Review

Network-based approaches for modeling disease regulation and progression

Gihanna Galindez [a,b,1], Sepideh Sadegh [c,d,1], Jan Baumbach [d,e], Tim Kacprowski [a,b,*,2], Markus List [c,**,2]

[a] Division Data Science in Biomedicine, Peter L. Reichertz Institute for Medical Informatics of Technische Universität Braunschweig and Hannover Medical School, Braunschweig, Germany
[b] Braunschweig Integrated Centre of Systems Biology (BRICS), TU Braunschweig, Braunschweig, Germany
[c] Chair of Experimental Bioinformatics, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany
[d] Institute for Computational Systems Biology, University of Hamburg, Hamburg, Germany
[e] Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark

ARTICLE INFO

ABSTRACT

Molecular interaction networks lay the foundation for studying how biological functions are controlled by the complex interplay of genes and proteins. Investigating perturbed processes using biological networks has been instrumental in uncovering mechanisms that underlie complex disease phenotypes. Rapid advances in omics technologies have prompted the generation of high-throughput datasets, enabling large-scale, network-based analyses. Consequently, various modeling techniques, including network enrichment, differential network extraction, and network inference, have proven to be useful for gaining new mechanistic insights. We provide an overview of recent network-based methods and their core ideas to facilitate the discovery of disease modules or candidate mechanisms. Knowledge generated from these computational efforts will benefit biomedical research, especially drug development and precision medicine. We further discuss current challenges and provide perspectives in the field, highlighting the need for more integrative and dynamic network approaches to model disease development and progression.

© 2022 Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

* Corresponding author at: Division Data Science in Biomedicine, Peter L. Reichertz Institute for Medical Informatics of Technische Universität Braunschweig and Hannover Medical School, Braunschweig, Germany.
** Corresponding author.
  E-mail addresses: t.kacprowski@tu-braunschweig.de (T. Kacprowski), markus.list@tum.de (M. List).
1 Co-first authors
2 Joint senior authors

1. Introduction

Various biomolecular components of a cell work together to carry out fundamental molecular processes and functions and ultimately influence clinical phenotypes and complex traits. Most diseases do not arise from a single gene mutation, but their etiology could be driven by multiple genes or perturbations that accumulate over time through interacting molecular components [1,2]. Complex biological systems can be represented in the form of networks or graphs, in which nodes represent molecules (e.g. genes or proteins) and edges indicate physical or functional relationships. Examples of biological networks include protein-protein interaction (PPI) [1,2], co-expression [3], metabolic [4], signaling [5,6], and gene regulatory networks (GRNs) [7]. A major challenge in biomedical research is to extract meaningful and interpretable knowledge from molecular profiling (omics) data [8]. Consequently, the field of systems medicine, which deals with network-based modeling of biological systems, has led to deeper and even mechanistic insights from the accumulating wealth of omics data (see Fig. 1 for an overview of these techniques). In particular, network-based approaches to study diverse disease contexts have important clinical and biomedical applications, including hypothesis generation, biomarker discovery, disease classification, target prioritization, and drug repurposing [9]. We review recent network-based approaches used to study diseases, focusing on methods proposed in the last five years and some classical methods, and offer perspectives in the field. First, we discuss *de novo* network enrichment, which overlays molecular profiles onto a biological network to obtain new candidate disease mechanisms. Next, we discuss gene regulatory network inference, which aims to recover the set of gene-gene relationships from high-throughput data. We further discuss the emerging areas of differential co-expression networks and sample-specific networks.
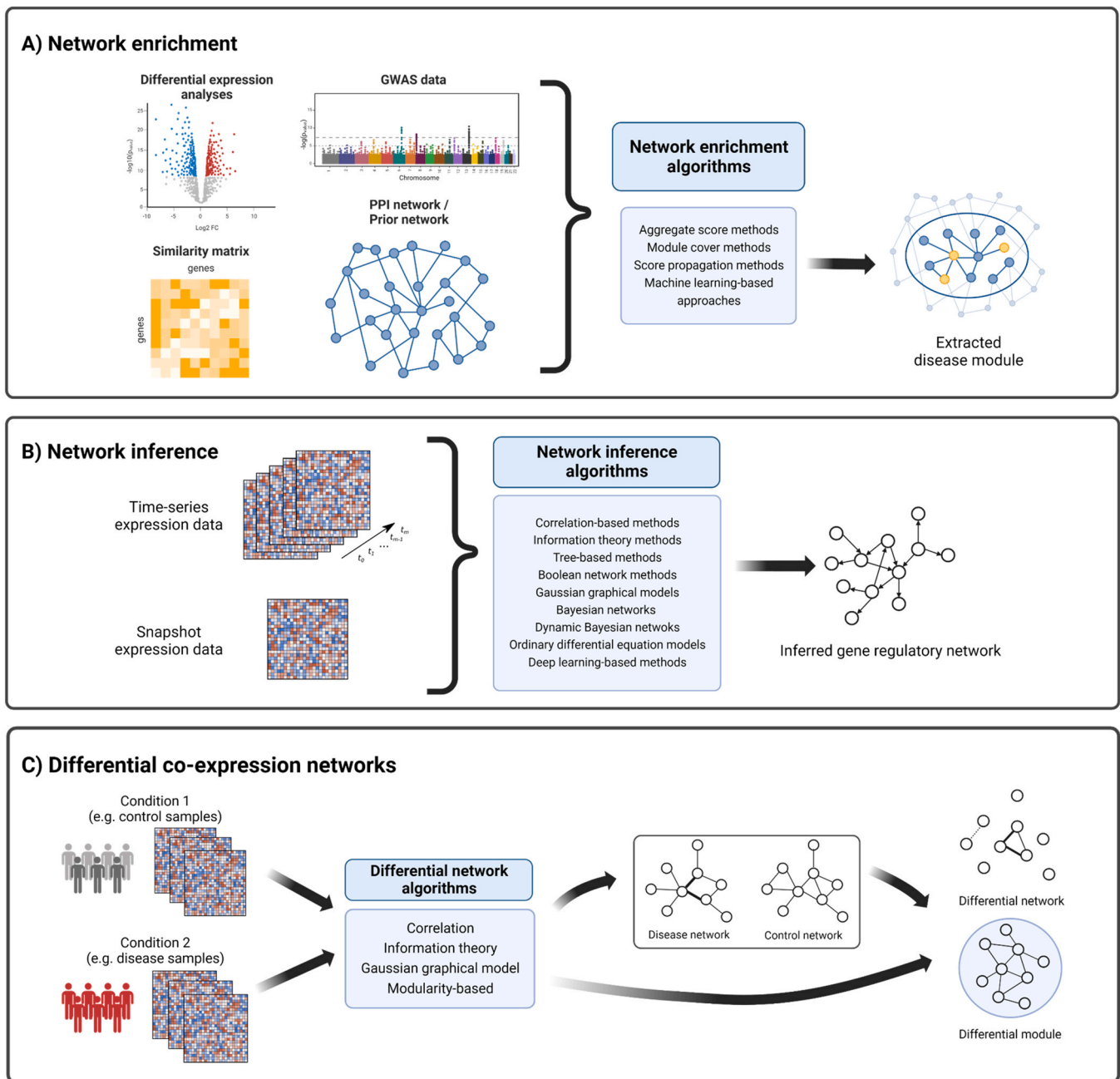
2. *De novo* network enrichment

De novo network enrichment methods (DNE), also referred to as active module identification methods, can be used to identify a disease module, i.e. a connected subnetwork of the human interactome that can be linked to a disease of interest [10]. The concept of disease modules is based on the observation that disease genes are not scattered randomly, but, due to their functional association, tend to be highly connected among themselves or in the same neighborhood in the interactome [11,12]. The accurate identification of disease modules can help to identify new disease genes and pathways and aid rational drug target identification [13,14].

In contrast to classical enrichment analysis, i.e. using Fisher's exact test or gene set enrichment analysis [15,16], DNE extracts condition-specific subnetworks in a more data-driven manner. Rather than relying on predefined pathways or curated gene sets [17], which are unlikely to reveal new mechanisms underlying diseases, DNE methods construct "active" subnetworks by projecting experimental data (mostly transcriptomic or genomic profiles) onto a molecular interaction network. Candidate subnetworks are then scored by an objective function and heuristics are implemented to identify locally optimal solutions efficiently. Batra et al. compared seven state-of-the-art DNE tools and concluded that the optimal subnetwork identification strategy depends on the specific application [17]. Expanding on the categories suggested by Batra et al., we review some of the recently published DNE methods (Table 1).

Aggregate score methods compute a summary score for a candidate subnetwork based on the assigned scores to individual genes. Scores are typically calculated using fold changes or *P*-values derived from differential expression analyses. SigMod takes gene-level *P*-values from genome-wide association studies (GWAS) and a network as input and implements a min-cut algorithm to identify optimally enriched disease modules [18]. Based on Kruskal's algorithm for obtaining a minimum spanning tree, IODNE scores nodes and edges based on differential expression and the topology of the PPI network [19]. Alternatively, several tools extract subnetworks by solving the prize-collecting Steiner forest (PCSF) problem. PCSF maximizes an objective function in which node scores (here called prizes) count positively, whereas edge weights, reflecting the confidence of an interaction, are counted negatively. In the R package PCSF, differential expression, mutation status, or copy number data can be used [20]. Similarly, the Omics Integrator package solves a variation of the PCSF problem with support for other omics types [21]. Menche et al. demonstrated that pairwise shortest paths of known disease genes show a significant left shift in their distribution compared to the random expectation [12]. Therefore, Minimum-weight Steiner trees (MWSTs), that can be viewed as generalizations of shortest paths with more than two endpoints (terminals), are a good choice to cover a large fraction of the disease-relevant molecular pathways. Using known disease genes as seeds (terminals), MuST identifies a disease module by aggregating several approximations of Steiner trees into a single subnetwork [22]. ROBUST is another method using MWST which offers improved robustness compared to MuST by enumerating pairwise diverse rather than pairwise non-identical disease modules [23]. Levi et al. demonstrated that most DNE methods using gene expression data do not fully exploit the information contained in this data [24]. They tried to address this issue in DOMINO by aiming to find disjoint connected Steiner trees in which the active genes (seeds) are over-represented [24].

Module cover approaches either accept a user-provided list of relevant (active) genes for a specific condition (e.g. the result of differential gene expression analysis) or implement a separate pre-processing step to determine such genes. These tools then extract subnetworks that "cover" a large number of the pre-selected active genes. For instance, KeyPathwayMiner expects molecular profiles encoded as binary indicator matrices as input to solve a variant of the maximal connected subnetwork problem [25]. ModuleDiscoverer identifies subnetworks based on the maximum clique enumeration problem. Starting from random seed proteins, the algorithm iteratively extends cliques of size 3, selects cliques that are significantly enriched in differentially expressed genes, and assembles them into regulatory modules [26]. NoMAS operates on mutation profiles and finds subnetworks that are enriched in mutations associated with survival time by solving the maximal connected *k*-set log-rank problem [27]. A related tool is nCOP, which utilizes individual mutation profiles based on the minimum connected set cover problem and finds a connected subnetwork that covers as many patients as possible who collectively harbor alterations in genes involved in pathways [28].

Score propagation methods assign initial scores to the nodes and propagate them through the network before extracting high-scoring subnetworks. NetDecoder extracts sparse subnetworks as directed graphs using information flow between sources and sinks that act as regulators. It first creates a weighted interactome by integrating a PPI network with a co-expression network generated

Fig. 1. Overview of main network-based techniques for modeling human diseases. A) Network enrichment aims to find active subnetworks associated with a specific disease state from arbitrary networks. B) Gene regulatory network inference methods infer relationships between a regulator and its target gene by using snapshot data (condition-specific) and/or time-series omics data. C) Differential network analysis identifies interactions that are rewired between two conditions. Created with BioRender.com.

from context-specific data. The weights represent Pearson correlation across samples of the same context, determining the capacity for an edge. Sources are defined as DEGs, mutated genes, drug targets or any custom gene set, while sinks are defined as genes involved in transcriptional regulation based on GO annotations [29]. Methods using heat diffusion models have the problem that highly mutated genes (hot nodes), such as TP53, propagate the heat to their neighboring genes which may not be mutated and are not biologically interesting. To overcome this issue, HotNet2 uses an insulated heat diffusion process and considers the directionality of heat flow in the identification of subnetworks. HotNet2 considers both the individual gene scores (e.g. mutation frequency) and the local network topology [30]. It has been extended with a random walk-based network clustering approach implemented in Hierarchical HotNet

[31]. The latter combines network interactions and vertex scores to construct a hierarchy of high-weight, topologically close subnetworks. Hierarchical HotNet is more flexible, more robust and less computationally intensive than HotNet2.

Machine learning-based approaches score subnetworks based on their predictive performance in supervised learning. Grand Forest implements a graph-constrained random forest algorithm, extracting subnetworks of high feature importance with respect to various response variables, such as phenotype, symptom, survival, or treatment response [32]. The N2V-HC approach constructs an integrated network using known PPIs, GWAS and expression Quantitative Trait Loci (eQTL) data and employs deep representation learning on this multi-layer network. The nodes are embedded using the node2vec algorithm, based on a biased random walk, to

Table 1
Recently developed network enrichment tools categorized by algorithm type. Links to code/tools are provided in Supplementary Table 1.

Tool	Input data type (s)	Algorithm Type	Example application	Reference
SigMod	GWAS	Aggregate score	Identify functionally and biologically relevant genes in childhood-onset asthma[18]	[18]
IODNE	Gene expression	Aggregate score	Identify potentially novel target genes for drug selection in triple-negative breast cancer[19]	[19]
PCSF	Multi-omics data (gene expression, mutation profiles, or copy number)	Aggregate score	Extract subnetworks of enriched metabolite interactions in multiple sclerosis[36]	[20]
Omics Integrator	Gene expression	Aggregate score	Link α-synuclein to multiple parkinsonism genes and druggable targets[37]	[21]
MuST	Disease-associated genes (derived from GWAS or DEG analyses)	Aggregate score	Investigation of coagulation pathway in COVID-19[38]	[22]
ROBUST	Disease-associated genes (derived from GWAS or DEG analyses)	Aggregate score	Identify an oxidative stress module in multiple sclerosis[23]	[23]
DOMINO	Disease-associated genes (derived from GWAS or DEG analyses)	Aggregate score	Integrated as the downstream analysis step in a splicing-aware framework for time course data analysis[39]	[24]
KeyPathwayMiner	Gene expression / multi-omics data	Module cover	Reveal epigenetic targets in SARS-CoV-2 infection, used together with gene co-expression networks[40]	[25]
ModuleDiscoverer	Gene expression	Module cover	Identify regulatory modules of the response to mesenchymal stromal cells treatment for mitigating liver damage[41]	[26]
NoMAS	Mutation profiles	Module cover	Identify subnetworks with strong association to survival in different cancer types[27]	[27]
nCOP	Mutation profiles	Module cover	Identify cancer genes including those with low mutation frequencies across 24 different cancer types[28]	[28]
NetDecoder	Gene expression; mutation profiles	Score propagation	Studying mechanism of CCN1-associated resistance to HSV-1-derived oncolytic immunovirotherapies for glioblastomas[42]	[29]
HotNet2	Mutation profiles	Score propagation	Identify mutated subnetworks in a genome-scale interaction network used for subtyping pancreatic cancer[43]	[30]
Hierarchical HotNet	Mutation profiles	Score propagation	Detect functional interactions that connect the cellular targets of viral proteins with the downstream changes in SARS-CoV and SARS-CoV2[44]	[31]
Grand Forest	Gene expression / methylation data	Machine learning	Stratify lung cancer patients into clinically relevant molecular subgroups[32]	[32]
N2V-HC	GWAS	Machine learning	Discover biologically meaningful modules related to pathways underlying Parkinson's disease and Alzheimer's diseases[33]	[33]
BiCoN	Gene expression / methylation data	Machine learning	Applied to TCGA breast cancer data for subtyping[34]	[34]
TiCoNE	Time-series expression data	Machine learning	Analysis of time-series lipidomics dataset of human mesenchymal stem cells after drug treatment[45]	[35]

incorporate global connectivity features in the network. Afterwards, the learned representations are subjected to hierarchical clustering. An iterative selection procedure using enrichment analysis is then applied to produce the final modules [33]. Most existing disease module mining methods rely on case-versus-control annotations. Unsupervised approaches have the advantage that they can not only return disease modules but simultaneously *de novo* cluster patients into subgroups. Biclustering Constrained by Networks (BiCoN) is one of the few such unsupervised approaches which is based on the Ant colony optimization algorithm. It uses a heterogeneous network of patients and genes, where genes are linked to patients via the expression data and are linked with each other via PPIs [34]. Time-course network enrichment (TiCoNE) identifies temporal response pathways by clustering genes with similar co-expression patterns. It uses as input time-series expression data and benefits from a human-augmented approach, i.e. it allows users to add, remove, merge, or split temporal patterns based on their domain knowledge. Induced subnetworks from the PPI network can then be generated from the identified gene clusters [35].

3. Gene regulatory network inference

Gene regulation is a fundamental process that influences cell fate and development, response to stimuli, and disease progression. In particular, transcription factors (TFs) regulate other genes and many methods have been developed to model these relationships as networks. Transcriptional gene regulation primarily involves the binding of transcription factors (TFs) at both promoter and enhancer regions and the physical interaction of these bound complexes via DNA looping. The problem of network inference aims to reconstruct the structure of the GRN, where a (typically directed) edge indicates an interaction between a regulator (e.g. transcription factors acting as activators or repressors) and its target gene. Network inference or "reverse-engineering" of biological networks using omics data (e.g. gene expression data) is useful for identifying genes that may influence the phenotypic changes associated with a disease of interest. Ideally, causal and direct relationships are reconstructed; however, regulatory interactions may be non-linear and complex. Reverse-engineering of GRNs remains a major challenge, as evident from large community efforts [46]. Sample sizes from biological experiments are typically small, ranging from tens to hundreds, relative to the number of variables, in the order of tens of thousands of genes or transcripts. Consequently, network inference is an underdetermined problem in which many models fit the data and an exponentially large space of networks needs to be considered [47]. Various computational GRN inference approaches have been developed, which include correlation-, information theory-, regression-based techniques, Gaussian graphical models, Boolean networks, Bayesian networks, and ordinary differential equation (ODE)-based methods. Depending on the approach, these methods could utilize either snapshot data (gene expression data sampled from a well-defined disease condition or phenotypic stage) or temporal data (samples are obtained from a series of time points). We provide an overview of these methods in Table 2.

3.1. Network inference from snapshot data

Traditional methods of constructing GRNs use correlation or mutual information as measures of edge confidence. Correlation-based methods are the most straightforward approach to infer pairwise relationships by calculating the correlation coefficient, resulting in weighted, undirected networks. Examples of correlation-based approaches include the popular weighted gene correlation network analysis (WGCNA) [48]. WGCNA constructs a correlation-based adjacency matrix, which is raised to a suitable power (ß parameter) such that the resulting network exhibits scale-free topology and ensures similar connectivity between neighboring genes. While inference using correlation-based network methods is fast, they typically produce a large number of false positives and output very dense correlation networks [49]. Information theoretic methods rely on statistical measures based on mutual information (MI), a non-linear correlation measure, to construct the networks. Popular MI-based tools include algorithms for the reconstruction of accurate cellular networks (ARACNE) [50,51], Context Likelihood of Relatedness (CLR) [52], and minimum-redundancy network (MRNET) [53]. In general, these methods calculate pairwise MI values, followed by a pruning step to eliminate indirect interactions. MI-based methods have low computational complexity and can better detect complex, non-linear dependencies, but they can only be applied to snapshot data.

Most regression based methods formalize the GRN inference problem as a feature selection problem and construct the GRN with an ensemble strategy. A popular approach is GENIE3, which builds ensembles of regression trees to predict the expression level of each target gene using the expression levels of TFs or other genes and subsequently aggregates the results to reconstruct the final GRN [54]. GRNBoost2 and Arboreto build on the GENIE3 approach by implementing gradient boosting of the trees to improve performance and runtime [54,55]. PLSNET is another ensemble approach that uses the partial least squares method (PLS) to select features that can predict the expression of the target genes [56].

GRNs are also often represented using Gaussian graphical models (GGM). GGMs are probabilistic models of conditional dependencies between variables and can be used to distinguish direct from indirect interactions. In a GGM, edges are defined as node pairs with non-zero partial correlation. Hence, an edge in the model indicates the dependency between two genes after removing the effects of all other genes. The GGM can be derived by calculating the inverse covariance matrix (precision matrix), with the nodes assumed to follow a multivariate normal distribution. One of the most widely used methods to estimate a GGM is the graphical LASSO, which uses L1 regularization [57]. FastGGM implements a penalized regression approach that outputs confidence intervals for each edge in the GRN [58,59].

A Bayesian network is a probabilistic graphical model that encodes the conditional dependencies between the random variables (genes) in a directed acyclic graph (DAG) structure. The Bayesian network formalism naturally encodes causality. Learning the optimal network from data requires estimating the network structure and parameters [60]is a non-trivial task because of the considerably large space of possible graphs. Various methods have been proposed for Bayesian network structure learning, which can be generally categorized as score-based or constraint-based [60]. Score-based methods use a score function (e.g. Bayesian Dirichlet equivalence score and the Bayesian information criterion score) to evaluate the structures over the search space of possible network structures. Heuristic methods, such as hill climbing, tabu search, and simulated annealing have been proposed to search for the optimal network structures [61]. On the other hand, constraint-based methods attempt to find the skeleton of the DAG by performing conditional dependence tests among the nodes. Afterwards, edge directions are determined by resolving directional constraints on the skeleton structure. However, Bayesian network methods are limited to up to hundreds of genes and are difficult to scale to thousands of genes. The BiDAG method uses the constraint-based PC algorithm [62] to restrict the search space and generate an initial skeleton. Afterwards, it implements a Markov chain Monte Carlo method to sample the network structures from the posterior distribution [63,64]. The local Bayesian network (LBN) method proceeds in four main steps. It first uses MI or conditional mutual information to generate a tentative skeleton structure and reduce the number of candidate regulators. Afterwards, a k-nearest neighbor algorithm is applied to decompose

Table 2
Recently developed network inference tools, input data, and their core algorithmic approaches. Links to code/tools are provided in Supplementary Table 2.

Tool	Input data	Algorithmic approach	Example application	Reference
WGCNA	Snapshot expression data	Correlation	Identify differences in brain co-expression networks underlying the pathology of autism[73]	[48]
ARACNE	Snapshot expression data	Information theory	Resolve the regulatory structure of co-expression modules and identify disease drivers in post-traumatic stress disorder[74]	[50,51]
CLR	Snapshot expression data	Information theory	Inference of miRNA co-expression network in type 2 diabetes[75]	[52]
MRNET	Snapshot expression data	Information theory	Network analysis of WNT1-regulated network in neurodegeneration[76]	[77]
GENIE3	Snapshot expression data	Regression trees	Identification of TEADs as regulators of invasive cell states in melanoma[78]	[54]
GRNBoost2	Snapshot expression data	Regression trees	Investigating the role of NKD2 in kidney fibrosis[79]	[55]
PLSNET	Snapshot expression data	Partial least squares regression	Identify new ERRα partners in breast cancer[80]	[56]
FastGM	Snapshot expression data	Gaussian graphical model	Applied to gene expression datasets in asthma and Alzheimer's disease[58]	[58]
BiDAG	Snapshot expression data	Bayesian network; Dynamic Bayesian network	Identify network transition structures in colorectal cancer[81]	[64]
LBN	Snapshot expression data	Bayesian network	Evaluated on DREAM3 challenge dataset and the E. coli SOS DNA repair network[65]	[65]
ATEN	Time-series expression data	Boolean network	Evaluated on Drosophila segment polarity gene regulatory network[82]	[82]
GABNI	Time-series expression data	Boolean network	Benchmarked on DREAM dataset and yeast cell cycle dataset[83]	[83]
dynGENIE3	Time-series expression data	Ordinary differential equations + Regression	Construction of GRN governing circadian clock[84]	[85]
Inferelator	Time-series expression data	Ordinary differential equations + Regression	Learning the Bacillus subtilis regulatory network[86]	[87–89]
Ma et al.	Snapshot and time-series expression data	Ordinary differential equations	Benchmarked on DREAM4 dataset and E. coli and S. cerevisiae datasets[65]	[90]
BETS	Time-series expression data	Granger causality	Predicting the transcriptional response of A549 cells to glucocorticoids[91]	[91]
SWING	Time-series expression data	Granger causality	Infer time-delayed edges in a SOS response network in E. coli following DNA damage[92]	[92]
iRafNet	Snapshot expression data; prior knowledge	Regression trees	Studying clear cell renal cell carcinoma using phosphopeptide data[93]	[94]
PostPLSR	Snapshot expression data; multi-omics data	Partial least squares regression	Studying the effects of circadian rhythm on gene expression[95]	[95]
PANDA	Snapshot expression data; prior knowledge	Message passing	Construction of sex-specific networks in chronic obstructive pulmonary disease[96]	[97]
Lemon-Tree	Snapshot expression data; multi-omics data	Bayesian network	Identifying changes in gene interactions during spontaneous migraine attack[98]	[99]
MERLIN+Prior	Snapshot expression data; prior knowledge	Bayesian network	Infer a general neural lineage network in early hindbrain and spinal cord development[100]	[101]
GRACE	Snapshot expression data; prior knowledge	Regression trees; Markov random fields	Construction of developmental GRNs in Arabidopsis thaliana and Drosophila melanogaster[102]	[102]
KiMONO	Snapshot expression data; prior knowledge	Penalized regression	Identification of host factors for COVID-19[103]	[104]
Inferelator-AMuSR	Snapshot expression data; multiple conditions	Penalized regression	Inferring shared and cell-type specific interactions in cortical interneuron development[105]	[106]
fused LASSO	Snapshot expression data; multiple conditions	Penalized regression	Construction of GRNs in early immune response to Avian influenza A viruses (IAV)[107]	[108]
GRADIS	Snapshot expression data; known TF-target interactions	Support vector machines	Evaluated on DREAM4 and DREAM5 datasets[68]	[68]
DeepSEM	Single-cell RNA sequencing (scRNA-seq) data	Deep learning	Identification of marker genes in the mouse cortex[72]	[72]
DeepDRIM	scRNA-seq data	Deep learning	Identifying genes relevant to COVID-19[70]	[70]
CNNC	scRNA-seq data	Deep learning	Evaluated on mouse embryonic stem cells, bone marrow-derived macrophages, and dendritic cells to predict new TF-target relationships[69]	[69]
GraphReg	DNA sequence data; Chromatin interaction data	Deep learning	Evaluated on ENCODE data from human cell lines and data from mouse embryonic stem cells to predict CAGE expression signals[109]	[109]
Enformer	DNA sequence data	Deep learning	Variant effect prediction[110]	[110]
Basenji	DNA sequence data	Deep learning	Identifying causal eQTLs[111]	[112]

the network into local networks (LBNs). For each LBN, the optimal structure is determined by searching through all possible configurations. A candidate GRN is assembled by integrating all the LBNs, followed by pruning of the edges using CMI. This procedure is iterated and the networks are refined to obtain the final GRN [65]. For further reading on causal structure learning, we refer the reader to [66,67].

Supervised machine learning approaches have also been developed for GRN inference. These require both gene expression data and a set of known regulatory interactions for predicting regulatory links between gene pairs. GRADIS first clusters samples and generates graph distance profiles from the gene expression data as features for training support vector machines to reconstruct the GRN [68]. Deep learning-based approaches have also been applied to the network inference problem using convolutional neural networks (CNNs) or variational graph autoencoders (VAEs). The convolutional neural network for coexpression (CNNC) approach uses known TF-target interactions in a supervised framework. CNNC transforms the joint expression of a gene pair into an image-like matrix, which can then be used for training a CNN [69]. DeepDRIM (deep learning-based direct regulatory interaction model) extends CNNC by considering both a target gene pair (primary image) and the gene pairs where one of the genes shows strong positive covariance to a gene that belongs to the known TF-target pairs (neighbor images). Both the primary and neighbor images are then used as inputs for training a CNN model. By incorporating neighborhood context, CNNC aims to remove false positives due to transitive interactions [70]. Recently, the Non-combinatorial Optimization via Trace Exponential and Augmented lagRangian for Structure learning (NOTEARS) algorithm has been developed, which is the first study to transform the structure learning problem from a combinatorial optimization into a continuous optimization problem as score-based learning of Bayesian networks from data [71]. This allows the use of well established numerical solvers to derive the DAG. The DAG-GNN approach adapts the NOTEARS algorithm by using a deep generative model under a VAE framework to model non-linear causal relationships. Based on the DAG-GNN formulation, DeepSEM was developed to infer GRNs from single-cell sequencing data [72]. DeepSEM implements a beta VAE and uses the weights of both the encoder and decoder functions to represent the adjacency matrix of the GRN.

3.2. Network inference from time-series data

Diseases can arise as the result of accumulated perturbations in regulatory processes that progress over time. The dynamic nature of biological phenomena can be investigated by taking "snapshots" of the system, i.e. measurements across multiple time points (e.g. through microarray or RNA-Seq experiments). The network inference methods discussed above assume static networks, which hampers the discovery of causal relationships and does not capture the temporal changes in network structure. Modeling dynamic GRNs allows for a more fine-grained investigation of molecular mechanisms and important regulators that govern these network transitions and drive topological changes over time. Various techniques have been implemented for modeling the time evolution of gene regulation, such as ODEs, Boolean networks, and dynamic Bayesian networks (DBNs).

Boolean networks describe regulatory interactions over time using Boolean variables as the states of genes and Boolean functions to determine the current gene states as a function of the previous states of other genes [113]. In a Boolean network, the expression of each gene is discretized to only "on" (expressed) or "off" (not expressed). Inferring a Boolean network corresponds to learning the Boolean functions that make up the regulatory rules. The And/Or tree ensemble (ATEN) algorithm infers a GRN from ensembles of and/or trees [82]. The MIBNI algorithm uses an MI-based feature selection step to identify potential regulators. This is followed by a subroutine that performs swapping to find optimal regulators for each target gene [114]. GABNI extends the MIBNI algorithm by further implementing a genetic algorithm to improve the set of candidate regulators and thus the dynamic consistency of the resulting Boolean network. An evaluation of Boolean methods for GRN inference and their implementations can be found in [115].

A DBN is a probabilistic graphical model for the joint distribution of random variables with observations across different time points [116]. DBNs are a widely used technique for modeling GRNs using time-series data. In contrast to traditional BNs, DBNs can handle self-loops, which are abundant in biological systems. DBNs can be learned from time-series data using the methods in BN literature. Learning DBN structures remains computationally challenging because of the extremely large search space. Various greedy and heuristic methods have been developed to restrict the network topologies. However, efficient modeling with DBNs is often limited to up to hundreds of nodes. The simplest DBN uses a first-order Markov model, where the value of a variable at the current time point is conditionally dependent on the values of a set of parent variables at the previous time point only. Classical homogeneous DBN models assume that the network topology and parameters characterizing the network are constant over time. BiDAG can also infer the DBN from time-series data [81]. On the other hand, non-homogeneous DBNs provide flexibility by allowing the network parameters and/or structure to vary.

ODEs describe the rate of change of a gene's expression levels as a function of its regulators. ODE models are a widely applied method to quantitatively describe the time evolution and regulatory dynamics of the gene expression levels. dynGENIE3 extends GENIE3 by modeling expression changes over time in an ODE. These are used as the inputs to the random forest and putative gene–gene interactions are next learned using random forest regression. A similar method uses nonlinear ODEs and gradient boosting trees [52] to infer the networks [53]. Using both snapshot and time-series data, Inferelator combines regularized regression and ODEs to model the time evolution of genes or pre-calculated gene clusters [88]. ODEs provide highly detailed models but require extensive parameterization and are computationally infeasible for genome-scale networks.

Other models that infer networks from time-series data build on Granger causality, which is based on the reasoning that if incorporating previous values of Y increases the prediction accuracy of X in an autoregression model, then Y is said to Granger-cause X. A causal relationship is then determined from the statistically significant coefficients of the model. BETS implements an elastic net autoregression model [91]. To increase robustness, the networks are inferred from bootstrapped samples from the data and stability selection is performed to identify the most frequent edges. This is combined with another bootstrapping procedure to control the false discovery rate. SWING improves on the standard pairwise Granger causality framework to simultaneously allow for multiple explanatory regulators by using user-specified sliding windows on temporally spaced observations and delay parameters to generate ensembles of training data. Potential predictors are then evaluated using supervised learning algorithms. A consensus network representing the final prediction is then aggregated from the high-confidence edges from the different windowed models [92].

The GRN inference tools discussed above typically require input from a single source or data type to reconstruct GRNs *de novo* and thus rely solely on the statistical properties of the data. In particular, most popular GRN inference tools utilize only gene expression data from an experiment as input.
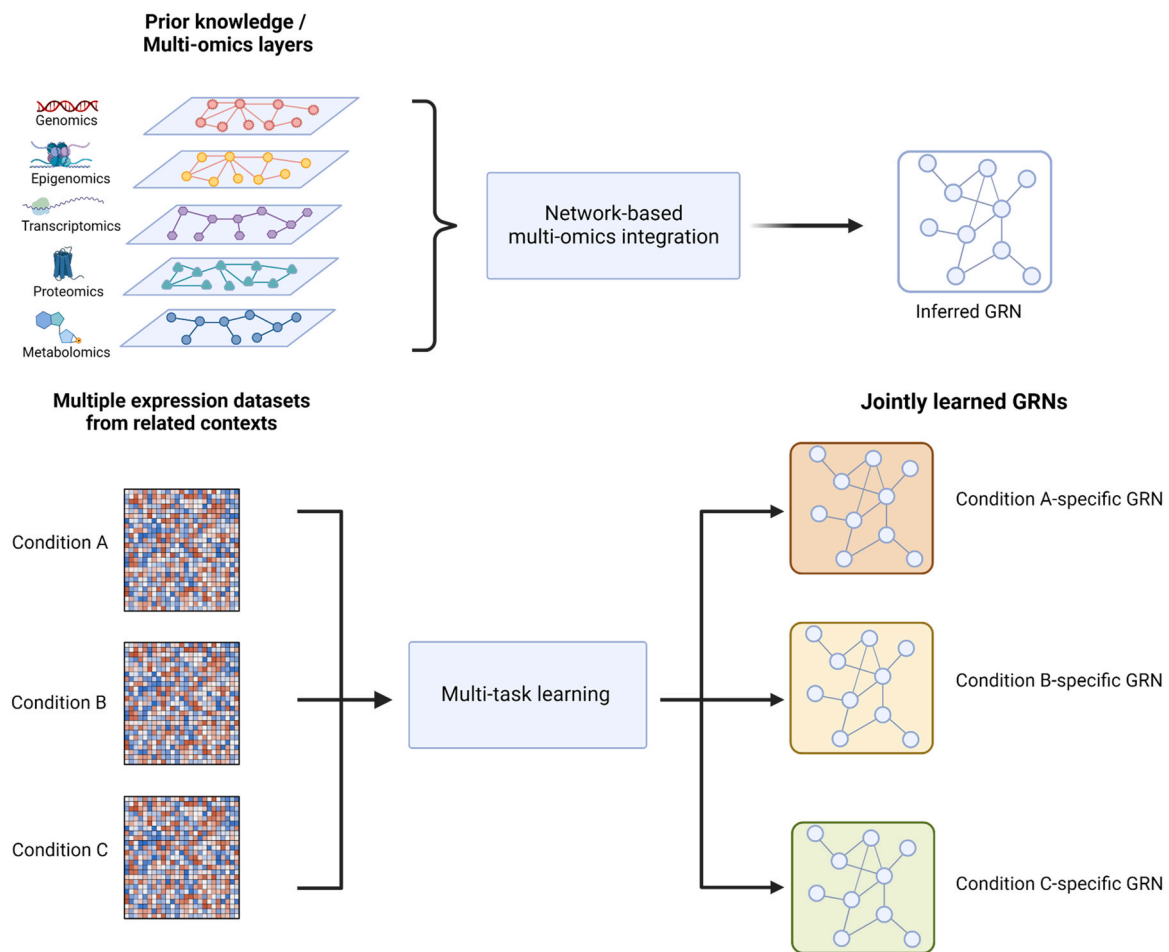
Fig. 2. Multi-modal and integrative approaches towards network inference. Depending on the integration strategy, network-based integration methods that incorporate prior information from biomedical databases or multi-omics data can inform biologically meaningful GRN reconstruction. Other approaches exploit the similarity of data from multiple related biological conditions in multi-task learning to simultaneously infer GRNs. Created with BioRender.com.

3.3. Multi-modal and integrative network inference approaches

Recent studies are increasingly integrative and incorporate prior biological information to facilitate more biologically meaningful GRN structures (Fig. 2). Combining data from multiple modalities in a principled manner can enhance our understanding of a disease [117,118]. Several other types of high-throughput genomics, proteomics, and metabolomics data are now being profiled over time, with studies profiling several omics levels simultaneously. Prior biological information can also be obtained from existing biomedical databases, including TF-target interactions and binding motifs, partially known networks from experimental data, DNase hypersensitivity data, ChIP-seq data, and GWAS. Therefore, the use of prior knowledge and biologically plausible assumptions with respect to the model structure can inform GRN reconstruction [119]. Various strategies that provide flexibility by allowing simultaneous analysis of multi-omics data or incorporating available prior knowledge are particularly desirable. For instance, the similarity network fusion (SNF) approach constructs a patient similarity network for each omics data type, where an edge indicates high pairwise similarity between individuals. The multiple networks are then combined into a fused network using an iterative message passing algorithm that updates the networks until they are harmonized [120]. The tool iR-afNet extends the GENIE3 regression tree approach by incorporating weighted scores from different evidence layers such as knockout expression data to influence the selection of potential regulators for a gene [94]. The PostPLSR approach uses variable importance derived from partial least squares regression from time-course gene expression data and partially known relationships from pathway databases to reconstruct GRNs [95]. PANDA implements a message passing algorithm that uses available TF-motif or known protein interactions to refine GRN reconstruction [97,121]. MERLIN+Prior constructs a prior network obtained by integrating non-expression data, including motif, knockout, and epigenomics data. The algorithm then learns a modular GRN using a Bayesian framework [101]. GRACE generates a meta-regulatory network from different biologically supported sources and filters for high-confidence edges using ensembles of Markov random fields [102]. Using a penalized regression approach, KiMONO can infer multi-level networks from any combination of omics types [104]. Lemon-Tree is a module network inference method [122] that employs a model-based Gibbs sampler [97,120] to infer co-expression modules and subsequently finds consensus modules using a spectral clustering algorithm. It then reconstructs regulatory programs by assigning a set of regulator genes to the network modules as learned through decision trees. A probabilistic score is calculated for each resulting regulatory program, incorporating other omics data, and high-scoring programs are considered relevant to the disease of interest [99].

Previous studies have shown considerable similarity in GRNs across tissues or conditions, with relatively smaller changes in local network structure driving differences in phenotypes or responses. Consequently, various methods have been developed to leverage experimental data from separate studies to improve GRN reconstruction, exploiting shared regulatory interactions across

different but related biological contexts. Inferelator-AMuSR implements multi-task learning to jointly infer the GRN from multiple experiments [106]. It enforces similarity across network models using an adaptive multiple regression technique. The fused LASSO approach infers GRNs from multiple time-series or perturbation conditions and employs regularization to promote the sparsity of the coefficients of the regression predicted TF-target relationships from multiple networks constructed from each input dataset. In addition, it enforces similarity of the GRNs reconstructed from multiple experiments [108].

4. Modeling gene regulation from sequence data

Deep learning approaches have also been developed for predicting gene regulation using DNA sequence data as input. For example, Basenji applies CNNs on sequence data to predict the effects of variants on gene expression and disease risk [112]. Newer architectures are also used, including graph neural networks (GNNs) that can handle graph-structured data and have emerged as part of standard analysis toolbox in various graph-related tasks, including node classification, link prediction, and graph classification [123,124]. GNNs work via a message passing algorithm to allow nodes in a network to exchange information. Graph attention networks are GNNs with attention, a mechanism which extracts useful context from long-range structural information [125,126]. In the field of machine translation and natural language processing, the addition of attention layers to deep learning models significantly improved performance [127,128]. In gene regulation, attention layers have been used to incorporate information from distal regulatory elements which has not been achieved using CNNs alone because of their limited local receptive field. For example, Enformer uses a transformer architecture which employs self-attention, thereby accurately inferring long-range enhancer-promoter interactions from up to 100 kb away from the transcription start site [110]. GraphReg predicts gene expression levels measured by Cap Analysis Gene Expression (CAGE)-seq by using CNNs to learn local representations from one-dimensional data (e.g. epigenomic data and/or sequence data) [109]. Afterwards, a graph attention network is applied to propagate the learned features over an interaction graph built using 3D chromatin data and can thus capture more distant regulatory features.

5. Differential co-expression networks

Differential co-expression (DC) represents changes in interactions between molecular components, also known as "network rewiring." DC can occur in response to cellular stimuli, developmental stages, drug treatments, or disease states and can also occur in the absence of differential expression [129–131]. Analyzing changes in the network structure of regulatory relationships can provide more context-specific insights into the mechanisms underlying diseases. Under the premise that co-expressed genes are more likely to be co-regulated, major changes in co-expression patterns between classes may indicate changes in regulation [132,133]. In a differential network, an edge thus represents a significant change in correlation or association strength in a pair of genes. Differential co-expression methods examine the differences in the network connectivity patterns by comparing two or more phenotypes or biological states (e.g., diseased vs. healthy individuals). This is different from disease module identification, where most methods focus on the individual genes (nodes) and typically use differentially expressed genes as starting points (seed genes). Differential co-expression analysis has been used to complement standard DE analysis [134].

A number of methods have been proposed to detect differential co-expression networks from data *de novo* (Table 3). Most methods use correlation as a measure of association, although others use

other metrics or tests, such as information theory, GGMs, linear models, or generalized linear models. Some methods implement a network-based approach, which aims to identify pairwise changes in the strength of co-expression between genes. DGCA calculates differential correlation by z-score transformation of Pearson correlation coefficient (PCC) [77]. Information theoretic approaches include MINDy, which uses MI as a measure, and CINDy, which is based on conditional mutual information [135]. Under the GGM framework, Danaher et al. proposed the joint graphical lasso, which employs fused lasso or group lasso penalties to jointly estimate the precision matrices from multiple classes. Assuming that the two conditions have shared characteristics, this method encourages similarity between the graphical models [136]. With the key assumption that the resulting differential network is sparse, the method of Yuan et al. directly estimates the difference of the precision matrices using a penalized loss function called the D-trace loss [137].

Other methods implement module-based approaches and thus focus on finding groups of differentially co-expressed genes, also known as DC modules. DICER calculates a probabilistic differential co-expression score called the T-score, which is based on the Pearson correlation. The T-score is then used to identify DC clusters (a differentially co-expressed gene set) or meta-modules (pairs of gene sets with DC between phenotypes) [138]. Diffcoex first constructs a differential adjacency matrix based on correlation. Afterwards, the topological overlap measure [139] is used to derive a dissimilarity matrix prior to applying WGCNA clustering to identify the DC modules [140]. ALtered Partitions Across Community Architectures (ALPACA) is another module-based method that extracts differential subnetworks by optimizing a measure called "differential modularity," which captures changes in network structure by comparing the density of edges in the diseased network relative to a control network [141].

A benchmarking study evaluated the performance of various differential network methods [142], indicating the need for further algorithmic improvements in handling more complex changes in co-expression structure. In addition, there is a need for the development of new methods that can better handle non-Gaussian data [143], which is abundant in biological systems.
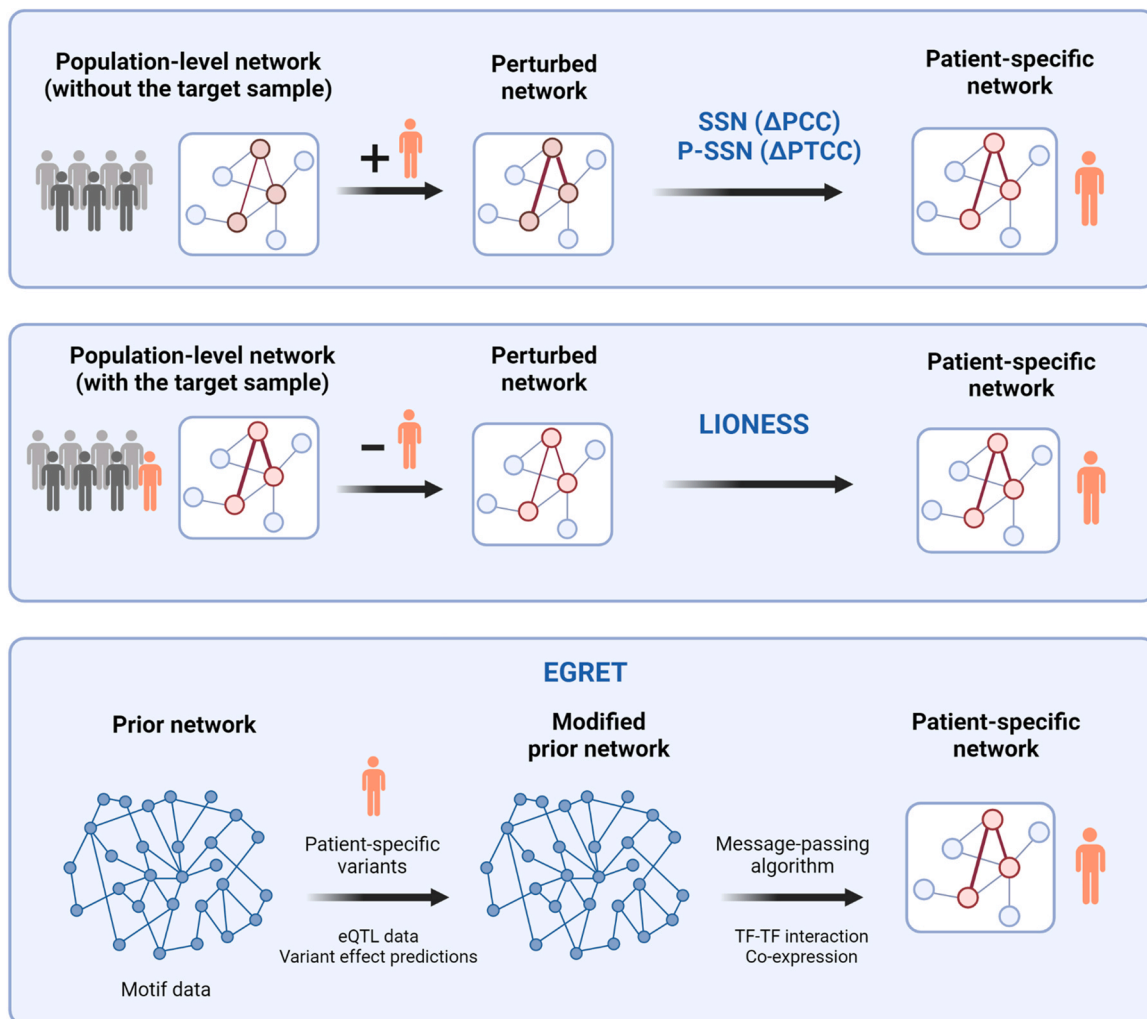
6. Sample-specific networks

The network inference methods discussed above assume an "aggregate" network representing a phenotype of interest and are mostly aimed to discover shared disease mechanisms in a given population or dataset. As such, they do not explicitly account for patient heterogeneity. In application cases related to personalized medicine, it may be desirable to construct networks at the single-sample level, which can aid in disease characterization and development of therapies tailored for a patient or subgroup of patients. Several methods have been proposed to derive patient-specific networks from gene expression data. Representative methods include LIONESS, SSN, and P-SSN (Fig. 3). Using a perturbation strategy, dysregulated edges in the GRN that are unique to an individual can be inferred by comparison to the genotype-agnostic baseline network. The main idea behind perturbation-based inference of sample-specific networks is to compare two networks, namely, a reference network and a perturbed network. The reference network is inferred from multiple samples in the population. Depending on the method, the perturbed network is constructed by adding or removing the sample of interest. The difference between the two networks is then used to derive the sample-specific network representing the characteristics of the individual patient. The single-sample network (SSN) method calculates the change in the pairwise Pearson's correlation coefficient values upon the addition of the sample [152]. To filter for direct interactions, the P-SSN method uses a partial correlation-based measure [153]. Linear Interpolation to

Table 3
Representative tools for deriving differential co-expression networks. Links to code/tools are provided in Supplementary Table 3.

Tool	Input data	Algorithmic approach	Example application	Reference
DGCA	Two conditions	Correlation	Differential connectivity analysis in myelin dysregulation in a mouse model of Alzheimer's disease[144]	[145]
MINDY	Two conditions	Information theory	Identify modulators of B cell signaling[146]	[147]
CINDy	Two conditions	Information theory	Infer candidate upstream modulators of master regulator proteins in various cancer states[148]	[135]
Joint Graphical LASSO	Multiple conditions	Gaussian graphical model	Network analysis of gut microbiome data related to pediatric obesity[149]	[136]
Yuan et al.	Multiple conditions	Gaussian graphical model	Identification of genes involved in microsatellite stable colorectal cancers[137]	[137]
DICER	Multiple conditions	Correlation	Identification of differentially correlated gene clusters in Alzheimer's disease[138]	[138,142]
ALPACA	Two conditions	Differential modularity	Identification of network modules associated with glioblastoma survival[150]	[141]
Diffcoex	Multiple conditions	Correlation	Investigating altered co-expression patterns following influenza virus infection[151]	[140,142]

Obtain Network Estimates for Single Samples (LIONESS) models the weights of the aggregate network as a linear combination of the weights of the individual sample networks. The change in the weight of an edge is attributed to the sample that was removed in generating the perturbed network. In other words, the sample-level network is considered the contribution of the individual sample to the reference network. The LIONESS approach is a framework to be used in tandem with standard network inference methods [154]. Estimating the Genetic Regulatory Effect on TFs (EGRET) is a multi-omics approach that is based on refinement of a prior network using a message-passing algorithm [155]. EGRET first constructs a prior network using motif data. A modified prior network is then derived by incorporating patient-specific genotype data, eQTLs, and variant effect predictions. Next, a message-passing algorithm is applied with

Fig. 3. General principles behind methods for inferring sample-specific networks. SSN, P-SSN and LIONESS apply perturbation analysis upon the addition or removal of a particular sample from the aggregate or reference network. SSN and P-SSN estimate the sample-level networks based on correlation measures. LIONESS uses linear interpolation to derive the sample-level networks but can also be extended with alternative measures. EGRET uses a message-passing algorithm to refine a prior network using patient-specific data. ΔPCC = differential Pearson's correlation; ΔPTCC = differential partial correlation; eQTL = expression quantitative trait loci; TF = Transcription factor. Created with BioRender.com.

Table 4
Representative tools for deriving sample-specific networks. Links to code/tools are provided in Supplementary Table 4.

Tool	Input data	Algorithmic approach	Example application	Reference
SSN	Expression data	Network perturbation	Construction of multicellular gene network to investigating the prognostic role of macrophage–tumor cell interactions in tumor progression in gliomas[156]	[152]
P-SSN	Expression data	Network perturbation	Identify individual-specific driver genes in cancer patients using the TCGA dataset[153]	[153]
LIONESS	Expression data	Network perturbation	Infer sample-specific GRNs from 29 healthy human tissue types to study sex differences in gene expression and regulatory networks[157]	[154]
EGRET	Snapshot expression data; multi-omics data	Message passing	Construction of genotype-specific GRNs in GM12878 and K562 cell lines; construction of patient-specific networks in Crohn's disease and coronary artery disease[155]	[155]

co-expression and TF cooperativity data to obtain the final patient-specific network. (((Table 4))).

7. Summary and Outlook

Biological networks provide a fundamental basis for modeling the complex and dynamic nature of biological processes. The increasing availability of extensive biomedical big data has significantly improved our ability to quantify and characterize biological networks. Consequently, network theory and related concepts have become indispensable for exploring the etiology of human diseases [158] and revealing complex patterns of gene expression regulation. In particular, network enrichment and network inference have become standard methods for analyzing omics data and have significantly improved our understanding of disease development and progression.

Despite the extensive use of DNE methods, they have several shortcomings. Lazareva et al. developed a test suite to evaluate the robustness of DNE methods against network randomization. They showed that many DNE methods mostly learn from the node degree distributions rather than from the encoded biological knowledge in the edges of PPI networks [159]. By assessing the gene ontology enrichment of DNE methods on randomly permuted input data, Levi et al. questioned the context-specificity of the existing methods [24]. Since diseases with higher morbidity or mortality and their associated proteins are studied more extensively, PPI networks are subject to study bias [160]. This study bias and the fact that a large fraction of PPIs are still not discovered have been found to significantly distort networks and consequently methods used in network-based techniques that rely on PPIs [161]. The full capacity of modern computational algorithms to predict protein structures and PPIs can be used to augment our currently limited knowledge of PPIs derived from experimental settings. For instance, the protein structures output by Alphafold can be further used to predict PPIs to reduce bias and complement existing PPI networks [162].

Network inference is a classical problem in computational biology that aims to reconstruct the set of gene-gene relationships that are dominant in a phenotype of interest or how they change through time. Despite considerable progress in the past decades, network inference is far from being fully resolved and remains an active area of research, as evidenced by ongoing efforts such as the Dialogue for Reverse Engineering Assessments and Methods (DREAM) project [163]. Previous benchmarking studies have shown varying results and that no single algorithmic class consistently outperforms all other methods [164,165]. The most suitable algorithmic approach may depend on various factors, such as the biological context being examined and the specific characteristics of the dataset. Furthermore, combining multiple network inference methods was generally found to perform better than using a single approach [46], highlighting an advantage of ensemble methods and multi-omics integration.

Earlier work on network inference focused on snapshot expression data. While these have undoubtedly advanced our knowledge on complex diseases, they provide a limited view of the underlying molecular processes. The next crucial step to understanding disease warrants new computational approaches that can more effectively incorporate and exploit the dynamic nature of biological phenomena. Time-series experimental data offer rich latent information that require more sophisticated network-based computational tools. However, recovering the network structures from temporal data at the genome scale is faced with several practical limitations. The still relatively high level of noise of current omics platforms combined with low numbers of samples and the sparsity of measured time points remain a major bottleneck. The complexity of the networks may be highly variable across phenotypes and inference approaches typically require a high number of parameters relative to the small

number of available samples. In addition, the appropriate time scales for measuring disease progression are unknown or very difficult to determine in many cases. Thus, new and more efficient computational strategies are needed. As the quantity, quality, and types of temporal omics data continue to increase, we expect that network inference methods would play more important roles to accurately model disease trajectories and discover new disease drivers.

The network inference tools we described were focused on bulk transcriptomic data. Along with the development and increasing popularity of single-cell technologies, various network inference methods have been developed for single-cell transcriptomics data. Compared to bulk data, which characterizes the expression profile of an average of the cells in a sample, single-cell technology can profile up to tens of millions of individual cells and thus capture a more detailed gene regulatory landscape. However, single-cell data pose additional challenges, such as noise and zero-inflation, and thus require more tailored analysis steps. Methods for dynamic network inference from single-cell data are also an active area of research and have been especially useful for reconstructing temporal ordering of cells in multiple developmental processes. For in-depth discussion and benchmarking of network inference tools dedicated to single-cell data, we refer the reader to focused benchmarks and reviews [166–169].

Multi-omics integration also represents a new challenge in the development of flexible tools that will enable researchers to more easily combine and customize their analysis or create workflows according to the nature of the biological process under study. Incorporating existing knowledge and diverse sources of information, including GWAS, molecular interaction databases (PPI, GRNs), pathway databases and literature knowledge, can further provide a more comprehensive picture of the determinants and drivers of complex diseases [170,171]. There are now more concerted efforts to simultaneously profile several omics types using the same patient groups/samples, such as The Cancer Genome Atlas (TGCA) [172], Genotype-Tissue Expression (GTEx) [173], and ENCODE [174]. While most existing methods focus exclusively on TFs as regulators, the importance of other regulators such as microRNAs in GRN inference is increasingly recognized [175]. Nevertheless, multi-omics analysis remains a relatively underexplored area of research, considering the extensive amount of data being generated [171,176]. Thus, we expect the development of more integrative approaches at different levels of organization, which can be much more powerful in capturing disease complexity and interplay between molecular layers.

Deep learning-based methods have also gained popularity in recent years because of their capability to handle large amounts of biological data and to achieve state-of-the-art performance in many prediction tasks in genomics. One limitation of the application of deep learning models in biology and medicine is the difficulty of obtaining mechanistic explanations behind their predictions. Nevertheless, interpretability of deep neural networks is a highly active area of research, and efforts are ongoing to systematically understand the regulatory grammar that governs gene expression regulation in various contexts [177–179].

The lack of gold standard datasets for the validation of methods in the field of computational biology is a well-known problem. Conventional validation is performed using proxy measures, such as gene set over-representation analyses against known disease genes from DisGeNET [180] or disease-associated pathways from databases such as KEGG [181], Reactome [182], and WikiPathways [183]. Competitions such as the DREAM challenges have become instrumental in harnessing the "wisdom of the crowds" to answer various computational problems in biomedicine, including challenges that address the network inference problem [184]. Importantly, they provide the community with high-quality datasets for benchmarking, including both simulated steady-state and time-series gene expression data, which can help researchers better understand

mechanisms behind disease trajectories. These reference datasets have also been extensively used for systematic evaluation of newly developed algorithms against state-of-the-art algorithms and have helped avoid the "self-assessment trap" [185]. Nevertheless, our current knowledge of disease mechanisms is far from complete [159]. We can additionally improve benchmarking of GRN methods using knowledge gained from unbiased experimental methods, such as CRISPR-based perturbation screens [186–188] that allow precise manipulation of genomic sequences. Results from such studies could provide the scientific community with more context-specific methods of elucidating gene regulation [189].

Another bottleneck is the lack of mechanistic disease definitions. Diseases are largely based on phenotypes (manifested symptoms) and the clinical definitions are generally not readily harmonized with new molecular findings [14]. This can result in symptom-wise homogeneous but molecularly and mechanistically heterogeneous patient cohorts. Thus, the mechanistic profiles extracted by network-based methods, particularly the supervised ones, may not be very reliable, since the input data based on disease annotation is flawed. The databases employed for validation also use ill-defined disease terms, further propagating the problem to the validation of computational approaches.

Some of the tools reviewed here have been successfully applied in diverse disease contexts, such as diabetes, coronary heart diseases, and lung diseases [158]. Further improvements in theoretical development, tool usability, and multi-omics or multi-study integration show great promise for obtaining a more comprehensive understanding of complex diseases. We envision that the combination of enrichment, differential networks, and more focus on inference tools from time-series experiments into comprehensive and holistic analysis pipelines will help capture more detailed molecular mechanisms underlying diseases. The integration of several different types of multi-omics time-series and snapshot data sets will be instrumental to reconstruct accurate GRN models and thus obtain a more complete picture of the dynamics and complexity of disease progression. Another emerging area of research is the estimation of sample-specific gene networks. These are important for investigating heterogeneity among individuals and characterizing their specific molecular alterations or disease states. Consequently, advances in this field will enable the identification of more suitable therapeutic targets or drug combinations, representing a significant step towards precision medicine [190]. Beyond GRNs, network inference methods can also be applied in related fields such as microbiome research where networks of co-occurring taxa are already constructed borrowing concepts originally proposed for GRN inference [191].

We provided an overview of the main types of network-based methods that can be applied to study the molecular basis of diseases, their core algorithmic concepts, and some representative tools for analysis. Each type of analysis captures different aspects of the biological system. The next crucial step to understanding disease warrants new computational strategies that can more effectively exploit the dynamic nature of biological phenomena. Developing more flexible methods can fill remaining gaps and pave the way towards more accurate identification of disease mechanisms and effective drug treatments.

Funding

Research (BMBF) within the framework of the e:Med research and funding concept (grant 01ZX1908A) (SS, JB and ML).

CRediT authorship contribution statement

Gihanna Galindez: Investigation, Writing – original draft, Visualization, Conceptualization. **Sepideh Sadegh:** Investigation, Writing – original draft, Visualization, Conceptualization. **Jan Baumbach:** Writing – review & editing, Conceptualization. **Tim Kacprowski:** Writing – review & editing, Conceptualization, Supervision. **Markus List:** Writing – review & editing, Conceptualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2022.12.022.

References

[1] Oughtred R, Stark C, Breitkreutz B-J, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung G, McAdam R, Zhang F, Dolma S, Willems A, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M. The BioGRID interaction database: 2019 update. Nucleic Acids Res 2019;47:D529–41.

[2] Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res 2020;49:D605–12.

[3] Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W. Learning from co-expression networks: possibilities and challenges. Front Plant Sci 2016;7:444.

[4] Watson E, Yilmaz LS, Walhout AJM. Understanding metabolic regulation at a systems level: metabolite sensing, mathematical predictions, and model organisms. Annu Rev Genet 2015;49:553–75.

[5] Papin JA, Hunter T, Palsson BO, Subramaniam S. Reconstruction of cellular signalling networks and analysis of their properties. Nat Rev Mol Cell Biol 2005;6:99–111.

[6] Lun X-K, Bodenmiller B. Profiling cell signaling networks at single-cell resolution. Mol Cell Proteomics 2020;19:744–56.

[7] Emmert-Streib F, Dehmer M, Haibe-Kains B. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. Front Cell Dev Biol 2014;2:38.

[8] Rintala TJ, Ghosh A, Fortino V. Network approaches for modeling the effect of drugs and diseases. Brief Bioinform 2022;23. https://doi.org/10.1093/bib/bbac229

[9] Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011;12:56–68.

[10] Sharma A, Menche J, Huang CC, Ort T, Zhou X, Kitsak M, Sahni N, Thibault D, Voung L, Guo F, Ghiassian SD, Gulbahce N, Baribaud F, Tocker J, Dobrin R, Barnathan E, Liu H, Panettieri jr RA, Tantisira KG, Qiu W, Raby BA, Silverman EK, Vidal M, Weiss ST, Barabási A-L. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. Hum Mol Genet 2015;24:3005–20.

[11] Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L. The human disease network. Proc Natl Acad Sci 2007;104:8685–90.

[12] Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabasi A-L. Uncovering disease-disease relationships through the incomplete interactome. 1257601–1257601 Science 2015;347. https://doi.org/10.1126/science.1257601

[13] Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol 2008;4:682–90.

[14] Nogales C, Mamdouh ZM, List M, Kiel C, Casas AI, Schmidt HHHW. Network pharmacology: curing causal mechanisms instead of treating symptoms. Trends Pharmacol Sci 2022;43:136–50.

[15] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;102:15545–50.

[16] Geistlinger L, Csaba G, Santarelli M, Ramos M, Schiffer L, Turaga N, Law C, Davis S, Carey V, Morgan M, Zimmer R, Waldron L. Toward a gold standard for benchmarking gene set enrichment analysis. Brief Bioinform 2020. https://doi.org/10.1093/bib/bbz158

[17] Batra R, Alcaraz N, Gitzhofer K, Pauling J, Ditzel HJ, Hellmuth M, Baumbach J, List M. On the performance of de novo pathway enrichment. NPJ Syst Biol Appl 2017;3:6.

[18] Liu Y, Brossard M, Roqueiro D, Margaritte-Jeannin P, Sarnowski C, Bouzigon E, Demenais F. SigMod: an exact and efficient method to identify a strongly

[19] interconnected disease-associated module in a gene network. Bioinformatics 2017;33:1536–44.

[19] S.M. Inavolu, J. Renbarger, M. Radovich, IODNE: An integrated optimization method for identifying the deregulated subnetwork for precision medicine in cancer CPT 2017.⟨https://ascpt.onlinelibrary.wiley.com/doi/abs/10.1002/psp4.12167%4010.1002/%28ISSN%292163-8306.Cancer⟩.

[20] Akhmedov M, Kedaigle A, Chong RE, Montemanni R, Bertoni F, Fraenkel E, Kwee I. PCSF: an R-package for network-based interpretation of high-throughput data. PLoS Comput Biol 2017;13:e1005694.

[21] Tuncbag N, Gosline SJC, Kedaigle A, Soltis AR, Gitter A, Fraenkel E. Network-based interpretation of diverse high-throughput datasets through the omics integrator software package. PLoS Comput Biol 2016;12:e1004879https://doi.org/10.1371/journal.pcbi.1004879

[22] Sadegh S, Skelton J, Anastasi E, Bernett J, Blumenthal DB, Galindez G, Salgado-Albarrán M, Lazareva O, Flanagan K, Cockell S, Nogales C, Casas AI, Schmidt HHHW, Baumbach J, Wipat A, Kacprowski T. Network medicine for disease module identification and drug repurposing with the NeDRex platform. Nat Commun 2021;12:6848.

[23] Bernett J, Krupke D, Sadegh S, Baumbach J, Fekete SP, Kacprowski T, List M, Blumenthal DB. Robust disease module mining via enumeration of diverse prize-collecting Steiner trees. Bioinformatics 2022. https://doi.org/10.1093/bioinformatics/btab876

[24] Levi H, Elkon R, Shamir R. DOMINO: a novel algorithm for network-based identification of active modules with reduced rate of false calls. Mol Syst Biol 2021;17. https://doi.org/10.1101/2020.03.10.984963

[25] Alcaraz N, List M, Dissing-Hansen M, Rehmsmeier M, Tan Q, Mollenhauer J, Ditzel HJ, Baumbach J. Robust de novo pathway enrichment with KeyPathwayMiner 5. F1000Res 2016;5:1531.

[26] Vlaic S, Conrad T, Tokarski-Schnelle C, Gustafsson M, Dahmen U, Guthke R, Schuster S. Modulediscoverer: Identification of regulatory modules in protein-protein interaction networks. Sci Rep 2018;8. https://doi.org/10.1038/s41598-017-18370-2

[27] Altieri F, Hansen TV, Vandin F. NoMAS: a computational approach to find mutated subnetworks associated with survival in genome-wide cancer studies. Front Genet 2019;10:265.

[28] Hristov BH, Singh M. Network-based coverage of mutational profiles reveals cancer genes. Cell Syst 2017;5:221–9. e4.

[29] da Rocha EL, Ung CY, McGehee CD, Correia C, Li H. NetDecoder: a network biology platform that decodes context-specific biological networks and gene activities. Nucleic Acids Res 2016;44:e100.

[30] Leiserson MDM, Vandin F, Wu H-T, Dobson JR, Eldridge JV, Thomas JL, Papoutsaki A, Kim Y, Niu B, McLellan M, Lawrence MS, Gonzalez-Perez A, Tamborero D, Cheng Y, Ryslik GA, Lopez-Bigas N, Getz G, Ding L, Raphael BJ. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nat Genet 2015;47:106–14.

[31] Reyna MA, Leiserson MDM, Raphael BJ. Hierarchical HotNet: identifying hierarchies of altered subnetworks. Bioinformatics 2018;34:i972–80.

[32] Larsen SJ, Schmidt HHHW, Baumbach J. De Novo and supervised endophenotyping using network-guided ensemble learning. Syst Med 2020;3:8–21.

[33] Wang T, Peng Q, Liu B, Liu Y, Wang Y. Disease Module Identification Based on Representation Learning of Complex Networks Integrated From GWAS, eQTL Summaries, and Human Interactome. Front Bioeng Biotechnol 2020;8:418.

[34] Lazareva O, Canzar S, Yuan K, Baumbach J, Blumenthal DB, Tieri P, Kacprowski T, List M. BiCoN: Network-constrained biclustering of patients and omics data. Bioinformatics 2020. https://doi.org/10.1093/bioinformatics/btaa1076

[35] Wiwie C, Kuznetsova I, Mostafa A, Rauch A, Haakonsson A, Barrio-Hernandez I, Blagoev B, Mandrup S, Schmidt HHHW, Pleschka S, Röttger R, Baumbach J. Time-resolved systems medicine reveals viral infection-modulating host targets. Syst Med 2019;2:1–9.

[36] Fitzgerald KC, Smith MD, Kim S, Sotirchos ES, Kornberg MD, Douglas M, Nourbakhsh B, Graves J, Rattan R, Poisson L, Cerghet M, Mowry EM, Waubant E, Giri S, Calabresi PA, Bhargava P. Multi-omic evaluation of metabolic alterations in multiple sclerosis identifies shifts in aromatic amino acid metabolism. Cell Rep Med 2021;2:100424.

[37] Khurana V, Peng J, Chung CY, Auluck PK, Fanning S, Tardiff DF, Bartels T, Koeva M, Eichhorn SW, Benyamini H, Lou Y, Nutter-Upham A, Baru V, Freyzon Y, Tuncbag N, Costanzo M, Luis B-JS, Schöndorf DC, Inmaculada Barrasa M, Ehsani S, Sanjana N, Zhong Q, Gasser T, Bartel DP, Vidal M, Deleidi M, Boone C, Fraenkel E, Berger B, Lindquist S. Genome-scale networks link neurodegenerative disease genes to α-synuclein through specific molecular pathways. e14 Cell Syst 2017;4:157–70. https://doi.org/10.1016/j.cels.2016.12.011

[38] Matschinske J, Salgado-Albarrán M, Sadegh S, Bongiovanni D, Baumbach J, Blumenthal DB. Individuating Possibly Repurposable Drugs and Drug Targets for COVID-19 Treatment Through Hypothesis-Driven Systems Medicine Using CoVex. Assay Drug Dev Technol 2020;18:348–55.

[39] Lio CT, Louadi Z, Fenn A, Baumbach J, Kacprowski T, List M, Tsoy O. Systematic analysis of alternative splicing in time course data using Spycone. bioRxiv 2022. https://doi.org/10.1101/2022.04.28.489857

[40] Salgado-Albarrán M, Navarro-Delgado EI, Del Moral-Morales A, Alcaraz N, Baumbach J, González-Barrios R, Soto-Reyes E. Comparative transcriptome analysis reveals key epigenetic targets in SARS-CoV-2 infection. NPJ Syst Biol Appl 2021;7:21.

[41] Nickel S, Vlaic S, Christ M, Schubert K, Henschler R, Tautenhahn F, Burger C, Kühne H, Erler S, Roth A, Wild C, Brach J, Hammad S, Gittel C, Baunack M, Lange U, Broschewitz J, Stock P, Metelmann I, Bartels M, Pietsch U-C, Krämer S, Eichfeld U, von Bergen M, Dooley S, Tautenhahn H-M, Christ B. Mesenchymal stromal cells mitigate liver damage after extended resection in the pig by modulating thrombospondin-1/TGF-β. NPJ Regen Med 2021;6:84.

[42] Monie DD, Correia C, Zhang C, Ung CY, Vile RG, Li H. Modular network mechanism of CCN1-associated resistance to HSV-1-derived oncolytic immunovirotherapies for glioblastomas. Sci Rep 2021;11:11198.

[43] Bailey P, Chang DK, Nones K, Johns AL, Patch A-M, Gingras M-C, Miller DK, Christ AN, Bruxner TJC, Quinn MC, Nourse C, Murtaugh LC, Harliwong I, Idrisoglu S, Manning S, Nourbakhsh E, Wani S, Fink L, Holmes O, Chin V, Anderson MJ, Kazakoff S, Leonard C, Newell F, Waddell N, Wood S, Xu Q, Wilson PJ, Cloonan N, Kassahn KS, Taylor D, Quek K, Robertson A, Pantano L, Mincarelli L, Sanchez LN, Evers L, Wu J, Pinese M, Cowley MJ, Jones MD, Colvin EK, Nagrial AM, Humphrey ES, Chantrill LA, Mawson A, Humphris J, Chou A, Pajic M, Scarlett CJ, Pinho EK, Giry-Laterriere M, Rooman I, Samra JS, Kench JG, Lovell JA, Merrett ND, Toon CW, Epari K, Nguyen NQ, Barbour A, Zeps N, Moran-Jones K, Jamieson NB, Graham JS, Duthie F, Oien K, Hair J, Grützmann R, Maitra A, Iacobuzio-Donahue CA, Wolfgang CL, Morgan RA, Lawlor RT, Corbo V, Bassi C, Rusev B, Capelli P, Salvia R, Tortora G, Mukhopadhyay D, Petersen GM, Australian Pancreatic Cancer Genome Initiative, Munzy DM, Fisher WE, Karim SA, Eshleman JR, Hruban RH, Pilarsky C, Morton JP, Sansom OJ, Scarpa A, Musgrove EA, Bailey U-MH, Hofmann O, Sutherland RL, Wheeler DA, Gill AJ, Gibbs RA, Pearson JV, Waddell N, Biankin AV, Grimmond SM. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 2016:47–52.

[44] Stukalov A, Girault V, Grass V, Karayel O, Bergant V, Urban C, Haas DA, Huang Y, Oubraham L, Wang A, Hamad MS, Piras A, Hansen FM, Tanzer MC, Paron I, Zinzula L, Engleitner T, Reinecke M, Lavacca TM, Ehmann R, Wölfel R, Jores J, Kuster B, Protzer U, Rad R, Ziebuhr J, Thiel V, Scaturro P, Mann M, Pichlmair A. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. Nature 2021;594:246–52.

[45] Chen Y, An X, Wang Z, Guan S, An H, Huang Q, Zhang H, Liang L, Huang B, Wang H, Lu M, Nie H, Wang J, Dai X, Lu X. Transcriptome and lipidome profile of human mesenchymal stem cells with reduced senescence and increased trilineage differentiation ability upon drug treatment. Aging 2021;13:9991–10014.

[46] Hill SM, Heiser LM, Cokelaer T, Unger M, Nesser NK, Carlin DE, Zhang Y, Sokolov A, Paull EO, Wong CK, Graim K, Bivol A, Wang H, Zhu F, Afsari B, Danilova LV, Favorov AV, Lee WS, Taylor D, Hu CW, Long BL, Noren DP, Bisberg AJ, HPN-DREAM Consortium, Mills GB, Gray JW, Kellen M, Norman T, Friend S, Qutub AA, Fertig EJ, Guan Y, Song M, Stuart JM, Spellman PT, Koeppl H, Stolovitzky G, Saez-Rodriguez J, Mukherjee S. Inferring causal molecular networks: empirical assessment through a community-based effort. Nat Methods 2016;13:310–8.

[47] Prill RJ, Marbach D, Saez-Rodriguez J, Sorger PK, Alexopoulos LG, Xue X, Clarke ND, Altan-Bonnet G, Stolovitzky G. Towards a rigorous assessment of systems biology models: the DREAM3 challenges. PLoS One 2010;5:e9202.

[48] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform 2008;9:559.

[49] Saint-Antoine MM, Singh A. Network inference in systems biology: recent developments, challenges, and applications. Curr Opin Biotechnol 2020;63:89–98.

[50] Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinform 2006;7(1):S7.

[51] Lachmann A, Giorgi FM, Lopez G, Califano A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. Bioinformatics 2016;32:2233–5.

[52] Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol 2007;5:e8.

[53] Liu W, Zhu W, Liao B, Chen H, Ren S, Cai L. Improving gene regulatory network structure using redundancy reduction in the MRNET algorithm. RSC Adv 2017;7:23222–33.

[54] Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. PLoS One 2010;5. https://doi.org/10.1371/journal.pone.0012776

[55] Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J, Aerts S. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. Bioinformatics 2019;35:2159–61.

[56] Guo S, Jiang Q, Chen L, Guo D. Gene regulatory network inference using PLS-based methods. BMC Bioinform 2016;17:545.

[57] Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 2008;9:432–41.

[58] Wang T, Ren Z, Ding Y, Fang Z, Sun Z, MacDonald ML, Sweet RA, Wang J, Chen W. FastGGM: an efficient algorithm for the inference of Gaussian graphical model in biological networks. PLoS Comput Biol 2016;12:e1004755.

[59] Ren Z, Sun T, Zhang C-H, Zhou HH. Asymptotic normality and optimalities in estimation of large Gaussian graphical models. Aos 2015;43:991–1026.

[60] Koller D, Friedman N. Probabilistic Graphical Models: Principles And Techniques. MIT Press; 2009.

[61] Beretta S, Castelli M, Gonçalves I, Henriques R, Ramazzotti D. Learning the structure of bayesian networks: a quantitative assessment of the effect of different algorithmic schemes. Complexity 2018;2018:1–12. https://doi.org/10.1155/2018/1591878

[62] P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction, and Search, (2001). ⟨https://doi.org/10.7551/mitpress/1754.001.0001⟩.

[63] Kuipers J, Thurnherr T, Moffa G, Suter P, Behr J, Goosen R, Christofori G, Beerenwinkel N. Mutational interactions define novel cancer subgroups. Nat Commun 2018;9:4353.

[64] P. Suter, J. Kuipers, G. Moffa, N. Beerenwinkel, Bayesian structure learning and sampling of Bayesian networks with the R package BiDAG, arXiv [stat.CO]. (2021). ⟨http://arxiv.org/abs/2105.00488⟩.

[65] Liu F, Zhang S-W, Guo W-F, Wei Z-G, Chen L. Inference of gene regulatory network based on local bayesian networks. PLoS Comput Biol 2016;12:e1005024.

[66] N.K. Kitson, A.C. Constantinou, Z. Guo, Y. Liu, K. Chobtham, A survey of Bayesian Network structure learning, arXiv [cs.LG]. (2021). ⟨http://arxiv.org/abs/2109.11415⟩.

[67] Vowels MJ, Camgoz NC, Bowden R. D'ya like DAGs? A survey on structure learning and causal discovery. ACM Comput Surv 2023;55:1–36. https://doi.org/10.1145/3527154

[68] Razaghi-Moghadam Z, Nikoloski Z. Supervised learning of gene-regulatory networks based on graph distance profiles of transcriptomics data. NPJ Syst Biol Appl 2020;6:21.

[69] Yuan Y, Bar-Joseph Z. Deep learning for inferring gene relationships from single-cell expression data. Proc Natl Acad Sci USA 2019;116:27151–8.

[70] Chen J, Cheong C, Lan L, Zhou X, Liu J, Lyu A, Cheung WK, Zhang L. DeepDRIM: a deep neural network to reconstruct cell-type-specific gene regulatory network using single-cell RNA-seq data. Brief Bioinform 2021;22. https://doi.org/10.1093/bib/bbab325

[71] X. Zheng, B. Aragam, P. Ravikumar, E.P. Xing, DAGs with NO TEARS: Continuous optimization for structure learning, arXiv [stat.ML]. (2018). ⟨https://proceedings.neurips.cc/paper/2018/hash/e347c51419ffb23ca3fd5050202f9c3d-Abstract.html⟩ (accessed November 30, 2022).

[72] Shu H, Zhou J, Lian Q, Li H, Zhao D, Zeng J, Ma J. Modeling gene regulatory networks using neural network architectures. Nat Comput Sci 2021;1:491–501. https://doi.org/10.1038/s43588-021-00099-8

[73] Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 2011;474:380–4.

[74] Girgenti MJ, Wang J, Ji D, Cruz DA, Traumatic Stress Brain Research Group, Stein MB, Gelernter J, Young KA, Huber BR, Williamson DE, Friedman MJ, Krystal JH, Zhao H, Duman RS. Transcriptomic organization of the human brain in post-traumatic stress disorder. Nat Neurosci 2021;24:24–33.

[75] Zampetaki A, Kiechl S, Drozdov I, Willeit P, Mayr U, Prokopi M, Mayr A, Weger S, Oberhollenzer F, Bonora E, Shah A, Willeit J, Mayr M. Plasma microRNA profiling reveals loss of endothelial miR-126 and other microRNAs in type 2 diabetes. Circ Res 2010;107:810–7.

[76] Wexler EM, Rosen E, Lu D, Osborn GE, Martin E, Raybould H, Geschwind DH. Genome-wide analysis of a Wnt1-regulated transcriptional network implicates neurodegenerative pathways. Sci Signal 2011;4:ra65.

[77] Meyer PE, Kontos K, Lafitte F, Bontempi G. Information-theoretic inference of large transcriptional regulatory networks. EURASIP J Bioinform Syst Biol 2007:79879.

[78] Verfaillie A, Imrichova H, Atak ZK, Dewaele M, Rambow F, Hulselmans G, Christiaens V, Svetlichnyy D, Luciani F, Van den Mooter L, Claerhout S, Fiers M, Journe F, Ghanem G-E, Herrmann C, Halder G, Marine J-C, Aerts S. Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. Nat Commun 2015;6:6683.

[79] Kuppe C, Ibrahim MM, Kranz J, Zhang X, Ziegler S, Perales-Patón J, Jansen J, Reimer KC, Smith JR, Dobie R, Wilson-Kanamori JR, Halder M, Xu Y, Kabgani N, Kaesler N, Klaus M, Gernhold L, Puelles VG, Huber TB, Boor P, Menzel S, Hoogenboezem RM, Bindels EMJ, Steffens J, Floege J, Schneider RK, Saez-Rodriguez J, Henderson NC, Kramann R. Decoding myofibroblast origins in human kidney fibrosis. Nature 2021;589:281–6.

[80] Cerutti C, Zhang L, Tribollet V, Shi J-R, Brillet R, Gillet B, Hughes S, Forcet C, Shi T-L, Vanacker J-M. Computational identification of new potential transcriptional partners of ERRα in breast cancer cells: specific partners for specific targets. Sci Rep 2022;12:3826.

[81] Suter P, Kuipers J, Beerenwinkel N. Discovering gene regulatory networks of multiple phenotypic groups using dynamic Bayesian networks. Brief Bioinform 2022. https://doi.org/10.1093/bib/bbac219

[82] Shi N, Zhu Z, Tang K, Parker D, He S. ATEN: And/Or tree ensemble for inferring accurate Boolean network topology and dynamics. Bioinformatics 2020;36:578–85.

[83] Barman S, Kwon Y-K. A Boolean network inference from time-series gene expression data using a genetic algorithm. Bioinformatics 2018;34:i927–33.

[84] Litovchenko M, Meireles-Filho ACA, Frochaux MV, Bevers RPJ, Prunotto A, Anduaga AM, Hollis B, Gardeux V, Braman VS, Russeil JMC, Kadener S, Dal Peraro M, Deplancke B. Extensive tissue-specific expression variation and novel regulators underlying circadian behavior. Sci Adv 2021;7. https://doi.org/10.1126/sciadv.abc3781

[85] Huynh-Thu VA, Geurts P. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. Sci Rep 2018;8:3384.

[86] Arrieta-Ortiz ML, Hafemeister C, Bate AR, Chu T, Greenfield A, Shuster B, Barry SN, Gallitto M, Liu B, Kacmarczyk T, Santoriello F, Chen J, Rodrigues CDA, Sato T, Rudner DZ, Driks A, Bonneau R, Eichenberger P. An experimentally supported model of the Bacillus subtilis global transcriptional regulatory network. Mol Syst Biol 2015;11:839.

[87] Gibbs CS, Jackson CA, Saldi G-A, Tjärnberg A, Shah A, Watters A, De Veaux N, Tchourine K, Yi R, Hamamsy T, Castro DM, Carriero N, Gorissen BL, Gresham D, Miraldi ER, Bonneau R. High performance single-cell gene regulatory network inference at scale: the inferelator 3.0. Bioinformatics 2022. https://doi.org/10.1093/bioinformatics/btac117

[88] Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, Thorsson V. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome Biol 2006;7:R36.

[89] Madar A, Greenfield A, Ostrer H, Vanden-Eijnden E, Bonneau R. The inferelator 2.0: a scalable framework for reconstruction of dynamic regulatory network models. Conf Pro. IEEE Eng Med Biol Soc 2009;2009:5448–51.

[90] Ma B, Fang M, Jiao X. Inference of gene regulatory networks based on nonlinear ordinary differential equations. Bioinformatics 2020;36:4885–93.

[91] Lu J, Dumitrascu B, McDowell IC, Jo B, Barrera A, Hong LK, Leichter SM, Reddy TE, Engelhardt BE. Causal network inference from gene transcriptional time-series response to glucocorticoids. PLoS Comput Biol 2021;17:e1008223.

[92] Finkle JD, Wu JJ, Bagheri N. Windowed Granger causal inference strategy improves discovery of gene regulatory networks. Proc Natl Acad Sci USA 2018;115:2252–7.

[93] Clark DJ, Dhanasekaran SM, Petralia F, Pan J, Song X, Hu Y, da Veiga Leprevost F, Reva B, Lih T-SM, Chang H-Y, Ma W, Huang C, Ricketts CJ, Chen L, Krek A, Li Y, Rykunov D, Li QK, Chen LS, Ozbek U, Vasaikar S, Wu Y, Yoo S, Chowdhury S, Wyczalkowski MA, Ji J, Schnaubelt M, Kong A, Sethuraman S, Avtonomov DM, Ao M, Colaprico A, Cao S, Cho K-C, Kalayci S, Ma S, Liu W, Ruggles K, Calinawan A, Gümüş ZH, Geiszler D, Kawaler E, Teo GC, Wen B, Zhang Y, Keegan S, Li K, Chen F, Edwards N, Pierorazio PM, Chen XS, Pavlovich CP, Hakimi AA, Brominski G, Hsieh JJ, Antczak A, Omelchenko T, Lubinski J, Wiznerowicz M, Linehan WM, Kinsinger CR, Thiagarajan M, Boja ES, Mesri M, Hiltke T, Robles AI, Rodriguez H, Qian J, Fenyö D, Zhang B, Ding L, Schadt E, Chinnaiyan AM, Zhang Z, Omenn GS, Cieslik M, Chan DW, Nesvizhskii AI, Wang P, Zhang H. Clinical proteomic tumor analysis consortium, integrated proteogenomic characterization of clear cell renal cell carcinoma. Cell 2020;180:207.

[94] Petralia F, Wang P, Yang J, Tu Z. Integrative random forest for gene regulatory network inference. Bioinformatics 2015;31:i197–205.

[95] Nguyen P, Braun R. Semi-supervised network inference using simulated gene expression dynamics. Bioinformatics 2018;34:1148–56.

[96] Glass K, Quackenbush J, Silverman EK, Celli B, Rennard SI, Yuan G-C, DeMeo DL. Sexually-dimorphic targeting of functionally-related genes in COPD. BMC Syst Biol 2014;8:118.

[97] Glass K, Huttenhower C, Quackenbush J, Yuan G-C. Passing messages between biological networks to refine predicted interactions. PLoS One 2013;8:e64832.

[98] Kogelman LJA, Falkenberg K, Buil A, Erola P, Courraud J, Laursen SS, Michoel T, Olesen J, Hansen TF. Changes in the gene expression profile during spontaneous migraine attacks. Sci Rep 2021;11:8294.

[99] Bonnet E, Calzone L, Michoel T. Integrative multi-omics module network inference with Lemon-Tree. PLoS Comput Biol 2015;11:e1003983.

[100] Chasman D, Iyer N, Fotuhi Siahpirani A, Estevez Silva M, Lippmann E, McIntosh B, Probasco MD, Jiang P, Stewart R, Thomson JA, Ashton RS, Roy S. Inferring regulatory programs governing region specificity of neuroepithelial stem cells during early hindbrain and spinal cord development. Cell Syst 2019;9:167–86. e12.

[101] Siahpirani AF, Roy S. A prior-based integrative framework for functional transcriptional regulatory network inference. Nucleic Acids Res 2017;45:2221.

[102] Banf M, Rhee SY. Enhancing gene regulatory network inference through data integration with markov random fields. Sci Rep 2017;7:41174.

[103] Hu Y, Rehawi G, Moyon L, Gerstner N, Ogris C, Knauer-Arloth J, Bittner F, Marsico A, Mueller NS. Network embedding across multiple tissues and data modalities elucidates the context of host factors important for COVID-19 infection. Front Genet 2022;13:909714.

[104] Ogris C, Hu Y, Arloth J, Müller NS. Versatile knowledge guided network inference method for prioritizing key regulatory factors in multi-omics data. Sci Rep 2021;11:6806.

[105] Allaway KC, Gabitto MI, Wapinski O, Saldi G, Wang C-Y, Bandler RC, Wu SJ, Bonneau R, Fishell G. Genetic and epigenetic coordination of cortical interneuron development. Nature 2021;597:693–7.

[106] Castro DM, de Veaux NR, Miraldi ER, Bonneau R. Multi-study inference of regulatory networks for more accurate models of gene regulation. PLoS Comput Biol 2019;15:e1006591.

[107] Sun J, Wang J, Yuan X, Wu X, Sui T, Wu A, Cheng G, Jiang T. Regulation of early host immune responses shapes the pathogenicity of avian influenza A virus. Front Microbiol 2019;10:2007.

[108] Omranian N, Eloundou-Mbebi JMO, Mueller-Roeber B, Nikoloski Z. Gene regulatory network inference using fused LASSO on multiple data sets. Sci Rep 2016;6:20533.

[109] A. Karbalayghareh, M. Sahin, C.S. Leslie, Chromatin interaction aware gene regulatory modeling with graph attention networks, (n.d.). ⟨https://doi.org/10.1101/2021.03.31.437978⟩.

[110] Avsec Ž, Agarwal V, Visentin D, Ledsam JR, Grabska-Barwinska A, Taylor KR, Assael Y, Jumper J, Kohli P, Kelley DR. Effective gene expression prediction from sequence by integrating long-range interactions. Nat Methods 2021;18:1196–203.

[111] Wang QS, Kelley DR, Ulirsch J, Kanai M, Sadhuka S, Cui R, Albors C, Cheng N, Okada Y, Biobank Japan Project, Aguet F, Ardlie KG, MacArthur DG, Finucane HK. Leveraging supervised learning for functionally informed fine-mapping of cis-eQTLs identifies an additional 20,913 putative causal eQTLs. Nat Commun 2021;12:3394.

[112] Kelley DR, Reshef YA, Bileschi M, Belanger D, McLean CY, Snoek J. Sequential regulatory activity prediction across chromosomes with convolutional neural networks. Genome Res 2018;28:739–50.

[113] Hemedan AA, Niarakis A, Schneider R, Ostaszewski M. Boolean modelling as a logic-based dynamic approach in systems medicine. Comput Struct Biotechnol J 2022;20:3161–72.

[114] Barman S, Kwon Y-K. A novel mutual information-based Boolean network inference method from time-series gene expression data. PLoS One 2017;12:e0171097.

[115] Pušnik Ž, Mraz M, Zimic N, Moškon M. Review and assessment of Boolean approaches for inference of gene regulatory networks. Heliyon 2022;8:e10222.

[116] Scutari M. Bayesian network models for incomplete and dynamic data. Stat Neerl 2020;74:397–419.

[117] Verleyen W, Ballouz S, Gillis J. Measuring the wisdom of the crowds in network-based gene function inference. Bioinformatics 2015;31:745–52.

[118] Marbach D, Roy S, Ay F, Meyer PE, Candeias R, Kahveci T, Bristow CA, Kellis M. Predictive regulatory models in *Drosophila melanogaster* by integrative inference of transcriptional networks. Genome Res 2012;22:1334–49. https://doi.org/10.1101/gr.127191.111

[119] Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R. Gene regulatory network inference: Data integration in dynamic models—a review. Biosystems 2009;96:86–103.

[120] Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B, Goldenberg A. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods 2014;11:333–7.

[121] Schlauch D, Paulson JN, Young A, Glass K, Quackenbush J. Estimating gene regulatory networks with pandaR. Bioinformatics 2017;33:2232–4.

[122] Segal E, Pe'er D, Regev A, Koller D, Friedman N, Jaakkola T. Learning module networks. J Mach Learn Res 2005;6⟨https://www.jmlr.org/papers/volume6/segal05a/segal05a.pdf⟩.

[123] Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS, Comprehensive A. Survey on graph neural networks. IEEE Trans Neural Netw Learn Syst 2021;32:4–24.

[124] Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The graph neural network model. IEEE Trans Neural Netw 2009;20:61–80.

[125] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph Attention Networks, arXiv [stat.ML]. (2017). ⟨http://arxiv.org/abs/1710.10903⟩.

[126] Brauwers G, Frasincar F. A general survey on attention mechanisms in deep learning. IEEE Trans Knowl Data Eng 2021. 1–1.

[127] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst 2017;30⟨https://proceedings.neurips.cc/paper/7181-attention-is-all-you-need⟩.

[128] D. Bahdanau, K. Cho, Y. Bengio, Neural Machine Translation by Jointly Learning to Align and Translate arXiv [cs.CL] 2014.⟨http://arxiv.org/abs/1409.0473⟩.

[129] Lai Y, Wu B, Chen L, Zhao H. A statistical method for identifying differential gene-gene co-expression patterns. Bioinformatics 2004;20:3146–55. https://doi.org/10.1093/bioinformatics/bth379

[130] Carter SL, Brechbühler CM, Griffin M, Bond AT. Gene co-expression network topology provides a framework for molecular characterization of cellular state. Bioinformatics 2004;20:2242–50.

[131] Kostka D, Spang R. Finding disease specific alterations in the co-expression of genes. Bioinformatics 2004;20(1):i194–9.

[132] Ideker T, Krogan NJ. Differential network biology. Mol Syst Biol 2012;8:565.

[133] Singh AJ, Ramsey SA, Filtz TM, Kioussi C. Differential gene regulatory networks in development and disease. Cell Mol Life Sci 2018;75:1013–25.

[134] Farahbod M, Pavlidis P. Differential coexpression in human tissues and the confounding effect of mean expression levels. Bioinformatics 2019;35:55–61.

[135] Giorgi FM, Lopez G, Woo JH, Bisikirska B, Califano A, Bansal M. Inferring protein modulation from gene expression data using conditional mutual information. PLoS One 2014;9:e109569.

[136] Danaher P, Wang P, Witten DM. The joint graphical lasso for inverse covariance estimation across multiple classes. J R Stat Soc Ser B Stat Methodol 2014;76:373–97.

[137] Yuan H, Xi R, Chen C, Deng M. Differential network analysis via lasso penalized D-trace loss. Biometrika 2017;104:755–70. https://doi.org/10.1093/biomet/asx049

[138] Amar D, Safer H, Shamir R. Dissection of regulatory networks that are altered in disease via differential co-expression. PLoS Comput Biol 2013;9:e1002955.

[139] Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. Hierarchical organization of modularity in metabolic networks. Science 2002;297:1551–5.

[140] Tesson BM, Breitling R, Jansen RC. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. BMC Bioinform 2010;11:497.

[141] Padi M, Quackenbush J. Detecting phenotype-driven transitions in regulatory network structure. NPJ Syst Biol Appl 2018;4:16.

[142] Bhuva DD, Cursons J, Smyth GK, Davis MJ. Differential co-expression-based detection of conditional relationships in transcriptional data: comparative analysis and application to breast cancer. Genome Biol 2019;20:236.

[143] Shojaie A. Differential network analysis: a statistical perspective. Wiley Interdiscip Rev Comput Stat 2021;13:e1508.

[144] McKenzie AT, Moyon S, Wang M, Katsyv I, Song W-M, Zhou X, Dammer EB, Duong DM, Aaker J, Zhao Y, Beckmann N, Wang P, Zhu J, Lah JJ, Seyfried NT, Levey AI, Katsel P, Haroutunian V, Schadt EE, Popko B, Casaccia P, Zhang B. Multiscale network modeling of oligodendrocytes reveals molecular components of myelin dysregulation in Alzheimer's disease. Mol Neurodegener 2017;12:82.

[145] McKenzie AT, Katsyv I, Song W-M, Wang M, Zhang B. DGCA: a comprehensive r package for differential gene correlation analysis. BMC Syst Biol 2016;10:106.

[146] Wang K, Alvarez MJ, Bisikirska BC, Linding R, Basso K, Dalla Favera R, Califano A. Dissecting the interface between signaling and transcriptional regulation in human B cells. Pac Symp Biocomput 2009:264–75.

[147] Wang K, Saito M, Bisikirska BC, Alvarez MJ, Lim WK, Rajbhandari P, Shen Q, Nemenman I, Basso K, Margolin AA, Klein U, Dalla-Favera R, Califano A. Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. Nat Biotechnol 2009;27:829–39.

[148] Paull EO, Aytes A, Jones SJ, Subramaniam PS, Giorgi FM, Douglass EF, Tagore S, Chu B, Vasciaveo A, Zheng S, Verhaak R, Abate-Shen C, Alvarez MJ, Califano A. A modular master regulator landscape controls cancer transcriptional identity. Cell 2021;184:334–51. e20.

[149] Riva A, Borgo F, Lassandro C, Verduci E, Morace G, Borghi E, Berry D. Pediatric obesity is associated with an altered gut microbiota and discordant shifts in Firmicutes populations. Environ Microbiol 2017;19:95–105.

[150] Lopes-Ramos CM, Belova T, Brunner TH, Ben Guebila M, Osorio D, Quackenbush J, Kuijjer ML. Regulatory network of PD1 signaling is associated with prognosis in glioblastoma multiforme. Cancer Res 2021;81:5401–12.

[151] Zhai Y, Franco LM, Atmar RL, Quarles JM, Arden N, Bucasas KL, Wells JM, Niño D, Wang X, Zapata GE, Shaw CA, Belmont JW, Couch RB. Host transcriptional response to influenza and other acute respiratory viral infections – a prospective cohort study. PLOS Pathogens 2015;11:e1004869https://doi.org/10.1371/journal.ppat.1004869

[152] Liu X, Wang Y, Ji H, Aihara K, Chen L. Personalized characterization of diseases using sample-specific networks. Nucleic Acids Res 2016;44:e164.

[153] Huang Y, Chang X, Zhang Y, Chen L, Liu X. Disease characterization using a partial correlation-based sample-specific network. Brief Bioinform 2021;22. https://doi.org/10.1093/bib/bbaa062

[154] M.L. Kuijjer, M. Tung, G. Yuan, J. Quackenbush, K. Glass, Estimating Sample-Specific Regulatory Networks, SSRN Electronic Journal. (n.d.). ⟨https://doi.org/10.2139/ssrn.3253573⟩.

[155] Weighill D, Ben Guebila M, Glass K, Quackenbush J, Platig J. Predicting genotype-specific gene regulatory networks. Genome Res 2022;32:524–33.

[156] Sun X, Liu X, Xia M, Shao Y, Zhang XD. Multicellular gene network analysis identifies a macrophage-related gene signature predictive of therapeutic response and prognosis of gliomas. J Transl Med 2019;17:159.

[157] Lopes-Ramos CM, Chen C-Y, Kuijjer ML, Paulson JN, Sonawane AR, Fagny M, Platig J, Glass K, Quackenbush J, DeMeo DL. Sex differences in gene expression and regulatory networks across 29 human tissues. Cell Rep 2020;31:107795.

[158] Silverman EK, Schmidt HHHW, Anastasiadou E, Altucci L, Angelini M, Badimon L, Balligand J-L, Benincasa G, Capasso G, Conte F, Di Costanzo A, Farina L, Fiscon G, Gatto L, Gentili M, Loscalzo J, Marchese C, Napoli C, Paci P, Petti M, Quackenbush J, Tieri P, Viggiano D, Vilahur G, Glass K, Baumbach J. Molecular networks in Network Medicine: Development and applications. Wiley Interdiscip Rev Syst Biol Med 2020;12:e1489.

[159] Lazareva O, Baumbach J, List M, Blumenthal DB. On the limits of active module identification. Brief Bioinform 2021;22. https://doi.org/10.1093/bib/bbab066

[160] Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, Duesbury M, Dumousseau M, Galeota E, Hinz U, Iannuccelli M, Jagannathan S, Jimenez R, Khadake J, Lagreid A, Licata L, Lovering RC, Meldal B, Melidoni AN, Milagros M, Peluso D, Perfetto L, Porras P, Raghunath A, Ricard-Blum S, Roechert B, Stutz A, Tognolli M, van Roey K, Cesareni G, Hermjakob H. The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res 2014;42:D358–63.

[161] Schaefer MH, Serrano L, Andrade-Navarro MA. Correcting for the study bias associated with protein-protein interaction measurements reveals differences between protein degree distributions from different cancer types. Front Genet 2015;6:260.

[162] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature 2021:583–9.

[163] Marbach D, Prill RJ, Schaffter T, Mattiussi C, Floreano D, Stolovitzky G. Revealing strengths and weaknesses of methods for gene network inference. Proc Natl Acad Sci USA 2010;107:6286–91.

[164] Choobdar S, The DREAM Module Identification Challenge Consortium, Ahsen ME, Crawford J, Tomasoni M, Fang T, Lamparter D, Lin J, Hescott B, Hu X, Mercer J, Natoli T, Narayan R, Subramanian A, Zhang JD, Stolovitzky G, Kutalik Z, Lage K, Slonim DK, Saez-Rodriguez J, Cowen LJ, Bergmann S, Marbach D. Assessment of network module identification across complex diseases. Nat Methods 2019;16:843–52. https://doi.org/10.1038/s41592-019-0509-5

[165] Emmert-Streib F, Glazko GV, Altay G, de Matos Simoes R. Statistical inference and reverse engineering of gene regulatory networks from observational expression data. Front Genet 2012;3:8.

[166] Nguyen H, Tran D, Tran B, Pehlivan B, Nguyen T. A comprehensive survey of regulatory network inference methods using single cell RNA sequencing data. Brief Bioinform 2021;22. https://doi.org/10.1093/bib/bbaa190

[167] Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol 2019;15⟨https://www.embopress.org/doi/abs/10.15252/msb.20188746⟩.

[168] Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat Methods 2020. https://doi.org/10.1038/s41592-019-0690-6

[169] Blencowe M, Arneson D, Ding J, Chen Y-W, Saleem Z, Yang X. Network modeling of single-cell omics data: challenges, opportunities, and progresses. Emerg Top Life Sci 2019;3:379–98.

[170] Baur B, Shin J, Zhang S, Roy S. Data integration for inferring context-specific gene regulatory networks. Curr Opin Syst Biol 2020;23:38–46.

[171] Yan J, Risacher SL, Shen L, Saykin AJ. Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. Brief Bioinform 2018;19:1370–81.

[172] Network TCGA. The cancer genome atlas network, comprehensive molecular portraits of human breast tumours. Nature 2012;490:61–70. https://doi.org/10.1038/nature11412

[173] Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, Gillard B, Ramsey K, Sullivan S, Bridge J, Harold Magazine, Syron J, Fleming J, Siminoff L, Traino H, Mosavel M, Barker L, Jewell S, Rohrer D, Maxim D, Filkins D, Harbach P, Cortadillo E, Berghuis B, Turner L, Hudson E, Feenstra K, Sobin L, Robb J, Branton P, Korzeniewski G, Shive C, Tabor D, Qi L, Groch K, Nampally S, Buia S, Zimmerman A, Smith A, Burges R, Robinson K, Valentino K, Bradbury D, Cosentino M, Diaz-Mayoral N, Kennedy M, Engel T, Williams P, Erickson K, Ardlie K, Winckler W, Getz G, DeLuca D, MacArthur D, Kellis M, Thomson A, Young T, Gelfand E, Donovan M, Meng Y, Grant G, Mash D, Marcus Y, Basile M, Liu J, Zhu J, Tu Z, Cox NJ, Nicolae DL, Gamazon ER, Im HK, Konkashbaev A, Pritchard J, Stevens M, Flutre T, Wen X, Dermitzakis ET, Lappalainen T, Guigo R, Monlong J, Sammeth M, Koller D, Battle A, Mostafavi S, McCarthy M, Pritchard M, Maller J, Rusyn I, Nobel A, Gelfand F, Shabalin A, Feolo M, Sharopova N, Sturcke A, Paschal J, Anderson JM, Wilder EL, Derr LK, Green ED, Struewing JP, Temple G, Volpi S, Boyer JT, Thomson EJ, Guyer MS, Ng C, Abdallah A, Colantuoni D, Insel TR, Koester SE, Little AR, Bender PK, Lehner T, Yao Y, Compton CC, Vaught JB, Sawyer S, Lockhart NC, Demchok J, Moore HF. The genotype-tissue expression (GTEx) project. Nat Genet 2013:580–5.

[174] Zhang J, Lee D, Dhiman V, Jiang P, Xu J, McGillivray P, Yang H, Liu J, Meyerson W, Clarke D, Gu M, Li S, Lou S, Xu J, Lochovsky L, Ung M, Ma L, Yu S, Cao Q, Harmanci A, Yan K-K, Sethi A, Gürsoy G, Schoenberg MR, Rozowsky J, Warrell J, Emani P, Yang YT, Galeev T, Kong X, Liu S, Li X, Krishnan J, Feng Y, Rivera-Mulia JC, Adrian J, Broach JR, Bolt M, Moran J, Fitzgerald D, Dileep V, Liu T, Mei S, Sasaki T, Trevilla-Garcia C, Wang S, Wang Y, Zang C, Wang D, Klein RJ, Snyder M, Gilbert DM, Yip K, Cheng C, Yue F, Liu XS, White KP, Gerstein M. An integrative ENCODE resource for cancer genomics. Nat Commun 2020;11:3696.

[175] List M, Dehghani Amirabad A, Kostka D, Schulz MH. Large-scale inference of competing endogenous RNA networks with sparse partial correlation. Bioinformatics 2019;35:i596–604.

[176] Hawe JS, Theis FJ, Heinig M. Inferring interaction networks from multi-omics data. Front Genet 2019;10:535.

[177] Talukder A, Barham C, Li X, Hu H. Interpretation of deep learning in genomics and epigenomics. Brief Bioinform 2021;22. https://doi.org/10.1093/bib/bbaa177

[178] Zrimec J, Buric F, Kokina M, Garcia V, Zelezniak A. Learning the regulatory code of gene expression. Front Mol Biosci 2021;8:673363.

[179] Zhao Y, Shao J, Asmann YW. Assessment and optimization of explainable machine learning models applied to transcriptomic data. Genom Proteom Bioinform 2022. https://doi.org/10.1016/j.gpb.2022.07.003

[180] Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res 2020;48:D845–55.

[181] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30.

[182] Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, Loney F, May B, Milacic M, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P. The reactome pathway knowledgebase. Nucleic Acids Res 2020;48:D498–503.

[183] Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: pathway editing for the people. PLoS Biol 2008;6:e184.

[184] Meyer P, Saez-Rodriguez J. Advances in systems biology modeling: 10 years of crowdsourcing DREAM challenges. Cell Syst 2021;12:636–53.

[185] Norel R, Rice JJ, Stolovitzky G. The self-assessment trap: can we all be better than average? Mol Syst Biol 2011;7:537.

[186] Jackson CA, Castro DM, Saldi G-A, Bonneau R, Gresham D. Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. Elife 2020;9. https://doi.org/10.7554/eLife.51254

[187] Lopes R, Sprouffske K, Sheng C, Uijttewaal ECH, Wesdorp AE, Dahinden J, Wengert S, Diaz-Miyar J, Yildiz U, Bleu M, Apfel V, Mermet-Meillon F, Krese R, Eder M, Olsen AV, Hoppe P, Knehr J, Carbone W, Cuttat R, Waldt A, Altorfer M, Naumann U, Weischenfeldt J, deWeck A, Kauffmann A, Roma G, Schübeler D, Galli GG. Systematic dissection of transcriptional regulatory networks by genome-scale and single-cell CRISPR screens. Sci Adv 2021;7. https://doi.org/10.1126/sciadv.abf5733

[188] Pierce SE, Granja JM, Greenleaf WJ. High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. Nat Commun 2021;12:2969.

[189] Akinci E, Hamilton MC, Khowpinitchai B, Sherwood RI. Using CRISPR to understand and manipulate gene regulation. Development 2021;148. https://doi.org/10.1242/dev.182667

[190] Ozturk K, Dow M, Carlin DE, Bejar R, Carter H. The emerging potential for network analysis to inform precision cancer medicine. J Mol Biol 2018;430:2875–99.

[191] Matchado MS, Lauber M, Reitmeier S, Kacprowski T, Baumbach J, Haller D, List M. Network analysis methods for studying microbial communities: a mini review. Comput Struct Biotechnol 2021;19:2687–98.