obtained from JIA and pediatric Lupus patients from the rheumatology clinic in CHOA. **(A)** Uniform Manifold Approximation and Projection (UMAP) embedding of scRNA-seq data from both methods across all patients consisting of >41,000 high-quality single-cells distributed into 11 cell types. Canonical cell types are based on the expression of marker genes that include: B-Cells (*MS4A1*$^+$, *CD79A*$^+$), Memory B-cells (*CD19*$^+$, *IGLC2*$^+$), NK Cells (*NKG7*$^+$, *KLRD1*$^+$, *CD3D*$^-$), Myeloid cells (CD14$^+$, MNDA$^+$, FCGR3A$^+$, FCN$^+$), CD4+ Naïve T-cells (*CD3D*$^+$, *CD4*$^+$, *CCR7*$^+$, *LEF1*$^+$), CD4$^+$ Memory T-cells (*CD3D*$^+$, *CD4*$^+$, *TRAAD*$^+$, *TNFRSF4*$^+$), IFN T cells (*CD3D*$^+$, ISG15$^+$, STAT1$^+$, IFI6$^+$), CD8$^+$ Effector T-cells (*GZMA*$^+$, *GZMB*$^+$, *CD8A*$^+$), CD8$^+$ Naive T-cells (*CD8A*$^+$, *CCR7*$^+$, *LEF1*$^+$, *TCF7*$^+$), CD8$^+$ Memory T-cells (*CCL5*$^+$, *GZMB*$^+$, *CD8A*$^+$), and platelets (SNCA$^+$). **(B)** Dot Plot depicting expression profile of markers genes used for annotating different cell type clusters. The relative expression and percent of cells expressing specific markers are shown by shades of red color and the size of the dot respectively. **(C)** PBMC and SENSE single-cell method-based split UMAP showing the distribution of cell types. There are slightly elevated differences in T-cells subtypes in the SENSE group, while PBMC samples showed higher levels of myeloid cells. **(D)** Stacked bar plot showing the relative patient contribution in each individual cell type cluster. The samples from PBMC and SENSE methods are shown with shades of blue and red respectively. Each cluster depicted the varying levels of contribution from individual patients. The contribution of cells from each sample is shown using a pie graph with orange and purple colors representing SENSE and PBMC profiled samples respectively. **(E)** Heatmap displaying the top two gene markers expressed by each cell type. Columns represent individual cells, grouped by cell type, while rows display individual genes. Horizontal colored bars above the heatmap indicate the different cell types. Relative gene expression is shown in pseudo color, where blue represents low expression, and red represents high expression. Top markers generally correlate with well-established canonical markers for each cell type. **(F, G)** Comparative analysis proportions of cell types in the PBMC and SENSE methods. The proportion of Total **(F)** and CD45+ **(G)** cells per sample between PBMC and SENSE methods are shown. Each bar plot depicts the mean proportions and ± standard error of the mean. Each dot represents an individual sample. The significance of the difference in the mean in the groups was tested using paired Student's t-test, with significant differences being indicated with * (P<0.05) and ** (P<0.01).

which may be a result of the clinical condition of that patient. Patient 1 is a recently diagnosed JIA patient as compared to other JIA patients in this study who have been undergoing treatment for quite some time (Table 1).

To further dissect the expression profiles of different cell types, we compared the gene expression profile of the target cluster with the other cells in the sample based on the non-parametric Wilcoxon Rank Sum test (average log-fold change (FC) ≥0.25, > 25% of cells expressing gene, and $P < .01$). This analysis allowed the identification of a gene signature for each cell type (Figure 3E). The SENSE method captured a higher number of T cells per sample as compared to the PBMCs method indicating its advantage in T-cell repertoire characterization (Figure 3F; Supplementary Figure 2). T cell subtype comparison depicted that CD8+ T cells (Naïve, Memory) are significantly elevated ($P<.05$) in the SENSE method profiled samples (Figure 3F; Supplementary Figure 3). On the other hand, the PBMC-based method depicted significant enrichment ($P<.05$) of the myeloid cells (Figure 3F). Additionally, PBMC method also depicted significant enrichment of platelets as compared to SENSE ($P<.01$) (Figure 3F). Similar observations were made while considering only CD45+ immune cells of samples, where the PBMC-based method illustrated significantly better capture for the memory B-Cells in addition to myeloid cells (Figure 3G). The disparity observed in myeloid cells and platelets may be attributed to the CD15+ granulocyte removal steps employed in the SENSE method, while the disparity in T cells might be due to density gradient step in the PBMC method.

2.3 Cell types exhibited similar transcriptome profiles for SENSE and PBMC methods

To assess the sample processing method-induced technical variations in the overall expression profiles, we studied the clustering based on cell types split on SENSE and PBMC methods. Most of the matching cell types, irrespective of the processing methods, depicted similar clustering patterns except for subtle variations in the myeloid cells compartment (Figure 4A). The hierarchical clustering based on the cell types/subtypes markers genes identified based on the Wilcoxon Rank Sum test (average log-FC ≥0.25, >25% of cells expressing gene, and

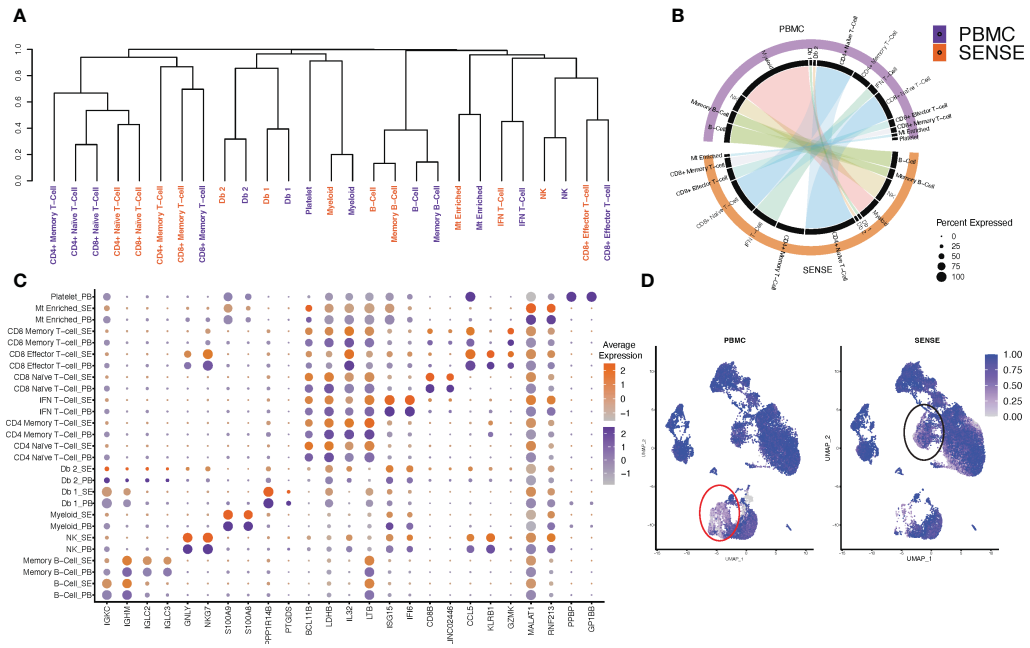


FIGURE 4
Comparison of single cell profiles of samples processed using SENSE and PBMC methods. The scRNA-seq data from blood samples processed using SENSE and PBMC methods were analyzed using a uniform bioinformatics workflow for comparative analysis. **(A)** Dendrogram showing the distances between cell types from each method based on the differentially expressed genes for each cell type computed independently. The differentially expressed genes were identified by comparing the target cell type with others based on an average log FC > 0.25 and Wilcoxon Rank Sum test $P<.01$ as well as genes expressed in > 25% of a given cell population. **(B)** A Circos plot showing the correlation between expression profiles of cell types profiled using SENSE and PBMC methods. The individual cell types between profiling methods depict significant similarities in the expression profiles. Some cell subtypes within the T cell compartment depicted lower correlations. **(C)** Comparative analysis of canonical cell type-specific markers between the two methods. Most of the cell type defining markers are concordantly expressed across corresponding cell types indicating strong similarity in the SCPs generated by SENSE and PBMC methods. The color scales on the right show the gene expression levels in samples processed using PBMC (purple) and SENSE (orange) methods. The size of the dot represents the percent of cells expressing specific markers. The Y-axis shows the cell types with SE indicating samples processed using the SENSE method and PB representing samples processed using the PBMC method. The X-axis shows the gene names. **(D)** Shannon's entropy-based batch effect estimation. The UMAP plot shows Shannon's entropy of different clusters calculated based on the distribution of SCP protocol labels (i.e., SENSE, PBMC) among the cell's 100 nearest neighbors. The analysis was performed on normalized data without any batch effect correction. Low entropy values were observed in myeloid cell clusters (marked with red lasso) and an IFN T-cell cluster (marked with black lasso), indicating poor mixing and method-based batch effect.

$P < .01$) depicted that naive and memory T-cells formed one distinct cluster, while doublets and myeloid cells formed another cluster. B-cells formed a separate cluster, and natural killer (NK) cells and effector T-cells clustered together in a different group. For most of the cellular compartments, the same cell types/subtypes depicted the highest correlation except the T cells compartment (Figure 4A). For example, the NK, $CD8^+$ effector T-cells, and myeloid cells depicted the highest correlation between transcriptome profiles from SENSE and PBMCs method (Figure 4A). In the T cells compartment, some of the cell types depicted a weaker correlation between the matching cell types from the two methods indicating some variation. This finding aligns with a higher proportion of the T cells captured using the SENSE method as compared to PBMCs based method. To further validate the consistency of cell type labeling across methods, we assessed the similarity of the differentially expressed markers for each of the 11 cell types from the two methods. To achieve this, cells from each method were subsetted, and the top differentially expressed markers for each cell type with respect to all other cells from the same method were identified based on the Wilcoxon Rank Sum Test (average log-FC ≥0.25, >25% of cells expressing gene, and $P < .01$) and visualized using Circos plots generated using ClusterMap (26) R package. The cell types from the SENSE method depicted high transcriptome correlation with matching cell types from the PBMCs method, again indicating strong concordance among the methods (Figure 4B). Next, we assessed the similarity in canonical/top markers expression for various cell types based on processing protocol. The markers for each cell type depicted similar expressions irrespective of the processing method (split dot plot in Figure 4C). The consistency of key marker genes expression establishes the transcriptome similarity of cellular profiles from the SENSE and PBMCs methods. To further assess and quantify batch effects due to processing methods, we calculated Shannon's entropy/cell to assess the degree of mixing of samples from the two methods (27). Low entropy values indicating poor mixing of a cells from different samples and methods were observed mainly in myeloid (0.526 ± 0.005) (Figure 4D, red lasso) and T cells (0.701 ± 0.004) (Figure 4D, black lasso). The rest of the cell types depicted high entropy (0.872 ± 0.001) indicating no batch effects (Figure 4D).

2.4 SENSE enables deep profiling of immune repertoire by capturing profiles of T cell subtypes

T cells are highly diverse and play a critical role in eliciting immune responses against antigens. To further investigate the different T cell subtypes captured by SENSE and PBMC methods, we performed a focused analysis after subsetting out and reclustering the T and NK cell clusters. The analysis included 27,982 cells that were annotated into 10 distinct T and NK cell subtypes (Figures 5A, B) based on the expression of marker genes that include Naïve T-cells (*CD3D*$^+$, *CCR7*$^+$, *LEF1*$^+$), Effector T-cells (*CD3D*$^+$, *GNLY*$^+$, *GZMK*$^+$), $CD4^+$ Naive T-cells ($CD4^+$, $CCR7^+$, $LEF1^+$), $CD4^+$ memory T-cells (*CD4*$^+$, *TCF7*$^+$, *TNFRSF4*$^+$), $CD4^+$ Memory IFN T-cells (*CD4*$^+$, *TNFRSF4*$^+$, *ISG15*$^+$, *MX1*$^+$), $CD8^+$ Naive T-cells (*CD8*$^+$, *CCR7*$^+$, *LEF1*$^+$), $CD8^+$ Memory/Effector T-cells (*CD8*$^+$, *TCF7*$^+$, *TNFRSF4*$^+$*GZMK*$^+$), $CD8^+$ IFN T-cells (*CD8*$^+$, *ISG15*$^+$, *MX1*$^+$), IFN NK/T cells (*CD3D*$^-$, *GNLY*$^+$, NKG7$^+$, *ISG15*$^+$, *MX1*$^+$) and NK cells (*CD3D*$^-$, *GNLY*$^+$, NKG7$^+$). In both SENSE and PBMC methods various subtypes of $CD8^+$ T-cells were the dominant T-cells (~41%), with the remaining cells consisting primarily of $CD4^+$ T-cells (~37%) along with NK cells (~8%), other T cells (~8.5%) and NK/T (specific to patient 1, ~4.5%) (Supplementary Figure 3). The Naïve T cells formed 51% of the total T cells captured in the assays. Most of the T cell clusters had contributions from all the patients except IFN-stimulated clusters that are patient-specific (Figure 5C). Overall, the SENSE-based method captured a significantly (P=.007) higher number of T cells as compared to the PBMC method (Supplementary Figure 2), but the relative proportion of T cell subtypes is similar in both methods (Figure 5D; Supplementary Figure 3). Further to explore the functional landscape of T/NK we performed a comparative analysis of cellular communication based on the expression of ligands and receptors (28). Comparison of the overall number of interactions and their strengths revealed them to be similar between SENSE and PBMC methods (Figure 5E; Supplementary Figure 4). Further communication analysis depicted similar communication patterns among cell types, with CD8+ Naïve T-cells with the highest incoming interactions and Effector T-cells with the highest outgoing interactions (Figure 5F). Further, we explored the information flow of the signaling pathways based on the sum of communication probability among cell types of SENSE and PBMC methods. We observed that most of the pathways showed a similar information flow pattern, including CLEC, MHC-I, LCK, IL16, ICAM, and ITGB2 (Figure 5G). Some pathways including MIF, and CD99 depicted different signaling between cell types from SENSE and PBMC methods. These pathways typically involve myeloid, platelets, and dendritic cells (29–31). Therefore, the differential signaling observed in these pathways may be attributed to the differences in myeloid cells and platelets captured by the two methods. These results indicate a common signaling network operates between the cells processed using either of the two methods, indicating that the SENSE method yields similar results to the PBMC method and is suitable for analysis of the T cells landscape in whole blood samples.

2.5 Myeloid lineage cell types have lower enrichment but similar profiles between single-cell preparation methods

The myeloid cell compartment is the second largest compartment of cell types observed in both sample processing methods. Overall, the SENSE method captured a significantly lower number of myeloid cells (Figure 3F; Supplementary Figure 2) which might be due to the filtering out of CD15+ myeloid/granulocytes to enhance single cell capture efficacy and generate high-quality cells for capturing. The sub-clustering and annotation on the myeloid and platelet compartments comprising 7,599 cells yielded seven distinct cell types or subtypes. Among these, six were identified as

10.3389/fimmu.2023.1271800

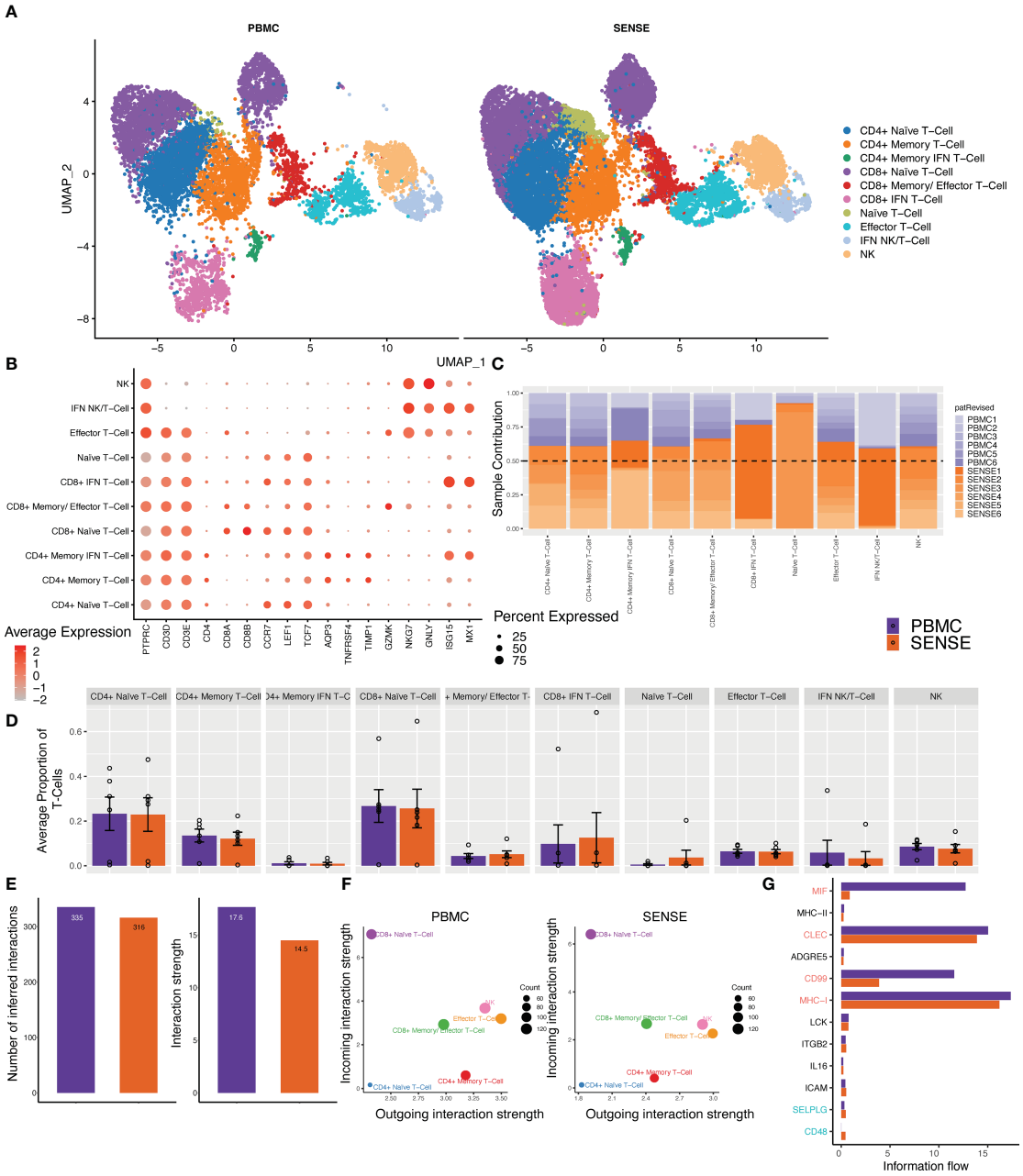


FIGURE 5
Focused analysis on the T cell clusters to estimate the impact of blood samples processing protocols. **(A)** A UMAP displaying the T-cell subclusters split based on single-cell processing protocols (i.e., SENSE, PBMC). Subclusters were manually labeled as T/NK cells (Naïve, Effector, IFN+, NK), $CD4^+$ T-cells (Naïve, Memory, Memory IFN+), $CD8^+$ T-cells (Naïve, Memory, IFN+) based on the expression of specific markers. The counts depicted slightly better capture for T cell sub-clusters in the SENSE method as compared to the PBMC method. **(B)** Dot plot demonstrating the expression profile of key markers for each T-cell subtype. The gradient of red color and size of dot represent the relative expression and percent cells expressing specific markers, respectively. **(C)** Stacked bar plot showing the relative patient contribution in each individual T- cell sub-cluster. The samples from PBMC and SENSE methods are shown with shades of blue and red respectively. Each cluster depicted the varying levels of contribution from individual patients. **(D)** Comparative analysis proportions of cell types in the PBMC and SENSE methods for T-cell subclusters. The proportion of total T-cells cells per sample between PBMC and SENSE methods for each sub-cluster is shown. **(E-G)** CellChat based analysis of cell-cell communication. **(E)** Total number of interactions and interaction strength of the inferred cell-cell communication networks for T-Cells from different methods, PBMC (purple) and SENSE (orange). **(F)** Scatter plot to compare the major sources and targets of interaction on the 2D space where the incoming and outgoing strength for each T cluster along the y-axis and x-axis, respectively. **(G)** Bar graph to compare the overall information flow of each signaling pathway between PBMC and SENSE methods.

myeloid subtypes and a single cluster as platelets. Both methods captured every cell type from the myeloid lineage as evident from the split UMAP plot (Figure 6A). These clusters look like classical and non-classical monocytes with the expression of *CD14* and *FCGR3A* genes along with the expression of other markers (Figure 6B). The analysis of patients' contribution to different clusters depicted that some of the clusters like cluster A (*LYZ*+, *CD14*+, *PABPC1*+) had contributions from multiple patients

10.3389/fimmu.2023.1271800

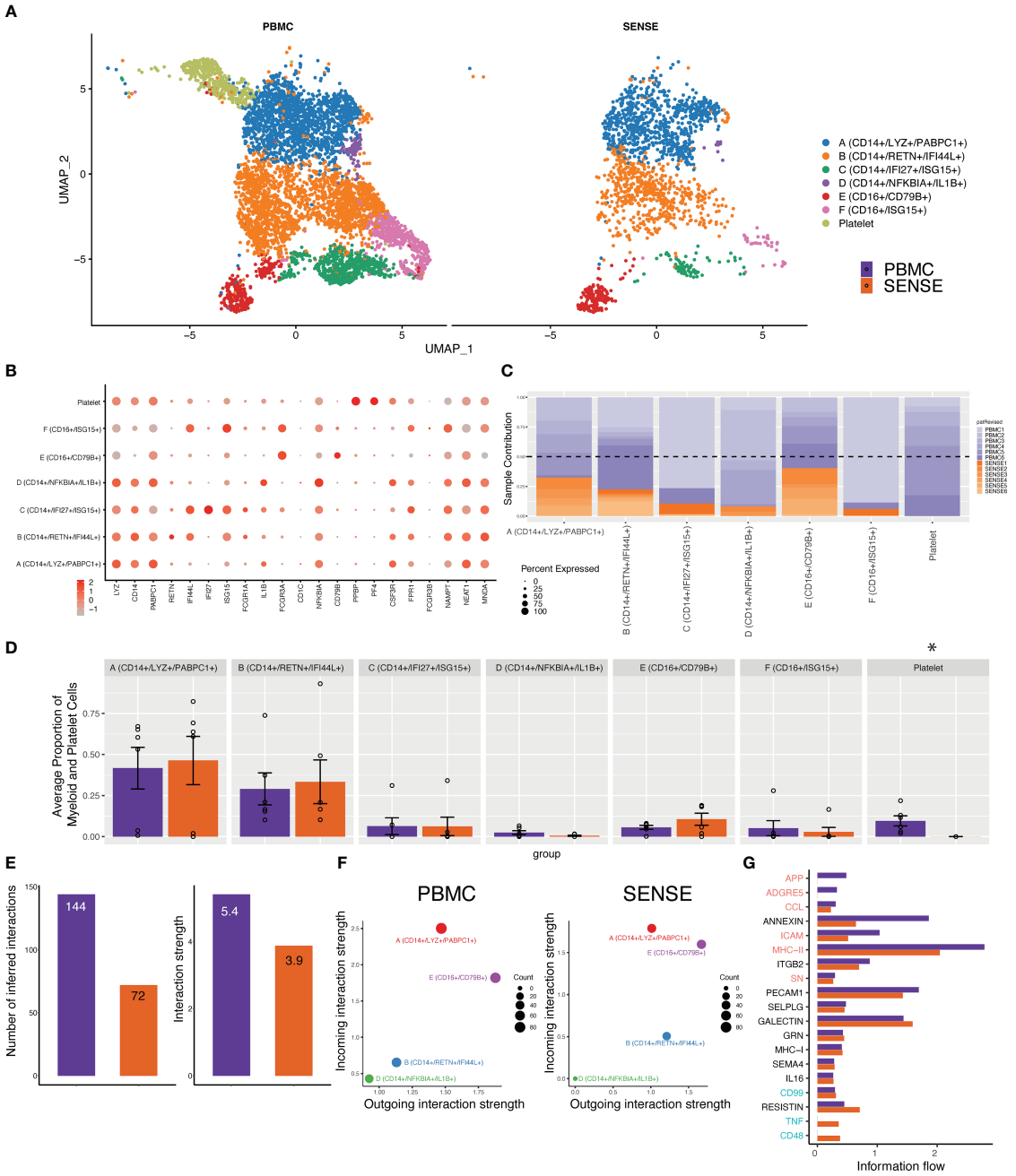


FIGURE 6
Focused analysis on the Myeloid cell clusters to estimate the impact of single-cell processing protocols. **(A)** A UMAP displaying the myeloid and platelet subclusters split based on single-cell processing protocols (i.e., SENSE, PBMC). Subclusters were manually labeled as Clusters A (*CD14+, LYZ +, PABPC1+*), B (*CD14+, RETN+, IFI44L+*), C (*CD14+, IFI27+, ISG15+*), D (*CD14+, NFKBIA+, IL1B+*), E (*CD16+, CD79B+*), F (*CD16+, ISG15+*), and Platelets (*PPBP+, PF4+*) based on the expression of top markers. The counts were lower for myeloid cell sub-clusters in the SENSE method as compared to the PBMC method. The platelets were present only in the PBMC samples. **(B)** Dot plot demonstrating the expression profile of common myeloid, neutrophils, and platelet markers. The gradient of red color and size of the dot represents the relative expression and percent cells expressing specific markers, respectively. **(C)** Stacked bar plot showing the relative patient contribution in each individual sub-cluster. The samples from PBMC and SENSE methods are shown with shades of blue and red respectively. Each cluster depicted the varying levels of contribution from individual patients. **(D)** Comparative analysis proportions of cell types in the PBMC and SENSE methods for myeloid and platelet subclusters. The proportion of total myeloid and platelet cells per sample between PBMC and SENSE methods for each sub-cluster is shown. **(E–G)**. CellChat based analysis of cell-cell communication for Myeloid clusters. **(E)** Total number of interactions and interaction strength of the inferred cell-cell communication networks for myeloid cells from different methods, PBMC (purple) and SENSE (orange). **(F)** Scatter plot to compare the major sources and targets of interaction on the 2D space where the incoming and outgoing strength for each T cluster along the y-axis and x-axis respectively, **(G)** Bar graph to compare overall information flow of each signaling pathway between PBMC and SENSE methods.

whereas others like cluster F (*CD16+, ISG15+*) were patient specific (Figure 6C). Although the SENSE method captured fewer cells in this compartment, comparative analysis of relative proportions of myeloid cell subtypes depicts concordance of single cells captured using the two methods (Figure 6D). To better understand the communication patterns among different subtypes of the myeloid lineage, we conducted cell communication analysis. The analysis was performed only on the clusters that have cells from multiple patients. Our analysis revealed lower interactions and weaker strengths of interaction between the subtypes of the myeloid lineage using the SENSE method (Figure 6E). One possible explanation for these findings is that there were fewer overall cells present during the analysis in this compartment from the SENSE method (PBMC: n=5,401, SENSE: n=1,733). Comparison of cellular communications patterns among the myeloid cells depicted significant concordance with cluster A with characteristics like classical monocytes with most incoming interactions (Figure 6F). Further analysis of key pathways associated with the flow of information among cells also depicted multiple similar pathways including PECAM1, GRN, MHC-I, SELPLG, and Galectin (Figure 6G). In summary, although there is concordance in the myeloid cell sub types captured by the two methods, the lower capture of myeloid cells in the SENSE method is leading to some cellular communication differences that is one of its limitations.

2.6 SENSE method generated transcriptome profile similar to publicly available PBMC transcriptome profile

To further evaluate the transcriptome profile of SENSE method WB generated data, we performed a comparative analysis with publicly available PBMC dataset. This PBMC dataset (32) was obtained from the 10x Genomics Inc. website and processed uniformly and integrated with our data using integration anchors-based batch correction. The comparative analysis of cellular profiles based on split UMAP depicted co-embedding of major cell types indicating similarity in transcriptome profiles (Figure 7A). In line with the publicly available 10x Genomics PBMC dataset (10x PBMC), the SENSE method also captured T cells as the most abundant cell types from the whole blood profiling. Shannon's entropy was computed per cell to assess the degree of mixing of samples from three datasets (i.e., 10x PBMC, PBMC, SENSE). Most clusters from different datasets depicted high entropy indicating the mixing of cells from different datasets in respective clusters (Figure 7B). We observed low entropy in the myeloid cell clusters (i.e., poor mixing) which might be due to lower capture of myeloid cells using the SENSE method. Further comparative analysis of data quality by measuring proportions of cytoplasmic, extracellular, membrane, ribosomal, and mitochondrial genes depicted similar profiles indicating the similar quality of single-cell data (Figure 7C). The assessment of the similarity in canonical marker expression distribution from 10X PBMC dataset and our cells from SENSE method depicted similar expressions for most cell types (Figure 7D), with the primary exception being the previously noted myeloid cells. The consistency of key marker expression demonstrates that cell types can be identified reliably using the SENSE method, and comparative analysis can be performed among the samples profiled using different methods.

3 Discussion

Analysis of blood samples is the most direct and least invasive approach (33) to decipher disease mechanisms and identify biomarkers (34). SCP of blood samples is ideal for characterizing how the profiles and characteristics of different immune cells in the blood change in response to disease or therapy, however, the need for immediate sample processing to prepare and preserve viable single cells is a major deterrent towards implementing this on samples collected in a clinic or hospital setting. The traditional method for isolating PBMCs using the Ficoll-Paque density gradient method for SCP is cumbersome and its implementation is challenging due to the lack of time, equipment, and trained personnel in most clinics. This may also be partly responsible for the limited implementation of single-cell profiling in clinical trials. Direct cryopreservation of blood samples without pre-processing has been reported to result in cell death and RNA/DNA degradation, hampering molecular profiling (35). To address these limitations, we have developed the SENSE method for viably freezing WB collected in EDTA tubes without any need for centrifugation steps, special reagents, and trained personnel. The one-step addition of FBS/DMSO freezing solution assists in the cryopreservation of WB cells by preventing the formation of intracellular ice crystals, minimizing cell stress, and thereby maintaining cell integrity/preventing senescence. Granulocytes depletion post-thawing of WB samples enables the recovery of high-quality mononuclear cells as granulocytes are poorly cryopreserved in freezing media and release DNA and lysosomal enzymes promoting cellular damage/clumping (36).

Validation of one-step SENSE method for cryopreservation of WB will jump-start clinical implementation of SCP as well as advance single-cell research. To validate the SENSE method and demonstrate its suitability for cryopreserving high-quality single cells for SCP, we processed freshly collected blood samples with both the SENSE and the traditional density gradient isolation of PBMCs methods. The initial step following the procurement of blood samples, i.e., cryopreservation was much faster and easier with the SENSE method compared to the more time-consuming and complex density-gradient isolation of PBMCs. Although there were slight differences in the viability of cells after the thawing and washing steps, they were not significant and did not affect the quality of the single-cell profiles. We tested multiple quality metrics to evaluate the quality of cells prepared using the SENSE method in comparison to the PBMC isolation method. Quality metrics like median gene counts and unique molecular identifiers (UMIs), were found to be comparable between SENSE and PBMC methods. High mitochondrial content is indicative of poor-quality cells that are either undergoing apoptosis or have lyzed (37). The median representation of <10% mitochondrial genes in WB and PBMC samples confirms a similar proportion of high-quality single cells obtained with both methods. The SENSE method depicted a slight

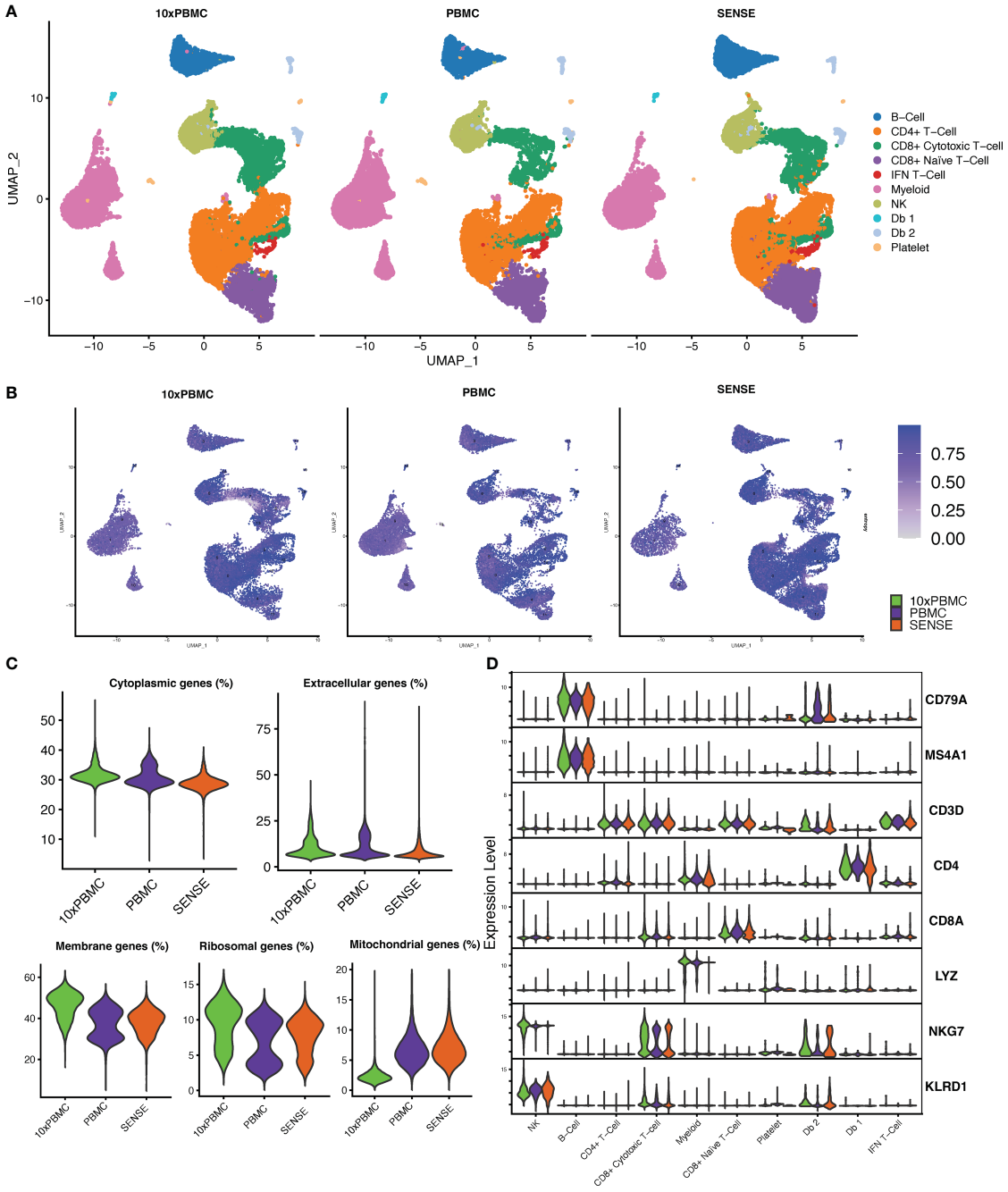


FIGURE 7
Comparing the transcriptomic profile of SENSE method samples with PBMC method samples from 10x Genomics legacy datasets. **(A)** Split UMAP of SENSE and PBMC methods data from our experiments and legacy data, 10x PBMC, from 10x genomics. Cluster labels are based on transferring the labels from SENSE, PBMC data. **(B)** Shannon's entropy-based batch effect estimation. The UMAP plot shows Shannon's entropy of different clusters calculated based on the distribution of SCP protocol labels (i.e., SENSE, PBMC, 10x PBMCs) among the cell's 100 nearest neighbors. The analysis was performed on normalized data with batch effect correction using integration anchors. **(C)** Proportion of group-wise genes in cytoplasmic, membrane, extracellular, and ribosomal gene ontology categories, along with percent mitochondrial genes (green - 10x PBMC, purple-PBMC, orange - SENSE). **(D)** Violin Plots comparing the expression of various cell markers among our SENSE, PBMC data with 10x PBMC dataset.

advantage in capturing the profile of the higher number of cells as compared to the PBMCs method from similar cell suspensions (concentration and viability). The high quality of cells was further confirmed by the lack of cellular damage artifacts with the SENSE method. On closer inspection of the percentage genes in the cytoplasmic ontology category, we see bimodal distribution in PBMC method, in contrast to a unimodal distribution in SENSE method (Figure 2D). Myeloid and platelet cells were found to have a higher percentage of cytoplasmic genes (~32% - 40%) as compared to the rest of the cell types (~20% - 30%) (Supplementary Figure 5). Myeloid and platelets cells were captured more in the PBMC method as compared to SENSE method, resulting in the bimodal

distribution observed in Figure 2D with the former method. Processing times and cryopreservation can result in changes in cell cycle stages when performing SCP which will affect the transcriptome (38). We observed no significant differences in the cell cycle phases between samples processed using the two methods. Multiplets can be biologically misleading and arise when two or more cells are captured in single droplets during encapsulation steps (39). Doublets can occur due to poor quality of cells resulting in two cells clumping together, dying cells and/or broken cells resulting in misleading hybrid transcriptomes (25, 39). We did not observe any significant differences in the number of doublets present between the samples processed using the SENSE method and the PBMCs method. The SENSE method was shown to have slightly fewer doublets compared to the traditional PBMC method (doublet cells: PBMC method, n=916; SENSE method, n=539). Therefore, comparative analysis of the single-cell quality of cells obtained using the one-step cryopreservation SENSE method revealed striking similarities to the traditional multi-step cryopreservation PBMCs method, reinforcing its utility as a method of choice for ease of cryopreservation and single-cell profiling of clinical WB samples.

Cellular landscape revealed by clustering enables identification of cell types and their individual biological states and specific functional roles in disease development and progression (40). All major cell types were represented in UMAPs generated from scRNA-seq data of samples prepared with both methods. The marker genes were shown to have similar expressions for each cluster/cell type or subtype from both methods. Hierarchical clustering demonstrated that different immune cell types have unique transcriptomes that enable their classification into distinct clusters irrespective of sample processing method. The relative cellular abundance analysis revealed that while dominant cell clusters were similar in samples processed using either of the two methods, there were differences in the myeloid and T cells subtype clusters. Although the density gradient centrifugation method to get PBMCs should remove heavier granulocytes, there are instances of incomplete removal of granulocytes, especially in certain pathological conditions like sepsis (41) and autoimmune disorders (42) where there is increased amounts of low-density granulocytes (43). Also, delay in processing of blood can result in granulocyte activation; resultant degranulation gives rise to low density granulocytes that will not be separated out efficiently by density gradient methods (44). In SENSE method, the $CD15^+$ granulocytes are selected and removed to obtain high quality $CD15^-$ mononuclear cells. CD15+ cell depletion was combined with density gradient centrifugation to effectively purify PBMCs from sepsis patients with high percentage of low-density granulocytes (43). Though there was some resultant loss of additional cells other than granulocytes during the $CD15^+$ cells depletion steps, the functional cellular properties were not compromised (43). The observed differences in myeloid clusters in this study might be attributed to the SENSE method's removal of $CD15^+$ cells to filter out sticky granulocytes that might have also filtered out aggregating monocytes and platelets. Importantly, the SENSE method was able to recapitulate the myeloid compartment associated with disease as we observed similar patient-wise differences of cell-type distribution in scRNA-seq data from both methods. In summary, even though SENSE captures fewer myeloid cells compared to the PBMC method, it is still capable of revealing differences in the proportions of myeloid cells within given samples.

On the other hand, we observed more efficient capture of T cells in SENSE method cryopreserved WB samples. Importantly, our focused analysis on the T cells compartment showed that the SENSE method captured a significantly higher number of T cells representing various types and subtypes, including Naïve, Effector, and Memory T-cells. The possible cause of lower T-cell enrichment with the PBMCs method might be due to some T-cells being lost during the Ficoll-Paque density gradient centrifugation step due to the difference in density of these cells (45). These results make our simple WB sample cryopreservation combined with $CD15^+$ granulocyte removal method especially suitable for immune repertoire profiling using VDJ enrichment to explore the association of T cell clonality with disease or therapeutic outcomes analyses. Based on the single-cell quality metrics, the cryopreserved WB using the SENSE method yielded high-quality single cells similar to cryopreserved PBMCs isolated using the traditional density-gradient method. Furthermore, the SENSE method can be extended for more granular characterization of immune repertoire using single cell proteomics/multidimensional profiling.

Comparison analysis of our data with an external 10x Genomics PBMC dataset (32) revealed concordance between the three datasets as all cell types were consistently identified in all three datasets. The high quality of cells obtained with the SENSE method was further demonstrated by quality metrics like lower % membrane genes and higher % extracellular genes compared to the external PBMC dataset. This analysis further validates the robustness of the SENSE method to acquire high quality single-cells for single cell profiling.

4 Methods

4.1 Sample collection

Informed consent according to Emory University IRB protocol (IRB00079391 Determinants of Childhood Autoimmunity) was obtained from Juvenile Idiopathic Arthritis (JIA) (n=5) and pediatric Lupus (n=1) patients being treated in the rheumatology clinic in CHOA prior to sample collection. Blood samples were collected in lavender top EDTA tubes and transported to the lab from the clinic at room temperature. Samples were cryopreserved within 2h post-collection.

4.2 PBMCs isolation and whole blood cryopreservation

The freshly collected blood was split into two equal aliquots, with one aliquot processed for isolation of PBMCs while the other aliquot was frozen directly using a cryopreservation solution. PBMCs were isolated using the standard Ficoll-Paque density-gradient method according to manufacturer's instructions

(Cytiva). Briefly, blood diluted in phosphate buffer saline (PBS) (1:1) was gently layered onto Ficoll-Paque PLUS (Cytiva, 17144002) and spun at 500g for 30 minutes at 21°C. The top layer (plasma) was removed and discarded. The layer containing the mononuclear cells was then carefully removed and diluted with 3x volume of PBS, mixed well, and spun at 500g, for 10 minutes at 21°C. The pellet was resuspended in PBS and washed again by spinning for 10 minutes at 500g and 21°C. The PBMCs pellet was then resuspended in 1 ml recovery cell culture freezing media (Fisher Scientific, 12648010) at a concentration of $<10X10^6$ cells/ml. The second set of blood samples (by the SENSE method) was viably preserved by mixing whole blood 1:1 with freezing solution made up of 80% heat-inactivated fetal bovine serum (hiFBS) and 20% dimethyl sulfoxide (DMSO). Samples were gradually frozen by placing in Mr. Frosty freezing container (Fisher Scientific, 5100-0001) and stored at -80 °C till further use.

4.3 Single cells preparation

Frozen PBMC samples were thawed and washed with wash buffer (PBS containing 1% BSA) to prepare viable single cell suspensions (14). Frozen whole blood samples were also thawed, and cells were pelleted (380g, RT, 6.5 minutes), the supernatant was gently removed so as not to disturb the pellet, which was then resuspended in EasySep buffer (STEMCELL technologies, 20144) and filtered through 100 µm filter mesh (Fisherbrand, 22363549). The EDTA concentration of EasySep buffer used for washing and diluting the cells was modified. The amount of EDTA in the EasySep buffer containing 1mM EDTA was increased to 4mM by adding an additional 3mM EDTA. EDTA is known to rapidly reverse the preferential binding of platelets to monocytes (46). Therefore, the presence of higher EDTA concentration in the buffer results in increased capture of high-quality mononuclear cells. Granulocytes were removed using a modified EasySep CD15 selection (EasySep™ Human CD15 Positive Selection Kit; STEMCELL technologies, 18651) protocol. Also, RBC depletion beads (STEMCELL Technologies, 18170) were added following the CD15 cocktail mix and RapidSpheres incubation steps to remove red blood cells. Following EasySep magnetic separation of $CD15^+$ antibody-bound granulocytes, the $CD15^-$ mononuclear cells supernatant was collected, cells pelleted and resuspended in PBS containing 1% BSA for generating viable cells for scRNA-seq libraries. A detailed stepwise protocol for the SENSE method is included as Supplementary document 1.

4.4 Single-cell assays and sequencing

ScRNA-seq libraries were prepared from viably thawed WB and PBMCs single-cell samples prepared in the previous section according to manufacturer's (10x Genomics) instructions. CellPlex kit (10x Genomics, 1000261), which allows the pooling of samples prior to GEMs generation by labeling samples with unique cell multiplexing oligos (CMOs), was used to multiplex samples. The pooled samples were used to generate GEMs, followed by RT-PCR steps and cDNA amplification using Next GEM single cell 3'v3.1 kits (10x Genomics, 1000268). Following the cDNA amplification step, size selection beads were used to separate the CMO and gene expression (GEX) cDNAs that were then used to prepare the CMO and GEX libraries respectively. The final CMO and GEX libraries were then pooled and sequenced according to 10x Genomics sequencing parameters using Novaseq S4 PE100 (Illumina) kits for comprehensive transcriptome profiling.

4.5 Single-cell profiling analysis

The raw FASTQ files from each multiplex sample were aligned using 10x Genomics Cell Ranger (32) 6.1.2 to align against a reference human genome (GRCh38) for generating raw cell-gene count matrices. The count and CMO matrices from the samples were analyzed with R (v 4.2.2) using Seurat (47) (v 4.0.4) and other Bioconductor packages. Low-quality cells were filtered using Seurat to keep only cells with >200 unique genes, >600 UMI reads, and < 20% mitochondrial UMIs. Potential doublets were marked using the doubletFinder (25) algorithm that identifies doublets based on neighborhood search on principal component analysis (PCA). Assuming 3.5% of doublet formation from the 10x multiplexing experiment, we performed analysis with top 10 principal components with a neighborhood size of 0.1(pK) to predict doublets. The count matrices were normalized using the SCTransform algorithm, regressing out the per-cell UMI count, the number of unique features per cell, and the percent mitochondrial reads mapped to a cell. The normalized cell count was used for selecting the top 2,000 variable genes for principal component analysis (PCA) to identify the principal components capturing the most variance across the samples. Similar cells were clustered together via Louvain clustering on the top principal components using the Seurat package that was visualized Uniform Manifold Approximation and Projection (UMAP) to determine the overall relationship among the cells. The cell clusters were manually annotated based on canonical cell markers described in our previous studies. The cell markers for the different cell clusters were identified by comparing target cell types with others captured in the assay using the Wilcoxon Rank Sum test (adjusted $P<.10$, average log2FC > 0.25, and percent cell expression > 25%).

4.6 Entropy calculation and Gene-Ontology based cellular component enrichment

Shannon entropy was calculated per cell for assessing the batch effect due to method variation using 100 neighbors and 20 principal components using the CellMixS (27) R package. The gene signatures for cellular components (extracellular region, cytoplasm, membrane, ribosome) were sourced from Gene Ontology (48) database (GO:0005576, GO:0005737, GO:0016020, GO:0005840). PercentageFeatureSet function of Seurat was used to calculate the percentage of all UMIs that belong to the gene signature per cell.

4.7 Comparative analysis of cell types across single-cell processing methods

We performed a comparative analysis of cell type abundance as well as gene expression between SENSE and PBMCs methods. The cellular proportion per patient were calculated and compared between methods using paired t-tests. The cell types with a P value <.05 were considered significantly differently enriched between methods. To determine the correlation between cell types between methods we implemented the ClusterMap package in R designed to compare cellular profiles across multiple single-cell datasets (26). Initially differentially expressed genes (DEGs) for each cell type in a method-specific manner (i.e., SENSE, PBMC) were identified based on the fold change and Wilcoxon Rank Sum Test (average log-FC ≥0.25, >25% of cells expressing gene, and P <.01). DEGs were computed using Seurat's "FindAllMarkers" function. This was followed by hierarchical clustering of DEGs using their presence or absence (binary expression) in different cell types to generate a cluster dendrogram. The relative distance of cell types on the cluster dendrogram can be quantified by the similarity of the cell types. The similarity of the cell types is measured based on their Jaccard index. To match a cell type profile with another cell type ClusterMap introduced a purity tree cut algorithm (26). The algorithm uses the origin of cell types, clustering pattern on the dendrogram, and similarity to match the cell types from different methods. This results in matching cell types as well as merging cell types in a group if cell types depict >90% similarity within a method. The results from the analysis are displayed as a Circos plot summarizing the similarity in cell types and subtypes similarity.

4.8 Cellular communication and interaction analysis

Cellular communication analysis was performed using the CellChat platform (28). Cells from each processing method were isolated, and ligand-receptor (L-R) analysis was performed on the SENSE and PBMC methods independently using the standard CellChat analysis. Differentially expressed signaling genes were identified using the Wilcoxon rank sum test ($P< 0.05$), which was followed by communication probability/strength calculation between any interacting cell types. The cell-cell communications were filtered out if they were present in a cell type/subtype with less than 10 cells. The number of interactions and their strengths were aggregated for each method. To compare the overall signaling structure between cells in SENSE and PBMC samples, interaction weights were used, which sum the information flow of all L-R interactions between two cell types of lymphoid and myeloid lineages. The sum of outgoing or incoming communication probability associated with each cell group was visualized on a scatter plot showing the dominant senders (sources) and receivers (targets) cell types. The size of the data points on the scatter plot corresponded to the number of inferred links, both outgoing and incoming, connected to each specific cell type. Information flow/interaction strength characterizes the likelihood of cell-cell interaction occurring through a given pathway. Cells with high expression of a known ligand will have high information flow scores with cells that have high expression of the matching receptor. The conserved or processing method-specific pathways were evaluated by comparing the sum of communication probability among cell-type pairs for each pathway.

4.9 Comparing the SENSE and PBMC data with external PBMC dataset

Single-cell gene expression dataset for frozen PBMC samples (10x PBMC) from 3 donors (Donor A, B, and C) were downloaded from 10x Genomics datasets (32). The filtered gene expression matrices were merged with the PBMC and SENSE samples. The count matrices were again normalized using the SCTransform algorithm, regressing out the per-cell UMI count, the number of unique features per cell, and the percent mitochondrial reads mapped to a cell. The top 2,000 variable genes were found, and further Louvain clustering was performed on the top principal components using the Seurat package that generated a UMAP to visualize the overall relationship among the cells. To correct for any batch effect the samples count matrices from 10x PBMC and PBMC, SENSE datasets were normalized and integrated using integration anchors-based batch correction approach of the Seurat package. The cell clusters were manually annotated by transferring cluster labels from PBMC, SENSE to 10x PBMC samples. Based on distribution of existing labels on new clustering, some clusters were merged like B-Cell (B-Cell and Memory B-Cell), CD4+ T-Cell (CD4+ Naïve and Memory T-Cell) and $CD8^+$ Cytotoxic T-Cell ($CD8^+$ Effector and Memory T-Cell).

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 below: https://www.ncbi.nlm.nih.gov/geo/, GSE226557(GEO).

Ethics statement

The studies involving humans were approved by Emory University IRB protocol (IRB00079391 Determinants of Childhood Autoimmunity). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

SS: Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. BT: Formal analysis,

Investigation, Methodology, Writing – original draft, Writing – review & editing. WP: Formal analysis, Software, Writing – original draft, Writing – review & editing. MB: Investigation, Writing – review & editing. LP: Investigation, Writing – review & editing. RP: Conceptualization, Writing – review & editing. SP: Investigation, Writing – review & editing. SB: Formal analysis, Writing – original draft, Writing – review & editing. DM: Conceptualization, Writing – review & editing. MB: Conceptualization, Formal analysis, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. MaB acknowledges funding from National Institutes of Health (R01HL086741), CURE Childhood foundation and Emory Startup to Bhasin lab. SP is supported in part by a grant from the Marcus Foundation Inc., Atlanta.

Acknowledgments

Sequencing was carried out at the NPRC genomics core at Emory University. The Emory NPRC Genomics Core is supported in part by NIH P51 OD011132. Sequencing data was acquired on an Illumina NovaSeq 6000 funded by NIH S10 OD026799.

Conflict of interest

Author MaB serves on the board of Canomiks Inc. as chief scientific advisor and has equity in it. SB serves as CEO of Anxomics LLC and has equity in it.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1271800/full#supplementary-material

References

1. Tang X, Huang Y, Lei J, Luo H, Zhu X. The single-cell sequencing: new developments and medical applications. *Cell Biosci* (2019) 9:53. doi: 10.1186/s13578-019-0314-y

2. Xu J, Fang Y, Chen K, Li S, Tang S, Ren Y, et al. Single-cell RNA sequencing reveals the tissue architecture in human high-grade serous ovarian cancer. *Clin Cancer Res* (2022) 28(16):3590–602. doi: 10.1158/1078-0432.CCR-22-0296

3. Angarica VE, Del Sol A. Bioinformatics tools for genome-wide epigenetic research. *Adv Exp Med Biol* (2017) 978:489–512. doi: 10.1007/978-3-319-53889-1_25

4. Trapnell C. Defining cell types and states with single-cell genomics. *Genome Res* (2015) 25(10):1491–8. doi: 10.1101/gr.190595.115

5. Kim N, Eum HH, Lee HO. Clinical perspectives of single-cell RNA sequencing. *Biomolecules* (2021) 11(8):1161. doi: 10.3390/biom11081161

6. Radpour R, Forouharkhou F. Single-cell analysis of tumors: Creating new value for molecular biomarker discovery of cancer stem cells and tumor-infiltrating immune cells. *World J Stem Cells* (2018) 10(11):160–71. doi: 10.4252/wjsc.v10.i11.160

7. Gohil SH, Iorgulescu JB, Braun DA, Keskin DB, Livak KJ. Applying high-dimensional single-cell technologies to the analysis of cancer immunotherapy. *Nat Rev Clin Oncol* (2021) 18(4):244–56. doi: 10.1038/s41571-020-00449-x

8. Lim B, Lin Y, Navin N. Advancing cancer research and medicine with single-cell genomics. *Cancer Cell* (2020) 37(4):456–70. doi: 10.1016/j.ccell.2020.03.008

9. Pilcher W, Thomas BE, Bhasin SS, Jayasinghe RG, Yao L, Gonzalez-Kozlova E, et al. Cross center single-cell RNA sequencing study of the immune microenvironment in rapid progressing multiple myeloma. *NPJ Genom Med* (2023) 8(1):3. doi: 10.1038/s41525-022-00340-x

10. Yao L, Jayasinghe RG, Lee BH, Bhasin SS, Pilcher W, Doxie DB, et al. Comprehensive characterization of the multiple myeloma immune microenvironment using integrated scRNA-seq, cyTOF, and CITE-seq analysis. *Cancer Res Commun* (2022) 2(10):1255–65. doi: 10.1158/2767-9764.CRC-22-0022

11. Mumme HL, Bhasin SS, Thomas BE, Dwivedi B, DeRyckere D, Castellino SM, et al. Pediatric single cell cancer atlas: an integrative web-based resource for single cell transcriptome data from pediatric leukemias. *Blood* (2021) 138:3488. doi: 10.1182/blood-2021-154282

12. Tellechea A, Bai S, Dangwal S, Theocharidis G, Nagai M, Koerner S, et al. Topical application of a mast cell stabilizer improves impaired diabetic wound healing. *J Invest Dermatol* (2020) 140(4):901–11 e11. doi: 10.1016/j.jid.2019.08.449

13. Theocharidis G, Baltzis D, Roustit M, Tellechea A, Dangwal S, Khetani RS, et al. Integrated skin transcriptomics and serum multiplex assays reveal novel mechanisms of wound healing in diabetic foot ulcers. *Diabetes* (2020) 69(10):2157–69. doi: 10.2337/db20-0188

14. Theocharidis G, Thomas BE, Sarkar D, Mumme HL, Pilcher WJR, Dwivedi B, et al. Single cell transcriptomic landscape of diabetic foot ulcers. *Nat Commun* (2022) 13(1):181. doi: 10.1038/s41467-021-27801-8

15. Wylezinski LS, Shaginurova GI, Spurlock Iii CF. Longitudinal assessment and stability of long non-coding RNA gene expression profiles measured in human peripheral whole blood collected into PAXgene blood RNA tubes. *BMC Res Notes* (2020) 13(1):531. doi: 10.1186/s13104-020-05360-3

16. Vaught JB. Blood collection, shipment, processing, and storage. *Cancer Epidemiol Biomarkers Prev* (2006) 15(9):1582–4. doi: 10.1158/1055-9965.EPI-06-0630

17. Wang X, Park J, Susztak K, Zhang NR, Li M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* (2019) 10(1):380. doi: 10.1038/s41467-018-08023-x

18. Efroni I, Ip PL, Nawy T, Mello A, Birnbaum KD. Quantification of cell identity from single-cell gene expression profiles. *Genome Biol* (2015) 16(1):9. doi: 10.1186/s13059-015-0580-x

19. Haque A, Engel J, Teichmann SA, Lonnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* (2017) 9(1):75. doi: 10.1186/s13073-017-0467-4

20. Betsou F, Gaignaux A, Ammerlaan W, Norris PJ, Stone M. Biospecimen science of blood for peripheral blood mononuclear cell (PBMC) functional applications. *Curr Pathobiol Rep* (2019) 7(2):17–27. doi: 10.1007/s40139-019-00192-8

21. Massoni-Badosa R, Iacono G, Moutinho C, Kulis M, Palau N, Marchese D, et al. Sampling time-dependent artifacts in single-cell genomics studies. *Genome Biol* (2020) 21(1):112. doi: 10.1186/s13059-020-02032-0

22. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol* (2016) 17 (1):1–15. doi: 10.1186/s13059-016-0888-1

23. DePasquale EAK, Schnell DJ, Van Camp PJ, Valiente-Alandi I, Blaxall BC, Grimes HL, et al. DoubletDecon: deconvoluting doublets from single-cell RNA-sequencing data. *Cell Rep* (2019) 29(6):1718–27 e8. doi: 10.1016/j.celrep.2019.09.082

24. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell* (2021) 184(13):3573–87.e29. doi: 10.1016/j.cell.2021.04.048

25. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst* (2019) 8 (4):329–37.e4. doi: 10.1016/j.cels.2019.03.003

26. Gao X, Hu D, Gogol M, Li H. ClusterMap: compare multiple single cell RNA-Seq datasets across different experimental conditions. *Bioinformatics* (2019) 35(17):3038–45. doi: 10.1093/bioinformatics/btz024

27. Lutge A, Zyprych-Walczak J, Brykczynska Kunzmann U, Crowell HL, Calini D, Malhotra D, et al. CellMixS: quantifying and visualizing batch effects in single-cell RNA-seq data. *Life Sci Alliance* (2021) 4(6):e202001004. doi: 10.26508/lsa.202001004

28. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun* (2021) 12(1):1088. doi: 10.1038/s41467-021-21246-9

29. Farr L, Ghosh S, Moonah S. Role of MIF cytokine/CD74 receptor pathway in protecting against injury and promoting repair. *Front Immunol* (2020) 11. doi: 10.3389/fimmu.2020.01273

30. Cham LB, Gunst JD, Schleimann MH, Frattari GS, Rosas-Umbert M, Vibholm LK, et al. Single cell analysis reveals a subset of cytotoxic-like plasmacytoid dendritic cells in people with HIV-1. *iScience* (2023) 26(9):107628. doi: 10.1016/j.isci.2023.107628

31. Li W, Zhang B, Cao W, Zhang W, Li T, Liu L, et al. Identification of potential resistance mechanisms and therapeutic targets for the relapse of BCMA CAR-T therapy in relapsed/refractory multiple myeloma through single-cell sequencing. *Exp Hematol Oncol* (2023) 12(1):44. doi: 10.1186/s40164-023-00402-5

32. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* (2017) 8:14049. doi: 10.1038/ncomms14049

33. Koh W, Pan W, Gawad C, Fan HC, Kerchner GA, Wyss-Coray T, et al. Noninvasive in *vivo* monitoring of tissue-specific global gene expression in humans. *Proc Natl Acad Sci USA* (2014) 111(20):7361–6. doi: 10.1073/pnas.1405528111

34. Taguchi A, Hanash SM. Unleashing the power of proteomics to develop blood-based cancer markers. *Clin Chem* (2013) 59(1):119–26. doi: 10.1373/clinchem.2012.184572

35. Fowke KR, Behnke J, Hanson C, Shea K, Cosentino LM. Apoptosis: a method for evaluating the cryopreservation of whole blood and peripheral blood mononuclear cells. *J Immunol Methods* (2000) 244(1):139–44. doi: 10.1016/S0022-1759(00)00263-5

36. Bakken AM. Cryopreserving human peripheral blood progenitor cells. *Curr Stem Cell Res Ther* (2006) 1(1):47–54. doi: 10.2174/157488806775269179

37. Wang X, He Y, Zhang Q, Ren X, Zhang Z. Direct comparative analyses of 10X genomics chromium and smart-seq2. *Genom Proteomics Bioinf* (2021) 19(2):253–66. doi: 10.1016/j.gpb.2020.02.005

38. Devine RD, Sekhri P, Behbehani GK. Effect of storage time and temperature on cell cycle analysis by mass cytometry. *Cytometry A* (2018) 93(11):1141–9. doi: 10.1002/cyto.a.23630

39. Sun B, Bugarin-Estrada E, Overend LE, Walker CE, Tucci FA, Bashford-Rogers RJM. Double-jeopardy: scRNA-seq doublet/multiplet detection using multi-omic profiling. *Cell Rep Methods* (2021) 1(1):100008. doi: 10.1016/j.crmeth.2021.100008

40. Ke M, Elshenawy B, Sheldon H, Arora A, Buffa FM. Single cell RNA-sequencing: A powerful yet still challenging technology to study cellular heterogeneity. *Bioessays* (2022) 44(11):e2200084. doi: 10.1002/bies.202200084

41. van den Akker ELT, Baan CC, van den Berg B, Russcher H, Joosten K, Hokken-Koelega ACS, et al. Ficoll-separated mononuclear cells from sepsis patients are contaminated with granulocytes. *Intensive Care Med* (2008) 34(5):912–6. doi: 10.1007/s00134-007-0989-0

42. Carmona-Rivera C, Kaplan MJ. Low-density granulocytes: a distinct class of neutrophils in systemic autoimmunity. *Semin Immunopathol* (2013) 35(4):455–63. doi: 10.1007/s00281-013-0375-7

43. Schenz J, Obermaier M, Uhle S, Weigand MA, Uhle F. Low-density granulocyte contamination from peripheral blood mononuclear cells of patients with sepsis and how to remove it – A technical report. *Front Immunol* (2021) 12. doi: 10.3389/fimmu.2021.684119

44. McKenna KC, Beatty KM, Vicetti Miguel R, Bilonick RA. Delayed processing of blood increases the frequency of activated CD11b+ CD15+ granulocytes which inhibit T cell function. *J Immunol Methods* (2009) 341(1-2):68–75. doi: 10.1016/j.jim.2008.10.019

45. Appay V, Reynard S, Voelter V, Romero P, Speiser DE, Leyvraz S. Immuno-monitoring of CD8+ T cells in whole blood versus PBMC samples. *J Immunol Methods* (2006) 309(1-2):192–9. doi: 10.1016/j.jim.2005.11.007

46. Sarma J, Laan CA, Alam S, Jha A, Fox KA, Dransfield I. Increased platelet binding to circulating monocytes in acute coronary syndromes. *Circulation* (2002) 105 (18):2166–71. doi: 10.1161/01.CIR.0000015700.27754.6F

47. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive integration of single-cell data. *Cell* (2019) 177(7):1888–902 e21. doi: 10.1016/j.cell.2019.05.031

48. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* (2000) 25(1):25–9. doi: 10.1038/75556

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 25 June 2024
DOI 10.3389/fimmu.2024.1309916

OPEN ACCESS

EDITED BY
Jennie R. Lill,
Genentech Inc., United States

REVIEWED BY
Marvin Thielert,
Max Planck Institute of Biochemistry, Germany
Paola de Candia,
University of Naples Federico II, Italy

*CORRESPONDENCE
Alvaro Martinez Barrio
Alvaro.Martinez.Barrio@pixelgen.com

RECEIVED 08 October 2023
ACCEPTED 09 April 2024
PUBLISHED 25 June 2024

CITATION
Rhomberg-Kauert J, Karlsson M, Thiagarajan D, Kallas T, Karlsson F, Fredriksson S, Dahlberg J and Martinez Barrio A (2024) Using adjusted local assortativity with Molecular Pixelation unveils colocalization of membrane proteins with immunological significance. *Front. Immunol.* 15:1309916. doi: 10.3389/fimmu.2024.1309916

Using adjusted local assortativity with Molecular Pixelation unveils colocalization of membrane proteins with immunological significance

Jan Rhomberg-Kauert[1,2], Max Karlsson[1], Divya Thiagarajan[1], Tomasz Kallas[1], Filip Karlsson[1], Simon Fredriksson[1,3], Johan Dahlberg[1] and Alvaro Martinez Barrio[1]*

[1]Pixelgen Technologies AB, Stockholm, Sweden, [2]Department of Geodesy and Geoinformation, TU Wien, Vienna, Austria, [3]Department of Protein Science, Royal Institute of Technology, Stockholm, Sweden

Advances in spatial proteomics and protein colocalization are a driving force in the understanding of cellular mechanisms and their influence on biological processes. New methods in the field of spatial proteomics call for the development of algorithms and open up new avenues of research. The newly introduced Molecular Pixelation (MPX) provides spatial information on surface proteins and their relationship with each other in single cells. This allows for *in silico* representation of neighborhoods of membrane proteins as graphs. In order to analyze this new data modality, we adapted local assortativity in networks of MPX single-cell graphs and created a method that is able to capture detailed information on the spatial relationships of proteins. The introduced method can evaluate the pairwise colocalization of proteins and access higher-order similarity to investigate the colocalization of multiple proteins at the same time. We evaluated the method using publicly available MPX datasets where T cells were treated with a chemokine to study uropod formation. We demonstrate that adjusted local assortativity detects the effects of the stimuli at both single- and multiple-marker levels, which enhances our understanding of the uropod formation. We also applied our method to treating cancerous B-cell lines using a therapeutic antibody. With the adjusted local assortativity, we recapitulated the effect of rituximab on the polarity of CD20. Our computational method together with MPX improves our understanding of not only the formation of cell polarity and protein colocalization under stimuli but also advancing the overall insight into immune reaction and reorganization of cell surface proteins, which in turn allows the design of novel therapies. We foresee its applicability to other types of biological spatial data when represented as undirected graphs.

KEYWORDS

molecular pixelation, single cell, spatial proteomics, graph theory, topological data analysis, local assortativity, uropod formation, Rituximab

Introduction

The spatial organization of proteins governs a number of complex cellular processes such as cell signaling, cell–cell communication, and mobility. To enable the detection of proteins in cells and tissues, affinity reagents have remained the mainstay in the field. They have been used extensively in fluorescence microscopy tagged with fluorophores, typically providing fluorescence intensity data from each channel in one focal plane. The generation of three-dimensional information at high throughput and multiplexity is thus limited by the need for microscopy imaging. Imaging flow cytometry overcomes this throughput limitation by coupling traditional flow cytometers with the acquisition of an image of each cell (1). Fluorescence resonance energy transfer (FRET) microscopy measures the transfer of energy from an excited molecular fluorophore (the donor) to another fluorophore (the acceptor) (2). FRET microscopy imaging can achieve colocalization of labeled pairs of probes within sub-micron distances. However, the limitation of all microscopy techniques in terms of dimensionality and high-plexity information remains, as only a few antibodies can be acquired at the same time on the different microscope channels. Furthermore, the signal to noise is also hampered by auto-fluorescence, detector noise, optical noise, and spectral bleed-through between channels. Super-resolution imaging methods have provided groundbreaking insights in three-dimensional (3D) but are yet limited in multiplexing and throughput (3). Furthermore, super-resolution instrumentation is expensive and requires advanced training to even analyze the data.

To overcome the multiplexity problems, mass cytometry coupled antibodies to isotopes of different atomic weights that are detected by a mass spectrometer, such that the quantity of detected ions in a particular mass channel becomes a proxy for molecular detection (4). Although imaging mass cytometry (IMC) has been used with success in tissues, still, the multiplexity reported is still as high as 80 proteins (5). However, the application of IMC to tissues of 1-mm thickness (6) holds promise for 3D resolution on single cells one day. With the advent of next-generation sequencing (NGS), the tagging of antibodies with DNA oligos coupled to NGS readout has the potential of unlimited multiplexing. Although reading protein tags does not provide any spatial information (7), it has been shown to multiplex to 273 proteins (8). Other methods with different tagging strategies have been successful in showing antibody specificity to antigens by sequencing paired B-cell receptor (BCR) clonotypes interacting with DNA-barcoded antigens (9) or a recent proximity ligation assay by sequencing that is able to infer protein complexes (10). Similar to FRET, the drawback of the Prox-seq approach is that only proteins of interest are found in pairs, not larger constellations, and the location of proteins in the cell membrane is not achieved, as it lacks the relationship context.

As a result, developing a novel method to study spatial protein organization in a single cell has gained enormous significance in the past decade (11–13). A new emerging NGS-based method, Molecular Pixelation (MPX), provides spatial information on surface protein abundance and their relationship with each other on single cells in a three-dimensional field of view. Every single cell in MPX high-throughput datasets is encoded as a bipartite graph, which in turn can be analyzed to gain new insights into the colocalization of cell surface proteins (14). As graph metrics are commonly employed in social and biological networks, there are many analysis methods with potential applications for this new type of single-cell proteomics input data. In this study, we have adapted the application of local assortativity (15) to not only compare sets of proteins per node in the graph of each single cell but also numerically compare all nodes of the graph in terms of attribute distribution. Although MPX is not able to infer direct protein interactions in its current state, it enables the discovery of protein constellations of biological significance and allows the exploration of protein colocalization as a novel therapeutic target.

Materials and methods

Molecular Pixelation

The MPX workflow builds an amplicon in three steps: the first step involves staining the cells with antibody–oligonucleotide conjugates (AOCs). In the next step, a set of DNA pixels, each containing a unique sequence identifier so-called A-pixel, hybridize into a group of spatially proximal AOCs each, and a gap-fill ligation reaction adds the unique sequence identifier to the AOC, imprinting AOCs with the same A-pixel neighborhood tag. Next, a second reaction is performed with a set of B-pixels connecting several A-pixel areas. The combined spatial information imprinted by A- and B-pixels preserves the information of which protein molecules were spatially adjacent on the original cell surface (14).

MPX data from any immune cell in solution can be represented as a bipartite graph G, where A- and B-pixels are nodes interconnected by a set of AOCs as edges. We transformed each bipartite graph G into an A-node projection, where edge attributes of the bipartite graph become node attributes of the projected graph in the A-node of the A–B parts (14). Subsequently, the A-nodes become directly connected following the original connections of the B-nodes. We used A-node projected graphs from the original bipartite graph G throughout this study to move the antibody edge labels and counts into A-nodes and to be able to use local assortativity. Local assortativity only works for MPX if protein labels and counts are projected to the A-node. For the rest of this study, when we used the concept of node or vertex on a cell graph, and we referred to an A-node with antibody labels and counts.

MPX can record the counts of each protein molecule, which can be used to assess differences in protein abundance between cell states or conditions. However, the two most important features of this data type are to be able to study the relative positioning of individual protein markers, as well as their colocalization. First, the Jaccard Index and Pearson's correlation across different proteins in the same single-cell graph were used in order to ascertain if two proteins tend to colocalize or not upon stimulation. MPX global measure to study homophily/heterophily in single-cell graphs currently requires the definition of a local neighborhood parameter to identify molecules present in pixels assigned to a

given antibody (14), and it would be desirable to have a parameter-free definition of local neighborhoods.

Molecular Pixelation datasets

Karlsson and colleagues, by applying the MPX workflow, generated several datasets[1] demonstrating the technological capabilities of MPX for different applications (14).

One of those applications is stimulating human T cells with phytohemagglutinin (PHA) followed by IL2 for 5 days into the formation of uropods. Leukocyte migration prompts the formation of distinct structures in cells in order to follow chemotactic gradients and reach the target tissue. Leukocytes polarize and convert mechanical force into forward locomotion by coordinating a regulated bidirectional cycle: the leading edge pushes the cell forward, whereas the plasma membrane moves to the rear (16). The leukocyte uropod formation was first described during studies of the interactions between T lymphoblasts and macrophages (17). Irrespective of the cell type, the uropod trailing protrusion, referred to as the "uropod knob" (18), involves intracellular actin polymerization and actomyosin contraction providing the force that creates the protrusion. CD50 (ICAM3) and several proteins are supposed to colocalize on the uropod structure (Figure 1A) (19) with CD50 polarization being validated by microscopy (Figure 1B) (14).

Karlsson and colleagues fixed and cultured PHA blasts on plates coated with either 5 µg/mL of CD54Fc antibody alone or with two different chemotactic cytokines in solution or 10 ng/mL of CCL5 (RANTES) in one condition or CCL2 (MCP1) in another at 37°C for 1 h. We downloaded the output dataset PXL files from three of the conditions in the experiment[2], one with cells fixed with CD54 and stimulated within solution RANTES ("uropod CD54 fixed RANTES stimulated", 657 cells), a second one with cells fixed with CD54 ("uropod CD54 fixed", 733 cells), and the last with cells in solution without the stimulation as a control ("uropod control", 658 cells); for the rest of this paper, we will refer to these datasets as stimulated cells, fixed control, and control, respectively.

In another MPX application, Raji cells (ATCC, Manassas, VA, USA)[3] were Fc-receptor blocked with 50 µg/mL of human IgG for 15 min at 4°C and washed. Cells were then either fixed directly with paraformaldehyde (PFA) ("control", 607 cells) or incubated with 20 µg/mL of rituximab (RTX) (ProteoGenix, Schiltigheim, France) with a specific AOC ("treated", 873 cells) in Roswell Park Memorial Institute (RPMI) media for 60 min at 37°C, followed by PFA fixation and washing. RTX, a monoclonal antibody therapy approved for medical use in 1997, targets CD20 primarily on the surface of B cells. RTX mediates antibody-dependent cellular cytotoxicity (ADCC), allowing specific NK-cell killing (23) (Figure 1C) by polarizing CD20 on a cap at the surface of B cells (Figure 1D).

We downloaded the output dataset PXL files from the two conditions in the Raji cell MPX experiment[4] (Figure 1E) and applied adjusted local assortativity to them. Comparing both populations of treated and control cells allowed us to recapitulate the already described RTX mechanism of action (23).

Local assortativity

Assortativity is a well-known concept in graph theory and network science, which compares the patterns of vertex attributes across the network. The most common version is the global assortativity where the whole graph is considered and the vertices are compared on a global scale (24, 25). A downside of the global measurement is that it does not account for local heterogeneity in subregions of the network. This problem was overcome by different versions of the local assortativity, which focused on studying the homogeneity in communities of labeled networks. The advantage of the local assortativity is that each vertex gets assigned a score based on the attribute of interest, and thus, one can analyze in detail the distribution of the network properties (15, 26, 27). A classic example of this is calculating the local assortativity for the degree of each vertex (Figure 1F). We transformed every cell bipartite graph into their A-node projected graph and transferred the labels from the edges to the vertices to be able to apply local assortativity and calculate an assortativity score for each vertex.

Herein, we used local assortativity as defined by Peel and colleagues (15) and applied it to two of the MPX public datasets (Figure 2) generated with slight modifications in the PageRank threshold (Supplementary Data 1.1).

In addition to the 76 antibodies targeting specific protein epitopes in the panel (14), three mouse isotype control antibodies were included (mIgG2b, mIgG1, and mIgG2a). With the information provided by these control proteins, we determined a lower boundary required in order to calculate the colocalization score on the other 76 proteins. This "isotype threshold" is set by the maximum number of isotype protein counts per cell based on the three mouse controls.

$$threshold_{proteins} = max(N_{proteins}(mIgG2b), N_{proteins}(mIgG1), N_{proteins}(mIgG2a))$$

If the number of raw molecule counts for a given protein in a cell is below that threshold, the assortativity scores are set to zero for that protein in all nodes of the A-node graph. A second filter ("vertex threshold") will require each protein to be present in more than 10 vertices on the A-node graph. If those limits are not met, we consider that there is too little information in the cell to create a high-quality local assortativity distribution and all the A-nodes for that protein on that cell are also initialized to zero.

1 https://software.pixelgen.com/datasets/.

2 https://software.pixelgen.com/datasets/uropod-t-cells-v1.0-immunology-l.

3 https://www.atcc.org/products/ccl-86.

4 https://software.pixelgen.com/datasets/cd20-rituximab-v1.0-immunology-l.

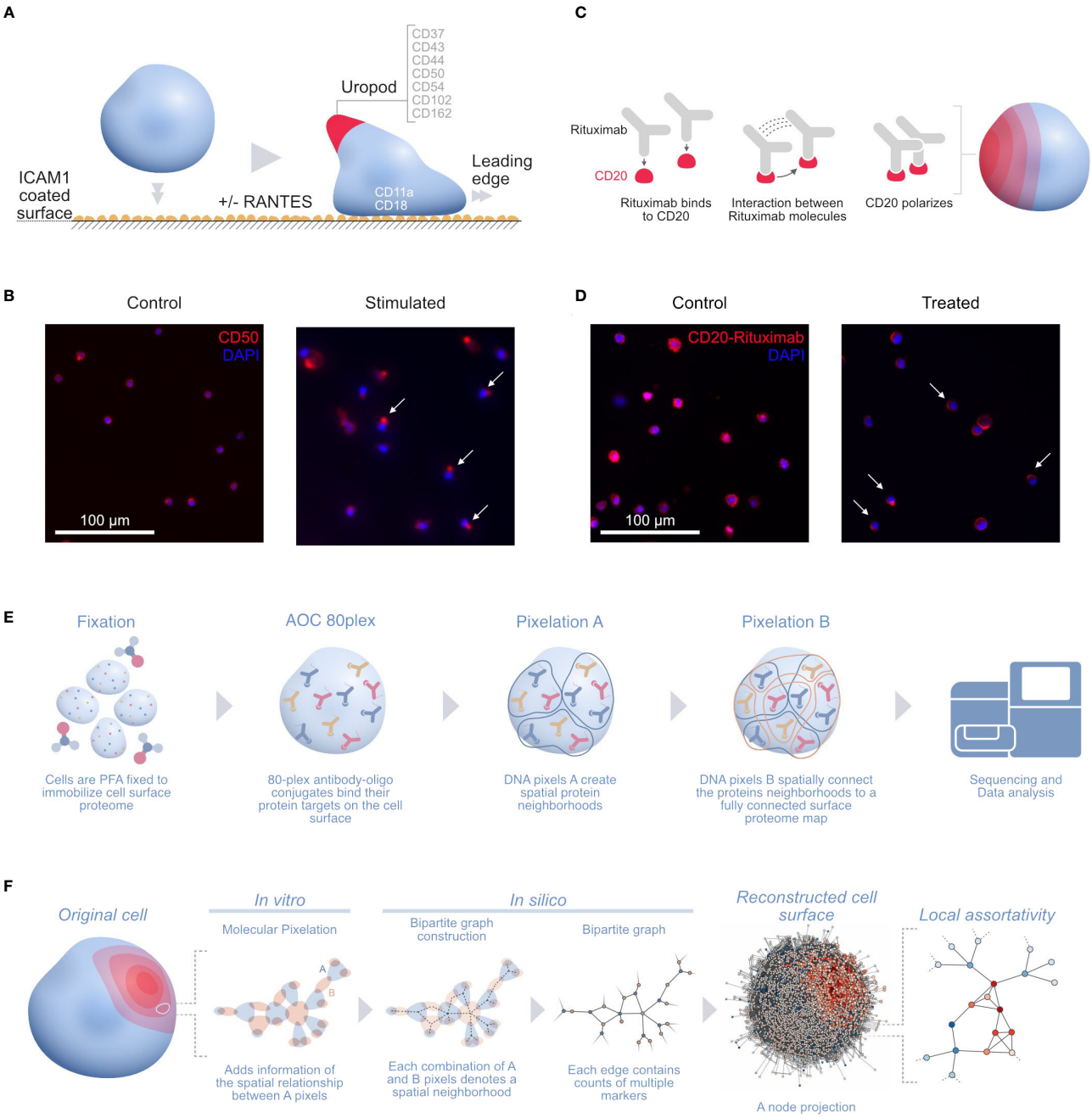


FIGURE 1
(A) Illustration of the uropod formation on the CD54 coated surface and proteins previously associated with T-cell uropods: ICAMs (ICAM1–3: CD54, CD102, and CD50), mucins (CD43 and CD162), and integrins (CD11a/CD18; αLβ2 integrin or LFA-1) (16); CD44 (19, 20); and CD37 (21, 22). **(B)** Widefield immunofluorescence microscopy picture of the uropod formation in both control (left) and stimulated cells (right) with CD50 (red, phycoerythrin fluorophore) and stained nuclei (blue, DAPI). Some of the polarized CD50 pertaining to the uropods upon stimulation are marked by white arrows in the picture. **(C)** Illustration of the stimulated CD20 receptors on the RTX-treated B-cell sample. Here, the RTX monoclonal antibodies interact with each other, thus creating a strong polarization cap on one side of the cell. **(D)** Widefield immunofluorescence microscopy picture of the RTX experiment with control (left) and treated (right) Raji cells with CD20-RTX (red, phycoerythrin fluorophore) and stained nuclei (blue, DAPI). Cells polarized after RTX capping are marked by white arrows in the picture. **(E)** The MPX workflow starts with cell fixation to immobilize the proteome on the surface of the cell, followed by staining with the AOC panel and two steps of Molecular Pixelation before a library is prepared for sequencing. **(F)** Illustration of the cell-to-graph transformation as explained in the MPX study (14). The double MPX workflow step carried out *in vitro* denotes a spatial neighborhood represented by a bipartite graph with AOC molecule counts associated with the edges. This bipartite graph can then be represented into its A-nodes. The A-node projection results in a shift of information from the edges to the vertices. Based on these vertex attributes, we can now compute the (adjusted) local assortativity for each vertex and color the nodes on a scale from assortative (red) to uniform mixing (white) to disassortative (blue), as seen in the last step of the panel (15). ICAMs, intercellular adhesion molecules; RTX, Rituximab; MPX, Molecular Pixelation; AOC, antibody–oligonucleotide conjugate.

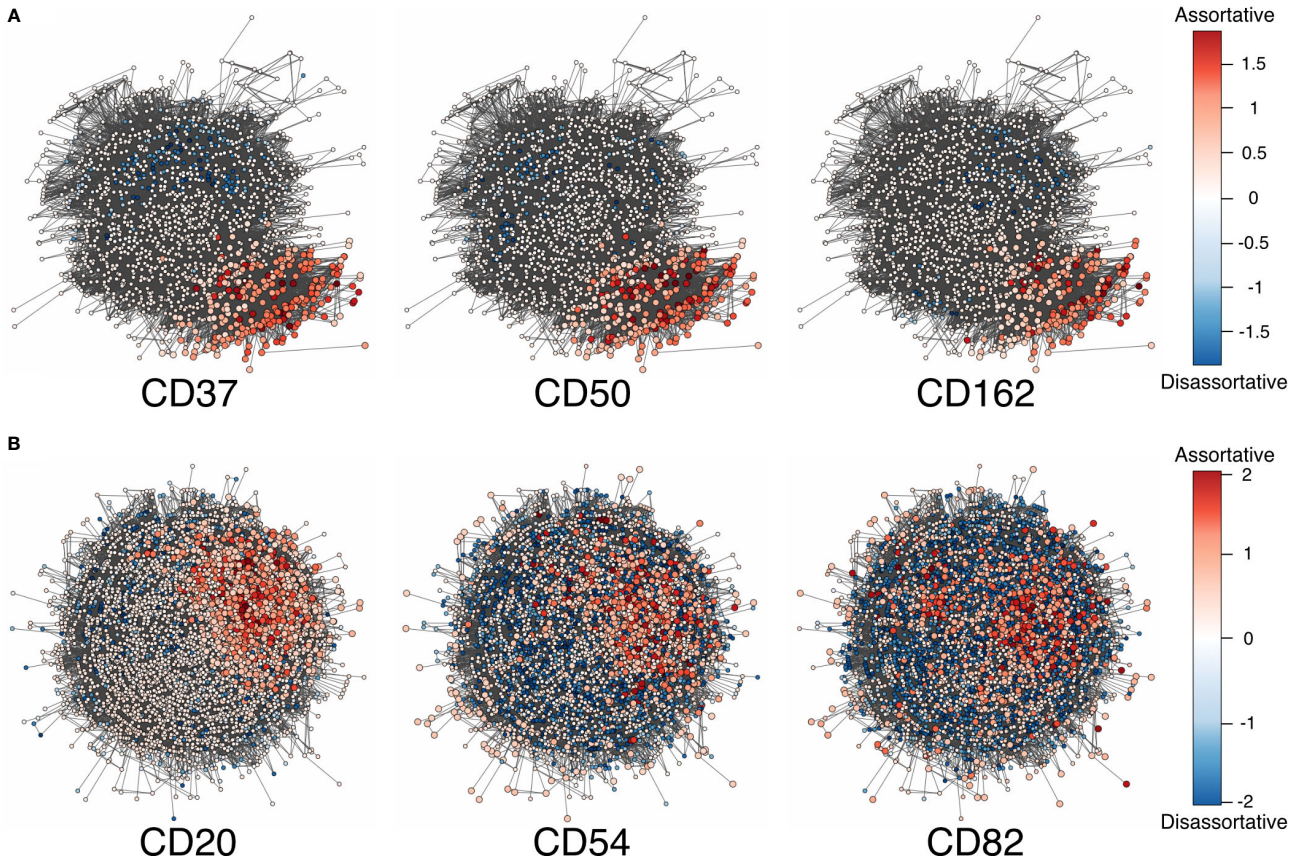


FIGURE 2
(A) Adjusted local assortativity scores for the CD50, CD162, and CD37 displaying the characteristic uropod in one of the cells from the stimulated chemotactic experiment. The color scheme here is a gradient from high local assortativity in red to high local disassortativity in blue with uniform mixing in white. **(B)** Stimulated cells from the RTX-treated experiment where CD20, CD54, and CD82 are colored by the adjusted local assortativity. RTX, Rituximab.

Bound version of local assortativity

The original version of local assortativity defined by Peel and colleagues (15) did not have explicit boundaries, and the distribution of values was not generally comparable across different graphs. Therefore, in this study, we had to improve the score by creating an adjusted version, which improves the general comparison across graphs and values. With this in mind, first, we created a distribution that would have similar maximum and minimum values for the same marker when looking at different cells. These values would imply a boundary for each distribution, allowing us to compare the scores across cells and samples.

To account for the different scores across graphs, we adjusted local assortativity to have zero mean by reweighting the positive and negative scores separately and preserving zero as uniform mixing. This also has the advantage of creating a boundary in both directions. In practice, we therefore compute first the local assortativity as defined by Peel et al. and divide each positive value by the sum of all positive values to normalize the data. The same is done for the negatives by the sum of all negatives.

In mathematical notation, this is equivalent to the following statements.

Let x_j be the unadjusted local assortativity scores as calculated by Peel et al. (15), and then normalized assortativity $f(x_j)$ is defined as

$$f(x_j) = \frac{x_j}{\sum_{i=1, x_i>0}^{n} x_i} \text{ for } x_j \geq 0 \text{ and } \sum_{i=1}^{n} |x_i| > 0,$$

$$f(x_j) = \frac{x_j}{\sum_{i=1, x_i<0}^{n} |x_i|} \text{ for } x_j < 0 \text{ and } \sum_{i=1}^{n} |x_i| > 0$$

and $f(x_j) = 0$ for and $\sum_{i=1}^{n} |x_i| = 0$.

After the first step of normalization, we adjust the created score to have one standard deviation dividing each value of the scores from the previous equation by the standard deviation of the distribution. This results in global upper and lower limits for the normalized local assortativity distribution. Rewriting now the first equation for simpler notation, we get $\widetilde{x_j} = f(x_j)$, which when divided by the standard deviation gives us the normalized standardized local assortativity of the workflow $g(\widetilde{x_j})$.

$$g(\widetilde{x_j}) = \frac{\widetilde{x_j}}{\sum_{i=1}^{n} (\widetilde{x_i} - \mu)^2} = \frac{\widetilde{x_j}}{\sum_{i=1}^{n} \widetilde{x_i}^2} \text{ for } \sum_{i=1}^{n} |\widetilde{x_i}| > 0,$$

and $g(\widetilde{x_j}) = 0$ for $\sum_{i=1}^{n} |\widetilde{x_i}| = 0$.

In order to correct for outliers and homogenize the scale of this distribution, akin to standard single-cell methods (28), we used the log transformation to create a more uniform distribution

(Supplementary Figure 2). Here, we observed similar minima and maxima across multiple proteins, assuring the comparability of the scores across different cells or, more generally, different networks containing attribute information. Therefore, our work improves the previous local assortativity method (15) specifying the upper and lower bounds and enabling representative comparisons and aggregations.

$$h(z_j) = \mathrm{sgn}(z_j) \cdot log(|z_j| + 1)$$

$$\Rightarrow \varrho(x_j) = (h \circ g \circ f)(x_j).$$

By combining all these transformations, we obtained the adjusted version of local assortativity, and proof for its bounds can be found in Supplementary Data 1.2.

Pairwise colocalization

Our aim was to look at any combination of proteins colocalizing, but initially, we created a metric that outputs the colocalization of two proteins by combining the newly introduced adjusted local assortativity measurements. With local assortativity, we had positive and negative values for each node; thus, colocalization would translate to the correlation of vertex values. Therefore, we can apply Spearman's correlation to create a metric that yields the desired colocalization for the two given proteins of interest.

Let n be the number of vertices in the graph, x_i and y_i be the local assortativity scores for two proteins on a vertex in the graph with $0 \leq i \leq n$, and R the rank transformation. Then, the colocalization score of two proteins $X = \{x_0, x_1, \ldots, x_n\}$ and $Y = \{y_0, y_1, \ldots, y_n\}$can be expressed using $\varrho(x)$ and Spearman's correlation (29–31) as

$$coloc(X, Y) = \frac{\sum_{i=1}^{n} R(\varrho(x_i)) \cdot R(\varrho(y_i))}{\sqrt{\sum_{i=1}^{n} R(\varrho(x_i))^2 \cdot \sum_{i=1}^{n} R(\varrho(y_i))^2}}$$

Proteins that failed to pass our filters ("isotype and vertex thresholds") were zeroed for every node in the A-node graph of the cell. Additionally, in the special case that one of the two adjusted local assortativity scores was zero for every node, the pairwise colocalization score would be defined as zero to avoid edge cases with Spearman's correlation. This zeroing in the colocalization measure was well aligned with the local assortativity distribution where random noise could be thought of as a case of uniform mixing.

Proteins measured with AOCs give a relative measurement per cell, making the pairwise local assortativity scores difficult to interpret in terms of absolute values. Therefore, a more robust approach is to compare among experimental conditions, i.e., the uropod-stimulated sample to the control sample. Therefore, we calculated differential colocalization by comparing the scores of the uropod-stimulated sample with both control samples, fixed and in solution. In the RTX experiment, the treated sample was compared to a corresponding control. All statistical tests were performed using Wilcoxon rank sum tests of different contrasts.

Higher-order colocalization

We ultimately aimed to assess proteins that colocalize in groups larger than in pairs. However, we only performed pairwise protein comparisons at the moment. Therefore, a new kind of similarity measure is required to calculate colocalization for multiple proteins. In an ecosystem, multi-species interactions can be measured in multiple sites at the same time using specific scores (32). We adapted this measure to reflect the overlap of local assortativity regions and compare the colocalization of multiple proteins at the same time.

The multiple-site similarity measure (32) is defined as

$$C_s^T = \frac{T}{T-1}\left(\frac{\sum_{i<j}|A_i \cap A_j| - \sum_{i<j<k}|A_i \cap A_j \cap A_k| + \ldots + |A_i \cap \ldots \cap A_T|}{\sum_i |A_i|}\right)$$

$$= \frac{T}{T-1}\left(1 - \frac{\left|\bigcup_{i=1}^{T} A_i\right|}{\sum_{i=1}^{T} |A_i|}\right)$$

In the simplest case, where $T = 2$, this simplifies the Sørensen similarity index (Supplementary Figure 3A) (32, 33).

In the cases where $T \geq 3$, we can apply this similarity measure to protein colocalization of multiple proteins (Supplementary Figure 3B). The multiple-site similarity is well suited for hypothesis testing on the putative colocalization of a group of proteins, as the comparison of proteins increases factorially.

Here, the exact selection of the sets that should be compared was made based on the adjusted local assortativity scores. First, we selected a threshold for the set of nodes we wanted to compare as the scores are numerical values. Given that local assortativity was centered around zero, we decided to select a threshold centered in zero and proceeded with all values with a score greater than zero. Effectively, this means a selection on all the nodes displaying assortativity rather than a random distribution of proteins (i.e., uniform mixing).

Results

Pairwise colocalization

The improved pairwise analysis of protein regions on single cells across the control, fixed control, and stimulated sample (Figure 3) replicates the results found by Karlson and colleagues (14) using different computational methods. Three proteins (CD162, CD37, and CD44) strongly colocalize with CD50 on the uropod (Figure 3). Specifically, we can observe that CD50, CD162 (P-selectin glycoprotein ligand 1 or PSGL1), and CD37, a member of the tetraspanin family, show the highest pairwise colocalization when stimulated with RANTES. The pair CD50 and CD162 is only second to CD45 and CD18 in the stimulated condition. Other known adhesion molecules such as CD102 (ICAM2) did not show a

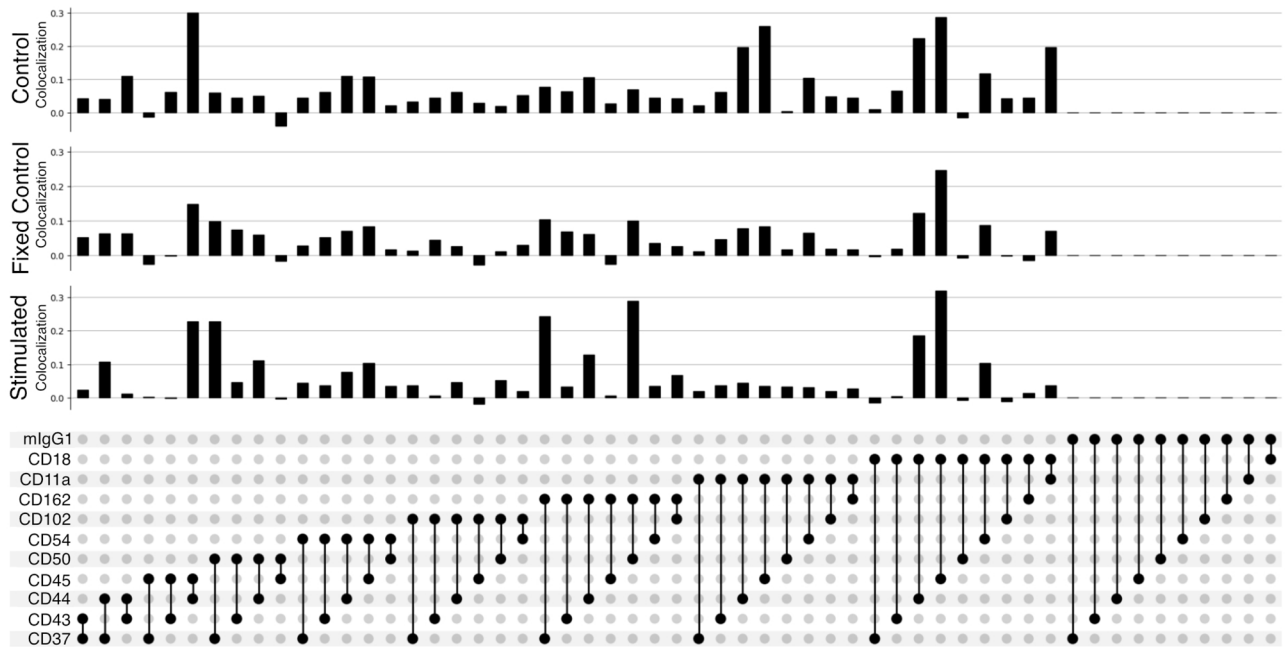


FIGURE 3
Pairwise colocalization of selected proteins shown by an UpSet plot (34). Each barplot represents the colocalization score of two proteins on the control, fixed control, and stimulated samples. The link panel at the bottom shows what pairs of proteins are interrogated in each respective barplot above.

high colocalization with either of these proteins in the stimulated condition. We also noticed that some of the highly abundant protein pairs, such as CD18 and CD45, colocalized in all the conditions: control, fixed control, and stimulated cells (Figure 3). To account for biases toward highly abundant proteins, we assessed whether they could be overcome by employing permutation testing (Supplementary Data 1.3 and 1.4). However, the correction from permutation per vertex is on average less than 1% from the original adjusted score in important uropod proteins (CD50) at the expense of a much longer runtime. Therefore, in the end, we decided to omit permutation testing for the results presented in this paper.

When fixing cells, we expected to observe CD11a/CD18 complex (αLβ2 integrin or LFA-1) binding to the CD54Fc antibody coated in the plates. However, the pairwise colocalization, although present and uniformly mixed, is not as strong as expected (Figure 3).

Furthermore, the pairwise comparison of CD50 with the colocalization of one of the isotype controls (mIgG1) (Figure 3) can be interpreted as background noise level and shows the significance of the colocalization on pairwise combinations among CD50, CD162, and CD37.

Pairwise differential colocalization

Our differential colocalization analysis compares first the scores of the uropod-stimulated sample against the two control samples (Figure 4A and Supplementary Figure 7).

When comparing experimental conditions, a pronounced increase in colocalization of the uropod structural proteins in stimulated cells could be observed when compared to the unstimulated control cells, both fixed and in solution, that cannot be associated with experimental fixation (Figure 4A). This is especially striking in the colocalization scores in all pairwise comparisons of CD50, CD162, and CD37 (Figure 4A) (p-value ≤ 0.0001, Wilcoxon rank test). Otherwise, we observed significant differences (p-value ≤ 0.0001, Wilcoxon rank test) at that level in mean colocalization scores across the three experimental conditions involving one of those three proteins in the pair and highly abundant proteins (HLA-ABC, B2M, CD2, and CD3E). However, the mean difference to the control conditions was small in all those cases (<0.05). By taking CD82 or CD44 (P-glycoprotein 1) proteins combined with CD50, we observed a consistent difference in mean colocalization (>0.05) and very significant at the same time (p-value ≤ 0.0001, Wilcoxon rank test) (Figure 4A).

More interestingly, there were some proteins showing high colocalization with the same sign only in stimulated cells, such as CD26 and CD29 (Figure 4A), compared to the control condition samples (p-value ≤ 0.0001, Wilcoxon rank test). CD29 was also colocalized with HLA-ABC with a mean difference larger than 0.05. The only two proteins with such a significant difference and opposite signs against each contrast were CD18 and CD44.

RTX induces the capping of CD20 on the surface of B cells (35) (23). In our pairwise analysis with CD20, there was a strong increase of colocalization with CD54 (ICAM1) or CD82 when comparing treated and control conditions (Supplementary Figure 8). When compared to the control experiment, those two pair combinations showed a stronger signal than when comparing CD20 and other highly abundant proteins, such as HLA-DR or HLA-ABC/B2M with high significance (p-value ≤ 0.0001, Wilcoxon rank test)

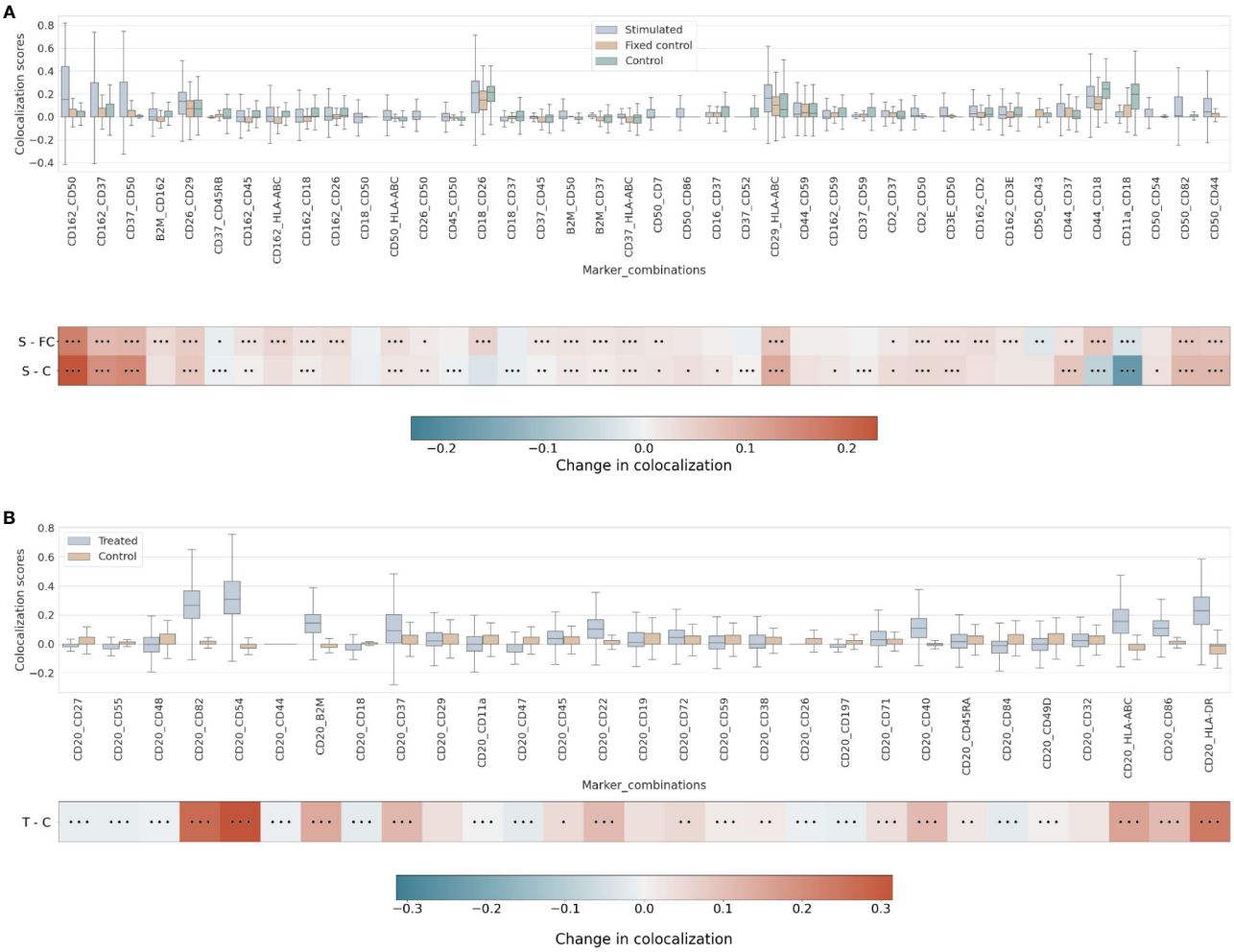


FIGURE 4
(A) Differential pairwise colocalization on a group of selected proteins compared for the three experimental conditions in the uropod experiment (see "Molecular Pixelation datasets" in the Materials and Methods section for a description). First, distributions of all pairwise values per cell are visualized as boxplots per condition side by side. Then, in the first row, differential colocalization between the stimulated (S) and fixed control (FC) samples is measured per pairwise comparison as mean(S) – mean(FC) scores per protein. In the second row, differential colocalization between the stimulated (S) and control (C) samples. The dots indicate p-value ranges generated by a Wilcoxon rank test: 0 dots (p-value > 0.01), 1 dot (0.001< p-value ≤ 0.01), 2 dots (0.0001< p-value ≤ 0.001), and 3 dots (p-value ≤ 0.0001). The scale bar of these differences is shown at the bottom.
(B) Differential pairwise colocalization of CD20 with selected proteins from the treated (T) and control (C) samples of the RTX experiment. Analog to panel A, the differential colocalization is given by the mean difference of both samples, mean(T) – mean(C), and the dots indicate the same p-value changes as before. RTX, Rituximab.

(Figure 4B and Supplementary Figure 9). CD82 is a membrane glycoprotein of the tetraspanin family found associated with both B-cell MHC class II compartments (36) and CD20 in supramolecular complexes (37). Other proteins found with slightly lower pairwise differential colocalization (>0.75) but high significance (p-value ≤ 0.0001, Wilcoxon rank test) were CD37, CD22, CD40, and CD86 (Figure 4B and Supplementary Figure 9).

Higher-order colocalization

When assessing combinations in the stimulated condition of three proteins (i.e., trios), higher-order colocalization allows us to specifically test multiple proteins combined and their colocalization relationship in the same cell graph. Higher-order colocalization was applied to the adjusted local assortativity values of the different experiments and conditions and only calculated on a subset of proteins of interest from the pairwise results shown above in the uropod and RTX experiments.

On the uropod datasets, we observed the highest colocalization score on the stimulated sample among the trio comparison (order of 3) containing CD44, CD45, and CD162 (Supplementary Figure 10). Furthermore, we observed the second-highest colocalization on CD44, CD45, and CD54 (Supplementary Figure 10), which are also well-known uropod proteins (16). One of these proteins (CD44) is specifically involved in the uropod formation (16, 38), and another (CD45) is a widely abundant pan-lymphocyte signaling molecule. Furthermore, all combinations of order 3 containing two out of CD43, CD44, CD50, and CD54 produce high colocalization scores (>0.10) (Supplementary Figure 10).

Strikingly, the combination of CD50 and CD162 with CD44 has one of the highest scores of colocalization in the stimulated sample as well as the largest mean differential colocalization with the controls (Figure 5A and Supplementary Figure 11) on all trios displayed compared to CD50 and CD162 with CD37, which had the highest pairwise colocalization scores between them behind CD18 and CD45 (Figure 3). As expected, these proteins (CD50, CD162, CD44, and CD37) colocalized in the same cellular region

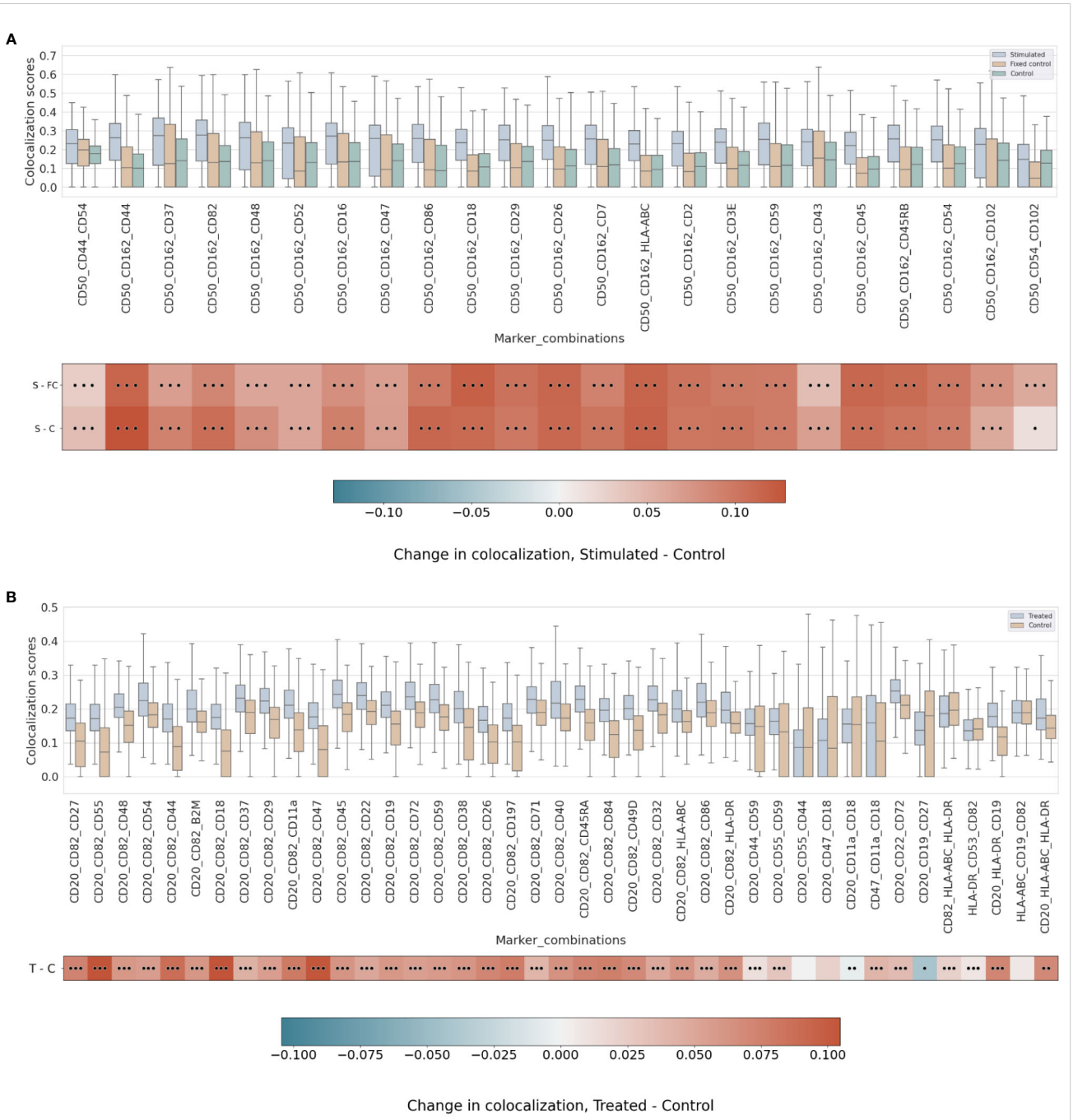


FIGURE 5
(A) Differential higher-order colocalization of trios (order 3) on a group of selected proteins for the three experimental conditions in the uropod experiment (see "Molecular Pixelation datasets" in the Materials and Methods section for a description). First, distributions of all protein comparison values per cell are visualized as boxplots per condition side by side. Then, in the first row, differential colocalization between the stimulated sample (S) and the fixed control (FC) is measured per pairwise comparison as mean(S) – mean(FC) scores. In the second row, differential colocalization between the stimulated (S) and control (C) samples can be found. The dots indicate p-value ranges generated by a Wilcoxon rank test: 0 dots (p-value > 0.01), 1 dot (0.001< p-value ≤ 0.01), 2 dots (0.0001< p-value ≤ 0.001), and 3 dots (p-value ≤ 0.0001). The scale bar of these differences is shown at the bottom. **(B)** Differential higher-order colocalization of CD20 and CD54 with different proteins of relevance on a trio (order 3). The boxplots display the two experimental conditions, RTX-treated (T) and control (C), and compare these by subtracting mean(T) – mean(C) for each protein distribution. The significance of the differences in the differential analysis was analogously computed to panel A using the Wilcoxon rank test and displaying significance using the same dot nomenclature as before. RTX, Rituximab.

(Supplementary Figure 12) and were in alignment with scientific knowledge about the uropod formations (19). However, the CD50, CD162, and CD44 trios demonstrated that our higher-order colocalization method was able to improve scoring even when pairs had shown lower pairwise colocalization scores than others.

In the RTX experiment, by combining pairwise scores into a higher order of 3, the scores of CD20, CD54, and CD82 were expected to be at the top. Surprisingly, our colocalization method for higher orders detects new trios with mean colocalization larger than the score of those three proteins. The combination of CD20 and CD82 with CD45, CD22, CD72, or CD37 in the treated sample produced a higher mean colocalization of order 3 with high significance (p-value ≤ 0.0001, Wilcoxon rank sum) (Figure 5B and Supplementary Figure 13). Interestingly, CD45 or CD37 pairwise colocalized with CD20 had both lower significance and mean differential colocalization to controls compared to CD54 or CD82 with CD20 (Figure 4B). However, when CD45 or CD37 was combined together with CD20 and CD82, they scored two of the 10 highest mean colocalization scores with very high significance (p-value ≤ 0.0001, Wilcoxon rank sum) (Figure 5B). Again, this is another observation that our method was able to improve scoring in cases of pairs with lower pairwise colocalization scores.

Other proteins of interest in the order of 3 that colocalized with CD20 and CD82, albeit with lower colocalization in the treatment but with larger differential mean colocalization (>0.07) and high significance (p-value ≤ 0.0001, Wilcoxon rank sum), were CD55 (DAF), CD44, CD18, CD11a, CD47, CD197 (CCR7), and CD84. DAF regulates the complement system on the cell surface that impairs the formation of the membrane attack complex (MAC), and another protein, CD59, is the MAC-inhibitory protein. CD59 scored higher in colocalization with CD20 and CD82 than CD55, but the mean difference against the control experiment was smaller.

Finally, we calculated the colocalization of order 4 for CD20, CD82, and CD37 with all other non-control proteins (Supplementary Figure 14). Unexpectedly, the three proteins (i.e., CD82, CD54, and CD37) obtaining the highest pairwise colocalization with CD20 (Figure 4B), not counting in the major histocompatibility proteins, were colocalized with high scores in both the control and treatment (>0.20), thus achieving lower significance (0.001< p-value ≤ 0.01, Wilcoxon rank test) (Supplementary Figure 14). Any of CD86, HLA-ABC, or HLA-DR that were high pairwise scoring with CD20 failed to achieve any significance (p-value > 0.01, Wilcoxon rank test) with CD20, CD82, and CD37 when compared to the control experiment (Supplementary Figure 14).

Abundance and colocalization provide different biological aspects

To understand how protein abundance and colocalization measure different aspects of cellular responses to the environment and stimuli, we compared pairwise colocalization and protein abundance as raw molecule counts. We chose different pairs of proteins of interest in both experiments and plotted both the most abundant protein of the pair and the pairwise colocalization averaged across cells. Figure 6 shows no inflation in our pairwise colocalization measurements by abundance.

In the uropod experiment (Figure 6A), all the pairs that we found to be highly colocalized (CD37, CD162, CD50, CD44, and CD54) were not among the highly or mid-abundant proteins. Another highly scoring pair in colocalization, CD29 and HLA-ABC, is also highly colocalized due to the abundance of one of them. However, CD37, CD50, and CD162 colocalization with HLA-ABC was not influenced by its abundance, as they were confined to the bulge of the uropod.

In the RTX experiment (Figure 6B), CD20 was, on average, the third most abundant protein in the Raji cells after HLA-DR and HLA-ABC. Therefore, it was difficult not to perceive dependence on abundance, as CD20 pairwise colocalization was high with those two proteins (>0.10). The other two most colocalized proteins, CD54 and CD82, were the third and fifth most abundant, respectively. Also, B2M, CD40, and CD86 were some of the most abundant proteins with high colocalization to CD20.

Discussion

We analyzed two publicly available MPX experiments with our adjusted local assortativity algorithm for the detection of polarized and colocalized proteins on the surface of single cells.

Cells that were stimulated to form uropods after fixation of PHA-stimulated blasts and treated with RANTES (CCL5) showed a high colocalization score in pairwise comparison of proteins associated with the uropod (CD50, CD162, and CD37) (16). Notably, a member of the tetraspanin family, CD37, has been described as playing a role in the cytoskeleton remodeling of actin filaments but has never colocalized with other uropod proteins such as CD50 or CD162 (21). CD37 is necessary for leukocytes to follow a CXCL1 chemotactic gradient as tested in CD37-deficient mice (39).

On the attachment side of the stimulated cells, αLβ2 integrin (CD11a/CD18 or LFA-1) pairwise colocalization is not as significant as expected. This is mostly due to the low abundance of CD11a, which is often at the threshold level of control isotypes and leads, therefore, to generally lower scores in the stimulated cells. It is possible that the experimental conditions by fixing CD54 coating and posterior cleavage by enzymatic reaction may have affected the protein complex structure as well as epitope availability of CD11a.

Intercellular adhesion molecules (ICAMs) are arguably some of the best-annotated proteins in migrating immune cells (16). At a higher order of magnitude, we found that ICAMs scored much more significantly at order 3 and beyond. However, our colocalization method was able to distinguish that ICAM1–3 (CD54, CD102, and CD50) together at order 3 were not highly significant compared to the control. It is possible that ICAMs selectively group together and become more structurally

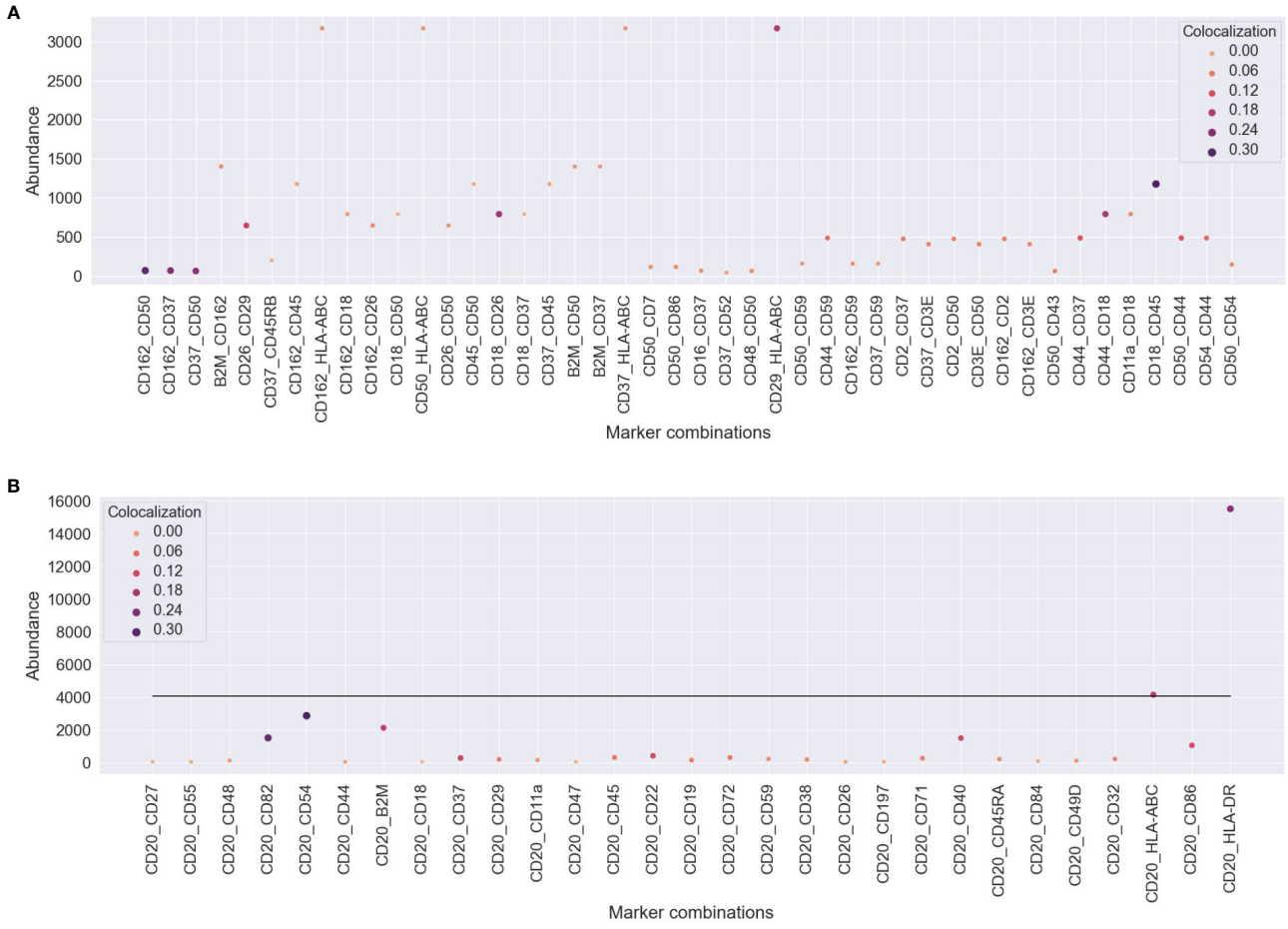


FIGURE 6
Comparison of pairwise colocalization and abundance on the proteins of interest for both experiments. **(A)** In the uropod stimulation experiment, the abundance is given by the maximum number of counts of the two compared proteins (y-axis), and colocalization is given as Spearman's correlation (size and color). Some protein pairs indicate that high colocalization may be found occasionally when one of the proteins is highly abundant, e.g., CD29 and HLA-ABC, but not always, e.g., CD37 and HLA-ABC. **(B)** For the RTX-treated sample, all comparisons are made between CD20 and the proteins of interest. The abundance axis reflects the counts of the proteins of interest, and the line shows the averaged abundance of CD20 across all treated cells. RTX, rituximab.

significant in a larger cell membrane area that punctuates the colocalization of other proteins in pairs. Direct colocalization of CD18 with ICAM1 or with ICAM3 in *trans*-interactions has been reported through microscopy (40), but we cannot discard that *cis*-interactions may occur in our migratory model system. It has been observed that β2 integrin bending on human neutrophils rolling on a microfluidic device coupled to advanced microscopy facilitates interaction with ICAMs in *cis*-, thus inhibiting leukocyte adhesion *in vitro* and *in vivo* (41). On that system, they are able to prove that ICAM3 is the dominant LFA-1 ligand in *cis*- and that inhibition of the interaction between Mac-1 (C11b/CD18) and ICAM1 in *cis*- limits significantly neutrophil accumulation.

Pairwise colocalization signals on CD26 and CD29 have been reported in healthy mouse myofibroblasts in the past (42). Being present in most cell types, CD26 plays a double functionality as an immune-regulatory and proteolytic enzyme. CD26 can be found integral to both the membrane and its soluble form (43). This multifunctional protein is able to influence T-cell proliferation and chemotaxis but also truncate RANTES and alter the sub-receptor specificity of the cleaved chemokine (44). CD26 has a key role in adhesion and invasion for several cancer cells and has therefore become an established cell surface marker in serum (45). The extracellular matrix (ECM) is able to provide cells with co-stimulatory signals through different receptor–ligand interactions. Collagen has been described to provide proliferation signals to CD4 + cells via the CD3 pathway with the mediation of VLA-3 (CD49c/CD29) and CD26 receptors (46). Different adhesion factors of the very late activation antigen (VLA) family, sharing a common β1 subunit (CD29 or ITGB1), are able to receive signals either directly or indirectly to different proteins of the ECM and CD26 to collagen type I, IV, and fibronectin (47). Furthermore, on the pairwise colocalization effect of CD29 with HLA-ABC, certain isotypes of HLA-B are able to decrease ITGB1 expression and affect pancreatic cancer cell migration with contrasting effects (48).

CD44 is a transmembrane glycoprotein presenting ubiquitous expression and is able to bind to several ECM proteins (49). Some sources suggest that CD44 and CD18 may colocalize to mediate lymphocyte rolling and adhesion (50) and that CD44 interacts with the β2 subunit (CD18) of the LFA-1 integrin in lymphocytes (51) (52) and in colon cancer cells (53). CD44 is known to be expressed on cancer stem cells and implicated in many cancers as a marker of tumor burden and metastatic potential due to its numerous variant isoforms (49). Also, CD44 is a signaling partner in relation to cell growth, survival, and differentiation (54). As a therapeutic target, CD44 has held some promise in the past, e.g., anti-CD44 mAb therapy in breast cancer xenografts, reducing tumor growth and relapse post-chemotherapy (55). Despite recent disappointments in late-phase trials (56), still, new avenues are explored, e.g., nanoparticles (57) or carbon nanotubes (58), and hope remains on CD44 as a target as well as on better stratification of the patient population (56).

RTX is one of the pioneer biological therapies effective in many B-cell malignancies, such as chronic lymphocytic leukemias, non-Hodgkin's, and Burkitt's lymphomas. The human IgG1 Fc portion of RTX is capable of activating several mechanisms to cause cell death: complement-dependent cytotoxicity (CDC), complement-dependent cellular cytotoxicity, antibody-dependent cellular phagocytosis, or antibody-dependent cellular cytotoxicity (59). The relative killing efficiencies of RTX have been well studied *in vitro*, but the *in vivo* precise mechanism of action remains elusive (60), and better understanding is still needed to impede disease relapsing. In order to design for improved effects, different IgG subtypes have been engineered and studied both *in vitro* with Ramos cells (61) and lymphoma B-cell organoids (62).

Our analysis shows that, upon RTX treatment, CD55 and CD59 are colocalized with CD20 via the CD82 tetraspanin, whereas the direct pairwise colocalization with CD20 of both proteins was not significantly differentiated from controls. It suggests that CD55 and CD59 are indirectly associated with CD20 via a tetraspanin network, resembling the CD46 association with many β1 integrins and tetraspanins (63). This may indicate that targeting inhibitors of CDC may achieve superior killing, as it has been suggested by others (64).

Our data also support that CD82, but not CD9, colocalizes with CD19 and CD20 (65). Unfortunately, at the time of writing, some important proteins that play a crucial part in the CD20 therapeutic "enigma" (59) are not present on the current MPX AOC panel, among them, CD46, another complement inhibitory component; CD21 (CR2), the complement C3d receptor; and CD81 (TAPA-1), another tetraspanin. The trio of proteins, CD21, CD19, and CD81, form the CR2–CD19–CD81 complex, often called the B-cell co-receptor complex that enhances BCR signaling (66).

We also found another tetraspanin, CD37, suggested to be part of a multicomponent supramolecular complex, so-called "tetraspans-DR complexes". After solubilization of membranes of human B-cell lines and tonsillar B cells, seven components were discovered by co-precipitation together with HLA-DR antigens: four of the tetraspanins present in B cells (CD37, CD53, TAPA-1, and R2/C33), as well as CD19 and CD21 (67). The same laboratory employed later another technique, flow cytometric energy transfer, to find three tetraspan molecules (CD53, CD81, and CD82) complexed with MHC class I, MHC class II, and CD20 on the surface of a human B-cell line (37). Recently, CD20 and CD37 have been confirmed to form a complex by a proximity ligation assay (68). In this preprint, it is hypothesized that the presence of CD20 stabilizes CD37 in the cell membrane as increased internalization of anti-CD37 is measured on deficient CD20 lymphoma B-cell lines (68).

The potential of CD37 as a therapeutic target has been recognized by developing biparatopic antibodies with engineered Fc chains that form IgG hexamers (69) and, in clinical trials (NCT01317901), exploring combinatorial therapies for relapsed patients and good overall response rate (70). Bobrowicz and colleagues recently tested that upon diminished levels of CD37 in different cell lines, even with downregulation of CD20, cytotoxicity of CAR-T cells was not significantly impaired. Therefore, in their opinion, CD37 remains an attractive therapeutic target (68).

Overall, we want to highlight the complexity and dynamism of the cellular membrane driven by tetraspanins, integrins, and adhesion molecules. We find several molecules in common to both datasets that colocalize together upon very different stimuli. Tetraspanin-enriched microdomains facilitate the compartmentalization of specialized receptors and adhesion molecules in membrane domains that connect to the underlying intracellular architecture of the cell (71, 72).

One of the main caveats of colocalization analysis is the difference between the abundance and true signal. CD20 is the third molecule in mean abundance in the RTX experiment and presents high pairwise colocalization with, e.g., HLA-DR. These macro-complexes have been well described in the literature (37), but highly abundant proteins may colocalize with all other proteins by chance. In the uropod experiment, the highly colocalizing pairs have low mean abundances compared to the highly abundant proteins distributed uniformly across the area of the cell, e.g., HLA-ABC, whereas the CD20 cap on a Raji cell after RTX treatment is likely a much larger area than the smaller and well-constrained uropod bulge and, also, more prone to accidental overlap by low- and high-abundance proteins. While these are two very different cellular responses and biological systems, the area of polarization and overlap may warrant different interpretations.

Even though the local assortativity is improving on this by taking the spatial aspect of the graph into account, there is still some bias toward abundant proteins on the cell surface. Although the interesting signal in our method is likely going to be indicated by low abundant proteins showing high colocalization, inflation of colocalization scores at higher orders of comparison has been observed, and strategies to ameliorate this may use corrections from lower orders of comparison (i.e., correct scores on order 3 with scores from order 2). We think that the major confounding factor to colocalization measurements is abundant proteins. Improvements to our method in this area may consider the idea of richness of species when comparing values across sites (73). However, we have shown how the abundance and colocalization of proteins measure different aspects of cell biology, both equally important.

MPX capabilities in throughput, sensitivity, and three-dimensional field of view (14) create exceptional conditions to study protein constellations at single-cell resolution. Here, we expand on the concept of local assortativity as presented in the Peel et al. study (15) for labeled nodes in networks to capture the influence of both the structure of the cell represented in the graph and the location distribution of each protein. Adjusted local assortativity provides a parameter-free algorithm that calculates the colocalization of molecules, avoiding the complex problem of finding neighborhoods of clustered features by different approaches (74, 75).

Local assortativity could be used with other single-cell technologies and experimental designs. However, it is required for it to work in that data were processed in the form of networks with labels and features in the vertices. With the MPX technology, cells and their protein molecules are modeled in the form of graphs. With this method, we improved the global scores provided in the Karlsson et al. study in terms of polarization and colocalization to find continuous and bounded measures of the biological phenomena assayed by MPX. Furthermore, we built on the pairwise local assortativity using a multi-site similarity method used from an ecological context (32) in order to achieve multiple protein comparisons from groups of more than two proteins. With pairwise and multi-way comparison, we aimed to capture detailed structural properties of the cell graphs and facilitate the comparison of molecules colocalizing among groups of proteins in a more detailed way. The scoring methods devised for pairwise and higher-order colocalization are different, so a direct comparison of score levels across those two is not possible at the moment. The higher order of colocalization is very useful for hypothesis testing, but we foresee its use rather as a tool for specific *in silico* experiments guided by prior knowledge.

MPX with pairwise and a higher order of colocalization yields deep phenotyping not achievable with other assays by measuring 76 proteins (and four controls) at the same time in a single experimental workflow. This is a throughput of several orders of magnitude higher in plexity at a reduced experimental time from sample to processed data over what can be achieved with confocal and super-resolution microscopy. The experiments presented in this report took approximately 2 weeks to complete. However, given the exponential number of protein combinations, there is a clear need to develop algorithms and tools to exploit this novel data type.

Another advantage of MPX data is that they achieve single-cell resolution in one experiment of 300–1,000 cells to study variability in response to stimuli. We have measured effect sizes against controlled experiments but have not explored yet the complexity of responses in terms of dividing treatment and controls in different subgroups. We should also consider the combinatorial multiplexity of single-cell studies, and the comparison of experimental conditions is therefore a vital tool in the analysis of colocalization, which begins with a good study design.

Our findings underpinned by MPX together with our novel computational method may provide avenues for hypothesis-driven therapeutic design that explores spatially colocalized protein constellations in the cell.

Data availability statement

Pixelator output data from MPX libraries are made publicly available under a Creative Commons Share Alike License at https://software.pixelgen.com/datasets. Accompanying analysis and data for the purpose of this paper is released under GPL-2 license at https://github.com/PixelgenTechnologies/adjusted-local-assortativity-paper. Pixelator is available under MIT license at https://github.com/PixelgenTechnologies/pixelator and the pipeline orchestrator is released as a nf-core workflow at https://github.com/nf-core/pixelator under MIT license.

Ethics statement

Only publicly available MPX datasets were used for this study. The microscopy pictures shown in Figure 1 were obtained in the Karlsson et al. (14) study, where all the ethics statements can be found and cell lines providers are listed.

Author contributions

JR: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. MK: Writing – review & editing. DT: Writing – review & editing. TK: Writing – review & editing. FK: Writing – review & editing. SF: Writing – review & editing. JD: Conceptualization, Writing – review & editing. AB: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The presented work was funded by the Wellcome Leap ΔTissue Program and Stiftelsen för Strategisk Forskning, SSF.

Acknowledgments

We would like to express our most sincere gratitude to Leto Peel for licensing the code under an MIT license[5] and helping to understand the code of the original local assortativity paper (15). Furthermore, we would like to thank him for extending the code and integrating the permutation testing, even if in the end it was not used for the data in this paper.

5 https://github.com/piratepeel/MultiscaleMixing.

Conflict of interest

All authors are current or past employees of Pixelgen Technologies AB commercializing products based on Molecular Pixelation.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1309916/full#supplementary-material

References

1. Rees P, Summers HD, Filby A, Carpenter AE, Doan M. Imaging flow cytometry. *Nat Rev Methods Primers*. (2022) 2:1–13. doi: 10.1038/s43586-022-00167-x

2. Sekar RB, Periasamy A. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J Cell Biol*. (2003) 160:629–33. doi: 10.1083/jcb.200210140

3. Navikas V, Leitao SM, Grussmayer KS, Descloux A, Drake B, Yserentant K, et al. Correlative 3D microscopy of single cells using super-resolution and scanning ion-conductance microscopy. *Nat Commun*. (2021) 12:4565. doi: 10.1038/s41467-021-24901-3

4. Spitzer MH, Nolan GP. Mass cytometry: single cells, many features. *Cell*. (2016) 165:780–91. doi: 10.1016/j.cell.2016.04.019

5. Hosogane T, Santana LS, Eling N, Moch H, Bodenmiller B. Compressed sensing expands the multiplexity of imaging mass cytometry. *bioRxiv*. (2023). doi: 10.1101/2023.11.06.565119

6. Kuett L, Catena R, Özcan A, Plüss ACancer Grand Challenges IMAXT Consortium, , Schraml P, et al. Three-dimensional imaging mass cytometry for highly multiplexed molecular and cellular mapping of tissues and the tumor microenvironment. *Nat Cancer*. (2022) 3:122–33. doi: 10.1038/s43018-021-00301-w

7. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*. (2017) 14:865–8. doi: 10.1038/nmeth.4380

8. Liu Y, DiStasio M, Su G, Asashima H, Enninful A, Qin X, et al. High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nat Biotechnol*. (2023) 41:1405–9. doi: 10.1038/s41587-023-01676-0

9. Walker LM, Shiakolas AR, Venkat R, Liu ZA, Wall S, Raju N, et al. High-throughput B cell epitope determination by next-generation sequencing. *Front Immunol*. (2022) 13:855772. doi: 10.3389/fimmu.2022.855772

10. Vistain L, Van Phan H, Keisham B, Jordi C, Chen M, Reddy ST, et al. Quantification of extracellular proteins, protein complexes and mRNAs in single cells by proximity sequencing. *Nat Methods*. (2022) 19(12):1578–89. doi: 10.1038/s41592-022-01684-z

11. Lundberg E, Borner GHH. Spatial proteomics: a powerful discovery tool for cell biology. *Nat Rev Mol Cell Biol*. (2019) 20:285–302. doi: 10.1038/s41580-018-0094-y

12. Paul I, White C, Turcinovic I, Emili A. Imaging the future: the emerging era of single-cell spatial proteomics. *FEBS J*. (2021) 288:6990–7001. doi: 10.1111/febs.15685

13. Dayao MT, Brusko M, Wasserfall C, Bar-Joseph Z. Membrane marker selection for segmenting single cell spatial proteomics data. *Nat Commun*. (2022) 13:1999. doi: 10.1038/s41467-022-29667-w

14. Karlsson F, Kallas T, Thiagarajan D, Karlsson M, Schweitzer M, Navarro JF, et al. Molecular pixelation: spatial proteomics of single cells by sequencing. *Nat Methods*. (2024). doi: 10.1038/s41592-024-02268-9

15. Peel L, Delvenne J-C, Lambiotte R. Multiscale mixing patterns in networks. *Proc Natl Acad Sci U.S.A*. (2018) 115:4057–62. doi: 10.1073/pnas.1713019115

16. Sánchez-Madrid F, Serrador JM. Bringing up the rear: defining the roles of the uropod. *Nat Rev Mol Cell Biol*. (2009) 10:353–9. doi: 10.1038/nrm2680

17. Mcfarland W, Heilman DH. Lymphocyte Foot Appendage: Its Role in Lymphocyte Function and in Immunological Reactions. *Nature*. (1965) 205:887–8. doi: 10.1038/205887a0

18. Hind LE, Vincent WJB, Huttenlocher A. Leading from the back: the role of the uropod in neutrophil polarization and migration. *Dev Cell*. (2016) 38:161–9. doi: 10.1016/j.devcel.2016.06.031

19. Krüger K, Büning C, Schriever F. Activated T lymphocytes bind in *situ* to stromal tissue of colon carcinoma but lack adhesion to tumor cells. *Eur J Immunol*. (2001) 31:138–45. doi: 10.1002/1521-4141(200101)31:1%3C138::aid-immu138%3E3.0.co;2-p

20. Mrass P, Kinjyo I, Ng LG, Reiner SL, Puré E, Weninger W. CD44 mediates successful interstitial navigation by killer T cells and enables efficient antitumor immunity. *Immunity*. (2008) 29:971–85.

21. Yeung L, Hickey MJ, Wright MD. The many and varied roles of tetraspanins in immune cell recruitment and migration. *Front Immunol*. (2018) 9:1644. doi: 10.3389/fimmu.2018.01644

22. Draber P, Vonkova I, Stepanek O, Hrdinka M, Kucova M, Skopcova T, et al. SCIMP, a transmembrane adaptor protein involved in major histocompatibility complex class II signaling. *Mol Cell Biol*. (2011) 31:4550–62.

23. Rudnicka D, Oszmiana A, Finch DK, Strickland I, Schofield DJ, Lowe DC, et al. Rituximab causes a polarization of B cells that augments its therapeutic function in NK-cell-mediated antibody-dependent cellular cytotoxicity. *Blood*. (2013) 121:4694–702. doi: 10.1182/blood-2013-02-482570

24. Newman MEJ. Assortative mixing in networks. *Phys Rev Lett*. (2002) 89:208701. doi: 10.1103/PhysRevLett.89.208701

25. Newman MEJ. Mixing patterns in networks. *Phys Rev E Stat Nonlin Soft Matter Phys*. (2003) 67:026126. doi: 10.1103/PhysRevE.67.026126

26. Khanam KZ, Srivastava G, Mago V. The homophily principle in social network analysis: A survey. *Multimed Tools Appl*. (2023) 82:8811–54. doi: 10.1007/s11042-021-11857-1

27. Citraro S, Rossetti G. Identifying and exploiting homogeneous communities in labeled networks. *Appl Netw Sci*. (2020) 5:1–20. doi: 10.1007/s41109-020-00302-1

28. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol*. (2019) 15:e8746. doi: 10.15252/msb.20188746

29. Spearman C. The proof and measurement of association between two things. *Am J Psychol*. (1904) 15:72–101. doi: 10.2307/1412159

30. de Winter JCF, Gosling SD, Potter J. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychol Methods*. (2016) 21:273–90. doi: 10.1037/met0000079

31. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. (2020) 17:261–72. doi: 10.1038/s41592-020-0772-5

32. Diserud OH, Odegaard F. A multiple-site similarity measure. *Biol Lett*. (2007) 3:20–2. doi: 10.1098/rsbl.2006.0553

33. Sorensen T. A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons. Biologiske Skrifter/Kongelige Danske Videnskabernes Selskab (1948). scientific research publishing. Available at: https://www.scirp.org/(S(351jmbntvnsjt1aadkozje))/reference/referencespapers.aspx?referenceid=2200146 (Accessed July 13, 2023).

34. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: visualization of intersecting sets. *IEEE Trans Vis Comput Graph*. (2014) 20:1983–92.

35. Beers SA, Chan CHT, French RR, Cragg MS, Glennie MJ. CD20 as a target for therapeutic type I and II monoclonal antibodies. *Semin Hematol*. (2010) 47:107–14. doi: 10.1053/j.seminhematol.2010.01.001

36. Hammond C, Denzin LK, Pan M, Griffith JM, Geuze HJ, Cresswell P. The tetraspan protein CD82 is a resident of MHC class II compartments where it associates with HLA-DR, -DM, and -DO molecules. *J Immunol*. (1998) 161:3282–91. doi: 10.4049/jimmunol.161.7.3282

37. Szöllósi J, Horejsí V, Bene L, Angelisová P, Damjanovich S. Supramolecular complexes of MHC class I, MHC class II, CD20, and tetraspan molecules (CD53, CD81, and CD82) at the surface of a B cell line JY. *J Immunol*. (1996) 157:2939–46. doi: 10.4049/jimmunol.157.7.2939

38. Mempel TR, Marangoni F. CD44 keeps tumor killers polarized. *Immunity*. (2008) 29:843–5. doi: 10.1016/j.immuni.2008.12.004

39. Wee JL, Schulze KE, Jones EL, Yeung L, Cheng Q, Pereira CF, et al. Tetraspanin CD37 regulates β2 integrin-mediated adhesion and migration in neutrophils. *J Immunol*. (2015) 195:5770–9. doi: 10.4049/jimmunol.1402414

40. Costantini C, Calzetti F, Perbellini O, Micheletti A, Scarponi C, Lonardi S, et al. Human neutrophils interact with both 6-sulfo LacNAc+ DC and NK cells to amplify NK-derived IFN{gamma}: role of CD18, ICAM-1, and ICAM-3. *Blood*. (2011) 117:1677–86. doi: 10.1182/blood-2010-06-287243

41. Fan Z, McArdle S, Marki A, Mikulski Z, Gutierrez E, Engelhardt B, et al. Neutrophil recruitment limited by high-affinity bent β2 integrin binding ligand in cis. *Nat Commun*. (2016) 7:12658. doi: 10.1038/ncomms12658

42. Haas MR, Nguyen DV, Shook BA. Recovery of altered diabetic myofibroblast heterogeneity and gene expression are associated with CD301b+ Macrophages. *Biomedicines*. (2021) 9(12):1752. doi: 10.3390/biomedicines9121752

43. Gorrell MD, Gysbers V, McCaughan GW. CD26: a multifunctional integral membrane and secreted protein of activated lymphocytes. *Scand J Immunol*. (2001) 54:249–64. doi: 10.1046/j.1365-3083.2001.00984.x

44. Mentlein R. Dipeptidyl-peptidase IV (CD26)–role in the inactivation of regulatory peptides. *Regul Pept*. (1999) 85:9–24. doi: 10.1016/S0167-0115(99)00089-0

45. Havre PA, Dang LH, Ohnuma K, Iwata S, Morimoto C, Dang NH. CD26 expression on T-anaplastic large cell lymphoma (ALCL) line Karpas 299 is associated with increased expression of versican and MT1-MMP and enhanced adhesion. *BMC Cancer*. (2013) 13:517. doi: 10.1186/1471-2407-13-517

46. Dang NH, Torimoto Y, Schlossman SF, Morimoto C. Human CD4 helper T cell activation: functional involvement of two distinct collagen receptors, 1F7 and VLA integrin family. *J Exp Med*. (1990) 172:649–52. doi: 10.1084/jem.172.2.649

47. Gorski A, Kupiec-Weglinski JW. Extracellular matrix proteins, regulators of T-cell functions in healthy and diseased individuals. *Clin Diagn Lab Immunol*. (1995) 2:646–51. doi: 10.1128/cdli.2.6.646-651.1995

48. Sliker BH, Goetz BT, Barnes R, King H, Maurer HC, Olive KP, et al. HLA-B influences integrin beta-1 expression and pancreatic cancer cell migration. *Exp Cell Res*. (2020) 390:111960. doi: 10.1016/j.yexcr.2020.111960

49. Senbanjo LT, Chellaiah MA. CD44: A multifunctional cell surface adhesion receptor is a regulator of progression and metastasis of cancer cells. *Front Cell Dev Biol*. (2017) 5:18. doi: 10.3389/fcell.2017.00018

50. Bonder CS, Clark SR, Norman MU, Johnson P, Kubes P. Use of CD44 by CD4+ Th1 and Th2 lymphocytes to roll and adhere. *Blood*. (2006) 107:4798–806. doi: 10.1182/blood-2005-09-3581

51. Koopman G, van Kooyk Y, de Graaff M, Meyer CJ, Figdor CG, Pals ST. Triggering of the CD44 antigen on T lymphocytes promotes T cell adhesion through the LFA-1 pathway. *J Immunol*. (1990) 145:3589–93. doi: 10.4049/jimmunol.145.11.3589

52. Vermot-Desroches C, Wijdenes J, Valmu L, Roy C, Pigott R, Nortamo P, et al. A CD44 monoclonal antibody differentially regulates CD11a/CD18 binding to intercellular adhesion molecules CD54, CD102 and CD50. *Eur J Immunol*. (1995) 25:2460–4. doi: 10.1002/eji.1830250908

53. Fujisaki T, Tanaka Y, Fujii K, Mine S, Saito K, Yamada S, et al. CD44 stimulation induces integrin-mediated adhesion of colon cancer cell lines to endothelial cells by up-regulation of integrins and c-Met and activation of integrins. *Cancer Res*. (1999) 59:4427–34.

54. Ponta H, Sherman L, Herrlich PA. CD44: from adhesion molecules to signalling regulators. *Nat Rev Mol Cell Biol*. (2003) 4:33–45. doi: 10.1038/nrm1004

55. Marangoni E, Lecomte N, Durand L, de Pinieux G, Decaudin D, Chomienne C, et al. CD44 targeting reduces tumour growth and prevents post-chemotherapy relapse of human breast cancers xenografts. *Br J Cancer*. (2009) 100:918–22. doi: 10.1038/sj.bjc.6604953

56. Sahin IH, Klostergaard J. CD44 as a drug delivery target in human cancers: where are we now? *Expert Opin Ther Targets*. (2015) 19:1587–91. doi: 10.1517/14728222.2015.1088834

57. Zhang M, Ma Y, Wang Z, Han Z, Gao W, Zhou Q, et al. A CD44-targeting programmable drug delivery system for enhancing and sensitizing chemotherapy to drug-resistant cancer. *ACS Appl Mater Interf*. (2019) 11:5851–61. doi: 10.1021/acsami.8b19798

58. Jain Singhai N, Ramteke S. CNTs mediated CD44 targeting; a paradigm shift in drug delivery for breast cancer. *Genes Dis*. (2020) 7:205–16. doi: 10.1016/j.gendis.2019.10.009

59. Pavlasova G, Mraz M. The regulation and function of CD20: an "enigma" of B-cell biology and targeted therapy. *Haematologica*. (2020) 105:1494–506. doi: 10.3324/haematol.2019.243543

60. Weiner GJ. Rituximab: mechanism of action. *Semin Hematol*. (2010) 47:115–23. doi: 10.1053/j.seminhematol.2010.01.011

61. Könitzer JD, Sieron A, Wacker A, Enenkel B. Reformatting rituximab into human IgG2 and IgG4 isotypes dramatically improves apoptosis induction in vitro. *PloS One*. (2015) 10:e0145633. doi: 10.1371/journal.pone.0145633

62. Lara S, Heilig J, Virtanen A, Kleinau S. Exploring complement-dependent cytotoxicity by rituximab isotypes in 2D and 3D-cultured B-cell lymphoma. *BMC Cancer*. (2022) 22:678. doi: 10.1186/s12885-022-09772-1

63. Lozahic S, Christiansen D, Manié S, Gerlier D, Billard M, Boucheix C, et al. CD46 (membrane cofactor protein) associates with multiple beta1 integrins and tetraspans. *Eur J Immunol*. (2000) 30:900–7. doi: 10.1002/1521-4141(200003)30:3<900::AID-IMMU900>3.3.CO;2-O

64. Harjunpää A, Junnikkala S, Meri S. Rituximab (anti-CD20) therapy of B-cell lymphomas: direct complement killing is superior to cellular effector mechanisms. *Scand J Immunol*. (2000) 51:634–41. doi: 10.1046/j.1365-3083.2000.00745.x

65. Horváth G, Serru V, Clay D, Billard M, Boucheix C, Rubinstein E. CD19 is linked to the integrin-associated tetraspans CD9, CD81, and CD82. *J Biol Chem*. (1998) 273:30537–43. doi: 10.1074/jbc.273.46.30537

66. Janeway CA Jr, Travers P, Walport M, Shlomchik MJ. Antigen receptor structure and signaling pathways. *Garland Sci*. (2001).

67. Angelisová P, Hilgert I, Horejsí V. Association of four antigens of the tetraspans family (CD37, CD53, TAPA-1, and R2/C33) with MHC class II glycoproteins. *Immunogenetics*. (1994) 39:249–56. doi: 10.1007/BF00188787

68. Bobrowicz M, Kusowska A, Krawczyk M, Slusarczyk A, Barankiewicz J, Domagala J, et al. CD20 expression regulates CD37 levels in B-cell lymphoma – implications for immunotherapies. *bioRxiv*. (2023). doi: 10.1101/2023.12.06.570441

69. Oostindie SC, van der Horst HJ, Kil LP, Strumane K, Overdijk MB, van den Brink EN, et al. DuoHexaBody-CD37®, a novel biparatopic CD37 antibody with enhanced Fc-mediated hexamerization as a potential therapy for B-cell Malignancies. *Blood Cancer J*. (2020) 10:30. doi: 10.1038/s41408-020-0292-7

70. Gopal AK, Tarantolo SR, Bellam N, Green DJ, Griffin M, Feldman T, et al. Phase 1b study of otlertuzumab (TRU-016), an anti-CD37 monospecific ADAPTIRTM therapeutic protein, in combination with rituximab and bendamustine in relapsed indolent lymphoma patients. *Invest New Drugs*. (2014) 32:1213–25. doi: 10.1007/s10637-014-0125-2

71. Rocha-Perugini V, Sánchez-Madrid F, Martínez Del Hoyo G. Function and dynamics of tetraspanins during antigen recognition and immunological synapse formation. *Front Immunol*. (2015) 6:653. doi: 10.3389/fimmu.2015.00653

72. Saiz ML, Rocha-Perugini V, Sánchez-Madrid F. Tetraspanins as organizers of antigen-presenting cell function. *Front Immunol*. (2018) 9:1074. doi: 10.3389/fimmu.2018.01074

73. Baselga A, Jiménez-Valverde A, Niccolini G. A multiple-site similarity measure independent of richness. *Biol Lett*. (2007) 3:642–5. doi: 10.1098/rsbl.2007.0449

74. Premachandran, Kakarala. (2013). Consensus of k-NNs for robust neighborhood selection on graph-based manifolds, in: Portland, OR, USA: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1594–601.

75. Levet F, Julien G, Galland R, Butler C, Beghin A, Chazeau A, et al. A tessellation-based colocalization analysis approach for single-molecule localization microscopy. *Nat Commun*. (2019) 10:2379. doi: 10.1038/s41467-019-10007-4

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 19 July 2024
DOI 10.3389/fimmu.2024.1425466

OPEN ACCESS

EDITED BY
Veronica Anania,
Genentech Inc., United States

REVIEWED BY
Alexander Mann,
Harvard Medical School, United States
Jennifer L. Larson-Casey,
University of Alabama at Birmingham,
United States

*CORRESPONDENCE
Alessandro Venosa
alessandro.venosa@pharm.utah.edu

RECEIVED 29 April 2024
ACCEPTED 28 June 2024
PUBLISHED 19 July 2024

CITATION
Moos PJ, Cheminant JR, Cowman S, Noll J, Wang Q, Musci T and Venosa A (2024) Spatial and phenotypic heterogeneity of resident and monocyte-derived macrophages during inflammatory exacerbations leading to pulmonary fibrosis.
Front. Immunol. 15:1425466.
doi: 10.3389/fimmu.2024.1425466

Spatial and phenotypic heterogeneity of resident and monocyte-derived macrophages during inflammatory exacerbations leading to pulmonary fibrosis

Philip J. Moos, Jenna R. Cheminant, Sophie Cowman, Jessica Noll, Qiuming Wang, Teresa Musci and Alessandro Venosa*

Department of Pharmacology and Toxicology, University of Utah College of Pharmacy, Salt Lake City, UT, United States

Introduction: Genetic mutations in critical nodes of pulmonary epithelial function are linked to the pathogenesis of pulmonary fibrosis (PF) and other interstitial lung diseases. The slow progression of these pathologies is often intermitted and accelerated by acute exacerbations, complex non-resolving cycles of inflammation and parenchymal damage, resulting in lung function decline and death. Excess monocyte mobilization during the initial phase of an acute exacerbation, and their long-term persistence in the lung, is linked to poor disease outcome.

Methods: The present work leverages a clinical idiopathic PF dataset and a murine model of acute inflammatory exacerbations triggered by mutation in the alveolar type-2 cell-restricted Surfactant Protein-C [SP-C] gene to spatially and phenotypically define monocyte/macrophage changes in the fibrosing lung.

Results: SP-C mutation triggered heterogeneous $CD68^{+}$ macrophage activation, with highly active peri-injured cells relative to those sampled from fully remodeled and healthy regions. Ingenuity pathway analysis of sorted $CD11b^{-}SigF^{+}CD11c^{+}$ alveolar macrophages defined asynchronous activation of extracellular matrix re-organization, cellular mobilization, and Apolipoprotein E (*Apoe)* signaling in the fibrosing lung. Cell-cell communication analysis of single cell sequencing datasets predicted pro-fibrogenic signaling (*fibronectin/Fn1, osteopontin/Spp1,* and *Tgfb1*) emanating from *Trem2/TREM2*$^{+}$ interstitial macrophages. These cells also produced a distinct lipid signature from alveolar macrophages and monocytes, characterized by *Apoe* expression. Mono- and di-allelic genetic deletion of ApoE in SP-C mutant mice had limited impact on inflammation and mortality up to 42 day after injury.

Abbreviations: AT2, alveolar type-2 cell; $Sftpc^{I73T}$ surfactant protein-C I73T mutant; IPF, idiopathic pulmonary fibrosis; AIE, acute inflammatory exacerbations; AMs, alveolar macrophages; MoDMs, monocyte-derived macrophages; ApoE, apolipoprotein-E; FN1, fibronectin1; SPP1, osteopontin.

10.3389/fimmu.2024.1425466

Discussion: Together, these results provide a detailed spatio-temporal picture of resident, interstitial, and monocyte-derived macrophages during SP-C induced inflammatory exacerbations and end-stage clinical PF, and propose ApoE as a biomarker to identify activated macrophages involved in tissue remodeling.

KEYWORDS

alveolar type-2 cell, surfactant protein-C I73T mutant, pulmonary fibrosis, alveolar macrophages, monocyte-derived macrophages, apolipoprotein-E, fibronectin1, osteopontin1

Introduction

Acute inflammatory exacerbations represent a key feature in the evolution of interstitial lung diseases (ILD). This cluster of chronic progressive pathologies includes idiopathic pulmonary fibrosis, non-specific interstitial pneumonia, connective tissue disease associated ILD, chronic hypersensitivity pneumonitis, pneumoconiosis, sarcoidosis, and more (1, 2). The exact sequence of events igniting these flare-ups have yet to be fully understood. However, epidemiological and experimental evidence suggest that a combination of genetic susceptibility, preexistent pulmonary and systemic conditions, biological aging/senescence of the alveolar compartment, and external stressors (environmental toxic exposure) contribute to the repeated cycles of focal inflammation, spatially heterogenous injury, and aberrant repair characteristic of fibrotic disease (3–7).

To date, over 60 mutations in the alveolar epithelial type 2 specific gene encoding for the surfactant protein C (SP-C) have been linked to the development of a fibrotic phenotype, with the isoleucine-to-threonine missense substitution at position 73 of the SP-C proprotein (SP-C^{I73T}) representing the most common (8–10). Our group has previously characterized endoplasmic reticulum toxicity and macroautophagy block resulting from aberrant processing and trafficking of the SP-C proprotein and described the sequelae of immunological events accompanying tissue remodeling (8, 9, 11, 12). Here, we expand on this prior investigation by defining the spatial and phenotypic distribution of monocytes and macrophages responding to the initial epithelial injury and establish patterns of intercellular communication among cellular species in the lung.

Mounting evidence supports the notion that inflammatory monocyte mobilization in the fibrosing lung represents a valid indicator of poor disease prognosis (13, 14). Yet, experimental modeling and clinical trials designed to non-specifically target inflammation (e.g., corticosteroids, broad spectrum cytokine modulation) have revealed low efficacy or even harmful toxicities (15). Among the reasons for this therapeutic failure is the absence of a nuance approach capable of controlling the maturation, activation, and persistence of ontologically and phenotypically heterogeneous cellular entities, and the relatively fragmented understanding of the spatial distribution of cells and signals in a temporally extended pathology such as fibrosis (16–18). Experimental evidence highlights distinct transcriptomes emanating from tissue-resident and monocyte-derived macrophages (19, 20), with the latter generating a complex fibrogenic signature (10, 21, 22). The fibronectin/FN1, osteopontin/SPP1, tumor growth factor (TGF)β1, and interleukin 4/13 signaling pathways represent the most studied networks mediating tissue remodeling (23–27), while metabolic networks have gained traction as potential targets in chronic inflammatory diseases (28, 29). *In vitro* systems establish a reliance on glycolysis in response to canonical pro-inflammatory signals (IFNγ, LPS), juxtaposed to fatty acid oxidation, tricarboxylic acid cycle, and mitochondrial oxidative phosphorylation following challenge with anti-inflammatory and pro-remodeling signals (IL-10, IL-4/13, TGFβ1) (30–32). This evidence emphasizes the importance of factors governing lipid synthesis, handling, and metabolism (*PPARγ, LXR, FXR, SREBP1*) in regulating macrophage function (33, 34). The cholesterol and phospholipid transporter apolipoprotein E (ApoE) has been linked to monocyte-derived macrophage activation in chemical-induced injury, though there is limited evidence that these responses are consistent across the spectrum of fibrosis (35, 36).

Through a combination of bulk, single cell, and spatial transcriptomics we show a hyperactive niche of macrophages surrounding fully remodeled lung regions during fibrogenic exacerbations triggered by mutant SP-C induction. Cellular annotation and communication analysis identify time-related changes in intercellular networks, with alveolar and interstitial/*Trem2*$^{+}$ macrophages and inflammatory monocytes responsible for *Spp1, Fn1*, and *Tgfβ1* pro-fibrotic signaling. Analysis of human idiopathic PF confirms the presence of an interstitial population responsible for pro-fibrotic signaling in the diseased lung. Our findings also highlight disease-related shifts in lipid transcriptional signatures and interstitial/*Trem2*$^{+}$ macrophages as the sole cellular cluster expressing ApoE. While genetic ablation of ApoE in SP-C^{I73T} induced injury did not significantly impact fibrotic disease outcome, our results pinpoint this molecule as a potential biomarker identifying fibrogenic myeloid populations. Together, these results reveal temporal and phenotypic heterogeneity in the macrophage compartment and implicates *Trem2*$^{+}$ interstitial macrophages and their monocytic precursors as viable targets for anti-fibrogenic therapy.

Materials and methods

Sex as a biological variable

These studies utilized both male and female mice.

Murine model of SP-C^{I73T} induced lung injury

Tamoxifen inducible SP-C^{I73T} mice were generated as previously reported (9). For studies investigating the role of the apoliproprotein E in SP-C^{I73T} induced injury, a parallel line was crossed with ApoE knock out mice purchased from Jackson laboratories (Strain #002052, The Jackson Laboratory, Bar Harbor, ME). Briefly, an estrogen receptor (ER)-2 controlled Flp-O recombinase strain knocked into the Rosa26 locus (The Jackson Laboratory). Adult homozygote SP-C^{I73T}Flp mice (8-12 weeks) received two tamoxifen oral gavages three days apart (90 mg/kg each) to excise a neomycin cassette placed within the *Sftpc* gene. Both male and female animals were used for the studies. For studies involving apoE mutants, a SP-C^{I73T}ApoE$^{WT/KO}$ breeding pair was utilized (SP-C^{I73T}ApoE^{HET}) so as to generate wild type, heterozygous, and homozygous experimental littermates. Control groups mice are represented as pooled data from tamoxifen treated SP-C^{I73T} not expressing Flp-O recombinase or oil (vehicle) treated Flp-O expressing SP-C^{I73T} mice. All mice were housed under pathogen free conditions in AALAC approved barrier facilities at the Skaggs College of Pharmacy, University of Utah. All experiments were approved by the Institutional Animal Care and Use Committee, University of Utah.

Reagents

Tamoxifen (non-pharmaceutical grade) was purchased from Sigma-Aldrich (St Louis, MO). Giemsa cytological stain was purchased from Sigma-Aldrich. Antibody list: *Spp1* (RNAscope® Probe Green, Ref#435191); *Tgfb1* (RNAscope® Probe Red, Ref# 407751-C2); *Apoe* (for *in situ* hybridization – Advanced Cell Diagnostics, RNAscope® LS 2.5 Probe #313278; for immunohistochemical - Abcam; Cat #ab183597, 1:500; for western blot – Cell Signaling Technology, Cat #49285, 1:1000), CD68 for immunohistochemistry (Cell Signaling Technology; Cat # 97778; 1:1000). Flow cytometric panel for cell sorting and bulk sequencing of macrophages CD16/32 (clone 93; eBiosciences, San Diego, CA), CD11b (clone # M1/70; eFluo450, eBiosciences); Fixable Viability dye (Cat # 65-0865-14; eFluo780, eBiosciences); SigF (clone S17007L; PE-CF594, BD Biosciences, San Jose, CA); CD45 (clone 30-F11; PerCP5.5, Biolegend, San Diego, CA); CD11c (clone # N418; BV705, Biolegend); Ly6G (clone # 1A8; AF700, Biolegend); CD64 (clone X54-5/7.1; PE/Cy7, Biolegend); CD3 (clone # 17A2; BUV395, Biolegend). All other reagents were purchased from Thermo Fisher Scientific, Inc. (Waltham, MA) or Sigma-Aldrich.

Bronchoalveolar lavage, cell counts, ELISA, and western blot analysis

Following terminal anesthesia, inert mice were subject to cannulation. Bronchoalveolar lavage (BAL) fluid was collected from mice using 1 mL sterile saline lavage and collected into a microcentrifuge tube. Four additional lavages were performed and collected in a separate container. The two fractions were then spun at 400 × g, 6 min. Supernatant from the first lavage was collected and immediately frozen at -80°C for ELISA and western blot analysis, while the two cellular pellets were combined and suspended in 1 ml of saline solution for cell counts, flow cytometric, or RNA sequencing analysis. BAL cells were enumerated using a NucleoCounter (New Brunswick Scientific, Edison, NJ). Aliquots of first lavage were analyzed for IL-4 and IL-13 levels using the Luminex platform (Panel MCYTOMAG-70K-17) following Thermo Fisher's protocol. For western blot, equal volumes (15 µl) of thawed BAL fluid were loaded onto 4-12% NuPage Bis-Tris gels (ThermoFisher Scientific) with NuPage 4X LDS sample buffer (ThermoFisher Scientific) and then electrophoresed approximately 90 minutes using a constant voltage of 100V. Proteins from the gels were then transferred to a 0.45 µm PVDF membrane at 30V and 4°C for one hour and blocked in 5% non-fat dried milk (NFDM). The membranes were subsequently probed with primary ApoE antibodies. The SuperSignal West Dura Chemiluminescent Substrate detection system was applied before exposing the membrane on the ProteinSimple FluorChem M imager (BioTechne).

Histology, histochemical and *In situ* hybridization analysis

For histological and histochemical analysis, lungs from unresponsive anesthetized mice were cleared of excess blood through cardiac perfusion with 0.9% sodium chloride solution. A 20-gauge cannula was inserted in the trachea for tissue fixation with 10% neutral buffer formalin at constant pressure (25 cm H_2O). A suture was used to seal the tracheal opening upon cannula removal, thus avoiding pulmonary deflation during the fixation process. Tissue was placed in a histology cassette and submerged in 10% neutral buffer formalin for 72 h. The suture was then removed, and the lung sequentially moved to a 2% sucrose solution (in PBS, two washes of 5 min) and 70% ethyl alcohol. The submerged cassettes were submitted to the University of Utah histology core (Associated Regional and University Pathologists Inc.) for embedding. Paraffin blocks were sectioned at 6 µm thickness and used for Hematoxylin & Eosin (H&E) or immunohistochemical staining, alone or in combination with *in situ* hybridization as previously described (37). For protein staining, paraffin was removed using xylene solutions followed by gradient alcohol washes (100-50%). Citrate antigen retrieval (10.2 mM sodium citrate, pH 6.0, for 20 min) and endogenous peroxidase quenching with 3% hydrogen peroxide in methanol (30 min) were performed. Serum-based blocking (10% goat serum in PBS) preceded the overnight incubation with anti-rabbit primary antibody. In all studies, a serum/IgG control was used. During the second day, slides

underwent incubation with a biotinylated secondary antibody (30 minutes, Vectastain Elite ABC kit, Vector Labs, Burlingame, CA) and chromogenic reaction achieved using a Peroxidase Substrate Kit DAB (Vector Labs). Counterstain was accomplished with Harris Modified Hematoxylin (Thermo Fisher Scientific, Inc.). For *in situ* hybridization studies, after paraffin removal with xylene/alcohol solutions, slides were air dried. Peroxidase quenching (10 min, away from light) was followed by antigen retrieval (20 min, RNAscope® Target Retrieval Reagent, ACD) and incubation with protease IV (30min, RNAscope® Protease IV Reagent, ACD). Excess solution was then washed off. Slides were then incubated for 2 hours in a 40° C hybridization oven with a *Spp1*, *Tgfb1, or Apoe* probe. A series of signal amplification steps (6 for single color detection, 10 for double-staining assay) and washes were followed by chromogen development. At this point the experiment was either concluded with counterstain and toluene-based permount coverslip placement, or the immunohistochemistry protocol resumed from the blocking and primary antibody step as described above.

Fluorescence activated cell sorting

In some studies, following cardiac perfusion the left lobe was tied off with a suture and removed for flow cytometric and FACS analysis. Tissue was minced with surgical scissors and transferred into a 50 ml conical tube and incubated (37°C, 30 minutes) in DMEM + 5% FBS + 2 mg/ml Collagenase D (Cat #11088866001, Roche, Indianapolis, IN). Digested lungs were passed through 70-µm nylon mesh to obtain a single-cell suspension, counted and mixed with ACK Lysis Buffer (Thermo Fisher Scientific, Inc.) to remove any remaining red blood cells. The single cell suspension was counted and resuspended to yield 1 x 10^6 cells per 100µl of flow cytometry staining buffer (PBS+0.1% sodium azide). Cells were then incubated with anti-mouse CD16/32 antibody for 10 min at 4°C to block nonspecific binding. This was followed by 30-minute incubation with fluorescently-tagged antibodies or appropriate isotype controls (0.25–1.5 µg/10^6 cells) for 30 minutes (4°C). Cells were then spun and resuspended in staining buffer for viability staining (30 minutes at 4°C). Cells were fixed in 2% paraformaldehyde and sorted using a FACS ARIA (BD Biosciences). Alveolar macrophages (SigF$^+$CD11b$^-$ CD11c$^+$) were identified following forward and side scatter selection of singlet CD45$^+$ viable cells. To ensure cell sorting of a purified population, a series of exclusion gates were designed to remove eosinophils (SigF^{int}CD11b$^+$CD11c$^-$), neutrophils (Ly6G$^+$) and lymphocytes (CD3$^+$). All analysis was performed using FlowJo software (FlowJo, LLC, Ashland, Oregon).

Bulk and single-cell RNA sequencing preparation and analysis

For bulk RNA sequencing studies (deposited on NCBI GEO GSE166300), sorted SigF$^+$CD11b$^-$CD11c$^+$ macrophages underwent RNA extraction using Qiagen RNeasy Plus Universal mini kit following manufacturer's instructions (Qiagen, Hilden, Germany). Extracted RNA samples were quantified using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and RNA integrity was checked using Agilent TapeStation 4200 (Agilent Technologies, Palo Alto, CA, USA). RNA sequencing libraries were prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina following manufacturer's instructions (NEB, Ipswich, MA, USA). Briefly, mRNAs were first enriched with Oligo(dT) beads. Enriched mRNAs were fragmented for 15 minutes at 94°C. First-strand and second strand cDNAs were subsequently synthesized. cDNA fragments were end-repaired and adenylated at 3'ends, and universal adapters were ligated to cDNA fragments, followed by index addition and library enrichment by limited-cycle PCR. The sequencing libraries were validated on the Agilent TapeStation (Agilent Technologies, Palo Alto, CA, USA), and quantified by using Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) as well as by quantitative PCR (KAPA Biosystems, Wilmington, MA, USA). The sequencing libraries were pooled and clustered on 1 lane of a flow cell. After clustering, the flowcell was loaded on the Illumina HiSeq4000 instrument according to manufacturer's instructions. The samples were sequenced using a 2x150bp Paired End (PE) configuration. Image analysis and base calling were conducted by the HiSeq Control Software (HCS). Raw sequence data (bcl files) generated from Illumina HiSeq was converted into fastq files and de-multiplexed using Illumina's bcl2fastq 2.17 software. One mismatch was allowed for index sequence identification. Analysis of RNA counts was performed using R (3.6.3) (38). Differential gene expression analysis was conducted using the hciR package (39). Further pathway analysis was conducted using IPA (QIAGEN Inc). Datasets were filtered using log2 fold change (minimum -1 or 1) and p-value cut offs (minimum $p<0.05$) to ensure an appropriate number of molecules (200-3000) were used in the IPA pipeline.

Single-cell RNA sequencing raw data of SP-C^{I73T} model were deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession numbers GSE247520 and GSE196657. Tissue collection and single cell suspension were achieved using mechanical mincing, digestion in Collagenase D, red blood cell lysis, and suspension created by using a 70-µm strainer. RNA extraction and library preparation are described in the published manuscript (10). Mining of human control and IPF lungs are accessible through GSE136831 (40). As described by Adams et al., representative apical and basal segments of explanted lungs were minced mechanically, digested [elastase (30 U/ml) + deoxyribonuclease I (0.2 mg/ml) + liberase (0.3 mg/ml) + 1% penicillin/streptomycin], cleared of red blood cells, and single cell suspension created using a 100/70/40 strainers (40). Re-analysis of each single-cell dataset included dimension reduction and clustering by SCTransformation (0.3.5) using the Gamma-Poisson generalized linear model method (glmGamPoi, 1.8.0) and were performed using the Seurat (4.0.4) package (41–45). Multiple levels of resolution were evaluated using Clustree (0.5.0). The data was assessed for cell cycle effects using CellCycleScoring and regressed for uneven cell cycle expression across clusters. Cell types were identified using differential gene expression and all manual annotations were compared to those produced through automated classification using SingleR (1.10.0). A specific R package was used to interface with enrichR database (46). For pseudotime analysis, monocle-3 software was used, while cell-cell communication analysis was conducted with CellChat software (47–51).

Spatial transcriptomic analysis

For spatial analysis (deposited on NCBI GEO GSE264128), neutral buffered formalin-fixed lungs were inflated without bronchoalveolar lavage. Paraffin-embedded sections were baked for 1 hour at 60°C and stained according to the NanoString Leica BOND RX RNA FFPE Semi-Automatic protocol. Following Proteinase K (1.0ug/mL, 15 min) and ER2 (20 min) processing, sections were incubated with fluorescently tagged antibodies against PanCytokeratin (PanCK, Novus Biologics, NBP2-33200AF488; 1:400), Syto83 (S11364; 1:10, Thermo Fisher Scientific, Inc.), CD45 (Nanostring Technologies, 121302304; 1:5), and CD68 (Abcam, ab125212; 1:25). Regions of interest (ROI; N = 3, healthy; N = 4, peri-injured, N = 5, injured/remodeled) were selected based on histopathological assessment of inflammation and epithelial thickening and remodeling. Tissue segmentation selectively identified $CD45^+CD68^+$ macrophages. The collection plate was then removed from the GeoMx instrument and prepared for sequencing. The GeoMx library was prepared, processed, and sequenced according to the NanoString NGS Readout User Manual Protocol (ref. MAN-10153-04). Sequencing was performed on the Illumina NovaSeq 6000, S2 v1.5 with a 100 cycles flowcell at 27 bp pair-end reads. Generated FASTQ files were then processed to DCC files utilizing the NanoString GeoMx NGS Pipeline according to manufacturer's instructions. Gene expression was analyzed using Nanostring DSP analysis software (NanoString Technologies) with built-in statistical analyses. Raw data counts were run through internal quality control and biological probe quality control.

Statistical analysis

Unless otherwise indicated, all data are presented as group mean ± SE. Statistical analyses were performed with Prism GraphPad 9.4 (GraphPad Software, San Diego, CA). Student's t-tests were used for paired data; for analyses involving multiple groups, one-way or two-way analysis of variance (ANOVA) was performed with *post hoc* testing as indicated. Survival analyses were performed using Log Rank (Mantel-Cox) test. For spatial analysis, segments were filtered to 55% of the limit of quantitation (LOQ) to render the top 6,000 expressed genes. Filtered genes were normalized to Q3 (3^{rd} quartile of all selected targets). Hierarchical clustering was performed as quality control. CD68 ROIs were compared across regions using a linear mixed model (LMM) with Benjamini-Hochberg (BH) correction and a random effect for the region. For RNA sequencing pathway analysis, Wilcoxon rank sum test was used. In all cases, statistical significance was considered at $p \leq 0.05$.

Results

Spatial discrimination of activated macrophages during SP-C^{I73T}-induced injury

To spatially define the phenotype of macrophages involved in fibrotic lung injury, fluorescent antibodies recognizing DNA (green), CD45 (yellow) and CD68 (red) were used to perform bulk sequencing in the injured/remodeled or peri-injured alveolar regions of the lung 14 days post SP-C^{I73T} injury. A control (non-induced) lung was used to define baseline activation. Identification of injured regions was based on histopathological evaluation of all five lobes in the tissue sections. Four peri-injured areas were selected: two in relative proximity to the injury (areas of interest 006 and 010) and two distal region (areas of interest 009 and 012) (Figure 1A). Cell-type deconvolution analysis was used to resolve monocyte/macrophage specific expression signature from the dataset (52) (Supplementary Table 1). Principal component analysis identified clustering of transcriptomes based on sampling annotation, with separation in the peri-injured regions based on proximity to injury (Figure 1B; Supplementary Table 1). Global pathway map showed extensive expression of genes involved in metabolism (including RNA metabolism), signal transduction, transcriptional regulation, immune system function and cell cycle (Supplementary Figure 1A). Reactome-based analysis revealed 20.9% (148/712 pathways) and 34.9% (249/712 pathways) of pathways as differentially expressed when comparing peri-injury vs. healthy macrophages and peri-injury vs. injured cells, respectively. Notably, macrophages isolated form injured regions did not produce a particularly strong signaling signature and demonstrated a considerable degree of similarity to healthy macrophages (≈90% of pathways) (Figure 1C; Supplementary Figures 1B–D). Peri-injured macrophages were enriched in pathways associated with neutrophil degranulation, ROS/RNS production and release, activation in oxidative stress-induced senescence, pro-remodeling functions (TGF-beta and GPVI cascade), and metabolic alterations (citric acid cycle, gluconeogenesis, lipoprotein assembly/remodeling/clearance) compared to healthy- and injured-derived $CD68^+$ macrophages (Figures 1D, E; Supplementary Table 1). When compared to controls, injured region macrophages were defective in programmed cell death signaling, ROS detoxification, and elastic fiber formation, but displayed enhanced DNA damage-induced senescence, degradation of extracellular matrix (ECM), and metalloproteinase function (Figure 1F). Hierarchical gene expression analysis of pathways related to degradation of ECM (Figure 1G), assembly of collagen fibrils (Figure 1H), complement system, adaptive immunity, and L13a-mediated ceruloplasmin expression all showed robust clustering based on region of origin (Supplementary Figures 1B–D, Supplementary Table 1).

Among the top 6,000 genes expressed genes in the dataset, genes associated with complement responses (*C1qa, C1qb, C1qb*), myeloid cell recruitment (*Cxcl14, Cxcl15*), inflammation (*Csf1r, Ctss, App*) and metalloproteinases *(Mmp13, Mmp14, Mmp15*) were most abundant in peri-injured macrophages (Figure 1I). Despite increases in normalized counts for fibrogenic genes and significant enrichments in the associated pathways, expression of the individual genes did not meet significance (*Fn1, Spp1*) (Figure 1J; Supplementary Figures 1E–G).

Temporal heterogeneity of resident alveolar macrophages during SP-C^{I73T}-induced injury

To better understand macrophage behavior in the events leading up to end-stage fibrosis, we sought to examine transcriptional

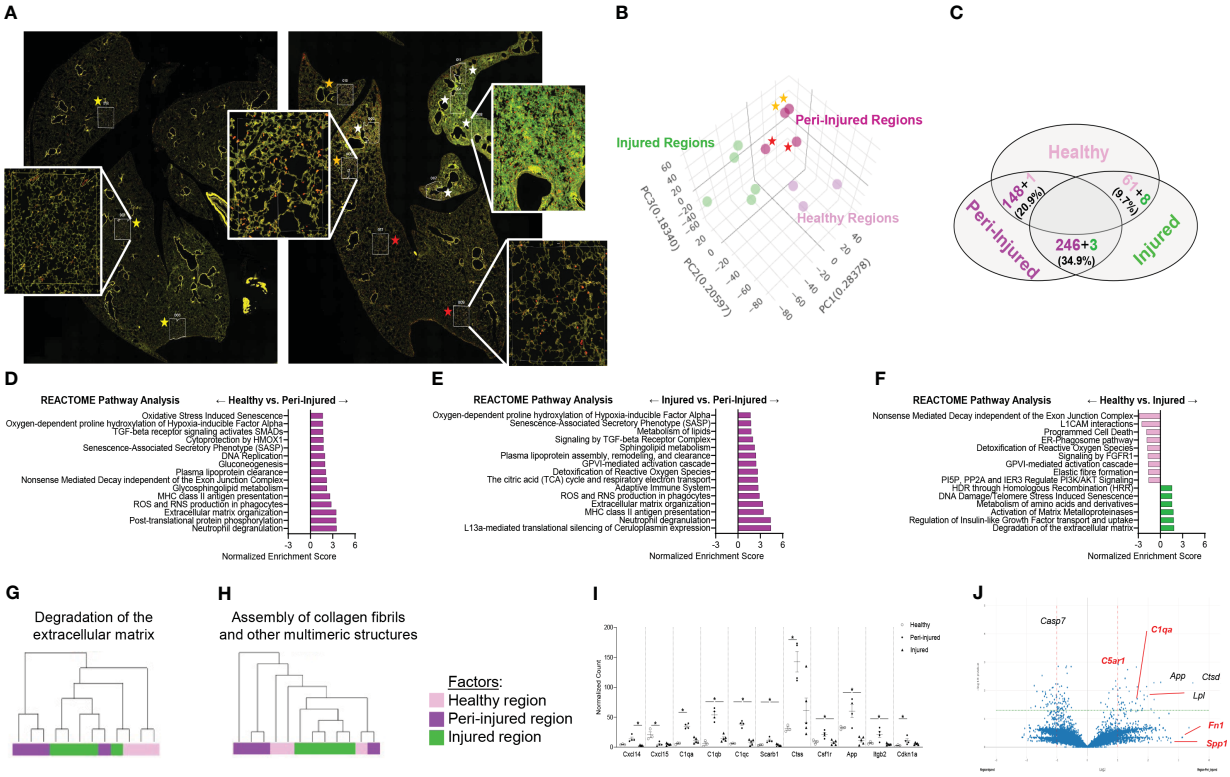


FIGURE 1
Spatial analysis of CD68$^+$ macrophages in SP-C^{I73T} induced injury. Paraffin-embedded sections from controls or 14 days post SP-C^{I73T} injury were stained with fluorescent antibodies against CD45, CD68, and DNA. **(A)** Twelve alveolar regions were selected for CD68$^+$ macrophage RNA sequencing (cells labeled in red). Regions from an untreated mouse were used as controls (three technical replicates; yellow asterisks). Injured/remodeled (five technical replicates; white asterisks) and peri-injured (two proximal technical replicates, orange asterisks; two distal technical replicates; red asterisks) region selection was based on histopathological assessment of inflammation and loss of alveolar architecture structure. Insets show representative regions. **(B)** Three dimensional PCA plot shows the clustering of samples based on sampling regions. Orange asterisks indicate two proximal technical replicates; red asterisks indicate two distal technical replicates. **(C)** Venn diagram of Reactome-based pathway analysis shows differentially regulated pathways between each pairing. **(D–F)** Bar graphs representing normalized enrichment scores for selected pathways. **(G-H)** Hierarchical clustering of 'Degradation of the extracellular matrix' and 'Assembly of collagen fibrils and other multimeric structures' signaling pathways in healthy (dark purple), peri-injured (pink), and injured (green) regions of the lung 14 days after SP-C^{I73T} induced injury. **(I)** Box plots of selected genes associated with macrophage activation (chemokine, metalloproteinases, complement cascade). The Y-axis represents normalized counts. A p-value ≤0.05 was considered significant using Linear Mixed Model. **(J)** Volcano plot comparing gene expression between peri-injured and injured macrophages. Fold changes are represented on log2 scale. Significance is shown as -log10(pvalue) using linear mixed model. In red are representative complement and fibrosis-associated genes.

changes in resident macrophages using a model of fibrogenic injury triggered by mutant SP-C^{I73T}. For these studies, we leveraged bulk RNA sequencing of flow cytometry sorted CD11b$^-$ SigF$^+$CD11c$^+$CD64$^+$ resident mature macrophages isolated from naïve (controls) and inflamed lungs. Exclusion gate ensured no contamination from Ly6G$^+$ neutrophils, B220$^+$ B cells, and CD3$^+$ lymphocytes. Ingenuity Pathway Analysis highlighted distinct gene and signaling signatures in lung macrophages during inflammatory initiation (3 days post mutant induction) and early remodeling (14 days). Principal component analysis highlighted transcriptional variance across the study groups (Figure 2A). Differential gene expression analysis revealed a relatively small set (48 genes) between controls and 3-day injury, while these responses were more pronounced between controls vs. 14-day post induction (3393) or 3-day vs. 14-day comparison (1446) (Figure 2B). Volcano plots showed increases in genes linked to innate immunity (the hematopoietic transcription factor, *Gata2;* histidine decarboxylase, *Hdc; interferon induced transmembrane protein 1, Ifitm1;* colony stimulating factor 1, *Csf1;* chemokine ligand, *Ccl17;* immunoglobulin epsilon receptor, *Ms4a2) and* metabolism *(*cholesterol side-chain cleavage enzyme, *Cyp11a1;* ATPase Na+/K+ transporting subunit alpha 3, *Atp1a3;* adenylate cyclase 6, *adcy6;* peptidyl arginine deiminase 2, *Padi2;* and myristoylated alanine rich protein kinase C substrate, *Marcks)* 3 days post injury (Figure 2C). By comparison, macrophage expression profile was bidirectional at 14 days, and demonstrated more sizable changes (as represented by adjusted p-values and log fold changes). Immunity, cell cycle, and metabolism genes were among the most significantly altered genes (complement C1q C-chain, *C1qc*; peptidoglycan recognition protein 1, *Pglyrp1*; ADAM metallopeptidase domain 19, *Adam19*; secreted protein acidic and cysteine rich, *Sparc*; apolipoprotein E, *Apoe*; insulin like growth factor 1 receptor, *Igf1r*; toll-like receptor 7, *Tlr7*; cyclin D2, *Ccnd2* (Figure 2D). Ingenuity Pathways Analysis predicted signaling related to activation, proliferation, and apoptosis of leukocyte predominantly at 3 days, while distinct cell movement and chemotaxis pathways, proliferation, angiogenesis and fibrogenesis were projected to be

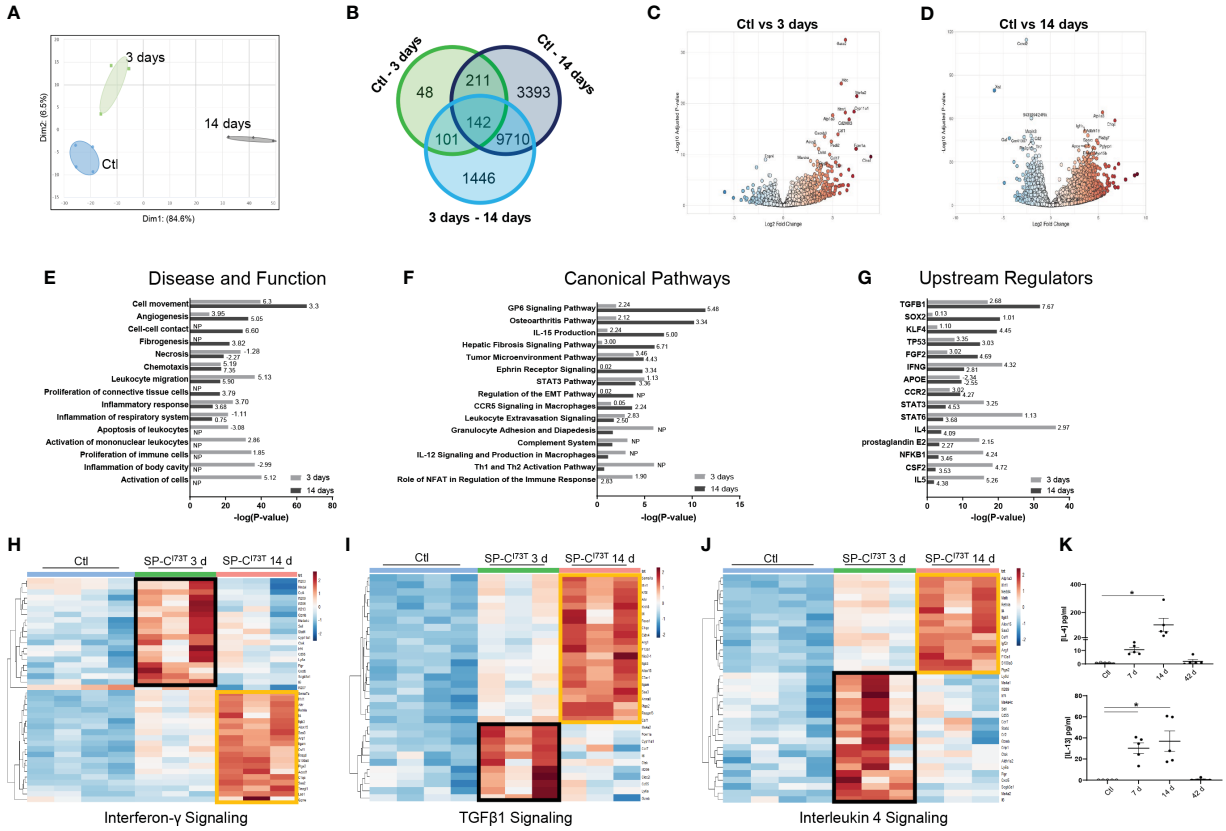


FIGURE 2
Transcriptional analysis of flow cytometry sorted alveolar macrophages following SP-C^{I73T} induced injury. Bulk RNA sequencing of flow cytometry sorted CD11b^{-}SigF^{+}CD11c^{+}CD64^{+} resident lung macrophages from control (N=4; Ctl, oil treated SP-C^{I73T} mice) or tamoxifen-treated SP-C^{I73T} mice at 3 d and 14 d (N=3 for each condition). **(A)** Principal Component Analysis (PCA) plot showing transcriptome variance in Ctl (blue), 3 days (black), and 14 days post SP-C^{I73T} induced injury. **(B)** Venn diagram breaking down significantly expressed genes among groups. **(C, D)** Volcano plots comparing fold change expression between Ctl vs. 3 days and Ctl vs. 14 days. **(E-G)** Ingenuity Pathway Analysis (IPA) of enriched Diseases and Functions, Canonical Pathways, and Upstream Regulators in lung macrophages 3- and 14-days after injury relative to Ctl. Bars represent enrichment [−log(pvalue)] 3 days (gray) and 14 days (black) after injury. Z-scores indicate predicted activation and inhibition respectively (N.P. not predictable). **(H-J)** Heat maps depicting significantly altered genes associated with Interferon-γ, TGFβ1, and Interleukin-4 signaling 3- and 14-days after injury relative to Ctl; criteria for significance was a 5% false discovery rate. Note that orange and black boxes highlight signatures specific to a given time point. **(K)** ELISA for IL-4 and IL-13 from SP-C^{I73T} BAL fluid collected from controls, 7 days, 14 days, or 42 days post injury. Dot plots with Mean + SE are shown. *p < 0.05 versus control group using One Way ANOVA followed by Tukey *post-hoc* test.

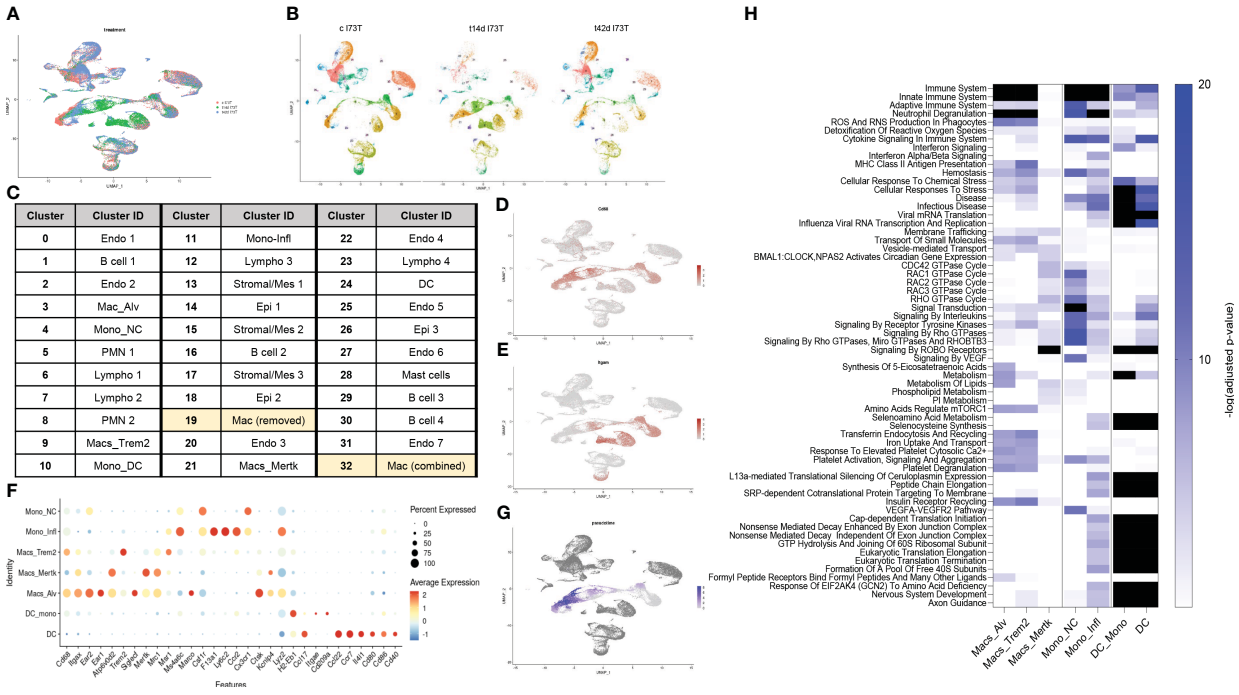
induced 14 days post-induction (Figure 2E; Supplementary Table 2). Specific pathways linked to NFAT dependent regulation of immune responses (directionality/z-score 1.90), Th1 and Th2 activation (no predicted z-score), granulocyte adhesion and diapedesis (no predicted z-score), and STAT3 signaling (z-score 1.13) were noted 3 days post injury. Comparatively, glycoprotein 6 (z-score 5.48), IL-15 production (z-score 5.00), fibrosis (z-score 6.71) and epithelial-mesenchymal transition signaling (no z-score), and osteoarthritis (z-score 3.34) were identified in macrophages at 14 days (Figure 2F). Notably, IPA's Upstream Regulators analysis predicted activation of pro-inflammatory signals at 3 days (NFKB1, IL5, IFNG, prostaglandin E2, STAT3 and STAT6), and conventional fibrogenic pathways (TGFB1, TP53, FGF2, and SOX2) 14 days (Figure 2G). Reactome-based breakdown of top differentially expressed genes from the Interferon-γ pathway identified distinct gene-sets expressed during inflammatory initiation (*Ifi203, Ifi206, Ifi209, Ifi213, Stat4)* and 14 days post injury (*S100a8, Arg1, Csf1, Alox15, Retnla, Il4)* (Figure 2H; Supplementary Table 2). A comparable dual response was observed in the TGF-β1 (3 days: *Ms4a2, Ctsk, Clec2i* and Cd55; 14 days: *Il1rl1*, *Arg1, Alox15, F13a1, Itgam*, P*tgs2, C1qc*, and A*nxa8*) and IL-4 signaling pathways (3 days: *Ms4a1, Cxcl5, Il6, Stat4*; 14 days: *S100a8, Agf2r, Csf1, Retnla, Il4)* (Figures 2I, J). ELISA-based validation of this established pro-fibrotic pathway confirmed increases in IL-4 and IL-13 expression in the bronchoalveolar lavage fluid 7-14 days post SP-C^{I73T} injury (Figure 2K). Notably, IL-15 and Glycoprotein-6 signaling, as well as angiogenesis displayed time restricted enrichments (14 days post-induction) (Supplementary Figures 2A, B, and not shown).

Single-cell RNA sequencing reveals activation of distinct monocyte/macrophage clusters responding to SP-C^{I73T} induced injury

We then employed single-cell RNA sequencing to overcome the constraints (and therefore bias) of antibody-based analysis of macrophages involved in SP-C^{I73T} induced injury. A 59,440-cell

dataset including healthy controls (29,213 cells), and two injury times representing peak inflammation (14 days - 14,266 cells) and established fibrosis (42 days - 15,961 cells) were studied. Population clustering using SCTransformation with the glmGamPoi method at a resolution of 0.4 yielded 33 clusters (Supplementary Figure 3A). Partition of the clusters based on origin/identity (Ctl, 14 days, 42 days) highlighted shifts in endothelial cells (cluster 25), eosinophils (cluster 8), and macrophages (cluster 9) composition after SP-C^{I73T} injury (Figures 3A, B). A combination of SingleR, manual annotation, and the top three non-redundant genes from each populations was used for identification of epithelial (3), endothelial (6), mesenchymal/stromal (3), megakaryocytes, granulocytes (3), B cells (4), lymphocytes (4), and mononuclear myeloid populations (9) (Figure 3C; Supplementary Figures 3B–I). Macrophage Cluster 19 (identified in only one of the eleven specimens) and Cluster 32 (of low abundance and merged with Cluster 3 after pseudobulk analysis determined high degree of transcriptional overlap).

Uniform Manifold Approximation and Projection (UMAP) analysis for the pan-macrophage gene *Cd68* and the mobilization marker *Itgam/Cd11b* were utilized to locate resident and recruited macrophages, monocyte-derived, and by exclusion monocytes and dendritic cells (Figures 3D, E). A curated gene set (mobility, maturity, and activation markers) combined with pseudotime analysis was used to clearly differentiate mononuclear myeloid clusters (Supplementary Table 3). Cluster 3 was annotated as alveolar macrophages since it expressed a combination of *Cd68, Itgax, Ear1*, *Ear2, Siglecf*, and moderate levels of *Mertk* (Figure 3F). Reconstructed trajectory analysis also predicted terminal differentiation within the alveolar macrophage cluster (Figure 3G). Cluster 9, identified as interstitial macrophages (*Cd68*, *Msr1, and C1qa*), was labeled based on its specific expression of *Trem2* (Figure 3F and not shown). Cluster 21 (macs_mertk) was identified as a mature subset based on pseudotime analysis and distinctive expression of *Mertk* alongside *Itgax*, *Mrc1*, *Atp6v0d2*, *Kcnip4.* This cluster was found solely in controls and 42 days post injury (Figures 3F, G). Two monocyte-like subsets were annotated as non-classical/NC (cluster 4 - *Csf1r, Cx3cr1*) and classical/inflammatory (cluster 11 - *Ccr2, Ly6c2, Lyz2, Ms4a6c, F13a1*) (Figure 3F; Supplementary Table 3). Cluster 10 was



Cluster	Cluster ID	Cluster	Cluster ID	Cluster	Cluster ID
0	Endo 1	11	Mono-Infl	22	Endo 4
1	B cell 1	12	Lympho 3	23	Lympho 4
2	Endo 2	13	Stromal/Mes 1	24	DC
3	Mac_Alv	14	Epi 1	25	Endo 5
4	Mono_NC	15	Stromal/Mes 2	26	Epi 3
5	PMN 1	16	B cell 2	27	Endo 6
6	Lympho 1	17	Stromal/Mes 3	28	Mast cells
7	Lympho 2	18	Epi 2	29	B cell 3
8	PMN 2	19	Mac (removed)	30	B cell 4
9	Macs_Trem2	20	Endo 3	31	Endo 7
10	Mono_DC	21	Macs_Mertk	32	Mac (combined)

FIGURE 3
Single-cell RNA sequencing of SP-C^{I73T} induced injury and fibrosis. Single-cell RNA sequencing analysis was performed on collagenase digested single cell suspensions from controls (Ctl, oil treated SP-C^{I73T} mice), 14 days or 42 days post SP-C^{I73T} induction. **(A)** UMAP analysis overlaying cells from controls (cI73T, pink), 14 days (t14d I73T, green), and 42 days (t42d I73T, blue). **(B)** Split identify UMAP depiction of all cellular clusters identified in controls (cI73T), 14 days (t14d I73T), and 42 days (t42d I73T) at 0.4 resolution. **(C)** Cluster breakdown identifying 7 endothelial, 3 mesenchymal, 3 epithelial, 2 polymorphonuclear, mast cells, 4 B cell, 4 lymphocytic, and 9 mononuclear myeloid clusters. Note that mononuclear myeloid clusters were renamed based on single R and manual annotation: cluster 3 = Mac_Alv; cluster 4 = Mono_non classical/NC; cluster 9 = Mac_Trem2; cluster 10 = Mono_DC; cluster 11 = Mono_inflammatory/infl; cluster 19 = was removed from further analysis as it was identified solely in one of the controls; cluster 21 = Mac_MerTK; cluster 24 = DC; cluster 32 = was combined with cluster 3/Mac_Alv since pseudobulk analysis revealed analogous gene expression and pathway activation. **(D, E)** UMAP analysis for *Cd68* and *Itgam/Cd11b* defines distribution and cellular expression in the SP-C^{I73T} lung. **(F)** Bubble plot of selected genes associated with macrophage maturation (*Trem2, Itgax, Ear1, Ear2, Atp6v0d2, Siglecf, Mertk*), recruitment and activation (*F13a1, Ly6c2, Ccr2, Cx3cr1, Ctsk, Lyz2, H2-eb1, CcL22, Ccl17, Cd80, Cd86)* in all annotated monocytes/macrophages/dendritic cell clusters (Mo/Mac/DCs). **(G)** Pseudotime UMAP analysis. Note color intensity defining cellular maturation among monocyte/macrophage clusters. **(H)** Enrichr-based Reactome pathway analysis for the top-15 predicted pathways in the 7 resulting monocyte/macrophage/DC clusters. Single gradient color heatmap shows logarithmic adjusted p-values. Values above the arbitrary threshold (-log of adjusted p-value) of 20 were color-coated in black.

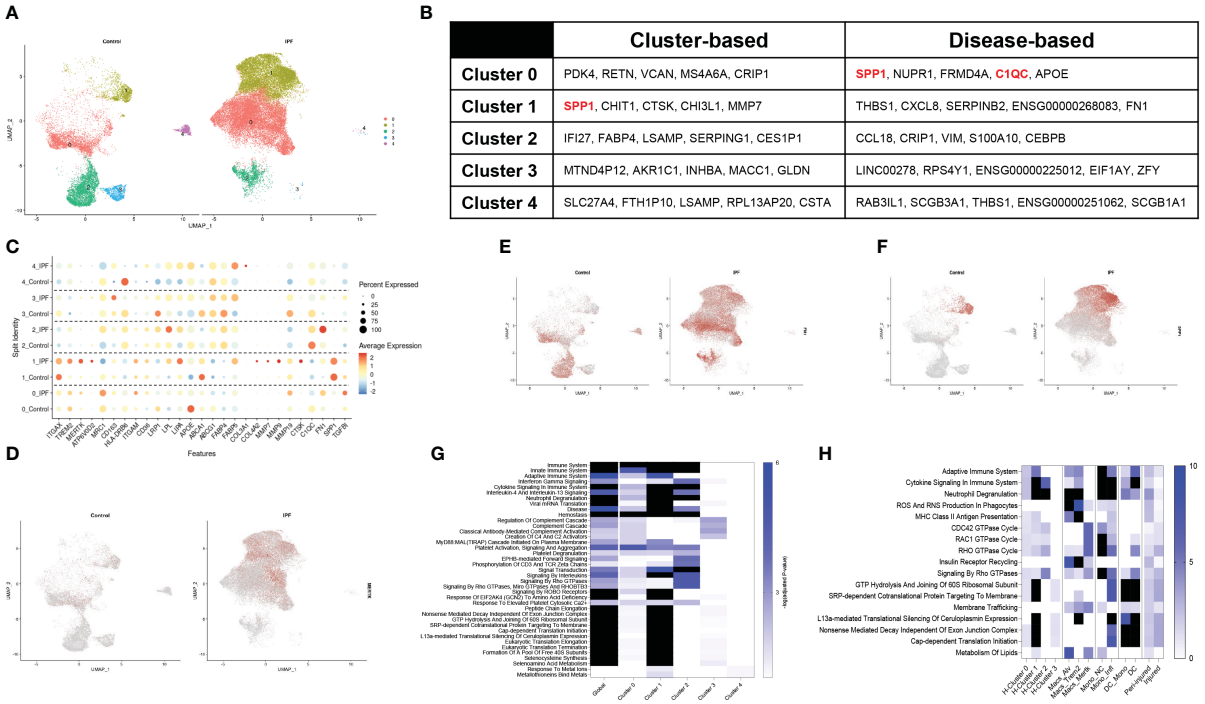
annotated as monocyte-derived DCs based on a mixed myeloid and lymphoid signature (*Ccr2*, *Itgae/Cd103*, *Cd209/Dc-sign)* and major histocompatibility complex genes. A cluster of lymphoid-derived DCs (cluster 24) lacking all classical myeloid features and expressing *Cd80, Cd86, Cd40, Il4i, Ccr7, Ccl17*, and *Ccl22* was also annotated (Figure 3F). Split identity analysis of these curated gene sets also highlighted time related changes in the expression of established macrophage maturity and identity genes (e.g. *Atp6v0d2, Itgax*, and *Cd68* expression increases, downregulation of *Siglecf, Ear1*, and *Ear2*) (Supplementary Figure 4A). Similar trends were observed for neutrophils (PMN 1 lost *Ptgs2* and *Tgm2* in favor of *Retnlg* and *Lcn2), e*osinophils (PMN 2 - *Siglecf, Itgam, Cxcr2, Cd33, Csf3r, Tgfbr1*), epithelial cells (Epi1 - *Cldn18, Ager, Hopx*, *Krt8, Krt19, Sftpb*), mesenchymal/stromal cells (Mes 2 - *Tgfbr3, Ccn1, Pdgfra, Npnt*, *Loxl1, Ecm1, Fgf2*), and lymphatic endothelial cells (Endo 6 - *Nrgn, Itga2b, Gp1bb* increased 14 days post injury). By comparison, B and T cells showed limited transcriptional fluctuation over the 42-day injury (Supplementary Figures 4B–G).

Pseudobulk differential expression analysis of *Trem2*$^+$ macrophages surveyed gene expression and patterns during injury. Approximately 700 genes demonstrated a transient drop in abundance at 14 days (expression pattern group-1), while the abundance of 621 genes was significantly increased following SP-C^{I73T} induction (expression pattern group-2) (Supplementary Figure 5). Smaller gene sets were shown to transiently increase at 14 days (122 genes, expression pattern group-3) or steadily decrease after SP-C induction (61 genes, expression pattern group-4) (Supplementary Figure 5). Notably, Reactome-based analysis predicted no significant pathway to be altered for genes annotated in expression pattern-1, while cytokine and interleukin signaling, immunity, antigen presentation, and lipid and carbohydrate metabolism were among significant pathways for expression pattern-2 genes. By comparison, transcript abundance in inflammatory monocytes followed two expression patterns: transient decrease (237 genes) or increase (158) at 14 days post injury. The latter predicted engagement of pathways related to complement cascade, extracellular matrix organization and immune responses (Supplementary Table 3).

Reactome-based analysis of the top-15 most significantly regulated pathways revealed shared inflammatory signature ('immune system', 'neutrophil degranulation', 'ROS and RNA production in phagocytes') between alveolar macrophages and *Trem2*$^+$ interstitial macrophages, though at considerably different adjusted *p*-values (Alv macs *p*-value $<10^{-25}$ vs. *Trem2*$^+$ *p*-value $<10^{-103}$) (Figure 3H; Supplementary Table 3). By comparison, *Mertk*$^+$ cells were not predicted to engage in innate or adaptive immunity or cytokine mediated signaling, but generated a strong GTPase signature (CDC42, RAC, RHO, and ROBO/SLIT). Examination of monocyte-like clusters highlighted analogous 'cytokine signaling in immune system', 'innate immunity', and 'neutrophil degranulation' engagement. *Cx3cr1*$^+$ non-classical cells excelled in 'VEGF signaling', 'signaling by tyrosine kinases', and 'Rho GTPase signaling', while *Ccr2*$^+$*Ly6c2*$^+$ classical/inflammatory monocytes produced a signature related to RNA synthesis and translation and pathways involved in 'cellular response to stress'. Lastly, monocyte-derived and lymphoid DCs produced comparable activation profile (infectious disease response, translational elongation and termination, and ribosomal homeostasis) (Figure 3H).

Macrophages from human IPF display a comparable phenotype as murine SP-C^{I73T} injury.

A recent dataset published by Adams and colleagues provided in depth assessment of epithelial, endothelial, mesenchymal, and immune populations in healthy and IPF tissue explants (40). The mononuclear myeloid compartment from this set (44,226 cells; 12,514 from controls and 31,712 IPF, GSE136831) was re-examined to define similarities between clinical and experimental fibrosis. To avoid overfitting the data, the lowest resolution was utilized (0.1), ultimately generating 5 major subsets (Supplementary Figure 6A). UMAP analysis identified clusters 3 and 4 solely in the controls, with an expansion in clusters 0 and 1 was noted in IPF (Figure 4A). Cluster-based analysis of top-5 non-redundant genes revealed differential expression for *SPP1, CTSK, MMP7* (cluster 1) and *FABP4* (cluster *2*) (Figure 4B; Supplementary Table 4). Disease-based analysis highlighted notable changes in cluster 0 (*SPP1, C1QC, APOE*) and cluster 1 (*THBS1, FN1*), consistent with our murine dataset, including, (Figures 4B, C, Supplementary Table 5). A curated set of 25 identity and lipid associated genes, used to annotate this clusters, identified cluster 0 as an interstitial macrophage population (widespread expression of *ITGAX*, *TREM2*, *MRC1, ITGAM*, *APOE)* and Cluster 1 as alveolar macrophages (*ITGAX/CD11C, ITGAM*, *ABCA1*). Notably, cluster 1 also displayed a signature characteristic of a potentially pro-fibrotic monocyte-derived population in IPF lungs (*TREM2*, *MERTK*, *ATP6V0D2, CTSK*, *COL4A2, MMP7, MMP9)* paired with a shift in metabolic function (*FABP5, LPL* and *LIPA* increases). Clusters 2, 3 and 4 did not exhibit distinguishing signatures besides overexpression of major histocompatibility complex genes (*HLA-DRB6*) (Figure 4C; Supplementary Figure 6B). UMAP analysis of the distribution of *MERTK* and fibronectin1/*FN1* demonstrated widespread presence in cluster 0 and 1 of the IPF lung, while ostepontin1/*SPP1* was restricted to cluster 1 (Figures 4D–F). Tumor growth factor β1/*TGFβ1* was expressed in all cells regardless of disease state (Supplementary Figures 6C, D). Enrichr-based pathway analysis using the Reactome database highlighted IPF induced changes (annotated as "global") in all facets of immune cell behavior, including "Immunity", "Cellular Responses to Stress", "Cytokine Signaling in Immune System", complement activation, and GTPase signaling, and control of transcription and translation. Clusters 0, 1, 2 were responsible for the majority of these signals (Figure 4G; Supplementary Table 4). Direct comparison of Reactome pathways from the spatial analysis and both the murine and human single-cell sequencing datasets highlighted similarities between cluster 1 annotated in the human IPF dataset and inflammatory monocytes from the SP-C^{I73T} mouse model. No distinctive similarities were noted with respect to CD68$^+$ macrophages sampled from peri-injured and injured regions of the lung during our spatial analysis (Figure 4H).



	Cluster-based	Disease-based
Cluster 0	PDK4, RETN, VCAN, MS4A6A, CRIP1	SPP1, NUPR1, FRMD4A, C1QC, APOE
Cluster 1	SPP1, CHIT1, CTSK, CHI3L1, MMP7	THBS1, CXCL8, SERPINB2, ENSG00000268083, FN1
Cluster 2	IFI27, FABP4, LSAMP, SERPING1, CES1P1	CCL18, CRIP1, VIM, S100A10, CEBPB
Cluster 3	MTND4P12, AKR1C1, INHBA, MACC1, GLDN	LINC00278, RPS4Y1, ENSG00000225012, EIF1AY, ZFY
Cluster 4	SLC27A4, FTH1P10, LSAMP, RPL13AP20, CSTA	RAB3IL1, SCGB3A1, THBS1, ENSG00000251062, SCGB1A1

FIGURE 4
Single-cell RNA sequencing of annotated macrophages from human IPF. Macrophages from a published IPF dataset (GSE136831) were re-mined. **(A)** Split identity UMAP analysis of clusters annotated at 0.1 resolution using the Seurat analysis package. **(B)** Cluster-based and disease-based top-5 non-redundant genes from annotated clusters. **(C)** Split identity UMAP shows expression distribution for *TREM2* in control and IPF lungs. **(D)** Split identity bubble plot of selected genes associated with macrophage maturation (*TREM2, MERTK, ATP6V0D2, ITGAX*), recruitment and activation (*ITGAM, CD163, HLA-DRB6, CTSK, COL3A1, COL4A2, MMP7, MMP9, FN1, SPP1, TGFBI*), and lipid homeostasis (*APOE, CD36, LRP1, LPL, LIPA, ABCA1, ABCG1, FABP5, PPARG*). **(E)** Split identity UMAP showing expression distribution for *MERTK* in control and IPF lungs. **(F)** Split identity UMAP showing expression distribution for *FN1* in control and IPF lungs. **(G)** Split identity UMAP showing expression distribution for *SPP1* in control and IPF lungs. **(H)** Enrichr-based Reactome pathway analysis for the top-15 predicted pathways between controls and IPF samples. Single gradient color heatmap shows logarithmic adjusted p-values. Values above the arbitrary threshold (-log of adjusted p-value) of 6 were color-coated in black. **(I)** Enrichr-based Reactome pathway analysis comparing significant pathways in mononuclear myeloid clusters from the human IPF dataset, SP-C^{I73T} spatial transcriptomics, and SP-C^{I73T} single-cell sequencing datasets. Single gradient color heatmap shows logarithmic adjusted p-values. Values above the arbitrary threshold (-log of adjusted p-value) of 10 were color-coated in black.

Osteopontin1 and fibronectin1 intercellular communication during SP-C^{I73T} induced injury

CellChat software was utilized to predict communication among cellular clusters during SP-C^{I73T} induced fibrotic injury. Initial analysis grouped all 33 clusters into 7 macro-groups (epithelial, endothelial, stromal, B cells, granulocytes, lymphocytes, and Mo/Mac/DCs) (Supplementary Figure 7A) (51, 53). CellChat based aggregation of ligand: receptor expression estimated increases in differential number and strength of interaction between mesenchymal-epithelial and mesenchymal-Mo/Mac/DCs and within epithelial clusters 14 day post-injury, relative to controls. Within the Mo/Mac/DC macro-groups the communication was estimated to increase in strength (Supplementary Figure 7B). Interrogation of the inter-cluster network 42 days post-induction showed increased interactions in stromal and Mo/Mac/DC macro clusters relative to controls and 14 days (Supplementary Figures 7C, D), with incoming and outcoming signals from the Mo/Mac/DC macro-cluster associated with SPP1 and FN1 pathway (Supplementary Figures 7E–G, Supplementary Tables 6–9).

To pinpoint the exact cellular origin of these signals, CellChat analysis was carried out after split of the Mo/Mac/DC macro-cluster. This analysis revealed alveolar macrophages as a dynamic cluster at baseline (both incoming/receptor-based and outgoing/ligand-based interactions), while *Trem2*$^{+}$ macrophages produced high volume of incoming/outgoing signals 14 days post injury and *Mertk*$^{+}$ cells became active at 42 days. Classical/inflammatory monocytes maintained comparable network profile throughout the 42-day time course, with non-classical monocytes effectively reducing outgoing signals after SP-C^{I73T} injury (Figure 5A; Supplementary Figures 8A–C). Split identity pathway analysis across the 42-day time course demonstrated the origin of the osteopontin1/SPP1, fibronectin1/FN1, chemokine ligand/CCL, laminin, semaphorin3/SEMA3, galectin, complement, and collagen signaling to be driven by *Trem2*$^{+}$ macrophages at 14 days (Figure 5B). By 42 days, the predicted interaction strength was comparable across the monocyte and macrophage clusters (Supplementary Figures 8D–F). Connectome ring plots were used to capture the changes and directionality of these interactions during injury and fibrosis. Analysis of CCL signaling predicted baseline crosstalk among all annotated mononuclear myeloid cells and granulocytes. While these responses were unaffected 14 days

post SP-C^{I73T} injury, by 42 days there was a global activation of this signaling network (Figure 5C). Differential expression of chemokines/cytokines/interleukins and their receptors highlighted distinct signatures in alveolar macrophages (*Ccl6*, *Il18*, *Cxcl2*, and *Il1a)*, *Trem2*$^{+}$ macrophages (*Ccl12, Ccl2, Ccl24, Ccl9, Cxcl16, Ccr5, Il10rb, Il11ra1), while Mertk*$^{+}$ macrophages and monocyte-derived DCs were relatively quiescent. Classical and inflammatory monocytes presented a receptor dominant repertoire (*Cx3cr1, Il10ra, Il17ra, Il6ra* – inflammatory monocytes exclusively expressed *Ccr2)* (Figures 5D, E). CellChat-based network analysis for complement signaling predicted outgoing communication from classical monocytes to other myeloid clusters in all conditions, with a transient activation originating from stromal cells 14 days post injury (Figure 5F). Differential expression revealed distinct signatures in *trem2*$^{+}$ macrophages (*C1qa, C1qb, C1qc, C3ar1*) and stromal cells (stromal 2 - *C1qtnf7, C1ra, C2, C3, C4b, C7*), with granulocytes also expressing *C3ar1*, *C5ar1*, and *C5ar2* (Figure 5G).

Trem2^{+} macrophages coordinate pro-fibrotic communication during SP-C^{I73T} induced injury and fibrosis

We then query the dataset for distinctive extracellular matrix reorganization gene signatures. Unsurprisingly, mesenchymal cells produced signals from collagen genes, metalloproteinases, laminins, and platelet-derived growth factor receptor alpha and beta (Supplementary Figure 9). Expression of *Timp1, Mmp12*, and *Mmp14 were restricted to Trem2*$^{+}$ cells, while *Spp1, Fn1*, and *Mmp19* were shared with alveolar macrophages (Figure 6A). *Tgfb1* transcripts were maximal in inflammatory monocytes, though UMAP analysis of distribution suggested widespread expression in the lung (Figures 6A, B). Notably, *Spp1*, and *Fn1* expression was also found in the endothelial and mesenchymal compartment (Figures 6C, D). Signaling connectome ring for *Tgfb1* highlighted baseline signals emanating primarily from alveolar

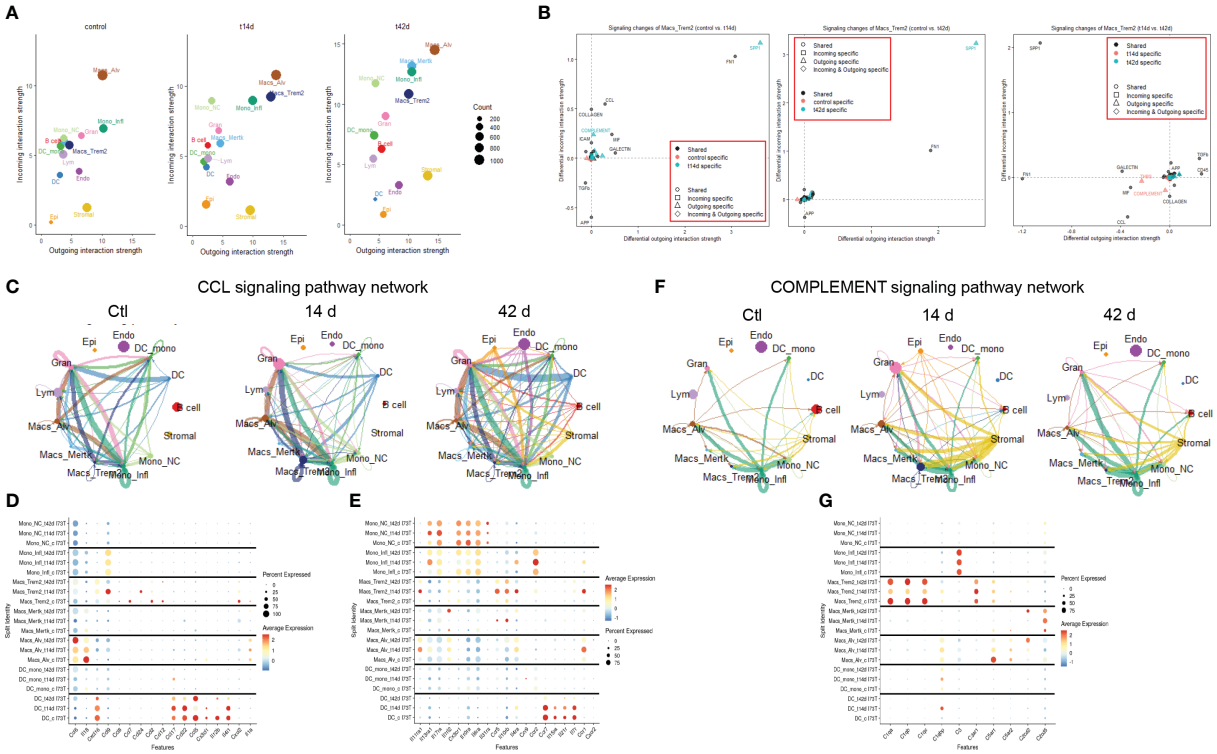


FIGURE 5
Cell-cell communication analysis in SP-C^{I73T} induced injury and fibrosis. CellChat software was used to estimate cell-cell communication in controls (Ctl, oil treated SP-C^{I73T} mice), 14 days or 42 days post SP-C^{I73T} induction. **(A)** Ligand:receptor expression analysis examining incoming (y-axis) and outgoing (x-axis) signals in controls, 14 days or 42 days post injury. Note that individual endothelial, mesenchymal, epithelial, granulocyte, B cell, and lymphocyte clusters were combined into "macro-clusters", while mononuclear myeloid cells remained split. The size of the circles is representative of population size. **(B)** Prediction of signaling changes in *Trem2*$^{+}$ macrophages between control and 14-day, control and 42-day, and 14-d and 42 d. Plot legend describes directionality (circle - shared by both groups, square - incoming in a specific group, triangle - outgoing in a specific group, diamond - incoming and outgoing in a specific group) and signal specificity (black - shared by both groups, orange - control specific, cyan - injury/tamoxifen specific). **(C)** Connectome ring plots for chemokine ligand/CCL signaling pathway network predict directionality and communication strength among clusters. **(D, E)** Split identity bubble plot for chemokine/cytokine/interleukin ligands and receptors among mononuclear myeloid clusters. Note that the size of the bubble indicates the relative abundance of the population expressing target gene. Color coating indicates average expression. **(F)** Connectome ring plots for Complement signaling network predicting directionality and strength of communication among clusters. **(G)** Split identity bubble plot for complement associated genes among mononuclear myeloid clusters. Note that the size of the bubble indicates the relative abundance of the population expressing target gene. Color coating indicates average expression.

macrophages, non-classical monocytes, monocyte-derived DCs, and B cells. Despite transient drop 14 days post SP-C^{I73T} induced injury, the network returned to control levels by 42 days, with involvement of *Mertk*$^{+}$ macrophages (Figure 6E; Supplementary Tables 6–9). Osteopontin1 signaling network revealed alveolar macrophages as the sole driver in control lungs. SP-C^{I73T} induced injury produced increases in alveolar and *Trem2*$^{+}$ macrophages (14 days), and ultimately global expression in all lung cells (42 days) (Figure 6F; Supplementary Tables 6–9). Inflammatory monocytes and alveolar macrophages were shown to engage in fibronectin 1 signaling in the control and injured lung. Notably, *Trem2*$^{+}$ macrophages and stromal cells were predicted to partake in FN1 communication 14 days post-injury, while *Mertk*$^{+}$ macrophages were shown at 42 days (Figure 7G; Supplementary Tables 6–9). Although spatial examination did not detect significant differences among CD68^{+} macrophages collected from healthy, peri-injured, and injured regions of the lung 14 days post injury, we found that proximity to the remodeled tissue (areas of interest 006 and 010) resulted in higher expression of *Tgfb1*, *Spp1*, and *Fn1* (Figures 6H–J).

Genetic ablation of ApoE does not impact SP-C^{I73T} induced injury

Curated analysis for genes involved in lipid handling and metabolism showed heterogeneous expression among mononuclear myeloid cells during SP-C^{I73T} induced injury and fibrosis. Monocytes exhibited a limited repertoire of lipid-associated genes. Alveolar macrophages displayed an extensive signature at baseline (*Fabp1, Plin2, Lipa, Lpl, Lpin1, Abcg1, Acox1*) which expanded after injury (*Nr1h3, Marco, Fabp4*). *Mertk*$^{+}$ macrophages developed distinct signatures at 14 days (*Nr1h3, Marco, Plin2, Abca1*) and 42 days (*Pparg Ldlr, Lrp1, Lrp12*, and *Acox1*), a time coordinated with their maximal abundance (Figure 7A). *Trem2*$^{+}$ gene expression was restricted to ATP binding cassette transporters (*Abca1, Abca9*) and exclusive *Apoe* expression. Analysis of the distribution of *Apoe* and two of its established receptors (*Lrp1, Ldlr*) confirmed predominant expression in mononuclear myeloid and mesenchymal clusters (Figures 7B–D) (54). Immunohistochemical analysis of ApoE showed increases in parenchymal and immune cells 7 days and 14 days post induction, in particular within injury foci (Figure 7E). *In situ* hybridization analysis validated the origin of *Apoe* in CD68^{+} macrophages up to 42 days after SP-C^{I73T} injury, while western blot analysis of BAL fluid also confirmed protein increases after SP-C^{I73T} induced injury (Figures 7F, G). Our bulk RNA sequencing predicted activation of the ApoE signaling pathway in SigF^{+}CD11c^{+}CD64^{+} alveolar macrophages, however its expression (or that of its receptors) was not spatially restricted (Figures 7H, I). To test the hypothesis that ApoE is directly implicated in Mo/Mac/DC function during pulmonary remodeling, mono- and bi-allelic deletion of ApoE (SP-C^{I73T}ApoE^{HET} and ApoE^{HOM}) was designed (Figure 7J). Neither ApoE hypofunctional or null mice impacted accumulation of *Spp1*$^{+}$ macrophages within areas of injury 14 days post SP-C^{I73T} induction (Figure 7K). Similarly, there was no improvements in pulmonary histopathological scoring, BAL cell counts, or mortality (Figures 7L–O).

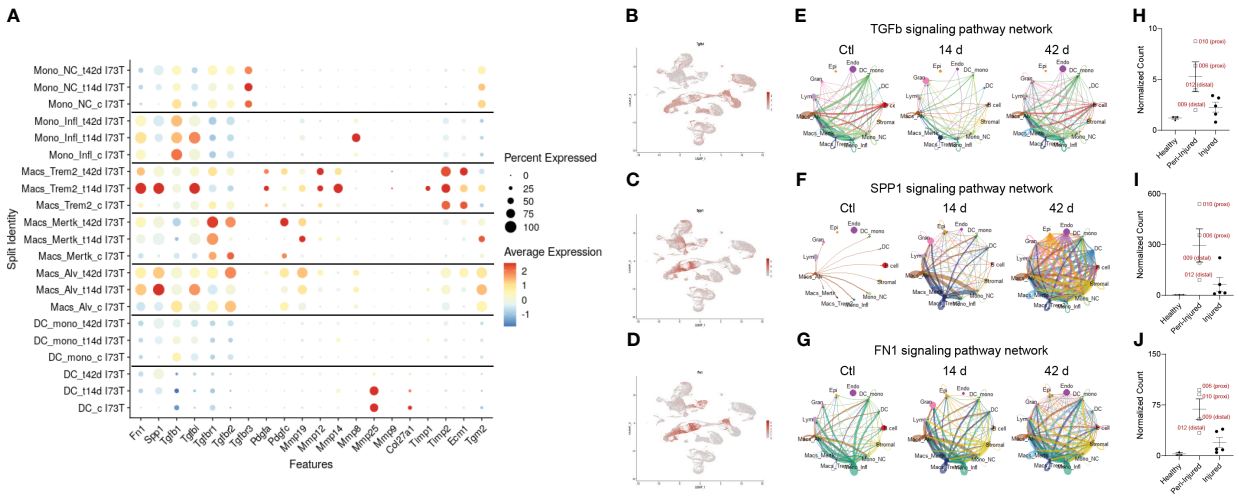


FIGURE 6
Pro-fibrotic cell-cell communication in SP-C^{I73T} induced injury and fibrosis. CellChat-based analysis of fibrogenic signaling from controls (Ctl, oil treated SP-C^{I73T} mice), 14 days or 42 days post SP-C^{I73T} induction. **(A)** Split identity bubble plot for fibrosis-associated genes among mononuclear myeloid clusters. Note that the size of the bubble indicates the relative abundance of the population expressing target gene. Color coating indicates average expression. **(B-D)** UMAP analysis for *tgfb1, spp1, fn1*. **(E-G)** Connectome ring plots for TGFb, osteopontin/SPP1 fibronectin/FN1 signaling pathway network predicting directionality and communication strength among clusters. Note that only Mo/Mac/DC clusters are split. **(H-J)** Box plots for *Tgfb1, Spp1, Fn1*. CD68^{+} macrophages isolated from healthy regions are shown in dark purple, peri-injured macrophages in pink, and injured macrophages are shown in green. The Y-axis represents normalized counts.

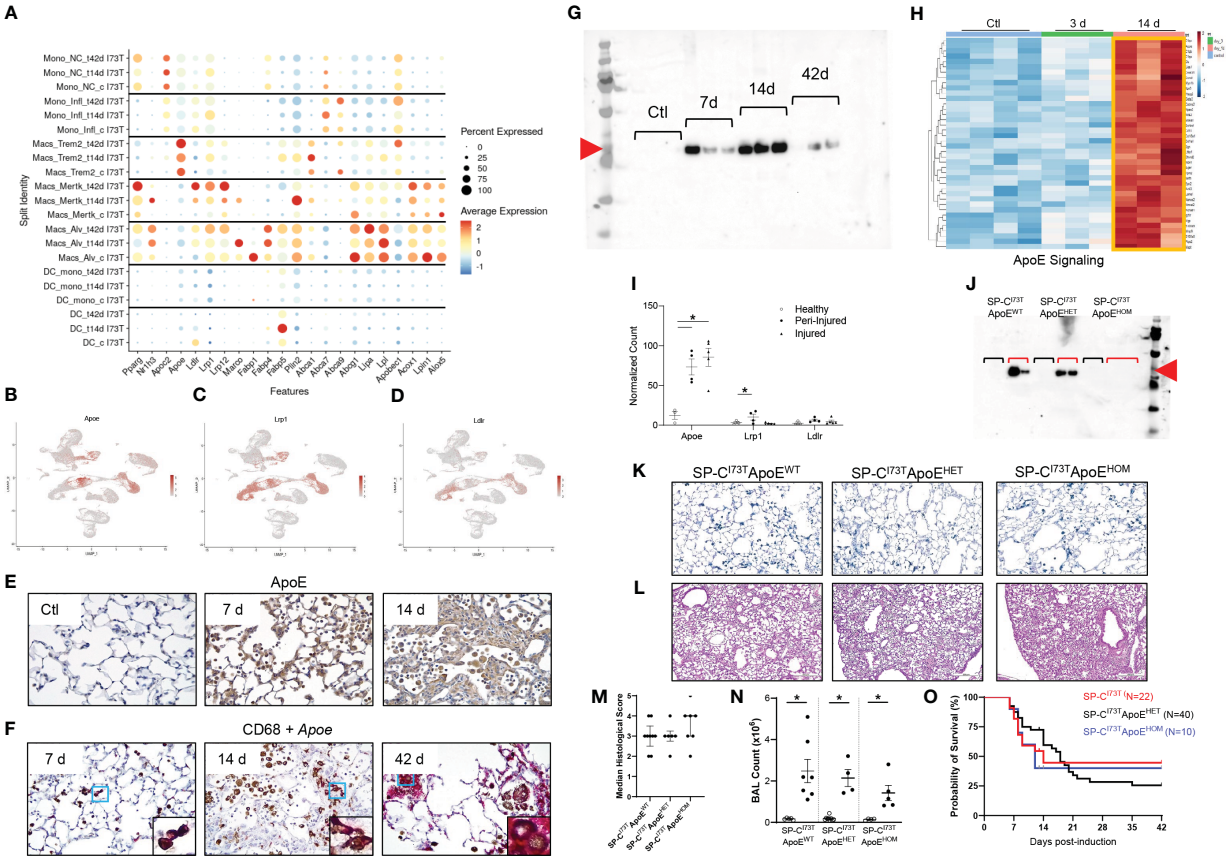


FIGURE 7
Lipid signature in SP-C^{I73T} induced injury and fibrosis. Lipid-associated gene expression was assessed in controls (Ctl, oil treated SP-C^{I73T} mice), 14 days or 42 days post SP-C^{I73T} induction. **(A)** Split identity bubble plot for lipid-associated genes among mononuclear myeloid clusters. Note that the size of the bubble indicates the relative abundance of the population expressing target gene. Color coating indicates average expression. **(B-D)** UMAP analysis for *Apoe, Lrp1, Ldlr.* **(E)** Immunohistochemical analysis for ApoE in controls, 14 days or 42 days post SP-C^{I73T} induced injury. A representative image is shown. **(F)** Combination of immunohistochemistry (CD68, brown) and *in situ* hybridization (*Apoe*, magenta). Protein expression was visualized using a DAB Vectastain kit. Original magnification, 400x. Insets, 1000x. **(G)** ApoE western blot analysis of BAL fluid from SP-C^{I73T}control (Ctl), 7 days, 14 days, and 42 days post induction (N=3). Red arrowhead indicates molecular weight band for ApoE (34 KDa). **(H)** Heat maps of significantly altered genes associated with ApoE signaling in flow cytometry sorted CD11b^{-}SigF^{+}CD11c^{+}CD64^{+} resident lung macrophages 3- and 14-days after injury. The criteria for significance was a 5% false discovery rate. Note that the orange box highlights signatures specific to a given time point. **(I)** Box plots for *Apoe, Lrp1, Ldlr*. The Y-axis represents normalized counts. A p-value ≤0.05 was considered significant using Linear Mixed Model. **(J)** ApoE western blot analysis of BAL fluid from control and 14 days SP-C^{I73T}ApoE^{WT} (controls N=2; 14 days N=2), SP-C^{I73T}ApoE^{HET} (controls N=2; 14 days N=2), and SP-C^{I73T}ApoE^{HOM} (controls N=2; 14 days N=3). Red arrowhead indicates molecular weight band for ApoE (34 KDa). Black brackets represent ApoE^{WT}, Apoe^{HET}, ApoE^{HOM} BAL fluid from control SP-C^{I73T} mice. Red brackets represent ApoE^{WT}, Apoe^{HET}, ApoE^{HOM} BAL fluid 14 days post SP-C^{I73T} induced injury. **(K)** Duplex *in situ* hybridization analysis for *Tgfb1* (pink) *Spp1* (blue) in SP-C^{I73T}ApoE^{WT}, SP-C^{I73T}ApoE^{HET}, and SP-C^{I73T}ApoE^{HOM} lungs 14 days post injury. Representative images are shown. **(L-M)** H&E staining and histological scoring of SP-C^{I73T}ApoE^{WT}, SP-C^{I73T}ApoE^{HET}, and SP-C^{I73T}ApoE^{HOM} lungs 14 days post injury. Pathological scoring included counting of foci of injury, extent of inflammation, edema, hemorrhage, and alveolar architecture remodeling. Representative images are shown. **(N)** Bronchoalveolar lavage fluid (BAL) cell counts of SP-C^{I73T}ApoE^{WT}, SP-C^{I73T}ApoE^{HET}, and SP-C^{I73T}ApoE^{HOM} samples (controls and 14 days post injury). Data are presented as mean ± SEM (n = 4-9 mice/group), analyzed using two-way ANOVA. A p<0.05 (*) was considered significant. Lines mark significant groups. **(O)** Kaplan–Meier survival curve of SP-C^{I73T}ApoE^{WT}, SP-C^{I73T}ApoE^{HET}, and SP-C^{I73T}ApoE^{HOM} lungs 14 days post injury. Analysis includes mice found dead or displaying ≥20% body weight from study initiation. Log-Sum (Mantel-Cox) Rank test was used.

Discussion

Genetic mutations in key nodes of pulmonary epithelial function are intertwined with familial forms of pulmonary fibrosis (PF) and other interstitial lung diseases (55). Dysbiosis among parenchymal cells promote repeated cycles of injury, inflammation, epithelial-to-mesenchymal transition, and the proliferation of polyclonal fibroblast clusters, leading to spatially and temporally heterogenous tissue remodeling (56–58). In the current paradigm of PF, the role of immune cells remains ambiguous and primarily circumscribed to "acute inflammatory exacerbations", sporadic events that severely worsen disease phenotype and accelerate patient mortality. This paradigm is supported by clinical and experimental evidence linking excess inflammatory monocyte mobilization and their retention as pro-fibrotic monocyte-derived macrophages to poor prognosis (13, 14, 20, 22). Our group has previously defined the involvement of mononuclear myeloid cells at all stages of inflammatory exacerbations triggered by a clinically relevant single point mutation in the alveolar epithelial cell specific gene encoding for the surfactant protein C (SP-C^{I73T}) (59). Using this experimental platform and available human IPF datasets, this work aimed to

expand the current understanding of PF and fill an important gap pertaining to macrophage phenotypic heterogeneity and aberrant cell-cell crosstalk in fibrotic disease. Specifically, we spatially characterized the transcriptional profile of CD68$^+$ macrophages within and surrounding fibrotic foci of injury; defined incoming and outgoing pro-fibrotic communications among pulmonary cells; identified a dynamic lipid signature across all annotated macrophage and monocyte clusters, with a sole population producing *Apoe*; and showed that SP-C^{I73T} induced fibrosis is marginally impacted by genetic *Apoe* deletion.

Initial work was designed to spatially localize activated macrophages in the healthy, inflamed, and fibrosing lung. This approach aligns with a handful of recent reports that elegantly resolve epithelial, mesenchymal, and immune cell identity and abundance in clinical and experimental fibrosis (60, 61). Our studies go beyond the focus of these publications by placing emphasis on the distribution of macrophages from spatially diverse regions of the injured lung. CD68 represents a compromise allowing to sample both control/healthy and injured macrophages. Our gene expression and hierarchical clustering analysis demonstrate heterogenous transcriptional signatures among macrophages sampled from the fibrosing lung, thus suggesting a gradient of activation based on spatial localization. Pathway analysis also painted an unexpected picture characterized by transcriptionally quiescent CD68$^+$ cells sampled from fully remodeled regions, contrasting a hyperactive phenotype emanating from peri-injured macrophages. Within this region of interest our findings also support the notion that macrophages activation is linked to their proximity to the injury. Findings that innate and adaptive immunity, matrix remodeling, senescence, and redox balance signals are predicted to be driven primarily by peri-injured macrophages (compared to all other sampled regions) adds depth to previous reports showing unequal expression of inflammatory proteins among immune cells found in healthy and fibrotic regions of the clinically diseased lung (62). Our results also support and complement evidence showing downregulation of inflammatory signaling (TNFα) in IPF immune infiltrates compared to healthy/unaffected regions (63). The comprehensive nature of our transcriptional analysis is novel in the context of fibrosis, particularly in pulmonary injury triggered by genetic susceptibility of the epithelial compartment and could provide insights in the development of more targeted and effective therapeutics in the future.

To further appreciate the role of the resident alveolar compartment during inflammatory exacerbations progressing to fibrosis, flow cytometry-based sorting of CD11b$^-$SigF$^+$CD11c$^+$ cells was performed 3 days and 14 days after SP-C^{I73T} induced injury. Our results allowed us to place these cells as players of all phases of the SP-C^{I73T} injury, through temporally restricted activation of canonical inflammatory pathways (IL6 and prostaglandin E2 exclusively shown 3 days post injury) and pro-fibrotic pathways (glycoprotein/GP6, SOX2, KLF4). At the same time, we show a subset of signaling networks sustained over our 14 day analysis, albeit driven by distinct transcriptional signatures (IFNγ, TGFβ1, IL-4) (64, 65). While these signals have been previously reported in the fibrotic lung, specific annotation in resident alveolar macrophages strengthens the notion that this population functions as a pleiotropic effector of inflammation and tissue remodeling (66, 67), and further substantiates our prior histopathological investigation of the SP-C^{I73T} injury (dysregulated inflammation and increased mortality) following pharmacological depletion of alveolar macrophages with clodronate liposomes (10).

Due to the phenotypic heterogeneity intrinsic of monocytes/macrophages in chronic disease, we built on our previous characterization of monocyte-derived macrophages through single-cell RNA sequencing (10, 22, 68). Albeit a low-resolution clustering analysis, we identified three macrophage (alveolar macrophages, *Trem2*$^+$, and *Mertk*$^+$), as well as transcriptionally distinct populations of monocytes (classical - *Itgam/cd11b*, *Cx3cr1*; inflammatory - *Ccr2, Ly6c, Lyz2*) (69). Our findings that complement signaling is involved in SP-C^{I73T} injury and fibrosis is consistent with evidence that soluble defense collagens support activation, proliferation, and tissue-repair functions of macrophages (70–72). Alongside these signals, our results define inflammatory monocytes and monocyte-derived moieties (interstitial macrophages) as centrally involved in fibronectin/FN1 and osteopontin/SPP1 signaling in the fibrosing lung, a finding consistent with bleomycin-induced injury (42, 73–78).

Mertk has been used to identify alveolar macrophages due to its role in mediating phagocytosis of apoptotic cells (79). Our annotation of a *Mertk^{int}* cluster in control lungs is consistent to their identity as alveolar macrophages. However, identification of a second population of *Mertk^{high}* cells appearing 42 days post injury and displaying a unique transcriptional signatures suggest these may be monocyte-derived moieties settling within the lung following the end of the inflammatory exacerbation. Further work needs to establish their identity since our pseudotime analysis did not fully recapitulate their origin.

Mining of a publicly available human IPF dataset (GSE136831, www.ipfcellatlas.com) offered the opportunity to add a translational value to these findings (40). Our analysis identified three macrophage clusters in the IPF lung that generate a transcriptional signature comparable to our murine results (innate immunity, interleukin mediated signaling, fibrogenic processes, and heightened transcriptional and translational control). In particular, this work confirmed the presence of *TREM2*$^+$ interstitial macrophages (cluster 0) and a population of alveolar macrophages (cluster 1) in late stage IPF (22, 80, 81). Due to the low resolution of the clustering, our findings do show a single cluster co-expressing *MERTK* and *TREM2*. At this stage, it is unclear if such population appears in established fibrosis as clinical literature seldom cites both markers. Independent of the nomenclature, our assessment is consistent with reports showing osteopontin/SPP1, fibronectin/FN1, and TGFβ1 signaling in the fibrotic niche (73, 82–84).

The importance of defining the metabolic signatures accompanying acute and chronic inflammation has great therapeutic potential (85). Our analysis identifies a robust lipid signature in steady-state alveolar macrophages, a notion consistent with their role in surfactant lipid (fluid) recycling and maintenance (86). The changes observed during SP-C^{I73T} injury are consistent with the notion that pro-fibrotic reprogramming requires a metabolic shift towards lipid consumption (87, 88).

Therapeutically noteworthy is the transcriptional signature produced by *Mertk*^{high} macrophages 42 days post injury, which included *Pparg* and several apolipoprotein receptors. Indeed, there is a large body of evidence showing that engaging this transcription factor is effective in attenuating the fibrotic phenotype (89–91). By comparison, this work leveraged the specificity of the *Apoe* signal originating from *Trem2*$^{+}$ cells and recent experimental evidence linking this lipoprotein to pro-fibrotic monocyte-derived macrophages during chemical-induced fibrosis (36, 92–94). Our findings that ApoE deletion does not provide overt benefits on *Spp1*$^{+}$ cell accumulation in the foci of injury, total inflammation, or survival is somewhat surprising. Confounding elements related to the compensatory effects of a global knock out, or the impact of ApoE deletion on a surfactant impaired system may need further examination. Despite these incongruences, our results support the value of ApoE as a biomarker indicative of the presence of *Spp1*$^{+}$ and *Fn1*$^{+}$ pro-fibrotic macrophages in the lung.

Though comprehensive, any experimental modeling of disease has limitations. While pairing murine sequencing data with human single-cell datasets offers translational value, elements related to disease staging (early inflammatory exacerbation vs. end-stage disease) and heterogeneity of human IPF etiology make this assessment less obvious. Furthermore, it is well established that use of antibody-based sorting introduces bias to the analysis (e.g., CD11b^{-}SigF^{+}CD11c^{+} in our bulk sequencing, CD45^{+}CD68^{+} for spatial analysis). To a lesser degree, clustering analysis of single-cell RNA sequencing data and cell-communication predictive tools introduce bias related to pathway annotation, and therefore any analysis attempting to describe non-canonical signaling (e.g. ApoE signaling in inflammation rather than lipid homeostasis) may not find a fitting match.

Despite any potential drawback, this work comprehensively assesses the spatial and phenotypic distribution of macrophages in pulmonary fibrosis triggered by a fibrogenic mutation in the alveolar epithelial cell restricted gene encoding for the SP-C. Our data finds peri-injury macrophages to produce an extremely active phenotype, while CD68^{+} cells localized within the fibrotic foci appear transcriptionally dormant at the peak of an inflammation exacerbation. Single cell analysis elucidated the intercellular communications occurring in the lung, while identifying *Ccr2*$^{+}$*Ly6c*$^{+}$ inflammatory monocytes and *trem2*$^{+}$ interstitial macrophages as distinct fibrogenic populations in SP-C^{I73T} induced injury. Furthermore, our work defined distinct lipid signatures among macrophage populations and propose ApoE as a potential biomarker to identify SPP1- and FN1-producing macrophages. Taken together, this work provides an essential framework for the identification (and future targeting) of deleterious macrophage populations in the early and late stages of the fibrogenic process.

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 approved by the Institutional Animal Care and Use Committee, University of Utah. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

PM: Data curation, Formal analysis, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. JC: Conceptualization, Data curation, Methodology, Validation, Writing – review & editing. SC: Data curation, Formal analysis, Methodology, Software, Visualization, Writing – review & editing. JN: Data curation, Formal analysis, Methodology, Visualization, Writing – review & editing. QW: Data curation, Methodology, Visualization, Writing – review & editing. TM: Formal analysis, Methodology, Visualization, Writing – review & editing. AV: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Grant Support: NIEHS R01ES032553 (AV); ALSAM Foundation Research Grant (AV).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1425466/full#supplementary-material

References

1. Cottin V, Hirani NA, Hotchkin DL, Nambiar AM, Ogura T, Otaola M, et al. Presentation, diagnosis and clinical course of the spectrum of progressive-fibrosing interstitial lung diseases. *Eur Respir Rev*. (2018) 27:180076. doi: 10.1183/16000617.0076-2018

2. Wijsenbeek M, Cottin V. Spectrum of fibrotic lung diseases. *N Engl J Med*. (2020) 383:958–68. doi: 10.1056/NEJMra2005230

3. Conti S, Harari S, Caminati A, Zanobetti A, Schwartz JD, Bertazzi PA, et al. The association between air pollution and the incidence of idiopathic pulmonary fibrosis in Northern Italy. *Eur Respir J*. (2018) 51(1):1700397. doi: 10.1183/13993003.00397-2017

4. Winterbottom CJ, Shah RJ, Patterson KC, Kreider ME, Panettieri RA Jr., Rivera-Lebron B, et al. Exposure to ambient particulate matter Is associated with accelerated functional decline in idiopathic pulmonary fibrosis. *Chest*. (2018) 153:1221–8. doi: 10.1016/j.chest.2017.07.034

5. Spagnolo P, Kropski JA, Jones MG, Lee JS, Rossi G, Karampitsakos T, et al. Idiopathic pulmonary fibrosis: Disease mechanisms and drug development. *Pharmacol Ther*. (2021) 222:107798. doi: 10.1016/j.pharmthera.2020.107798

6. Cui F, Sun Y, Xie J, Li D, Wu M, Song L, et al. Air pollutants, genetic susceptibility and risk of incident idiopathic pulmonary fibrosis. *Eur Respir J*. (2023) 61(2):2200777. doi: 10.1183/13993003.00777-2022

7. Sesé L, Harari S. Now we know: chronic exposure to air pollutants is a risk factor for the development of idiopathic pulmonary fibrosis. *Eur Respir J*. (2023) 61(2):2202113. doi: 10.1183/13993003.02113-2022

8. Mulugeta S, Nureki S, Beers MF. Lost after translation: insights from pulmonary surfactant for understanding the role of alveolar epithelial dysfunction and cellular quality control in fibrotic lung disease. *Am J Physiol Lung Cell Mol Physiol*. (2015) 309:L507–25. doi: 10.1152/ajplung.00139.2015

9. Venosa A, Katzen J, Tomer Y, Kopp M, Jamil S, Russo SJ, et al. Epithelial expression of an interstitial lung disease–associated mutation in surfactant protein-C modulates recruitment and activation of key myeloid cell populations in mice. *J Immunol*. (2019) 202:2760–71. doi: 10.4049/jimmunol.1900039

10. Venosa A, Cowman S, Katzen J, Tomer Y, Armstrong BS, Mulugeta S, et al. Role of CCR2(+) myeloid cells in inflammation responses driven by expression of a surfactant protein-C mutant in the alveolar epithelium. *Front Immunol*. (2021) 12:665818–8. doi: 10.3389/fimmu.2021.665818

11. Hawkins A, Guttentag SH, Deterding R, Funkhouser WK, Goralski JL, Chatterjee S, et al. A non-BRICHOS SFTPC mutant (SP-C(I73T)) linked to interstitial lung disease promotes a late block in macroautophagy disrupting cellular proteostasis and mitophagy. *Am J Physiol Lung Cell Mol Physiol*. (2015) 308:L33–47. doi: 10.1152/ajplung.00217.2014

12. Katzen J, Wagner BD, Venosa A, Kopp M, Tomer Y, Russo SJ, et al. An SFTPC BRICHOS mutant links epithelial ER stress and spontaneous lung fibrosis. *JCI Insight*. (2019) 4(6):e126125. doi: 10.1172/jci.insight.126125

13. Teoh AKY, Jo HE, Chambers DC, Symons K, Walters EH, Goh NS, et al. Blood monocyte counts as a potential prognostic marker for idiopathic pulmonary fibrosis: analysis from the Australian IPF registry. *Eur Respir J*. (2020) 55(4):1901855. doi: 10.1183/13993003.01855-2019

14. Kreuter M, Lee JS, Tzouvelekis A, Oldham JM, Molyneaux PL, Weycker D, et al. Monocyte count as a prognostic biomarker in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. (2021) 204:74–81. doi: 10.1164/rccm.202003-0669OC

15. Gibson CD, Kugler MC, Deshwal H, Munger JS, Condos R. Advances in targeted therapy for progressive fibrosing interstitial lung disease. *Lung*. (2020) 198:597–608. doi: 10.1007/s00408-020-00370-1

16. Janssen WJ, Barthel L, Muldrow A, Oberley-Deegan RE, Kearns MT, Jakubzick C, et al. Fas determines differential fates of resident and recruited macrophages during resolution of acute lung injury. *Am J Respir Crit Care Med*. (2011) 184:547–60. doi: 10.1164/rccm.201011-1891OC

17. Gibbings SL, Goyal R, Desch AN, Leach SM, Prabagar M, Atif SM, et al. Transcriptome analysis highlights the conserved difference between embryonic and postnatal-derived alveolar macrophages. *Blood*. (2015) 126:1357–66. doi: 10.1182/blood-2015-01-624809

18. Tan SY, Krasnow MA. Developmental origin of lung macrophage diversity. *Development*. (2016) 143:1318–27. doi: 10.1242/dev.129122

19. Reyfman PA, Walter JM, Joshi N, Anekalla KR, Mcquattie-Pimentel AC, Chiu S, et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am J Respir Crit Care Med*. (2019) 199:1517–36. doi: 10.1164/rccm.201712-2410OC

20. Joshi N, Watanabe S, Verma R, Jablonski RP, Chen C-I, Cheresh P, et al. A spatially restricted fibrotic niche in pulmonary fibrosis is sustained by M-CSF/M-CSFR signalling in monocyte-derived alveolar macrophages. *Eur Respir J*. (2020) 55(1):1900646. doi: 10.1183/13993003.00646-2019

21. Young LR, Gulleman PM, Short CW, Tanjore H, Sherrill T, Qi A, et al. Epithelial-macrophage interactions determine pulmonary fibrosis susceptibility in Hermansky-Pudlak syndrome. *JCI Insight*. (2016) 1:e88947. doi: 10.1172/jci.insight.88947

22. Misharin AV, Morales-Nebreda L, Reyfman PA, Cuda CM, Walter JM, Mcquattie-Pimentel AC, et al. Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J Exp Med*. (2017) 214:2387–404. doi: 10.1084/jem.20162152

23. Takahashi F, Takahashi K, Okazaki T, Maeda K, Ienaga H, Maeda M, et al. Role of osteopontin in the pathogenesis of bleomycin-induced pulmonary fibrosis. *Am J Respir Cell Mol Biol*. (2001) 24:264–71. doi: 10.1165/ajrcmb.24.3.4293

24. Dong J, Ma Q. Osteopontin enhances multi-walled carbon nanotube-triggered lung fibrosis by promoting TGF-β1 activation and myofibroblast differentiation. *Particle Fibre Toxicol*. (2017) 14:18. doi: 10.1186/s12989-017-0198-0

25. Upagupta C, Shimbori C, Alsilmi R, Kolb M. Matrix abnormalities in pulmonary fibrosis. *Eur Respir Rev*. (2018) 27:180033. doi: 10.1183/16000617.0033-2018

26. Gao X, Jia G, Guttman A, Depianto DJ, Morshead KB, Sun K-H, et al. Osteopontin links myeloid activation and disease progression in systemic sclerosis. *Cell Rep Med*. (2020) 1:100140. doi: 10.1016/j.xcrm.2020.100140

27. Hatipoglu OF, Uctepe E, Opoku G, Wake H, Ikemura K, Ohtsuki T, et al. Osteopontin silencing attenuates bleomycin-induced murine pulmonary fibrosis by regulating epithelial–mesenchymal transition. *Biomedicine Pharmacotherapy*. (2021) 139:111633. doi: 10.1016/j.biopha.2021.111633

28. Conroy LR, Clarke HA, Allison DB, Valenca SS, Sun Q, Hawkinson TR, et al. Spatial metabolomics reveals glycogen as an actionable target for pulmonary fibrosis. *Nat Commun*. (2023) 14(1):2759. doi: 10.1038/s41467-023-38437-1

29. Rajesh R, Atallah R, Bärnthaler T. Dysregulation of metabolic pathways in pulmonary fibrosis. *Pharmacol Ther*. (2023) 246:108436. doi: 10.1016/j.pharmthera.2023.108436

30. Kelly B, O'neill LA. Metabolic reprogramming in macrophages and dendritic cells in innate immunity. *Cell Res*. (2015) 25:771–84. doi: 10.1038/cr.2015.68

31. Liu Y, Xu R, Gu H, Zhang E, Qu J, Cao W, et al. Metabolic reprogramming in macrophage responses. *biomark Res*. (2021) 9:1. doi: 10.1186/s40364-020-00251-y

32. Wang S, Liu G, Li Y, Pan Y. Metabolic reprogramming induces macrophage polarization in the tumor microenvironment. *Front Immunol*. (2022) 13:840029. doi: 10.3389/fimmu.2022.840029

33. Dorotea D, Koya D, Ha H. Recent insights into SREBP as a direct mediator of kidney fibrosis via lipid-independent pathways. *Front Pharmacol*. (2020) 11. doi: 10.3389/fphar.2020.00265

34. Jaroonwitchawan T, Arimochi H, Sasaki Y, Ishifune C, Kondo H, Otsuka K, et al. Stimulation of the farnesoid X receptor promotes M2 macrophage polarization. *Front Immunol*. (2023) 14. doi: 10.3389/fimmu.2023.1065790

35. Baitsch D, Bock HH, Engel T, Telgmann R, Muller-Tidow C, Varga G, et al. Apolipoprotein E induces antiinflammatory phenotype in macrophages. *Arterioscler Thromb Vasc Biol*. (2011) 31:1160–8. doi: 10.1161/ATVBAHA.111.222745

36. Cui H, Jiang D, Banerjee S, Xie N, Kulkarni T, Liu RM, et al. Monocyte-derived alveolar macrophage apolipoprotein E participates in pulmonary fibrosis resolution. *JCI Insight*. (2020) 5(5):e134539. doi: 10.1172/jci.insight.134539

37. Venosa A, Malaviya R, Gow AJ, Hall L, Laskin JD, Laskin DL. Protective role of spleen-derived macrophages in lung inflammation, injury, and fibrosis induced by nitrogen mustard. *Am J Physiol Lung Cell Mol Physiol*. (2015) 309:L1487–1498. doi: 10.1152/ajplung.00276.2015

38. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria (2013). Available at: https://www.R-project.org/.

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

40. Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv*. (2020) 6:eaba1983. doi: 10.1126/sciadv.aba1983

41. Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience*. (2018) 7(7):giy083. doi: 10.1093/gigascience/giy083

42. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*. (2019) 20:163–72. doi: 10.1038/s41590-018-0276-y

43. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. (2019) 20:296. doi: 10.1186/s13059-019-1874-1

44. Ahlmann-Eltze C, Huber W. glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinformatics*. (2021) 36:5701–2. doi: 10.1093/bioinformatics/btaa1009

45. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell*. (2021) 184:3573–3587.e3529. doi: 10.1016/j.cell.2021.04.048

46. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. (2016) 44:W90–97. doi: 10.1093/nar/gkw377

47. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. (2014) 32:381–6. doi: 10.1038/nbt.2859

48. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. (2017) 14:979–82. doi: 10.1038/nmeth.4402

49. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. (2019) 566:496–502. doi: 10.1038/s41586-019-0969-x

50. Dimitrov D, Türei D, Garrido-Rodriguez M, Burmedi PL, Nagai JS, Boys C, et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nat Commun*. (2022) 13(1):3224. doi: 10.1038/s41467-022-30755-0

51. Dimitrov D, Schäfer PSL, Farr E, Mier PR, Lobentanzer S, Dugourd A, et al. LIANA+: an all-in-one cell-cell communication framework. *bioRxiv*. (2023) 2023.08.19.553863. doi: 10.1101/2023.08.19.553863

52. Tu J-J, Li H-S, Yan H, Zhang X-F. EnDecon: cell type deconvolution of spatially resolved transcriptomics data via ensemble learning. *Bioinformatics*. (2023) 39:btac825. doi: 10.1093/bioinformatics/btac825

53. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. (2021) 12(1):1088. doi: 10.1038/s41467-021-21246-9

54. Lane-Donovan C, Herz J. ApoE, apoE receptors, and the synapse in alzheimer's disease. *Trends Endocrinol Metab*. (2017) 28:273–84. doi: 10.1016/j.tem.2016.12.001

55. Kropski JA, Blackwell TS, Loyd JE. The genetic basis of idiopathic pulmonary fibrosis. *Eur Respir J*. (2015) 45:1717–27. doi: 10.1183/09031936.00163814

56. Cook DN, Brass DM, Schwartz DA. A matrix for new ideas in pulmonary fibrosis. *Am J Respir Cell Mol Biol*. (2002) 27:122–4. doi: 10.1165/ajrcmb.27.2.f245

57. Hardie WD, Hagood JS, Dave V, Perl AK, Whitsett JA, Korfhagen TR, et al. Signaling pathways in the epithelial origins of pulmonary fibrosis. *Cell Cycle*. (2010) 9:2769–76. doi: 10.4161/cc.9.14.12268

58. Winters NI, Burman A, Kropski JA, Blackwell TS. Epithelial injury and dysfunction in the pathogenesis of idiopathic pulmonaryFibrosis. *Am J Med Sci*. (2019) 357:374–8. doi: 10.1016/j.amjms.2019.01.010

59. Nureki SI, Tomer Y, Venosa A, Katzen J, Russo SJ, Jamil S, et al. Expression of mutant Sftpc in murine alveolar epithelia drives spontaneous lung fibrosis. *J Clin Invest*. (2018) 128:4008–24. doi: 10.1172/JCI99287

60. Ciccimarra R, Bolognesi MM, Zoboli M, Cattoretti G, Stellari FF, Ravanetti F. The normal and fibrotic mouse lung classified by spatial proteomic analysis. *Sci Rep*. (2022) 12(1):8742. doi: 10.1038/s41598-022-12738-9

61. Blumhagen RZ, Kurche JS, Cool CD, Walts AD, Heinz D, Fingerlin TE, et al. Spatially distinct molecular patterns of gene expression in idiopathic pulmonary fibrosis. *Respir Res*. (2023) 24:287. doi: 10.1186/s12931-023-02572-6

62. Nuovo GJ, Hagood JS, Magro CM, Chin N, Kapil R, Davis L, et al. The distribution of immunomodulatory cells in the lungs of patients with idiopathic pulmonary fibrosis. *Modern Pathol*. (2012) 25:416–33. doi: 10.1038/modpathol.2011.166

63. Eyres M, Bell JA, Davies ER, Fabre A, Alzetani A, Jogai S, et al. Spatially resolved deconvolution of the fibrotic niche in lung fibrosis. *Cell Rep*. (2022) 40:111230. doi: 10.1016/j.celrep.2022.111230

64. Wu Q, Zhang K-J, Jiang S-M, Fu L, Shi Y, Tan R-B, et al. p53: A key protein that regulates pulmonary fibrosis. *Oxid Med Cell Longevity*. (2020) 2020:6635794. doi: 10.1155/2020/6635794

65. Chandran RR, Xie Y, Gallardo-Vara E, Adams T, Garcia-Milian R, Kabir I, et al. Distinct roles of KLF4 in mesenchymal cell subtypes during lung fibrogenesis. *Nat Commun*. (2021) 12(1):7179. doi: 10.1038/s41467-021-27499-8

66. Lee C-M, Cho SJ, Cho W-K, Park JW, Lee J-H, Choi AM, et al. Laminin α1 is a genetic modifier of TGF-β1–stimulated pulmonary fibrosis. *JCI Insight*. (2018) 3(18): e99574. doi: 10.1172/jci.insight.99574

67. Uhl K, Paithankar S, Leshchiner D, Jager TE, Abdelgied M, Dixit B, et al. Differential transcriptomic signatures of small airway cell cultures derived from IPF and COVID-19-induced exacerbation of interstitial lung disease. *Cells*. (2023) 12 (20):2501. doi: 10.3390/cells12202501

68. Nguyen J, Deering-Rice CE, Armstrong BS, Massa C, Reilly CA, Venosa A. Parenchymal and inflammatory cell responses to single and repeated ozone exposure in healthy and surfactant protein-C mutant lung. *Toxicol Sci*. (2022) 189:107–23. doi: 10.1093/toxsci/kfac074

69. Trzebanski S, Jung S. Plasticity of monocyte development and monocyte fates. *Immunol Lett*. (2020) 227:66–78. doi: 10.1016/j.imlet.2020.07.007

70. Minutti CM, Jackson-Jones LH, García-Fojeda B, Knipper JA, Sutherland TE, Logan N, et al. Local amplifiers of IL-4Rα-mediated macrophage activation promote repair in lung and liver. *Science*. (2017) 356:1076–80. doi: 10.1126/science.aaj2067

71. Casals C, García-Fojeda B, Minutti CM. Soluble defense collagens: Sweeping up immune threats. *Mol Immunol*. (2019) 112:291–304. doi: 10.1016/j.molimm.2019.06.007

72. Sikkeland LIB, Ueland T, Lund MB, Durheim MT, Mollnes TE. A role for the terminal C5-C9 complement pathway in idiopathic pulmonary fibrosis. *Front Med*. (2023) 10. doi: 10.3389/fmed.2023.1236495

73. Morse C, Tabib T, Sembrat J, Buschur KL, Bittar HT, Valenzi E, et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J*. (2019) 54(2):1802441. doi: 10.1183/13993003.02441-2018

74. Matsubara E, Komohara Y, Esumi S, Shinchi Y, Ishizuka S, Mito R, et al. SPP1 derived from macrophages is associated with a worse clinical course and chemoresistance in lung adenocarcinoma. *Cancers (Basel)*. (2022) 14(18):4374. doi: 10.3390/cancers14184374

75. Wang J, Zhang L, Luo L, He P, Xiong A, Jiang M, et al. Characterizing cellular heterogeneity in fibrotic hypersensitivity pneumonitis by single-cell transcriptional analysis. *Cell Death Discovery*. (2022) 8:38. doi: 10.1038/s41420-022-00831-x

76. Han H, Ge X, Komakula SSB, Desert R, Das S, Song Z, et al. Macrophage-derived osteopontin (SPP1) protects from nonalcoholic steatohepatitis. *Gastroenterology*. (2023) 165:201–17. doi: 10.1053/j.gastro.2023.03.228

77. Hoeft K, Schaefer GJL, Kim H, Schumacher D, Bleckwehl T, Long Q, et al. Platelet-instructed SPP1(+) macrophages drive myofibroblast activation in fibrosis in a CXCL4-dependent manner. *Cell Rep*. (2023) 42:112131. doi: 10.1016/j.celrep.2023.112131

78. Moore PK, Anderson KC, Mcmanus SA, Tu TH, King EM, Mould KJ, et al. Single-cell RNA sequencing reveals unique monocyte-derived interstitial macrophage subsets during lipopolysaccharide-induced acute lung inflammation. *Am J Physiol Lung Cell Mol Physiol*. (2023) 324:L536–l549. doi: 10.1152/ajplung.00223.2022

79. Chakraborty S, Singh A, Wang L, Wang X, Sanborn MA, Ye Z, et al. Trained immunity of alveolar macrophages enhances injury resolution via KLF4-MERTK-mediated efferocytosis. *J Exp Med*. (2023) 220(11):e20221388. doi: 10.1084/jem.20221388

80. Kotlyarov S. Participation of ABCA1 transporter in pathogenesis of chronic obstructive pulmonary disease. *Int J Mol Sci*. (2021) 22(7):3334. doi: 10.3390/ijms22073334

81. Yang HQ, Sun H, Li K, Shao MM, Zhai K, Tong ZH. Dynamics of host immune responses and a potential function of Trem2(hi) interstitial macrophages in Pneumocystis pneumonia. *Respir Res*. (2024) 25:72. doi: 10.1186/s12931-024-02709-1

82. Cai B, Dongiovanni P, Corey KE, Wang X, Shmarakov IO, Zheng Z, et al. Macrophage merTK promotes liver fibrosis in nonalcoholic steatohepatitis. *Cell Metab*. (2020) 31:406–421.e407. doi: 10.1016/j.cmet.2019.11.013

83. Lv J, Gao H, Ma J, Liu J, Tian Y, Yang C, et al. Dynamic atlas of immune cells reveals multiple functional features of macrophages associated with progression of pulmonary fibrosis. *Front Immunol*. (2023) 14. doi: 10.3389/fimmu.2023.1230266

84. She Y, Xu X, Yu Q, Yang X, He J, Tang XX. Elevated expression of macrophage MERTK exhibits profibrotic effects and results in defective regulation of efferocytosis function in pulmonary fibrosis. *Respir Res*. (2023) 24:118. doi: 10.1186/s12931-023-02424-3

85. Li J, Zhai X, Sun X, Cao S, Yuan Q, Wang J. Metabolic reprogramming of pulmonary fibrosis. *Front Pharmacol*. (2022) 13:1031890. doi: 10.3389/fphar.2022.1031890

86. Poelma DL, Ju MR, Bakker SC, Zimmermann LJ, Lachmann BF, Van Iwaarden JF. A common pathway for the uptake of surfactant lipids by alveolar cells. *Am J Respir Cell Mol Biol*. (2004) 30:751–8. doi: 10.1165/rcmb.2003-0127OC

87. Wculek SK, Dunphy G, Heras-Murillo I, Mastrangelo A, Sancho D. Metabolism of tissue macrophages in homeostasis and pathology. *Cell Mol Immunol*. (2022) 19:384–408. doi: 10.1038/s41423-021-00791-9

88. Vassiliou E, Farias-Pereira R. Impact of lipid metabolism on macrophage polarization: implications for inflammation and tumor immunity. *Int J Mol Sci*. (2023) 24(15):12032. doi: 10.3390/ijms241512032

89. Chandran S, Schilke RM, Blackburn CMR, Yurochko A, Mirza R, Scott RS, et al. Lipin-1 contributes to IL-4 mediated macrophage polarization. *Front Immunol*. (2020) 11. doi: 10.3389/fimmu.2020.00787

90. Zhou F, Fan X, Miao Y. LPIN1 promotes triglycerides synthesis and is transcriptionally regulated by PPARG in buffalo mammary epithelial cells. *Sci Rep*. (2022) 12(1):2390. doi: 10.1038/s41598-022-06114-w

91. Zeng Q, Zhou T-T, Huang W-J, Huang X-T, Huang L, Zhang X-H, et al. Asarinin attenuates bleomycin-induced pulmonary fibrosis by activating PPARγ. *Sci Rep*. (2023) 13:14706. doi: 10.1038/s41598-023-41933-5

92. Venosa A, Malaviya R, Choi H, Gow AJ, Laskin JD, Laskin DL. Characterization of distinct macrophage subpopulations during nitrogen mustard-induced lung injury and fibrosis. *Am J Respir Cell Mol Biol*. (2016) 54:436–46. doi: 10.1165/rcmb.2015-0120OC

93. Gordon EM, Yao X, Xu H, Karkowsky W, Kaler M, Kalchiem-Dekel O, et al. Apolipoprotein E is a concentration-dependent pulmonary danger signal that activates the NLRP3 inflammasome and IL-1β secretion by bronchoalveolar fluid macrophages from asthmatic subjects. *J Allergy Clin Immunol*. (2019) 144:426–441.e423. doi: 10.1016/j.jaci.2019.02.027

94. Venosa A, Smith LC, Murray A, Banota T, Gow AJ, Laskin JD, et al. Regulation of macrophage foam cell formation during nitrogen mustard (NM)-induced pulmonary fibrosis by lung lipids. *Toxicol Sci*. (2019) 172:344–58. doi: 10.1093/toxsci/kfz187

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 30 October 2024
DOI 10.3389/fimmu.2024.1393017

OPEN ACCESS

EDITED BY
Ying Zhu,
Genentech Inc., United States

REVIEWED BY
Diego Velasco,
Universidad Carlos III de Madrid de Madrid, Spain
Shiang-Jong Tzeng,
National Taiwan University, Taiwan

*CORRESPONDENCE
Seçil Vural
sevural@ku.edu.tr
Atay Vural
atayvural@ku.edu.tr

†These authors have contributed equally to this work and share first authorship

‡These authors have contributed equally to this work and share senior authorship

RECEIVED 28 February 2024
ACCEPTED 08 October 2024
PUBLISHED 30 October 2024

CITATION
Khoshbakht S, Albayrak Ö, Tiryaki E, Ağcaoğlu O, Öktem A, Pınar Sun G, Er Gülbezer E, Ertekin SS, Boyvat A, Vural A and Vural S (2024) A cost-effective protocol for single-cell RNA sequencing of human skin.
Front. Immunol. 15:1393017.
doi: 10.3389/fimmu.2024.1393017

A cost-effective protocol for single-cell RNA sequencing of human skin

Saba Khoshbakht[1†], Özgür Albayrak[2†], Ergün Tiryaki[1], Orhan Ağcaoğlu[3], Ayşe Öktem[4], Gizem Pınar Sun[5], Elif Er Gülbezer[6], Sümeyre Seda Ertekin[7], Ayşe Boyvat[4], Atay Vural[2,8*‡] and Seçil Vural[2,7*‡]

[1]Graduate School of Health Sciences, Koç University, Istanbul, Türkiye, [2]Koç University Research Center for Translational Medicine, Koç University, Istanbul, Türkiye, [3]Department of Surgery, Koç University School of Medicine, Istanbul, Türkiye, [4]Department of Dermatology, Ankara University Faculty of Medicine, Ankara, Türkiye, [5]Department of Dermatology, Başakşehir Çam ve Sakura Şehir Hastanesi, Istanbul, Türkiye, [6]Department of Rheumatology, Koç University School of Medicine, Istanbul, Türkiye, [7]Department of Dermatology, Koç University School of Medicine, Istanbul, Türkiye, [8]Department of Neurology, Koç University School of Medicine, Istanbul, Türkiye

Introduction: Single-cell RNA sequencing (scRNAseq) and flow cytometry studies in skin are methodologically complex and costly, limiting their accessibility to researchers worldwide. Ideally, RNA and protein-based analyses should be performed on the same lesion to obtain more comprehensive data. However, current protocols generally focus on either scRNAseq or flow cytometry of healthy skin.

Methods: We present a novel label-free sample multiplexing strategy, building on the souporcell algorithm, which enables scRNAseq analysis of paired blood and skin samples. Additionally, we provide detailed instructions for simultaneous flow cytometry analysis from the same sample, with necessary adaptations for both healthy and inflamed skin specimens.

Results: This tissue multiplexing strategy mitigates technical batch effects and reduces costs by 2-4 times compared to existing protocols. We also demonstrate the effects of varying enzymatic incubation durations (1, 3, and 16 hours, with and without enzyme P) on flow cytometry outcomes. Comprehensive explanations of bioinformatic demultiplexing steps and a detailed step-by-step protocol of the entire experimental procedure are included.

Discussion: The protocol outlined in this article will make scRNAseq and flow cytometry analysis of skin samples more accessible to researchers, especially those new to these techniques.

KEYWORDS

skin, inflammation, single cell RNA sequencing (scRNA), flow cytometry, souporcell, multiplexing, skin dissociation

Introduction

Analyzing RNA and protein at the single-cell level in lesional skin is crucial for understanding the immune mechanisms underlying dermatological disorders and identifying novel therapeutic targets. Recent advancements in standardized tissue dissociations and single-cell RNA sequencing (scRNAseq) systems have facilitated studies on diseases such as psoriasis and atopic dermatitis using single-cell analysis methods (1–4). These studies have significantly enhanced our understanding of the immune mechanisms involved in these disorders. However, the complex methodology and high costs associated with these studies limit their accessibility, preventing many laboratories from conducting such experiments and hindering the application of these studies to other dermatological disorders. Therefore, there is a pressing need for cost-effective and detailed experimental protocols to make these techniques more accessible to researchers.

Obtaining high-quality cells with intact RNA and protein epitopes from solid tissues has been a significant challenge for single-cell studies. Numerous protocols for tissue dissociation have been documented in the scientific literature (5–7). The introduction of automated tissue dissociator systems has further facilitated flow cytometry and scRNAseq analysis of dissociated skin cells (8, 9). Recently, comprehensive methodological papers have provided efficient and optimized protocols for scRNAseq and flow cytometry analysis of human and pig skin, which we recommend for further reading (5, 10–12). However, these papers predominantly focus on either scRNAseq or flow cytometry, without testing both methods on the same sample, and primarily concentrate on healthy skin. Inflamed skin, which typically has a higher cell count than healthy skin, may require different experimental conditions.

High-throughput single-cell multi-omics methods, such as CITE-Seq, have recently been developed to simultaneously study RNA and protein. However, these techniques are less sensitive to dim cell surface markers. Conventional and spectral flow cytometry, as well as mass cytometry, remain the gold standards for single-cell protein-based studies. Preservation of cell surface epitopes is essential for these methodologies and can be influenced by enzyme selection and prolonged enzymatic incubation periods. Studies systematically comparing the effects of enzymatic incubation durations are scarce, and none have examined the impact of enzyme P, commonly used in cell dissociation protocols to increase cell yield in scRNAseq studies.

Sample multiplexing is vital for reducing technical batch effects and experimental costs in scRNAseq studies. Traditional sample multiplexing relies on oligonucleotide-conjugated hashtag antibodies. However, this approach has limitations, including the risk of inefficient antibody binding to all cells, additional material requirements for both multiplexing and demultiplexing steps, and increased experimental complexity and costs. Existing protocols often lack detailed steps for multiplexing samples, library generation for barcodes, and demultiplexing methods. In our experience, initial attempts at sample multiplexing in our laboratory faced significant challenges due to insufficient detail in existing protocols, risking the loss of samples and materials, which prompted us to look for alternative sample multiplexing and demultiplexing methods.

Recently, algorithms that recognize individual-specific single nucleotide polymorphism signatures have been developed for label-free demultiplexing of pooled samples. One such method, souporcell, has shown excellent benchmarking results against antibody-based cell hashing techniques (13). However, the complex experimental design proposed in the original article, which involves splitting the same sample into multiple experimental batches, has limited its usage in research projects. A novel experimental design is needed to make this method more widely applicable.

In this study, we aimed to establish a practical and cost-effective protocol for single-cell protein and RNA analysis. We provide tailored instructions for both healthy and lesional skin, demonstrating the influence of various enzymatic incubation conditions on flow cytometry analysis. Importantly, we introduce a novel, two-layered sample multiplexing design that combines two complementary strategies from the literature, along with demultiplexing steps, for scRNAseq experiments. This approach can lower the experimental costs 2-4 fold. Detailed explanations for each optimization step are provided, allowing researchers to adapt and customize protocols according to their specific needs (Supplementary Material).

Materials and methods

Experimental design

Our experimental design consists of two main parts. The first part includes the acquisition of skin samples, followed by tissue dissociation and flow cytometry analysis of freshly isolated cells. A fraction of dissociated skin cells are cryopreserved for future scRNAseq analysis at the end of this step (Figure 1). The second part begins with sample preparation for scRNAseq analysis by cell sorting, followed by gel beads-in-emulsion (GEM) generation, library construction, sequencing, and data analysis (Figure 1). A step-by-step protocol is presented in the Supplementary Materials.

Sample collection

The study was approved by the Koç University Committee on Human Research (protocol number 2022.058.IRB2.007). Informed consent was obtained from all participants.

Healthy skin samples were obtained by surgical excision of excess skin tissue from individuals undergoing surgery (n=6). Tissues were kept in cold tissue storage solution (130-100-008, Miltenyi Biotec, Germany) until delivered to the laboratory on ice. From each large surgical excision, eight 6 mm punch biopsies were obtained. These punch biopsies were then distributed as two pieces for each experimental condition, including three different enzyme incubation periods and treatment with enzyme P. Lesional skin biopsy specimens were collected from patients with Behçet's disease (BD, n=12). A 4 mm punch was used to collect skin biopsy samples

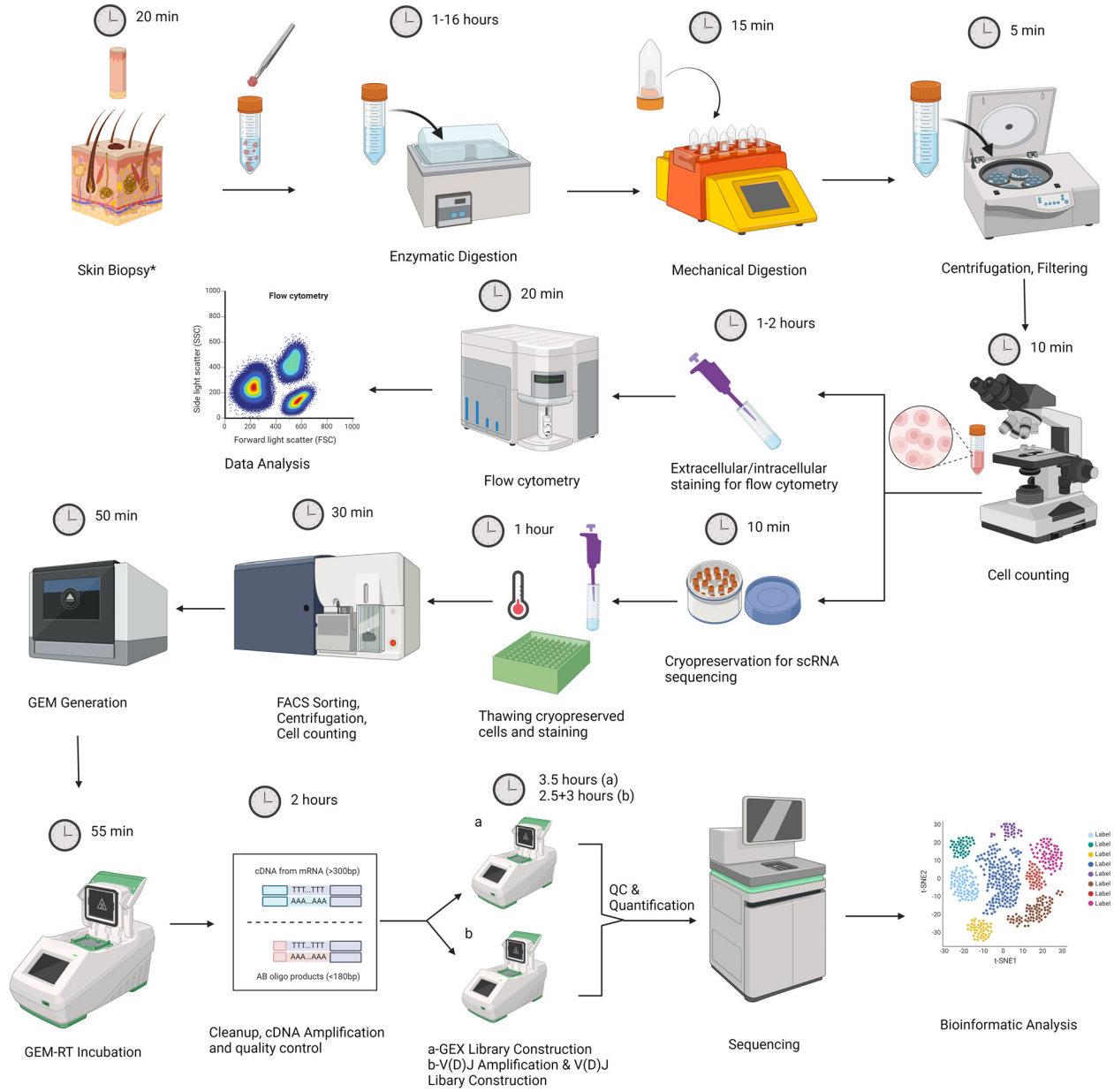


FIGURE 1
Graphic representation of our experimental workflow that involves skin biopsy, tissue dissociation using the Whole Skin Dissociation Kit (Miltenyi Biotec, Germany) in 2 steps of enzymatic and mechanical dissociation, staining part of the dissociated cells for flow cytometry analysis, and cryopreservation part of cells for subsequent scRNAseq analysis. Second part includes thawing of cell suspensions, sample preparation by Fluorescence-activated cell sorting (FACS), cell counting by trypan blue exclusion method, Gel Beads-in-emulsion (GEM) generation, library preparation, sequencing and data analysis. The figure was created with BioRender.com. *For lesional skin, one 4mm punch biopsy was obtained. For healthy skin samples, two 6 mm specimens were obtained.

from the lesional area. Demographic and clinical characteristics of the study samples are demonstrated in Table 1.

For the isolation of peripheral blood mononuclear cells (PBMCs), venous blood was diluted 1:1 with phosphate-buffered saline (PBS) and spread over an equal volume of Lymphoprep 1.077g/ml density gradient (Axis-Shield, Norway) in 50 ml Falcon tubes. The tubes were centrifuged at 500g for 30 minutes at room temperature without brakes. Following centrifugation, the PBMC layer was transferred into another 50 ml sterile tube and washed with PBS containing an equal volume of 1% BSA. A fraction of cells were analyzed freshly by flow cytometry and the remaining cells were frozen in the cryopreservation solution (10% DMSO, 90% FBS) for later usage.

Skin dissociation

After the delivery of skin samples, the tissue was washed thoroughly with PBS, and subcutaneous tissue was removed with a scalpel, paying attention to keeping the dermis intact. Then, the

TABLE 1 Demographic and clinical characteristics of the study samples, including Behcet's disease (BD) patients and healthy controls.

Characteristic	BD (n=12)	HC (n=6)
Age: median (Interquartile Range)	35 (30-41)	70 (65-73)
Gender (Male: Female)	6:6	0:6
Smoking	3/12	1/6
HLA-B51 positivity	4/9	–
Lesion Type:		
Erythema Nodosum-like lesions (EN)	3	–
Papulopustular Eruptions (PPE)	6	–
Genital Ulcers (GU)	3	–

Data for age are presented as median with interquartile range (IQR). Gender is reported as a ratio of male to female participants. Smoking status and HLA-B51 positivity are also included. For patients, lesion types are categorized as erythema nodosum-like lesions (EN), papulopustular eruptions (PPE), and genital ulcers (GU).

skin tissues were dissociated enzymatically and mechanically using the Whole Skin Dissociation Kit (130-101-540, Miltenyi Biotec, Germany) and gentleMACS Octo Dissociator with Heaters (Miltenyi Biotec, Germany) following the manufacturer's protocol. The kit recommends the use of three enzymes (enzymes A, D, and P) for enzymatic dissociation, with the usage of enzyme P being optional depending on the subsequent analysis method. Enzyme P is known to cause cleavage of some extracellular epitopes, which may interfere with flow cytometry analysis. The duration of enzymatic incubation can be either 3 hours or overnight as per the manufacturer's manual. To investigate the effect of enzyme P and various incubation durations (1h, 3h, and 16h) on flow cytometry results, we first conducted a systematic study, as these factors have not been studied comprehensively before.

The enzymatic steps were carried out as follows: 435 µL of Buffer L and, if applicable, 12.5 µL of Enzyme P were combined in a gentleMACS C tube. Subsequently, 50 µL of Enzyme D and 2.5 µL of Enzyme A were added to the mixture, which was then thoroughly mixed. The tissue-enzyme mixture was then placed in a 37°C water bath for incubation. We compared three different incubation durations: 1, 3, and 16 hours.

After incubation, 500µL of cold DMEM (Gibco, USA) was added to the mixture, and tubes were placed onto the gentleMACS system. The "h_skin_01" program was initiated to mechanically dissociate the tissues into cell suspension. Following completion of the program, samples were briefly centrifuged and filtered using a 70 µm cell strainer (83.3945.070, Sarstedt, Germany), with 4 ml of DMEM used to wash the cells.

After centrifugation at 350g for 10 minutes, the cell pellet was counted using a hemocytometer with 0.4% Trypan blue to exclude dead cells. Approximately 10^5 cells (in 100 µL buffer) were stained and analyzed by flow cytometry on the same day, while the remaining cells were cryopreserved (in 1 mL fetal bovine serum plus 10% DMSO) for subsequent scRNAseq analysis.

Flow cytometry

All antibodies and the fixable viability dye used in this study are listed in Supplementary Table 1. Initially, 10^5 freshly isolated skin cells or PBMCs were incubated with Zombie NIR fixable viability dye for 10 minutes on ice. Subsequently, 2 ml of FACS buffer (PBS+ 1% BSA) was added to the tubes, followed by centrifugation at 500 g for 5 minutes. After centrifugation, the supernatant was decanted, and a cell surface antibody cocktail was added. Samples were then incubated for 20 minutes on ice. Post-incubation, samples were washed again with 2 ml of FACS buffer and the supernatant was removed. For intracellular staining, cells were fixed with 500 µL of Fixation Buffer (420801, BioLegend, USA) for 20 minutes at room temperature and directly centrifuged at 500 g for 5 minutes. Following the removal of the supernatant, cells were washed with Intracellular Staining Permeabilization Wash Buffer (421002, BioLegend, USA) and incubated with an intracellular antibody cocktail for 20 minutes at room temperature. After the incubation period, cells were washed with the Permeabilization Wash Buffer and the pellet was resuspended with 500 µL of FACS buffer. The acquisition was performed using a CytoFLEX SRT (Beckman Coulter) flow cytometer, and the results were analyzed using FlowJo v.10.9.0 (BD Biosciences, USA). The markers used to identify the cell populations of interest are listed in Supplementary Table 2.

Sample preparation for single-cell RNA sequencing

For successful single-cell RNA sequencing using the 10X Chromium system, proper sample preparation is paramount. The viability and concentration of cells for GEM generation must adhere to the manufacturer's protocol. A live cell ratio exceeding 90% is highly advisable, although ratios above 60-70% can also be attempted with potentially reduced success rates.

In this study, we adopted a sample pooling strategy to mitigate batch effects and lower experimental costs by combining two or more samples. Initially, cryopreserved cell suspensions from skin and peripheral blood were thawed. Subsequently, cells were stained with fixable viability dye (Zombie NIR) in PBS, followed by anti-human CD45 PE-Cy5 in FACS buffer (PBS+ 1% BSA). CD45+ live cells were then simultaneously sorted from two samples using the CytoFLEX SRT (Beckman Coulter) and FACS Aria III cell sorters (BD Biosciences, USA). During sorting, a 100 µm nozzle was utilized, with a constant pressure of 20 psi, and the sample chamber was maintained at 4°C. The sorting speed was kept low (1500 events/seconds) to ensure high viability and purity.

The manufacturer's protocol for the Chromium Next GEM Single Cell 5' Reagent Kit V2 (Dual Index, 10X Genomics, USA) used for this study suggests an ideal cell concentration of 700 to 1200 cells per microliter for GEM generation. However, cell sorting from skin tissue often results in concentrations lower than this range. In such cases, it is necessary to concentrate the cell

suspension after sorting. In the current study, this was accomplished by an additional centrifugation step at 850 g for 5 minutes. After carefully removing the supernatant, the cells were reconstituted in the desired volume.

Library preparation and sequencing

Once the sample preparation step was completed by reaching the optimum cell concentration, steps for cDNA preparation, amplification, and library preparation were performed by using the Chromium Controller and the Chromium Next GEM Single Cell 5' Reagent Kits V2 (Dual Index, 10X Genomics, USA).

First, GEMs were generated by loading a master mix containing cells, gel beads, and partitioning oil on the Chromium Next GEM Chip K and running the Chromium Controller system. Next, gel beads were dissolved, cells were lysed and 10X barcoded cDNAs were produced from poly-adenylated mRNAs. The barcoded cDNAs were purified from the reaction mixture using Silane magnetic beads and amplified via PCR. The quality control (QC) and quantification were performed by Agilent 2100 Bioanalyzer (Agilent Technologies, USA). For Gene Expression (GEX) library construction enzymatic fragmentation, size selection was performed to get the optimal cDNA amplicon lengths. Finally, sequencing-ready dual index libraries were prepared by end repair, A-tailing, adaptor ligation, and sample index PCR.

Paired-end sequencing was performed on the Illumina NovaSeq 6000 platform by outsourcing to a service provider. A minimum reading depth of 20,000 reads per cell was utilized.

Alignment and demultiplexing of single-cell RNAseq data

The raw FASTQ files were aligned to the reference genome (GRCh38) using Cell Ranger (v.7.1.0) multi pipeline. Both reference genome and Cell Ranger software were obtained from the official 10X Genomics website.

We used souporcell for demultiplexing of pooled samples (13). The BAM file generated by Cell Ranger was plugged into the souporcell pipeline. This method utilizes single nucleotide polymorphisms (SNPs) detected in scRNA-seq reads to demultiplex scRNA-seq data originating from different individuals (13). This *in silico* method is freely available and label-free. The souporcell pipeline was executed on a high-performance computing cluster using the singularity image provided by souporcell authors. The analysis employed the "*souporcell_pipeline.py*" script and utilized the same reference transcriptome used during alignment.

Identification of paired samples and donor sex in dual reactions

After demultiplexing of pooled samples by souporcell, the identity of each cluster was determined based on: 1) The presence of matched skin and PBMC pairs, which have the same genotype in dual reactions, and 2) the Identification of the donor's sex. Skin and PBMC samples from the same donor were identified using the "*shared_samples.py*" module of the souporcell. The donor's sex for each souporcell cluster was determined by using a subset of Y chromosome genes including ZFY, RPS4Y1, EIF1AY, KDM5D, NLGN4Y, TMSB4Y, UTY, DDX3Y, and USP9Y (14).

Analysis of single-cell RNAseq data

After alignment and demultiplexing, count matrices were processed using Scanpy. Cells with less than 200 genes and genes expressed in fewer than 3 cells were filtered. Moreover, cells containing more than 4000 genes were excluded. In the doublet detection part, cross-genotype doublets, as detected by souporcell, were eliminated, and Scrublet was employed to identify doublets based on expression profiles.

Subsequently, the data was normalized to 10,000, and the log(x +1) transformation was applied. Feature selection was performed using the "sc.pp.highly_variable_genes" function, and principal component analysis (PCA) was computed on the scaled expression matrix of the highly variable genes. BBKNN's ridge regression function was utilized to eliminate technical confounders such as donor-specific variation and count depth, while preserving biological variation such as cell types and disease effects (15, 16). Following this preprocessing step, different pools were integrated using the Harmony algorithm (17). A neighborhood graph and UMAP embedding were computed based on the harmony-corrected principal components.

Cell type annotation was done by CellTypist using pre-trained built-in models such as "Immune_All_High.pkl" and "Immune_All_Low.pkl" (18). During cell type prediction, the majority voting classifier is enabled to increase the accuracy of annotation.

Statistical analysis

Repeated Measures One-Way ANOVA was used to compare different incubation durations. Paired T-test was used to assess the effect of enzyme P. Statistical analysis was conducted by GraphPad Prism v8 (GraphPad Software, USA).

Results

Identification of immune cell subsets and intracellular cytotoxic molecules in skin by flow cytometry

We initially dissociated healthy skin samples by incubating them with enzymes A and D (without enzyme P) for three hours and analyzed them freshly after staining with flow cytometry. Our gating strategy and fluorescence-minus-one (FMO) controls used for dimly expressed markers are illustrated in Supplementary

Figures 1, 2, respectively. Using the antibody panel in this study, we could readily identify T cell subsets, including CD4+ T helper cells, CD8+ cytotoxic T cells, CD69+ resident memory T cells, as well as natural killer (NK) cells. Additionally, we were able to detect intracellular granzyme B and granzyme K expression in cytotoxic T cells.

The impact of enzymatic incubation duration and enzyme P on cell counts and cell type frequencies

The impact of varying incubation periods and the presence of enzyme P on selected extracellular and intracellular markers is illustrated in Figure 2. We observed a significant increase in the yield of trypan blue-negative live cells with longer incubation durations (Figure 3A). Specifically, samples subjected to a 1-hour incubation displayed a live cell count ranging from 5.5×10^4 to 1.22×10^5 per one 6 mm punch specimen, which was deemed suboptimal for subsequent applications. In contrast, samples incubated for 3 hours yielded a live cell count ranging from 7.5×10^4 to 1.5×10^5 per one 6 mm punch specimen, while a 16-hour incubation resulted in a live cell count ranging from 1.15×10^5 to 2.5×10^5 per one 6 mm punch specimen.

The percentage of CD45+ leukocytes was higher after 16 hours of incubation compared to 3 hours. The frequency of T cells, NK cells, CD4+ T helper lymphocytes (THL), CD8+ cytotoxic T lymphocytes (CTL), double negative cells, and CD69+ resident memory T (TRM) cells remained similar between groups (Figure 3A). However, there was a decline in the percentage of granzyme B-positive cytotoxic T lymphocytes (CTLs) after 16 hours of incubation compared to 1 hour (median, range: 24.8, 14.5-30.3 vs 13.8, 11.6-24.0; p=0.027), while the ratio of granzyme K-expressing CTLs remained unchanged. The mean fluorescent intensity (MFI) of CD3, CD8, CD56, CD69, and granzyme K was similar between groups. However, there was a significant decrease in the MFI of CD4 with longer enzymatic incubation periods (1h vs 16h: 33169, 13566-45794 vs 6387, 4494-12979; p=0.0077, 3h vs 16h: 21511, 9628-28753 vs 6387, 4494-12979; p=0.0071), and a decline was seen in Granzyme B MFI after 16 hours of incubation (1h vs 16h, p=0.042) (Figure 3B).

We investigated the impact of enzyme P during a 3-hour incubation period. The live cell count per one 6 mm punch specimen after 3 hours of incubation without enzyme P ranged from 7.5×10^4 to 1.5×10^5, while with the addition of enzyme P, it increased significantly to 8.5×10^4 to 1.85×10^5 ($p<0.0001$) (Figure 4A). The use of enzyme P did not significantly alter the percentages of T cells, CD4+ T cells, granzyme K positive CTLs, or granzyme B positive CTLs. However, the percentages of CD8+ CTLs (41.95, 27.2-71.7 vs 1.44, 0.83-2.67; p=0.005) and CD69+ TRM cells (92.7, 81.8-93.8 vs 11.27, 0.52-22; p=0.044) decreased dramatically. Additionally, there was a statistically significant reduction in the percentage of NK cells after incubation with enzyme P (5.03, 1.90-6.48 vs 1.25, 0.82-4.33; p=0.026) (Figure 4A). Moreover, as anticipated, there was a significant decrease in the MFI of CD8 (42108, 39568-49811 vs 18090, 16887-23848; p=0.0009), CD56 (7239, 4347-9927 vs 3964, 3716-4110; p=0.039), and CD69 (2598, 1606-3370 vs 1515, 1339-1911; p=0.038) in the group treated with enzyme P compared to the group without enzyme P (Figure 4B).

Based on the results of these experiments, we decided to use a 3-hour enzymatic incubation without enzyme P for the analysis of lesional skin for simultaneous flow cytometry and scRNAseq analysis.

Determination of tissue size required for flow cytometry and scRNAseq analysis in the inflamed skin

In prior methodological studies, the necessary size of skin tissue for scRNAseq analysis was determined. Given the higher concentration of inflammatory cells in inflamed skin, we hypothesized that a smaller tissue sample would be adequate for subsequent analysis compared to healthy skin.

For healthy skin specimens, a single 6 mm punch biopsy provided an average cell count of 1.125×10^5 after a three-hour enzymatic incubation without enzyme P (range: 0.75×10^5 to 1.5×10^5, n=6). This quantity is sufficient for flow cytometry analysis of lymphocytes. However, if scRNAseq is also intended, we recommend using a second biopsy specimen.

Conversely, we discovered that after three hours of enzyme incubation without enzyme P, inflamed skin tissue measuring 4 mm in diameter yielded an average of 4.6×10^5 live cells (range: 8×10^4 - 9.7×10^5, n=12). Despite some variability between samples, this amount is typically adequate for both flow cytometry and scRNAseq analysis.

The effect of cryopreservation on dissociated skin cells

Sample multiplexing can help avoid technical batch effects and reduce the cost of single-cell RNA sequencing (scRNAseq). However, multiplexing freshly isolated cells can be challenging due to the need for synchronized sample collection from different individuals. Cryopreservation of isolated cell suspensions offers a solution to this issue. In our study, we stored dissociated skin cells in liquid nitrogen for subsequent cell sorting and scRNAseq analyses. This approach allowed us to compare the viability of CD45+ cells in freshly isolated and thawed samples. We found that the live CD45+ cell percentage was consistently above 85% in both freshly isolated and thawed skin cells, with no statistically significant difference between the two groups (Supplementary Figure 3).

Sample multiplexing strategy for batch scRNAseq analysis of paired blood and skin samples

Current sample multiplexing method for scRNAseq analysis rely on the usage of sample barcoding kits. In this study, our aim was to

10.3389/fimmu.2024.1393017

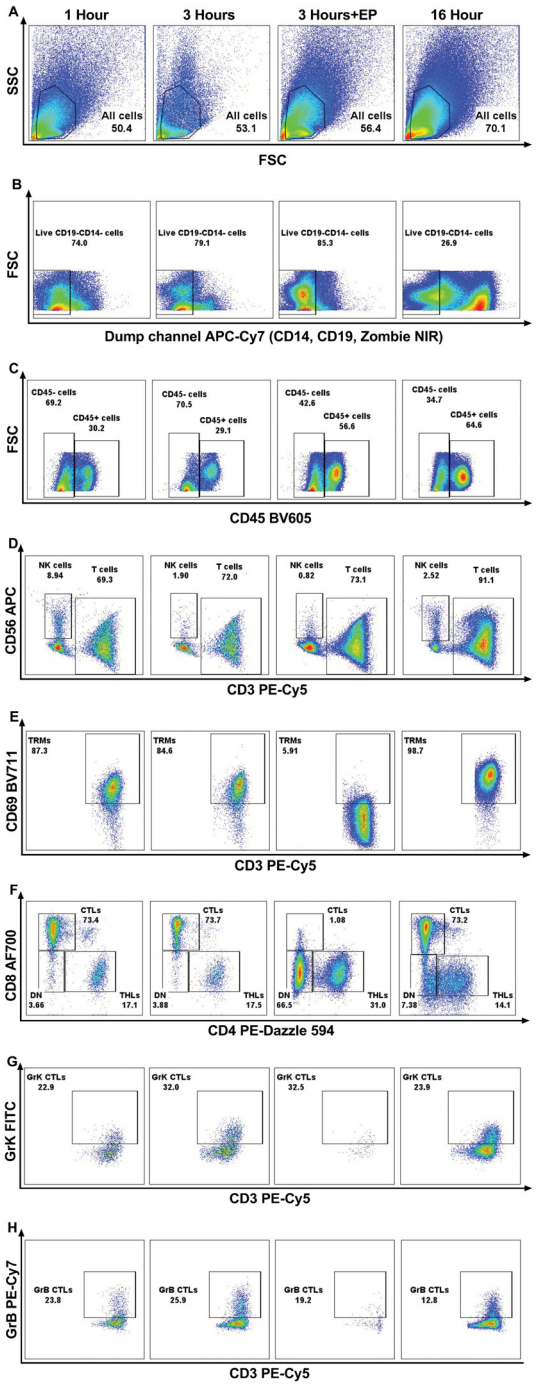


FIGURE 2
Representative flow cytometry graphs demonstrating the effect of enzymatic treatment for 1 hour, 3 hours, 16 hours; and 3 hours of incubation using enzyme P on all cells in healthy skin **(A)**, live CD19-CD14- cells **(B)**, CD45+ and CD45- cells **(C)**, CD3+ T cells and CD56+ NK cells **(D)**, CD69+ resident memory T cells (TRMs) **(E)** CD4+ T helper cells (THLs) and CD8+ cytotoxic T cells (CTLs) **(F)**, Granzyme K+ (GrK) CTLs **(G)**, and Granzyme B+ (GrB) CTLs **(H)**. It is demonstrated that, due to the cleavage of CD69 and CD8 antigens by enzyme P, the population of TRMs **(E)** and CTLs **(F)** decreased dramatically.

conduct single-cell RNA sequencing (scRNAseq) analysis of matched skin and peripheral blood samples from various subjects in batches. To achieve this, we developed a novel sample multiplexing strategy that is based on a recently published label-free demultiplexing algorithm called "souporcell," which utilizes distinct single nucleotide polymorphism (SNP) patterns unique to genetically different individuals (13). Our strategy represents a modified and enhanced version of the multiplexing strategy described in the original paper and allows us to multiplex two or more samples without using any additional tissue barcoding steps.

In the first strategy, two samples can be multiplexed using a relatively simple design, similar to the approach outlined in the

10.3389/fimmu.2024.1393017

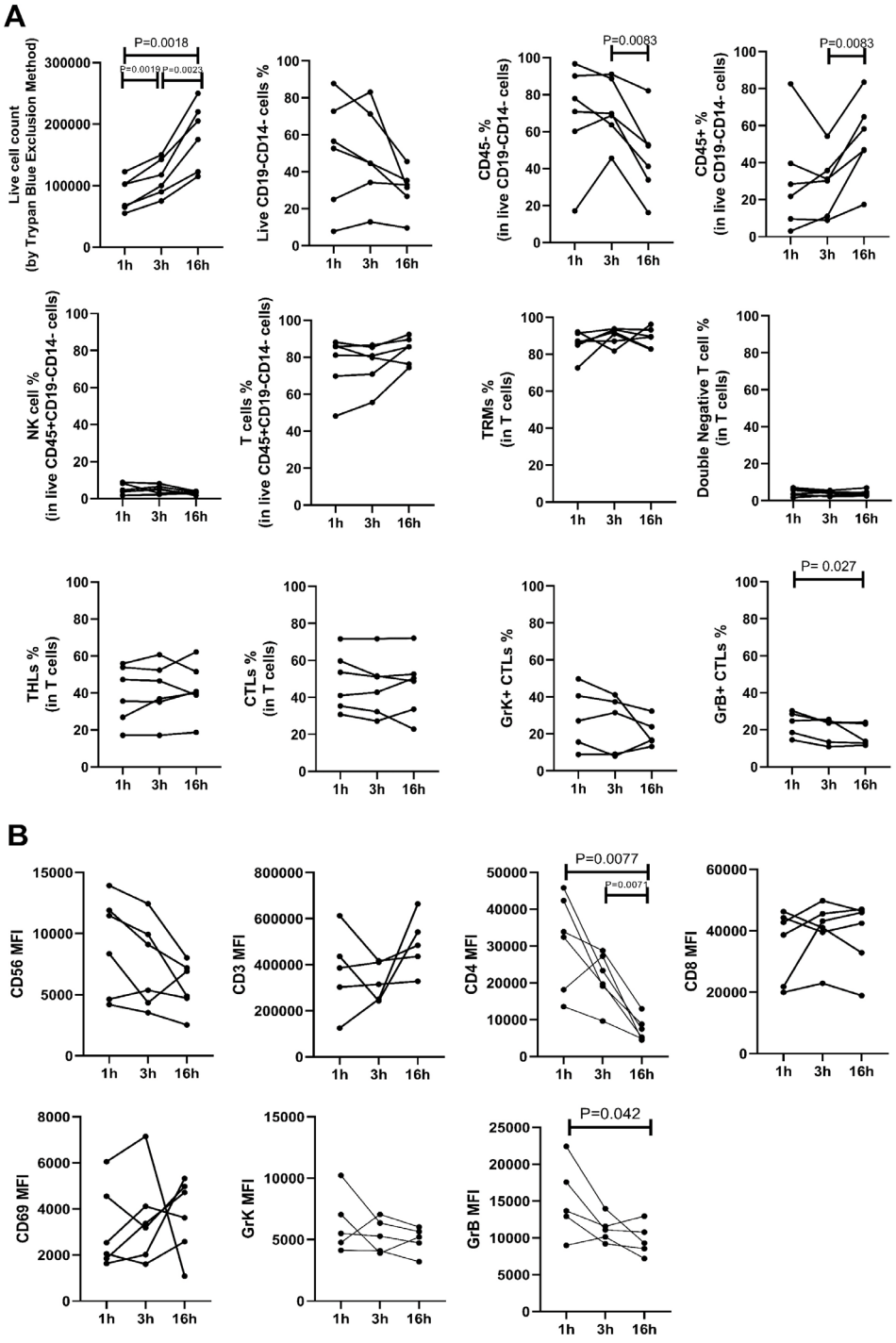


FIGURE 3
The impact of various enzymatic incubation periods (1 hour, 3 hours, 16 hours) on cell percentages **(A)** and mean fluorescence intensity (MFI) measurements **(B)** is shown. A) Live cell counts (determined by trypan blue exclusion method) increased with longer incubation times, indicating enhanced cell yield. The percentage of viable CD14-CD19- leukocytes, assessed by flow cytometry, showed no significant difference across the three incubation periods. However, the percentage of CD45- cells decreased, leading to a significant increase in CD45+ percentage after 16 hours compared to 3 hours of incubation. There was no notable difference in the percentage of lymphocyte subtypes across different incubation times, except for a decrease in CD8+Granzyme B+ cytotoxic T lymphocytes (CTLs) after 16 hours compared to 1 hour incubation. B) MFI values of CD3, CD8, CD56, CD69, and Granzyme K (GrK) showed no significant difference following enzymatic incubation for different durations. However, CD4 and Granzyme B (GrB) MFI decreased significantly after 16 hours of incubation.

10.3389/fimmu.2024.1393017

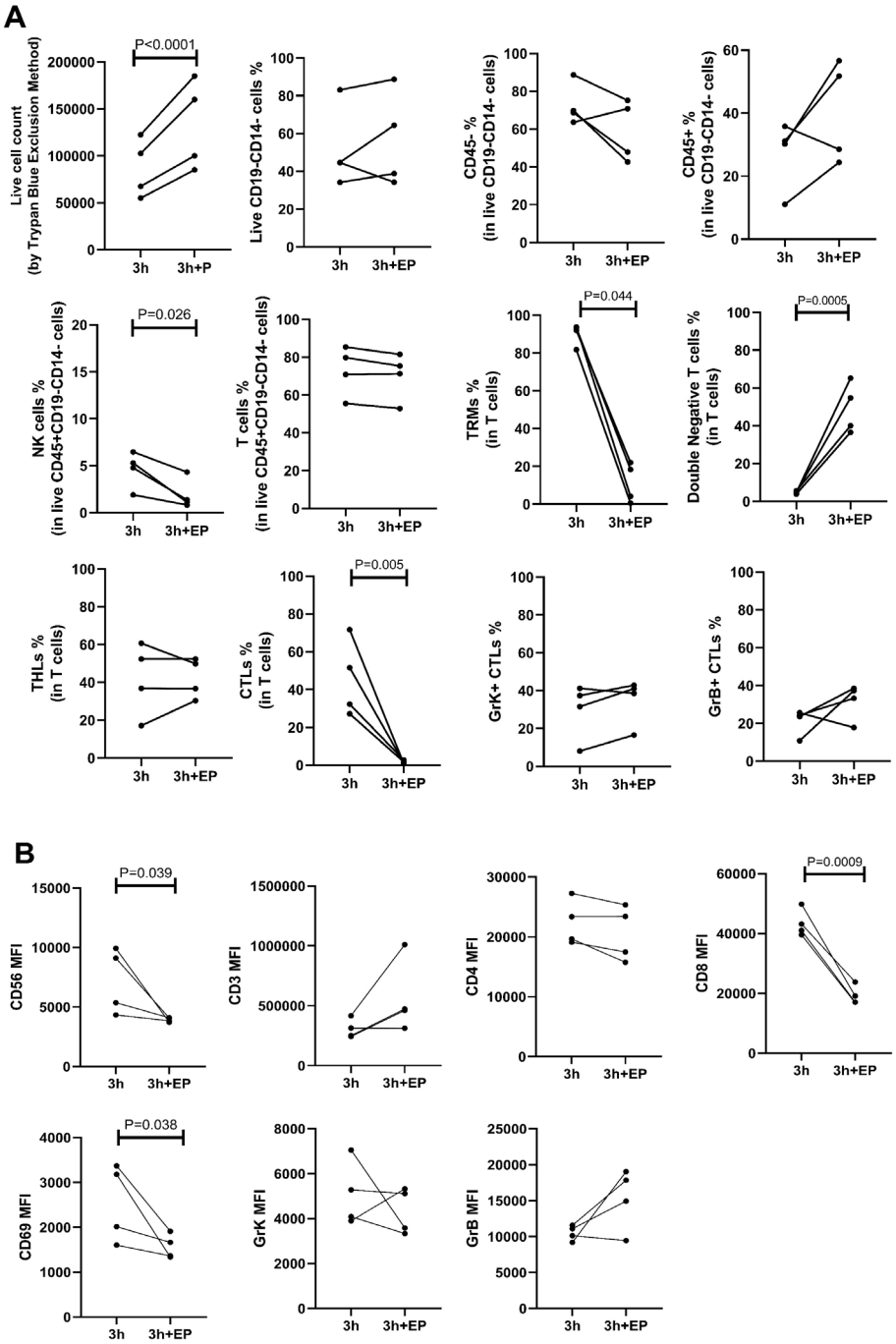


FIGURE 4
The impact of enzyme P on all cell percentages **(A)** and mean fluorescence intensity (MFI) measurements **(B)** is shown. A) Live cell counts demonstrated increased cell yield with the use of enzyme P. However, the percentage of NK cells, CD69+ cells, and CD8+ cytotoxic T lymphocytes (CTLs) significantly declined when enzyme P was used. B) The MFI of CD8, CD56, and CD69 significantly decreased when enzyme P was used in the enzymatic incubation. Statistical analyses were performed using Repeated Measures ANOVA (to compare different time points) and paired ratio T-test (for comparison of the effect of enzyme P).

original article (13). To implement this strategy, the PBMC and dissociated skin samples from the same subject are placed in two separate tubes. Additionally, one genetically unrelated sample is added to each tube (these additional samples should not be paired samples) (Table 2). Subsequently, scRNAseq analysis of these four samples is conducted in two separate reactions. The demultiplexing of samples in each reaction is performed with souporcell, where the paired samples in different tubes belonging to the same individual can be readily identified based on their common SNP pattern. As the identity of paired samples is known beforehand, these clusters can be annotated with their respective sample identities. Once one of the sample identities is revealed, it becomes straightforward to identify the remaining sample in each tube. Using this strategy, it is possible to process nine pairs of PBMC and skin samples in 9

TABLE 2 Multiplexing strategy for scRNAseq analysis 18 paired skin and peripheral blood samples obtained from nine different individuals.

Tube/ Reaction no:	Tissue:	Subject code:								
		S1	S2	S3	S4	S5	S6	S7	S8	S9
I	Skin	X	X							
II	PBMC		X	X						
III	Skin				X	X				
IV	PBMC					X	X			
V	Skin							X	X	
VI	PBMC								X	X
VII	Skin			X			X			
VIII	PBMC	X			X					
IX	Skin + PBMC							X (PBMC)		X (Skin)

The PBMC and dissociated skin sample pairs from the same subject were placed separately in two tubes/reactions. Then, one genetically unrelated sample was added to each tube.

reactions, instead of 18, effectively reducing the number of necessary reactions by half.

Our second, more advanced strategy incorporates two sources of genetic information: the common individual SNP fingerprint of paired samples and the donor sex (Table 3, Figure 5). In this strategy, paired PBMC and dissociated skin samples are again placed separately in dual reactions. Additionally, two genetically unrelated samples are added to each tube, resulting in three samples per tube. The key point in this step is to select samples from subjects of different sexes for each tube. Following scRNAseq analysis and sample demultiplexing with souporcell, matched sample pairs with identical genetic SNP patterns are identified similar to the first strategy. Subsequently, the sex information of the remaining two samples in each tube is determined. Specifically, the percentage of Y chromosome gene expression among the whole transcriptome is used as a proxy for male sex. With this sex information, the identities of the remaining samples can be easily determined. With a carefully designed setup as proposed in Table 3, it is feasible to analyze nine pairs of matched PBMC and skin tissues in 6 reactions, instead of 18.

Bioinformatics analysis and representative results

The individual steps of the bioinformatics analysis pipeline are summarized in Figure 6. First, alignment was done with Cell Ranger. Next, demultiplexing was performed with souporcell and the identification of samples was done with the strategy explained above. Then, anndata object was formed and quality control, normalization, feature selection, principal component analysis, and batch correction were done with Scanpy (19). Quality metrics of the scRNAseq data is presented at Supplementary Figure 4. UMAP graph showing representative results belonging to the analysis of inflamed skin samples obtained from patients with Behçet's disease (n=4) and healthy skin (n=3) is presented in Figure 7A. Cell type distribution (Figure 7B) and percentages (Figure 7C), key genetic markers of each cell subset (Figure 7G), and cumulative number and percentage of each cell subset (Figure 7E) are shown. Also, cell type distribution (Figure 7D), cumulative number, and percentage of each cell subset (Figure 7F) from 4 PBMC samples of BD patients are illustrated.

TABLE 3 An example of the label-free triplet pooling strategy that allows for the analysis of nine pairs of matched PBMC and skin tissues in six reactions, instead of 18.

Reaction no:	Tissue:	Subject code:								
		S1	S2	S3	S4	S5	S6	S7	S8	S9
I	Skin			MP1 PBMC						
II	PBMC			MP1 Skin						
III	Skin								MP2 PBMC	
IV	PBMC								MP2 Skin	
V	Skin									
VI	PBMC									

Blue color shows samples from males and pink shows samples from females. MP: matched pair of samples obtained from the same subject. Detailed descriptions of the first two reactions in this table are illustrated in Figure 5.

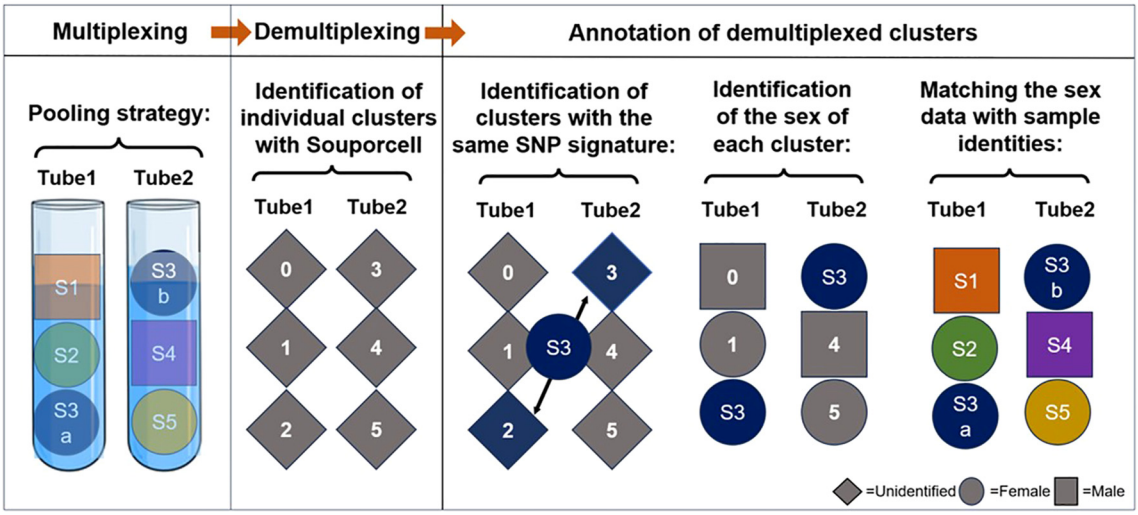


FIGURE 5
Schematic representation of our label-free triplet pooling strategy. First, we segregate paired PBMC (S3a) and dissociated skin samples (S3b) from a selected individual into two distinct tubes. Additionally, two genetically unrelated samples, each from subjects of different sexes (Female is demonstrated as a circle and Male as a square), are introduced to each tube, resulting in three samples per tube. Following scRNA-seq and sample demultiplexing using souporcell, we first identify the clusters exhibiting identical genetic single nucleotide polymorphism (SNP) patterns in dual reactions and then annotate their sample identity. Subsequently, Y chromosome genes are utilized to ascertain the sex information of the remaining two samples in each tube. By using this information the sample identities of the remaining clusters can be determined readily. S, sample. Different colors denote samples from separate subjects.

Supplementary Figure 5 demonstrates single-cell expression of marker genes across different samples. Comparison of lymphocyte subset frequencies detected by scRNAseq and flow cytometry analyses is shown in Supplementary Figure 6. Six samples (3 from Behçet's disease patients and 3 from healthy controls) illustrated in Figure 7 were analyzed. No statistically significant differences were observed between the two methods for T cell, T helper cell, cytotoxic T cell, NK cell, Granzyme K+ T cell, and Granzyme B+

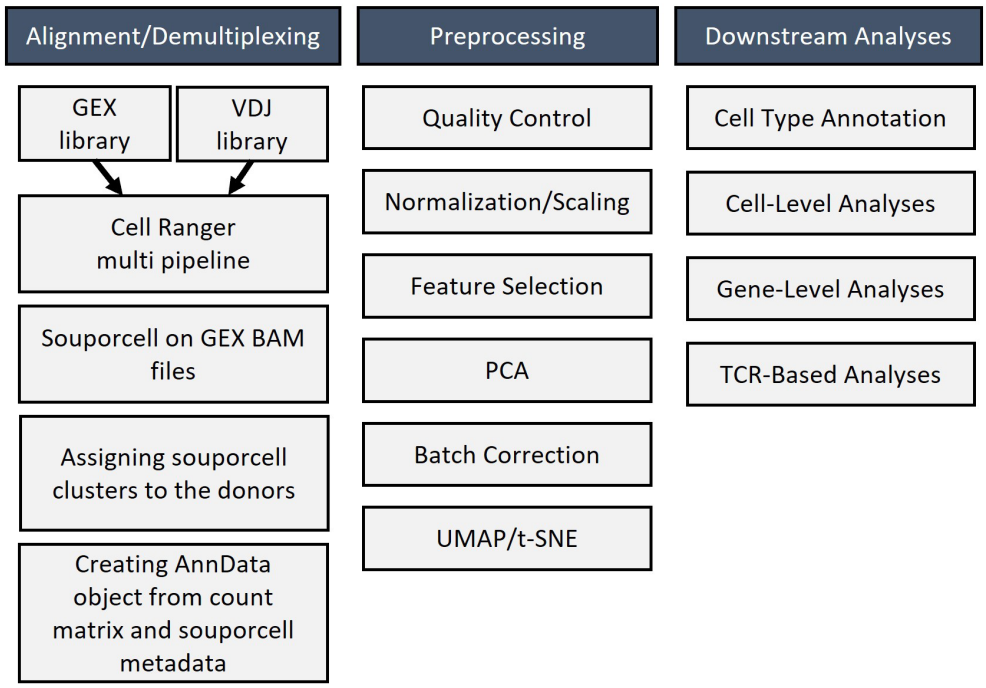
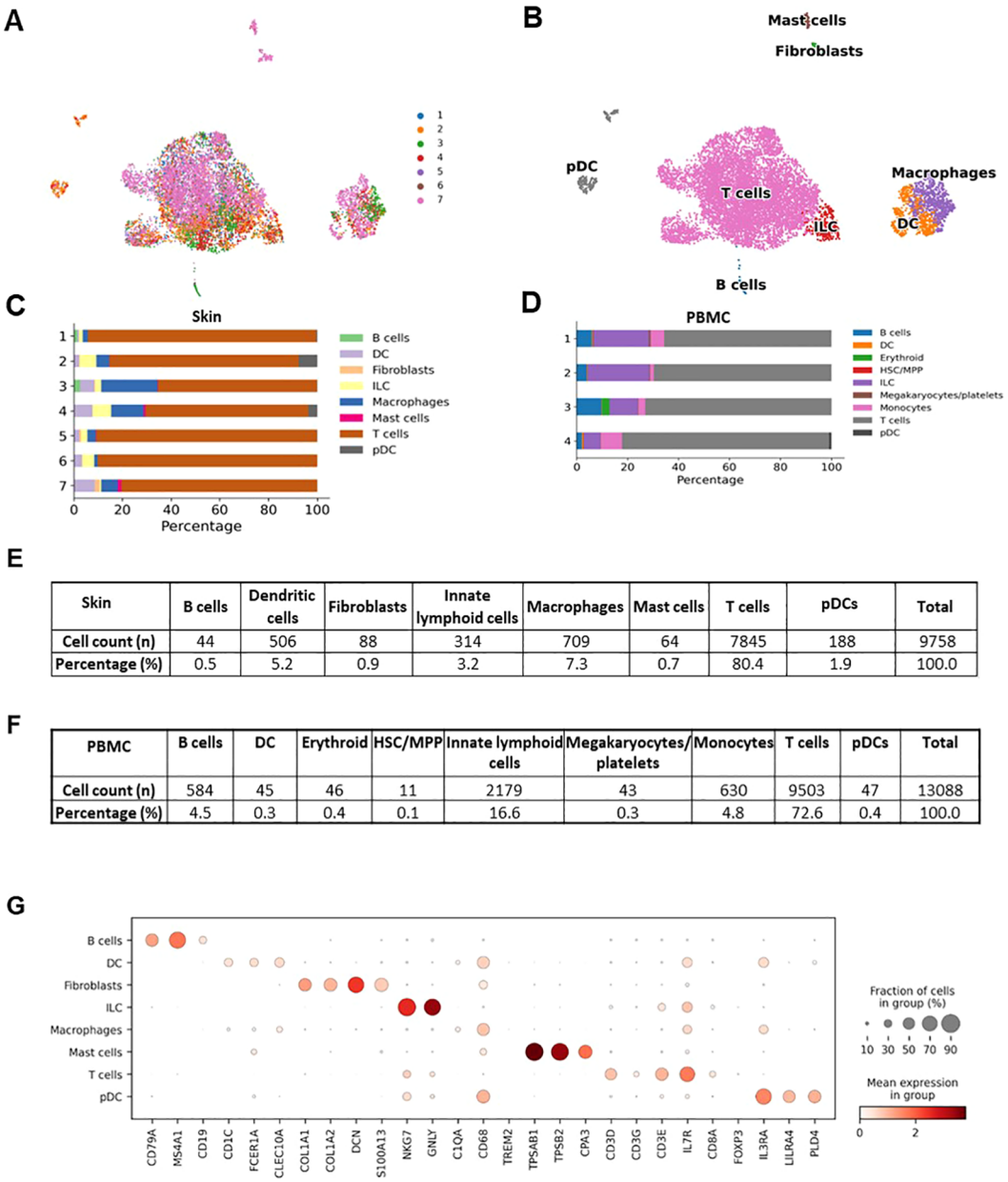


FIGURE 6
Bioinformatic workflow for analyzing single-cell RNA sequencing results that include alignment, demultiplexing, preprocessing, and downstream analysis of sequencing data.



Skin	B cells	Dendritic cells	Fibroblasts	Innate lymphoid cells	Macrophages	Mast cells	T cells	pDCs	Total
Cell count (n)	44	506	88	314	709	64	7845	188	9758
Percentage (%)	0.5	5.2	0.9	3.2	7.3	0.7	80.4	1.9	100.0

PBMC	B cells	DC	Erythroid	HSC/MPP	Innate lymphoid cells	Megakaryocytes/ platelets	Monocytes	T cells	pDCs	Total
Cell count (n)	584	45	46	11	2179	43	630	9503	47	13088
Percentage (%)	4.5	0.3	0.4	0.1	16.6	0.3	4.8	72.6	0.4	100.0

FIGURE 7
Representative results of the single cell RNAseq analysis of skin samples (n=7) obtained from patients with Behçet's disease (n=4) and healthy controls (n=3). **(A)** Demultiplexed skin samples are highlighted on UMAP embedding. **(B)** Identified high-hierarchy cell types are visualized. **(C)** Cell type proportion for skin samples from each donor is illustrated with a stacked bar plot. **(D)** Cell type proportions for PBMC samples from 4 Behçet's disease patients are illustrated with a stacked bar plot. **(E)** The cell counts and proportions of cell types in skin samples are shown. The ratio of fibroblasts is less than 1%, indicating the high purity of the sorted CD45+ cells. **(F)** The cell counts and percentages of cell types in PBMC samples are illustrated. **(G)** Marker genes to characteristic for each cell type are shown as a dot plot.

T cell percentages (Supplementary Figure 6A). Spearman correlation analysis indicated a strong trend towards correlation of T cell percentages measured by these two methods (Supplementary Figure 6B).

Discussion

In this study, we present a novel approach for multiplexing skin and blood samples for single-cell RNA sequencing, significantly reducing costs. Additionally, we provide detailed optimization steps and a step-by-step protocol that enables both scRNAseq and protein-based single-cell analyses from the same tissue sample.

Sample multiplexing is a key strategy to enhance experimental feasibility, mitigate technical batch effects, and significantly reduce the costs associated with scRNAseq. Several demultiplexing methods exist in the literature, such as the use of oligonucleotide-labeled hashtag antibodies to uniquely barcode cells (20). In the scRNAseq protocol for human skin developed by Saluzzo et al., hashtag antibodies were employed for sample de-multiplexing (10).

Compared to our label-free de-multiplexing strategy, this method adds additional experimental steps and extra costs due to antibodies, cDNA library preparation, and sequencing and it is limited to multiplexing of two samples.

Another de-multiplexing method is the demuxlet method that leverages single-nucleotide polymorphisms (SNPs) from a genotype reference obtained through whole-genome or exome sequencing, which was previously regarded as the gold-standard de-multiplexing method (21). Recently, the souporcell algorithm was introduced for de-multiplexing samples based on scRNAseq data without requiring a genotype reference or any other label (13). Souporcell generated benchmarking results that surpasses the demuxlet method, showing its huge potential in scRNAseq studies (13). Based on this, we decided to adopt souporcell as the basis of our sample multiplexing strategy.

Although multiplexed samples can be successfully separated using souporcell, the challenge lies in matching these samples to the correct donor. The original souporcell paper proposed a solution by using the same donor-specific sample in multiple reactions. This approach encodes the inclusion or exclusion of each donor as a bit (0 or 1), and samples in the mixtures are then assigned to corresponding donors using this information-theoretic method (13). In our study, we modified this approach by placing matched skin and PBMC samples from the same donor into different reactions. This provided a canonical way to identify the donors of origin. To further enhance this strategy, we incorporated genetic sex information as a second layer of encoding. This two-step approach allowed for successful demultiplexing and donor identification, reducing the number of required reactions and experimental costs by two-thirds.

Both protein-based methods, such as flow cytometry and CyTOF, and RNA-based single-cell studies conducted on lesional skin are surprisingly scarce. One of the main reasons for this is the challenge associated with skin dissociation compared to other tissues. The skin's dense collagenous structure necessitates more rigorous enzymatic and mechanical processing to obtain a viable cell suspension, which can sometimes compromise cell viability, degrade epitopes, and decrease RNA quality. The recent introduction of automated tissue dissociator systems and skin dissociation kits has provided a standardized and reliable way to perform this step. However, these kits offer a general protocol that needs to be optimized based on the intended experimental methods, such as cell culturing, or ex vivo single-cell RNA or protein analysis from fresh or frozen cells. While recent protocol papers have detailed the use of these kits to prepare skin cell suspensions for scRNAseq analysis of healthy human and pig skin (10, 11), our study expands on these methodologies by optimizing the protocol for inflamed skin. Additionally, we introduce a method suitable for simultaneous flow cytometry analysis from the same tissue, which can also be adapted to other protein-based analyses such as CyTOF and CITE-seq.

The amount of tissue required for analysis is an important consideration. Our findings align with prior research indicating that a 6 mm punch biopsy is optimal for obtaining sufficient skin cells from healthy skin for subsequent scRNAseq analysis (10, 11). Previously, Saluzzo et al. showed that a 6 mm punch biopsy specimen gives a yield of 1-2 $\times$ 10^5 skin cells after incubating with enzyme P for 3 hours. In our study, we found that 3 hours of enzymatic digestion without enzyme P yielded an average of 1.125 $\times$ 10^5 cells, in line with Saluzzo et al. We recommend the usage of two pieces of 6 mm biopsy from healthy skin if both scRNAseq and flow cytometry analysis are planned. For inflamed skin samples, a 4 mm punch biopsy specimen yielded an average of 4.6 $\times$ 10^5 cells, significantly higher than healthy skin, and sufficient for both flow cytometry and scRNAseq analysis. It should be noted that in this study we have only tested Behçet's Disease skin lesions which are acute inflammatory lesions by their nature. There is a possibility that in more chronic lesions such as psoriasis, systemic sclerosis etc., the degree of inflammation and the number of inflammatory cells may be lower. Therefore, we recommend prior determination of the size of required skin specimen for other lesion types in future studies.

Another important consideration for single-cell studies is whether to analyze cells freshly or after cryopreservation. In a recent study, scRNAseq was performed on freshly isolated cells from 4 mm punch biopsies obtained from patients with atopic dermatitis and psoriasis (14). While fresh analysis of skin samples may seem ideal, it is challenging to coordinate, as all experimental steps—beginning with the collection of biopsy specimens—must be conducted in parallel across multiple patients. In our study, we opted to use cryopreserved cells, which offered several advantages, such as greater flexibility in experimental design, simplified sample multiplexing, and reduced technical batch effects, ultimately lowering scRNAseq costs. Importantly, our assessment of CD45+ lymphocyte viability before and after cryopreservation showed no significant differences in live cell count compared to freshly isolated cells. This finding aligns with previous studies that evaluated the effects of cryopreservation using a 90% FBS + 10% DMSO solution on pig skin cells (11), where cryopreservation did not significantly affect cell viability, aggregation, or gene expression profiles, as demonstrated by scRNAseq analysis.

Custom protocols for skin dissociation have also been used previously. Burja et al. developed a method for skin tissue dissociation for scRNAseq analysis using dispase II, collagenase IV, and trypsin on 4 mm punch biopsies (5). This method yielded a total of 24,053 skin cells per sample from fresh healthy skin and 18,535 cells per sample from skin explants obtained from systemic sclerosis patients cultured for 24 hours. A comparison between cultured skin explants and freshly dissociated samples revealed no significant differences in scRNAseq quality metrics. While the authors observed some alterations in the expression of marker genes in fibroblasts, no such changes were detected in immune cells between freshly dissociated and explant-cultured samples. These findings indicate that the explant method can be an alternative approach for studying the composition of skin immune cells.

The successful generation of GEMs is a key step in scRNAseq using the 10X Chromium system and relies on the presence of high-

quality cells in the right concentration. It is important to note that the quality requirements for scRNAseq analysis are higher than those for flow cytometry analysis, and obtaining high-quality cells within the recommended concentration range after fluorescence-activated cell sorting can be challenging. Based on our experience, minimizing the time between cell sorting and GEM generation is essential, a point also noted by Saluzzo et al. (10). In our experimental design, we pooled 2-3 samples for each reaction and conducted two reactions simultaneously using two cell sorters to reduce post-sorting handling time. Careful preparation is crucial to ensure that all necessary materials for GEM generation are ready immediately after cell sorting. Often, the concentration of cells after sorting falls below the recommended range. In such cases, Saluzzo et al. suggest using a volume reduction device (10); however, this equipment may not be available in all laboratories. To address this limitation, we recommend centrifuging the cells at 850 g for 5 minutes after sorting, then reconstituting the cell pellet at a higher concentration than initially desired before recounting the cells, as some may be lost during centrifugation and supernatant removal.

Enzyme P was used in previous tissue dissociation protocols for scRNAseq to increase the cellular yield. However, we do not recommend using enzyme P when protein-based analysis such as flow cytometry or CITE-Seq is planned as this enzyme cleaves cell surface epitopes, including CD8, CD56, and CD69. Of note, CD45 is not significantly affected by enzyme P, therefore cell sorting of live CD45+ leukocytes can still be performed successfully even if this enzyme is used during dissociation.

The duration of enzymatic incubation is an important consideration for single cell studies. We tested the effect of various enzymatic incubation periods on major lymphocyte markers and found that 1-hour, 3-hour, and 16-hour incubations produced similar results. However, we observed that prolonged incubation resulted in a reduction in the percentage of granzyme B-expressing CD8+ cytotoxic T cells, as well as a decrease in CD4 expression on the surface of T cells. Since granzyme B is typically expressed by effector cytotoxic lymphocytes, we concluded that a 3-hour incubation provides an optimal balance between cellular yield and minimizing damage to effector cells. We recommend refraining from longer incubation durations for single cell studies whenever possible.

Recently, Polakova et al. (12) published a protocol for rapid flow cytometry analysis using 4-mm punch biopsies and collagenase IV and DNase I. In this comprehensive study, the authors compared different incubation times of enzymatic digestion, and compared their tissue dissociation method with the whole skin dissociation kit, which included usage of enzyme P. Using their protocol 12,000 CD45+ cells could be obtained from a 4 mm biopsy after 30 minutes of enzymatic incubation. The percentage of major immune cell types (T cells, B cells, NK cells, and ILCs) were similar between both skin dissociation techniques. Moreover, the authors report that 30 minutes of tissue dissociation can be more reliable for the staining of chemokine receptors compared to longer incubation periods.

In conclusion, in this method paper, we describe a novel cost-effective sample multiplexing approach for scRNAseq studies that can be used with both healthy and inflamed skin and can be combined with single-cell protein analysis from the same tissue. The step-by-step protocol and critical optimization steps reported in this paper can be used to design customized single-cell omics experiments by using skin and other solid tissues.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://figshare.com/articles/dataset/Skin_scRNA-seq_Data/24079362.

Ethics statement

The studies involving humans were approved by Koç University Committee on Human Research (protocol number 2022.058. IRB2.007). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SK: Writing – original draft, Formal analysis, Investigation, Methodology, Visualization. ÖA: Writing – review & editing, Formal analysis, Investigation, Methodology, Visualization. ET: Writing – original draft, Formal analysis, Investigation, Methodology, Visualization. OA: Writing – review & editing, Resources. AO: Writing – review & editing, Resources. GP: Writing – review & editing, Resources. EG: Writing – review & editing, Resources. SE: Writing – review & editing, Resources. AB: Writing – review & editing, Resources. AV: Writing – review & editing, Conceptualization, Formal analysis, Methodology, Supervision, Visualization. SV: Writing – review & editing, Funding acquisition, Resources, Supervision, Visualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by European Academy of Dermatology and Venereology (EADV, grant no: PPRC-2021-31), and the Scientific and Technological Research Council of Turkey (TUBITAK, grant no: 221S877).

Acknowledgments

The authors gratefully acknowledge the use of the services and facilities of the Koç University Research Center for Translational Medicine (KUTTAM), funded by the Presidency of Turkey, Presidency of Strategy and Budget. We would like to acknowledge the assistance of ChatGPT 3.5 for English language editing and revision of this article.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1393017/full#supplementary-material

References

1. He H, Suryawanshi H, Morozov P, Gay-Mimbrera J, Del Duca E, Kim HJ, et al. Single-cell transcriptome analysis of human skin identifies novel fibroblast subpopulation and enrichment of immune subsets in atopic dermatitis. *J Allergy Clin Immunol.* (2020) 145:1615–28. doi: 10.1016/j.jaci.2020.01.042

2. Rojahn TB, Vorstandlechner V, Krausgruber T, Bauer WM, Alkon N, Bangert C, et al. Single-cell transcriptomics combined with interstitial fluid proteomics defines cell type–specific immune regulation in atopic dermatitis. *J Allergy Clin Immunol.* (2020) 146:1056–69. doi: 10.1016/j.jaci.2020.03.041

3. Liu J, Chang HW, Huang ZM, Nakamura M, Sekhon S, Ahn R, et al. Single-cell RNA sequencing of psoriatic skin identifies pathogenic TC17 cell subsets and reveals distinctions between CD8+ T cells in autoimmunity and cancer. *J Allergy Clin Immunol.* (2021) 147:2370–80. doi: 10.1016/j.jaci.2020.11.028

4. Kim J, Lee J, Kim HJ, Kameyama N, Nazarian R, Der E, et al. Single-cell transcriptomics applied to emigrating cells from psoriasis elucidate pathogenic versus regulatory immune cell subsets. *J Allergy Clin Immunol.* (2021) 148:1281–92. doi: 10.1016/j.jaci.2021.04.021

5. Burja B, Paul D, Tastanova A, Edalat SG, Gerber R, Houtman M, et al. An optimized tissue dissociation protocol for single-cell RNA sequencing analysis of fresh and cultured human skin biopsies. *Front Cell Dev Biol.* (2022) 10:872688. doi: 10.3389/fcell.2022.872688

6. Veelken H, Jesuiter H, Mackensen A, Kulmburg P, Schultze J, Rosenthal F, et al. Primary fibroblasts from human adults as target cells for ex vivo transfection and gene therapy. *Hum Gene Ther.* (1994) 5:1203–10. doi: 10.1089/hum.1994.5.10-1203

7. Clark RA, Chong B, Mirchandani N, Brinster NK, Yamanaka K, Dowgiert RK, et al. The vast majority of CLA+ T cells are resident in normal skin. *J Immunol.* (2006) 176:4431–9. doi: 10.4049/jimmunol.176.7.4431

8. Du W, Lenz D, Kohler R, Zhang E, Cendon C, Li J, et al. Rapid Isolation of Functional ex vivo Human Skin Tissue-Resident Memory T Lymphocytes. *Front Immunol.* (2021) 12:624013. doi: 10.3389/fimmu.2021.624013

9. Mashiko S, Bouguermouh S, Rubio M, Baba N, Bissonnette R, Sarfati M. Human mast cells are major IL-22 producers in patients with psoriasis and atopic dermatitis. *J Allergy Clin Immunol.* (2015) 136:351–9.e1. doi: 10.1016/j.jaci.2015.01.033

10. Saluzzo S, Gail LM, Neuwirth T, Pandey RV, Farlik M, Stary G. Processing human skin samples for single-cell assays. *STAR Protoc.* (2022) 3:101470. doi: 10.1016/j.xpro.2022.101470

11. Han L, Jara CP, Wang O, Shi Y, Wu X, Thibivilliers S, et al. Isolating and cryopreserving pig skin cells for single-cell RNA sequencing study. *PloS One.* (2022) 17:e0263869. doi: 10.1371/journal.pone.0263869

12. Polakova A, Hudemann C, Wiemers F, Kadys A, Gremke N, Lang M, et al. Isolation of lymphocytes from human skin and murine tissues: A rapid and epitope-preserving approach. *JID Innov.* (2023) 3:100155. doi: 10.1016/j.xjidi.2022.100155

13. Heaton H, Talman AM, Knights A, Imaz M, Gaffney DJ, Durbin R, et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat Methods.* (2020) 17:615–20. doi: 10.1038/s41592-020-0820-1

14. Zhang B, Roesner LM, Traidl S, Koeken V, Xu CJ, Werfel T, et al. Single-cell profiles reveal distinctive immune response in atopic dermatitis in contrast to psoriasis. *Allergy.* (2023) 78:439–53. doi: 10.1111/all.15486

15. Polanski K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park JE. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics.* (2020) 36:964–5. doi: 10.1093/bioinformatics/btz625

16. Park JE, Botting RA, Dominguez Conde C, Popescu DM, Lavaert M, Kunz DJ, et al. A facscell atlas of human thymic development defines T cell repertoire formation. *Science.* (2020) 367:eaay3224. doi: 10.1126/science.aay3224

17. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods.* (2019) 16:1289–96. doi: 10.1038/s41592-019-0619-0

18. Dominguez Conde C, Xu C, Jarvis LB, Rainbow DB, Wells SB, Gomes T, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science.* (2022) 376:eabl5197. doi: 10.1126/science.abl5197

19. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* (2018) 19:15. doi: 10.1186/s13059-017-1382-0

20. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM 3rd, et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* (2018) 19:224. doi: 10.1186/s13059-018-1603-1

21. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol.* (2018) 36:89–94. doi: 10.1038/nbt.4042

frontiers | Frontiers in Immunology

TYPE Review
PUBLISHED 09 January 2025
DOI 10.3389/fimmu.2024.1520860

OPEN ACCESS

EDITED BY
Jeon Woong Kang,
Massachusetts Institute of Technology, United States

REVIEWED BY
Aitzol Garcia-Etxarri,
Donostia International Physics Center (DIPC), Spain
Sonia Cohen,
Massachusetts General Hospital and Harvard Medical School, United States

*CORRESPONDENCE
Jennifer Dionne
jdionne@stanford.edu
Amanda Kirane
akirane@stanford.edu

†These authors have contributed equally to this work

RECEIVED 31 October 2024
ACCEPTED 25 November 2024
PUBLISHED 09 January 2025

CITATION
Chadokiya J, Chang K, Sharma S, Hu J, Lill JR, Dionne J and Kirane A (2025) Advancing precision cancer immunotherapy drug development, administration, and response prediction with AI-enabled Raman spectroscopy.
Front. Immunol. 15:1520860.
doi: 10.3389/fimmu.2024.1520860

Advancing precision cancer immunotherapy drug development, administration, and response prediction with AI-enabled Raman spectroscopy

Jay Chadokiya[1], Kai Chang[2], Saurabh Sharma[1], Jack Hu[3], Jennie R. Lill[4], Jennifer Dionne[3,5,6*†] and Amanda Kirane[1*†]

[1]Department of Surgery, Stanford School of Medicine, Stanford University Medical Center, Stanford, CA, United States, [2]Department of Electrical Engineering, Stanford University, Stanford, CA, United States, [3]Pumpkinseed Technologies, Palo Alto, CA, United States, [4]Genentech, South San Francisco, CA, United States, [5]Department of Materials Science and Engineering, Stanford University, Stanford, CA, United States, [6]Department of Radiology, Molecular Imaging Program at Stanford (MIPS), Stanford University School of Medicine, Stanford, CA, United States

Molecular characterization of tumors is essential to identify predictive biomarkers that inform treatment decisions and improve precision immunotherapy development and administration. However, challenges such as the heterogeneity of tumors and patient responses, limited efficacy of current biomarkers, and the predominant reliance on single-omics data, have hindered advances in accurately predicting treatment outcomes. Standard therapy generally applies a "one size fits all" approach, which not only provides ineffective or limited responses, but also an increased risk of off-target toxicities and acceleration of resistance mechanisms or adverse effects. As the development of emerging multi- and spatial-omics platforms continues to evolve, an effective tumor assessment platform providing utility in a clinical setting should i) enable high-throughput and robust screening in a variety of biological matrices, ii) provide in-depth information resolved with single to subcellular precision, and iii) improve accessibility in economical point-of-care settings. In this perspective, we explore the application of label-free Raman spectroscopy as a tumor profiling tool for precision immunotherapy. We examine how Raman spectroscopy's non-invasive, label-free approach can deepen our understanding of intricate inter- and intra-cellular interactions within the tumor-immune microenvironment. Furthermore, we discuss the analytical advances in Raman spectroscopy, highlighting its evolution to be utilized as a single "Raman-omics" approach. Lastly, we highlight the translational potential of Raman for its integration in clinical practice for safe and precise patient-centric immunotherapy.

KEYWORDS

Raman spectroscopy, label-free analysis, immunotherapy, time analysis, multiomics

1 Introduction

The immune system plays a vital role in detecting cancer by recognizing neoantigens generated by tumor cells that can initiate immune responses (1). However, tumors have evolved several strategies to evade immune detection (2). These include downregulating antigen presentation, which impairs the ability of immune cells to recognize and attack tumor cells, and expression of surface protein ligands, such as Programmed Death-Ligand 1 (PD-L1), that interact with immune checkpoint proteins, such as Programmed Death Protein 1 (PD-1), on immune cells (3). Tumor-secreted factors modulate the tumor immune microenvironment through several mechanisms, including: i) releasing immunosuppressive cytokines such as IL-2, TGF-β, IL-10, IL-35 and VEGF, which inhibit various immune cell activities (4); ii) releasing tumor-derived exosomes which contain immunosuppressive molecules, including TRAIL, Fas-L, PGE-2, etc (5); and iii) recruiting regulatory immune cells such as regulatory-T cells, tumor associated macrophages, and myeloid-derived suppressor cells to the tumor site (6). Epigenetic modulation within cancer cells can also silence genes related to antigen presentation (7). To effectively deploy immunotherapy, it is essential to accurately detect and classify the evasion tactics of cancer cells. Our manuscript discusses how Raman spectroscopy, as a label-free, reliable, and cost-effective technology, can sense these tactics across the immunological synapse.

Various immunotherapy strategies currently utilized include immune checkpoint inhibitors (ICIs), cancer vaccines, adoptive cellular therapies (ACT), cytokines, targeted antibodies including T cell-engaging bispecifics, and adjuvants & immunostimulants. Although these approaches have led to improved outcomes for some patients, their benefits are often limited to a small and unpredictable segment of cancer patients. This has led to increased cases of immune-related adverse events (irAEs) (8, 9). For example, in melanoma, where ICIs are the mainstay treatment, the overall response rate is only 30-45% for the most common single-agent anti-PD-1 approach (10). Further, many cancers, such as pancreatic adenocarcinoma, have unique biologic environments such as high levels of fibrosis, contributing to immune cell resistance and evasion that render these immunotherapeutic agents significantly less effective (11, 12). Therefore, accurately assessing a patient's tumor microenvironment (TiME) and predicting their response to immunotherapy are essential for maximizing treatment effectiveness. An important step towards this is precise biomarker prediction which helps in establishing more accurate, individualized profiles to guide immunotherapeutic selection (13, 14). As many existing biomarker predictive models rely on single-omics data, which may not capture the complex biological interactions involved in tumor immunology, their predictive power has been limited (15, 16). Multi-omics approaches that combine genomic (17), transcriptomic (18), proteomic (19) lipidomic, and metabolomic data can improve the accuracy of response predictions (20–23). In a recent study, Kong et al. utilized a machine learning framework that integrated various -omics data to predict responses to ICIs in melanoma, gastric cancer, and bladder cancer, demonstrating superior predictive capabilities compared to traditional biomarkers. Investigators curated data from more than 700 ICI-Treated patients' samples with clinical outcomes and transcriptomic data. Their network-based ML algorithm showed significantly better performance in predicting ICI treatment responses in all the above-mentioned types of cancers compared to existing models, demonstrating network biology as a powerful means to identify robust biomarkers (16).

Multi-omics technologies have increased our understanding of the complex inter- and intra- molecular cross-talk between immune cells and tumor cells within TiME. However, working with large analytical and statistical datasets generated by single or spatial technologies presents significant computational hurdles (24). One major issue is the batch effects caused by using different analytical techniques employed in -omics data collection (25). These techniques are costly, time consuming, and require extensive labeling steps which may require disruption of native biological environments for the cells of focus (26, 27). Raman spectroscopy can effectively harmonize all the -omics techniques for analyzing TiME interplay and its intricate changes under a single platform. Additionally, a combination of Raman and traditional multi-omics can also leverage the strengths of both methodologies, including the high sensitivity, multiplexing capabilities, rapid analysis, and non-destructive, label-free nature of Raman, alongside the specificity and extensive data provided by traditional -omics approaches. In the past decade, label-free Raman spectroscopy has found significant applications in cancer diagnostics, particularly in cell type differentiation (28–30) and metabolite characterization (31–33). It also allows for the identification of biochemical changes within tumors, enhancing our ability to monitor responses to therapies more efficiently (34). These studies provide the foundation for deploying Raman spectroscopy as a platform for immunotherapy development, administration, and response monitoring.

In this perspective, we discuss the principles and role of Raman spectroscopy in immunotherapy. In section 2, we describe advances in nanophotonics which render Raman suitable for non-invasive, label-free detection of the TiME at the single-cell to few-molecule level. We also discuss the role of machine learning and artificial intelligence (ML/AI) in Raman spectral analysis and data interpretation. Section 3 describes the role of Raman spectroscopy in identifying, characterizing, and analyzing the complex inter- and intra- metabolic and phenotypic changes occurring within TiME, as well as Raman spectrosocpy's role in predicting responses to various immunotherapeutic treatments. Section 4 outlines the current analytical advancements in Raman spectroscopy within the field of immunology. Finally, Section 5 explores how Raman spectroscopy can serve as a unifying, multi-omic technique that stitches genomic, transcriptomic, proteomic, and metabolomic data, as well as a potentially low-cost tool with translational potential in clinical settings.

2 Nanophotonic-enhanced Raman spectroscopy and AI-enabled interpretation

Raman spectroscopy (RS) is a non-invasive, vibrational spectroscopic method that examines the composition, structure, and vibrational energy states of materials (including molecules and

cells). In RS, a sample is illuminated with monochromatic light. When the incident light interacts with molecular vibrations in the sample, photons can be inelastically scattered and re-emitted with either lower or higher energy (Figure 1). This energy difference, known as a Raman shift, provides a distinct molecular "fingerprint" of the material (35). By analyzing the unique spectral fingerprints of molecules fundamental in cellular biology, RS can provide detailed insight into the molecular composition and the structural and functional makeup of cells and tissues, both *in vivo* and *ex vivo* (36, 37). For example, there are biologically-relevant windows (38, 39) that elucidate biomarkers spanning lipids (40, 41), proteins and peptides (42, 43), metabolites (44–46) and nucleic acids (47, 48) (Figure 1). In turn, these markers can demarcate normal and malignant cells (49, 50) and stratify cancer types (51) or pathologic grades (52, 53), facilitating potential early diagnosis and intervention pathways. As a non-destructive optical technique, Raman spectroscopy can be seamlessly integrated with other modalities on the same sample, allowing for multi-omic resolution in a single measurement.

Although Raman spectroscopy is non-invasive and highly specific in providing molecular and structural information, a major challenge of spontaneous RS lies in its intrinsically weak scattering process. Because of the low likelihood of a Raman scattering event [roughly 1 in 10E6-7 incident photons (54–56)], complementary strategies have been adopted to address its signal intensity and enhance sensitivity. The emergence in the fields of nanophotonic materials and machine learning models, in particular, have improved Raman sensitivity and resolution and to enable deeper spectral interpretation.

One strategy to amplify the signal-to-noise ratio of Raman is through surface-enhanced Raman scattering (SERS), which uses optically resonant surfaces or nanoparticles (NPs) to increase the Raman cross-section (Figure 1). Vast literature has been published using metallic nanostructures for SERS. When light interacts with these metallic nanostructures, the electrons in the metal oscillate in resonating manner, creating an intensified electromagnetic field known as a plasmon resonance on the surface. This additional field strength localization intensifies the light interaction that occurs between molecules, with enhancement coefficients ranging from 10^4-10^8, and as high as 10^{11} (45–47). The resulting process generates highly-detailed, vibrational spectra, making it particularly useful in fields like cancer immunotherapy (48, 49), biochemistry (50, 51), medical diagnosis (52), and surgical treatments (53). SERS studies employing colloidal NPs have shown extensive success in cancer biological interrogation, from Liu et al. exploiting Au/Ag nanostar geometries to quantify BRAF gene mutations in colorectal cancer with comparable LOD to qPCR, to Sun et al. leveraging Au nanorods as a multifunctional agent to identify and induce photothermal ablation of tumor margins (45, 46). Recent advances in large-area nanoarray fabrication leveraging self-assembled NP aggregation or nanolithography have led to the rise and potential of SERS-active substrates. Zhao et al. designed one such substrate by fabricating

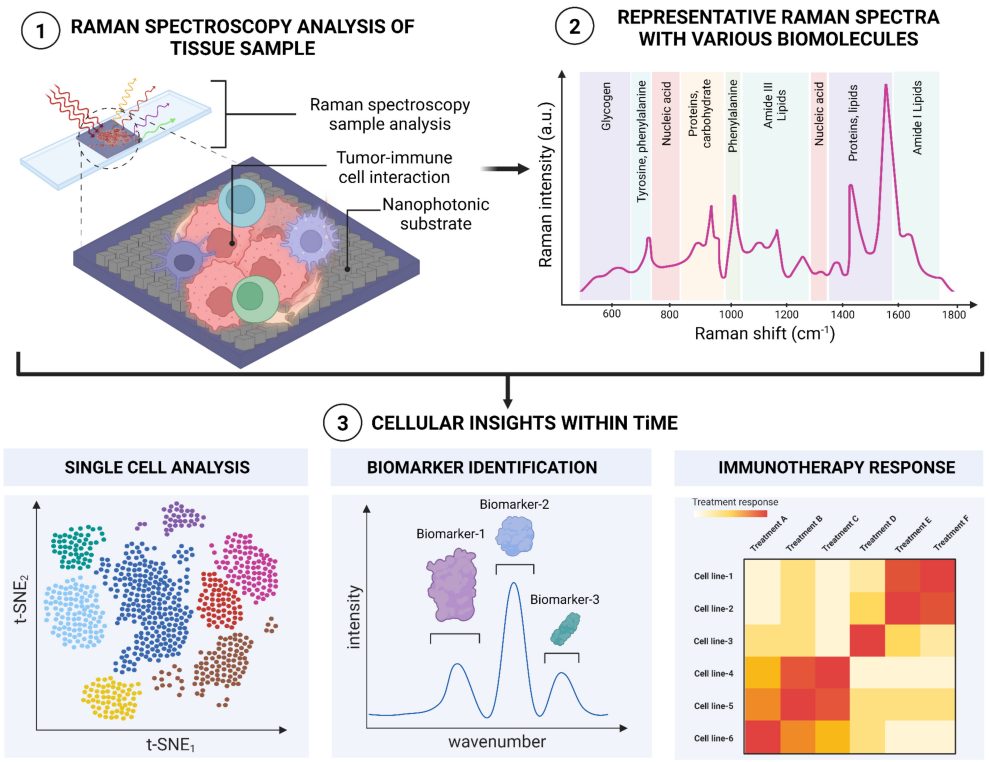


FIGURE 1
A schematic illustration of Raman spectroscopic workflow for analyzing and observing inter and intra cellular interactions within TiME.

nanoarrays of plasmonic trimers to successfully label adenocarcinoma, squamous carcinoma and benign tumor samples across fresh lung tissues (57). These SERS-active devices can yield comparable enhancements to colloidal NPs, all while improving sample adhesion and hotspot uniformity and distribution.

Although there is less literature, recent innovations in dielectric-based substrates for SERS present an advantageous opportunity for material and biological characterization. Unlike metallic nanostructures, which exhibit high photothermal effects damaging cells or altering biomolecular structures, dielectric nanostructures undergo minimal heat conversion, making them highly suitable for biological preservation and measurement reliability (58, 59). Advancements in highly resonant, high quality-factor (Q) metasurfaces have also overcome conventionally limited electromagnetic field enhancements (60–62), yielding Raman scattering efficiencies comparable to plasmonic counterparts (58, 63, 64). In work by Cambiasso et al. and Romano et al., for example, dielectric nanodimers and photonic crystals were utilized to demonstrate Raman spectral amplification across β-carotenal monolayers and Raman analytes with minimal absorption loss (65, 66). Silicon-based designs, in particular, can further leverage the device footprint scaling of matured CMOS infrastructure (67). Barkey et al. demonstrated one such design by pixelating 2D arrays of Si-ellipse pairs to resolve real-time conformational dynamics of photoswitchable lipid membranes representative of cell membrane behavior (68). These large-area fabricated arrays can enable homogenous SERS regions for rapid spatial profiling all while providing compatibility to assess the same sample with other modalities.

Enhancing the utility of Raman spectroscopy can be achieved by incorporating machine learning (ML) and artificial intelligence (AI), which can extract underlying spectral features linked to biological and chemical responses. Spectral information from RS is often feature-rich, but the unprocessed information can be complex and noisy. As a result, employment of both more traditional statistical approaches and newer deep learning algorithms can be utilized to isolate pertinent information from background and extract insights in an otherwise opaque spectra. Dimension reduction techniques adopted prior to analysis can improve feature selection, reduce overfitting, and improve computational runtime, all while preserving original spectra information. Linear techniques such as principal component analysis (PCA) can decompose large feature sets into smaller ones encapsulating the most significant spectral patterns and differentiators, while nonlinear reduction methods like t-distributed stochastic neighbor embeddings (t-SNE) or uniform manifold approximation (UMAP) can help contextualize the local and global structural relationship of Raman spectra datasets. Classification algorithms can further intake the Raman spectra and provide distinct cell type labeling to predict post-treatment outcomes in untested samples. Support vector machines (SVMs) and Random decision forests (RFs) can be used to robustly classify cancer subtypes as recently demonstrated in brain tissue (69, 70) and in breast cancer garnering an accuracy of +97% (69, 70). Advances in multilayer architectures such as convolutional neural networks (CNN) and residual neural networks (ResNet) have further increased the predictive capacity of RS, even against high inter-patient variability and complex background sources. For example, in melanoma, where the clinical diagnostic sensitivity and specificity ranges from 40-80%, the implementation of artificial neural networks on Raman spectra resulted in an improved sensitivity and specificity of 85% and 99%, respectively (53). Such integration of deep learning and the continued advancements in AI can stand to provide a powerful opportunity to analyze Raman spectra beyond single cells and across the tissue domain. Further, as discussed later in the perspective, integration of RS with existing multi-omics and spatial-omics data, using existing AI models, could offer a more comprehensive understanding of tumor heterogeneity.

3 Role of Raman spectroscopy in characterizing tumor-immune microenvironment

The TiME is a complex and diverse ecosystem containing a variety of immunosuppressive cells, including tumor cells, cancer-associated fibroblasts (CAFs), vascular endothelial cells, suppressive myeloid cells, regulatory T (Treg) cells, and regulatory B cells. Increasing evidence strongly suggests that TiME plays a significant role in immune checkpoint inhibitors' responses, tumor immune surveillance, and immunological evasion (71, 72). Paidi et al. showed evidence that label-free Raman spectroscopy can show TiME compositional changes in response to ICIs. Using CT26 murine colorectal tumor xenografts, they compared tumor responses with treatment across three doses of anti–CTLA4 and anti–PD-L1 antibodies each. They determined that ICI exposure significantly changes the composition of the TiME independent of conventional cellular, molecular, or proteomic characterizations (34). This ability to assess multiple biomolecular changes simultaneously adds significant depth in understanding the TiME and response to therapies. Figure 2 highlights the multitude of signals that Raman spectroscopy can provide about the TiME. As seen, Raman spectroscopy can be used in differentiating various cancer and immune cell types, including B cells, cytotoxic T cells, helper T cells, NK cells, and dendritic cells. For instance, Chen et al. employed Raman spectroscopy to accurately identify various subsets of immune cells, including T-lymphocytes, dendritic cells, and natural killer (NK) cells, distinguishing CD56+ NK cells from CD4+ and CD8+ T cells with specificities reaching 93% and 96%, respectively. The differentiation between CD4+ and CD8+ T cells was less effective, yielding a specificity of 68% and a sensitivity of 69%, suggesting that these closely related cell types present more challenges in their identification (73). Conventional techniques for immune cell identification and complex classification of the TiME currently relies on extensive labeling for label-based techniques, due to the need to both "rule-in" and "rule-out" broad cell surface markers and utilize multiple labels related to functional behavior and activation status. The exploration of RS to distinguish cell types has been provocative, here we highlight several critical cell types that have been shown to be highly distinguishable by RS (73). While the Raman spectra of these immune cells may appear quite similar,

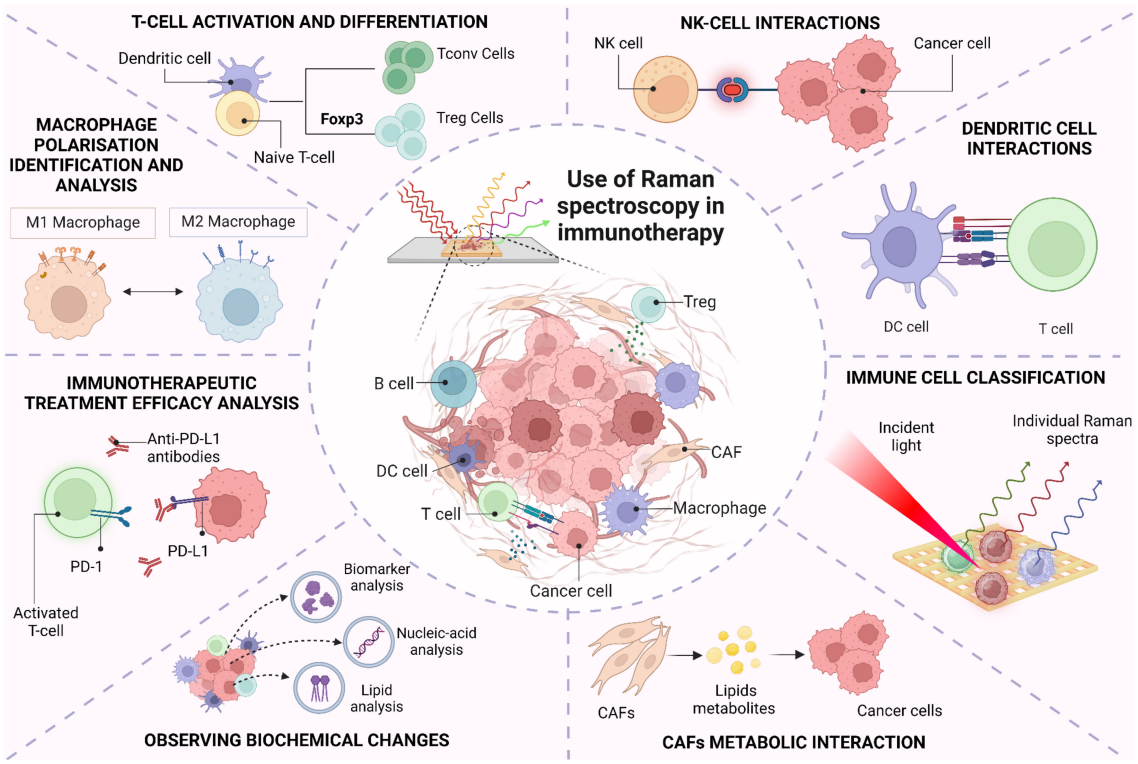


FIGURE 2
Use of Raman spectroscopy in immune cell classification and its interaction within tumor-immune microenvironment.

data analysis techniques can reveal the subtle distinctions among them (73–75). Raman spectra can also provide information about the activation states of these cells, including macrophage polarization and T-cell state responses (eg, from activated to exhausted.) Single-cell Raman analysis can further reveal how different cell types interact within the TiME. Finally, Raman can help elucidate tumor heterogeneity and how the spatial structure of the tumor impacts immune responses, currently a major obstacle for effective immunotherapy (76). In this section, we will explore the utility of RS in characterizing, classifying and analyzing different inter- and intra-molecular interactions between immune cells within the TiME.

3.1 Macrophages

Macrophages, essential phagocytic and antigen-presenting cells, exhibit a diverse functional spectrum from immunosuppressive, tumor-promoting behaviors to highly inflammatory responses. Their role in the tumor microenvironment is pivotal, as they can either support tumor control or contribute to autoimmune toxicities. Conventionally, differential expression levels of surface polarization markers, such as CD11b, CD80, CD54, CD163 and CD206, are used to differentiate macrophage phenotypes, however the transition from inflammatory to immunosuppressive behavior is highly linked to metabolic switching that can be detected by Raman spectroscopy. In a study by Naumann et al., distinct features of monocyte-derived macrophages, including naïve M0, classically activated M1, and alternatively activated M2 phenotypes were detected by analyzing 65 chemically fixed primary human monocyte-derived macrophages from three donors in combination with N-FINDR spectral unmixing. The authors identified polarization-dependent spectral features associated with the chemical composition of lipids, proteins, and nucleic acids across macrophage phenotypes. Pro-inflammatory M1 macrophages displayed a significantly higher lipid content compared to M0 and M2 phenotypes. M2 macrophages exhibited reduced triacylglycerol content but increased fatty acids. These spectral distinctions facilitated the development of models for automated classification of M1 macrophages, achieving a classification accuracy of 86%, with a sensitivity of 93% and specificity of 85% (77). In another study by Lu et al., macrophage response to biomaterial implants was examined to gain insights into the immune system's foreign body reaction. Two types of macro-encapsulation pouches (PVDF and TPU-chronoflex) were implanted in streptozotocin-induced diabetic rat models for 15 days. Their research demonstrated that label-free Raman microspectroscopy could effectively identify extracellular matrix (ECM) components within the fibrotic capsule and distinguish between pro-inflammatory M1 and anti-inflammatory M2 macrophage activation states. Significant spectral changes in the nuclei of M1 and M2 macrophages indicated variations in nucleic acid methylation, a key process in fibrosis progression. Specifically, increased peak intensities at 857 cm^{-1} and 879 cm^{-1} in M2 macrophages were linked to proline, hydroxyproline, tryptophan, and tyrosine, suggesting that M2 macrophages have lower

methylation levels than M1 macrophages (78). Thus, RS plays an important role in analyzing biochemical changes in lipids, proteins, and nucleic acids across macrophage phenotypes and identifies extracellular matrix (ECM) components.

3.2 T-lymphocytes

T cells are important effector cells in the TiME, including cytotoxic and regulatory subtypes that attack cancers or suppress immune responses to cancers, respectively. T cell classifications, like macrophages, generally require multiple labels, such as CD3, CD4, and CD8, to define subtype in addition to a multitude of co-stimulatory signals, such as activating ligands or regulating checkpoints to modulate the degree of amplification for T cell responses. Authors Pavillon et al. leveraged the non-tissue destructive nature of RS to monitor live T cell development *in vitro*, demonstrating that without directly describing the cell surface features of these traditional labels, other nuanced molecular changes related to cell state development and activation had high correlation with the transition points identified by label-based assays (29). The sensitivity in this assay also successfully delineated between activation and differentiation by detecting differences in the *in vitro* stimulated cells versus ex vivo activated T cells that otherwise would have required multiple additional labeling steps to define naive versus effector cells. Regulatory T cells (Tregs) are crucial for maintaining immunological self-tolerance and have been identified as having an important role in immunotherapeutic failures. The findings by Pavillon et al. indicated that Raman could distinguish Treg subpopulations without altering cell integrity (29) by the detection of intracellular transcription factor Foxp3, a specific Treg marker. Since Foxp3 is not detectable in live cells, the authors employed RS to reliably identify and isolate functional Treg populations. They sorted conventional T (Tconv) and Treg cells using FACS with Foxp3-hCD2 surface staining, followed by Raman measurements on the isolated populations. A ML model was then developed to differentiate between Tconv and Treg cells, achieving an accuracy of 78.3% on test data, comparable between models trained on naive cells and those based on FACS-sorted data (78.25% for FACS vs. 77.9% for naive cells). When they applied confident learning (CL) to filter out samples with low-probability labels, the model achieved a remarkable 92% accuracy. Figure 3A illustrates the classification of human Tconv/Treg using the CL model transformation. Here, negative bands observed can be linked to specific protein structures, such as the amide III α-helix (at 1340 cm^{-1} and 1286 cm^{-1}) and amide I (at 1619 cm^{-1} and 1669 cm^{-1}. Conversely, the primary positive bands appear to be associated with DNA/RNA, indicated by cytosine/uracil rings indicated at 785 cm^{-1}. This approach also enabled the distinction of human Tconv and Tregs from PBMCs with similar accuracy despite donor variability. However, a notable limitation of this method is its throughput;

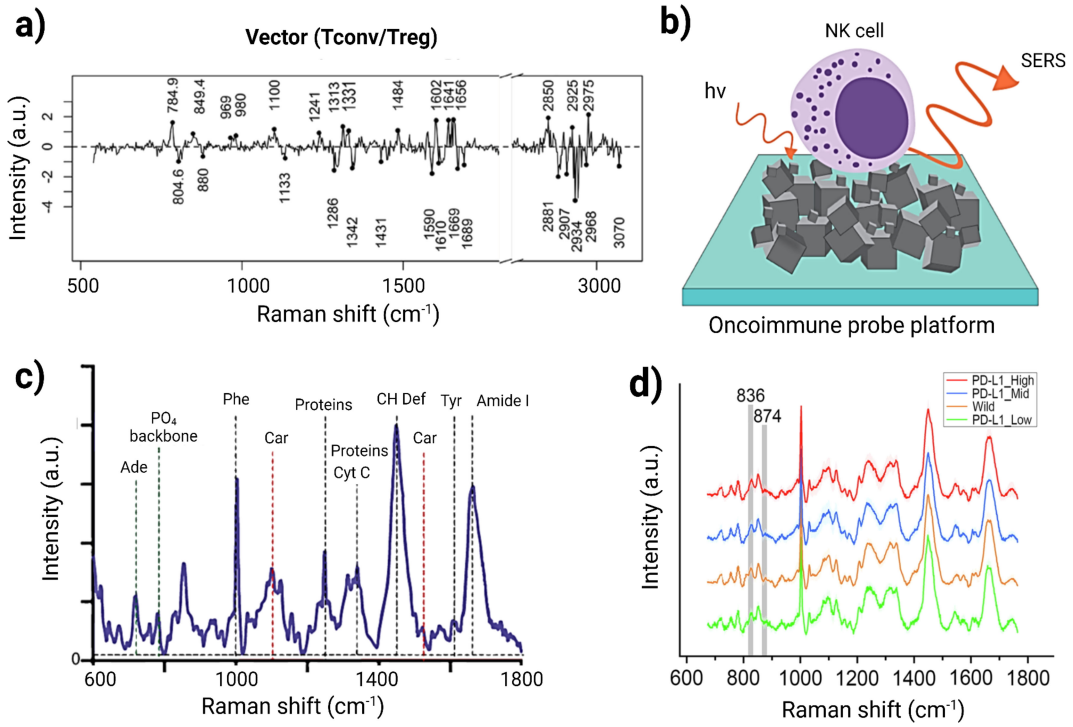


FIGURE 3
(A) Classification efficacy for human Tconv/Treg cells utilizing a separation vector to detect human Treg cells. Adapted with permission under a Creative Commons CC-BY License from ref (79). **(B)** Schematic illustration of NK cells on the OncoImmune probe platform, synthesized with 3D networks of nickel- nickel oxide nanocubiforms. **(C)** Representative Raman spectra of NK cells illustrating the presence of several biomolecules within NK cells. Adapted with permission under a Creative Commons CC-BY License from ref (80). **(D)** Average Raman spectra for PD-L1 expression in cancer cells. Adapted with permission under a Creative Commons CC-BY License from ref (81).

the automated sequential detection system currently processes approximately 1,000 cells per hour, which is insufficient for applications requiring millions of cells (79).

3.3 Natural Killer cells

Natural Killer (NK) cells are lymphocytes that play a crucial role in targeting viruses and cancer cells, particularly cancer stem cells (CSCs), which are linked to therapeutic resistance and tumor relapse (82, 83). Ishwar et al. explored the profiling of circulating NK cells as a diagnostic tool using SERS-driven liquid biopsy. The authors specifically synthesized an OncoImmune probe platform to detect metabolic changes in NK cells when they interact with tumor cells, illustrated in Figure 3B. Raman spectra of tumor-free NK cells exhibited characteristic bands associated with carbohydrates, proteins, and lipids, including peaks at 1450 cm^{-1} (CH deformation), 1661 cm^{-1} (amide I), 1555 cm^{-1} (amide II), and 1337 cm^{-1} (amide III) (Figure 3C). In contrast, tumor-associated NK cells showed altered spectral intensities, indicating an active response to tumor recognition. A decrease in the peak at 520 cm^{-1} suggested changes in Killer Immunoglobulin Receptor (KIR) expression due to CSC interaction. PCA revealed distinct clustering of NK cell signatures associated with breast, lung, and colon CSCs compared to non-cancer-associated NK cells. Utilizing machine learning, the study demonstrated that features of NK cell activity could accurately identify cancer from non-cancer samples using just 5 µL of peripheral blood, achieving 100% accuracy for cancer detection and 93% for localization. This research also highlights the importance of material advances for amplifying the SERS signal, where hybrid material consisting of nickel and nickel oxide produced an enhanced and reproducible SERS signal. This marker-free method generated a detailed NK cell metabolic profile that could be highly advantageous for cellular diagnostic applications. Thus, label-free SERS technique can be used for profiling immune cells and their metabolic changes in difficult to detect tumors such as small-cell lung cancer, triple-negative breast cancer, and colorectal adenocarcinoma (80).

3.4 Dendritic cell interactions

Dendritic cells (DCs) play a crucial role in cancer immunotherapy by interacting with cancer cells and presenting tumor antigens to T cells. When DCs capture antigens from cancer cells, their maturation status determines the immune response. Fully mature DCs effectively present these antigens on major histocompatibility complex (MHC) molecules, activating both CD4+ helper and CD8+ cytotoxic T cells. Enhancing DC function and antigen presentation is a key strategy in developing effective cancer immunotherapies (84). T cell receptors (TCRs) form an immunological synapse (IS) with antigen–MHC complexes and co-stimulatory ligands on dendritic cells (DCs), characterized by a distinct "bull's-eye" structure known as the supramolecular activation cluster (SMAC). Zoladek et al. employed label-free confocal Raman micro-spectroscopy (CRMS) to analyze the IS formed between laminin-treated DCs and T cells *in vitro*. They compared Raman spectral images with immunofluorescence imaging to identify signatures of key macromolecules, including nucleic acids, lipids, and proteins. Using a 785 nm laser, the study assessed the impact of laminin treatment on the DC–T cell junction by capturing images of control and treated DCs stained with phalloidin. Laminin treatment enhanced actin filament polarization and improved IS formation at the DC–T cell interface. The Raman spectra revealed detailed actin distribution in the IS, with characteristic peaks at 1450 cm−1 (CH deformation), 1661 cm^{-1} (amide I), 1555 cm^{-1} (amide II), and 1337 cm^{-1} (amide III). A significant band at 1003 cm^{-1} correlated to histone proteins present in the nucleus. For both DC and T cells, Raman spectral images in the 788 cm^{-1} band exhibit good concordance with the DAPI image, demonstrating the potential of CRMS for non-invasive imaging of live immune cell interactions and providing insights into the dynamics of the immunological synapse (85). This research plays an important role in designing dendritic cell based immunotherapies by providing real time data regarding DC-T cell interactions within TiME.

3.5 Cancer-associated fibroblasts

Cancer-associated fibroblasts (CAFs) are integral to the tumor microenvironment, often contributing to immunosuppression by stromal remodeling that protects cancer cells or communication with multiple immune cells via secreted factors. CAFs undergo metabolic changes that aid in tumor growth through interactions with cancer and stromal cells, their inherent plasticity leads to dynamic shifts in the fibroblast population. This emphasizes the need for precise evaluation of CAF's phenotypic and functional heterogeneity (86). Lipid metabolites released by CAFs not only facilitate metastasis but also serve as indicators of aggressive cancer types (87). The accumulation of lipids within the tumor microenvironment provides fatty acids to nearby tumor cells, fueling their energy needs. Since obesity is characterized by high levels of fatty acid, its impact on CAF's lipid metabolism remains poorly understood. Yeu et al. investigated this relationship using Raman spectroscopy as a non-invasive technique to analyze lipid metabolite changes in CAFs from endometrial cancer (EC) patients having different BMI. The study focused on Raman spectral regions associated with lipid biochemical changes (600–1800 cm^{-1} and 2800–3200 cm^{-1}). Through direct band and ratiometric analyses, researchers observed slight shifts in the CH2 symmetric stretch of lipids at 2879 cm−1 and CH3 asymmetric stretching from proteins at 2932 cm^{-1} in overweight or obese patient CAFs compared to non-obese patients. These shifts indicated a higher lipid content and increased lipid saturation in the obese CAFs and, with the help of PCA, metabolic phenotypes linked to obesity and cancer progression were effectively differentiated. The identification of specific Raman spectral signatures in CAFs offers valuable insights into the tumor microenvironment's influence on EC progression (88).

3.6 Tumor-immune microenvironment biomarker prediction

Designing effective studies to evaluate immunotherapeutic treatment efficacy poses a significant challenge, particularly regarding immune cell interactions and its characterization. The interactions within the TiME are intricate and dynamic, and understanding these interactions are essential towards predicting immunotherapy response. For instance, in Merkel cell carcinoma (MCC), research has shown that tumor-associated macrophages (TAMs) can express immunosuppressive markers that inhibit T-cell function. TAMs exhibit an immunosuppressive gene profile typical of monocytic MSDCs and notably express several immune checkpoint molecules that are potential therapeutic targets, such as PD-L1 and LILRB receptors (89, 90), which are absent on tumor cells. A study analyzing 54 tumor samples prior to immunotherapy revealed that a specific subset of TAMs (characterized by CD163+, S100A8+, CD14+) preferentially infiltrate tumors with a higher presence of CD8+ T-cells. Furthermore, a higher density of these TAMs was linked to resistance against PD-1 blockade therapies (91). In another study, single-cell RNA sequencing (scRNA-seq) revealed that a lower immune-cell infiltration (CD8 T-cell, NK cells, and a complete absence of γδ T-cells) was more common in acral melanoma when compared to non-acral melanoma (92). Tumor heterogeneity not only affects initial responses but also contributes to acquired resistance to immunotherapies which takes the form of immunosuppression and antigen escape. As tumors undergo immunotherapeutic treatments, they may develop subpopulations of cells that are resistant to immune-mediated cell death (76). These cases have been noted in melanoma (93) and breast cancer (94) studies and highlight the necessity of characterizing immune cell subsets and their activation states to tailor immunotherapy approaches effectively.

Raman spectroscopy has shown to be effective in immunological whole-tumor profiling, with Ou et al. showing the simultaneous detection of PET and SERS in monitoring the dynamics of tumor cell compositions *in vivo*. Currently, PD-L1 expression in TiME is the most important clinical biomarker assessed prior to immunotherapy use. High levels of PD-L1 have been associated with better outcomes in various cancers, including melanoma (95), lung cancer (96), and metastatic renal cell carcinoma (97). However, due to tumor heterogeneity, the relationship between PD-L1 expression levels in tissues and therapeutic responses to anti-PD-1/PD-L1 treatments is not always consistent (3, 98). This variability can be partially attributed to the influence of N-linked glycosylation on PD-L1, which may hinder the binding of commonly used anti-PD-L1 antibodies, thus the rapid glycosylation assessment possible with RS could enhance the reliability of PD-L1 as a biomarker for predicting responses to immune checkpoint therapies (99). Additionally, the expression of PD-L1 in both tumor and immune cells has been correlated to ICI clinical responses, making accurate PD-L1 characterization a valuable companion diagnostic for PD-1/PD-L1 inhibitors.

To assess PD-L1 expression, Zhou et al. developed an intra-operative technique using label-free Raman spectroscopy combined with ML for data analysis and visualizing PD-L1 in glioma cells, macrophages, CD8+ T cells, and normal cells. They employed stainless steel and Calcium Fluoride substrates to minimize background signals. Principal component analysis (PCA) was first utilized to differentiate Raman spectra between PD-L1$_G$ (high PD-L1 expression in glioma cells) and PD-L1$_L$ (Low PD-L1 expression in glioma cells) subgroups. Random Forest (RF) analysis identified five significant peaks at 723, 783, 837, 874, and 1437 cm^{-1}. PD-L1$_G$ exhibited stronger intensities at 837, 874, and 1437 cm^{-1} compared to PD-L1$_L$, which showed weaker intensities at 724 and 783 cm^{-1}. Figure 3D represents the average Raman spectra for PD-L1 expression in cancer cells. The peak intensities at 837 cm^{-1} and 834 cm^{-1} showed a positive linear correlation with PD-L1 levels. This is correlated with the increased expression levels of PD-L1 in glioma cells. The study also explored spectral differences among PD-L1$_G$, PD-L1$_T$ (high PD-L1 expression in T-cell), and PD-L1$_M$ (High PD-L1 expression in macrophage) subgroups, revealing biological correlations between cell types and their Raman spectral features. Notably, ganglioside, phosphatidylcholine (PC), and cytochrome-c contributed to PD-L1$_T$, while sphingomyelin and oleic acid were linked to PD-L1$_M$. The relationship between spectral features and biomolecule levels were qualitatively assessed across different cell types. Multiple ML algorithms—including CLS, HCA, SVM, and SA—were employed to analyze Raman spectra for model training and visualize PD-L1 expression in the glioblastoma immune microenvironment. This method for detecting the PD-L1 biomarker can be extended to other tumor biomarkers or target cells of interest, enhancing intra-operative diagnostics for surgical guidance and post-operative immunotherapy (81).

3.7 Predicting response to immunotherapeutic treatment

The current clinical metrics for prediction and evaluation of response to anti–CTLA4 and anti–PD-L1 immune checkpoint inhibitors (ICIs) in the TiME are not very effective (100, 101). PD-L1 score of 0, for example, can still demonstrate response to therapy and score is not currently utilized as a selection criteria for therapy (102). A liquid biopsy strategy combining blood count parameters, clinical characteristics, and serum lactate dehydrogenase predicted the response of patients without metastatic disease to anti–PD-1 therapy with about 60% accuracy (103). Studies have also leveraged PD-1/PD-L1 and CTLA4– targeting antibodies radiolabeled with 89Z for evaluating the tumor uptake of therapeutics using PET imaging; however, such measurements are associated with challenges (104). To address the challenges in predicting immunotherapy responses, Paidi et al. employed label-free Raman spectroscopy to monitor compositional changes in the tumor immune microenvironment (TiME). Using a CT26 murine model of colorectal cancer, tumors were treated with anti–CTLA-4 or anti–PD-L1 antibodies. Snap-frozen tumors were thawed, flattened, and positioned between a quartz

coverslip and an aluminum block with PBS to prevent dehydration, with the quartz selected for its low fluorescence interference. The team utilized a fiber-optic probe connected to a portable Raman system (830 nm diode laser) on a motorized translational stage to gather data. They collected 7,500 spectra from 25 tumors over a 5-second acquisition time. Ex vivo Raman mapping conducted three days post-treatment yielded 7,585 spectra from approximately 300 spatially distinct points across the tumors. Key Raman peaks identified included 849 cm^{-1} (C–O–C skeletal mode of polysaccharides), 1,260 cm^{-1} (amide III of proteins), 1,301 cm^{-1} (lipid and collagen bending), 1,448 cm^{-1} (lipid and collagen CH_2 bending), and 1,657 cm^{-1} (amide I of proteins). Comparisons between treatment groups revealed subtle yet statistically significant differences in lipid, nucleic acid, and collagen value, suggesting that responses to anti–CTLA-4 and anti–PD-L1 therapies influence TiME composition (34). These findings align with emerging research on the role of metabolism and the tumor microenvironment in shaping immune responses. Variations in lipid-based metabolites between treatments are likely to reflect differential lipid metabolism within the TiME due to immunotherapy (105). The machine learning analyses in this study demonstrated high prediction accuracy for treatment responses, highlighting precise spectral markers for each therapy. This study demonstrates that label-free Raman spectroscopy can sensitively detect early biomolecular changes in tumors. This is advantageous in offering valuable insights for clinical monitoring of immunotherapy responses in cancer patients.

4 Raman spectroscopy for drug response and metabolomic monitoring

The past years have seen breakthrough achievements in immunotherapeutic interventions including checkpoint inhibitors, cytokine-based immunotherapy, vaccines, and cell therapy (eg, CAR-T cell, CAR-NK cell and TIL therapy). However, the response to immunotherapeutic treatment has been variable among patients, and only a small percentage of cancer patients benefit from this treatment depending on the histological type of tumor and other host factors. In clinical practice, immunohistochemistry (IHC) typically serves as the initial method for assessing patient biomarkers. However, this approach has several limitations, including variability in assay results, ambiguous positivity thresholds, and instances where patients with low expression levels still show therapeutic benefits. It is also heavily dependent on the pathologist's judgment and experience (106). For patients suffering from cancer, imaging techniques like FDG-PET scans enhance understanding of metabolic changes during immunotherapy (107). Furthermore, radiolabeling checkpoint inhibitors with radioactive isotopes like 89Z allows for PET imaging to track the biodistribution of these inhibitors (108). Despite their utility, these methods often come with challenges related to cost, time, and the need for specialized personnel (109). As shown by some recent studies, researchers can leverage Raman spectroscopy to assess responses to immunotherapeutic drugs while simultaneously examining cancer cell differentiation (69), drug uptake within cells (110), and patterns of cancer metastasis (111, 112).

For drug response monitoring, techniques like colorimetric analysis, fluoroimmunoassay, ELISA, and radioimmunoassays are employed, each with distinct advantages and limitations. For instance, the complexity of ELISA protocols often involves multiple incubation and washing steps, making them time-consuming (113). This is especially challenging when working with large sample sizes. Furthermore, the reagents used are costly and can have lot-to-lot variability (114). In immunological studies, researchers commonly use techniques like flow cytometry, ELISA, and confocal microscopy to study the activation, polarization, and plasticity of immune cells along with their cytokine profiles. However, these methods often require the fixation of cells with paraformaldehyde (PFA), the addition of chemical dyes for labeling, and fluorescent tagging with antibodies—either conjugated or unconjugated. Such procedures can be invasive, costly, time consuming and may disrupt biological processes. One notable advancement for label-free drug screening is the Thermostable Raman Interaction Profiling (TRIP) method developed by Altangerel et al. (115). TRIP enables efficient screening of protein-ligand binding at low concentrations and doses under physiologically relevant conditions, as illustrated in Figure 4A. TRIP has been successfully applied to eight different protein-ligand systems which demonstrates excellent reproducibility in Raman measurements. The technique requires only a small 10 μL droplet of protein solution on a gold-coated glass slide which dissipates heat from the excitation laser while blocking fluorescent interference. Key applications of TRIP include time-dependent protein-drug binding using 2,4-dinitrophenol (DNP) with transthyretin (TTR), static protein-drug binding involving the streptavidin-biotin complex, and antigen-antibody binding detection with protein A and various antibodies, including those targeting the SARS-CoV-2 spike protein. TRIP is advantageous because of its cost-effectiveness and rapid detection capabilities. This eliminates the need for extensive sample preparation. Future enhancements could enable high-throughput drug screening and real-time monitoring of drug-target interactions, potentially improving drug development processes for complex immunotherapeutic interventions (115).

Single-cell RNA sequencing and other profiling methods allow researchers to study cells in detail, but these techniques destroy the cells during the several processing steps (116). On the other hand, Raman microscopy can analyze the vibrational energy of proteins and metabolites without damaging the cells, achieving a very fine resolution. However, it doesn't provide genetic information. Raman2RNA (R2R) is a new method that can predict single-cell expression profiles in living cells using label-free hyperspectral Raman images (Figure 4B). Either by combining Raman data with single-molecule fluorescence *in situ* hybridization or using advanced machine learning techniques. This kind of approach performed much better than traditional brightfield imaging, with cosine similarities of R2R > 0.85 compared to brightfield < 0.15. When reprogramming mouse fibroblasts into induced pluripotent stem cells, R2R effectively predicted the expression profiles of different cell states. Additionally, while tracking mouse embryonic stem cell differentiation, R2R identified early signs of lineage divergence and development paths (116).

Fluorescence-Activated Cell Sorting (FACS) has been a cornerstone for immunophenotyping and the detailed analysis of immune cell interactions. While FACS bridges the gap between

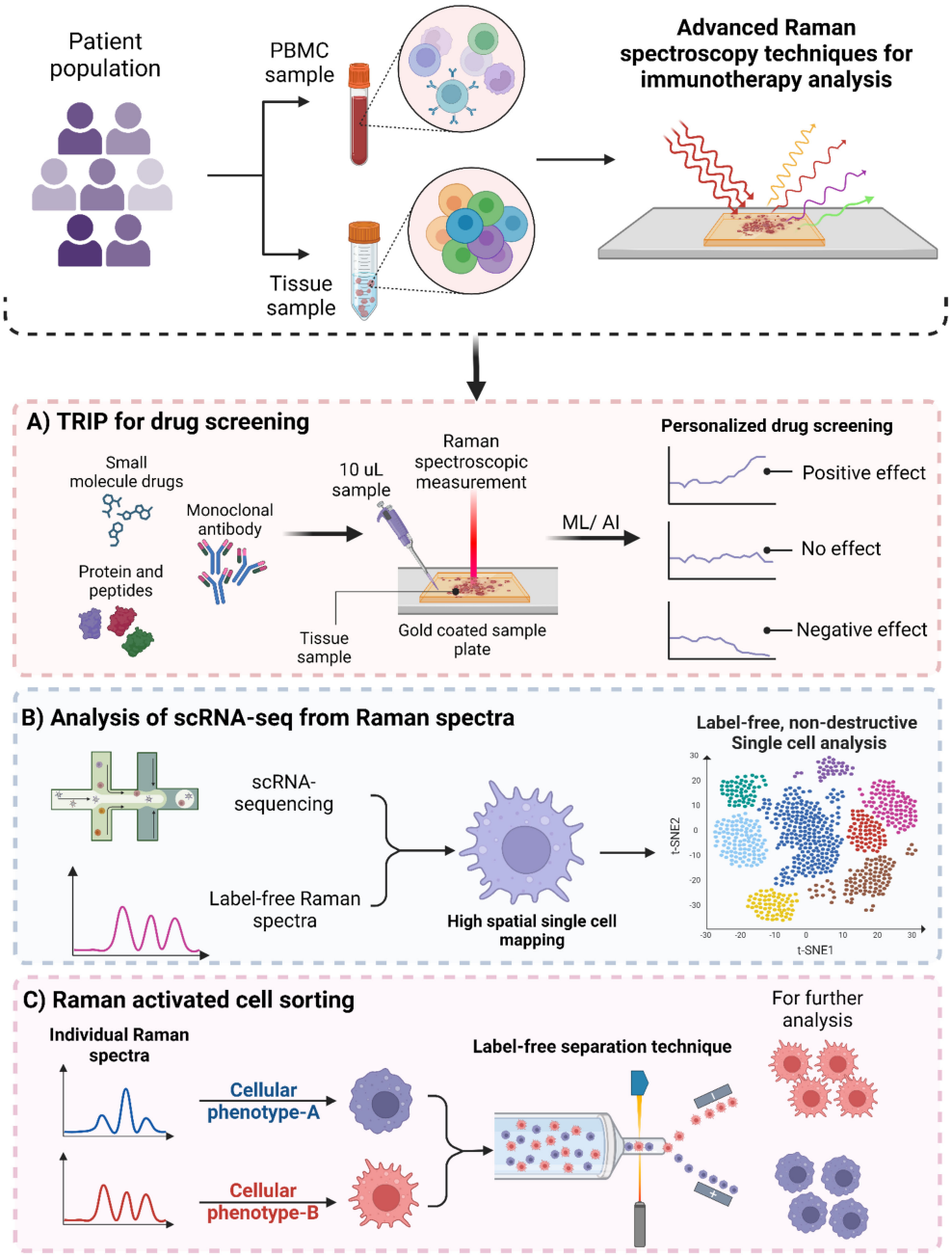


FIGURE 4
Advanced Raman spectroscopy based techniques for immunotherapy. **(A)** Schematic illustration of Thermostable Raman Interaction Profiling (TRIP) for personalized drug screening. **(B)** Raman spectroscopy based single-cell RNA sequencing for providing high spatial single-cell analysis. **(C)** Raman activated cell sorting (RACS) for label-free cell sorting.

genetic, cellular and population analyses, its reliance on fluorescent probes can interfere with cell metabolism and introduce reliability issues (117) and spectral spillovers (118, 119). Staining the cells with fluorescent dyes also impart cytotoxicity (120), alter the behavior of cells being analyzed (121), and breakdown of dyes which can result in reliability issues. It also limits its application in *in vivo* cell therapies such as stem cell therapy (122) and CAR-T cells (123). In contrast, Raman-Activated Cell Sorting (RACS) presents an exciting alternative. It allows for label-free immunophenotyping by measuring the emitted molecular vibrations of cells as illustrated in Figure 4C. RACS integrates multiple technologies to obtain single-cell Raman spectra using different cell-isolation techniques. These methods include operating in a flow environment with microfluidic systems, utilizing Raman tweezers for cellular analysis in solution, and employing Raman Activated Cell Ejection (RACE) for surface-based applications In a study by Wu et al. (124) they developed a novel approach using SERS combined with microfluidic technology to observe real-time interactions between cancer cells and the immune system. This platform is fully automated and integrates optofluidic systems which allows for

effective monitoring of these intercellular communications. This integrated system offers several key benefits. Firstly, it facilitates direct on-chip communication between cells. This helps to maintain the bioactivity and concentration of proteins released during interactions, thus closely mimicking the *in vivo* conditions. Secondly, a quantitative SERS immunoassay was employed to evaluate how various drugs influence the secretion patterns of cancer cells and the functionality of immune cells by utilizing an SERS-enhanced 3D barcode immunoassay. Moreover, this automated system significantly minimizes human error and simplifies operational complexity, enhances the reliability of results in drug screening and immunotherapy research.

Amongst most critical applications to date, Raman spectroscopy can probe tumor metabolism in the TiME (125, 126) as growing evidence suggests that the metabolic state of the TiME plays a crucial role in the success of cancer immunotherapy. The TiME can significantly influence the energy consumption and metabolic reprogramming of immune cells, often causing them to become tolerogenic and less effective at eliminating cancer cells. Understanding these metabolic interactions is key to improving immunotherapy outcomes. Unlike mass spectrometry-based single-cell metabolomics, which requires destructive sample preparation (127), label-free Raman spectroscopy can analyze metabolites in living cells and tissues in a non-invasive manner. This makes it well-suited for *in vivo* investigations of tumor metabolism. Recent studies have utilized Raman confocal microscopy combined with ML algorithms to analyze the activation of immune cells such as T cells, B cells, and monocytes (28). For example, Chaudhary et al. employed Raman micro-spectroscopy to identify activated immune cells. Their study included both cell lines and primary cells consisting of purified subgroups of monocytes and lymphocytes, as well as mixed populations of peripheral blood mononuclear cells (PBMCs), all obtained from healthy donors. ML models were designed for cell differentiation and evaluated against flow cytometry data. Spectral signatures of T-cell, B-cell and monocytes before and after activation were also determined using high performance classification models, including spectral fitting to identify spectral biomarkers (28). Importantly, these analyses were conducted alongside traditional methods like flow cytometry and ELISA in both *in vitro* and ex vivo models. The findings indicate that immune cells exhibit unique spectral profiles in response to different stimuli, highlighting the critical roles of both cell type and specific activating signals in shaping their responses. For instance, upon activation, T cells may undergo significant changes in lipid metabolism and protein synthesis, while monocytes might show alterations in cytoskeletal dynamics. These biochemical shifts vary among different immune cell types and are indicative of the complex signaling pathways that govern their activation and differentiation. By examining these spectral changes through Raman spectroscopy, researchers can gain valuable insights into the mechanisms driving immune responses (28). This understanding could pave the way for developing targeted therapeutic strategies aimed at effectively modulating immune function. For example, if specific spectral signatures are associated with effective T cell activation against tumors, therapies could be designed to enhance these pathways for improved cancer treatment outcomes.

5 Integrating Raman spectroscopy with -omics approaches and progress towards clinical use

The "one-size-fits-all" model in immunotherapy often fails to account for individual variations in genetics, environment, and lifestyle, limiting the effectiveness of immunotherapy for many patients (128). Multi-omic approaches that synthesize divergent tumor features such as genomics, transcriptomics, proteomics and metabolomics have significantly advanced the detailed description of heterogeneous tumors and facilitated better understanding of immunotherapy responses (129–131). Integrating multi-dimensional data from various -omics layers remains a significant challenge, and translating these data into precise drug selection for clinical applications has yet to be realized. Additionally, the high costs and labor-intensive nature of genomics, transcriptomics, proteomics, lipidomics, and metabolomics studies require sophisticated analytical and statistical methods. Consequently, these factors have limited the longitudinal capture of events across clinical studies (104, 105). Raman spectroscopy presents a crucial opportunity to harmonize these -omics into a single phenotypic, "Raman-omic" technique. Figure 5A illustrates the role of Raman spectroscopy in multi-omics approaches in immunotherapy, to delineate patient heterogeneity, reduce time for analysis, reduce cost associated with those analyses, and harmonize data for better ML/AI analysis by reducing heterogenous data incompatibility. In this section we discuss how Raman spectroscopy can be used to complement and augment genomics, transcriptomics, proteomics, and metabolomics in immunotherapy.

5.1 Raman spectroscopy in genomics and transcriptomics

Detecting specific DNA sequences and identifying single-nucleotide polymorphisms (SNPs) are vital for cancer diagnostics and in predicting immunotherapy treatment outcome (134). Next-generation sequencing (NGS) highlights the potential of somatic DNA markers as both independent indicators and novel therapeutic targets (135, 136). Raman spectroscopy has significant potential for studying genomic and transcriptomic alterations. In particular, changes in the vibrational modes of DNA and RNA, including miRNA, can indicate mutations or epigenetic modifications relevant to cancer. Studies have indicated that the activation state of T cells is primarily linked to alterations within DNA rather than proteins (137–139). Chromosomal DNA degradation of activated mature T cells when stimulated via the CD3/T cell receptor complex experience rapid apoptosis. This DNA degradation plays a crucial role in eliminating autoreactive T cells in the thymus (140, 141). In a study by Lee et al., they focused on the Raman spectral analysis of activated mature $CD8^+$ T cells and their DNA changes during apoptosis. They noted a decrease in Raman spectral intensities related to DNA, specifically at 768, 1071, and 1463 cm^{-1}. These

10.3389/fimmu.2024.1520860

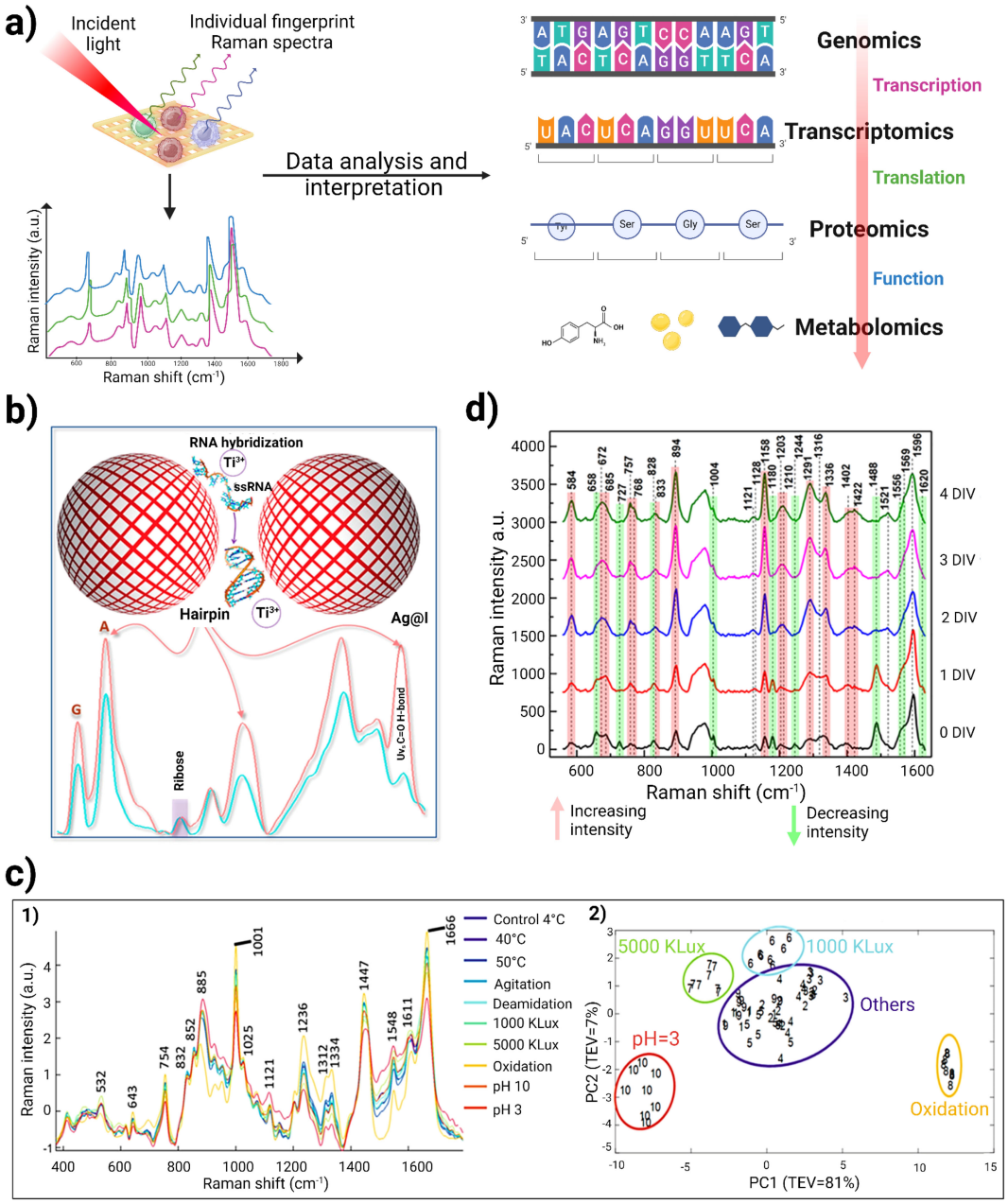


FIGURE 5
(A) Schematic representation of utility of Raman spectroscopy in multi-omics study. When incident light strikes the cells of interest, it generates individual fingerprint Raman spectra. This provides information regarding molecular and chemical composition within cells. Raman spectroscopic data analysis and interpretation using various ML/AI techniques can provide insights for genomics, proteomics, transcriptomics and metabolomics. **(B)** Schematic representation of label-free miRNA identification, using Titanium ions to induce silver nanoparticle "hotspots" to identify RNA sequences of homopolymeric bases and to locate the peak position of each base in the Raman spectrum. Adapted with permission under a Creative Commons CC-BY License from ref (48). **(C)** 1) Raman spectra obtained for 8 different degradation studies of therapeutic monoclonal antibodies was validated against conventional size-exclusion chromatography and peptide mapping. 2) represents the PCA analysis of RS, which can clearly demarcate samples from different degradation clusters (pH 3, oxidation, 5000 kLux·h and 1000 kLux·h) from the control group to allow rapid analysis for therapeutic quality control (132). **(D)** Raman spectra of the DMEM culture medium recorded at various Days *in vitro* (DIV). The red and green lines in the spectra highlight peaks that show increasing and decreasing intensities, respectively. Adapted with permission under a Creative Commons CC-BY License from reference (133).

intensity reductions likely reflect the breakdown of the DNA's ring structure, signaling its disintegration during apoptosis. Notably, significant changes were observed in the O-P-O region of the DNA backbone (around 780 to 800 cm^{-1}) and in PO_2 (around 1053 to 1087 cm^{-1}). This suggests a correlation with internucleosomal DNA cleavage progression. The differences in Raman spectra between resting and activated mature $CD8^+$ T cells were analyzed using PCA which revealed a clear discrimination of DNA from activated T cells compared to resting T cells. Thus, this study infers that the decreased Raman intensities in activated mature $CD8^+$ T cell

DNA are indicative of apoptosis, highlighting the utility of label-free Raman spectroscopy as a tool for assessing the activation status of these immune cell (142).

In parallel, Li et al. developed a unique detection method for capturing SERS signal from unlabeled RNA without hampering its structural integrity. They utilized titanium ions as an aggregating agent along with silver nanoparticles. This formed electromagnetic "hot-spots" for non-destructive and label-free single molecule detection of miRNA molecules. Unlike traditional metal cation aggregators (like Al^{3+} and Mg^{2+}), the acidic titanium ions helped stabilize RNA molecules. The researchers conducted SERS analysis on homopolymeric sequences of the four RNA bases (A, G, C, and U) and examined the secondary hairpin structure (Figure 5B). The ribose peak at 959 cm^{-1} was used for normalization, revealing distinct peak positions for each base: A at 731 cm^{-1}, G at 665 cm^{-1}, C at 789 cm^{-1}, and U at 795 cm^{-1}. To check the robustness of their system, they designed RNA sequences of IL10 and 1HP3 which contained the same bases but in a different sequential manner. A peak at 1446 cm^{-1} corresponded to U vibrations in AU base pairs, while increases in peak intensities at 1314 cm^{-1} (G in GC pairs) and 1635 cm^{-1} (C in GC pairs) indicated complementary pairing. This label-free detection method for miRNA demonstrated a high signal-to-noise ratio with remarkable sensitivity while preserving the original structure of miRNA. This research reduces the analysis cost of miRNA characterization as well as supporting the development of miRNA therapeutics in the future (48).

5.2 Raman spectroscopy in proteomics and peptidomics

In the context of cancer diagnosis and new therapeutic development, proteomics plays a valuable role for identifying biomarkers. By analyzing proteins expressed in cancerous tissues compared to healthy tissues, researchers can discern proteins that are uniquely or differentially expressed in either state. Label-free Raman spectroscopy can characterize proteins and their conformational states, providing insights into their roles in cancer. Uzunbajakava et al. demonstrated the first successful use of nonresonant Raman imaging to analyze protein distribution in cells. This study compared Raman images of two cell types: peripheral blood lymphocytes (PBLs) and lens epithelial cells (LECs). The Raman images revealed distinct differences in protein distribution within the nuclei of PBLs and LECs, with clear contrasts in protein intensity visible in the PBL nucleus (near 3000 cm^{-1}) (143). Raman scattering can also be utilized to study the α and β-sheets conformations and changes in proteins. Rygula et al. explored the secondary structures of 26 different proteins (including hemoglobin (Hb), cytochrome c, peroxidase, albumin, collagens, lectins, glucose oxidase, proteinase, ubiquitin, and heme protein) using Raman spectroscopy by analyzing their Amide I and III vibrations, which reveal the ratios of α-helices and β-sheets (144). This research suggests that proteoforms may each have their own vibrational fingerprint. Therefore, even when specific binders are unavailable to discern, eg, post-translationally-modified proteins, Raman can prove specific information about modifications or changes in secondary or tertiary structure.

Peptidomics can also benefit from Raman spectroscopy. For example Raman spectroscopy can help in understanding the roles of specific peptides involved in tumor cell signaling and immune responses (145). Raman spectroscopy has also emerged as a promising tool for detecting post-translational modifications (PTMs) and assessing degradation in monoclonal antibody (mAb) therapeutics (132). These modifications, which occur after protein synthesis, can significantly impact the structure and properties of antibodies, leading to issues like aggregation and fragmentation. PTMs are classified based on the modified amino acids or the enzymes involved, with common modifications including phosphorylation, glycation, acylation, alkylation, glycosylation, deamidation, and oxidation. This is particularly important in mABs, where structural changes can result in unwanted immune reactions (146), decreased effectiveness (147), and material loss during production (148). Monoclonal antibodies are especially vulnerable to aggregation and fragmentation due to various processing conditions with soluble mAB aggregates posing a significant risk for triggering unwanted immune responses (149). A label-free and high throughput Raman spectroscopy can aid in identifying these PTMs in real-time. Due to rapid spectral data collection, little to no sample preparation, and without any interference due to water, Raman spectroscopy emerges as an outstanding candidate for real-time Process Analytical Technology analysis in biotherapeutic production (150). For instance, McAvan et al. studied the effectiveness of label-free RS in detecting PTMs in IgG4 mAbs under various degradation conditions, such as changes in pH (3 and 10), temperature (4, 40, and 50°C), light stresses (1000 and 5000 kLux·h), and agitation. By integrating principal component analysis (PCA) with RS and circular dichroism (CD) spectroscopy, they differentiated mABs based on their PTMs and degradation states. Figure 5C–1 represents Raman spectra which were obtained for 8 different degradation data. Notably, spectral peaks at 1666 cm^{-1} and 532 cm^{-1} remained stable which indicates that β-sheet and disulfide bonds were largely unaffected by these conditions. However, significant changes were observed in the amide III region (1312 to 1334 cm^{-1}), suggesting alterations in the protein's tertiary structure linked to the degradation conditions. Additionally, RS detected shifts at 885, 1121, and 1450 cm^{-1} associated with tryptophan and other molecular components, showing that both tryptophan and C–H vibrations increased in wavenumber with larger aggregates. Conversely, the C–N backbone exhibited a decrease in wavenumber as aggregation increased. This research highlights the potential of RS for monitoring PTMs in mAb which were subjected to various forced degradation conditions. The PCA analysis revealed that the data with identical conditions group together. This indicated that the data is consistent and reproducible. Notably, the samples that form distinct clusters apart from the control group include those subjected to oxidation, pH 3, and light exposure at 5000 kLux·h and 1000 kLux·h which is represented in Figure 5C–2 (132). Furthermore, Zhang et al. used a label-free RS approach along with SVM and PCA model for

quantitative prediction of protein aggregation in Antibody Drug Conjugates. Additionally, they have also investigated the impact of temperature and humidity (40°C/75% RH/1 month) on aggregation of proteins that mimics long-term storage conditions (151). These studies suggest that label-free Raman spectroscopy can be used to monitor real-time PTMs during biotherapeutic production.

5.3 Raman spectroscopy in metabolomics

Immunometabolomics has become a vital area of study by providing detailed insights into the metabolic interactions within the TiME. The transfer of metabolites between cancer cells and nearby immune cells can shape immune responses, indicating that these metabolic exchanges are key to both immune surveillance and evasion. Research is focused on understanding the vital contribution of metabolic communication between these cells, particularly how tumor metabolism contributes to immune evasion and resistance to immunotherapy (152). Tumor metabolism leads to the buildup of metabolites such as lipids, carbohydrates etc. that regulate immune responses within the TiME (153). These metabolites not only serve as signals but also interfere with the development of immune cells such as CAFs, T-cells and macrophages (154–156). There is an urgent need for new techniques that allow for single-cell metabolic interaction analysis in a quick and cost-effective way. To overcome these hurdles, researchers have utilized Raman spectroscopy for understanding these intricate immune-cell metabolic cross talks. For example, Shalabaeva et al. used a time-resolved method for metabolite tracking in cell culture using label-free SERS, allowing simultaneous analysis of multiple molecules without any sample processing. The method used Ag nanostructures integrated in cell culture medium in a four day study involving NIH/3T3 cells, with Raman spectra collected from media. The analysis of specific peaks revealed temporal changes in metabolic components such as L-tyrosine, L-tryptophan, glycine, L-phenylalanine, L-histidine, and proteins from fetal bovine serum (FBS), as seen in Figure 5D. The observed trends for L-tyrosine and its degradation products-acetoacetate and fumarate signified the consumption of L-tyrosine and simultaneously the production of its breakdown products. The decreasing intensity of certain peaks likely indicates exhaustion of these cell medium components over time. This method was also applied to analyze LPS-driven differentiation of Raw 264.7 macrophage cells. Analysis of the Raman spectra collected over a 24hr period reflected macrophage transition from quiescent to an activated pro-inflammatory state. This research indicates that label-free SERS could identify different metabolites at various time points, thereby providing insights into the immune cell states (133).

In cancer metabolomics, lipid metabolism plays a crucial role in cancer development, progression and also influences tumor growth mechanisms, including support for metastasis, ferroptosis-mediated cell death, and interactions between tumor and immune cells (157). Abnormal lipid levels and disrupted metabolic pathways contribute to cancer growth, metastasis, and treatment resistance. As Raman vibrational peaks are exceptionally sensitive for observing lipid content, Raman spectroscopy is increasingly applied for lipidomic analysis across a wide range of cancers (157–159). Lipid droplets (LDs) are dynamic organelles primarily involved in lipid storage and metabolism, and dictating cellular energy balance and signaling. Their significance in cancer biology has garnered attention, particularly regarding resistance to chemotherapy, their interactions with immune cells within the TiME, and implications for immunotherapy (160). It was found that a significantly higher quantity of lipid droplets was present in high-grade glioblastoma and colorectal cancer when compared with low-grade cancers and normal tissues (161). Ben et al. utilized Multiplex coherent anti-Stokes Raman scattering (MCARS), a label-free technique, to detect lipid droplets in colon cancer cell lines expressing the neurotrophin receptor TrkB. The overexpression of TrkB subsequently activates the PI3K/Akt signaling pathway and phosphorylation of Akt (P-Akt), leading to lipid droplet formation in cells. The MCARS technique focused on the 2500–3200 cm−1 spectral range, where the CH2 (2850 cm−1) and CH3 (2930 cm−1) vibrational signatures are primarily associated with lipid and protein contents respectively. MCARS images of cells generated from signal integration of CH2 stretching modes allowed researchers to discriminate between lipid accumulation in the endoplasmic reticulum and the formation of cytoplasmic lipid droplets. This approach tracked the changes in lipid metabolism in both TrkB high-expressing HT29 cells and low-expressing HEK293 cells following treatment with brain-derived neurotrophic factor (BDNF), demonstrating that BDNF-induced TrkB activation leads to lipid droplet formation in HT29 cells. Thus, with MCARS along with data processing, researchers were able to a) detect cancerous cells, b) assess the tumor progression, and c) predict the resistivity of cancer cells by analyzing the content of cytoplasmic lipid droplets (162).

5.4 Translational potential of Raman spectroscopy in cancer diagnosis and treatment

Raman spectroscopy is increasingly recognized for its clinical utility in cancer diagnosis and therapy (163, 164). One notable application of label-free Raman spectroscopy is intraoperative margin assessment of brain tumors, particularly glioblastomas. Studies have shown that Raman spectroscopy can differentiate between tumor and healthy brain tissue in real-time during surgical procedures, potentially improving surgical outcomes by ensuring complete tumor resection while preserving surrounding healthy tissue. Jermyn et al. utilized a handheld Raman spectroscopy probe, without any labeling, for real-time detection of cancer cells in human brain tissue during surgery. This technique achieved a sensitivity of 93% and specificity of 91%, effectively distinguishing between normal brain tissue, dense cancer, and cancer-invaded areas. The probe illuminated a 0.5-mm tissue area, sampling up to 1 mm deep in just 0.2 seconds, integrating seamlessly into neurosurgical workflows. The spectra covered a range of shifts from 381 to 1653 cm−1. The Raman spectra revealed

distinct differences in lipid bands and nucleic acid content between cancerous and normal tissues. Specifically, variations at 700 cm−1 and 1142 cm−1 indicated changes in cholesterol and phospholipids, while increased bands from 1540 to 1645 cm−1 suggested higher nucleic acid levels in cancerous tissues. With ML analysis, they were able to classify the samples with an overall accuracy of 92% (107). This portable Raman technology enhances intraoperative decision-making by providing quick, reliable identification of invasive cancer, minimizing residual tumor volume and improving patient survival outcomes. Raman spectroscopy has also now been used for real time cancer cell differentiation and diagnosis in oral cancer (165, 166), gastric cancer (167), and skin cancer (168).

Furthermore, Raman spectroscopy is gaining momentum as a non-invasive diagnostic tool in oncology-based clinical trials (Table 1). In a recent clinical investigation by Wang et al., serum samples from 729 patients diagnosed with either prostate cancer (PC) or benign prostatic hyperplasia (BPH) were analyzed. The researchers utilized SERS combined with an AI model based on convolutional neural networks (CNN) for diagnostic purposes. Their findings indicate an accuracy of ~85% for distinguishing between PC and BPHBy integrating patient age and prostate-specific antigen (PSA) levels into their multimodal CNN approach, the classification accuracy improved significantly to over 88% (169). Encouraged by these results, the researchers have initiated a clinical trial to explore this diagnostic technique, registered under NCT05854940 (170).

In another example, label-free RS has been used for diagnosis and staging of diffuse large B-cell lymphoma (DLBCL) and chronic lymphocytic leukemia (CLL) (171, 172). Label-free Raman spectroscopy (RS) has emerged as a valuable tool for diagnosing and staging diffuse large B-cell lymphoma (DLBCL) and chronic lymphocytic leukemia (CLL). In a study conducted by Chen et al. (2022), label-free SERS spectra were obtained from 47 healthy controls and 53 DLBCL patients. AgNPs was used as a substrate for SERS

TABLE 1 Current clinical trials of Raman spectroscopy for cancer diagnosis and treatment.

Sr. no	NCT number	Study title	Conditions	Interventions	Brief summary
1	NCT04162431	DOLPHIN-VIVO: Diagnosis Of LymPHoma IN Vivo (Ex Vivo Phase)	Lymphoma; Head and Neck Cancer	Combined FNA/Raman spectroscopy needle probe	Study for the use of Raman spectroscopy for non-invasive analysis of lymph node tissue (*x-vivo* and *in-vivo*) for providing immediate diagnostic results without laboratory delays. It aims to streamline the biopsy process by integrating fine needle aspiration during routine procedures, maintaining clinical standards.
2	NCT05010369	DOLPHIN-VIVO: Diagnosis of LymPHoma in Vivo (*In Vivo* Phase)			
3	NCT06384924	Raman Spectroscopy and Skin Cancer	Skin Cancer; Basal Cell Carcinoma; Squamous Cell Carcinoma	Handheld Raman Spectroscopy probe	Retrospective trial investigating the effectiveness of Raman Spectroscopy in assessing skin cancer tumor size and spread using a handheld probe that gently contacts the skin with laser light. This method aims to enhance diagnostic accuracy and efficiency.
4	NCT06394050	Label-free Raman Spectroscopy for Discrimination Between Breast Cancer Tumor and Adjacent Tissues After Neoadjuvant Treatment	Breast cancer	Label-free Raman spectroscopy based diagnosis	This trial aims to utilize label-free Raman spectroscopy to distinguish between cancerous cells and adjacent tissues in breast cancer patients' post-treatment.
5	NCT04817449	Spectroscopy in Ovarian Cancer	Ovarian Cancer; Ovarian Neoplasms	Raman spectroscopy	This trial investigates the utility of label-free RS for early detection of ovarian cancer, by analyzing blood plasma (from ovarian cancer patients) and fibrotic tissue (post-chemotherapy) with label-free RS to identify active cancer.
6	NCT04869618	Validation of an Artificial Intelligence System Based on Raman Spectroscopy for Diagnosis of Gastric Premalignant Lesions and Early Gastric Cancer	Gastric Intestinal Metaplasia; Gastric Cancer	AI and Raman spectroscopy-based device (SPECTRA IMDx)	Study for using Raman spectroscopy based artificial intelligence system (SPECTRA IMDx) for early detection and treatment of gastric premalignant lesions and early gastric cancer (EGC).
7	NCT05854940	Multicenter, Prospective Clinical Study of the Serum Raman Spectroscopy Intelligent System for the Diagnosis of Prostate Cancer	Prostate Cancer	Serum Raman spectroscopy intelligent diagnostic system	Trial for validating the effectiveness of RS at screening prostate cancer by detecting prostate-specific antigen (PSA)focusing on early prostate cancer diagnosis.
8	NCT05995990	Raman Spectroscopy for Liver Tumours Following Liver Surgery	Colorectal Cancer Metastatic	Raman Spectrometry	Trial utilizing both RS and multivariate spectral analysis to develop a quick and reliable method for evaluating tissue sections for residual tumors in liver samples after surgery.

analysis. Their analysis revealed that DLBCL samples exhibited higher spectral intensities at 725, 1093, 1329, 1371, and 1444 cm^{-1}, indicating the presence of biomolecules such as hypoxanthine, adenine, thymine, collagen, and phospholipids. While lower intensities were observed at 493, 636, 888, 1003, 1133, 1580, and 1652 cm^{-1} which relate to ergothioneine, uric acid, tyrosine, lactose, phenylalanine, acetoacetate, amide I, and alpha-helix. They also found distinct spectral variations between early-stage (I and II) and late-stage (III and IV) DLBCL. To analyze the complex SERS data effectively, multivariate techniques were employed. The k-nearest neighbors (kNN) model demonstrated better results in both diagnosing and staging DLBCL with an accuracy of 87.3%, sensitivity of 0.921 and specificity rates of 0.809 for diagnosis (171). In another patient-centric study, Bai et al. explored the potential of RS to create a blood test for the noninvasive detection of DLBCL and CLL through biomarker analysis. They examined blood plasma samples from 33 DLBCL patients, 39 CLL patients, and 30 healthy individuals. Their analysis revealed that the intensity at 1445 cm^{-1}, associated with collagen and lipids, was notably higher in DLBCL samples. Conversely, the intensity at 1655 cm^{-1}, linked to proteins and alpha-helix structures decreased in CLL samples while increasing in DLBCL samples. By combining RS with orthogonal partial least squares discriminant analysis (OPLS-DA), the researchers were able to differentiate the blood plasma of CLL and DLBCL patients from that of healthy donors. This integrated approach achieved sensitivity rates of 92.86% for CLL and 80% for DLBCL along with specificity rates of 100% and 92.31%, respectively (172). To further this research, various ongoing clinical trials are investigating both ex vivo and *in vivo* diagnostic methods for lymphoma detection. These trials highlight the current clinical need in cancer diagnostic approaches, especially in cancer immunotherapy. With the advancement of ML and AI technology, integrating RS in biomarker prediction as a diagnostic tool can be crucial for a personalized approach in immunotherapy. This will help solve many current limitations which are present in immunotherapeutic treatment.

6 Future directions in label-free assays to develop personalized therapeutic approaches

Label-free Raman spectroscopy in cancer diagnosis and immunotherapy is poised to revolutionize the landscape of oncological care. As a non-invasive diagnostic tool, label-free Raman offers advantages in terms specificity and throughput, enabling the detection of molecular signatures associated with various cancers directly from biofluids such as blood, urine, and saliva (173), distinguishing various tissue types, and detecting pathological alterations across a multitude of diseases. Preclinical, translational, and clinical *in vivo* applications have significantly enhanced Raman spectroscopy's role in bridging crucial knowledge gaps, especially in the complex analysis of whole-tissue to accurately describe tumor microenvironments. However, several challenges persist in utilizing Raman spectroscopy as a standalone multi-omic test or as a complementary tool to existing multi-omics. Achieving the ambitious goal of entirely label-free assays that are low-cost and high-throughput is essential for accelerating clinical patient studies.

One important step to advancing Raman application for cancers is increasing utilization of formalin-fixed paraffin-embedded (FFPE) specimens, where currently-described studies predominantly concentrate on fresh or frozen tissue samples. FFPE is the conventional method used for the preservation and storage of tissue, especially tumor sampling that is a very small size such as in melanoma or biopsies of metastatic tissues. Due to the dominant vibrational signal native in paraffin, deconvoluting the relatively weak signature of the tissue spectra from paraffin spectra remains a persistent challenge. Robust suppression of the background signal from the paraffin, either through chemical dewaxing demonstrated by Ning et al. and Gaifulina et al (174, 175), digital processing as shown by Tfayli et al. and Ibrahim et al. (176, 177), or vibrational mode-suppressing SERS devices as shown by Kurouski et al. (178), can greatly increase the possible patient data banks available to process and construct the necessary library for the integration of Raman into multi-omic studies. A notable study by Lewis et al. exemplifies this potential by utilizing label-free Raman spectroscopy to compare findings with immunohistochemistry (IHC). They generated Raman spectral maps from FFPE colonic tissues obtained from healthy individuals and used principal components analysis (PCA) to validate their findings against several IHC markers. Their results demonstrated the ability to differentiate mucin based on glycosylation patterns, identify nuclear regions through DNA content analysis, and categorize various tissues according to their amino acid compositions. Their results confirm excellent correlation between the IHC markers and label-free RS. This assures that label-free RS could be utilized in detecting compositional changes, thus eliminating the use of expensive antibodies (179). Ability to access the wealth of banked and stored FFPE could facilitate the next leap in biologic study by greatly expanding available specimens.

A second step for widespread adoption of RS in clinical care, particularly at point of care sites, is efficient sample pre-processing and data post-processing. Clinical integration of sample preprocessing techniques prior would greatly facilitate Raman analysis by eliminating unwanted background and noise. Common sample preparation materials and ubiquitously-present chemical molecules can often obscure and influence relevant functional group vibrational signals. Strategic suppression of non-relevant chemical groups or biological bands either chemically (180) or through Raman-active platforms (181) can greatly improve functional group targeting and better map them to observable biomarker differences. Additional construction of a global spectra library would further assist in signal deconvolution and aid in standardization across samples. Timely tumor profiling will also require rapid integration at subcellular resolution over the entire tissue sample. As such, utilization of higher-throughput Raman systems enabling line- or image-based spectral collection pathways can greatly improve spectral acquisition throughput and capacity. Higher collection bandwidth can aid in the population of a data bank derived from historical samples.

Further advancements in Raman-based tumor investigations necessitate continuous enhancements in the technical performance

of spectral acquisition and the resolution of signals. While current SERS devices strategically drive enhancements at the incident light source, signal intensity can further be amplified by additional consideration of enhancements in the scattering wavelength regime. Design of doubly resonant platforms, with the second broader resonance directing Raman scattered light towards the detector, can yield increased spectra intensity and sensitivity. Further, multi-resonant platforms accounting for polarization dependency can enable sample filtering by polarizability. Careful considerations will need to address spectral fidelity associated with fabrication imperfections and hotspot intensities variations across regions. Finally, as tumors and the TiME are most faithfully depicted as a three-dimensional ecosystem, future SERS designs should extend applicability to include all three degrees of spatial freedom. Although confocal RS has been utilized as a 3D molecular contrasting tool (180, 182), similar applications have not yet been applied in SERS-based tumor studies. Potential adoption of suspended or resonantly stratified NPs could provide z-stacking capabilities, while maintaining high-sensitivity. Similarly, considerations will need to be taken to address hotspot uniformity and off-focus signal contributions.

The field of AI and machine learning in Raman spectroscopy data analysis has revolutionized the way we approach real-time data interpretation, particularly in single-cell and multi-omics studies. These LLM models have shown remarkable potential in integrating diverse data types, allowing researchers to simultaneously characterize different cellular processes. However, the journey from laboratory research to clinical application of Raman spectroscopy to immunotherapy principles faces several hurdles. One significant challenge lies in the data acquisition process, which often lacks standardization. Researchers employ varying methods for sample preparation, instrument operation, and data labeling. This leads to inconsistencies across different studies. To address this, the scientific community could benefit from establishing a global, public database for Raman spectroscopy data. This repository would not only store data from laboratories worldwide but also implement standardized normalization and preprocessing techniques, paving the way for more robust AI and ML method development. Another pressing issue is the "black box" nature of many AI models. While these algorithms excel at producing results, the opacity of their decision-making processes can be a stumbling block for clinical adoption. Healthcare professionals understandably hesitate to rely on tools they cannot fully comprehend or explain. Therefore, enhancing the transparency and interpretability of these models is crucial for their acceptance in medical settings. Looking ahead, the field of immunotherapy applications using Raman spectroscopy and AI has several promising fields for growth. Multi-center studies should be prioritized to improve data consistency and reliability, as current research often relies on single-center data. Additionally, the development of semi-supervised or unsupervised machine learning models could unlock new possibilities beyond current applications. These advanced models could potentially uncover hidden correlations between various omics data sets, opening doors for innovative hypothesis testing, drug discovery, and personalized medicine approaches in immunotherapy (35). By addressing these challenges and exploring new frontiers, the integration of AI, Raman spectroscopy, and immunotherapy holds promise for advancing our understanding of cellular processes and improving patient outcomes.

These technical advancements are crucial not only for studying therapeutic responses and discovering biomarkers but also for achieving precision immunotherapeutics. Accurate intraoperative diagnosis for complete tumor resection is essential for improving prognosis and determining optimal surgical approaches in multi-modal care settings.Raman spectroscopy has demonstrated the ability to distinguish malignant tissue from healthy tissue in real-time that can facilitate margin assessment and *in vivo* pathologic classification (183, 184). For example, one recent study applied label-free visible resonance Raman spectroscopy to enhance the precision of tumor boundary identification during glioma surgeries, with remarkable sensitivity, specificity, and accuracy rates reaching 100%, 96.3%, and 99.6%, respectively (185). Looking ahead, the integration of label-free Raman spectroscopy into surgical practice holds significant promise for improving cancer surgery outcomes. As this technology matures, it is expected to facilitate real-time assessments of tumor margins during surgical procedures. This will aid surgeons in achieving complete tumor resections. The development of portable Raman analytical techniques and advanced algorithms for data analysis will further enhance the utility of *in-situ* applications. This will make label-free Raman spectroscopy an invaluable tool in the future landscape of oncological surgery.

7 Conclusion

Label-free Raman spectroscopy could transform cancer diagnosis and immunotherapy by offering a non-invasive, high-throughput method for detecting molecular signatures in biofluids and tissue specimens. The studies outlined here highlight the myriad of challenges in multifaceted profiling of complex and heterogeneous tumors that can be addressed with technical innovations in Raman spectroscopy to transcend traditional single-omic strategies. The analytical advancements in Raman technologies, encompassing enhanced spectral isolation and refined data processing capabilities, establish it as a crucial instrument for elucidating the intricate mechanisms by which tumors circumvent immune detection—a critical stride towards precision medicine. Coupled with machine learning for real-time data analysis, these techniques position Raman technology as a disruptive tool throughout the continuum of oncological intervention.As techniques for suppressing background signals improve and as the construction of global spectral libraries advances, the accuracy and efficiency of Raman spectroscopy in clinical settings will be enhanced. The potential integration of Raman spectroscopy with existing multi-omics platforms could harmonize diverse datasets, facilitating a more comprehensive characterization of tumors and better predictive biomarker identification. Moreover, the potential for real-time tumor boundary identification during surgeries positions Raman technology at the forefront of precision immunotherapeutics. The ongoing development of portable systems and sophisticated data analysis algorithms promises to further embed label-free Raman spectroscopy within surgical practice, ultimately improving patient outcomes through more precise and informed interventions. By

enabling timely, personalized, and precise immunotherapy strategies, this technology could ultimately transform the landscape of oncological care, reducing reliance on a "one size fits all" treatment paradigm and enhancing patient outcomes.

Author contributions

JC: Writing – original draft. KC: Writing – original draft, Writing – review & editing. SS: Writing – original draft, Writing – review & editing. JH: Writing – original draft, Writing – review & editing. JL: Writing – original draft, Writing – review & editing. JD: Writing – original draft, Writing – review & editing. AK: Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

1. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell*. (2010) 140:883–99. doi: 10.1016/j.cell.2010.01.025

2. Beatty GL, Gladney WL. Immune escape mechanisms as a guide for cancer immunotherapy. *Clin Cancer Res*. (2015) 21:687–92. doi: 10.1158/1078-0432.CCR-14-1860

3. Sánchez-Magraner L, Miles J, Baker CL, Applebee CJ, Lee D-J, Elsheikh S, et al. High PD-1/PD-L1 checkpoint interaction infers tumor selection and therapeutic sensitivity to anti-PD-1/PD-L1 treatment. *Cancer Res*. (2020) 80:4244–57. doi: 10.1158/0008-5472.CAN-20-1117

4. Labani-Motlagh A, Ashja-Mahdavi M, Loskog A. The tumor microenvironment: A milieu hindering and obstructing antitumor immune responses. *Front Immunol*. (2020) 11:940. doi: 10.3389/fimmu.2020.00940

5. Yang C, Robbins PD. The roles of tumor-derived exosomes in cancer pathogenesis. *Clin Dev Immunol*. (2011) 2011:842849. doi: 10.1155/2011/842849

6. Iglesias-Escudero M, Arias-González N, Martínez-Cáceres E. Regulatory cells and the effect of cancer immunotherapy. *Mol Cancer*. (2023) 22:26. doi: 10.1186/s12943-023-01714-0

7. Yang J, Xu J, Wang W, Zhang B, Yu X, Shi S. Epigenetic regulation in the tumor microenvironment: molecular mechanisms and therapeutic targets. *Signal Transduct Target Ther*. (2023) 8:210. doi: 10.1038/s41392-023-01480-x

8. Martins F, Sofiya L, Sykiotis GP, Lamine F, Maillard M, Fraga M, et al. Adverse effects of immune-checkpoint inhibitors: epidemiology, management and surveillance. *Nat Rev Clin Oncol*. (2019) 16:563–80. doi: 10.1038/s41571-019-0218-0

9. Das S, Johnson DB. Immune-related adverse events and anti-tumor efficacy of immune checkpoint inhibitors. *J Immunother Cancer*. (2019) 7:306. doi: 10.1186/s40425-019-0805-8

10. Eggermont AMM, Crittenden M, Wargo J. Combination immunotherapy development in melanoma. *Am Soc Clin Oncol Educ Book*. (2018) 38:197–207. doi: 10.1200/EDBK_201131

11. Aamdal E, Jacobsen KD, Straume O, Kersten C, Herlofsen O, Karlsen J, et al. Ipilimumab in a real-world population: A prospective Phase IV trial with long-term follow-up. *Int J Cancer*. (2022) 150:100–11. doi: 10.1002/ijc.v150.1

12. Royal RE, Levy C, Turner K, Mathur A, Hughes M, Kammula US, et al. Phase 2 trial of single agent Ipilimumab (anti-CTLA-4) for locally advanced or metastatic pancreatic adenocarcinoma. *J Immunother*. (2010) 33:828–33. doi: 10.1097/CJI.0b013e3181eec14c

13. Krzyszczyk P, Acevedo A, Davidoff EJ, Timmins LM, Marrero-Berrios I, Patel M, et al. The growing role of precision and personalized medicine for cancer treatment. *Technol (Singap World Sci)*. (2018) 6:79–100. doi: 10.1142/S2339547818300020

14. Wang J, Ma Y, Lin H, Wang J, Cao B. Predictive biomarkers for immune-related adverse events in cancer patients treated with immune-checkpoint inhibitors. *BMC Immunol*. (2024) 25:8. doi: 10.1186/s12865-024-00599-y

15. Kovács SA, Fekete JT, Győrffy B. Predictive biomarkers of immunotherapy response with pharmacological applications in solid tumors. *Acta Pharmacol Sin*. (2023) 44:1879–89. doi: 10.1038/s41401-023-01079-6

16. Kong J, Ha D, Lee J, Kim I, Park M, Im S-H, et al. Network-based machine learning approach to predict immunotherapy response in cancer patients. *Nat Commun*. (2022) 13:3703. doi: 10.1038/s41467-022-31535-6

17. Chakravarty D, Solit DB. Clinical cancer genomic profiling. *Nat Rev Genet*. (2021) 22:483–501. doi: 10.1038/s41576-021-00338-8

18. Kovács SA, Győrffy B. Transcriptomic datasets of cancer patients treated with immune-checkpoint inhibitors: a systematic review. *J Transl Med*. (2022) 20:249. doi: 10.1186/s12967-022-03409-4

19. Thakur SS, Gowda H, Mani DR, Chatterjee B. Editorial: proteomics and its applications in cancer. *Front Oncol*. (2021) 11:772811. doi: 10.3389/fonc.2021.772811

20. Baysoy A, Bai Z, Satija R, Fan R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol*. (2023) 24:695–713. doi: 10.1038/s41580-023-00615-w

21. Menyhárt O, Győrffy B. Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. *Comput Struct Biotechnol J*. (2021) 19:949–60. doi: 10.1016/j.csbj.2021.01.009

22. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br J Cancer*. (2021) 124:686–96. doi: 10.1038/s41416-020-01122-x

23. Zhang S, Deshpande A, Verma BK, Wang H, Mi H, Yuan L, et al. Integration of clinical trial spatial multiomics analysis and virtual clinical trials enables immunotherapy response prediction and biomarker discovery. *Cancer Res*. (2024) 84:2734–48. doi: 10.1158/0008-5472.CAN-24-0943

24. Li Y, Wu X, Fang D, Luo Y. Informing immunotherapy with multi-omics driven machine learning. *NPJ Digit Med*. (2024) 7:67. doi: 10.1038/s41746-024-01043-6

25. Zeng Z, Li Y, Li Y, Luo Y. Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome Biol*. (2022) 23:83. doi: 10.1186/s13059-022-02653-7

26. Ma A, McDermaid A, Xu J, Chang Y, Ma Q. Integrative methods and practical challenges for single-cell multi-omics. *Trends Biotechnol*. (2020) 38:1007–22. doi: 10.1016/j.tibtech.2020.02.013

27. Tarazona S, Arzalluz-Luque A, Conesa A. Undisclosed, unmet and neglected challenges in multi-omics studies. *Nat Comput Sci*. (2021) 1:395–402. doi: 10.1038/s43588-021-00086-z

28. Chaudhary N, Nguyen TNQ, Cullen D, Meade AD, Wynne C. Discrimination of immune cell activation using Raman micro-spectroscopy in an *in-vitro* & ex-vivo model. *Spectrochim Acta A Mol Biomol Spectrosc*. (2021) 248:119118. doi: 10.1016/j.saa.2020.119118

29. Pavillon N, Smith NI. Non – invasive monitoring of T cell differentiation through Raman spectroscopy. *Sci Rep*. (2023) 13:3129. doi: 10.1038/s41598-023-29259-8

30. Borek-dorosz A, Maria A, Leszczenko P, Adamczyk A, Sowinski K, Gruszecki WI, et al. Raman-based spectrophenotyping of the most important cells of the immune system. *J Adv Res*. (2022) 41:191–203. doi: 10.1016/j.jare.2021.12.013

31. Tratwal J, Falgayrac G, During A, Bertheaume N, Bataclan C, Tavakol DN, et al. Raman microspectroscopy reveals unsaturation heterogeneity at the lipid droplet level and validates an *in vitro* model of bone marrow adipocyte subtypes. *Front Endocrinol (Lausanne)*. (2022) 13:1001210. doi: 10.3389/fendo.2022.1001210

32. Chen Y, Chen Z, Su Y, Lin D, Chen M, Feng S, et al. Metabolic characteristics revealing cell differentiation of nasopharyngeal carcinoma by combining NMR spectroscopy with Raman spectroscopy. *Cancer Cell Int*. (2019) 19:37. doi: 10.1186/s12935-019-0759-4

33. Greig JC, Tipping WJ, Graham D, Faulds K, Gould GW. New insights into lipid and fatty acid metabolism from Raman spectroscopy. *Analyst*. (2024) 149:4789–810. doi: 10.1039/D4AN00846D

34. Paidi SK, Rodriguez Troncoso J, Raj P, Monterroso Diaz P, Ivers JD, Lee DE, et al. Raman spectroscopy and machine learning reveals early tumor microenvironmental changes induced by immunotherapy. *Cancer Res*. (2021) 81:5745–55. doi: 10.1158/0008-5472.CAN-21-1438

35. Zhang Y, Chang K, Ogunlade B, Herndon L, Tadesse LF, Kirane AR, et al. From genotype to phenotype: raman spectroscopy and machine learning for label-free single-cell analysis. *ACS Nano*. (2024) 18:18101–17. doi: 10.1021/acsnano.4c04282

36. Miao Y, Wu L, Qiang J, Qi J, Li Y, Li R, et al. The application of Raman spectroscopy for the diagnosis and monitoring of lung tumors. *Front Bioeng Biotechnol*. (2024) 12:1–21. doi: 10.3389/fbioe.2024.1385552

37. Xu J, Morten KJ. Raman micro-spectroscopy as a tool to study immunometabolism. *Biochem Soc Trans*. (2024) 52:733–45. doi: 10.1042/BST20230794

38. Camp CHJr, Lee YJ, Heddleston JM, Hartshorn CM, Walker ARH, Rich JN, et al. High-speed coherent Raman fingerprint imaging of biological tissues. *Nat Photonics*. (2014) 8:627–34. doi: 10.1038/nphoton.2014.145

39. Pezzotti G. Raman spectroscopy in cell biology and microbiology. *J Raman Spectrosc*. (2021) 52:2348–443. doi: 10.1002/jrs.v52.12

40. Köhler M, Machill S, Salzer R, Krafft C. Characterization of lipid extracts from brain tissue and tumors using Raman spectroscopy and mass spectrometry. *Anal Bioanal Chem*. (2009) 393:1513–20. doi: 10.1007/s00216-008-2592-9

41. Gualerzi A, Niada S, Giannasi C, Picciolini S, Morasso C, Vanna R, et al. Raman spectroscopy uncovers biochemical tissue-related features of extracellular vesicles from mesenchymal stromal cells. *Sci Rep*. (2017) 7:9820. doi: 10.1038/s41598-017-10448-1

42. Contorno S, Darienzo RE, Tannenbaum R. Evaluation of aromatic amino acids as potential biomarkers in breast cancer by Raman spectroscopy analysis. *Sci Rep*. (2021) 11:1698. doi: 10.1038/s41598-021-81296-3

43. Bruzas I, Lum W, Gorunmez Z, Sagle L. Advances in surface-enhanced Raman spectroscopy (SERS) substrates for lipid and protein characterization: sensing and beyond. *Analyst*. (2018) 143:3990–4008. doi: 10.1039/C8AN00606G

44. Lu Y, Lin L, Ye J. Human metabolite detection by surface-enhanced Raman spectroscopy. *Mater Today Bio*. (2022) 13:100205. doi: 10.1016/j.mtbio.2022.100205

45. Cutshaw G, Uthaman S, Hassan N, Kothadiya S, Wen X, Bardhan R. The emerging role of raman spectroscopy as an omics approach for metabolic profiling and biomarker detection towards precision medicine. *Chem Rev*. 123:8297–346. doi: 10.1021/acs.chemrev.2c0089

46. Hu D, Xu X, Zhao Z, Li C, Tian Y, Liu Q, et al. Detecting urine metabolites of bladder cancer by surface-enhanced Raman spectroscopy. *Spectrochim Acta A Mol Biomol Spectrosc*. (2021) 247:119108. doi: 10.1016/j.saa.2020.119108

47. Wu L, Dias A, Diéguez L. Surface enhanced Raman spectroscopy for tumor nucleic acid: Towards cancer diagnosis and precision medicine. *Biosens Bioelectron*. (2022) 204:114075. doi: 10.1016/j.bios.2022.114075

48. Li D, Xia L, Zhou Q, Wang L, Chen D, Gao X, et al. Label-free detection of miRNA using surface-enhanced raman spectroscopy. *Anal Chem*. (2020) 92:12769–73. doi: 10.1021/acs.analchem.0c03335

49. Mo W, Ke Q, Yang Q, Zhou M, Xie G, Qi D, et al. A dual-modal, label-free Raman imaging method for rapid virtual staining of large-area breast cancer tissue sections. *Anal Chem*. (2024) 96:13410–20. doi: 10.1021/acs.analchem.4c00870

50. Brozek-Pluska B, Miazek K, Musiał J, Kordek R. Label-free diagnostics and cancer surgery Raman spectra guidance for the human colon at different excitation wavelengths. *RSC Adv*. (2019) 9:40445–54. doi: 10.1039/C9RA06831G

51. Du Y, Hu L, Wu G, Tang Y, Cai X, Yin L. Diagnoses in multiple types of cancer based on serum Raman spectroscopy combined with a convolutional neural network: Gastric cancer, colon cancer, rectal cancer, lung cancer. *Spectrochim Acta A Mol Biomol Spectrosc*. (2023) 298:122743. doi: 10.1016/j.saa.2023.122743

52. Conti F, D'Acunto M, Caudai C, Colantonio S, Gaeta R, Moroni D, et al. Raman spectroscopy and topological machine learning for cancer grading. *Sci Rep*. (2023) 13:7282. doi: 10.1038/s41598-023-34457-5

53. Gniadecka M, Philipsen PA, Wessel S, Gniadecki R, Wulf HC, Sigurdsson S, et al. Melanoma diagnosis by Raman spectroscopy and neural networks: Structure alterations in proteins and lipids in intact cancer tissue. *J Invest Dermatol*. (2004) 122:443–9. doi: 10.1046/j.0022-202X.2004.22208.x

54. Jones RR, Hooper DC, Zhang L, Wolverson D, Valev VK. Raman techniques: fundamentals and frontiers. *Nanoscale Res Lett*. (2019) 14:1–34. doi: 10.1186/s11671-019-3039-2

55. Langer J, de Aberasturi DJ, Aizpurua J, Alvarez-Puebla RA, Auguié B, Baumberg JJ, et al. Present and future of surface-enhanced raman scattering. *ACS Nano*. (2020) 14:28–117. doi: 10.1021/acsnano.9b04224

56. Shvalya V, Filipič G, Zavašnik J, Abdulhalim I, Cvelbar U. Surface-enhanced Raman spectroscopy for chemical and biological sensing using nanoplasmonics: The relevance of interparticle spacing and surface morphology. *Appl Phys Rev*. (2020) 7:031307. doi: 10.1063/5.0015246

57. Zhao X, Liu X, Chen D, Shi G, Li G, Tang X, et al. Plasmonic trimers designed as SERS-active chemical traps for subtyping of lung tumors. *Nat Commun*. (2024) 15:1–11. doi: 10.1038/s41467-024-50321-0

58. Černigoj J, Silvestri F, Stoevelaar LP, Berzinš J, Gerini G. Lattice resonances and local field enhancement in array of dielectric dimers for surface enhanced raman spectroscopy. *Sci Rep*. (2018) 8:1–7. doi: 10.1038/s41598-018-33941-7

59. Caldarola M, Albella P, Cortés E, Rahmani M, Roschuk T, Grinblat G, et al. Non-plasmonic nanoantennas for surface enhanced spectroscopies with ultra-low heat conversion. *Nat Commun*. (2015) 6:1–8. doi: 10.1038/ncomms8915

60. Dolia V, Balch HB, Dagli S, Abdollahramezani S, Carr Delgado H, Moradifar P, et al. Very-large-scale-integrated high quality factor nanoantenna pixels. *Nat Nanotechnol*. (2024) 19:1290–8. doi: 10.1038/s41565-024-01697-z

61. Chen X, Zhang Y, Cai G, Zhuo J, Lai K, Ye L. All-dielectric metasurfaces with high Q-factor Fano resonances enabling multi-scenario sensing. *Nanophotonics*. (2022) 11:4537–49. doi: 10.1515/nanoph-2022-0394

62. Huang L, Jin R, Zhou C, Li G, Xu L, Overvig A, et al. Ultrahigh-Q guided mode resonances in an All-dielectric metasurface. *Nat Commun*. (2023) 14:1–9. doi: 10.1038/s41467-023-39227-5

63. Chen W, Li Y, Liu Y, Gao Y, Yan Y, Dong Z, et al. All-dielectric SERS metasurface with strong coupling quasi-BIC energized by transformer-based deep learning. *Advanced Optical Materials*. (2024) 12:2301697. doi: 10.1002/adom.202301697

64. Xu Y, Wu S, Hong S, Wang X, Liu X, Wang J, et al. Raman enhancement via double optical resonances in all-dielectric photonic crystal slabs. *Appl Phys Lett*. (2024) 125:081704. doi: 10.1063/5.0226837

65. Cambiasso J, König M, Cortés E, Schlücker S, Maier SA. Surface-enhanced spectroscopies of a molecular monolayer in an all-dielectric nanoantenna. *ACS Photonics*. (2018) 5:1546–57. doi: 10.1021/acsphotonics.7b01604

66. Romano S, Zito G, Managò S, Calafiore G, Penzo E, Cabrini S, et al. Surface-enhanced raman and fluorescence spectroscopy with an all-dielectric metasurface. *J Phys Chem C*. (2018) 122:19738–45. doi: 10.1021/acs.jpcc.8b03190

67. Li N, Xu Z, Dong Y, Hu T, Zhong Q, Fu YH, et al. Large-area metasurface on CMOS-compatible fabrication platform: driving flat optics from lab to fab. *Nanophotonics*. (2020) 9:3071–87. doi: 10.1515/nanoph-2020-0063

68. Barkey M, Büchner R, Wester A, Pritzl SD, Makarenko M, Wang Q, et al. Pixelated high-Q metasurfaces for in situ biospectroscopy and artificial intelligence-enabled classification of lipid membrane photoswitching dynamics. *ACS Nano*. (2024) 18:11644–54. doi: 10.1021/acsnano.3c09798

69. Zhang L, Li C, Peng D, Yi X, He S, Liu F, et al. Raman spectroscopy and machine learning for the classification of breast cancers. *Spectrochim Acta A Mol Biomol Spectrosc*. (2022) 264:120300. doi: 10.1016/j.saa.2021.120300

70. Zhang W, Giang CM, Cai Q, Badie B, Sheng J, Li C. Using random forest for brain tissue identification by Raman spectroscopy. *Mach Learn Sci Technol*. (2023) 4:045053. doi: 10.1088/2632-2153/ad1349

71. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med*. (2018) 24:541–50. doi: 10.1038/s41591-018-0014-x

72. Wilkinson RW, Leishman AJ. Further advances in cancer immunotherapy: going beyond checkpoint blockade. *Front Immunol*. (2018) 9:1082. doi: 10.3389/fimmu.2018.01082

73. Chen M, McReynolds N, Campbell EC, Mazilu M, Barbosa J, Dholakia K, et al. The use of wavelength modulated raman spectroscopy in label-free identification of T lymphocyte subsets, natural killer cells and dendritic cells. *PloS One*. (2015) 10:1–14. doi: 10.1371/journal.pone.0125158

74. Hobro AJ, Kumagai Y, Akira S, Smith NI. Raman spectroscopy as a tool for label-free lymphocyte cell line discrimination. *Analyst*. (2016) 141:3756–64. doi: 10.1039/C6AN00181E

75. Ichimura T, Chiu L, Fujita K, Machiyama H, Yamaguchi T, Watanabe TM, et al. Non-label immune cell state prediction using Raman spectroscopy. *Sci Rep*. (2016) 6:37562. doi: 10.1038/srep37562

76. El-Sayes N, Vito A, Mossman K. Tumor heterogeneity: A great barrier in the age of cancer immunotherapy. *Cancers (Basel)*. (2021) 13:1–14. doi: 10.3390/cancers13040806

77. Naumann M, Arend N, Guliev RR, Kretzer C, Rubio I, Werz O, et al. Label-free characterization of macrophage polarization using raman spectroscopy. *Int J Mol Sci*. (2023) 24:1–10. doi: 10.3390/ijms24010824

78. Lu C-E, Levey RE, Ghersi G, Schueller N, Liebscher S, Layland SL, et al. Materials Today Bio Monitoring the macrophage response towards biomaterial implants using label-free imaging. *Mater Today Bio*. (2023) 21:100696. doi: 10.1016/j.mtbio.2023.100696

79. Pavillon N, Lim EL, Tanaka A, Hori S, Sakaguchi S, Smith NI. Non – invasive detection of regulatory T cells with Raman spectroscopy. *Sci Rep*. (2024) 14:14025. doi: 10.1038/s41598-024-64536-0

80. Ishwar D, Haldavnekar R. Minimally invasive detection of cancer using metabolic changes in tumor-associated nat- ural killer cells with Oncoimmune probes. *Nat Commun*. (2022) 13:4527. doi: 10.1038/s41467-022-32308-x

81. Zhou Q-Q, Guo J, Wang Z, Li J, Chen M, Xu Q, et al. Rapid visualization of PD-L1 expression level in glioblastoma immune microenvironment via machine learning cascade-based Raman histopathology. *J Adv Res*. (2024) 65:257–71. doi: 10.1016/j.jare.2023.12.002

82. Bugide S, Janostiak R, Wajapeyee N. Epigenetic mechanisms dictating eradication of cancer by natural killer cells. *Trends Cancer*. (2018) 4:553–66. doi: 10.1016/j.trecan.2018.06.004

83. Ames E, Canter RJ, Grossenbacher SK, Mac S, Chen M, Smith RC, et al. NK cells preferentially target tumor cells with a cancer stem cell phenotype. *J Immunol*. (2015) 195:4010–9. doi: 10.4049/jimmunol.1500447

84. Chen MY, Zhang F, Goedegebuure SP, Gillanders WE. Dendritic cell subsets and implications for cancer immunotherapy. *Front Immunol*. (2024) 15:1393451. doi: 10.3389/fimmu.2024.1393451

85. Zoladek AB, Johal K, Garcia-nieto S, Pascut F, Shakesheff KM. Label-free molecular imaging of immunological synapses between dendritic and T cells by Raman micro-spectroscopy †. (2010). *RSC. Analyst*. pp. 3205–12. Available at: https://pubs.rsc.org/en/content/articlelanding/2010/an/c0an00508h.

86. Yang D, Liu J, Qian H, Zhuang Q. Cancer-associated fibroblasts: from basic science to anticancer therapy. *Exp Mol Med*. (2023) 55:1322–32. doi: 10.1038/s12276-023-01013-0

87. Gong J, Lin Y, Zhang H, Liu C, Cheng Z, Yang X, et al. Reprogramming of lipid metabolism in cancer-associated fibroblasts potentiates migration of colorectal cancer cells. *Cell Death Dis*. (2020) 11:267. doi: 10.1038/s41419-020-2434-z

88. Yeu TH, Omar IS, Sani SFA, Pathmanathan D, Goh BT, Ravindran N, et al. Distinct lipid phenotype of cancer-associated fibroblasts (CAFs) isolated from overweight/obese endometrial cancer patients as assessed using raman spectroscopy. *Appl Spectrosc*. (2023) 77:723–33. doi: 10.1177/00037028231182721

89. Duan Z, Luo Y. Targeting macrophages in cancer immunotherapy. *Signal Transduct Target Ther*. (2021) 6:127. doi: 10.1038/s41392-021-00506-6

90. Zhang CC. A perspective on LILRBs and LAIR1 as immune checkpoint targets for cancer treatment. *Biochem Biophys Res Commun*. (2022) 633:64–7. doi: 10.1016/j.bbrc.2022.09.019

91. Tabachnick-Cherny S, Pulliam T, Rodriguez HJ, Fan X, Hippe DS, Jones DC, et al. Characterization of immunosuppressive myeloid cells in merkel cell carcinoma: correlation with resistance to PD-1 pathway blockade. *Clin Cancer Res*. (2024) 30:1189–99. doi: 10.1158/1078-0432.CCR-23-1957

92. Li J, Smalley I, Chen Z, Wu J-Y, Phadke MS, Teer JK, et al. Single-cell characterization of the cellular landscape of acral melanoma identifies novel targets for immunotherapy. *Clin Cancer Res*. (2022) 28:2131–46. doi: 10.1158/1078-0432.CCR-21-3145

93. Roesch A. Tumor heterogeneity and plasticity as elusive drivers for resistance to MAPK pathway inhibition in melanoma. *Oncogene*. (2015) 34:2951–7. doi: 10.1038/onc.2014.249

94. Janiszewska M, Stein S, Metzger Filho O, Eng J, Kingston NL, Harper NW, et al. The impact of tumor epithelial and microenvironmental heterogeneity on treatment responses in HER2+ breast cancer. *JCI Insight*. (2021) 6:9–12. doi: 10.1172/jci.insight.147617

95. Madore J, Strbenac D, Vilain R, Menzies AM, Yang JYH, Thompson JF, et al. PD-L1 negative status is associated with lower mutation burden, differential expression of immune-related genes, and worse survival in stage III melanoma. *Clin Cancer Res*. (2016) 22:3915–23. doi: 10.1158/1078-0432.CCR-15-1714

96. Xu Y, Wan B, Chen X, Zhan P, Zhao Y, Zhang T, et al. The association of PD-L1 expression with the efficacy of anti-PD-1/PD-L1 immunotherapy and survival of non-small cell lung cancer patients: a meta-analysis of randomized controlled trials. *Transl Lung Cancer Res*. (2019) 8:413–28. doi: 10.21037/tlcr.2019.08.09

97. Incorvaia L, Fanale D, Badalamenti G, Porta C, Olive D, De Luca I, et al. Baseline plasma levels of soluble PD-1, PD-L1, and BTN3A1 predict response to nivolumab treatment in patients with metastatic renal cell carcinoma: a step toward a biomarker for therapeutic decisions. *Oncoimmunology*. (2020) 9:1832348. doi: 10.1080/2162402X.2020.1832348

98. Terry S, Dalban C, Rioux-Leclercq N, Adam J, Meylan M, Buart S, et al. Association of AXL and PD-L1 expression with clinical outcomes in patients with advanced renal cell carcinoma treated with PD-1 blockade. *Clin Cancer Res*. (2021) 27:6749–60. doi: 10.1158/1078-0432.CCR-21-0972

99. Lee H-H, Wang Y-N, Xia W, Chen C-H, Rau K-M, Ye L, et al. Removal of N-linked glycosylation enhances PD-L1 detection and predicts anti-PD-1/PD-L1 therapeutic efficacy. *Cancer Cell*. (2019) 36:168–178.e4. doi: 10.1016/j.ccell.2019.06.008

100. Gibney GT, Weiner LM, Atkins MB. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet Oncol*. (2016) 17:e542–51. doi: 10.1016/S1470-2045(16)30406-5

101. Topalian SL, Taube JM, Anders RA, Pardoll DM. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat Rev Cancer*. (2016) 16:275–87. doi: 10.1038/nrc.2016.36

102. Larijani B, Miles J, Ward SG, Parker PJ. Quantification of biomarker functionality predicts patient outcomes. *Br J Cancer*. (2021) 124:1618–20. doi: 10.1038/s41416-021-01291-3

103. Weide B, Martens A, Hassel JC, Berking C, Postow MA, Bisschop K, et al. Baseline biomarkers for outcome of melanoma patients treated with pembrolizumab. *Clin Cancer Res*. (2016) 22:5487–96. doi: 10.1158/1078-0432.CCR-16-0127

104. van de Donk PP, Kist de Ruijter L, Lub-de Hooge MN, Brouwers AH, van der Wekken AJ, Oosting SF, et al. Molecular imaging biomarkers for immune checkpoint inhibitor therapy. *Theranostics*. (2020) 10:1708–18. doi: 10.7150/thno.38339

105. Shi R, Tang Y-Q, Miao H. Metabolism in tumor microenvironment: Implications for cancer immunotherapy. *MedComm*. (2020) 1:47–68. doi: 10.1002/mco2.v1.1

106. Kim S-W, Roh J, Park C-S. Immunohistochemistry for pathologists: Protocols, pitfalls, and tips. *J Pathol Transl Med*. (2016) 50:411–8. doi: 10.4132/jptm.2016.08.08

107. Jin P, Li J, Meng Y, Wu L, Bai M, Yu J, et al. PET/CT metabolic patterns in systemic immune activation: A new perspective on the assessment of immunotherapy response and efficacy. *Cancer Lett*. (2021) 520:91–9. doi: 10.1016/j.canlet.2021.06.028

108. Radaram B, Glazer SE, Yang P, Li C-W, Hung M-C, Gammon ST, et al. Evaluation of 89Zr-labeled anti-PD-L1 monoclonal antibodies using DFO and novel HOPO analogues as chelating agents for immuno-PET. *ACS Omega*. (2023) 8:17181–94. doi: 10.1021/acsomega.3c01547

109. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging*. (2015) 42:328–54. doi: 10.1007/s00259-014-2961-x

110. LaLone V, Smith D, Diaz-Espinosa J, Rosania GR. Quantitative Raman chemical imaging of intracellular drug-membrane aggregates and small molecule drug precipitates in cytoplasmic organelles. *Adv Drug Delivery Rev*. (2023) 202:115107. doi: 10.1016/j.addr.2023.115107

111. Paidi SK, Troncoso JR, Harper MG, Liu Z, Nguyen KG, Ravindranathan S, et al. Raman spectroscopy reveals phenotype switches in breast cancer metastasis. *Theranostics*. (2022) 12:5351–63. doi: 10.7150/thno.74002

112. Zhang X, Yu F, Li J, Song D, Li H, Wang K, et al. Investigation on the cancer invasion and metastasis of skin squamous cell carcinoma by Raman spectroscopy. *Molecules*. (2019) 24:2059. doi: 10.3390/molecules24112059

113. Tongdee M, Yamanishi C, Maeda M, Kojima T, Dishinger J, Chantiwas R, et al. One-incubation one-hour multiplex ELISA enabled by aqueous two-phase systems. *Analyst*. (2020) 145:3517–27. doi: 10.1039/D0AN00383B

114. Luo Y, Pehrsson M, Langholm L, Karsdal M, Bay-Jensen A-C, Sun S. Lot-to-lot variance in immunoassays-causes, consequences, and solutions. *Diagnostics (Basel)*. (2023) 13:1–14. doi: 10.3390/diagnostics13111835

115. Altangerel N, Neuman BW, Hemmer PR, Yakovlev VV, Rajil N, Yi Z, et al. Label-free drug interaction screening via Raman microscopy. *Proc Natl Acad Sci U.S.A*. (2023) 120:e2218826120. doi: 10.1073/pnas.2218826120

116. Kobayashi-Kirschvink KJ, Comiter CS, Gaddam S, Joren T, Grody EI, Ounadjela JR, et al. Prediction of single-cell RNA expression profiles in live cells by Raman microscopy with Raman2RNA. *Nat Biotechnol*. (2024) 42:1726–34. doi: 10.1038/s41587-023-02082-2

117. Hulspas R, Dombkowski D, Preffer F, Douglas D, Kildew-Shah B, Gilbert J. Flow cytometry and the stability of phycoerythrin-tandem dye conjugates. *Cytometry A*. (2009) 75:966–72. doi: 10.1002/cyto.a.v75a:11

118. Bhowmick D, van Diepen F, Pfauth A, Tissier R, Ratliff ML. A gain and dynamic range independent index to quantify spillover spread to aid panel design in flow cytometry. *Sci Rep*. (2021) 11:20553. doi: 10.1038/s41598-021-99831-7

119. Bhowmick D, Bushnell TP. How to measure "spillover spread. *Methods Mol Biol*. (2024) 2779:69–83. doi: 10.1007/978-1-0716-3738-8

120. Jiang L, Tixeira R, Caruso S, Atkin-Smith GK, Baxter AA, Paone S, et al. Monitoring the progression of cell death and the disassembly of dying cells by flow cytometry. *Nat Protoc*. (2016) 11:655–63. doi: 10.1038/nprot.2016.028

121. Hassdenteufel S, Schuldiner M. Show your true color: Mammalian cell surface staining for tracking cellular identity in multiplexing and beyond. *Curr Opin Chem Biol*. (2022) 66:102102. doi: 10.1016/j.cbpa.2021.102102

122. Lassailly F, Griessinger E, Bonnet D. Microenvironmental contaminations" induced by fluorescent lipophilic dyes used for noninvasive *in vitro* and *in vivo* cell tracking. *Blood*. (2010) 115:5347–54. doi: 10.1182/blood-2009-05-224030

123. Eyles JE, Vessillier S, Jones A, Stacey G, Schneider CK, Price J. Cell therapy products: focus on issues with manufacturing and quality control of chimeric antigen receptor T-cell therapies: Key issues for manufacturing and quality control of cell therapies. *J Chem Technol Biotechnol*. (2019) 94:1008–16. doi: 10.1002/jctb.2019.94.issue-4

124. Wu L, Wang Z, Zhang Y, Fei J, Chen H, Zong S, et al. *In situ* probing of cell–cell communications with surface-enhanced Raman scattering (SERS) nanoprobes and microfluidic networks for screening of immunotherapeutic drugs. *Nano Res*. (2017) 10:584–94. doi: 10.1007/s12274-016-1316-2

125. Xiong CC, Zhu SS, Yan DH, Dong Y, Zhe Y, Guo Z, et al. Rapid and precise detection of cancers via label – free SERS and deep learning. *Anal Bioanal Chem*. (2023) 415:3449–62. doi: 10.1007/s00216-023-04730-7

126. Wurm LM, Fischer B, Neuschmelting V, Reinecke D, Fischer I, Croner RS, et al. Rapid, label-free classification of glioblastoma differentiation status combining confocal

Raman spectroscopy and machine learning. *Analyst*. (2023) 148:6109–19. doi: 10.1039/D3AN01303K

127. Wevers D, Ramautar R, Clark C, Hankemeier T, Ali A. Opportunities and challenges for sample preparation and enrichment in mass spectrometry for single-cell metabolomics. *Electrophoresis*. (2023) 44:2000–24. doi: 10.1002/elps.202300105

128. Nakhoda SK, Olszanski AJ. Addressing recent failures in immuno-oncology trials to guide novel immunotherapeutic treatment strategies. *Pharmaceut Med*. (2020) 34:83–91. doi: 10.1007/s40290-020-00326-z

129. Yuan Q, Deng D, Pan C, Ren J, Wei T, Wu Z, et al. Integration of transcriptomics, proteomics, and metabolomics data to reveal HER2-associated metabolic heterogeneity in gastric cancer with response to immunotherapy and neoadjuvant chemotherapy. *Front Immunol*. (2022) 13:951137. doi: 10.3389/fimmu.2022.951137

130. Anagnostou V, Landon BV, Medina JE, Forde P, Velculescu VE. Translating the evolving molecular landscape of tumors to biomarkers of response for cancer immunotherapy. *Sci Transl Med*. (2022) 14:eabo3958. doi: 10.1126/scitranslmed.abo3958

131. Raufaste-Cazavieille V, Santiago R, Droit A. Multi-omics analysis: Paving the path toward achieving precision medicine in cancer treatment and immuno-oncology. *Front Mol Biosci*. (2022) 9:962743. doi: 10.3389/fmolb.2022.962743

132. McAvan BS, Bowsher LA, Powell T, O'Hara JF, Spitali M, Goodacre R, et al. Raman spectroscopy to monitor post-translational modifications and degradation in monoclonal antibody therapeutics. *Anal Chem*. (2020) 92:10381–9. doi: 10.1021/acs.analchem.0c00627

133. Shalabaeva V, Lovato L, La Rocca R, Messina GC, Dipalo M, Miele E, et al. Time resolved and label free monitoring of extracellular metabolites by surface enhanced Raman spectroscopy. *PloS One*. (2017) 12:1–16. doi: 10.1371/journal.pone.0175581

134. Nomizo T, Ozasa H, Tsuji T, Funazo T, Yasuda Y, Yoshida H, et al. Clinical impact of single nucleotide polymorphism in PD-L1 on response to nivolumab for advanced non-small-cell lung cancer patients. *Sci Rep*. (2017) 7:45124. doi: 10.1038/srep45124

135. Findlay JM, Middleton MR, Tomlinson I. A systematic review and meta-analysis of somatic and germline DNA sequence biomarkers of esophageal cancer survival, therapy response and stage. *Ann Oncol*. (2015) 26:624–44. doi: 10.1093/annonc/mdu449

136. Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet*. (2013) 45:478–86. doi: 10.1038/ng.2591

137. Hu Q, Xie Y, Ge Y, Nie X, Tao J, Zhao Y. Resting T cells are hypersensitive to DNA damage due to defective DNA repair pathway. *Cell Death Dis*. (2018) 9:662. doi: 10.1038/s41419-018-0649-z

138. Mogil RJ, Shi Y, Bissonnette RP, Bromley P, Yamaguchi I, Green DR. Role of DNA fragmentation in T cell activation-induced apoptosis. *Vitro vivo. J Immunol*. (1994) 152:1674–83. doi: 10.4049/jimmunol.152.4.1674

139. Bediaga NG, Coughlan HD, Johanson TM, Garnham AL, Naselli G, Schröder J, et al. Multi-level remodelling of chromatin underlying activation of human T cells. *Sci Rep*. (2021) 11:528. doi: 10.1038/s41598-020-80165-9

140. Alderson MR, Tough TW, Davis-Smith T, Braddy S, Falk B, Schooley KA, et al. Fas ligand mediates activation-induced cell death in human T lymphocytes. *J Exp Med*. (1995) 181:71–7. doi: 10.1084/jem.181.1.71

141. Owen DL, Sjaastad LE, Farrar MA. Regulatory T cell development in the thymus. *J Immunol*. (2019) 203:2031–41. doi: 10.4049/jimmunol.1900662

142. Lee YJ, Ahn HJ, Lee G-J, Jung GB, Lee G, Kim D, et al. Investigation of biochemical property changes in activation-induced CD8+ T cell apoptosis using Raman spectroscopy. *J BioMed Opt*. (2015) 20:75001. doi: 10.1117/1.JBO.20.7.075001

143. Uzunbajakava N, Lenferink A, Kraan Y, Willekens B, Vrensen G, Greve J, et al. Nonresonant Raman imaging of protein distribution in single human cells. *Biopolymers*. (2003) 72:1–9. doi: 10.1002/bip.10246

144. Rygula A, Majzner K, Marzec KM, Kaczor A, Pilarczyk M, Baranska M. Raman spectroscopy of proteins: a review: Raman spectroscopy of proteins. *J Raman Spectrosc*. (2013) 44:1061–76. doi: 10.1002/jrs.v44.8

145. Das G, La Rocca R, Lakshmikanth T, Gentile F, Tallerico R, Zambetti LP, et al. Monitoring human leukocyte antigen class I molecules by micro-Raman spectroscopy at single-cell level. *J BioMed Opt*. (2010) 15:027007. doi: 10.1117/1.3368687

146. Hansel TT, Kropshofer H, Singer T, Mitchell JA, George AJT. The safety and side effects of monoclonal antibodies. *Nat Rev Drug Discovery*. (2010) 9:325–38. doi: 10.1038/nrd3003

147. Liu H, May K. Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function: Structural variations, chemical modifications and possible impacts to stability and biological function. *MAbs*. (2012) 4:17–23. doi: 10.4161/mabs.4.1.18347

148. Wang X, An Z, Luo W, Xia N, Zhao Q. Molecular and functional analysis of monoclonal antibodies in support of biologics development. *Protein Cell*. (2018) 9:74–85. doi: 10.1007/s13238-017-0447-x

149. Elgundi Z, Reslan M, Cruz E, Sifniotis V, Kayser V. The state-of-play and future of antibody therapeutics. *Adv Drug Delivery Rev*. (2017) 122:2–19. doi: 10.1016/j.addr.2016.11.004

150. Herrington WFJr, Singh GP, Wu D, Barone PW, Hancock W, Ram RJ. Optical detection of degraded therapeutic proteins. *Sci Rep*. (2018) 8:5089. doi: 10.1038/s41598-018-23409-z

151. Zhang C, Springall JS, Wang X, Barman I. Rapid, quantitative determination of aggregation and particle formation for antibody drug conjugate therapeutics with label-free Raman spectroscopy. *Anal Chim Acta*. (2019) 1081:138–45. doi: 10.1016/j.aca.2019.07.007

152. Kao K-C, Vilbois S, Tsai C-H, Ho P-C. Metabolic communication in the tumour-immune microenvironment. *Nat Cell Biol*. (2022) 24:1574–83. doi: 10.1038/s41556-022-01002-x

153. DeBerardinis RJ, Chandel NS. Fundamentals of cancer metabolism. *Sci Adv*. (2016) 2:e1600200. doi: 10.1126/sciadv.1600200

154. Zhu Y, Li X, Wang L, Hong X, Yang J. Metabolic reprogramming and crosstalk of cancer-related fibroblasts and immune cells in the tumor microenvironment. *Front Endocrinol (Lausanne)*. (2022) 13:988295. doi: 10.3389/fendo.2022.988295

155. Soongsathitanon J, Jamjuntra P, Sumransub N, Yangngam S, de la Fuente M, Landskron G, et al. Crosstalk between tumor-infiltrating immune cells and cancer-associated fibroblasts in tumor growth and immunosuppression of breast cancer. *J Immunol Res*. (2021) 2021:8840066. doi: 10.1155/2021/8840066

156. Liu J, Gao M, Yang Z, Zhao Y, Guo K, Sun B, et al. Macrophages and metabolic reprograming in the tumor microenvironment. *Front Oncol*. (2022) 12:795159. doi: 10.3389/fonc.2022.795159

157. Broadfield LA, Pane AA, Talebi A, Swinnen JV, Fendt S-M. Lipid metabolism in cancer: New perspectives and emerging mechanisms. *Dev Cell*. (2021) 56:1363–93. doi: 10.1016/j.devcel.2021.04.013

158. Roman M, Wrobel TP, Panek A, Paluszkiewicz C, Kwiatek WM. Lipid droplets in prostate cancer cells and effect of irradiation studied by Raman microspectroscopy. *Biochim Biophys Acta Mol Cell Biol Lipids*. (2020) 1865:158753. doi: 10.1016/j.bbalip.2020.158753

159. Kopec M, Beton-Mysur K. The role of glucose and fructose on lipid droplet metabolism in human normal bronchial and cancer lung cells by Raman spectroscopy. *Chem Phys Lipids*. (2024) 259:105375. doi: 10.1016/j.chemphyslip.2023.105375

160. Cruz ALS, Barreto E de A, Fazolini NPB, Viola JPB, Bozza PT. Lipid droplets: platforms with multiple functions in cancer hallmarks. *Cell Death Dis*. (2020) 11:105. doi: 10.1038/s41419-020-2297-3

161. Geng F, Guo D. Lipid droplets, potential biomarker and metabolic target in glioblastoma. *Intern Med Rev (Wash DC)*. (2017) 3:1–4. doi: 10.18103/imr.v3i5.443

162. Guerenne-Del Ben T, Couderc V, Duponchel L, Sol V, Leproux P, Petit JM. Multiplex coherent anti-Stokes Raman scattering microspectroscopy detection of lipid droplets in cancer cells expressing TrkB. *Sci Rep*. (2020) 10:1–12. doi: 10.1038/s41598-020-74021-z

163. Doran CE, Frank CB, McGrath S, Packer RA. Use of handheld raman spectroscopy for intraoperative differentiation of normal brain tissue from intracranial neoplasms in dogs. *Front Vet Sci*. (2022) 8:1–6. doi: 10.3389/fvets.2021.819200

164. Hollon T, Orringer DA. Label – free brain tumor imaging using Raman – based methods. *J Neurooncol*. (2021) 151:393–402. doi: 10.1007/s11060-019-03380-z

165. Li L, Yu M, Li X, Ma X, Zhu L, Zhang T. A deep learning method for multi-task intelligent detection of oral cancer based on optical fiber Raman spectroscopy. *Anal Methods*. (2024) 16:1659–73. doi: 10.1039/D3AY02250A

166. Li X, Li L, Sun Q, Chen B, Zhao C, Dong Y, et al. Rapid multi-task diagnosis of oral cancer leveraging fiber-optic Raman spectroscopy and deep learning algorithms. *Front Oncol*. (2023) 13:1–15. doi: 10.3389/fonc.2023.1272305

167. Duraipandian S, Bergholt MS, Zheng W, Ho KY, Teh M, Yeoh KG, et al. Real-time Raman spectroscopy for *in vivo*, online gastric cancer diagnosis during clinical endoscopic examination. *J BioMed Opt*. (2012) 17:81418. doi: 10.1117/1.JBO.17.8.081418

168. Lui H, Zhao J, McLean D, Zeng H. Real-time raman spectroscopy for *in vivo* skin cancer diagnosis. *Cancer Res*. (2012) 72:2491–500. doi: 10.1158/0008-5472.CAN-11-4061

169. Wang Y, Qian H, Shao X, Zhang H, Liu S, Pan J, et al. Multimodal convolutional neural networks based on the Raman spectra of serum and clinical features for the early diagnosis of prostate cancer. *Spectrochim Acta A Mol Biomol Spectrosc*. (2023) 293:122426. doi: 10.1016/j.saa.2023.122426

170. ClinicalTrials.gov. Available online at: https://clinicaltrials.gov/study/NCT05854940 (Accessed October 23, 2024).

171. Chen X, Li X, Yang H, Xie J, Liu A. Diagnosis and staging of diffuse large B-cell lymphoma using label-free surface-enhanced Raman spectroscopy. *Spectrochim Acta A Mol Biomol Spectrosc*. (2022) 267:120571. doi: 10.1016/j.saa.2021.120571

172. Bai Y, Yu Z, Yi S, Yan Y, Huang Z, Qiu L. Raman spectroscopy-based biomarker screening by studying the fingerprint characteristics of chronic lymphocytic leukemia and diffuse large B-cell lymphoma. *J Pharm BioMed Anal*. (2020) 190:113514. doi: 10.1016/j.jpba.2020.113514

173. Constantinou M, Hadjigeorgiou K, Abalde-Cela S, Andreou C. Label-free sensing with metal nanostructure-based surface-Enhanced Raman Spectroscopy for cancer diagnosis. *ACS Appl Nano Mater*. (2022) 5:12276–99. doi: 10.1021/acsanm.2c02392

174. Ning T, Li H, Chen Y, Zhang B, Zhang F, Wang S. Raman spectroscopy based pathological analysis and discrimination of formalin fixed paraffin embedded breast cancer tissue. *Vib Spectrosc*. (2021) 115:103260. doi: 10.1016/j.vibspec.2021.103260

175. Gaifulina R, Caruana DJ, Oukrif D, Guppy NJ, Culley S, Brown R, et al. Rapid and complete paraffin removal from human tissue sections delivers enhanced Raman spectroscopic and histopathological analysis. *Analyst*. (2020) 145:1499–510. doi: 10.1039/C9AN01030K

176. Tfayli A, Gobinet C, Vrabie V, Huez R, Manfait M, Piot O. Digital dewaxing of Raman signals: discrimination between nevi and melanoma spectra obtained from paraffin-embedded skin biopsies. *Appl Spectrosc*. (2009) 63:564–70. doi: 10.1366/000370209788347048

177. Ibrahim O, Maguire A, Meade AD, Flint S, Toner M, Byrne HJ, et al. Improved protocols for pre-processing Raman spectra of formalin fixed paraffin preserved tissue sections. *Anal Methods*. (2017) 9:4709–17. doi: 10.1039/C6AY03308C

178. Kurouski D, Postiglione T, Deckert-Gaudig T, Deckert V, Lednev IK. Amide I vibrational mode suppression in surface (SERS) and tip (TERS) enhanced Raman spectra of protein specimens. *Analyst*. (2013) 138:1665–73. doi: 10.1039/c2an36478f

179. Lewis AT, Gaifulina R, Guppy NJ, Isabelle M, Dorney J, Lloyd GR, et al. Developing Raman spectroscopy as a diagnostic tool for label-free antigen detection. *J Biophotonics*. (2018) 11:e201700028. doi: 10.1002/jbio.2018.11.issue-2

180. Yakubovskaya E, Zaliznyak T, Martínez Martínez J, Taylor GT. Tear down the fluorescent curtain: A new fluorescence suppression method for raman microspectroscopic analyses. *Sci Rep*. (2019) 9:1–9. doi: 10.1038/s41598-019-52321-3

181. Wyss RM, Kewes G, Marabotti P, Koepfli SM, Schlichting K-P, Parzefall M, et al. Bulk-suppressed and surface-sensitive Raman scattering by transferable plasmonic membranes with irregular slot-shaped nanopores. *Nat Commun*. (2024) 15:1–11. doi: 10.1038/s41467-024-49130-2

182. Mukherjee A, Wackenhut F, Dohare A, Horneber A, Lorenz A, Müchler H, et al. Three-dimensional (3D) surface-enhanced Raman spectroscopy (SERS) substrates: Fabrication and SERS applications. *J Phys Chem C Nanomater Interfaces*. (2023) 127:13689–98. doi: 10.1021/acs.jpcc.3c02410

183. Wen Y, Liu R, Xie Y, Liu X, Li M. SERS surgical navigation with postsurgical immunotherapy of local microtumors and distant metastases for improved anticancer outcomes. *Sci Adv*. (2024) 10:eado2741. doi: 10.1126/sciadv.ado2741

184. Zúñiga WC, Jones V, Anderson SM, Echevarria A, Miller NL, Stashko C, et al. Raman spectroscopy for rapid evaluation of surgical margins during breast cancer lumpectomy. *Sci Rep*. (2019) 9:14639. doi: 10.1038/s41598-019-51112-0

185. Zhou Y, Liu C-H, Wu B, Yu X, Cheng G, Zhu K, et al. Optical biopsy identification and grading of gliomas using label-free visible resonance Raman spectroscopy. *J BioMed Opt*. (2019) 24:1–12. doi: 10.1117/1.JBO.24.9.095001

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 13 January 2025
DOI 10.3389/fimmu.2024.1522076

OPEN ACCESS

EDITED BY
Ying Zhu,
Genentech Inc., United States

REVIEWED BY
Roshni Roy,
Emory University, United States
Xiao Long,
Peking Union Medical College Hospital (CAMS), China

*CORRESPONDENCE
Bin Yang
yangbin1@smu.edu.cn
Ding-Heng Zhu
z127140@sina.com

†These authors have contributed equally to this work

RECEIVED 03 November 2024
ACCEPTED 16 December 2024
PUBLISHED 13 January 2025

CITATION
Deng C-C, Xu X-Y, Zhang Y, Liu L-C, Wang X, Chen J-Y, Yao L-Y, Zhu D-H and Yang B (2025) Single-cell RNA-seq reveals immune cell heterogeneity and increased Th17 cells in human fibrotic skin diseases.
Front. Immunol. 15:1522076.
doi: 10.3389/fimmu.2024.1522076

Single-cell RNA-seq reveals immune cell heterogeneity and increased Th17 cells in human fibrotic skin diseases

Cheng-Cheng Deng†, Xue-Yan Xu†, Yan Zhang†, Long-Can Liu, Xuan Wang, Jun-Yi Chen, Liu-Yi Yao, Ding-Heng Zhu* and Bin Yang*

Dermatology Hospital, Southern Medical University, Guangzhou, China

Background: Fibrotic skin disease represents a major global healthcare burden, characterized by fibroblast hyperproliferation and excessive accumulation of extracellular matrix components. The immune cells are postulated to exert a pivotal role in the development of fibrotic skin disease. Single-cell RNA sequencing has been used to explore the composition and functionality of immune cells present in fibrotic skin diseases. However, these studies detected the gene expression of all cells in fibrotic skin diseases and did not enrich immune cells. Thus, the precise immune cell atlas in fibrotic skin diseases remains unknown. In this study, we plan to investigate the intricate cellular landscape of immune cells in keloid, a paradigm of fibrotic skin diseases.

Methods: $CD45^+$ immune cells were enriched by fluorescence-activated cell sorting. Single-cell RNA sequencing was used to analyze the cellular landscape of immune cells in keloid and normal scar tissues. Ki-67 staining, a scratch experiment, real-time PCR, and Western blotting were used to explore the effect of the Th17 cell supernatant on keloid fibroblasts.

Results: Our findings revealed the intricate cellular landscape of immune cells in fibrotic skin diseases. We found that the percentage of Th17 cells was significantly increased in keloids compared to normal scars. All the subclusters of macrophages and dendritic cells (DCs) showed similar proportions between keloid samples and normal scar samples. However, upregulated genes in keloid M1 macrophages, M2 macrophages, and cDC2 are associated with the MHC class II protein complex assembly and antigen assembly, indicating that macrophages and cDC2 are active in keloids. Functional studies suggested that the supernatant of Th17 cells could promote proliferation, collagen expression, and migration of keloid fibroblasts through interleukin 17A. Importantly, increased Th17 cells are also found in other fibrotic skin diseases, such as hypertrophic scars and scleroderma, suggesting this represents a broad mechanism for skin fibrosis.

Conclusion: In summary, we built a single-cell atlas of fibrotic skin diseases in this study. In addition, we explored the function of Th17 cell-fibroblast interaction in skin fibrosis. These findings will help to understand fibrotic skin disease pathogenesis in depth and identify potential targets for fibrotic skin disease treatment.

KEYWORDS

immune cell, Th17 cell, fibrotic skin diseases, keloid, macrophage, dendritic cell, IL-17

1 Introduction

Fibrosis is a condition that is characterized by fibroblast proliferation and excessive accumulation of extracellular matrix components (1, 2). Fibrosis contributes to a high level of morbidity and mortality worldwide and can lead to progressive tissue scarring and organ dysfunction (1–3). Fibrotic skin diseases are characterized by the accumulation of extracellular matrix components in the dermis and include hypertrophic scars, keloids, scleroderma, and graft-vs.-host diseases (4–8). Studies have indicated a correlation between the development of fibrotic skin diseases and genetic predisposition, tissue tension, aberrant collagen synthesis and degradation processes, inflammatory responses, and immune dysregulation (3, 6, 7). However, the precise underlying pathogenesis of fibrotic skin diseases remains elusive, and radical treatments are still lacking.

The immune response is postulated to exert a pivotal role in the occurrence and progression of fibrotic skin diseases (6–10). A substantial infiltration of immune cells is observed within fibrotic skin diseases, and these cells potentially influence the development of fibrotic skin disease lesions through the release of inflammatory mediators and the modulation of extracellular matrix synthesis. Furthermore, it has been demonstrated that immune cells occupy a pivotal position in regulating the aberrant behavior exhibited by fibroblasts in fibrotic skin diseases (6, 7, 9, 11, 12). Single-cell RNA sequencing (scRNA-seq) has been used to explore the composition and functionality of immune cells present in fibrotic skin diseases, such as in keloids and scleroderma (13–15). However, these studies detected the gene expression of all cells in fibrotic skin diseases and did not enrich immune cells. Most of the cells in these single-cell RNA sequencing studies were keratinocytes, fibroblasts, and vascular endothelial cells, and the proportion of immune cells was low in the results (13–15). We need to enrich immune cells in single-cell RNA sequencing studies of fibrotic skin diseases to get a more precise immune cell atlas for fibrotic skin diseases.

Th17 cells are a lineage of $CD4^+$ T helper cells. Th17 cells have been implicated in numerous inflammatory diseases, including Crohn's disease, psoriasis, multiple sclerosis, rheumatoid arthritis, and inflammatory bowel disease (16–18). The pro-inflammatory cytokines derived from Th17 cells, including interleukin 17A (IL-17A), IL-17F, IL-21, IL-22, and IL-26, play crucial roles in the pathogenesis of these diseases (16–18). It has been observed that the inhibition of Th17 cell differentiation leads to a downregulation of IL-17A expression, subsequently mitigating hepatic fibrosis and pulmonary fibrosis (19, 20). It has been discovered that IL-17A secreted by Th17 cells augments the release of pro-inflammatory chemokines, including monocyte chemoattractant protein (MCP)-1 and IL-8, from dermal fibroblasts in systemic sclerosis. This, in turn, exerts a profound impact on the remodeling of the extracellular matrix (21, 22). These comprehensive investigations have established a link between Th17 cells and fibrotic diseases. However, the specific role and function of Th17 cells in fibrotic skin diseases are not fully understood.

In this study, we isolated $CD45^+$ cells from keloids, a paradigm of fibrotic skin diseases, using fluorescence-activated cell sorting (FACS) and performed single-cell RNA sequencing analysis. Our results revealed the intricate cellular landscape of immune cells in keloids. Compared to normal scar tissue, the percentage of Th17 cells was significantly increased in keloids. Further functional studies revealed that Th17 cells promote the proliferation, collagen expression, and migration of keloid fibroblasts by secreting IL-17A. Importantly, increased Th17 cells were also found in other fibrotic skin diseases, such as hypertrophic scars and scleroderma, suggesting this represents a broad mechanism for skin fibrosis. These findings will help us more thoroughly understand the pathogenesis of fibrotic skin diseases and provide potential targets for therapies for fibrotic skin diseases.

2 Materials and methods

2.1 Sample preparation and tissue dissociation

This study was approved by the Medical and Ethics Committees of Dermatology Hospital, Southern Medical University, and each patient signed an informed consent form before participating in this study. Keloid tissues were harvested during plastic surgery from three patients confirmed to have clinical evidence of keloid (Supplementary Table S1). Normal scar tissue was obtained from

three patients who underwent elective scar resection surgery (Supplementary Table S1). The tissues were washed with PBS on ice and the fat tissue was removed. The tissue samples were then cut into 1 cm^2 pieces in a digestion medium composed of 2.5 mg/ml Dispase II (Roche, USA, 04942078001) in PBS and incubated at 37°C for 2h. After removing the epidermis, the dermal portion was further cut and digested in 2.5 mg/ml collagenase IV (Yeasen Biotechnology, China; 40510ES60) at 37°C for 2h. The cell suspension was filtered through a 70μm-cell strainer, and then the enzymes were neutralized with buffer (PBS with 1% fetal bovine serum). The cells were centrifuged at 2,000 rpm for 10 min at 4°C and resuspended in buffer (PBS with 1% FBS). We then sorted the $CD45^+$ immune cells and constructed scRNA-seq libraries.

2.2 Single-cell cDNA and library preparation

Single-cell cDNA, library preparation, and 3′-end single-cell RNA sequencing were performed by Novogene (Beijing, China). For experiments using the 10×Genomics platform, the Chromium Single Cell 3′ Library and Gel Bead Kit v3.1, Chromium Single Cell 3′ Chip Kit v3.1, and Chromium i7 Multiplex Kit were used according to the manufacturer's instructions in the Chromium Single Cell 3′ Reagents Kits v3.1 User Guide. The single-cell suspension was washed twice with 1×PBS + 0.04% BSA. The cell number and concentration were confirmed using a TC20™ Automated Cell Counter.

Approximately 10,000 cells were immediately subjected to the 10×Genomics Chromium Controller machine for Gel Beads-in-Emulsion (GEM) generation. mRNA was prepared using 10×Genomics Chromium Single Cell 3′ reagent kits (V3 chemistry). During this step, cells were partitioned into the GEMs along with gel beads coated with oligos. These oligos provide poly-dT sequences to capture mRNAs released after cell lysis inside the droplets and cell-specific and transcript-specific barcodes (16 bp 10×barcodes and 10 bp unique molecular identifiers (UMIs), respectively).

After RT-PCR, cDNA was recovered, purified, and amplified to generate sufficient quantities for library preparation. Library quality and concentration were assessed using an Agilent Bioanalyzer 2100. Libraries were run on the Novaseq 6000 for Illumina PE150 sequencing.

2.3 Single-cell RNA-sequence data processing

The 10×Genomics Cell Ranger toolkit (v6.1.2) was used to process 10×Genomics raw data for read alignment and UMI matrix generation. Reads were aligned to the human reference genome (GRCh38) downloaded from the 10×Genomics official website with the STAR algorithm. The aligned reads were quantified as a gene expression matrix based on the number of UMIs detected in individual cells. Filtered gene-cell UMI matrices were generated for further single-cell analysis.

Low-quality cells that expressed fewer than 200 genes or more than 4,000 genes and more than 10% mitochondrial gene expression were eliminated. The R package DoubletFinder (v2.0.3) was applied to filter doublets. After removing low-quality cells and doublets, R package Seurat (v4.0.3) was used for unsupervised clustering of individual cells. First, a global-scaling normalization method, LogNormalize, that normalizes the gene expression measurements for each cell by the total number of UMIs in single cells and multiplied by a scaling factor of 10000 was used. After log-transformation, the top 2,000 highly variable genes were detected and principal component analysis (PCA) was performed for downstream unsupervised clustering analysis. The Louvain algorithm was adopted to cluster individual cells based on the top 30 PCs and the identified clusters were visualized with Uniform Manifold Approximation and Projection (UMAP). The accuracy of single-cell analysis can be affected by batch effects, and the canonical correlation analysis (CCA) method was applied based on the top 30 PCs with the default parameters for batch correction.

2.4 Gene signature scores

To assist in the identification of subpopulations of CD4 and CD8 T cells, we downloaded the sets of gene signatures associated with CD4 and CD8 T cells from the literature (23) and calculated functional signature scores for each cell with the AddModuleScore function in the Seurat package to illustrate the functional properties of each cell type.

To assign M1/M2 polarization estimates to the macrophage cells, we applied the AddModuleScore function in the Seurat package. The gene sets associated with M1 and M2 polarization were obtained from Sun et al. (24).

2.5 Functional enrichment analysis

Differentially expressed genes (DEGs) were identified using the FindMarkers function, implemented in the Seurat package, with the Wilcoxon rank sum test with the following criteria: log-scaled fold change ≥ 0.25 and P value < 0.05. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed using the clusterProfiler (v4.11.0) package based on the upregulated genes and downregulated genes. Pathways with adjusted $P < 0.05$ were considered significant.

2.6 Cell-cell interactions analysis

To investigate the cell-cell interactions between different cell types in the normal scar samples and keloid samples, cellular spatial organization mapper (CSOmap) software (v1.0) was used to identify ligand-receptor pairs. CSOmap was used to construct a three-dimensional (3D) pseudo space and infer the cell-cell interactions based on scRNA-seq data. CSOmap combined the gene expression data of single cells with prior knowledge of signaling and gene regulatory networks. FANTOM5, a human

ligand-receptor interaction database, was used to combine immune-associated chemokines, cytokines, costimulators, coinhibitors, and their receptors to estimate the cell-cell affinity matrix. The contribution of each L-R pair to the cell-cell affinity can provide clues to highlight important LR pairs underlying the cellular interactions.

2.7 Immunofluorescence staining

Human skin biopsies (Supplementary Table S2) were submerged in 4% paraformaldehyde for 24h at room temperature. The samples were dehydrated in gradient alcohol and embedded in paraffin according to standard protocols. Samples were sectioned at 4μm thickness and then incubated at 75℃ for 20min. The sections were deparaffinized with environmental dewaxling dip wax transparentize solution (Bioshap, China, 22181809) and rehydrated in 95% alcohol. The sections were placed with high-pH repair buffer (GeneTech, China; GT102410) in a 95℃ water bath for 20 min with a microwave. After overnight incubation at 4℃ with rabbit anti-IL-17A (Santa Cruz, sc-374218) and mouse anti-CD4 (Abcam, ab183685), sections were washed thrice with PBS and treated with 1:1000 diluted anti-rabbit Alexa Flour 488 (Abcam, ab150113) and anti-mouse Alexa Flour555 (Abcam, ab150110), and conjugated for 1 h at room temperature. After three washes with PBS, counterstaining of cell nuclei was performed using DAPI (Beyotime, China, P0131). Images were taken using a Nikon A1 confocal laser-scanning microscope.

2.8 Th17 cells polarizing

Peripheral blood mononuclear cells (PBMCs) were isolated from keloid patients' whole blood by centrifugation in a density gradient medium (Ficoll-PaqueTM Plus, Cytiva, 17144003). The cells were resuspended at a concentration of 5×10^7 cells/mL in buffer (PBS containing 2% fetal bovine serum and 1 mM EDTA). Naïve $CD4^+$ T cells (purity >99%) were isolated using an EasySep™ Human Naïve $CD4^+$ T Cell Isolation Kit II (Stemcell, 17555). Purified naïve T cells obtained as described above were cultured in a Th17-polarizing medium for 7 days to induce Th17 cells. The Th17-polarizing medium contained anti-CD3 Ab (2μg/mL, OKT-3; BioLegend), anti-CD28 Ab (1μg/mL, OKT-3, BioLegend), IL2 (10ng/ml,PeproTech), IL6 (20ng/mL, R&D Systems), TGF-β1 (10ng/mL, R&D Systems), IL1β (10ng/mL, PeproTech), IL23 (10ng/mL, PeproTech), anti-IL4 Ab (10μg/mL, BioLegend), and anti-IFN-γ Ab (10μg/mL, BioLegend).

2.9 Real-time quantitative PCR

RNA extraction from cells was performed using TRIzol Reagent (Invitrogen, Life Technologies, USA) according to the manufacturer's instructions. 1μg of RNA fraction was reverse transcribed to cDNA using PrimeScript™ RT Master Mix (Takara, Dalian, China). qRT-PCR was conducted using a BIO-RAD CFX Connect Real-time PCR Detection System and primers and templates mixed with SYBR Premix Ex Taq II (Vazyme, Nanjing, China). Threshold cycle (CT) values were used to calculate the fold change using the $2^{-\Delta\Delta CT}$ method. The relative mRNA expression was normalized to the *GAPDH* gene. Gene-specific primer pairs were designed with Primer Premier 5.0 software (Supplementary Table S3).

2.10 Western blot

The cells were washed once with ice-cold PBS and lysed with chilled RIPA buffer containing protease inhibitors. Cell lysates were separated by 10% SDS-PAGE (Bio-Rad) and then transferred from the gel to 0.45 um polyvinylidene difluoride membranes (Millipore, Billerica, USA). Page Ruler Plus Prestained Protein Ladder (Fermentas, Hanover, USA) was used to confirm protein electrophoresis and transferring. After blocking in a solution of 5% non-fat dry milk diluted in tris-buffered saline/Tween (TBST), the membranes were washed with TBST and then incubated with primary antibodies overnight at 4°C. The following antibodies were used for signaling pathway analysis: rabbit anti-Collagen I (Abcam, ab270993), rabbit anti-Collagen III (Abcam, ab184993), rabbit anti-alpha smooth muscle actin (Abcam, ab124964), and mouse anti-GAPDH (Beijing Ray Antibody Biotech, RM2002). After washing, the membranes were incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies [Goat anti-Mouse IgG (Beijing Ray Antibody Biotech, RM3001); Goat Anti-Rabbit IgG (Beijing Ray Antibody Biotech, RM3002)] for 1 h at 37°C. Bound antibodies were detected using the ECL Western blotting detection system.

2.11 Statistical analysis

All experiments were performed in triplicate and repeated at least three times. Statistical analyses were performed using SPSS software, version 19.0 (IBM, Armonk, NY, USA). Data represent mean ± standard deviation. A two-tailed, unpaired Student's t-test or the Mann-Whitney U test was employed to compare the values between subgroups for quantitative data. $P < 0.05$ was considered to be statistically significant.

3 Results

3.1 Single-cell RNA-seq reveals immune cell heterogeneity of fibrotic skin diseases and normal scar dermis tissues

To explore the immunological profile of fibrotic skin disease, we used FACS to isolate $CD45^+$ cells from keloid, a paradigm of fibrotic skin diseases, and normal scar dermis tissues for scRNA-seq (Figure 1A). We chose CD45 to enrich immune cells because CD45 has been suggested to express on almost all hematopoietic cells except for mature erythrocytes (25–27). We only used the dermis for scRNA-seq analysis because keloid is a skin dermis

10.3389/fimmu.2024.1522076

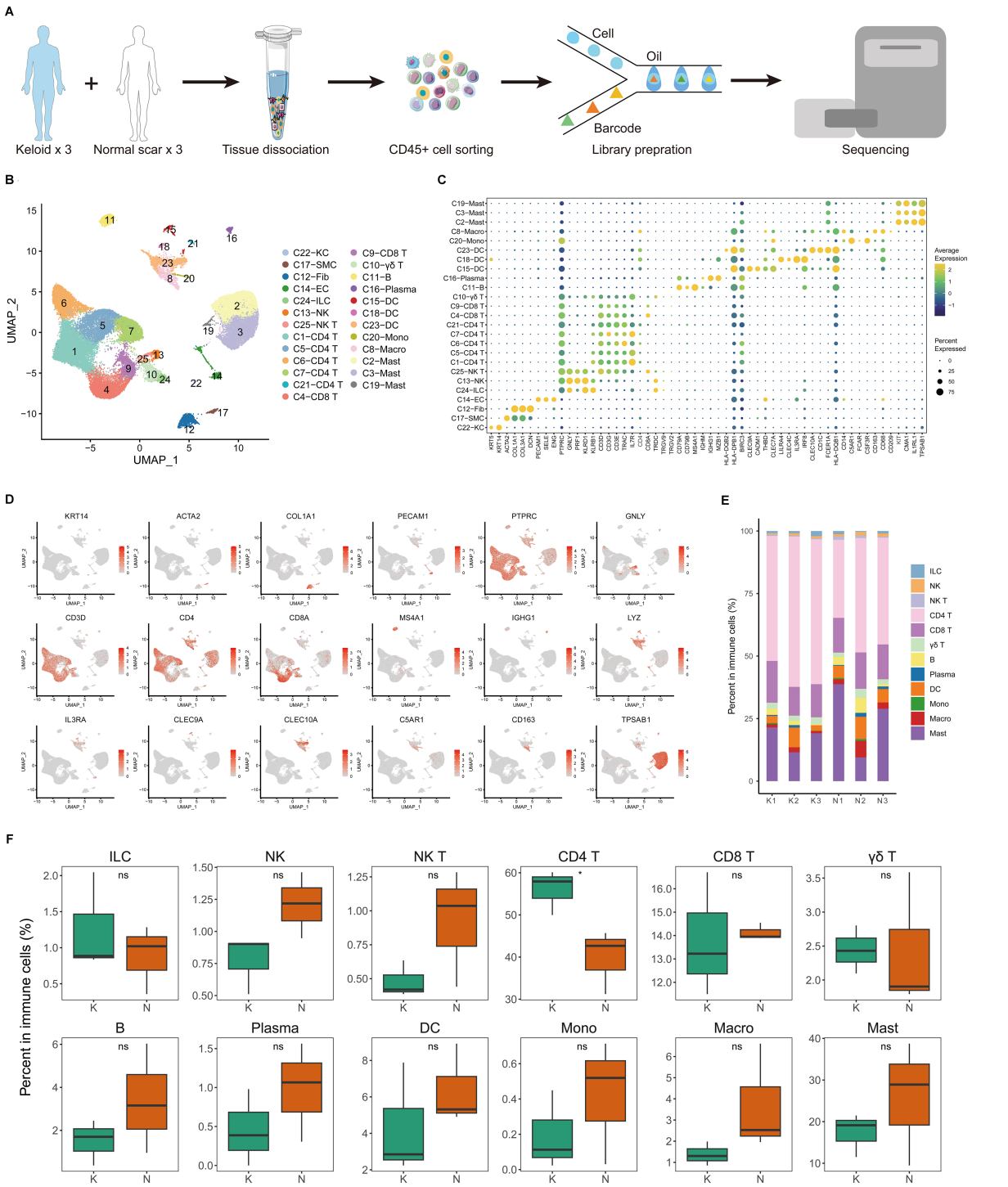


FIGURE 1
Single-cell transcriptome map of immune cells in fibrotic skin disease and normal scar dermis samples. **(A)** Workflow depicting the collection and processing of keloid, a paradigm of fibrotic skin diseases, and normal scar CD45$^+$ cells for scRNA-seq. **(B)** Unsupervised clustering of the 41,084 single cells from three keloid samples and three normal scar samples, including 25 clusters and 16 major clusters. KC, keratinocyte; SMC, smooth muscle cell; Fib, fibroblast; EC, endothelial cell; ILC, innate lymphoid cell; NK, natural killer; DC, dendritic cell; Mono, monocyte; Macro, macrophage. **(C)** Dot plot of the expression of key cell type marker genes in each cell cluster. Bubble size is proportional to the percentage of cells expressing a gene in a cluster and color intensity is related to the average scaled gene expression. **(D)** Feature plots of expression distribution for cell type-specific markers. **(E)** The proportion of each cell type in three keloid samples and three normal scar samples. K, keloid; N, normal scar. **(F)** The percentage of cells for each immune cell type in keloids and normal scars. Ns, not significant; *, P<0.05; K, keloid; N, normal scar.

fibrotic disease. After stringent quality control (Supplementary Figures S1A, B), we obtained the transcriptomes of 41,084 cells. Unsupervised UMAP clustering revealed 25 cell clusters (Figure 1B; Supplementary Figure S1C). Based on established lineage-specific marker genes (Figures 1C, D; Supplementary Figure S1D), we assigned these clusters to multiple cell lineages. The immune cell lineage was identified by PTPRC (Figure 1D). T cells, macrophages, dendritic cells (DCs), and mast cells accounted for the majority of the sequenced cells. Some cells expressed non-immune cell markers, such as fibroblast or endothelial cell markers, which may have resulted from the incomplete removal of these cells by FACS.

We next analyzed the proportions of immune cell lineages in keloids and normal scars. The immune cell lineages in the dermis of keloids and normal scars showed distinct relative cell number ratios (Figures 1E, F). The proportion of $CD4^+$ T cells increased significantly in keloids compared to normal scars, suggesting that $CD4^+$ T cells may play an important role in keloid development (Figure 1F). The proportions of natural killer (NK) cells, NK T cells, B cells, and macrophages were decreased in keloid tissues compared to normal scar tissues, although the difference is not significant. Some other T cells, such as CD8 T cells and γδ T cells, showed similar proportions in keloids and normal scars (Figure 1F).

3.2 T cell subclustering into distinct cell populations and Th17 cells are increased in fibrotic skin disease

Because $CD4^+$ T cells undergo significant changes in keloids compared to normal scars (Figure 1F), and T cells are important for keloid pathogenesis, we next performed unsupervised clustering of all keloid and normal scar T cells (Figure 2A; Supplementary Figure S2A). Based on DEGs, canonical immune markers, and curated gene signatures (Figures 2B–D; Supplementary Figure S2B), we defined 13 transcriptional states: naïve T (C9 and C10), CD8 Teff (C2, C7, and C12), Th17(C3), CD4 Trm (C4), CD4 Tmet (C5), CD4 Treg (C6), CD4 Tcm (C8), MAIT (C11), and Cycling T (C13) (Figures 2A–D). Figures 2E, F show the cell proportions of the T cell subclusters in keloids and normal scars. From the results, we can see that the proportion of Th17 cells and CD4 Tcm cells was consistently increased in the keloid samples compared to the normal scar samples (Figures 2E, F).

Because Th17 cells play an important role in the pathogenesis of a diverse group of inflammation-mediated skin diseases, and inflammation is important for keloid pathogenesis, we next focused on Th17 cells. We compared differences between keloid Th17 cells and normal scar Th17 cells. We identified genes associated with the IL-17 and TNF signaling pathways, such as IL-17A, TNFAIP3, and CCL20, which were significantly increased in keloid Th17 cells (Figure 2G). KEGG pathway analysis also suggested that the IL-17 signaling pathway, TNF signaling pathway, and Th17 cell differentiation-associated pathway were enriched in the keloid Th17 cells (Figure 2H; Supplementary Figure S2C). These results suggest that not only was the proportion of Th17 cells increased but also the identities of Th17 cells changed in keloids compared to normal scars.

We next integrated the scRNA-seq data from $CD45^+$ cells in healthy skin tissues (26) into our study. We integrated and analyzed 73,597 single cells from three keloid, three normal scar, and seven healthy control skin samples (Supplementary Figures S3A, B). The results also showed that the proportion of $CD4^+$ T cells increased significantly in keloids compared to normal scars and healthy skin (Supplementary Figure S3C). We next performed unsupervised clustering of all T cells in the keloid, normal scar, and healthy skin samples. Based on canonical immune markers and curated gene signatures, we defined 15 subclusters (Supplementary Figures S3D–F). The proportion of Th17 cells was increased in keloids compared to healthy skin (Supplementary Figure S3G), which is consistent with the finding in normal scars. We also found that the proportion of CD8 Teff (IFNG+) was consistently increased in keloids and normal scars compared to healthy skin (Supplementary Figure S3G), suggesting that the cells may play a role in scar formation. We next compared differences in the keloid Th17 cells, normal scar Th17 cells, and healthy skin Th17 cells. IL-17A, TNFAIP3, and CCL20 were found to be significantly increased in the keloid Th17 cells compared to the healthy skin Th17 cells (Supplementary Figure S3H). We also found that the Th17-type immune response and IL17-mediated signaling pathway were enriched in the keloid Th17 cells compared to the healthy skin Th17 cells (Supplementary Figure S3I).

We also analyzed Tregs, another important cell in immune regulation. We performed unsupervised clustering on all Tregs in keloids and normal scars. We observed further heterogeneity with two subclusters, $KLF2^+$ Tregs and $LAIR2^+$ Tregs (Supplementary Figure S4A). Pathway analysis suggested that the upregulated genes in the $LAIR2^+$ Tregs were associated with a response to the interleukin-2 and interleukin-15-mediated signaling pathway and the upregulated genes in the $KLF2^+$ Tregs were associated with the regulation of protein stability and fibrillar center (Supplementary Figure S4B). Both Treg subclusters showed similar proportions in the keloid and normal scar samples (Supplementary Figure S4C). We next compared the differences between the keloid Tregs and normal scar Tregs. GO analysis showed that the upregulated genes in the keloid $KLF2^+$ Tregs were associated with the platelet-derived growth factor receptor signaling pathway and the regulation of B cell activation, and the upregulated genes in the keloid $LAIR2^+$ Tregs were associated with the toll-like receptor 2 signaling pathway (Supplementary Figure S4D).

3.3 Transcriptional landscapes reveal the heterogeneity of mono-macrophages and increased macrophage activity in fibrotic skin disease

Because mono-macrophages are reported to play an important role in fibrotic skin disease pathogenesis (9, 11), we next performed unsupervised clustering of all mono-macrophages. Based on DEGs and canonical mono-macrophage markers, we observed further heterogeneity with five subclusters, C1 through C5 (Figures 3A–C). C1, C2, and C3 were macrophages, and C4 and C5 were monocytes. All the mono-macrophage subclusters expressed canonical CD14,

10.3389/fimmu.2024.1522076

CD68, and CD163 except the C4 subcluster, which highly expressed C5AR1 and CSF3R. Figure 3D shows the cell proportions of the mono-macrophage subclusters in keloids and normal scars. From the results, we can see that all five subclusters showed similar proportions in the keloid samples and normal scar samples.

Macrophages can be divided into M1 macrophages and M2 macrophages (28). By calculating M1 and M2 polarization scores using related gene sets, we found that the C1 macrophages were more like M1 macrophages, and the C2 macrophages were more like M2 macrophages. The C3 macrophages were like the intermediate state

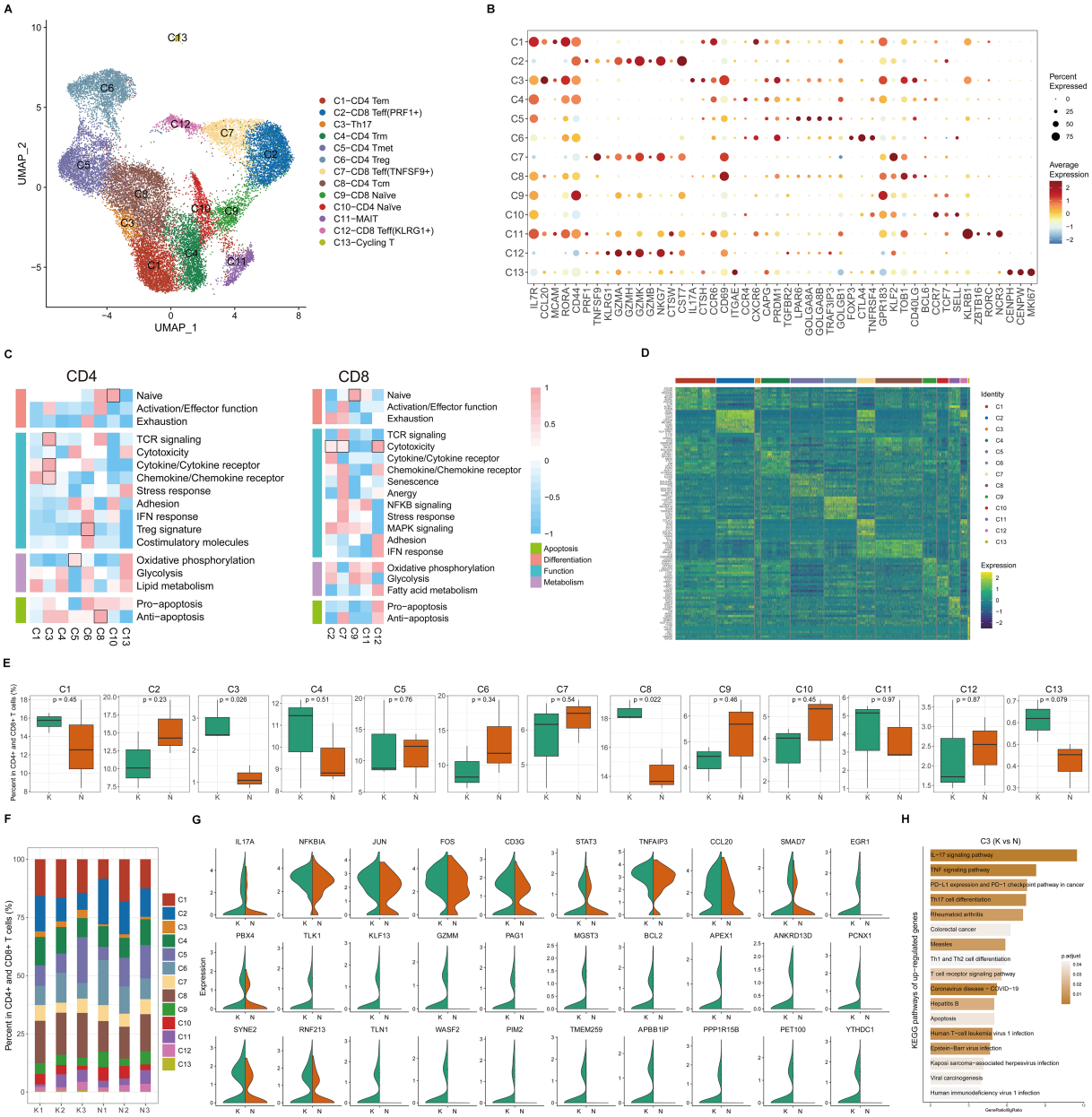


FIGURE 2
Transcriptional diversity of CD4$^+$ and CD8$^+$ T cells. **(A)** Uniform Manifold Approximation and Projection (UMAP) of 13 subclusters identified in CD4$^+$ and CD8$^+$ T cells. **(B)** Dot plots showing distinct expressions of the selected marker genes in each subcluster. **(C)** Heatmap illustrating the scaled score calculated based on the expression of curated gene signatures across CD4$^+$ T cell subclusters (left) and CD8$^+$ T cell subclusters (right). **(D)** Heatmap of the top 10 differentially expressed genes (ranked by log-transformed fold change in descending order) in the CD4$^+$ and CD8$^+$ T cell subclusters. **(E)** Box plots showing the percentage of cells for each T cell subcluster in the keloid and normal scar samples. The p-value indicated in the plot was calculated by unpaired two-tailed t-tests. **(F)** The proportion of each T cell subcluster in three keloid samples and three normal scar samples. K, keloid; N, normal scar. **(G)** Violin plots showing differentially expressed genes in the Th17 cells in keloids and normal scars. K, keloid; N, normal scar. **(H)** Functional KEGG pathway enrichment of the upregulated genes (keloid vs. normal scar, avg_logFC > 0.25, and p value < 0.05) in Th17 cells. The p-value was calculated using the hypergeometric distribution and corrected using the Benjamini and Hochberg method. Pathways with an adjusted p-value of <0.05 are considered significant.

10.3389/fimmu.2024.1522076

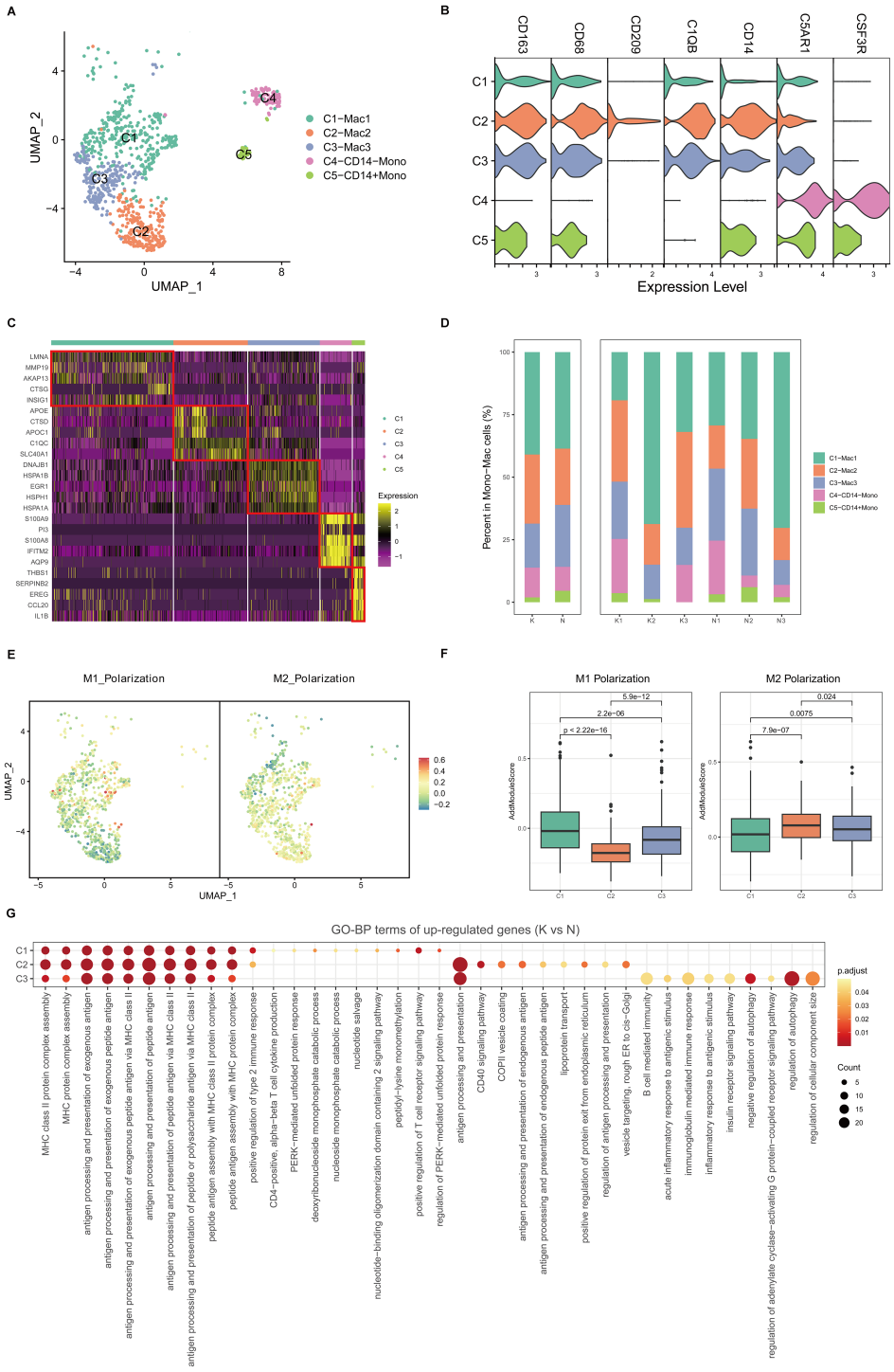


FIGURE 3
Fibrotic skin disease and normal scar mono-macrophages subclustered into distinct cell populations. **(A)** Uniform Manifold Approximation and Projection (UMAP) plots of the mono-macrophage subpopulations. **(B)** Violin plot showing key marker gene expression between the mono-macrophage subpopulations. **(C)** Heatmap showing the expression of the top 5 differentially expressed genes in the mono-macrophage subpopulations. **(D)** Bar plots showing the percentage of each mono-macrophage subpopulation in the keloid and normal scar samples. **(E)** UMAP plots showing the M1 and M2 scores for each cell in the macrophages. **(F)** Box plots showing the M1 and M2 scores for each subpopulation of macrophages. Significance was determined by the unpaired two-tailed t test. **(G)** GO terms enrichment of the upregulated genes (keloid vs. normal scar, avg_logFC > 0.25, and p-value < 0.05) in macrophages. GO terms with an adjusted p-value of <0.05 are considered significant.

between M1 and M2 macrophages (Figures 3E, F). We next compared differences between keloid mono-macrophages and normal scar mono-macrophages (Figure 3G; Supplementary Figures S5A–C). GO analysis showed that the upregulated genes in the keloid C1, C2, and C3 macrophages compared to the normal scar macrophages were all associated with MHC class II protein complex assembly and antigen processing and presentation (Figure 3G), suggesting the consistent active state of macrophages in keloids.

3.4 Transcriptional landscapes reveal heterogeneity of dendritic cells and increased cDC2 and migDC cell activity in fibrotic skin disease

DCs are important antigen-presenting cells (APCs) in the skin. We next performed unsupervised clustering of all DCs in keloids and normal scars. Consistent with previous reports, DCs in the skin were clustered into cDC1, cDC2, pDC, migDC, and Langerhans cells (LCs) (Figures 4A–C). Most of the DCs in keloids and normal scars were cDC2 (Figures 4A–C). There were several LCs in the results, which may have resulted from the incomplete removal of the epidermis. Cell proportion analysis suggested that the five subclusters showed similar proportions in the keloid and normal scar samples (Figures 4D, E).

We next compared differentially expressed genes in the keloid DCs and normal scar DCs (Figure 4F; Supplementary Figures S6A–D). GO analysis showed that the upregulated genes in keloid cDC2 and migDC were associated with MHC class II protein complex assembly and peptide antigen assembly (Figure 4F; Supplementary Figure S6B), suggesting the active state of cDC2 and migDC in keloids.

3.5 Transcriptional landscapes reveal heterogeneity of mast cells in keloids and increased IL-17 signaling in mast cells in fibrotic skin disease

Mast cells are reported to play an important role in fibrotic skin disease pathogenesis (7, 9, 11). We next performed unsupervised clustering of all mast cells in keloids and normal scars (Figure 5A). Based on differentially expressed genes (Figures 5B–D), we observed further heterogeneity with four subclusters, C1 through C4 (Figures 5A–D). The C1 subcluster constitutes the majority of the mast cells and highly expressed GLUL, RRAD, DUSP14, and so on. Figures 5E and F show the cell proportions of the mast cell subclusters in keloids and normal scars. From the results, we can see that all four subclusters showed similar proportions in the keloid and normal scar samples.

We next compared differentially expressed genes in the keloid mast cells and normal scar mast cells (Figure 5G; Supplementary Figures S7A–D). KEGG analysis showed that the upregulated genes in the keloid C1 and C2 mast cell subclusters were associated with the IL-17 signaling pathway (Figure 5G; Supplementary Figure S7A), suggesting that activating IL-17 signaling in the keloid microenvironment may act on mast cells.

3.6 Transcriptional landscapes reveal heterogeneity of B cells and increased plasma cell activity in fibrotic skin disease

Like T cells, B cells are important lymphocytes in the immune system. We next performed unsupervised clustering of all B cells in keloids and normal scars (Supplementary Figure S8A). Consistent with previous reports, B cells in the skin can be clustered into naïve B cells, activated B cells, and plasma cells (Supplementary Figure S8A). A dot plot shows the expression of specific markers in the B cell subpopulations (Supplementary Figure S8B). Cell proportion analysis suggested that the three subclusters showed similar proportions in the keloid and normal scar samples (Supplementary Figure S8C).

We next compared differentially expressed genes in keloid B cells and normal scar B cells (Supplementary Figure S8D). GO analysis showed that the upregulated genes in the keloid plasma cells were associated with antigen binding, the immunoglobulin complex, and the MHC protein complex (Supplementary Figure S8D), suggesting the active state of plasma cells in keloids.

3.7 Potential ligand–receptor interactions in fibrotic skin disease and normal scars

The single-cell dataset provided us with a unique chance to analyze cell-cell communication mediated by ligand-receptor interactions. To define the cell-cell communication landscape in keloids and normal scar immune cells, we used CSOmap, a bioinformatics tool to infer the spatial organization of tissues and molecular determinants of cellular interaction (29). We observed a significant increase in cell-cell communications in keloids compared to normal scars (Supplementary Figure S9A). Interestingly, the cell-cell communications between Th17 cells and other cells increased significantly in keloids compared to normal scars (Figure 6A), suggesting the active cell communication of Th17 cells and its important role in keloids. The main ligand-receptor pairs contributing to the cell-cell communications between Th17 cells and other cells were IL-17A, IL-17F, TNFα, and their receptors (Figure 6B).

3.8 Th17 cells promote proliferation, collagen expression, and migration of fibrotic skin disease fibroblasts by secreting IL-17A

Based on the scRNA-seq analysis, the percentage of Th17 cells was significantly increased in keloids compared to normal scars. To validate this finding, we performed immunofluorescence (IF) staining on skin tissues derived from normal controls and keloids.

10.3389/fimmu.2024.1522076

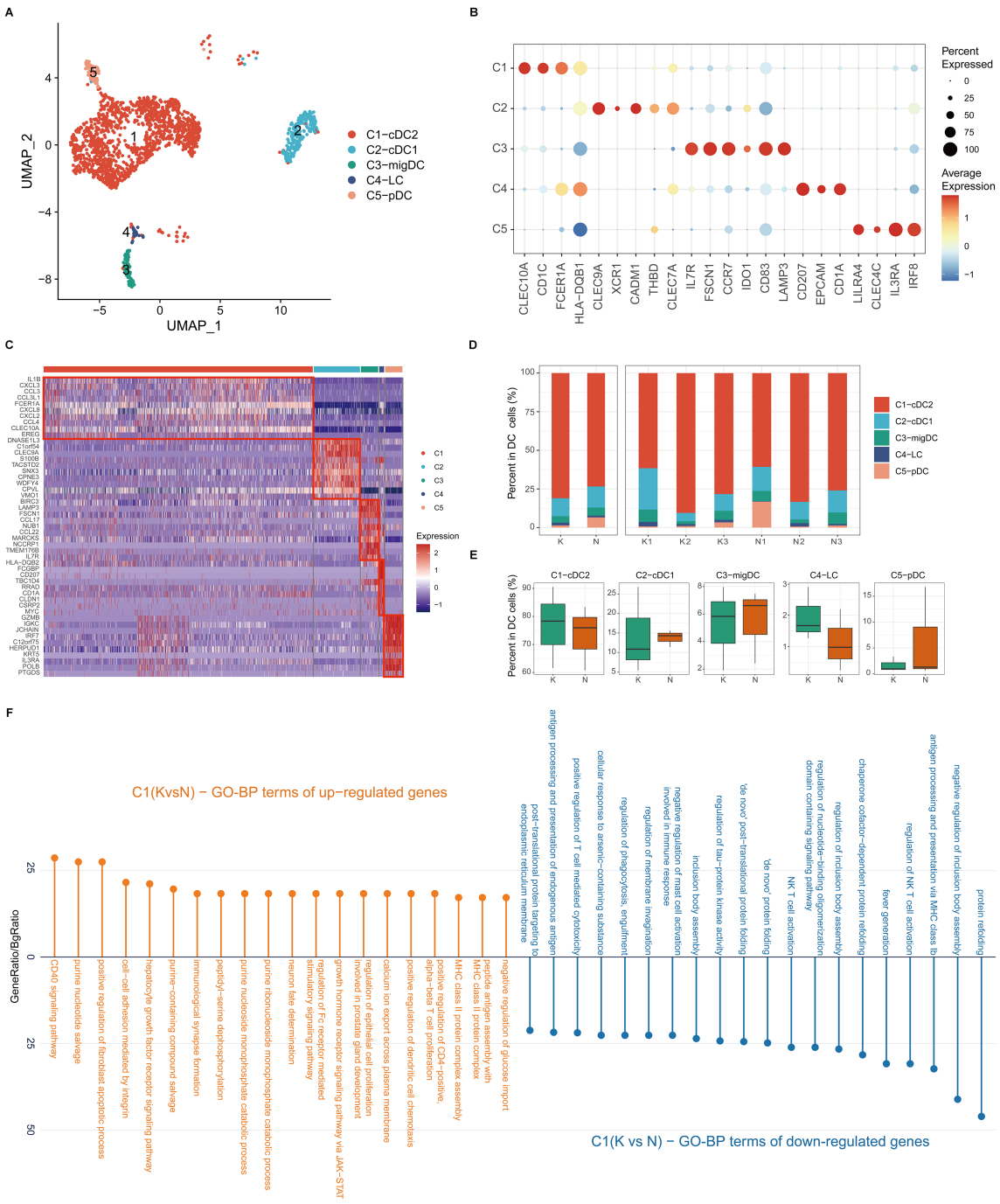


FIGURE 4
Fibrotic skin disease and normal scar dendritic cells (DCs) subclustered into distinct cell populations. **(A)** Uniform Manifold Approximation and Projection (UMAP) plot showing the annotation and color codes for subclusters of DCs. **(B)** Dot plot of representative genes of cell types in DCs. **(C)** Heatmap showing the expression of the top 10 differentially expressed genes in each subcluster of DCs. **(D)** Percentage distribution of each subcluster in the keloid and normal scar samples. **(E)** Box plots showing the percentage of DC subclusters in keloids and normal scars. **(F)** GO terms enrichment of differentially expressed genes in cDC2 cells. GO terms with an adjusted p-value of <0.05 are considered significant.

Th17 cells were identified based on CD4 and IL-17A expression (Figure 7A). The IF staining results showed that the proportion of IL-17A$^+$/CD4$^+$ cells was higher in keloids than in the normal controls (Figures 7A, B). This result is consistent with the scRNA-seq transcriptomics analysis. To explore the function of Th17 cells in keloids, we induced Th17 cells from keloid patients *in vitro* (Supplementary Figure S10A) and subsequently co-cultured them with primary keloid fibroblasts (KF) isolated from keloid patients. After co-culturing with Th17 cells, the KFs exhibited a significant increase in collagen I/III and α-SMA expression and in their proliferative and migratory capabilities, compared to the control groups (Figures 7C–F; Supplementary Figures S10B, C).

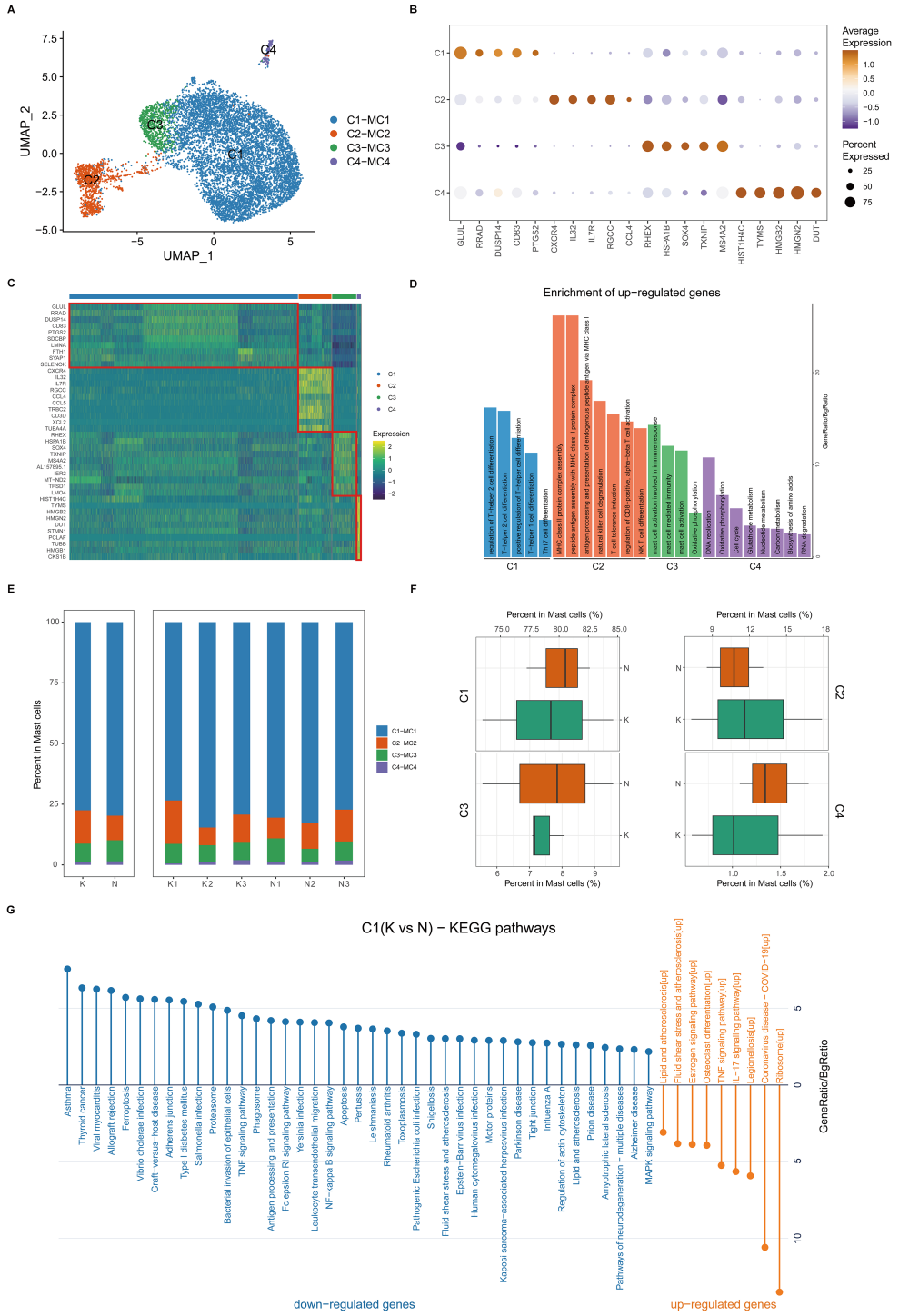


FIGURE 5
Fibrotic skin disease and normal scar mast cells subcluster into distinct cell populations. **(A)** Uniform Manifold Approximation and Projections (UMAPs) of subclustered mast cells, labeled in different colors. Cell type annotations are provided in the figure. **(B)** Dot plot indicating the expression of selected gene sets in mast subclusters. **(C)** Scaled expression of differentially expressed genes in mast subclusters. **(D)** Enrichment of differentially expressed genes in one mast subcluster compared to other mast subclusters. Results with adjusted P-value of <0.05 are considered significant. **(E)** Bar plot showing the fraction of mast subcluster in keloid and normal scar samples. **(F)** Boxplot showing the fraction of mast subclusters in keloid and normal scar. **(G)** KEGG pathway enrichment of differentially expressed genes in C1 subcluster. GO terms with adjusted P-value of <0.05 are considered significant.

10.3389/fimmu.2024.1522076

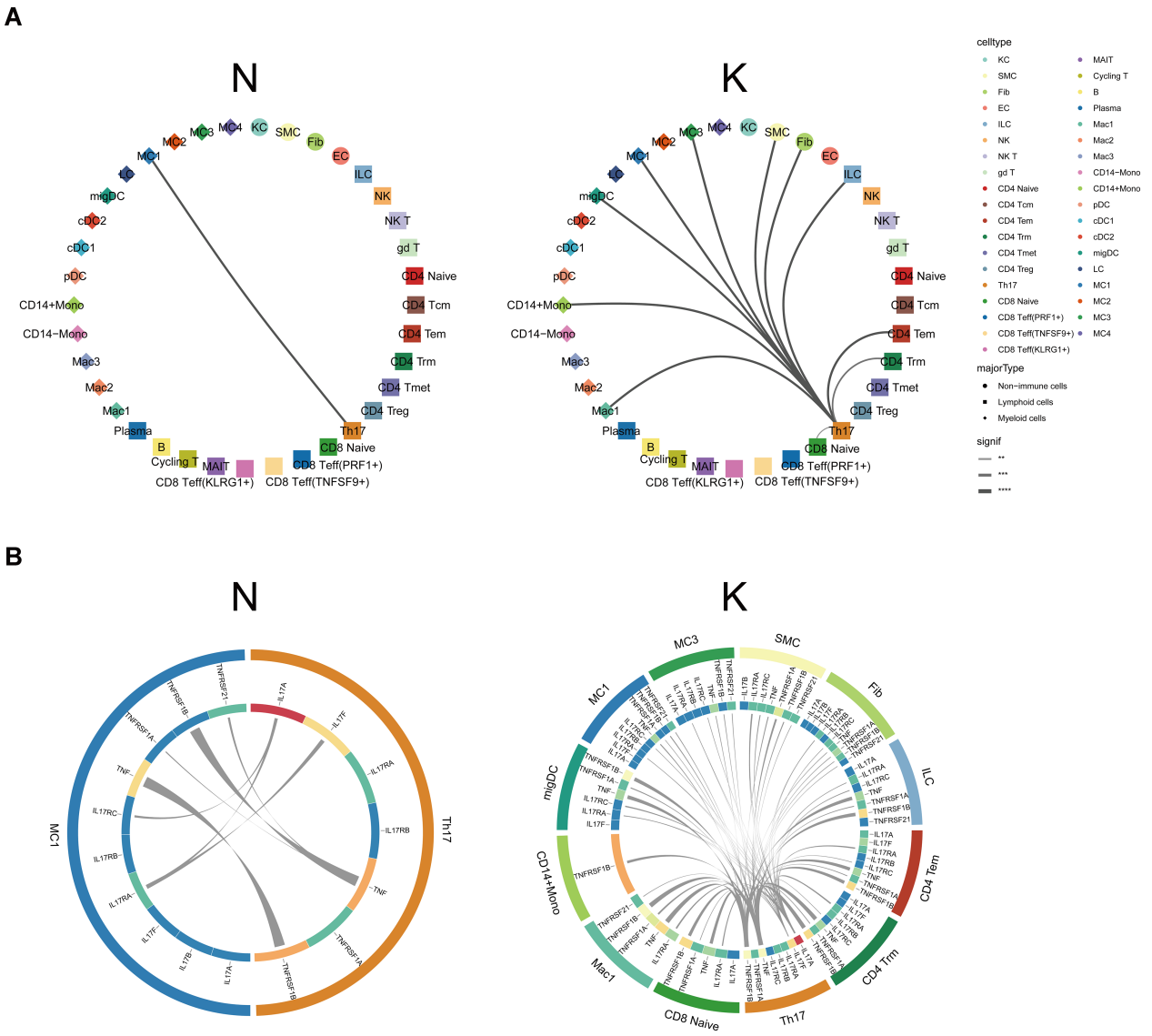


FIGURE 6
Cellular interactions between Th17 cells and other cell types. **(A)** CSOmap analysis showing the interaction between Th17 cells and other cell subsets in normal scars (left) and keloids (right). Line thickness represents the significance of the cell-cell interaction. **(B)** Putative ligand and receptor pairs related to IL-17 and TNF within the Th17 cells and other cell populations in normal scars (left) and in keloids (right). The color of the middle ring is related to the average expression of genes in the cell types, with red representing a high expression and blue representing a low expression. The thicker the line, the greater the contribution of the ligand-receptor pairs. **, p<0.01, ***, p<0.001, ****, p<0.0001.

IL-17A has been reported to be the key molecule for Th17 cells' functions in fibrotic diseases. To ascertain whether the increased collagen I/III expression and proliferative and migratory capabilities of KFs had resulted from IL-17A secreted by Th17 cells, we next introduced a neutralizing antibody against IL-17A into the co-culture system of Th17 cells and KFs. The IL-17A neutralizing antibody inhibited the increased expression of collagen I and III and the proliferative and migratory capabilities of the KFs co-cultured with Th17 (Figures 7G–J; Supplementary Figures S10D, E), suggesting that Th17 cells promote proliferation, collagen expression, and migration of keloid fibroblasts through IL-17A.

3.9 Th17 cells are increased in hypertrophic scars and scleroderma

To examine the consistency of our findings in other fibrotic skin diseases, we performed immunofluorescence staining in hypertrophic scar and scleroderma tissues. The immunofluorescence staining results showed that the proportion of IL-17A$^+$/CD4$^+$ cells was higher in hypertrophic scar and scleroderma tissues than in normal control tissues (Figures 8A–D). Taken together, these results indicated that increasing Th17 cells may be a universal mechanism in fibrotic skin diseases.

10.3389/fimmu.2024.1522076

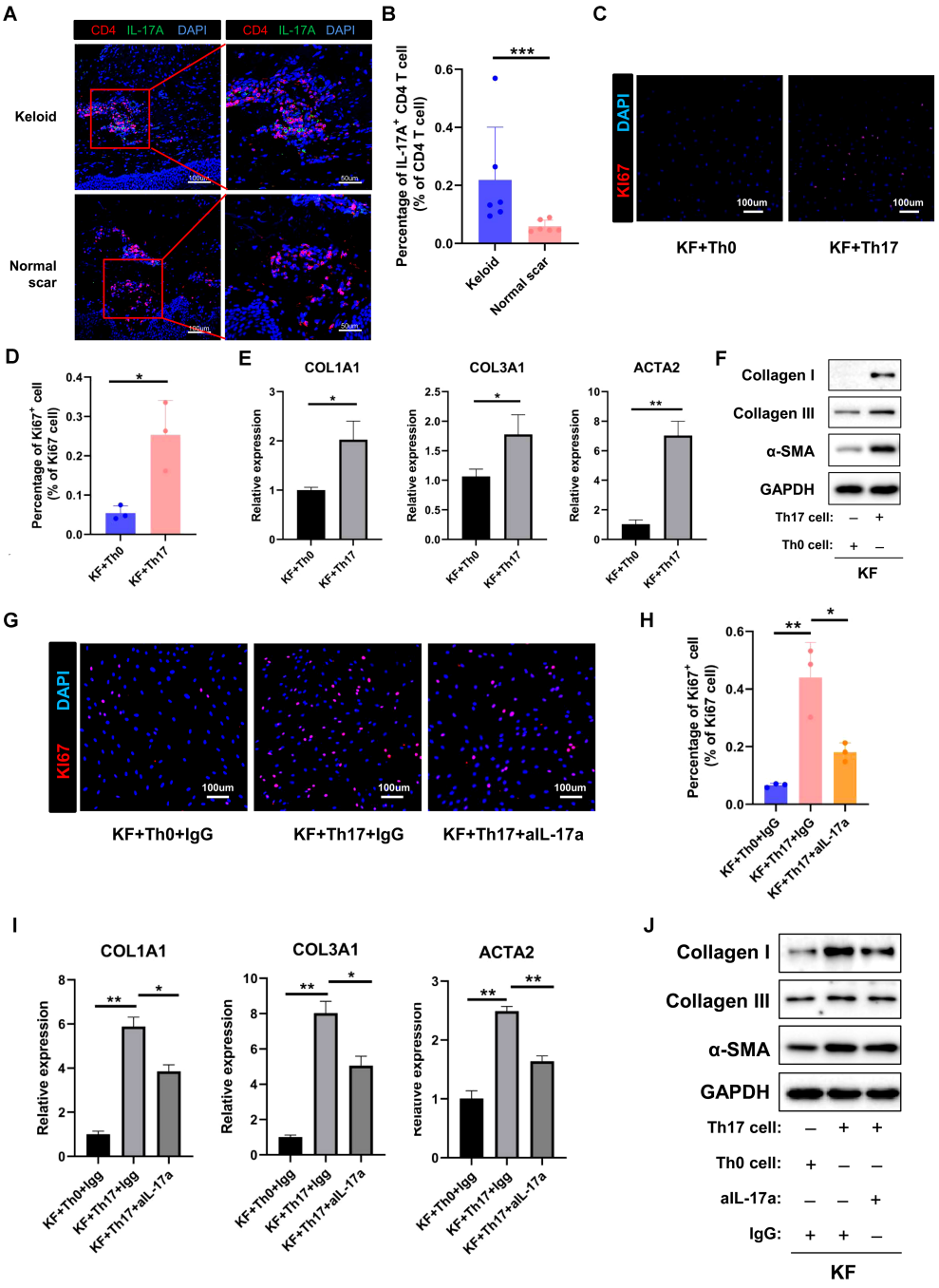


FIGURE 7
Th17 cell promotes the proliferation and collagen expression of keloid fibroblasts by secreting IL-17A. **(A)** Immunofluorescence staining of IL-17A and CD4 in keloid and normal scar tissues. The right panels are the insets of the left panels. Scale bar = 100 μm (left panel) and 50 μm (right panel). **(B)** Percentage of IL-17A$^+$/CD4$^+$ cells in normal and keloid tissues. Error bars represent SD (n=6). ***, p<0.001. **(C, D)** Ki67 staining analysis of fibroblasts co-cultured with Th0 or Th17 cells. Scale bar = 100 μm. Error bars represent SD (n=3). *, P<0.05. **(E, F)** qRT-PCR and Western blot analysis of collagen I, collagen III, and α-SMA expression in fibroblast co-cultured with Th0 or Th17 cells. Error bars represent SD (n=3). *, p<0.05; **, P<0.01. **(G, H)** Ki67 staining analysis of fibroblasts co-cultured with Th0 or Th17 cells in the presence or absence of anti-IL-17A antibody. Error bars represent SD (n=3). *, p<0.05; **, p<0.01. **(I, J)** qRT-PCR and Western blot analysis of collagen I, collagen III, and α-SMA expression in fibroblasts co-cultured with Th0 or Th17 cells in the presence or absence of anti-IL-17A antibody. Error bars represent SD (n=3). *p<0.05; **, p<0.01.

4 Discussion

Immune cells and inflammation have been reported to be important for the pathogenesis of fibrotic skin diseases (7, 9, 11). Although there have been some studies exploring the composition of immune cells in fibrotic skin diseases, these studies detected the gene expression of all cells in fibrotic skin disease tissues and did not enrich immune cells (13–15). In this study, we built a single-cell atlas of fibrotic skin disease and normal scar immune cells using FACS-enriched CD45$^+$ cells and explored the function of the Th17

cell-fibroblast interaction in the pathogenesis of fibrotic skin disease. These findings will help us understand fibrotic skin disease pathogenesis in depth, and provide potential targets for clinical therapies for fibrotic skin diseases.

Mast cells, Treg cells, and M2 macrophages have been reported to play important roles in the pathogenesis of fibrotic skin disease (6, 7, 11, 12). These cells have been suggested to be increased and promote extracellular matrix deposition in keloid tissues. However, in our single-cell atlas of keloid and normal scar immune cell research, we found that there were no differences in the proportions of these cells in keloids and normal scars. The gene expression of these cells had significant differences in keloids and normal scars. The upregulated genes in the keloid C1 mast cell subcluster, the major subcluster of mast cells, were associated with lipids and atherosclerosis, the TNF signaling pathway, and the IL-17 signaling pathway (Figure 5G). The upregulated genes in the keloid Treg cells, compared to normal scar Treg cells, were associated with the TNF signaling pathway, IL-17 signaling pathway, and apoptosis (Supplementary Figure S2C). The upregulated genes in the keloid M2 macrophages, compared to normal scar M2 macrophages, were

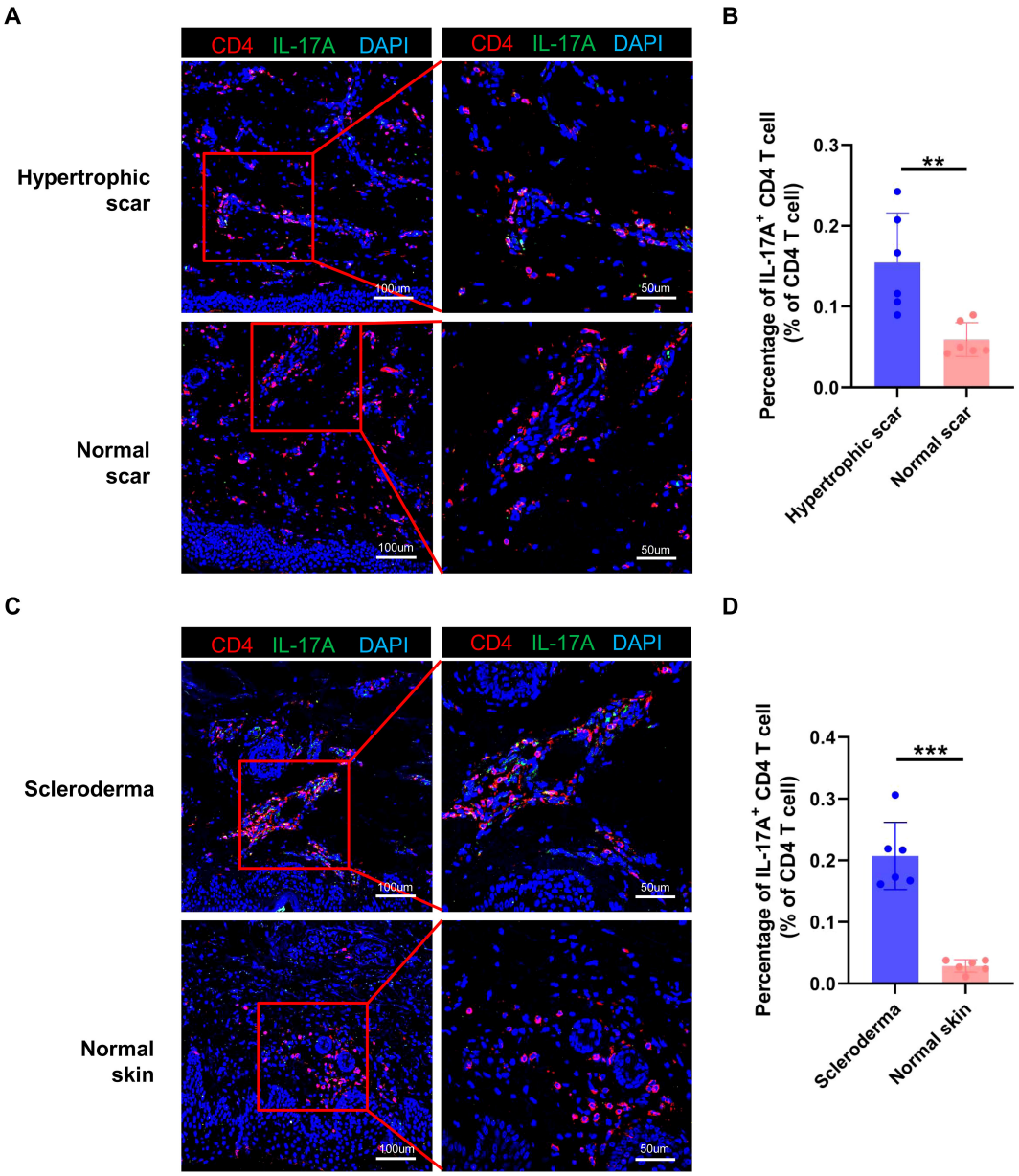


FIGURE 8
Th17 cells are increased in hypertrophic scars and scleroderma. **(A)** Immunofluorescence staining of IL-17A and CD4 in hypertrophic scar and normal scar tissues. The right panels are the insets of the left panels. Scale bar = 100 μm (left panel) and 50 μm (right panel). **(B)** Percentage of IL-17A$^+$/CD4$^+$ cells in hypertrophic scar and normal scar tissues. Error bars represent SD (n=6). **, p<0.01. **(C)** Immunofluorescence staining of IL-17A and CD4 in scleroderma and normal skin tissues. The right panels are the insets of the left panels. Scale bar = 100 μm (left panel) and 50 μm (right panel). **(D)** Percentage of IL-17A$^+$/CD4$^+$ cells in scleroderma and normal skin tissues. Error bars represent SD (n=6). ***, p<0.001.

associated with MHC class II protein complex assembly and antigen processing and presentation (Figure 3G). These findings suggested that not the change in cell proportions but the change in gene expression of mast cells, Treg cells, and M2 macrophages may contribute to fibrotic skin disease development.

Th17 cells are key cells for host protection against mucosal infections and are major pathogenic cells in multiple autoimmune and inflammatory diseases, including psoriasis and systemic lupus erythematosus (16, 30, 31). IL-17 has been reported to be the major effector molecule of Th17 in the aforementioned functions (32, 33). In recent years, the roles of Th17 cells in fibrotic diseases have been paid increasing attention. Th17 cells have been reported to play important roles in intestinal fibrosis, lung fibrosis, and myocardial fibrosis (18, 34, 35). However, the functions of Th17 cells in keloids are still unknown. In this study, we found that the percentage of Th17 cells was significantly increased in keloids compared to normal scars and Th17 cells promoted the collagen expression, proliferation, and migration of keloid fibroblast (Figures 2E, 7, Supplementary Figure S10). Mechanism studies showed that the Th17 cells performed the above functions by secreting IL-17A (Figure 7, Supplementary Figure S10), which is consistent with previous findings (36, 37). Importantly, we also found an increased number of Th17 cells in hypertrophic scars and scleroderma compared to normal controls (Figure 8). These results suggested that Th17 cells may have an important role in multiple skin fibrosis diseases, and may serve as target cells for fibrosis treatment.

The mechanism that IL-17A promotes fibrotic diseases is complex, organ-specific, and disease-specific. IL-17A has been reported to promote the fibrosis of systemic sclerosis by increasing inflammation and the proliferation and collagen deposition of fibroblasts (38, 39). The increased level of IL-17A in liver fibrosis facilitates the influx of inflammatory cells, drives the expression of profibrogenic growth factors, and activates hepatic stellate cells in the liver (40–42). The liver-infiltrating inflammatory cells in turn induce the production of profibrotic cytokines such as TNF-α, IL-6, IL-1, and TGF-β1 to accelerate fibrosis (40). Several studies have suggested that IL-17 directly interacts with colonic IL-17R, expressing myofibroblasts and contributing significantly to stricture development in Crohn's disease (43–45). IL-6, IL-8, and MCP-1 secretions were rapidly induced by IL-17 in colonic subepithelial myofibroblasts (43). *In vitro* stimulation of IL-17 induced HSP47 and type I collagen in human intestinal myofibroblasts (45). The mechanism by which IL-17A facilitates fibrosis in keloids is still unclear. We will explore the mechanism using high-throughput sequencing methods and molecular biology experiments. Illustrating the downstream signaling pathways activated by IL-17A in fibroblasts of keloids can supply new targets for keloid therapy.

Both macrophages and DCs are important antigen-presenting cells in the skin. Many studies suggest that M2 macrophages are increased and play an important role in keloid development (12, 46), but the roles of DCs are still unclear in keloids. In our findings, all the subclusters of macrophages and DCs showed similar proportions between the keloid and normal scar samples (Figures 3, 4). However, the upregulated genes in the keloid M1 macrophages, M2 macrophages, and cDC2 were all associated with MHC class II protein complex assembly and antigen assembly (Figures 3G, 4F). These results indicate that macrophages and cDC2 are active in fibrotic skin diseases and may serve as target cells for fibrotic skin disease therapy.

In conclusion, we provided a systematic analysis of immune cell heterogeneity in fibrotic skin disease at single-cell resolution in this study. In addition, we identified that increased Th17 cells in fibrotic skin disease are involved in the proliferation, collagen expression, and migration of fibrotic skin disease fibroblasts. These findings will help us to understand fibrotic skin disease pathogenesis in depth and identify potential targets for fibrotic skin disease treatment.

Data availability statement

Single-cell RNA-seq data have been deposited in the Gene Expression Omnibus database under accession codes "GSE270438".

Ethics statement

The studies involving humans were approved by the Medical and Ethics Committees of Dermatology Hospital, Southern Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

C-CD: Validation, Visualization, Data curation, Formal analysis, Methodology, Resources, Software, Writing – original draft. X-YX: Data curation, Methodology, Software, Validation, Visualization, Writing – original draft. YZ: Methodology, Software, Writing – original draft, Formal analysis. L-CL: Methodology, Writing – original draft, Resources, Validation, Visualization. XW: Methodology, Resources, Validation, Visualization, Writing – original draft. J-YC: Methodology, Data curation, Writing – original draft. L-YY: Data curation, Methodology, Writing – original draft. D-HZ: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing. BY: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article.

This work was supported by grants from the National Natural Science Foundation of China (82373457), the Natural Science Foundation of Guangdong Province (2023A1515010120) and University Joint Laboratory of Guangdong, Hong Kong and Macao (2021LSYS010).

Acknowledgments

We sincerely thank the researchers and study participants for their contributions to this study. We also thank Prof. Zhili Rong and Prof. Yingping Xu for helpful scientific discussions.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1522076/full#supplementary-material

References

1. Henderson NC, Rieder F, Wynn TA. Fibrosis: from mechanisms to medicines. *Nature*. (2020) 587:555–66. doi: 10.1038/s41586-020-2938-9

2. Antar SA, Ashour NA, Marawan ME, Al-Karmalawy AA. Fibrosis: types, effects, markers, mechanisms for disease progression, and its relation with oxidative stress, immunity, and inflammation. *Int J Mol Sci*. (2023) 24. doi: 10.3390/ijms24044004

3. Younesi FS, Miller AE, Barker TH, Rossi FMV, Hinz B. Fibroblast and myofibroblast activation in normal tissue repair and fibrosis. *Nat Rev Mol Cell Biol*. (2024) 25:617–38. doi: 10.1038/s41580-024-00716-0

4. Talbott HE, Mascharak S, Griffin M, Wan DC, Longaker MT. Wound healing, fibroblast heterogeneity, and fibrosis. *Cell Stem Cell*. (2022) 29:1161–80. doi: 10.1016/j.stem.2022.07.006

5. Griffin MF, desJardins-Park HE, Mascharak S, Borrelli MR, Longaker MT. Understanding the impact of fibroblast heterogeneity on skin fibrosis. *Dis Models Mech*. (2020) 13. doi: 10.1242/dmm.044164

6. Do NN, Eming SA. Skin fibrosis: models and mechanisms. *Curr Res Trans Med*. (2016) 64:185–93. doi: 10.1016/j.retram.2016.06.003

7. Cohen AJ, Nikbakht N, Uitto J. Keloid disorder: genetic basis, gene expression profiles, and immunological modulation of the fibrotic processes in the skin. *Cold Spring Harbor Perspect Biol*. (2023) 15. doi: 10.1101/cshperspect.a041245

8. Li DJ, Berry CE, Wan DC, Longaker MT. Clinical, mechanistic, and therapeutic landscape of cutaneous fibrosis. *Sci Trans Med*. (2024) 16:eadn7871. doi: 10.1126/scitranslmed.adn7871

9. Lee CC, Tsai CH, Chen CH, Yeh YC, Chung WH, Chen CB. An updated review of the immunological mechanisms of keloid scars. *Front Immunol*. (2023) 14:1117630. doi: 10.3389/fimmu.2023.1117630

10. Hong YK, Chang YH, Lin YC, Chen B, Guevara BEK, Hsu CK. Inflammation in wound healing and pathological scarring. *Adv Wound Care*. (2023) 12:288–300. doi: 10.1089/wound.2021.0161

11. Yu Y, Wu H, Zhang Q, Ogawa R, Fu S. Emerging insights into the immunological aspects of keloids. *J Dermatol*. (2021) 48:1817–26. doi: 10.1111/1346-8138.16149

12. Wang ZC, Zhao WY, Cao Y, Liu YQ, Sun Q, Shi P, et al. The roles of inflammation in keloid and hypertrophic scars. *Front Immunol*. (2020) 11:603187. doi: 10.3389/fimmu.2020.603187

13. Direder M, Weiss T, Copic D, Vorstandlechner V, Laggner M, Pfisterer K, et al. Schwann cells contribute to keloid formation. *Matrix biology: J Int Soc Matrix Biol*. (2022) 108:55–76. doi: 10.1016/j.matbio.2022.03.001

14. Feng C, Shan M, Xia Y, Zheng Z, He K, Wei Y, et al. Single-cell rna sequencing reveals distinct immunology profiles in human keloid. *Front Immunol*. (2022) 13:940645. doi: 10.3389/fimmu.2022.940645

15. Gaydosik AM, Tabib T, Domsic R, Khanna D, Lafyatis R, Fuschiotti P. Single-cell transcriptome analysis identifies skin-specific T-cell responses in systemic sclerosis. *Ann rheumatic Dis*. (2021) 80:1453–60. doi: 10.1136/annrheumdis-2021-220209

16. Akhter S, Tasnim FM, Islam MN, Rauf A, Mitra S, Emran TB, et al. Role of th17 and il-17 cytokines on inflammatory and auto-immune diseases. *Curr Pharm design*. (2023) 29:2078–90. doi: 10.2174/1381612829666230904150808

17. Yan JB, Luo MM, Chen ZY, He BH. The function and role of the th17/treg cell balance in inflammatory bowel disease. *J Immunol Res*. (2020) 2020:8813558. doi: 10.1155/2020/8813558

18. Zhao X, Yang W, Yu T, Yu Y, Cui X, Zhou Z, et al. Th17 cell-derived amphiregulin promotes colitis-associated intestinal fibrosis through activation of mtor and mek in intestinal myofibroblasts. *Gastroenterology*. (2023) 164:89–102. doi: 10.1053/j.gastro.2022.09.006

19. Chen W, Lin F, Feng X, Yao Q, Yu Y, Gao F, et al. Msc-derived exosomes attenuate hepatic fibrosis in primary sclerosing cholangitis through inhibition of th17 differentiation. *Asian J Pharm Sci*. (2024) 19:100889. doi: 10.1016/j.ajps.2024.100889

20. Park SJ, Hahn HJ, Oh SR, Lee HJ. Theophylline attenuates blm-induced pulmonary fibrosis by inhibiting th17 differentiation. *Int J Mol Sci*. (2023) 24. doi: 10.3390/ijms24021019

21. Brembilla NC, Montanari E, Truchetet ME, Raschi E, Meroni P, Chizzolini C. Th17 cells favor inflammatory responses while inhibiting type I collagen deposition by dermal fibroblasts: differential effects in healthy and systemic sclerosis fibroblasts. *Arthritis Res Ther*. (2013) 15:R151. doi: 10.1186/ar4334

22. Vettori S, Barra G, Russo B, Borgia A, Pasquale G, Pellecchia L, et al. T-cell proapoptotic and antifibrotic activity against autologous skin fibroblasts *in vitro* is associated with il-17a axis upregulation in systemic sclerosis. *Front Immunol*. (2020) 11:220. doi: 10.3389/fimmu.2020.00220

23. Chu Y, Dai E, Li Y, Han G, Pei G, Ingram DR, et al. Pan-cancer T cell atlas links a cellular stress response state to immunotherapy resistance. *Nat Med*. (2023) 29:1550–62. doi: 10.1038/s41591-023-02371-y

24. Sun Y, Wu L, Zhong Y, Zhou K, Hou Y, Wang Z, et al. Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. *Cell*. (2021) 184:404–21.e16. doi: 10.1016/j.cell.2020.11.041

25. Hailer AA, Wu D, El Kurdi A, Yuan M, Cho RJ, Cheng JB. Isolation of human cutaneous immune cells for single-cell rna sequencing. *STAR Protoc*. (2023) 4:102239. doi: 10.1016/j.xpro.2023.102239

26. Liu Y, Wang H, Taylor M, Cook C, Martínez-Berdeja A, North JP, et al. Classification of human chronic inflammatory skin disease based on single-cell immune profiling. *Sci Immunol*. (2022) 7:eabl9165. doi: 10.1126/sciimmunol.abl9165

27. Al Barashdi MA, Ali A, McMullin MF, Mills K. Protein tyrosine phosphatase receptor type C (Ptprc or cd45). *J Clin Pathol*. (2021) 74:548–52. doi: 10.1136/jclinpath-2020-206927

28. Yunna C, Mengru H, Lei W, Weidong C. Macrophage M1/M2 polarization. *Eur J Pharmacol*. (2020) 877:173090. doi: 10.1016/j.ejphar.2020.173090

29. Ren X, Zhong G, Zhang Q, Zhang L, Sun Y, Zhang Z. Reconstruction of cell spatial organization from single-cell rna sequencing data based on ligand-receptor mediated self-assembly. *Cell Res*. (2020) 30:763–78. doi: 10.1038/s41422-020-0353-2

30. Li B, Huang L, Lv P, Li X, Liu G, Chen Y, et al. The role of th17 cells in psoriasis. *Immunologic Res*. (2020) 68:296–309. doi: 10.1007/s12026-020-09149-1

31. Muhammad Yusoff F, Wong KK, Mohd Redzwan N. Th1, th2, and th17 cytokines in systemic lupus erythematosus. *Autoimmunity*. (2020) 53:8–20. doi: 10.1080/08916934.2019.1693545

32. Mills KHG. Il-17 and il-17-producing cells in protection versus pathology. *Nat Rev Immunol*. (2023) 23:38–54. doi: 10.1038/s41577-022-00746-9

33. McGeachy MJ, Cua DJ, Gaffen SL. The il-17 family of cytokines in health and disease. *Immunity*. (2019) 50:892–906. doi: 10.1016/j.immuni.2019.03.021

34. Guo Y, He Z, Chen Z, Chen F, Wang C, Zhou W, et al. Inhibition of th17 cells by donepezil ameliorates experimental lung fibrosis and pulmonary hypertension. *Theranostics*. (2023) 13:1826–42. doi: 10.7150/thno.82069

35. Lu M, Qin X, Yao J, Yang Y, Zhao M, Sun L. Th17/treg imbalance modulates rat myocardial fibrosis and heart failure by regulating lox expression. *Acta physiologica (Oxford England)*. (2020) 230:e13537. doi: 10.1111/apha.13537

36. Lee SY, Kim EK, Seo HB, Choi JW, Yoo JH, Jung KA, et al. Il-17 induced stromal cell-derived factor-1 and profibrotic factor in keloid-derived skin fibroblasts via the stat3 pathway. *Inflammation*. (2020) 43:664–72. doi: 10.1007/s10753-019-01148-1

37. Lee SY, Lee AR, Choi JW, Lee CR, Cho KH, Lee JH, et al. Il-17 induces autophagy dysfunction to promote inflammatory cell death and fibrosis in keloid fibroblasts via the stat3 and hif-1α Dependent signaling pathways. *Front Immunol*. (2022) 13:888719. doi: 10.3389/fimmu.2022.888719

38. Ahmed S, Misra DP, Agarwal V. Interleukin-17 pathways in systemic sclerosis-associated fibrosis. *Rheumatol Int*. (2019) 39:1135–43. doi: 10.1007/s00296-019-04317-5

39. Yang X, Yang J, Xing X, Wan L, Li M. Increased frequency of th17 cells in systemic sclerosis is related to disease activity and collagen overproduction. *Arthritis Res Ther*. (2014) 16:R4. doi: 10.1186/ar4430

40. Meng F, Wang K, Aoyama T, Grivennikov SI, Paik Y, Scholten D, et al. Interleukin-17 signaling in inflammatory, kupffer cells, and hepatic stellate cells exacerbates liver fibrosis in mice. *Gastroenterology*. (2012) 143:765–76.e3. doi: 10.1053/j.gastro.2012.05.049

41. Hara M, Kono H, Furuya S, Hirayama K, Tsuchiya M, Fujii H. Interleukin-17a plays a pivotal role in cholestatic liver fibrosis in mice. *J Surg Res*. (2013) 183:574–82. doi: 10.1016/j.jss.2013.03.025

42. Tan Z, Qian X, Jiang R, Liu Q, Wang Y, Chen C, et al. Il-17a plays a critical role in the pathogenesis of liver fibrosis through hepatic stellate cell activation. *J Immunol (Baltimore Md: 1950)*. (2013) 191:1835–44. doi: 10.4049/jimmunol.1203013

43. Hata K, Andoh A, Shimada M, Fujino S, Bamba S, Araki Y, et al. Il-17 stimulates inflammatory responses via nf-kappab and map kinase pathways in human colonic myofibroblasts. *Am J Physiol Gastrointestinal liver Physiol*. (2002) 282:G1035–44. doi: 10.1152/ajpgi.00494.2001

44. Andoh A, Hata K, Araki Y, Fujiyama Y, Bamba T. Interleukin (Il)-4 and il-17 synergistically stimulate il-6 secretion in human colonic myofibroblasts. *Int J Mol Med*. (2002) 10:631–4. doi: 10.3892/ijmm.10.5.631

45. Honzawa Y, Nakase H, Shiokawa M, Yoshino T, Imaeda H, Matsuura M, et al. Involvement of interleukin-17a-induced expression of heat shock protein 47 in intestinal fibrosis in crohn's disease. *Gut*. (2014) 63:1902–12. doi: 10.1136/gutjnl-2013-305632

46. Xu X, Gu S, Huang X, Ren J, Gu Y, Wei C, et al. The role of macrophages in the formation of hypertrophic scars and keloids. *Burns Trauma*. (2020) 8:tkaa006. doi: 10.1093/burnst/tkaa006

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 14 January 2025
DOI 10.3389/fimmu.2024.1439082

OPEN ACCESS

EDITED BY
Veronica Anania,
Genentech Inc., United States

REVIEWED BY
Ting Zhao,
Yunnan University of Traditional Chinese Medicine, China
Lih-Ling Lin,
PharmaEssentia, Taiwan

*CORRESPONDENCE
Ho-Youn Kim
ho0919@gmail.com
Wan-Uk Kim
wan725@catholic.ac.kr

†These authors have contributed equally to this work

RECEIVED 27 May 2024
ACCEPTED 16 December 2024
PUBLISHED 14 January 2025

CITATION
Hong B-K, You S, Kim JG, Kim M, Lee N, Lee K, Baek I-P, Ju JH, Kim W-U and Kim H-Y (2025) Upregulation of interferon-γ response genes in monocytes and T cells identified by single-cell transcriptomics in patients with anti-citrullinated peptide antibody-positive early rheumatoid arthritis.
Front. Immunol. 15:1439082.
doi: 10.3389/fimmu.2024.1439082

Upregulation of interferon-γ response genes in monocytes and T cells identified by single-cell transcriptomics in patients with anti-citrullinated peptide antibody-positive early rheumatoid arthritis

Bong-Ki Hong[1†], Sungyong You[2†], Jung Gon Kim[1,3†], Minhyung Kim[2], Naeun Lee[1], Kijun Lee[4,5], In-Pyo Baek[5], Ji Hyeon Ju[4,5,6], Wan-Uk Kim[1,6*] and Ho-Youn Kim[7*]

[1]Center for Integrative Rheumatoid Transcriptomics and Dynamics, The Catholic University of Korea, Seoul, Republic of Korea, [2]Urology and Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA, United States, [3]Division of Rheumatology, Department of Internal Medicine, Inje University Ilsan Paik Hospital, Goyang, Republic of Korea, [4]Catholic iPSC Research Center, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea, [5]YiPSCELL, Inc., Seoul, Republic of Korea, [6]Department of Internal Medicine, The Catholic University of Korea, Seoul St. Mary's Hospital, Seoul, Republic of Korea, [7]The Catholic University of Korea and Ho-Youn Kim's Clinic for Arthritis Rheumatism, Seoul, Republic of Korea

Introduction: Our aim was to investigate the insufficiently understood differences in the immune system between anti-citrullinated peptide antibody (ACPA)-positive ($ACPA^+$) and ACPA-negative ($ACPA^-$) early rheumatoid arthritis (eRA) patients.

Methods: We performed multiple cytokine assays using sera from drug-naïve $ACPA^+$ and $ACPA^-$ eRA patients. Additionally, we conducted single-cell RNA sequencing of $CD45^+$ cells from peripheral blood samples to analyze and compare the distribution and functional characteristics of the cell subsets based on the ACPA status.

Results: Serum concentrations of interferon-γ (IFN-γ) and interleukin (IL)-12 were higher in $ACPA^+$ eRA than in $ACPA^-$ eRA. Single-cell transcriptome analysis of 37,318 cells identified 17 distinct cell types and revealed the expansion of $IL1B^+$ proinflammatory monocytes, $IL7R^+$ T cells, and $CD8^+$ $CCL4^+$ T cells in $ACPA^+$ eRA. Furthermore, we observed an enrichment of IFN-γ response genes in nearly all monocytes and T cells of $ACPA^+$ eRA subsets. Heightened interactions between IFN-γ and IFN-γ receptors were observed in $ACPA^+$ eRA, particularly between monocytes and T cells. We examined *IFITM2* and *IFITM3* as potential key markers in $ACPA^+$ eRA given their pronounced upregulation and association with the IFN response. Specifically, the expression of these genes was elevated in $IL1B^+$ proinflammatory monocytes (likely M1 monocytes), correlating with serum IFN-γ levels.

Discussion: Compared to ACPA$^-$ eRA, ACPA$^+$ eRA showed higher serum IFN-γ and IL-12 levels, upregulated IFN-γ response genes, and enhanced IFN-γ-driven monocyte-T cell interactions. These distinct immune features of the peripheral circulation in ACPA$^+$ eRA suggest a role for type 1 helper T cell-related immunity in its pathogenesis.

KEYWORDS

single-cell transcriptomics, peripheral blood mononuclear cells, anti-citrullinated peptide antibody, rheumatoid arthritis, rheumatoid arthritis pathogenesis, Th1 immunity, interferon signature, IFITM2/3

Introduction

Rheumatoid arthritis (RA) is a chronic autoimmune disease characterized by inflammation of the joint synovium (1). It results from complex interplays of synovial T cells, B cells, macrophages, dendritic cells (DCs), and fibroblast-like synoviocytes (FLSs) leading to joint destruction via autoreactive antibodies, chemokines, and pathogenic cytokines (2). Over recent decades, targeted biologics against lymphocytes and key pathogenic cytokines have achieved great success in managing RA (1). Nevertheless, 6% of patients with RA in Japan and 10% in the United Kingdom are refractory to these therapies, highlighting the limitations of the current treatment strategies (3, 4).

RA is a heterogeneous disease with various endo-phenotypes, for which personalized medicine is desirable (5, 6). Although a personalized approach has not been established in RA, there have been efforts to guide therapy using the anti-citrullinated peptide antibody (ACPA), the most commonly used diagnostic and prognostic biomarker (7). Clinical studies have suggested a better response to Abatacept than a tumor necrosis factor inhibitor in ACPA$^+$ RA (8). Moreover, longer drug retention of a Janus kinase inhibitor (JAK) was observed in ACPA$^+$ RA than in ACPA$^-$ RA (9). Together, these earlier reports indicate that the presence or absence of ACPA could significantly shape the most effective treatment strategy for RA, underscoring the importance of a patient-centric approach to RA treatment by considering each patient's ACPA status.

To utilize ACPA as a biomarker in guiding treatment, it is crucial to comprehend the immunologic difference according to the presence of ACPA. Previously known, immune complexes of ACPAs and citrullinated peptides can promote pro-inflammatory reactions of macrophages through binding to Fc receptors (10). Antibodies against mutated citrullinated vimentin, a highly specific ACPA for RA, can activate osteoclastogenesis and bone resorption (11). Transcriptome analysis has revealed that chemokine profiles of myeloid cells are altered and cytotoxic properties of T cells are differentially upregulated in ACPA$^+$ RA (12). Despite previous investigations, we have insufficient understanding of the differentiating immunologic characteristics between endotypes of ACPA$^+$ and ACPA$^-$ RA.

Interferon-gamma (IFN-γ) holds significant interest in the field of autoimmunity research due to its influential role in promoting and regulating inflammation (13). This aspect is particularly critical in the context of RA, where the dominant pathogenic cells are type 1 helper T (Th1) cells, known for their primary production of IFN-γ (2, 13, 14). A number of studies have documented elevated levels of IFN-γ in patients with RA, as well as in mouse models of autoimmune arthritis (15–17). JAK inhibitor targeting the IFN pathway, as well as other pathogenic cytokines, have shown excellent efficacy in the treatment of RA (18). Considering the highly heterogeneous nature of RA, an in-depth understanding of the level of IFN-γ expression in individual RA patients is needed.

Here, to gain insight into the immunological background for tailored medicine, we aimed to comparatively study immunologic characteristics according to ACPA status. To this end, we performed multiplex cytokine assay (MCA) demonstrating increased serum levels of Th1 cell-related cytokines, specifically IL-12 and IFN-γ, in ACPA$^+$ early RA (eRA). Subsequently, we performed single-cell RNA sequencing (scRNA-seq) analysis of peripheral blood mononuclear cells (PBMCs) from eRA patients and then compared gene expression and cell-cell interaction patterns between ACPA$^+$ and ACPA$^-$ eRA. It revealed that interferon response genes (IRGs), particularly *IFITM2* and *IFITM3*, were distinctly upregulated in monocytes and T cells of ACPA$^+$ eRA compared to those of ACPA$^-$ eRA. Such upregulation in ACPA$^+$ eRA might have resulted from Th1-skewed antigen-specific T-cell immunity and its related activation of monocytes involved in RA. Furthermore, we found a positive correlation between expression levels of the major IRGs in monocytes and T cells and levels of serum IL-6 and IFN-γ. Collectively, these findings provide novel insights into the immuno-pathogenic mechanisms underlying RA, potentially contributing to the development of more effective, personalized treatments for this complex disease.

Materials and methods

Patient recruitment and sample processing

Untreated (no current or prior use of glucocorticoids or disease-modifying anti-rheumatic drugs) patients with early and active RA who met the 2010 ACR/EULAR RA classification criteria (19) were recruited from Seoul St. Mary's Hospital in Korea. Unclassified arthritis patients were recruited based on the following inclusion criteria: (1) at least one swollen joint in the wrists or hands; (2) negative result for ACPA; (3) symptom duration of less than 12 months. The exclusion criteria for unclassified arthritis were: (1) meeting the 2010 ACR/EULAR RA classification criteria (19); (2) presence of other connective tissue diseases; (3) acute trauma; and (4) current or previous use of glucocorticoids or disease-modifying anti-rheumatic drugs (20, 21). Healthy volunteers were also recruited as controls. Peripheral blood samples were obtained from the participants for scRNA-seq and cytokine assay at the time of recruitment. Patient information, including demographic profile, laboratory markers, and disease activity scores, was collected at the time of blood sampling (Supplementary Table S1). Peripheral blood mononuclear cells were isolated using Ficoll-Paque gradient centrifugation. Cell quantity and viability were then determined by Trypan Blue staining. This study was approved by the Institutional Review Board of Seoul St Mary's Hospital (approval number: KC14TIMI0697). All participants provided written informed consent.

Multiplex cytokine assay

Concentrations of IFN-γ, IL-12 and IL-6, in serum samples of eRA patients were measured from using Millipore's MILLIPLEX MAP High Sensitivity Human Cytokine multiplex kit (cat. no. HSTCMAG-28SK; Merck, Billerica, MA, USA) according to the manufacturer's instructions. The minimum detection limits for the MCA were established at 0.61 pg/mL for IFN-γ, 0.49 pg/mL for IL-12, and 0.18 pg/mL for IL-6.

Single cell preparation and multiplexing individual samples for scRNA-seq

Cell stocks were thawed in 37°C 10% FBS/DMEM. The samples were washed twice with cold, Ca^{2+}- and Mg^{2+}-free 0.04% BSA/PBS at 300 × g for 5 min at 4°C. They were then gently resuspended in cold staining buffer (BD Biosciences, catalog no. 554656) and counted using a LUNA-FX7 Automated Fluorescence Cell Counter (Logos Biosystems) with AO/PI staining. To multiplex the samples, each sample was tagged with an antibody-polyadenylated DNA barcode specific for human cells (BD Biosciences, catalog no. 633781). Briefly, the cells were stained with the multiplexing antibody for 20 min at room temperature, followed by three washes with staining buffer (BD Biosciences, catalog no. 554656). After the final wash, the samples were gently resuspended in cold Sample Buffer (BD Biosciences, catalog no. 664887), counted using a LUNA-FX7 Automated Fluorescence Cell Counter (Logos Biosystems), and pooled.

Single cell capture and cDNA synthesis

Single-cell capture was performed using a BD Rhapsody Express instrument according to the manufacturer's instructions (BD Biosciences). Briefly, pooled cells from each sample were suspended in cold sample buffer and loaded into a BD Rhapsody cartridge (BD Biosciences, catalog no. 633731). After cell separation, cell-barcode magnetic beads were added to the cartridge. The cells were then lysed and the mRNA capture beads were retrieved. cDNA synthesis and Exonuclease I treatment were performed on the mRNA capture beads using a BD Rhapsody cDNA Kit (BD Biosciences, catalog no. 633773).

Library preparation and scRNA-seq

According to the 'mRNA Whole Transcriptome Analysis (WTA) and Sample Tag Library Preparation' protocol, scRNA-seq libraries were constructed using the BD Rhapsody WTA amplification kit (BD Biosciences, catalog no. 633801). For the WTA library, cDNA was sequentially subjected to random priming and extension (RPE), RPE amplification, and index PCR. For the sample tag library, cDNA was sequentially subjected to nested PCR (PCR 1 and PCR 2) and index PCR. The purified WTA and sample tag libraries were quantified using qPCR according to the qPCR Quantification Protocol Guide (KAPA) and assessed using the 4200 TapeStation System (Agilent Technologies, catalog no. 5067-4626). The libraries were sequenced using the HiSeq platform (Illumina).

Preprocessing of sequencing data

The raw sequencing data were processed using the BD Rhapsody WTA Analysis Pipeline v1.8 (BD Biosciences) and aligned against the human reference genome (GRCh38) obtained from the Ensembl database. The resulting gene expression matrices were converted to individual Seurat objects using the Seurat package in R (v3.8.0) (22). For each object, we filtered data based on the number of unique molecular identifiers (UMIs) and the number of genes detected. The genes that were expressed in at least five cells, and cells with gene detection between 500 to 2000 were retained. The filtered objects were normalized and their variance stabilized using the SCTransform function of Seurat. We reduced batch effects and performed combined analysis by integrating individual Seurat objects from various batches using the FindIntegrationAnchors and IntegrateData functions in Seurat. We addressed batch effects through a confirmation process as shown in Supplementary Figure S1.

Dimension reduction and major cell type annotation

The number of UMIs, percentage of mitochondrial genes, and cell cycle genes were regressed out, and genes were scaled to unit

variance. Principle component analysis (PCA) was performed. Clusters were then identified using UMAP. Cell identity was assigned using known cell markers shown in Supplementary Figure S2. We compared gene expression levels between cells in the cluster and those in all the other clusters to determine cluster marker genes. Clusters were manually annotated based on known marker genes. Thereafter, we validated annotations by referring to results from "seurat_annotation," "human cell atlas," and "Z_annotation" (Supplementary Table S2) (23). Adjacent clusters were merged if they were regarded as identical entities according to the similarity of transcriptomes.

Detection of differentially expressed genes and pathway analysis

Differential gene expression testing was performed using the 'FindMarkers' function within Seurat, employing the Wilcoxon test. All *p*-values were adjusted using Bonferroni correction. Differentially expressed genes (DEGs) were filtered using a minimum log2(fold change) of 0.5 and a maximum adjusted *p*-value of 0.05. They were then ranked by average log2(fold change) and false discovery rate (FDR). Enrichment analysis for functions of the DEGs was conducted using the clusterProfiler package and DAVID (https://david.ncifcrf.gov/) (24). Gene sets were based on Gene Ontology terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

Analysis of cell-cell interaction

To comprehensively analyze cell-to-cell interactions between immune cells, we used SingleCellSignalR (25). We derived potential ligand-receptor interactions based on the expression of a receptor by one cell subpopulation and ligand expression by another. We separately fetched normalized counts from healthy controls, $ACPA^+$, and $ACPA^-$ eRA patients and used them as input for the algorithm. To validate the cell-to-cell interactions and ligand-receptor interaction result from SinglCellSignalR, we performed the same analysis with CellChat (v.1.0) (26) and CellphoneDB (v.4.0) (27).

Inferring differentially expressed transcription factors

To determine the relationship between IFN signaling activity and anti-CCP antibodies, we used transcriptomics data to estimate the overall expression of IFN signaling genes for each sample. The decoupleR (v1.6.0) 'wmean' method and 'SCTransform' normalized data were used to calculate normalized gene expression levels of IFNA, IFNG, IFNAR1, IFANR2, IFNGR1, and IFNGR2 per cell for each unfiltered slide (28).

Chromatin binding profiles

We searched for chromatin binding sites of IFITM2 and IFITM3 based on chromatin binding profiles provided by ReMap 2022. Detected transcription factors (TFs), which were matched to upregulated TFs derived from master regulator analysis, were visualized with Integrative Genomics Viewer (IGV) (29).

Data visualization

All plots were generated using the ggplot2 (v3.2.1), pheatmap (v1.0.12), and EnhancedVolcano (v1.2.0) packages in R v4.0.0. Box plots are defined as follows: the middle line corresponds to the median; lower and upper hinges correspond to the first and third quartiles, respectively; the upper whisker extends from the hinge to the largest value, reaching no more than 1.5× the interquartile range (or the distance between the first and third quartiles) from the hinge; and the lower whisker extends from the hinge to the smallest value, not exceeding 1.5× the interquartile range from the hinge. Data beyond the end of whiskers were designated as "outliers". They were plotted individually.

Results

Serum cytokine profiles of $ACPA^+$ and $ACPA^-$ eRA

To elucidate differences between $ACPA^+$ eRA patients and $ACPA^-$ eRA patients, we recruited 37 eRA patients, 16 unclassified arthritis patients, and 21 healthy participants. Based on the experimental design presented in Figure 1A, their serum cytokine levels and transcriptome were examined using MCA and scRNA-seq, respectively. Among eRA patients, the $ACPA^+$ eRA group displayed elevated serum levels of IFN-γ and IL-12, the hallmark cytokines of type 2 interferon signaling, in comparison with the $ACPA^-$ eRA group (Figures 1B, C). There was a strong positive correlation between the two cytokine levels (Figure 1D). These findings highlight that increased serum IFN-γ concentrations are closely related to the seropositivity of RA.

Single-cell RNA-seq analysis landscape of eRA PBMCs

We next conducted scRNA-seq to characterize and compare transcriptome profiles of $CD45^+$ cells from PBMCs obtained from healthy controls (n = 4), drug-naïve $ACPA^-$ eRA patients (n = 6), and drug-naïve $ACPA^+$ eRA patients (n = 6). We initially addressed batch effects through a QC confirmation process (Supplementary Figure S1) and subsequently performed filtering procedures. A total of 37,318 immune cells were analyzed and segregated into 21

distinct clusters based on their transcriptomic profiles (Figures 2A, B). To identify the cell types within each cluster, we analyzed expression levels of marker genes (*CD14*, *MS4A1*, *CD3E*, *FCGR3A*, *FCER1A*, *CD8A*, *PCNA*, *CD38*, and *CD4*). The results are presented in Figure 2C and Supplementary Figure S2A. Those levels were cross-referenced with immune cell data predicted from singleR (Supplementary Figure S2B). Additionally, we consulted canonical cell marker expression (Supplementary Figure 3), Human Cell Atlas annotations, and gene sets extracted from publicly available scRNA-seq data on RA synovial cells (https://www.immport.org/shared/study/SDY998) (23) (Supplementary Table S2). As a result, we finally identified 17 unique cell types within 21 clusters.

A Uniform Manifold Approximation and Projection (UMAP) plot demonstrated six clusters for T cells, four clusters for monocytes, three clusters for B cells/plasmablasts, two clusters for dendritic cells, one cluster for natural killer (NK) cells, and one cluster for progenitor cells (Figures 2A, D, E). Proportions of B cells, dendritic cells, monocytes, and T cells were comparable between ACPA$^-$ RA and ACPA$^+$ eRA (Supplementary Figure S2C). However, we observed a substantial expansion of IL1B$^+$ proinflammatory monocytes, CD8$^+$ CCL4$^+$ T cells, and IL7R$^+$ T cells in eRA patients compared to healthy controls. Furthermore, these three subsets were significantly frequent in ACPA$^+$ eRA than in ACPA$^-$ eRA (Figures 2F–H).

Conclusively, through global transcriptome profiling, we identified 17 unique cell types in human PBMCs, including three subsets of immune cells, IL1B$^+$ proinflammatory monocytes, CD8$^+$ CCL4$^+$ T cells, and IL7R$^+$ T cells, presumably representing the peripheral landscape of immuno-pathology of ACPA$^+$ eRA.

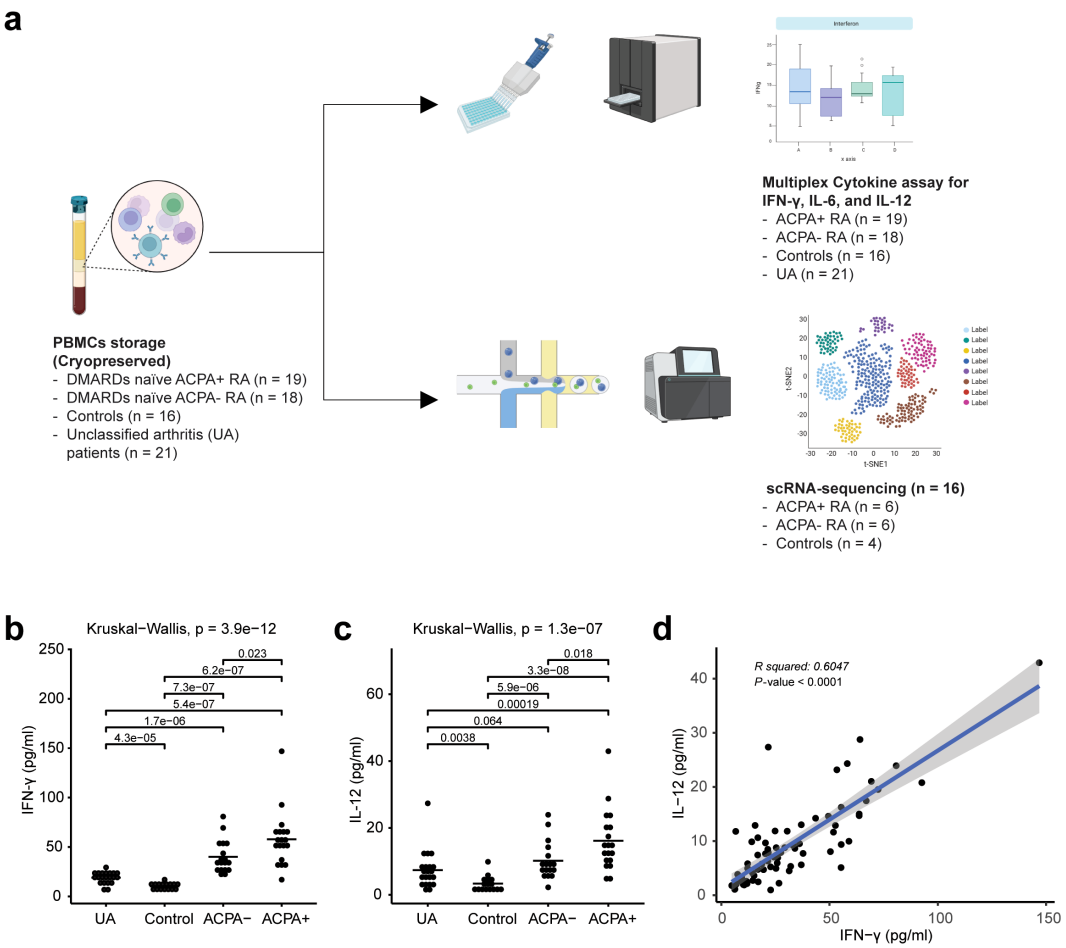


FIGURE 1
Overall study design and multiplex cytokine analysis of rheumatoid arthritis (RA) peripheral blood mononuclear cells (PBMCs). **(A)** Workflow chart outlining the overall study design, illustrating each step from patient selection to data analysis, and highlighting the methodologies used in the extraction and analysis of PBMCs from unclassified arthritis (UA), healthy controls, ACPA-negative early RA (ACPA$^-$ eRA), and ACPA-positive early RA (ACPA$^+$ eRA). **(B)** Dot plot showing the serum concentration levels of interferon-γ (IFN-γ) in UA (n = 21), controls (n = 16), ACPA$^-$ eRA (n = 18), and ACPA$^+$ eRA (n = 19). The horizontal bar indicates the mean value. **(C)** Dot plot showing the serum concentration levels of interleukin-12 (IL-12) in UA (n = 21), controls (n = 16), ACPA$^-$ eRA (n = 17), and ACPA$^+$ eRA (n = 19). The horizontal bar represents the mean value. **(D)** Scatter plot illustrating the correlation between IFN-γ and IL-12 concentrations in serum samples. Statistical significance was assessed using the Kruskal–Wallis test for **(B, C)** and Pearson's correlation coefficient for **(D)**. P-values less than 0.05 were considered significant.

10.3389/fimmu.2024.1439082

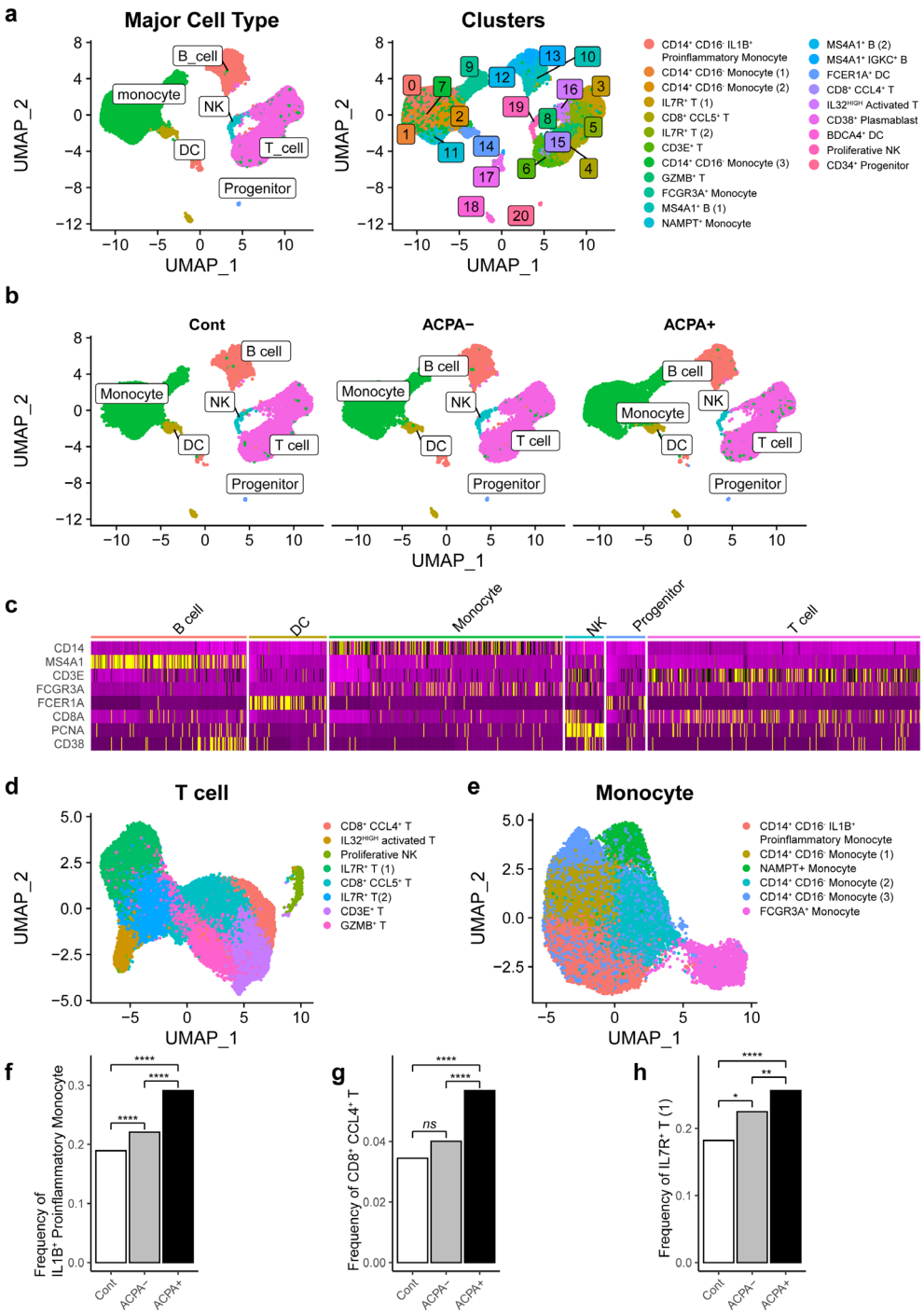


FIGURE 2
Cellular fractions of PBMCs reconstructed from single-cell RNA-sequencing (scRNA-seq) analysis in RA patients. **(A)** Uniform Manifold Approximation and Projection (UMAP) based on scRNA-seq data from PBMCs of drug naïve RA patients (n =12) and age-sex matched healthy controls (n = 4). A total of 37,318 cells were classified into 21 clusters (left panel) and differentiated into major cell types, including monocytes, dendritic cells (DCs), T cells, B cells, natural killer (NK) cells, and progenitors. **(B)** UMAP illustrates the distribution of the main cell types in PBMCs of control subjects, ACPA$^-$ RA patients, and ACPA$^+$ RA patients. **(C)** Heatmap representing marker gene expression patterns for major cell types across different clusters, providing a detailed view of gene expression signatures characteristic of each cell type. **(D, E)** Subsequent UMAP plots denoting sub-clusters and specific cell types for main cellular subsets, focusing on T cells **(D)** and monocytes **(E)**. **(F-H)** Bar graphs showcasing the relative proportion of specific cell subsets in PBMCs, including IL1B$^+$ proinflammatory monocytes **(F)**, CD8$^+$ CCL4$^+$ T cells **(G)**, and IL7R$^+$ T cells **(H)**, in control, ACPA$^-$ RA, and ACPA$^+$ RA group. P-values from z-tests for population proportions indicate significant cell type expansions. *P-value<0.05, **P-value<0.01, and ****P-value<0.0001.

10.3389/fimmu.2024.1439082

Increased expression of interferon response genes in ACPA$^+$ eRA

To gain a deeper understanding of the alterations in gene expression linked to ACPA positivity and to explore the underlying mechanisms of RA associated with these changes, we next analyzed DEGs in each cell type of PBMCs from ACPA$^+$ and ACPA$^-$ eRA patients. Given the multiplex cytokine data in Figure 1, we sought to focus on the IFN-γ and IL-12-JAK pathway for the analysis. The top 20 DEGs from the scRNA-seq analysis are listed, which included IRGs such as *IER3*, *JUNB*, and *IFITM2*, and *IFITM3*, and among them, *IFITM3* showed nearly the highest fold change (Table 1). Notably, volcano plots of cell subsets demonstrated that differential expression of *IFITM3* was mainly observed in monocytes and T cells, not in B cells (Figure 3A).

To compare functional characteristics of each cell subset between ACPA$^+$ and ACPA$^-$ eRA, we also performed Gene Set Enrichment Analysis (GSEA) using Hallmark gene sets provided by MsigDB. The results indicate that all monocyte subsets of ACPA$^+$ eRA had higher transcriptional profiles for "IFN-γ response" and "IFN-α response" than those of ACPA$^-$ eRA (Figure 3B). Most T cell subsets, with the exception of IL32^{high} T cells, exhibited strong enrichment for these IFN responses (Figures 3B, C). In particular, in ACPA$^+$ eRA, the three cell subsets-IL1B$^+$ proinflammatory monocytes, CD8$^+$ CCL4$^+$ T cells, and IL7R$^+$ T cells-exhibited increased expansion with heightened enrichment profiles for IFN-γ and IFN-α responses compared to those in ACPA$^-$ eRA (Figure 3C).

The hierarchical clustering and heat map analysis illustrate that upregulated IRGs in monocytes were different from those in T cells, indicating that transcriptional responses to IFN are different according to cell type (Figures 3D, E). Through pseudo-bulk analysis of scRNA-seq data, we also observed the increases in JAK-STAT pathway-related genes in IL7R$^+$ T cells and IL1B$^+$ proinflammatory monocytes of ACPA$^+$ eRA patients as compared to those of ACPA$^-$ eRA patients (Supplementary Figures S4A, B, D, E). Subsequent GSEA analysis revealed upregulation of the genes associated with the IL2-STAT5 signaling pathway (Supplementary Figure S4C). Given that IL-2 plays a role in Th1 differentiation by inducing the expression of IL-12 receptor and T-bet in a STAT5-dependent manner (30), this finding supports the notion that ACPA$^+$ eRA has upregulated Th1 immunity. Interestingly, there was a strong positive correlation between ESR levels and IFN-γ signature genes in RA, which were obtained from the previously published data (Supplementary Figures S5A, B) (31–33). In summary, ACPA$^+$ eRA showed increased activity of the IFN-JAK-STAT pathway as compared to ACPA$^-$ eRA, which was more prominent in the cell types of IL1B$^+$ proinflammatory monocytes and IL7R$^+$ T cells.

TABLE 1 The top 20 differentially expressed genes between ACPA$^-$ and ACPA$^+$ from PBMC scRNA-seq.

	p-value	avg_log2FC	pct.1	pct.2	*p*-value_(adj)
HLA-DQA2	7.97E-266	1.892949811	0.132	0.037	1.29E-261
IFITM3	1.79E-205	0.965439862	0.407	0.289	2.91E-201
HLA-DRB5	7.71E-191	0.729622267	0.313	0.189	1.25E-186
JUNB	6.23E-163	0.339508966	0.861	0.804	1.01E-158
IFITM2	6.14E-135	0.467551747	0.58	0.482	9.96E-131
ERAP2	3.73E-122	0.663834203	0.258	0.166	6.05E-118
BTG2	9.67E-112	0.405070841	0.605	0.519	1.57E-107
IFI44L	4.05E-110	0.903485471	0.179	0.105	6.57E-106
TNFSF10	3.71E-108	0.818914251	0.199	0.123	6.02E-104
IFI6	6.41E-108	0.789436412	0.311	0.227	1.04E-103
FOSB	7.27E-106	0.362939434	0.737	0.664	1.18E-101
IER3	1.07E-96	0.805034193	0.38	0.304	1.73E-92
XAF1	2.85E-96	0.612964022	0.313	0.23	4.62E-92
MNDA	4.82E-90	0.601017761	0.38	0.303	7.82E-86
HLA-C	6.54E-85	0.171143633	0.99	0.984	1.06E-80
FOS	1.16E-84	0.274752943	0.933	0.907	1.89E-80
TSC22D3	6.51E-84	0.370209625	0.56	0.481	1.06E-79
TNFAIP3	1.29E-79	0.383694023	0.657	0.589	2.09E-75
IER2	8.44E-75	0.358348283	0.567	0.498	1.37E-70
IFIT3	2.25E-73	1.163609645	0.075	0.035	3.64E-69

10.3389/fimmu.2024.1439082

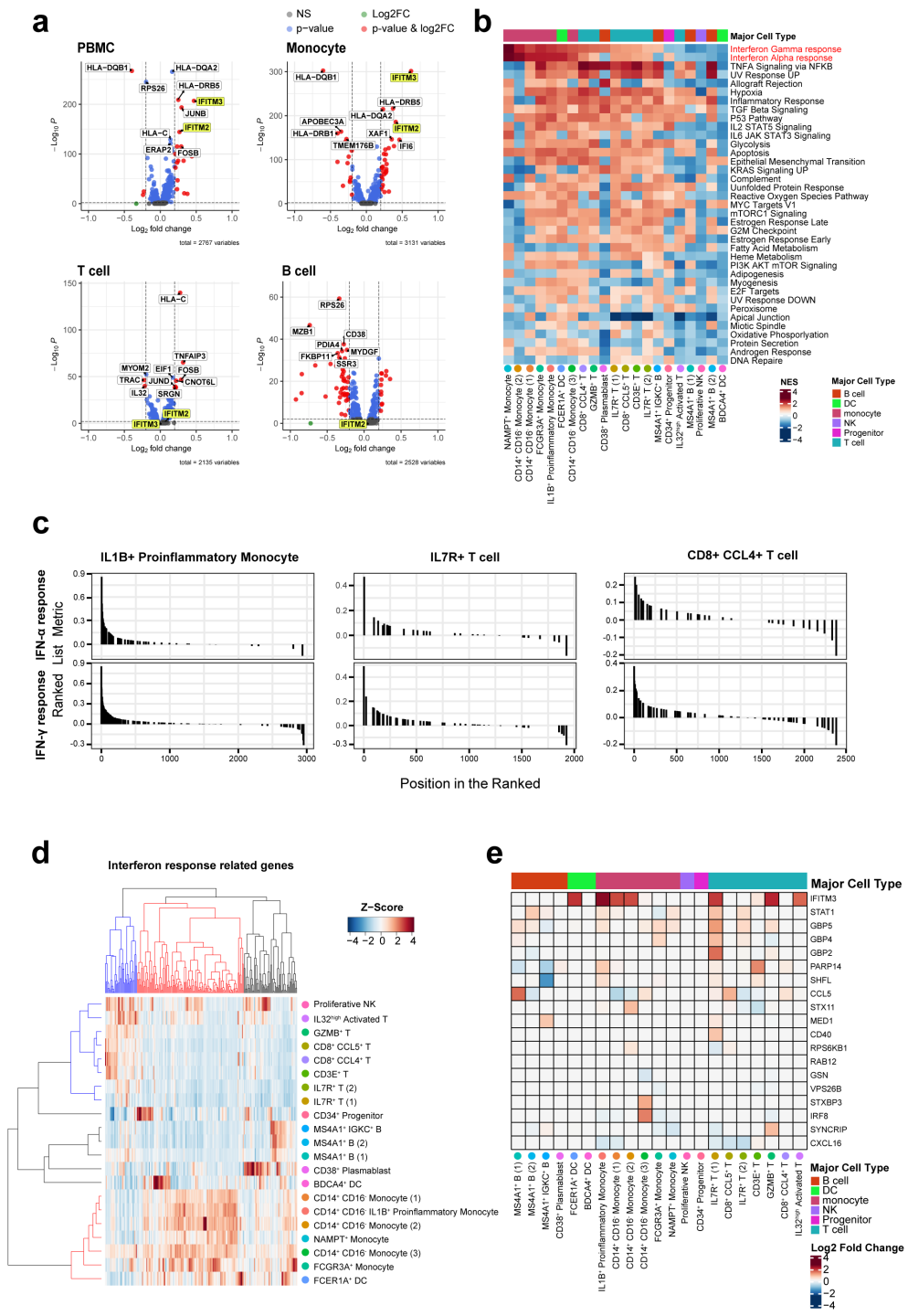


FIGURE 3
Differentially expressed gene sets in each cell subset based on gene ontology terms and Kyoto Encyclopedia of Genes and Genomes. **(A)** Volcano plot comparing ACPA$^+$ and ACPA$^-$ group for PBMCs, T cells, B cells, and monocytes, revealing differentially expressed genes. IFITM2/3 are highlighted in yellow. **(B)** Heat map displaying gene set enrichment analysis of genes with changing expression levels in ACPA$^-$ and ACPA$^+$ across 17 cell types. Column annotations on the heatmap show major cell types. Red color indicates the enrichment score increased in ACPA$^+$ RA, and blue color indicates an increased enrichment score in ACPA$^-$ RA. **(C)** GSEA plots of IFN-α and IFN-γ response gene sets in the three subsets expanded in ACPA$^+$ patients, including IL1B$^+$ proinflammatory monocytes, CD8$^+$ CCL4$^+$ T cells, and IL7R$^+$ T cells. **(D, E)** Heatmap representing differences in expression of genes related to IFN-γ by cell type. The heatmap in **(D)** shows the expression patterns of IFN-α and IFN-γ response genes in each of the 21 cell clusters. In the dendrogram of both columns and rows, blue represents genes highly expressed in T cells, while red denotes those highly expressed in monocytes. The color matrix represents the Z-score of relative IRG expression levels within cells, where red indicates high expression and blue indicates low expression. The heatmap in **(E)** shows the log2 scaled fold change values of 19 IFN-γ response genes in ACPA$^+$ patients compared to those in ACPA$^-$ patients. The highly expressed genes in monocytes were selected for this heatmap. Red indicates an increase in ACPA$^+$ patients, while blue indicates a decrease.

Upregulation of interferon-activated transcription factors in ACPA$^+$ eRA

IFN-γ is primarily produced by Th1 cells. It is a critical activator of immune response, promoting the killing of intracellular microbes by macrophages and dendritic cells (13, 34). In RA patients, IFN-γ is known to be produced due to Th1 skewing (14). Here, we focused on identifying TFs induced by IFN-γ signaling, including signal activators of transcription (STATs) and interferon regulatory factor (IRFs) (34, 35), in T cells and monocytes of PBMCs obtained from ACPA$^+$ eRA versus ACPA$^+$ eRA patients. To address this, we performed a master regulator analysis using VIPER score, which enabled us to identify differentially activated TFs in each cell type between healthy controls, ACPA$^+$, and ACPA$^-$ eRA. As a result, we found that IL1B$^+$ proinflammatory monocytes and CD14$^+$ CD16$^-$ monocytes exhibited higher STAT1, 2, and 3 transcriptional activities in ACPA$^+$ eRA patients than in ACPA$^+$ eRA (Figure 4A). Moreover, IL1B$^+$ proinflammatory monocytes, NAMPT$^+$ monocytes, GZMB$^+$ T cells, IL32^{high} T cells, and IL7R$^+$ T cells in ACPA$^+$ eRA exhibited higher STAT3 activity than other cell subsets (Figure 4A).

We further investigated the upstream reactions governing IRGs by quantitatively assessing the interaction of IFN-γ with its receptors using ligand-receptor (LR) scores. Our results indicate that the LR scores for IFN-γ-IFN-γ receptors were significantly higher in ACPA$^+$ eRA than in ACPA$^-$ eRA, as seen in the dot plot of LR scores above 0.5 (Figure 4B). Based on this interaction data, we next compared cell-to-cell interactions in ACPA$^-$ eRA versus ACPA$^+$ eRA, focusing on monocytes and T cell interaction. As shown in Figure 4C, the total number of IFN-γ-IFN-γ receptor interactions within diverse monocyte and T cell subsets was substantially higher in ACPA$^+$ eRA than in ACPA$^-$ eRA. Most strikingly, the interaction direction was entirely from T cells (→) to monocytes in ACPA$^-$ eRA. In a sharp contrast, there were bidirectional and even multidirectional interactions between T cells and monocytes in ACPA$^+$ eRA (Figure 4C, lower panel). Particularly, many interactions from monocytes (→) to T cells, in addition to those from T cells (→) to monocytes, were observed in ACPA$^+$ eRA, and they were primarily driven by a subset of CD14$^+$ CD16$^-$ monocytes (*See* green arrows in the upper panel of Figure 4C). Moreover, we also detected interaction between CD4$^+$ and CD8$^+$ T cells, as well as between different subsets of monocytes in ACPA$^+$ eRA, which were rarely found in ACPA$^-$ eRA. To avoid any differences between the algorithms and sources of LR interaction information, we performed a cell-to-cell interaction analysis using CellPhoneDB (v.4.0) (27) and obtained the same results as previously obtained (Supplementary Figure S6). The cell-to-cell communications between CD4$^+$, CD8$^+$ T cells and monocytes were also detected in the analysis using CellChat (26) (Figure 4D).

Taken together, we observed the elevated STAT activity in IL1B$^+$ proinflammatory monocytes in ACPA$^+$ eRA, which seems to be associated with more interactions between IFN-γ-IFN-γ receptors. Notably, there were discernible patterns in monocyte and T cell interactions that appear to distinguish ACPA$^+$ eRA (bidirectional) from ACPA$^-$ eRA (unidirectional).

Association of IFITM2/IFITM3 expression with STAT3, IL-6, and IFN-γ level in eRA patients

It is well known that IFITM2 and IFITM3 are induced by interferon stimulation (36, 37). As seen in Table 1, *IFITM2* and *IFITM3* were found to be the top 5 DEGs as IRGs (Table 1). Elevated levels of *IFITM2* and *IFITM3* were observed in the monocytes of ACPA$^+$ eRA patients, especially in IL1B$^+$ proinflammatory monocytes, which are presumably the M1 monocyte subset (Figures 5A, B). These findings have sparked our curiosity to explore further the regulatory mechanisms that control the transcription of IFITM2 and IFITM3 in ACPA$^+$ eRA. To address this, we searched for the chromatin binding profiles of *IFITM2* and *IFITM3* (Figure 6A) using the ReMap2022 database (38). We identified 27 binding regions for 13 TFs from public data produced by 11 independent studies (Supplementary Table S3). The results showed that *STAT3*, which is known as an IFN-activated TF (39), was identified as one of the regulatory TFs for *IFITM2* and *IFITM3* transcription. These results, together with the data in Figures 4A and B, suggest that increased activation of STAT3 is involved in regulation of IFITM2 and IFITM3 in ACPA$^+$ eRA.

Finally, we examined the relationships of *IFITM2* and *IFITM3* expression in monocytes and T cells with pathogenic cytokines of eRA (Figures 6B, C). We found a moderate positive correlation between serum IL-6 concentrations and expression levels of *IFITM2* in T/NK cells ($R = 0.65$, $p = 0.03$). The serum IL-6 level showed a strong positive association with expression levels of *IFITM3* in T/NK cells ($R = 0.77$, $p = 0.0051$) and *IFITM3* in monocytes ($R = 0.8$, $p = 0.0031$) (Figure 6B). Serum levels of IFN-γ had moderate to strong positive correlations with expression levels of *IFITM2* in T/NK cells ($R = 0.68$, $p = 0.02$), *IFITM3* in T/NK cells ($R = 0.77$, $p = 0.0052$), and *IFITM3* in monocytes ($R = 0.65$, $p = 0.031$) (Figure 6C).

Together, these data suggest that *IFITM2* and *IFITM3*, the highly upregulated IRGs in ACPA$^+$ eRA, are associated with *STAT3* activation and increased serum levels of IL-6 and IFN-γ.

Discussion

ACPA-based stratification is the most widely accepted method for classifying RA. Our study elaborated contrasting immunologic features depending on ACPA status by mainly investigating global transcriptome profile of RA PBMCs and serum cytokines. Here, we demonstrated that serum IFN-γ and IL-12 levels were higher in in ACPA$^+$ eRA than in ACPA$^-$ eRA and healthy controls, indicating a skewing towards a Th1 phenotype. Moreover, IL1B$^+$ proinflammatory monocytes (most strikingly), CD8$^+$ CCL4$^+$ T cells, and IL7R$^+$ T cells were expanded in ACPA$^+$ eRA. In

ACPA$^+$ eRA, most of monocyte and T cell subsets, including the three expanded subsets of ACPA$^+$ eRA, had upregulated expressions of IRGs related to IFN-γ and IFN-α. The observed elevation in the transcriptional activity of IRGs is supported by increased expression of STATs, the IFN-driven TFs, in monocytes of ACPA$^+$ eRA. Notably, IFN-γ and its receptor interaction between monocytes and T cells in ACPA$^+$ eRA was markedly increased and displayed a multi-directional pattern, contrasting with the unidirectional pattern observed in ACPA$^-$ eRA. Together, these findings suggest that IFN-mediated responses are overactive in ACPA$^+$ eRA. In support of this, *IFITM2* and *IFITM3* expression levels of in monocytes and T/NK cells had positive correlations with circulatory IL-6 and/or IFN-γ levels.

This study initially identified concurrent increases in serum levels of IFN-γ and IL-12, the Th1 cytokines, in ACPA$^+$ eRA compared to ACPA$^-$ eRA. Furthermore, our transcriptome analysis demonstrated an elevation of IRGs, suggesting substantial IFN-γ exposure in monocytes and T cells of ACPA$^+$ eRA. Consistent with this, recent research shows that ACPA$^+$ RA patients have a significantly higher count and proportion of

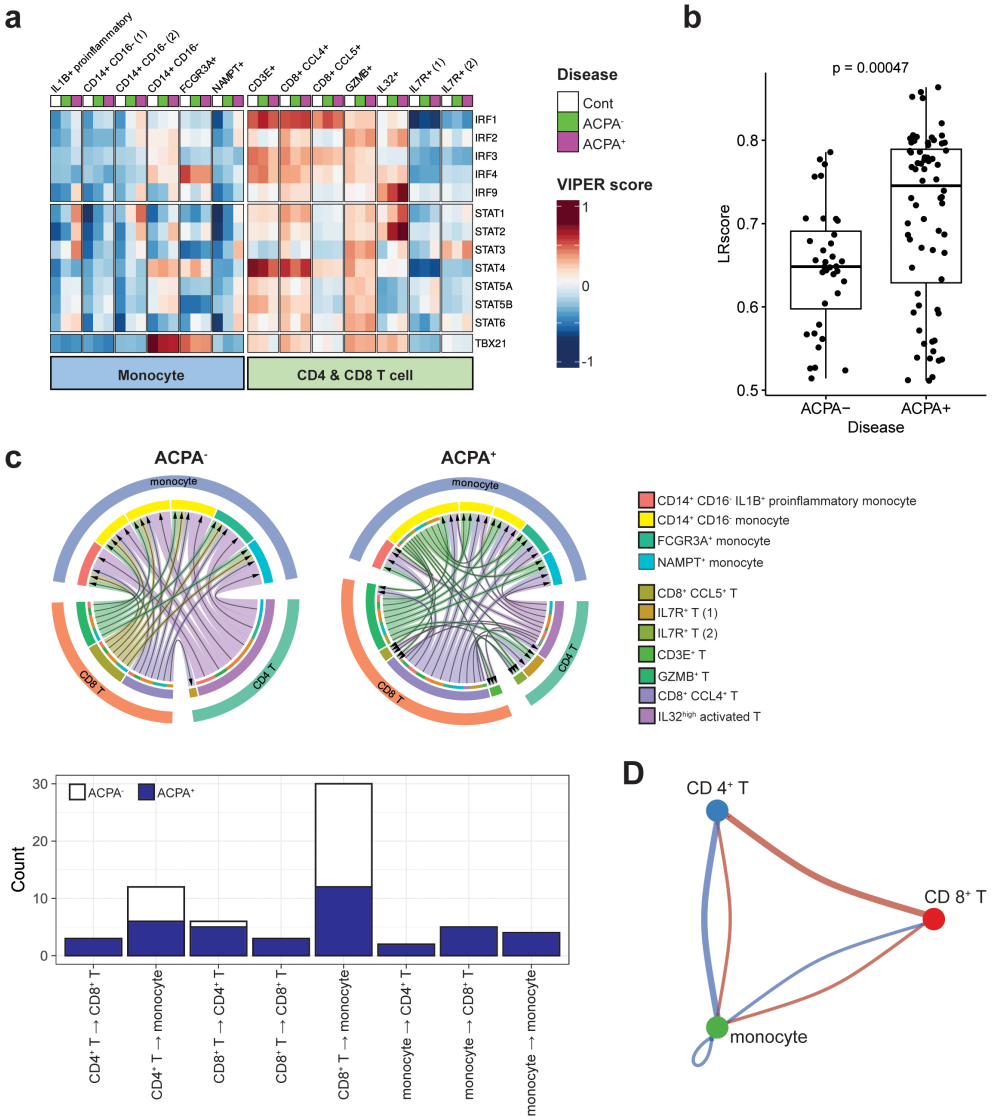
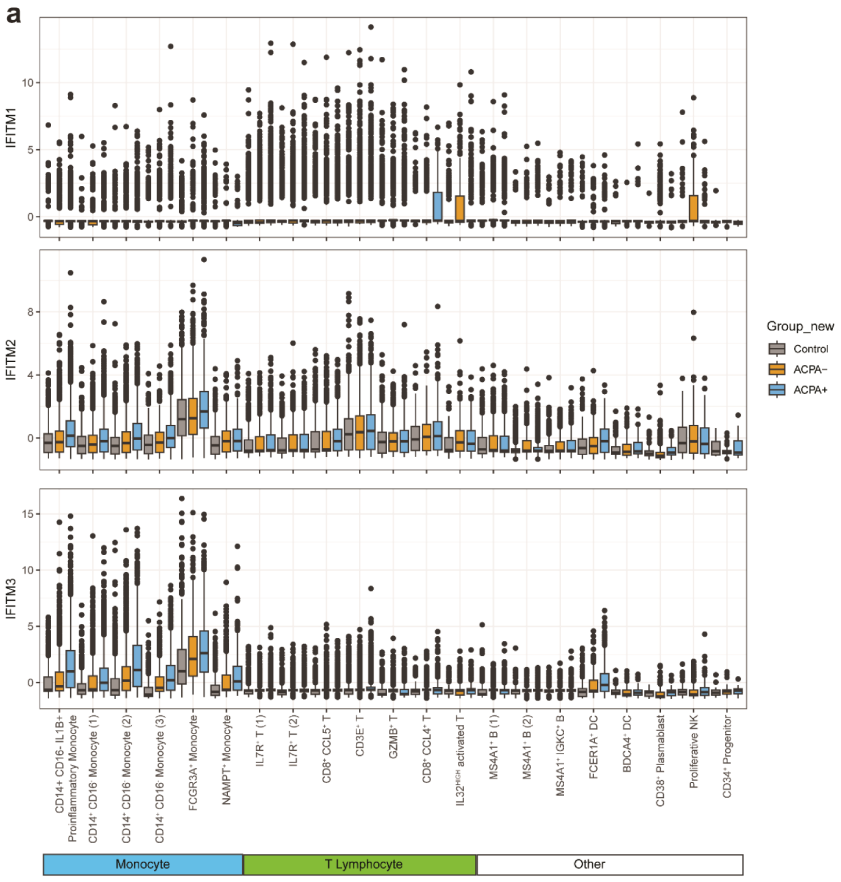


FIGURE 4
Downstream regulators of IFN-γ signaling and IFN-γ and its receptor interaction in monocytes and T cells of RA. **(A)** Heatmap displaying the activity scores of STAT and IRF family members predicted as downstream regulators of IFN-γ signaling. The scores were calculated by the DoRothEA package, a gene regulatory network containing signed transcription factor-target gene interactions. **(B)** Bar graph illustrating ligand-receptor (LR) scores for the interaction between IFN-γ (IFNG) and its receptor IFNGR1, derived from a ligand-receptor analysis. **(C)** Chord diagram illustrating variations in interactions between IFN-γ and IFNGR1 in monocytes and T cell subsets within ACPA$^+$ and ACPA$^-$ patient groups (upper panel). The outermost layer of the diagram represents the original major cell types, including monocytes, CD4+ T cells, and CD8+ T cells. Moving inward, the diagram shows the subsets of each cell type, which are color-coded according to the legend on the right side of the diagram. This color-coding is to facilitate the identification and differentiation of the various cell subsets within each primary cell category. The bar graph in the bottom panel depicts the directional variability of IFN-γ—IFN-γ receptor signaling in monocyte and T cell subsets from both ACPA$^+$ and ACPA$^-$ patients, showing the quantity of signals according to their signaling direction. **(D)** Cell-to-cell communication diagram depicts differential number of interactions between the three major cell types. Red indicates high in ACPA$^+$ and blue indicates low in ACPA$^-$.

circulating Th1 cells relative to ACPA$^-$ RA patients (40). James et al. reported that Th1 cells are the most abundant subset within CD4$^+$ T cells that are specific to citrullinated peptides in RA patients (41). Although we did not provide direct evidence of Th1 cell expansion in ACPA$^+$ eRA cells, we observed an increase in the number of IL7R$^+$ T cells. Notably, a large proportion of the IL7R^{high} CD62L^{low} T cells were Th1 cells (42). IL-7 potently stimulated IFN-γ production in synovial CD4$^+$ T cells, suggesting a link between IL7R$^+$ T cells and the Th1 response (43). Collectively, our findings not only support previous observations, but also underscore the critical role of Th1 immunity in the peripheral circulation of ACPA$^+$ eRA through a systemic approach.

Here, we found that IFN-γ response genes were upregulated in ACPA$^+$ eRA. IFN-γ both induces Th1 cell differentiation and is



b

Symbol	Major Cell Type	Cell Type	Average log2 Fold Change	Raw P-Value	Adjusted P-value	PCT 1	PCT 2
IFITM2	Monocyte	CD14+ CD16- IL1B+ Proinflammatory Monocyte	0.808	1.83E-60	2.97E-56	0.707	0.524
		CD14+ CD16- Monocyte	0.653	1.85E-21	3.00E-17	0.704	0.553
		CD14+ CD16- Monocyte (1)	0.664	1.92E-36	3.12E-32	0.602	0.457
		CD14+ CD16- Monocyte (2)	0.635	1.02E-24	1.66E-20	0.639	0.507
		FCGR3A+ Monocyte	0.339	1.30E-15	2.11E-11	0.931	0.876
		NAMPT+ Monocyte	0.310	2.16E-01	1.00E+00	0.558	0.524
	T cell	CD3E+ T	-0.005	8.77E-01	1.00E+00	0.679	0.7
		CD8+ CCL4+ T	0.225	5.47E-02	1.00E+00	0.642	0.598
		CD8+ CCL5+ T	0.204	6.53E-04	1.00E+00	0.532	0.478
		GZMB+ T	0.197	9.69E-03	1.00E+00	0.595	0.554
		IL32HIGH activated T	-0.116	4.11E-01	1.00E+00	0.506	0.551
		IL7R+ T (1)	0.166	8.80E-03	1.00E+00	0.38	0.339
		IL7R+ T (2)	0.194	3.00E-03	1.00E+00	0.453	0.402
IFITM3	Monocyte	CD14+ CD16- IL1B+ Proinflammatory Monocyte	1.305	7.59E-125	1.23E-120	0.728	0.465
		CD14+ CD16- Monocyte	1.031	8.55E-36	1.39E-31	0.658	0.462
		CD14+ CD16- Monocyte (1)	0.880	4.23E-50	6.86E-46	0.576	0.396
		CD14+ CD16- Monocyte (2)	0.975	9.52E-56	1.54E-51	0.741	0.569
		FCGR3A+ Monocyte	0.424	7.39E-19	1.20E-14	0.9	0.831
		NAMPT+ Monocyte	1.001	2.66E-07	4.31E-03	0.522	0.371
	T cell	CD3E+ T	0.044	9.27E-01	1.00E+00	0.223	0.226
		CD8+ CCL4+ T	-0.261	1.98E-01	1.00E+00	0.159	0.193
		CD8+ CCL5+ T	0.293	1.55E-02	1.00E+00	0.143	0.115
		GZMB+ T	0.036	7.84E-01	1.00E+00	0.167	0.163
		IL32HIGH activated T	0.151	6.61E-01	1.00E+00	0.16	0.146
		IL7R+ T (1)	0.232	1.15E-01	1.00E+00	0.113	0.097
		IL7R+ T (2)	0.341	2.13E-02	1.00E+00	0.121	0.095

FIGURE 5
Comparison of expressions of *IFITM1*, *IFITM2*, and *IFITM3* across different cell types in the control, ACPA$^-$ eRA, and ACPA$^+$ eRA groups. **(A)** A boxplot showing the expression of *IFITM1*, *IFITM2*, and *IFITM3* in each cell type of control subjects, ACPA$^-$ eRA patients, and ACPA$^+$ eRA patients. **(B)** A table showing the differences in expression of *IFITM2* and *IFITM3* in monocytes and T cells, which was analyzed across control subjects, ACPA$^-$ eRA patients, and ACPA$^+$ eRA patients. The meanings of each statistical value are as follows: Average log2 Fold Change: The log fold-change of the average expression between the two groups. Positive value indicates a higher expression in ACPA$^+$ eRA; Raw P-Value: The unadjusted *P*-value; Adjusted P-Value: The adjusted *P*-value, based on the Bonferroni correction using all features in the dataset. PCT1: The percentage of cells in which the feature is detected in the ACPA$^+$ eRA; PCT2: The percentage of cells in which the feature is detected in the ACPA$^-$ eRA.

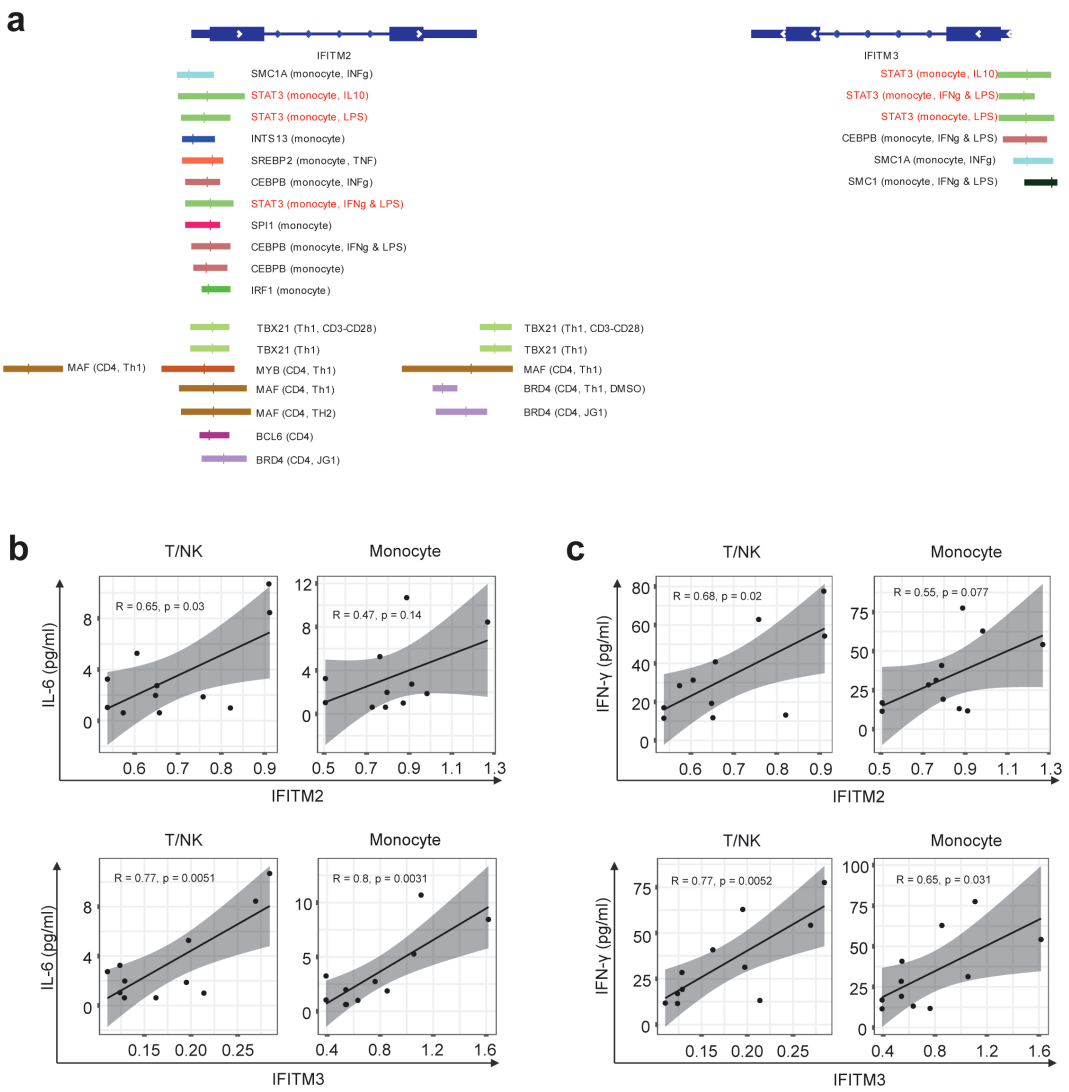


FIGURE 6
Differences in IFITM2/3 expression in monocytes and T cells of RA patients depending on the presence of anti-CCP antibodies. **(A)** Illustration depicting the transcription factors predicted to bind to the promoters of *IFITM2* and *IFITM3* genes, along with their specific binding locations. The binding sites for the same TF are indicated by the same color. The first parenthesis indicates which cell line it is from and the second indicates the treatment. The absence of these indicates the basal condition with no treatment. **(B, C)** Graphs correlating mRNA expression levels of *IFITM2* **(B)** and *IFITM3* **(C)** in T/NK cells and monocytes with the concentrations of IL-6 and interferons simultaneously measured in the serum of RA patients. The correlation coefficient (*R*) is indicated, and a linear regression line is fitted to the data points. Statistical significance of the correlation is assessed using Pearson's correlation coefficient. P-values less than 0.05 were considered significant.

subsequently secreted by them, initiating a cascade of immunological responses (44). IFN-γ drives the polarization of macrophages into the M1 subtype and enhances antigen presentation via MHC class II (14). IFN-γ also induces synovial fibroblasts to express MHC class II, significantly enhancing their interaction with citrullinated vimentin upon autophagy induction (45). Beyond its association with autoimmunity, IFN-γ is involved with synovial inflammation in RA. An omics study revealed a notable expansion of HLA-DRA^{hi} sublining fibroblasts, enriched with *HLA-DR*, *HLA-DP*, and *IFN-γ-inducible protein 30*, in the leukocyte-rich RA synovium (23). IFN-activated monocytes, another highly expanded subset of these tissues, exhibit elevated IRG transcription (23). IFN-γ has a complex role, as mice lacking it are more prone to arthritis (46) and an anti-IFN-γ antibody was ineffective in an RA trial (14), yet excess levels are likely to exacerbate autoimmune diseases (14). This study suggests that IFN-γ response genes are enriched in ACPA^{+} eRA and may serve as a therapeutic target specific to this population, as earlier findings support that excess of IFN-γ is likely pathogenic.

Type I IFN, known for their antiviral role, also contribute to autoimmunity through maladaptive lymphocyte activation (47). Our scRNA-seq analysis of drug-naïve RA PBMCs suggested that type I IFN response genes were upregulated in monocytes and T cells of ACPA^{+} eRA. An earlier study supported this finding by showing a positive correlation between type I IFN signature gene expression and ACPA levels in patients with RA (31). Consistent with this, persistent stimulation of type I IFN can trigger the enhancement of T and B cell effector functions, resulting in the

synthesis of autoantibodies (35, 47). IFN gene signatures and IFN-α levels are associated with RA disease activity (48). Additionally, IFN-activated monocytes are more abundant in leukocyte-rich synovial tissues compared to those with fewer leukocytes (23). Given the association of type I IFN with autoimmunity and RA severity, this study emphasizes the relevance of type I IFN in $ACPA^+$ eRA and its potential as a therapeutic target.

According to the Accelerating Medicines Partnership (AMP) publication, bulk-RNA seq using leukocyte-rich RA synovium shows upregulation of IL1B and CCL4 in monocytes and CD8 T cells, respectively (23). IL1B is regarded as a conventionally important pathogenic cytokine of RA (2). Thus, increased $IL1B^+$ monocytes support a more aggressive phenotype of $ACPA^+$ eRA. CCL4, also known as macrophage inflammatory protein 1-β (MIP-1β), is amplified in the joint tissues and peripheral circulation of patients with RA. This amplification of CCL4 facilitates the migration of inflammatory cells and osteoclasts, positioning it as a significant pathogenic chemokine in RA (49). Our scRNA-seq data on the higher proportion of $IL1B^+$ proinflammatory monocytes and $CD8^+$ $CCL4^+$ T cells in $ACPA^+$ eRA are compatible to the previous reports (23), which indicates that the peripheral scRNA-seq landscape may be a molecular reflection of immunologic dysregulation in synovial compartment of RA patients, suggesting a possible communication between the periphery and the joints in establishing RA pathology.

Of note, our scRNA-seq analysis revealed that *IFITM2* and *IFITM3* belonging to IRGs were included in the top 20 DEGs. The human genome encodes at least five IFITM proteins. In particular, IFITM1, IFITM2, and IFITM3 have antiviral activities by inhibiting viral entry into human cells and some other pathways (37). Both type I and II IFNs can increase the expression of IFITM1, IFITM2, and IFITM3. In mice, among those IFITM proteins, *Ifitm3* is most strongly induced by IFN (37). Despite the proven role of IFITMs in defense against viral infection and in the mouse system, little is known about their role in human RA. *IFITM3* is one of the marker genes of the IFN-activated monocyte subset in synovial tissues; however, how the expression of *IFITM3* affects RA pathogenesis remains unclear (23). Here, we found that expression levels of *IFITM3* and *IFITM2* in monocytes and T/NK cells had a positive linear relation with the serum levels of IL-6 and/or IFN-γ, suggesting a possible induction of *IFITM2* and *IFITM3* by a cytokine-rich microenvironment. Notably, the specific linkage of IFITMs with IFN-γ points to a possible role for IFITMs in influencing Th1 skewing in the immunological dynamics of $ACPA^+$ eRA. This finding is particularly intriguing as it offers a new perspective on the functions of IFITMs, extending beyond their established antiviral roles. Given its promise, this hypothesis warrants further detailed exploration.

In sum, our data, along with the earlier studies, indicate that targeting type 1 and 2 IFN signaling may be a patient-centric approach for $ACPA^+$ RA patients. In this regard, the JAK inhibitor, which targets type I and type II IFNs (18), can be more effective in treating $ACPA^+$ eRA than $ACPA^-$ eRA. In support of this notion, two clinical studies have shown that ACPA positivity leads to a higher rate of drug retention of JAK inhibitor, an indicator for overall effectiveness and safety of a drug (9, 50). Although earlier findings suggest that an IFN-targeting strategy may hold promise for better efficacy in patients with $ACPA^+$ RA, further research is needed to confirm its superiority in this population.

This study has several limitations. First, we did not perform experimental validation of the findings from the scRNA-seq analysis. Second, the sample size for scRNA-seq analysis was small. Third, this study provides only an immunological basis that supports the potential for greater efficacy of interferon-targeting strategies in $ACPA^+$ eRA than in $ACPA^-$ eRA, which was not examined here. Therefore, these findings require further confirmation in follow-up studies.

To summarize, we observed differences in cytokine profiles, cell subset abundance, and gene expression patterns within the peripheral landscape between $ACPA^+$ eRA and $ACPA^-$ eRA. In $ACPA^+$ eRA, serum IFN-γ levels were elevated, and peripheral blood T cells and monocytes exhibited upregulated IFN-γ response genes and IFN-γ-mediated cell-cell interactions, suggesting Th1 skewing. Moreover, $ACPA^+$ eRA patients showed an expanded population of $IL1B^+$ proinflammatory monocytes, $CD8^+$ $CCL4^+$ T cells, and $IL7R^+$ T cells, in which IRGs were upregulated. Particularly, IFITM 2 and 3, which are associated with IRGs, could be new biomarkers for $ACPA^+$ RA, offering promising avenues for future research and treatment strategies in eRA.

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 below: GSE260796 (GEO).

Ethics statement

The studies involving humans were approved by Institutional Review Board of Seoul St Mary's Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

B-KH: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Visualization, Writing – review & editing, Data curation, Resources, Writing – original draft. SY: Methodology, Software, Supervision, Validation, Visualization, Writing – review & editing. JK: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. MK: Methodology, Software, Validation, Visualization, Writing – review & editing. NL: Formal analysis, Project administration, Writing – review & editing. KL: Software, Writing – review & editing. I-PB: Resources, Visualization, Writing – review & editing. JJ: Data curation, Resources, Writing

– review & editing. W-UK: Conceptualization, Data curation, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. H-YK: Conceptualization, Data curation, Funding acquisition, Investigation, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00442793 to W-UK, Republic of Korea) and Korea Drug Development Fund funded by Ministry of Science and ICT, Ministry of Trade, Industry, and Energy, and Ministry of Health and Welfare (No. RS-2024-00462449 to W-UK, Republic of Korea).

Acknowledgments

Special thanks are due to ROKIT Genomics for their sequencing services and for generating the single-cell RNA-sequencing data. We used the large language model ChatGPT-4 to refine the grammar and language of the paper.

Conflict of interest

KL and I-PB are employed by YiPSCELL, Inc. Seoul, Republic of Korea. JJ is the CEO and founder of YiPSCELL, Inc. SY has been an editorial board member but has no role in the decision to publish this article.

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1439082/full#supplementary-material

References

1. Smolen JS, Aletaha D, Barton A, Burmester GR, Emery P, Firestein GS, et al. Rheumatoid arthritis. *Nat Rev Dis Primers*. (2018) 4:18001. doi: 10.1038/nrdp.2018.1

2. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. *N Engl J Med*. (2011) 365:2205–19. doi: 10.1056/NEJMra1004965

3. Takanashi S, Kaneko Y, Takeuchi T. Characteristics of patients with difficult-to-treat rheumatoid arthritis in clinical practice. *Rheumatol (Oxford)*. (2021) 60:5247–56. doi: 10.1093/rheumatology/keab209

4. Kearsley-Fleet L, Davies R, De Cock D, Watson KD, Lunt M, Buch MH, et al. Biologic refractory disease in rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. *Ann Rheum Dis*. (2018) 77:1405–12. doi: 10.1136/annrheumdis-2018-213378

5. Rivellese F, Pitzalis C. Cellular and molecular diversity in Rheumatoid Arthritis. *Semin Immunol*. (2021) 58:101519. doi: 10.1016/j.smim.2021.101519

6. Derksen VF, Ajeganova S, Trouw LA, van der Helm-van Mil AH, Hafström I, Huizinga TW, et al. Rheumatoid arthritis phenotype at presentation differs depending on the number of autoantibodies present. *Ann Rheum Dis*. (2017) 76:716–20. doi: 10.1136/annrheumdis-2016-209794

7. Seegobin SD, Ma MH, Dahanayake C, Cope AP, Scott DL, Lewis CM, et al. ACPA-positive and ACPA-negative rheumatoid arthritis differ in their requirements for combination DMARDs and corticosteroids: secondary analysis of a randomized controlled trial. *Arthritis Res Ther*. (2014) 16:R13. doi: 10.1186/ar4439

8. Kim MJ, Lee SK, Oh S, Kim HA, Park YB, Lee SS, et al. Efficacy of abatacept versus tumor necrosis factor inhibitors in anti-citrullinated protein antibody-positive patients with rheumatoid arthritis: results from a Korean nationwide biologics registry. *Rheumatol Ther*. (2022) 9:1143–55. doi: 10.1007/s40744-022-00467-4

9. Jung JY, Lee E, Kim JW, Suh CH, Kim HA. Efficacy and drug retention of tofacitinib in rheumatoid arthritis: from the nationwide Korean College of Rheumatology Biologics registry. *Clin Exp Rheumatol*. (2023) 41:1034–41. doi: 10.55563/clinexprheumatol/6fcyza

10. Laurent L, Clavel C, Lemaire O, Anquetil F, Cornillet M, Zabraniecki L, et al. Fcγ receptor profile of monocytes and macrophages from rheumatoid arthritis patients and their response to immune complexes formed with autoantibodies to citrullinated proteins. *Ann Rheum Dis*. (2011) 70:1052–9. doi: 10.1136/ard.2010.142091

11. Harre U, Georgess D, Bang H, Bozec A, Axmann R, Ossipova E, et al. Induction of osteoclastogenesis and bone loss by human autoantibodies against citrullinated vimentin. *J Clin Invest*. (2012) 122:1791–802. doi: 10.1172/jci60975

12. Wu X, Liu Y, Jin S, Wang M, Jiao Y, Yang B, et al. Single-cell sequencing of immune cells from anticitrullinated peptide antibody positive and negative rheumatoid arthritis. *Nat Commun*. (2021) 12:4977. doi: 10.1038/s41467-021-25246-7

13. Lees JR. Interferon gamma in autoimmunity: A complicated player on a complex stage. *Cytokine*. (2015) 74:18–26. doi: 10.1016/j.cyto.2014.10.014

14. Kato M. New insights into IFN-γ in rheumatoid arthritis: role in the era of JAK inhibitors. *Immunol Med*. (2020) 43:72–8. doi: 10.1080/25785826.2020.1751908

15. Reyes-Pérez IV, Sánchez-Hernández PE, Muñoz-Valle JF, Martínez-Bonilla GE, García-Iglesias T, González-Díaz V, et al. Cytokines (IL-15, IL-21, and IFN-γ) in rheumatoid arthritis: association with positivity to autoantibodies (RF, anti-CCP, anti-MCV, and anti-PADI4) and clinical activity. *Clin Rheumatol*. (2019) 38:3061–71. doi: 10.1007/s10067-019-04681-4

16. Han JH, Suh CH, Jung JY, Ahn MH, Han MH, Kwon JE, et al. Elevated circulating levels of the interferon-γ-induced chemokines are associated with disease activity and cutaneous manifestations in adult-onset Still's disease. *Sci Rep*. (2017) 7:46652. doi: 10.1038/srep46652

17. Sarkar S, Cooney LA, White P, Dunlop DB, Endres J, Jorns JM, et al. Regulation of pathogenic IL-17 responses in collagen-induced arthritis: roles of endogenous interferon-gamma and IL-4. *Arthritis Res Ther*. (2009) 11:R158. doi: 10.1186/ar2838

18. Tanaka Y, Luo Y, O'Shea JJ, Nakayamada S. Janus kinase-targeting therapies in rheumatology: a mechanisms-based approach. *Nat Rev Rheumatol*. (2022) 18:133–45. doi: 10.1038/s41584-021-00726-8

19. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum*. (2010) 62:2569–81. doi: 10.1002/art.27584

20. de Pablo P, Dinnes J, Berhane S, Osman A, Lim Z, Coombe A, et al. Systematic review of imaging tests to predict the development of rheumatoid arthritis in people with unclassified arthritis. *Semin Arthritis Rheum*. (2022) 52:151919. doi: 10.1016/j.semarthrit.2021.10.003

21. Lei X, Li H, Zhan Y, Qu J. Predict rheumatoid arthritis conversion from undifferentiated arthritis with dynamic contrast-enhanced MRI and laboratory indexes. *Clin Exp Rheumatol*. (2018) 36:552–8.

22. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell*. (2021) 184:3573–87.e29. doi: 10.1016/j.cell.2021.04.048

23. Zhang F, Wei K, Slowikowski K, Fonseka CY, Rao DA, Kelly S, et al. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat Immunol*. (2019) 20:928–42. doi: 10.1038/s41590-019-0378-1

24. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)*. (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141

25. Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, Fau C, Lacroix M, Colinge J. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res*. (2020) 48:e55. doi: 10.1093/nar/gkaa183

26. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. (2021) 12:1088. doi: 10.1038/s41467-021-21246-9

27. Garcia-Alonso L, Lorenzi V, Mazzeo CI, Alves-Lopes JP, Roberts K, Sancho-Serra C, et al. Single-cell roadmap of human gonadal development. *Nature*. (2022) 607:540–7. doi: 10.1038/s41586-022-04918-4

28. Badia IMP, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinform Adv*. (2022) 2:vbac016. doi: 10.1093/bioadv/vbac016

29. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. (2013) 14:178–92. doi: 10.1093/bib/bbs017

30. Liao W, Lin JX, Wang L, Li P, Leonard WJ. Modulation of cytokine receptors by IL-2 broadly regulates differentiation into helper T cell lineages. *Nat Immunol*. (2011) 12:551–9. doi: 10.1038/ni.2030

31. Castañeda-Delgado JE, Bastián-Hernandez Y, Macias-Segura N, Santiago-Algarra D, Castillo-Ortiz JD, Alemán-Navarro AL, et al. Type I interferon gene response is increased in early and established rheumatoid arthritis and correlates with autoantibody production. *Front Immunol*. (2017) 8:285. doi: 10.3389/fimmu.2017.00285

32. Raterman HG, Vosslamber S, de Ridder S, Nurmohamed MT, Lems WF, Boers M, et al. The interferon type I signature towards prediction of non-response to rituximab in rheumatoid arthritis patients. *Arthritis Res Ther*. (2012) 14:R95. doi: 10.1186/ar3819

33. Lübbers J, Brink M, van de Stadt LA, Vosslamber S, Wesseling JG, van Schaardenburg D, et al. The type I IFN signature as a biomarker of preclinical rheumatoid arthritis. *Ann Rheum Dis*. (2013) 72:776–80. doi: 10.1136/annrheumdis-2012-202753

34. Gocher AM, Workman CJ, Vignali DAA. Interferon-γ: teammate or opponent in the tumour microenvironment? *Nat Rev Immunol*. (2022) 22:158–72. doi: 10.1038/s41577-021-00566-3

35. Theofilopoulos AN, Baccala R, Beutler B, Kono DH. Type I interferons (alpha/beta) in immunity and autoimmunity. *Annu Rev Immunol*. (2005) 23:307–36. doi: 10.1146/annurev.immunol.23.021704.115843

36. Friedman RL, Manly SP, McMahon M, Kerr IM, Stark GR. Transcriptional and posttranscriptional regulation of interferon-induced gene expression in human cells. *Cell*. (1984) 38:745–55. doi: 10.1016/0092-8674(84)90270-8

37. Diamond MS, Farzan M. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat Rev Immunol*. (2013) 13:46–57. doi: 10.1038/nri3344

38. Hammal F, de Langen P, Bergon A, Lopez F, Ballester B. ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res*. (2022) 50:D316–d25. doi: 10.1093/nar/gkab996

39. Reid LE, Brasnett AH, Gilbert CS, Porter AC, Gewert DR, Stark GR, et al. A single DNA response element can confer inducibility by both alpha- and gamma-interferons. *Proc Natl Acad Sci U.S.A*. (1989) 86:840–4. doi: 10.1073/pnas.86.3.840

40. Liang Z, Wang N, Shang L, Wang Y, Feng M, Liu G, et al. Evaluation of the immune feature of ACPA-negative rheumatoid arthritis and the clinical value of matrix metalloproteinase-3. *Front Immunol*. (2022) 13:939265. doi: 10.3389/fimmu.2022.939265

41. James EA, Rieck M, Pieper J, Gebe JA, Yue BB, Tatum M, et al. Citrulline-specific Th1 cells are increased in rheumatoid arthritis and their frequency is influenced by disease duration and therapy. *Arthritis Rheumatol*. (2014) 66:1712–22. doi: 10.1002/art.38637

42. Colpitts SL, Dalton NM, Scott P. IL-7 receptor expression provides the potential for long-term survival of both CD62Lhigh central memory T cells and Th1 effector cells during Leishmania major infection. *J Immunol*. (2009) 182:5702–11. doi: 10.4049/jimmunol.0803450

43. van Roon JA, Glaudemans KA, Bijlsma JW, Lafeber FP. Interleukin 7 stimulates tumour necrosis factor alpha and Th1 cytokine production in joints of patients with rheumatoid arthritis. *Ann Rheum Dis*. (2003) 62:113–9. doi: 10.1136/ard.62.2.113

44. Wenner CA, Güler ML, Macatonia SE, O'Garra A, Murphy KM. Roles of IFN-gamma and IFN-alpha in IL-12-induced T helper cell-1 development. *J Immunol*. (1996) 156:1442–7. doi: 10.4049/jimmunol.156.4.1442

45. Sugawara E, Kato M, Kudo Y, Lee W, Hisada R, Fujieda Y, et al. Autophagy promotes citrullination of VIM (vimentin) and its interaction with major histocompatibility complex class II in synovial fibroblasts. *Autophagy*. (2020) 16:946–55. doi: 10.1080/15548627.2019.1664144

46. Lee SH, Jy K, Kim S-Y, Jung K, Cho M-L. Interferon-gamma regulates inflammatory cell death by targeting necroptosis in experimental autoimmune arthritis. *Sci Rep*. (2017) 7:10133. doi: 10.1038/s41598-017-09767-0

47. Kiefer K, Oropallo MA, Cancro MP, Marshak-Rothstein A. Role of type I interferons in the activation of autoreactive B cells. *Immunol Cell Biol*. (2012) 90:498–504. doi: 10.1038/icb.2012.10

48. Cooles FAH, Anderson AE, Lendrem DW, Norris J, Pratt AG, Hilkens CMU, et al. The interferon gene signature is increased in patients with early treatment-naive rheumatoid arthritis and predicts a poorer response to initial therapy. *J Allergy Clin Immunol*. (2018) 141:445–8.e4. doi: 10.1016/j.jaci.2017.08.026

49. Elemam NM, Hannawi S, Maghazachi AA. Role of chemokines and chemokine receptors in rheumatoid arthritis. *Immunotargets Ther*. (2020) 9:43–56. doi: 10.2147/itt.S243636

50. Sebastiani M, Zabotti A, Biasi B, Cacioppo S, Sandri G, Giovannini I, et al. Factors associated to long-term retention rate of Janus kinase inhibitors in a multi-failure rheumatoid arthritis population. *Clin Exp Rheumatol*. (2024) 42:1416–20. doi: 10.55563/clinexprheumatol/za0hpu

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 13 May 2025
DOI 10.3389/fimmu.2025.1534768

OPEN ACCESS

EDITED BY
Zora Modrusan,
Genentech Inc., United States

REVIEWED BY
Divya Jha,
Icahn School of Medicine at Mount Sinai, United States
Liming Mao,
Nantong University, China
Zongqi He,
Suzhou TCM Hospital Affiliated to Nanjing University of Chinese Medicine, China

*CORRESPONDENCE
Lijiang Ji
fsyy01040ss@njucm.edu.cn
Yan Xu
13862313063@163.com

RECEIVED 26 November 2024
ACCEPTED 17 April 2025
PUBLISHED 13 May 2025

CITATION
Huang H, Ma J, Kang A, Guo T, Sun W, Xu Y and Ji L (2025) Investigating the molecular mechanisms associated with ulcerative colitis through the application of single-cell combined spatial transcriptome sequencing. *Front. Immunol.* 16:1534768. doi: 10.3389/fimmu.2025.1534768

Investigating the molecular mechanisms associated with ulcerative colitis through the application of single-cell combined spatial transcriptome sequencing

Hua Huang[1], Jiaze Ma[2], An Kang[3], Tianwei Guo[4], Wei Sun[5], Yan Xu[6]* and Lijiang Ji[1]*

[1]Department of Anorectal Surgery, Changshu Hospital Affiliated to Nanjing University of Chinese Medicine, Changshu, Jiangsu, China, [2]No. 1 Clinical Medical College, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China, [3]School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China, [4]Department of Pathology, Changshu Hospital Affiliated to Nanjing University of Chinese Medicine, Changshu, Jiangsu, China, [5]Department of Oncology, Changshu Hospital Affiliated to Nanjing University of Chinese Medicine, Changshu, Jiangsu, China, [6]Department of Pharmacy, Affiliated Changshu Hospital of Nantong University, Changshu No. 2 People's Hospital, Changshu, Jiangsu, China

Background: Ulcerative colitis (UC) is a chronic inflammatory bowel disease marked by dysregulated immune responses, resulting in sustained inflammation and ulceration of the colonic and rectal mucosa. To elucidate the cellular subtypes and gene expression profiles implicated in the pathogenesis of UC, we utilized single-cell and spatial transcriptomic analyses.

Methods: We conducted an analysis of single-cell data to identify cell types involved in the pathogenesis of UC. Employing machine learning methodologies, we screened for key genes implicated in UC and validated these findings through spatial transcriptomics. Additionally, immunohistochemistry was performed on UC lesion samples to investigate the expression patterns of the identified key genes. In an animal model, we utilized immunofluorescence and western blotting to validate the expression of these genes in the affected intestinal segments.

Results: Our investigation identified specific monocyte subtypes associated with UC through a comprehensive analysis involving cell communication, Least Absolute Shrinkage and Selection Operator (LASSO), and Support Vector Machine (SVM) methodologies. Notably, two genes, G protein subunit gamma 5 (*GNG5*) and tissue inhibitor of metalloproteinase 1 (*TIMP1*), were identified as key regulators of UC development. Spatial transcriptomic indicated a downregulation of *GNG5* expression in UC, whereas *TIMP1* expression was upregulated. Furthermore, a significant correlation was detected between *TIMP1* and T cell exhaustion-related genes such as genes related to T cell exhaustion, including T cell immunoreceptor with Ig and ITIM domains (*TIGIT*) and cytotoxic T-lymphocyte-associated protein 4 (*CTLA4*). Immunohistochemical analysis of UC lesion samples revealed diminished expression levels of *GNG5* and elevated expression levels of *TIMP1*. A dextran

sulfate sodium (DSS)-induced colitis mouse model was developed, demonstrating that the protein expression levels of *GNG5* in the colonic tissue of model mice were significantly decreased compared to controls w)ile the expression levels of *TIMP1* were increased ($p < 0.01$). Furthermore, immunofluorescence staining indicated co-localization of *TIMP1* with the macrophage marker F4/80 in monocytes.

Conclusion: Our research delineated distinct monocyte subtypes correlated with UC and identified two pivotal genes, *GNG5* and *TIMP1*, that contribute to the disease's pathogenesis. These insights offer a significant theoretical basis for enhancing the clinical diagnosis and therapeutic strategies for patients with UC.

KEYWORDS

immune infiltration, single-cell transcriptome sequencing, spatial transcriptome sequencing, ulcerative colitis, T cell exhaustion

1 Background

Ulcerative colitis (UC) is a chronic, idiopathic form of inflammatory bowel disease (IBD) that predominantly affects the mucosal and submucosal layers of the colorectal region. The pathogenesis of UC is characterized by dysregulated immune responses, resulting in persistent inflammation and ulceration of the colonic and rectal mucosa. Contributing factors include genetic predisposition, environmental influences—including infections and dietary components—and an exaggerated immune response to gut microbiota. These factors collectively undermine the integrity of the mucosal barrier, facilitate the infiltration of inflammatory cells, and promote the release of pro-inflammatory mediators. Epidemiological evidence suggests that UC is relatively prevalent in developed countries, with high incidence rates in North America and Europe (1). In recent years, however, there has been an observable increase in the incidence of UC in many newly industrialized countries, including China, coinciding with global economic development and dietary changes (2). This trend is particularly concerning given the generally reduced life expectancy of UC patients, alongside their heightened risk of requiring colectomy and progression to colorectal cancer. Therefore, the active investigation of UC pathogenesis and the formulation of precise therapeutic strategies have become urgent research imperatives.

Currently, the management of UC primarily involves the administration of 5-aminosalicylic acid (5-ASA) preparations and glucocorticoids. While these pharmacological agents frequently offer prompt alleviation of symptoms, they are also linked to considerable toxic side effects and low patient adherence. Immunosuppressants are primarily employed for maintenance therapy following the remission of symptoms induced by glucocorticoids, with the objective of minimizing glucocorticoid dosage. Additionally, biological agents, specifically monoclonal antibodies that target distinct inflammatory mediators such as tumor necrosis factor or integrins, are incorporated into the therapeutic regimen. The American Gastroenterological Association (AGA) guidelines (3) advocate for the initiation of biologic therapy as a first-line treatment and suggest early step-down strategies, thereby surpassing traditional treatment approaches (4). In the context of selecting biologics for UC, current clinical guidelines endorse the use of vedolizumab (VDZ) or anti-tumor necrosis factor alpha (TNF-α) agents (5). It is important to highlight that over 30% of patients demonstrate resistance to TNF-α therapies, with a subset eventually necessitating intestinal or colon resection surgery (6). In China, over 50% of patients with IBD show suboptimal responses to treatment after approximately one year of first-line anti-TNF-α therapy (7). This secondary dysregulation may be attributed to the immunogenicity of TNF-α antibodies, leading to the development of drug-resistant antibodies (8). Therefore, a deeper investigation into the intricate biological mechanisms underlying UC is essential for advancing the development of effective therapeutic strategies.

Single-cell transcriptome sequencing (scRNA-seq) is a sophisticated technique employed to examine RNA expression at the individual cell level, revealing cellular heterogeneity and the transcriptional profiles of specific cell types. Our preliminary single-cell analysis revealed significant increases in the populations of Plasma cells, activated memory $CD4^+$ T cells, resting Natural Killer (NK) cells, M0 Macrophages, M1 Macrophages, activated Dendritic cells, activated Mast cells, and Neutrophils in patients with UC compared to healthy controls. Subsequently, we conducted an in-depth investigation into the expression of target genes within immune cells, taking into account the complex interactions between bile acid metabolism and immune cell dynamics (9). Spatial transcriptome sequencing, which retains the spatial context of tissues, quantifies gene expression through methods such as microarrays applied to tissue sections or spatial

fluorescence *in situ* hybridization. This methodology facilitates the examination of gene expression within specific tissue regions. The integration of single-cell and spatial transcriptome sequencing techniques permits the concurrent exploration of mechanisms at both the cellular and tissue levels, providing novel insights into the complex mechanisms underlying diseases (10). In this study, we integrated single-cell and spatial transcriptome sequencing to identify target genes associated with UC and conducted a preliminary investigation into the interrelationships among these target genes, immune cells, and the microenvironment. This approach was designed to advance our understanding of UC therapeutic targets and the underlying mechanisms.

2 Materials and methods

2.1 Study design

The study design is presented in Figure 1.

2.2 Data acquisition

The Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/info/datasets.html), curated by the National Center for Biotechnology Information (NCBI), functions

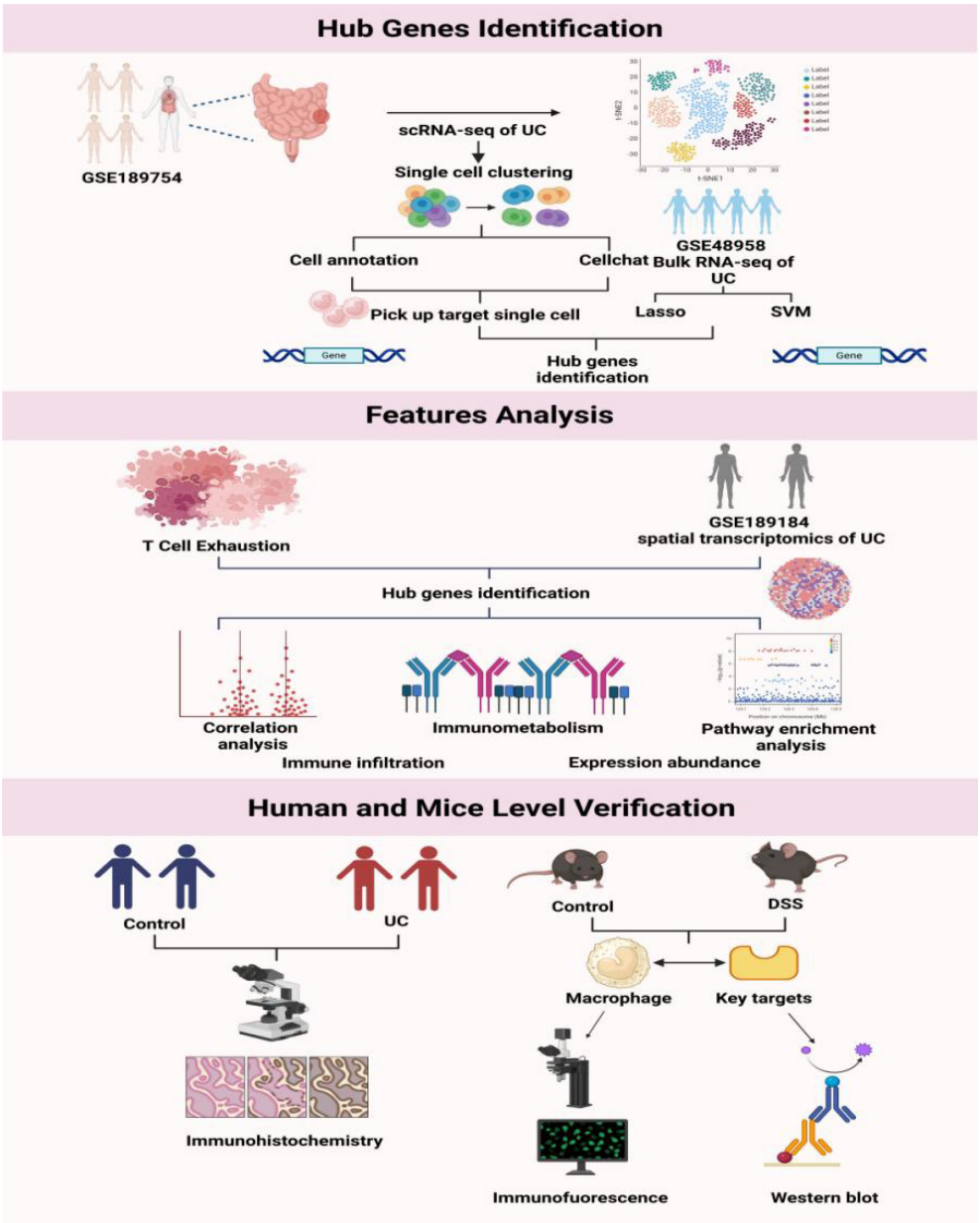


FIGURE 1
Research flowchart.

as a comprehensive repository for gene expression data. From this resource, we acquired the single-cell transcriptome data corresponding to GSE189754, concentrating on 11 samples that provided complete single-cell expression profiles for single-cell analysis. Additionally, we downloaded the spatial transcriptome data for GSE189184, selecting two control groups (B10, C5) and two disease groups (B8, B4) for analysis. Furthermore, we procured the transcriptome data for GSE48958, encompassing data from 8 controls and 13 disease samples.

2.3 Quality control and data standardization

In this study, the processes of quality control and data standardization are essential to ensure the accuracy of subsequent analyses. We employed the Seurat package (11) for initial data processing. For cell quality control, we conducted screening based on the total number of unique molecular identifiers (UMIs) per cell, the number of expressed genes, and the proportion of mitochondrial gene expression. Typically, a high proportion of mitochondrial gene expression in a cell indicates low RNA expression levels and potential progression towards cell death, warranting the exclusion of such cells. Additionally, we utilized the median absolute deviation (MAD) for quality control, removing outliers that deviate from the median by more than three times the MAD to maintain data reliability. Subsequently, we applied DoubletFinder (version 2.0.4) (12) to individually filter doublet cells in each sample, thereby completing the comprehensive cell quality control process.

In the data standardization process, the LogNormalize method of global normalization is employed. This technique mitigates the impact of variations in total RNA content between cells on gene expression analysis by scaling the total expression level of each cell with a coefficient \(s_0\), adjusting it to 10,000, and subsequently normalizing it through logarithmic transformation. Cell cycle scores are computed using the CellCycleScoring function, and highly variable genes are identified via the FindVariableFeatures function. The ScaleData function is utilized to eliminate gene expression fluctuations attributable to mitochondrial gene expression, ribosomal gene expression ratios, and cell cycle differences. Linear dimensionality reduction is conducted on the expression matrix using the RunPCA function, with 20 principal components selected for further analysis. The Harmony algorithm is applied with default parameters to correct for batch effects, and finally, the RunUMAP function is employed with default parameters for nonlinear dimensionality reduction.

2.4 Identification of cell clusters

Cell types and corresponding marker genes were identified using CellMarker (13), PanglaoDB (14), and literature, supplemented by automated annotation with SingleR (15) software. The FindAllMarkers function was employed to filter marker genes within each category, with only positive markers expressed in at least 25% of the cells retained (only.pos = TRUE, min.pct = 0.25).

2.5 Ligand receptor interaction analysis (Cellchat)

CellChat (16) is a sophisticated tool designed for the quantitative inference and analysis of intercellular communication networks derived from single-cell data. Employing network analysis and pattern recognition methodologies, CellChat facilitates the prediction of principal signaling inputs and outputs of cells, elucidating the mechanisms by which these cells and signals orchestrate their functions. In this study, we employed standardized single-cell expression profiles as input data, alongside cell subtype classifications obtained through single-cell analysis, to serve as cell-specific information. We conducted an in-depth examination of cell-related interactions, quantifying the strength and frequency of cell-to-cell interactions to observe the activity and impact of each cell type in the disease.

2.6 Feature selection process of LASSO regression and SVM algorithm

We utilized the Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression and Support Vector Machine (SVM) algorithms to select features for diagnostic markers of diseases. The LASSO algorithm utilizes the “glmnet” package, while SVM-Recursive Feature Elimination (SVM-RFE) is a machine learning method based on support vector machines (17). SVM-RFE removes feature vectors generated by SVM to identify optimal variables, and establishes a support vector machine model through the “e1071” package to further assess the diagnostic value of these biomarkers in disease contexts.

2.7 Immune infiltration analysis

The CIBERSORT method is a prevalent technique for assessing immune cell types within microenvironments (18). In this study, utilized the CIBERSORT algorithm was employed to analyze patient data, allowing for the inference of the relative proportions of 22 immune-infiltrating cell types. Furthermore, a correlation analysis was conducted to examine the relationship between gene expression and immune cell content.

2.8 GSEA analysis

Patients were categorized into high and low-expression groups based on the expression of key genes. Subsequently, Gene Set Enrichment Analysis (GSEA) was utilized to examine disparities

in signaling pathways between these cohorts. The annotation gene set employed for the subtype pathway analysis was derived from version 7.0 of the Molecular Signatures Database (MsigDB). Differential expression analysis of pathways between the groups was conducted, and significantly enriched gene sets (adjusted *p-value* < 0.05) were ranked by consistency score. GSEA is frequently used to explore the correlation between disease classification and biological significance.

2.9 GSVA analysis

Gene Set Variation Analysis (GSVA) is a nonparametric, unsupervised method for assessing gene set enrichment in transcriptome data. GSVA assigns a comprehensive score to each gene set of interest, converting gene-level changes into pathway-level changes. This allows for the identification of potential biological function changes in different samples. In this study, gene sets were downloaded from MsigDB, and the GSVA algorithm was applied to comprehensively score each gene set, enabling the evaluation of potential biological function differences among the samples.

2.10 Non-coding RNA network associated with key genes

MicroRNAs (miRNAs) are small non-coding RNAs known to regulate gene expression by facilitating mRNA degradation or inhibiting mRNA translation. Consequently, we conducted an in-depth analysis to determine the presence of miRNAs associated with key genes involved in the transcriptional regulation or degradation of potentially deleterious genes. We identified miRNAs related to these key genes using the miRcode database and subsequently visualized the miRNA-gene interaction network utilizing Cytoscape software (19).

2.11 Transcription factor regulatory network

This study utilized the R package "RcisTarget" to predict transcription factors, with all computations conducted by RcisTarget being predicated on motif analysis. The normalized enrichment score (NES) of a motif depended on the total number of motifs in the database. In addition to the motifs annotated by the source data, we inferred further annotation files based on motif similarity and gene sequences. To estimate the overrepresentation of each motif in the gene set, we initially calculated the area under the curve (AUC) for each pair of motif-motif set. This was performed based on the recovery curve calculation of the gene set ranking of the motifs. The NES of each motif was calculated based on the AUC distribution of all motifs in the gene set.

2.12 Source of human sample

To verify the expression of target genes in the diseased colon tissue of UC patients, tissue biopsy samples were collected from UC patients within the research cohort at the Digestive Endoscopy Center of Changshu Hospital Affiliated to Nanjing University of Chinese Medicine (Ethical Number: CZYLS-2024120). Patients with UC secondary to other diseases or with differing pathological results were excluded. Normal tissue samples for the control group were obtained from the periphery of pathological specimens diagnosed with colon cancer and subjected to Miles surgery in the General Surgery Department of Changshu Hospital Affiliated to Nanjing University of Chinese Medicine. The collection of all samples was approved by the hospital's Ethics committee, and written informed consent was obtained from the patients.

2.13 Immunohistochemistry

Colon tissue sections fixed with paraformaldehyde were deparaffinized using xylene and incubated with primary antibodies (tissue inhibitor of metalloproteinase 1 (*TIMP1*):1:200, Absin, Shanghai, China; G protein subunit gamma 5 (*GNG5*):1:200, Abcam, Shanghai, China) at 37°C for 1.5 hours. After three washes with PBS, immunocomplex detection was performed using diaminobenzidine, and nuclei were counterstained with hematoxylin. The sections were examined under a microscope (Leica, Wetzlar, Germany) (20).

2.14 Animals and treatment

Male C57BL/6 mice (weighing 18–20 grams) were obtained from Beijing Vital River Laboratory Animal Technology Co., Ltd. (SCXK-2021-0011, Beijing, China). Before the experiments, the mice were provided with standard laboratory chow and water ad libitum under controlled conditions of 60 ± 5% humidity, 23 ± 1°C temperature, and a 12-hour light/dark cycle. The experimental protocol was approved by the Ethics Committee of the Experimental Animal Center at Nanjing University of Chinese Medicine (Ethical Number: NJUCCSHAE-2021-1123). The mice were randomly divided into two groups: a control group and a dextran sulfate sodium (DSS) group, each with 6 mice. The control group received drinking water, while the DSS group was given 3% DSS in drinking water for 7 days (21). On the 8th day, all mice were euthanized, and their colon samples were collected for further analysis.

2.15 Hematoxylin and eosin staining

Colon tissue was fixed in 4% paraformaldehyde, subsequently embedded in dehydrated paraffin, and sectioned at a thickness of

4.5μm. The sections were then stained with H&E. Pathological alterations in the tissue samples were examined using an optical microscope (Leica, Wetzlar, Germany).

2.16 Enzyme-linked immunosorbent assay

Accurately weigh the colon tissue to achieve a weight (mg) to volume (μL) ratio of 1:9. Add nine times the volume of physiological saline and homogenize the mixture mechanically under ice water bath conditions to prepare a 10% homogenate. Centrifuge the homogenate at 2500–3000 rpm for 10 minutes and collect the supernatant for subsequent ELISA analysis. Following the manufacturer's protocol, the concentrations of TNF-α (mIC50536-1, Mlbio, Shanghai, China) and interleukin-6 (IL-6) (ml098430, Mlbio, Shanghai, China), were quantified using a commercially available ELISA kit.

2.17 Immunofluorescence staining for co-localization validation

The sample slices were fixed in 10% formalin, embedded in paraffin, dewaxed, and subjected to antigen retrieval. After a one-hour blocking step at room temperature, the slices were incubated overnight at 4°C with primary antibodies *TIMP1* (1:200, Absin, Shanghai, China) and F4/80 (1:50, Abcam, Shanghai, China). Following three 10-minute washes with PBS, the slices were incubated for one hour at room temperature with Alexa Fluor 488 and Alexa Fluor 594 secondary antibodies. After another three PBS washes, an anti-quenching medium was used to mount the cover glass onto the slide. The sections were then examined under a fluorescence microscope (Leica, Wetzlar, Germany) at a magnification of 80 for microscopic analysis and imaging (22).

2.18 Western blot for expression validation

Total protein was extracted from colon tissue samples of human or mouse origin using RIPA lysis buffer (Beyotime, Nanjing, China) and quantified using the Bicinchoninic Acid (BCA) protein assay kit (Beyotime, Nanjing, China). Subsequently, 20 micrograms of protein were separated on a 10% SDS-PAGE gel and transferred to a polyvinylidene fluoride (PVDF) membrane. The membrane was then blocked with 5% (w/v) bovine serum albumin (BSA) or skim milk at room temperature for 1 hour. Following the blocking step, the membrane was incubated overnight at 4°C with primary antibodies targeting *GNG5* (1:1000, Absin, Shanghai, China), *TIMP1* (1:1000, Absin, Shanghai, China), and GAPDH (1:5000, Proteintech, Wuhan, China). On the next day, the membrane was incubated with secondary antibodies (horseradish peroxidase-conjugated goat anti-rabbit or anti-mouse IgG, 1:5000, Cell Signaling Technology, Danvers, MA, USA) at room temperature for 1 hour. Visualization of the protein bands was performed using an ECL detection kit and a gel imaging system (Tanon, Shanghai, China). The intensity of the bands was then quantified using the densitometric analysis feature of Gel Pro 4.0 software (Tanon, Shanghai, China) (23).

2.19 Statistical analysis

All statistical analyses were performed using the R programming language (version 4.3.0), with a significance threshold set at $p < 0.05$.

3 Results

3.1 Preliminary processing of single-cell expression profile data

During the initial processing of single-cell expression profile data, rigorous adherence to established quality control and standardized procedures was maintained. Following the screening process, cells with fewer than 200 captured genes were excluded, while those meeting the criteria were retained based on indicators such as nFeature-RNA, nCount-RNA, and percent.mt, resulting in a dataset of 22,345 high-quality cells. Concurrently, doublets were removed, and the top 2,000 highly variable genes were selected for subsequent analysis. The processed data demonstrated favorable distribution characteristics, as evidenced by violin plots and scatter plots, thereby establishing a robust foundation for precise cell subpopulation annotation and gene expression analysis in future studies (Supplementary Figure S1A, C). This approach effectively mitigates analysis bias associated with data quality issues.

3.2 Single-cell data cell subpopulation annotation and ligand-receptor interaction analysis (Cellchat)

The data underwent standardization, homogenization, and subsequent analysis using Principal Component Analysis (PCA), Harmony, and Uniform Manifold Approximation and Projection (UMAP) (Supplementary Figures S1D-F, Figure 2A). Each subtype was annotated to one of seven cell categories: $CD4^+$ T cells, B cells, $CD8^+$ T cells, Fibroblasts, Monocytes, Mast cells, and NK cells (Figure 2A). A bubble plot and histogram were generated to visualize the expression of classic markers and cell proportions for these categories (Figures 2B, C). The software package Cellchat was employed to examine ligand-receptor interactions within the single-cell expression profile, revealing intricate relationships between the cell subtypes (Figure 2D). Notably, Monocytes demonstrated a closer potential interaction with other cell types (Figures 2E, F).

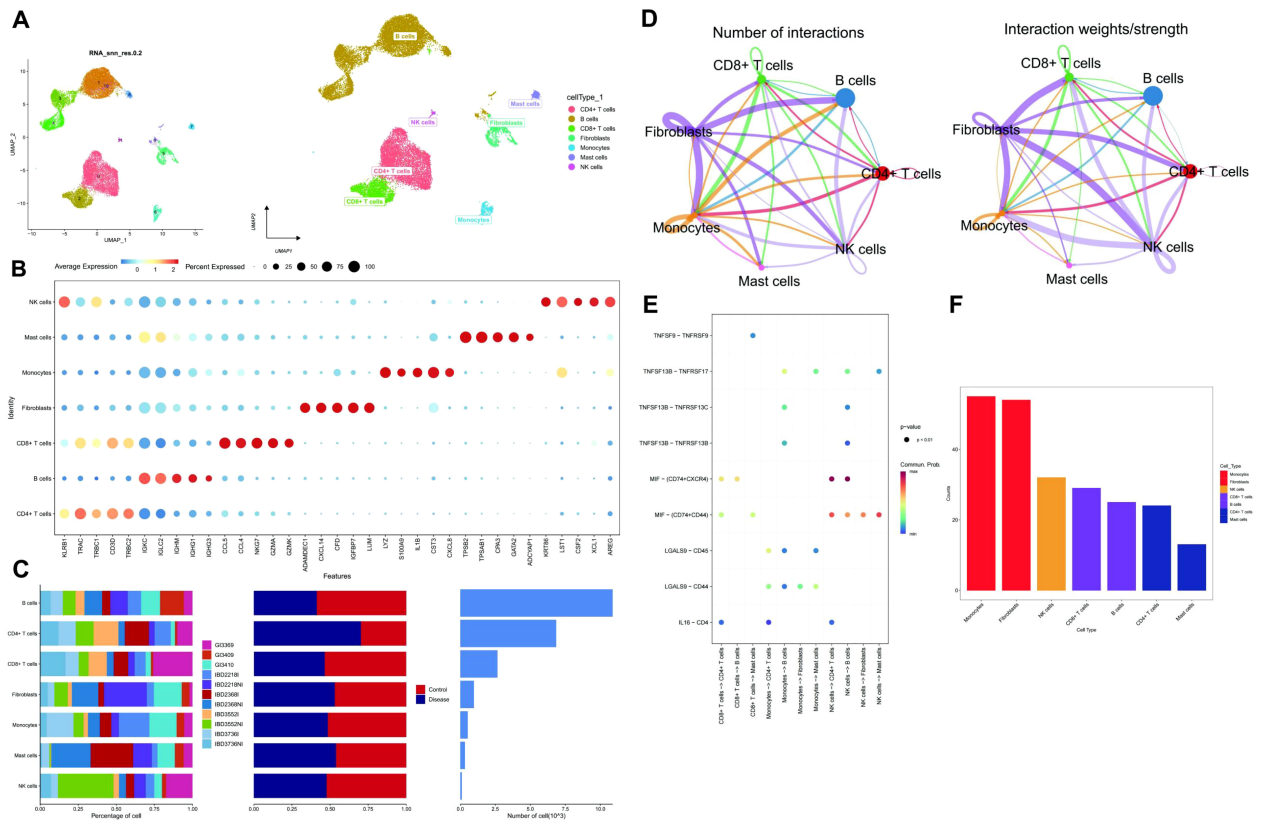


FIGURE 2
Cell annotation and communication. **(A)** UMAP-based clustering of cells into 12 groups. Classification of clusters into 7 cell types: $CD4^+$ T cells, B cells, $CD8^+$ T cells, Fibroblasts, Monocytes, Mast cells, and NK cells. **(B)** A Dotplot visualization of cell type markers and their expression levels. **(C)** Bar charts displaying the proportions and content of 7 cell types in the sample. **(D)** Cell interaction network among 7 cell types based on communication probability and strength. **(E)** Bubble plot of receptor-ligand interactions between cells, with colors showing communication probabilities. **(F)** Comparison of total interactions among 7 cell types, showing a decreasing trend from left to right, with Monocytes having the strongest interactions.

3.3 Feature selection process of LASSO regression and SVM algorithm

To investigate the genetic underpinnings of UC, we retrieved expression profile data from the GEO database (GSE48958). This dataset comprised 21 patient samples. To identify key genes associated with this condition, we employed a two-step approach. First, we utilized a combination of LASSO regression and SVM algorithms to screen the 386 monocyte marker genes ($p\_adj < 0.05$ & $LogFC > 0.585$) previously identified. LASSO regression yielded 18 characteristic genes (Figures 3A, B), while SVM identified 4 (Figure 3C). By intersecting these gene sets, we identified 2 genes, *GNG5* and *TIMP1*, as the most promising candidates for further exploration in our research on UC (Figure 3D).

3.4 Immune infiltration analysis

The microenvironment, a pivotal factor in disease progression, consists of a complex interplay between cellular and extracellular components. This intricate ecosystem includes fibroblasts, immune cells, extracellular matrix, growth factors, inflammatory factors, and unique physical and chemical properties. The microenvironment exerts a substantial influence on disease diagnosis, prognosis, and therapeutic response. Our investigation revealed distinct patterns of immune cell infiltration and correlation in various disease states (Figures 4A, B). Compared to the control group, the disease group exhibited significantly elevated levels of M1 Macrophages, resting NK cells, $CD4^+$ memory activated T cells, and $CD4^+$ memory resting T cells. Conversely, resting Mast cells and NK cells activated were significantly reduced in the disease group (Figure 4C). Further analysis of the relationship between key genes and immune cells demonstrated a strong positive correlation between *TIMP1* and several immune cell types, including $CD4^+$ memory resting T cells, $CD4^+$ memory activated T cells, follicular helper T cells, resting NK cells, M0 Macrophages, M1 Macrophages, activated Dendritic cells, and Neutrophils. Conversely, *TIMP1* was negatively correlated with Plasma cells, $CD8^+$,T cells regulatory T cells (Tregs), activated NK cells, and resting Mast cells (Figure 4D). Moreover, our analysis of the correlation between key genes and different immune factors, including immunosuppressive factors, immunostimulatory factors, chemokines, and receptors, suggests that these genes are intimately involved in shaping the immune microenvironment (Supplementary Figures S2A-E).

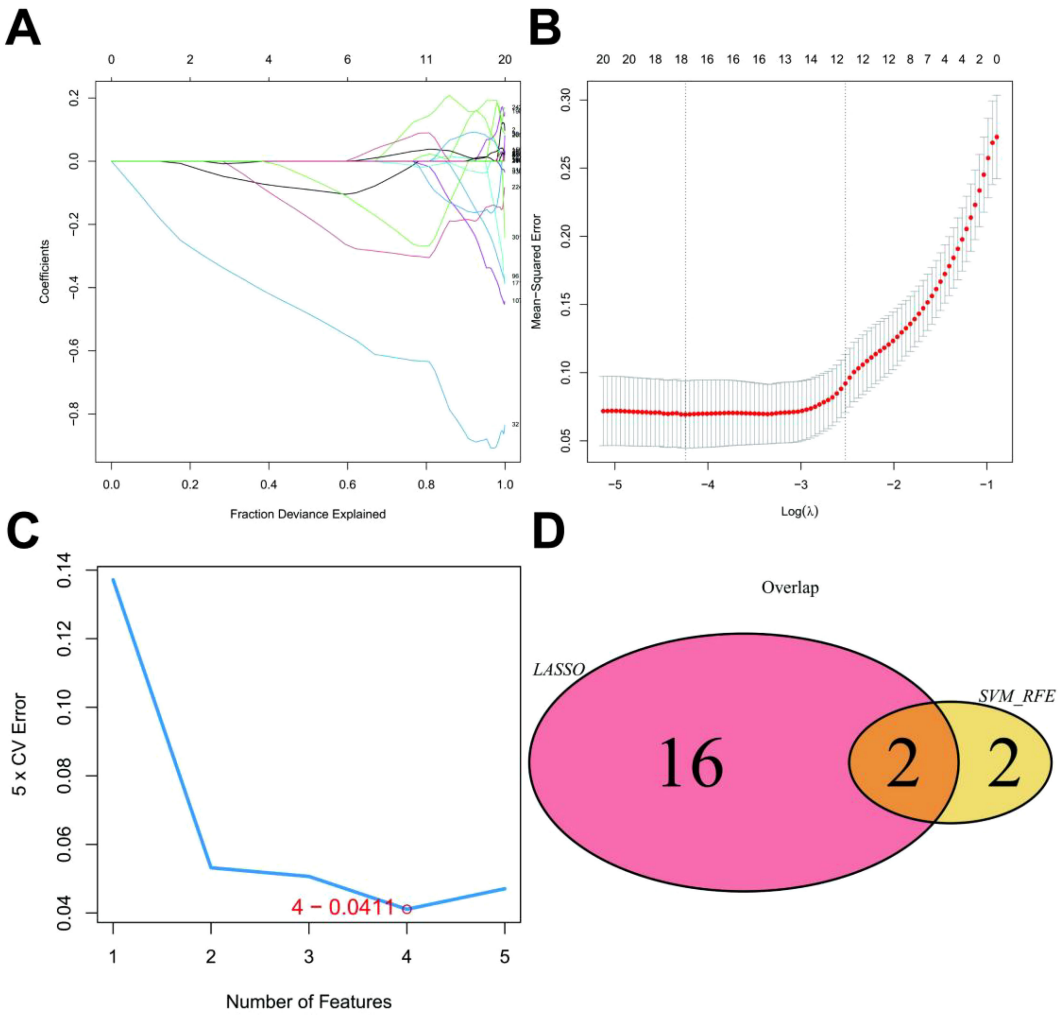


FIGURE 3
LASSO Model Construction. **(A)** LASSO coefficient distribution and gene combination at the minimum lambda value. **(B)** Ten-fold cross-validation for tuning parameter selection to find the minimum lambda. **(C)** Top four feature genes with the lowest error rate in the SVM algorithm. **(D)** Venn plot showing two overlapping genes selected by both LASSO regression and SVM algorithms.

3.5 Signaling pathways involved in key genes

To elucidate the specific signaling pathways involved in the key genes and explore their potential molecular mechanisms in disease progression, we conducted a comprehensive analysis. GSEA revealed that *GNG5* was significantly enriched in signaling pathways such as propanoate metabolism, butanoate metabolism, and peroxisome proliferator-activated receptor (PPAR) signaling (Figures 5A, B). *TIMP1*, on the other hand, was enriched in pathways including B cell receptor signaling, interleukin-17 (IL-17) signaling, and NF-κB signaling (Figures 5D, E). Additionally, GSVA identified *GNG5* as being enriched in pathways associated with protein secretion and adipogenesis (Figure 5C). *TIMP1* was found to be enriched in pathways related to hedgehog signaling and epithelial-mesenchymal transition (Figure 5F). These findings collectively suggest that the key genes may influence disease progression through these identified signaling pathways.

3.6 Non-coding RNA network and transcriptional regulatory network related to key genes

Subsequently, we employed the miRcode database to conduct a reverse prediction of the key genes, resulting in the identification of 20 miRNAs and a total of 23 mRNA-miRNA regulatory relationships. These interactions were visualized using Cytoscape (Supplementary Figure S3A). By utilizing the key genes as a gene set for this analysis, we discovered that these genes were subject to regulation by common mechanisms, such as multiple transcription factors. To identify these transcription factors, we employed cumulative recovery curves and conducted motif-transcription factor annotation and selection analysis on the key genes. The motif with the highest standardized enrichment score (NES: 14) was determined to be cisbp:M6056. We have provided a comprehensive visualization of all the enriched motifs and their corresponding transcription factors associated with the key genes (Supplementary Figure S3B, C).

10.3389/fimmu.2025.1534768

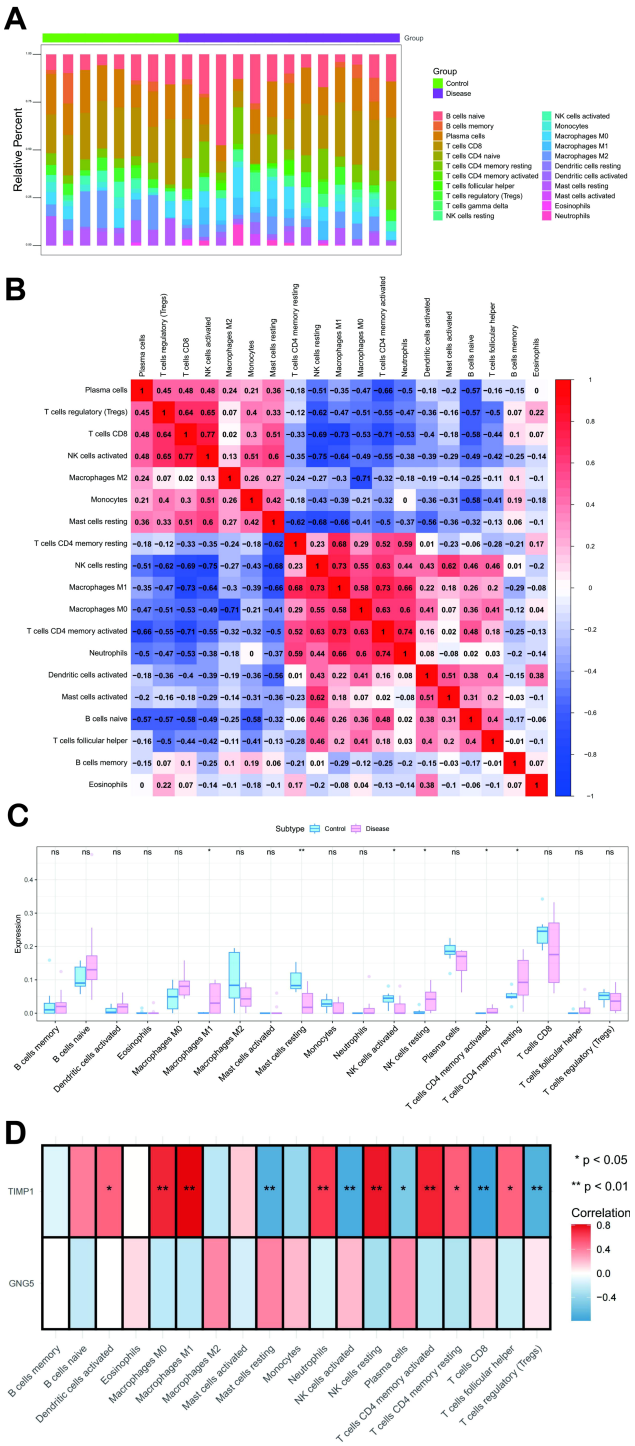


FIGURE 4
Immune Infiltration Analysis. **(A)** Comparison of immune cell subset percentages between control and disease groups, with immune cells on the x-axis and relative percentages on the y-axis. **(B)** Correlation of immune cell infiltration, showing cell types on both axes; red indicates positive correlation, blue indicates negative, and darker colors signify stronger associations. **(C)** Blue and pink bars show the immune cell content differences between control and disease groups, respectively, with cell types on the x-axis, scores on the y-axis, and * indicating statistical significance. **(D)** The x-axis shows immune cell types, the y-axis shows two key genes, and an asterisk marks statistical significance in their correlation. **: $p < 0.01$.

10.3389/fimmu.2025.1534768

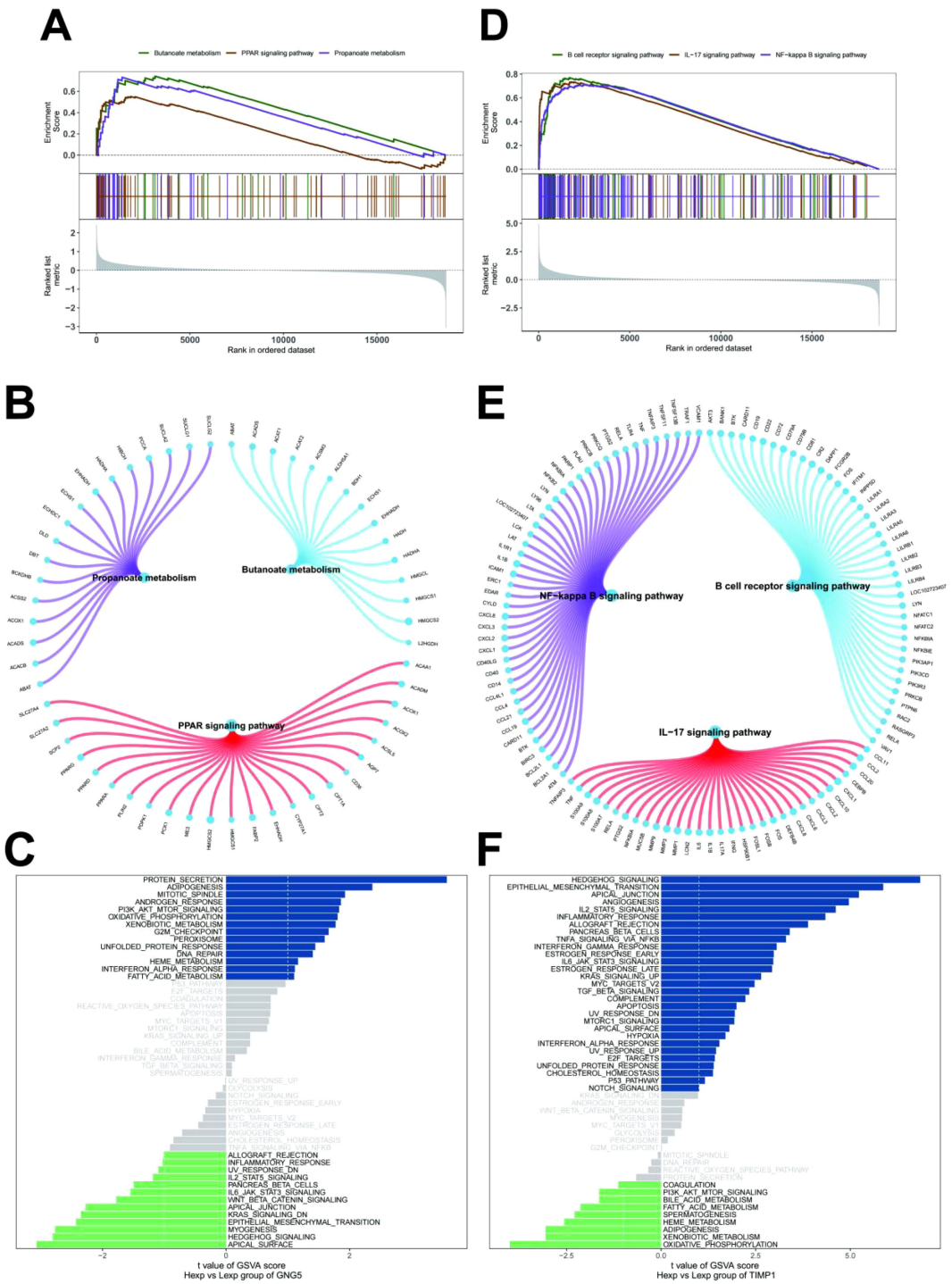


FIGURE 5
Signal pathways of key genes. **(A, B)** *GNG5* in the KEGG signaling pathway and their regulatory roles. **(C)** GSVA analysis showing signaling pathways for high expression *GNG5* (blue) and low expression *GNG5* (green), using the Hallmark gene set as a reference. **(D, E)** *TIMP1* in the KEGG signaling pathway and their regulatory roles. **(F)** GSVA analysis showing signaling pathways for high expression *TIMP1* (blue) and low expression *TIMP1* (green), using the Hallmark gene set as a reference.

3.7 Relationship between key genes and disease-related genes

In the current study, the GeneCards database (https://www.genecards.org/) was utilized to identify genes potentially implicated in disease regulation. To assess inter-group expression differences amongst these genes, we analyzed the expression levels of 20 highly ranked genes (based on the Relevance*score) with confirmed expression within the transcriptome data. This analysis revealed significant expression differences between the two patient groups for

genes including interleukin 23 receptor (*IL23R*), interferon gamma (*IFNG*), nucleotide-binding oligomerization domain 2 (*NOD2*), tumor protein 53 (*TP53*), transforming growth factor beta 1 (*TGFB1*), interleukin 1 receptor antagonist (*IL1RN*), interleukin-1 beta (*IL1B*), interleukin 8 (*CXCL8*), tumor necrosis factor (*TNF*) and ATP-binding cassette subfamily B member 1 (*ABCB1*) (Figure 6A). Furthermore, a correlation analysis was performed to investigate the relationship between key genes and disease regulation genes. The expression levels of these key genes demonstrated statistically significant correlations with the expression levels of disease regulation genes. Notably, *TIMP1* exhibited a strong positive correlation (cor = 0.949) with *IL1RN* while displaying a significant negative correlation (cor = -0.807) with *ABCB1* (Figure 6B).

3.8 Expression profile of key genes in spatial transcriptome and validation of pathological tissues derived from human sources

We analyzed the spatial transcriptome data to assess the expression levels of two key genes. Compared with the control group, *GNG5* expression was inhibited in the disease group, while *TIMP1* expression was significantly upregulated in the disease group (Figure 7A). We assessed the differential expression levels of key genes across various groups utilizing bubble and violin plot visualizations. Our analysis revealed a downregulation of *GNG5* and an upregulation of *TIMP1* in UC (Figure 7B). IHC analysis was conducted on colon lesions from patients with UC to assess the expression levels of *GNG5* and *TIMP1*. Results demonstrated a significant upregulation of *TIMP1*, and a significant downregulation of *GNG5* in the disease group compared to the control group ($p < 0.001$), aligning with the previously presented spatial transcriptome data (Figure 7C).

3.9 Validation of key genes in tissues from a DSS-induced colitis mouse model

A DSS-induced colitis mice model was generated successfully (Figure 8A). The distal colon tissue was collected and its length was measured and then imaged. It was found that compared with the control group, the colon of the DSS-induced colitis mice model was significantly shortened under inflammatory stimulation ($p < 0.001$)

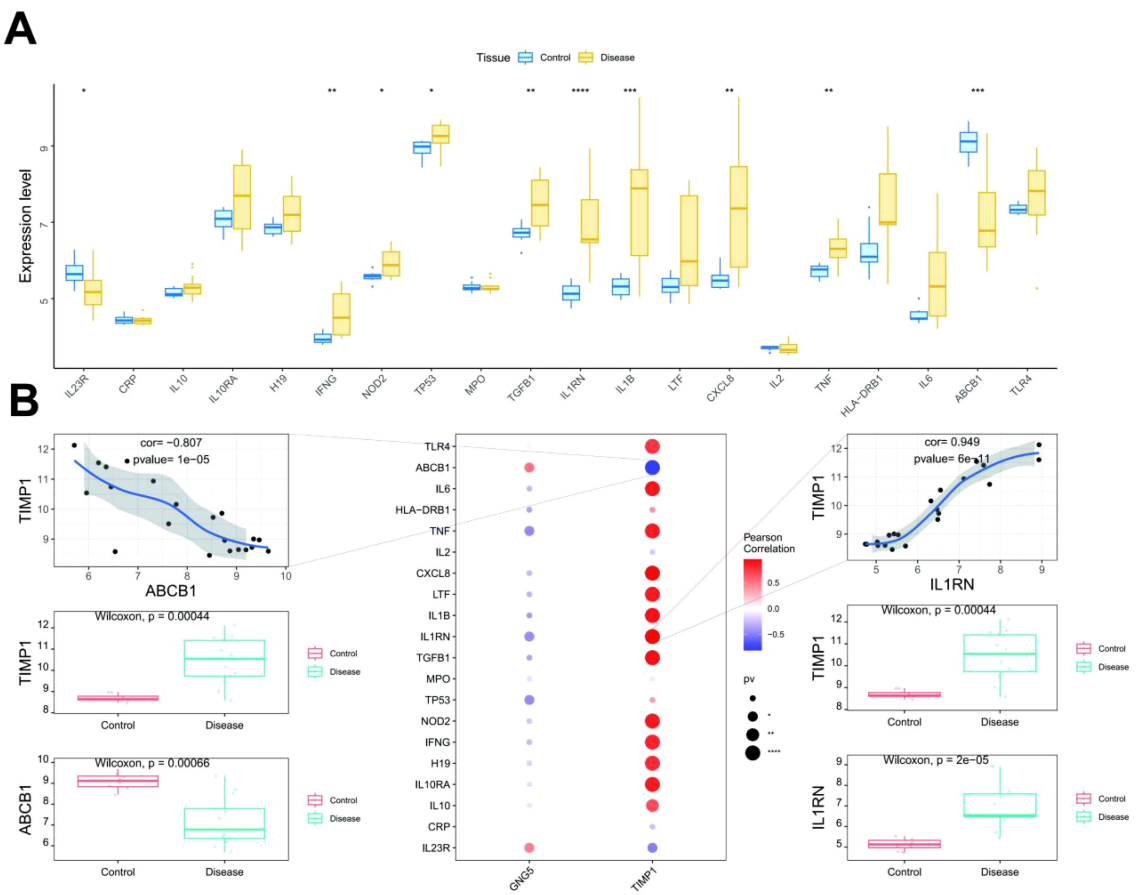
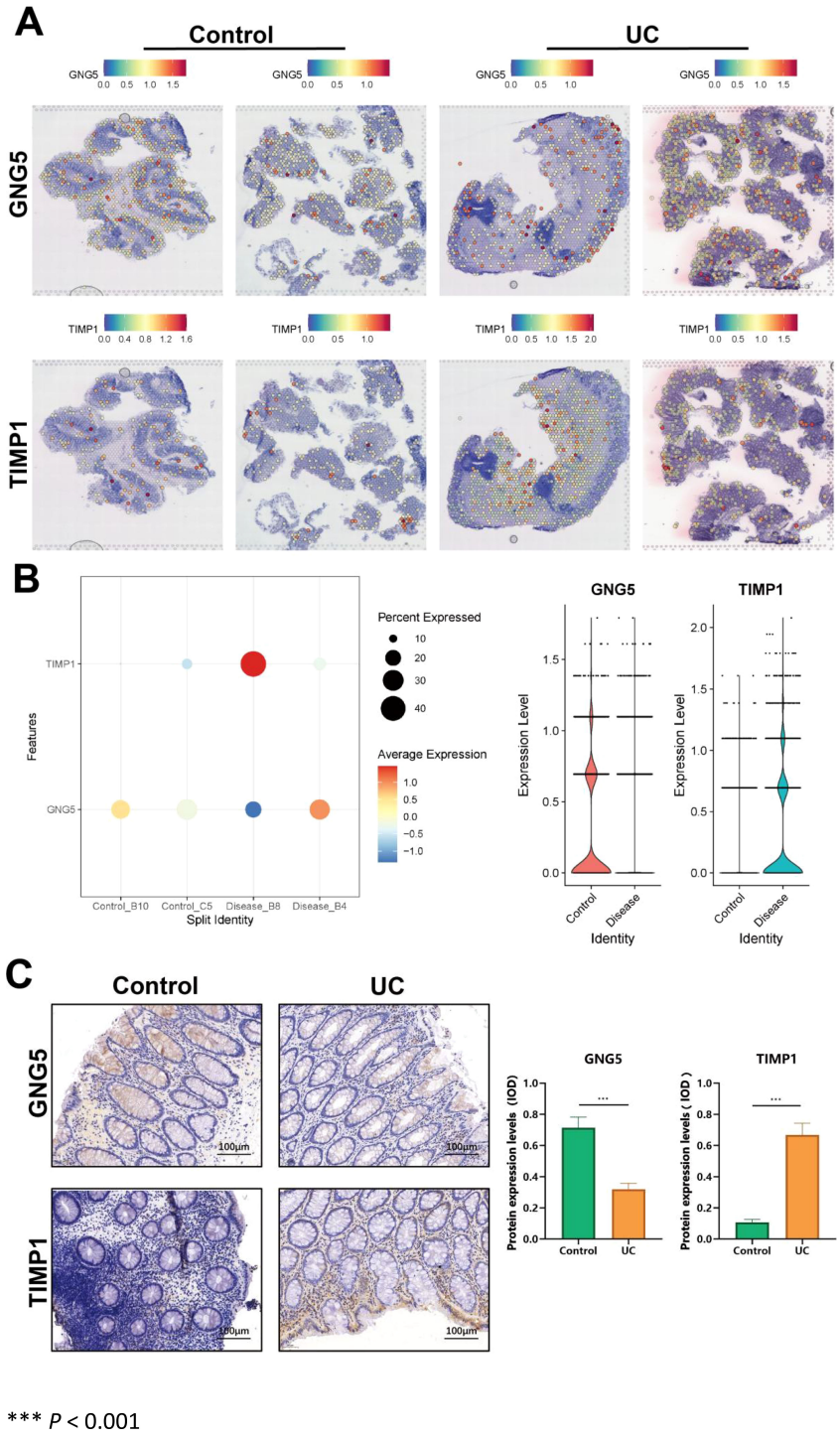


FIGURE 6
Correlation between key genes and disease genes. **(A)** The top figure illustrates the differential expression of disease regulatory genes, with blue for control patients and yellow for disease patients. **(B)** The bottom figure presents correlation analysis, where blue denotes negative and red denotes positive correlations. *: p < 0.05; **: p < 0.01; ***: p < 0.001; ****: p < 0.0001.

10.3389/fimmu.2025.1534768



*** $P < 0.001$

FIGURE 7
Key gene expression in spatial transcriptome and lesion tissues of UC patients. **(A)** Scatter plot showing key gene expression levels in single-cell idling. **(B)** Up: Bubble plot depicting key gene expression levels (blue = low, red = high). Down: Violin plot illustrating key gene expression distribution in single-cell idling. **(C)** Differential expression of *GNG5* and *TIMP1* between control and UC groups. *** $P < 0.001$.

(Figure 8B). The colon tissue of mice was collected for H&E staining. As illustrated in Figure 8C, DSS induction resulted in the desquamation and necrosis of colonic epithelial cells, infiltration of inflammatory cells within the mucosal layer, and loss of crypt structures in DSS-induced colitis model mice. The levels of IL-6 and TNF-α in colon tissue were measured utilizing ELISA. The results revealed a significant increase in the levels of inflammatory cytokines IL-6 and TNF-α in the colon tissue of DSS-induced colitis model mice ($p < 0.001$) compared with the control group (Figures 8D, E). To further elucidate the expression of key genes in

10.3389/fimmu.2025.1534768

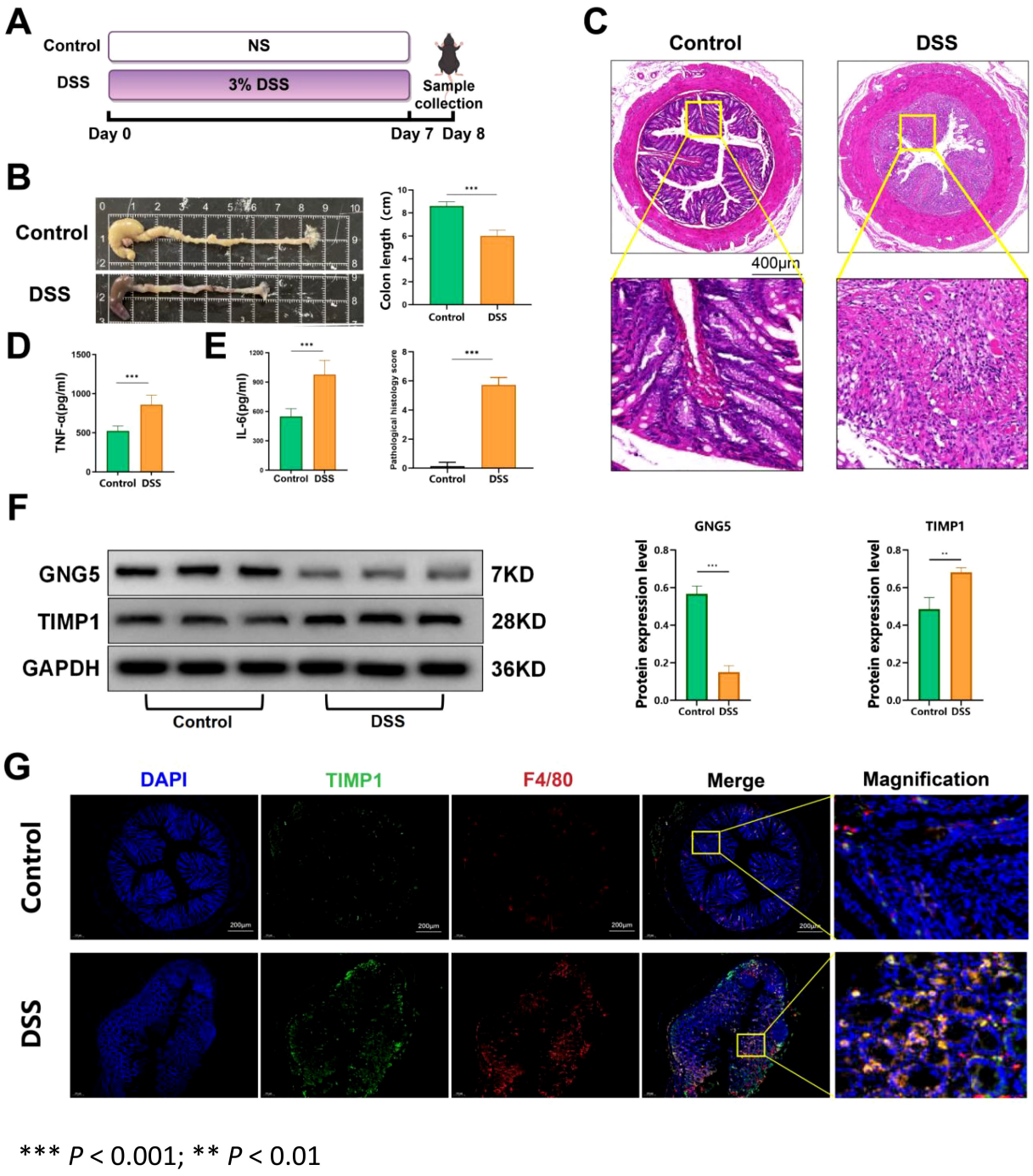


FIGURE 8
Key gene expression in DSS-induced colonic lesions in mice model. **(A)** The animal experimental protocol. **(B)** H&E staining images of colon tissue from the indicated groups. **(C)** Comparison of colon length in the indicated groups. **(D, E)** The levels of IL-6 and TNF-α in the colon homogenate of each group. F Protein levels of *GNG5* and *TIMP1* in the colon. **(G)** Co-localization of *TIMP1* with the macrophage marker F4/80 in the colon. *** $P < 0.001$; ** $P < 0.01$.

UC, we conducted Western blot to evaluate corresponding protein expression levels in colon samples obtained from a DSS-induced colitis mouse model. Results indicated that, compared to the control group, *GNG5* protein levels were significantly downregulated and *TIMP1* levels were significantly upregulated in the DSS group ($p < 0.001$, $p < 0.01$), which confirmed our spatial transcriptome predictions (Figure 8G).

Our single-cell sequencing data showed high monocyte expression in UC samples, with macrophages being crucial monocyte components. Lots of references indicates that macrophages are vital in UC inflammation and tissue repair (24). Immune infiltration analysis revealed a stronger correlation between *TIMP1* and UC-related immune cells compared to *GNG5*. Thus, we used immunofluorescence co-localization to assess *TIMP1* expression

10.3389/fimmu.2025.1534768

in macrophages. The results indicated that *TIMP1* co-localizes with macrophage marker F4/80 in colon tissue, suggesting that *TIMP1* may affect disease progression through functional expression in UC colon macrophages (Figure 8G).

3.10 The link between key genes and immune metabolic pathways along with T cell exhaustion correlation analysis

To quantitatively assess the activity of immune metabolism genes in individual cells, we utilized AUCell. Bubble plots were employed to visualize the differential activity of key genes within these pathways. Our findings revealed that *GNG5* and *TIMP1* were significantly upregulated in oxidative phosphorylation, the unfolded protein response, and related pathways (Figure 9A). Furthermore, an analysis of classical exhaustion-related genes (*LAG3*, *PDCD1*, *TIGIT*, *HAVCR2*, *CTLA4*) in single cells indicated a pronounced T cell exhaustion phenotype (Figures 9B, C). We also investigated the correlation between exhaustion-related genes and immune infiltration as well as their differential expression in the transcriptome. *CTLA4*, *LAG3* and *TIGIT* were found to be significantly up-regulated in UC (Figure 9D).Our analysis identified a significant positive correlation between five exhaustion-related genes and activated memory $CD4^+$ T cells, T follicular helper cells, as well as other cell types. (Figure 9E). To delve deeper into the relationship between *GNG5*, *TIMP1*, and cellular depletion, a correlation analysis was conducted involving five depletion-related genes. This analysis identified a significant positive correlation between *TIMP1* and *TIGIT* as well as *CTLA4* (Figure 9F) (p = 3.5e-06, 1.1e-07).

4 Discussion

The integration of single-cell and spatial transcriptomic analyses in this study has yielded novel insights into the cellular heterogeneity and molecular dynamics underlying the pathogenesis of UC. Our findings underscore the pivotal roles of specific monocyte subtypes and two key genes, *GNG5* and *TIMP1*, in modulating the inflammatory microenvironment and driving disease progression. These discoveries not only enhance our

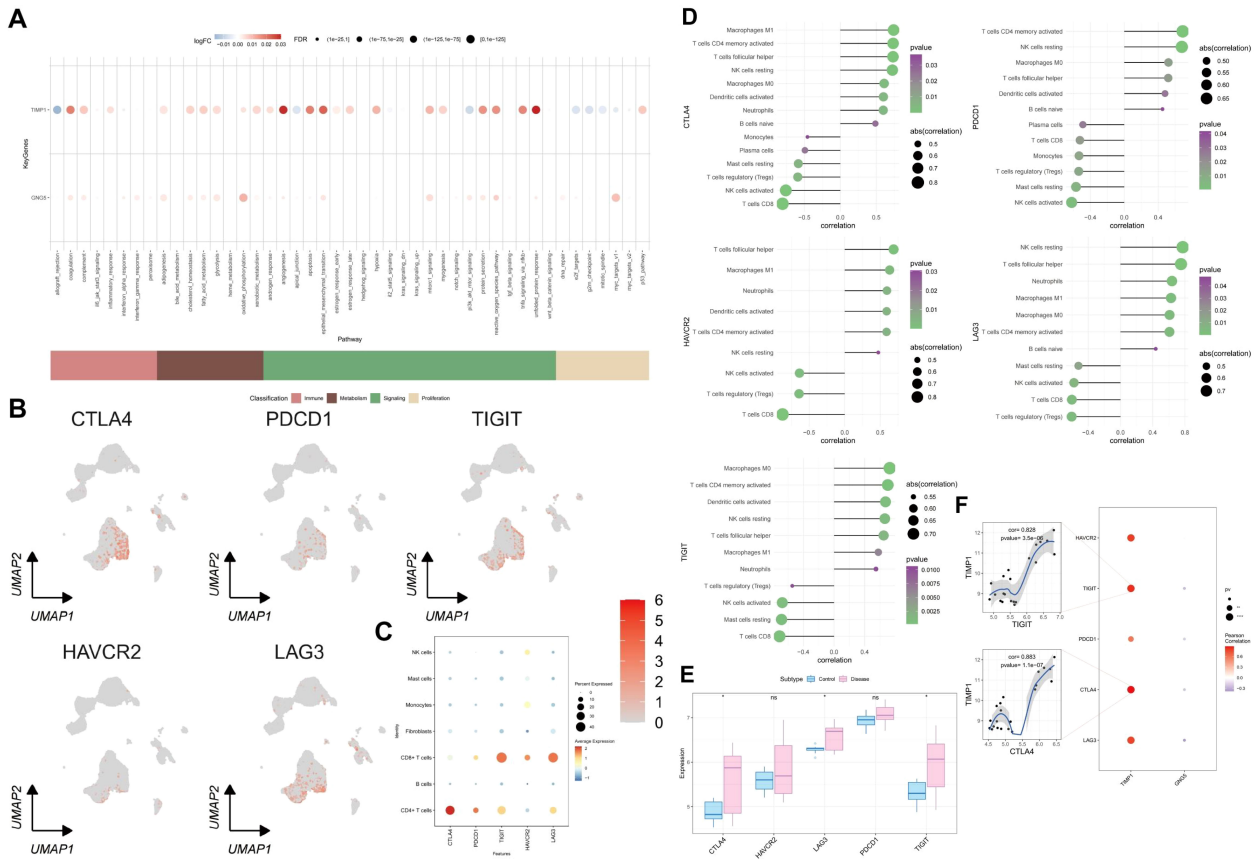


FIGURE 9
Link between key genes and immune metabolic pathways, and T cell exhaustion analysis. **(A)** Correlation bubble plot: x-axis shows immune metabolism pathways, y-axis shows two key genes, blue indicates low expression, red indicates high expression. **(B)** UMAP diagram: expression profile of exhaustion-related genes in single cells. **(C)** Bubble plot showing single-cell expression of exhausted phase genes, with blue for low and red for high expression. **(D)** Circle size and color represent the correlation coefficient and p-value between exhaustion-related genes and immune cells. **(E)** Differential expression of exhausted genes in UC. **(F)** Blue and red circles indicate negative and positive correlations between key and exhaustion-related genes, with circle size showing statistical significance.

understanding of UC immunopathology but also suggest potential therapeutic targets for precision medicine. Utilizing single-cell transcriptomics, we identified distinct subpopulations of monocytes characterized by altered communication networks in UC, and monocytes demonstrated the most robust ligand-receptor interactions. The application of machine learning techniques, specifically LASSO and SVM algorithms, further identified *GNG5* and *TIMP1* as central regulatory elements in UC. *GNG5*, a G protein subunit involved in signal transduction, was significantly downregulated in UC lesions, whereas *TIMP1*, a metalloproteinase inhibitor associated with extracellular matrix remodeling, was markedly upregulated. These findings are corroborated by spatial transcriptomic and immunohistochemical analyses in human UC tissues, as well as in DSS-induced murine colitis models, thereby confirming their consistent dysregulation across species.

In UC, the disproportionate distribution of monocyte subtypes—characterized by a predominance of classical and intermediate subtypes with pro-inflammatory tendencies, alongside functional impairments in non-classical subtypes—contributes to the pathogenesis of intestinal inflammation and fibrosis. Monocytes are capable of further differentiation into macrophages within specific tissues, including the intestinal and dermal regions (25). Classical monocytes, identified by the $CD14^{++}$/CD16 phenotype, engage C-C motif chemokine receptor 2 (CCR2) signaling, which is pivotal for their function, thereby activating downstream NF-κB and MAPK pathways. This activation facilitates their differentiation into pro-inflammatory M1 macrophages and augments the secretion of inflammatory mediators, including interleukin-1β (IL-1β) and reactive oxygen species (ROS) (26). CCR2 signaling has the potential to enhance TIMP1 expression, potentially through the activation of the PI3K/Akt pathway, inhibit MMP-9 activity, and consequently exacerbate extracellular matrix (ECM) deposition and fibrosis (27). Intermediate monocytes, identified by the $CD14^{+}$/$CD16^{+}$, undergo differentiation into M2 macrophages in response to the influence of TGF-β. The GNG5 protein is involved in G protein-coupled receptor (GPCR) signaling pathways, including those mediated by CCR2 and chemokine (C-X3-C motif) receptor 1 (CX3CR1). The absence of GNG5 may impede the migration of monocytes to the intestinal environment and their subsequent differentiation into anti-inflammatory phenotypes, such as M2 macrophages (28).

GNG5, a member of the G protein gamma subunit family, is a component of the glutamate transporter family. G proteins are essential signaling molecules involved in various physiological processes. *GNG5* plays a regulatory role in the body, influencing cell proliferation, differentiation, and metabolism. Previous studies (29) have implicated *GNG5* in glioma cell proliferation, migration, and macrophage infiltration. *GNG5* is involved in cell cycle regulation and promotes cell proliferation, potentially through the modulation of growth factor receptor-associated signaling pathways (30). Additionally, *GNG5* has been linked to apoptosis in human chondrocytes (31) and lung cancer cells (32). However, the correlation between *GNG5*, UC, and immune cells remains understudied. In this study, we present novel findings identifying *GNG5* as a target gene associated with monocyte markers and characteristics of UC. The observed downregulation of *GNG5* in UC tissues, as evidenced by spatial transcriptomic, immunohistochemical, and murine model analyses, indicates its potential involvement in maintaining mucosal homeostasis. *GNG5* is implicated in modulating intracellular signaling pathways, including those mediated by G Protein-Coupled Receptors (GPCRs), which are essential for immune cell activation and epithelial repair. Our analyses utilizing GSEA and GSVA have linked *GNG5* to propanoate metabolism and the PPAR signaling pathways, both of which are known for their roles in regulating anti-inflammatory responses and preserving epithelial barrier integrity. The reduced expression of *GNG5* may undermine these protective mechanisms, potentially exacerbating inflammation and tissue damage.

TIMP1 is a zinc and calcium-containing proteolytic enzyme secreted by neutrophils and lymphocytes. Its primary function is to inhibit matrix metalloproteinases (MMPs), which are crucial for extracellular matrix (ECM) degradation. MMPs are upregulated after tissue injury and are involved in cytokine activation, cell migration, and ECM remodeling. TIMPs balance MMP activity, promoting tissue wound healing (33). MMPs and TIMPs are key regulators in IBD pathogenesis. Their imbalance is associated with inflammation and intestinal fibrosis in IBD (34). MMPs also modulate the inflammatory response by cleaving and activating cytokines, intensifying inflammation (35). Despite *TIMP1*'s inhibitory effect on MMPs, it does not exhibit the expected anti-inflammatory properties in inflammatory diseases. Elevated *TIMP1* levels have been associated with poor prognosis in various inflammatory conditions (36). Schoeps et al. (37) found that high *TIMP1* expression can activate neutrophils to release neutrophil extracellular traps (NETs). In patients with IBD, *TIMP1* expression is significantly elevated in colon tissue and serum, correlating with disease severity (38). These findings suggest that *TIMP1*'s pro-inflammatory properties outweigh its MMP inhibitory effects in inflammatory diseases. In this study, we hypothesized that *TIMP1* is a target gene associated with monocyte markers and UC characteristics, with more significant interaction with immune cells than *GNG5*. *TIMP1* expression is markedly upregulated in UC samples and positively correlates with various immune cell populations, including M0 and M1 macrophages, activated dendritic cells, and neutrophils. Human sample analysis corroborated these findings. A DSS-induced acute colitis model in mice revealed a significant increase in *TIMP1* protein levels in colonic tissue compared to controls. Immunofluorescence demonstrated *TIMP1* co-localization with the macrophage marker F4/80, suggesting that *TIMP1*'s pro-inflammatory effects in mice may be mediated through its influence on macrophage function in the colon.

UC is often associated with dysregulated intestinal immune cells, particularly T cell activation and functional alterations. T cell exhaustion, a progressive decline in T cell function under prolonged immune stimulation, is common in chronic infections and tumor microenvironments. This phenomenon limits T cell antigen response capacity and effector functions, reducing overall immune response

efficacy (39). Exhausted T cells exhibit decreased proliferative capacity, upregulated inhibitory receptors(programmed cell death protein 1 (*PD-1*), *CTLA4*, and *LAG3*), and loss of effector function (40). *LAG3* and *PD-1* co-expression drives T cell exhaustion and regulates the expression of thymus high mobility group box protein (TOX). Ectopic TOX expression in effector T cells induces T cell exhaustion transcriptional programs (41). Slevin et al. (42) reported *LAG3* upregulation in mucositis, primarily on effector memory T cells, correlating with disease activity, suggesting *LAG3* as a potential therapeutic target for UC. However, a clinical trial by D'Haens et al. (43) showed that *LAG3*-depleting monoclonal antibody GSK2831781 did not reduce colonic mucosal inflammation despite successful *LAG3* depletion. Therefore, the mechanisms linking T cell exhaustion to UC prognosis require further exploration. Our analysis of immune infiltration has identified a skewed microenvironment in UC, characterized by a predominance of M1 Macrophages, activated $CD4^+$ T cells, and Neutrophils, alongside a reduction in regulatory T cells and resting Mast cells. This imbalance reflects the pro-inflammatory environment observed in advanced stages of UC, where persistent inflammation contributes to tissue damage. Importantly, *TIMP1* was identified as a central gene positively associated with markers of T cell exhaustion, such as *TIGIT* and *CTLA4*. T cell exhaustion, indicative of chronic antigen exposure, is increasingly recognized in UC and may account for the limited effectiveness of current immunotherapies. The co-enrichment of *TIMP1* with the IL-17 and NF-κB pathways further implicates it in sustaining Th17-driven inflammation, a critical pathway in UC pathogenesis. These findings suggest that targeting *TIMP1* or its downstream effectors could alleviate both inflammation and immune exhaustion, providing a dual therapeutic approach.

5 Conclusion

This study represents the inaugural integration of single-cell and spatial transcriptomics methodologies to examine the spatial distribution characteristics of cellular heterogeneity and gene expression in UC. The investigation elucidates the metabolic regulatory role of *GNG5* and the pro-inflammatory and depletion-promoting effects of *TIMP1*, identifying them as potential novel biomarkers or targets for therapeutic intervention. The study further identified that *TIMP1* facilitates the progression of ulcerative colitis via a dual mechanism involving T cell depletion and macrophage activation. This finding offers a theoretical foundation for developing therapeutic strategies aimed at targeting *TIMP1* or other associated immune checkpoints.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Ethics statement

To validate the expression of target genes in the diseased colonic tissue of patients with UC, biopsy samples were obtained from individuals within the research cohort at the Digestive Endoscopy Center of Changshu Hospital, affiliated with Nanjing University of Chinese Medicine. All procedures involving patient specimens received approval from the Ethics Committee of Changshu Hospital Affiliated to Nanjing University of Chinese Medicine (Approval No. CZYLS-2024120). Written informed consent for participation in this study was obtained from the participants or their legal guardians. The animal studies were approved by the Ethics Committee of the Experimental Animal Center at Nanjing University of Chinese Medicine (Approval No. NJUCCSHAE-2021-1123) and were conducted in compliance with local legislation and institutional requirements.

Author contributions

HH: Data curation, Formal analysis, Software, Writing – original draft, Writing – review & editing. JM: Formal analysis, Methodology, Software, Writing – original draft. AK: Data curation, Project administration, Software, Writing – original draft. TG: Conceptualization, Supervision, Writing – original draft. WS: Methodology, Writing – original draft. YX: Writing – review & editing. LJ: Supervision, Writing – review & editing, Project administration, Resources.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by Jiangsu Province Traditional Chinese Medicine Technology Development Plan Project (MS2023099), Research Plan Project of Jiangsu Society of Traditional Chinese Medicine (CYTF2024048, PDJH2024045), Natural Science Foundation of Nanjing University of Traditional Chinese Medicine (XZR2023076), Suzhou Science and Technology Bureau's "Science and Education Revitalization of Health" Science and Technology Project (KJXW2023069, MSXM2024053), Changshu Science and Technology Bureau Medical Application Basic Research Project (CY202340, KCH202406).

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.

The reviewer ZH declared a shared parent affiliation with the authors HH, JM, AK, TG, WS, LJ to the handling editor at the time of the review.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2025.1534768/full#supplementary-material

References

1. Burisch J, Zhao M, Odes S, De Cruz P, Vermeire S, Bernstein CN, et al. The cost of inflammatory bowel disease in high-income settings: a Lancet Gastroenterology & Hepatology Commission. *Lancet Gastroenterol Hepatol.* (2023) 8:458–92. doi: 10.1016/S2468-1253(23)00003-1

2. Li Z, Qi H. Li Maojuan, Miao YingleiResearch progress in the economic burden of inflammatory bowel disease. *Chin J Inflamm Bowel Dis.* (2023) 07:286–91. doi: 10.1097/CM9.0000000000003345

3. Singh S, Allegretti JR, Siddique SM, Terdiman JP. AGA technical review on the management of moderate to severe ulcerative colitis. *Gastroenterology.* (2020) 158 (5):1465-1496.e17. doi: 10.1053/j.gastro.2020.01.007

4. Feuerstein JD, Isaacs KL, Schneider Y, Siddique SM, Falck-Ytter Y, Singh S. AGA clinical practice guidelines on the management of moderate to severe ulcerative colitis. *Gastroenterology.* (2020) 158:1450–61. doi: 10.1053/j.gastro.2020.01.006

5. Huang H, Ma J, Hua Q, Jin C, Ji L. Clinical study on the treatment of distal ulcerative colitis with Ganyu-Pixu by the therapeutic method of Rougan Jianpi and mesalazine enema. *Int J Tradit Chin Med.* (2024) 46:444–50. doi: 10.1093/ecco-jcc/jjaf052

6. Bitoun S, Hässler S, Ternant D, Szely N, Gleizes A, Richez C, et al. Response to biologic drugs in patients with rheumatoid arthritis and antidrug antibodies. *JAMA Netw Open.* (2023) 6:e2323098. doi: 10.1001/jamanetworkopen.2023.23098

7. Li J, Liu Z, Hu P, Wen Z, Cao Q, Zou X, et al. Indicators of suboptimal response to anti-tumor necrosis factor therapy in patients from China with inflammatory bowel disease: results from the EXPLORE study. *BMC Gastroenterol.* (2022) 22:44. doi: 10.1186/s12876-021-02074-z

8. Cheah E, Huang JG. Precision medicine in inflammatory bowel disease: Individualizing the use of biologics and small molecule therapies. *World J Gastroenterol.* (2023) 29:1539–50. doi: 10.3748/wjg.v29.i10.1539

9. Huang H, Yan S, Guo T, Hua Q, Wang Y, Xu S, et al. Bile acid metabolism modulates intestinal immunity involved in ulcerative colitis progression. *Heliyon.* (2024) 10:e34352. doi: 10.1016/j.heliyon.2024.e34352

10. Hu X, Xu W, Ren Y, Wang Z, He X, Huang R, et al. Spinal cord injury: molecular mechanisms and therapeutic interventions. *Signal Transduct Target Ther.* (2023) 8:245. doi: 10.1038/s41392-023-01477-6

11. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell.* (2021) 184:3573–3587.e29. doi: 10.1016/j.cell.2021.04.048

12. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* (2019) 8:329–337.e4. doi: 10.1016/j.cels.2019.03.003

13. Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, et al. CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res.* (2023) 51:D870–6. doi: 10.1093/nar/gkac947

14. Franzén O, Gan LM, Björkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford).* (2019) 2019: baz046. doi: 10.1093/database/baz046

15. Cao G, Xuan X, Li Y, Hu J, Zhang R, Jin H, et al. Single-cell RNA sequencing reveals the vascular smooth muscle cell phenotypic landscape in aortic aneurysm. *Cell Commun Signal.* (2023) 21:113. doi: 10.1186/s12964-023-01120-5

16. Denier C, Labauge P, Bergametti F, Marchelli F, Riant F, Arnoult M, et al. Genotype-phenotype correlations in cerebral cavernous malformations patients. *Ann Neurol.* (2006) 60:550–6. doi: 10.1002/ana.20947

17. Mundra PA, Rajapakse JC. SVM-RFE with MRMR filter for gene selection. *IEEE Trans Nanobiosci.* (2010) 9:31–7. doi: 10.1109/TNB.2009.2035284

18. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol.* (2018) 1711:243–59. doi: 10.1007/978-1-4939-7493-1_12

19. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/gr.1239303

20. Wang Y, Zhang B, Liu S, Xu E, Wang Z. The traditional herb Sargentodoxa cuneata alleviates DSS-induced colitis by attenuating epithelial barrier damage via blocking necroptotic signaling. *J Ethnopharmacol.* (2024) 319:117373. doi: 10.1016/j.jep.2023.117373

21. He YX, Li YY, Wu YQ, Ren LZ, Wang Y, Wang YM, et al. Huanglian Ganjiang decoction alleviates ulcerative colitis by restoring gut barrier via APOC1-JNK/P38 MAPK signal pathway based on proteomic analysis. *J Ethnopharmacol.* (2024) 318:116994. doi: 10.1016/j.jep.2023.116994

22. Chen C, Bao Y, Ju S, Jiang C, Zou X, Zhang X, et al. Single-cell and bulk RNA-seq unveils the immune infiltration landscape associated with cuproptosis in cerebral cavernous malformations. *Biomark Res.* (2024) 12:57. doi: 10.1186/s40364-024-00603-y

23. Liu H, Feng X, Wang D, Liu L, Liu Y, Liu B, Zhu L, Zhang C, Yang W. Mechanism of Sishen Pills-Tongxie Yaofang in the treatment of ulcerative colitis based on network pharmacology and experimental verification. *Sci Tradit Chin Med.* (2024) 2 (3):224–36. doi: 10.1097/st9.0000000000000038

24. Zhang J, Zhao Y, Hou T, Zeng H, Kalambhe D, Wang B, et al. Macrophage-based nanotherapeutic strategies in ulcerative colitis. *J Control Release.* (2020) 320:363–80. doi: 10.1016/j.jconrel.2020.01.047

25. Liu Z, Gu Y, Chakarov S, Bleriot C, Kwok I, Chen X, et al. Fate mapping via ms4a3-expression history traces monocyte-derived cells. *Cell.* (2019) 178:1509–1525.e19. doi: 10.1016/j.cell.2019.08.009

26. Yang J, Guan X, He S, Ge L, Gao Q, Wu X. FTY720 attenuates acute colitis via colonic T cells reduction and inhibition of M1 macrophages polarization independent of CCR2-mediated monocytes input. *Int Immunopharmacol.* (2023) 123:110731. doi: 10.1016/j.intimp.2023.110731

27. Xi C, Zhang G, Sun N, Liu M, Ju N, Shen C, et al. Repurposing homoharringtonine for thyroid cancer treatment through TIMP1/FAK/PI3K/AKT signaling pathway. *iScience.* (2024) 27:109829. doi: 10.1016/j.isci.2024.109829

28. Shroka TM, Kufareva I, Salanga CL, Handel TM. The dual-function chemokine receptor CCR2 drives migration and chemokine scavenging through distinct mechanisms. *Sci Signal.* (2023) 16:eabo4314. doi: 10.1126/scisignal.abo4314

29. Zhang W, Liu Z, Liu B, Jiang M, Yan S, Han X, et al. GNG5 is a novel oncogene associated with cell migration, proliferation, and poor prognosis in glioma. *Cancer Cell Int.* (2021) 21:297. doi: 10.1186/s12935-021-01935-7

30. Liu B, Su Q, Xiao B, Zheng G, Zhang L, Yin J, et al. RAB42 promotes glioma pathogenesis via the VEGF signaling pathway. *Front Oncol.* (2021) 11:657029. doi: 10.3389/fonc.2021.657029

31. Shen X-F, Cheng Y, Dong Q-R, Zheng M-Q. MicroRNA-675-3p regulates IL-1β-stimulated human chondrocyte apoptosis and cartilage degradation by targeting GNG5. *Biochem Biophys Res Commun.* (2020) 527:458–65. doi: 10.1016/j.bbrc.2020.04.044

32. Peng J, Li S, Li B, Hu W, Ding C. Exosomes derived from M1 macrophages inhibit the proliferation of the A549 and H1299 lung cancer cell lines via the miRNA-let-7b-5p-GNG5 axis. *PeerJ.* (2023) 11:e14608. doi: 10.7717/peerj.14608

33. Aksnes M, Schibstad MH, Chaudhry FA, Neerland BE, Caplan G, Saltvedt I, et al. Differences in metalloproteinases and their tissue inhibitors in the cerebrospinal fluid are associated with delirium. *Commun Med (Lond).* (2024) 4:124. doi: 10.1038/s43856-024-00558-z

34. de Bruyn M, Vandooren J, Ugarte-Berzal E, Arijs I, Vermeire S, Opdenakker G. The molecular biology of matrix metalloproteinases and tissue inhibitors of metalloproteinases in inflammatory bowel diseases. *Crit Rev Biochem Mol Biol.* (2016) 51:295–358. doi: 10.1080/10409238.2016.1199535

35. Zhang K, Yang J, Chen L, He J, Qu D, Zhang Z, et al. Gut microbiota participates in polystyrene microplastics-induced hepatic injuries by modulating the gut-liver axis. *ACS Nano.* (2023) 17:15125–45. doi: 10.1021/acsnano.3c04449

36. Schoeps B, Frädrich J, Krüger A. Cut loose TIMP-1: an emerging cytokine in inflammation. *Trends Cell Biol*. (2023) 33:413–26. doi: 10.1016/j.tcb.2022.08.005

37. Schoeps B, Eckfeld C, Prokopchuk O, Böttcher J, Häußler D, Steiger K, et al. TIMP1 triggers neutrophil extracellular trap formation in pancreatic cancer. *Cancer Res*. (2021) 81:3568–79. doi: 10.1158/0008-5472.CAN-20-4125

38. Meijer MJW, Mieremet-Ooms MAC, van der Zon AM, van Duijn W, van Hogezand RA, Sier CFM, et al. Increased mucosal matrix metalloproteinase-1, -2, -3 and -9 activity in patients with inflammatory bowel disease and the relation with Crohn's disease phenotype. *Dig Liver Dis*. (2007) 39:733–9. doi: 10.1016/j.dld.2007.05.010

39. Belk JA, Yao W, Ly N, Freitas KA, Chen Y-T, Shi Q, et al. Genome-wide CRISPR screens of T cell exhaustion identify chromatin remodeling factors that limit T cell persistence. *Cancer Cell*. (2022) 40(7):768-86. doi: 10.1016/j.ccell.2022.06.001

40. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. *Nat Rev Immunol*. (2015) 15:486–99. doi: 10.1038/nri3862

41. Andrews LP, Butler SC, Cui J, Cillo AR, Cardello C, Liu C, et al. LAG-3 and PD-1 synergize on CD8+ T cells to drive T cell exhaustion and hinder autocrine IFN-γ-dependent anti-tumor immunity. *Cell*. (2024) 187(16):4355-72.e22. doi: 10.1016/j.cell.2024.07.016

42. Slevin SM, Garner LC, Lahiff C, Tan M, Wang LM, Ferry H, et al. Lymphocyte activation gene (LAG)-3 is associated with mucosal inflammation and disease activity in ulcerative colitis. *J Crohns Colitis*. (2020) 14:1446–61. doi: 10.1093/ecco-jcc/jjaa054

43. D'Haens G, Peyrin-Biroulet L, Marks DJB, Lisi E, Liefaard L, Beaton A, et al. A randomised, double-blind, placebo-controlled study of the LAG-3-depleting monoclonal antibody GSK2831781 in patients with active ulcerative colitis. *Aliment Pharmacol Ther*. (2023) 58:283–96. doi: 10.1111/apt.17557

Frontiers in Immunology

Explores novel approaches and diagnoses to treat immune disorders.

The official journal of the International Union of Immunological Societies (IUIS) and the most cited in its field, leading the way for research across basic, translational and clinical immunology.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact