Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Gene regulatory network inference in single-cell biology

Kyle Akers[1] and T. M. Murali[2]

Abstract

Gene regulatory networks record relationships between transcription factors and the genes whose expression they control. Recent computational methods have been developed to predict these regulatory interactions based on gene expression data generated by single-cell sequencing technologies. In this review, we summarize these gene regulatory network inference algorithms, methods for evaluating predicted regulatory interactions, and approaches to simulate scRNA-seq data. We conclude by discussing developing trends in single-cell multiomics that we expect to influence future research on network inference.

Addresses

[1] Interdisciplinary Ph.D. Program in Genetics, Bioinformatics, and Computational Biology, Virginia Tech, Blacksburg, VA 24061, USA
[2] Department of Computer Science, Virginia Tech, Blacksburg, VA 24061, USA

Corresponding author: Murali, T.M (murali@cs.vt.edu)

Current Opinion in Systems Biology 2021, **26**:87–97

This review comes from a themed issue on **Theoretical approaches to analyze single-cell data (2021)**

Edited by **Olivier Gandrillon** and **Michael Stumpf**

For complete overview of the section, please refer the article collection - Theoretical approaches to analyze single-cell data (2021)

Available online 24 April 2021

https://doi.org/10.1016/j.coisb.2021.04.007

Keywords

Gene regulatory network, Network inference, Single-cell data, RNA-sequencing data, Multiomics data.

Introduction

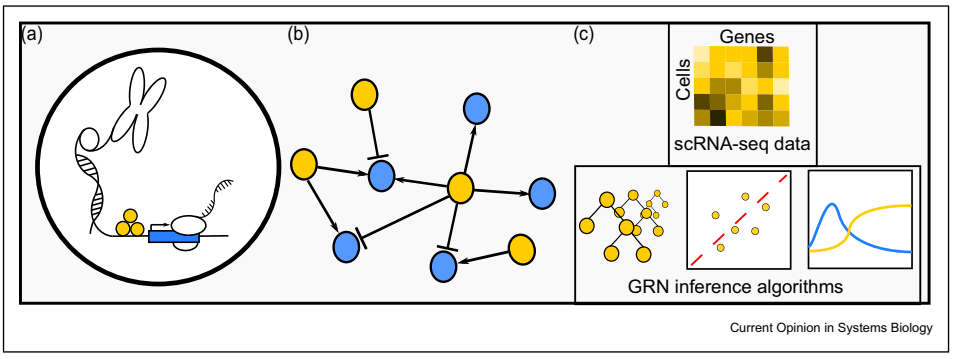
Transcription factors (TFs) regulate cellular identity as cells differentiate into specialized types. TFs also control changes in gene expression in response to environmental signals. A gene regulatory network (GRN) identifies and records these mechanisms, thus enabling further research into tissue composition and function in the context of health and disease [1]. Specifically, a GRN models the regulatory effects of TFs on the expression of target genes (Fig. 1(a)). Each node in a GRN represents a TF or a gene and each edge corresponds to a regulatory relationship between a TF and a target gene; the edge is directed from the TF to the target (Fig. 1(b)). An edge may also have a sign denoting whether it is an activating or an inhibiting regulatory interaction.

GRNs can be constructed by carefully reviewing the published literature, which can be a time-consuming and labor intensive process. This observation has inspired the development of computational methods that can automatically infer GRNs, especially from transcriptional measurements. Because GRNs represent regulation of transcription, several of these techniques have used DNA microarray and next-generation RNA-sequencing (RNA-seq) data [2]. These types of RNA-seq data sets average measurements of gene expression over a heterogeneous population of cell types, obscuring biological signals in the gene expression profiles of individual cells. The advent of single-cell RNA-sequencing (scRNA-seq) [3] has dramatically changed this landscape by enabling the characterization of gene expression in individual cells without the need to purify each cell type. However, single-cell transcriptomics data pose unique challenges for computational analysis in general, and GRN inference in particular, such as cell-to-cell stochastic variation in gene expression, changes in gene expression resulting from different stages of the cell cycle, and high sparsity due to insufficient sensitivity in the sequencing of transcripts in individual cells for genes with low expression. In spite of these technical difficulties, the increasing availability of scRNA-seq data sets has inspired the development of a wide variety of new algorithms to analyze these data sets, including methods for GRN inference. This review focuses on recent developments in GRN inference algorithms, on methods for evaluating them, and on strategies for simulating scRNA-seq data sets. It also discusses continuing advances in single-cell experimental methods that promise to enable research into the next generation of network inference in single-cell biology.

GRN inference from single-cell RNA-seq data

There are several approaches that have been applied to the problem of GRN inference based on scRNA-seq data. The input to each algorithm includes a matrix quantifying the level of expression for each gene across a number of individual cells (Fig. 1(c)). The output of a

Figure 1



GRN inference from single-cell RNA-seq data. (**a**) A TF (yellow) binds to DNA sequences in the promoter region of a target gene (blue) and has an activating or inhibiting regulatory effect on its transcription. (**b**) A GRN in which each node represents a TF (yellow) or a gene (blue), and each edge represents a regulatory relationship between a TF and a target gene. (**c**) Gene expression data sets generated by scRNA-seq contain levels of expression of each gene in each cell. GRN inference algorithms predict regulatory interactions based on patterns in the gene expression profiles using methods such as random forests, linear regression, and ODEs. GRN, gene regulatory network; TF, transcription factors; ODEs, ordinary differential equations.

GRN inference algorithm is a network of predicted TF–gene relationships (Fig. 1(b)). Each edge may have a weight representing a level of confidence in the corresponding prediction.

In the rest of this section, we summarize GRN inference techniques and refer the reader to other reviews for more detailed information [4]. A significant development that has inspired renewed interest in GRN inference using scRNA-seq data is the concept of pseudotime, which often differentiates these algorithms from earlier implementations based on bulk transcriptomic data. Cells in a sample may be in various states and transitioning between these states. Hence, it is possible to place cells in a pseudotemporal ordering based on differences in their gene expression patterns [5]. GRN inference methods developed specifically for scRNA-seq data are able to use this ordering as time-series data to inform their prediction of regulatory interactions.

Correlation

Pearson's correlation, a well-established statistic for calculating the association between two variables, has been applied to measure the coexpression of TFs and target genes in both RNA-seq and scRNA-seq data sets. Being symmetric in its arguments, this correlation does not predict the directionality of the regulatory interaction. It may identify associations between pairs of genes that do not necessarily have a direct regulatory relationship. Methods such as ppcor [6] account for the influence of other genes by computing semi-partial correlations. LEAP [7], an algorithm developed specifically for analyzing single-cell data, calculates the maximum Pearson's correlation between each pair of genes over windows in the pseudotime ordering. Since this type of correlation is not symmetric, LEAP is capable of predicting directed regulatory networks.

Mutual information

Information theoretic approaches use the mutual information, which measures the reduction in entropy for one variable (e.g., the scRNA-seq measurements for one gene) given the value of another variable (e.g., the measurements for another gene). To reduce false positives resulting from indirect relationships between two genes, methods such as PIDC [8] use partial information decomposition to calculate the unique contribution for a pair of genes that cannot be explained by the expression of a third gene. Because this relationship is symmetric, the predicted edges are undirected. The Scribe [9] algorithm uses pseudotime to calculate conditioned restricted directed information. This quantity measures the mutual information between a TF's past level of expression and the current level of a target gene's expression, conditioned on the target's expression earlier in the pseudotemporal ordering. Because the mutual information between past and current expression may not be symmetric for a TF and target gene, Scribe can infer directed edges.

Regression

In principle, the expression level of a target gene can be predicted by the levels of its TFs. Under this assumption, another approach for GRN inference models the expression of each gene as a function of the level of expression of other genes and uses regression-based methods to solve the resulting system of equations. GENIE3 [10], a notable GRN inference algorithm developed for bulk RNA-seq measurements that has also been applied to scRNA-seq data, uses a random forest method based on an ensemble of regression trees.

The weight of an edge from a TF to a target gene arises from the importance of the TF in predicting the target's expression, averaged over all regression trees in the random forest. The GRNBoost2 [11] software improves the scalability of GENIE3 especially with a view to processing large single-cell data sets efficiently. The SINCERITIES [12] algorithm instead solves a regression model based on the difference between the distribution of each gene's expression profiles at consecutive points in experimental time or pseudotime.

Bayesian networks

Another method for GRN inference models regulatory interactions between in a Bayesian network. The GRNVBEM [13] algorithm divides pseudotime into intervals. It then models the fold change in the expression of a gene between consecutive intervals as a linear combination of the expression in the previous time interval of the gene's parents in the Bayesian network. The HBFM [14] method represents gene coexpression using a sparse hierarchical Bayesian factor model to reduce the effects of high cell-to-cell variability and noise in single-cell data sets on the predicted network.

Boolean networks

While the previously introduced methods predict networks describing the regulatory effects of individual TFs, they do not account for logical rules governing the combinatorial effect of multiple TFs on the expression of a target gene. Biological regulatory mechanisms may involve the activation of a gene only in the presence of multiple TFs, or alternatively its inhibition by the expression of another gene irrespective of other factors. Boolean networks are able to characterize these combinations of interactions by representing the active or inactive state of a gene as a binary variable discretized using a threshold on gene expression, and combining these states using AND, OR, and NOT operations to explain the expression of other genes. SCNS [15] computes logical rules explaining the progression of gene expression from one point in pseudotime to the next. More recently, the LogicNet [16] algorithm uses probabilistic continuous logic to build a Boolean network in which gene expression is modeled as a continuous, rather than binary, variable between 0 and 1 to construct a GRN with directed and signed edges.

Differential equations

The availability of pseudotemporal information in single-cell data sets also enables gene expression to be modeled using ordinary differential equations (ODEs). Here, the rate of change in the expression of a target gene with respect to time is a function of the expression of its TFs. By solving this system of equations, regulatory relationships can be identified based on the weight of each TF in the function describing a gene's change in expression. The SCODE [17] algorithm makes the simplifying assumption that changes in gene expression can be defined as a linear combination of reduced dimensions to efficiently solve a less complex system of equations using linear regression. Alternatively, GRISLI [18] estimates the velocity of each cell, i.e., the rate at which each gene's expression changes according to the dynamical process in each cell. Subsequently, it simplifies the system of equations based on the assumption the resulting GRN has few regulatory edges relative to the number of genes, thereby creating a sparse regression problem. A valuable property of GRISLI is that it allows cells to follow multiple differentiation trajectories. The dynGENIE3 [19] algorithm applies the GENIE3 random forest approach to solve a system of ODEs in which the change in expression of one gene is defined as a potentially nonlinear combination of the expression of other genes.

Another class of approaches is based on the observation that cell-to-cell variation in gene expression can arise from the stochastic nature of molecular regulatory interactions. The Piecewise-Deterministic Markov Process (PDMP) model formulates the ODE for a gene's expression as a function of a stochastic two-state Markov process indicating whether or not the gene's transcription is activated, instead of directly as a function of the expression of the regulating TFs [20]. For each gene, the probability function representing the transitions between the active and inactive states includes a weight for each potential regulator. PDMP uses maximum likelihood estimation to determine these weights and thereby infer the edges of the GRN. The WASABI framework implements an alternative maximum likelihood estimation based on the concept that observed increases or decreases in gene expression should be preceded by transitions between active and inactive states in an earlier temporal window [21].

TF binding motifs

Additional regulatory information can also be integrated with the output of each of the aforementioned methods to increase confidence in the predicted interactions. For example, SCENIC [22] uses a database of TF binding motifs to filter predicted regulatory interactions identified by GENIE3. It includes only those interactions in which motifs for the TF are enriched in the target gene's promoter region. The more recent pySCENIC [23] implementation uses parallelization to improve on the efficiency of SCENIC.

Inference of multiple GRNs

Each of the aforementioned algorithms assume that the patterns of gene expression in a scRNA-seq data set can be described by a single GRN. This assumption might not be valid in the case of disparate experimental conditions, or to represent a variety of cell types in a

heterogeneous population of single cells in which different biological pathways may become activated. There has been some research into the development of algorithms which simultaneously infer multiple GRNs from a given data set.

CSN constructs undirected gene association networks specific to each cell in the scRNA-seq data set [24]. CSN computes probabilities for the observed expression for each gene occurring in an individual cell based on the frequency of that gene's expression in neighboring cells with similar gene expression. CSN compares the marginal probability for each pair of genes to their joint probability in the same cell. It adds an undirected edge connecting the two genes if the marginal probabilities are not independent. The probabilities for each pair of genes are unique for every cell, resulting in the construction of cell-specific networks.

The joint Gaussian copula graphical model (JGCGM) algorithm clusters cells into subgroups and infers a GRN for each subgroup, under the assumption that every GRN can be partitioned into a subgraph specific to the corresponding group of cells and another subgraph that is shared across all groups [25]. Based on Gaussian graphical models, JGCGM constructs an undirected graph by adding an edge between two genes if the probabilities of their expression in a normal distribution are not conditionally independent. JGCGM models the gene expression for each cell cluster as a separate Gaussian distribution. It uses maximum likelihood estimation to determine the parameters for each distribution that maximize the probability of the observed expression across all cells.

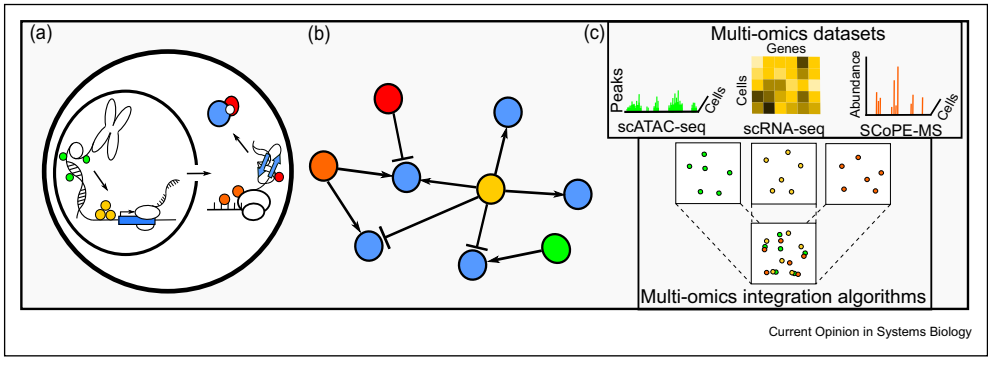
GRN inference from single-cell multiomics data

Cells have evolved multiple, diverse layers of regulatory mechanisms, including at the levels of epigenetics, transcription, translation, and post-translational modification (Fig. 2(a)). A natural generalization of a GRN would include edges representing the activating or inhibiting effects of combinations of these different types of regulatory interactions, in addition to the effect of TFs on the expression of target genes (Fig. 2(b)). The transcriptional information that is present in scRNA-seq data is not by itself sufficient to predict regulatory interactions that do not affect transcription. Fortunately, continuing and rapid innovation in experimental technology is enabling the generation of other types of omics data sets with single-cell granularity. These advancements present an opportunity to develop a next generation of GRN inference algorithms that model heterogeneous regulatory relationships and provide a more holistic view of the complex biological networks governing differences in gene expression and function between cells and cell types.

Multiomics integration

One strategy that has been applied in this context is the integration of different types of single-cell multiomics data sets independently sampled from the same cell types or under similar experimental conditions (Fig. 2(c)). Many of these methods profile the

Figure 2



GRN inference from single-cell multiomics data. (**a**) Multiple layers of regulation occur in each cell in addition to TFs (yellow) controlling the transcription of target genes (blue). Epigenetic regulation (green) such as chromatin packaging and DNA methylation affects the accessibility of target genes for transcription. Translational regulation (orange) such as mRNA degradation and initiation factors affect the level of protein abundance for the target gene. Post-translational modifications (red) such as phosphorylation can regulate the function and activity of a target protein through conformational changes. (**b**) A generalization of a GRN could additionally include edges representing the activating or inhibiting effects of different types of regulatory interactions inferred using single-cell multiomics data. (**c**) Single-cell sequencing technologies such as scATAC-seq, scRNA-seq, and SCoPE-MS measure omics data such as chromatin accessibility, gene expression, and protein abundance, respectively. Multiomics integration algorithms project omics data into a shared lower-dimensional space to enable GRN inference algorithms to identify patterns in the combination of multiple data sets. GRN, gene regulatory network.

epigenome in individual cells to assess the accessibility of genes for transcription, including the identification of open chromatin regions (scATAC-seq [26]), histone modifications (scChIP-seq [27] and scChIC-seq [28]), and DNA methylation states (scCGI-seq [29] and scBS-seq [30]). Meanwhile, advances in mass spectrometry such as SCoPE-MS [31] have been used to quantify the levels of protein abundance in individual cells.

A major challenge in the integration of these additional data types for GRN inference is associating corresponding clusters of cells between the separate data sets to identify patterns of coexpression between genes in the combined multiomics data. While GRN inference algorithms for multiomics data are still in their infancy, the methods we review in the following suggest promising directions for how these diverse data may be integrated (Fig. 2(c)).

Canonical correlation analysis

This method seeks to integrate two types of single-cell omics data sets by embedding them into a shared low-dimensional space. Canonical correlation analysis (CCA) can be used to find a linear combination of the features from each data set that maximizes their correlation. The Seurat software package for gene expression analysis provides an implementation of CCA [32]. It identifies pairs of cells between data sets that are mutual nearest neighbors in the low-dimensional space. Seurat uses these associations between cells to compute a transformation of the scATAC-seq data to predict the gene expression in each cell. The predicted gene expression matrix is then integrated with other scRNA-seq data to be used in downstream analysis. In addition to a similar integration algorithm based on CCA, MAESTRO [33] provides implementations of quality control for both scRNA-seq and scATAC-seq data to filter out unrepresentative cells prior to integration.

Non-negative matrix factorization

Non-negative matrix factorization (NMF) is an alternative method that has been applied to multiomics integration, in which expression matrices are decomposed into two lower-dimensional and non-negative factor matrices. LIGER [34] formulates a NMF problem in which a low-dimensional factor matrix is shared in the factorization of multiple data sets. Coupled NMF [35] is an alternative approach that first uses a regression model to compute a coupling matrix representing a linear transformation of the feature matrices between data sets and then adds a term to the optimization function to penalize inconsistency between data sets based on this coupling matrix.

Benchmarking of GRN inference algorithms

Given the wide variety of approaches to GRN inference that we have discussed, it is challenging to select an algorithm that is best suited to a research application studying GRNs without a standardized assessment of the performance of available methods. Independently published benchmarking results indicate that there is tremendous scope for improvement in the accuracy of GRN inference methods [36,37]. A framework for evaluating GRN inference methods can also be a useful tool by which to iteratively test and verify improvements to algorithms during development.

Ground-truth networks

The performance of GRN algorithms can be evaluated along multiple dimensions, including the accuracy of predicted regulatory relationships, the consistency of an algorithm's output across multiple simulations, and the scalability of the method for processing increasingly large single-cell data sets. The lack of well-established and commonly agreed upon ground-truth networks is by far the biggest challenge to estimating the accuracy of GRN inference algorithms reliably [37]. Data sets arising from ChIP-seq experiments are often used to create ground-truth networks. One strategy is to manually match ChIP-seq experiments and scRNA-seq data based on the cell type in which they are collected [9,37]. A less common approach is to use integrated databases that span multiple cell types [37]. While these data sets may not be specific to any cell type, they are often built by curation of the literature [38] and measure the reliability of individual experiments [39]. Finally, networks of physical or functional interactions among proteins may also be used for evaluating the accuracy of GRN inference algorithms [37,40].

Simulation of single-cell data sets

Owing to the challenges inherent in obtaining high-quality ground-truth data sets to evaluate GRNs inferred from experimental scRNA-seq data, it is also useful to supplement them with simulated inputs using GRNs based on careful literature curation with known structures. Alternatively, synthetic networks can be an effective method for evaluating the performance of GRN inference algorithms over a wider range of well-defined network topologies than are available or can be efficiently obtained from literature curation (Fig. 3(a)). The BEELINE [37] benchmarking framework includes a collection of synthetic networks, curated models, and experimentally derived networks for evaluation of GRN inference algorithms.

GeneNetWeaver [41] converts a known GRN into a system of ODEs in which gene expression is modeled as a function of TF protein abundance, which is in turn defined as a function of TF−gene expression. The method adds random noise to each equation to model the stochasticity of gene expression. GeneNetWeaver then outputs measurements of gene expression sampled at simulated points in time using this system of ODEs.

While GeneNetWeaver was originally developed to simulate bulk gene expression data, it has been recently used for generating synthetic scRNA-seq data [36,42,37]. A drawback of GeneNetWeaver is that it combines the regulators of each gene using a randomly generated logic function. Hence, it may synthesize data that do not accurately or realistically capture the dynamics of a GRN.

BoolODE [37] extends GeneNetWeaver by preserving the Boolean logic function expressing the combined regulatory effect of multiple TFs in the specification of the GRN. Cells in data sets simulated by BoolODE can be ordered along trajectories that end in clusters that are consistent with the steady states of the GRN. Moreover, there is a continuous pattern of gene expression along each trajectory that spans the length of the simulation (Fig. 3(b)). As a result, BoolODE may be more appropriate than GeneNetWeaver for the evaluation of GRN inference methods that rely on scRNA-seq data [37].

A significant difficulty presented by the generation of gene expression data is accurately simulating the complexity of cell-to-cell variation in gene expression, high dropout rates and technical noise related to sequencing technologies, and heterogeneity in the population of cell types in actual experimental data. Splatter uses the alternative approach of analyzing real scRNA-seq data sets to learn the parameters for a distribution of gene expression from which simulated data is generated [43]. The SERGIO simulator [44] generates gene expression profiles representing different cell types and models additional sources of technical variation.

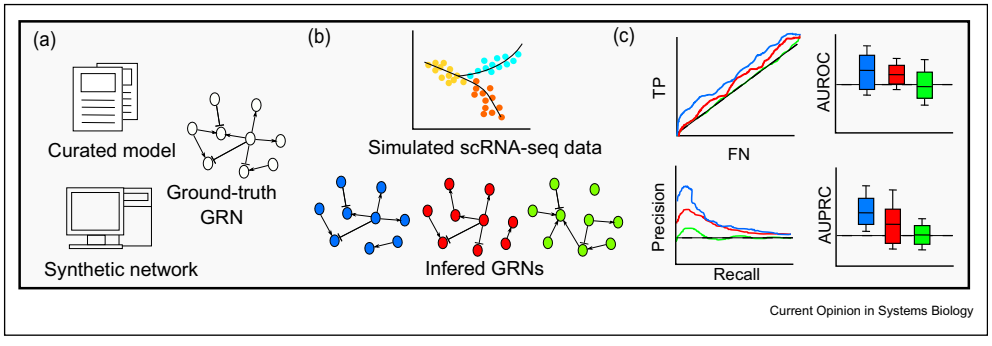
Each of these methods focuses on the generation of gene expression data. As more GRN inference methods that integrate multiomics data are developed, there will be a need to implement simulations for other single-cell data sets and meet the challenge of modeling additional stochastic effects specific to different sequencing technologies. The recently developed dyngen package can simulate multimodal regulatory events including transcription, splicing, translation, and degradation and stochastically generate the corresponding abundances of pre-mRNA, mRNA, and protein with single-cell granularity (Cannoodt et al., bioRxiv https://doi.org/10.1101/2020.02.06.936971).

Evaluation strategies

Once a ground-truth network has been decided upon, the area under precision–recall curve and the area under receiver-operating characteristic curve have both been used as measures by which to compare a GRN inference algorithm to other methods or to a random predictor (Fig. 3(c)) [36]. GRNs are generally expected to be sparse networks, with relatively few regulatory interactions between all possible pairs of genes. Owing to the resulting class imbalance, the area under precision–recall curve may be the more appropriate indicator for use in evaluation [45]. Because the highest confidence edges predicted by the algorithm may be of most interest to an experimentalist, the early precision, which considers only the top-ranking predictions, can also be used as an evaluatio measure [37].

Several methods have been used to characterize the biological relevance of predicted edges, including enrichment of GO or KEGG pathways and identification of literature support for regulatory relationships among genes connected by an inferred GRN [42]. TF binding motifs enriched in the promoter region of predicted target genes are a notable example of experimental evidence used for validation [40]. The distributions of

Figure 3



Benchmarking of GRN inference algorithms. (**a**) Ground-truth GRNs constructed from curated models and synthetic networks have been used to evaluate GRN inference algorithms. (**b**) Simulated scRNA-seq data sets are generated with gene expression profiles and trajectories based on the regulatory interactions specified in the ground-truth GRN. GRN inference algorithms predict regulatory edges in inferred GRNs using the simulated data. (**c**) Receiver-operating characteristic and precision–recall curves have been used to visualize the relative performance of different methods. The distribution of metrics such as AUROC and AUPRC across multiple simulations can be used to benchmark a GRN inference algorithm against other methods or a random predictor. GRN, gene regulatory network; AUROC, area under receiver-operating characteristic curve.;AUPRC: area under the precision–recall curve

node degrees for TFs and target genes in inferred GRN networks have been compared with corresponding distributions in ground-truth networks [36]. Another strategy is to check whether the total weight of the edges outgoing from TFs is larger than for their targets [9].

Given the high cell-to-cell variability in gene expression and the low sequencing depth inherent to single-cell data sets, the stability of an algorithm's predictions in response to variations in input data can also be a useful measurement to report [37]. Execution time and memory utilization of an algorithm given a range of input sizes can be also reported as a practical consideration for applicability to analyzing large data sets [37].

Performance results

Recent independent benchmarking results have indicated that the accuracy of most current GRN inference methods, by and large, is close to that of a random predictor [36,37]. Moreover, methods developed specifically for single-cell data sets do not consistently outperform earlier methods designed for bulk RNA-seq data, such as GENIE3 [37]. The relative performance of GRN inference algorithms can vary significantly depending on the test dataset and methodology [36]. In general, existing methods tend to perform better given simulated data as input, compared to their performance on experimental scRNA-seq data sets [37].

Future perspectives

We now turn our attention to the latest trends in experimental single-cell techniques. Because the technologies we discuss are very recent, none of them has yet become as widely used as scRNA-seq. We present these technologies with a view to discussing how they can inspire a new generation of GRN inference methods.

Simultaneous multiomics sequencing

There has been a spate of innovative technologies that simultaneously measure multiple omics data types in the same individual cells. For example, scNMT-seq [46], SNARE-seq [47], sci-CAR [48], scMT-seq [49], and scM&T-seq [50] enable the simultaneous profiling of epigenomics and transcriptomics. Technologies including RAID [51], CITE-seq [52], and REAP-seq [53] have been applied to simultaneously measure the proteome and the transcriptome. To study diseases such as cancers, especially to capture single nucleotide variations between somatic cells, there are methods that simultaneously sequence DNA and RNA in single cells, including G&T-seq [54], DR-Seq [55], SIDR [56], and TARGET-seq [57]. The scNT-seq [58] method enables the simultaneous sequencing of newly transcribed and pre-existing mRNAs to profile temporal gene expression dynamics. As more data sets generated by using these recent advances in single-cell sequencing become available, there will be a need to develop tools for the computational analysis of these integrated multiomics data with the potential to revolutionize GRN inference.

Knockout screens and single-cell multiomics

The increase or decrease in expression of other genes in response to a gene knockout is an alternative approach that can be used to predict regulatory interactions. In this context, another category of single-cell experimental methods combine CRISPR-Cas9 knockout screening with single-cell sequencing to measure the changes in the gene expression of individual cells in response to genetic perturbations. The technologies include CRISP-seq [59], Perturb-seq [60], and CROP-seq [61]. Similarly, methods such as Perturb-ATAC [62] enable the profiling of chromatin in response to CRISPR-induced cell-specific knockouts. The activating or inhibiting impact of a gene knockout on the expression of other genes can be determined by measuring the fold change of each gene in cells for a given knockout relative to the average expression for each gene in other cells [59]. Alternatively, scMAGeCK [63] uses a linear regression-based approach to model gene expression as a function of the knockout of other genes and to compute scores for the effect of each knockout on the expression of other genes.

Even more recent advancements in experimental techniques, such as ECCITE-seq [64], are enabling the simultaneous measurement of multi-omics data in response to knockouts. The measurement of changes in gene expression in individual cells, integrated with associations between expression profiles in a heterogeneous population of cell types, could be used to identify different regulatory impacts from silenced genes occurring in the context of the other genes expressed in the cell. This additional information has the potential to improve predictions for the interaction of multiple factors on gene expression, and distinguish between direct and indirect regulatory relationships.

Supervised GRN inference

By and large, GRN inference methods for scRNA-seq data are unsupervised, in the sense that they do not exploit known TF—gene relationships to predict new connections. Very recently, supervised techniques for predicting GRNs have appeared in the literature [40,65]. For example, convolutional neural network for coexpression (CNNC) computes a normalized empirical probability distribution function for each gene pair based on their expression profiles [40]. CNNC inputs this function as an image to a conventional convolutional neural network to predict regulatory relationships. To train the neural network, CNNC uses TF-target gene pairs identified by ChIP-seq experiements as positive examples and randomly-selected gene pairs as negative examples. Another approach by the same authors applies

graph convolutional neural networks on spatial transcriptomics data to predict interacting genes between cells [66]. Although this approach does not infer a GRN directly, it highlights the potential for supervised techniques for this problem.

Conclusions

Despite more than 20 years of research, inference of GRNs continues to remain a challenging problem. The BEELINE evaluation of a dozen methods showed that even the best methods had an accuracy only marginally better than a random predictor, whether they were applied to simulated or to experimental scRNA-seq data [37], underscoring the need for continued research in this area. There are several directions for future development of GRN inference algorithms.

The lack of high-quality, cell-type ground-truth data sets continues to bedevil the evaluation of GRN inference methods. Therefore, community-wide efforts to define such networks will benefit the field. Curated GRNs that describe developmental processes usually contain only tens of nodes, which may represent a simplification of real regulatory networks. Therefore, additional research is needed to create larger-scale GRNs with well-defined multimodal regulatory logic that correspond to complex and realistic trajectories. A new generation of simulators that can synthesize multimodal single-cell data sets that correspond to these GRNs can further drive the development of inference algorithms.

Statistical relationships between expression profiles of genes may be imperfect indicators of regulatory relationships. Therefore, integrating these data with other types of single-cell omics measurements is of paramount importance. Approaches that combine genetic perturbations with molecular profiling are very promising in this regard. Single-cell multimodal GRN inference is still in its infancy, especially because the development of experimental techniques is very rapid. Methods that incorporate other types of omics data sets for GRN inference will also need to address inherent difficulties in processing single-cell data such as cell-to-cell variability in sequencing depth, as well as new challenges specific to each sequencing methodology.

In conclusion, there is a critical need and tremendous scope for new ideas and directions in GRN inference. We look forward to these developments over the next few years.

Conflict of interest statement

Nothing declared.

Acknowledgements

The authors acknowledge support from National Science Foundation grants CCF-1617678 and DBI-1759858.

References

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Wagner A, Regev A, Yosef N: **Revealing the vectors of cellular identity with single-cell genomics**. *Nat Biotechnol* 2016, **34**: 1145–1160, https://doi.org/10.1038/nbt.3711.

2. Marbach D, Costello JC, Kuffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Kellis M, Collins JJ, Stolovitzky G, Aderhold A, Allison KR, Bonneau R, Camacho DM, Chen Y, Collins JJ, Cordero F, Costello JC, Crane M, Dondelinger F, Drton M, Esposito R, Foygel R, de la Fuente A, Gertheiss J, Geurts P, Greenfield A, Grzegorczyk M, Haury AC, Holmes B, Hothorn T, Husmeier D, Huynh-Thu VA, Irrthum A, Kellis M, Karlebach G, Kuffner R, Lebre S, De Leo V, Madar A, Mani S, Marbach D, Mordelet F, Ostrer H, Ouyang Z, Pandya R, Petri T, Pinna A, Poultney CS, Prill RJ, Rezny S, Ruskin HJ, Saeys Y, Shamir R, Sirbu A, Song M, Soranzo N, Statnikov A, Stolovitzky G, Vega N, Vera-Licona P, Vert JP, Visconti A, Wang H, Wehenkel L, Windhager L, Zhang Y, Zimmer R: **Wisdom of crowds for robust gene network inference**. *Nat Methods* 2012, **9**:796–804, https://doi.org/10.1038/nmeth.2016.

3. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA: **mRNA-Seq whole-transcriptome analysis of a single cell**. *Nat Methods* 2009, **6**:377–382, https://doi.org/10.1038/nmeth.1315.

4. Hu X, Hu Y, Wu F, Leung RWT, Qin J: **Integration of single-cell multi-omics for gene regulatory network inference**. *Comput Struct Biotechnol J* 2020, **18**:1925–1938, https://doi.org/10.1016/j.csbj.2020.06.033.

5. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL: **The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells**. *Nat Biotechnol* 2014, **32**:381, https://doi.org/10.1038/nbt.2859.

6. Kim S: **Ppcor: an R package for a fast calculation to semi-partial correlation coefficients**. *Commun Stat Appl Methods* 2015, **22**:665–674, https://doi.org/10.5351/CSAM.2015.22.6.665.

7. Specht AT, Li J: **LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering**. *Bioinformatics* 2017, **33**:764–766, https://doi.org/10.1093/bioinformatics/btw729.

8. Chan TE, Stumpf MPH, Babtie AC: **Gene regulatory network inference from single-cell data using multivariate information measures**. *Cell Syst* 2017, **5**:251–267, https://doi.org/10.1016/j.cels.2017.08.014.

9. Qiu X, Rahimzamani A, Wang L, Ren B, Mao Q, Durham T, McFaline-Figueroa JL, Saunders L, Trapnell C, Kannan S: **Inferring causal gene regulatory networks from coupled single-cell expression dynamics using Scribe**. *Cell Syst* 2020, **10**: 265–274, https://doi.org/10.1016/j.cels.2020.02.003.

10. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P: **Inferring regulatory networks from expression data using tree-based methods**. *PloS One* 2010, **5**, e12776, https://doi.org/10.1371/journal.pone.0012776.

11. •• Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J, Aerts S: **GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks**. *Bioinformatics* 2019, **35**:2159–2161, https://doi.org/10.1093/bioinformatics/bty916.

GRNBoost2 was identified through independent benchmarks as one of the most accurate GRN inference algorithms. Based on the GENIE3

algorithm (also a top performer), GRNBoost2 improves runtime performance for processing larger single-cell data sets by applying stochastic gradient boosting.

12. Papili Gao N, Ud-Dean SMM, Gandrillon O, Gunawan R: **SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles**. *Bioinformatics* 2018, **34**:258–266, https://doi.org/10.1093/bioinformatics/btx575.

13. Sanchez-Castillo M, Blanco D, Tienda-Luna IM, Carrion MC, Huang Y: **A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data**. *Bioinformatics* 2018, **34**:964–970, https://doi.org/10.1093/bioinformatics/btx605.

14. Sekula M, Gaskins J, Datta S: **A sparse Bayesian factor model for the construction of gene co-expression networks from single-cell RNA sequencing count data**. *BMC Bioinf* 2020, **21**:361, https://doi.org/10.1186/s12859-020-03707-y.

15. Woodhouse S, Piterman N, Wintersteiger CM, Göttgens B, Fisher J: **SCNS: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data**. *BMC Syst Biol* 2018, **12**:59, https://doi.org/10.1186/s12918-018-0581-y.

16. Malekpour SA, Alizad-Rahvar AR, Sadeghi M: **LogicNet: probabilistic continuous logics in reconstructing gene regulatory networks**. *BMC Bioinf* 2020, **21**:318, https://doi.org/10.1186/s12859-020-03651-x.

17. Matsumoto H, Kiryu H, Furusawa C, Ko MSH, Ko SBH, Gouda N, Hayashi T, Nikaido I: **SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation**. *Bioinformatics* 2017, **33**:2314–2321, https://doi.org/10.1093/bioinformatics/btx194.

18. Aubin-Frankowski PC, Vert JP: **Gene regulation inference from single-cell RNA-seq data with linear differential equations and velocity inference**. *Bioinformatics* 2020, **36**:4774–4780, https://doi.org/10.1093/bioinformatics/btaa576.

19. Huynh-Thu VA, Geurts P: **dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data**. *Sci Rep* 2018, **8**:3384, https://doi.org/10.1038/s41598-018-21715-0.

20. Herbach U, Bonnaffoux A, Espinasse T, Gandrillon O: **Inferring gene regulatory networks from single-cell data: a mechanistic approach**. *BMC Syst Biol* 2017, **11**:105, https://doi.org/10.1186/s12918-017-0487-0.

21. Bonnaffoux A, Herbach U, Richard A, Guillemin A, Gonin-Giraud S, Gros PA, Gandrillon O: **WASABI: a dynamic iterative framework for gene regulatory network inference**. *BMC Bioinf* 2019, **20**:220, https://doi.org/10.1186/s12859-019-2798-1.

22. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J, van den Oord J, Atak ZK, Wouters J, Aerts S: **SCENIC: single-cell regulatory network inference and clustering**. *Nat Methods* 2017, **14**:1083–1086, https://doi.org/10.1038/nmeth.4463.

23. •• Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, Seurinck R, Saelens W, Cannoodt R, Rouchon Q, Verbeiren T, De Maeyer D, Reumers J, Saeys Y, Aerts S: **A scalable SCENIC workflow for single-cell gene regulatory network analysis**. *Nat Protoc* 2020, **15**:2247–2276, https://doi.org/10.1038/s41596-020-0336-2.
The SCENIC algorithm combines regression based predictions made by GENIE3 from scRNA-seq data with transcription factor binding motif information. The pySENIC workflow described here provides a more efficient parallelized implementation of this method, improving scalability for single-cell analysis.

24. Dai H, Li L, Zeng T, Chen L: **Cell-specific network constructed by single-cell RNA sequencing data**. *Nucleic Acids Res* 2019, **47**:e62, https://doi.org/10.1093/nar/gkz172.

25. Wu N, Yin F, Ou-Yang L, Zhu Z, Xie W: **Joint learning of multiple gene networks from single-cell gene expression data**. *Comput Struct Biotechnol J* 2020, **18**:2583–2595, https://doi.org/10.1016/j.csbj.2020.09.004.

26. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ: **Single-cell chromatin accessibility reveals principles of regulatory variation**. *Nature* 2015, **523**:486–490, https://doi.org/10.1038/nature14590.

27. Rotem A, Ram O, Shoresh N, Sperling RA, Goren A, Weitz DA, Bernstein BE: **Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state**. *Nat Biotechnol* 2015, **33**:1165–1172, https://doi.org/10.1038/nbt.3383.

28. Ku WL, Nakamura K, Gao W, Cui K, Hu G, Tang Q, Ni B, Zhao K: **Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification**. *Nat Methods* 2019, **16**:323–325, https://doi.org/10.1038/s41592-019-0361-7.

29. Han L, Wu HJ, Zhu H, Kim KY, Marjani SL, Riester M, Euskirchen G, Zi X, Yang J, Han J, Snyder M, Park IH, Irizarry R, Weissman SM, Michor F, Fan R, Pan X: **Bisulfite-independent analysis of CpG island methylation enables genome-scale stratification of single cells**. *Nucleic Acids Res* 2017, **45**:e77, https://doi.org/10.1093/nar/gkx026.

30. Clark SJ, Smallwood SA, Lee HJ, Krueger F, Reik W, Kelsey G: **Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq)**. *Nat Protoc* 2017, **12**:534–547, https://doi.org/10.1038/nprot.2016.187.

31. Budnik B, Levy E, Harmange G, Slavov N: **SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation**. *Genome Biol* 2018, **19**:161, https://doi.org/10.1186/s13059-018-1547-5.

32. • Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R: **Comprehensive integration of single-cell data**. *Cell* 2019, **177**:1888–1902, https://doi.org/10.1016/j.cell.2019.05.031.
This paper describes an algorithm for the integration of scRNAseq and scATAC-seq data using canonical correlation analysis. This method converts chromatin accessibility data from scATAcseq data into a predicted gene expression matrix that can then be processed by existing GRN inference methods based on scRNAseq. This approach may serve as a model for the integration of other multi-omics data sets for GRN inference.

33. Wang C, Sun D, Huang X, Wan C, Li Z, Han Y, Qin Q, Fan J, Qiu X, Xie Y, Meyer CA, Brown M, Tang M, Long H, Liu T, Liu XS: **Integrative analyses of single-cell transcriptome and regulome using MAESTRO**. *Genome Biol* 2020, **21**:198, https://doi.org/10.1186/s13059-020-02116-x.

34. Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ: **Single-cell multi-omic integration compares and contrasts features of brain cell identity**. *Cell* 2019, **177**:1873–1887, https://doi.org/10.1016/j.cell.2019.05.006.

35. Duren Z, Chen X, Zamanighomi M, Zeng W, Satpathy AT, Chang HY, Wang Y, Wong WH: **Integrative analysis of single-cell genomics data by coupled nonnegative matrix factorizations**. *Proc Natl Acad Sci U S A* 2018, **115**:7723–7728, https://doi.org/10.1073/pnas.1805681115.

36. Chen S, Mar JC: **Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data**. *BMC Bioinf* 2018, **19**:232, https://doi.org/10.1186/s12859-018-2217-z.

37. • Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM: **Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data**. *Nat Methods* 2020, **17**:147–154, https://doi.org/10.1038/s41592-019-0690-6.
BEELINE provides a benchmarking framework including simulated single-cell gene expression data generated by the BoolODE algorithm, curated gene regulatory networks, and experimental data sets against which GRN inference algorithms can be evaluated. BEELINE presents benchmarking results for a representative set of GRN inference algorithms for scRNA-seq data.

38. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, Lee S, Kang B, Jeong D, Kim Y, Jeon HN, Jung H, Nam S, Chung M, Kim JH, Lee I: **TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions**. *Nucleic Acids Res* 2018, **46**:D380–D386, https://doi.org/10.1093/nar/gkx1013.

39. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J: **Benchmark and integration of resources for the estimation of human transcription factor activities**. *Genome Res* 2019, **29**:1363–1375, https://doi.org/10.1101/gr.240663.118.

40. • Yuan Y, Bar-Joseph Z: **Deep learning for inferring gene relationships from single-cell expression data**. *Proc Natl Acad Sci U S A* 2019, **116**:27151–27158, https://doi.org/10.1073/pnas.1911536116.
One of the first supervised methods for GRN inference from scRNA-seq data. It converts the joint distribution of the expression of a pair of genes into an image that is input to a convolutional neural network.

41. Schaffter T, Marbach D, Floreano D: **GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods**. *Bioinformatics* 2011, **27**:2263–2270, https://doi.org/10.1093/bioinformatics/btr373.

42. H. Nguyen, D. Tran, B. Tran, B. Pehlivan, T. Nguyen, A comprehensive survey of regulatory network inference methods using single-cell RNA sequencing data, *Brief Bioinform, 2020,* https://doi.org/10.1093/bib/bbaa190.

43. Zappia L, Phipson B, Oshlack A: **Splatter: simulation of single-cell RNA sequencing data**. *Genome Biol* 2017, **18**:174, https://doi.org/10.1186/s13059-017-1305-0.

44. Dibaeinia P, Sinha S: **SERGIO: a single-cell expression simulator guided by gene regulatory networks**. *Cell Syst* 2020, **11**:252–271, https://doi.org/10.1016/j.cels.2020.08.003.

45. Saito T, Rehmsmeier M: **The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets**. *PloS One* 2015, **10**, e0118432, https://doi.org/10.1371/journal.pone.0118432.

46. Clark SJ, Argelaguet R, Kapourani CA, Stubbs TM, Lee HJ, Alda-Catalinas C, Krueger F, Sanguinetti G, Kelsey G, Marioni JC, Stegle O, Reik W: **scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells**. *Nat Commun* 2018, **9**:781, https://doi.org/10.1038/s41467-018-03149-4.

47. Chen S, Lake BB, Zhang K: **High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell**. *Nat Biotechnol* 2019, **37**:1452–1457, https://doi.org/10.1038/s41587-019-0290-0.

48. Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, Daza RM, McFaline-Figueroa JL, Packer JS, Christiansen L, Steemers FJ, Adey AC, Trapnell C, Shendure J: **Joint profiling of chromatin accessibility and gene expression in thousands of single cells**. *Science* 2018, **361**:1380–1385, https://doi.org/10.1126/science.aau0730.

49. Hu Y, Huang K, An Q, Du G, Hu G, Xue J, Zhu X, Wang CY, Xue Z, Fan G: **Simultaneous profiling of transcriptome and DNA methylome from a single cell**. *Genome Biol* 2016, **17**:88, https://doi.org/10.1186/s13059-016-0950-z.

50. Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX, Krueger F, Smallwood S, Ponting CP, Voet T, Kelsey G, Stegle O, Reik W: **Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity**. *Nat Methods* 2016, **13**:229–232, https://doi.org/10.1038/nmeth.3728.

51. Gerlach JP, van Buggenum JAG, Tanis SEJ, Hogeweg M, Heuts BMH, Muraro MJ, Elze L, Rivello F, Rakszewska A, van Oudenaarden A, Huck WTS, Stunnenberg HG, Mulder KW: **Combined quantification of intracellular (phospho-)proteins and transcriptomics from fixed single cells**. *Sci Rep* 2019, **9**:1469, https://doi.org/10.1038/s41598-018-37977-7.

52. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R, Smibert P: **Simultaneous epitope and transcriptome measurement in single cells**. *Nat Methods* 2017, **14**:865–868, https://doi.org/10.1038/nmeth.4380.

53. Peterson VM, Zhang KX, Kumar N, Wong J, Li L, Wilson DC, Moore R, McClanahan TK, Sadekova S, Klappenbach JA: **Multiplexed quantification of proteins and transcripts in single cells**. *Nat Biotechnol* 2017, **35**:936–939, https://doi.org/10.1038/nbt.3973.

54. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, Goolam M, Saurat N, Coupland P, Shirley LM, Smith M, Van der Aa N, Banerjee R, Ellis PD, Quail MA, Swerdlow HP, Zernicka-Goetz M, Livesey FJ, Ponting CP, Voet T: **G&T-seq: parallel sequencing of single-cell genomes and transcriptomes**. *Nat Methods* 2015, **12**:519–522, https://doi.org/10.1038/nmeth.3370.

55. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A: **Integrated genome and transcriptome sequencing of the same cell**. *Nat Biotechnol* 2015, **33**:285–289, https://doi.org/10.1038/nbt.3129.

56. Han KY, Kim KT, Joung JG, Son DS, Kim YJ, Jo A, Jeon HJ, Moon HS, Yoo CE, Chung W, Eum HH, Kim S, Kim HK, Lee JE, Ahn MJ, Lee HO, Park D, Park WY: **SIDR: simultaneous isolation and parallel sequencing of genomic DNA and total RNA from single cells**. *Genome Res* 2018, **28**:75–87, https://doi.org/10.1101/gr.223263.117.

57. Rodriguez-Meira A, Buck G, Clark SA, Povinelli BJ, Alcolea V, Louka E, McGowan S, Hamblin A, Sousos N, Barkas N, Giustacchini A, Psaila B, Jacobsen SEW, Thongjuea S, Mead AJ: **Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing**. *Mol Cell* 2019, **73**:1292–1305, https://doi.org/10.1016/j.molcel.2019.01.009.

58. Qiu Q, Hu P, Qiu X, Govek KW, Cámara PG, Wu H: **Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq**. *Nat Methods* 2020, **17**:991–1001, https://doi.org/10.1038/s41592-020-0935-4.

59. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, Salame TM, Tanay A, van Oudenaarden A, Amit I: **Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq**. *Cell* 2016, **167**:1883–1896, https://doi.org/10.1016/j.cell.2016.11.039.

60. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, Adamson B, Norman TM, Lander ES, Weissman JS, Friedman N, Regev A: **Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens**. *Cell* 2016, **167**:1853–1866, https://doi.org/10.1016/j.cell.2016.11.038.

61. Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler A, Alpar D, Bock C: **Pooled CRISPR screening with single-cell transcriptome readout**. *Nat Methods* 2017, **14**:297–301, https://doi.org/10.1038/nmeth.4177.

62. Rubin AJ, Parker KR, Satpathy AT, Qi Y, Wu B, Ong AJ, Mumbach MR, Ji AL, Kim DS, Cho SW, Zarnegar BJ, Greenleaf WJ, Chang HY, Khavari PA: **Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks**. *Cell* 2019, **176**:361–376, https://doi.org/10.1016/j.cell.2018.11.022.

63. Yang L, Zhu Y, Yu H, Cheng X, Chen S, Chu Y, Huang H, Zhang J, Li W: **scMAGeCK links genotypes with multiple phenotypes in single-cell CRISPR screens**. *Genome Biol* 2020, **21**:19, https://doi.org/10.1186/s13059-020-1928-4.

64. • Mimitou EP, Cheng A, Montalbano A, Hao S, Stoeckius M, Legut M, Roush T, Herrera A, Papalexi E, Ouyang Z, Satija R, Sanjana NE, Koralov SB, Smibert P: **Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells**. *Nat Methods* 2019, **16**:409–412, https://doi.org/10.1038/s41592-019-0392-0.
ECCITE-seq is a recent advancement in sequencing methods in which CRISPR knockout screening is combined with the simultaneous multi-omics measurement of gene expression and protein abundance in single-cells. As more data sets using this technology are published, a next generation of GRN inference methods may make use of the additional information from the response of gene expression and protein abundance to genetic perturbations in order to improve the prediction regulatory interactions.

65. Turki T, Taguchi YH: **SCGRNs: novel supervised inference of single-cell gene regulatory networks of complex diseases**. *Comput Biol Med* 2020, **118**:103656, https://doi.org/10.1016/j.compbiomed.2020.103656.

66. Yuan Y, Bar-Joseph Z: **GCNG: graph convolutional networks for inferring gene interaction from spatial transcriptomics data**. *Genome Biol* 2020, **21**:300, https://doi.org/10.1186/s13059-020-02214-w.