



## OPEN ACCESS

## EDITED BY

Mahak Singh,  
ICAR Research Complex for NEH Region,  
Nagaland, India

## REVIEWED BY

Muhammet Rasit Ugur,  
IVF Michigan Fertility Centers, United States

## \*CORRESPONDENCE

Olga Amelkina  
✉ amelkina@izw-berlin.de

RECEIVED 20 October 2025

REVISED 16 December 2025

ACCEPTED 29 December 2025

PUBLISHED 12 January 2026

## CITATION

Amelkina O and Comizzoli P (2026) Making sense of expanding transcriptomic data: network-based approaches for studying reproduction in domestic and wild animal species.

*Front. Vet. Sci.* 12:1728981.

doi: 10.3389/fvets.2025.1728981

## COPYRIGHT

© 2026 Amelkina and Comizzoli. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Making sense of expanding transcriptomic data: network-based approaches for studying reproduction in domestic and wild animal species

Olga Amelkina<sup>1,2\*</sup> and Pierre Comizzoli<sup>2</sup>

<sup>1</sup>Department of Reproduction Biology, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany, <sup>2</sup>Smithsonian's National Zoo and Conservation Biology Institute, Washington, DC, United States

Transcriptomic datasets in animal reproductive biology are expanding rapidly, creating more opportunities to explore genome-phenome relationships, uncover biological mechanisms, and improve assisted reproductive technologies. This mini-review emphasizes the shift from single-gene analyses to a systems biology approach, where genes and pathways are studied within networks to capture their interactions and better understand biological systems. We show how network visualization can help synthesize knowledge from complex RNA-seq outputs and provide examples of tools and workflows suitable for species with different levels of data availability and annotation. Best practices for data generation and integration from various databases are discussed, highlighting the importance of high quality well-annotated datasets, transparent reporting, and the pitfalls of overinterpretation. Machine learning methods are explored as an analysis option for experiments with hundreds of data points. Ultimately, expanding available expression datasets for non-model species, combined with rigorous data processing and interpretation, will enable reproductive biologists to integrate network-based strategies into their research and advance reproductive science as well as conservation programs.

## KEYWORDS

machine learning, network visualization, non-model animal species, pathway enrichment, reproduction, transcriptomics

## 1 Introduction

High-throughput sequencing technologies are becoming affordable and accessible, resulting in an unprecedented rate of data accumulation, especially from transcriptomics studies. These datasets can help researchers investigate cellular mechanisms, uncover novel biological functions, identify associated molecular markers, and explore genome-phenome relationships (1–3). However, the analysis and interpretation of such complex, multidimensional data, defined by numerous interconnected variables, remain a major challenge, particularly in fields outside of cancer research where machine learning (ML) methods and network-based approaches have already been widely adopted (4, 5).

Genomics and systems biology share a common ground in understanding biological processes through integrative analysis, using experimental and computational methods (6). In essence, systems biology aims to study biological systems by first perturbing them, then measuring resulting gene, protein, or pathway responses, integrating the observed data, and ultimately modeling these data to describe the structure of the system and its

response to perturbations (7). Tools such as pathway enrichment, gene regulatory networks, and knowledge graphs offer powerful means to visualize data structure and interpret it within the broader context of biological function, while ML methods such as artificial neural networks can add predictive power (the ability to predict outcomes) and identify key drivers of observed transcriptional changes. Introducing these strategies more broadly into the field of animal reproduction biology can advance understanding of reproductive mechanisms and accelerate the development of assisted reproductive technologies (ARTs) in domestic and wildlife species. Our recent studies demonstrate the use of network visualization in interpreting transcriptomic data to study gonadal tissue development and response to preservation protocols in the domestic cat (8–10), and combination of transcriptomic and proteomic data to analyze semen composition and environmental response in the endangered black-footed ferret (*Mustela nigripes*) (11), highlighting the potential of this approach in fields of biobanking and conservation. Meanwhile, mice and cattle benefit from extensive annotation, vast tissue expression databases and high sample size to leverage ML methods to generate novel reproductive insights and fertility predictions (12–16).

The objective of this mini-review is to introduce network-based strategies for interrogating and integrating transcriptomic data to community of reproductive biologists, as well as advice on best practices and potential pitfalls. We provide an overview of databases, visualization tools and workflows suitable for model and non-model animal species, with a primary focus on bulk RNA sequencing (RNA-seq) studies. References to comprehensive reviews and analysis pipelines are included throughout, and readers are encouraged to consult these sources prior to undertaking large-scale analyses. While this review focuses on transcriptome, introduced strategies and tools can also be used when working with other types of omics data, such as genome, epigenome, proteome, metabolome, and lipidome. Ultimately, integrating various layers of data makes the research even more powerful (17, 18).

## 2 Transcriptomic data layers in reproduction

Transcriptomic data generated from RNA-seq provides a dynamic snapshot of gene activity and gives the ground for exploring various reproductive processes (19–21). Before introducing network-based analysis tools, it is essential to consider what each transcriptomic layer can reveal about the biological system and to outline practical strategies for generating high-quality data or sourcing relevant datasets from public repositories.

### 2.1 Coding transcriptome

The coding transcriptome represents mRNAs that are ultimately translated into proteins and drive cellular function. While bulk RNA-seq is used widely to profile gene expression across reproductive tissues and developmental stage, single cell (sc) and spatial RNA-seq (22) provides cell-level resolution and allows to dissect heterogeneity of reproductive tissues (21).

#### 2.1.1 Generating your own data

Careful experimental design is critical when generating your own transcriptomic data. Biological replication is one of the key considerations and should reflect expected variability in the population (23). Todd et al. provide an excellent guide for organizing your RNA-seq experiment in non-model species, connecting sample size, statistical power and effect size for various degrees of genetic variation (24). Sequencing depth is another important consideration and a good rule of thumb for mammalian transcriptome would be 150 paired-end sequencing with 30 million read depth per each biological replicate/library to get a comprehensive transcriptome data, while 10 million read depth may be enough for simple differential expression comparison of highly expressed genes (23, 25). Finally, RNA quality itself must be good enough as low-quality RNA results in uneven gene coverage, higher false-positive rates during differential expression analysis, higher duplication rates and negative correlation with library complexity (26, 27). For mammalian RNA, solutions to stabilize RNA such as RNAlater can help with preserving RNA integrity, especially during field collections (28). If there is no possibility of obtaining a good quality RNA, there are library preparation methods that were adapted to reduce the effect of RNA degradation (29, 30), as well as bioinformatic tools to account for bias toward shorter RNA species (31). Well-annotated reference genomes boost accurate alignment and interpretation (32), but if they are not available for particular wild species, comparing results of de-novo transcriptome assembly and alignment to the related species genome can help with initial investigation (33, 34).

#### 2.1.2 Utilizing publicly available datasets

Public repositories such as GEO (35) and SRA (36) host numerous RNA-seq datasets for reproductive tissues. For model species, there are also species-specific resources available, such as gene expression database GXD (37) and recount3 (38) for mouse, and cattle genotype-tissue expression atlas CattleGTEx (39). For non-model species, data availability is more limited and often requires searching individual publications, highlighting that comprehensive expression atlases and so called digital biobanks (40) for at least domestic models for wildlife species (such as domestic cat) are critically needed.

When reusing public data, good processing step is essential. To minimize batch effects and annotation drifts, raw reads can be re-quantified using a consistent pipeline and current reference genome. Converting raw reads to Transcripts Per Million (TPM) normalizes relative molar concentration of transcripts per sample and it is a baseline quantification recommended when working with RNA-seq (41), unless the method specifically uses raw reads (42). When integrating datasets from various sources, it's important to apply metadata standardization, such as tissue type, developmental stage, sequencing parameters, as well as batch correction methods, for example ComBat-seq (43) and svaseq (44).

### 2.2 Non-coding transcriptome

Beyond coding mRNAs, non-coding RNAs add critical regulatory layers in reproductive processes. Among these, microRNAs (miRNAs),

long non-coding (lnc) RNAs, and resulting competing endogenous (ce) RNAs are particularly relevant for network-based analyses.

### 2.2.1 MicroRNAs

miRNAs regulate gene expression primarily through mRNA degradation or translational inhibition (45). While degradation effects can be inferred from measured mRNA levels, translational inhibition often requires proteomic data for confirmation. Although miRNAs are highly tissue- and stage-specific (46), their high sequence and target conservation status across species helps with miRNA identification in non-model animals by utilizing orthologs of well-studied species (47).

While miRBase database is the most commonly used for identifying miRNAs (48), it is limited in its incomplete species coverage and naming inconsistencies. Another database MirGeneDB is hand curated and extremely reliable (49), however, the miRNA names from this database are rarely used in publications and chemical company catalogs. Tools like miRDeep that identify known and novel miRNAs, as well as orthologs based on the databases provided, are very useful in miRNA studies for both domestic and wild animal species (50). Separately, for mouse a whole miRNA tissue expression atlas (miRNATissueAtlas) is already available (51). Validated and predicted targets of miRNAs can either be sources from existing databases such as miRTarBase (52), miRDB (53), TargetScan (54) and miRWalk (55), or predicted for specific species based on uploaded 3'UTR sequences (54). miRmapper is a useful R package that allows to collect targets from most of the available databases (56).

### 2.2.2 Competing endogenous (ce) RNAs

The ceRNA hypothesis, first proposed by Salmena et al. in 2011 (57), describes how lncRNAs, circular (circ) RNAs and pseudogenes compete for shared miRNAs, indirectly regulating each other's expression. lncRNAs can act as miRNA sponges to prevent them from binding to their target mRNA, while circRNAs influence transcriptional and post-transcriptional regulation by interacting with spliceosomal components (58).

Databases for lncRNAs and circRNAs exist primarily for model species (59), and low sequence conservation limits cross-species applicability. However, computational prediction of miRNA binding sites on lncRNAs and circRNAs allows to construct ceRNA network in non-model species (60, 61).

## 3 Network-based approaches for biological interpretation

Network-based approaches provide a powerful tool for interpreting complex transcriptomic data by representing biological systems as interconnected entities (62). In these models, nodes usually correspond to genes, proteins, or pathways, while edges represent relationships between the nodes such as co-expression, physical or functional interaction, or regulation. Edges can be undirected, e.g., correlation based, or directed, indicating causal interactions with a sign (activation or inhibition) and context (phosphorylation, transcriptional activation, repression) (63). Visualization tools such as Cytoscape allow one to build their own networks, explore them with various layouts and add additional data layers, integrating topology with functional annotations (64, 65).

## 3.1 Pathway enrichment and visualization

RNA-seq experiments often produce extensive lists of differentially expressed genes (DEGs), which are difficult to interpret through manual literature review alone. Functional enrichment analysis provides a systematic approach by identifying statistically overrepresented biological pathways and function (66, 67). However, enrichment outputs can themselves be overwhelming, often comprising long, redundant lists of terms. Network visualization mitigates this by clustering related gene sets into coherent themes, reducing redundancy and helping to grasp the overall picture of enrichment results.

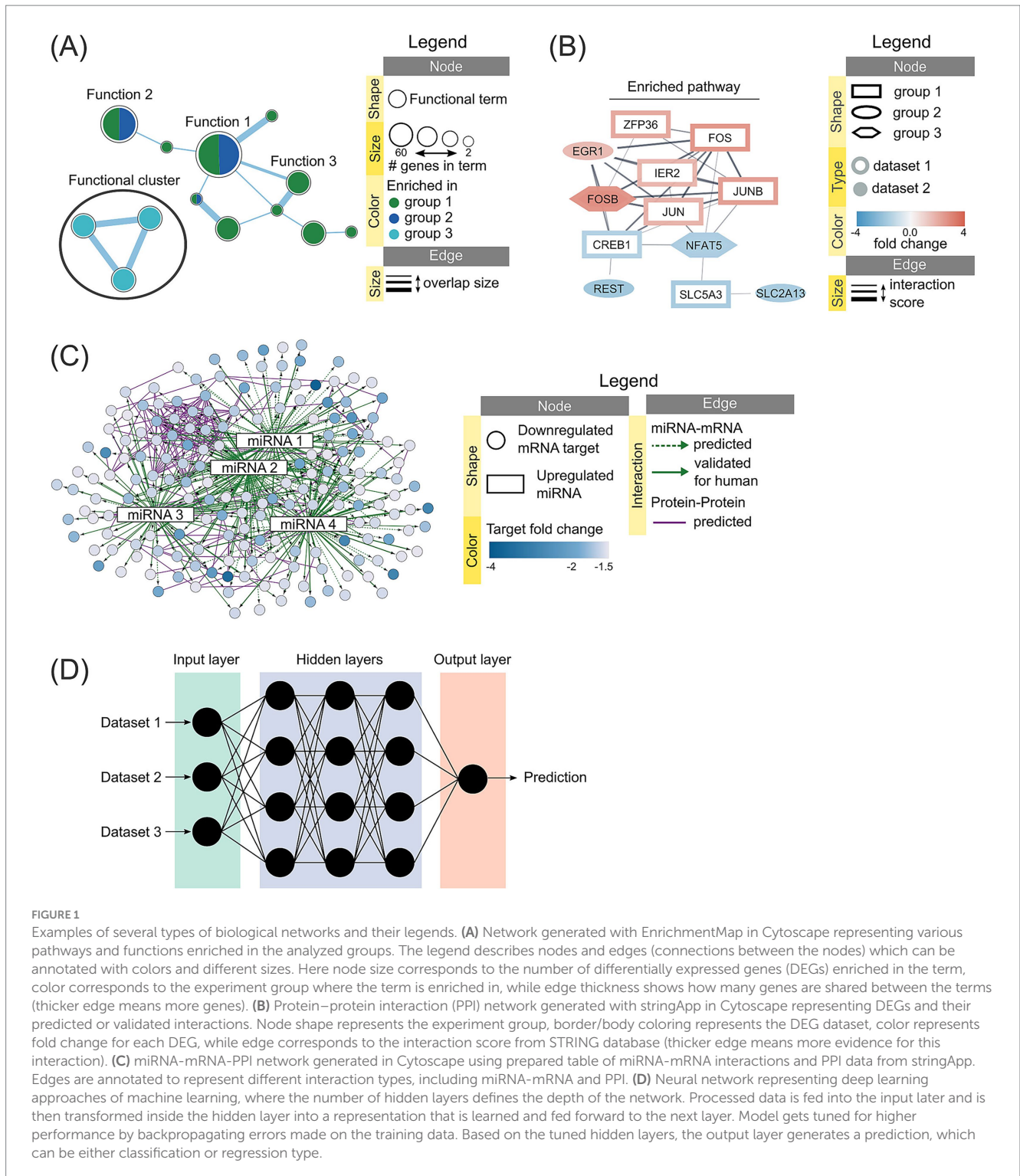
EnrichmentMap (68), available as a Cytoscape app (69) and also as a web tool (70), is a tool of choice for this. It organizes enrichment results into networks where nodes are enriched terms and edges represent gene overlap between the terms, see example in Figure 1A. Additional Cytoscape apps such as clusterMaker (71), WordCloud (72), and AutoAnnotate (73) further enhance interpretability by grouping and labeling clusters. This protocol (74) provides a great example of enrichment analysis pipeline with EnrichmentMap visualization.

The choice of enrichment tool depends on species coverage and the statistical approach, i.e., predefined gene set vs. ranked gene list from differential expression analysis. For instance, DAVID (75) performs gene set enrichment and supports annotations for over 50 species, including domestic cat and ferret, making it a good choice for non-model species analysis. Meanwhile, GSEA (66) operates on ranked gene list but is limited to human and mouse annotations, relying on the MSigDB database; therefore, analysis in other species require orthology mapping. Finally, g:Profiler (76) supports over 800 species via Ensembl and allows custom annotations, making it another great choice for non-model species analysis. Outputs from all three of the above enrichment tools can be then visualized with EnrichmentMap. Ideally, species-specific databases should be used to avoid missing lineage-specific genes. On the other hand, human and mouse annotation databases provide more functional information and can be used to expand the results and generate additional hypotheses. However, this should be done carefully and only in addition to the species-specific annotation to not over-interpret your results: some species-specific genes, particularly those involved in immune and reproductive functions, may lack direct counterparts in human or mouse, limiting transferability. For a broader overview of gene set enrichment analysis see (77).

Beyond standard enrichment analysis, topology-aware methods consider the actual structure and direction of pathways. For example SPIA (78) combines classical enrichment with the analysis of how a pathway is perturbed under specific conditions, and it is available for numerous species annotated in KEGG database, including domestic cat and ferret. Other approaches, such as PROGENy (79), estimate pathway activity based on downstream gene responses, while CARNIVAL (80) reconstructs causal networks by linking interactions that have both direction and sign. These methods are particularly useful for studying signaling pathways, but are often limited to human and mouse, requiring orthology mapping for other species.

## 3.2 Protein–protein interaction (PPI) networks

Unlike traditional approaches that focus on individual DEGs, network-based strategies capture gene interactions, which have been



shown to be more predictive of phenotype than single-gene markers (81). PPI networks position DEGs within interaction maps, revealing functional clusters and signaling cascades, and including such analysis on top of the initial pathway enrichment can add confidence in the observed data trends. STRING database integrates experimental and predictive evidence for both functional and physical interactions of proteins for many species (82, 83), and StringApp in Cytoscape (84) allows to map additional information to the network, see example in Figure 1B. Mice studies can utilize GeneMANIA (85) and its

Cytoscape app (86) which collects numerous interactions from various databases, and expand it by addition of transcription factor binding information from manually curated TRRUST v2 database (87). For cattle, domestic cat and ferret, AnimalTFDB 4.0 database (88) can be used for predicted transcription factors.

When PPI networks are too large they resemble so called “hairballs” and require the use of clustering methods to improve network readability (89). Cytoscape apps such as clusterMaker (71) and MCODE (90) can extract PPI clusters based on their interaction

score, which can then be analyzed for functional enrichment (e.g., integrated function in StringApp) to improve biological interpretability. Utilizing human orthologs, CORUM database (91) can also be used to extract known mammalian protein complexes.

### 3.3 Gene regulatory networks (GRNs) and integration strategies

GRNs can model relationships between regulators (transcription factors, miRNA) and targets (mRNA) and are a powerful abstraction of biological systems (62, 63); see example in Figure 1C. Zhao et al. (92) provide an in-depth overview of tools for inferring GRNs from various types of expression data, dividing them into model-based, information-based and machine learning-based methods. Model-based methods, including differential equation, Boolean and Bayesian methods, are suitable for inferring small real networks, while information theory-based methods, including Pearson correlation coefficient and (conditional/part) mutual information, are suitable for steady-state data (92). Machine learning-based methods, such as widely used tool GENIE3 that utilizes random forests to infer GRN (93) and its adaptation for time-series expression data dynGENIE3 (94), provide best results when reconstructing large-scale networks. In addition, reverse engineering approaches have been developed to trace back the initial relationships between genes that resulted in the observed gene expression (62). Mercatelli et al. (95) go through a variety of tools available for GRN inference, and highlight that the optimal tool selection ultimately relies on the biological context studied and data availability on the species, transcription factors, cellular context or specific perturbation.

Integrating strategies can extend GRNs to multi-layer networks by incorporating miRNA-mRNA interactions, ceRNA relationships, genomic variants and other omics data layers. Morabito et al. (96) provide a useful overview of recent algorithms and tools applied to integrate genomics, transcriptomics, proteomics, and metabolomics, while a more ML oriented approach can be found in the review of Picard et al. (97).

### 3.4 Machine learning and neural networks

Machine learning uses computation to recognize patterns in the data by fitting predictive models to it or identifying informative clustering within data (98). This approach allows scientists to make predictions where experimental data is lacking to guide future research and to improve the understanding of biological systems. Greener et al. (98) provide an essential guide to ML for biologists that is a good starting point, which can further be expanded with practical advices from Chicco (99) and more in-depth review of use for biological networks from Camacho et al. (100).

ML methods are divided into supervised, where a model is fitted to labeled data, and unsupervised, where a model identifies patterns in unlabeled data (98). Apart from traditional ML, an area of deep learning that relies on neural networks has been rapidly developing for genomics analyses (101, 102), see example in Figure 1D. While deep learning is a powerful tool, it is limited to specific applications where a large amount of highly structured data is available, i.e., each data point has many features with clear relationship (103). Generally, traditional ML should be the starting point to find the most

appropriate method for a given analysis and in some cases outperforms deep representation learning in phenotype prediction from transcriptomics data (81). At the same time, interpretable ML for omics data is becoming more prevalent in systems biology (104).

Because ML typically requires large amount of data points, from hundreds and thousands for traditional ML to millions for deep learning, as well as training and validation sets for supervised methods, this approach may not always be applicable to bulk transcriptomic data in non-model species. When sample size is limited, applying unsupervised ML methods, such as t-SNE (105) and UMAP (106), to scRNA-seq data can enable detailed profiling of reproductive tissues and their cellular environment, providing foundation for *in vitro* system development and generation of further hypotheses. Here single-nuclei RNA-seq approaches are particularly useful in field conditions, as tissues can be snap-frozen in liquid nitrogen without immediate dissociation (107). In addition, genomics data is growing fast for wild species and opens possibilities to study evolutionary trends in reproduction with the use of ML in the future, including miRNA emergence in placental animals (108) and effect of deleterious mutations on species fitness (109, 110).

## 4 Limitations when working with omics data

Despite the promise of transcriptomics, ML and network-based approaches, several challenges constrain their interpretability and predictive power. Figure 2 illustrates how workflows differ between a well-studied species (mouse and cattle) and non-model species (domestic cat and black-footed ferret), highlighting the impact of database availability.

### 4.1 “Garbage in, garbage out”: the incentive for high-quality data

The principle of “garbage in, garbage out” is borrowed from computer science and can be applied directly to omics analysis: flawed input data inevitably produce misleading predictions and interpretations. For transcriptomics-based phenotype prediction, proper normalization and robust regression methods are essential, but the factors that have the biggest impact still consist of adequate biological replications and sequencing depth, ensuring accurate sample metadata, incorporating complementary data types such as proteomics, and improved prior knowledge. Paton et al. demonstrated that preprocessing choices significantly affect downstream enrichment results (111). In order to trust the results of network inference and be able to apply ML methods in a study, researchers must produce/utilize high-quality raw data and provide transparent reporting of quality control and processing.

### 4.2 The “curse of dimensionality” and strategies to overcome it

Termed as the “curse of dimensionality,” a common problem in omics datasets is that the number of features is much higher than the number of samples. This leads to overfitting and poor generalization in ML models (104). Dimensionality reduction through feature extraction, selection or engineering can reduce the number of

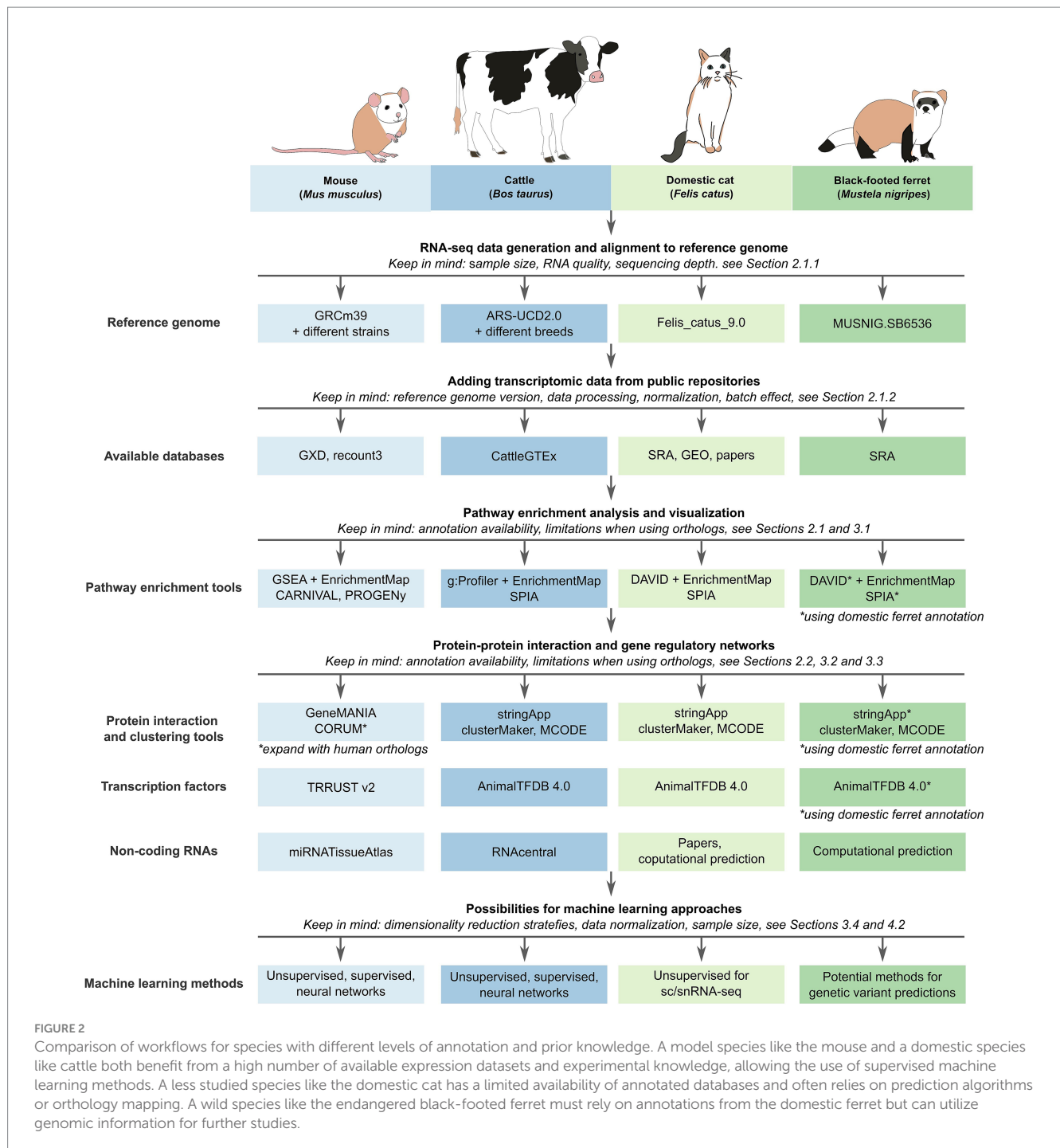


FIGURE 2

Comparison of workflows for species with different levels of annotation and prior knowledge. A model species like the mouse and a domestic species like cattle both benefit from a high number of available expression datasets and experimental knowledge, allowing the use of supervised machine learning methods. A less studied species like the domestic cat has a limited availability of annotated databases and often relies on prediction algorithms or orthology mapping. A wild species like the endangered black-footed ferret must rely on annotations from the domestic ferret but can utilize genomic information for further studies.

variables (data sparsity) and improve prediction reliability (112). Lasso regularized models and nearest shrunken centroids introduce sparsity by removing contributions of unimportant features (113), as used in the example of combining transcriptomics and genetic variants to predict phenotype from genotype (114).

### 4.3 Overinterpretation problem

While omics data and its integration into network-based or ML approaches are powerful methods to infer observed changes in

biological systems, they can also lead to overinterpretation when patterns are mistaken for causal relationships. Hub nodes, for example, often reflect annotation density rather than true biological centrality. Similarly, inferred regulatory relationships of nodes via edges should be treated as hypotheses rather than definitive interactions, particularly in non-model species where prior knowledge is sparse. It is always best practice to report data quality and confidence scores, be careful when discussing results and, when possible, validate predictions experimentally. Ultimately, insight gained from computational analysis and network visualization should rather guide hypothesis formulation and future functional studies but never replace them.

## 5 Future directions

Future progress in reproductive biology depends on closer integration with the genomics community to connect fertility and reproductive adaptations to genomic variation. While genome-to-phenome mapping and fertility prediction are rapidly advancing in model species and livestock through genome/transcriptome-wide association studies and (e)QTL approaches, non-model species and wildlife remain behind, with genomic data rarely linked to reproductive phenotypes. Emerging genomics studies are applying genetic load estimates to guide breeding strategies and reduce inbreeding depression in captive populations (115). Integrating fertility and expression data into such approaches will enhance their efficiency and strengthen conservation breeding efforts. Network-based approaches provide a natural framework for this integration.

To achieve this, we will need more reproductive expression datasets and species-specific digital biobanks where genomics meets transcriptomics, proteomics, and ultimately fertility phenomics – “fertilomics.” These resources, combined with multi-omics integration, advanced network visualization tools, and ML, will enable predictive modeling and comparative analyses across evolutionary contexts.

## 6 Conclusion

Network-based approaches provide a powerful framework for interpreting transcriptomic data, enabling integration of complex datasets, hypothesis generation, and guiding functional studies. Adopting the strategies discussed in this mini-review would drive the transition from single-gene interpretations to systems-level approaches in reproductive studies. Expanding available expression datasets for non-model species, combined with best practices in data quality, processing and careful interpretation of results will allow the reproductive biology community to efficiently integrate network-based and ML strategies into their research and advance evolutionary and reproduction research, ARTs and conservation programs.

## Author contributions

OA: Conceptualization, Writing – original draft, Writing – review & editing, Visualization. PC: Writing – original draft, Writing – review & editing.

## References

- Wang, Z, Gerstein, M, and Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* (2009) 10:57–63. doi: 10.1038/nrg2484
- Stark, R, Grzelak, M, and Hadfield, J. RNA sequencing: the teenage years. *Nat Rev Genet.* (2019) 20:631–56. doi: 10.1038/s41576-019-0150-2
- Karczewski, KJ, and Snyder, MP. Integrative omics for health and disease. *Nat Rev Genet.* (2018) 19:299–310. doi: 10.1038/nrg.2018.4
- Hussen, BM, Abdullah, SR, Hidayat, HJ, Samsami, M, and Taheri, M. Integrating AI and RNA biomarkers in cancer: advances in diagnostics and targeted therapies. *Cell Commun Signal.* (2025) 23:430. doi: 10.1186/s12964-025-02434-2
- Wekesa, JS, and Kimwele, M. A review of multi-omics data integration through deep learning approaches for disease diagnosis, prognosis, and treatment. *Front Genet.* (2023) 14:1199087. doi: 10.3389/fgene.2023.1199087
- Conesa, A, and Mortazavi, A. The common ground of genomics and systems biology. *BMC Syst Biol.* (2014) 8:S1. doi: 10.1186/1752-0509-8-S2-S1
- Ideker, T, Galitski, T, and Hood, L. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet.* (2001) 2:343–72. doi: 10.1146/annurev.genom.2.1.343
- Amelkina, O, and Comizzoli, P. Initial response of ovarian tissue transcriptome to vitrification or microwave-assisted dehydration in the domestic cat model. *BMC Genomics.* (2020) 21:828. doi: 10.1186/s12864-020-07236-z
- Amelkina, O, Silva, A. M.da, Silva, A. R., and Comizzoli, P. Transcriptome dynamics in developing testes of domestic cats and impact of age on tissue resilience to cryopreservation *BMC Genomics* 2021 22:847 doi: 10.1186/s12864-021-08099-8 PMID: 34814833
- Amelkina, O, da Silva, AM, Silva, AR, and Comizzoli, P. Feline microRNAome in ovary and testis: exploration of in-silico miRNA-mRNA networks involved in gonadal function and cellular stress response. *Front Genet.* (2022) 13:1009220. doi: 10.3389/fgene.2022.1009220

## Funding

The author(s) declared that financial support was not received for this work and/or its publication.

## Acknowledgments

We are grateful to IZW library for providing journal access. We thank Margot Shimoura for inspiring discussions and guidance.

## Conflict of interest

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

## Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by *Frontiers* with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## 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.

11. Ali, N, Amelkina, O, Santymire, RM, Koepfli, KP, Comizzoli, P, and Vazquez, JM. Semen proteome and transcriptome of the endangered black-footed ferret (*Mustela nigripes*) show association with the environment and fertility outcome. *Sci Rep.* (2024) 14:7063. doi: 10.1038/s41598-024-57096-w
12. Ito, K, Hirakawa, T, Shigenobu, S, Fujiyoshi, H, and Yamashita, T. Mouse-Geneformer: a deep learning model for mouse single-cell transcriptome and its cross-species utility. *PLoS Genet.* (2025) 21:e1011420. doi: 10.1371/journal.pgen.1011420
13. Finnerty, RM, Carulli, DJ, Hedge, A, Wang, Y, Boadu, F, Winuthayanon, S, et al. Multi-omics analyses and machine learning prediction of oviductal responses in the presence of gametes and embryos. *eLife.* (2025) 13:RP100705. doi: 10.7554/eLife.100705.3
14. Wang, X, Shi, S, Ali Khan, MY, Zhang, Z, and Zhang, Y. Improving the accuracy of genomic prediction in dairy cattle using the biologically annotated neural networks framework. *J Anim Sci Biotechnol.* (2024) 15:87. doi: 10.1186/s40104-024-01044-1
15. Rabaglino, MB, Salilew-Wondim, D, Zolini, A, Tesfaye, D, Hoelker, M, Lonergan, P, et al. Machine-learning methods applied to integrated transcriptomic data from bovine blastocysts and elongating conceptuses to identify genes predictive of embryonic competence. *FASEB J.* (2023) 37:e22809. doi: 10.1096/fj.202201977R
16. Hoorn, QA, Rabaglino, MB, Amaral, TF, Maia, TS, Yu, F, Cole, JB, et al. Machine learning to identify endometrial biomarkers predictive of pregnancy success following artificial insemination in dairy cow†. *Biol Reprod.* (2024) 111:54–62. doi: 10.1093/biolre/iaoe052
17. Okamoto, J, Yin, X, Ryan, B, Chiou, J, Luca, F, Pique-Regi, R, et al. Multi-INTACT: integrative analysis of the genome, transcriptome, and proteome identifies causal mechanisms of complex traits. *Genome Biol.* (2025) 26:19. doi: 10.1186/s13059-025-03480-2
18. Han, B, Tian, D, Li, X, Liu, S, Tian, F, Liu, D, et al. Multiomics analyses provide new insight into genetic variation of reproductive adaptability in Tibetan sheep. *Mol Biol Evol.* (2024) 41:msae058. doi: 10.1093/molbev/msae058
19. Suzuki, S, Diaz, VD, and Hermann, BP. What has single-cell RNA-seq taught us about mammalian spermatogenesis? *Biol Reprod.* (2019) 101:617–34. doi: 10.1093/biolre/iz088
20. He, C, Wang, K, Gao, Y, Wang, C, Li, L, Liao, Y, et al. Roles of noncoding RNA in reproduction. *Front Genet.* (2021) 12:777510. doi: 10.3389/fgene.2021.777510
21. Zhang, X, Cao, Q, Rajachandran, S, Grow, EJ, Evans, M, and Chen, H. Dissecting mammalian reproduction with spatial transcriptomics. *Hum Reprod Update.* (2023) 29:794–810. doi: 10.1093/humupd/dmad017
22. Islam, S, Zeisel, A, Joost, S, La Manno, G, Zajac, P, Kasper, M, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods.* (2014) 11:163–6. doi: 10.1038/nmeth.2772
23. Conesa, A, Madrigal, P, Tarazona, S, Gomez-Cabrero, D, Cervera, A, McPherson, A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* (2016) 17:13. doi: 10.1186/s13059-016-0881-8
24. Todd, EV, Black, MA, and Gemmill, NJ. The power and promise of RNA-seq in ecology and evolution. *Mol Ecol.* (2016) 25:1224–41. doi: 10.1111/mec.13526
25. Liu, Y, Zhou, J, and White, KP. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics.* (2014) 30:301–4. doi: 10.1093/bioinformatics/btt688
26. Wang, L, Nie, J, Sciotte, H, Li, Y, Eckel-Passow, JE, Dasari, S, et al. Measure transcript integrity using RNA-seq data. *BMC Bioinformatics.* (2016) 17:58. doi: 10.1186/s12859-016-0922-z
27. Gallego Romero, I, Pai, AA, Tung, J, and Gilad, Y. RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol.* (2014) 12:42. doi: 10.1186/1741-7007-12-42
28. Hatzis, C, Sun, H, Yao, H, Hubbard, RE, Meric-Bernstam, F, Babiera, GV, et al. Effects of tissue handling on RNA integrity and microarray measurements from resected breast cancers. *JNCI J Natl Cancer Inst.* (2011) 103:1871–83. doi: 10.1093/jnci/djr438
29. Adiconis, X, Borges-Rivera, D, Satija, R, DeLuca, DS, Busby, MA, Berlin, AM, et al. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods.* (2013) 10:623–9. doi: 10.1038/nmeth.2483
30. Schuierer, S, Carbone, W, Knehr, J, Petitjean, V, Fernandez, A, Sultan, M, et al. A comprehensive assessment of RNA-seq protocols for degraded and low-quantity samples. *BMC Genomics.* (2017) 18:442. doi: 10.1186/s12864-017-3827-y
31. Xiong, B, Yang, Y, Fineis, FR, and Wang, JP. DegNorm: normalization of generalized transcript degradation improves accuracy in RNA-seq analysis. *Genome Biol.* (2019) 20:75. doi: 10.1186/s13059-019-1682-7
32. Hall, NAL, Carlyle, BC, Haerty, W, and Tunbridge, EM. Roadblock: improved annotations do not necessarily translate into new functional insights. *Genome Biol.* (2021) 22:320. doi: 10.1186/s13059-021-02542-5
33. Hölzer, M. A decade of de novo transcriptome assembly: are we there yet? *Mol Ecol Resour.* (2021) 21:11–3. doi: 10.1111/1755-0998.13268
34. Li, B, and Dewey, CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* (2011) 12:323. doi: 10.1186/1471-2105-12-323
35. Clough, E, Barrett, T, Wilhite, SE, Ledoux, P, Evangelista, C, Kim, IF, et al. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res.* (2024) 52:D138–44. doi: 10.1093/nar/gkad965
36. Sayers, EW, O'Sullivan, C, and Karsch-Mizrachi, I. Using GenBank and SRA. *Methods Mol Biol Clifton NJ.* (2022) 2443:1–25. doi: 10.1007/978-1-0716-2067-0\_1
37. Baldarelli, RM, Smith, CM, Finger, JH, Hayamizu, TF, McCright, JJ, Xu, J, et al. The mouse gene expression database (GXD): 2021 update. *Nucleic Acids Res.* (2021) 49:D924–31. doi: 10.1093/nar/gkaa914
38. Wilks, C, Zheng, SC, Chen, FY, Charles, R, Solomon, B, Ling, JP, et al. recount3: summaries and queries for large-scale RNA-seq expression and splicing. *Genome Biol.* (2021) 22:323. doi: 10.1186/s13059-021-02533-6
39. Liu, S, Gao, Y, Canela-Xandri, O, Wang, S, Yu, Y, Cai, W, et al. A multi-tissue atlas of regulatory variants in cattle. *Nat Genet.* (2022) 54:1438–47. doi: 10.1038/s41588-022-01153-5
40. Brancato, V, Esposito, G, Coppola, L, Cavaliere, C, Mirabelli, P, Scapicchio, C, et al. Standardizing digital biobanks: integrating imaging, genomic, and clinical data for precision medicine. *J Transl Med.* (2024) 22:136. doi: 10.1186/s12967-024-04891-8
41. Wagner, GP, Kin, K, and Lynch, VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci Theor Den Biowissenschaften.* (2012) 131:281–5. doi: 10.1007/s12064-012-0162-3
42. Love, MI, Huber, W, and 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
43. Zhang, Y, Parmigiani, G, and Johnson, WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genomics Bioinform.* (2020) 2:lqaa078. doi: 10.1093/nargab/lqaa078
44. Leek, JT. Svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Res.* (2014) 42:e161. doi: 10.1093/nar/gku864
45. O'Brien, J, Hayder, H, Zayed, Y, and Peng, C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Front Endocrinol.* (2018) 9:402. doi: 10.3389/fendo.2018.00402
46. Reza, AMMT, Choi, YJ, Han, SG, Song, H, Park, C, Hong, K, et al. Roles of microRNAs in mammalian reproduction: from the commitment of germ cells to peri-implantation embryos. *Biol Rev Camb Philos Soc.* (2019) 94:415–38. doi: 10.1111/brv.12459
47. Friedman, RC, Farh, KKH, Burge, CB, and Bartel, DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* (2009) 19:92–105. doi: 10.1101/gr.082701.108
48. Kozomara, A, Birgaoanu, M, and Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* (2019) 47:D155–62. doi: 10.1093/nar/gky1141
49. Fromm, B, Høye, E, Domanska, D, Zhong, X, Aparicio-Puerta, E, Ovchinnikov, V, et al. MirGeneDB 2.1: toward a complete sampling of all major animal phyla. *Nucleic Acids Res.* (2022) 50:D204–10. doi: 10.1093/nar/gkab1101
50. An, J, Lai, J, Lehman, ML, and Nelson, CC. miRDeep\*: an integrated application tool for miRNA identification from RNA sequencing data. *Nucleic Acids Res.* (2013) 41:727–37. doi: 10.1093/nar/gks1187
51. Rishik, S, Hirsch, P, Grandke, F, Fehlmann, T, and Keller, A. miRNATissueAtlas 2025: an update to the uniformly processed and annotated human and mouse non-coding RNA tissue atlas. *Nucleic Acids Res.* (2025) 53:D129–37. doi: 10.1093/nar/gkae1036
52. Cui, S, Yu, S, Huang, HY, Lin, YCD, Huang, Y, Zhang, B, et al. miRTarBase 2025: updates to the collection of experimentally validated microRNA–target interactions. *Nucleic Acids Res.* (2025) 53:D147–56. doi: 10.1093/nar/gkae1072
53. Chen, Y, and Wang, X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* (2020) 48:D127–31. doi: 10.1093/nar/gkz757
54. Agarwal, V, Bell, GW, Nam, JW, and Bartel, DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife.* (2015) 4:e05005. doi: 10.7554/eLife.05005
55. Sticht, C, De La Torre, C, Parveen, A, and Gretz, N. miRWalk: An online resource for prediction of microRNA binding sites. *PLoS One.* (2018) 13:e0206239. doi: 10.1371/journal.pone.0206239
56. da Silveira, WA, Renaud, L, Simpson, J, Glen, WB, Hazard, ES, Chung, D, et al. miRMapper: a tool for interpretation of miRNA–mRNA interaction networks. *Genes.* (2018) 9:458. doi: 10.3390/genes9090458
57. Salmena, L, Poliseno, L, Tay, Y, Kats, L, and Pandolfi, PP. A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? *Cell.* (2011) 146:353–8. doi: 10.1016/j.cell.2011.07.014
58. Gao, Y, Takenaka, K, Xu, SM, Cheng, Y, and Janitz, M. Recent advances in investigation of circRNA/lncRNA-miRNA-mRNA networks through RNA sequencing data analysis. *Brief Funct Genomics.* (2025) 24:elaf005. doi: 10.1093/bfpg/elaf005
59. Sweeney, BA, Tagmazian, AA, Ribas, CE, Finn, RD, Bateman, A, and Petrov, AI. Exploring non-coding RNAs in RNAcentral. *Curr Protoc Bioinformatics.* (2020) 71:e104. doi: 10.1002/cpbi.104
60. Dori, M, Caroli, J, and Forcato, M. Circr, a computational tool to identify miRNA:circRNA associations. *Front Bioinform.* (2022) 2:852834. doi: 10.3389/fbinf.2022.852834
61. Yang, T, He, Y, and Wang, Y. Introducing TEC-LncMir for prediction of lncRNA-miRNA interactions through deep learning of RNA sequences. *Brief Bioinform.* (2025) 26:bbaf046. doi: 10.1093/bib/bbaf046
62. Cutello, V, Pavone, M, and Zito, F. Inferring a gene regulatory network from gene expression data. An overview of best methods and a reverse engineering approach In:

- D Cantone and A Pulvirenti, editors. From computational logic to computational biology: Essays dedicated to Alfredo Ferro to celebrate his scientific career [internet]. Cham: Springer Nature Switzerland (2024)
63. Huynh-Thu, VA, and Sanguinetti, G. Gene regulatory network inference: An introductory survey In: G Sanguinetti and VA Huynh-Thu, editors. Gene regulatory networks: Methods and protocols [internet]. New York, NY: Springer (2019)
64. Cline, MS, Smoot, M, Cerami, E, Kuchinsky, A, Landys, N, Workman, C, et al. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc.* (2007) 2:2366–82. doi: 10.1038/nprot.2007.324
65. 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
66. Subramanian, A, Tamayo, P, Mootha, VK, Mukherjee, S, Ebert, BL, Gillette, MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
67. Zhao, K, and Rhee, SY. Interpreting omics data with pathway enrichment analysis. *Trends Genet.* (2023) 39:308–19. doi: 10.1016/j.tig.2023.01.003
68. Merico, D, Isserlin, R, Stueker, O, Emili, A, and Bader, GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One.* (2010) 5:e13984. doi: 10.1371/journal.pone.0013984
69. Isserlin, R, Merico, D, Voisin, V, and Bader, GD. Enrichment map - a Cytoscape app to visualize and explore OMICs pathway enrichment results. *F1000Res.* (2014) 3:141. doi: 10.12688/f1000research.4536.1
70. Franz, M, Lopes, CT, Kucera, M, Voisin, V, Isserlin, R, and Bader, GD. Gene-set enrichment analysis and visualization on the web using EnrichmentMap:RNASeq. *Bioinforma Adv.* (2025) 5:vbafl78. doi: 10.1093/bioadv/vbaf178
71. Morris, JH, Apeltsin, L, Newman, AM, Baumbach, J, Wittkop, T, Su, G, et al. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics.* (2011) 12:436. doi: 10.1186/1471-2105-12-436
72. Oesper, L, Merico, D, Isserlin, R, and Bader, GD. WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. *Source Code Biol Med.* (2011) 6:7. doi: 10.1186/1751-0473-6-7
73. Kucera, M, Isserlin, R, Arkhangorodsky, A, and Bader, GD. AutoAnnotate: a Cytoscape app for summarizing networks with semantic annotations. *F1000Res.* (2016) 5:1717. doi: 10.12688/f1000research.9090.1
74. Reimand, J, Isserlin, R, Voisin, V, Kucera, M, Tannus-Lopes, C, Rostamianfar, A, et al. Pathway enrichment analysis and visualization of omics data using g:profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc.* (2019) 14:482–517. doi: 10.1038/s41596-018-0103-9
75. Huang, DW, Sherman, BT, Tan, Q, Kir, J, Liu, D, Bryant, D, et al. DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* (2007) 35:W169–75. doi: 10.1093/nar/gkm415
76. Reimand, J, Kull, M, Peterson, H, Hansen, J, and Vilo, J. G:profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* (2007) 35:W193–200. doi: 10.1093/nar/gkm226
77. Candia, J, and Ferrucci, L. Assessment of gene set enrichment analysis using curated RNA-seq-based benchmarks. *PLoS One.* (2024) 19:e0302696. doi: 10.1371/journal.pone.0302696
78. Tarca, AL, Draghici, S, Khatri, P, Hassan, SS, Mittal, P, Kim, J, et al. A novel signaling pathway impact analysis. *Bioinformatics.* (2009) 25:75–82. doi: 10.1093/bioinformatics/btn577
79. Schubert, M, Klinger, B, Klünemann, M, Sieber, A, Uhlitz, F, Sauer, S, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun.* (2018) 9:20. doi: 10.1038/s41467-017-02391-6
80. Liu, A, Trairatphisan, P, Gjerga, E, Didangelos, A, Barratt, J, and Saez-Rodriguez, J. From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. *NPJ Syst Biol Appl.* (2019) 5:40. doi: 10.1038/s41540-019-0118-z
81. Smith, AM, Walsh, JR, Long, J, Davis, CB, Henstock, P, Hodge, MR, et al. Standard machine learning approaches outperform deep representation learning on phenotype prediction from transcriptomics data. *BMC Bioinformatics.* (2020) 21:119. doi: 10.1186/s12859-020-3427-8
82. Szklarczyk, D, Morris, JH, Cook, H, Kuhn, M, Wyder, S, Simonovic, M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* (2017) 45:D362–8. doi: 10.1093/nar/gkw937
83. Szklarczyk, D, Franceschini, A, Wyder, S, Forslund, K, Heller, D, Huerta-Cepas, J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* (2015) 43:D447–52. doi: 10.1093/nar/gku1003
84. Doncheva, NT, Morris, JH, Gorodkin, J, and Jensen, LJ. Cytoscape StringApp: network analysis and visualization of proteomics data. *J Proteome Res.* (2019) 18:623–32. doi: 10.1021/acs.jproteome.8b00702
85. Franz, M, Rodriguez, H, Lopes, C, Zuberi, K, Montojo, J, Bader, GD, et al. GeneMANIA update 2018. *Nucleic Acids Res.* (2018) 46:W60–4. doi: 10.1093/nar/gky311
86. Montojo, J, Zuberi, K, Rodriguez, H, Kazi, F, Wright, G, Donaldson, SL, et al. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics.* (2010) 26:2927–8. doi: 10.1093/bioinformatics/btq562
87. Han, H, Cho, JW, Lee, S, Yun, A, Kim, H, Bae, D, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* (2018) 46:D380–6. doi: 10.1093/nar/gkx1013
88. Shen, WK, Chen, SY, Gan, ZQ, Zhang, YZ, Yue, T, Chen, MM, et al. AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. *Nucleic Acids Res.* (2023) 51:D39–45. doi: 10.1093/nar/gkac907
89. Brohée, S, and van Helden, J. Evaluation of clustering algorithms for protein–protein interaction networks. *BMC Bioinformatics.* (2006) 7:488. doi: 10.1186/1471-2105-7-488
90. Bader, GD, and Hogue, CWV. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* (2003) 4:2. doi: 10.1186/1471-2105-4-2
91. Tsitsiridis, G, Steinkamp, R, Giurgiu, M, Brauner, B, Fobo, G, Frishman, G, et al. CORUM: the comprehensive resource of mammalian protein complexes-2022. *Nucleic Acids Res.* (2023) 51:D539–45. doi: 10.1093/nar/gkac1015
92. Zhao, M, He, W, Tang, J, Zou, Q, and Guo, F. A comprehensive overview and critical evaluation of gene regulatory network inference technologies. *Brief Bioinform.* (2021) 22:bbab009. doi: 10.1093/bib/bbab009
93. Huynh-Thu, VA, Irrthum, A, Wehenkel, L, and Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One.* (2010) 5:e12776. doi: 10.1371/journal.pone.0012776
94. Huynh-Thu, VA, and Geurts, P. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. *Sci Rep.* (2018) 8:3384. doi: 10.1038/s41598-018-21715-0
95. Mercatelli, D, Scalambra, L, Triboli, L, Ray, F, and Giorgi, FM. Gene regulatory network inference resources: a practical overview. *Biochimica et Biophysica Acta (BBA).* (2020) 1863:194430. doi: 10.1016/j.bbaggm.2019.194430
96. Morabito, A, De Simone, G, Pastorelli, R, Brunelli, L, and Ferrario, M. Algorithms and tools for data-driven omics integration to achieve multilayer biological insights: a narrative review. *J Transl Med.* (2025) 23:425. doi: 10.1186/s12967-025-06446-x
97. Picard, M, Scott-Boyer, MP, Bodein, A, Périn, O, and Droit, A. Integration strategies of multi-omics data for machine learning analysis. *Comput Struct Biotechnol J.* (2021) 19:3735–46. doi: 10.1016/j.csbj.2021.06.030
98. Greener, JG, Kandathil, SM, Moffat, L, and Jones, DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol.* (2022) 23:40–55. doi: 10.1038/s41580-021-00407-0
99. Chicco, D. Ten quick tips for machine learning in computational biology. *BioData Min.* (2017) 10:35. doi: 10.1186/s13040-017-0155-3
100. Camacho, DM, Collins, KM, Powers, RK, Costello, JC, and Collins, JJ. Next-generation machine learning for biological networks. *Cell.* (2018) 173:1581–92. doi: 10.1016/j.cell.2018.05.015
101. Zou, J, Huss, M, Abid, A, Mohammadi, P, Torkamani, A, and Telenti, A. A primer on deep learning in genomics. *Nat Genet.* (2019) 51:12–8. doi: 10.1038/s41588-018-0295-5
102. Tang, B, Pan, Z, Yin, K, and Khateeb, A. Recent advances of deep learning in bioinformatics and computational biology. *Front Genet.* (2019) 10:214. doi: 10.3389/fgene.2019.00214
103. Jones, DT. Setting the standards for machine learning in biology. *Nat Rev Mol Cell Biol.* (2019) 20:659–60. doi: 10.1038/s41580-019-0176-5
104. Sidak, D, Schwarzerová, J, Weckwerth, W, and Waldherr, S. Interpretable machine learning methods for predictions in systems biology from omics data. *Front Mol Biosci.* (2022) 9:926623. doi: 10.3389/fmolb.2022.926623
105. Kobak, D, and Berens, P. The art of using t-SNE for single-cell transcriptomics. *Nat Commun.* (2019) 10:5416. doi: 10.1038/s41467-019-13056-x
106. Becht, E, McInnes, L, Healy, J, Dutertre, CA, Kwok, IWH, Ng, LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol.* (2018) 37:38–44. doi: 10.1038/nbt.4314
107. Minati, MA, Fages, A, Dauguet, N, Zhu, J, and Jacquemin, P. Optimized nucleus isolation protocol from frozen mouse tissues for single nucleus RNA sequencing application. *Front Cell Dev Biol.* (2023) 11:1243863. doi: 10.3389/fcell.2023.1243863
108. Taylor, AS, Tinning, H, Ovchinnikov, V, Edge, J, Smith, W, Pullinger, AL, et al. A burst of genomic innovation at the origin of placental mammals mediated embryo implantation. *Commun Biol.* (2023) 6:459. doi: 10.1038/s42003-023-04809-y
109. Peers, JA, Nash, WJ, and Haerty, W. Gene pseudogenization in fertility-associated genes in cheetah (*Acinonyx jubatus*), a species with long-term low effective population size. *Evol Int J Org Evol.* (2025) 79:574–85. doi: 10.1093/evolut/qpaf005
110. Yuan, J, Wang, G, Zhao, L, Kitchener, AC, Sun, T, Chen, W, et al. How genomic insights into the evolutionary history of clouded leopards inform their conservation. *Sci Adv.* (2023) 9:eadh9143. doi: 10.1126/sciadv.adh9143
111. Paton, V, Ramirez Flores, RO, Gabor, A, Badia-i-Mompel, P, Tanevski, J, Garrido-Rodriguez, M, et al. Assessing the impact of transcriptomics data analysis pipelines on downstream functional enrichment results. *Nucleic Acids Res.* (2024) 52:8100–11. doi: 10.1093/nar/gkac552

112. Bommert, A, Welchowski, T, Schmid, M, and Rahnenführer, J. Benchmark of filter methods for feature selection in high-dimensional gene expression survival data. *Brief Bioinform.* (2022) 23:bbab354. doi: 10.1093/bib/bbab354
113. Murdoch, WJ, Singh, C, Kumbier, K, Abbasi-Asl, R, and Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci USA.* (2019) 116:22071–80. doi: 10.1073/pnas.1900654116
114. Nguyen, ND, Jin, T, and Wang, D. Varmole: a biologically drop-connect deep neural network model for prioritizing disease risk variants and genes. *Bioinforma Oxf Engl.* (2021) 37:1772–5. doi: 10.1093/bioinformatics/btaa866
115. Speak, SA, Birley, T, Bortoluzzi, C, Clark, MD, Percival-Alwyn, L, Morales, HE, et al. Genomics-informed captive breeding can reduce inbreeding depression and the genetic load in zoo populations. *Mol Ecol Resour.* (2024) 24:e13967. doi: 10.1111/1755-0998.13967