Statistical Science
2021, Vol. 36, No. 1, 89–108
https://doi.org/10.1214/20-STS792

Network Modeling in Biology: Statistical Methods for Gene and Brain Networks

Y. X. Rachel Wang, Lexin Li, Jingyi Jessica Li and Haiyan Huang

Abstract. The rise of network data in many different domains has offered researchers new insights into the problem of modeling complex systems and propelled the development of numerous innovative statistical methodologies and computational tools. In this paper, we primarily focus on two types of biological networks, gene networks and brain networks, where statistical network modeling has found both fruitful and challenging applications. Unlike other network examples such as social networks where network edges can be directly observed, both gene and brain networks require careful estimation of edges using measured data as a first step. We provide a discussion on existing statistical and computational methods for edge estimation and subsequent statistical inference problems in these two types of biological networks.

Key words and phrases: Gene regulatory networks, brain connectivity networks, network reconstruction, network inference.

1. INTRODUCTION

Network structures exist everywhere in biology as many biological systems function via complex interactions among their individual components. In ecosystems, species interact in a number of different forms which are central to maintaining biodiversity, the most common being predator-prey relationships. In human brain, neurons communicate by passing electric and chemical signals through synapses. At the cellular level, DNA, RNA, proteins and other molecules participate in a variety of biochemical reactions that determine inner workings of a cell. Networks offer a succinct mathematical representation of these systems, with "sets of items, which we will call vertices or sometimes nodes, with connections between them, called edges" [151].

Network modeling has been successfully applied in many settings where the biological questions of interest have their counterparts in graph theory. For example, many biochemical networks have a scale-free topology with a few highly connected nodes [14], known as hubs in network analysis, which may correspond to key enzymes in biochemical processes. Another key goal in network analysis is to detect communities, which are groups of tightly connected nodes. These could be genes with related functionalities, or regions of brain with coordinated actions.

From the statistical point of view, another reason why biological systems are particularly amenable to network analysis lies in the richness of data made available by various technologies, especially for gene network and brain network modeling. In these areas, measurements of variables are not limited to observational settings, and extensive experimental studies can be performed to examine how variables respond under different conditions. One prominent example can be found in genomics studies, where numerous high-throughput, deep sequencing technologies have generated a staggering amount of data measuring gene expression levels and epigenetic interactions. One particularly relevant technology is RNA-seq, routinely used nowadays to characterize the transcriptome. In addition to observational data, gene intervention data can be obtained by performing gene knockout or knockdown experiments to study the effect of perturbations. Another example is the studies of brain, where numerous imaging technologies, such as fMRI, have collected a wide variety of brain images measuring distinct brain characteristics, ranging from brain structure and function to numerous chemical constituents. Such data can be collected under resting state or when the subjects are required to perform

Y. X. Rachel Wang is a Senior Lecturer in the School of Mathematics and Statistics, University of Sydney, Sydney, New South Wales, Australia (e-mail: rachel.wang@sydney.edu.au). Lexin Li is a Professor in the Department of Biostatistics and Epidemiology, and School of Public Health, University of California, Berkeley, California, USA (e-mail: lexinli@berkeley.edu). Jingyi Jessica Li is an Associate Professor, Department of Statistics, University of California, Los Angeles, California, USA (e-mail: jli@stat.ucla.edu). Haiyan Huang is a Professsor in the Department of Statistics, University of California, Berkeley, California, USA (e-mail: hhuang@stat.berkeley.edu).

cognitive tasks (e.g., task-based fMRI). For this reason, here we choose to discuss statistical methods for gene and brain networks, with more focus on the former.

The enormous wealth of data provides both opportunities and challenges for the analysis of the above two classes of networks. Unlike physical or social networks, interactions among genes are much harder to observe. Although experiments can be performed to search for and verify each gene–gene interaction, it is much more cost effective to infer these interactions and reconstruct network edges using statistical and computational tools on high-throughput gene expression data (more recently single cell expression data). The computational results can help narrow down possible candidates for further experimental validation. The computationally inferred networks may contain up to tens of thousands of nodes requiring efficient methods for network inference. In this article, we focus on a few specific problems involved in gene network analysis and mention other relevant applications in genomics beyond gene networks when appropriate. As another example in biology, we will also review statistical methods for constructing and analyzing brain connectivity networks.

Without claiming to be exhaustive, we will discuss the challenges in these networks considering the type and quality of data available, relevant biological questions to be addressed, and statistical and computational concerns. For gene networks, we will primarily focus on the use of gene expression data measured by RNA-seq or more traditional microarrays. We will also discuss RNA sequencing data obtained at the single-cell level, known as single-cell RNA-seq (scRNA-seq). For both gene and brain networks, we highlight the success and limitations of current network modeling paradigms and statistical methodologies, and propose possible directions for future development.

2. GENE NETWORKS

Gene regulatory networks play a fundamental role in defining cell structure and function. In such a network, transcription factors (TFs), RNA and other small molecules act as regulators to activate or repress the expression levels of genes, which in turn increase or decrease the production of proteins. Thus gene–gene interactions can occur in the form of direct physical binding of proteins (TFs) to their target sequences, which can be represented as directed graphs with causal relationships. In a broader sense, gene–gene interactions may also include indirect interactions when the expression of a gene influence those of others with regulations caused by one or more intermediaries, or when two genes are co-regulated thus showing similar expression profiles; these associations are generally reported as undirected graphs.

Despite all being the focus of studies in the network literature, biological networks such as gene networks differ from social networks in a few important aspects, which give rise to challenging situations for statistical modeling. Compared to relationship networks obtained from popular social media, gene networks are typically smaller in size. The former can additionally grow in size by including more users, whereas the size of gene networks is limited by the number of genes that exist in an organism and can be measured in an experiment. As will be explained in detail in Section 2.1, edges in gene networks need to be estimated from covariates. Since the measurements of these covariates rely on specific technologies, the number of samples one can take is often restricted by cost considerations and other practical constraints. Finally, since edges in these networks represent interactions between nodes, they are directly affected by the underlying dynamics in gene regulation. These biological processes are complicated in nature; gene regulatory mechanisms depend on tissue types, cellular environment and their activities can be changed by disease state. All of these factors can give rise to challenging situations for estimating and interpreting network structures. In the following sections, we will review existing approaches in the relevant literature with these limitations on biological data in mind.

2.1 Inferring Gene–Gene Relationships Using Expression Data

In the past two decades, estimating gene–gene interactions have primarily relied on gene expression data, which have been made readily available in the form of microarray or RNA-seq data. Coexpression is one of the earliest concepts proposed to infer edges in a gene network and is based on the concept of "guilt by association": genes that have similar expression profiles under different experimental conditions are likely to be coregulated, and hence functionally related. However, despite the extensive literature, many open questions remain due to the complex nature of gene interactions: in a broader sense, these coexpression relationships can be nonlinear, transient and subject to changes depending on the cellular environment. In this section, without claiming to be exhaustive, we discuss a few main approaches for inferring gene networks that differ in their considerations of how genes behave across the given samples. More detailed reviews can be found in, for example, [228].

Pairwise coexpression measure. Given an expression matrix with p genes arranged in rows and their expression levels measured under n experimental conditions in columns, computing the coexpression between genes i and j involves choosing a suitable similarity measure for estimating the association between two vectors. The choice of the measure crucially depends on a number

of factors including the nature of the interaction, experimental design, the number of samples available and other computational concerns.

Correlation measures based on Pearson's or Spearman's correlation are among the most popular methods used in the literature [51, 109, 197, 202, 236]. Either hard [25] or soft thresholding [112] is then applied to produce a binary or weighted network. These correlations are easy to compute and interpret but limited in the type of pairwise association they can detect, which is linear or monotonic. When the relationship between expression vectors is more complex, one commonly used class of methods is based on mutual information (MI). MI measures the general statistical dependence between gene expression levels and is thus able to capture nonlinear relationships. In the calculation of MI, marginal and joint entropies of the expression levels can be approximated via discretization [20] or using a smoothing kernel [15, 39, 139]. Other variants including MI with background, maximal MI and maximal information coefficient (MIC, [175]) have also been used in practice. For time-course data, techniques in times series analysis (e.g., time-frequency analysis) have been applied to improve the sensitivity of similarity measures [54, 172], often assuming explicit models for generating the observed data.

Other features intrinsic to the nature of gene interactions may create more complex situations. For instance, gene interactions may change as the intrinsic cellular state varies or only exist under a specific cellular condition [27, 257]. To detect local correlation patterns, spline regression models [160] and nonparametric methods based on comparing local expression patterns [178, 230] have been proposed. For time series data, another prevalent feature is the presence of time shifts between association patterns, reflecting the fact that regulation may take effect after a time delay. Methods for handling the time lag issue include time-shifted Pearson's correlation [100], time-shifted expression rank pattern analysis [229] and time sequence alignment algorithms [1, 69, 111, 252].

Partial correlation for group interactions. In a real biological pathway, a gene may interact with a group of genes but not possess a strong marginal relationship with any individual member of the group. Gaussian graphical models (GGM) offer a more realistic way to model these higher-level interactions. Assuming a multivariate normal distribution for the expression vectors for a set of genes W, this approach aims to estimate the partial correlation between genes i and j, that is, their correlation conditioning on $W \setminus \{i, j\}$. Since the partial correlations are proportional to their corresponding entries in the inverse covariance matrix Σ^{-1}, the inference problem amounts to estimating Σ^{-1}, or the precision matrix. The major difficulty of such an estimation arises from the high-dimensional nature of gene expression data, which naturally requires inbuilt sparsity in inference methods. A rich wealth of literature exists both in the context of gene expression analysis [119, 142, 161] and general high-dimensional inference [61, 81, 251, 259] to tackle this problem.

One limitation of this approach lies in the choice of the conditional set in the partial correlation calculation. As pointed out in [41] and [103], the inclusion of noisy genes in the set $W \setminus \{i, j\}$ may introduce spurious dependencies and consequently false edges in the estimated network. Instead of conditioning on the entire set $W \setminus \{i, j\}$, there have been efforts on using lower order partial correlations [41, 120, 121, 135, 234, 235], which condition on one or two other genes. Beyond lower order interactions, [103] proposed a semisupervised approach to screen for conditionally correlated genes with a small set of known pathway genes. An unsupervised approach involving applying sparse canonical correlation analysis coupled with repeated random partition and subsampling can be found in [231].

Adding causality and dynamics. A deeper understanding of the gene regulation system requires going beyond undirected relationships between genes and knowing the causal drivers behind them. Bayesian networks (BNs) use directed acyclic graphs (DAGs) to represent the joint distribution of nodes (genes) as a series of local probability distributions. In a BN, given its parents, each node is independent of its nondescendants. In this sense, each directed edge can be interpreted as a causal link. The difficulty of inferring BNs lies in the computational cost required to search through all possible graph structures, which is NP-hard [34]. In addition to greedy search [250], various heuristics have been proposed to increase the search efficiency [2, 128, 149]. One can gain further information from perturbation gene experiments (by knockout or RNA interference), with methods that provide causal bounds for direct and indirect effects based on skeleton graphs obtained from the path consistency algorithm [133], and estimate the posterior distribution of a causal ordering of genes with MCMC techniques [174]. Other studies have jointly modeled intervention and observational data using the maximum likelihood approach and greedy search [77], and utilizing the principle of invariance to model data from multiple experiments [143, 162].

BNs can be extended to capture temporal relationships between the variables [82, 104, 221, 261, 265]. In a dynamic BN, the joint probability factorizes into local probabilities of each node associated with every time point, where the parents of a node can include nodes from previous time points. Another class of methods based on differential equations (DEs), which models the rate of change in the expression level of a gene as a function of the expression of other genes with different functional forms

[12, 32, 209]. In addition to the issue of computational complexity, another drawback of these methods lies in the nature of data required to perform extensive inference. More sample measurements need taken on a slowly changing system or finely spaced in time in order to capture the underlying dynamics. Attempts to capture these causal and dynamic relationships in gene networks for higher organisms using expression data alone have had limited success; auxiliary information from other data sources (e.g., protein-protein interaction, ChIP-seq) can increase our chances in characterizing the complexity in a more realistic and accurate way [227, 243].

Beyond traditional studies: *Recent advancements and future trends*. Most of the methods discussed above focus on analyzing a single gene expression dataset, which suffers from the dimensionality curse ($p \gg n$ problem). To obtain a more complete picture, one step further is to perform integrated analysis on gene expression data generated by different groups [156, 225] to increase n, and even other types of data such as TF binding, protein-protein interaction (PPI), which provide direct physical evidence of regulatory interactions [11, 117, 138].

It is worth noting that recent advances in single-cell sequencing technology are offering a new perspective on studying gene pathway with gene expression information. For instance, single-cell RNA-seq (scRNA-seq) data can measure gene functional activities at individual cell resolution, and thus has potential to characterize gene regulatory actions with cell-to-cell variability [28, 56, 140]. However, despite these attractive and promising features, the high noise level in typical single-cell experiments as well as the dynamics of individual cells also present new challenges for developing statistical methods for data preprocessing and network/pathway inference. We will discuss single-cell data in more detail in Section 3.

In addition to computationally inferred gene–gene interactions, extensive experiments also have provided sets of "true" interactions in some species. For example, genome-wide experimental screening of gene–gene interactions have been carried out in yeast with high-throughput techniques (SGA, [212]; E-MAP, [187]), whereas screening these interactions in higher organisms require more restrictive techniques and have achieved much less coverage [46, 116, 124]. These experimentally validated interactions can be found in databases such as RegulonDB and KEGG [154, 183]. Comparing results from various computational methods against these validated interactions allows us to assess the performance of each method. Some of these comparisons also suggest different computational methods can lead to quite disparate sets of predicted interactions [137, 194]. As a result, the use of ensemble methods [86, 136] has been proposed to achieve more stable and accurate outcomes via bootstrapping and aggregation. Overall, constructing a complete catalog of gene interactions remains challenging due to the extensive scale of the problem. On the other hand, more specific biological goals and prior knowledge can help us narrow down possible approaches and lead to plausible simplifications.

2.2 Understanding Network Structures

Having reconstructed gene networks, it is now possible to systematically study the topological features of these graphs using graph-theoretical tools to understand and predict the underlying biological functions. While many local and global features can be extracted from the graphs, in this section we will focus on discussing node-level connectivity measures and how they reflect the functional importance of the nodes, and higher-level connectivity patterns including motifs and communities. Other extensive reviews on using graph-based methods for analyzing biological networks can be found in, for example, [6, 84, 159]. In this section, we will consider computationally reconstructed gene networks, which contain noise arising from estimation errors, and also known networks in genomics where edges are directly measured with biological assays, such as PPI networks.

Node-level connectivity. In gene and PPI networks, how a node is connected to the rest of the network can be an important indication of its biological role. Removing nodes with high connectivity or nodes between highly connected components can significantly affect the overall topology. In biological systems, this may correspond to malfunctioning of key genes or proteins which can cause serious perturbations. Different measures of node connectivity, or centrality, exist. In the simplest form, nodes with high degrees, which are also known as hubs, have been long studied in gene and PPI networks for model organisms, especially yeast [89, 264]. They have shown that hub genes and proteins encoded by them are essential to survival, and these genes tend to be older and evolve more slowly [60]. In human, hubs have been associated with cancer and other types of disease [13]—the protein products of disease-related genes tend to have high degrees [93, 223, 245]. Identifying hubs requires measuring node centrality, the simplest kind of which is node degrees. Reweighing each neighbor with their own degrees gives rise to another measure called eigenvector centrality. Similar to the PageRank algorithm, eigenvector centrality gives more weight to nodes connected to important neighbors and been used to distinguish essential proteins in yeast PPI networks [53] and mine gene–disease associations [157].

Another measure of connectivity, termed betweeness centrality, considers the number of times a node lies on the shortest path between two other nodes. Since nodes with high betweenness centrality act as bridges connecting subgraphs, they are also called bottlenecks. It has

been observed that many bottleneck nodes correspond to essential connector proteins and genes in directed regulatory networks [94, 249]. Further centrality measures and comparison of their performance in identifying essential or disease-causing genes/proteins can be found in [29, 158].

Higher-order structures—Motifs. At a higher level, biological networks are often decomposed into smaller functional modules in which individual nodes perform coordinated actions. The concept of motifs was introduced by [144] as simple building blocks of complex networks. A motif is a small connected subgraph, which occurs significantly more frequently in the given network than expected by chance. Commonly occurring motifs include positive and negative feedback loops, oscillators and bifans, and these have been associated with optimized biological functions in regulatory networks [13]. The statistical analysis of motifs amounts to a problem of subgraph counting: for a given subgraph, one needs to first obtain the frequency of all subgraphs which are topologically equivalent, then determine its statistical significance. The challenge of the first part lies in the computational challenge when the network is large. Since exhaustive enumeration is usually infeasible, sampling methods [99, 233] are needed for estimation. The second part depends on the random graph model used to determine the background frequency. [17] took a local graph alignment approach, which is conceptually similar to sequence alignment, with a scoring function measuring the significance of individual subgraphs and their similarity so that the aligned subgraphs are characterized by a consensus motif that has a high number of internal connections. [92] proposed a finite mixture model for random networks and used an EM algorithm to estimate the parameters and background probabilities. Other motif algorithms can be found in [36, 238].

Clustering and community detection. Since motifs tend to be small in size, another approach is to identify densely connected clusters of nodes which can correspond to genes with related functions or proteins involved in the same complex. Clustering can be applied to gene expression vectors directly using heuristic algorithms (Self-Organizing Maps [205]), genetic algorithms [44] or model based approaches (Expectation Maximization [148, 248]; variational Bayes [208]). Alternatively, noting that most methods in Section 2.1 give rise to a similarity matrix, k-means and hierarchical clustering have been widely used in gene expression studies [51, 207]. Taking into account that one gene can participate in multiple pathways, fuzzy versions of k-means have also been developed [42, 65]. One of the difficulties of these methods lies in the choice of number of clusters or where to cut the tree structure to produce the final clusters. This is usually done by considering the within cluster dispersion, or statistics derived from it including the gap statistic [210] and the silhouette measure [177].

Other methods operate on the given networks directly. Many heuristic algorithms have been developed for gene networks [16, 191] and PPI networks [67, 147] to identify tightly connected components in the graphs. In PPI networks, Markov Clustering (MCL, [219]) has been particularly popular. The algorithm simulates random walks on a given graph by iteratively taking powers of the underlying stochastic matrix and inflating each entry until the graph is partitioned into subsets. MCL has been widely applied to discover protein complexes and cluster protein sequences into families [52, 185, 222], but still lacks theoretical justification.

In statistical network analysis, identifying tightly connected clusters corresponds to the problem of community detection. Model-based community detection requires a generative probabilistic model for random graphs, one of the most popular being the stochastic block model [79]. In a SBM with K classes, for each node i, a latent class variable $Z_i \in \{1, \ldots, K\}$ is assigned according to some categorical distribution. Then the probability of an edge between nodes i and j is given by $P(A_{i,j} = 1 | Z_i = k, Z_j = l) = H_{k,l}$, where A is the adjacency matrix and H is the $K \times K$ connectivity probability matrix. The inference problems for SBMs involve both node classification and parameter estimation, and a block with a high internal edge probability can be considered as a potential functional module. An extensive literature on inference methods for SBM exists. On the other hand, although SBM and community detection have been applied to gene networks and PPI networks [40, 73], the vanilla model is too simplistic to account for real network features such as degree variation within blocks and overlapping blocks. These can be addressed to some extent using a degree-corrected SBM [98] and mixed membership SBM [5].

Going beyond gene and PPI networks, recent advances in chromatin conformation capture techniques open up new ground for applying community detection algorithms. Chromatin conformation capture experiments like Hi-C measure the frequency of interaction between pairs of genome loci in 3D space, thus providing insights the spatial organization of genomes. One specific feature of the 3D organization is known as topologically associating domains (TADs), which are densely interacting, contiguous chromatin regions playing important roles in regulating gene expression [45, 114, 189]. Treating genome loci as nodes and their interactions as edges, one can consider the structure of chromatin as an interaction network with TADs corresponding to dense communities. Methods based on mixed membership SBM [21] and modularity maximization [152, 247] have been proposed but do not enforce the constraint that the communities in this case have to be contiguous. [232] proposed a network

model that accounts for nonexchangeability of nodes (genome loci) and is capable of incorporating biological covariates at the TAD boundaries.

Gene prioritization—Semisupervised clustering. When a specific biological process or pathway is concerned with partial knowledge of the process/pathway known, a relevant question that has been considered extensively under a supervised or semisupervised setting in the literature is known as "gene prioritization" [146]. In general, gene prioritization refers to a computational analysis for ranking genes by their relevance to a disease or biomedical condition through a set of seed (bait) genes and some chosen relevance measure or criteria.

When attempting a whole genome analysis, a major challenge in gene prioritization analysis is how to extract sparse, true signals from large, heterogeneous, noisy data. For instance, when a particular pathway is targeted, the considered data would likely have a low signal-to-noise ratio since the great majority of genes may have no relation to the pathway of interest and the sheer number of pairs of such genes outweighs those that show patterned relations in data. Among many existing approaches, GIANT [70, 237] and ENDEAVOUR [213] have been widely used for a genome-wide gene prioritization analysis. They can accept a group of seed (or bait) genes that are believed to be related to the same biological process as input, and return a list of genes that have been ranked according to computed functional relevance by incorporating multiple sources of data. In particular, GIANT uses a data-driven Bayesian methodology to integrate diverse experiments and information such as genome-wide association study (GWAS) p-values and tissue-specific networks; ENDEAVOUR obtains a single global ranking of candidate genes by integrating their rankings associated with each data source using order statistics. These approaches have been found successful in many applications. However, the incorporation of multiple sources of information may bring both positive and negative effects to the analysis. On one side, more sources of information would allow assessing the interactions between candidate and seed (bait) genes from different perspectives and so may offer a more comprehensive portrait on the considered biological process. But on the other hand, information from multiple, heterogeneous sources could reflect different biology with diverse noise and so may dilute the strength in studying a specific biological process under a certain condition.

Given gene expression data, many available methods for gene prioritization analysis have pointed to the general "guilt by association" principle and its extensions by generating hypotheses about potential interactions between candidate genes and seed (bait) genes (e.g., [31, 72, 214]). For instance, GeneFishing [129] uses this strategy and identifies novel genes relevant to a biological process of interest under the guidance of seed (bait) genes utilizing a semisupervised, nonparametric clustering procedure coupled with a bagging-like majority voting approach. GeneFishing shares identical input-output schema with GIANT and ENDEAVOUR, but also differs from GIANT and ENDEAVOUR in key aspects. In particular, GeneFishing only uses gene-expression data. In a brief summary, the key features of GeneFishing include: (i) repeatedly, randomly splitting a large search-space into smaller ones and aggregating the results from all the sub-search-spaces (i.e., the bagging idea); (ii) adding known pathway genes into each sub-search-space to provide a focus for the search (making the method semi-supervised). Consequently, GeneFishing has been found to be advantageous in terms of being robust against noise in the seed (bait) genes and also being effective with handling a large noisy dataset with sparse signal in some applications.

As should be clear, false discoveries and missed discoveries are key issues with all the three methods mentioned above. One way to handle these issues is to use the irrelevance of most genes and replicability to deal with type I error, cross validation and stability for type II. Most importantly, if possible, it would be ideal to have results be guided by experimental validation.

3. SINGLE-CELL RNA SEQUENCING (SCRNA-SEQ) DATA

The recent advances of single-cell RNA-sequencing (scRNA-seq) technologies have revolutionized biomedical sciences by revealing genome-wide gene expression levels at an unprecedentedly individual cell level [50, 76, 88, 181, 190]. Most of the methods discussed in Section 2 were developed primarily for microarray and bulk RNA-seq technologies, which measure average gene expression levels across a collection of (from thousands to millions) cells and provide "coarse" tissue-level gene expression profiles. New scRNA-seq technologies have led to expression measurements at finer resolution and enabled researchers to confirm previously known cell types, to identify new cell types and to characterize gene–gene interactions within each cell type. Given that scRNA-seq data have revealed widespread heterogeneity among various cell types of the same tissue [18], gene networks inferred at the cell-type level are expected to uncover gene–gene relationships masked in tissue-level gene networks constructed using microarry and bulk RNA-seq data.

Conceptually, the aforementioned computational approaches for inferring gene networks from bulk gene expression data should still be relevant to scRNA-seq data if the data structure is compatible. The distinct characteristics of scRNA-seq data, however, have posed new computational challenges for gene network inference. Below we summarize the challenges and the state-of-the-art methodological development in three subsections. In Section 3.1,

we describe several computational issues in scRNA-seq data preprocessing, including the detection of "problematic cells," normalization of gene expression levels across cells, and imputation of missing gene expression levels in individual cells. In Section 3.2, we discuss identification of cell types from scRNA-seq data. In Section 3.3, we review existing studies on inferring cell-type-specific gene networks and discuss some open challenges and future research directions for network analysis using scRNA-seq data.

3.1 scRNA-seq Data Preprocessing

Both being high-throughput sequencing technologies for measuring gene expression, the preprocessing of scRNA-seq data shares some conceptual similarity with that of bulk RNA-seq data, but also presents unique challenges. While many well-studied preprocessing techniques are available for bulk RNA-seq, developing relevant methods for scRNA-seq data is still a very active research area. For this reason, we present here a discussion of issues arising from scRNA-seq preprocessing.

Similar to bulk RNA-seq, the existence of a variety of scRNA-seq platforms and protocols presents a hurdle for computational method development and cross-validation across datasets. Several published reviews have compared a portion of these platforms [30, 76, 108, 164, 263]. Certain data preprocessing issues are only specific to a particular type of platforms. For example, several platforms use unique molecular identifiers (UMIs) to remove polymerase chain reaction (PCR) amplification bias [134]. Preprocessing data generated by these platforms requires a step called UMI deduplication, which is to correct UMI errors that occur during amplification and sequencing. Multiple methods have been developed for this task [163, 195, 198]. Another issue is the detection of "problematic cells" including empty droplets (not an actual cell) and doublets (two cells are mistaken for one cell) in droplet-based platforms [107, 134, 258], and damaged cells in all platforms. Accordingly, multiple computational and experimental solutions have been proposed [85, 96, 200].

In bulk RNA-seq, the number of sequenced reads can vary widely among different samples. Analogously, individual cells may have vastly different numbers of sequenced reads in scRNA-seq. The reason is a combination of biological phenomena (e.g., some cells indeed have more mRNA transcripts than others) and technical artifacts (e.g., variations in cell capture efficiency). It is important to normalize scRNA-seq data so that gene expression levels are comparable across cells, a condition necessary for any downstream analyses. Existing scRNA-seq normalization methods belong to two major categories: spike-in dependent methods and direct normalization methods. In the former, spike-in RNA molecules with the same concentration are added to each cell prior to library preparation [199], and normalization is done through scaling so that spike-in read counts are equalized across cells. However, the addition of spike-in is only allowed for plate-based platforms such as STRT-seq, SMART-seq and SMART-seq2 [87, 165, 173, 206], and it does not apply to the more recently developed droplet-based platforms, which have advantages including a lower per-cell cost and a larger number of cells to sequence in parallel [107, 134]. The second and more dominant category, direct normalization methods, in contrast, do not require modification to experimental procedures and are thus more generally applicable. Direct normalization methods for scRNA-seq data are either adaptation of existing normalization methods for bulk RNA-seq data (e.g., DESeq [9], trimmed mean of M values (TMM) normalization [176], and the simple library size normalization so that all cells have the same total number of reads) or new methods that specifically account for distinct features of scRNA-seq data (e.g., excess zero counts [57, 101], more details below). Examples of new methods include scran, which uses cell pooling and subsequent deconvolution to estimate scale factors of individual cells [130], and SCnorm, which groups genes whose counts have similar dependence on sequencing depths and estimates a scaling factor for each gene group [10]. For a comprehensive review of scRNA-seq normalization methods, we refer interested readers to [217].

Finally, a concern unique in scRNA-seq data analysis is the presence of excess zero counts. This can be caused by a technical artefact, known as the "dropout" phenomenon, in which a gene is observed at a moderate expression level in one cell but is undetected in another cell of the same type [101, 108]. Dropouts occur because the current technologies do not reliably and consistently detect low levels of RNA, and consequently, genes may incorrectly appear to be inactive. Dropouts appear as excess zero or low counts in scRNA-seq data, obscuring downstream analyses such as the identification of differentially expressed genes between cell types and the inference of gene networks. To address this issue, multiple methods have been developed to improve the quality of scRNA-seq data from various perspectives. Examples of imputation or recovery methods include scImpute, which first identifies likely false zero and low counts and then imputes them by borrowing information from similar cells [123]; SAVER, which estimates unobserved true gene expression levels in a Bayesian model by borrowing information across genes [80]; and MAGIC, which alters gene expression levels by sharing information across similar cells based on the idea of heat diffusion [218]. A recent review of existing imputation methods is available at [254]. Alternatively, the presence of zero counts can be due to natural fluctuations in gene expression levels as cells go through different stages of the cell cycle [240].

3.2 Identification of Cell Types

After appropriate preprocessing, scRNA-seq data offer a new opportunity for inferring gene networks at the cell-type level. In order to do this, a key task is the identification of cell types, also known as cell subpopulations or cell states. There are two major approaches to identifying cell types from scRNA-seq data. The first approach leverages prior knowledge on cell-type marker genes. However, it cannot lead to the discovery of new cell types or subtypes. The second approach is based on unsupervised cell clustering. While it is useful for *de novo* discovery of new cell types and subtypes, unsupervised learning depends on many user-specific inputs, including which clustering algorithm to use (e.g., K-means clustering, hierarchical clustering, density-based clustering or graph-based clustering), the type of similarity or distance metric between two cells, and the number of clusters, which is a key parameter needed for many clustering algorithms. Taking into account the distinct features of scRNA-seq data, multiple cell clustering algorithms have been developed, including SNN-Cliq, which does not use conventional similarity measures but leverages the ranking of cells to construct a cell-cell graph for identifying cell clusters [244]; BiSNN-Walk, which extends SNN-Cliq and uses an iterative biclustering approach to return a ranked list of cell clusters, each associated with a set of ranked genes based on their levels of affiliation with the cluster [192]; CIDR, the first clustering method that incorporates imputation of dropout gene expression levels [125]; SC3, a widely-used ensemble method that combines multiple clustering algorithms [106]; and Seurat, which identifies cell clusters based on a shared nearest neighbor (SNN) clustering algorithm [184]. In addition to commonly used similarity metric including the Pearson correlation, Spearman correlation, Euclidean distance, other cell similarity measures can be found in, for example, [91, 186]. An evaluation study that compares multiple clustering methods is available in [48]. For a recent review of methods and challenges in unsupervised clustering of scRNA-seq data, please refer to [105].

3.3 Inference of Cell-Type-Specific Gene Networks and Its Challenges

Having identified cell types, one possible approach to gene network inference is to use existing inference or construction methods within each cell type (e.g., SINCERA [74]). Several other methods have been developed to incorporate scRNA-seq data characteristics. One study inferred gene co-expression networks by identifying significant pairwise gene associations using both continuous and binary components of linearly transformed scRNA-seq gene expression data [166]. Another study used Boolean regulatory network models with discretized single-cell expression profiles to construct a network of 20 transcription factors (TFs), which predicts direct regulation of the TF *Erg* in early blood development of mouse embryos [145]. More recently, SCENIC is a computational method that simultaneously reconstructs gene regulatory networks and identifies cell types from scRNA-seq data [4]. SCENIC defines regulons as TF-target gene (TG) co-expression modules with TF motif enrichment, and it then calculates regulon activity scores, which are robust against dropouts, for downstream analyses. PIDC is another computational method that infers gene regulatory networks using a multivariate information measure based on partial information decomposition, which captures higher-order information than pairwise mutual information [28]. PIDC is enabled by the large sample size (i.e., number of cells) in scRNA-seq data. These network inference methods together facilitate the investigation of gene regulatory relationships at the cell-type level.

Despite the existence of many gene network inference methods, scRNA-seq data still call for new computational methods for specific network analysis tasks. Given the high level of noise and excess zeros in scRNA-seq data, one limitation of many inferred cell-type-specific gene networks is that they are typically small in size and require TF-TG pairing information. Another difficulty in inferring cell-type-specific gene networks can arise when the subpopulation of cells under consideration has relatively small number of cells, especially for novel cell types. One reasonable approach is to consider joint network inference for multiple cell types, which borrows information across cell types to achieve more accurate inference. Relevant statistical approaches include a method that infers multiple Gaussian graphical models (GGMs) with a joint sparsity constraint [35] and a Bayesian nonparametric dynamic Poisson graphical model that combines information across biological conditions for joint inference of TF coactivation networks [131]. In some cases, there is a known or inferred temporal or spatial structure of cell types, such as a reconstructed cell lineage or pseudotime by computational methods including Monocle [215], Waterfull [193], Wishbone [188], TSCAN [90], Monocle2 [171], Slingshot [201] and CellRouter [38]. To incorporate such a cell type structure into network inference, one may leverage existing statistical methods in the network analysis literature, for example, a Bayesian neighborhood selection method that jointly estimates multiple GGMs with a spatial and/or temporal structure among these GGMs [126] and a group-fused graphical Lasso method for estimating piecewise constant time-evolving GGMs.

The availability of cell-type-specific gene networks has opened up new grounds for applications of statistical network inference. For example, an important statistical question is how to test for the differences between two networks or among multiple networks along a spatial or

temporal trajectory. For this task, there are multiple differential network analysis approaches that have been applied to studying protein-protein interaction networks and gene networks constructed from bulk gene expression data [8, 66, 68, 75, 83, 132, 253]. A new statistical challenge in extending these existing methods to scRNA-seq data is how to incorporate the uncertainty in cell type identification and/or cell type trajectory reconstruction.

As we discussed in Section 2.1, how to fairly evaluate network inference methods remains a critical challenge for computational biologists. Multiple steps can affect the network inference results, including the aforementioned complex data preprocessing, the choice of nodes (what genes to include), and the definition of edges (marginal or conditional associations, linear or nonlinear associations, directed or undirected associations, etc.). The lack of proper benchmarking data is the key reason behind this challenge, and it is necessary to have joint efforts with experimentalists to design reasonable benchmark standards.

4. BRAIN NETWORKS

Human brain is a complex, interconnected network. The study of brain is another important area where network analysis tools have proven extremely useful. One popular type of analysis is brain connectivity analysis, which aims to provide an accurate and informative mapping and signal extraction of the human brain by analyzing connectivities between different neurons or brain regions [19]. Results from such analysis can lead to crucial insights of pathologies of neurological disorders. For example, increasing amount of evidence suggests that compared to a healthy brain, a connectivity network changes in the presence of numerous neurological disorders [59]. There has been a fast development of brain connectivity analysis using graph theoretical tools [58]. At the heart of this endeavor is the notion that brain connectivity can be abstracted to a graph, with nodes representing neural elements, for example, neurons or brain regions, and links representing some measure of structural, functional or causal interaction between nodes. Such a representation brings the rich repository of graph theory and tools to the realm of brain connectivity analysis to characterize diverse anatomical, functional and dynamical properties of brain networks.

4.1 Basics

A graphical analysis of brain connectivity starts with defining nodes. This step is crucial and nontrivial, with different ways of defining nodes at different resolutions. At the microscopic level, nodes are neurons, with the number of neurons ranging at the order of 10^{13} to 10^{14}. At the macroscopic level, nodes can be individual image voxels, with the number of voxels ranging at the order of 10^5 to 10^7, or can be spatial brain regions-of-interest (ROIs), with the number of ROIs ranging at the order of 10^2. The ROIs can be defined anatomically, according to a brain atlas that is built on prior anatomical information such as sclcal and gyral landmarks [43, 216], or can be defined functionally, based on prior functional information such as coordinates of peak activation [47]. More recently, there have been proposals to parcellate the brain and define the ROIs according to data-driven clustering of resting-state functional or diffusion-weighted imaging measures [167].

The next step is to determine the edges between nodes, and we discuss network edge estimation in Section 4.2. It is useful to recognize that there are three broad classes of brain connectivity one can consider to define edges: functional connectivity, structural connectivity and effective connectivity [62, 63]. Different classes lead to different ways of defining network edges. Simply speaking, functional connectivity refers to statistical correlations and dependencies between spatially distinct neurophysiological recordings of brain activities. Structural connectivity refers to the anatomical connections and physical wirings between brain regions. Effective connectivity refers to the causal influence exerted amongst neural systems.

In this review, we primarily focus on statistical methods for functional connectivity analysis where the nodes are predefined brain regions. This is the area that has probably been most intensively studied in both neuroscience and statistics. See [196] for a recent review on functional connectivity analysis, and a discussion on blind spots and breakthroughs. We only briefly discuss structural connectivity analysis and effective connectivity analysis. Even though we attempt to cover a wide range of methods, we are sure to miss some important papers. See [220] and [58] for more discussion on node and edge definitions in brain connectivity analysis.

4.2 Network Estimation

Functional connectivity analysis. Two mainstream imaging modalities to study brain functional connectivity are electroencephalography (EEG) [153, 155, 224], and resting-state functional magnetic resonance imaging (fMRI) [113, 127]. For each study subject, EEG records the voltage values of multiple electrodes placed at various scalp locations over time, producing a spatial by temporal data matrix that can be used for downstream analysis. Resting-state fMRI measures changes in blood flow and oxygenation at individual voxels of brain over time, yielding a 4-way data array, which needs preprocessing first. Following a prespecified brain region parcellation, the time course data of the voxels within the same region are summarized, most often averaged, to represent the signal of that region. Alternatively, instead of using simple averages, [97] on proposed to use kernel canonical correlation coefficient between all the voxels from the two

regions to define the strength of connectivity. For both modalities, the resulting data is a spatial (location/region) temporal (time) matrix for each individual subject, from which a functional connectivity network is estimated. The most commonly used approach to construct a connectivity network is to treat the time series data of each spatial location as repeated measures to compute marginal correlations between every pair of nodes/locations. Some of these methods, for example, the Pearson correlation coefficient [71] and partial correlation [22, 180, 226], have been discussed in Section 2.1 in the context of gene networks. In addition to various connectivity network estimation solutions, [241] developed a formal inference approach to explicitly quantify the significance of individual links in a connectivity network. They adopted the matrix normal distribution, formulated the problem as precision matrix testing, and controlled the false discovery of multiple testing.

Alternatively, [168] treated the time-course data as continuous random functions, and developed a functional graphical model to estimate the connectivity network, based on functional conditional independence under a functional normal assumption. [118] further relaxed the normality assumption, and proposed the notion of functional additive conditional independence as a criterion for constructing functional graphical models. Their method requires neither parametric assumption, nor high-dimensional kernels, and thus avoids the curse of dimensionality and is able to scale to large networks.

In addition to measuring the correlation of two time series in the time domain, which shows how the signal changes over time, one can also measure the correlation in the frequency domain, which shows the signal within each given frequency band over a range of frequencies. Such frequency domain analyses help address two problems existing in time domain analyses: temporal inconsistency and noise sensitivity. Coherence is one correlation measure in the frequency domain, which is the analog of cross correlation in the time domain, and is a temporally invariant frequency-specific measure of linear association between signals. [55] studied the EEG data and used partial coherence as the measure of functional connectivity, which identifies the frequency bands that drive the direct linear association between any pair of nodes. They developed a generalized shrinkage estimator, a weighted average of a parametric and a nonparametric estimator, of the partial coherence matrix. Moreover, [122] employed time-series, clustering and functional data analysis to study spectral synchronicity and functional connectivity also using EEG data. [182] discussed using mutual information and partial mutual information to estimate functional connectivity network, and [26] further extended the method. Similar ideas apply to fMRI data as well; see [3].

Dynamic functional connectivity. Traditionally, functional connectivity analysis based on resting-state fMRI assumes that the functional connectivity network is static. Consequently, one often aggregates the time-course data over the entire duration of the scanning session and obtains a single estimate of the connectivity network. In recent years, emerging evidences have suggested that the network very likely changes dynamically over the scan [23]. [7] proposed to assess the functional connectivity dynamics based on spatial independent component analysis, sliding time window correlation, and k-means clustering of the windowed correlation matrices. [37] developed a dynamic connectivity regression approach to detect temporal change points in functional connectivity. [246] further extended this approach to handle large networks. [204] proposed a structured tensor factorization approach that encourages sparsity and smoothness in parameters along the specified tensor modes. They then built a dynamic tensor clustering method, and applied to brain dynamic functional connectivity analysis.

[170] developed a method to estimate individual graph given an external variable, for example, age and proposed a multistep procedure. They first obtained the sample covariance matrix estimates at the observed values of the external variable. They then constructed a smoothed covariance estimate through kernel smoothing for any value of the external variable in between. Finally, they plugged the covariance estimate into a sparse precision matrix estimation method such as CLIME in [22].

Beyond functional connectivity. As we have mentioned previously, aside from functional connectivity, structural connectivity and effective connectivity have been also been considered in the literature. While a vast number of papers on each of the topic is available, due to space limit, we only briefly discuss a few. More details can be found in the references therein.

Structural connectivity analysis aims to reconstruct white matter fiber tracts, which are large axonal bundles with similar destinations, in brain. Such a white fiber structure serves as a proxy to brain anatomical structure. It is indicative of brain abnormality in white matter due to axonal loss or deformation, and is thought to be related to many neural degenerative diseases. The white fiber structure can be deduced from the diffusion characteristics of water molecules in brain, as water tends to diffuse faster along the fiber bundles. Diffusion tensor imaging (DTI) is an in vivo and noninvasive medical imaging technology that measures water diffusion in brain. [239] developed a method for fiber direction estimation, smoothing and tracking. [256] developed a way to utilize multiple white matter features to construct structural connectivity across subjects.

Effective connectivity aims to model causal relationships between brain neurons or regions, and refers explicitly to the influence that one neural system exerts over another, either at a synaptic or population level. See [63] for a review. While functional connectivity is often encoded by an undirected graph, effective connectivity is encoded by a directed graph. Two common classes of effective connectivity modeling approaches are dynamic causal modeling (DCM, [64]), and structural equation modeling [141]. Notably, the DCM approach utilizes ordinary differential equation models for the neural dynamics and hemodynamic response. However, it is often computationally expensive and is often restricted to a relatively small number of nodes. [95] proposed a spatio-spectral mixed-effects model for effective connectivity analysis using task-based fMRI. [255] developed a dynamic directional model with block structure for effective connectivity using electrocorticographic (ECoG) data. [24] developed a causal dynamic network model to estimate brain activation and connection also using task-based fMRI data.

4.3 Network Comparison

Accumulated evidences have indicated that, compared to a healthy brain, a connectivity network alters in the presence of numerous neurological disorders, including Alzheimer's disease, attention deficit hyperactivity disorder, autism spectrum disorder, and many others [78, 179, 211]. Such alternations in brain connectivity are associated with cognitive and behavioral functions, and hold crucial insights of pathologies of neurological disorders [59]. As such, it is of paramount importance to compare brain connectivity networks under different physiological conditions, for example, the disorder diagnostic status.

The first question is to estimate multiple brain connectivity networks jointly under different conditions. [262] modeled the spatial temporal data as matrix-valued normal, then proposed a nonconvex penalization to simultaneously estimate multiple networks coded by precision matrices under different conditions. They assumed that not only each individual precision matrix is sparse, but also the difference of the precision matrices across the conditions is sparse. Both types of sparsity are biologically sensible. [3] approached the problem in the frequency domain, and developed a sparse reduced rank modeling framework for functional connectivity analysis across multiple groups.

The second question is to carry out formal statistical inference to compare brain connectivity networks under different conditions. [102] tackled this problem by first summarizing the network through a set of network metrics, then employing a standard two-sample test. This strategy is commonly employed in the neuroscience literature and is easy to implement. However, it remains unclear to what extent each network metric provides a meaningful representation of brain function and structure [58]. [33] developed a method to detect differentially expressed connectivity subnetworks under different conditions, by searching clusters of the graph, and resorting to a permutation test to obtain the p-value of the selected subnetwork. [49] developed a fully Bayesian solution for network comparison, under a series of prior distributions, and the solution is very flexible. [150] turned the matrix data into vector normal by whitening, and used bootstrap resampling method for inference. [242] adopted the matrix normal distribution, and developed an inferential procedure for testing equality of individual entries of partial correlation matrices across multiple groups while controlling for false discovery.

In addition to the element-wise comparison of multiple networks, there is another family of methods that use persistent homology and are built to take into account the network topology. Homology is an algebraic formalism to associate a sequence of objects with a topological space. Persistent homology is a technique of computational topology that charts the changes in topological network features over multiple resolutions and scales. In doing so, it reveals the most persistent topological features that are robust to noise. See [115, 169, 196] for more details.

5. CONCLUDING REMARKS

In this review, using gene networks and brain networks as primary examples, we have discussed statistical methods for constructing networks and how biological knowledge can be extracted from network topology. It is also possible to bring other biological covariates into the network analysis; one popular such example is to associate the estimated brain connectivity network with external phenotypes. [110] proposed a semiparametric Bayesian conditional graphical model for joint selection of important neuroimaging biomarkers such as the brain functional connectivity, as well as significant genetic biomarkers. [203] developed a class of tensor response regression models that associate a symmetric correlation matrix with a set of covariates such as age and sex. [260] developed intrinsic regression models that associate the diffusion tensor from structural connectivity analysis with the covariates.

It is useful to note that gene coexpression networks and brain connectivity networks share some conceptual similarities, both using some correlation measure to represent edges. However, they differ in how repeated measures are taken. In the latter, the replications are repeated measures of the time series, and a single correlation network can be constructed for every single subject/sample. Thus typically in a brain network study, multiple networks are

available for statistical analysis. In the case of gene networks with microarray or bulk RNA-seq data, the replications are individual samples, and usually only a single correlation network is constructed across all the samples. As discussed in Section 3, this perspective is changing with the availability of scRNA-seq, which allows for a network to be constructed for each cell type.

ACKNOWLEDGMENTS

Y. X. R. Wang is supported by the ARC DECRA Fellowship DE180101252.

L. Li is supported by NSF Grant DMS-1613137; NIH Grants R01AG034570 and R01AG061303.

J. J. Li is supported by NSF Grant DBI-1846216; NIH/NIGMS Grant R01GM120507; Johnson & Johnson WiSTEM2D Award; Sloan Research Fellowship.

REFERENCES

[1] Aach, J. and Church, G. M. (2001). Aligning gene expression time series with time warping algorithms. *Bioinformatics* **17** 495–508. https://doi.org/10.1093/bioinformatics/17.6.495

[2] Aghdam, R., Ganjali, M., Zhang, X. and Eslahchi, C. (2015). CN: A consensus algorithm for inferring gene regulatory networks using the SORDER algorithm and conditional mutual information test. *Mol. BioSyst.* **11** 942–949.

[3] Ahn, M., Shen, H., Lin, W. and Zhu, H. (2015). A sparse reduced rank framework for group analysis of functional neuroimaging data. *Statist. Sinica* **25** 295–312. MR3328816

[4] Aibar, S., González-Blas, C. B., Moerman, T., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J. et al. (2017). SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14** 1083–1086.

[5] Airoldi, E. M., Blei, D. M., Fienberg, S. E. and Xing, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

[6] Aittokallio, T. and Schwikowski, B. (2006). Graph-based methods for analysing networks in cell biology. *Brief. Bioinform.* **7** 243–255.

[7] Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T. and Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* **24** 663–676.

[8] Amar, D., Safer, H. and Shamir, R. (2013). Dissection of regulatory networks that are altered in disease via differential co-expression. *PLoS Comput. Biol.* **9** Art. ID e1002955. https://doi.org/10.1371/journal.pcbi.1002955

[9] Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* **11** Art. ID R106. https://doi.org/10.1186/gb-2010-11-10-r106

[10] Bacher, R., Chu, L.-F., Leng, N., Gasch, A. P., Thomson, J. A., Stewart, R. M., Newton, M. and Kendziorski, C. (2016). SCnorm: A quantile-regression based approach for robust normalization of single-cell RNA-seq data. bioRxiv 090167. https://doi.org/10.1101/090167

[11] Bar-Joseph, Z., Gerber, G. K., Lee, T. I., Rinaldi, N. J., Yoo, J. Y., Robert, F., Gordon, D. B., Fraenkel, E., Jaakkola, T. S. et al. (2003). Computational discovery of gene modules and regulatory networks. *Nat. Biotechnol.* **21** 1337–1342.

[12] Bar-Joseph, Z., Gitter, A. and Simon, I. (2012). Studying and modelling dynamic biological processes using time-series gene expression data. *Nat. Rev. Genet.* **13** 552–564. https://doi.org/10.1038/nrg3244

[13] Barabási, A.-L., Gulbahce, N. and Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12** 56–68. https://doi.org/10.1038/nrg2918

[14] Barabási, A.-L. and Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5** 101–113. https://doi.org/10.1038/nrg1272

[15] Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R. and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37** 382–390.

[16] Ben-Dor, A., Shamir, R. and Yakhini, Z. (1999). Clustering gene expression patterns. *J. Comput. Biol.* **6** 281–297.

[17] Berg, J. and Lässig, M. (2004). Local graph alignment and motif search in biological networks. *Proc. Natl. Acad. Sci. USA* **101** 14689–14694.

[18] Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., Teichmann, S. A., Marioni, J. C. and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33** 155–160. https://doi.org/10.1038/nbt.3102

[19] Bullmore, E. and Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev., Neurosci.* **10** 186–198. https://doi.org/10.1038/nrn2575

[20] Butte, A. J. and Kohane, I. S. (1999). Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. In *Biocomputing* 2000 418–429. World Scientific, Singapore.

[21] Cabreros, I., Abbe, E. and Tsirigos, A. (2016). Detecting community structures in hi-c genomic data. In 2016 *Annual Conference on Information Science and Systems* (*CISS*) 584–589. IEEE, Los Alamitos, CA.

[22] Cai, T., Liu, W. and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973 https://doi.org/10.1198/jasa.2011.tm10155

[23] Calhoun, V. D., Miller, R., Pearlson, G. and Adali, T. (2014). The chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron* **84** 262–274.

[24] Cao, X., Sandstede, B. and Luo, X. (2019). A functional data method for causal dynamic network modeling of task-related fMRI. *Front. Neurosci.* **13** Art. ID 127. https://doi.org/10.3389/fnins.2019.00127

[25] Carter, S. L., Brechbühler, C. M., Griffin, M. and Bond, A. T. (2004). Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* **20** 2242–2250.

[26] Cassidy, B., Rae, C. and Solo, V. (2014). Brain activity: Connectivity, sparsity, and mutual information. *IEEE Trans. Med. Imag.* **34** 846–860.

[27] Chahrour, M., Jung, S. Y., Shaw, C., Zhou, X., Wong, S. T., Qin, J. and Zoghbi, H. Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320** 1224–1229.

[28] Chan, T. E., Stumpf, M. P. and Babtie, A. C. (2017). Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.* **5** 251–267.

[29] CHAVALI, S., BARRENAS, F., KANDURI, K. and BENSON, M. (2010). Network properties of human disease genes with pleiotropic effects. *BMC Syst. Biol.* **4** Art. ID 78. https://doi.org/10.1186/1752-0509-4-78

[30] CHEN, G., NING, B. and SHI, T. (2019). Single-cell RNA-seq technologies and related computational data analysis. *Front. Genet.* **10** Art. ID 317. https://doi.org/10.3389/fgene.2019.00317

[31] CHEN, J., BARDES, E. E., ARONOW, B. J. and JEGGA, A. G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37** W305–W311. https://doi.org/10.1093/nar/gkp427

[32] CHEN, K.-C., WANG, T.-Y., TSENG, H.-H., HUANG, C.-Y. F. and KAO, C.-Y. (2005). A stochastic differential equation model for quantifying transcriptional regulatory network in *Saccharomyces cerevisiae*. *Bioinformatics* **21** 2883–2890.

[33] CHEN, S., KANG, J., XING, Y. and WANG, G. (2015). A parsimonious statistical method to detect groupwise differentially expressed functional connectivity networks. *Hum. Brain Mapp.* **36** 5196–5206.

[34] CHICKERING, D. M., HECKERMAN, D. and MEEK, C. (2004). Large-sample learning of Bayesian networks is NP-hard. *J. Mach. Learn. Res.* **5** 1287–1330. MR2248018

[35] CHUN, H., ZHANG, X. and ZHAO, H. (2015). Gene regulation network inference with joint sparse Gaussian graphical models. *J. Comput. Graph. Statist.* **24** 954–974. MR3432924 https://doi.org/10.1080/10618600.2014.956876

[36] CIRIELLO, G. and GUERRA, C. (2008). A review on models and algorithms for motif discovery in protein–protein interaction networks. *Brief. Funct. Genomics Proteomics* **7** 147–156.

[37] CRIBBEN, I., HARALDSDOTTIR, R., ATLAS, L. Y., WAGER, T. D. and LINDQUIST, M. A. (2012). Dynamic connectivity regression: Determining state-related changes in brain connectivity. *NeuroImage* **61** 907–920.

[38] DA ROCHA, E. L., ROWE, R. G., LUNDIN, V., MALLESHAIAH, M., JHA, D. K., RAMBO, C. R., LI, H., NORTH, T. E., COLLINS, J. J. et al. (2018). Reconstruction of complex single-cell trajectories using CellRouter. *Nat. Commun.* **9** Art. ID 892.

[39] DAUB, C. O., STEUER, R., SELBIG, J. and KLOSKA, S. (2004). Estimating mutual information using B-spline functions—An improved similarity measure for analysing gene expression data. *BMC Bioinform.* **5** Art. ID 118. https://doi.org/10.1186/1471-2105-5-118

[40] DAUDIN, J.-J., PICARD, F. and ROBIN, S. (2008). A mixture model for random graphs. *Stat. Comput.* **18** 173–183. MR2390817 https://doi.org/10.1007/s11222-007-9046-7

[41] DE LA FUENTE, A., BING, N., HOESCHELE, I. and MENDES, P. (2004). Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics* **20** 3565–3574.

[42] DEMBÉLÉ, D. and KASTNER, P. (2003). Fuzzy C-means method for clustering microarray data. *Bioinformatics* **19** 973–980.

[43] DESIKAN, R. S., SÉGONNE, F., FISCHL, B., QUINN, B. T., DICKERSON, B. C., BLACKER, D., BUCKNER, R. L., DALE, A. M., MAGUIRE, R. P. et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31** 968–980.

[44] DI GESÚ, V., GIANCARLO, R., BOSCO, G. L., RAIMONDI, A. and SCATURRO, D. (2005). GenClust: A genetic algorithm for clustering gene expression data. *BMC Bioinform.* **6** Art. ID 289.

[45] DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. and REN, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485** 376–380.

[46] DIXON, S. J., COSTANZO, M., BARYSHNIKOVA, A., ANDREWS, B. and BOONE, C. (2009). Systematic mapping of genetic interaction networks. *Annu. Rev. Genet.* **43** 601–625.

[47] DOSENBACH, N. U., NARDOS, B., COHEN, A. L., FAIR, D. A., POWER, J. D., CHURCH, J. A., NELSON, S. M., WIG, G. S., VOGEL, A. C. et al. (2010). Prediction of individual brain maturity using fMRI. *Science* **329** 1358–1361.

[48] DUÒ, A., ROBINSON, M. D. and SONESON, C. (2018). A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Res.* **7** Art. ID 1141. https://doi.org/10.12688/f1000research.15666.2

[49] DURANTE, D. and DUNSON, D. B. (2018). Bayesian inference and testing of group differences in brain networks. *Bayesian Anal.* **13** 29–58. MR3737942 https://doi.org/10.1214/16-BA1030

[50] EBERWINE, J., SUL, J.-Y., BARTFAI, T. and KIM, J. (2014). The promise of single-cell sequencing. *Nat. Methods* **11** 25–27.

[51] EISEN, M. B., SPELLMAN, P. T., BROWN, P. O. and BOTSTEIN, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95** 14863–14868.

[52] ENRIGHT, A. J., VAN DONGEN, S. and OUZOUNIS, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30** 1575–1584.

[53] ESTRADA, E. (2006). Virtual identification of essential proteins within the protein interaction network of yeast. *Proteomics* **6** 35–40. https://doi.org/10.1002/pmic.200500209

[54] FENG, J., BARBANO, P. E. and MISHRA, B. (2004). Time-frequency feature detection for time-course microarray data. In *Proceedings of the* 2004 *ACM Symposium on Applied Computing* 128–132. ACM, New York.

[55] FIECAS, M. and OMBAO, H. (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *Ann. Appl. Stat.* **5** 1102–1125. MR2840188 https://doi.org/10.1214/10-AOAS396

[56] FIERS, M. W. E. J., MINNOYE, L., AIBAR, S., GONZÁLEZ-BLAS, C. B., ATAK, Z. K. and AERTS, S. (2018). Mapping gene regulatory networks from single-cell omics data. *Brief. Funct. Genomics* **17** 246–254. https://doi.org/10.1093/bfgp/elx046

[57] FINAK, G., MCDAVID, A., YAJIMA, M., DENG, J., GERSUK, V., SHALEK, A. K., SLICHTER, C. K., MILLER, H. W., MCELRATH, M. J. et al. (2015). MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16** Art. ID 278.

[58] FORNITO, A., ZALESKY, A. and BREAKSPEAR, M. (2013). Graph analysis of the human connectome: Promise, progress, and pitfalls. *NeuroImage* **80** 426–444.

[59] FOX, M. D. and GREICIUS, M. (2010). Clinical applications of resting state functional connectivity. *Front. Syst. Neurosci.* **4** Art. ID 19.

[60] FRASER, H. B., HIRSH, A. E., STEINMETZ, L. M., SCHARFE, C. and FELDMAN, M. W. (2002). Evolutionary rate in the protein interaction network. *Science* **296** 750–752.

[61] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

[62] FRISTON, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Hum. Brain Mapp.* **2** 56–78.

[63] FRISTON, K. J. (2011). Functional and effective connectivity: A review. *Brain Connect*. **1** 13–36. https://doi.org/10.1089/brain.2011.0008

[64] FRISTON, K. J., HARRISON, L. and PENNY, W. (2003). Dynamic causal modelling. *NeuroImage* **19** 1273–1302.

[65] FU, L. and MEDICO, E. (2007). FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. *BMC Bioinform*. **8** Art. ID 3. https://doi.org/10.1186/1471-2105-8-3

[66] GAMBARDELLA, G., MORETTI, M. N., DE CEGLI, R., CARDONE, L., PERON, A. and DI BERNARDO, D. (2013). Differential network analysis for the identification of condition-specific pathway activity and regulation. *Bioinformatics* **29** 1776–1785.

[67] GAO, L., SUN, P.-G. and SONG, J. (2009). Clustering algorithms for detecting functional modules in protein interaction networks. *J. Bioinform. Comput. Biol*. **7** 217–242.

[68] GILL, R., DATTA, S. and DATTA, S. (2010). A statistical framework for differential network analysis from microarray data. *BMC Bioinform*. **11** Art. ID 95. https://doi.org/10.1186/1471-2105-11-95

[69] GOLTSEV, Y. and PAPATSENKO, D. (2009). Time warping of evolutionary distant temporal gene expression data based on noise suppression. *BMC Bioinform*. **10** Art. ID 353. https://doi.org/10.1186/1471-2105-10-353

[70] GREENE, C. S., KRISHNAN, A., WONG, A. K., RICCIOTTI, E., ZELAYA, R. A., HIMMELSTEIN, D. S., ZHANG, R., HARTMANN, B. M., ZASLAVSKY, E. et al. (2015). Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet*. **47** 569–576.

[71] GREICIUS, M. D., KRASNOW, B., REISS, A. L. and MENON, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. USA* **100** 253–258.

[72] GUALA, D. and SONNHAMMER, E. L. L. (2017). A large-scale benchmark of gene prioritization methods. *Sci. Rep*. **7** Art. ID 46598. https://doi.org/10.1038/srep46598

[73] GUIMERA, R. and AMARAL, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature* **433** 895–900.

[74] GUO, M., WANG, H., POTTER, S. S., WHITSETT, J. A. and XU, Y. (2015). SINCERA: A pipeline for single-cell RNA-seq profiling analysis. *PLoS Comput. Biol*. **11** Art. ID e1004575.

[75] HA, M. J., BALADANDAYUTHAPANI, V. and DO, K.-A. (2015). DINGO: Differential network analysis in genomics. *Bioinformatics* **31** 3413–3420.

[76] HAQUE, A., ENGEL, J., TEICHMANN, S. A. and LÖNNBERG, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Gen. Med*. **9** Art. ID 75. https://doi.org/10.1186/s13073-017-0467-4

[77] HAUSER, A. and BÜHLMANN, P. (2015). Jointly interventional and observational data: Estimation of interventional Markov equivalence classes of directed acyclic graphs. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **77** 291–318. MR3299409 https://doi.org/10.1111/rssb.12071

[78] HEDDEN, T., VAN DIJK, K. R., BECKER, J. A., MEHTA, A., SPERLING, R. A., JOHNSON, K. A. and BUCKNER, R. L. (2009). Disruption of functional connectivity in clinically normal older adults harboring amyloid burden. *J. Neurosci*. **29** 12686–12694.

[79] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

[80] HUANG, M., WANG, J., TORRE, E., DUECK, H., SHAFFER, S., BONASIO, R., MURRAY, J. I., RAJ, A., LI, M. et al. (2018). SAVER: Gene expression recovery for single-cell RNA sequencing. *Nat. Methods* **15** 539–542.

[81] HUANG, S., JIN, J. and YAO, Z. (2016). Partial correlation screening for estimating large precision matrices, with applications to classification. *Ann. Statist*. **44** 2018–2057. MR3546442 https://doi.org/10.1214/15-AOS1392

[82] HUSMEIER, D. (2003). Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics* **19** 2271–2282.

[83] IDEKER, T. and KROGAN, N. J. (2012). Differential network biology. *Mol. Syst. Biol*. **8** Art. ID 565.

[84] IDEKER, T. and SHARAN, R. (2008). Protein networks in disease. *Genome Res*. **18** 644–652.

[85] ILICIC, T., KIM, J. K., KOLODZIEJCZYK, A. A., BAGGER, F. O., MCCARTHY, D. J., MARIONI, J. C. and TEICHMANN, S. A. (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biol*. **17** Art. ID 29.

[86] IRRTHUM, A., WEHENKEL, L., GEURTS, P. et al. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5** Art. ID e12776.

[87] ISLAM, S., KJÄLLQUIST, U., MOLINER, A., ZAJAC, P., FAN, J.-B., LÖNNERBERG, P. and LINNARSSON, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res*. **21** 1160–1167.

[88] ISLAM, S., ZEISEL, A., JOOST, S., MANNO, G. L., ZAJAC, P., KASPER, M., LÖNNERBERG, P. and LINNARSSON, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11** 163–166. https://doi.org/10.1038/nmeth.2772

[89] JEONG, H., MASON, S. P., BARABÁSI, A-L. and OLTVAI, Z. N. (2001). Lethality and centrality in protein networks. *Nature* **411** 41–42.

[90] JI, Z. and JI, H. (2016). TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res*. **44** Art. ID e117. https://doi.org/10.1093/nar/gkw430

[91] JIANG, H., SOHN, L. L., HUANG, H. and CHEN, L. (2018). Single cell clustering based on cell-pair differentiability correlation and variance analysis. *Bioinformatics* **34** 3684–3694.

[92] JIANG, R., TU, Z., CHEN, T. and SUN, F. (2006). Network motif identification in stochastic networks. *Proc. Natl. Acad. Sci. USA* **103** 9404–9409.

[93] JONSSON, P. F. and BATES, P. A. (2006). Global topological features of cancer proteins in the human interactome. *Bioinformatics* **22** 2291–2297.

[94] JOY, M. P., BROCK, A., INGBER, D. E. and HUANG, S. (2005). High-betweenness proteins in the yeast protein interaction network. *BioMed Res. Int*. **2005** 96–103.

[95] KANG, H., OMBAO, H., LINKLETTER, C., LONG, N. and BADRE, D. (2012). Spatio-spectral mixed-effects model for functional magnetic resonance imaging data. *J. Amer. Statist. Assoc*. **107** 568–577. MR2980068 https://doi.org/10.1080/01621459.2012.664503

[96] KANG, H. M., SUBRAMANIAM, M., TARG, S., NGUYEN, M., MALISKOVA, L., MCCARTHY, E., WAN, E., WONG, S., BYRNES, L. et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol*. **36** 89–94.

[97] KANG, J., BOWMAN, F. D., MAYBERG, H. and LIU, H. (2016). A depression network of functionally connected regions discovered via multi-attribute canonical correlation graphs. *NeuroImage* **141** 431–441.

[98] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E*

(3) **83** Art. ID 016107. MR2788206 https://doi.org/10.1103/PhysRevE.83.016107

[99] KASHTAN, N., ITZKOVITZ, S., MILO, R. and ALON, U. (2004). Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics* **20** 1746–1758.

[100] KATO, M., TSUNODA, T. and TAKAGI, T. (2001). Lag analysis of genetic networks in the cell cycle of budding yeast. *Genome Inform*. **12** 266–267.

[101] KHARCHENKO, P. V., SILBERSTEIN, L. and SCADDEN, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11** 740–742. https://doi.org/10.1038/nmeth.2967

[102] KIM, J., WOZNIAK, J. R., MUELLER, B. A., SHEN, X. and PAN, W. (2014). Comparison of statistical tests for group differences in brain functional networks. *NeuroImage* **101** 681–694.

[103] KIM, K., JIANG, K., TENG, S. L., FELDMAN, L. J. and HUANG, H. (2012). Using biologically interrelated experiments to identify pathway genes in *Arabidopsis*. *Bioinformatics* **28** 815–822.

[104] KIM, S., IMOTO, S. and MIYANO, S. (2004). Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data. *Biosystems* **75** 57–65. https://doi.org/10.1016/j.biosystems.2004.03.004

[105] KISELEV, V. Y., ANDREWS, T. S. and HEMBERG, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* **20** 273–282. https://doi.org/10.1038/s41576-018-0088-9

[106] KISELEV, V. Y., KIRSCHNER, K., SCHAUB, M. T., ANDREWS, T., YIU, A., CHANDRA, T., NATARAJAN, K. N., REIK, W., BARAHONA, M. et al. (2017). SC3: Consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14** 483–486.

[107] KLEIN, A. M., MAZUTIS, L., AKARTUNA, I., TALLAPRAGADA, N., VERES, A., LI, V., PESHKIN, L., WEITZ, D. A. and KIRSCHNER, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161** 1187–1201.

[108] KOLODZIEJCZYK, A. A., KIM, J. K., SVENSSON, V., MARIONI, J. C. and TEICHMANN, S. A. (2015). The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58** 610–620.

[109] KUMARI, S., NIE, J., CHEN, H.-S., MA, H., STEWART, R., LI, X., LU, M.-Z., TAYLOR, W. M. and WEI, H. (2012). Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. *PLoS ONE* **7** Art. ID e50411.

[110] KUNDU, S. and KANG, J. (2016). Semiparametric Bayes conditional graphical models for imaging genetics applications. *Stat* **5** 322–337. MR3590884 https://doi.org/10.1002/sta4.119

[111] KWON, A. T., HOOS, H. H. and NG, R. (2003). Inference of transcriptional regulation relationships from gene expression data. In *Proceedings of the* 2003 *ACM Symposium on Applied Computing* 135–140. ACM, New York.

[112] LANGFELDER, P. and HORVATH, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9** Art. ID 559.

[113] LAZAR, N. (2008). *The Statistical Analysis of Functional MRI Data*. Springer, New York.

[114] LE DILY, F., BAÙ, D., POHL, A., VICENT, G. P., SERRA, F., SORONELLAS, D., CASTELLANO, G., WRIGHT, R. H., BALLARE, C. et al. (2014). Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev.* **28** 2151–2162.

[115] LEE, H., CHUNG, M. K., KANG, H., KIM, B.-N. and LEE, D. S. (2011). Discriminative persistent homology of brain networks. In 2011 *IEEE International Symposium on Biomedical Imaging*: *From Nano to Macro* 841–844. IEEE, Los Alamitos, CA.

[116] LEHNER, B., CROMBIE, C., TISCHLER, J., FORTUNATO, A. and FRASER, A. G. (2006). Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat. Genet.* **38** 896–903.

[117] LEMMENS, K., DE BIE, T., DHOLLANDER, T., DE KEERSMAECKER, S. C., THIJS, I. M., SCHOOFS, G., DE WEERDT, A., DE MOOR, B., VANDERLEYDEN, J. et al. (2009). DISTILLER: A data integration framework to reveal condition dependency of complex regulons in *Escherichia coli*. *Genome Biol.* **10** Art. ID R27.

[118] LI, B. and SOLEA, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fMRI. *J. Amer. Statist. Assoc.* **113** 1637–1655. MR3902235 https://doi.org/10.1080/01621459.2017.1356726

[119] LI, H. and GUI, J. (2005). Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks. *Biostatistics* **7** 302–317.

[120] LI, K.-C. (2002). Genome-wide coexpression dynamics: Theory and application. *Proc. Natl. Acad. Sci. USA* **99** 16875–16880.

[121] LI, K.-C., PALOTIE, A., YUAN, S., BRONNIKOV, D., CHEN, D., WEI, X., CHOI, O.-W., SAARELA, J. and PELTONEN, L. (2007). Finding disease candidate genes by liquid association. *Genome Biol.* **8** Art. ID R205.

[122] LI, Q., ŞENTÜRK, D., SUGAR, C. A., JESTE, S., DISTEFANO, C., FROHLICH, J. and TELESCA, D. (2019). Inferring brain signals synchronicity from a sample of EEG readings. *J. Amer. Statist. Assoc.* **114** 991–1001. MR4011753 https://doi.org/10.1080/01621459.2018.1518233

[123] LI, W. V. and LI, J. J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nat. Commun.* **9** Art. ID 997.

[124] LIN, A., WANG, R. T., AHN, S., PARK, C. C. and SMITH, D. J. (2010). A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. *Genome Res.* **20** 1122–1132.

[125] LIN, P., TROUP, M. and HO, J. W. K. (2017). CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biol.* **18** Art. ID 59. https://doi.org/10.1186/s13059-017-1188-0

[126] LIN, Z., WANG, T., YANG, C. and ZHAO, H. (2017). On joint estimation of Gaussian graphical models for spatial and temporal data. *Biometrics* **73** 769–779. MR3713111 https://doi.org/10.1111/biom.12650

[127] LINDQUIST, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. MR2530545 https://doi.org/10.1214/09-STS282

[128] LIU, F., ZHANG, S.-W., GUO, W.-F., WEI, Z.-G. and CHEN, L. (2016). Inference of gene regulatory network based on local Bayesian networks. *PLoS Comput. Biol.* **12** Art. ID e1005024.

[129] LIU, K., THEUSCH, E., ZHOU, Y., ASHUACH, T., DOSE, A., BICKEL, P. J., MEDINA, M. W. and HUANG, H. (2019). GeneFishing: A method to reconstruct context-specific portraits of biological processes and its application to cholesterol metabolism. *Proc. Natl. Acad. Sci. USA* **116** 18943–18950.

[130] LUN, A. T. L., BACH, K. and MARIONI, J. C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17** Art. ID 75. https://doi.org/10.1186/s13059-016-0947-7

[131] LUO, X. and WEI, Y. (2018). Nonparametric Bayesian learning of heterogeneous dynamic transcription factor networks. *Ann. Appl. Stat.* **12** 1749–1772. MR3852696 https://doi.org/10.1214/17-AOAS1129

[132] MA, C., XIN, M., FELDMANN, K. A. and WANG, X. (2014). Machine learning–based differential network analysis: A study of stress-responsive transcriptomes in *Arabidopsis*. *Plant Cell* **26** 520–537.

[133] MAATHUIS, M. H., KALISCH, M. and BÜHLMANN, P. (2009). Estimating high-dimensional intervention effects from observational data. *Ann. Statist.* **37** 3133–3164. MR2549555 https://doi.org/10.1214/09-AOS685

[134] MACOSKO, E. Z., BASU, A., SATIJA, R., NEMESH, J., SHEKHAR, K., GOLDMAN, M., TIROSH, I., BIALAS, A. R., KAMITAKI, N. et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161** 1202–1214.

[135] MAGWENE, P. M. and KIM, J. (2004). Estimating genomic coexpression networks using first-order conditional independence. *Genome Biol.* **5** Art. ID R100.

[136] MARBACH, D., COSTELLO, J. C., KÜFFNER, R., VEGA, N. M., PRILL, R. J., CAMACHO, D. M., ALLISON, K. R., ADERHOLD, A., BONNEAU, R. et al. (2012). Wisdom of crowds for robust gene network inference. *Nat. Methods* **9** 796–804.

[137] MARBACH, D., PRILL, R. J., SCHAFFTER, T., MATTIUSSI, C., FLOREANO, D. and STOLOVITZKY, G. (2010). Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. USA* **107** 6286–6291.

[138] MARBACH, D., ROY, S., AY, F., MEYER, P. E., CANDEIAS, R., KAHVECI, T., BRISTOW, C. A. and KELLIS, M. (2012). Predictive regulatory models in *Drosophila melanogaster* by integrative inference of transcriptional networks. *Genome Res.* **22** 1334–1349.

[139] MARGOLIN, A. A., NEMENMAN, I., BASSO, K., WIGGINS, C., STOLOVITZKY, G., DALLA FAVERA, R. and CALIFANO, A. (2006). ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform.* **7** Art. ID S7.

[140] MATSUMOTO, H., KIRYU, H., FURUSAWA, C., KO, M. S. H., KO, S. B. H., GOUDA, N., HAYASHI, T. and NIKAIDO, I. (2017). SCODE: An efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics* **33** 2314–2321. https://doi.org/10.1093/bioinformatics/btx194

[141] MCLNTOSH, A. R. and GONZALEZ-LIMA, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Hum. Brain Mapp.* **2** 2–22.

[142] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363 https://doi.org/10.1214/009053606000000281

[143] MEINSHAUSEN, N., HAUSER, A., MOOIJ, J. M., PETERS, J., VERSTEEG, P. and BÜHLMANN, P. (2016). Methods for causal inference from gene perturbation experiments and validation. *Proc. Natl. Acad. Sci. USA* **113** 7361–7368.

[144] MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D. and ALON, U. (2002). Network motifs: Simple building blocks of complex networks. *Science* **298** 824–827.

[145] MOIGNARD, V., WOODHOUSE, S., HAGHVERDI, L., LILLY, A. J., TANAKA, Y., WILKINSON, A. C., BUETTNER, F., MACAULAY, I. C., JAWAID, W. et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* **33** 269–276.

[146] MOREAU, Y. and TRANCHEVENT, L.-C. (2012). Computational tools for prioritizing candidate genes: Boosting disease gene discovery. *Nat. Rev. Genet.* **13** 523–536. https://doi.org/10.1038/nrg3253

[147] MOSCHOPOULOS, C. N., PAVLOPOULOS, G. A., SCHNEIDER, R., LIKOTHANASSIS, S. D. and KOSSIDA, S. (2009). GIBA: A clustering tool for detecting protein complexes. *BMC Bioinform.* **10** Art. ID S11.

[148] MURO, S., TAKEMASA, I., OBA, S., MATOBA, R., UENO, N., MARUYAMA, C., YAMASHITA, R., SEKIMOTO, M., YAMAMOTO, H. et al. (2003). Identification of expressed genes linked to malignancy of human colorectal carcinoma by parametric clustering of quantitative expression data. *Genome Biol.* **4** Art. ID R21.

[149] NAIR, A., CHETTY, M. and WANGIKAR, P. P. (2015). Improving gene regulatory network inference using network topology information. *Mol. BioSyst.* **11** 2449–2463.

[150] NARAYAN, M. (2015). Inferential methods to find differences in populations of graphical models with applications to functional connectomics. Ph.D. thesis, Rice Univ. MR3781788

[151] NEWMAN, M. E. J. (2010). *Networks*: *An Introduction*. Oxford Univ. Press, Oxford. MR2676073 https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

[152] NORTON, H. K., EMERSON, D. J., HUANG, H., KIM, J., TITUS, K. R., GU, S., BASSETT, D. S. and PHILLIPS-CREMINS, J. E. (2018). Detecting hierarchical genome folding with network modularity. *Nat. Methods* **15** 119–122.

[153] NUNEZ, M. D., NUNEZ, P. L., SRINIVASAN, R., OMBAO, H., LINQUIST, M., THOMPSON, W. and ASTON, J. (2016). Electroencephalography (EEG): Neurophysics, experimental methods, and signal processing. In *Handbook of Neuroimaging Data Analysis*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods* 175–197. CRC Press, Boca Raton, FL.

[154] OKUDA, S., YAMADA, T., HAMAJIMA, M., ITOH, M., KATAYAMA, T., BORK, P., GOTO, S. and KANEHISA, M. (2008). KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res.* **36** W423–W426. https://doi.org/10.1093/nar/gkn282

[155] OMBAO, H., SCHRODER, A. L., EUAN, C., TING, C. M. and SAMDIN, B. (2016). Advanced topics for modeling electroencephalograms. In *Handbook of Neuroimaging Data Analysis* (H. Ombao, M. Linquist, W. Thompson and J. Aston, eds.) 567–626. Chapman & Hall/CRC, Boca Raton, FL.

[156] OMRANIAN, N., ELOUNDOU-MBEBI, J. M. O., MUELLER-ROEBER, B. and NIKOLOSKI, Z. (2016). Gene regulatory network inference using fused LASSO on multiple data sets. *Sci. Rep.* **6** Art. ID 20533. https://doi.org/10.1038/srep20533

[157] OZGÜR, A., VU, T., ERKAN, G. and RADEV, D. R. (2008). Identifying gene–disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics* **24** i277–i285. https://doi.org/10.1093/bioinformatics/btn182

[158] PALADUGU, S. R., ZHAO, S., RAY, A. and RAVAL, A. (2008). Mining protein networks for synthetic genetic interactions. *BMC Bioinform.* **9** Art. ID 426. https://doi.org/10.1186/1471-2105-9-426

[159] PAVLOPOULOS, G. A., SECRIER, M., MOSCHOPOULOS, C. N., SOLDATOS, T. G., KOSSIDA, S., AERTS, J., SCHNEIDER, R. and BAGOS, P. G. (2011). Using graph theory to analyze biological networks. *BioData Min.* **4** Art. ID 10. https://doi.org/10.1186/1756-0381-4-10

[160] PEI, Y., GAO, Q., LI, J. and ZHAO, X. (2014). Identifying local co-regulation relationships in gene expression data. *J. Theoret. Biol.* **360** 200–207.

[161] PENG, J., ZHOU, N. and ZHU, J. (2009). Partial correlation estimation by joint sparse regression models. *J. Amer. Statist. Assoc.* **104** 735–746. MR2541591 https://doi.org/10.1198/jasa.2009.0126

[162] PETERS, J., BÜHLMANN, P. and MEINSHAUSEN, N. (2016). Causal inference by using invariant prediction: Identification and confidence intervals. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 947–1012. MR3557186 https://doi.org/10.1111/rssb.12167

[163] PETUKHOV, V., GUO, J., BARYAWNO, N., SEVERE, N., SCADDEN, D. T., SAMSONOVA, M. G. and KHARCHENKO, P. V. (2018). dropEst: Pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. *Genome Biol.* **19** Art. ID 78.

[164] PICELLI, S. (2017). Single-cell RNA-sequencing: The future of genome biology is now. *RNA Biol.* **14** 637–650. https://doi.org/10.1080/15476286.2016.1201618

[165] PICELLI, S., BJÖRKLUND, Å. K., FARIDANI, O. R., SAGASSER, S., WINBERG, G. and SANDBERG, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10** 1096–1098.

[166] PINA, C., TELES, J., FUGAZZA, C., MAY, G., WANG, D., GUO, Y., SONEJI, S., BROWN, J., EDÉN, P. et al. (2015). Single-cell network analysis identifies DDIT3 as a nodal lineage regulator in hematopoiesis. *Cell Rep.* **11** 1503–1510.

[167] POWER, J. D., COHEN, A. L., NELSON, S. M., WIG, G. S., BARNES, K. A., CHURCH, J. A., VOGEL, A. C., LAUMANN, T. O., MIEZIN, F. M. et al. (2011). Functional network organization of the human brain. *Neuron* **72** 665–678.

[168] QIAO, X., GUO, S. and JAMES, G. M. (2019). Functional graphical models. *J. Amer. Statist. Assoc.* **114** 211–222. MR3941249 https://doi.org/10.1080/01621459.2017.1390466

[169] QIU, A., LEE, A., TAN, M. and CHUNG, M. K. (2015). Manifold learning on brain functional networks in aging. *Med. Image Anal.* **20** 52–60.

[170] QIU, H., HAN, F., LIU, H. and CAFFO, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 487–504. MR3454206 https://doi.org/10.1111/rssb.12123

[171] QIU, X., HILL, A., PACKER, J., LIN, D., MA, Y.-A. and TRAPNELL, C. (2017). Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14** 309–315.

[172] RAMONI, M. F., SEBASTIANI, P. and KOHANE, I. S. (2002). Cluster analysis of gene expression dynamics. *Proc. Natl. Acad. Sci. USA* **99** 9121–9126. MR1909705 https://doi.org/10.1073/pnas.132656399

[173] RAMSKÖLD, D., LUO, S., WANG, Y.-C., LI, R., DENG, Q., FARIDANI, O. R., DANIELS, G. A., KHREBTUKOVA, I., LORING, J. F. et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30** 777–782.

[174] RAU, A., JAFFRÉZIC, F. and NUEL, G. (2013). Joint estimation of causal effects from observational and intervention gene expression data. *BMC Syst. Biol.* **7** Art. ID 111. https://doi.org/10.1186/1752-0509-7-111

[175] RESHEF, D. N., RESHEF, Y. A., FINUCANE, H. K., GROSSMAN, S. R., MCVEAN, G., TURNBAUGH, P. J., LANDER, E. S., MITZENMACHER, M. and SABETI, P. C. (2011). Detecting novel associations in large data sets. *Science* **334** 1518–1524.

[176] ROBINSON, M. D. and OSHLACK, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11** Art. ID R25. https://doi.org/10.1186/gb-2010-11-3-r25

[177] ROUSSEEUW, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20** 53–65.

[178] ROY, S., BHATTACHARYYA, D. K. and KALITA, J. K. (2014). Reconstruction of gene co-expression network from microarray data using local expression patterns. *BMC Bioinform.* **15** Art. ID S10. https://doi.org/10.1186/1471-2105-15-S7-S10

[179] RUDIE, J. D., BROWN, J. A., BECK-PANCER, D., HERNANDEZ, L. M., DENNIS, E. L., THOMPSON, P. M., BOOKHEIMER, S. Y. and DAPRETTO, M. J. N. C. (2013). Altered functional and structural brain network organization in autism. *NeuroImage Clin.* **2** 79–94.

[180] RYALI, S., CHEN, T., SUPEKAR, K. and MENON, V. (2012). Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *NeuroImage* **59** 3852–3861.

[181] SALIBA, A.-E., WESTERMANN, A. J., GORSKI, S. A. and VOGEL, J. (2014). Single-cell RNA-seq: Advances and future challenges. *Nucleic Acids Res.* **42** 8845–8860. https://doi.org/10.1093/nar/gku555

[182] SALVADOR, R., SUCKLING, J., SCHWARZBAUER, C. and BULLMORE, E. (2005). Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philos. Trans. R. Soc. B* **360** 937–946.

[183] SANTOS-ZAVALETA, A., SALGADO, H., GAMA-CASTRO, S., SÁNCHEZ-PÉREZ, M., GÓMEZ-ROMERO, L., LEDEZMA-TEJEIDA, D., GARCÍA-SOTELO, J. S., ALQUICIRA-HERNÁNDEZ, K., MUÑIZ-RASCADO, L. J. et al. (2018). RegulonDB v 10.5: Tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* **47** D212–D220.

[184] SATIJA, R., FARRELL, J. A., GENNERT, D., SCHIER, A. F. and REGEV, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33** 495–502. https://doi.org/10.1038/nbt.3192

[185] SATULURI, V., PARTHASARATHY, S. and UCAR, D. (2010). Markov clustering of protein interaction networks with improved balance and scalability. In *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology* 247–256. ACM, New York.

[186] SCHIFFMAN, C., LIN, C., SHI, F., CHEN, L., SOHN, L. and HUANG, H. (2017). SIDEseq: A cell similarity measure defined by shared identified differentially expressed genes for single-cell RNA sequencing data. *Stat. Biosci.* **9** 200–216.

[187] SCHULDINER, M., COLLINS, S. R., THOMPSON, N. J., DENIC, V., BHAMIDIPATI, A., PUNNA, T., IHMELS, J., ANDREWS, B., BOONE, C. et al. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123** 507–519.

[188] SETTY, M., TADMOR, M. D., REICH-ZELIGER, S., ANGEL, O., SALAME, T. M., KATHAIL, P., CHOI, K., BENDALL, S., FRIEDMAN, N. et al. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* **34** 637–645.

[189] SEXTON, T., YAFFE, E., KENIGSBERG, E., BANTIGNIES, F., LEBLANC, B., HOICHMAN, M., PARRINELLO, H., TANAY, A. and CAVALLI, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148** 458–472.

[190] SHAPIRO, E., BIEZUNER, T. and LINNARSSON, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14** 618–630. https://doi.org/10.1038/nrg3542

[191] SHARAN, R., MARON-KATZ, A. and SHAMIR, R. (2003). CLICK and EXPANDER: A system for clustering and visualizing gene expression data. *Bioinformatics* **19** 1787–1799.

[192] SHI, F. and HUANG, H. (2017). Identifying cell subpopulations and their genetic drivers from single-cell RNA-Seq data using a biclustering approach. *J. Comput. Biol.* **24** 663–674. MR3671106 https://doi.org/10.1089/cmb.2017.0049

[193] SHIN, J., BERG, D. A., ZHU, Y., SHIN, J. Y., SONG, J., BONAGUIDI, M. A., ENIKOLOPOV, G., NAUEN, D. W., CHRISTIAN, K. M. et al. (2015). Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* **17** 360–372.

[194] SMET, R. D. and MARCHAL, K. (2010). Advantages and limitations of current network inference methods. *Nat. Rev., Microbiol.* **8** 717–729. https://doi.org/10.1038/nrmicro2419

[195] SMITH, T., HEGER, A. and SUDBERY, I. (2017). UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27** 491–499. https://doi.org/10.1101/gr.209601.116

[196] SOLO, V., POLINE, J.-B., LINDQUIST, M. A., SIMPSON, S. L., BOWMAN, F. D., CHUNG, M. K. and CASSIDY, B. (2018). Connectivity in fMRI: Blind spots and breakthroughs. *IEEE Trans. Med. Imag.* **37** 1537–1550. https://doi.org/10.1109/TMI.2018.2831261

[197] SPELLMAN, P. T., SHERLOCK, G., ZHANG, M. Q., IYER, V. R., ANDERS, K., EISEN, M. B., BROWN, P. O., BOTSTEIN, D. and FUTCHER, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9** 3273–3297.

[198] SRIVASTAVA, A., MALIK, L., SMITH, T., SUDBERY, I. and PATRO, R. (2019). Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. *Genome Biol.* **20** Art. ID 65. https://doi.org/10.1186/s13059-019-1670-y

[199] STEGLE, O., TEICHMANN, S. A. and MARIONI, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16** 133–145. https://doi.org/10.1038/nrg3833

[200] STOECKIUS, M., ZHENG, S., HOUCK-LOOMIS, B., HAO, S., YEUNG, B. Z., MAUCK, W. M., SMIBERT, P. and SATIJA, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19** Art. ID 224.

[201] STREET, K., RISSO, D., FLETCHER, R. B., DAS, D., NGAI, J., YOSEF, N., PURDOM, E. and DUDOIT, S. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19** Art. ID 477.

[202] STUART, J. M., SEGAL, E., KOLLER, D. and KIM, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302** 249–255.

[203] SUN, W. W. and LI, L. (2017). STORE: Sparse tensor response regression and neuroimaging analysis. *J. Mach. Learn. Res.* **18** Art. ID 135. MR3763769

[204] SUN, W. W. and LI, L. (2019). Dynamic tensor clustering. *J. Amer. Statist. Assoc.* **114** 1894–1907. MR4047308 https://doi.org/10.1080/01621459.2018.1527701

[205] TAMAYO, P., SLONIM, D., MESIROV, J., ZHU, Q., KITAREEWAN, S., DMITROVSKY, E., LANDER, E. S. and GOLUB, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA* **96** 2907–2912.

[206] TANG, F., BARBACIORU, C., WANG, Y., NORDMAN, E., LEE, C., XU, N., WANG, X., BODEAU, J., TUCH, B. B. et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6** 377–382.

[207] TAVAZOIE, S., HUGHES, J. D., CAMPBELL, M. J., CHO, R. J. and CHURCH, G. M. (1999). Systematic determination of genetic network architecture. *Nat. Genet.* **22** 281–285.

[208] TESCHENDORFF, A. E., WANG, Y., BARBOSA-MORAIS, N. L., BRENTON, J. D. and CALDAS, C. (2005). A variational Bayesian mixture modelling framework for cluster analysis of gene-expression data. *Bioinformatics* **21** 3025–3033.

[209] TIAN, T. and BURRAGE, K. (2006). Stochastic models for regulatory networks of the genetic toggle switch. *Proc. Natl. Acad. Sci. USA* **103** 8372–8377.

[210] TIBSHIRANI, R., WALTHER, G. and HASTIE, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 411–423. MR1841503 https://doi.org/10.1111/1467-9868.00293

[211] TOMASI, D. and VOLKOW, N. D. (2012). Abnormal functional connectivity in children with attention-deficit/hyperactivity disorder. *Biol. Psychiatry* **71** 443–450.

[212] TONG, A. H. Y., LESAGE, G., BADER, G. D., DING, H., XU, H., XIN, X., YOUNG, J., BERRIZ, G. F., BROST, R. L. et al. (2004). Global mapping of the yeast genetic interaction network. *Science* **303** 808–813.

[213] TRANCHEVENT, L.-C., ARDESHIRDAVANI, A., ELSHAL, S., ALCAIDE, D., AERTS, J., AUBOEUF, D. and MOREAU, Y. (2016). Candidate gene prioritization with Endeavour. *Nucleic Acids Res.* **44** W117–W121. https://doi.org/10.1093/nar/gkw365

[214] TRANCHEVENT, L.-C., BARRIOT, R., YU, S. and VOOREN, S. V. (2006). Gene prioritization through genomic data fusion. *Nat. Biotechnol.* **24** 537–544.

[215] TRAPNELL, C., CACCHIARELLI, D., GRIMSBY, J., POKHAREL, P., LI, S., MORSE, M., LENNON, N. J., LIVAK, K. J., MIKKELSEN, T. S. et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32** 381–386.

[216] TZOURIO-MAZOYER, N., LANDEAU, B., PAPATHANASSIOU, D., CRIVELLO, F., ETARD, O., DELCROIX, N. et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15** 273–289.

[217] VALLEJOS, C. A., RISSO, D., SCIALDONE, A., DUDOIT, S. and MARIONI, J. C. (2017). Normalizing single-cell RNA sequencing data: Challenges and opportunities. *Nat. Methods* **14** 565–571. https://doi.org/10.1038/nmeth.4292

[218] VAN DIJK, D., SHARMA, R., NAINYS, J., YIM, K., KATHAIL, P., CARR, A. J., BURDZIAK, C., MOON, K. R., CHAFFER, C. L. et al. (2018). Recovering gene interactions from single-cell data using data diffusion. *Cell* **174** 716–729.

[219] VAN DONGEN, S. M. (2000). Graph clustering by flow simulation. Ph.D. thesis.

[220] VAROQUAUX, G. and CRADDOCK, R. C. (2013). Learning and comparing functional connectomes across subjects. *NeuroImage* **80** 405–415.

[221] VINH, N. X., CHETTY, M., COPPEL, R. and WANGIKAR, P. P. (2011). GlobalMIT: Learning globally optimal dynamic Bayesian network with the mutual information test criterion. *Bioinformatics* **27** 2765–2766.

[222] VLASBLOM, J. and WODAK, S. J. (2009). Markov clustering versus affinity propagation for the partitioning of protein interaction graphs. *BMC Bioinform.* **10** Art. ID 99. https://doi.org/10.1186/1471-2105-10-99

[223] WACHI, S., YONEDA, K. and WU, R. (2005). Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* **21** 4205–4208.

[224] WANG, Y., HU, L. and OMBAO, H. (2016). Statistical analysis of electroencephalograms. In *Handbook of Neuroimaging Data Analysis* (H. Ombao, M. Linquist, W. Thompson and J. Aston, eds.) 523–565. Chapman & Hall/CRC, Boca Raton, FL.

[225] WANG, Y., JOSHI, T., ZHANG, X.-S., XU, D. and CHEN, L. (2006). Inferring gene regulatory networks from multiple microarray datasets. *Bioinformatics* **22** 2413–2420.

[226] WANG, Y., KANG, J., KEMMER, P. B. and GUO, Y. (2016). An efficient and reliable statistical method for estimating functional connectivity in large scale brain networks using partial correlation. *Front. Neurosci.* **10** Art. ID 123.

[227] WANG, Y., ZHANG, X.-S. and XIA, Y. (2009). Predicting eukaryotic transcriptional cooperativity by Bayesian network integration of genome-wide data. *Nucleic Acids Res.* **37** 5943–5958.

[228] WANG, Y. R. and HUANG, H. (2014). Review on statistical methods for gene network reconstruction using expression data. *J. Theoret. Biol.* **362** 53–61.

[229] WANG, Y. R., LIU, K., THEUSCH, E., ROTTER, J. I., MEDINA, M. W., WATERMAN, M. S. and HUANG, H. (2017). Generalized correlation measure using count statistics for gene expression data with ordered samples. *Bioinformatics* **34** 617–624.

[230] WANG, Y. R., WATERMAN, M. S. and HUANG, H. (2014). Gene coexpression measures in large heterogeneous samples using count statistics. *Proc. Natl. Acad. Sci. USA* **111** 16371–16376.

[231] WANG, Y. X. R., JIANG, K., FELDMAN, L. J., BICKEL, P. J. and HUANG, H. (2015). Inferring gene–gene interactions and functional modules using sparse canonical correlation analysis. *Ann. Appl. Stat.* **9** 300–323. MR3341117 https://doi.org/10.1214/14-AOAS792

[232] WANG, Y. X. R., SARKAR, P., URSU, O., KUNDAJE, A. and BICKEL, P. J. (2019). Network modelling of topological domains using Hi-C data. *Ann. Appl. Stat.* **13** 1511–1536. MR4019148 https://doi.org/10.1214/19-AOAS1244

[233] WERNICKE, S. (2006). Efficient detection of network motifs. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **3** 347–359.

[234] WILLE, A. and BÜHLMANN, P. (2006). Low-order conditional independence graphs for inferring genetic networks. *Stat. Appl. Genet. Mol. Biol.* **5** Art. ID 1. MR2221304 https://doi.org/10.2202/1544-6115.1170

[235] WILLE, A., ZIMMERMANN, P., VRANOVÁ, E., FÜRHOLZ, A., LAULE, O., BLEULER, S., HENNIG, L., PRELIĆ, A., VON ROHR, P. et al. (2004). Sparse graphical Gaussian modeling of the isoprenoid gene network in *Arabidopsis thaliana*. *Genome Biol.* **5** Art. ID R92.

[236] WOLFE, C. J., KOHANE, I. S. and BUTTE, A. J. (2005). Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC Bioinform.* **6** Art. ID 227. https://doi.org/10.1186/1471-2105-6-227

[237] WONG, A. K., KRISHNAN, A. and TROYANSKAYA, O. G. (2018). GIANT 2.0: Genome-scale integrated analysis of gene networks in tissues. *Nucleic Acids Res.* **46** W65–W70. https://doi.org/10.1093/nar/gky408

[238] WONG, E., BAUR, B., QUADER, S. and HUANG, C.-H. (2011). Biological network motif detection: Principles and practice. *Brief. Bioinform.* **13** 202–215.

[239] WONG, R. K. W., LEE, T. C. M., PAUL, D., PENG, J. and ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE (2016). Fiber direction estimation, smoothing and tracking in diffusion MRI. *Ann. Appl. Stat.* **10** 1137–1156. MR3553216 https://doi.org/10.1214/15-AOAS880

[240] WU, Z., ZHANG, Y., STITZEL, M. L. and WU, H. (2018). Two-phase differential expression analysis for single cell RNA-seq. *Bioinformatics* **34** 3340–3348.

[241] XIA, Y. and LI, L. (2017). Hypothesis testing of matrix graph model with application to brain connectivity analysis. *Biometrics* **73** 780–791. MR3713112 https://doi.org/10.1111/biom.12633

[242] XIA, Y. and LI, L. (2019). Matrix graph hypothesis testing and application in brain connectivity alternation detection. *Statist. Sinica* **29** 303–328. MR3889369

[243] XIONG, Q., ANCONA, N., HAUSER, E. R., MUKHERJEE, S. and FUREY, T. S. (2012). Integrating genetic and gene expression evidence into genome-wide association analysis of gene sets. *Genome Res.* **22** 386–397.

[244] XU, C. and SU, Z. (2015). Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics* **31** 1974–1980.

[245] XU, J. and LI, Y. (2006). Discovering disease-genes by topological features in human protein–protein interaction network. *Bioinformatics* **22** 2800–2805.

[246] XU, Y. and LINDQUIST, M. A. (2015). Dynamic connectivity detection: An algorithm for determining functional connectivity change points in fMRI data. *Front. Neurosci.* **9** Art. ID 285. https://doi.org/10.3389/fnins.2015.00285

[247] YAN, K.-K., LOU, S. and GERSTEIN, M. (2017). MrTADFinder: A network modularity based approach to identify topologically associating domains in multiple resolutions. *PLoS Comput. Biol.* **13** Art. ID e1005647. https://doi.org/10.1371/journal.pcbi.1005647

[248] YEUNG, K. Y., FRALEY, C., MURUA, A., RAFTERY, A. E. and RUZZO, W. L. (2001). Model-based clustering and data transformations for gene expression data. *Bioinformatics* **17** 977–987.

[249] YU, H., KIM, P. M., SPRECHER, E., TRIFONOV, V. and GERSTEIN, M. (2007). The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *PLoS Comput. Biol.* **3** Art. ID e59. MR2373048 https://doi.org/10.1371/journal.pcbi.0030059

[250] YU, J., SMITH, V. A., WANG, P. P., HARTEMINK, A. J. and JARVIS, E. D. (2002). Using Bayesian network inference algorithms to recover molecular genetic regulatory networks. In *International Conference on Systems Biology* **2002**.

[251] YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824 https://doi.org/10.1093/biomet/asm018

[252] YUAN, Y., CHEN, Y.-P. P., NI, S., XU, A. G., TANG, L., VINGRON, M., SOMEL, M. and KHAITOVICH, P. (2011). Development and application of a modified dynamic time warping algorithm (DTW-S) to analyses of primate brain expression time series. *BMC Bioinform.* **12** Art. ID 347.

[253] ZHANG, B., LI, H., RIGGINS, R. B., ZHAN, M., XUAN, J., ZHANG, Z., HOFFMAN, E. P., CLARKE, R. and WANG, Y. (2008). Differential dependency network analysis to identify condition-specific topological changes in biological networks. *Bioinformatics* **25** 526–532.

[254] ZHANG, L. and ZHANG, S. (2018). Comparison of computational methods for imputing single-cell RNA-sequencing data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **17** 376–389.

[255] ZHANG, T., WU, J., LI, F., CAFFO, B. and BOATMAN-REICH, D. (2015). A dynamic directional model for effective brain connectivity using electrocorticographic (ECoG) time series. *J. Amer. Statist. Assoc.* **110** 93–106. MR3338489 https://doi.org/10.1080/01621459.2014.988213

[256] ZHANG, Z., DESCOTEAUX, M., ZHANG, J., GIRARD, G., CHAMBERLAND, M., DUNSON, D., SRIVASTAVA, A. and ZHU, H. (2018). Mapping population-based structural connectomes. *NeuroImage* **172** 130–145.

[257] ZHAO, F., MCCARRICK-WALMSLEY, R., ÅKERBLAD, P., SIGVARDSSON, M. and KADESCH, T. (2003). Inhibition of p300/CBP by early B-cell factor. *Mol. Cell. Biol.* **23** 3837–3846.

[258] ZHENG, G. X., TERRY, J. M., BELGRADER, P., RYVKIN, P., BENT, Z. W., WILSON, R., ZIRALDO, S. B., WHEELER, T. D., MCDERMOTT, G. P. et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8** Art. ID 14049.

[259] ZHOU, S., RÜTIMANN, P., XU, M. and BÜHLMANN, P. (2011). High-dimensional covariance estimation based on Gaussian graphical models. *J. Mach. Learn. Res.* **12** 2975–3026. MR2854354

[260] ZHU, H., CHEN, Y., IBRAHIM, J. G., LI, Y., HALL, C. and LIN, W. (2009). Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *J. Amer. Statist. Assoc.* **104** 1203–1212. MR2750245 https://doi.org/10.1198/jasa.2009.tm08096

[261] ZHU, J., CHEN, Y., LEONARDSON, A. S., WANG, K., LAMB, J. R., EMILSSON, V. and SCHADT, E. E. (2010). Characterizing dynamic changes in the human blood transcriptional network. *PLoS Comput. Biol.* **6** Art. ID e1000671.

[262] ZHU, Y. and LI, L. (2018). Multiple matrix Gaussian graphs estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 927–950. MR3874304 https://doi.org/10.1111/rssb.12278

[263] ZIEGENHAIN, C., VIETH, B., PAREKH, S., REINIUS, B., GUILLAUMET-ADKINS, A., SMETS, M., LEONHARDT, H., HEYN, H., HELLMANN, I. et al. (2017). Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* **65** 631–643.

[264] ZOTENKO, E., MESTRE, J., O'LEARY, D. P. and PRZYTYCKA, T. M. (2008). Why do hubs in the yeast protein interaction network tend to be essential: Reexamining the connection between the network topology and essentiality. *PLoS Comput. Biol.* **4** Art. ID e1000140. MR2443413 https://doi.org/10.1371/journal.pcbi.1000140

[265] ZOU, M. and CONZEN, S. D. (2004). A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics* **21** 71–79.