Briefings in Bioinformatics, 2026, 27, bbag152
https://doi.org/10.1093/bib/bbag152
Published: 14 April 2026
Review

OXFORD

Multimodal bioinformatic analyses of genome-scale expression beyond gene-centric differential expression

Jiratchaya Nuanpirom [1,2] and Varodom Charoensawan [2,3,4,5,6,7,*]

[1]Doctor of Philosophy Program in Biochemistry (International Program), Faculty of Science, Mahidol University, 272 Rama VI Rd, Ratchathewi, Bangkok 10400, Thailand
[2]Department of Biochemistry, Faculty of Science, Mahidol University, 272 Rama VI Rd, Ratchathewi, Bangkok 10400, Thailand
[3]Integrative Computational BioScience (ICBS) Center, Mahidol University, 999 Phutthamonthon Sai 4 Rd, Salaya, Nakhon Pathom 73170, Thailand
[4]Division of Medical Bioinformatics, Research Department, Faculty of Medicine Siriraj Hospital, 2 Wang Lang Rd, Bangkok Noi, Mahidol University, Bangkok 10700, Thailand
[5]Department of Biochemistry, Faculty of Medicine Siriraj Hospital, 2 Wang Lang Rd, Bangkok Noi, Mahidol University, Bangkok 10700, Thailand
[6]Siriraj Genomics, Faculty of Medicine Siriraj Hospital, Mahidol University, 2 Wang Lang Rd, Bangkok Noi, Bangkok 10700, Thailand
[7]School of Chemistry, Institute of Science, Suranaree University of Technology, 111, Maha Witthayalai Rd, Mueang Nakhon Ratchasima, Nakhon Ratchasima 30000, Thailand
*Corresponding author. Department of Biochemistry, Faculty of Science, Mahidol University, 272 Rama VI Rd, Ratchathewi, Bangkok 10400, Thailand. E-mail: varodom.cha@mahidol.ac.th

Abstract

Genome-scale gene expression analysis has become a standard approach for discovering biomarkers and understanding molecular mechanisms. Recent advances in omics technologies now enable investigations beyond conventional case–control comparison and standard gene-centric differential expression (DE) analyses. In this review, we highlight conceptual and methodological advances in using transcriptomic and multimodal omic data to elucidate diverse mechanisms of gene expression. We first provide a comprehensive overview of different types of gene expression study designs, along with suitable statistical testing, as well as key considerations. We then describe strategies for inferring gene co-expression and regulatory networks, with particular emphasis on context-specific network models and machine learning methods that capture the multifactorial nature of gene expression regulation. Finally, we present perspectives on emerging modalities such as single-cell and spatial transcriptomics, which enable unprecedented resolution in mapping regulatory complexity. We envisage that the concepts and examples described here will raise awareness and encourage the application of advanced network-based analyses.

Keywords **transcriptomics, differential gene expression analysis, gene regulatory network, machine learning, multi-omics data, multi-modal data**

Introduction

Gene expression is shaped by a complex interplay between internal and environmental factors [1]. Together, these multiple determinants shape gene regulatory responses and phenotypic outcomes, creating substantial challenges for both experimental design and data analysis and interpretation.

Omics-based approaches enable comprehensive investigation of the multimodal nature of gene regulatory mechanisms. In terms of study designs, the conventional case–control regime is the most popular one, yet frequently insufficient to comprehensively capture biological complexity [2]. Furthermore, advanced study designs such as time-series, multifactorial perturbation, and integration of multi-omics approaches are increasingly employed to uncover dynamic regulatory mechanisms that cannot be resolved by pairwise comparison [3–5]. Focusing on transcriptomics, one of the most common modalities of gene expression regulation profiling due to the extensive adoption of high-throughput sequencing technologies [6], differential expression (DE) analysis has long been and continues to be the primary bioinformatics strategy for identifying genes that exhibit significant changes in expression patterns. Despite well-documented strategies for statistical testing and multiple testing correction in DE analysis [7], suboptimal selection of analytical methods may result in spurious detection of differentially expressed genes (DEGs) [8, 9], resulting in misinterpretation and increased validation costs.

Gene-by-gene analyses of gene expression as DE frequently fail to capture the cooperative and coordinated behavior of genes within biological networks and pathways [10]. To overcome this limitation, systems-level approaches, including gene co-expression network

Received: November 28, 2025. **Revised:** February 11, 2026. **Accepted:** March 7, 2026

(GCN) and gene regulatory network (GRN) inference have been introduced [11]. GCNs are an effective framework for investigating coordinated gene expression patterns across conditions; however, co-expression networks alone cannot capture interaction and directions of TFs and their target genes. Conversely, static GRNs frequently fail to represent context-specific gene expression regulation. More recently, new machine learning (ML) approaches have been introduced for inference as well as in-depth analysis of GCNs and GRNs, including identification of nonlinear relationships [12]. Other studies also described integration of multimodal omics data, such as chromatin accessibility data and proteomics, to perform context-specific prediction, including transcription factor activity (TFA) [13, 14], regulatory relationships between genes and their regulatory elements [15], GRN inference from multimodal data integration [16], and higher-order network analyses [17]. The scale, dimensionality, and variability of biological data continue to pose challenges in selecting appropriate data integration and inference methods, highlighting the need for re-survey of emerging tools suitable for specific research questions.

In this review, we first revisited fundamental concepts of study designs of gene expression analyses, which can go beyond conventional case–control design and gene-centric studies (see overview in Fig. 1). We then outlined the strengths and limitations of existing approaches and assessed whether conventional gene-centric DE analysis is applicable for different types of experimental designs. We then described the concepts and update of systems-level strategies, including inferences of GCNs and GRNs. Finally, we described emerging multimodality analyses, including single-cell and spatial omics, and discussed how integration of multi-omics and AI-driven methods can offer unprecedented opportunities to dissect gene expression regulation in context-specific and dynamic manners.

Study designs and data modalities for gene expression analyses

Pairwise comparison study design

Pairwise DE analysis is already widely used to identify significant gene expression changes between two conditions (Fig. 1A and 2A), with technical concepts and foundations reviewed elsewhere, e.g. [7]. However, this approach is less suitable for more complex research questions, such as multifactorial and time-series designs.

Beyond standard gene-level DE analyses; however, pairwise comparison designs can be adapted to investigate gene expression mechanisms across several other aspects. For instance, transcript-level quantification enables DE analysis at the isoform level, e.g. [18]. Alignment-free algorithms can be used to bypass intensive mapping processes, while accounting for the read-to-transcript ambiguity (RTA) problem, arising from shared exons among transcript isoforms [19, 20]. Moreover, the variability from the RTA problem can be estimated and corrected [21], thereby enabling a wide range of transcript-resolution analyses, including differential transcript expression (DTE) and differential transcript/exon usage (DTU/DEU) [18, 22–24]. As an alternative to aggregation of transcript counts which may yield false positives due to transcript length effects [18], transcript-level DE followed by *P*-value aggregation has been proposed [23], and recent developments in edgeR further address RTA-induced variability through quasi-negative binomial modeling [20].

Other important aspects of pairwise comparison worth considering include a limited number of replicates and zero-count inflation in DE analysis [25, 26]. Under these conditions, robust parameter estimation in parametric DE methods becomes challenging, as gene or transcript counts often deviate from assumed probability distribution [25]. In such cases, non-parametric DE approaches can serve as alternatives or complementary methods to reduce false DEG detection [27]. Representative examples include signal-to-noise ratio in NOISeq [28], and rank-based tests implemented in SAMseq [29] and Swish [24]. While rank-based methods accommodate diverse experimental designs and transcript-level analyses, benchmarking studies have reported higher false positive rates for NOISeq and SAMseq relative to parametric methods in small-scale RNA-seq datasets [30]. Conversely, in population-scale studies, rank-based tests such as the Wilcoxon rank-sum have demonstrated improved false positive control, indicating greater robustness in cohort studies [8, 31, 32]. Although well suited for cross-platform data integration without uniform normalization, rank transformation discards expression magnitude and may reduce sensitivity to subtle but biologically meaningful changes [33].

Another important aspect of DE analyses but frequently overlooked is selection thresholds, which typically rely on an arbitrary user-defined false discovery rate (FDR) (e.g. 0.05 or 0.01). However, FDR alone may yield unequal power across comparisons due to differences in fold changes, variance structures, and sample sizes [34]. To address this, more recent ML-based meta-analyses have been applied to help prioritize DEGs across multiple comparison sets, e.g. using support vector machine (SVM) clustering to identify core stress-responsive genes across abiotic conditions [35].

Multifactorial study design

In contrast to pairwise designs, more recent genome-wide gene expression studies consider multiple factors simultaneously, such as genetic and environmental perturbations, e.g. [4, 36]. These factors may function independently or interact to influence gene expression outcomes (Fig. 2B).

A multifactorial study design has been introduced to investigate perturbation of several factors in multiple manners, including additive, antagonistic (epistatic), or synergistic effects (Fig. 1B) [37, 38]. Earlier studies often focused only on a primary/main factor, while treating other sources of variation, such as batch effects [39], sequencing platforms [40], or timepoints [41], as blocking or random effects. Consequently, multifactorial gene expression analysis critically depends on accurately specifying experimental factors and their interactions within the design matrix. However, in multifactorial settings, applying standard pairwise contrasts, such as testing one factor at a time and intersecting DEGs across the contrasts, can yield unreliable results, and still be confounded by the blocking effects (e.g. sequencing type) [42].

DE analysis in multifactorial design has several fundamental aspects in common with pairwise comparisons; however, successful analysis hinges upon the correct specification of factor structures, including primary, nested, and random factors. See earlier method articles for matrix design for multifactorial and compatible analytic pipeline [43]. Beyond conventional pipeline, Schrode and colleagues developed a useful framework to dissect multifactorial gene expression data by explicitly modeling synergistic, additive, and antagonistic interactions among genetic variants, environmental perturbations, and drug responses [4, 38].

Time-series study design

In time-series studies, experimental subjects are typically organized into treatment groups and sampled at predetermined time points (Fig. 2C). Time-series gene expression data can be analyzed by treating

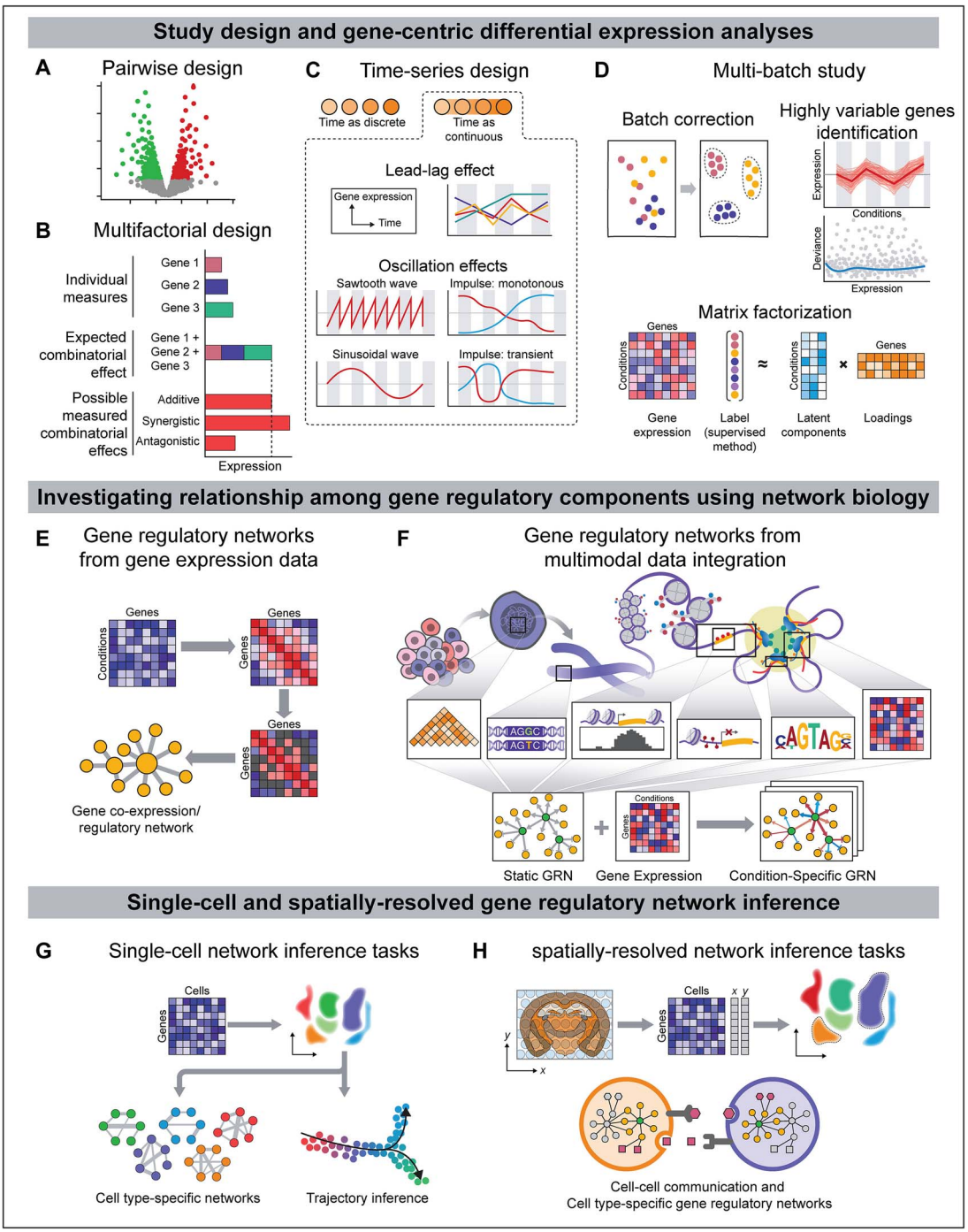

Figure 1 Overview of genome-wide gene expression analyses described in this review, covering analysis strategies from (A–D) gene-centric differential expression to (E–H) systems-level analyses, covering (E) gene regulatory network inference from gene expression data and (F) from integrating multimodal omics data, as well as more recent perspective of genome-scale gene expression analyses at (G) a single-cell and (H) spatially-resolved resolutions. Figure was created using Adobe Illustrator 30.2, with additional illustrations from the NIAID NIH BioArt Source (bioart.niaid.nih.gov/bioart/637; bioart.niaid.nih.gov/bioart/172).

time as discrete or continuous variables, thereby enable uncovering waves, periodic, or cyclic gene expression patterns (Fig. 1C).

Time as a discrete variable

Conventional DE methods that treat time as a discrete variable, including the Quasi-likelihood F-test in edgeR [44], likelihood ratio test (LRT) in DESeq2 [45], and linear models in limma [46], are often sufficient for time-series datasets with short timespans or a number of timepoints. Under this approach, DEGs are identified simply based on significant expression changes at one or more time points [47]. Existing tools such as DESeq2 and edgeR perform comparably for short time series ($t < 8$), but diverge in performance for longer series ($t \geq 8$) [48]. However, treating time as a discrete variable overlooks continuous transcriptional dynamics, limiting the detection of temporal dependencies,

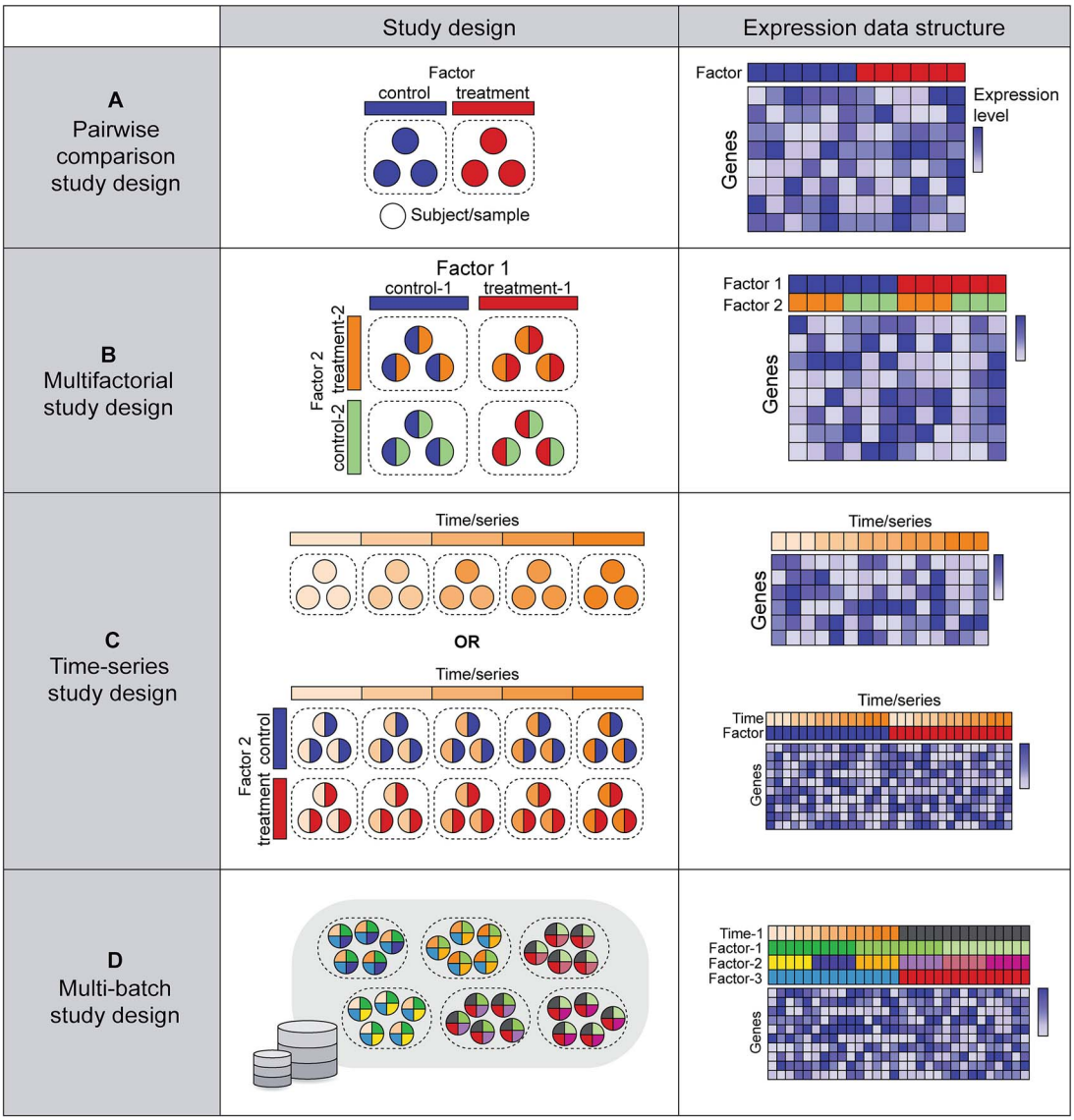

Figure 2 Study designs and examples of data structures of differential gene expression studies, including commonly known (A) pairwise comparison study design, as well as less widely employed (B) multifactorial, (C) time-series, and (D) multi-batch study designs.

delays, and non-linear trajectories. Such models are also sensitive to sampling frequency and may miss transient or phase-shifted expression patterns, thereby reducing power to resolve complex temporal regulation [49].

Time as a continuous variable

Expression analysis that treats time as a continuous variable typically requires three or a larger number of timepoints, which to investigate more complex expression dynamic patterns, such as lead–lag [47, 50], oscillation and cyclical patterns [51, 52].

Lead–lag effects describe how changes at earlier time points influence later expression after a delay, and can be inferred using regression-based models, such as EBSeq-HMM (hidden Markov model) [47] and Trendy [50]. While EBSeq-HMM considers temporal expression as monotonic or sporadic paths, it may scale poorly for longer time series due to an exponential increase in possible patterns. Trendy addresses this limitation by using segmented regression with model selection to identify breakpoints in gene expression trajectories [50].

Longer time-course gene expression designs enable the detection of oscillatory gene-expression patterns, which potentially underlie dynamic processes such as circadian regulation and metabolism. Existing approaches perform efficiently under the parametric assumption, including natural cubic spline regression [3], which captures non-linear expression patterns over time, and can be described by a polynomial function; and the quasi-likelihood F-test in edgeR [53], which models smoothed time point and estimates expression dispersion. However, polynomial-based methods often fail to capture asymmetric or complex oscillatory waveforms. To address this caveat, non-parametric methods such as RAIN [51] was introduced to detect symmetric and asymmetric umbrella-shaped oscillations, and is widely used for circadian analyses [54, 55]. However, a limitation of RAIN lies in its constrained assumptions, as it is applicable only to umbrella-shaped expression patterns and

is sensitive to overdispersion, which can lead to false detection of rhythmic genes [56, 57].

Other approaches widely used to investigate time-course gene expression patterns include MetaCycle [52], which employs three wave detection strategies, namely parametric autoregressive spectral estimation (ARSER), non-parametric Jonckheere-Terpstra-Kendall test (JTK_CYCLE), and Fast Fourier transform analysis (Lomb-Scargle), where optimal model is determined from the consensus across these three strategies. In this way, MetaCycle achieves robustness through integration of multiple rhythm-detection algorithms, but they were first developed for microarray data, and hence may not fully account for variability in RNA-seq data. In contrast, ImpulseDE2 [49], a more recent approach developed based on sequencing-based data, models time-dependent expression changes using a negative binomial-based impulse function, enabling the identification of transient versus permanent transcriptional responses through likelihood ratio testing, based on their relative fit to impulse versus sigmoid models.

Multi-batch study design

In multi-batch study design, primary 'experimental' factors normally include 'manipulated' variables, such as genetic, chemical, and environmental perturbations (Fig. 2D), and 'non-experimental' observational factors such as gender, age, developmental stage, batches of experiments, also known as confounding factors. Multi-batch design is also common in comparative expression studies (e.g. [35, 58]), and gene expression compendia (e.g. [27, 59, 60]. Here, we describe key aspects that should be considered when designing a multi-batch experiment.

Batch effect correction

Batch effects arise from technical variations introduced by various aspects of experimental settings, laboratory conditions, reagent lots, analytical pipelines, or even wet-lab proficiency and procedure of different individuals, which collectively can confound gene expression heterogeneity [61, 62]. When possible, batch effects can and should be controlled at the experimental design step, through randomization across batches and blocking strategy [63]. But if not possible and if the sources of batch effects are known, one can include specific factors or covariates into the design formula prior to gene count normalization, such as in edgeR [64], DESeq2 [45], and Combat-seq [65], or during the normalization process using the *removeBatchEffect* function of limma [46]. The Removal of Unwanted Variation (RUV) method uses negative controls, which are genes or sequences assumed to be unaffected by experimental factors, such as spike-ins, e.g. [66], or housekeeping genes, to mitigate the batch effects. Correcting batch effects using negative controls can be performed during between-sample normalization [67], or *post hoc* adjustment using factor-augmented regression models [68, 69]. However, specifying a suitable set of negative controls, or selecting suitable housekeeping genes for certain experiments, remains challenging and prone to mis-specification [69].

When known batch labels cannot sufficiently account for the non-experimental factor differences, computational frameworks such as surrogate variable analysis (SVA) in svaseq [70] and probabilistic estimation of expression residuals (PEERs) [71], can be used to estimate latent confounders. SVA infers hidden batch effects and incorporates them into linear models, but its accuracy may drop when batch effects are comparable in magnitude to biological signals and can scale poorly with too many surrogate variables. In contrast, PEER uses empirical Bayes matrix factorization, offering greater sensitivity, scalability, and robustness, and is widely applied in large-scale multi-center (e.g. [59, 60]) and expression quantitative trait loci (eQTL) studies (e.g. [72, 73]). When the contribution of each factor is unknown, estimating the magnitude of known factors and latent confounders can facilitate effective batch effect correction. Principal variance component analysis (PVCA) [74] and VariancePartition [75] quantify factors contributing to gene expression heterogeneity [76]. The resulting covariates can subsequently be applied to batch corrector such as limma's removeBatchEffect [46] or lme4 [77].

Over-correction of batch effects can result from inappropriate method of choice, group imbalance, or the removal of confounding factors that genuinely reflect biological variability [78]. Mitigating over-correction remains challenging due to the unsupervised nature of most batch correction methods. One practical diagnostic is to monitor the expression of 'reference genes,' presumed to be either constitutively expressed, as deviations in their stability or loss of DE, respectively, may indicate insufficient correction or over-correction [79]. For a more comprehensive review and practical guideline of transcriptomic data quality control and preprocessing, please refer to recent articles [80, 81].

Identification of highly variable genes

Highly variable genes (HVGs) are identified by their elevated expression variability across samples relative to mean expression levels. HVGs can be characterized using standard measurements of variations, including standard deviation (SD) and/or *P*-values from analysis of variance (ANOVA), as demonstrated in a comparative transcriptomic study of temperature-responsive arabidopsis HVGs [82], or alternative quantification strategies described below.

Coefficient of variation (CV), a ratio of the SD to the mean, provides a more flexible approach when comparing the measures across different conditions. However, a simple CV exhibits a strong negative heteroskedastic relationship when applied to zero-inflated gene expression data, and biases for genes with low mean expression [83, 84]. Residual CV serves as an alternative measure that applies local polynomial regression (LOESS) to the CV relative to the median expression, as seen in [84]. Another alternative is mean/median absolute deviation (MAD), or the average of absolute deviations, which is more robust to outliers and skewed gene expression data, as seen in a study of human gene expression variation related to methylation and aging, which utilized LOESS regression to estimate the expected MAD and adjust for it as a measure of expression variation [83].

Matrix factorization for feature/variable classification and dimensionality reduction

Multi-batch studies often involve a large number of variables that jointly drive gene expression heterogeneity, resulting in high-dimensional biologically relevant data, as well as unwanted technical variability. In such settings, matrix factorization (MF) can provide a flexible framework for integrative analyses.

For feature subtyping and biomarker discovery, unsupervised MF-based methods such as canonical correlation analysis (CCA) [85] and factor analysis (FA) [86], as well as supervised approaches such as sparse partial least squares discriminant analysis (sPLS-DA) [87], project data into low-dimensional spaces to extract biologically meaningful patterns. CCA maximizes the correlations between paired data, while FA captures shared variance across multiple modalities, and sPLS-DA selects features that best discriminate predefined

groups, such as in biomarker discovery between patient and healthy populations [88].

MF is also applicable for estimating TFA from the transcription factor (TF) repertoire where regulatory signals are mixed up, such as independent component analysis (ICA) [89], and network component analysis (NCA) [14]. Unlike other methods that maximize the common latent space, ICA aims at maximizing the statistical independence of each feature without requiring training datasets, whereas NCA requires topological information from prior knowledge to constrain the regulatory strength of each TF-target connectivity in TFA estimation.

Investigating relationships among gene regulatory components using network biology

While the experimental designs discussed above primarily focus on identifying changes at the level of individual genes or transcripts, biological regulation is inherently coordinated and operates across interacting molecular components. In addition, conventional DE analysis of individual genes often fails to capture relationships among gene regulatory components. For example, TFs often exhibit modest expression changes between conditions yet exert disproportionate effects on the expression of their target genes [7]. Furthermore, specific TF-target regulatory patterns can buffer or amplify regulatory signals [90, 91], which are often missed by conventional DE analysis, but can be detectable through systems-level and/or network-based analyses (Fig. 1E). Note that fundamental concepts of network biology, including what the nodes (vertices) connected by edges (links) represent, have been extensively reviewed elsewhere, e.g. [92–94]. In this review, we focus on practical concepts and tools for assessing expression correlation and regulatory relationship that provide insights into the modularity and systems-level dynamics of gene expression regulation.

Inferring gene co-expression network from gene expression data

Gene co-expression networks (GCNs) capture the relationships among genes with similar expression patterns through context-specific relationships (edges) between genes (nodes) and their correlation coefficients (edge weights), and the edges are normally undirected. Unlike investigation of individual DEGs, gene–gene relationships (adjacency) can be quantified using multiple types of matrices (Fig. 3A and Table 1), and correlation coefficient is one of the most efficient and widely-adopted approaches [11, 95]. There is; however, no consensus on optimal sample sizes in GCN inference due to differences in benchmarking strategies, data types (e.g. microarray or RNA-seq, experimental or synthetic), organisms of choice, and normalization methods to name a few. Saying that, it is thought that for bulk RNA-seq analyses, ≥20 samples with >10 million reads per sample can already achieve performance comparable to the best microarray-based GCNs. [96]. However, this benchmark relied on simple rather than normalized counts, and accounted for a small number of independent datasets. A comprehensive evaluation by [95] demonstrated that sample sizes ranging from 5 to 40 using counts adjusted with TMM-normalization factor (CTF) or upper-quartile-normalization factor (CUF) can achieve highest accuracy on large homogeneous and small-multiple heterogeneous datasets.

For noisy correlation matrices, variances can be stabilized using context-likelihood of relatedness (CLR) [97], by upweighting regulator-target gene pairs, and down-weighting spurious correlations [95]. However, such network transformation is only efficient when the dataset is >40 samples [95]. In addition, network thresholding can be applied to filter only top-ranked context-specific gene modules, based on absolute correlations [98], or significant *P*-values [99]. Additionally, one can examine the soft-thresholding power from the goodness of fit of the network degree distribution to the power law distribution [11]. Community detection methods can also be applied to remove sporadic connectivities in the graph [100, 101].

Inferring sample-specific co-expression networks from gene expression data

Most GCN inferences aggregate gene expression data across multiple conditions or samples, and this may not be appropriate when the differences between samples or conditions are of interest, such as in patient-specific drug response networks [102, 103]. Methods for inferring sample-specific GCNs, primarily aim to compare between aggregated and perturbed networks (Fig. 3B and Table 1) [104–106]. The key differences among these tools are how the perturbed networks are identified and the normalization parameters are utilized. For example, LIONESS constructs a perturbed network using an adjacency matrix of all samples except the one of interest, and subsequently employs linear interpolation to estimate the corresponding sample-specific network [105]. On the other hand, SWEET modifies the method implemented in LIONESS and also adjusts for the network size imbalance [104].

Inferring gene regulatory networks from gene expression data

GRN represent causal relationships (directed edges) between TFs (source nodes) and target genes (target nodes). Data from GCN can serve as a proxy for regulatory relationships inference; however, high correlation in expression of two genes may not reflect their biological relationship nor causality [107]. GRN inference from gene expression data shares fundamental concepts with GCN inference, including adjacency matrix generation (Fig. 3A). However, beyond correlation coefficients (Fig. 3C), alternative measures can represent gene–gene adjacency based on different assumptions and concepts as described below (see also Table 1).

In brief, mutual information (MI) (Fig. 3D) quantifies shared information between two genes, and can effectively capture non-linear aspects of relationships [108]. However, a study using a large number of samples (n > 200) showed inferior robustness compared with simple correlations [96]. Ensemble learning (Fig. 3E), such as random forests, is another alternative that captures non-linear and multifactorial relationships by regressing each target gene against all predictors [109]. These methods can be extended with ordinary differential equations (ODEs) to model steady-state and time-series data, though performance remains dataset- and parameter-dependent [110]. GENIE3, one of better-known GCN inference tools, aggregates predictions from all subtrees including those with weak predictive power, whereas another popular tool GRNBoost2 redesigns the inference pipeline by adopting stochastic gradient boosting machine (GBM) with early-stopping regularization, which improves its speed, memory efficiency, and scalability to large datasets [111, 112]. Conditional Gaussian, in turn, (Fig. 3F) addresses higher-order regulation by clustering genes

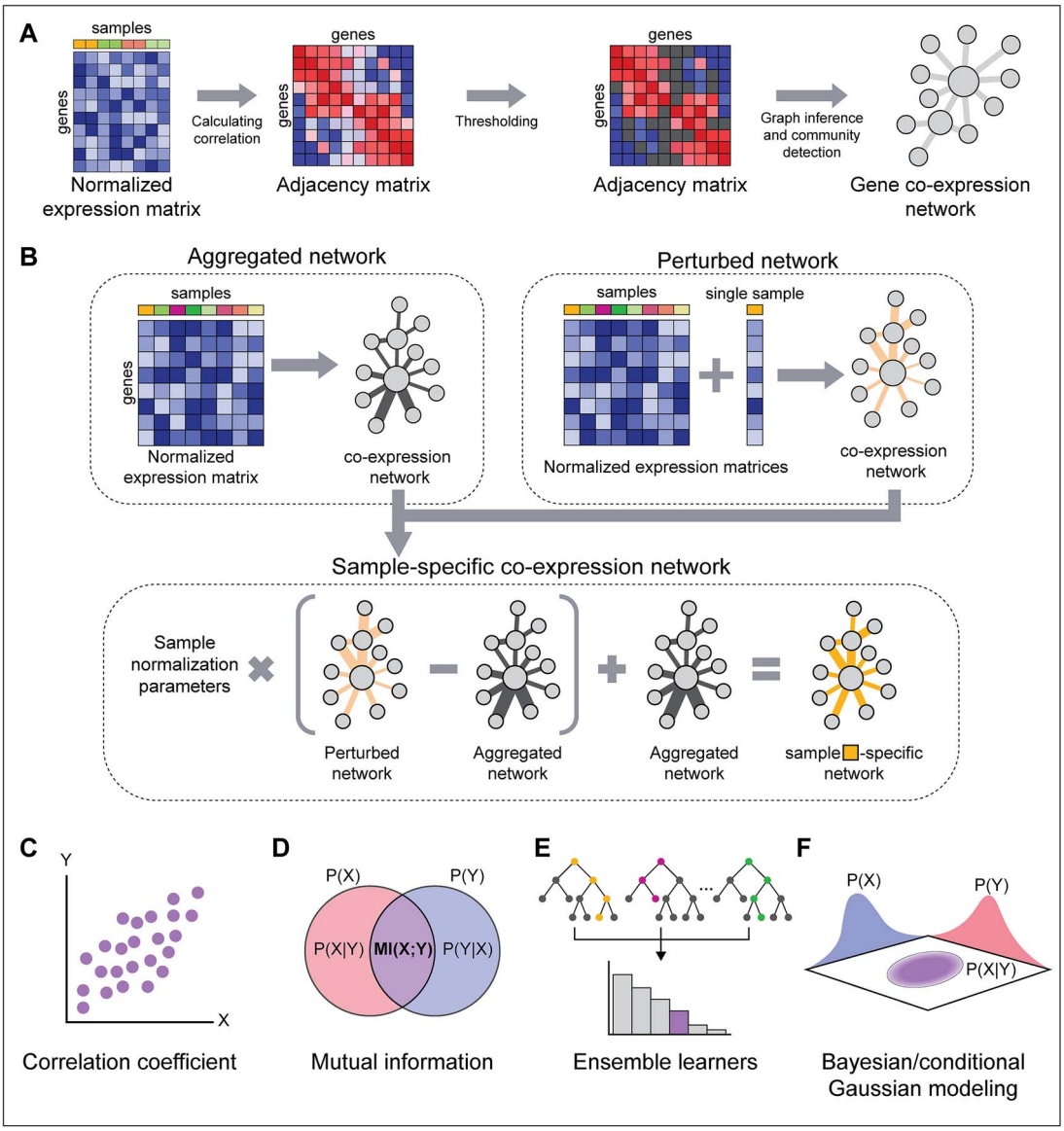

Figure 3 Network inference approaches based on genome-scale gene expression data. (A) Gene co-expression network inference starts with computing gene–gene expression adjacency matrix from normalized gene counts. Thresholding may be applied to the co-expression matrix. Clustering or community detection methods can be applied to this adjacency matrix to group co-expressed genes and simplify the network graph by retaining only meaningful connections. (B) Overview of sample-specific co-expression network analysis adapted from SWEET [104]. Aggregated and perturbed co-expression networks are constructed to interpolate sample-specific networks. The aggregated network uses all samples, while the perturbed network uses all samples and another sample of interest. Both networks are then interpolated, scaled, and normalized to infer the sample-specific co-expression networks. To infer gene co-expression and gene regulatory networks from gene expression data, commonly used metrics for quantifying gene–gene relationships including: (C) correlation coefficients, (D) mutual information, (E) ensemble learners, and (F) conditional probability from Bayesian/conditional Gaussian modeling.

with shared regulatory logic into modules and using them as probabilistic priors to infer putative regulators and co-regulators [113].

Expression-based GRN inference has been utilized broadly among gene expression data analysis to investigate context-specific gene communities. However, it less overlaps with GRNs derived from genomic-context information such as TF-binding assays and motif information especially when applied to infer eukaryotic GRNs [114]. Therefore, using genomic-context information as prior knowledge in combination with genome-wide gene expression data can improve the reliability of the GRN inference.

Inferring gene regulatory network from multimodal data integration

As thoroughly discussed in earlier sections, gene expression profiling provides a powerful and accessible readout of cellular states and responses; however, it cannot capture a complete view of regulatory processes and mechanisms. Combining gene expression with other omic layers, including genomics and epigenomics, through multimodal data integration, can uncover complex, non-linear relationships across molecular layers, although such data are often

Table 1 Co-expression and regulatory network inference tools based on gene expression data.

Tool name	Method name	Gene–gene adjacency measure	Application	Implementation/Platform	Software status (as of January 2026)	References
ARACNE	Algorithm for the Reconstruction of Accurate Cellular Networks	MI	GCN	C++ library 'ARACNe' https://califano.c2b2.columbia.edu/aracne	Unknown	[108]
CLR	Context likelihood of relatedness	MI	GCN	R package 'minet' (third-party) https://www.bioconductor.org/packages/release/bioc/html/minet.html; R package 'parmigene' (third-party) https://cran.r-project.org/web/packages/parmigene/index.html	minet: Maintained, parmigene: Maintained	[97, 180, 181]
WGCNA	Weighted gene co-expression network analysis	COR	GCN	R package 'WGCNA' https://cran.r-project.org/web/packages/WGCNA/index.html	Active	[11]
WTO	Weighted topological overlap	COR	GCN	R package 'wTO' https://cran.r-project.org/web/packages/wTO/index.html	Maintained	[182]
CEMiTool	Co-Expression Modules identification Tool	COR	GCN	R package 'CEMiTool' https://www.bioconductor.org/packages/release/bioc/html/CEMiTool.html and web application https://cemitool.sysbio.tools	Maintained	[183]
GWENA	Gene Whole co-Expression Network Analysis	COR	GCN	R package 'GWENA' https://bioconductor.org/packages/release/bioc/html/GWENA.html	Maintained	[184]
BONOBO	Bayesian optimized networks obtained by assimilating omic data	PGM	Sample-specific GCN	Python package 'netZooPy'https://netzoo.github.io/zooanimals/ss/bonobo	Active	[106]
LIONESS	Linear Interpolation to Obtain Network Estimates for Single Samples	COR	Sample-specific GCN	R package 'netZooR'; python package 'netZooPy'; MATLAB package 'netZooM' https://netzoo.github.io/zooanimals/ss/lioness	R: Active, Python: Active; MATLAB: Maintained	[105]
SWEET	Sample-specific-weighted correlation network	COR	Sample-specific GCN	Python source code 'SWEET' https://github.com/SysMednet/SWEET	Maintained	[104]
GENIE3	GEne Network Inference with Ensemble of trees	REG	GRN	R package 'GENIE3' https://bioconductor.org/packages/release/bioc/html/GENIE3.html	Maintained	[109]
MERLIN	Modular regulatory network learning with per gene information	PGM	GRN	C++ library as a module in GP-DREAM collection https://pages.discovery.wisc.edu/~sroy/merlin; https://github.com/liu/gpdream/wiki/Merlin	Deprecated, replaced by MERLIN-P-TFA	[113]
dynGENIE3	dynamical GENIE3	REG	Time-dependent GRN	Python, MATLAB and R packages 'dynGENIE3' https://github.com/vahuynh/dynGENIE3	Active	[110]
Inferelator	Inferelator	REG	Time-dependent GRN	Python library 'Inferlator' https://github.com/flatironinstitute/inferelator	Active	[145]
GRNBoost2	GRNBoost2	REG	GRN	Python library 'arboreto' https://arboreto.readthedocs.io/en/latest/	Maintained	[111]

Software status is classified into: **Active**: software regularly updated with expansion of new feature/architecture improvements, software capabilities, bug-fixing; **Maintained**: software updated for minor changes to preserve functionality, security, and compatibility, or updated documents only; **Deprecated**: software archived, discontinued, or superseded; and **Unknown**: unable to find related deployment date. Abbreviations: MI: mutual information, COR: correlation, PGM: probabilistic graphical model, REG: regression, GCN: gene co-expression network, GRN: gene regulatory network.

heterogeneous, high-dimensional, and sparse, introducing challenges related to the 'curse of dimensionality' [115].

Multimodal omic integration starts with converting different data forms into integratable features, such as assigning *cis*-regulatory elements (CREs) to corresponding genes, annotating TF-binding motifs from accessible chromatin regions, or jointly integrating multimodal data into a lower-dimensional space [116]. Multi-modal data integration has been applied to diverse objectives, yielding different forms of integration outputs, including classification labels (subtypes), predictor-outcome relationships, and heterogeneous networks [117]. Heterogeneous GRNs from multimodal omics data integration can have a single node type with multiple edge types, or multiple nodes with multiple edge types [94]. See also currently available computational methods for integration of multi-modal omic data [115, 117–119].

In this review, we focus on GRN inference from gene expression data, in combination with a prior knowledge to constrain the GRN, by removing connections that are not supported by evidence from another omic layer, such as conserved regulatory sequences in the genome. Such biological evidences include catalogues of characterized sequence-specific regulatory motifs in relevant databases (e.g. JASPAR [120] and AnimalTFDB [121]), high-throughput chromatin accessibility assay (e.g. Assay for Transposase-Accessible Chromatin with High-Throughput Sequencing (ATAC-seq) and Micrococcal Nuclease sequencing (MNase-seq)), TF and nucleosome binding (e.g. Chromatin Immunoprecipitation Sequencing (ChIP-seq)), DNA methylation (e.g. Bisulfite sequencing (BS-Seq)), transcriptomic studies following perturbation of TF genes as well as known biological networks (see also [122, 123]) (Fig. 4A). Incorporating prior knowledge in multimodal GRN inference improves inference accuracy, and enables several prediction tasks including individual-specific GRN [124, 125], subnetwork identification [126], time-dependent GRN [16], and TFA estimation [13, 14, 124, 127–129] (see also Table 2).

TFA quantifies the regulatory influence of a TF on the expression of its target genes, which can be activation or suppression to name a few, frequently in a context-specific manner [130, 131]. TFA can be measured experimentally, or computationally inferred using MF as a non-negative matrix [13, 14, 129] (Fig. 4B), ML [124, 127], or deep learning [128] models.

We further review GRN inference methods that integrate multimodal omics data and categorize them by their prior knowledge requirements, input types, prediction tasks, and their availability (see also Table 2).

Regression and machine learning

Links between TF and their targets derived from prior knowledge (see examples in Fig. 4A), and gene expression, can be formulated as a regression problem. Regression models such as the multi-task learning in AMuSR [127] and PSIONIC [124], and linear and logistic regression in TEPIC2 [132], can infer context-specific GRNs and TFA by modeling the gene expression as an outcome of the prior knowledge and TFA (Fig. 4C). Random forest function in iRafNet [133] can incorporate protein–protein interactions (PPI) to compute TF-specific prior weights during tree construction. Supervised nearest shrunken centroid (NSC) classification in iOmicsPASS [126] identifies predictive context-specific subnetworks, by applying z-transformation for edges in the heterogeneous networks, such as PPI and GRN. Unsupervised classification methods such as NCA in NetREX-CF [129], in combination with collaborative filtering, can reconcile edges to infer context-specific GRNs and predict TFA.

Probabilistic graphical models

Probabilistic graphical models (PGMs) infer context-specific GRNs by integrating the prior knowledge with gene expression data, and representing the TFA as conditional probabilities (Fig. 4D). Dependency network frameworks, such as in MERLIN+P [114], infer the conditional probability of each gene given a set of known regulators from prior knowledge, and incorporating with NCA and sparsity regularization in MERLIN+P + TFA [14] can improve the accuracy of TFA estimation and downstream GRN inference. Alternatively, Gaussian graphical models (GGMs) and covariance shrinkage implemented in DRAGON [134], infer heterogeneous GRNs from paired multimodal omics data, such as gene expression and methylation, and stabilize the noises across modalities. Despite these strengths, Gaussian and linear methods often fail to capture non-linear regulatory relationships, and scale poorly to large and complex networks [135].

Message passing

Message passing (Fig. 4E) is at the core of PANDA [136], PUMA [137], SPIDER [138], and EGRET [125], which were all adapted from an affinity propagation for iterative data clustering [139], and has been successfully utilized in yeast GRNs [140]. These methods make use of prior networks encompassing three edge types: binding or association (from PPI, miRNA binding, chromatin accessibility, SNPs information), regulatory (from TF-motif binding), and co-expression (from gene expression) edges. PANDA [136] iteratively propagates context-specific expression signals (a.k.a., 'message') into the prior network to refine TF-target interactions until reaching maximum convergence, resulting in downstream context-specific heterogeneous GRNs [136].

Deep learning

While message-passing and probabilistic graphical approaches integrate the prior knowledge through iterative refinement or conditional probability modeling, deep learning (DL) methods, particularly graph neural networks (GNNs) (Fig. 4F), offer a flexible alternative, which is capable of capturing nonlinear and higher-order interactions from graph representation [94]. These models typically transform gene expression and prior knowledge information to nonlinear latent embeddings. As examples, DeepFGRN [128] formulates GRN inference as an association-prediction task to infer edge direction and regulatory types. Recent advancements in neural networks extend to solving differential equations, e.g. NeuralODEs, and time-dependent GRN inference, e.g. PHOENIX [16].

Although DL models are suitable for complex and non-linear biological problems, many rely on supervised learning and therefore require large, well-annotated training datasets. Despite the abundance of existing bulk gene expression data, the lack of labels may limit the development of multimodal deep learning-based GRN prediction [128]. Nevertheless, DL remains effective beyond GRN construction, such as predicting regulatory relationships. For example, convolutional neural networks (CNNs) which integrate regulatory sequences, chromatin profiles, and long-range upstream regulatory sequence (e.g. Hi-C), can effectively predict gene expression and facilitate functional enhancer-gene interactions [15]. Conceptual reviews of leveraging other DL methods for predicting gene expression profiles from genomic and epigenetic features can be found in [12, 141].

Table 2 Current tools for gene regulatory network inference from multimodal data.

Tool	Method category	Prediction tasks	Prior knowledge requirement	Source of prior knowledge	Omics/modalities integrated	Implementation/platform	Software status (as of January 2026)	References
MERLIN-P	PGM	context-specific GRN	Yes	ChIP, Gene knockout, DNA binding motif	Transcriptomics, Prior knowledge/GRN	C++ library 'MERLIN+P' https://github.com/Roy-lab/merlin-p	Deprecated, replaced by MERLIN-P-TFA	[114]
DRAGON	PGM, ML	context-specific GRN	No	-	DNA methylation, Transcriptomics	R package 'netZooR' https://github.com/netZoo/netZooR; Python package 'netZooPy' https://github.com/netZoo/netZooPy	Active	[134]
MERLIN+P+TFA	PGM, MF	context-specific GRN; TFA estimation	Yes	DNA binding motif, GRN databases, DNase-seq	Transcriptomics, Prior knowledge/GRN	Analysis pipeline 'MERLIN-P-TFA' https://github.com/Roy-lab/MERLIN-P-TFA, use with MERLIN+P	Active	[14]
AMuSR	REG	context-specific GRN; TFA estimation	Yes	Chromatin accessibility, Gene knockout, Gene overexpression, DNA binding motif	Transcriptomics, Prior knowledge/GRN	Inferelator workflow AMuSR https://github.com/simonsfoundation/multitask_inferelator/tree/AMuS	Active	[127]
PANDA	MP	context-specific GRN	Yes	DNA binding motif	PPI, Transcriptomics, Prior knowledge/GRN	R package 'netZooR' https://github.com/netZoo/netZooR; Python package 'netZooPy' https://github.com/netZoo/netZooPy; MATLAB package 'netZooM' https://github.com/netZoo/netZooM; C++ library 'netZooC' https://github.com/netZoo/netZooC	R package: Active; Python package: Active; MATLAB package: Maintained; C++ library: Maintained	[136]
PUMA	MP	miRNA-mRNA regulatory network	Yes	miRNA-seq	Transcriptomics, miRNA-mRNA interaction prior knowledge	R package 'netZooR' https://github.com/netZoo/netZooR; Python package 'netZooPy' https://github.com/netZoo/netZooPy; MATLAB package 'netZooM' https://github.com/netZoo/netZooM; C++ library 'netZooC' https://github.com/netZoo/netZooC	R package: Active; Python package: Active; MATLAB package: Maintained; C++ library: Maintained	[137]
SPIDER	MP	context-specific GRN	Yes	DNA binding motif, ChIP-seq, DNase-seq	Transcriptomics, Prior knowledge/GRN	R package 'netZooR' https://github.com/netZoo/netZooR; MATLAB package 'netZooM' https://github.com/netZoo/netZooM	R package: Active; MATLAB package: Maintained	[138]
EGRET	MP	genotype/individual-specific GRN	Yes	SNP (genomics), SNP (eQTL), DNA binding motif	PPI, Transcriptomics, Prior knowledge/GRN	R package 'netZooR' https://github.com/netZoo/netZooR	Active	[125]

(continued)

Table 2 Continued.

Tool	Method category	Prediction tasks	Prior knowledge requirement	Source of prior knowledge	Omics/modalities integrated	Implementation/platform	Software status (as of January 2026)	References
TIGER	MF	context-specific GRN; sample-specific TFA estimation	Yes	GRN databases, ChIP-seq, DNA binding motif	Transcriptomics, Prior knowledge/GRN	R package 'netZooR' https://github.com/netZoo/netZooR	Active	[13]
iRafNet	ML	context-specific GRN	Yes	Gene knockout, PPI, Time-series gene expression	Transcriptomics, Prior knowledge/GRN	Discontinued R package 'iRafNet' https://github.com/cran/iRafNet	Deprecated	[133]
iOmicsPASS	ML	network-driven subnetwork identification	Yes	PPI, GRN	Prior knowledge/GRN, Transcriptomics, Proteomics	C++ source code 'iOmicsPASS' https://github.com/cssblab/iOmicsPASS; R source code 'iOmicsPASS+' https://github.com/cssblab/iOmicsPASSplus	Deprecated, replaced by iOmicsPASS+	[126]
PSIONIC	ML	Patient-specific GRN; TFA estimation	Yes	Chromatin accessibility, DNA binding motif	Transcriptomics, Prior knowledge/GRN	Matlab package 'PSIONIC' https://github.com/osmanbeyoglulab/PSIONIC	Maintained	[124]
NetREX-CF	ML	context-specific GRN; TFA estimation	Yes	ChIP, Gene knockout, GRN databases, DNA binding motif	Transcriptomics, Prior knowledge/GRN	Python source code 'NetREX_CF' https://github.com/EJIUB/NetREX_CF	Maintained	[129]
PHOENIX	DL, ODE	Time-dependent GRN; require time-series gene expression data	Yes	GRN databases	Transcriptomics, Prior knowledge/GRN	Python source code 'PHOENIX' https://github.com/QuackenbushLab/phoenix	Maintained	[16]
DeepFGRN	DL	context-specific GRN; TFA estimation	Yes	GRN, Transcriptomics	GRN databases	Python source code 'DeepFGRN' https://github.com/PhoebeGaoZhen/DeepFGRN/tree/master	Maintained	[128]
Inferelator	REG, ODE	context-specific GRN; time-dependent GRN	Yes	Chromatin accessibility	Transcriptomics, Prior knowledge/GRN	Python library 'Inferlator' https://github.com/flatironinstitute/inferelator	Active	[145]

Software status is classified as: **Active**: software regularly updated with addition and expansion of feature/architecture improvements, software capabilities, bug-fixing; **Maintained**: software updated for minor changes to preserve functionality, security, and compatibility, or update of documents only; **Deprecated**: software archived, discontinued, or superseded; and **Unknown**: unable to find related deployment date.
Abbreviations: PGM: Probabilistic graphical model, ML: Machine learning, MF: Matrix factorization, REG: Regression, MP: Message passing, DL: Deep learning, ODE: Ordinary differential equation, GRN: Gene regulatory network, ChIP: Chromatin immunoprecipitation, TFA: Transcription factor activity, PPI: Protein–protein interaction network, SNP: Single nucleotide polymorphisms, eQTL: Expression quantitative trait loci.

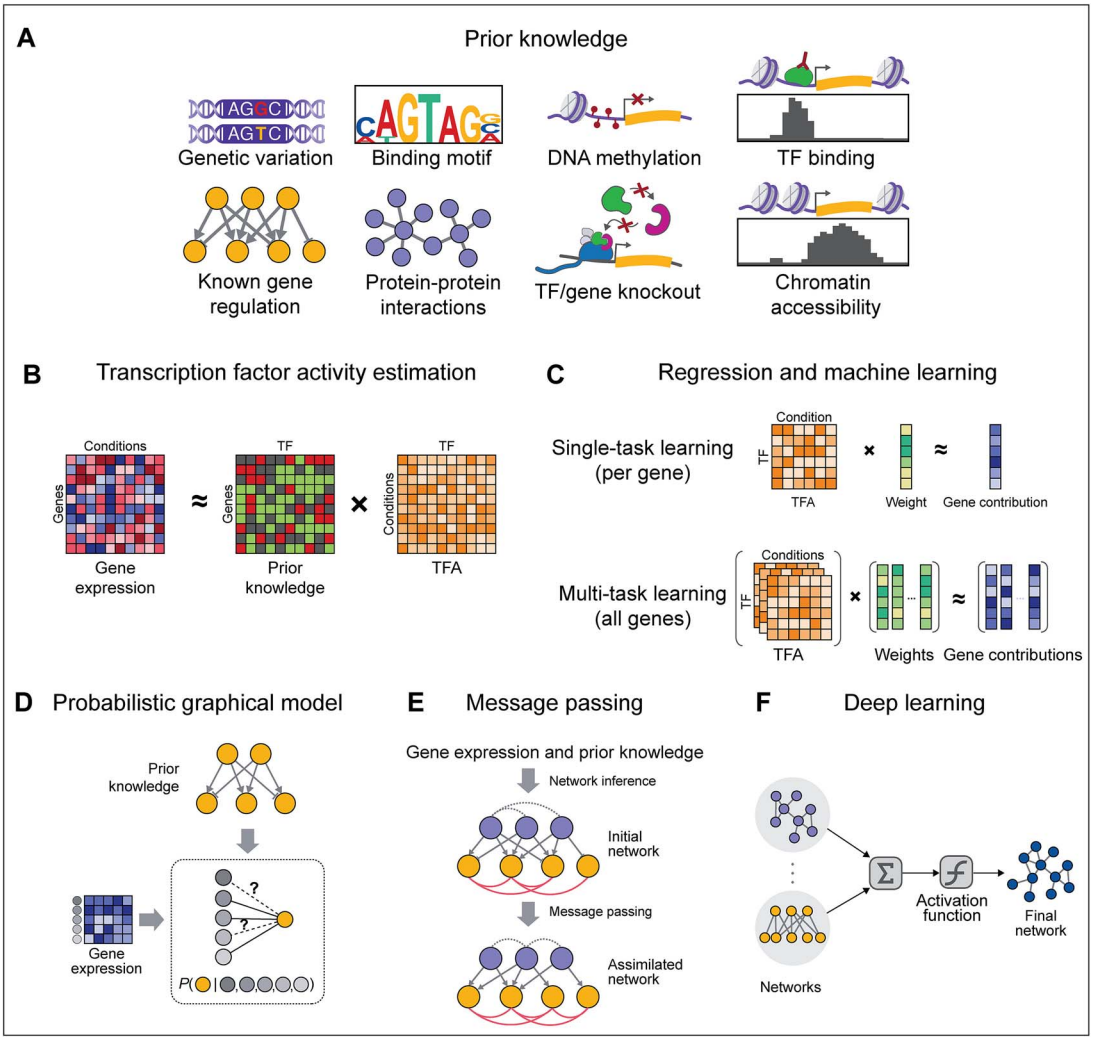

Figure 4 Gene regulatory network inference from multimodal omics data and prior knowledge integration. (A) Prior knowledge can be made up of genomic, epigenomic, and other genome-scale methods, such as genetic variation, DNA binding motif information, DNA methylation, TF binding information, known gene regulatory and protein–protein interaction networks, gene knockout experiments, and chromatin accessibility. Other types of prior knowledge might also be applicable, depending on the requirements and compatibility of the tools. Examples of key inference methods include (B) estimating transcription factor activity from gene expression data and prior knowledge integration using matrix factorization. Multimodal omics data can be integrated using a number of available methods, namely (C) regression and machine learning, (D) probabilistic graphical model, (E) message passing, and (F) deep learning to construct context-specific gene regulatory networks.

Performance evaluation strategies

The rapid expansion of high-throughput omics and continual advances in network inference methods, however, rigorous evaluation of inferred networks remains essential. Most GRN inference approaches are based on assumptions of co-expression or conditional dependent [107], thus requiring validation against ground truth information to ensure biologically meaningful conclusions. Accordingly, causality-informed molecular measurements can serve as both gold standard and as prior knowledge for multimodal GRN inference and TFA estimation (see also [123, 131, 142]).

Common performance measures for GRN inference include overall performance metrics (e.g. area under curves), and biological ground truth (e.g. number of predicted TFs). The area under the receiver operating characteristic (AUROC) curve can be used to evaluate the classification performance by contrasting true positive rate against false positive rate, but it can be misleading in imbalanced settings, where abundant true negatives inflate apparent performance despite poor precision. Alternatively, the area under the precision-recall (AUPRC) curve emphasizes the recovery of true positives while penalizing false positives. However, unlike the AUROC with a fixed base line of 0.5, the AUPRC baseline is a ratio of positive and negative classes, making it suitable for imbalanced data but incomparable across datasets.

Evaluation of GRN inference methods from gene expression data using the Dialogue on Reverse Engineering Assessment and Methods (DREAM) benchmark revealed varied performance across organisms and datasets in most inference methods due to different underlying mathematical assumptions [143]. Notably, integrating predictions from multiple methods (at least three) into community networks yielded substantially improved performance, as complementary strengths offset individual limitations [143]. In addition, ensemble learners, including GENIE3 [109] and GRNBoost2 [111], demonstrated robust performance on par in single-cell RNA-seq data [112, 144],

reflecting the generalizability of the methods. Furthermore, inference methods that incorporate the prior knowledge or estimated TFA outperform the approaches relying solely on expression in both bulk and single-cell RNA-seq data [144, 145].

Single-cell and spatially-resolved gene regulatory network inference

Beyond conventional bulk gene expression profiling, advances in single-cell sequencing and spatial-resolved transcriptomics (SRT) now enable capturing of cellular heterogeneity, spatial organization, and fine-scale regulatory dynamics (see also [146]).

Single-cell sequencing and SRT capture molecular readout as sparse cell-by-gene matrices with tens to hundreds of thousands of columns of cells, with SRT additionally providing spatial coordinates within tissue contexts. GRNs inferred from scRNA-seq hence enable the dissection of cell-type-specific regulatory programs, as well as regulatory changes along differentiation trajectories or state transitions [146, 147] (Fig. 1G). Extending GRN inference to spatially resolved gene expression data enables investigation of how cellular organization and local microenvironments drive intracellular signaling and gene regulation, as well as providing insights into cell–cell interactions and tissue organization [148, 149].

GRN inference from single-cell transcriptomic data

Gene–gene relationships in single-cell gene expression data can be quantified using measures similar to those applied to bulk gene expression data (Fig. 3C–F), with adaptations to account for the unique characteristics of single-cell data, such as cell type-specific expression and sparsity. According to the BEELINE benchmark, both bulk and single-cell GRN inference methods exhibit comparable performance [112]. Among the single-cell inference methods, PIDC [150] achieved the highest overall accuracy and stability, yet captured only co-expression rather than putative regulatory relationships [112]. Notably, tree-based ensemble learners originally developed for bulk RNA-seq, such as GENIE3 and GRNBoost2, demonstrated comparable accuracy to the top performer, suggesting the generalizability and accuracy of the regression methods.

Building on ensemble learners (see earlier section and Fig. 3E), the SCENIC pipeline [151]) was developed for inference of regulatory networks from scRNA-seq data. SCENIC infers regulatory relationships and subsequently refines network components by incorporating TF-binding motif enrichment to retain interactions supported by putative TF-binding evidence. However, its performance and computational complexity may vary substantially with gene numbers as regression trees must be constructed for each gene [152]. Furthermore, GRN inference methods that leverage trajectory inference (i.e. pseudotime information) are dependent on dimensionality reduction methods (e.g. Uniform Manifold Approximation and Projection (UMAP), or t-distributed Stochastic Neighbor Embedding (t-SNE)), which are sensitive to data sparsity and confounding factors [112, 144]). Trajectory inference itself is inherently limited by complex topologies beyond simple temporal lag effects, such as branching or cycling trajectories.

As also discussed earlier, GRN inference based on single-cell gene expression data alone can often lead to spurious connections, which then weaken the inference performance. Alternatively, multimodal experiments on the same samples, existing prior knowledge such as TF binding specificities, or pre-train datasets, can all improve predictive performance or generalizability of the models. Chromatin accessibility profiling via scATAC-seq has been widely adopted with scRNA-seq for inferring causal GRNs through both experimental and computational approaches [145, 153, 154]. Notably, SCENIC+ [153] integrates scRNA-seq and scATAC-seq by first inferring putative GRNs from gene expression using GENIE3 or GRNBoost2, and then incorporating TF-binding motifs from prior knowledge of promoter and enhancer sequences. While multi-omic assays are ideal for capturing multiple layers of regulation and readouts of gene expression in a context-specific manner, they can be cost-intensive and may not be suitable for certain studies. Alternatively, several single-cell GRN inference methods leverage publicly available omics data as prior knowledge, providing a practical and cost-effective strategy for multimodal integration, albeit with limited standardized benchmarking across methods (see [123]). More recently, a large collection of bulk and single-cell omics and causal regulatory datasets have enabled deep learning-based GRN inference through pre-training or fine-tuning, which then support other diverse downstream tasks, including cell type-specific GRN, feature embedding and cell type annotation, TFA estimation, in silico treatment and perturbation, gene dosage sensitivity, chromatin and network dynamics predictions [155–157]. Together, these developments highlight a rapidly evolving landscape for single-cell GRN inference, which is expected to further expand the scope and impact of GRN-based analyses.

Spatially-resolved GRNs and spatial-specific inference tasks

SRT adds an additional dimension beyond the cell-by-gene matrix by providing spatial coordinates of individual cells or clusters of cells in spots. SRT readout varies by technology (i.e. sequencing-based and image-based SRT), and typically pose tradeoffs between throughput and qualities (i.e. resolution and sensitivity) [148]. For more information about the updates on current SRT methods, their concepts and limitations, see [158].

Spatial transcriptomics captures cellular heterogeneity as well as their specific spatial niches, revealing that cells of the same types can adopt distinct GRN states depending on their spatial contexts [159]. These differences arise from the integration of intracellular GRNs with spatially patterned intercellular signaling, such as receptor–ligand interactions [149, 160]. At a larger scale, zonation represents continuous gene expression gradients across spatial domains driven by gradual changes in ligand availability that progressively modulate cell type-specific GRNs [161, 162]. A canonical example is liver zonation, where hepatocytes along the portal-central vein axis showing gradients of oxygen, nutrients and morphogens, leading to gradual reprogramming of metabolism and gene regulation [162].

Emerging spatial inference tasks can broadly be categorized into three primary domains: cell–cell communication (CCC) prediction, cell-specific GRN reconstruction, and joint inference of CCC and GRN. These inference tasks establish a new paradigm for interpreting the mechanistic microenvironment that can be thought of as a spatial neighborhood. For instance, a ligand X from cell type A binds to a receptor Y on neighboring cell type B, modulating the activation of the TF Z regulon, ultimately driving phenotypic changes [163, 164]. CCC focuses on identifying molecular interactions between cells that shape their phenotypes and cellular states. Representative tools such as CellPhoneDB [165], NicheNet [166], and CellChat [167] infer CCC primarily from scRNA-seq data by aggregating gene expression into

cell-specific 'pseudo-bulk' gene expression profiles and, quantifying pairwise relationships between ligand and receptor expression levels in the sender and receiving cells. Statistical testing, characterized marker genes, spatial coordinates, or prior knowledge of TFA, can all be incorporated to prioritize biologically meaningful ligand-receptor pairs.

Spatial GRN inference offers advantages over GRN inference from other transcriptomics modalities in the sense that it not only captures TF-gene regulation but also incorporates cell type specificity and spatial niches. To date, only a limited number of spatial GRN methods support true multimodal integration. As an example, MultiGATE [168] jointly infers heterogeneous GRNs from paired spatial RNA-seq and ATAC-seq through graph embedding, and subsequently integrates with spatial information using a two-level graph attention autoencoder, producing cross-modality heterogeneous networks, and suitable for downstream prediction tasks such as spatial domain prediction, *cis*- and *trans*-regulatory link prediction, and PPI inference. In contrast to earlier multimodal SRT integration methods that were originally developed for single-cell multimodal data integration [169], or incapable of inferring regulatory links [170], MultiGATE represents a dedicated multimodal spatial GRN inference method. However, it is currently limited to paired sequencing-based SRT, suggesting the need for future approaches that integrate unpaired multi-batch SRT data.

Most current spatial inference methods model these events independently, thereby posing some limitations. CCC predictions alone often infer ligand-receptor interaction through correlation and select meaningful interactions by differential correlation and statistical testing [165]. Furthermore, cell-specific GRN inference methods often overlook the influence of ligand-receptor binding events that modulate the signaling pathways governing the TFA. To date, joint inference methods including CLARIFY [171] and SpaGRN [164] model cell-specific GRNs coupled with CCC based on image-based SRT, enabling several prediction tasks, including cell-specific GRNs, spatial domains, cell type-specific regulons, and TFAs.

Summary and future outlook

Transcriptomics provides a genome-wide snapshot of gene expression profiles that enables systematic interrogation of transcriptional regulatory programs across biological conditions. However, similar to other omics approaches, transcriptomics poses substantial analytical challenges due to high dimensionality, sparsity, and complex variable relationships. Consequently, method selection must be matched to the data regime (as outlined in Fig. 2), as no single analytical approach is universally applicable.

Beyond DE analysis of individual genes and transcripts, transcriptomic analyses have expanded to capture multiple layers of regulation and modalities. While conventional DE analysis is sufficient for revealing transcriptional dysregulation between conditions, analyses of transcript isoforms or exons enable investigation of post-transcriptional splicing regulation, including isoform switching, intron retention, and exon inclusion [172, 173]. Furthermore, integrative gene expression data analysis with other omic modalities reveals additional regulatory perspectives, such as aberrant splicing [174], eQTL [73, 175], cell-type deconvolution [176].

Beyond gene-centric analysis, modern transcriptomics has shifted towards systems-level analyses to capture gene expression and regulatory networks. Network-based approaches also provide a framework to interpret complex transcriptional regulation and simulate biological perturbations [177], through investigation of established measures such as centrality, betweenness, cliqueness, differential coexpression, community detection, and signal propagation analysis. More recently, DL models have enabled learning of higher-order and intricate relationships within biological data, supporting predictive and graph-based analyses that extend beyond traditional gene-centric statistical frameworks [94].

Finally, beyond DE analyses of bulk transcriptomics, advances in single-cell and spatial approaches have enabled the investigation of cellular heterogeneity and tissue organization at unprecedented resolution. Both single-cell and spatial data lead to a greater amount of sparse, high-dimensional, and complex data, which may, at least in part, mitigate as well as exacerbate the 'large *p*, small *n*' problem, where the number of 'features' (*p*), such as genes and transcripts, exceeds the number of observations (*n*) [177, 178]. Despite the scale of these datasets that motivate the utilization of DL and foundational models to learn complex biological variable relationships or manifold structures, caution is required to avoid biases arising from intra-individual variability or pseudo-replication [179].

Key Points

- We provide a comprehensive overview of common genome-wide expression study designs and corresponding differential expression (DE) analysis methods.
- We survey current approaches to co-expression and regulatory network inference, which can resolve context-specific regulatory mechanisms and molecular cooperativity.
- We describe current machine learning (ML) and neural network applications for multimodal omics data integration to infer gene regulatory networks, enabling resolving regulatory complexity.
- Recent advances in single-cell and spatial omics provide additional layers for uncovering context-specific dynamics and spatiotemporal heterogeneity.

Acknowledgements

We thank Todsapol Techo, Kanjana Laosuntisuk, Phaewa Chaiwijit, and Yasumin Kerdsri for their feedback and valuable comments on the manuscript during the initial submission as well as revision.

Author contributions

Jiratchaya Nuanpirom (Writing—original draft, Writing—review & editing, Visualization) and Varodom Charoensawan (Conceptualization, Writing—original draft, Writing—review & editing, Supervision)

Conflicts of interest

None declared.

Funding

This work was supported by Scholarships for Ph.D. Student from Mahidol University to JN and VC; the Thailand Rice Science Research Hub of Knowledge (Grant Number: N35E680080), the joint research seeding grant from Macquarie University and Mahidol University (round 2024), and the mid-career researcher grant from National Research Council of Thailand (NRCT) and Mahidol University (Grant Number: N42A670557) to VC.

Data availability

Data sharing is not applicable to this article as no new data were generated or analyzed in this study.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the authors used ChatGPT, Claude and Google Gemini, exclusively to improve readability and for grammar check. After using these tools, the authors carefully reviewed and edited the content as needed, and took full responsibility for the contents of the publication.

References

1. Boye C, Nirmalan S, Ranjbaran A, *et al.* Genotype × environment interactions in gene regulation and complex traits. *Nat Genet* 2024;**56**:1057–68. https://doi.org/10.1038/s41588-024-01776-w.
2. Schmid PR, Palmer NP, Kohane IS, *et al.* Making sense out of massive data by going beyond differential expression. *Proc Natl Acad Sci* 2012;**109**:5594–9. https://doi.org/10.1073/pnas.1118792109.
3. Michna A, Braselmann H, Selmansberger M, *et al.* Natural cubic spline regression modeling followed by dynamic network reconstruction for the identification of radiation-sensitivity Gene Association networks from time-course transcriptome data. *PloS One* 2016;**11**:e0160791. https://doi.org/10.1371/journal.pone.0160791.
4. Schrode N, Ho SM, Yamamuro K, *et al.* Synergistic effects of common schizophrenia risk variants. *Nat Genet* 2019;**51**:1475–85. https://doi.org/10.1038/s41588-019-0497-5.
5. Zagorščak M, Abdelhakim L, Rodriguez-Granados NY, *et al.* Integration of multi-omics data and deep phenotyping provides insights into responses to single and combined abiotic stress in potato. *Plant Physiol* 2025;**197**:kiaf126. https://doi.org/10.1093/plphys/kiaf126.
6. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet* 2019;**20**:631–56. https://doi.org/10.1038/s41576-019-0150-2.
7. Van Den Berge K, Hembach KM, Soneson C, *et al.* RNA sequencing data: Hitchhiker's guide to expression analysis. *Annu Rev Biomed Data Sci* 2019;**2**:139–73. https://doi.org/10.1146/annurev-biodatasci-072018-021255.
8. Li Y, Ge X, Peng F, *et al.* Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biol* 2022;**23**:79. https://doi.org/10.1186/s13059-022-02648-4.
9. Corcelli M, Sagar R, Petzendorfer E, *et al.* Pleiotropic effects of a recessive Col1a2 mutation occurring in a mouse model of severe osteogenesis imperfecta. *PloS One* 2025;**20**:e0309801. https://doi.org/10.1371/journal.pone.0309801.
10. Wolfe CJ, Kohane IS, Butte AJ. Systematic survey reveals general applicability of 'guilt-by-association' within gene coexpression networks. *BMC Bioinformatics* 2005;**6**:227. https://doi.org/10.1186/1471-2105-6-227.
11. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;**9**:559. https://doi.org/10.1186/1471-2105-9-559.
12. Sokolova K, Chen KM, Hao Y, *et al.* Deep learning sequence models for transcriptional regulation. *Annu Rev Genomics Hum Genet* 2024;**25**:105–22. https://doi.org/10.1146/annurev-genom-021623-024727.
13. Chen C, Padi M. Flexible modeling of regulatory networks improves transcription factor activity estimation. *Npj Syst Biol Appl* 2024;**10**:58. https://doi.org/10.1038/s41540-024-00386-w.
14. Siahpirani AF, McCalla SG, Pyne S, *et al.* Uncovering functional gene regulatory networks in bulk and single-cell data through robust transcription factor activity estimation and model-guided experimental validation. Preprint, bioRxiv. 2025, 2025.06.09.658650. https://doi.org/10.1101/2025.06.09.658650.
15. Karbalayghareh A, Sahin M, Leslie CS. Chromatin interaction-aware gene regulatory modeling with graph attention networks. *Genome Res* 2022;**32**:930–44. https://doi.org/10.1101/gr.275870.121.
16. Hossain I, Fanfani V, Fischer J, *et al.* Biologically informed NeuralODEs for genome-wide regulatory dynamics. *Genome Biol* 2024;**25**:127. https://doi.org/10.1186/s13059-024-03264-0.
17. Liu R, Hirn M, Krishnan A. Accurately modeling biased random walks on weighted networks using node2vec+. *Bioinformatics* 2023;**39**:btad047. https://doi.org/10.1093/bioinformatics/btad047.
18. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences, 4:1521. *F1000Research* 2016;**4**:1521. https://doi.org/10.12688/f1000research.7563.2.
19. Trapnell C, Hendrickson DG, Sauvageau M, *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* 2013;**31**:46–53. https://doi.org/10.1038/nbt.2450.
20. Baldoni PL, Chen Y, Hediyeh-zadeh S, *et al.* Dividing out quantification uncertainty allows efficient assessment of differential transcript expression with edgeR. *Nucleic Acids Res* 2024;**52**:e13. https://doi.org/10.1093/nar/gkad1167.
21. Jiang H, Wong WH. Statistical inferences for isoform expression in RNA-seq. *Bioinformatics* 2009;**25**:1026–32. https://doi.org/10.1093/bioinformatics/btp113.
22. Pimentel H, Bray NL, Puente S, *et al.* Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods* 2017;**14**:687–90. https://doi.org/10.1038/nmeth.4324.
23. Yi L, Pimentel H, Bray NL, *et al.* Gene-level differential analysis at transcript-level resolution. *Genome Biol* 2018;**19**:53. https://doi.org/10.1186/s13059-018-1419-z.
24. Zhu A, Srivastava A, Ibrahim JG, *et al.* Nonparametric expression analysis using inferential replicate counts. *Nucleic Acids Res* 2019;**47**:e105–5. https://doi.org/10.1093/nar/gkz622.
25. Tarazona S, García-Alcalde F, Dopazo J, *et al.* Differential expression in RNA-seq: a matter of depth. *Genome Res* 2011;**21**:2213–23. https://doi.org/10.1101/gr.124321.111.
26. Hughes CJ, Fields KM, Danis EP, *et al.* SIX1 and EWS/FLI1 co-regulate an anti-metastatic gene network in Ewing sarcoma. *Nat Commun* 2023;**14**:4357. https://doi.org/10.1038/s41467-023-39945-w.
27. Melé M, Ferreira PG, Reverter F, *et al.* The human transcriptome across tissues and individuals. *Science* 2015;**348**:660–5. https://doi.org/10.1126/science.aaa0355.
28. Tarazona S, Furió-Tarí P, Turrà D, *et al.* Data quality aware analysis of differential expression in RNA-seq with NOISeq R/bioc package. *Nucleic Acids Res* 2015;**43**:e140. https://doi.org/10.1093/nar/gkv711.
29. Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-seq data. *Stat Methods Med Res* 2013;**22**:519–36. https://doi.org/10.1177/0962280211428386.
30. Schurch NJ, Schofield P, Gierliński M, *et al.* How many biological replicates are needed in an RNA-seq experiment and which

differential expression tool should you use? *RNA* 2016;**22**:839–51. https://doi.org/10.1261/rna.053959.115.
31. Braun DA, Hou Y, Bakouny Z, *et al.* Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. *Nat Med* 2020;**26**:909–18. https://doi.org/10.1038/s41591-020-0839-y.
32. Redmond EJ, Ronald J, Davis SJ, *et al.* Single-plant-omics reveals the cascade of transcriptional changes during the vegetative-to-reproductive transition. *Plant Cell* 2024;**36**:4594–606. https://doi.org/10.1093/plcell/koae226.
33. Lyu Y, Li Q. A semi-parametric statistical model for integrating gene expression profiles across different platforms. *BMC Bioinformatics* 2016;**17**:5. https://doi.org/10.1186/s12859-015-0847-y.
34. Higdon R, van Belle G, Kolker E. A note on the false discovery rate and inconsistent comparisons between experiments. *Bioinformatics* 2008;**24**:1225–8. https://doi.org/10.1093/bioinformatics/btn120.
35. Sanchez-Munoz R, Depaepe T, Samalova M, *et al.* Machine-learning meta-analysis reveals ethylene as a central component of the molecular core in abiotic stress responses in Arabidopsis. *Nat Commun* 2025;**16**:4778. https://doi.org/10.1038/s41467-025-59542-3.
36. Zandalinas SI, Sengupta S, Fritschi FB, *et al.* The impact of multifactorial stress combination on plant growth and survival. *New Phytol* 2021;**230**:1034–48. https://doi.org/10.1111/nph.17232.
37. Caesar LK, Cech NB. Synergy and antagonism in natural product extracts: when 1 + 1 does not equal 2. *Nat Prod Rep* 2019;**36**:869–88. https://doi.org/10.1039/C9NP00011A.
38. Schrode N, Seah C, Deans PJM, *et al.* Analysis framework and experimental design for evaluating synergy-driving gene expression. *Nat Protoc* 2021;**16**:812–40. https://doi.org/10.1038/s41596-020-00436-7.
39. Buechler MB, Pradhan RN, Krishnamurty AT, *et al.* Cross-tissue organization of the fibroblast lineage. *Nature* 2021;**593**:575–9. https://doi.org/10.1038/s41586-021-03549-5.
40. Brooks AN, Yang L, Duff MO, *et al.* Conservation of an RNA regulatory map between drosophila and mammals. *Genome Res* 2011;**21**:193–202. https://doi.org/10.1101/gr.108662.110.
41. Wu TY, Goh H, Azodi CB, *et al.* Evolutionarily conserved hierarchical gene regulatory networks for plant salt stress response. *Nat Plants* 2021;**7**:787–99. https://doi.org/10.1038/s41477-021-00929-7.
42. Holmes S, Huber W. *Modern Statistics for Modern Biology*. Cambridge, United Kingdom: Cambridge university press, 2019.
43. Law CW, Zeglinski K, Dong X, *et al.* A guide to creating design matrices for gene expression experiments, 9:1444. *F1000Research* 2020;**9**:1444. https://doi.org/10.12688/f1000research.27893.1.
44. Lund SP, Nettleton D, McCarthy DJ, *et al.* Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat Appl Genet Mol Biol* 2012;**11**:1–42. https://doi.org/10.1515/1544-6115.1826.
45. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550. https://doi.org/10.1186/s13059-014-0550-8.
46. Ritchie ME, Phipson B, Wu D, *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47. https://doi.org/10.1093/nar/gkv007.
47. Leng N, Li Y, McIntosh BE, *et al.* EBSeq-HMM: a Bayesian approach for identifying gene-expression changes in ordered RNA-seq experiments. *Bioinformatics* 2015;**31**:2614–22. https://doi.org/10.1093/bioinformatics/btv193.
48. Spies D, Renz PF, Beyer TA, *et al.* Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Brief Bioinform* 2019;**20**:288–98. https://doi.org/10.1093/bib/bbx115.
49. Fischer DS, Theis FJ, Yosef N. Impulse model-based differential expression analysis of time course sequencing data. *Nucleic Acids Res* 2018;**46**:e119. https://doi.org/10.1093/nar/gky675.
50. Bacher R, Leng N, Chu LF, *et al.* Trendy: segmented regression analysis of expression dynamics in high-throughput ordered profiling experiments. *BMC Bioinformatics* 2018;**19**:380. https://doi.org/10.1186/s12859-018-2405-x.
51. Thaben PF, Westermark PO. Detecting rhythms in time series with RAIN. *J Biol Rhythms* 2014;**29**:391–400. https://doi.org/10.1177/0748730414553029.
52. Wu G, Anafi RC, Hughes ME, *et al.* MetaCycle: an integrated R package to evaluate periodicity in large scale data. *Bioinformatics* 2016;**32**:3351–3. https://doi.org/10.1093/bioinformatics/btw405.
53. Zhao L, Hutchison AT, Liu B, *et al.* Time–restricted eating alters the 24-hour profile of adipose tissue transcriptome in men with obesity. *Obesity (Silver Spring)* 2023;**31**:63–74. https://doi.org/10.1002/oby.23499.
54. Malhan D, Basti A, Relógio A. Transcriptome analysis of clock disrupted cancer cells reveals differential alternative splicing of cancer hallmarks genes. *Npj Syst Biol Appl* 2022;**8**:17. https://doi.org/10.1038/s41540-022-00225-w.
55. Su Y, Hu J, Xia M, *et al.* An undiscovered circadian clock to regulate phytoplankton photosynthesis. *PNAS Nexus* 2024;**3**:pgae497. https://doi.org/10.1093/pnasnexus/pgae497.
56. Ray S, Valekunja UK, Stangherlin A, *et al.* Circadian rhythms in the absence of the clock gene Bmal1. *Science* 2020;**367**:800–6. https://doi.org/10.1126/science.aaw7365.
57. Abruzzi KC, Gobet C, Naef F, *et al.* Comment on "circadian rhythms in the absence of the clock gene Bmal1". *Science* 2021;**372**:eabf0922. https://doi.org/10.1126/science.abf0922.
58. Mure LS, Le HD, Benegiamo G, *et al.* Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *Science* 2018;**359**:eaao0318. https://doi.org/10.1126/science.aao0318.
59. Lappalainen T, Sammeth M, Friedländer MR, *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013;**501**:506–11. https://doi.org/10.1038/nature12531.
60. The GEUVADIS Consortium, 'T Hoen PAC, Friedländer MR, *et al.* Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nat Biotechnol* 2013;**31**:1015–22. https://doi.org/10.1038/nbt.2702.
61. Leek JT, Scharpf RB, Bravo HC, *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010;**11**:733–9. https://doi.org/10.1038/nrg2825.
62. Zheng Y, Liu Y, Yang J, *et al.* Multi-omics data integration using ratio-based quantitative profiling with quartet reference materials. *Nat Biotechnol* 2024;**42**:1133–49. https://doi.org/10.1038/s41587-023-01934-1.
63. Wagner MR, Kleiner M. How thoughtful experimental design can empower biologists in the omics era. *Nat Commun* 2025;**16**:7263. https://doi.org/10.1038/s41467-025-62616-x.
64. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**:139–40. https://doi.org/10.1093/bioinformatics/btp616.
65. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* 2020;**2**:lqaa078. https://doi.org/10.1093/nargab/lqaa078.

66. Jiang L, Schlesinger F, Davis CA, *et al.* Synthetic spike-in standards for RNA-seq experiments. *Genome Res* 2011;**21**:1543–51. https://doi.org/10.1101/gr.121095.111.
67. Risso D, Ngai J, Speed TP, *et al.* Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* 2014;**32**:896–902. https://doi.org/10.1038/nbt.2931.
68. Sun Y, Zhang NR, Owen AB. Multiple hypothesis testing adjusted for latent variables, with an application to the AGEMAP gene expression data. *Ann Appl Stat* 2012;**6**:1664–88. https://doi.org/10.1214/12-AOAS561.
69. Gerard D, Stephens M. Unifying and generalizing methods for removing unwanted variation based on negative controls. *Stat Sin* 2021;**31**:1145–66. https://doi.org/10.5705/ss.202018.0345.
70. Leek JT. Svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Res* 2014;**42**:e161. https://doi.org/10.1093/nar/gku864.
71. Stegle O, Parts L, Durbin R, *et al.* A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol* 2010;**6**:e1000770. https://doi.org/10.1371/journal.pcbi.1000770.
72. Qi T, Wu Y, Fang H, *et al.* Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat Genet* 2022;**54**:1355–63. https://doi.org/10.1038/s41588-022-01154-4.
73. Taylor DJ, Chhetri SB, Tassia MG, *et al.* Sources of gene expression variation in a globally diverse human cohort. *Nature* 2024;**632**:122–30. https://doi.org/10.1038/s41586-024-07708-2.
74. Li J, Bushel PR, Chu TM, *et al.* Principal variance components analysis: Estimating batch effects in microarray gene expression data. In: *Batch Effects and Noise in Microarray Experiments*. John Wiley & Sons, Ltd, 2009, 141–54. https://doi.org/10.1002/9780470685983.ch12.
75. Hoffman GE, Schadt EE. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* 2016;**17**:483. https://doi.org/10.1186/s12859-016-1323-z.
76. Boedigheimer MJ, Wolfinger RD, Bass MB, *et al.* Sources of variation in baseline gene expression levels from toxicogenomics study control animals across multiple laboratories. *BMC Genomics* 2008;**9**:285. https://doi.org/10.1186/1471-2164-9-285.
77. Bates D, Mächler M, Bolker B, *et al.* Fitting linear mixed-effects models using lme4. *J Stat Softw* 2015;**67**:1–48. https://doi.org/10.18637/jss.v067.i01.
78. Yu Y, Mai Y, Zheng Y, *et al.* Assessing and mitigating batch effects in large-scale omics studies. *Genome Biol* 2024;**25**:254. https://doi.org/10.1186/s13059-024-03401-9.
79. Hu X, Li H, Chen M, *et al.* Reference-informed evaluation of batch correction for single-cell omics data with overcorrection awareness. *Commun Biol* 2025;**8**:521. https://doi.org/10.1038/s42003-025-07947-7.
80. Tzec-Interián JA, González-Padilla D, Góngora-Castillo EB. Bioinformatics perspectives on transcriptomics: a comprehensive review of bulk and single-cell RNA sequencing analyses. *Quantitative Biology* 2025;**13**:e78. https://doi.org/10.1002/qub2.78.
81. Comendul A, Ruf-Zamojski F, Ford CT, *et al.* Comprehensive guide for epigenetics and transcriptomics data quality control. *STAR Protocols* 2025;**6**:103607. https://doi.org/10.1016/j.xpro.2025.103607.
82. Sriden N, Charoensawan V. Large-scale comparative transcriptomic analysis of temperature-responsive genes in Arabidopsis thaliana. *Plant Mol Biol* 2022;**110**:425–43. https://doi.org/10.1007/s11103-021-01223-y.
83. Bashkeel N, Perkins TJ, Kærn M, *et al.* Human gene expression variability and its dependence on methylation and aging. *BMC Genomics* 2019;**20**:941. https://doi.org/10.1186/s12864-019-6308-7.
84. Sigalova OM, Shaeiri A, Forneris M, *et al.* Predictive features of gene expression variation reveal mechanistic link with differential expression. *Mol Syst Biol* 2020;**16**:e9539. https://doi.org/10.15252/msb.20209539.
85. Tenenhaus A, Philippe C, Guillemot V, *et al.* Variable selection for generalized canonical correlation analysis. *Biostatistics* 2014;**15**:569–83. https://doi.org/10.1093/biostatistics/kxu001.
86. Argelaguet R, Arnol D, Bredikhin D, *et al.* MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol* 2020;**21**:111. https://doi.org/10.1186/s13059-020-02015-1.
87. Rohart F, Eslami A, Matigian N, *et al.* MINT: a multivariate integrative method to identify reproducible molecular signatures across independent experiments and platforms. *BMC Bioinformatics* 2017;**18**:128. https://doi.org/10.1186/s12859-017-1553-8.
88. Jousse C, Parry L, Cueff G, *et al.* Identification of a specific set of genes predicting obesity before phenotype appearance. *iScience* 2025;**28**:112377. https://doi.org/10.1016/j.isci.2025.112377.
89. McConn JL, Lamoureux CR, Poudel S, *et al.* Optimal dimensionality selection for independent component analysis of transcriptomic data. *BMC Bioinformatics* 2021;**22**:584. https://doi.org/10.1186/s12859-021-04497-7.
90. Chen D, Yan W, Fu LY, *et al.* Architecture of gene regulatory networks controlling flower development in Arabidopsis thaliana. *Nat Commun* 2018;**9**:4534. https://doi.org/10.1038/s41467-018-06772-3.
91. Li X, Zhang H, Hodder T, *et al.* Systems-level design principles of metabolic rewiring in an animal. *Nature* 2025;**640**:203–11. https://doi.org/10.1038/s41586-025-08636-5.
92. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004;**5**:101–13. https://doi.org/10.1038/nrg1272.
93. Jinawath N, Bunbanjerdsuk S, Chayanupatkul M, *et al.* Bridging the gap between clinicians and systems biologists: from network biology to translational biomedical research. *J Transl Med* 2016;**14**:324. https://doi.org/10.1186/s12967-016-1078-3.
94. Zitnik M, Li MM, Wells A, *et al.* Current and future directions in network biology. *Bioinformatics. Advances* 2024;**4**:vbae099. https://doi.org/10.1093/bioadv/vbae099.
95. Johnson KA, Krishnan A. Robust normalization and transformation techniques for constructing gene coexpression networks from RNA-seq data. *Genome Biol* 2022;**23**:1. https://doi.org/10.1186/s13059-021-02568-9.
96. Ballouz S, Verleyen W, Gillis J. Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics* 2015;**31**:2123–30. https://doi.org/10.1093/bioinformatics/btv118.
97. Faith JJ, Hayete B, Thaden JT, *et al.* Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol* 2007;**5**:e8. https://doi.org/10.1371/journal.pbio.0050008.
98. Punchkhon C, Plaimas K, Buaboocha T, *et al.* Drought-tolerance gene identification using genome comparison and Co-expression network analysis of chromosome substitution lines in Rice. *Genes* 2020;**11**:art. 10. https://doi.org/10.3390/genes11101197.
99. Jaemthaworn T, Kalapanulak S, Saithong T. Topological clustering of regulatory genes confers pathogenic tolerance to cassava brown streak virus (CBSV) in cassava. *Sci Rep* 2021;**11**:7872. https://doi.org/10.1038/s41598-021-86806-x.

100. Saelens W, Cannoodt R, Saeys Y. A comprehensive evaluation of module detection methods for gene expression data. *Nat Commun* 2018;**9**:1090. https://doi.org/10.1038/s41467-018-03424-4.
101. Choobdar S, Ahsen ME, Crawford J, *et al.* Assessment of network module identification across complex diseases. *Nat Methods* 2019;**16**:843–52. https://doi.org/10.1038/s41592-019-0509-5.
102. Liu X, Wang Y, Ji H, *et al.* Personalized characterization of diseases using sample-specific networks. *Nucleic Acids Res* 2016;**44**:e164–4. https://doi.org/10.1093/nar/gkw772.
103. Yin L, Traversa P, Elati M, *et al.* Sample-specific network analysis identifies gene co-expression patterns of immunotherapy response in clear cell renal cell carcinoma. *iScience* 2025;**28**:113061. https://doi.org/10.1016/j.isci.2025.113061.
104. Chen HH, Hsueh CW, Lee CH, *et al.* SWEET: a single-sample network inference method for deciphering individual features in disease. *Brief Bioinform* 2023;**24**:bbad032. https://doi.org/10.1093/bib/bbad032.
105. Kuijjer ML, Tung MG, Yuan G, *et al.* Estimating sample-specific regulatory networks. *iScience* 2019;**14**:226–40. https://doi.org/10.1016/j.isci.2019.03.021.
106. Saha E, Fanfani V, Mandros P, *et al.* Bayesian inference of sample-specific coexpression networks. *Genome Res* 2024;**34**:gr.279117.124. https://doi.org/10.1101/gr.279117.124.
107. Huynh-Thu VA, Sanguinetti G. Gene regulatory network inference: An introductory survey. In: Sanguinetti G, Huynh-Thu VA (eds). *Gene Regulatory Networks: Methods and Protocols*. New York, NY: Springer, 2019, 1–23.
108. Margolin AA, Nemenman I, Basso K, *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 2006;**7**:S7. https://doi.org/10.1186/1471-2105-7-S1-S7.
109. Huynh-Thu VA, Irrthum A, Wehenkel L, *et al.* Inferring regulatory networks from expression data using tree-based methods. *PloS One* 2010;**5**:e12776. https://doi.org/10.1371/journal.pone.0012776.
110. Huynh-Thu VA, Geurts P. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. *Sci Rep* 2018;**8**:3384. https://doi.org/10.1038/s41598-018-21715-0.
111. Moerman T, Aibar Santos S, Bravo González-Blas C, *et al.* GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* 2019;**35**:2159–61. https://doi.org/10.1093/bioinformatics/bty916.
112. Pratapa A, Jalihal AP, Law JN, *et al.* Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods* 2020;**17**:147–54. https://doi.org/10.1038/s41592-019-0690-6.
113. Roy S, Lagree S, Hou Z, *et al.* Integrated module and gene-specific regulatory inference implicates upstream signaling networks. *PLoS Comput Biol* 2013;**9**:e1003252. https://doi.org/10.1371/journal.pcbi.1003252.
114. Siahpirani AF, Roy S. A prior-based integrative framework for functional transcriptional regulatory network inference. *Nucleic Acids Res* 2017;**45**:e21. https://doi.org/10.1093/nar/gkw963.
115. Athieniti E, Spyrou GM. A guide to multi-omics data collection and integration for translational medicine. *Comput Struct Biotechnol J* 2023;**21**:134–49. https://doi.org/10.1016/j.csbj.2022.11.050.
116. Argelaguet R, Cuomo ASE, Stegle O, *et al.* Computational principles and challenges in single-cell data integration. *Nat Biotechnol* 2021;**39**:1202–15. https://doi.org/10.1038/s41587-021-00895-7.
117. Zitnik M, Nguyen F, Wang B, *et al.* Machine learning for integrating data in biology and medicine: principles, practice, and opportunities. *Inf Fusion* 2019;**50**:71–91. https://doi.org/10.1016/j.inffus.2018.09.012.
118. Baur B, Shin J, Zhang S, *et al.* Data integration for inferring context-specific gene regulatory networks. *Curr Opin Syst Biol* 2020;**23**:38–46. https://doi.org/10.1016/j.coisb.2020.09.005.
119. Baião AR, Cai Z, Poulos RC, *et al.* A technical review of multi-omics data integration methods: from classical statistical to deep generative approaches. *Brief Bioinform* 2025;**26**:bbaf355. https://doi.org/10.1093/bib/bbaf355.
120. Ovek Baydar D, Rauluseviciute I, Aronsen DR, *et al.* JASPAR 2026: expansion of transcription factor binding profiles and integration of deep learning models. *Nucleic Acids Res* 2026;**54**:D184–93. https://doi.org/10.1093/nar/gkaf1209.
121. Shen WK, Chen SY, Gan ZQ, *et al.* AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. *Nucleic Acids Res* 2023;**51**:D39–45. https://doi.org/10.1093/nar/gkac907.
122. Svinin G, Glaab E. Causal network analysis of omics data using prior knowledge databases. *Brief Bioinform* 2025;**26**:bbaf654. https://doi.org/10.1093/bib/bbaf654.
123. Stock M, Losert C, Zambon M, *et al.* Leveraging prior knowledge to infer gene regulatory networks from single-cell RNA-sequencing data. *Mol Syst Biol* 2025;**21**:214–30. https://doi.org/10.1038/s44320-025-00088-3.
124. Osmanbeyoglu HU, Shimizu F, Rynne-Vidal A, *et al.* Chromatin-informed inference of transcriptional programs in gynecologic and basal breast cancers. *Nat Commun* 2019;**10**:4369. https://doi.org/10.1038/s41467-019-12291-6.
125. Weighill D, Guebila MB, Glass K, *et al.* Predicting genotype-specific gene regulatory networks. *Genome Res* 2022;**32**:524–33. https://doi.org/10.1101/gr.275107.120.
126. Koh HWL, Fermin D, Vogel C, *et al.* iOmicsPASS: network-based integration of multiomics data for predictive subnetwork discovery. *Npj Syst Biol Appl* 2019;**5**:22. https://doi.org/10.1038/s41540-019-0099-y.
127. Castro DM, de Veaux NR, Miraldi ER, *et al.* Multi-study inference of regulatory networks for more accurate models of gene regulation. *PLoS Comput Biol* 2019;**15**:e1006591. https://doi.org/10.1371/journal.pcbi.1006591.
128. Gao Z, Su Y, Xia J, *et al.* DeepFGRN: inference of gene regulatory network with regulation type based on directed graph embedding. *Brief Bioinform* 2024;**25**:bbae143. https://doi.org/10.1093/bib/bbae143.
129. Wang Y, Lee H, Fear JM, *et al.* NetREX-CF integrates incomplete transcription factor data with gene expression to reconstruct gene regulatory networks. *Commun Biol* 2022;**5**:1282. https://doi.org/10.1038/s42003-022-04226-7.
130. Alvarez MJ, Shen Y, Giorgi FM, *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* 2016;**48**:838–47. https://doi.org/10.1038/ng.3593.
131. Hecker D, Lauber M, Behjati Ardakani F, *et al.* Computational tools for inferring transcription factor activity. *PROTEOMICS* 2023;**23**:e2200462. https://doi.org/10.1002/pmic.202200462.
132. Schmidt F, Kern F, Ebert P, *et al.* TEPIC 2—an extended framework for transcription factor binding prediction and integrative epigenomic analysis. *Bioinformatics* 2019;**35**:1608–9. https://doi.org/10.1093/bioinformatics/bty856.
133. Petralia F, Wang P, Yang J, *et al.* Integrative random forest for gene regulatory network inference. *Bioinformatics* 2015;**31**:i197–205. https://doi.org/10.1093/bioinformatics/btv268.

134. Shutta KH, Weighill D, Burkholz R, *et al.* DRAGON: determining regulatory associations using graphical models on multi-Omic networks. *Nucleic Acids Res* 2023;**51**:e15. https://doi.org/10.1093/nar/gkac1157.
135. Altenbuchinger M, Weihs A, Quackenbush J, *et al.* Gaussian and mixed graphical models as (multi-)omics data analysis tools. *Biochim Biophys Acta Gene Regul Mech* 2020;**1863**:194418. https://doi.org/10.1016/j.bbagrm.2019.194418.
136. Glass K, Huttenhower C, Quackenbush J, *et al.* Passing messages between biological networks to refine predicted interactions. *PloS One* 2013;**8**:e64832. https://doi.org/10.1371/journal.pone.0064832.
137. Kuijjer ML, Fagny M, Marin A, *et al.* PUMA: PANDA using MicroRNA associations. *Bioinformatics* 2020;**36**:4765–73. https://doi.org/10.1093/bioinformatics/btaa571.
138. Sonawane AR, DeMeo DL, Quackenbush J, *et al.* Constructing gene regulatory networks using epigenetic data. *Npj Syst Biol Appl* 2021;**7**:45. https://doi.org/10.1038/s41540-021-00208-3.
139. Frey BJ, Dueck D. Clustering by passing messages between data points. *Science* 2007;**315**:972–6. https://doi.org/10.1126/science.1136800.
140. Bailly-Bechet M, Braunstein A, Pagnani A, *et al.* Inference of sparse combinatorial-control networks from gene-expression data: a message passing approach. *BMC Bioinformatics* 2010;**11**:355. https://doi.org/10.1186/1471-2105-11-355.
141. Chantaraamporn J, Phumikhet P, Nguantad S, *et al.* Machine learning applications for transcription level and phenotype predictions. *IUBMB Life* 2022;**74**:1273–87. https://doi.org/10.1002/iub.2693.
142. Hegde A, Nguyen T, Cheng J. Machine learning methods for gene regulatory network inference. *Brief Bioinform* 2025;**26**:bbaf470. https://doi.org/10.1093/bib/bbaf470.
143. The DREAM5 Consortium, Marbach D, Costello JC, *et al.* Wisdom of crowds for robust gene network inference. *Nat Methods* 2012;**9**:796–804. https://doi.org/10.1038/nmeth.2016.
144. McCalla SG, Fotuhi Siahpirani A, Li J, *et al.* Identifying strengths and weaknesses of methods for computational network inference from single-cell RNA-seq data. *G3 (Bethesda)* 2023;**13**:jkad004. https://doi.org/10.1093/g3journal/jkad004.
145. Skok Gibbs C, Jackson CA, Saldi GA, *et al.* High-performance single-cell gene regulatory network inference at scale: the Inferelator 3.0. *Bioinformatics* 2022;**38**:2519–28. https://doi.org/10.1093/bioinformatics/btac117.
146. Badia-i-Mompel P, Wessels L, Müller-Dott S, *et al.* Gene regulatory network inference in the era of single-cell multi-omics. *Nat Rev Genet* 2023;**24**:739–54. https://doi.org/10.1038/s41576-023-00618-5.
147. Longo SK, Guo MG, Ji AL, *et al.* Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat Rev Genet* 2021;**22**:627–44. https://doi.org/10.1038/s41576-021-00370-8.
148. Rao A, Barkley D, França GS, *et al.* Exploring tissue architecture using spatial transcriptomics. *Nature* 2021;**596**:211–20. https://doi.org/10.1038/s41586-021-03634-9.
149. Armingol E, Baghdassarian HM, Lewis NE. The diversification of methods for studying cell–cell interactions and communication. *Nat Rev Genet* 2024;**25**:381–400. https://doi.org/10.1038/s41576-023-00685-8.
150. Chan TE, Stumpf MPH, Babtie AC. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Systems* 2017;**5**:251–267.e3. https://doi.org/10.1016/j.cels.2017.08.014.
151. Aibar S, González-Blas CB, Moerman T, *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 2017;**14**:1083–6. https://doi.org/10.1038/nmeth.4463.
152. Nguyen H, Tran D, Tran B, *et al.* A comprehensive survey of regulatory network inference methods using single cell RNA sequencing data. *Brief Bioinform* 2021;**22**:bbaa190. https://doi.org/10.1093/bib/bbaa190.
153. Bravo González-Blas C, De Winter S, Hulselmans G, *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* 2023;**20**:1355–67. https://doi.org/10.1038/s41592-023-01938-4.
154. Kamimoto K, Stringa B, Hoffmann CM, *et al.* Dissecting cell identity via network inference and in silico gene perturbation. *Nature* 2023;**614**:742–51. https://doi.org/10.1038/s41586-022-05688-9.
155. Theodoris CV, Xiao L, Chopra A, *et al.* Transfer learning enables predictions in network biology. *Nature* 2023;**618**:616–24. https://doi.org/10.1038/s41586-023-06139-9.
156. Kalfon J, Samaran J, Peyré G, *et al.* scPRINT: pre-training on 50 million cells allows robust gene network predictions. *Nat Commun* 2025;**16**:3607. https://doi.org/10.1038/s41467-025-58699-1.
157. Yuan Q, Duren Z. Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. *Nat Biotechnol* 2025;**43**:247–57. https://doi.org/10.1038/s41587-024-02182-7.
158. Liu Y, Dai Y, Wang L. Spatial omics at the forefront: emerging technologies, analytical innovations, and clinical applications. *Cancer Cell* 2026;**44**:24–49. https://doi.org/10.1016/j.ccell.2025.12.009.
159. Palla G, Fischer DS, Regev A, *et al.* Spatial components of molecular tissue biology. *Nat Biotechnol* 2022;**40**:308–18. https://doi.org/10.1038/s41587-021-01182-1.
160. Vento-Tormo R, Efremova M, Botting RA, *et al.* Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* 2018;**563**:347–53. https://doi.org/10.1038/s41586-018-0698-6.
161. Ben-Moshe S, Itzkovitz S. Spatial heterogeneity in the mammalian liver. *Nat Rev Gastroenterol Hepatol* 2019;**16**:395–410. https://doi.org/10.1038/s41575-019-0134-x.
162. Bravo González-Blas C, Matetovici I, Hillen H, *et al.* Single-cell spatial multi-omics and deep learning dissect enhancer-driven gene regulatory networks in liver zonation. *Nat Cell Biol* 2024;**26**:153–67. https://doi.org/10.1038/s41556-023-01316-4.
163. Zhang Y, Liu T, Hu X, *et al.* CellCall: integrating paired ligand–receptor and transcription factor activities for cell–cell communication. *Nucleic Acids Res* 2021;**49**:8520–34. https://doi.org/10.1093/nar/gkab638.
164. Li Y, Liu X, Guo L, *et al.* SpaGRN: investigating spatially informed regulatory paths for spatially resolved transcriptomics data. *Cell Systems* 2025;**16**:101243. https://doi.org/10.1016/j.cels.2025.101243.
165. Troulé K, Petryszak R, Cakir B, *et al.* CellPhoneDB v5: inferring cell–cell communication from single-cell multiomics data. *Nat Protoc* 2025;**20**:3412–40. https://doi.org/10.1038/s41596-024-01137-1.
166. Sang-aram C, Browaeys R, Seurinck R, *et al.* Unraveling cell–cell communication with NicheNet by inferring active ligands from transcriptomics data. *Nat Protoc* 2025;**20**:1439–67. https://doi.org/10.1038/s41596-024-01121-9.
167. Jin S, Plikus MV, Nie Q. CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. *Nat Protoc* 2025;**20**:180–219. https://doi.org/10.1038/s41596-024-01045-4.
168. Miao J, Li J, Xin J, *et al.* MultiGATE: integrative analysis and regulatory inference in spatial multi-omics data via graph representation learning. *Nat Commun* 2025;**16**:9403. https://doi.org/10.1038/s41467-025-63418-x.

169. Hao Y, Hao S, Andersen-Nissen E, *et al.* Integrated analysis of multimodal single-cell data. *Cell* 2021;**184**:3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
170. Long Y, Ang KS, Sethi R, *et al.* Deciphering spatial domains from spatial multi-omics with SpatialGlue. *Nat Methods* 2024;**21**:1658–67. https://doi.org/10.1038/s41592-024-02316-4.
171. Bafna M, Li H, Zhang X. CLARIFY: cell–cell interaction and gene regulatory network refinement from spatially resolved transcriptomics. *Bioinformatics* 2023;**39**:i484–93. https://doi.org/10.1093/bioinformatics/btad269.
172. Love MI, Soneson C, Patro R. Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification. *F1000Research* 2018;**7**:952. https://doi.org/10.12688/f1000research.15398.3.
173. Baldoni PL, Chen L, Li M, *et al.* Dividing out quantification uncertainty enables assessment of differential transcript usage with limma and edgeR. *Nucleic Acids Res* 2025;**53**:gkaf1305. https://doi.org/10.1093/nar/gkaf1305.
174. Yépez VA, Mertes C, Müller MF, *et al.* Detection of aberrant gene expression events in RNA sequencing data. *Nat Protoc* 2021;**16**:1276–96. https://doi.org/10.1038/s41596-020-00462-5.
175. Dobbyn A, Huckins LM, Boocock J, *et al.* Landscape of conditional eQTL in dorsolateral prefrontal cortex and Co-localization with schizophrenia GWAS. *Am J Hum Genet* 2018;**102**:1169–84. https://doi.org/10.1016/j.ajhg.2018.04.011.
176. Newman AM, Steen CB, Liu CL, *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019;**37**:773–82. https://doi.org/10.1038/s41587-019-0114-2.
177. Ham L, Woodward TE, Coomer MA, *et al.* Mapping, modeling, and reprogramming cell-fate decision-making systems. *Annu Rev Biomed Data Sci* 2025;**8**:537–62. https://doi.org/10.1146/annurev-biodatasci-101424-121439.
178. Angerer P, Simon L, Tritschler S, *et al.* Single cells make big data: new challenges and opportunities in transcriptomics. *Curr Opin Syst Biol* 2017;**4**:85–91. https://doi.org/10.1016/j.coisb.2017.07.004.
179. Zimmerman KD, Espeland MA, Langefeld CD. A practical solution to pseudoreplication bias in single-cell studies. *Nat Commun* 2021;**12**:738. https://doi.org/10.1038/s41467-021-21038-1.
180. Meyer PE, Lafitte F, Bontempi G. minet: A R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics* 2008;**9**:461. https://doi.org/10.1186/1471-2105-9-461.
181. Sales G, Romualdi C. parmigene—a parallel R package for mutual information estimation and gene network reconstruction. *Bioinformatics* 2011;**27**:1876–7. https://doi.org/10.1093/bioinformatics/btr274.
182. Gysi DM, Voigt A, Fragoso T de M, *et al.* wTO: an R package for computing weighted topological overlap and a consensus network with integrated visualization tool. *BMC Bioinformatics* 2018;**19**:392. https://doi.org/10.1186/s12859-018-2351-7.
183. Russo PST, Ferreira GR, Cardozo LE, *et al.* CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. *BMC Bioinformatics* 2018;**19**:56. https://doi.org/10.1186/s12859-018-2053-1.
184. Lemoine GG, Scott-Boyer MP, Ambroise B, *et al.* GWENA: gene co-expression networks analysis and extended modules characterization in a single Bioconductor package. *BMC Bioinformatics* 2021;**22**:267. https://doi.org/10.1186/s12859-021-04179-4.

Briefings in Bioinformatics, 2026, **27**, bbag152
https://doi.org/10.1093/bib/bbag152
Review