Technische Universität München
TUM School of Life Sciences

Incorporation of prior knowledge to improve gene regulatory network inference from single-cell transcriptomics data

Marco Elmar Stock, M.Sc.

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Dmitrij Frischmann

Prüfende der Dissertation:

1. Prof. Dr. Maria Colomé-Tatché

2. Prof. Dr. Markus List

Die Dissertation wurde am 27.03.2025 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 14.07.2025 angenommen.

Abstract

Gene regulatory networks (GRNs) model regulatory interactions among genes as edges in a graph. They offer mechanistic insights into cellular functions and provide a robust foundation for targeted disease modeling and drug discovery. When inferred primarily from large-scale single-cell RNA sequencing (*scRNA-Seq*) data, several independent benchmarking studies have demonstrated that the performance of current GRN inference algorithms is, at best, moderate. Notably, correlation-based approaches continue to achieve state-of-the-art performance despite their inability to distinguish causal interactions, and they are generally not consistently outperformed by more complex inference methods. Current evaluation methods used in the benchmarking studies formulate the network inference task as a binary classification problem at the individual edge level, thus neglecting the intrinsic graph structure of regulatory networks. Therefore, it is essential to develop evaluation techniques that assess both the accuracy of individual edge predictions and the overall quality of the network topology. Furthermore, while research to improve inference performance using prior knowledge has grown rapidly, a comprehensive review and benchmarking framework remains to be developed to systematically guide both users and developers.

In the first part of this thesis, we perform the first topological benchmarking study for GRN inference from single-cell transcriptomics to assess the ability of existing algorithms to accurately infer network structure. This is achieved by quantifying both node-wise and network-wide topological properties of the GRNs. The evaluation uncovers a general insensitivity of benchmarked algorithms to topological features inherently associated with, for example, GRN stability.

In the second part of the thesis, topological and other types of priors that can additionally inform the GRN inference process are compiled and systematically compared. For priors that can be represented as graphs, different strategies for handling the priors in the inference process are examined. Because performance improvements from the priors cannot be easily disentangled from the intrinsic performance of the algorithms, benchmarking algorithms that leverage prior knowledge is typically challenging. By leveraging graph representations of priors, it becomes possible to separate these contributions and enable a fair comparison.

This thesis deepens our understanding of the capabilities and limitations of current GRN inference algorithms while providing guidance for the development of future methods with improved performance and a rigorous benchmarking framework to quantify those improvements.

Zusammenfassung

Genregulatorische Netzwerke (GRNs) modellieren regulatorische Interaktionen zwischen Genen als Kanten in einem Graphen. Sie bieten mechanistische Einblicke in zelluläre Funktionen und eine robuste Grundlage für gezielte Krankheitsmodellierung und Medikamentenforschung. Mehrere unabhängige Studien haben gezeigt, dass die Leistung aktueller GRN-Inferenzalgorithmen, wenn sie hauptsächlich aus umfangreichen Einzelzell-RNA-Sequenzierungs-(*scRNA-Seq*)-Daten abgeleitet werden, bestenfalls moderat ist. Bemerkenswerterweise werden korrelationsbasierte Ansätze trotz ihrer Unfähigkeit kausale Wechselwirkungen zu unterscheiden, weiterhin von komplexeren Methoden nicht konsistent übertroffen. Allerdings wird die Netzwerkinferenz häufig als binäres Klassifikationsproblem auf Kantenebene formuliert und vernachlässigt dabei die strukturellen Eigenschaften regulatorischer Netzwerke. Daher ist es wichtig, Evaluierungstechniken zu entwickeln, die sowohl die Genauigkeit einzelner Kantenvorhersagen als auch die Gesamtqualität der Netzwerktopologie bewerten. Während die Forschung zur Verbesserung der Inferenzleistung unter Verwendung von Vorwissen rapide gewachsen ist, fehlt hier noch ein umfassender Überblick und systematische Vergleiche der Algorithmen für Anwender und Entwickler.

Im ersten Teil dieser Arbeit führen wir die erste topologische Benchmarking-Studie zur GRN-Inferenz aus Einzelzell-Transkriptomik durch, um die Fähigkeit bestehender Algorithmen zur präzisen Netzwerkstrukturinferenz zu bewerten. Dies wird durch die Quantifizierung sowohl knotenweiser als auch netzwerkweiter topologischer Eigenschaften der GRNs erreicht. Die Auswertung deckt eine allgemeine Unempfindlichkeit der getesteten Algorithmen gegenüber topologischen Merkmalen auf, die inhärent mit beispielsweise der GRN-Stabilität verbunden sind.

Im zweiten Teil der Arbeit werden topologische und andere Arten von vorhandenem Wissen, die den GRN-Inferenzprozess zusätzlich informieren können, zusammengestellt und systematisch verglichen. Für Vorwissen, das als Graph dargestellt werden kann, werden verschiedene Strategien zur Handhabung im Inferenzprozess vorgestellt. Da Leistungsverbesserungen durch das Vorwissen nicht einfach von der eigentlichen Leistung der Algorithmen zu trennen ist, ist das Benchmarking dieser Algorithmen herausfordernd. Graph-Repräsentationen ermöglichen diese Trennung und somit faire Vergleiche.

Diese Arbeit vertieft unser Verständnis der Möglichkeiten und Grenzen aktueller GRN-Inferenzalgorithmen und bietet gleichzeitig Orientierung für die Entwicklung zukünftiger Methoden mit verbesserter Leistung sowie ein Benchmarking-Entwurf zur Quantifizierung dieser Verbesserungen.

Acknowledgements

I would like to thank Antonio Scialdone and Eva Hörmanseder for offering me a position in their research teams. Their consistent support, our productive scientific exchanges, the engaging projects, and the opportunity to attend international conferences with their funding have been fundamental to my work.

I thank Maria Colomé-Tatché, Markus List, and Dmitrij Frishman for serving on my PhD thesis committee. I also value the thoughtful feedback from Carsten Marr and Maria-Elena Torres-Padilla during my Thesis Advisory Committee meetings.

It has been rewarding to guide several talented students through their thesis and study projects. I especially thank Niclas Popp, Erik Steiger, and Harshit Pateria for their confidence in my supervision, which has been a significant learning experience for me as well.

My sincere thanks go to Meghana Oak, Tomas Zikmund, Clara Hermant, Corinna Losert, Kim Job, Matteo Zambon, Florin Ratajczak, Martin Miranda, Gabriele Lubatti, Jonathan Fiorentino, Stefan Hamperl, Ana Janeva, Veronica Finazzi and Matthias Heinig for their insightful scientific discussions and valuable contributions to my projects.

I am grateful to Alexander Tong, Jason Hartford, Yoshua Bengio, Mathieu Seyfrid, and Nicole Zhang for the opportunity to visit MILA in Montreal during my PhD studies—an experience that substantially influenced my personal and research perspective.

I also thank the members of both the Hörmanseder and Scialdone labs and colleagues throughout the IES, for creating an excellent working environment. Additionally, I acknowledge the Joachim Herz Foundation and the Munich School of Data Science for providing enriching seminars, networking opportunities, and supplementary funding that supported important research activities during my PhD.

Finally, I thank my friends and family, particularly Julia, Michael, Matthias, Luisa, Steven, Christine and Mailin, for making this period of my life memorable. I look forward to the adventures that lie ahead.

List of contributed publications

This thesis is based on the following core publications as main author:

Core publications as main author

I) Stock, M., Popp, N., Fiorentino, J. and Scialdone, A. (2024). Topological benchmarking of algorithms to infer gene regulatory networks from single-cell RNA-seq data. *Bioinformatics*, *40*(5), btae267,
doi: 10.1093/bioinformatics/btae267
(See also publication [1] in the bibliography and in A.1.)

II) Stock, M., Losert, C., Zambon, M., Popp, N., Lubatti, G., Hörmanseder, E., Heinig, M. and Scialdone, A. (2025). Leveraging prior knowledge to infer Gene Regulatory Networks from single-cell RNA-sequencing data. *Molecular Systems Biology*, *21*(3), 214-230,
doi: 10.1038/s44320-025-00088-3
(See also publication [2] in the bibliography and in A.2.)

Further main author manuscripts not included in this thesis

I) *Oak, M.S., *Stock, M., Mezes, M., Straub, T., Hynes-Allen, A.M., van den Ameele, J., Forne, I., Ettinger, A., Imhof, A., Scialdone, A., Hörmanseder, E.B. (2025). H3K4 methylation-promoted transcriptional memory ensures faithful zygotic genome activation and embryonic development. *bioRxiv*,
doi: 10.1101/2025.01.20.633863
(See also manuscript [3] in the bibliography.)

Further co-author manuscripts not included in this thesis

I) Lubatti, G., Stock, M., Iturbide, A., Ruiz Tejada Segura, M.L., Riepl, M., Tyser, R.C.V., Danese, A., Colomé-Tatché, M., Theis, F.J., Srinivas, S., Torres-Padilla, M.E., Scialdone, A. (2023). CIARA: a cluster-independent algorithm for identifying markers of rare cell types from single-cell sequencing data. *Development*, *150* (11), dev201264,
doi: 10.1242/dev.201264
(See also publication [4] in the bibliography.)

II) Zikmund, T., Fiorentino, J., Penfold, C., Stock, M., Agarwal, G., Langfeld, L., Hamperl, S., Scialdone, A., Hörmanseder, E. (2025). Differentiation success of reprogrammed cells is heterogeneous in vivo and modulated by somatic cell identity memory. *Stem Cell Reports*,
doi: 10.1016/j.stemcr.2025.102447
(See also publication [5] in the bibliography.)

*equal contribution

III) Janeva, A., Penfold, C., Llorente-Armij, S., Li, H., Zikmund, T., Stock, M., Jullien, J., Straub, T., Forne, I., Imhof, A., Vaquerizas, J.M., Gurdon, J., Hörmanseder, E. (2025). 'Digital Reprogramming' Decodes Epigenetic Barriers of Cell Fate Changes. *bioRxiv*, doi: 10.1101/2025.01.22.634227
(See also manuscript [6] in the bibliography.)

IV) Nourisa, J., Passemiers, A., Stock, M., Zeller-Plumhoff, B., Cannoodt, R., Arnold, C., Tong, A., Hartford, J., Scialdone, A., Moreau, Y., Li, Y., Lücken, M. (2025). geneRNIB: a living benchmark for gene regulatory network inference. *bioRxiv*, doi: 10.1101/2025.02.25.640181
(See also manuscript [7] in the bibliography.)

I, Marco Stock, am the main author of the publications [1, 2]. A summary of these articles can be found in Section 1.6 and in full length, including a detailed description of my contributions in Sections 2.1, 2.2, respectively.

Contents

1. Introduction

1.1. Introduction to gene regulatory networks

Gene regulatory networks (GRNs) are a framework to model gene interactions in a cell, where genes can either repress or activate the expression of other genes. Thus, they provide mechanistic representations of cellular behavior that hold significant promise for advancing our understanding of diseases and driving drug development [8, 9]. This section introduces the biological foundations of gene expression regulation and describes single-cell RNA-sequencing, a quantification method used to measure gene expression at the single-cell level. It further presents two distinct representations of regulatory interactions: GRNs and enhancer gene regulatory networks (eGRNs).

1.1.1. Regulation of gene expression

All cells in the human body contain a near-identical DNA sequence and gene set. However, multi-cellular organisms comprise numerous distinct cell types, each performing specialized functions. Different cell types can be distinguished based on the frequency at which a gene's DNA sequence is transcribed into RNA. Consequently, differences in RNA transcript abundance yield unique gene expression profiles. In many cases, cells preferentially transcribe specialized cell type–specific marker genes, while genes involved in fundamental processes, such as cell cycle regulation, are expressed across most cell types [10, 11]. For protein-coding genes, the transcribed RNAs serve as templates for protein synthesis. However, additional regulatory layers in the protein synthesis process can decorrelate protein levels and RNA abundance [12].

Gene expression is controlled at multiple levels, including transcriptional, post-transcriptional, translational, and post-translational regulation [13–15]. For example, post-transcriptional mechanisms influence RNA splicing, transport, and degradation, whereas post-translational modifications affect protein stability, localization, and activity. Central to transcriptional regulation are transcription factors, which modulate gene expression by either facilitating or inhibiting the recruitment of RNA polymerase II to target genes, a multiprotein complex responsible for transcription [16]. These transcription factors generally bind to specific DNA sequences within cis-regulatory elements (CREs), which are regions located either in close proximity to, or at some distance from, the transcription start site (TSS) of a gene. Two principal types of CREs include promoters, found near the TSS, and enhancers, which may be located at a distance yet brought into proximity through DNA looping [17]. The regulatory influence exerted by a CRE may depend on a single transcription factor or a combination of factors. In some cases, individual transcription factors act independently, analogous to a logical OR operation, whereas in other cases, a cooperative interaction among multiple factors, resembling a logical AND, is necessary to achieve the desired regulatory outcome [18–20].

Gene expression is further modulated by the chromatin configuration and epigenetic modifications that determine the accessibility of genomic regions [21]. Factors such as DNA methylation and histone modifications influence transcription factor binding and gene activity [22]. Notably, some transcription factors demonstrate a higher binding affinity for specific DNA sequence motifs, which can assist in identifying distant enhancer regions [23]. Precise regulation of gene expression is critical for proper development and cell differentiation. While the gene expression profile of a fully differentiated cell tends to be stable, the precise timing and extent of gene expression changes during development are essential for proper cellular differentiation [24, 25].

1.1.2. Single-cell RNA-Sequencing

High-throughput sequencing technologies now enable the simultaneous quantification of RNA expression levels for all annotated genes in a given sample. Initially, these methods were applied to bulk samples, in which RNA is isolated from many cells pooled together. More recently, single-cell RNA sequencing (scRNA-seq) techniques have been developed that capture gene expression levels across individual cells, which is important for GRN inference, since it reveals the heterogeneity and variability between the single cells [26]. Common scRNA-seq platforms include plate-based methods such as Smart-seq2 [27] and droplet-based approaches exemplified by the 10x Chromium 3´ protocol [28]. While Smart-seq2 offers full-length transcript capture and higher sequencing depth per cell, 10x Chromium relies on 3´-end tagging and is considerably more scalable, often processing roughly two orders of magnitude more cells per experiment [28, 29].

Technical challenges Despite the advantages of single-cell resolution, several technical challenges remain. It is important to note that the quantifications obtained from sequencing experiments can be influenced by batch effects [30]. Therefore, when comparing two conditions, samples should ideally be sequenced concurrently to mitigate such technical variability. In some protocols, spike-in RNAs, which are synthetic molecules of known concentration, are incorporated to normalize for differences in sequencing depth. However, many modern workflows also rely on computational normalization methods, such as scaling library sizes or incorporating unique molecular identifiers (UMIs) [31].
In addition to batch effects, scRNA-seq data are often characterized by dropout events, in which transcripts present in the cell are not detected, resulting in sparse datasets [32]. The inherent high dimensionality of the data further complicates analysis, necessitating the application of dimensionality reduction or feature selection techniques. Another common challenge is the occurrence of doublets in droplet-based methods, where two (or more) cells are inadvertently encapsulated within a single droplet. Various computational tools have been developed to detect and remove these doublets, i.e., DoubletFinder [33].

Analysis of scRNA-Seq data The standard analysis pipeline for scRNA-seq data begins with rigorous quality control of the raw sequencing data. Once quality thresholds are met, the

sequencing reads are aligned to an annotated reference genome, yielding a count matrix in which each entry represents the number of reads mapped to a given gene for each cell. This matrix is typically further filtered to exclude cells that do not meet minimum count criteria, and the remaining counts are normalized, often by logarithmic transformation after the addition of a pseudocount [34]. The resulting normalized matrix serves as the foundation for downstream analyses including cell clustering and type annotation, batch correction, differential expression analysis, trajectory inference, and dimensionality reduction [30]. Clustering methods such as the Leiden or Louvain algorithms are commonly applied, and subsequent analysis may include the identification of marker genes characteristic of each cluster [35]. Batch correction methods vary in their handling of the tradeoff to retain biological signal while reducing technical noise [36]. In comparative studies, differential expression analysis is used to statistically evaluate the differences between groups, such as cell types or conditions [37]. Trajectory inference aims to computationally reconstruct differentiation timing of the heterogeneous set of single cells [38]. It may be derived from pseudo time ordering or by leveraging additional data, for example, the ratio of unspliced to spliced transcripts as implemented in RNA velocity analysis [39]. Some GRN inference algorithms also leverage this computed information for the inference process, since differentiation dynamics are another source of variability among the observed single cells [40]. Dimensionality reduction techniques such as principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE) [41], and uniform manifold approximation and projection (UMAP) [42] are then used to visualize the data.

Downstream tasks The advent of scRNA-seq has opened up numerous avenues for more sophisticated analyses. Beyond standard clustering and differential expression, advanced approaches include cell–cell communication analysis, spatial transcriptomics, and regulatory network inference. Cell–cell communication analysis explores the signaling interactions between neighboring cells that contribute to the formation of organized three-dimensional tissue structures [43, 44]. Spatial transcriptomics integrates spatial information with gene expression data, thereby preserving the cellular microenvironment context [45]. Finally, computational inference of regulatory networks, which aims to model the interactions between regulators and their target genes, is a key focus of this work.

1.1.3. Gene regulatory networks

This work focuses on the inference of gene regulatory networks (GRNs), which serve as abstract representations of the regulatory interactions occurring within cells [46]. In principle, a fully elucidated GRN could act as a mechanistic model for cellular behavior, enabling predictions regarding the cell's response to perturbations such as gene knockouts [47]. Typically, GRNs are depicted as graphs in which nodes represent regulators and target genes and edges denote the regulatory interactions. A comprehensive regulatory network might include multiple node types, for instance, proteins and RNAs, as well as various edge types, such as activating and repressing interactions. In fact, because multiple transcription factors can cooperatively regulate a single

target, the network is more accurately described as a hypergraph [48]. In the literature, however, GRN inference is often limited to simplified models in which nodes correspond solely to genes and edges represent either directed or undirected regulatory interactions. In this framework, transcription factors are treated as genes, resulting in a homogeneous graph, and in some cases, the direction or nature (e.g., activating or repressing) of the regulation is also inferred [49].

Well-characterized GRNs tend to exhibit several common topological properties. Often, the node-degree distribution follows a scale-free pattern, in which a small number of transcription factors serve as hubs with many connections while most genes have only a few interactions [50]. In addition, conserved network motifs such as feed-forward and feedback loops are frequently observed. These structural features are believed to contribute to the robustness of the network, helping to maintain a stable cellular identity once differentiation has occurred [51, 52].

GRNs can also vary in scope. Some networks are specific to particular cell types [22, 53], patients [54], or developmental stages and capture the dynamic processes of cell differentiation [55]. Inference of the regulatory interactions often relies on scRNA-Seq data. Such data must exhibit sufficient variability, whether temporal or reflecting cell-to-cell heterogeneity, to enable the elucidation of the underlying regulatory mechanisms. Variability may be intrinsic [26] or introduced experimentally through targeted gene overexpression, knockdowns, knockouts, or drug perturbations [56, 57].

The complexity of GRN inference is compounded by the rapid growth of the solution space with the number of genes considered. Even for simplified homogeneous graphs with n nodes and undirected edges, the number of possible networks becomes extremely large with $2^{\binom{n}{2}}$, for example, a set of 10 genes can yield already approx. $3.5 \cdot 10^{13}$ potential graphs, while, i.e., humans already have approximately 20,000 annotated protein-coding genes [58]. For directed graphs, the number of potential graphs is even higher with $4^{\binom{n}{2}}$ and for signed directed edges it is already $9^{\binom{n}{2}}$. In practice, observational data typically result in a set of plausible networks, with further refinement toward a unique solution often requiring additional interventional data [59].

Due to their potential to provide mechanistic insights into cellular behavior and responses to perturbations [60], GRNs are of great interest across various research domains. They are particularly relevant in drug discovery for identifying and prioritizing new therapeutic targets, as well as in personalized medicine when GRNs are inferred on a patient-specific basis [61, 62].

1.1.4. Enhancer Gene regulatory networks

Recent advancements in GRN inference have led to the development of algorithms that incorporate additional biological context [9, 63, 64]. Traditional methods often assume that transcription factors regulate target genes solely via direct binding to promoters. However, it is well established

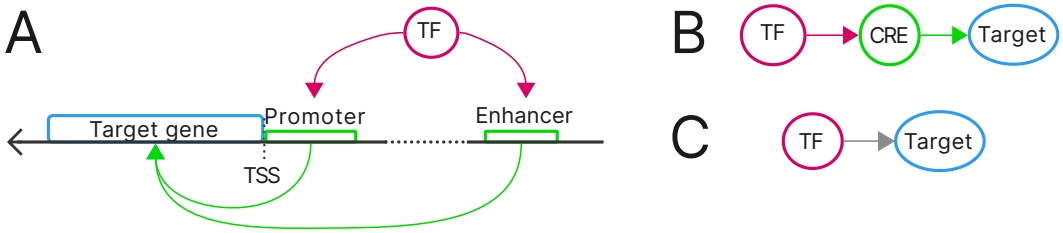


Figure 1.1: Schematic representation of an edge in a gene regulatory network. (A) Schematic denoting biological ground truth and (B) their representation in an enhancer gene regulatory network and (C) gene regulatory network respectively. TF = Transcription factor, TSS = Transcription start site, CRE = Cis-regulatory element

that transcription factors can also influence gene expression indirectly through interactions with other cis-regulatory elements, such as enhancers [65, 66]. In these enhancer gene regulatory networks (eGRNs), a regulatory connection from a transcription factor to a target gene is decomposed into two distinct interactions: one between the transcription factor and a cis-regulatory element (CRE), and another between the CRE and the target gene. This framework more accurately reflects the actual cellular mechanisms and may thereby improve inference performance [67].

For ease of comparison with traditional homogeneous GRN models, the eGRN can be collapsed into a standard GRN by introducing direct or skip connections between transcription factors and target genes whenever an indirect regulatory pathway via a CRE is identified [67]. Figure 1.1A shows an example of a TF with two possible binding sites, the promoter close to the transcription start site of the target gene, and the more distant enhancer region. Figure 1.1B shows the related abstract representation in an eGRN, where the CRE mediating the regulatory effect is included as a node, whereas there are different types of edges connecting either transcription factors and CREs, or CREs with target genes. Figure 1.1C shows the same collapsed regulatory link in a GRN where CREs are not explicitly modelled.

1.2. Methods for inference of GRNs from scRNA-Seq data

Given the importance of GRNs as mechanistic models, their inference has become a central research focus. Significant effort has been dedicated to inferring GRNs from scRNA-seq data, which provides an unbiased quantification of the transcriptome at single-cell resolution. This per-cell data captures variability across individual cells, which can enhance the accuracy of GRN inference. However, in addition to the conventional challenges of GRN inference (such as the large solution space and ambiguity in regulatory interactions [68, 69]), scRNA-seq data introduce additional issues related to noise and sparsity [70].

One straightforward method for inferring regulatory interactions is to study gene co-expression, for instance by computing linear correlations between gene expression levels across cells. This approach, however, has several drawbacks. In particular, it may fail to capture complex non-linear dependencies that arise from intricate gene interaction patterns, and it does not elucidate the directionality or causality of regulatory effects, as it relies solely on statistical correlations.

To overcome these limitations, more advanced, data-driven GRN inference algorithms have been developed. These methods vary in their learning paradigms, inference techniques, and strategies for handling genes, leading to differences in performance across various settings. For example, many algorithms employ regression-based techniques to infer regulatory networks from expression data [71, 72]. Moreover, different types of neural networks and probabilistic graphical models are commonly applied to the GRN inference problem [73, 74]. Other methods draw on concepts from information theory by using measures such as mutual information, conditional mutual information, multivariate information, or partial information decomposition [75–77]. Additional approaches include ordinary differential equation (ODE)-based methods [78, 79]. Finally, community-based approaches aggregate results from several algorithms to yield a consensus GRN [80].

Furthermore, these algorithms differ in how they model genes during inference. The most common approach is to handle each gene individually, which achieves high resolution but is also more prone to noise. To reduce the complexity and enhance robustness, some methods infer regulatory networks at the level of gene modules, which are groups of genes with similar expression patterns that are treated as single units [81]. Other algorithms combine both single-gene and gene-module strategies to balance resolution and noise reduction [82].

1.2.1. Correlation

Some of the simplest approaches used for inference of GRNs rely on the analysis of correlations among gene expression levels. Two common correlation coefficients used in this context are

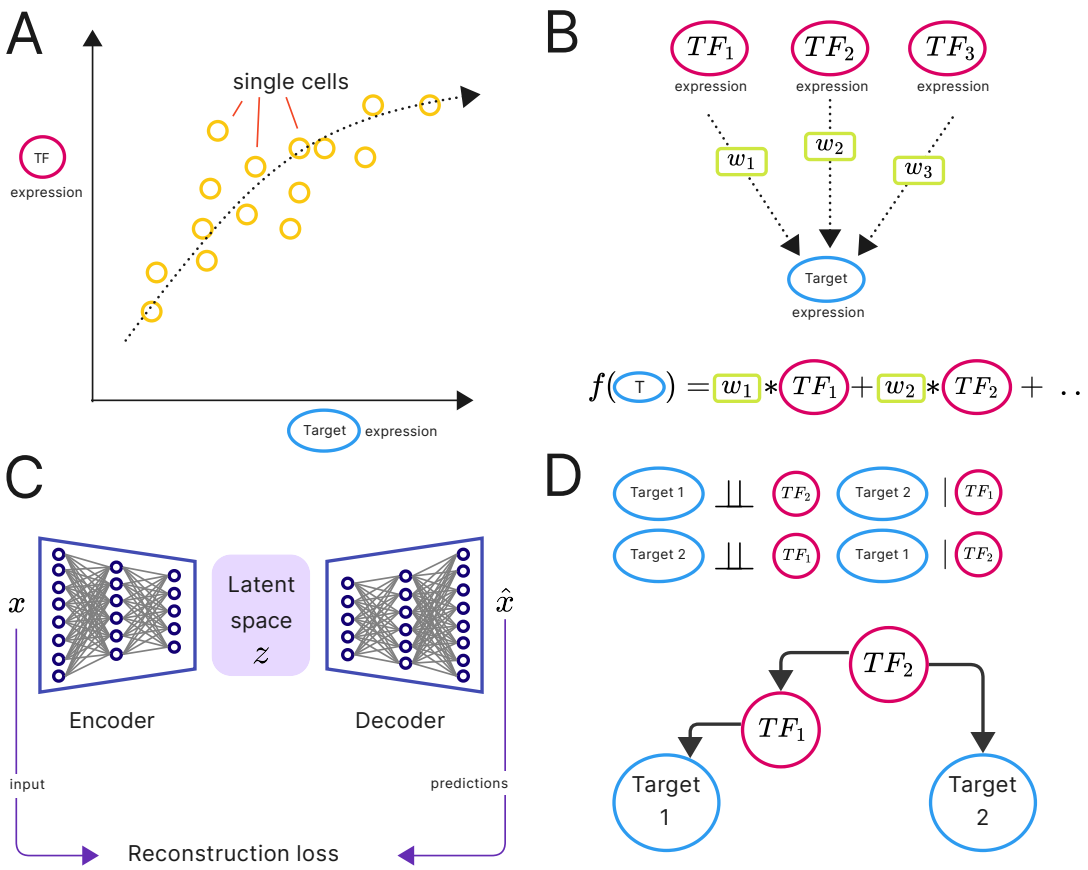


Figure 1.2: Schematic representation of categories of GRN inference algorithms. (A) Correlation-based algorithms (B) Linear regression as an example of regression-based algorithms (C) Autoencoder Model as an example of neural network algorithms (D) Bayesian Network as an example of Probabilistic Graphical Models. TF = Transcription factor, T = Target gene

Pearson's correlation coefficient and Spearman's rank correlation coefficient [83, 84].
Pearson's correlation coefficient [85], denoted by r_p, quantifies the strength of a linear relationship between two variables. It is defined as

$$r_p = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}, \tag{1.1}$$

where x_i and y_i are the expression levels of two genes across n samples (cells), as visualized in Figure 1.2A, and $\bar{x}$ and $\bar{y}$ represent their respective means. The coefficient ranges from -1 (indicating perfect negative linear correlation) to +1 (indicating perfect positive linear correlation), with 0 implying no linear correlation.

In contrast, Spearman's rank correlation coefficient, denoted by r_s, assesses the strength of a monotonic relationship between two variables. Unlike Pearson's coefficient, Spearman's ρ is based on the ranked values of the data rather than their raw values [86]. It is given by

$$r_p = \frac{\sum_{i=1}^{n} \left(R_{i1} - \bar{R_1}\right)\left(R_{i2} - \bar{R_2}\right)}{\sqrt{\sum_{i=1}^{n} \left(R_{i1} - \bar{R_1}\right)^2 \sum_{i=1}^{n} \left(R_{i2} - \bar{R_2}\right)^2}}, \tag{1.2}$$

where R_1 and R_2 are the ranks of the expression levels of two genes across n samples. This measure is particularly useful for detecting relationships that are monotonic but not necessarily linear.

It is important to note that while these correlation measures provide useful approximations for inferring GRNs, they only capture statistical associations rather than true causal relationships. As a result, networks constructed solely from correlation data often include indirect or second-order interactions as false positive edges [87]. For example, if transcription factor A regulates transcription factor B, which in turn regulates gene C, then the expression levels of A, B, and C are likely to be correlated, complicating the identification of the underlying causal structure. Moreover, because both correlation coefficients are symmetric, they do not indicate the direction of regulatory interactions, yielding an undirected network.

A commonly used tool that implements correlation-based network analysis in the context of GRNs is WGCNA (Weighted Correlation Network Analysis) [88], an R package designed to construct and analyze gene co-expression networks.

1.2.2. Regression

Regression models represent a fundamental class of methods for inferring gene regulatory networks (GRNs). In these models, the expression level of a target gene is typically predicted from the expression levels of potential regulators, mostly transcription factors [89, 90]. When a curated list of transcription factors is available, the regression model can focus on these regulators and yield more robust results. In contrast, for less well-studied organisms where such prior information is lacking, all genes may be considered as potential regulators, thereby increasing the model's complexity.

In general, the regression task for GRN inference involves modeling the expression level of a target gene as a function of the expression levels of potential regulators. Mathematically, consider a set of M samples (usually cells), where each sample j has a target gene expression y_j and regulator expression levels denoted by x_{ij} for regulator $i = 1, \ldots, N$. The linear regression model, as depicted in Figure 1.2B, is a simple regression model that predicts y_j by associating a weight W_i with each regulator in a linear combination. A common formulation for the predicted expression $\hat{y}_j$ is given by

$$\hat{y}_j = \sum_{i=1}^{N} x_{ij} W_i. \tag{1.3}$$

The objective is to determine the weights W_i such that the predicted values $\hat{y}_j$ closely match the

observed values y_j. A typical loss function is defined as the mean squared error (MSE) combined with a regularization term, which yields

$$\mathcal{L}(W) = \frac{1}{M}\sum_{j=1}^{M}\left(y_j - \hat{y}_j\right)^2 + L(W). \tag{1.4}$$

In this loss function, the first term $\frac{1}{M}\sum_{j=1}^{M}\left(y_j - \hat{y}_j\right)^2$ quantifies the discrepancy between the observed y_j and predicted $\hat{y}_j$ gene expression across all M samples (cells). The regularization term $L(W)$ is added to prevent overfitting by penalizing overly complex models; common choices include L_1 (Lasso) and L_2 (Ridge) regularization.
For example, the L_1 regularization term is defined as

$$L_1(W) = \lambda\sum_{i=1}^{N}|W_i|, \tag{1.5}$$

where λ is a hyperparameter that determines the strength of regularization and encourages sparsity in the weight coefficients, effectively performing feature selection. Similarly, the L_2 regularization term is given by

$$L_2(W) = \lambda\sum_{i=1}^{N}W_i^2. \tag{1.6}$$

Variations of linear regression models are widely used for GRN inference [91, 92], for example, in the popular CellOracle inference algorithm [93].

Random forest models are an example of more complex regression models commonly used for the GRN inference task [89, 94]. A random forest is an ensemble learning approach that constructs multiple regression trees during training. Each tree is built on a bootstrap sample of the data, and at every split, a random subset of predictors is considered. This randomness in both data sampling and feature selection helps to reduce the variance of the model and mitigates overfitting, as the predictions from individual trees are aggregated, typically by averaging, to yield a robust final prediction [95]. Random forests capture nonlinear relationships and interactions among the variables, while also providing measures of feature importance that can be valuable in GRN inference and used as weights for the edges of the network.
A notable example is GRNBoost2 [89], an optimized extension of the GENIE3 algorithm [72]. GRNBoost2 employs a random forest model to predict target gene expression from the expression of potential regulators. The use of random forest models in this context leverages their ability to handle high-dimensional data, capture nonlinear interactions, and provide robust performance with reduced computational cost.

1.2.3. Neural Networks

With the continuous expansion of available datasets, increasingly advanced machine learning techniques, including a variety of neural network architectures, have been applied to the inference of gene regulatory networks (GRNs). The networks are trained from large data using backpropagation, a gradient-based optimization method that computes the error gradient with respect to each parameter and iteratively updates the weights to minimize a chosen loss function [96].

The neural network architectures used for GRN inference extend beyond simple multilayer perceptrons and include, for example, autoencoder networks. Autoencoders, as depicted in Figure 1.2C, are trained to reconstruct the input data via a lower-dimensional latent space, thereby serving as an effective method for dimensionality reduction. An autoencoder is composed of two main parts: an encoder and a decoder. The encoder, defined as a function $f(x) = z$, compresses the high-dimensional input data x into a latent representation z. Conversely, the decoder, defined as $g(z) = \hat{x}$, attempts to reconstruct the original input from the latent representation. The network is trained using backpropagation to minimize a reconstruction loss. This is typically the mean squared error (MSE) for regression tasks or cross-entropy loss for classification tasks thereby ensuring that the output $\hat{x}$ closely approximates the input x. For instance, the mean squared error is defined as

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{x}_i)^2 , \tag{1.7}$$

where x_i is the true value of the i^{th} of N samples, $\hat{x}_i$ is the corresponding reconstructed value, and N denotes the total number of samples. In the context of scRNA-Seq data, this can be used to learn the reconstruction of the gene expression values [97]. For binary classification problems, which is commonly used for the binary prediction of edges in GRNs [73, 98], the binary cross-entropy loss (BCE) is given by

$$\text{BCE} = -\frac{1}{N} \sum_{i=1}^{N} \left[x_i \log(\hat{x}_i) + (1 - x_i) \log\left(1 - \hat{x}_i\right) \right] , \tag{1.8}$$

where $x_i \in 0, 1$ represents the true binary label for the i^{th} sample and $\hat{x}_i$ is the predicted probability of the positive class (regulatory edge exists in GRN).
The bottleneck imposed by the reduced dimensionality prevents the network from merely learning the identity function, which encourages the extraction of salient features that generalize well to unseen data. Although autoencoders are primarily trained in an unsupervised manner, the learned latent representations can subsequently be utilized for downstream supervised tasks.

A prominent generative extension of the standard autoencoder is the variational autoencoder (VAE) [99], in which the latent space is further constrained to follow a specified probability distribution, most commonly a Gaussian. In a VAE, instead of encoding each input as a single fixed

vector, the encoder produces parameters of a probability distribution (typically the mean and variance) describing the latent space. The training objective of a VAE comprises two terms: the reconstruction error and a regularization term. The latter is formulated as the Kullback-Leibler (KL) divergence, which measures the difference between the approximate posterior distribution $q(z \mid x)$, where z represents the latent variable, and a predefined prior distribution $p(z)$ (often chosen as a Gaussian with zero mean and standard deviation of 1). This regularization term encourages $q(z \mid x)$ to be close to the prior $p(z)$, thereby enhancing the model's generative and generalization capabilities.

In the context of GRNs, the graph structure plays a critical role because it captures the underlying regulatory interactions among genes. Treating each edge independently can result in the loss of valuable contextual information provided by the network topology. Several strategies have been developed to incorporate graph structure into neural network models. One method involves embedding the graph into a feature space using algorithms such as `node2vec` [100] and subsequently processing these features with conventional neural networks. Alternatively, message-passing neural networks (MPNNs) explicitly take advantage of the graph's connectivity by guiding the exchange of information between neighboring nodes.

Notable examples of such frameworks include graph convolutional networks (GCNs) [101] and graph attention networks (GATs) [102]. Graph Convolutional Networks (GCNs) are designed to perform convolution operations over graph-structured data. In a GCN, the representation of each node is updated by aggregating and transforming the features of its neighbors. A common formulation for a GCN layer is given by

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}\right), \tag{1.9}$$

where $\tilde{A} = A + I$ is the adjacency matrix A of the graph with added self-loops (with I being the identity matrix), $\tilde{D}$ is the degree matrix corresponding to $\tilde{A}$, $H^{(l)}$ is the matrix of node features at layer l, $W^{(l)}$ is the learnable weight matrix at layer l, and σ is a nonlinear activation function such as the rectified linear unit (ReLU). This formulation normalizes the aggregation, ensuring that each node's representation is updated based on a weighted average of its neighbors' features.

Graph Attention Networks (GATs) extend this concept by incorporating an attention mechanism that assigns different weights to different neighbors, thus allowing the network to focus on the most informative parts of the graph. For a pair of nodes (genes) i and j, the attention coefficient α_{ij} is computed as

$$\alpha_{ij}^{(l)} = \frac{\exp\left(\sigma\left(a^{(l)^T}[W^{(l)}h_i^{(l)} \,\|\, W^{(l)}h_j^{(l)}]\right)\right)}{\sum_{k\in\mathcal{N}(i)}\exp\left(\sigma\left(a^{(l)^T}[W^{(l)}h_i^{(l)} \,\|\, W^{(l)}h_k^{(l)}]\right)\right)}, \tag{1.10}$$

where l indicates the current layer, h_i and h_j are the feature vectors for nodes i and j, respectively, W is a shared linear transformation applied to the node features, a is a learnable weight vector that parametrizes the attention mechanism, $\|$ denotes concatenation, and $\mathcal{N}(i)$ represents the set of neighboring nodes of node i. The resulting node feature vector $h_i^{(l+1)}$ is then computed as a weighted sum of the transformed features of its neighbors:

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)}\right), \tag{1.11}$$

with σ being a nonlinear activation function. Notably, transformer models [103], which have significantly advanced the field of natural language processing, rely on a similar attention mechanism.

Graph neural networks (GNNs) have been applied to a wide range of tasks, including graph classification, node classification, and link prediction. For GRN inference, link prediction is particularly relevant, as it involves determining the presence or absence of regulatory interactions between genes. Moreover, integrating GNN layers into an autoencoder architecture yields graph autoencoder models [104]. These models have shown promising results in the link prediction setting for GRN inference, as exemplified by methods such as `GENELINK` [73] and `GNNLINK` [98].

1.2.4. Probabilistic graphical models

Another class of algorithms for inferring gene regulatory networks (GRNs) relies on probabilistic graphical models, which inherently account for the network structure of the problem. Probabilistic graphical models represent a broad class of statistical frameworks that explicitly capture complex dependency structures among variables. In the context of GRNs, this allows one to represent genes as nodes and their regulatory interactions as edges, thereby providing a natural and interpretable framework for modeling the underlying biological processes.

These models can be broadly divided into two categories: directed models and undirected models. Directed models encode asymmetric relationships by specifying a direction for each edge, while undirected models, exemplified by Markov random fields capture symmetric dependencies without imposing a direction. Each category has its own way of factorizing the joint probability distribution over all variables [105].

A prominent example of a directed model is the Bayesian network [106] (see Figure 1.2D). In a Bayesian network, each node denoted by x_i and representing a gene is connected by directed edges that encode probabilistic dependencies. The network is structured as a directed acyclic graph (DAG), meaning that there are no cycles in the graph. This structure ensures that each gene is modeled to be conditionally independent of its non-descendant nodes given its parent nodes. This conditional independence assumption allows the joint probability distribution of n genes to be factorized as follows:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{Parents}(x_i)), \tag{1.12}$$

where $\text{Parents}(x_i)$ denotes the set of genes that have a direct regulatory influence on x_i. In this factorization, each conditional probability term $P(x_i \mid \text{Parents}(x_i))$ models the likelihood of gene x_i given the expression of its regulators, thereby capturing causal or directional regulatory relationships.

In contrast, undirected models such as Markov random fields (MRFs) [107, 108] represent the joint distribution of genes without assuming any inherent directionality. Instead of conditional probability distributions based on a DAG, MRFs express the joint probability distribution using potential functions defined over cliques (fully connected subgraphs) of the network. The joint distribution in an MRF is typically written as:

$$P(x_1, x_2, \ldots, x_n) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C), \tag{1.13}$$

where $\mathcal{C}$ is the set of all cliques in the graph, $\psi_C(x_C)$ is the potential function associated with clique C (which encapsulates the compatibility among the genes in C), and $Z = \sum_{x_1, x_2, \ldots, x_n} \prod_{C \in \mathcal{C}} \psi_C(x_C)$ is the partition function that normalizes the distribution.

The inherent flexibility of probabilistic graphical models makes them highly relevant for modeling GRNs. Directed models, such as Bayesian networks offer the advantage of representing clear causality in the edges of the graphical model, a highly desired property for GRNs, while undirected models, such as MRFs, do not. However, Bayesian networks assume a DAG, which is a constraint that is not necessarily correct for known GRNs, which are characterized to also contain cycles [109, 110].

1.3. Priors for incorporation

Recent studies have demonstrated that most standard GRN inference algorithms from scRNA-Seq data perform poorly and often fail to robustly outperform even simple linear correlation methods [83, 111]. One promising strategy for improving GRN inference is to integrate additional information beyond gene expression alone. By incorporating complementary information, deeper insights into gene regulation can be achieved, which in turn can significantly enhance the accuracy of GRN construction. For example, the integration of multi-omics data can narrow down the solution space by providing constraints on which regulatory interactions are possible and which should be excluded [9]. These complementary datasets can validate or invalidate potential interactions inferred from expression data, thereby reducing ambiguity and focusing the search on biologically plausible connections.

Additional data sources may include protein-protein interactions, curated knowledge from databases, priors on the network topology and high-throughput experimental techniques, usually referred to as multi-omic approaches. Protein-protein interaction can help to find protein complexes involved in regulatory interactions [112]. Curated prior knowledge, usually in the form of abstract weighted regulatory links, can be derived from several databases. Popular examples are ENCODE [113], Biogrid [114] and the yeast specific YEASTRACT [115]. Moreover, priors derived from known network topologies, such as expected node-degree distributions and modularity, contribute unique constraints that help exclude unlikely regulatory links, further refining the inference process [116, 117].

Specific high-throughput experimental techniques play a critical role in directly informing transcription factor–target relationships. Single-cell Perturb-Seq experiments, for example, combine genetic perturbations with RNA sequencing at the single-cell level, enabling systematic inference of regulatory effects by observing changes in gene expression following targeted perturbations [112, 118–120]. Chromatin immunoprecipitation followed by sequencing (ChIP-Seq) can be used to identify in vivo binding sites of transcription factors across the genome, furnishing location-specific evidence of regulatory interactions [121, 122]. The same technique can also be used to quantify histone modification abundance for CRE identification [67]. Similarly, (single cell) ATAC-Seq assesses chromatin accessibility, thus highlighting potential regulatory regions such as enhancers and promoters that are available for binding [123]. Hi-C and related chromatin conformation capture techniques reveal the three-dimensional organization of the genome, linking distant regulatory elements with their corresponding target genes [67, 121]. These diverse experimental approaches provide complementary layers of evidence that, when integrated, create a more comprehensive picture of the regulatory landscape.

A common experimental strategy involves pairing scRNA-Seq with single-cell ATAC sequencing (scATAC-Seq) data, often supplemented by known transcription factor motifs from databases.

When scRNA-Seq and scATAC-Seq data are obtained from the same cell, for example, using 10X Multiome technology, each cell is characterized by both a transcriptomic and an epigenomic profile, enabling a direct integration of gene expression with chromatin accessibility. This concurrent measurement facilitates the mapping of accessible cis-regulatory elements (CREs) to the expression levels of potential target genes. In contrast, when these datasets are collected in separate experiments, computational methods must be employed to integrate and pair them effectively. Techniques such as canonical correlation analysis (CCA), mutual nearest neighbors (MNN), or other alignment algorithms are typically used to match cells across modalities based on shared features [124, 125]. In a standard workflow, accessible CREs are identified via ATAC-Seq peaks and then linked to target genes by correlating the accessibility levels of the CREs with the expression levels of nearby genes. Promoters are typically annotated based on their proximity to the transcription start site (TSS) of the gene, whereas enhancers are inferred from their correlation with gene expression [67, 126–129]. Subsequently, these regulatory elements are cross-referenced with transcription factor motif databases to assign potential regulators with high binding affinity. By combining the inferred transcription factor–CRE associations with the CRE–target gene links, transcription factors can be connected to their potentially regulated target genes, thereby tightening the constraints on the predicted network.

Priors may be categorized based on how they are incorporated into the inference process. One possible classification distinguishes three types of priors based on the timing of incorporation in the inference process [116]. The first type includes those that are incorporated as inputs to the inference process (e.g., experimental priors). The second type consists of priors integrated during the inference process (e.g., structural priors such as topological constraints). Lastly, there are priors that can be applied post hoc to refine the inferred network, such as by fine-tuning the network and removing false positive edges. Although this post-inference adjustment is less commonly used in published algorithms, it holds the potential for further improving GRN inference.

A notable challenge for all inference methods is the correlation pitfall arising from regulatory cascades [130]. This applies, for example, to a scenario in which transcription factor A regulates transcription factor B, which in turn regulates target gene C. Many algorithms struggle to determine whether a direct regulatory interaction from A to C exists or if the observed correlation is solely due to the indirect effect via B, despite the correlated expression values. Post hoc fine-tuning of the network to eliminate such false positive "skip" connections represents an additional approach to integrating prior knowledge into the GRN inference process.

1.4. Evaluation of gene regulatory network inference

An essential criterion for selecting a gene regulatory network (GRN) inference algorithm from the wide range of available methods is an objective performance evaluation, as demonstrated by independent benchmarking studies. In this section, we first present the statistical metrics commonly used for this purpose and then discuss topological properties that are not directly captured by these statistical measures. Finally, we describe the unique setting of GRN inference as a positive-unlabeled problem and examine its implications.

1.4.1. Statistical evaluation metrics

A standard method for evaluating gene regulatory network (GRN) inference performance is to frame the task as a binary classification problem [83, 111]. In this formulation, every possible edge in the network is assigned one of two ground truth classes: either a regulatory interaction exists, or it does not. The output of an inference algorithm typically consists of either logits or probabilities for each edge. By applying a threshold and with the help of a ground truth GRN, one can construct a binary confusion matrix that reports the number of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) edges. From this confusion matrix, common performance metrics, such as precision, recall, and the F1-score, can be calculated.

$$precision = \frac{TP}{TP + FP} \tag{1.14}$$

$$recall = \frac{TP}{TP + FN} \tag{1.15}$$

$$F_1 = 2\frac{precision \cdot recall}{precision + recall} \tag{1.16}$$

This evaluation strategy is particularly relevant when only a fixed number of edges are of interest. For instance, in GRN inference, the top-scoring predicted edges might be selected for validation in wet lab experiments, which can only test a limited number of edges. In such cases, the threshold is set to select the required number of top edges and a metric known as Precision at top k can be reported [131].

Alternatively, when the overall performance of the model is of interest, the threshold becomes an arbitrary parameter. To address this issue, area under the curve (AUC) metrics are commonly employed because they summarize performance across all possible thresholds into a single value. Two popular metrics in this category are the area under the precision-recall curve (AUPR) and the area under the receiver operating characteristic curve (AUROC). The AUPR represents the integral of the precision versus recall curve, capturing the trade-off between these metrics at various thresholds. The AUROC, on the other hand, is computed as the area under the curve that plots the true positive rate (TPR) against the false positive rate (FPR). Both the AUPR and AUROC values range between 0 and 1, with higher values indicating better performance. It

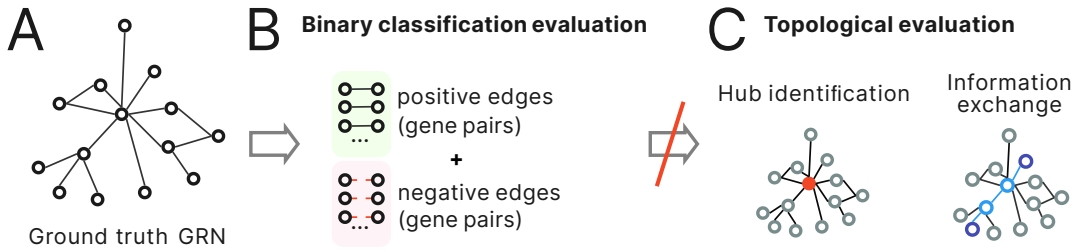


Figure 1.3: Different evaluation metrics for GRN inference (A) Example for a ground truth GRN for evaluation (B) When framed as a binary classification problem, the gene pairs (all possible edges) of the graph are split into positives and negatives. It loses the information about the graph structure. (C) The identification of hubs and general information exchange metrics in a topological evaluation provides complementary information on the performance not captured by the binary classification approach

is important to note that AUPR is sensitive to class imbalance, resulting in lower baseline scores for random predictors in scenarios with high imbalance, whereas a random predictor would yield an AUROC of 0.5. In many studies, both metrics are reported since they reflect different aspects of the classification performance and can lead to different model rankings and none of them are clearly superior [132].

In the above metrics, each edge is treated with equal importance; such measures are referred to as edge-equality metrics. Alternatively, some metrics take into account the degree of nodes by computing a metric for each node and then aggregating these node-level scores over the entire graph [131]. These are termed node-equality metrics because they assign equal weight to every node. For example, the mean average precision (MAP) metric computes the average precision for the connections of each node individually and then averages these values across all nodes. This approach is particularly relevant in GRN evaluation since GRNs often exhibit scale-free degree distributions, where a few hub transcription factors dominate edge-equality metrics. Despite their potential advantages, node-equality metrics are less commonly used in GRN evaluation studies.

1.4.2. Topological properties of a graph

Since both node- and edge-equality metrics introduced above treat every edge or node neighborhood as a separate entity for evaluation purposes, they do not capture the different structural features of the regulatory graphs, as visualized in Figure 1.3. However, these are biologically an important aspect of the regulatory network, since the structure is, for example, predictive of the stability of the GRN [133]. The structural organization can be quantified by different topological properties of a graph. These properties can be computed at both the node level and the network level.

At the node level, various centrality measures are used to quantify the importance of a node within the network. The simplest and most common centrality is the degree centrality. For a

graph with n nodes and without self-loops, the degree centrality of a node v is defined as

$$C_{Degree}(v) = \frac{\deg(v)}{n-1} \ , \tag{1.17}$$

where $\deg(v)$ represents the number of edges incident to v. The normalization by $n-1$ indicates the fraction of possible connections that node v actually has.

Another important measure is betweenness centrality [134]. This metric quantifies the frequency with which a node lies on the shortest paths connecting pairs of other nodes. It is defined as

$$C_{Betweenness}(v) = \frac{1}{p_v} \sum_{u \neq v \neq w} \frac{\eta_v(u,w)}{\eta(u,w)} \ , \tag{1.18}$$

where $\eta(u,w)$ denotes the total number of shortest paths between nodes u and w, $\eta_v(u,w)$ is the number of these paths that pass through node v, and p_v is a normalization constant, typically chosen so that the maximum possible value is 1.

A further measure is radiality centrality [135], which assesses how well a node is connected to all other nodes in the network. It is given by

$$C_{Radiality}(v) = \max_{x,y \in N}, d(x,y) + 1 - \frac{1}{n-1} \sum_{w \in N \ w \neq v} d(v,w) \ , \tag{1.19}$$

where $d(v,w)$ denotes the shortest path distance between nodes v and w, and N is the set of all nodes in the graph. The term $\max_{x,y \in N} d(x,y)$ represents the maximum shortest path distance in the network, and the addition of 1 ensures that the measure is properly normalized.

Nodes that exhibit high centrality values are generally referred to as hubs. Although different centrality measures are often positively correlated, the choice of metric can alter the ranking of nodes. In the context of gene regulatory networks, hubs are typically transcription factors acting as master regulators and thus are of particular interest for further analysis.

At the network level, several global topological properties can be evaluated. One common metric is the average shortest path length, which quantifies the typical number of hops between any two nodes in the graph. It is defined as

$$\overline{l_{sp}}(G) = \frac{1}{n \cdot (n-1)} \sum_{v,w \in N} d(v,w) \ , \tag{1.20}$$

where $d(v,w)$ is the shortest distance between nodes v and w, and $n = |N|$ is the total number of nodes in the network G.

Another global metric is global efficiency, which quantifies the capability for information exchange in the network. Global efficiency is defined as

$$E_{glob}(G) = \frac{1}{n \cdot (n-1)} \sum_{v,w \in N \; v \neq w} \frac{1}{d(v,w)} \; . \tag{1.21}$$

Here, the inverse of the shortest path length, $\frac{1}{d(v,w)}$, measures the efficiency of information transfer between nodes v and w. A higher value indicates better connectivity.

In addition to the overall global efficiency, local efficiency can be computed for subsets of the network. For each node v, the subgraph G_v induced by its direct neighbors is considered, and the global efficiency of this subgraph, $E_{glob}(G_v)$, is computed. The local efficiency of the network is then given by

$$E_{loc}(G) = \frac{1}{n} \sum_{v \in N} E_{glob}(G_v) \; . \tag{1.22}$$

To assess the network structure in terms of hub formation, several measures are used. One such metric is degree centralization [136], which focuses on the extent to which the network is dominated by the node with the highest degree. It is defined as

$$H(G) = \frac{1}{H_{max}} \sum_{v \in N} \left[\deg(v^*) - \deg(v) \right] \; , \tag{1.23}$$

where v^* is the node with the maximum degree and H_{max} is a normalization factor representing the maximum possible sum of degree differences, given by

$$H_{max,\text{undirected}} = (n-1)(n-2) \; , \tag{1.24}$$

for undirected graphs, and

$$H_{max,\text{directed}} = (n-1)^2 \; , \tag{1.25}$$

for directed graphs.

Degree assortativity [137] provides another perspective by quantifying the tendency of nodes to connect with other nodes of the same degree. This measure is defined as

$$r_{deg}(G) = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} \; , \tag{1.26}$$

where e_{ii} is the fraction of edges connecting nodes of degree i to nodes of degree i, and a_i and b_i denote the fraction of edges incident on nodes with degree i at the two ends of an edge, respectively.

The clustering coefficient [138] measures the degree to which nodes in a graph tend to form

clusters. The local clustering coefficient for a node v is defined as

$$CC_{\text{loc}}(v) = \frac{2 \cdot L_v}{k_v(k_v - 1)} \ , \tag{1.27}$$

where L_v is the number of edges between the immediate neighbors of v, and k_v is the degree of v (with $k_v \geq 2$). The global clustering coefficient of the network is then obtained by averaging the local coefficients:

$$CC_{\text{glob}}(G) = \frac{1}{n} \sum_{v \in N} CC_{\text{loc}}(v) \ . \tag{1.28}$$

Other structural metrics include the frequency of specific network motifs, such as feedback and feedforward loops, which can provide additional insights into the network's regulatory architecture.

1.4.3. Positive Unlabeled setting

Gene regulatory network (GRN) inference is often evaluated by framing the task as a binary classification problem and using area-under-the-curve (AUC) metrics. However, this approach is a simplification that comes with several caveats. Besides the missing quantification of graph topology discussed above, the standard binary classifiers require a ground truth label for every edge in order to compute a confusion matrix, which includes counts of true positives, false positives, true negatives, and false negatives.

In the case of GRN inference, the ground truth is considered to be incomplete. In standard binary classification, it is assumed that every potential regulatory interaction (edge) is labeled either positive or negative. In GRN inference some regulatory interactions are well supported by validation experiments and can be reliably labeled as positive. In contrast, the absence of an edge in the current dataset does not necessarily indicate a true negative, as the interaction may simply have not been validated yet [139, 140]. This ambiguity can lead to misrepresentation in the confusion matrix by mistaking unvalidated true interactions for negatives, thus diluting the reliability of metrics such as precision and recall. The situation where only positive examples are confidently known while negative labels remain uncertain is referred to as a positive-unlabeled (PU) setting, a variant of semi-supervised learning [141–143].

In a PU setting, only confidently validated positive interactions are available as labels, while the remaining unlabeled edges may include both undetected true negatives and actual false negatives. This heterogeneous composition of the unlabeled set can skew standard performance metrics. For example, if a predicted interaction is not present in the current database, it might be erroneously counted as a false positive. Such misclassification inflates the false positive rate, yielding a lower precision value even when some of these predicted interactions are, in reality, true regulatory relationships that have not yet been validated. Therefore, this issue necessitates the

development or adoption of tailored evaluation metrics that can adjust for the uncertainty in labeling, such as those employed in PU learning frameworks or modified precision-recall estimations.

Although for certain well-studied cell types the available GRN information is more comprehensive, it remains debatable whether the corresponding network can be considered entirely correct. Thus, instead of referring to it as the definitive ground truth, researchers typically describe these datasets as "gold standard" [144, 145] or "silver standard" [146] to acknowledge their inherent uncertainty.

Another way to eliminate this uncertainty from benchmarking settings is the use of synthetic data. In this case, the ground truth graph is given and the transcriptomic data is generated from it in silico. However, this approach has other drawbacks. The quality of the simulated scRNA-Seq data depends on the degree of realism of the simulation tool used to generate it from the ground truth graph. This leads to the same problem again, that evaluation scores, obtained with the pair of ground truth network and transcriptomic data, contain a degree of uncertainty [147].

1.5. Research questions

Gene regulatory network (GRN) inference holds significant promise for constructing mechanistic models that drive advances in areas such as drug discovery. Consequently, the development of more advanced inference algorithms is a crucial step toward building reliable mechanistic models. In this thesis, the following research questions are addressed:

I) To what extent do existing GRN inference algorithms accurately recover the underlying network topology?

II) How can the frequently moderate performance of GRN inference be improved by incorporating additional prior knowledge?

III) How can we establish fair benchmarking protocols for comparing inference algorithms that utilize different types of prior knowledge?

1.6. Summary of results

Core publications as main author

- [1] in section 2.1:

  Topological benchmarking of algorithms to infer gene regulatory networks from single-cell RNA-seq data

  Several benchmarking studies have evaluated published algorithms for GRN inference using the statistical evaluation metrics introduced earlier. However, the capability of the inference algorithm to correctly reconstruct topological properties in their inferred GRNs has not been evaluated for scRNA-Seq GRN inference yet. The correct topology of the predicted GRNs is biologically important, for example for the prediction of the robustness of the GRN. We developed STREAMLINE, a scalable benchmarking pipeline that is compatible with BEELINE and is designed to assess how accurately the inferred GRN captures the overall network topology, providing insights into properties such as network stability. By applying STREAMLINE to commonly used GRN inference algorithms, we demonstrate systematic biases in the topological features recovered by different algorithms. Our results indicate that there is potential to improve performance by integrating prior information about the expected network topology during the inference process.

- [2] in section 2.2:

  Leveraging prior knowledge to infer Gene Regulatory Networks from single-cell RNA-sequencing data

  Numerous methods for GRN inference have been published, and more recently, many algorithms have sought to leverage additional prior knowledge in order to enhance inference performance. Existing reviews have focused on specific types of priors, missing to provide a unifying overview of all possible types of priors. Furthermore, directly comparing methods with different types of priors has proven challenging, since the overall performance can not easily disentangled into the performance gain from prior and the intrinsic algorithm performance itself. In this paper, we provide a comprehensive summary of the available sources of prior knowledge and a categorization of the inference methods that incorporate graph representations of such priors. Moreover, based on this categorization, we propose a new benchmarking scheme that enables fair testing of these methods within each group. Our comprehensive review provides guidance for both users and developers of GRN inference algorithms.

2. Methods: Summary of Contributed Articles

This chapter summarizes the contributed articles and my individual contributions.

2.1. Topological benchmarking of algorithms to infer gene regulatory networks from single-cell RNA-seq data

By combining different graph topology metrics described in section 1, we built an easy-to-use benchmarking pipeline to analyze the topological accuracy of inferred gene regulatory networks (GRNs), called STREAMLINE [1] and applied it to evaluate the capability of commonly used GRN inference methods to correctly infer different network topologies.

Research problem

Available benchmarking studies evaluate the performance of GRN inference algorithms most commonly by treating the problem as a binary classification problem and calculating statistical performance metrics for this task, such as the area under the precision-recall curve (AUPR) or the area under the receiver operator curve (AUROC) [83, 111]. However, this evaluation metric is maximized by correctly predicting the maximum number of links. It treats every individual edge equally, thus not evaluating the network structures emerging from the set of predicted edges, which are known to show interesting topological features, such as a scale-free node degree distribution, an important property for the robustness of regulatory networks.

Approach

We set up a benchmarking pipeline, as shown in Figure 2.1A, to compare the network topology of predicted gene regulatory networks to ground truth graphs. For this purpose, different node-based and network-based metrics are used. The evaluation is based on synthetic as well as experimental scRNA-Seq data.

The pipeline is compatible with the BEELINE benchmarking framework [111] for statistical evaluation of GRN inference algorithms from single-cell RNA-sequencing data. This allows for easy integration of STREAMLINE within the benchmarking community built around the popular BEELINE framework.

Results

STREAMLINE results on the topological performance of predicted GRNs from different algorithms show little correlation with the statistical performance of the same networks. This indicates that good statistical performance, i.e., recovery of correct edges in general, is not necessarily linked to recovery of the overall network topology (Figure 2.1B). On the other hand, even networks with lower edge-level recovery can still provide better predictions of the network topology, which might be interesting for some downstream analyses, such as estimating the perturbation stability of the GRN. Overall, no single algorithm consistently outperforms others in terms of the topological accuracy of the inferred GRN, as the ranking of the algorithms depends

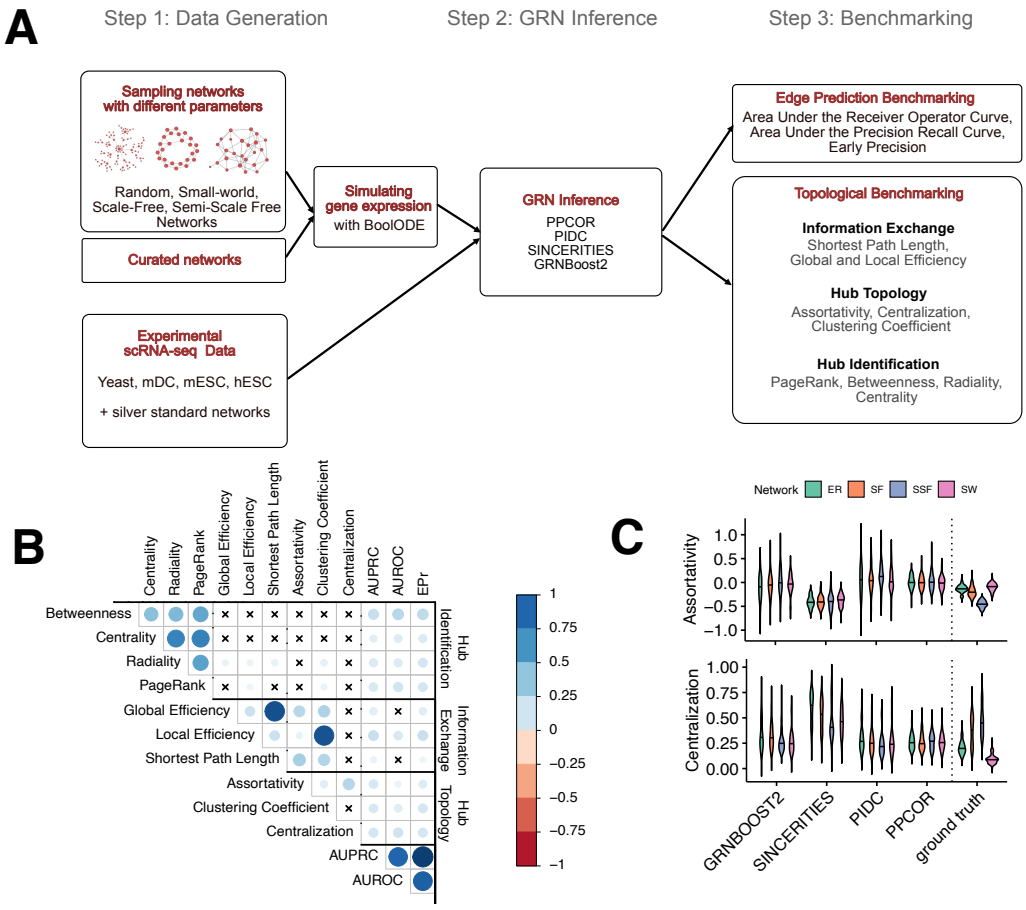


Figure 2.1: Summary of results for STREAMLINE benchmarking pipeline taken from [1] (A) Overview of STREAMLINE benchmarking pipeline for topological evaluation of gene regulatory network inference algorithms. (B) Spearman correlation coefficient shows little correlation between different topological evaluation metrics and statistical edge-based metrics (AUROC, AUPRC, Early Precision *EPr*). (C) Centralization and Assortativity for inferred GRNs of different algorithms and the ground truth networks to be inferred (ER = Erdős–Rényi, SF = Scale-free, SSF = Semi-scale-free, SW = Small world) reveal algorithm-specific biases.

strongly on the specific metric and dataset evaluated. STREAMLINE also reveals algorithmic biases in the predicted networks that are independent of the actual GRN target to be inferred. For example, the predicted networks by the SINCERITIES algorithm [148] generally overestimate the centralization and underestimate the assortativity of the ground truth network (Figure 2.1C). Another example is GRNBoost2 [89], which generally underestimates the global efficiency of GRNs.

Conclusion

The results of STREAMLINE reveal topological biases and a lack of performance in the topological accuracy of predicted GRNs by currently available inference algorithms. However, this lack of sensitivity to the topology of GRN also offers an opportunity to improve network predictions by incorporating prior information about the network topology, such as the commonly assumed

scale-free degree distribution.

Individual contribution: Antonio Scialdone, Jonathan Fiorentino, and I had the idea to investigate the topology of inferred GRNs from different inference algorithms. Together, we developed the scope of STREAMLINE. I implemented the code of the benchmarking pipeline in Python. I authored the evaluation scripts and the visualizations. I optimized the speed of the pipeline to enable running a large set of algorithms on large scRNA-Seq datasets. Finally, I wrote the manuscript and created the figures of the manuscript together with Antonio Scialdone, Jonathan Fiorentino, and Niclas Popp.
I am the shared main author of this publication with Niclas Popp.

2.2. Leveraging prior knowledge to infer Gene Regulatory Networks from single-cell RNA-sequencing data.

This review summarizes the available types and sources of prior knowledge that can be leveraged in GRN inference and proposes a new framework for graph-based priors that enables standardized comparisons of different algorithms and prior types.

Research problem

Recently, many algorithms for GRN inference have been developed, however, their performance is not optimal [1, 83, 111]. Therefore, the research question arises of how to improve their moderate performance. Instead of optimizing the algorithms themselves, one option is to add additional data input to the inference problem. For these algorithms leveraging prior knowledge, the lack of a comprehensive overview of current approaches hinders both the application and improvement of these algorithms. Moreover, the missing overview leads to a lack of unbiased benchmarking of their performance, as well as the improvement gained from each prior.

Approach

We systematically review available types and sources of prior knowledge for the inference process, as well as different modes of incorporation by current algorithms. Based on this comprehensive review, we then classify the algorithms and propose a novel framework for unbiased benchmarking of these algorithms.

Results

In the paper, we characterize 12 different experimental sources of prior knowledge for the inference process of GRNs. Based on a set of representative algorithms that leverage the different sources of prior, we then propose a novel classification of the algorithms: those that use the knowledge directly in a data-driven way and those that, in the first step, extract a graph prior representation from the data before the actual GRN inference step (Figure 2.2A). The relevant

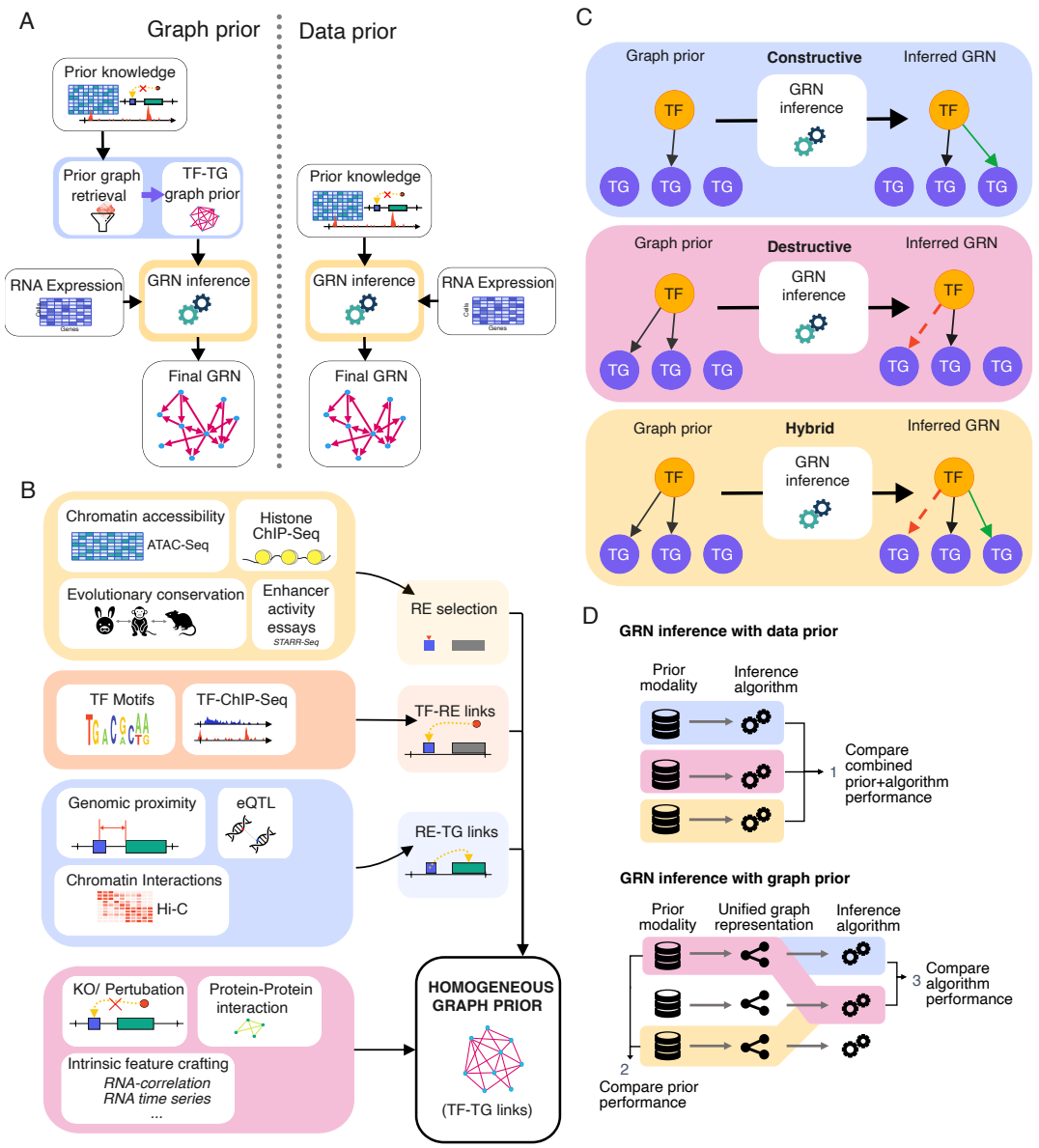


Figure 2.2: Summary of main results taken from [2]. (A) shows a schematic for the two different ways to incorporate data as prior knowledge. Graph priors are graph representations built from the data as an intermediate step before GRN inference. (B) summarizes different ways to build transcription factor to target graph priors from data. (C) shows the different ways to use graph priors in inference algorithms by adding (constructive) or removing (destructive) edges of the graph prior. Hybrid algorithms add and remove simultaneously. (D) visualizes the possibility of disentangling prior from algorithmic performance when using graph priors. TF = Transcription factor, TG = Target gene, RE = Regulatory element

experimental sources that can be used to build such a prior graph are visualized in Figure 2.2B. We then split these graph prior-based algorithms further into three operation modes, based on their preference to either add (constructive) or remove (destructive) edges from the prior graph or do both (hybrid). Schematics for each operation mode can be found in Figure 2.2C. Finally, we propose a new benchmarking framework for algorithms leveraging graph priors that

allows disentangling the contribution to performance due to the prior from the one due to the inference algorithm (see Figure 2.2D). Without the unifying graph representation, only the combination of the algorithm and respective input prior type can be benchmarked together. On the other hand, with the unifying graph prior representation, a modular benchmarking can be achieved, that separately evaluates the contribution of the algorithm and the prior to the performance. This allows for a more detailed analysis of the performance driver in the inference of GRNs.

Conclusion

We present the first comprehensive overview of algorithms leveraging different types of prior knowledge for GRN inference from single-cell transcriptomics, including topological priors. We outline available sources of prior knowledge for GRN inference and categorize existing algorithms based on their incorporation mode of prior knowledge. We further propose a new benchmarking framework for algorithms leveraging prior knowledge. This work aims to assist users in choosing algorithms and to help developers improve future inference algorithms.

Individual contribution: Antonio Scialdone and I had the idea of systematically reviewing the many approaches developed and used by the GRN inference community. We wanted to provide an overview of the available algorithms and types of priors to both the community of users and developers. I came up with the idea of a graph-based benchmarking framework that would allow for unbiased comparisons of the contribution of different types of prior and different algorithms to the overall performance. I produced the figures and I wrote most of the manuscript.
I am the main author of this publication.

3. Discussion

In this thesis, I have established a benchmarking pipeline to evaluate the topological properties of gene regulatory networks (GRNs). The pipeline is designed to be compatible with previous benchmarking efforts, thereby enabling direct comparisons across studies and facilitating reproducibility. Beyond offering a standardized framework, the results provide practical guidance to users of popular inference algorithms by indicating which methods are most promising for specific metrics of interest and under which data regimes they tend to perform reliably.

A central finding is that topological performance results are only weakly correlated with traditional binary evaluation outcomes, such as edge-level area under the receiver operating characteristic curve (AUROC) or the area under the precision-recall curve (AUPRC). This discrepancy underscores the necessity of explicitly assessing topological metrics, capturing properties such as information exchange and hub identification alongside binary correctness. Notably, for some metrics, including hub identification, several methods outperform random predictors, indicating that meaningful higher-order structure can be recovered even when edge-level performance appears modest. Taken together, these observations argue for a multi-criteria evaluation strategy and caution against relying exclusively on binary metrics when selecting or tuning GRN inference methods.

I also confirm systematic differences between topological performance in inference from experimental versus synthetic data. Although synthetic networks offer perfectly known ground truth, the generation of single-cell RNA-sequencing data relies on simplified ordinary differential equation models and stylized noise processes that may not fully capture the heterogeneity, nonlinear dynamics, and context-specific regulation present in real systems. Conversely, experimental ground truths are often incomplete or condition-specific, even for comparatively well-studied organisms such as yeast, which complicates objective evaluation and may penalize correct but undocumented edges. As a result, it is not straightforward to claim that either synthetic or experimental benchmarks provide a universally more objective estimate of inference performance. Bringing synthetic and experimental results into closer alignment will require improved RNA-seq simulation frameworks that better match empirical distributions and regulatory dynamics, together with continued curation and expansion of high-confidence ground truth networks, ideally incorporating perturbation evidence.

Finally, the topological analyses reveal inherent biases in current inference algorithms, where algorithms output predictions with similar topological properties while not being sensitive to the topological differences of the ground truth networks to be inferred. The results suggest that incorporating topological priors and constraints can help mitigate these biases and improve both the robustness and accuracy of inferred GRNs.

I have further provided a structured overview of available prior types intended to guide researchers both in selecting suitable GRN inference algorithms and in acquiring complementary data that can materially enhance the inference process. The overview clarifies how commonly

used data modalities contribute to identifying regulatory elements, linking those elements to their target genes, and to their regulating transcription factors. Popular examples include the use of chromatin accessibility and histone marks to nominate candidate regulatory elements; Hi-C and eQTL to link regulatory elements to target genes; and transcription factor motifs and ChIP-seq data to link regulatory regions to their regulating transcription factors. Prior knowledge from perturbation experiments can be used to extract causal relationships between transcription factors and target genes directly. Moreover, the overview highlights for method developers how each of the priors is commonly used in the inference process.

This work organizes algorithms that exploit prior knowledge into methodological families, which are correlation-based approaches, regression and regularized regression, probabilistic graphical models, and neural network–based methods. Beyond methodology, I distinguish two broad integration strategies of directly injecting additional data into the inference process versus first extracting a graph-based representation of the prior knowledge that is then refined by a core inference algorithm based on the RNA-sequencing data.

Among algorithms that begin with a graph representation of the prior, I identify subtypes based on weighted versus binary edges and homogeneous networks consisting of only genes as nodes versus heterogeneous graphs that include various node types.

This work is, to my knowledge, the first to explicitly differentiate inference strategies that add edges to a prior graph from those that remove edges, an operational distinction with clear implications for the benchmarking of these methods. For methods that rely on unifying graph representations, I propose a framework for fair benchmarking that disentangles the contribution of the prior from the contribution of the algorithm. The graph prior interface allows for the modularization of the benchmarking and enables head-to-head comparisons across algorithms under fixed priors and across priors under fixed algorithms. This addresses the previous limitation where only algorithm–prior bundles could be benchmarked as a single, inseparable entity. Together, these contributions provide actionable guidance for practitioners and a principled path for developers to design the next generation of GRN inference methods.

Moving forward, I want to highlight four additional important challenges that need to be addressed.

The first concerns the limitations of current benchmarking approaches, whose results quickly become outdated as methods, datasets, and evaluation standards evolve. Initiatives such as the OpenProblems platform [149] demonstrate a dynamic alternative, offering continuously updated and transparent benchmarking with versioned datasets and reproducible pipelines. I have contributed to bringing the GRN inference task to the OpenProblems platform [7]. Its design prioritizes extensibility by the addition of new inference algorithms, but also the incorporation of diverse datasets, tasks, and metrics. This is supported by containerized submissions, and rigourous documentation. In terms of metrics, the addition of standard area-under-the-curve metrics (e.g., AUROC and AUPRC) and the inclusion of topological metrics to the the platform is critical for a more complete assessment of GRN quality. Such metrics may be similar to the

ones used in the topological evaluation presented in this thesis. Equally important is curating a representative and regularly refreshed collection of benchmark datasets, especially with more large scale Perturb-seq datasets becoming available. Together, these elements will yield benchmarking results that remain current, reproducible, and decision-relevant for both practitioners and method developers.

This is also important in light of a second challenge, which is the choice of an evaluation metric that objectively best captures inference performance. Different evaluation metrics emphasize different aspects of the graph inference performance whose relevance depends on the goal of the downstream analysis, and there is no single best metric to use in general. For close to perfect inference performance, both binary, as well as topological metrics are simultaneously optimized. However, since the overall performance is currently still moderate, topological and binary edge based performance may very well deviate, as shown by the low correlation observed between these metrics in the topological benchmarking presented in this theses. Depending on the downstream application, e.g., local or global network topology may be critical and is not reflected in the commonly used statistical metrics. Even subtypes of the same metric can lead to different evaluation results. For example, the statistical evaluation as a binary classification task with the area under the curve metrics is often sensitive to the chosen sampling strategy for negative examples (non-edges) used in the evaluation. Thus, it is advantageous to combine multiple evaluation metrics to obtain a more comprehensive picture of the inference performance or to choose a metric to evaluate the performance for a specific downstream task.

A third challenge that remains is the limiting representation of the GRN as a simple abstract graph $G = (V, E)$ (where V denotes the set of genes and E represents the regulatory interactions between a single regulator and its target), even if the edges are annotated with edge weights. This representation omits the possible co-regulation which requires multiple regulators acting as a complex since the regulating edges are usually assumed to be additive. It further misses functional annotations on each edge that are necessary to describe dynamic parent–target relationships, which are crucial for simulating gene expression behavior. Consequently, GRNs should be regarded as intermediate representations. When the goal of the downstream analysis is to predict gene expression values upon perturbations, a promising alternative direction is to train an end-to-end model that directly maps inputs to expression outcomes, thereby bypassing the explicit graph prediction step [150, 151].

The fourth remaining challenge to highlight is the scalability of GRN inference to the full genome. In summary, while GRN inference is approaching a performance threshold for effective use in mechanistic modeling, its current application—particularly on a genome-wide scale—is limited. Many studies target small networks with only a handful of genes [109]. However, there are several promising directions investigated currently to enable large-scale inference of causal GRNs.

One example is the advance in causal inference research that has yielded promising methods with strong theoretical bases, but these approaches still face scalability challenges for whole-transcriptome GRN inference, an issue that ongoing research continues to address [59, 152]. Another example is technical advances that are promising for the future of GRN inference. For instance, the simultaneous measurement of multiple omics modalities (e.g., transcriptomics and epigenomics) in the same cells can provide enhanced resolution for network inference [153–155]. Similarly, laboratory automation by AI and robotics [156] and large-scale perturbation assays [157] offer the possibility of obtaining higher-quality and larger interventional datasets, which are optimal for disentangling causal effects while reducing batch effects.

My work aims to enhance the understanding of GRN inference performance as a basis for the development of improved algorithms in the future. Combined with better and larger datasets in the near future, the holy grail remains to construct large-scale causal GRN models that can be leveraged to accelerate drug development and deepen our understanding of complex cellular processes. An interesting exemplary field of application is cellular reprogramming. Cellular reprogramming can be either achieved by overexpression of transcription factors (direct reprogramming) [158] or by somatic cell nuclear transfer (SCNT) [159]. For nuclear transfer, a donor cell is reprogrammed to give rise to a whole embryo by injecting the donor nucleus into an enucleated egg. In both cases, understanding the mechanistic GRN helps optimize the perturbations of the original cell type-specific network that lead to the most efficient reprogramming of the cells. Optimizing these reprogramming techniques holds potential for applications in regenerative medicine [160–162]. I started to complement ongoing work on SCNT reprogramming in *Xenopus laevis* models [3, 5, 6] with GRN analysis, trying to find differences in the causal GRNs of cloned embryos compared to in vitro fertilized embryos. This promising approach aims to explain the reprogramming inefficiencies based on the network differences and proposes to target them to increase reprogramming efficiency.

Bibliography

[1] M. Stock et al. "Topological benchmarking of algorithms to infer gene regulatory networks from single-cell RNA-seq data". *Bioinformatics* 40.5 (Apr. 2024), btae267. DOI: 10.1093/bioinformatics/btae267.

[2] M. Stock et al. "Leveraging prior knowledge to infer gene regulatory networks from single-cell RNA-sequencing data". *Molecular Systems Biology* 21.3 (2025), pp. 214–230. DOI: 10.1038/s44320-025-00088-3.

[3] M. S. Oak et al. "H3K4 methylation-promoted transcriptional memory ensures faithful zygotic genome activation and embryonic development". *bioRxiv* (2025). DOI: 10.1101/2025.01.20.633863.

[4] G. Lubatti et al. "CIARA: a cluster-independent algorithm for identifying markers of rare cell types from single-cell sequencing data". *Development* 150.11 (June 2023). DOI: 10.1242/dev.201264.

[5] T. Zikmund et al. "Differentiation success of reprogrammed cells is heterogeneous in vivo and modulated by somatic cell identity memory". *Stem Cell Reports* (Mar. 2025). DOI: 10.1016/j.stemcr.2025.102447.

[6] A. Janeva et al. "'Digital Reprogramming' Decodes Epigenetic Barriers of Cell Fate Changes". *bioRxiv* (2025). DOI: 10.1101/2025.01.22.634227.

[7] J. Nourisa et al. "geneRNIB: a living benchmark for gene regulatory network inference". *bioRxiv* (2025). DOI: 10.1101/2025.02.25.640181.

[8] F. Gao et al. "Decoding the IGF1 signaling gene regulatory network behind alveologenesis from a mouse model of bronchopulmonary dysplasia". *eLife* 11 (Oct. 2022), e77522. DOI: 10.7554/eLife.77522.

[9] P. Badia-i-Mompel et al. "Gene regulatory network inference in the era of single-cell multi-omics". *Nature Reviews Genetics* 24.11 (Nov. 2023), pp. 739–754. DOI: 10.1038/s41576-023-00618-5.

[10] A. Regev et al. "Science Forum: The Human Cell Atlas". *eLife* 6 (Dec. 2017), e27041. DOI: 10.7554/eLife.27041.

[11] R. C. Jones et al. "The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans". *Science* 376.6594 (2022), eabl4896. DOI: 10.1126/science.abl4896.

[12] Y. Liu, A. Beyer, and R. Aebersold. "On the Dependency of Cellular Protein Levels on mRNA Abundance". *Cell* 165.3 (Apr. 2016), pp. 535–550. DOI: 10.1016/j.cell.2016.03.014.

[13] M. Furlan, S. de Pretis, and M. Pelizzola. "Dynamics of transcriptional and post-transcriptional regulation". *Briefings in Bioinformatics* 22.4 (Dec. 2020), bbaa389. DOI: 10.1093/bib/bbaa389.

[14] A. H. Corbett. “Post-transcriptional regulation of gene expression and human disease”. *Current Opinion in Cell Biology* 52 (2018), pp. 96–104. DOI: 10.1016/j.ceb.2018.02.011.

[15] P. S. Bruno et al. “Post-Translational Modifications of Proteins Orchestrate All Hallmarks of Cancer”. *Life* 15.1 (2025). DOI: 10.3390/life15010126.

[16] S. R. Archuleta, J. A. Goodrich, and J. F. Kugel. “Mechanisms and Functions of the RNA Polymerase II General Transcription Machinery during the Transcription Cycle”. *Biomolecules* 14.2 (2024). DOI: 10.3390/biom14020176.

[17] J. L. Plank and A. Dean. “Enhancer Function: Mechanistic and Genome-Wide Insights Come Together”. *Molecular Cell* 55.1 (July 2014), pp. 5–14. DOI: 10.1016/j.molcel.2014.06.015.

[18] J. Kim et al. “The co-regulation mechanism of transcription factors in the human gene regulatory network”. *Nucleic Acids Research* 40.18 (July 2012), pp. 8849–8861. DOI: 10.1093/nar/gks664.

[19] S. Balaji et al. “Comprehensive Analysis of Combinatorial Regulation using the Transcriptional Regulatory Network of Yeast”. *Journal of Molecular Biology* 360.1 (2006), pp. 213–227. DOI: 10.1016/j.jmb.2006.04.029.

[20] S. Rao, K. Ahmad, and S. Ramachandran. “Cooperative binding between distant transcription factors is a hallmark of active enhancers”. *Molecular Cell* 81.8 (Apr. 2021), 1651–1665.e4. DOI: 10.1016/j.molcel.2021.02.014.

[21] S. L. Klemm, Z. Shipony, and W. J. Greenleaf. “Chromatin accessibility and the regulatory epigenome”. *Nature Reviews Genetics* 20.4 (Apr. 2019), pp. 207–220. DOI: 10.1038/s41576-018-0089-8.

[22] D. Chasman and S. Roy. “Inference of cell type specific regulatory networks on mammalian lineages”. *Current Opinion in Systems Biology* 2 (2017), pp. 130–139. DOI: 10.1016/j.coisb.2017.04.001.

[23] J. A. Castro-Mondragon et al. “JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles”. *Nucleic Acids Research* 50.D1 (Nov. 2021), pp. D165–D173. DOI: 10.1093/nar/gkab1113.

[24] D. K. Goode et al. “Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation”. *Developmental Cell* 36.5 (Mar. 2016), pp. 572–587. DOI: 10.1016/j.devcel.2016.01.024.

[25] B. J. Strober et al. “Dynamic genetic regulation of gene expression during cellular differentiation”. *Science* 364.6447 (2019), pp. 1287–1290. DOI: 10.1126/science.aaw0040.

[26] A. Gupta et al. “Inferring gene regulation from stochastic transcriptional variation across single cells at steady state”. *Proceedings of the National Academy of Sciences* 119.34 (2022), e2207392119. DOI: 10.1073/pnas.2207392119.

[27] S. Picelli et al. “Full-length RNA-seq from single cells using Smart-seq2”. *Nature Protocols* 9.1 (Jan. 2014), pp. 171–181. DOI: 10.1038/nprot.2014.006.

[28] G. X. Y. Zheng et al. “Massively parallel digital transcriptional profiling of single cells”. *Nature Communications* 8.1 (Jan. 2017), p. 14049. DOI: 10.1038/ncomms14049.

[29] B. Hwang, J. H. Lee, and D. Bang. “Single-cell RNA sequencing technologies and bioinformaticspipelines”. *Experimental & Molecular Medicine* 50.8 (Aug. 2018), pp. 1–14. DOI: 10.1038/s12276-018-0071-8.

[30] M. D. Luecken and F. J. Theis. “Current best practices in single-cell RNA-seq analysis: a tutorial”. *Molecular Systems Biology* 15.6 (2019), e8746. DOI: 10.15252/msb.20188746.

[31] Y. You et al. “Benchmarking UMI-based single-cell RNA-seq preprocessing workflows”. *Genome Biology* 22.1 (Dec. 2021), p. 339. DOI: 10.1186/s13059-021-02552-3.

[32] R. Jiang et al. “Statistics or biology: the zero-inflation controversy about scRNA-seq data”. *Genome Biology* 23.1 (Jan. 2022), p. 31. DOI: 10.1186/s13059-022-02601-5.

[33] C. S. McGinnis, L. M. Murrow, and Z. J. Gartner. “DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors”. *Cell Systems* 8.4 (Apr. 2019), 329–337.e4. DOI: 10.1016/j.cels.2019.03.003.

[34] C. Ahlmann-Eltze and W. Huber. “Comparison of transformations for single-cell RNA-seq data”. *Nature Methods* 20.5 (May 2023), pp. 665–672. DOI: 10.1038/s41592-023-01814-1.

[35] I. N. Grabski, K. Street, and R. A. Irizarry. “Significance analysis for clustering with single-cell RNA-sequencing data”. *Nature Methods* 20.8 (Aug. 2023), pp. 1196–1202. DOI: 10.1038/s41592-023-01933-9.

[36] M. D. Luecken et al. “Benchmarking atlas-level data integration in single-cell genomics”. *Nature Methods* 19.1 (Jan. 2022), pp. 41–50. DOI: 10.1038/s41592-021-01336-8.

[37] X. Guo et al. “Recent advances in differential expression analysis for single-cell RNA-seq and spatially resolved transcriptomic studies”. *Briefings in Functional Genomics* 23.2 (Apr. 2023), pp. 95–109. DOI: 10.1093/bfgp/elad011.

[38] W. Saelens et al. “A comparison of single-cell trajectory inference methods”. *Nature Biotechnology* 37.5 (May 2019), pp. 547–554. DOI: 10.1038/s41587-019-0071-9.

[39] G. La Manno et al. “RNA velocity of single cells”. *Nature* 560.7719 (Aug. 2018), pp. 494–498. DOI: 10.1038/s41586-018-0414-6.

[40] Y. Zhang, X. Chang, and X. Liu. “Inference of gene regulatory networks using pseudo-time series data”. *Bioinformatics* 37.16 (Feb. 2021), pp. 2423–2431. DOI: 10.1093/bioinformatics/btab099.

[41] L. Van der Maaten and G. Hinton. “Visualizing data using t-SNE.” *Journal of machine learning research* 9.11 (2008).

[42] L. McInnes, J. Healy, and J. Melville. “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”. *arXiv* (2020). DOI: 10.48550/arXiv.1802.03426.

[43] M. Efremova et al. “CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes”. *Nature Protocols* 15.4 (Apr. 2020), pp. 1484–1506. DOI: 10.1038/s41596-020-0292-x.

[44] S. Jin et al. “Inference and analysis of cell-cell communication using CellChat”. *Nature Communications* 12.1 (Feb. 2021), p. 1088. DOI: 10.1038/s41467-021-21246-9.

[45] C. G. Williams et al. “An introduction to spatial transcriptomics for biomedical research”. *Genome Medicine* 14.1 (June 2022), p. 68. DOI: 10.1186/s13073-022-01075-1.

[46] K. Akers and T. Murali. “Gene regulatory network inference in single-cell biology”. *Current Opinion in Systems Biology* 26 (2021), pp. 87–97. DOI: 10.1016/j.coisb.2021.04.007.

[47] C. Li et al. “Benchmarking AI Models for In Silico Gene Perturbation of Cells”. *bioRxiv* (2025). DOI: 10.1101/2024.12.20.629581.

[48] S. Feng et al. “Hypergraph models of biological networks to identify genes critical to pathogenic viral response”. *BMC Bioinformatics* 22.1 (May 2021), p. 287. DOI: 10.1186/s12859-021-04197-2.

[49] B. He and K. Tan. “Understanding transcriptional regulatory networks using computational models”. *Current Opinion in Genetics & Development* 37 (2016), pp. 101–108. DOI: 10.1016/j.gde.2016.02.002.

[50] I. R. Wolf, R. P. Simões, and G. T. Valente. “Three topological features of regulatory networks control life-essential and specialized subsystems”. *Scientific Reports* 11.1 (Dec. 2021), p. 24209. DOI: 10.1038/s41598-021-03625-w.

[51] B. Clauss and M. Lu. “A quantitative evaluation of topological motifs and their coupling in gene circuit state distributions”. *iScience* 26.2 (2023), p. 106029. DOI: 10.1016/j.isci.2023.106029.

[52] L. T. MacNeil and A. J. M. Walhout. “Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression.” *Genome research* 21 5 (2011), pp. 645–57. DOI: 10.1101/gr.097378.109.

[53] M. Bafna, H. Li, and X. Zhang. “CLARIFY: cell–cell interaction and gene regulatory network refinement from spatially resolved transcriptomics”. *Bioinformatics* 39.Supplement 1 (June 2023), pp. i484–i493. DOI: 10.1093/bioinformatics/btad269.

[54] J. Cha and I. Lee. “Single-cell network biology for resolving cellular heterogeneity in human diseases”. *Experimental & Molecular Medicine* 52.11 (Nov. 2020), pp. 1798–1808. DOI: 10.1038/s12276-020-00528-0.

[55] W. J. Longabaugh, E. H. Davidson, and H. Bolouri. “Computational representation of developmental genetic regulatory networks”. *Developmental Biology* 283.1 (2005), pp. 1–16. DOI: 10.1016/j.ydbio.2005.04.023.

[56] S. Peidli et al. “scPerturb: harmonized single-cell perturbation data”. *Nature Methods* 21.3 (Mar. 2024), pp. 531–540. DOI: 10.1038/s41592-023-02144-y.

[57] Q. Hu et al. “Gene regulatory network inference during cell fate decisions by perturbation strategies”. *npj Systems Biology and Applications* 11.1 (Mar. 2025), p. 23. DOI: 10.1038/s41540-025-00504-2.

[58] S. L. Salzberg. “Open questions: How many genes do we have?” *BMC Biology* 16.1 (Aug. 2018), p. 94. DOI: 10.1186/s12915-018-0564-x.

[59] M. Chevalley et al. “CausalBench: A Large-scale Benchmark for Network Inference from Single-cell Perturbation Data”. *arXiv* (2023). DOI: 10.48550/arXiv.2210.17283.

[60] J. Han et al. “Mechanistic gene networks inferred from single-cell data with an outlier-insensitive method”. *Mathematical Biosciences* 342 (2021), p. 108722. DOI: 10.1016/j.mbs.2021.108722.

[61] A. N. Forbes et al. “Discovery of therapeutic targets in cancer using chromatin accessibility and transcriptomic data”. *Cell Systems* 15.9 (2024), 824–837.e6. DOI: 10.1016/j.cels.2024.08.004.

[62] M. A. Nakazawa et al. “Novel cancer subtyping method based on patient-specific gene regulatory network”. *Scientific Reports* 11.1 (Dec. 2021), p. 23653. DOI: 10.1038/s41598-021-02394-w.

[63] O. Cassan et al. “Optimizing data integration improves gene regulatory network inference in Arabidopsis thaliana”. *Bioinformatics* 40.7 (June 2024), btae415. DOI: 10.1093/bioinformatics/btae415.

[64] D. Kim et al. “Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data”. *npj Systems Biology and Applications* 9.1 (Oct. 2023), p. 51. DOI: 10.1038/s41540-023-00312-6.

[65] T. Zhu et al. “Integrated enhancer regulatory network by enhancer–promoter looping in gastric cancer”. *NAR Cancer* 6.2 (May 2024), zcae020. DOI: 10.1093/narcan/zcae020.

[66] X. Wang et al. “Exploring the Role of Enhancer-Mediated Transcriptional Regulation in Precision Biology”. *International Journal of Molecular Sciences* 24.13 (2023). DOI: 10.3390/ijms241310843.

[67] A. Kamal et al. “GRaNIE and GRaNPA: inference and evaluation of enhancer‐mediated gene regulatory networks”. *Molecular Systems Biology* 19.6 (2023), e11627. DOI: 10.15252/msb.202311627.

[68] S. M. M. Ud-Dean and R. Gunawan. “Ensemble Inference and Inferability of Gene Regulatory Networks”. *PLoS One* 9.8 (2014), e103812. DOI: 10.1371/journal.pone.0103812.

[69] S. Hu et al. “Inferring circadian gene regulatory relationships from gene expression data with a hybrid framework”. *BMC Bioinformatics* 24 (2023), p. 362. DOI: 10.1186/s12859-023-05458-y.

[70] V. D and A. S. “PEPN-GRN: A Petri net-based approach for the inference of gene regulatory networks from noisy gene expression data”. *PLoS One* 16.5 (2021), pp. 1–27. DOI: 10.1371/journal.pone.0251666.

[71] R. Bonneau et al. “The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo”. *Genome Biology* 7 (2006). DOI: 10.1186/gb-2006-7-5-r36.

[72] V. Huynh-Thu et al. “Inferring Regulatory Networks from Expression Data Using Tree-Based Methods”. *PLoS One* 5.9 (2010), pp. 1–10. DOI: 10.1371/journal.pone.0012776.

[73] G. Chen and Z.-P. Liu. “Graph attention network for link prediction of gene regulations from single-cell RNA-sequencing data”. *Bioinformatics* 38.19 (Aug. 2022), pp. 4522–4529. DOI: 10.1093/bioinformatics/btac559.

[74] S. Hill et al. “Bayesian Inference of Signaling Network Topology in a Cancer Cell Line”. *Bioinformatics* 28.21 (2012), pp. 2804–2810. DOI: 10.1093/bioinformatics/bts514.

[75] A. Margolin et al. “ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context”. *BMC Bioinformatics* 7 Suppl 1 (2006), S7. DOI: 10.1186/1471-2105-7-S1-S7.

[76] X. Zhang et al. “Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information”. *Bioinformatics* 28.1 (2011), pp. 98–104. DOI: 10.1093/bioinformatics/btr626.

[77] T. Chan, M. Stumpf, and A. Babtie. “Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures”. *Cell Systems* 5.3 (2017), 251–267.e3. DOI: 10.1016/j.cels.2017.08.014.

[78] H. Matsumoto et al. “SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation”. *Bioinformatics* 33.15 (2017), pp. 2314–2321. DOI: 10.1093/bioinformatics/btx194.

[79] A. Ocone, A. Millar, and G. Sanguinetti. “Hybrid regulatory models: a statistically tractable approach to model regulatory network dynamics”. *Bioinformatics* 29.7 (2013), pp. 910–916. DOI: 10.1093/bioinformatics/btt069.

[80] A. C. Babtie, T. E. Chan, and M. P. Stumpf. “Learning regulatory models for cell development from single cell transcriptomic data”. *Current Opinion in Systems Biology* 5 (2017), pp. 72–81. DOI: 10.1016/j.coisb.2017.07.013.

[81] E. Segal et al. “Learning Module Networks”. *arXiv* (2012). DOI: 10.48550/arXiv.1212.2517.

[82] N. Novershtern, A. Regev, and N. Friedman. “Physical Module Networks: an integrative approach for reconstructing transcription regulation”. *Bioinformatics* 27.13 (2011), pp. i177–i185. DOI: 10.1093/bioinformatics/btr222.

[83] S. G. McCalla et al. “Identifying strengths and weaknesses of methods for computational network inference from single-cell RNA-seq data”. *G3: Genes, Genomes, Genetics* 13.3 (Jan. 2023), jkad004. DOI: 10.1093/g3journal/jkad004.

[84] Z.-P. Liu. “Quantifying Gene Regulatory Relationships with Association Measures: A Comparative Study”. *Frontiers in Genetics* 8 (2017). DOI: 10.3389/fgene.2017.00096.

[85] K. Pearson and F. Galton. “VII. Note on regression and inheritance in the case of two parents”. *Proceedings of the Royal Society of London* 58.347-352 (1895), pp. 240–242. DOI: 10.1098/rspl.1895.0041.

[86] M. Kutner et al. *Applied Linear Statistical Models*. Jan. 2005.

[87] D. Marbach et al. “Revealing strengths and weaknesses of methods for gene network inference”. *Proceedings of the National Academy of Sciences* 107.14 (2010), pp. 6286–6291. DOI: 10.1073/pnas.0913357107.

[88] P. Langfelder and S. Horvath. “WGCNA: an R package for weighted correlation network analysis”. *BMC Bioinformatics* 9.1 (Dec. 2008), p. 559. DOI: 10.1186/1471-2105-9-559.

[89] T. Moerman et al. “GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks”. *Bioinformatics* 35.12 (Nov. 2018), pp. 2159–2161. DOI: 10.1093/bioinformatics/bty916.

[90] C. Ogris et al. “Versatile knowledge guided network inference method for prioritizing key regulatory factors in multi-omics data”. *Scientific Reports* 11.1 (Mar. 2021), p. 6806. DOI: 10.1038/s41598-021-85544-4.

[91] N. P. Gao, S. Minhaz Ud-Dean, and R. Gunawan. “Gene Regulatory Network Inference Using Time-Stamped Cross-Sectional Single Cell Expression Data”. *IFAC-PapersOnLine* 49.26 (2016), pp. 147–152. DOI: 10.1016/j.ifacol.2016.12.117.

[92] F. H. M. Salleh, S. Zainudin, and S. M. Arif. “Multiple Linear Regression for Reconstruction of Gene Regulatory Networks in Solving Cascade Error Problems”. *Advances in Bioinformatics* 2017.1 (2017), p. 4827171. DOI: 10.1155/2017/4827171.

[93] K. Kamimoto et al. “Dissecting cell identity via network inference and in silico gene perturbation”. *Nature* 614.7949 (Feb. 2023), pp. 742–751. DOI: 10.1038/s41586-022-05688-9.

[94] D. M. Alawad et al. “AGRN: accurate gene regulatory network inference using ensemble machine learning methods”. *Bioinformatics Advances* 3.1 (Apr. 2023), vbad032. DOI: 10.1093/bioadv/vbad032.

[95] L. Breiman. “Random Forests”. *Machine Learning* 45.1 (Oct. 2001), pp. 5–32. DOI: 10.1023/A:1010933404324.

[96] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. “Learning representations by back-propagating errors”. *Nature* 323.6088 (Oct. 1986), pp. 533–536. DOI: 10.1038/323533a0.

[97] Z. Fang, R. Zheng, and M. Li. "scMAE: a masked autoencoder for single-cell RNA-seq clustering". *Bioinformatics* 40.1 (Jan. 2024), btae020. DOI: 10.1093/bioinformatics/btae020.

[98] G. Mao et al. "Predicting gene regulatory links from single-cell RNA-seq data using graph neural networks". *Briefings in Bioinformatics* 24.6 (Nov. 2023), bbad414. DOI: 10.1093/bib/bbad414.

[99] D. P. Kingma and M. Welling. "Auto-Encoding Variational Bayes". *arXiv* (2022). DOI: 10.48550/arXiv.1312.6114.

[100] A. Grover and J. Leskovec. "node2vec: Scalable Feature Learning for Networks". *arXiv* (2016). DOI: 10.48550/arXiv.1607.00653.

[101] T. N. Kipf and M. Welling. "Semi-Supervised Classification with Graph Convolutional Networks". *arXiv* (2017). DOI: 10.48550/arXiv.1609.02907.

[102] P. Veličković et al. "Graph Attention Networks". *arXiv* (2018). DOI: 10.48550/arXiv.1710.10903.

[103] A. Vaswani et al. "Attention Is All You Need". *arXiv* (2017). DOI: 10.48550/arXiv.1706.03762.

[104] T. N. Kipf and M. Welling. "Variational Graph Auto-Encoders". *arXiv* (2016). DOI: 10.48550/arXiv.1611.07308.

[105] J. Pearl. "Chapter 3 - MARKOV AND BAYESIAN NETWORKS: Two Graphical Representations of Probabilistic Knowledge". *Probabilistic Reasoning in Intelligent Systems*. San Francisco (CA): Morgan Kaufmann, 1988, pp. 77–141. DOI: 10.1016/B978-0-08-051489-5.50009-6.

[106] N. Friedman et al. "Using Bayesian Networks to Analyze Expression Data". *Journal of Computational Biology* 7.3-4 (2000), pp. 601–620. DOI: 10.1089/106652700750050961.

[107] M. Banf and S. Y. Rhee. "Enhancing gene regulatory network inference through data integration with markov random fields". *Scientific Reports* 7.1 (Feb. 2017), p. 41174. DOI: 10.1038/srep41174.

[108] V. Ravikumar et al. "Efficient Inference of Spatially-varying Gaussian Markov Random Fields with Applications in Gene Regulatory Networks". *arXiv* (2022). DOI: 10.48550/arXiv.2206.10174.

[109] L. Atanackovic et al. "DynGFN: Towards Bayesian Inference of Gene Regulatory Networks with GFlowNets". *arXiv* (2023). DOI: 10.48550/arXiv.2302.04178.

[110] A. Tejada-Lapuerta et al. "Causal machine learning for single-cell genomics". *arXiv* (2023). DOI: 10.48550/arXiv.2310.14935.

[111] A. Pratapa et al. "Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data". *Nature Methods* 17.2 (Feb. 2020), pp. 147–154. DOI: 10.1038/s41592-019-0690-6.

[112] F. Petralia et al. “Integrative random forest for gene regulatory network inference”. *Bioinformatics* 31.12 (June 2015), pp. i197–i205. DOI: 10.1093/bioinformatics/btv268.

[113] ENCODE Consortium. “An integrated encyclopedia of DNA elements in the human genome”. *Nature* 489.7414 (Sept. 2012), pp. 57–74. DOI: 10.1038/nature11247.

[114] R. Oughtred et al. “The BioGRID interaction database: 2019 update”. *Nucleic Acids Research* 47.D1 (Nov. 2018), pp. D529–D541. DOI: 10.1093/nar/gky1079.

[115] P. T. Monteiro et al. “YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts”. *Nucleic Acids Research* 48.D1 (Oct. 2019), pp. D642–D649. DOI: 10.1093/nar/gkz859.

[116] A. Nair. “Incorporating and Generating Prior Knowledge to Improve Gene Regulatory Network Inference”. PhD thesis. Monash University, 2017. DOI: 10.4225/03/59bf0bad8745a.

[117] W. Liu et al. “NSCGRN: a network structure control method for gene regulatory network inference”. *Briefings in Bioinformatics* 23.5 (May 2022), bbac156. DOI: 10.1093/bib/bbac156.

[118] D. Yao et al. “Scalable genetic screening for regulatory circuits using compressed Perturb-seq”. *Nature Biotechnology* 42.8 (Aug. 2024), pp. 1282–1295. DOI: 10.1038/s41587-023-01964-9.

[119] J. M. Replogle et al. “Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing”. *Nature Biotechnology* 38.8 (Aug. 2020), pp. 954–961. DOI: 10.1038/s41587-020-0470-y.

[120] A. Dixit et al. “Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens”. *Cell* 167.7 (Dec. 2016), 1853–1866.e17. DOI: 10.1016/j.cell.2016.11.038.

[121] Z.-J. Cao and G. Gao. “Multi-omics single-cell data integration and regulatory inference with graph-linked embedding”. *Nature Biotechnology* 40.10 (Oct. 2022), pp. 1458–1466. DOI: 10.1038/s41587-022-01284-4.

[122] Y. Wang et al. “NetREX-CF integrates incomplete transcription factor data with gene expression to reconstruct gene regulatory networks”. *Communications Biology* 5.1 (Nov. 2022), p. 1282. DOI: 10.1038/s42003-022-04226-7.

[123] J. D. Buenrostro et al. “Single-cell chromatin accessibility reveals principles of regulatory variation”. *Nature* 523.7561 (July 2015), pp. 486–490. DOI: 10.1038/nature14590.

[124] M. Y. Y. Lee, K. H. Kaestner, and M. Li. “Benchmarking algorithms for joint integration of unpaired and paired single-cell RNA-seq and ATAC-seq data”. *Genome Biology* 24.1 (Oct. 2023), p. 244. DOI: 10.1186/s13059-023-03073-x.

[125] T. Stuart et al. “Comprehensive Integration of Single-Cell Data”. *Cell* 177.7 (June 2019), 1888–1902.e21. DOI: 10.1016/j.cell.2019.05.031.

[126] J. S. Fleck et al. “Inferring and perturbing cell fate regulomes in human brain organoids”. *Nature* 621.7978 (Sept. 2023), pp. 365–372. DOI: 10.1038/s41586-022-05279-8.

[127] C. Bravo González-Blas et al. “SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks”. *Nature Methods* 20.9 (Sept. 2023), pp. 1355–1367. DOI: 10.1038/s41592-023-01938-4.

[128] V. K. Kartha et al. “Functional inference of gene regulation using single-cell multi-omics”. *Cell genomics* 2.9 (2022). DOI: 10.1016/j.xgen.2022.100166.

[129] Z. Li et al. “scMEGA: single-cell multi-omic enhancer-based gene regulatory network inference”. *Bioinformatics Advances* 3.1 (2023), vbad003. DOI: 10.1093/bioadv/vbad003.

[130] M. Saint-Antoine and A. Singh. “Limits on Inferring Gene Regulatory Networks Subjected to Different Noise Mechanisms”. *bioRxiv* (2023). DOI: 10.1101/2023.01.23.525259.

[131] G. Crichton et al. “Neural networks for link prediction in realistic biomedical graphs: a multi-dimensional evaluation of graph embedding-based approaches”. *BMC Bioinformatics* 19.1 (May 2018), p. 176. DOI: 10.1186/s12859-018-2163-9.

[132] M. B. A. McDermott et al. “A Closer Look at AUROC and AUPRC under Class Imbalance”. *arXiv* (2025). DOI: 10.48550/arXiv.2401.06091.

[133] Y. Guo and A. Amir. “Exploring the effect of network topology, mRNA and protein dynamics on gene regulatory network stability”. *Nature communications* 12.1 (2021), p. 130. DOI: 10.1038/s41467-020-20472-x.

[134] L. C. Freeman. “A set of measures of centrality based on betweenness”. *Sociometry* (1977), pp. 35–41. DOI: 10.2307/3033543.

[135] T. W. Valente and R. K. Foreman. “Integration and radiality: Measuring the extent of an individual’s connectedness and reachability in a network”. *Social networks* 20.1 (1998), pp. 89–105. DOI: 10.1016/S0378-8733(97)00007-5.

[136] L. C. Freeman et al. “Centrality in social networks: Conceptual clarification”. *Social network: critical concepts in sociology. Londres: Routledge* 1.3 (2002), pp. 238–263. DOI: 10.1016/0378-8733(78)90021-7.

[137] M. E. Newman. “Mixing patterns in networks”. *Physical review E* 67.2 (2003), p. 026126. DOI: 10.1103/PhysRevE.67.026126.

[138] D. J. Watts and S. H. Strogatz. “Collective dynamics of ‘small-world’networks”. *Nature* 393.6684 (1998), pp. 440–442. DOI: 10.1038/30918.

[139] S. Chen and J. C. Mar. “Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data”. *BMC Bioinformatics* 19.1 (June 2018), p. 232. DOI: 10.1186/s12859-018-2217-z.

[140] L. Cerulo, C. Elkan, and M. Ceccarelli. “Learning gene regulatory networks from only positive and unlabeled data”. *BMC Bioinformatics* 11.1 (May 2010), p. 228. DOI: 10.1186/1471-2105-11-228.

[141] C. Elkan and K. Noto. "Learning classifiers from only positive and unlabeled data". *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '08. Las Vegas, Nevada, USA: Association for Computing Machinery, 2008, pp. 213–220. DOI: 10.1145/1401890.1401920.

[142] L. Nunes, T. Faleiros, and R. Rossi. "Positive Unlabeled Learning: Adapting NMF for text classification". *Anais do XX Encontro Nacional de Inteligência Artificial e Computacional*. Belo Horizonte/MG: SBC, 2023, pp. 697–711. DOI: 10.5753/eniac.2023.234337.

[143] J. Bekker and J. Davis. "Learning from positive and unlabeled data: a survey". *Machine Learning* 109.4 (Apr. 2020), pp. 719–760. DOI: 10.1007/s10994-020-05877-5.

[144] C. Skok Gibbs et al. "High-performance single-cell gene regulatory network inference at scale: the Inferelator 3.0". *Bioinformatics* 38.9 (Feb. 2022), pp. 2519–2528. DOI: 10.1093/bioinformatics/btac117.

[145] S. Zhang et al. "Inference of cell type-specific gene regulatory networks on cell lineages from single cell omic datasets". *Nature Communications* 14.1 (May 2023), p. 3064. DOI: 10.1038/s41467-023-38637-9.

[146] J. Tang et al. "Single-cell gene regulatory network analysis for mixed cell populations". *Quantitative Biology* 12.4 (2024), pp. 375–388. DOI: 10.1002/qub2.64.

[147] H. L. Crowell et al. "The shaky foundations of simulating single-cell RNA sequencing data". *Genome Biology* 24.1 (Mar. 2023), p. 62. DOI: 10.1186/s13059-023-02904-1.

[148] N. Papili Gao et al. "SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles". *Bioinformatics* 34.2 (2018), pp. 258–266. DOI: 10.1093/bioinformatics/btx575.

[149] M. D. Luecken et al. "Defining and benchmarking open problems in single-cell analysis". *Research Square* (2024), rs–3. DOI: 10.21203/rs.3.rs-4181617/v1.

[150] G. I. Gavriilidis et al. "A mini-review on perturbation modelling across single-cell omic modalities". *Computational and Structural Biotechnology Journal* 23 (2024), pp. 1886–1896. DOI: 10.1016/j.csbj.2024.04.058.

[151] W. Wang et al. "RegVelo: gene-regulatory-informed dynamics of single cells". *bioRxiv* (2024). DOI: 10.1101/2024.12.11.627935.

[152] T. Nguyen et al. "Causal Inference in Gene Regulatory Networks with GFlowNet: Towards Scalability in Large Systems". *arXiv* (2023). DOI: 10.48550/arXiv.2310.03579.

[153] H. Xiong et al. "Single-cell joint profiling of multiple epigenetic proteins and gene transcription". *Science Advances* 10.1 (2024), eadi3664. DOI: 10.1126/sciadv.adi3664.

[154] C. Ding et al. "Simultaneous profiling of chromatin-associated RNA at targeted DNA loci and RNA-RNA Interactions through TaDRIM-seq". *Nature Communications* 16.1 (Feb. 2025), p. 1500. DOI: 10.1038/s41467-024-53534-5.

[155] Z. Liu et al. “Linking genome structures to functions by simultaneous single-cell Hi-C and RNA-seq”. *Science* 380.6649 (2023), pp. 1070–1076. DOI: 10.1126/science.adg3797.

[156] A. I. Cooper et al. “Accelerating Discovery in Natural Science Laboratories with AI and Robotics: Perspectives and Challenges from the 2024 IEEE ICRA Workshop, Yokohama, Japan”. *arXiv* (2025). DOI: 10.48550/arXiv.2501.06847.

[157] X. Zheng et al. “Massively parallel in vivo Perturb-seq reveals cell-type-specific transcriptional networks in cortical development”. *Cell* 187.13 (June 2024), 3236–3248.e21. DOI: 10.1016/j.cell.2024.04.050.

[158] K. Takahashi and S. Yamanaka. “Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors”. *Cell* 126.4 (2006), pp. 663–676. DOI: 10.1016/j.cell.2006.07.024.

[159] J. B. Gurdon, T. R. Elsdale, and M. Fischberg. “Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei”. *Nature* 182.4627 (1958), pp. 64–65. DOI: 10.1038/182064a0.

[160] H. Wang et al. “Direct cell reprogramming: approaches, mechanisms and progress”. *Nature Reviews Molecular Cell Biology* 22.6 (June 2021), pp. 410–424. DOI: 10.1038/s41580-021-00335-z.

[161] Y. Kim, J. Jeong, and D. Choi. “Small-molecule-mediated reprogramming: a silver lining for regenerative medicine”. *Experimental & Molecular Medicine* 52.2 (Feb. 2020), pp. 213–226. DOI: 10.1038/s12276-020-0383-3.

[162] J. Kuang, T. Huang, and D. Pei. “The Art of Reprogramming for Regenerative Medicine”. *Frontiers in Cell and Developmental Biology* Volume 10 - 2022 (2022). DOI: 10.3389/fcell.2022.927555.

A. Appendix

A.1. Topological benchmarking of algorithms to infer gene regulatory networks from single-cell RNA-seq data

Stock, M., Popp, N., Fiorentino, J. and Scialdone, A.

This is a published version of the article in Bioinformatics following peer review. The article is open-access thus the published version is inserted here.

Bioinformatics, 2024, **40(5)**, btae267
https://doi.org/10.1093/bioinformatics/btae267
Advance Access Publication Date: 16 April 2024
Original Paper

OXFORD

Gene expression

Topological benchmarking of algorithms to infer gene regulatory networks from single-cell RNA-seq data

Marco Stock [1,2,3,4,†], **Niclas Popp**[1,2,3,†], **Jonathan Fiorentino** [1,2,3,5,*], **Antonio Scialdone** [1,2,3,*]

[1]Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München—German Research Center for Environmental Health, Munich 81377, Germany
[2]Institute of Functional Epigenetics, Helmholtz Zentrum München—German Research Center for Environmental Health, Munich 85764, Germany
[3]Institute of Computational Biology, Helmholtz Zentrum München—German Research Center for Environmental Health, Munich 85764, Germany
[4]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Munich 85354, Germany
[5]Present address: Center for Life Nano- & Neuro-Science, Fondazione Istituto Italiano di Tecnologia, Viale Regina Elena 291, Rome 00161, Italy.
[†]These authors contributed equally to this work.
*Corresponding authors. Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München—German Research Center for Environmental Health, Munich 81377, Germany. E-mails: jonathan.fiorentino@iit.it (J.F.) and antonio.scialdone@helmholtz-munich.de (A.S.).

Associate Editor: Anthony Mathelier

Abstract

Motivation: In recent years, many algorithms for inferring gene regulatory networks from single-cell transcriptomic data have been published. Several studies have evaluated their accuracy in estimating the presence of an interaction between pairs of genes. However, these benchmarking analyses do not quantify the algorithms' ability to capture structural properties of networks, which are fundamental, e.g., for studying the robustness of a gene network to external perturbations. Here, we devise a three-step benchmarking pipeline called STREAMLINE that quantifies the ability of algorithms to capture topological properties of networks and identify hubs.

Results: To this aim, we use data simulated from different types of networks as well as experimental data from three different organisms. We apply our benchmarking pipeline to four inference algorithms and provide guidance on which algorithm should be used depending on the global network property of interest.

Availability and implementation: STREAMLINE is available at https://github.com/ScialdoneLab/STREAMLINE. The data generated in this study are available at https://doi.org/10.5281/zenodo.10710444.

1 Introduction

Single-cell transcriptomics techniques allow probing patterns of gene expression on an increasingly larger scale, with recent studies including millions of cells and thousands of genes (Svensson *et al.* 2020). Such rapid progress in expanding the scale of available data makes single-cell datasets more appealing for tasks like the inference of gene regulatory networks (GRNs), with the goal of achieving a mechanistic understanding of the systems at hand and going beyond purely descriptive characterizations (Akers and Murali 2021, Stumpf 2021, Saint-Antoine and Singh 2020). However, GRN inference from single-cell data entails many computational challenges, such as high levels of technical noise in the data (Brennecke *et al.* 2013), the extreme sparsity of the ground truth network to be inferred (Banf and Rhee 2017) and the increasing scale of gene expression data (Hillerton *et al.* 2022). For this reason, many algorithms for GRN inference from single-cell data have been published in the last few years. The increasingly large number of such algorithms demands benchmarking studies that can guide the user in the choice of the best-performing methods under various conditions (Chen and Mar 2018, Pratapa *et al.* 2020, Kang *et al.* 2021).

While the benchmarking studies that have been published offer some guidance for users, they are affected by important limitations. First, the quantification of the performance is obtained for a limited number and types of networks. Furthermore, the available benchmarking studies mostly focus on the ability of the GRN algorithms to predict local features of networks, like the interactions between pairs of genes, using, e.g., area under the curve metrics, or the presence of specific sub-graphs (network motifs). These metrics do not assess the algorithms' ability to infer the structural properties of the GRN, which can quantify important features like the robustness to perturbations (Guo and Amir 2021) and the presence of network hubs representing master regulators. Robustness is one of the main characteristics of GRNs (Noman *et al.* 2015), and for this reason, their topology is also studied to improve the robustness of general network structures in other fields, such as wireless sensor networks (Kamapantula *et al.* 2014, Roy *et al.* 2018).

Received: 16 May 2023; **Revised:** 28 February 2024; **Editorial Decision:** 5 April 2024; **Accepted:** 16 April 2024

Moreover, the inclusion of network topology in GRN inference methods has also been shown to improve their performance, e.g., using microarray and bulk RNA-seq data (Villaverde *et al.* 2014, Zhang *et al.* 2015). Recently, an algorithm based on a global network centrality measure and local network motifs has been introduced (Liu *et al.* 2022), showing an improvement in inference performance due to the reduction of network redundancy. This class of inference methods is still lacking for single-cell RNA-seq data.

The structural properties of networks can be quantified by topological measurements (Koutrouli *et al.* 2020), including, for instance, the network efficiency and the assortativity. So far, the performance of GRN inference algorithms on the estimation of topological properties has only been assessed with bulk RNA-seq data (Kiani *et al.* 2016, Escorcia-Rodríguez *et al.* 2023), and by employing a limited number of synthetic networks (Kiani *et al.* 2016), which makes it hard to reach robust conclusions for single-cell data.

In this work, we developed STREAMLINE, a three-step benchmarking framework to score the performance of GRN inference algorithms in estimating structural properties of networks from single-cell RNA-seq (scRNA-seq) datasets. The structural properties we considered quantify the information exchange efficiency, which is related to the network's robustness to perturbations, and the presence and identification of hubs. We used data simulated from hundreds of networks belonging to four classes with different structural properties (Watts and Strogatz 1998, Ouma *et al.* 2018), as well as from a set of curated (Cur) networks extracted from real GRNs (Pratapa *et al.* 2020). In addition to simulated data, we also used real datasets from yeast, mouse, and human (McCalla *et al.* 2023).

We applied STREAMLINE to four GRN inference algorithms chosen among the top-performing ones in predicting gene-gene interactions (Pratapa *et al.* 2020). Our benchmarking analysis provides guidance in the choice of the algorithm for the prediction of network robustness and the identification of hubs. Moreover, our results point to systematic biases in some algorithms, which could indicate ways of improving them.

To facilitate the use of our benchmarking framework, we made it compatible with an existing pipeline (BEELINE (Pratapa *et al.* 2020)), and we made all the code available in a GitHub repository (https://github.com/ScialdoneLab/STREAMLINE).

2 Methods

2.1 Ground truth networks

2.1.1 Synthetic networks

We use parameter-controlled networks from four different classes as well as the Cur GRNs that have been used in BEELINE (Pratapa *et al.* 2020). The output of the network samplers is a graph G with n nodes and m edges.

2.1.1.1 *Random networks*

Random networks were created with the Erdös–Renyi $G(n, p)$ model, which outputs a graph with n nodes where each pair is connected with probability p (Erdős and Rényi 1959). We set p so that the expected number of edges equals m.

2.1.1.2 *Scale-Free networks*

Networks with a degree distribution that follows a power law are classified as Scale-Free (Cho *et al.* 2009). Given the parameter α, the expected degree distribution follows:

$$P(d) \sim d^{-\alpha}. \quad (1)$$

For directed networks, the in-degree distribution and the out-degree distribution can feature different parameters α_{in} and α_{out}. We applied combinations of different in- and out-degrees. The exact values can be found in Supplementary Table S1.

2.1.1.3 *Semi-Scale-Free networks*

Following the analysis of the degree distributions in known GRNs (Ouma *et al.* 2018), we sampled Semi-Scale-Free networks which feature an out-degree distribution that follows a power law but a uniform in-degree distribution. Additionally, only 50% of the nodes have outgoing edges.

2.1.1.4 *Small-World networks*

We used the Watts–Strogatz model to sample networks that feature Small-World topology (Watts and Strogatz 1998). The algorithm starts with n nodes with degree k in a regular lattice and then rewires edges with probability p.

2.1.1.5 *Curated networks*

Curated networks are four known GRNs that were used in BEELINE to evaluate the statistical performance of the GRN inference algorithms (Pratapa *et al.* 2020). These networks are simple models for mammalian cortical area development (mCAD), ventral spinal cord development (VSC), hematopoietic stem cell differentiation (HSC), and gonadal sex determination (GSD).

2.1.1.6 *Network sampling*

We use the Julia package LightGraphs.jl https://github.com/JuliaGraphs/Graphs.jl/ to sample the networks explained above. The parameters were chosen such that a large variety of structurally different networks is covered.

2.1.1.7 *Simulation of single-cell RNA-sequencing data*

We simulate single-cell RNA-sequencing data for the synthetic networks using BoolODE (Pratapa *et al.* 2020), a recently developed method that first converts a Boolean model into a set of ordinary differential equations (ODEs), and then, after adding a noise term, performs stochastic simulations of genes' expression levels. The method was shown to outperform previously developed algorithms to simulate gene expression from ground truth GRNs like GeneNetWeaver (Schaffter *et al.* 2011). In our simulations, we used the same BoolODE parameters and settings that were extensively tested in Pratapa *et al.* (2020). Specifically, we converted each generated synthetic network into a text file containing a set of Boolean rules, which is given as input to the BoolODE Python script ("path data" parameter). For every parameter set of the synthetic networks, we generated data from 100 cells ("num-cells" parameter) with a simulation time of five steps ("max-steps" parameter) for multiple networks using BoolODE. We used the parameter "sample-cells" to sample

one cell per simulation. The number of networks and associated parameters can be found in Supplementary Table S1.

2.1.2 Experimental networks

For the benchmarking of GRN inference on experimental single-cell RNA-sequencing data we selected four datasets from human (Han *et al.* 2018), mouse (Shalek *et al.* 2014, Tran *et al.* 2019), and yeast (Gasch *et al.* 2017) and compared the output networks to different types of silver standard networks that were collected by McCalla *et al.* (2023). The properties of the networks and the number of corresponding silver standards can be found in Supplementary Table S2. The silver standard networks were obtained from public databases and the literature. They were derived from ChIP-chip, ChIP-seq, or gene perturbations followed by bulk sequencing, which yielded multiple networks for each organism (McCalla *et al.* 2023). The ESC silver standard network was obtained from manual curation of GRNs found in the literature, as explained in full detail in McCalla *et al.* (2023).

2.2 Inference algorithms

We selected the four top-performing algorithms from BEELINE (Pratapa *et al.* 2020) and examined the results using our three-step benchmarking pipeline. Below, we added a brief description of each algorithm.

GRNBoost2: GRNBoost2 (Moerman *et al.* 2019) infers a GRN independently for each gene, by identifying the most important regulators using a regression model. It is an alternative to GENIE3, which uses a similar inference scheme but does not scale to larger datasets due to its runtime. The output of GRNBoost2 is a directed network.

SINCERITIES: SINCERITIES (Papili Gao *et al.* 2018) is a causality-based method that computes temporal changes in the expression of each gene. The GRN is inferred by solving a specifically formulated ridge regression problem. SINCERITIES outputs a directed network.

PIDC: The PIDC inference scheme (Chan *et al.* 2017) is based on partial information decomposition, which is a multivariate information-theoretic measure for triplets of random variables. Since it is symmetric, the resulting network is undirected.

PPCOR: PPCOR (Kim 2015) calculates the partial and semi-partial correlation coefficients for every possible pair of genes. Edges are ranked according to these values and they are undirected. By using the correlation as a sign, it is possible to obtain activatory and inhibitory interactions. However, we did not use this information in our benchmarking framework.

2.3 Evaluation of inferred networks

2.3.1 Processing of GRNs

2.3.1.1 *Processing of ground truth networks*

In all the ground truth networks, self-loops and duplicate edges are removed. For the experimental datasets, genes in the silver standard networks were subset to the genes that appear in the related gene expression dataset.

In directed networks, the largest weakly connected subgraph of the ground truth was used in the analysis. To perform the analysis on undirected networks, the directed ground truth networks were converted to undirected graphs by ignoring the direction of the edges and then the largest connected subgraph was extracted.

2.3.1.2 *Processing of inferred networks*

For analysis on undirected networks, the inferred directed networks of GRNBoost2 and SINCERITIES are first converted to undirected networks by ignoring the information about the directionality of the edges.

In both the undirected and directed evaluations, duplicate edges and self-loops in the inferred network are removed. Afterwards, the top k edges with the highest absolute predicted weight are used to construct the graph for evaluation. The parameter k is chosen to be the same as the number of edges in each associated processed ground truth or silver standard network.

If the resulting graph is not weakly connected in the directed evaluation, the largest weakly connected subgraph is extracted. When evaluating undirected networks, the largest connected subgraph is extracted, if the resulting graph is not connected.

2.3.2 Binary edge detection

To statistically benchmark the edge prediction we followed BEELINE in evaluating the *EPr*, defined as the fraction of true positives among the top k edges, ranked according to the weight returned by the inference algorithm, where k is the number of interactions in the corresponding ground truth or silver standard network. The *EPr* is better suited to classification accuracy on large datasets where the reference networks do not represent the entire ground truth. For the synthetic networks, we additionally report the *AUPRC* and the *AUROC*, as also commonly done in previous benchmarking (Pratapa *et al.* 2020, Chen and Mar 2018).

2.3.3 Graph properties related to information exchange efficiency

For evaluation of the information exchange efficiency in the graphs, we chose three topological properties. The properties are only computed in the evaluation of undirected networks, thus we assume an undirected graph G with a set $\{N\}$ of n nodes and m edges for the following definitions.

2.3.3.1 *Average shortest path length*

The average shortest path length $\bar{l_{sp}}(G)$ measures by how many links two random nodes are connected on average:

$$\bar{l_{sp}}(G) = \sum_{v,w \in \{N\}} \frac{d(v,w)}{n \cdot (n-1)}, \tag{2}$$

where $d(v, w)$ denotes the distance between two nodes v and w.

2.3.3.2 *Global efficiency*

The global efficiency $E_{glob}(G)$ estimates how efficiently information is exchanged in the network on a global scale. This is related to the concept of the vulnerability of networks to the decrease in networks efficiency in case some of the components malfunction (Latora and Marchiori 2001, Boccaletti *et al.* 2006). It is given by:

$$E_{glob}(G) = \frac{1}{n \cdot (n-1)} \sum_{\substack{v \neq w \\ v,w \in \{N\}}} \frac{1}{d(v,w)}. \tag{3}$$

Since gene regulation can be interpreted as information exchange between nodes in GRN, E_{glob} is a meaningful quantity to estimate. A more detailed explanation of the relationship between global efficiency and network vulnerability can be found in Latora and Marchiori (2007).

2.3.3.3 *Local efficiency*

The local efficiency $E_{loc}(G)$ describes the resistance of the network to perturbation on a small scale (Latora and Marchiori 2001). Similarly to global efficiency, it is linked to the behavior of the network when some of its constituents fail. A more detailed derivation can be found in Latora and Marchiori (2007). The quantity is defined as:

$$E_{loc}(G) = \frac{1}{n} \sum_{v \in \{N\}} E_{glob}(G_v), \tag{4}$$

where G_v is the subgraph of G that only consists of the direct neighbors of v. In practice, perturbations are more likely to be caused by changes in neighboring genes, thus the local efficiency can provide valuable information.

2.3.3.4 *Evaluation score for graph properties*

Since we wanted to preserve the information on whether certain topological features are over- or underestimated, we employed the *MSE* as an evaluation metric. For a property P_x which is being analyzed on ground truth networks $G_1, G_2, \ldots, G_k$ and predicted networks $G_{1,\text{inferred}}$, $G_{2,\text{inferred}}, \ldots, G_{k,\text{inferred}}$, *MSE* is computed by:

$$MSE(P_x) = \frac{1}{n} \sum_{i=1}^{k} P_x(G_{i,\text{inferred}}) - P_x(G_i). \tag{5}$$

Therefore, a positive value of the MSE refers to an overestimation of the property compared to the ground truth and a negative MSE to an underestimation of the ground truth property.

2.3.4 Topological properties related to the hub analysis

For our hub analysis, we selected different graph and node properties that allow for a meaningful structural characterization of hubs in a network. The graph properties defined below are evaluated with the MSE, as introduced before for the information exchange quantities. The node properties are evaluated with the Jaccard coefficient of the detected hubs, as explained below.

2.3.4.1 *Graph properties related to hub topology*

2.3.4.1.1 Degree assortativity

The preference for a network's node to attach to others that have a similar degree is captured by the degree assortativity (Newman 2003). It is quantified by the assortativity coefficient $r_{\text{deg}}(G)$:

$$r_{\text{deg}}(G) = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}, \tag{6}$$

with $a_i = \sum_j e_{ij}$, $b_j = \sum_i e_{ij}$ and e_{ij} is the fraction of edges from a node with degree i to a node with degree j from all edges m of the graph. For undirected networks, i and j are total degrees of nodes, whereas for directed networks we report the Assortativity based on the in-degrees i and j of the nodes. Networks with a negative assortativity coefficient are called disassortative, and networks with a positive $r_{\text{deg}}(G)$ are called assortative. Disassortative networks have a higher tendency to possess hubs, which is an important feature of GRNs that we examine in Section 2.3.

2.3.4.1.2 Degree centralization

The goal of the degree centralization $H(G)$ is to provide an estimate of how centralized a graph is around the node v^* which has the highest degree in the graph (Freeman 1978). It is defined as:

$$H(G) = \frac{1}{H_{max}} \cdot \sum_{v \in \{N\}} (\deg(v^*) - \deg(v)), \tag{7}$$

with

$$H_{max,\text{undirected}} = (n-1)(n-2), \tag{8}$$

$$H_{max,\text{directed}} = (n-1)(n-1), \tag{9}$$

for undirected and directed networks, where $deg(v)$ refers to the total degree of a node v for undirected networks and to the in-degree for directed networks, where the in centralization is reported. A highly centralized network is focused around a small number of nodes, which could be identified as biologically important.

2.3.4.1.3 Clustering coefficient

The clustering coefficient measures the extent to which nodes in a graph tend to cluster together. It is quantified by the local clustering coefficient $CC_{\text{loc}}(v)$, which measures the fraction of triangles that exist over all possible triangles in the neighborhood of a node v:

$$CC_{\text{loc}}(v) = \frac{2 \cdot L_v}{k_v \cdot (k_v - 1)}, \tag{10}$$

and the global clustering coefficient $CC_{\text{glob}}(G)$ (Watts and Strogatz 1998):

$$CC_{\text{glob}}(G) = \frac{1}{n} \sum_{v \in \{N\}} CC_{\text{loc}}(v), \tag{11}$$

which is the average of the local clustering coefficient over the whole network. L_v represents the number of links between the k_v neighbors of node v. For directed graphs, k_v includes both parents of the node with edges going from the parent to the node of interest, and children of the node, with edges going from the node of interest to the child node. For our analysis, we focus on the global clustering coefficient, since it captures clustering on a global scale. A network with a larger global clustering coefficient is more interconnected, which can result in more complex gene regulations.

2.3.4.2 *Node properties related to hub identification*

2.3.4.2.1 Degree centrality

The degree centrality $C_D(v)$ evaluates the degree d of a node v in a network, scaled by the maximum possible degree:

$$C_D(v) = \frac{\deg(v)}{n-1}. \tag{12}$$

For undirected graphs, the total degree centrality is reported, where $deg(v)$ represents the total degree of node v. For directed graphs, the out centrality is reported, where $deg(v)$ refers to the out-degree of node v.

2.3.4.2.2 Betweenness centrality

The betweenness centrality $C_B(v)$ describes the extent to which nodes are on the shortest path between other nodes. We use a normalized version of the definition by Freeman (1977):

$$C_B(v) = \frac{1}{p_v} \sum_{u \neq v \neq w} \frac{\eta_v(u,w)}{\eta(u,w)}, \tag{13}$$

where $\eta(u,w)$ is the number of shortest paths from u to w and $\eta_v(u,w)$ is the number of shortest paths from u to w passing through v. It is normalized by dividing by the number of pairs of vertices not including v, which is different for undirected and directed graphs:

$$p_{v,\text{undirected}} = (n-1)(n-2)/2, \tag{14}$$

$$p_{v,\text{directed}} = (n-1)(n-2). \tag{15}$$

2.3.4.2.3 PageRank centrality

The PageRank centrality is the output of the PageRank algorithm which is focused on link analysis. The output is a distribution that models the likelihood of reaching any particular node when randomly moving along edges. Details of the algorithm can be found in Page *et al.* (1999).

2.3.4.2.4 Radiality centrality

The radiality centrality $C_R(v)$ (Valente and Foreman 1998) considers the global structure of the networks and indicates how connected an individual is in the entire network structure. It is defined as:

$$C_R(v) = \max_{x,y \in \{N\}} d(x,y) + 1 - \frac{1}{n-1} \sum_{\substack{w \in \{N\} \\ w \neq v}} d(v,w). \tag{16}$$

2.3.4.2.5 Evaluation score for hub identification

First, we ranked the common nodes between the inferred networks and the associated ground truth network according to the centrality metrics defined above. We then labeled the 10% nodes with the highest values of the metrics as hubs. Then, we analyzed the set similarity between the hubs in the ground truth or silver standard Ω_{true} and the inferred networks Ω_{inf}. To this aim, we computed the Jaccard coefficient J (Jaccard 1912) for every network, which is given by:

$$J(\Omega_{true}, \Omega_{inf}) = \frac{|\Omega_{true} \cap \Omega_{inf}|}{|\Omega_{true} \cup \Omega_{inf}|}. \tag{17}$$

To compare the performance between different types of networks, we used as an evaluation score the ratio between the Jaccard coefficient J computed on the inferred networks and the expected coefficient J_{rand} for a random predictor. J_{rand} can be calculated explicitly from the probability P of obtaining a given number of hubs, x, among randomly selected nodes Ω_{rand} (i.e. $x = |\Omega_{true} \cap \Omega_{rand}|$):

$$P(x) = \frac{\binom{n_0}{x}\binom{n-n_0}{n_0-x}}{\binom{n}{n_0}}, \tag{18}$$

where n is the total number of nodes and $n_0 = |\Omega_{true}| = |\Omega_{rand}|$ is the number of nodes selected as hubs from the ground truth network or from the random predictor. Using the above expression of $P(x)$, the expected value of J_{rand} obtained with a random predictor can be computed as:

$$J_{rand} = \sum_x J(x)P(x), \tag{19}$$

where the sum runs over all the possible values of $x \in [0, n_0]$, and $J(x)$ is the value of the Jaccard coefficient when the size of the intersection is x, namely $J(x) = \frac{x}{2n_0 - x}$ (as it can be easily obtained from the definition of the Jaccard coefficient, Eq. (17)). With large values of n and $n_0 \ll n$, it can be shown that the following approximation holds true (Chung *et al.* 2019): $J_{rand} \sim \frac{n_0/n}{2-n_0/n}$.

2.3.5 Aggregated ranking of the algorithms

To provide an overall ranking of the algorithms, we first computed a max-scaled score for each metric, over different network realizations for each network type, and then we aggregated the scores into overall scores for the information exchange efficiency, the hub topology, and the hub identification. We also computed a final overall topology score. Below, we describe in detail how the scores are computed.

For the hub topology and the information exchange efficiency, the average of the MSE of each metric P_x is computed for each network type t and algorithm a, where $|t|$ refers to the numbers of networks G that are part of t:

$$S_1(P_x, a, t) = \frac{1}{|t|} \sum_{G \in t} \text{MSE}_G(P_x, a). \tag{20}$$

Then, for each network type, the absolute values of these averages for all algorithms A are max-scaled by dividing the values by the score of the best-performing algorithm on this network type:

$$S_2(P_x, a, t) = \frac{S_1(P_x, a, t)}{\max_{i \in A} S_1(P_x, i, t)}. \tag{21}$$

These scores are then again averaged for each algorithm over the $|T|$ different network types T, subtracted from 1 to produce a score that increases with the performance. Finally, the scores are max-scaled again between all algorithms by dividing by the highest score of the algorithms to produce the final scores $F(P_x, a)$ for each metric:

$$S_3(P_x, a) = 1 - \frac{1}{|T|} \sum_{t \in T} S_2(P_x, a, t), \tag{22}$$

$$F(P_x, a) = \frac{S_3(P_x, a)}{\max_{i \in A} S_3(P_x, i)}. \tag{23}$$

For the hub identification measures P_y, the Jaccard ratios to a random predictor J/J_{rand} are also averaged for every network type and algorithm. From the average, we subtract 1, to produce a score that sets the performance of a random predictor to 0. Afterwards, the scores are also max-scaled by

dividing by the best-performing algorithm per network type, summed over the different network types. The remaining negative values are replaced with 0. These scores are then again max-scaled to produce the final scores $F(P_x, a)$ for the individual hub identification measures:

$$J_1(P_y, a, t) = \frac{1}{|t|}\sum_{G\in t}\frac{J(P_y, a)}{J_{rand}(P_y)} - 1, \tag{24}$$

$$J_2(P_y, a, t) = \frac{J_1(P_y, a, t)}{\max\limits_{i\in A} J_1(P_y, i, t)}. \tag{25}$$

$$J_3(P_y, a) = \max\left(0, \frac{1}{|t|}\sum_{t\in T} J_2(P_y, a, t)\right), \tag{26}$$

$$F(P_y, a) = \frac{J_3(P_y, a)}{\max\limits_{i\in A} J_3(P_y, i)}. \tag{27}$$

With the final scores for each metric, the overall information exchange, overall hub topology, and overall hub identification scores T_2 are calculated by summing up the final scores for the associated metrics P for each algorithm a and max-scaling by the best algorithm

$$T_1(P, a) = \sum_{p\in P} F(p, a). \tag{28}$$

$$T_2(P, a) = \frac{T_1(P, a)}{\max\limits_{i\in A} T_1(P, i)}. \tag{29}$$

The overall topology score is calculated similarly by calculating T_2 with P as all metrics together.

2.3.6 Correlation analysis of the different metrics

To analyze the relationship between the performance in the evaluations of the different metrics, we computed Spearman's correlation coefficient ρ between the performance scores computed as detailed below.

We pooled all results from the synthetic data or the experimental data. Then, for the information exchange and hub topology metrics, we used the negative absolute mean signed error (−|MSE|), to have a score that increases with the performance. For the hub identification metrics, we used as a score the Jaccard coefficient ratio to a random predictor (J/J_{rand}). Finally, we defined as performance scores relative to edge prediction the values of *EPr*, *AUROC*, *AUPRC*. In the heatmaps of Supplementary Fig. S1B and C, we crossed out the correlation values corresponding to a P value above .01.

3 Results

3.1 Overview of STREAMLINE

The steps involved in STREAMLINE are schematically represented in Fig. 1. We consider three types of datasets: simulated, Cur, and experimental datasets.

With the simulated datasets, we generated scRNA-seq data *in silico* from four classes of networks with well-defined and different structural properties, to be able to test the algorithms in different scenarios. The classes of networks we consider are Random, Small-World (SW), Scale-Free (SF), and Semi-Scale-Free (SSF) Networks. Random or Erdös–Renyi (ER) networks include a set of nodes in which each node pair has the same probability of being connected by an edge (Erdős and Rényi 1959). We include this class of networks as a control. In SF networks, the edges are drawn such that the degree distribution follows a power law (Barabási *et al.* 2000). SF networks have been considered ubiquitous in cell biology (Albert 2005), but their presence, at least on a global network level, is still debated (Broido and Clauset 2019). For this reason, we also employ SSF networks, in which only the out-degree distribution follows a power law, while the in-degree distribution is uniform. Such networks were introduced by Ouma *et al.* (2018) to model real GRNs. SW networks have the property that the neighbors of any given node are likely to be neighbors of each other (Watts and Strogatz 1998). The SW property has been observed, for instance, in yeast (Van Noort *et al.* 2004) and human lung cancer (Sun *et al.* 2006) GRNs. In addition to these, we included four Cur networks that consist of sub-networks of known GRNs (Pratapa *et al.* 2020).

Networks from each class are defined by a set of parameters. To make our results independent of specific instances of networks, we sampled multiple networks from each class with different combinations of parameters and two sizes: a smaller (15 nodes and 50 edges) and a larger (25 nodes and 100 edges) size. All the results shown below are averaged over all the instances of networks generated for a given class. Details about the network classes and the parameters used for network sampling are provided in the Methods section, Supplementary Section S1 and Supplementary Table S1. From each of these networks, we simulated scRNA-seq datasets using BoolODE, a recently developed software based on ordinary differential equations (Pratapa *et al.* 2020) (see Methods section).

In addition to simulated datasets, we also considered four real scRNA-seq datasets generated from different organisms and cell types: yeast (Gasch *et al.* 2017), mouse dendritic cells (mDC) (Shalek *et al.* 2014), mouse embryonic stem cells (mESC) (Tran *et al.* 2019), and human embryonic stem cells (hESC) (Han *et al.* 2018). These datasets were used in a previous benchmarking study (McCalla *et al.* 2023), where the authors also provide estimations of silver standard networks. We report the details about the experimental datasets in Supplementary Table S2.

The second step of our pipeline involves running the algorithms to infer GRNs from each of the datasets. We chose the four top-performing algorithms according to a recent study where the accuracy in predicting gene–gene interactions was evaluated (Pratapa *et al.* 2020): PIDC (Chan *et al.* 2017), PPCOR (Kim 2015), SINCERITIES (Papili Gao *et al.* 2018), and GRNBoost2 (Moerman *et al.* 2019). Two of these methods (PIDC and PPCOR) give output as undirected networks, while SINCERITIES and GRNBoost2 provide directed networks. A brief description of each algorithm is included in the Methods section. To make the results comparable between the different algorithms, we considered the undirected version of the networks inferred by SINCERITIES and GRNBoost2. We show the effect of taking the edge direction into account in the Supplementary Material.

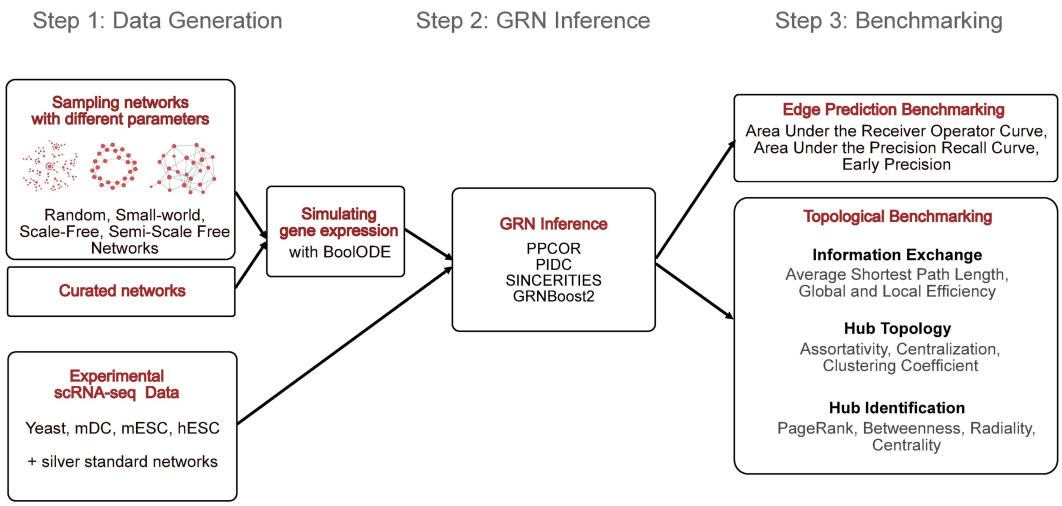


Figure 1. Schematic overview of STREAMLINE. STREAMLINE consists of three steps: first, synthetic scRNA-seq data are generated from different classes of networks (Step 1). Then, GRN inference methods are applied to synthetic as well as real data (Step 2). Finally, the methods' performance on the predictions of edges and of structural network properties (quantifying the network robustness and hub presence) is evaluated (Step 3).

In our analysis, we first scored each method's ability to predict the presence of edges. Specifically, we calculated the early precision (EPr) on the experimental and synthetic data. For the synthetic data, we computed also the area under the receiver operator curve (AUROC) and the area under the precision-recall curve (AUPRC) (Pratapa *et al.* 2020). The results are then grouped for each network class or organism, and we found results in line with previous studies (see (Pratapa *et al.* 2020, Chen and Mar 2018), Supplementary Section S2 and Supplementary Fig. S1).

Then, we analyzed the ability of each method to predict global properties of networks, which are defined at a graph level. In particular, we computed topological properties that quantify how efficiently the information is exchanged in the network and the tendency of networks to include hubs.

The efficiency of information exchange measures how the behavior of a network can change following variations in its topology due to, e.g., the failure of some of its constituents (Latora and Marchiori 2001). In this context, it can be used to assess the stability of a GRN, as it is subject to random errors due to mutations and extreme conditions that can hinder regulatory interactions (Boccaletti *et al.* 2006). The following topological measures can quantify the efficiency of information exchange: the Global Efficiency, the Local Efficiency, and the Average Shortest Path Length (see Methods section). The Global and Local Efficiency measures quantify the fault tolerance of a system (Latora and Marchiori 2007) and they have already been used to study the relationship between evolutionary and topological properties of human GRNs (Szedlak *et al.* 2016). The Average Shortest Path Length has been widely adopted as a measure of biological network navigability (defined as the ability to efficiently move from a source to a target node through short communication paths), which is crucial for information distribution (Barabasi and Oltvai 2004, Boguñá *et al.* 2009). Another biologically important property of networks is the presence of hubs, i.e., nodes that have a degree much larger than the average. In GRNs, hubs are genes that regulate the expression levels of many other genes and can represent master regulators of a biological process. Through structural network analysis, it has been shown that the presence of hubs is highly sensitive to perturbations in network topology (Ghoshal and Barabasi 2011), and it is linked to global topological quantities like the Centralization, the Assortativity, and the Clustering Coefficient (Sporns *et al.* 2007, Pechenick *et al.* 2012), which we compute in STREAMLINE.

In addition to quantifying the tendency of networks to possess hubs, it is important to identify them correctly. Hence, we tested the GRN inference algorithms for their ability to predict which nodes constitute hubs. To do so, we computed four local metrics used to detect hubs (Koschützki and Schreiber 2008)—Page Rank Centrality, Betweenness Centrality, Centrality, and Radiality (see Methods section)—and we compared the values obtained from the ground truth networks versus those calculated from the inferred networks.

Below, we describe the detailed results of each of these benchmarking analyses.

3.2 Estimation of information exchange efficiency

To quantify the efficiency of information exchange, we evaluated the Average Shortest Path Length, the Global Efficiency, and the Local Efficiency (Fig. 2A) of the inferred and ground truth networks. Then, we quantified the accuracy of the estimations obtained from the inferred networks by calculating the mean signed error (MSE) between these quantities computed on the ground truth networks and the networks inferred from each of the algorithms (see Methods section).

First, we considered the simulated datasets generated from different classes of networks. The different structural properties of each class of networks are reflected by different values of these topological measures, as shown in Supplementary Fig. S2A. For example, the SW networks are characterized by a larger Clustering Coefficient and higher Global and Local Efficiency compared to ER networks, as expected based on their properties (Watts and Strogatz 1998).

In Fig. 2B, we report the MSE for all the topological measures computed on the simulated datasets. With Global Efficiency, we observed relatively accurate estimations for all the synthetic networks with all the algorithms (the absolute value of the average MSE per network type is $\sim 3\%$ of the ground truth values; see Supplementary Fig. S2A). However, we observed that GRNBoost2 and PPCOR tend to

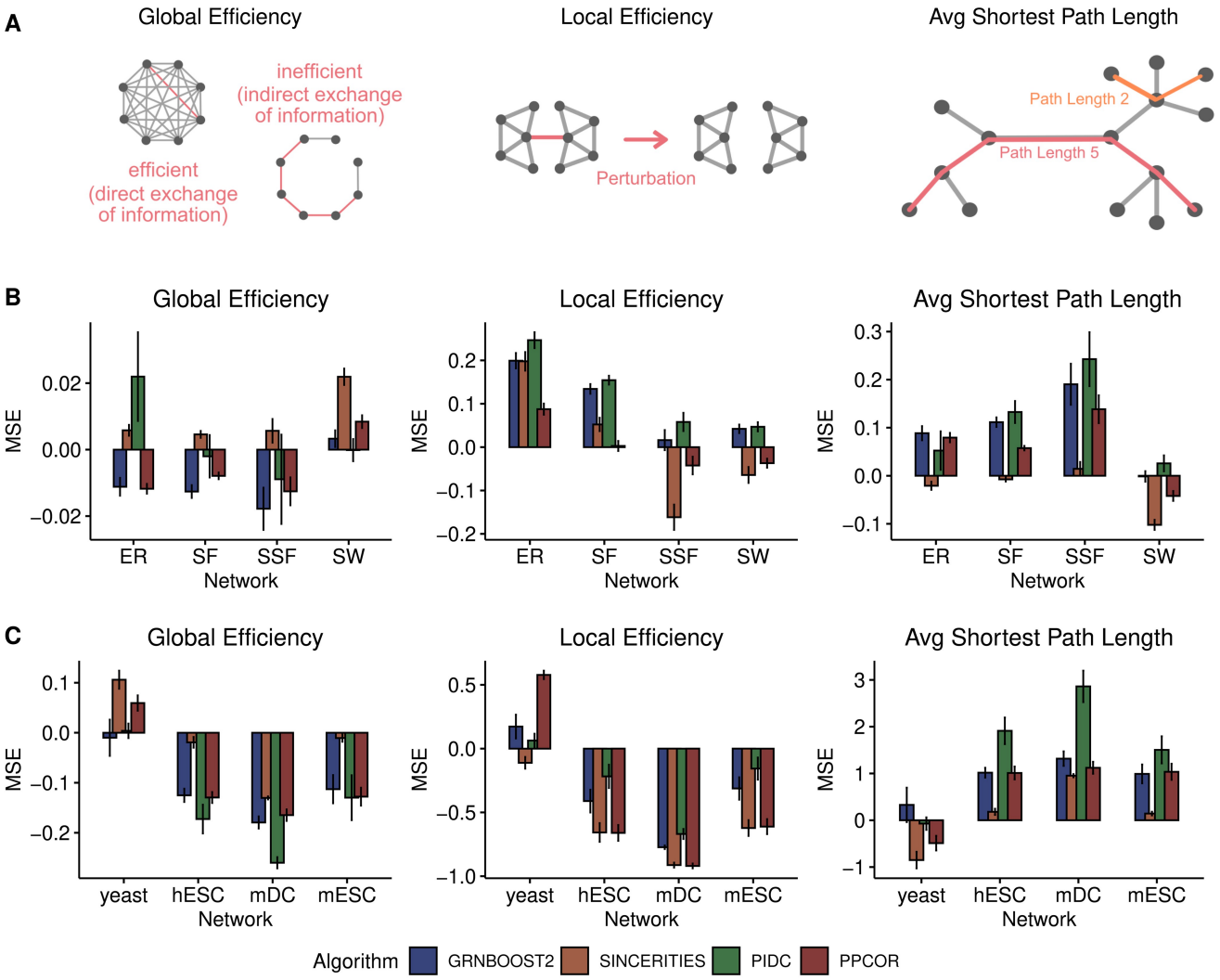


Figure 2. Results of the topological benchmarking of GRN inference algorithms with respect to information exchange both on synthetic and experimental scRNA-seq datasets. (A) Schematic representations of the three topological measures we computed (see Methods section). Global Efficiency quantifies how well the information can be distributed in the entire network. Local Efficiency measures how robust the network is to perturbation on a small scale. The Average Shortest Path Length specifies how many links are necessary to go from one node to another on average. (B) Barplots showing the mean signed error (MSE) for the estimations of the topological properties written at the top in different types of synthetic networks (indicated on the *x*-axis) and for different algorithms (marked by colors). (C) Same as B, for networks estimated from real scRNA-seq datasets (indicated on the *x*-axis). The heights and the error bars display the mean of the MSE values and the standard error of the mean, respectively, computed across datasets and networks in panel B and across networks in panel C.

underestimate it in ER, SF, and SSF networks, while SINCERITIES always overestimates it. The dependence on the type of network is particularly evident for some algorithms, like PIDC: while it provides the most accurate estimation of Global Efficiency in SW networks (MSE ~ 0), it shows the worst performance with ER networks (where the associated MSE is the largest).

We saw similar variability in the estimations of Local Efficiency. All the tested algorithms tend to overestimate it (MSE > 0), except for SINCERITIES and PPCOR, which underestimate it in SSF, SW, and Cur networks (Supplementary Fig. S3 and Supplementary Table S3). The best predictions (corresponding to MSE~ 0) are obtained on SW and SSF networks.

For the Average Shortest Path Length (Fig. 2B) SINCERITIES and PPCOR provide the best estimations, especially in the ER, SF, and SSF networks, while PIDC and GRNBoost2 perform better for SW networks. In the Cur networks, for which the Average Shortest Path Length is greater than for SSF graphs (Supplementary Fig. S2A and Supplementary Table S3), the algorithms tend to underestimate this property.

We performed the same analysis on four real scRNA-seq datasets from three species (Fig. 2C). The corresponding silver standard networks have lower Global Efficiency, similar Local Efficiency, and larger Average Shortest Path Length compared to the synthetic networks we considered, except for the yeast dataset that stands out for its lower values of the Local and Global Efficiency (Supplementary Fig. S2B).

The corresponding values of MSE are reported in Fig. 2C. Interestingly, we found an overall tendency of all algorithms to underestimate both the Local and Global Efficiency, except for the yeast dataset. SINCERITIES provides the most accurate predictions of the Global Efficiency in the hESC,

mDC, and mESC networks, while PIDC outperforms the other algorithms for the yeast dataset. PIDC is also the top-performing algorithm for Local Efficiency (Fig. 2C).

As for the Average Shortest Path Length, the MSE is mostly positive, indicating an overestimation, and it is smallest for SINCERITIES, which is the best-performing algorithm for all networks except for the yeast dataset, where it is outperformed by PPCOR.

Overall, the analysis above shows that the accuracy of the estimations of the topological properties measuring the information exchange depends both on the type of network and the algorithm. Moreover, at least with synthetic networks, properties like Global Efficiency are estimated with a small relative error ($\sim 3\%$), while the relative errors on the other properties and with experimental networks can be much larger (Fig. 2B and C and Supplementary Fig. S2A).

Next, we checked whether the performance of the algorithms in predicting these topological properties correlated with their ability to predict edges in the network. Interestingly, we found that the correlation between the performance measured in these two tasks is either statistically non-significant or very small (Supplementary Fig. S1B and C).

3.3 Hub analysis

One important downstream analysis on GRNs is the identification of genes with a number of links much larger than the average. These are known as network hubs, which can play key roles in differentiation and reprogramming (Kim *et al.* 2021) and have been identified as potential disease regulators or drug targets (Åkesson *et al.* 2021). The presence of hubs depends on several topological properties that change with the type of network. For example, we expect the hubs in SF and SSF networks to be more easily identifiable due to their node-degree distribution. Such a feature of SF and SSF networks is reflected by their higher Centralization values compared to other classes of networks (Supplementary Fig. S2A).

Here, we analyzed how accurately the algorithms can predict the values of two groups of topological properties: the first group of graph properties quantifies the tendency of networks to include hubs (Assortativity, Clustering Coefficient, and Centralization), and the second group of node properties is used to identify hubs (Betweenness, Centrality, Radiality, and PageRank). For hub identification, we use random networks for baseline predictions.

3.3.1 Hub-related topological quantities

The topological measures we chose to quantify the presence of hubs are Assortativity, Clustering Coefficient, and Centralization (Fig. 3A). In networks with negative and larger absolute values of Assortativity, nodes with lower degrees tend to be linked to nodes that feature a higher degree; hence, in these networks, hubs tend to be present and clearly identifiable. Networks with a large Clustering Coefficient feature groups of nodes with high interconnectivity that, thus, have similar node degrees. In this situation, hubs are less dispersed. The Centralization quantifies how centralized a graph is around a small number of nodes, which will have a large number of links, and will therefore tend to be strong and clearly identifiable hubs.

The synthetic networks we simulated data from are characterized by different values of the above topological measures due to their properties (see (Ouma *et al.* 2018), Supplementary Fig. S2A for undirected networks and Supplementary Fig. S4A for directed networks). This allowed us to explore the performance of the algorithms under different conditions. The three network properties assessed during this step uncovered multiple algorithm-specific behaviors (Fig. 3B). The most evident involves the algorithm SINCERITIES, which yielded GRNs with lower Assortativity (Supplementary Fig. S2A), which leads to an underestimation of this property for almost all types of networks, including the Cur networks (Fig. 3B and Supplementary Table S3). The results for the Clustering Coefficient are similar to those obtained for the Local Efficiency, with an overestimation of this property in ER and SF networks, and better performance of the algorithms in the case of SSF and SW networks (Fig. 3B). This result is in line with a known general relationship between these two metrics (Strang *et al.* 2018). In the case of the Cur networks, we also observed a tendency to overestimate the Clustering Coefficient, although the behavior is more dataset-specific (Supplementary Table S3).

When using directed networks, we found that the estimations of the Clustering Coefficient by GRNBoost2 and SINCERITIES change only slightly, while for the Assortativity we observed marked differences for SINCERITIES, which overestimates this property in all networks (Supplementary Fig. S4C).

As in the previous section, we repeated the analysis using real datasets (Fig. 3C). The corresponding silver standard networks have Assortativity values that are lower than those of most synthetic networks and are in line with the SF hypothesis for GRNs (Ouma *et al.* 2018). On average, the Clustering Coefficients are similar to those of SF and SSF networks (Supplementary Fig. S2B), except for the yeast dataset that shows smaller values of this metric. Such values of the topological properties indicate that these networks display a higher tendency to contain hubs that are more clustered together when compared to random networks.

Similarly to what happens with the synthetic datasets, here, the Centralization is overestimated by SINCERITIES (Fig. 3C), except for the mDC networks that are more centralized (Supplementary Fig. S2B). In contrast to the synthetic networks, the Assortativity is now overestimated by SINCERITIES rather than being underestimated.

GRNBoost2, PIDC, and PPCOR show similar performances. These three algorithms overestimate the Assortativity and underestimate the Centralization. Furthermore, the Clustering Coefficient is underestimated by all algorithms for the hESC, mDC, and mESC datasets, while the estimations are more accurate for the yeast dataset, whose silver standard networks have smaller values of this metric (Supplementary Fig. S2B).

Overall, we found differences in performance between the real and the synthetic data, which might be due to a number of factors. First, the silver standards provide only estimates for GRNs of the three organisms, while the synthetic data are simulated from fully specified networks. Furthermore, we found that the algorithms tend to output networks that feature similar values for the topological properties, regardless of the type of network they are run on (Supplementary Fig. S2). This might explain, e.g., the opposite trends in the estimations of the Assortativity and the Clustering Coefficient with the synthetic versus the real datasets, as observed in particular for the SINCERITIES algorithm.

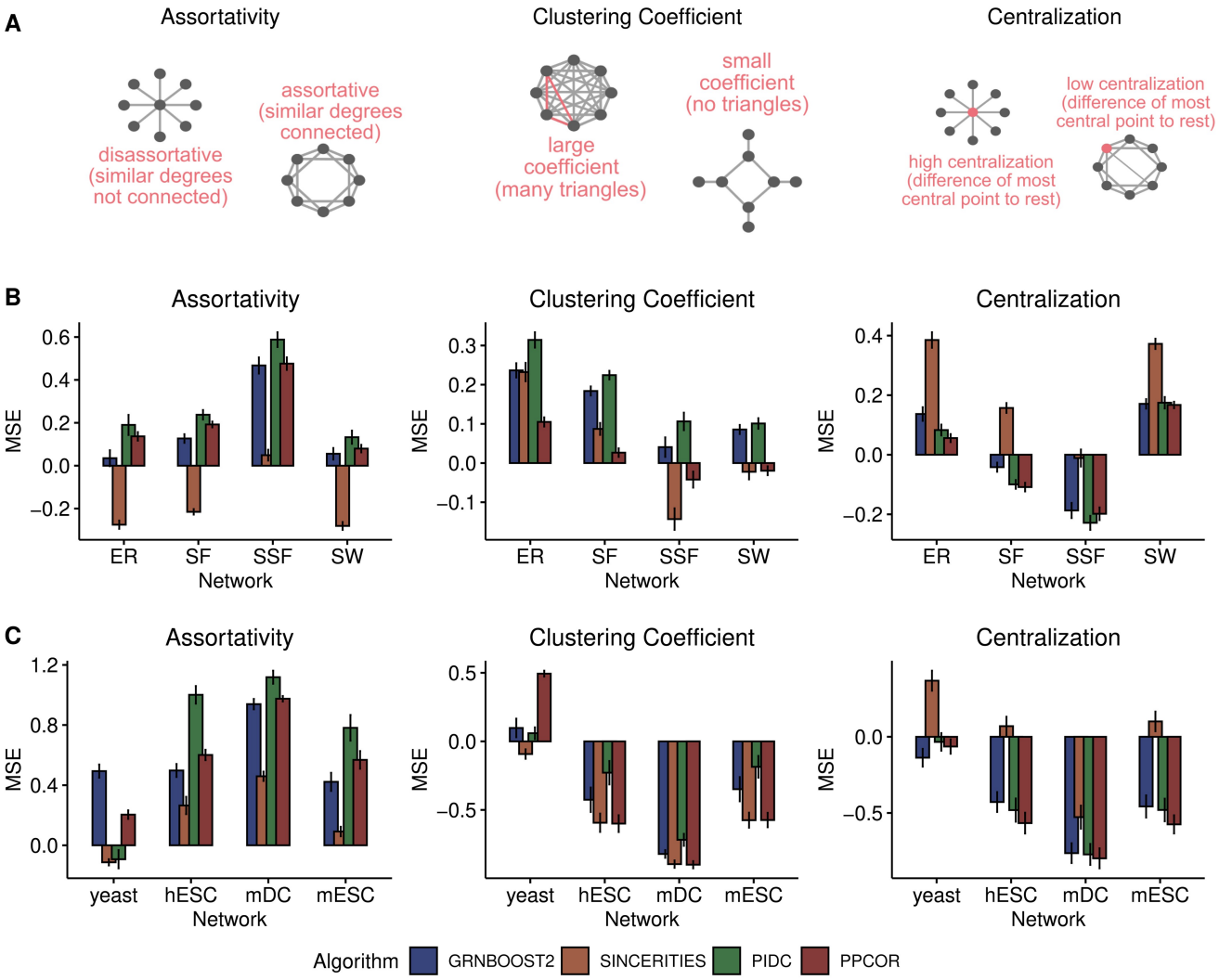


Figure 3. Results for the topological benchmarking of GRN inference assessing the presence of hubs. (A) Schematic representation of the three topological measures considered here (see Methods section). The Assortativity quantifies the tendency of nodes in the networks to attach to others with similar degrees. The Clustering Coefficient reflects how much the nodes in a graph tend to cluster together. The Centralization indicates how strongly the network is arranged around a single center. (B) Barplots showing the mean signed error (MSE) for the estimations of the topological properties written at the top in different types of synthetic networks (indicated on the *x*-axis) and for different algorithms (marked by colors). (C) Same as B, for networks estimated from real scRNA-seq datasets (indicated on the *x*-axis). The heights and the error bars display the mean of the MSE values and the standard error of the mean, respectively, computed across datasets and networks in panel B and across networks in panel C.

Finally, we checked whether the algorithms' ability to estimate the hub-related topological quantities correlates with their performance in predicting network edges. Consistently with what we observed before when looking at information exchange (see previous section), we found little to no correlation (Supplementary Fig. S1B and C).

3.3.2 Hub identification

While hubs are loosely defined as nodes having degrees higher than average, there is no consensus on the best metric to identify them. For this reason, here we compute four centrality measures that have been previously adopted to find hubs in GRNs (Koschützki and Schreiber 2008): the Betweenness (Freeman 1977), the Centrality (Freeman 1978), the Radiality (Valente and Foreman 1998), and the Page Rank (Page *et al.* 1999) (see Methods section). Among these, the Page Rank and the Centrality metrics are conserved along evolution and relevant in pluripotent cells (Wolf *et al.* 2021). Moreover, they were proposed as metrics to distinguish life-essential versus specialized subsystems (Wolf *et al.* 2021).

We verified how accurately the hub identification measures are estimated by the four inference algorithms introduced above. More specifically, we selected the set of top 10% nodes according to the centrality measure computed in the ground truth network, Ω_{true}, and in the inferred network, Ω_{inf}. Then, we quantified the similarity between the two sets of nodes with the Jaccard coefficient, J (Jaccard 1912) (see Methods section). Finally, we computed the ratio between J and J_{rand}, i.e., the Jaccard index between Ω_{true} and a set of randomly selected nodes Ω_{rand} (see Methods section). Hence, the ratio J/J_{rand} shown in Fig. 4 represents how well the hubs can be predicted from the inferred networks with respect to a random guess for synthetic (panel A) and experimental (panel B) networks.

In synthetic networks, we obtain similar values of the Jaccard coefficient on average (Supplementary Fig. S5A) but lower values of J/J_{rand} (Fig. 4A), compared to the

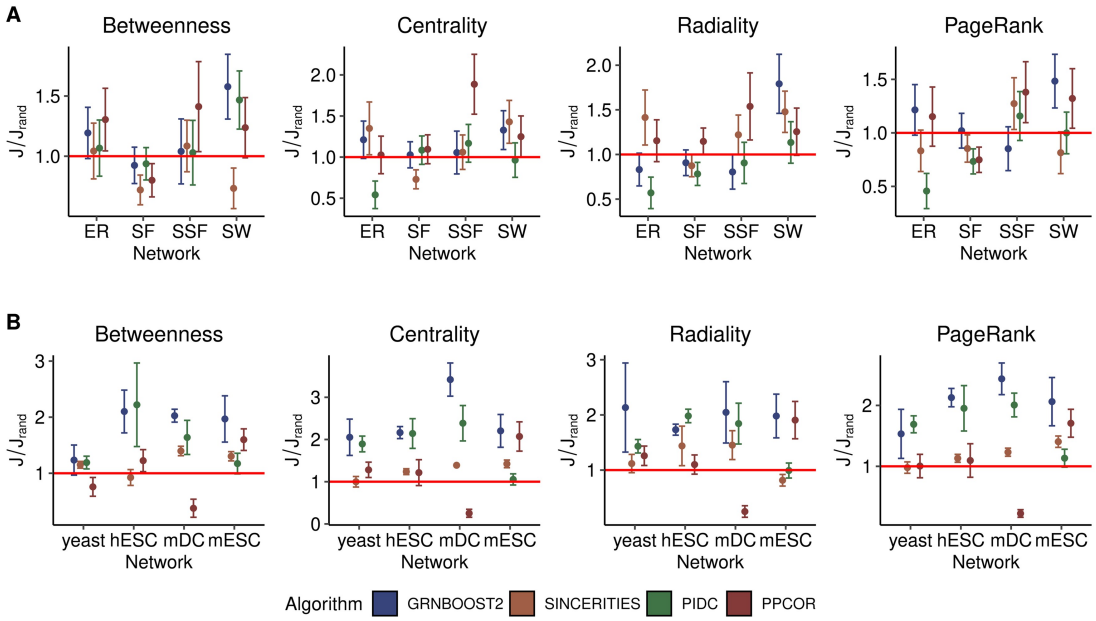


Figure 4. Accuracy of hub detection. The accuracy is measured by the Jaccard coefficient ratio, J/J_{rand}, using a random predictor as a reference (see Methods section). (A) J/J_{rand} is plotted for various hub metrics (written on top) as a function of the type of synthetic network (indicated on the *x*-axis) for different algorithms. (B) Same as (A), plotted as a function of the networks inferred from real scRNA-seq datasets (indicated on the *x*-axis). The Betweenness estimates the influence that a node has on the information exchange in a graph based on path lengths. The Centrality is the normalized total degree of a node. The Radiality assigns high centrality values to nodes with a short distance to all vertices in their reachable neighborhood compared to the graph diameter. PageRank is a generalization of the degree centrality that considers the eigenvalues of a modified adjacency matrix. We provide a detailed definition of the hub metrics in the text and the Methods section. The points and the error bars display the mean of the Jaccard coefficient ratio and the standard error of the mean, respectively, computed across datasets and networks in panel A and across networks in panel B.

experimental networks (Fig. 4B). Better performances are obtained in SSF networks, which are the most centralized, and SW networks, while the performances are generally poor for the SF networks. These results are likely related to the underestimation for SF networks and overestimation for the SW networks of the values of the Centralization that we previously observed (Fig. 3B and Supplementary Fig. S2A).

For the experimental networks, we find that higher values of J/J_{rand} are achieved by GRNBoost2 and PIDC in networks with stronger hubs (mDC, hESC, and mESC) (Fig. 4B), which features larger values of the Centralization and Clustering Coefficient compared to the yeast network (Supplementary Fig. S2B). However, the values of the Jaccard coefficients, J, are smaller than 0.2, indicating an overall poor performance of the algorithms (Supplementary Fig. S5B). PIDC and GRNBoost2 emerge as the top-performing algorithms, especially in the hESC and mDC networks, depending on the hub identification metric.

When we ran the analysis on directed networks, we found similar performance on synthetic networks, while for the experimental networks we observed a large increase of J/J_{rand}, especially for GRNBoost2 when using the Betweenness or the Out Centrality (Supplementary Fig. S6). This result is in line with recent results on hub identification from experimental scRNA-seq data (Kim *et al.* 2021, Fiorentino *et al.* 2024).

4 Discussion

Here, we performed benchmarking analyses to evaluate how well GRN inference algorithms can estimate the structural properties of the networks. More specifically, we quantify the ability of the algorithms to infer their robustness to perturbations as well as the presence and identification of network hubs. For this purpose, we computed six topological measures and tested four metrics for hub identification that are widely known and used in network theory. Moreover, we considered scRNA-seq data simulated from different types of networks as well as real data collected from different organisms.

In this extensive benchmarking, we focused on network properties (i.e. robustness and hubs) that are not taken into account in currently available benchmarking studies performed on scRNA-seq data, despite being considered important when studying GRNs and, more in general, biological networks (see, e.g. Noman *et al.* 2015, MacNeil and Walhout 2011, Winterbach *et al.* 2011). For example, the identification of putative master regulators via degree-based measures on GRNs is a commonly used practice (see, e.g. Koschützki and Schreiber 2008, Cholley *et al.* 2018, Padi and Quackenbush 2015). We chose to focus on general topological properties of networks whose definition and interpretation do not require assumptions on the biological process under study. However, more targeted approaches to investigate, e.g., network robustness, such as the *in silico* perturbation of specific genes (Theodoris *et al.* 2023) might be included in STREAMLINE in the future.

The benchmarking results are summarized in Fig. 5 (for undirected networks) and Supplementary Fig. S7 (for directed networks).

Interestingly, we found that the algorithms' performance in edge detection (Fig. 5 and Supplementary Fig. S1) has weak or no correlation with their performance in estimating the topological properties of networks, which indicates the need for a targeted benchmarking analysis like STREAMLINE.

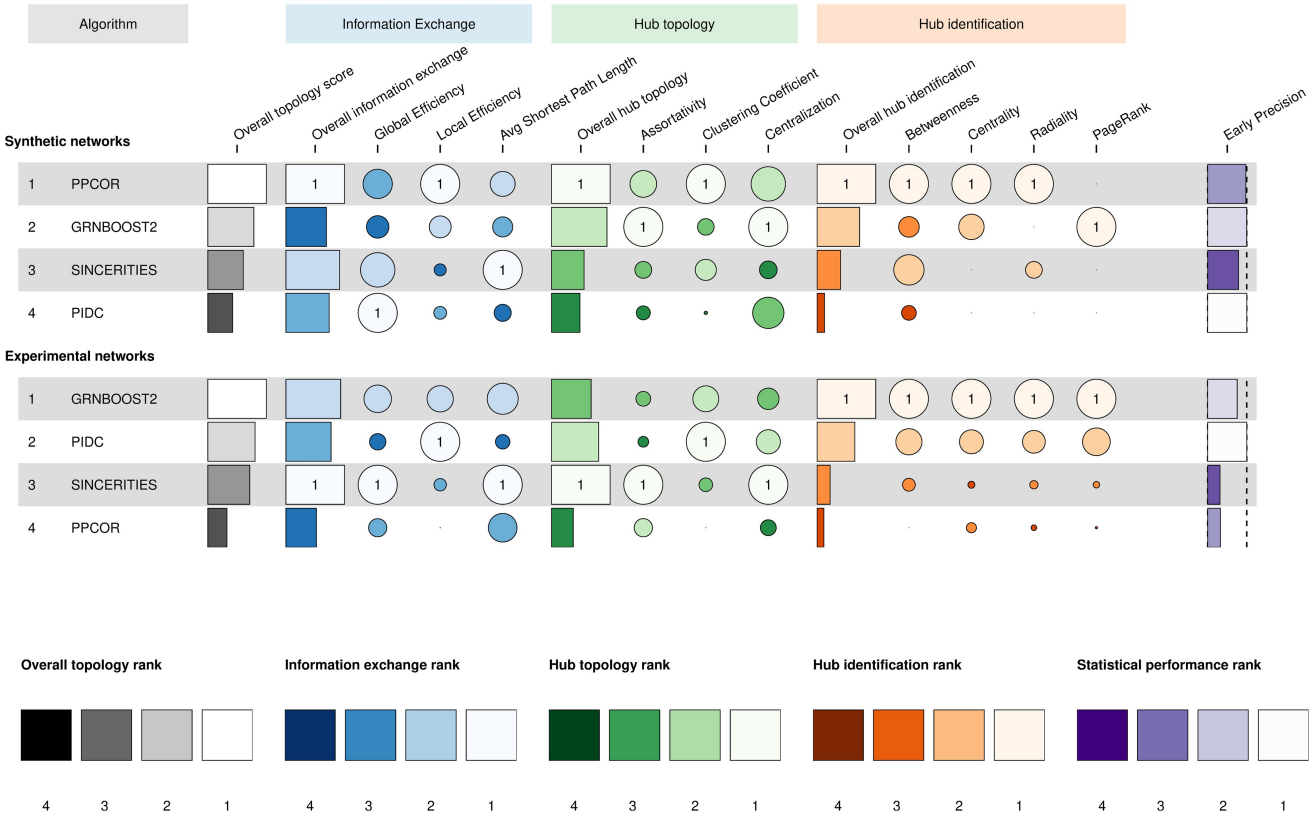


Figure 5. Ranking of the GRN inference algorithms. We report the overall performance of the algorithms on each topological metric for synthetic (top rows) and experimental (bottom rows) datasets. The algorithms are ranked according to an overall topology score (see Methods section). We also show the ranking for each group of topological metrics (Information Exchange, Hub Topology, and Hub Identification) and we report the performance in binary edge detection in the last column. The legend at the bottom shows the association between the colors and the ranks for each group of metrics.

Moreover, this result also implies that a GRN inference algorithm with poor performance in edge prediction can still provide accurate estimates of global network properties.

For synthetic datasets, we found that PPCOR is the best-performing algorithm in the three tasks that we benchmarked for topological metrics: Information Exchange, Hub Topology, and Hub Identification. However, we highlight that other algorithms might be preferred for the estimation of specific topological metrics. Indeed, SINCERITIES emerges as the top-performing algorithm for estimating the Average Shortest Path Length and GRNBoost2 has the best performance in estimating the Assortativity and the Centralization in the Hub Topology task, and the PageRank metric in the Hub Identification task.

In the case of experimental networks, we found that the best-performing algorithm is more dependent on the specific topological metrics (Fig. 5). While GRNBoost2 provides the best estimates for all the metrics related to Hub Identification, SINCERITIES is the top-performing one for the metrics related to the Information Exchange and Hub Topology. The Local Efficiency and Clustering Coefficient, for which we observed that the estimates are closely related for all algorithms (Figs. 2 and 3), are an exception since they are estimated better by PIDC.

We also found that, with directed networks, GRNBoost2 overall performs better than SINCERITIES (Supplementary Fig. S7) and the identification of hub genes is generally more accurate, as suggested by the higher values of J/J_{rand} (Fig. 4 and Supplementary Fig. S6).

The benchmarking done with synthetic networks allowed us to check the performance of algorithms with networks having specific and tunable properties. In some cases, this has brought to light specific biases present in the networks estimated by each algorithm. In particular, for most of the algorithms the inferred average values of some metrics (e.g. Assortativity and Local Efficiency) for different types of synthetic networks are close and do not show any clear trend, unlike their ground truth values (Supplementary Fig. S2). In other cases, the trend can even be inverted, as in the case of the Clustering Coefficients estimated by SINCERITIES, which are highest on average for ER networks and decrease in SF and SSF networks, while the ground truth values show the opposite trend (Supplementary Fig. S2).

Some algorithms show specific features. For example, SINCERITIES produces more disassortative and centralized networks (i.e. networks with relatively low Assortativity and high Centralization), which causes an underestimation of Assortativity and overestimation of Centralization for all types of synthetic networks (Fig. 3A). Similar observations can be made, e.g., with GRNBoost2, which tends to generate networks with lower Global Efficiency (Fig. 2A). While the underlying reasons for these observations remain elusive, we speculate that they might be caused by differences in the specific designs of the algorithms. For instance, SINCERITIES is a causality-based method that uses a linear regression model on temporal data, similar to Granger causality, which is known to have high false positive rates when its underlying assumptions are violated, as is the case in complex datasets

with nonlinear dynamics (Yuan and Shou 2022). More specifically, Granger causality fails when the system's dynamics are deterministic (or have a low noise level) or when pairs of variables have a common unobserved cause (Yuan and Shou 2022). The reason why SINCERITIES tends to output disassortative and centralized networks might come from the preferential attribution of false positive edges to specific genes, due to their specific dynamics.

Importantly, the knowledge of such biases can guide the effort to improve current algorithms, e.g., by assisting in the design of objective functions that can lead to networks with global properties closer to real GRNs. This approach can be justified by the observation that GRNs share certain topological features, such as an SF (Lopes *et al.* 2014) or SSF (Ouma *et al.* 2018) node-degree distributions, which could be assumed as prior knowledge during the inference process.

In this study, we chose to focus on unsupervised GRN inference algorithms, which are currently widely used by the community and are well characterized in their performance with local metrics (Pratapa *et al.* 2020, Chen and Mar 2018). It will be interesting to include in future versions of STREAMLINE the benchmarking of (semi-)supervised methods for GRN inference and the calculation of global metrics taking into account edge weights.

The precise definition of the ground truth has crucial importance in benchmarking studies. Current studies rely on either simulated data or experimental silver standard networks. While these methods represent the state-of-the-art, it is essential to acknowledge the limitations associated with both types of ground truths. Simulated data faces constraints related to the parametrization of ground truth networks into ODE or SDE models, as exemplified by BoolODE, which could affect network identifiability (Erbe *et al.* 2023). Nonetheless, innovative approaches are emerging, leveraging mechanistic models of gene regulation (Erbe *et al.* 2023), or deep-learning-based models (Zinati *et al.* 2023). These advancements aim to directly generate scRNA-seq datasets that encode direct causal regulatory relationships. Conversely, experimental silver standard networks are typically derived from ChIP-seq experiments. These experiments, while valuable, come with known limitations, including sequencing errors and GC bias (Park 2009). Such limitations may result in missing or inaccurate edges within the ground truth network. Furthermore, the incompleteness of these networks can affect the estimation of topological properties. These limitations are shared across benchmarking studies. However, the development of robust, widely applicable computational pipelines, such as STREAMLINE, is essential for ongoing enhancements in ground truth network generation and paves the way for more accurate assessments in the evolving landscape of benchmarking studies.

Finally, the topological quantities we considered can also be used to optimize community-based inference schemes. Currently, consensus networks are derived from the outputs of different methods by taking into account only their performance in estimating edges. Instead, new strategies could be devised that also consider the estimation of the network's topological properties.

Author contributions

Marco Stock and Niclas Popp generated data, analyzed data, and produced the figures. Jonathan Fiorentino and Antonio Scialdone conceived the study and provided project supervision. All authors interpreted the results and wrote the manuscript.

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest

No competing interest is declared.

Funding

Jonathan Fiorentino and Marco Stock were supported by a Joachim Herz Stiftung Add-on Fellowship for Interdisciplinary Life Science. Marco Stock was supported by the Helmholtz Association under the joint research school "Munich School for Data Science—MUDS". Work in the Scialdone lab is funded by the Helmholtz Association.

Code availability

The STREAMLINE pipeline in Python, the code for reproducing the figures in R and a tutorial for using the pipeline are available at https://github.com/ScialdoneLab/STREAMLINE. The python pipeline is using Networkx (Hagberg *et al.* 2008) and graph-tool (http://figshare.com/articles/graph_tool/1164194) for graph analysis and handling.

Data availability

We used publicly available experimental single-cell RNA-sequencing data from yeast (Gasch *et al.* 2017), mouse (Shalek *et al.* 2014, Tran *et al.* 2019), and human (Han *et al.* 2018) with Cur silver standard networks from McCalla *et al.* (2023). The normalized gene expression, the pseudo time data and silver standard networks can be downloaded from https://doi.org/10.5281/zenodo.5907527. We made available all the data generated in this study at the Zenodo repository https://doi.org/10.5281/zenodo.10710444. This repository includes the synthetic and Cur ground truth networks, the associated simulated scRNA-seq data and pseudo time files, the networks inferred by the four algorithms that we used, and the raw evaluation results (i.e. the values of the topological metrics) for all the networks we analyzed.

References

Akers K, Murali T. Gene regulatory network inference in single-cell biology. *Curr Opin Syst Biol* 2021;**26**:87–97.

Åkesson J, Lubovac-Pilav Z, Magnusson R *et al.* Comhub: community predictions of hubs in gene regulatory networks. *BMC Bioinformatics* 2021;**22**:58.

Albert R. Scale-free networks in cell biology. *J Cell Sci* 2005;**118**:4947–57.

Banf M, Rhee SY. Computational inference of gene regulatory networks: approaches, limitations and opportunities. *Biochim Biophys Acta Gene Regul Mech* 2017;**1860**:41–52.

Barabasi A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004;**5**:101–13.

Barabási A-L, Albert R, Jeong H. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Stat Mechan Appl* 2000;**281**:69–77.

Boccaletti S, Latora V, Moreno Y *et al.* Complex networks: structure and dynamics. *Phys Rep* 2006;**424**:175–308.
Boguñá M, Krioukov D, Claffy KC. Navigability of complex networks. *Nature Phys* 2009;**5**:74–80.
Brennecke P, Anders S, Kim JK *et al.* Accounting for technical noise in single-cell rna-seq experiments. *Nat Methods* 2013;**10**:1093–5.
Broido AD, Clauset A. Scale-free networks are rare. *Nat Commun* 2019;**10**:1017.
Chan TE, Stumpf MP, Babtie AC. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst* 2017;**5**:251–67.e3.
Chen S, Mar JC. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics* 2018;**19**:232–21.
Cho YS, Kim JS, Park J *et al.* Percolation transitions in scale-free networks under the achlioptas process. *Phys Rev Lett* 2009; **103**:135702.
Cholley P-E, Moehlin J, Rohmer A *et al.* Modeling gene-regulatory networks to describe cell fate transitions and predict master regulators. *NPJ Syst Biol Appl* 2018;**4**:29.
Chung NC, Miasojedow B, Startek M *et al.* Jaccard/tanimoto similarity test and estimation methods for biological presence–absence data. *BMC Bioinformatics* 2019;**20**:644–11.
Erbe R, Stein-O'Brien G, Fertig E. A mechanistic simulation of molecular cell states over time. *bioRxiv* 2023.
Erdős P, Rényi A. On random graphs I. *Publ Math Debrecen* 1959; **6**:290–7.
Escorcia-Rodríguez JM, Gaytan-Nuñez E, Hernandez-Benitez EM *et al.* Improving gene regulatory network inference and assessment: the importance of using network structure. *Front Genet* 2023; **14**:1143382.
Fiorentino J, Armaos A, Colantoni A *et al.* Prediction of protein–RNA interactions from single-cell transcriptomic data. *Nucleic Acids Res* 2024;**52**:gkae076.
Freeman LC. A set of measures of centrality based on betweenness. *Sociometry* 1977;**40**:35–41.
Freeman LC. Centrality in social networks conceptual clarification. *Social Networks* 1978;**1**:215–39.
Gasch AP, Yu FB, Hose J *et al.* Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. *PLoS Biol* 2017;**15**:e2004050.
Ghoshal G, Barabasi A-L. Ranking stability and super-stable nodes in complex networks. *Nat Commun* 2011;**2**:394.
Guo Y, Amir A. Exploring the effect of network topology, mRNA and protein dynamics on gene regulatory network stability. *Nat Commun* 2021;**12**:130–10.
Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using networkx. In Varoquaux G, Vaught T, Millman J (eds.), *Proceedings of the 7th Python in Science Conference*, p. 11–15. CA USA: Pasadena, 2008.
Han X, Chen H, Huang D *et al.* Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell RNA-sequencing. *Genome Biol* 2018;**19**:47.
Hillerton T, Seçilmiş D, Nelander S *et al.* Fast and accurate gene regulatory network inference by normalized least squares regression. *Bioinformatics* 2022;**38**:2263–8.
Jaccard P. The distribution of the flora in the alpine zone. *New Phytol* 1912;**11**:37–50.
Kamapantula BK, Abdelzaher A, Ghosh P *et al.* Leveraging the robustness of genetic networks: a case study on bio-inspired wireless sensor network topologies. *J Ambient Intell Human Comput* 2014; **5**:323–39.
Kang Y, Thieffry D, Cantini L. Evaluating the reproducibility of single-cell gene regulatory network inference algorithms. *Front Genet* 2021;**12**:617282.
Kiani NA, Zenil H, Olczak J *et al.* Evaluating network inference methods in preserving the topology and complexity of reconstructed genetic networks. *Semin Cell Dev Biol* 2016;**51**:44–52.
Kim J, Jakobsen ST, Natarajan KN *et al.* Tenet: gene network reconstruction using transfer entropy reveals key regulatory factors from single cell transcriptomic data. *Nucleic Acids Res* 2021;**49**:e1.
Kim S. ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Commun Stat Appl Methods* 2015;**22**:665–74.
Koschützki D, Schreiber F. Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regul Syst Bio* 2008;**2**:193–201.
Koutrouli M, Karatzas E, Paez-Espino D *et al.* A guide to conquer the biological network era using graph theory. *Front Bioeng Biotechnol* 2020;**8**:34.
Latora V, Marchiori M. Efficient behavior of small-world networks. *Phys Rev Lett* 2001;**87**:198701.
Latora V, Marchiori M. A measure of centrality based on network efficiency. *New J Phys* 2007;**9**:188.
Liu W, Sun X, Yang L *et al.* NSCGRN: a network structure control method for gene regulatory network inference. *Brief Bioinform* 2022;**23**:bbac156.
Lopes FM, Martins DC, Jr, Barrera J *et al.* A feature selection technique for inference of graphs from their known topological properties: revealing scale-free gene regulatory networks. *Inform Sci* 2014; **272**:1–15.
MacNeil L, Walhout A. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res* 2011;**21**:645–57.
McCalla SG, Fotuhi Siahpirani A, Li J *et al.* Identifying strengths and weaknesses of methods for computational network inference from single-cell RNA-seq data. *G3: Genes, Genomes, Genetics* 2023; **13**:jkad004.
Moerman T, Aibar Santos S, Bravo González-Blas C *et al.* Grnboost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* 2019;**35**:2159–61.
Newman ME. Mixing patterns in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2003;**67**:026126.
Noman N, Monjo T, Moscato P *et al.* Evolving robust gene regulatory networks. *PLoS One* 2015;**10**:e0116258.
Ouma WZ, Pogacar K, Grotewold E. Topological and statistical analyses of gene regulatory networks reveal unifying yet quantitatively different emergent properties. *PLoS Comput Biol* 2018;**14**:e1006098.
Padi M, Quackenbush J. Integrating transcriptional and protein interaction networks to prioritize condition-specific master regulators. *BMC Syst Biol* 2015;**9**:80–17.
Page L, Brin S, Motwani R *et al.* The pagerank citation ranking: bringing order to the web. Technical report, Stanford InfoLab, 1999.
Papili Gao N, Ud-Dean SM, Gandrillon O *et al.* Sincerities: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. *Bioinformatics* 2018;**34**:258–66.
Park PJ. Chip–seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 2009;**10**:669–80.
Pechenick DA, Payne JL, Moore JH. The influence of assortativity on the robustness of signal-integration logic in gene regulatory networks. *J Theor Biol* 2012;**296**:21–32.
Pratapa A, Jalihal AP, Law JN *et al.* Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods* 2020;**17**:147–54.
Roy S, Shah VK, Das SK. Design of robust and efficient topology using enhanced gene regulatory networks. *IEEE Trans Mol Biol Multi-Scale Commun* 2018;**4**:73–87.
Saint-Antoine MM, Singh A. Network inference in systems biology: recent developments, challenges, and applications. *Curr Opin Biotechnol* 2020;**63**:89–98.
Schaffter T, Marbach D, Floreano D. Genenetweaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics* 2011;**27**:2263–70.
Shalek AK, Satija R, Shuga J *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* 2014;**510**:363–9.
Sporns O, Honey CJ, Kötter R. Identification and classification of hubs in brain networks. *PLoS One* 2007;**2**:e1049.

Strang A, Haynes O, Cahill ND *et al.* Generalized relationships between characteristic path length, efficiency, clustering coefficients, and density. *Soc Netw Anal Min* 2018;**8**:1–6.
Stumpf MP. Inferring better gene regulation networks from single-cell data. *Curr Opin Syst Biol* 2021;**27**:100342.
Sun L, Jiang L, Li M *et al.* Statistical analysis of gene regulatory networks reconstructed from gene expression data of lung cancer. *Physica A: Stat Mechan Appl* 2006;**370**:663–71.
Svensson V, da Veiga Beltrame E, Pachter L. A curated database reveals trends in single-cell transcriptomics. *Database* 2020; **2020**:baaa073.
Szedlak A, Smith N, Liu L *et al.* Evolutionary and topological properties of genes and community structures in human gene regulatory networks. *PLoS Comput Biol* 2016;**12**:e1005009.
Theodoris CV, Xiao L, Chopra A *et al.* Transfer learning enables predictions in network biology. *Nature* 2023;**618**:616–24.
Tran KA, Pietrzak SJ, Zaidan NZ *et al.* Defining reprogramming checkpoints from single-cell analyses of induced pluripotency. *Cell Rep* 2019;**27**:1726–41.e5.
Valente TW, Foreman RK. Integration and radiality: measuring the extent of an individual's connectedness and reachability in a network. *Soc Netw* 1998;**20**:89–105.
Van Noort V, Snel B, Huynen MA. The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. *EMBO Rep* 2004;**5**:280–4.
Villaverde AF, Ross J, Morán F *et al.* Mider: network inference with mutual information distance and entropy reduction. *PLoS One* 2014;**9**:e96732.
Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature* 1998;**393**:440–2.
Winterbach W, Wang H, Reinders M *et al.* Metabolic network destruction: relating topology to robustness. *Nano Commun Netw* 2011; **2**:88–98.
Wolf IR, Simões RP, Valente GT. Three topological features of regulatory networks control life-essential and specialized subsystems. *Sci Rep* 2021;**11**:24209.
Yuan AE, Shou W. Data-driven causal analysis of observational biological time series. *Elife* 2022;**11**:e72518.
Zhang X, Zhao J, Hao J-K *et al.* Conditional mutual inclusive information enables accurate quantification of associations in gene regulatory networks. *Nucleic Acids Res* 2015;**43**:e31.
Zinati Y, Takiddeen A, Emad A. Groundgan: Grn-guided simulation of single-cell RNA-seq data using causal generative adversarial networks. *bioRxiv* 2023.

Bioinformatics, 2024, 40, 1–15
https://doi.org/10.1093/bioinformatics/btae267
Original Paper

A.2. Leveraging prior knowledge to infer Gene Regulatory Networks from single-cell RNA-sequencing data

Stock, M., Losert, C., Zambon, M., Popp, N., Lubatti, G., Hörmanseder, E., Heinig, M. and Scialdone, A.

This is a published version of the article in Molecular Systems Biology following peer review. The article is open-access thus the published version is inserted here.

Review Article

TRANSPARENT PROCESS OPEN ACCESS CHECK FOR UPDATES

molecular systems biology

Leveraging prior knowledge to infer gene regulatory networks from single-cell RNA-sequencing data

Marco Stock[1,2,3,4], Corinna Losert[2,5,7], Matteo Zambon[1,2,3,7], Niclas Popp[1,2,3], Gabriele Lubatti[1,2,3], Eva Hörmanseder[1], Matthias Heinig[2,5,6] & Antonio Scialdone[1,2,3]✉

Abstract

Many studies have used single-cell RNA sequencing (scRNA-seq) to infer gene regulatory networks (GRNs), which are crucial for understanding complex cellular regulation. However, the inherent noise and sparsity of scRNA-seq data present significant challenges to accurate GRN inference. This review explores one promising approach that has been proposed to address these challenges: integrating prior knowledge into the inference process to enhance the reliability of the inferred networks. We categorize common types of prior knowledge, such as experimental data and curated databases, and discuss methods for representing priors, particularly through graph structures. In addition, we classify recent GRN inference algorithms based on their ability to incorporate these priors and assess their performance in different contexts. Finally, we propose a standardized benchmarking framework to evaluate algorithms more fairly, ensuring biologically meaningful comparisons. This review provides guidance for researchers selecting GRN inference methods and offers insights for developers looking to improve current approaches and foster innovation in the field.

Keywords Gene Regulatory Network Inference; Prior Knowledge; Single-cell Transcriptomics; Single-cell Multiomics; Graph Learning
Subject Categories Chromatin, Transcription & Genomics; Computational Biology
https://doi.org/10.1038/s44320-025-00088-3
Received 10 October 2024; Revised 29 January 2025; Accepted 30 January 2025
Published online: 12 February 2025

Introduction

The increasing availability of large-scale single-cell RNA-sequencing (scRNA-seq) datasets (Svensson et al, 2020) has driven the development of numerous computational methods to infer gene regulatory networks (GRNs). scRNA-seq data offer unique insights into cell-to-cell variability that are obscured in bulk RNA-seq datasets, making them particularly valuable for understanding the intricate regulatory mechanisms underlying biological processes. The identification of GRNs provides critical insights into the causal relationships that govern gene interactions, revealing which genes are pivotal in specific contexts. However, constructing GRNs from scRNA-seq data is challenging for several reasons, including the presence of biological (e.g., cell-cycle-related (Buettner et al, 2015; Liu et al, 2021)) and technical confounding factors (e.g., high sparsity and high level of intrinsic noise and dropouts (Qiu, 2020)), as well as GRN properties, like the presence of feedback loops (Atanackovic et al, 2023). Many algorithms have been proposed to tackle this complex problem (Hawe et al, 2019; Pratapa et al, 2020). Nevertheless, the consensus of recently published benchmarking studies is that the performance of GRN inference algorithms is still limited (Kang et al, 2021; McCalla et al, 2023; Pratapa et al, 2020; Stock et al, 2024). (Pratapa et al, 2020) find highly variable and overall poor performance of the algorithms across several datasets. (Kang et al, 2021) highlight poor reproducibility of inferred GRNs, even from independent datasets collected under the same biological condition. (McCalla et al, 2023) demonstrate that advanced approaches cannot consistently outperform simple linear correlation in their analysis. (Stock et al, 2024) show that the available algorithms include topological biases in their inferred GRNs.

A promising strategy to improve this is the incorporation of prior knowledge into the inference process (Liu, 2018; McCalla et al, 2023). For instance, prior knowledge could involve known regulatory interactions between specific gene pairs, or the use of multi-omic data, which can provide information on chromatin accessibility, maps of DNA physical contacts, etc. Integrating such information can enhance GRN inference, e.g., by constraining the solution space or by providing labeled examples that the algorithms can use for learning.

Numerous sources of prior knowledge are available, as well as various approaches for incorporating them into the inference process. This diversity has led to the rapid development of a wide range of algorithms, each offering distinct combinations of strategies. As a result, it has become increasingly challenging to keep track of the available algorithms, the types of prior knowledge they utilize, their computational methodologies, and their performance across different datasets. For users, this complexity poses challenges in selecting the most suitable algorithm for their needs, while developers face obstacles in identifying weaknesses and areas for improvement. These issues are further compounded by the lack

[1]Helmholtz Center Munich Institute of Epigenetics und Stem Cells, Munich, Germany. [2]Helmholtz Center Munich Institute of Computational Biology, Munich, Germany. [3]Helmholtz Center Munich Institute of Functional Epigenetics, Munich, Germany. [4]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany. [5]Department of Computer Science, TUM School of Computation, Information and Technology, Technical University of Munich, Garching, Germany. [6]German Centre for Cardiovascular Research (DZHK), Munich Heart Association, Partner Site Munich, Berlin, Germany. [7]These authors contributed equally: Corinna Losert, Matteo Zambon. ✉E-mail: antonio.scialdone@helmholtz-munich.de

of an objective and unbiased benchmarking framework for evaluating GRN algorithms using prior knowledge, an inherently difficult task given the algorithms' variety and the heterogeneity of their strategies.

Recent reviews on GRN inference have predominantly focused on the use of single-cell multiomics data (e.g., Badia-i Mompel et al, 2023), a pivotal approach in this field. However, this represents just one among many methodologies currently under development. A broader and more versatile area of GRN inference involves incorporating prior knowledge, which spans diverse strategies, including data-type-independent methods such as the use of topological priors and generalized graph priors (Nair, 2017; Stock et al, 2024). While many algorithms have been designed to leverage these approaches, a comprehensive review that systematically presents and compares these strategies is still lacking.

In this review, we provide an overview of prior-knowledge-informed GRN inference from scRNA-seq data from three perspectives. First, we categorize the available sources of prior knowledge and the strategies that algorithms use to incorporate them into the inference process. Secondly, we classify the algorithms based on their computational approaches, their ability to incorporate flexible graph priors and their capability to handle the specific features of scRNA-seq data. Based on this classification, we propose a framework for benchmarking algorithms using graph representations of prior knowledge, which offer flexibility in utilizing diverse sources of priors. This framework aims to disentangle the contributions of the prior knowledge and the algorithm itself to the overall performance, addressing the confounding factors present in current benchmarking studies. By emphasizing these three aspects of GRN inference with prior knowledge, this review is designed to address the needs of two primary audiences: researchers applying GRN inference algorithms and those interested in developing novel computational methodologies. For the former, we provide insights into the types of prior knowledge that current algorithms can handle, and which algorithms are most suitable based on their data. For the latter, we offer a comprehensive overview of the available computational strategies, emphasizing critical aspects and opportunities for combining approaches to enhance performance. By balancing practical application insights with detailed descriptions of algorithmic strategies, the review aims to serve as a valuable resource for both groups, reflecting its dual focus on application and innovation.

Gene regulatory networks and their inference

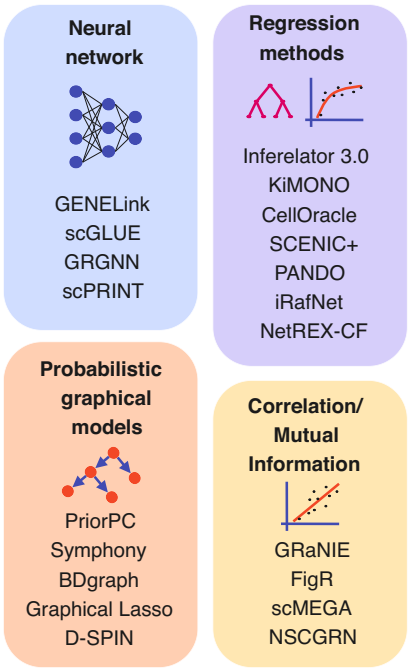
Gene regulatory networks

The deeper understanding of mechanistic processes in a cell is still a focus of ongoing research in systems biology. These processes are governed by a large network of interactions between the DNA, proteins, different RNAs and small molecules. They control cell proliferation and apoptosis, but can also help explain the stability of cell fate and other processes in each cell (Mukhopadhyay et al, 2020). The entirety of these regulatory interactions can be represented as a network (de Jong, 2002), which can be thought of as the fundamental organizational scheme for a cellular system. Due to the high complexity of networks that describe all DNA, RNA, protein and small molecule interactions, the regulatory network is often decomposed into several different layers. The most fundamental one, based on gene expression, describes the binding of regulating proteins, called transcription factors (TFs), to the promoter of their regulated target genes (TGs) on the DNA. The resulting networks are also called transcription regulation networks, or TF-TG-Networks (Vázquez et al, 2004). This approach simplifies the perspective of understanding gene regulation by only focusing on transcriptional regulation, which describes the levels of transcription of a target gene depending on the regulation by a transcription factor, compared to a broader definition of a network that also includes other regulatory effects, such as epigenetic modifications (Erbe et al, 2023; Tejada-Lapuerta et al, 2023). Such a mechanistic model of regulatory interactions can, for instance, help predict how cells react to perturbations, such as drug treatments, which could be valuable for drug target identification and understanding therapeutic effects (Park et al, 2022). In addition, by identifying key regulators for specific cell types, it becomes possible to reprogram cells into different types, a prospect of high interest in regenerative medicine, where cell reprogramming could aid in tissue repair or the treatment of degenerative diseases (Aguirre et al, 2023). Other downstream analyses include identifying functional modules from the GRN, comparative analysis between multiple GRNs, and conducting in-silico perturbation experiments aimed at simulating the effects of potential interventions on cellular behavior (Badia-i Mompel et al, 2023). A TF-TG GRN is usually not strictly bipartite since transcription factors can also regulate other transcription factors. Therefore, we consider the resulting GRN as a homogeneous graph of interactions between different genes.

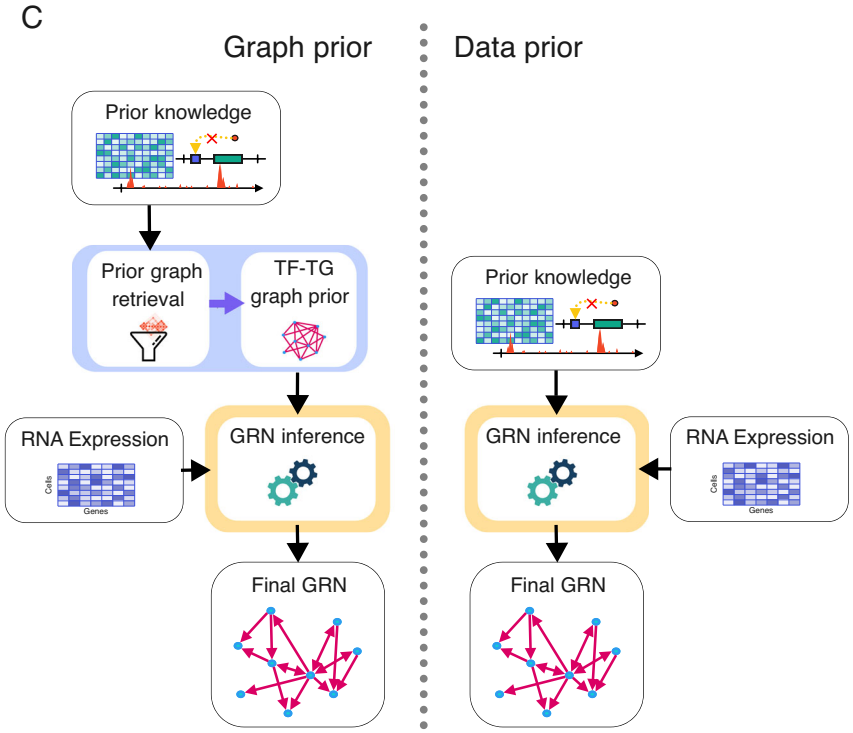
Recently, extended versions of GRNs that include regulatory elements (RE) as nodes have been constructed (Bravo González-Blas et al, 2023; Kamal et al, 2023). REs are regions on the DNA where TFs bind and regulate the expression of the target gene, such as promoters, enhancers, silencers, insulators and locus control regions. Since most approaches mainly focus on the incorporation of enhancers into the GRN, this kind of GRN is often called enhancer GRN (eGRN). It introduces multiple types of edges to the network, namely those going from transcription factors to the regulatory elements (TF-RE) and edges going from regulatory elements to the target genes (RE-TG). To compare them to other regular GRNs, a TF-TG network can be extracted from the eGRN in a post-processing step. In both TF-TG and eGRNs, the graph representation offers the advantage of leveraging theoretical frameworks and the wide range of analytical methods developed within the broader field of (biological) network science (Hetzel et al, 2021).

There are further characteristics that distinguish different types of GRNs. One example is resolution, where GRNs are either inferred from multiple cell types or individuals (Cha and Lee, 2020; Chasman and Roy, 2017), or at the highest resolution specifying personalized and cell type-specific GRNs (Bafna et al, 2023). Another feature is the size of the gene sets of interest, where some GRNs span as little as 5 genes (Atanackovic et al, 2023), whereas other GRNs include thousands of genes, which can be selected, for example, based on their highly variable expression levels (McCalla et al, 2023). In general, the larger the set of genes and the resolution of the GRN, the harder the inference problem becomes. To properly leverage the input scRNA-seq datasets, these should show

A



C



B

	Algorithms
scRNA-Seq	GENELink, scGLUE, Symphony, D-SPIN, CellOracle, FigR, Inferelator 3.0, SCENIC+, Pando, scMEGA, NetREX-CF, scPRINT
Bulk RNA-Seq	GRaNIE, KiMONO, Graphical Lasso
Microarray	GRGNN, PriorPC, BDgraph, iRafNet, NSCGRN

D

		Weighted edges	Binary edges
Graph prior	Homogeneous	NetREX-CF, D-SPIN, iRafNet, PriorPC	CellOracle, Inferelator 3.0, GRGNN, GENELink, Symphony, scPRINT
	Heterogeneous	scGLUE, BDgraph, Graphical Lasso	KiMONO
Data prior	FigR, GRaNIE, SCENIC+, Pando, scMEGA, NSCGRN		

Figure 1. Overview of available GRN inference algorithms leveraging different types of transcriptomic data and representations of prior knowledge.

(**A**) Classification of GRN inference methods according to the type of algorithm they employ. (**B**) GRN inference algorithms that leverage prior knowledge were designed to work on different types of gene expression data (first column). Methods tailored for scRNA-seq data are expected to perform better on this data type because they address its unique challenges, such as sparsity and high intrinsic noise. (**C**) Schematic representation of the main steps of algorithms leveraging "graph-based" (left) or "data-based" (right) priors. With graph-based priors, the algorithms distill all prior knowledge in a TF-TG prior graph, which is then combined with gene expression data to generate a final GRN. The algorithms that use data-based priors, instead, combine prior knowledge and transcriptomic data directly during the inference to build a GRN, without generating any prior graph. (**D**) Classification of GRN inference methods based on the type of integrated prior.

high variability in the gene set of interest, which might, for example, be caused by different external conditions or perturbations (Aguirre et al, 2024; Saint-Antoine and Singh, 2023). However, some work suggests that the heterogeneity within scRNA-seq data from a single condition provides enough variability to infer meaningful regulatory relationships (Gupta et al, 2022). This review is not restricted to a specific resolution or size but addresses the general task of GRN inference.

Types of inference algorithms

Various computational approaches have been applied to GRN inference, each leveraging different strategies and techniques. Broadly, GRN inference algorithms can be grouped into four categories based on their computational approach: correlation-based methods, regression methods, probabilistic graphical models, and neural network-based approaches (see Fig. 1A). Each of these categories offers unique strengths and limitations, depending on the underlying assumptions and techniques employed.

Correlation- and mutual information-based methods rely on statistical dependencies to infer regulatory relationships between genes. FigR (Kartha et al, 2022) employs the non-parametric Spearman's rank correlation coefficient to detect monotonic relationships between gene expression levels and chromatin accessibility scores. In contrast, GRaNIE (Kamal et al, 2023) and scMEGA (Li et al, 2023b) use the linear Pearson's correlation coefficient for the same purpose. While correlation-based approaches are fast, scalable, and robust (Li et al, 2023b), they have notable limitations, particularly in detecting more complex non-linear relationships. To address this, NSCGRN (Liu et al, 2022) utilizes mutual information, which, however, cannot distinguish between positive and negative relationships. More generally, simple statistical measures struggle to capture complex causal interactions, often producing dense networks with numerous indirect edges that

can obscure true regulatory signals (Hawe et al, 2019). In addition, these methods cannot establish directed edges without prior knowledge of transcription factors.

Probabilistic graphical models (PGMs) use directed or undirected graphs to denote probabilistic relationships between variables. PriorPC (Ghanbari et al, 2015), for example, utilizes Bayesian networks, which encode conditional dependencies between variables as directed acyclic graphs, offering a more refined approach that reduces the prevalence of indirect edges. However, the assumption of acyclicity in Bayesian networks is a major limitation since it prevents the representation of feedback loops, a common feature in GRNs (Osella et al, 2011). Other PGM methods such as BDgraph (Mohammadi and Wit, 2019) or Graphical Lasso (Friedman et al, 2008), circumvent this issue by using undirected graphs. Both are based on Gaussian graphical models, which typically assume that the data follows a Gaussian distribution. This does not hold for single-cell RNA-seq data, which are overdispersed and may be zero-inflated (Chen et al, 2018; Durif et al, 2019). To cope with this, BDgraph uses specific copula approaches to handle non-Gaussian data (Choudhary and Satija, 2022). Graphical Lasso (Friedman et al, 2008) focuses on sparsity, a well-known feature of GRNs, by introducing an L1-regularization penalty that enforces a sparser graph estimation. Importantly, both methods can incorporate priors on the graph structure either globally or for individual edges. Symphony (Burdziak et al, 2019) leverages a Bayesian hierarchical multi-view mixture model that allows the inference of cell type clusters and GRNs simultaneously. D-SPIN (Jiang et al, 2024) leverages spin networks, a type of maximum entropy model.

Regression-based approaches infer statistical relationships between genes using feature selection techniques. These techniques enable the detection of more complex dependencies and accommodate feedback loops and cycles, making them more versatile compared to simple correlation methods. The ability to model nonlinear dependencies makes these methods particularly powerful for GRN reconstruction. Pando (Fleck et al, 2023) showcases a basic linear regression model. iRafNet (Petralia et al, 2015) employs random forest regression, a tree-based machine learning strategy. SCENIC+ (Bravo González-Blas et al, 2023) utilizes a gradient boosting algorithm, another ensemble method using decision trees. KiMONO (Ogris et al, 2021) applies LASSO-penalized regression to enforce sparser predictions. Inferelator 3.0 (Gibbs et al, 2022) tested three different regression approaches and achieved the best results with Adaptive Multiple Sparse Regression, a multitask learning approach that also enforces sparse predictions. CellOracle (Kamimoto et al, 2023) uses ridge regression, which adds an L2-regularization, either as Bayesian ridge regression or as Bagging ridge regression.

Neural network-based approaches, particularly deep learning models, leverage loss optimization via backpropagation to learn complex relationships from large data. GRGNN (Wang et al, 2020) employs Graph Neural Networks in a (semi-) supervised graph classification setting. GENELink (Chen and Liu, 2022) utilizes a Graph Autoencoder architecture in a self-supervised link prediction setting. scGLUE (Cao and Gao, 2022) combines the Graph Autoencoder with multiple variational autoencoders for the different data types, and adds a discriminator to align the different latent spaces. scPRINT (Kalfon et al, 2024) uses a transformer architecture, training a foundation model on millions of single cells, and relies on the learned attention heads to predict the final GRN. One advantage of DL methods is that they do not rely on many assumptions about the underlying graph structure, which allows for improved performance compared to other methods (Shu et al, 2021; Yuan and Bar-Joseph, 2019). However, these approaches typically require significant computational resources, making them less practical for large-scale applications without extensive computational infrastructure.

Each of these categories of algorithms offers distinct advantages and challenges and requires specific strategies to incorporate prior knowledge. The choice of algorithm depends on the specific requirements of the GRN reconstruction task, such as data type, scale, and the desired balance between accuracy and computational efficiency.

All algorithms included in this review use gene expression data (Fig. 1B) combined with prior knowledge to infer GRNs. However, not all of them have been designed specifically for scRNA-seq data. Reconstructing GRNs from scRNA-seq data can potentially uncover new insights by modeling and exploiting the single-cell level heterogeneity. By modeling GRNs on a high-resolution cell type level or even per individual, it is possible to investigate context-specific differences in regulation between cell types or individuals. McCalla et al (McCalla et al, 2023) outline that GRN inference from one scRNA-seq dataset can reach the same performance as the GRN inference from multiple collected bulk RNA-seq datasets. At the same time, this data type poses new challenges, for example, due to its sparsity and noise (Akers and Murali, 2021). As a result, an algorithm designed for bulk sequencing data might need to be revised to work as efficiently on scRNA-seq data. A careful preprocessing of scRNA-seq data is also essential for its use in GRN inference. The raw output of RNA-seq experiments is read sequences that have to be aligned to the respective genome, resulting in a count score per annotated gene. Standard preprocessing steps include quality control and count normalization, followed by feature selection, batch integration and potentially further dimensionality reduction (Heumos et al, 2023). The quality control includes the correction for ambient RNA, the filtering of low-quality libraries, and the detection and removal of doublets. Recommended normalization techniques depend on the downstream task; an overview of different methods is available in a recent comparison of Ahlmann-Eltze and Huber (2023). The feature selection is usually done by subsetting the genes of interest to the highly variable genes. For batch integration, several algorithms including linear-embedding models like Harmony (Korsunsky et al, 2019) and Scanorama (Hie et al, 2019), and deep-learning approaches such as scANVI (Xu et al, 2021), scVI (Lopez et al, 2018), and scGen (Lotfollahi et al, 2019) were recently benchmarked by (Luecken et al, 2022). Finally, dimensionality reduction techniques like Principal Components Analysis, t-SNE (van der Maaten and Hinton, 2008), UMAP (McInnes et al, 2020) and PHATE (Moon et al, 2019) can be applied for visualization purposes. Algorithms presented in this study assume already preprocessed scRNA-seq data as an input since the exact choice of preprocessing depends on the specific dataset and thus cannot be easily automated.

Leveraging prior knowledge for GRN inference

As discussed above, GRN inference typically relies on measured gene expression data obtained from bulk RNA-seq or scRNA-seq experiments. When additional data beyond gene expression is incorporated to support the inference process, the algorithms are regarded as leveraging prior knowledge. Below, we detail the various types of prior knowledge, the sources from which they can be derived, and the algorithms capable of integrating these different types of priors.

Transcription factor databases

One commonly available source of additional information is the list of transcription factors in the model organism. This information can be used to filter out interactions that do not involve at least one TF and to infer the directionality of regulation based on co-expression patterns. Lists of TFs can be obtained from various resources, depending on the organism under study. For example, TF lists can be derived from gene annotation databases such as Ensembl (Harrison et al, 2024) or Uniprot (The UniProt Consortium, 2023), where genes annotated as TFs or associated with the relevant Gene Ontology terms can be identified. Dedicated TF databases, like PlantTFDB (Jin et al, 2017) for plants or AnimalTFDB (Shen et al, 2022) for animals, provide organism-specific collections of TFs. Alternatively, experimental evidence-based databases, such as JASPAR (Rauluseviciute et al, 2024) or the Encyclopedia of DNA elements (The ENCODE project consortium, 2012), often include ChIP-seq data and other experimental evidence of TF binding. These resources not only supply TF lists but also additional regulatory information that can be leveraged by GRN inference algorithms, as described below in further detail. A simple TF list, classifying genes as TFs, is widely used by many state-of-the-art algorithms, such as GENIE3 (Huynh-Thu et al, 2010). This review focuses exclusively on algorithms that incorporate at least one additional source of prior knowledge, leveraging these resources to enhance the inference process.

Experimental sources of prior knowledge

Most of the prior knowledge that can be integrated into the GRN inference process is derived from biological experiments. Here, we outline the different types of experimental prior knowledge, the experimental protocols with which those priors are built, the computational strategies that are applied to incorporate them and their specific characteristics and limitations. In Table 1, we summarize the most common sources of experimental prior knowledge.

Very often, the algorithms leverage multiple prior sources together. The most common combination is to use chromatin accessibility data from (single-cell) ATAC-seq and TF-binding motif enrichment. In this case, the regulatory elements inferred from accessibility peaks are then linked with matching TF motif enrichment. Methods that leverage both accessibility and motif enrichment data are Inferelator 3.0, FigR, SCENIC+, GRaNIE, scMEGA, Pando, Symphony, and CellOracle. Other algorithms, instead, leverage data from knockout experiments and protein–protein interaction data (iRafNet) from binding motifs, knockout experiments and ChIP-seq data (NetREX-CF), or from an arbitrary number of unpaired data modalities, like scGLUE or KiMONO. Hawe et al (2022) construct a prior network from a combination of protein–protein interactions, transcription factor and histone ChIP-seq, eQTL and SNP data to then run BDgraph, Graphical Lasso and iRafNet with the prior network.

As often there is no ground truth available for validating the resulting GRNs, experimental data that can be integrated as prior knowledge is also, at the same time, the best resource for validating inferred links. For example, data from knockout experiments is used for prior construction in some methods (iRafNet, NetREX-CF), whereas it is used for validation of inferred GRN links in others (SCENIC+, GRaNIE).

In this section, we discuss each experimental source by first outlining the primary goal of the experiment, such as identifying regulatory elements or specific types of interactions. This is followed by a brief explanation of the experiment and its resulting data output. We then describe how this data is currently utilized in published GRN inference algorithms. Lastly, we highlight any limitations or provide additional insights where relevant.

Chromatin accessibility data (ATAC-seq)

Chromatin accessibility can provide information on transcription factor binding sites and the regulatory potential of a genetic locus. Therefore, it can help identify regulatory elements (mainly enhancers) as GRN nodes that, in turn, link the transcription factors to their target genes.

Bulk or single-cell ATAC-seq (Buenrostro et al, 2015) is one of the most commonly used experimental techniques to quantify chromatin accessibility, even though other assays like DNase-seq or FAIRE-seq could also be used. In an ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) experiment, nuclei are treated with a transposase enzyme (e.g., hyperactive Tn5 transposase), which inserts sequencing adapters into accessible DNA regions. This is followed by sequencing. The locations with a high amount of sequencing reads (peaks) represent the regions of increased accessibility.

The candidate REs identified from peaks in ATAC-seq data are usually linked to potential target genes by calculating whether the peak overlaps with a pre-specified region around the target gene (see the section "Genomic distance"). If paired data is available, where RNA-seq and ATAC-seq are performed on the same single cells (Chen, 2019; Ma et al, 2020), the peaks might be associated with target genes by computing the correlation between chromatin accessibility at the peak and target gene expression or by more sophisticated methods like tree-based regression or others (e.g., GRaNIE, Pando, SCENIC+, FigR, scMEGA). For the RE-TG links based on the correlation of chromatin accessibility and RNA expression, GRaNIE filters for positive correlation, since negative correlation does not have a biological explanation and thus is considered noise (Kamal et al, 2023). In the same way, TFs also might be linked to the regulatory regions by associating the transcription factor expression to the peak accessibility (e.g., FigR, GRaNIE). Another way to link TFs to REs is to combine the chromatin accessibility data with binding motif data, which will be outlined in more detail in the next section "TF-Motif Enrichment".

As especially single-cell ATAC data is usually very sparse, most GRN inference methods pre-process the data to reduce sparsity (e.g., by aggregating the data at the cell-type level) before

Table 1. Overview of experimental data ("Source") from which prior knowledge ("Prior") useful for GRN inference can be extracted.

Source	Prior knowledge extracted	Input	Validation
Chromatin accessibility (sc/bulk ATAC-seq)	RE RE-TG TF-RE	Inferelator 3.0, scGLUE, CellOracle, Symphony GRaNIE, Pando, SCENIC+, FigR, scMEGA FigR, GRaNIE	
TF-Motif enrichment (DNA sequencing)	TF-RE	Inferelator 3.0, CellOracle, scMEGA, Pando, scMEGA, SCENIC+, GRaNIE, FigR,Symphony, NetREX-CF	
Protein-DNA binding (TF ChIP-sequencing)	TF-RE	scGLUE, NetREX-CF	*SCENIC+,scPRINT,GRaNIE,FigR,GRGNN,iRafNet,GENELink,D-SPIN,CellOracle*
Perturbation/Knockout Experiments (sc/bulk RNA expression)	TF-TG	iRafNet, NetREX-CF, D-SPIN	*SCENIC+,GRaNIE,scPRINT,GENELink*
Protein–protein interaction networks	TF-TG	iRafNet	
Protein-DNA binding (Histone modification ChIP-sequencing)	RE RE-TG TF-RE	GRaNIE	*SCENIC+,Pando*
Enhancer activity essays (STARR-seq)	RE		*SCENIC+*
DNA sequence conservation	RE	Pando	*GRGNN,iRafNet*
Genomic distance	RE-TG	scMEGA, NetREX-CF, scGLUE, Symphony,KiMONO, SCENIC+, CellOracle, Pando, GRaNIE, Inferelator 3.0, FigR	
Chromatin interaction (Hi-C)	RE-TG	GRaNIE, scGLUE	*SCENIC+*
Expression quantitative trait loci (eQTLs)	RE-TG	scGLUE	*KiMONO,GRaNIE*
Intrinsic feature extraction (sc/bulk RNA expression)	TF-TG	GRGNN, iRafNet, NetREX-CF	

The prior knowledge can consist of a list of regulatory elements (RE), links between regulatory elements and target genes (RE-TG), links between transcription factors and regulatory elements (TF-RE), and links between transcription factors and target genes (TF-TG). The table also includes a list of algorithms that leverage each type of prior extracted from a given type of data. These algorithms can then utilize this information either as input ("Input" column) to enhance GRN inference or as a way to evaluate their results ("Validation").

computing the correlation between chromatin accessibility and gene expression. With the recent emergence of paired scATAC- and scRNA-seq datasets, the performance of these methods could be further improved compared to computationally paired GRN reconstruction. SCENIC+, scMEGA and scGLUE already showcase the application on a 10x Genomics multiome dataset, where RNA expression and chromatin accessibility are simultaneously assessed.

While leveraging chromatin accessibility data has been shown to improve GRN inference (Alanis-Lobato et al, 2023; Argelaguet et al, 2022), one major limitation to consider is that chromatin accessibility does not always imply activity. Thus, it can be misleading when used to derive information about TFs, REs, and TGs interactions (Kamal et al, 2023).

TF-Motif Enrichment

The use of transcription factor binding motifs is another potential prior that can be used to build links between TFs and TGs via REs.

The binding motifs of TFs are usually collected in motif-binding databases. Those databases differ by whether they include binding motifs based only on experimental data from ChIP-seq, DNase-seq, or SELEX (e.g., JASPAR (Rauluseviciute et al, 2024), TRANFAC (Matys et al, 2006)) or whether they also include computational predictions of motif-binding sites (e.g., HOCOMOCO (Kulakovskiy et al, 2018)). Usually, they contain position frequency matrices that summarize the occurrence of each nucleotide at each position in a set of experimentally observed TF-DNA interactions. Those position frequency matrices can then be transformed into position weight matrices (PWMs) via probabilistic or energistic models.

To link TFs to TGs via motif databases, in the first step, a set of candidate REs needs to be identified. This can be done, for example, by choosing a fixed region upstream and downstream of the transcription start site of the target genes (see the section "Genomic distance"), as in Inferelator 3.0 or NetREX-CF. Alternatively, REs can be inferred first from other types of data (e.g., ATAC-seq, DNase-seq, ChIP-seq, Hi-C), or from databases, like ENCODE (Fleck et al, 2023; Gibbs et al, 2022; The ENCODE project consortium, 2012). Once the candidate REs are identified, a motif enrichment analysis is performed to infer potential links between them and TFs. In many GRN inference algorithms, this is done by using the R package motifmatchr (Schep, 2024) in combination

with chromVar (Schep et al, 2017) to calculate the enrichment scores of motifs in accessible regions (e.g., Pando, scMEGA, FigR). The Inferelator 3.0 method (Gibbs et al, 2022) was tested with different motif databases (CisBP, JASPAR, TRANSFAC) and it was shown that the employed motif library significantly affects the network output, but none of the databases was shown to outperform the others. In the SCENIC+ (Bravo González-Blas et al, 2023) method, this ambiguity is addressed by building a consensus position weight matrix from more than 40,000 PWMs from 29 collections of binding motifs to generate the most extensive TF binding motif database to date containing about 1.5k human, 1.3k mouse, and 467 drosophila TFs. The authors showed that keeping the diversity of different motifs for a TF results in a higher accuracy than having only one archetype motif per TF. The evaluation of three different motif enrichment methods further showed that cisTarget and DEM outperform Homer.

Leveraging TF binding motifs knowledge for refining TF-TG links can help distinguish direct from indirect targets and define the directionality of gene-gene interactions for GRN inference. However, TF motifs can also be very similar across TFs, and in general, motif-derived prior knowledge GRNs are noisy and influenced by choice of the motif library (Gibbs et al, 2022).

Protein-DNA binding assays (TF ChIP-seq)

Similarly to ATAC-seq data, ChIP-seq data can reveal potential interactions between TFs and REs by identifying potential TF binding sites.

With ChIP-seq (Chromatin Immunoprecipitation followed by sequencing) experiments, it is possible to identify DNA regions bound by specific proteins, such as transcription factors or histones. In these experiments, chromatin is cross-linked in the cell to stabilize the interactions between the TF and the DNA. Then the DNA is fragmented and a specific antibody that binds to the TF is added. Subsequently, the antibody-TF-DNA complexes are immunoprecipitated, the DNA is purified and the bound DNA fragments are sequenced. By analyzing the sequencing data, regions in the DNA where the TF was binding can be identified (ChIP-seq peaks). From this, TF-RE links can be inferred. Databases like ENCODE (The ENCODE project consortium, 2012), Remap (Hammal et al, 2022) or Unibind (Puig et al, 2021) collect and process information from ChIP-seq experiments to provide transcription factor binding sites predictions.

ChIP-seq data can be used as prior in different ways. NetREX-CF (Wang et al, 2022) generates TF-TG priors by combining TF-RE connections from TF ChIP-seq with RE-TG connections from a genomic distance prior (see the section "Genomic distance"), by assuming that a RE-TG link occurs whenever the RE falls within a gene body or up to 1 kb upstream of the gene. scGlUE (Cao and Gao, 2022) leverages ChIP-seq data from ENCODE in combination with ATAC-seq data to establish a ranking of the most likely target genes for each TF by overlapping ChIP-seq peaks with ATAC-peaks. Methods like GRaNIE (Kamal et al, 2023) and PriorPC (Ghanbari et al, 2015) outline the possibility of using ChIP-seq data priors as an input to the algorithm to identify regulatory regions, but do not showcase it in their applications.

TF ChIP-seq data was also used to build curated ground truth GRNs, as in the DREAM5 challenge yeast dataset (Marbach et al, 2012), which is used by GRGNN (Wang et al, 2020) and iRafNet (Petralia et al, 2015), and the curated ground truths included in the benchmarking analysis BEELINE (Pratapa et al, 2020), which are used, for example, by GENELink (Chen and Liu, 2022). Such curated ground truth networks are split by GRGNN and GENELink into a training part that serves as a prior and a validation part that is used to validate the predictions. Similarly, scPRINT (Kalfon et al, 2024) uses an intersection of perturbation- and ChIP-seq-based ground truth network from McCalla et al. (McCalla et al, 2023) for validation. Also in FigR, GRaNIE, CellOracle, D-SPIN and SCENIC +, ChIP-seq data is used for validating the results of the GRN inference.

The use of ChIP-seq to identify TF binding sites is limited by the lack of high-quality antibodies for some TFs, and the need for large amounts of homogeneous cells (e.g., cell lines, bulk tissues). Experimentally mapping TF binding sites on tissues with a high diversity like, for example, the brain, remains challenging (Bravo González-Blas et al, 2023).

Knockout/perturbation experiments

Knockout and, more in general, perturbation experiments enable the discovery of TF-TG interactions by screening for genes that are affected following the perturbation of specific transcription factors (Petralia et al, 2015).

One example of such experiments is CRISPR-based Perturb-seq assays (Replogle et al, 2022), as used by D-SPIN (Jiang et al, 2024). In these experiments, a library of guide RNAs is designed to target specific genes (in this case TFs) to knockout (CRISPR/Cas9), knockdown (CRISPRi) or activate (CRISPRa) the transcription of the gene. These guide RNAs are then inserted into the single cells, which subsequently leads to a cellular response to the perturbation. Subsequently, scRNA-seq is utilized to measure the impact of the perturbation on gene expression. By performing differential expression analysis between perturbed and unperturbed cells, the TGs influenced by the perturbation of the TF can be identified. Notably, perturbations are not restricted to genetic mutations; they can also involve drug-based interventions that disrupt diverse cellular mechanisms (Peidli et al, 2024). Simulations have demonstrated that perturbation data enhance the ability to distinguish direct gene-gene interactions more effectively. In addition, these perturbations introduce extra variability into expression data, which can be harnessed to improve the accuracy of GRN inference (Aguirre et al, 2024; Saint-Antoine and Singh, 2023). A large amount of knockdown experiment data is, for example, provided by ENCODE (The ENCODE project consortium, 2012).

Several algorithms utilize perturbation data to construct priors that are incorporated into GRN inference. iRafNet (Petralia et al, 2015) integrates the information of knockdown experiments provided within the DREAM5 (Marbach et al, 2012) challenge as prior knowledge by calculating a weight for each transcription factor to target interaction based on the p-value resulting from a t-test run on the expression levels of a given gene before and after the TF knockout. Similarly, NetREX-CF (Wang et al, 2022) also infers prior information on network edges by considering as potential target genes only those with a statistically significant difference in expression level between the control and the perturbed condition.

Other methods use the results of perturbation experiments to validate the inferred interactions (GRaNIE, SCENIC+). SCENIC+ calculates the gene set enrichment of the inferred targets for each

TF in relation to the differentially expressed genes observed in perturbation experiments of the same TF, using perturbation data from the ENCODE (The ENCODE project consortium, 2012) database. As mentioned already in the TF ChIP-seq section, scPRINT (Kalfon et al, 2024) validates the predicted GRN against the intersection of ChIP-seq- and perturbation-based ground truth provided by (McCalla et al, 2023). By leveraging the curated ground truths by BEELINE (Pratapa et al, 2020), GENELink (Chen and Liu, 2022) also makes use of loss-of-function experiments.

While perturbation experiments provide valuable insights into gene regulation, the availability of such data remains limited due to the high cost, time investment, and cell-type specificity of these experiments.

Protein–protein interaction networks

Protein–protein interaction (PPI) networks represent another type of experimentally derived prior, with many potential TF-TG interactions curated in large databases. These databases typically report physical interactions between protein pairs and sometimes functional interactions as well. A notable example is the Bioplex interactome (Huttlin et al, 2021), which was generated using Affinity-Purification Mass Spectrometry on human cell proteins. Another widely used example is HuRI (Luck et al, 2020), a reference interactome map comprising approximately 53,000 human binary protein interactions, created through systematic yeast two-hybrid assays.

iRafNet (Petralia et al, 2015) incorporates TF-TG interactions included in PPI networks as priors, applying a diffusion kernel transformation to pre-process the PPI data. While PriorPC (Ghanbari et al, 2015) also acknowledges the potential of PPIs as priors, it does not showcase any application. (Hawe et al, 2022) uses PPIs from BioGrid (Oughtred et al, 2019) to construct gene-gene edges that are then used as an input to the BDgraph (Mohammadi and Wit, 2019) and Graphical Lasso (Friedman et al, 2008) algorithms.

Protein-DNA binding assays (histone modification ChIP-seq)

Histone modifications such as H3K27ac are well-established markers of active regulatory elements (Creyghton et al, 2010). Consequently, assessing the presence of this modification via ChIP-seq serves as a valuable prior for selecting potential REs. In this case, in the ChIP-seq experiment, the fragmented chromatin is incubated with an antibody specific to H3K27ac (see ChIP-seq section "Protein-DNA binding assays (TF ChIP-seq)") to identify the regions with high levels of this histone modification.

For example, the authors of GRaNIE (Kamal et al, 2023) suggest using this kind of data as an alternative to ATAC-seq. By correlating the H3K27ac mark with RNA expression data, GRaNIE infers TF-RE and RE-TG interactions. Similarly, SCENIC+ (Bravo González-Blas et al, 2023) employs H3K27ac ChIP-seq for validating identified active regulatory regions. The authors of Pando (Fleck et al, 2023) also used a related method (Cut&Tag; (Kaya-Okur et al, 2019)) to validate their inferred regulatory regions by assessing the levels of H3K27ac modifications.

Enhancer activity assays

Another method to identify REs like active enhancers is massively parallel reporter assays (MPRAs). MPRAs are a powerful tool used to evaluate how DNA sequences, particularly regulatory elements like enhancers and promoters, influence gene expression. In an MPRA experiment, a library of DNA sequences is designed to include candidate REs, mutated variants, and randomized controls, each tagged with a unique barcode for identification. These sequences are cloned into reporter constructs that contain a minimal promoter and a reporter gene, and then introduced into cells. Inside the cells, the regulatory sequences drive the expression of the reporter gene. RNA sequencing is subsequently performed to quantify the barcode abundance, providing a direct link between the activity of each sequence and its influence on gene expression. The STARR-seq protocol, a specific type of MPRA (Arnold et al, 2013) designed for enhancer identification, utilizes high-throughput sequencing to quantify the activity of candidate enhancers. SCENIC+ (Bravo González-Blas et al, 2023) uses STARR-seq data to validate the regulatory regions identified by the algorithm.

Conservation of DNA sequences

After an initial list of candidate REs is identified, for instance through chromatin accessibility data, further filtering can be performed based on the evolutionary conservation of DNA sequences. This approach, as used by Pando (Fleck et al, 2023), involves overlapping the candidate REs with a set of highly conserved DNA regions across 30 mammalian species. The resulting list of candidate REs is considered more reliable due to their conservation over evolution.

Similarly, the curated yeast network used for the DREAM5 challenge (Marbach et al, 2012) applies evolutionary conservation as a key criterion to refine the list of TF binding sites derived from ChIP-seq data. This curated network serves as a benchmark for validating the iRafNet algorithm (Petralia et al, 2015) and is also utilized by GRGNN (Wang et al, 2020), both as a prior and for validation purposes.

Genomic distance

Genomic distance is a widely used and readily available metric for constructing priors on the relationships between candidate REs and TGs. REs, often identified from ATAC-seq data, are typically associated with regions around the transcription start site (TSS) of target genes. Depending on the method, the overlap window for linking REs to TGs ranges from 1 kb upstream of the TSS (Wang et al, 2022) to 250 kb (Kamal et al, 2023), covering potential cis-regulatory elements. Inferelator defines the regulatory region as 200 bp upstream and 50 bp downstream of the TSS, while Symphony (Burdziak et al, 2019) simply assigns each ATAC-seq peak to the nearest gene. Several methods that link regulatory regions to target genes include scGLUE, GRaNIE, SCENIC+, FigR, Pando, scMEGA, and CellOracle. NetREX-CF (Wang et al, 2022) associates candidate transcription factor binding sites and TF ChIP peaks with target genes by considering the gene body plus 1 kb upstream of the TSS. KiMONO (Ogris et al, 2021) uses genomic distance to link SNPs to methylation peaks and methylation peaks to genes. In addition, scGLUE (Cao and Gao, 2022) integrates genomic distance with Hi-C data, as discussed in the following section.

Chromatin interactions (Hi-C)

Another source of experimental data that can be used to derive potential links between REs and TGs is data from Hi-C experiments.

This chromosome conformation capture technology measures the frequency at which two DNA fragments physically interact in the 3D space and can be valuable for detecting long-range gene regulation.

The scGLUE algorithm (Cao and Gao, 2022) integrates data from promoter capture Hi-C (pcHi-C), along with other priors, measured in 17 human primary blood cell types to calculate a weight defining pcHi-C-supported interactions between REs (peaks) and TGs. In this application, pcHi-C support for a candidate RE-TG pair was determined based on three criteria: (1) the gene promoter must be located within 1 kb of a bait fragment; (2) the peaks must be located within 1 kb of another-end fragment; and (3) significant contact between the bait and other-end fragment must be observed. In addition to scGLUE (Cao and Gao, 2022), the integration of chromatin conformation data for prior construction is also supported by GRaNIE (Kamal et al, 2023), which uses this data to define RE-TG pairs for further testing, although no specific application is provided in the paper. SCENIC+ (Bravo González-Blas et al, 2023) employs this data to validate inferred RE-TG links.

Expression quantitative trait loci (eQTLs)

eQTL data is a valuable resource for identifying potential links between REs and TGs. eQTL data is typically obtained from experiments that measure both gene expression (e.g., via RNA-seq) and genomic variation. These studies assess the impact of single nucleotide polymorphisms (SNPs) on gene expression, often employing statistical models such as linear regression, where gene expression serves as the dependent variable and SNP genotype as the independent variable. When an SNP shows a statistically significant association with expression variation, it is classified as an eQTL, establishing a link between specific REs and their TGs. This approach helps identify genetic variants that influence gene expression through regulatory mechanisms. One resource that collects eQTL data across multiple human tissues is the Genotype-Tissue Expression (GTEx) project (The GTEx Consortium, 2020). With the increasing availability of single-cell data, recent efforts have focused on mapping cell-type-specific eQTLs, which could serve as valuable prior knowledge (Perez et al, 2022; van der Wijst et al, 2018; Yazar et al, 2022). The sc-eQTLGen Consortium (van der Wijst et al, 2020) was established to perform large-scale, systematic eQTL mapping across a wide range of single-cell datasets. In addition to mapping potential RE-TG links, a specific type of eQTL mapping, known as co-expression QTL mapping, can reveal potential TF-TG relationships by linking upstream TF regulators to downstream TGs (Li et al, 2023a).

The links between REs and TGs identified by eQTL mapping can be integrated as prior for RE-TG interactions. For instance, scGLUE (Cao and Gao, 2022) leverages eQTL data to identify peak-target gene interactions by verifying whether a peak derived from ATAC-seq data overlaps with an eQTL locus, and that locus is associated with the expression of the target gene. The eQTL data utilized by scGLUE is from the GTEx project (The GTEx Consortium, 2020). Other algorithms such as KiMONO (Ogris et al, 2021) and GRaNIE (Kamal et al, 2023) use eQTL data to validate predictions.

Intrinsic feature extraction (RNA-seq)

Besides using additional data in combination with gene expression data, which we consider leveraging prior knowledge in this review, there is also the possibility to enrich the inference by using algorithms to extract additional information from the same expression data. This process can also be seen as a kind of feature crafting from the expression data. An example of this approach is applying existing non-prior-based GRN inference methods, such as mutual information, Pearson correlation, Spearman Rank, or GENIE3 (Huynh-Thu et al, 2010), to gene expression data. The resulting graphs are then utilized as prior knowledge for GRN inference on either the same or a different set of gene expression data. This strategy is exemplified by the GRGNN algorithm (Wang et al, 2020), which constructs a 'noisy starting skeleton' using mutual information or correlation metrics to represent the relationships between gene expression profiles. NetREX-CF (Wang et al, 2022) constructs a co-expression network in order to extend the prior network to span more of the transcription factors. Another example is the use of multiple RNA-seq datasets as a time series, or the inference of pseudo-time (Matsumoto et al, 2017; Qiu et al, 2020; Singh et al, 2024) or RNA velocity (Singh et al, 2024) from the expression data, which can then be used as a prior. For example, iRafNet (Petralia et al, 2015) integrates time-series data and links two genes, i and k, in the prior graph if the past values of gene k are predictive of future values of gene i.

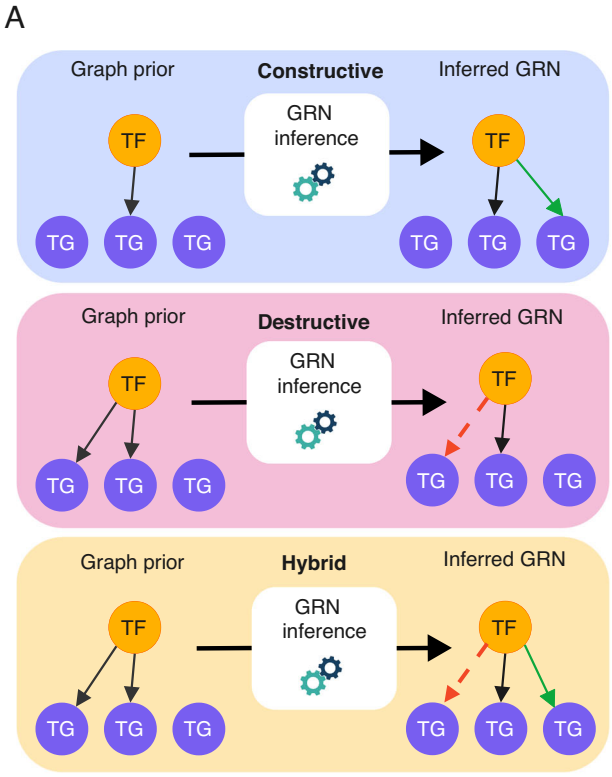
Further experimental priors

In addition to the most commonly used prior sources as outlined in Table 1, KiMONO and scGLUE are designed to handle further input data in the form of a feature matrix. KiMONO (Ogris et al, 2021) mentions protein expression data, which can be added to the prior graph by establishing links to genes and other proteins via protein annotation data. KiMONO further mentions methylation, SNP and mutation data to extend the heterogeneous prior graph by leveraging the genomic distance as a prior to link these additional features to genes. scGLUE (Cao and Gao, 2022) similarly leverages snmC-seq DNA-methylation data. Hawe et al. (2022) extend the genomic entities to CpG sites and construct CpG-to-gene priors using histone mark combinations (measured by ChIP-seq) and genomic distance.

Topological prior

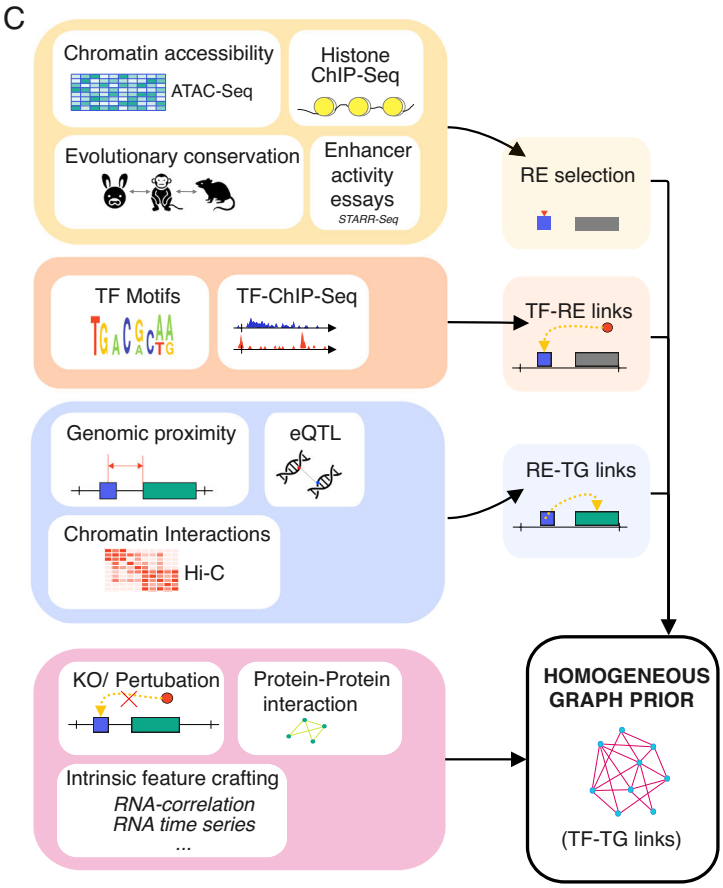
GRNs are characterized by specific structural properties, which can be quantified by topological measurements (Zhivkoplias et al, 2022). Recent studies showed that current GRN inference algorithms produce networks that lack the typical topological properties of GRNs, which can be important in predicting various biologically relevant properties (Stock et al, 2024), such as robustness to perturbations, etc. A common characteristic of GRNs is their scale-free node-degree distribution (Zhivkoplias et al, 2022). This implies that the node degree follows a power-law distribution, where a small number of nodes exhibit exceptionally high connectivity, serving as hubs with numerous interactions. Such hub genes can represent master regulators of a biological process.

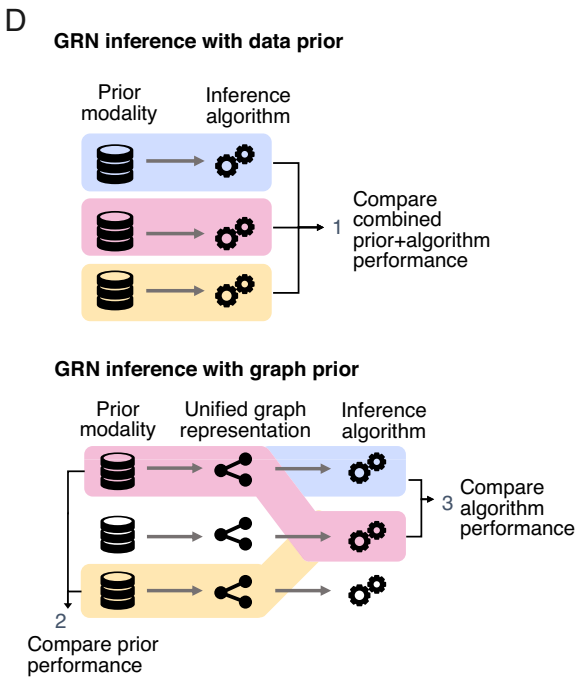
To enhance accuracy, some algorithms leverage known GRN properties to establish topological priors and introduce constraints on graph topology. For example, constraints on the node-degree distribution are employed in certain methods (Nair, 2017). The BDgraph algorithm (Mohammadi and Wit, 2019) takes a related approach by using priors such as the expected sparsity coefficient, the G-Wishart distribution for the null adjacency matrix, and the



B

		Operation mode	Hard positives	Hard negatives	Soft prior
homog. graph	GENELink	Constructive	✓	✓	X
	Inferelator 3.0	Constructive	✓	✓	X
	scPRINT	Constructive	✓	✓	X
	GRGNN	Constructive	✓	X	✓
	CellOracle	Destructive	X	✓	X
	Symphony	Destructive	X	✓	X
	PriorPC	Destructive	X	✓	✓
	iRafNet	Hybrid	X	X	✓
	NetREX-CF	Hybrid	X	X	✓
	D-SPIN	Hybrid	X	X	✓
heterog. graph	KiMONO	Destructive	X	✓	X
	scGLUE	Hybrid	X	X	✓
	BDgraph	Hybrid	X	X	✓
	Graphical Lasso	Hybrid	X	X	✓





Bernoulli distribution for link inclusion. Another notable method is Graphical Lasso (Friedman et al, 2008), which promotes sparsity by applying a lasso penalty to the inverse covariance matrix. In addition, NSCGRN (Liu et al, 2022) incorporates commonly recurring GRN motifs as structural priors, enforcing local patterns like cascades and feedforward loops in the inferred networks.

In their preprint (Aguirre et al, 2024) recently summarized additional common attributes of GRNs, such as modular structure (gene modules), bidirectional and overall sparsity, as well as asymmetric degree distributions (in-compared to out-degree) that could be leveraged as further structural priors for directed GRNs.

Figure 2. Algorithms leveraging homogeneous graph priors enable disentangling the contributions of priors and inference methods in benchmarking studies.

(A) Classification of GRN inference methods that leverage a homogeneous graph prior knowledge in terms of their operation mode. Constructive algorithms (top panel) add edges to the prior network, destructive algorithms (middle panel) remove edges from the prior network, and hybrid algorithms (bottom panel) can add and remove edges simultaneously. **(B)** This panel displays the operation mode of each algorithm, with an indication of whether they can incorporate hard positives, hard negatives, and soft prior. **(C)** Schematic illustration of the different approaches found among the reviewed methods to construct a homogeneous graph prior before the inference process that incorporates the RNA expression data. Some approaches directly infer TF-TG link priors (bottom panel). Other approaches build an enhancer GRN including regulatory elements first and then extracting the TF-TG graph from it. In this case, some priors are leveraged for selecting candidate REs ("RE selection"), for the definition of TF-RE links and for the definition of RE-TG links. **(D)** Benchmarking of GRN inference with prior knowledge. The upper panel shows a standard benchmarking setup, where the combination of a specific prior modality and algorithm are evaluated against each other. In this scenario, the effect of the prior and algorithm cannot be distinguished. The lower panel shows the two additional options when benchmarking methods with a unified graph prior input. This enables a comparative evaluation of the contribution of priors and algorithms separately.

Direct data prior incorporation and graph prior extraction

In GRN inference, the way prior knowledge is incorporated can vary depending on the algorithm but generally falls into two main approaches. Some algorithms use prior knowledge to construct a graph, which serves as a "prior graph" during the inference process. In contrast, other methods integrate prior knowledge directly during the inference without building any prior graph first. We categorize the first approach as "graph-based" priors and the second as "data-based" priors (see Fig. 1C).

Algorithms employing "data-based" priors frequently combine transcriptomic and epigenomic data, especially in single-cell contexts (e.g., FigR, GRaNIE, SCENIC+, Pando, scMEGA). These typically establish correlations between chromatin accessibility at regulatory regions, such as promoters and enhancers, and gene expression. For instance, FigR, GRaNIE, SCENIC+, Pando, and scMEGA all construct enhancer-based GRNs, where TF-RE and RE-TG links are inferred first. Then, a TF-to-TG network is derived through additional post-processing steps. Other algorithms in this category incorporate different types of prior knowledge, such as NSCGRN (Liu et al, 2022), which uses topological priors on GRNs (see section above).

On the other hand, algorithms that utilize "graph-based" priors follow a two-step process. First, a prior graph is constructed, either using additional data sources (e.g., Inferelator 3.0, iRafNet, CellOracle, scGLUE, NetREX-CF, KiMONO, Symphony, BDgraph, Graphical Lasso, D-SPIN) or by downsampling curated ground truth GRNs (e.g., GENELink, PriorPC, GRGNN, scPRINT). In the second step, this prior graph serves as input for inferring the final GRN.

A key advantage of this approach is its flexibility. The second step can accommodate various types of priors because the prior graph is presented in an abstract, unified format. This allows researchers to construct the prior graph from diverse experimental data sources, unlike algorithms that rely on data-based priors, which require specific experimental inputs. In addition, the two-step process enables benchmarking studies to independently evaluate the contributions of performance gains due to the prior from the gains due to the actual inference algorithm.

The following sections provide an overview of the various subtypes of "graph-based" priors and the algorithms designed to operate with this type of input.

GRN inference with graph priors

Graph-based priors are commonly represented using adjacency matrices B, where each entry $b_{i,j}$ denotes the weight of the edge between nodes i and j. In certain algorithms (e.g., CellOracle, GRGNN, GENELink, Inferelator 3.0, Symphony, scPRINT, KiMONO), these weights are binary, representing merely the presence or absence of a link. In other cases, the weights encode a confidence score, quantifying the uncertainty of the link, often due to inconsistencies across different sources of prior knowledge (e.g., PriorPC, iRafNet, NetREX-CF, scGLUE, BDgraph, Graphical Lasso, D-SPIN).

Graph priors can be categorized as either "homogeneous" or "heterogeneous", depending on their composition. Homogeneous graph priors consist solely of TF-TG interactions, as utilized in algorithms such as Inferelator 3.0, PriorPC, iRafNet, GENELink, CellOracle, GRGNN, Symphony, NetREX-CF, D-SPIN and scPRINT. In contrast, heterogeneous graph priors incorporate interactions with additional elements, such as REs in scGLUE or proteins in KIMONO. Hawe et al (2022) applies BDgraph, Graphical Lasso, and iRafNet to a heterogeneous graph prior. However, since iRafNet was originally presented to operate with homogeneous networks, we classify it under the homogeneous prior category. Figure 1D provides a detailed overview of the types of graph priors employed by various algorithms.

During the GRN inference process, graph prior edges are handled according to the confidence level attributed to them. If edges are considered certain, they are retained in the final GRN and referred to as "hard positives." Algorithms that accept graph priors with "hard positives" work in a "constructive" manner, adding missing edges to the prior graph to generate the final GRN. In contrast, some algorithms operate in a "destructive" fashion, generating the final GRN by removing edges from the graph prior. Here, the prior is treated as a list of potential interactions to be filtered, with the absence of an edge serving as evidence of no interaction (i.e., the prior contains "hard negatives"). A third approach is a "hybrid" method that allows both the addition and removal of edges from the graph prior, treating it as a "soft prior". Figure 2A provides a graphical summary of these three modes of operation.

As noted earlier, algorithms using graph-based priors offer a modular and flexible framework, allowing for customized graph priors as input. However, for effective results, it is critical to align the choice of algorithm with the nature of the graph prior (i.e., hard positives, hard negatives, or soft prior). Figure 2B summarizes the operational modes of the algorithms reviewed here. For instance, GENELink and GRGNN subsample the ground truth graph to obtain hard positive and hard negative edges (in the case of GENELink). In scPRINT, a curated set of hard positives and hard negatives is utilized to train a regression model, enabling the identification of the most effective attention heads in a transformer

model for predicting the rest of the GRN. In addition, GRGNN utilizes a soft prior on the remaining edges. Inferelator 3.0 removes certain genes from the prior network to assess performance on these genes using the ground truth later. Algorithms that operate in a "destructive" manner by removing edges from the graph prior include CellOracle, Symphony, KiMONO, and PriorPC, with PriorPC being the only one of these able to leverage a soft prior for potential interactions. Finally, algorithms capable of adding or removing edges based on a soft prior include scGLUE, iRafNet, NetREX-CF, BDgraph, D-SPIN and Graphical Lasso.

Figure 2C provides an overview of how graph priors are constructed using the types of data discussed in the section "Experimental sources of prior knowledge". Protein–protein interactions, perturbation data, DNA binding motifs, and the intrinsic feature extraction from RNA-seq are direct sources for TF-TG link priors. Alternatively, an enhancer GRN can serve as an intermediate step, consisting of TF-RE and RE-TG links, which can then be collapsed into a homogeneous TF-TG prior graph, as done in CellOracle (Kamimoto et al, 2023). Selecting functionally relevant regulatory elements involves data such as chromatin accessibility, TF ChIP-seq, evolutionary conservation across species, and enhancer activity assays. TF-RE links are typically derived from either TF binding motif enrichment or TF ChIP-seq data, while RE-TG links are estimated using genomic proximity, chromatin interaction data like Hi-C, or eQTL data.

In addition to being constructed, graph priors can also be retrieved from databases. These databases (reviewed in (Baltoumas et al, 2021)) combine inferred prior connections from various sources, including manually curated literature or predicted TF-TG links. Other databases store information critical for building priors, such as sets of known regulatory elements, like the ENCODE database (The ENCODE project consortium, 2012). GRaNIE (Kamal et al, 2023) is another example, offering a precompiled resource of potential TF binding sites for humans and mice, instead of computing motif enrichment scores directly from selected peaks. The choice of database often depends on the organism, cell type, and specific information required. For example, in iRafNet (Petralia et al, 2015), protein–protein interactions are sourced from BioGrid (Oughtred et al, 2019), DIP (Xenarios et al, 2000), and MINT (Zanzoni et al, 2002) to construct a GRN prior for *S. cerevisiae*. In KiMONO (Ogris et al, 2021), BioGrid is used to create gene-gene links, while Inferelator 3.0 (Gibbs et al, 2022) builds a GRN prior for *S. cerevisiae* based on the YEASTRACT network (Monteiro et al, 2020).

For certain well-studied models and cell types, databases of TF-TG interactions are considered "ground truth" networks due to their comprehensiveness. These "silver" or "gold" standard networks are frequently used to benchmark GRN inference methods. Algorithms that do not utilize prior knowledge compare their predicted GRN against these ground truth networks. For algorithms that do incorporate prior knowledge, part of the ground truth is typically subsampled as input, and the remaining hold-out interactions are used to evaluate the algorithm's performance (e.g., GRGNN, PriorPC, GENELink, scPRINT). One prominent example of a ground truth network is the one provided by the DREAM5 challenge (Marbach et al, 2012), which uses RegulonDB (Salgado et al, 2024) to curate a ground truth for *E. coli* (used by GRGNN and PriorPC). For *S. cerevisiae*, ChIP-seq data combined with binding motif conservation is used to construct a ground truth (used by GRGNN and iRafNet). Other ground truth networks, such as those on mouse and human embryonic stem cells, are curated in the BEELINE benchmarking study (Pratapa et al, 2020), and used by GENELink (Chen and Liu, 2022) and (McCalla et al, 2023), and by scPRINT (Kalfon et al, 2024), which also relies on Omnipath (Türei et al, 2016) and ENCODE (The ENCODE project consortium, 2012) for additional validation. While gold standard networks are widely used, they tend to have low complexity, cover a limited range of interactions, and are often unavailable for higher eukaryotes. Consequently, benchmarking against these networks has inherent limitations (Kamal et al, 2023). For less-studied organisms, databases from closely related species may be used as a substitute, although this approach can introduce additional uncertainty (Nair, 2017).

Benchmarking GRN inference algorithms that incorporate prior knowledge

The diversity of GRN inference algorithms and the various ways they incorporate prior knowledge provide researchers with many options. However, this variety can complicate the selection process for users and developers alike. As the field continues to grow, there is an increasing need for robust and transparent benchmarking studies that can guide the selection of methods. These studies should focus on comparing algorithms fairly, particularly those that leverage prior knowledge, as the quality and type of prior knowledge used can significantly influence the algorithm's performance. Several existing benchmarking efforts have compared GRN inference algorithms based on their ability to infer TF-TG links (Marbach et al, 2012; McCalla et al, 2023; Pratapa et al, 2020) or predict GRN topology (Stock et al, 2024). However, these studies often overlook the critical role of prior knowledge in influencing outcomes. Thus, future benchmarking studies should account for both the computational strategies and the nature of the prior knowledge used. Disentangling these factors will be essential for identifying algorithmic weaknesses and advancing the field.

To achieve a more objective comparison, we recommend standardizing the representation of prior knowledge across algorithms. For instance, the use of a homogeneous graph prior could allow for a fair evaluation of both algorithms and the quality of their prior knowledge sources (see Fig. 2D). Such an approach would enable not only the standard comparison of algorithm-prior combinations (Fig. 2D, top panel) but also two additional evaluations: a ranking of the value of different prior types and a comparable assessment of algorithmic performance using the same prior as input (Fig. 2D, bottom panel). Furthermore, this framework could enable the integration of multiple graph priors into a single input, allowing researchers to assess whether the combined use of different priors leads to improved GRN inference accuracy. For such a benchmarking study, the operation mode of the benchmark algorithms has to be taken into account, since destructive algorithms cannot be directly benchmarked against constructive algorithms. Sets of algorithms suitable for benchmarking, which share the same type of RNA expression input (e.g., scRNA-seq) and compatible operational modes (e.g., constructive or hybrid), can be identified from Table 2.

Table 2. Summary overview of the presented methods that leverage prior knowledge as a graph representation.

Overview of methods including a graph prior					
Algorithm	**Type**	**Graph**	**Mode**	**Exp. Input**	**Code**
GENELink (Chen and Liu, 2022)	Neural network (graph autoencoder)	Homog. (binary)	Constr.	scRNA-Seq, TF ChIP-Seq	Python
Inferelator 3.0 (Gibbs et al, 2022)	Regression (BBSR, StARS-LASSO, Multi-task Learning)	Homog. (binary)	Constr.	scRNA-Seq, ATAC-Seq, Motifs, Gen. distance	Python
scPRINT(preprint: (Kalfon et al, 2024))	Neural network (Transformer)	Homog. (binary)	Constr.	scRNA-Seq, Arbitr. graph	Python
GRGNN (Wang et al, 2020)	Neural Network (Graph convolution)	Homog. (binary)	Constr.	Microarray, Arbitr. graph	Python
CellOracle (Kamimoto et al, 2023)	Regression (Bayesian Ridge or Bagging Ridge)	Homog. (binary)	Destr.	scRNA-Seq, ATAC-Seq, Motifs, Gen. distance	Python
Symphony (preprint: (Burdziak et al, 2019))	PGM (Bayesian hierarchical model)	Homog. (binary)	Destr.	scRNA-Seq, ATAC-Seq, Motif, Gen. distance	Python (Jupyter)
PriorPC (Ghanbari et al, 2015)	PGM (Bayesian Network)	Homog. (weights)	Destr.	Microarray, Arbitr. graph	No code
iRafNet (Petralia et al, 2015)	Regression (Random Forest)	Homog. (weights)	Hybrid	Microarray, Knockouts, PPI networks	R
NetREX-CF (Wang et al, 2022)	Regression (Collaborative filtering, Network Component Analysis)	Homog. (weights)	Hybrid	scRNA-Seq, Motifs, TF ChIP-Seq, Knockouts, Gen. distance	Python (Jupyter)
D-SPIN (preprint: (Jiang et al, 2024))	PGM (Max Entropy)	Homog. (weights)	Hybrid	scRNA-Seq, Knockouts, Arbitr. graph	Python +Matlab
KiMONO (Ogris et al, 2021)	Regression (Sparse Lasso)	Heterog. (binary)	Destr.	bulk RNA-Seq, Gen. distance, Protein, Methylation, SNPs, Mutation, ...	R
scGLUE (Cao and Gao, 2022)	Neural network (Graph autoencoder)	Heterog. (weights)	Hybrid	scRNA-Seq, ATAC-Seq, TF ChIP-Seq, Gen. distance, Hi-C, eQTLs, Methylation, ...	Python
BDgraph (Mohammadi and Wit, 2019)	PGM (Gaussian Graphical Models)	Heterog. (weights)	Hybrid	Microarray, Topology, Arbitr. graph	R
Graphical Lasso (Friedman et al, 2008)	PGM (Regularized Graphical Models)	Heterog. (weights)	Hybrid	bulk RNA-Seq, Topology, Arbitr. graph	R

For each method, we also indicate the type of graph priors it utilizes (homogeneous or heterogeneous) and its operation mode (constructive, destructive, or hybrid).

Table 3. Summary overview of the presented methods that leverage a data prior without an intermediate graph representation of the prior

Overview of methods including a data prior			
Algorithm	Type	Exp. Input	Code
GRaNIE (Kamal et al, 2023)	Correlation	bulk RNA-Seq, ATAC-Seq, Motifs, Histone ChIP-Seq, Gen. distance, Hi-C	R
FigR (Kartha et al, 2022)	Correlation	scRNA-Seq, ATAC-Seq, Motifs, Gen. distance	R
SCENIC+ (Bravo González-Blas et al, 2023)	Regression (Gradient boosting)	scRNA-Seq, ATAC-Seq, Motifs, Gen. distance	Python
Pando (Fleck et al, 2023)	Regression (linear)	scRNA-Seq, ATAC-Seq, Motifs, Sequence conservation., Gen. distance	R
scMEGA (Li et al, 2023b)	Correlation	scRNA-Seq, ATAC-Seq, Gen. distance	R
NSCGRN (Liu et al, 2022)	Mutual Information	Microarray, Topology	No code

Conclusion and perspectives

In this review, we provided a comprehensive overview of the current strategies employed by GRN inference algorithms that integrate scRNA-seq data with prior knowledge. We discussed the various types of prior knowledge that can be utilized, the diverse ways these priors are incorporated into algorithms, and how these approaches affect the performance of GRN inference. The accompanying tables summarize key information about the reviewed algorithms, with those incorporating graph representations of prior knowledge listed in Table 2, and those relying on other types of priors in Table 3. A detailed description of recent representative inference algorithms that leverage graph priors is in the Appendix Texts S1–S4.

By examining how different algorithms utilize diverse sources of prior knowledge and proposing a unified representation of these priors as graphs, we aim to encourage the integration of multiple resources to improve the accuracy of inferred GRNs. While scRNA-seq data serves as a rich and valuable source of biological insight, it has inherent limitations that can impact GRN inference. At the same time, the availability and quality of complementary data sources, such as curated databases and other omics data, are continually improving. Leveraging these diverse datasets alongside scRNA-seq creates a powerful opportunity to enhance GRN inference, paving the way for deeper insights into gene regulatory mechanisms, the identification of therapeutic targets, and other transformative discoveries.

One underexplored but promising aspect is the potential of graph priors to provide more detailed GRN models. Current GRN inference methods have progressed from undirected to directed networks and have begun incorporating edge annotations representing activating and inhibiting regulatory relationships. However, further refinement is possible. For instance, self-loops, which are currently underrepresented in GRN models, could offer additional mechanistic insights. Similarly, current models often reduce regulatory interactions to simple OR-gate representations, neglecting cases where two TFs jointly regulate a target gene, which would be better represented by an AND-gate. These complexities, which remain unaccounted for in both TF-TG networks and enhancer-based GRNs, could be more effectively modeled using hypergraphs. Prior knowledge, for example from perturbation experiments, could help distinguish between co-regulation of a target by two TFs and independent regulation of the target by each TF. Advances in machine learning methods for hypergraph-based inference, combined with comprehensive perturbation datasets-such as combinatorial knockout assays-will be essential to achieve this level of detail.

Building on our analysis and classification of GRN inference algorithms, we proposed a benchmarking strategy specifically designed for algorithms that incorporate prior knowledge (see the section "Benchmarking GRN inference algorithms that incorporate prior knowledge"). This approach aims to disentangle the individual contributions of the type of prior knowledge and the algorithm itself to overall performance. Such a benchmarking study is currently lacking but is needed to address critical gaps in the field. Another essential element of future benchmarking strategies is the emphasis on openness, reproducibility, and transparency, which are vital for driving progress in GRN inference research. Initiatives like the Open Problems framework (Luecken et al, 2024) embody these principles, offering a strong foundation and setting an important precedent for future efforts. By adhering to these values, benchmarking studies can deliver robust, comparable, and broadly applicable evaluations of GRN inference methods, ultimately advancing the field.

Looking forward, we anticipate that future GRN inference methods will draw upon and combine ideas from the approaches we reviewed, addressing their current limitations while leveraging increasingly standardized benchmarking frameworks. As high-quality single-cell datasets become more available, especially with the rise of pre-trained foundation models (Cui et al, 2024; Theodoris et al, 2023), the inclusion of RNA-independent priors will further enhance our understanding of GRNs.

Expanded view data, supplementary information, appendices are available for this paper at https://doi.org/10.1038/s44320-025-00088-3.

Peer review information

A peer review file is available at https://doi.org/10.1038/s44320-025-00088-3

References

Aguirre M, Escobar M, Forero Amézquita S, Cubillos D, Rincón C, Vanegas P (2023) Application of the Yamanaka transcription factors Oct4, Sox2, Klf4 and c-Myc from the laboratory to the clinic. Genes 14(9):1697

Aguirre M, Spence J, Sella G, Pritchard J (2024) Gene regulatory network structure informs the distribution of perturbation effects. Preprint at https://doi.org/10.1101/2024.07.04.602130

Ahlmann-Eltze C, Huber W (2023) Comparison of transformations for single-cell RNA-seq data. Nat Methods 20:665–672

Akers K, Murali T (2021) Gene regulatory network inference in single-cell biology. Curr Opin Syst Biol 26:87–97

Alanis-Lobato G, Bartlett T, Huang Q, Simon C, McCarthy A, Elder K (2023) Mica: a multi-omics method to predict gene regulatory networks in early human embryos. Life Sci Alliance 7(1):e202302415

Argelaguet R, Lohoff T, Li J, Nakhuda A, Drage D, Krueger F et al (2022) Decoding gene regulation in the mouse embryo using single-cell multi-omics. Preprint at https://doi.org/10.1101/2022.06.15.496239

Arnold C, Gerlach D, Stelzer C, Boryń M, Rath M, Stark A (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 339(6123):1074–1077

Atanackovic L, Tong A, Wang B, Lee L, Bengio Y, Hartford J (2023) Dyngfn: towards Bayesian inference of gene regulatory networks with GFlowNets. Preprint at https://arxiv.org/abs/2302.04178

Badia-i Mompel P, Wessels L, Müller-Dott S, Trimbour R, Ramirez Flores R, Argelaguet R (2023) Gene regulatory network inference in the era of single-cell multi-omics. Nat Rev Genet 24(11):739–754

Bafna M, Li H, Zhang X (2023) Clarify: cell-cell interaction and gene regulatory network refinement from spatially resolved transcriptomics. Bioinformatics 39(Suppl 1):i484–i493

Baltoumas F, Zafeiropoulou S, Karatzas E, Koutrouli M, Thanati F, Voutsadaki K (2021) Biomolecule and bioentity interaction databases in systems biology: a comprehensive review. Biomolecules 11(8):1245

Bravo González-Blas C, De Winter S, Hulselmans G, Hecker N, Matetovici I, Christiaens V (2023) SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Nat Methods 20:1355–1367

Buenrostro J, Wu B, Litzenburger U, Ruff D, Gonzales M, Snyder M (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523(7561):486–490

Buettner F, Natarajan K, Casale F, Proserpio V, Scialdone A, Theis F (2015) Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat Biotechnol 33(2):155–160

Burdziak C, Azizi E, Prabhakaran S, Pe'er D (2019) A nonparametric multi-view model for estimating cell type-specific gene regulatory networks. Preprint at https://arxiv.org/abs/1902.08138

Cao Z, Gao G (2022) Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. Nat Biotechnol 40(10):1458–1466

Cha J, Lee I (2020) Single-cell network biology for resolving cellular heterogeneity in human diseases. Exp Mol Med 52:1798–1808

Chasman D, Roy S (2017) Inference of cell type specific regulatory networks on mammalian lineages. Curr Opin Syst Biol 2:130–139

Chen G, Liu Z (2022) Graph attention network for link prediction of gene regulations from single-cell RNA-sequencing data. Bioinformatics 38(19):4522–4529

Chen W, Li Y, Easton J, Finkelstein D, Wu G, Chen X (2018) Umi-count modeling and differential expression analysis for single-cell RNA sequencing. Genome Biol 19(1):70

Chen S, Lake BB, Zhang K (2019) High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nat Biotechnol 37(12):1452–1457

Choudhary S, Satija R (2022) Comparison and evaluation of statistical error models for scRNA-seq. Genome Biol 23(1):27

Creyghton M, Cheng A, Welstead G, Kooistra T, Carey B, Steine E (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA 107(50):21931–21936

Cui H, Wang C, Maan H, Pang K, Luo F, Duan N (2024) scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nat Methods 21(8):1470–1480

de Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. J Comput Biol 9(1):67–103

Durif G, Modolo L, Mold J, Lambert-Lacroix S, Picard F (2019) Probabilistic count matrix factorization for single cell expression data analysis. Bioinformatics 35:4011–4019

Erbe R, Stein-O'Brien G, Fertig E (2023) A mechanistic simulation of molecular cell states over time. Preprint at https://doi.org/10.1101/2023.02.23.529720

Fleck J, Jansen S, Wollny D, Seimiya M, Zenk F, Santel M (2023) Inferring and perturbing cell fate regulomes in human cerebral organoids. Nature 621:365–372

Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9(3):432–441

Ghanbari M, Lasserre J, Vingron M (2015) Reconstruction of gene networks using prior knowledge. BMC Syst Biol 9:84

Gibbs C, Jackson C, Saldi G, Tjarnberg A, Shah A, Watters A (2022) High performance single-cell gene regulatory network inference at scale: the Inferelator 3.0. Bioinformatics 38(9):2519

Gupta A, Martin-Rufino J, Jones T, Subramanian V, Qiu X, Grody E (2022) Inferring gene regulation from stochastic transcriptional variation across single cells at steady state. Proc Natl Acad Sci USA 119(34):e2207392119

Hammal F, de Langen P, Bergon A, Lopez F, Ballester B (2022) ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. Nucleic Acids Res 50(D1):D316–D325

Harrison P, Amode M, Austine-Orimoloye O, Azov A, Barba M, Barnes I (2024) Ensembl 2024. Nucleic Acids Res 52:D891–D899

Hawe J, Theis F, Heinig M (2019) Inferring interaction networks from multi-omics data. Front Genet 10:535

Hawe J, Saha A, Waldenberger M, Kunze S, Wahl S, Müller-Nurasyid M (2022) Network reconstruction for trans acting genetic loci using multi-omics data and prior information. Genome Med 14(1):125

Hetzel L, Fischer D, Günnemann S, Theis F (2021) Graph representation learning for single-cell biology. Curr Opin Syst Biol 28:100347

Heumos L, Schaar A, Lance C, Litinetskaya A, Drost F, Zappia L (2023) Best practices for single-cell analysis across modalities. Nat Rev Genet 24:550–572

Hie B, Bryson B, Berger B (2019) Efficient integration of heterogeneous single-cell transcriptomes using scanorama. Nat Biotechnol 37:685–691

Huttlin E, Bruckner R, Navarrete-Perea J, Cannon J, Baltier K, Gebreab F (2021) Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell 184(11):3022–3040.e28

Huynh-Thu V, Irrthum A, Wehenkel L, Geurts P (2010) Inferring regulatory networks from expression data using tree-based methods. PLoS ONE 5(9):1–10

Jiang J, Chen S, Tsou T, McGinnis C, Khazaei T, Zhu Q et al (2024) D-spin constructs gene regulatory network models from multiplexed scRNA-seq data revealing organizing principles of cellular perturbation response. Preprint at https://doi.org/10.1101/2023.04.19.537364

Jin J, Tian F, Yang D, Meng Y, Kong L, Luo J (2017) Planttfdb 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 45:D1040–D1045

Kalfon J, Samaran J, Peyré G, Cantini L (2024) scPRINT: pre-training on 50 million cells allows robust gene network predictions. Preprint at https://doi.org/10.1101/2024.07.29.605556

Kamal A, Arnold C, Claringbould A, Moussa R, Servaas N, Kholmatov M (2023) GRaNIE and GRaNPA: inference and evaluation of enhancer-mediated gene regulatory networks. Mol Syst Biol 19(6):e11627

Kamimoto K, Stringa B, Hoffmann C, Jindal K, Solnica-Krezel L, Morris S (2023) Dissecting cell identity via network inference and in silico gene perturbation. Nature 614(7949):742–751

Kang Y, Thieffry D, Cantini L (2021) Evaluating the reproducibility of single-cell gene regulatory network inference algorithms. Front Genet 12:617282

Kartha V, Duarte F, Hu Y, Ma S, Chew J, Lareau C (2022) Functional inference of gene regulation using single-cell multi-omics. Cell Genomics 2(9):100166

Kaya-Okur H, Wu S, Codomo C, Pledger E, Bryson T, Henikoff J (2019) Cut&tag for efficient epigenomic profiling of small samples and single cells. Nat Commun 10(1):1930

Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K (2019) Fast, sensitive and accurate integration of single-cell data with harmony. Nat Methods 16:1289–1296

Kulakovskiy I, Vorontsov I, Yevshin I, Sharipov R, Fedorova A, Rumynskiy E (2018) Hocomoco: towards a complete collection of transcription factor binding models for human and mouse via large-scale chip-seq analysis. Nucleic Acids Res 46:D252–D259

Li S, Schmid K, de Vries D, Korshevniuk M, Losert C, Oelen R (2023a) Identification of genetic variants that impact gene co-expression relationships using large-scale single-cell data. Genome Biol 24:80

Li Z, Nagai J, Kuppe C, Kramann R, Costa I (2023b) scMEGA: single-cell multi-omic enhancer-based gene regulatory network inference. Bioinforma Adv 3(1):vbad003

Liu J, Yang M, Zhao W, Zhou X (2021) Ccpe: cell cycle pseudotime estimation for single cell RNA-seq data. Nucleic Acids Res 50(2):704–716

Liu W, Sun X, Yang L, Li K, Y Y, Fu X (2022) NSCGRN: a network structure control method for gene regulatory network inference. Brief Bioinform 23(5):bbac156

Liu Z (2018) Towards precise reconstruction of gene regulatory networks by data integration. Quant Biol 6(2):113–128

Lopez R, Regier J, Cole M, Jordan M, Yosef N (2018) Deep generative modeling for single-cell transcriptomics. Nat Methods 15:1053–1058

Lotfollahi M, Wolf F, Theis F (2019) scGen predicts single-cell perturbation responses. Nat Methods 16:715–721

Luck K, Kim D, Lambourne L, Spirohn K, Begg B, Bian W (2020) A reference map of the human binary protein interactome. Nature 580(7803):402–408

Luecken M, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller M (2022) Benchmarking atlas-level data integration in single-cell genomics. Nat Methods 19:41–50

Luecken M, Gigante S, Burkhardt D, Cannoodt R, Strobl D, Markov N et al (2024) Defining and benchmarking open problems in single-cell analysis. Preprint at https://doi.org/10.21203/rs.3.rs-4181617/v1.

Ma S, Zhang B, LaFave L, Earl A, Chiang Z, Hu Y (2020) Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell 183(4):1103–1116.e20

Marbach D, Costello J, Küffner R, Vega N, Prill R, Camacho D (2012) Wisdom of crowds for robust gene network inference. Nat Methods 9(8):796–804

Matsumoto H, Kiryu H, Furusawa C, Ko M, Ko S, Gouda N (2017) Scode: an efficient regulatory network inference algorithm from single-cell RNA-seq during differentiation. Bioinformatics 33(15):2314–2321

Matys V, Kel-Margoulis O, Fricke E, Liebich I, Land S, Barre-Dirrie A (2006) Transfac and its module transcompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 34:D108–10

McCalla S, Fotuhi Siahpirani A, Li J, Pyne S, Stone M, Periyasamy V (2023) Identifying strengths and weaknesses of methods for computational network inference from single-cell RNA-seq data. G3 13(3):jkad004

McInnes L, Healy J, Melville J (2020) UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at https://arxiv.org/abs/1802.03426

Mohammadi R, Wit E (2019) Bdgraph: an r package for Bayesian structure learning in graphical models. J Stat Softw 89:1–30

Monteiro P, Oliveira J, Pais P, Antunes M, Palma M, Cavalheiro M (2020) YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts. Nucleic Acids Res 48(D1):D642–D649

Moon K, van Dijk D, Wang Z, Gigante S, Burkhardt D, Chen W (2019) Visualizing structure and transitions in high-dimensional biological data. Nat Biotechnol 37(12):1482–1492

Mukhopadhyay P, Greene R, Pisano M (2020) MicroRNA targeting of the non-canonical planar cell polarity pathway in the developing neural tube. Cell Biochem Funct 38(7):905–920

Nair A (2017) Incorporating and generating prior knowledge to improve gene regulatory network inference. Ph.D. thesis, Monash University

Ogris C, Hu Y, Arloth J, Müller N (2021) Versatile knowledge guided network inference method for prioritizing key regulatory factors in multi-omics data. Sci Rep 11:6806

Osella M, Bosia C, Corá D, Caselle M (2011) The role of incoherent microRNA-mediated feedforward loops in noise buffering. PLoS Comput Biol 7(3):1–16

Oughtred R, Stark C, Breitkreutz B, Rust J, Boucher L, Chang C (2019) The biogrid interaction database: 2019 update. Nucleic Acids Res 47(D1):D529–D541

Park H, Yamaguchi R, Imoto S, Miyano S (2022) Xprediction: explainable EGFR-TKIs response prediction based on drug sensitivity specific gene networks. PLoS ONE 17(5):e0261630

Peidli S, Green T, Shen C, Gross T, Min J, Garda S (2024) scPerturb: harmonized single-cell perturbation data. Nat Methods 21(3):531–540

Perez R, Gordon M, Subramaniam M, Kim M, Hartoularos G, Targ S (2022) Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. Science 376:eabf1970

Petralia F, Wang P, Yang J, Tu Z (2015) Integrative random forest for gene regulatory network inference. Bioinformatics 31(12):i197–i205

Pratapa A, Jalihal A, Law J, Bharadwaj A, Murali T (2020) Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat Methods 17(2):147–154

Puig R, Boddie P, Khan A, Castro-Mondragon J, Mathelier A (2021) UniBind: maps of high-confidence direct TF-DNA interactions across nine species. BMC Genomics 22(1):482

Qiu P (2020) Embracing the dropouts in single-cell RNA-seq analysis. Nat Commun 11:1169

Qiu X, Rahimzamani A, Wang L, Ren B, Mao Q, Durham T (2020) Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. Cell Syst 10(3):265–274.e11

Rauluseviciute I, Riudavets-Puig R, Blanc-Mathieu R, Castro-Mondragon J, Ferenc K, Kumar V (2024) JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. Nucleic Acids Res 52:D174–D182

Replogle J, Saunders R, Pogson A, Hussmann J, Lenail A, Guna A (2022) Mapping information-rich genotype-phenotype landscapes with genome-scale perturb-seq. Cell 185(14):3022–3040.e28

Saint-Antoine M, Singh A (2023) Limits on inferring gene regulatory networks subjected to different noise mechanisms. Preprint at https://doi.org/10.1101/2023.01.23.525259

Salgado H, Gama-Castro S, Lara P, Mejia-Almonte C, Alarcón-Carranza G, López-Almazo AG, Betancourt-Figueroa F, Peña-Loredo P, Alquicira-Hernández S, Ledezma-Tejeida D et al (2024) RegulonDB v12. 0: a comprehensive resource of transcriptional regulation in *E. coli* K-12. Nucleic Acids Res 52(D1):D255–D264

Schep A (2024) motifmatchr: fast motif matching in R. R package version 1.27.0

Schep A, Wu B, Buenrostro J, Greenleaf W (2017) chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods 14(10):975–978

Shen W, Chen S, Gan Z, Zhang Y, Yue T, Chen M (2022) Animaltfdb 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. Nucleic Acids Res 51(D1):D39–D45

Shu H, Zhou J, Lian Q, Li H, Zhao D, Zeng J (2021) Modeling gene regulatory networks using neural network architectures. Nat Comput Sci 1:491–501

Singh R, Wu A, Mudide A, Berger B (2024) Causal gene regulatory analysis with RNA velocity reveals an interplay between slow and fast transcription factors. Cell Syst 15(5):462–474.e5

Stock M, Popp N, Fiorentino J, Scialdone A (2024) Topological benchmarking of algorithms to infer gene regulatory networks from single-cell RNA-seq data. Bioinformatics 40(5):btae267

Svensson V, da Veiga Beltrame E, Pachter L (2020) A curated database reveals trends in single-cell transcriptomics. Database 2020:baaa073

Tejada-Lapuerta A, Bertin P, Bauer S, Aliee H, Bengio Y, Theis F (2023) Causal machine learning for single-cell genomics. Preprint at https://arxiv.org/abs/2310.14935.

The ENCODE project consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489(7414):57–74

The GTEx Consortium (2020) The GTEx consortium atlas of genetic regulatory effects across human tissues. Science 369(6509):1318–1330

The UniProt Consortium (2023) UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res 51:D523–D531

Theodoris C, Xiao L, Chopra A, Chaffin M, Al Sayed Z, Hill M (2023) Transfer learning enables predictions in network biology. Nature 618:616–624

Türei D, Korcsmáros T, Saez-Rodriguez J (2016) OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat Methods 13(12):966–967

van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605

van der Wijst M, Brugge H, de Vries D, Deelen P, Swertz M, Franke L (2018) Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. Nat Genet 50:493–497

van der Wijst M, de Vries D, Groot H, Trynka G, Hon C, Bonder M (2020) The single-cell eqtlgen consortium. eLife 9:e52155

Vázquez A, Dobrin R, Sergi D, Eckmann J, Oltvai Z, Barabási A (2004) The topological relationship between the large-scale attributes and local interaction patterns of complex networks. Proc Natl Acad Sci USA 101(52):17940–17945

Wang J, Ma A, Ma Q, Xu D, Joshi T (2020) Inductive inference of gene regulatory network using supervised and semi-supervised graph neural networks. Comput Struct Biotechnol J 18:3335–3343

Wang Y, Lee H, Fear J, Berger I, Oliver B, Przytycka T (2022) Netrex-cf integrates incomplete transcription factor data with gene expression to reconstruct gene regulatory networks. Commun Biol 5:1282

Xenarios I, Rice D, Salwinski L, Baron M, Marcotte E, Eisenberg D (2000) DIP: the database of interacting proteins. Nucleic Acids Res 28(1):289–291

Xu C, Lopez R, Mehlman E, Regier J, Jordan M, Yosef N (2021) Probabilistic harmonization and annotation of single cell transcriptomics data with deep generative models. Mol Syst Biol 17:e9620

Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon M, Andersen S (2022) Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. Science 376:eabf3041

Yuan Y, Bar-Joseph Z (2019) Deep learning for inferring gene relationships from single-cell expression data. Proc Natl Acad Sci USA 116(52):27151–27158

Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G (2002) MINT: a molecular interaction database. FEBS Lett 513(1):135–140

Zhivkoplias E, Vavulov O, Hillerton T, Sonnhammer E (2022) Generation of realistic gene regulatory networks by enriching for feed-forward loops. Front Genet 13:815692

Acknowledgements

MS and CL were supported by the Helmholtz Association under the joint research school "Munich School for Data Science - MUDS". MS was supported by a Joachim Herz Stiftung Add-on Fellowship for Interdisciplinary Life Science. This work was supported by the Helmholtz Association (AS, EH) and the Deutsche Forschungsgemeinschaft (SC280/2-1 to AS, HO 6864/2-1 and CRC1064 Chromatin Dynamics Project-ID 213249687 to EH). This project has been made possible in part by the Chan Zuckerberg Foundation (2019-202666, 2021-237882 to MH) and was supported by the DZHK partner site project 81Z0600106 (MH).

Author contributions

Marco Stock: Conceptualization; Formal analysis; Investigation; Visualization; Methodology; Writing—original draft; Writing—review and editing. **Corinna Losert**: Formal analysis; Investigation; Methodology; Writing—original draft; Writing—review and editing. **Matteo Zambon**: Formal analysis; Investigation; Methodology; Writing—original draft. **Niclas Popp**: Formal analysis; Investigation; Methodology; Writing—original draft. **Gabriele Lubatti**: Formal analysis; Investigation; Methodology; Writing—original draft. **Eva Hörmanseder**: Supervision; Writing—review and editing. **Matthias Heinig**: Supervision; Writing—review and editing. **Antonio Scialdone**: Conceptualization; Supervision; Funding acquisition; Project administration; Writing—review and editing.

Disclosure and competing interests statement

The authors declare no competing interests.