Title

Gene Regulatory Networks in Coronary Artery Disease

Journal

Current Atherosclerosis Reports, 25(12)

ISSN

1523-3804

Authors

Cheng, Jenny
Cheng, Michael
Lusis, Aldons J
et al.

Publication Date

2023-12-01

DOI

10.1007/s11883-023-01170-7

Peer reviewed

HHS Public Access

Author manuscript

Curr Atheroscler Rep. Author manuscript; available in PMC 2024 October 10.

Published in final edited form as:

Curr Atheroscler Rep. 2023 December ; 25(12): 1013–1023. doi:10.1007/s11883-023-01170-7.

Gene Regulatory Networks in Coronary Artery Disease

Jenny Cheng[1,2], **Michael Cheng**[1,3], **Aldons J. Lusis**[4,5], **Xia Yang**[1,2,3,6]

[1]Department of Integrative Biology and Physiology, University of California, Los Angeles, 610 Charles E. Young Drive East, Los Angeles, CA 90095, USA

[2]Molecular, Cellular and Integrative Physiology Interdepartmental Program, University of California, Los Angeles, 610 Charles E. Young Drive East, Los Angeles, CA 90095, USA

[3]Bioinformatics Interdepartmental Program, University of California, Los Angeles, 610 Charles E. Young Drive East, Los Angeles, CA 90095, USA

[4]Department of Medicine, Division of Cardiology, University of California, Los Angeles, 650 Charles E Young Drive South, Los Angeles, CA 90095, USA

[5]Departments of Human Genetics & Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, 650 Charles E. Young Drive South, Los Angeles, CA 90095, USA

[6]Department of Molecular and Medical Pharmacology, University of California, Los Angeles, 610 Charles E. Young Drive East, Los Angeles, CA 90095, USA

Abstract

Purpose of Review—Coronary artery disease is a complex disorder and the leading cause of mortality worldwide. As technologies for the generation of high-throughput multiomics data have advanced, gene regulatory network modeling has become an increasingly powerful tool in understanding coronary artery disease. This review summarizes recent and novel gene regulatory network tools for bulk tissue and single cell data, existing databases for network construction, and applications of gene regulatory networks in coronary artery disease.

Recent Findings—New gene regulatory network tools can integrate multiomics data to elucidate complex disease mechanisms at unprecedented cellular and spatial resolutions. At the same time, updates to coronary artery disease expression data in existing databases have enabled researchers to build gene regulatory networks to study novel disease mechanisms. Gene regulatory networks have proven extremely useful in understanding CAD heritability beyond what is explained by GWAS loci and in identifying mechanisms and key driver genes underlying disease onset and progression.

Aldons J. Lusis, jlusis@mednet.ucla.edu; Xia Yang, xyang123@ucla.edu.
Jenny Cheng and Michael Cheng share equal contribution.

Author contributions J.C. and M.C. wrote the main manuscript text and prepared figures. A.J.L and X.Y. supervised and revised the manuscript. All authors reviewed the manuscript.

Conflict of Interest The authors have no competing interests to report.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

Summary—Gene regulatory networks can holistically and comprehensively address the complex nature of coronary artery disease. In this review, we discuss key algorithmic approaches to construct gene regulatory networks and highlight state-of-the-art methods that model specific modes of gene regulation. We also explore recent applications of these tools in coronary artery disease patient data repositories to understand disease heritability and shared and distinct disease mechanisms and key driver genes across tissues, between sexes, and between species.

Keywords

Coronary artery disease; Atherosclerosis; Gene regulatory networks; Network modeling; RNA-sequencing; Single-cell RNA-sequencing

Introduction

Coronary artery disease (CAD), characterized by arterial plaque buildup, is the leading cause of mortality worldwide [1]. It is a complex disorder, involving numerous genetic and environmental factors, including hyperlipidemia, lack of exercise, and smoking [2, 3]. Lesion initiation and development are characterized by low-density lipoprotein (LDL) accumulation in the subendothelial space, leading to endothelial cell monolayer activation and monocyte infiltration, smooth muscle cell phenotypic switching, lipid-laden foam cell development, and ultimately the formation of necrotic debris in the intimal arterial layer [4]. Eventually, lesions can grow to hinder blood flow in the vessel lumen and become unstable, increasing the likelihood of plaque rupture and clinical complications, including myocardial infarction and stroke.

The complex nature of CAD onset and progression, involving numerous tissues, cell types, molecules, and pathways, is best addressed by a comprehensive and holistic approach [5, 6]. Network modeling represents such an approach to delineate the complexity, as networks capture relationships and interactions, depicted as network edges or connections, between large numbers of molecules represented as network nodes. Over the past decades, network approaches have been particularly empowered by the advances in technologies that have enabled the generation of high throughput biological data at varying molecular levels, termed multiomics data [7]. Multiomics data encompasses genomics, epigenomics, transcriptomics, proteomics, metabolomics, microbiome, and metagenomics, each representing a fragmented view of the whole system [6, 8, 9]. By modeling omics data at individual layers or by integrating multiple layers of omics data, various types of networks, such as gene regulatory networks (GRNs), coexpression networks, inter- and intracellular gene signaling networks, protein–protein interaction (PPI) networks, microbial community networks, and metabolic networks, can be derived [10]. Depending on the network type, edges may represent directed regulation such as those in GRNs, statistical correlation such as those in coexpression networks, physical binding such as those in PPIs, and metabolic reaction cascades such as those in metabolic networks, whereas network nodes may represent genes, proteins, metabolites, microbial species, or phenotypes [10].

Among the various types of networks, GRNs carry unique mechanistic information because they reflect causal and regulatory relationships between genes. Additionally, the genetic

architecture of CAD may be explained by the omnigenic disease model, which proposes that complex diseases can be explained by both core network genes that have large numbers of connections or edges and large effect sizes, and peripheral genes that have few connections and moderate to small effect sizes in interconnected disease networks [11]. Genome-wide association studies (GWAS) and meta-analyses of CAD have identified over 200 genetic loci contributing to disease, but together these loci account for less than half of CAD heritability, highlighting the urgent need to pinpoint additional genetic factors contributing to disease onset [12, 13]. Therefore, here we focus on GRNs as a powerful, graphical tool to uncover novel biological processes in atherosclerosis and CAD. GRNs provide an approach to understand the underlying interactions and to identify hub genes, termed key drivers, that contribute most importantly to the network.

In this review, we discuss current bulk tissue and single-cell GRN methodologies along with their corresponding tools, advantages, and limitations. We next describe available databases that may be used for network construction. Lastly, we discuss the application of GRNs in CAD studies.

GRN Methodologies and Tools

General GRN Modeling Algorithms

GRNs apply mostly to transcriptomics data with or without additional information (e.g., transcription factor-target relationships) or data (e.g., genetics) to detect regulatory patterns between genes. Varied gene expression and biological function by tissue or cell type warrant the use of corresponding transcriptomic data to construct networks that most accurately reflect context-specific gene-gene regulation. Common network construction methods utilize various modeling algorithms, such as Boolean, ordinary differential equations (ODE), machine learning (ML), and Bayesian inference, to infer gene-gene regulation. Numerous previous reviews have extensively discussed various types of GRN methods[10, 14-19]. Here, we only briefly outline a few key algorithms underlying the different methods, with more detailed descriptions of top-performing GRN methods based on benchmarking studies provided in Table 1 [20, 21, 22•].

Boolean methods discretize gene expression on a learned threshold into an active or inactive state and generate network regulations with logical or Boolean functions [18, 22•, 23]. Logical statements allow Boolean models to infer gene activity based on the combined states of multiple regulators. As a result, these networks can model a variety of regulatory dynamics like negative feedback loops and are helpful for simulating network structures for GRN benchmarking [18, 22•]. However, the thresholding of genes in Boolean networks ignores more complex relationships that rely on the dynamic fluctuations in gene abundance.

ODE-based GRNs use a system of differential equations to quantify gene expression on a continuous scale [18, 22•, 24]. When timepoint or cell pseudotime information is available, these networks can model time-dependent gene expression dynamics [14, 22•, 24]. However, many of these methods also impose simplifying assumptions about the ODEs (e.g., linear equations) since estimating very sophisticated equations can be computationally intractable [14].

ML-based methods implement regression models, often ensemble trees, to predict each given target gene expression from select feature genes and extract those with the strongest predictive power as its candidate regulators [14, 16, 22•]. Ensemble regression trees, such as random forest and gradient boosting, are a common choice for GRN algorithms because they can powerfully predict complex nonlinear relationships in the data [25-31]. However, ML methods are subject to false positive regulations given the high dimensional feature space in transcriptomics data, so these tools are accompanied by various feature selection strategies to mitigate spurious edges [25, 28].

Bayesian networks (BNs) are directed acyclic graphs that represent genes as random variables and edges as conditional probabilities of the state of a target gene given the states of the parent genes [18, 32, 33]. Applying Bayes' theorem, these methods can initialize networks based on prior information such as known gene regulatory principles, and continually update the structure to find the one with the highest probability of modeling the observed expression data. However, this probabilistic framework is computationally intensive and heavily reliant on high quality prior information [10].

These various algorithms underlie most of the GRN methods and can be performed on bulk or single cell omics datasets to construct knowledge-based or global GRNs. However, it is important to note that existing GRN methods for bulk tissue transcriptomics data do not perform well for single cell omics datasets due to differences in data sparsity and distribution, and it is recommended to use methods appropriate for the given dataset [10, 20, 22•].

Knowledge-Based vs Global/Unbiased GRNs

To avoid intractable computational run times and limit spurious interactions, many GRN methods select a subset of relevant genes to build the network. Such GRNs can explain different modes of gene regulation based on the function of the select genes of interest (Fig. 1). Below we introduce two common types of GRNs based on the types of candidate genes or knowledge graph of focus, namely transcription factor (TF)- and ligand-receptor (LR)-based GRNs, as well as unbiased global GRNs which do not rely on known regulatory relations or databases.

TF-Based Networks

TF-based gene regulation provides a strong biological basis for GRN inference since TFs create promoter complexes to initiate target gene transcription [16, 34]. Thus, they serve as candidate regulators for many GRN methods. Incorporating this information in GRNs requires restricting parents of a target gene to TFs. In addition to the transcriptomics input, TF databases, such as JASPAR, are needed to identify them in the data [16, 35]. GRN tools like GENIE3 that cater to bulk RNA sequencing (RNAseq) estimate TF activity at the tissue level [25]. The advent of single cell RNAseq (scR-NAseq) enabled new software to dissect cell type specific TF programs in the case of GRNBoost2 and SCENIC, as well as TF regulation that drives cell state progression along cell pseudotime in the case of SCODE and SINCERITIES [24, 27, 28, 36]. Multiomic profiling of single cell ATAC-seq (scATACseq) and scRNAseq recently introduced a new wave of GRNs that deduce gene regulation from

TF binding sites, accessible enhancer or promoter regions, and downstream gene expression [26, 30, 37]. CellOracle and SCENIC + are two emerging methods that incorporate enhancer information to bolster TF-gene relationships [26, 37].

GRNs that are subset to TF signaling are helpful in reducing spurious network edges and computational run times. Benchmarking methods report that they have great merit in capturing well documented TF signaling pathways [16, 22•]. However, there are some significant limitations to this approach. Firstly, TF regulation is complex and time dependent such that its expression can precede that of the target gene [38]. This presents a challenge as networks are built on static RNA-seq datasets where TF expression is likely to be much lower than the target gene's and, therefore, subject to noisier or missing data collection on TFs [27, 39]. Secondly, these networks ignore non-TF forms of gene regulation, including long noncoding RNAs and chromatin modifier genes [40, 41]. Regardless, these networks are still important to discover novel functions of TFs in disease.

LR-Based Networks

Inter-cell type communication GRNs have risen in popularity with scRNAseq data to understand how cell signaling affects downstream gene activity [21]. Similar to TF-based GRNs, they rely on curated LR and target gene information to locate interacting cell pairs and inform network topology. CellChat, for example, calculates LR interaction probabilities for cell types based on their overall LR coexpression and generates a network from significant interactions through a permutation test framework [42]. NicheNet generates a prior knowledge communication network with its own LR databases, trains random forest models for each target gene based on expression of upstream network genes, and prioritizes LR and target gene paths that are most accurately modelled [31]. SpaTalk uses spatial transcriptomics (ST) data to locate adjacent pairs of cells on which it identifies co-expressed ligands and receptors and deduces downstream signaling genes based on a database-driven knowledge graph [43]. Recent multiomics approaches like SpaOTsc resolve cellular location for scRNAseq by aligning expression profiles with ST data, which enables GRN inference based on ligand-receptor expression and cell proximity [44]. LR-based networks carry similar advantages and disadvantages to TF-based. They can identify novel functions or disease contribution of known ligands and targets but will miss other important modes of cell communication, such as metabolite-mediated signaling [45].

Global Unsupervised Networks

Other GRN methods gather potential parent genes in an unsupervised manner, surveying all genes rather than a specific subset. Empirical feature selection methods use expression data alone to identify related genes. SCING, a gradient boosting machine learning based GRN approach designed for scRNAseq and spatial transcriptomics data, first implements gene-wise K-nearest neighbors to partition genes into distinct sets of similarly expressed genes, with genes in each set subject to modeling regulator-target relations using gradient boosting [29]. Coexpression networks that capture significant statistical gene correlations, like Weighted Gene Coexpression Network Analysis (WGCNA), can also provide potential associations as a basis for subsequent GRN inference [46, 47, 48••, 49]. Bayesian network methods like RIMBANET use multiple types of information to cover various types

of potential gene regulation, including TF-target databases, genetic priors derived from expression quantitative trait loci (eQTL), gene-gene correlation, and mutual information to assign prior parent and child node properties, followed by adding, flipping, or removing genes to derive GRNs that best explain the data [32, 33, 49]. Datadriven feature selection grant unsupervised GRNs the potential to model non-TF mediated gene regulation and incorporate other forms of regulatory genomic information, such as eQTLs. However, this implies they are prone to reporting spurious and confounding edges. Bootstrapping and posthoc network edge pruning strategies can help reduce the noise in these GRNs [29].

Databases for GRN Construction

Several existing databases and study cohorts offer human or mouse gene expression data relevant for atherosclerosis that may be employed for GRN construction. Depending on the resource, the data may be publicly available or obtained through application.

Plaqview

Plaqview is an interactive, user-friendly webserver that hosts various human and mouse cardiovascular single cell datasets that may be queried online for cell type gene expression, trajectory analysis, and drug/gene interactions [50•]. Additionally, Plaqview directs users to the original publications for each dataset where raw single cell data files may be accessed. The raw count matrix and cell type information from each dataset may be inputted to GRN methods designed for scRNAseq (Table 1), such as SCENIC for TF-based GRNs or SCING for global GRNs to generate cell type-specific GRNs [27, 29].

Hybrid Mouse Diversity Panel

The Hybrid Mouse Diversity Panel (HMDP) consists of a collection of over 100 inbred strains of mice that have been used to study various genetic and environmental contributions to disease [51]. Advantages of the HMDP include control over environmental factors and increased statistical power compared to most other strain panels. Thus far, HMDP has been leveraged to study a wide range of traits relevant in atherosclerosis, fatty liver disease, obesity, and heart failure [52-56]. With regards to the atherosclerosis HMDP, mice with a hyperlipidemic background were established via transgenic expression of human apolipoprotein E-Leiden (APOE-Leiden) and human cholesteryl ester transfer protein (CETP) [53]. Various atherosclerotic characteristics were measured, including lesion size and morphology, plasma lipid, insulin, and glucose levels, plasma metabolite and cytokine levels, and uptake of acetylated LDL by macrophages. Global transcript levels in aorta and liver tissues were also measured which can be used for GRN construction using bulk RNAseq based methods such as RIMBANET.

STARNET

The Stockholm-Tartu Atherosclerosis Reverse Networks Engineering Task (STARNET) study has collected gene expression data via RNA-sequencing for patients with CAD matched to patients without CAD for seven tissues, including blood, mammary artery, atherosclerotic aortic root, subcutaneous fat, visceral abdominal fat, skeletal muscle, and liver [57]. The data provided by STARNET may be used to build bulk tissue GRNs, and

the STARNET webserver may also be queried to visualize coexpression modules built from WGCNA that contain genes of interest [48••]. The STARNET study was preceded by its pilot study, the Stockholm Atherosclerosis Gene Network (STAGE) study [58].

Framingham Heart Study

The Framingham Heart Study (FHS) is a large, longitudinal cohort study spanning over 70 years with approximately 15,000 participants across three generations in Framingham, MA [59]. The Original cohort was examined for 32 cycles at approximately 2-year intervals; the Offspring cohort was examined for nine cycles, every 4—7 years; the Third-Generation cohort has undergone 3 examination cycles thus far. Two minority cohorts from Framingham (Omni 1–1995 and Omni 2–2002) were later included in the study to reflect the changing demographics of the town. Blood samples, electrocardiograms, family history, echocardiograms, cognitive function, and circulating biomarkers have been collected for the earlier cohorts of the FHS. More recently, genome-wide DNA methylation and transcriptomic data have been generated in the Offspring and Third-Generation cohorts. If approved for data access by the FHS, the gene expression data from the Offspring and Third-Generation cohorts may be used for GRN construction.

TOPMed

The Trans-Omics for Precision Medicine (TOPMed) initiative aims to integrate whole-genome sequencing (WGS) with other omics datasets, such as metabolomics, epigenomics, proteomics, and transcriptomics, using ~200,000 participants from over 85 cohorts to offer patient-specific disease treatments [60]. While various existing human databases consist primarily of participants with European ancestry, TOPMed has achieved ancestral and ethnic diversity, with ~60% of sequenced participants of non-European ancestry. Disease focuses of TOPMed primarily include those related to heart, lung, and blood, and the omics data processing is performed by several sequencing centers. TOPMed includes human RNA-sequencing datasets relevant for building CAD GRNs, including Multi-Ethnic Study of Atherosclerosis (MESA) and FHS.

GRN Application in Atherosclerosis and CAD

Tissue level and cell-cell interactions GRNs have been employed to understand genetic factors and to dissect mechanisms contributing to CAD. The use of GRNs not only has helped identify CAD-associated subnetworks and pathways, but ligand-receptor pairs that mediate cell-cell communications in CAD as well as predicted key driver genes, defined as network hubs or core genes with large numbers of network connections to genes associated with CAD. To better understand the topology of networks, Key Driver Analysis (KDA) has been a commonly used downstream method that can identify key regulatory genes, or key drivers, that lie central within a network [61, 62].

Bulk RNA-seq Studies

Several studies have demonstrated the utility of GRNs in understanding CAD heritability beyond what has been discovered by CAD GWAS studies. Specifically, transforming coexpression networks depicting gene-gene relationships into GRNs has thus far proven

to be particularly insightful. For example, Zeng et al. used STAGE and STARNET data to infer coexpression networks with gene-gene relationships within and across seven metabolic and vascular tissues, and a linear Gaussian Bayesian algorithm was subsequently used to infer GRNs for each coexpression network with eQTLs and TFs as priors [49]. Their GRNs contributed to 10% of CAD heritability beyond what had been previously attributed to GWAS risk loci, and the GRNs identified causal biological functions in CAD pathogenesis, including DNA binding, RNA metabolism, and blood coagulation. Their tissue-specific analyses further found that fat and arterial wall GRNs influenced CAD risk most robustly. Koplev et al. also inferred GRNs from gene-gene coexpression networks; they integrated STARNET DNA genotype and RNA-seq data from seven CAD-relevant tissues to infer 224 WGCNA coexpression modules, which were then transformed into GRNs using random forest-based regression method GENIE3 [48••]. They further integrated eQTLs with the 224 GRNs, calculating the total contribution of the GRNs to CAD heritability to be 59.8%. Mendelian randomization confirmed central key driver genes in 218 of the 224 GRNs to be causal for CAD. They also investigated both intra- and inter-organ interactions by applying Bayesian network modeling to the WGCNA eigengene values for each tissue-specific and cross-tissue coexpression module to create a representative supernetwork, finding cross-tissue GRNs contribute to nearly threefold more CAD heritability than tissue-specific GRNs.

Transforming gene-gene coexpression networks into GRNs has been employed in further understanding sex differences in CAD mechanisms as well. Hartman et al. constructed sex-specific atherosclerotic arterial wall coexpression modules using STARNET data, and Bayesian network inference was performed with the Fast Greedy Equivalence Search for continuous data algorithm [63]. Comparing GRNs between sexes, they concluded that immune-related genes were more active in men, whereas genes associated with mesenchymal and endothelial cells were more active in women. Further, based on integration with both human and mouse scRNAseq data, they found female GRN key driver genes were expressed in phenotypically switched SMCs and modulated by the transcription factor Klf4.

RIMBANET, a Bayesian network method that can incorporate various prior information including TFs, eQTLs, correlation, mutual information etc., has been used by numerous cardiovascular researchers to construct GRNs. For example, Makinen et al. employed Bayesian GRNs constructed using RIMBANET from genetics and gene expression data from previously published human and mouse studies [46]. They integrated 16 GWAS studies from CARDIoGRAM and the Ottawa Heart Institute with (1) eQTLs from human CAD-relevant tissues, (2) known metabolic and signaling pathways, and (3) their constructed tissue-specific GRNs. KDA revealed that a GRN implicated in antigen processing was strongly associated with CAD, with *GLO1* and *PPIL1* as key drivers central within this network. Similarly, Zhao et al. integrated top CAD candidate genes and the 1000 genomes-based CAD GWAS from the CARDIoGRAMplusC4D (Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) Genetics) consortium with tissue-specific GRNs constructed with RIMBANET and protein-protein interaction networks to identify top key regulators of CAD subnetworks [64]. Their data-driven approach confirmed previously known CAD GWAS genes, such as *COL4A2* and *CXCL12*, as key drivers and also pin-pointed novel key drivers, including

LUM and *STAT3*. Von Scheidt and Zhao et al. combined individual liver Bayesian GRNs constructed from various published mouse and human liver datasets using RIMBANET [65]. They mapped mouse atherosclerosis and human CAD GWAS genes to the GRN to predict key driver genes functioning within disease-relevant liver subnetworks. KDA revealed transcription factor *Maff* as a top mouse key driver that was also directly connected to known human CAD GWAS loci, including *LDLR, MCL1*, and *TRIB1*. They confirmed *Maff* coexpression with lipid metabolism and inflammatory processes in the atherosclerosis HMDP and *MAFF* coexpression with CAD in human liver data from STARNET. Finally, Kurt and Cheng et al. studied the shared and distinct pathways and networks between human and mouse in CAD, integrating human CAD GWAS from the CARDIoGRAMplusC4D consortium and mouse atherosclerosis GWAS from the HMDP with eQTLs from human (STARNET and Genotype-tissue Expression project (GTEx)) and mouse (HMDP) datasets [66]. They found that human and mouse shared >75% of tissue-specific CAD causal pathways. They constructed Bayesian GRNs with RIMBANET using mouse aorta and liver gene expression data from HMDP and human aorta, coronary artery, and liver gene expression data from GTEx. Using these GRNs for KDA, they identified various key drivers shared and distinct across species, further validating their network connections with single cell data. Overall, Bayesian GRN construction via RIMBANET and further identification of network key drivers have identified various key mechanisms and genes involved in CAD onset and progression in both human and mouse. As Bayesian networks are directed acyclic graphs, additional GRN construction methods described in our review may also be implemented to study CAD mechanisms and key regulatory genes to gain further insight into the disease.

scRNAseq Studies

With advances in single cell omics technologies to study cellular heterogeneity within a single tissue, use of cell-cell interaction networks has elucidated the key mechanisms in CAD microenvironments. Ma et al. employed a human coronary artery scRNAseq dataset to examine cell-cell interaction within the lesional microenvironment using CellChat [50•]. From CellChat, they found most cell types within arterial lesions contribute to macrophage activation via ligand-receptor interactions including HLA-DPA1:CD4, HLA-DMB:CD4, and HLA-DRA:CD4. Additionally, the CellChat analysis revealed SMCs signal strongly in fibronectin and collagen pathways, fibroblasts strongly contribute to larninin and complement pathways, and almost all lesional cell types signal to NK and T cells, primarily through HLA-A. They also integrated their identified ligand-receptor interaction pairs with the Drug-Gene Interaction database (DGIdb 3.0) to determine candidate drugs that may target harmful cell-cell interactions and also queried various druggable genome databases. Application of additional GRN methods to this and future single cell datasets will help gain additional biological insight.

Future Directions

GRNs are proving extremely useful in understanding CAD, as they have elucidated the interactions among many physiological pathways related to its pathology and prognosis in cardiovascular tissues and cell types and further identified key driver genes that likely

regulate disease networks and tissue/cell crosstalk. They also provide functional genomic interpretation to genetic risk variants in CAD GWAS risk loci, and eQTLs prioritize disease subnetworks that correspond to immune and metabolic pathways.

It is important to note that these driver genes were discovered with tissue-level GRNs and cannot be attributed to any particular cell type. The rise of scRNAseq datasets and the corresponding GRN tools present exciting opportunities to resolve cell-level contributions to CAD by identifying shared and distinct disease mechanisms between cell types as well as novel key drivers in rare cell populations masked in bulk tissue GRNs. Cell-cell communication networks will expand on how cellular systems interact with each other within and between tissues. Additionally, the increasing prevalence of single cell and spatial multiomics technologies, such as scATACseq and spatial transcriptomics, and integrative GRN methods emphasizes a new frontier for CAD research to understand how epigenetic, transcriptomic, and spatially dependent signaling mechanisms interact with each other in this complex disease.

The variability in network approaches greatly influence the similarities and differences observed in these studies' findings. Since network structures are heavily dependent on data artifacts, prior gene information, and the inference algorithm, there is a need to harmonize the architectures from multiple networks. Systematic benchmarking of all these methods on gold standard bulk and single cell multiomics and multi-tissue datasets are necessary to provide insight on the strengths and weaknesses of each method in different experimental contexts. The different modes of gene regulation highlighted by these methods also give merit to incorporating the complementary information across multiple GRN approaches to produce a holistic picture of all forms of gene regulation present in the data. Overall, unifying these different GRN approaches will give greater insight into the causal mechanisms in CAD. As GRN modeling and analysis continue to elucidate novel CAD mechanisms and regulatory genes and quantify disease heritability, experimental validation of GRN findings remains important and necessary. Once validated, GRNs can serve as an important discovery tool to facilitate precision medicine.

Funding

This work was supported by NIH grants DK120342, HL148577, and HL147883 (AJL). This work was also supported by NIH grants NS111378, HL147883, and NS117148 and JDRF 2-SRA-2021-1028-S-B (XY).

Glossary

Boolean Network
Deterministic networks that state logical statements to describe network connectivity.

Ordinary Differential Equation
Mathematical equation that uses a derivative of a dependent variable as a function of a single independent variable.

Regression

Supervised machine learning paradigm for predicting a continuous outcome variable from a set of observed feature variables.

Ensemble Trees
A type of regression model using an aggregate of decision tree regressors to predict a consensus outcome.

Random Forest
Ensemble method that trains small decision trees based on a subset of samples and feature and aggregates the predictions across the trees as the final prediction.

Gradient Boosting
Ensemble method that aggregates small decision trees sequentially, training each subsequent tree to predict the errors of the previous trees until the final ensemble yields the most accurate prediction.

Bayesian Inference
Statistical inference principle based on Bayes' Theorem to update the probability for a hypothesis (posterior) based on the observed data (prior), and the likelihood of observing the hypothesis given the data (likelihood).

Directed Acyclic Graphs
Graphs with nodes and directional edges such that no cycles exist in the network.

Expression Quantitative Trait Loci (eQTL)
Genetic analysis that identifies genetic loci that affect the gene expression.

Mutual Information
Statistical measure of mutual dependence between two variables.

Partial Information Decomposition
A method to decompose the mutual information of random variables into unique, redundant, and synergistic information.

Key Drivers
Network hub or core genes with large numbers of connections to disease-associated genes.

References

Papers of particular interest, published recently, have been highlighted as:

• Of importance

•• Of major importance

1. Tsao CW, et al. Heart Disease and Stroke Statistics—2023 Update: A Report From the American Heart Association. Circulation. 2023;147(8):e93–621. 10.1161/CIR.0000000000001123. [PubMed: 36695182]
2. Watkins H, Farrall M. Genetic susceptibility to coronary artery disease: from promise to progress. Nat Rev Genet. 2006;7(3). 10.1038/nrgl805.

3. Pencina MJ, et al. Quantifying Importance of Major Risk Factors for Coronary Heart Disease. Circulation. 2019;139(13):1603–11. 10.1161/CIRCULATIONAHA.117.031855. [PubMed: 30586759]

4. Björkegren JLM, Lusis AJ. Atherosclerosis: Recent developments. Cell. 2022;185(10):1630–45. 10.1016/j.cell.2022.04.004. [PubMed: 35504280]

5. Seldin M, Yang X, Lusis AJ. Systems genetics applications in metabolism research. Nat Metab. 2019;1(11). 10.1038/s42255-019-0132-x.

6. Yang X. Multitissue Multiomics Systems Biology to Dissect Complex Diseases. Trends Mol Med. 2020;26(8):718–28. 10.1016/j.molmed.2020.04.006. [PubMed: 32439301]

7. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multiomics Data Integration, Interpretation, and Its Application. Bioinforma Biol Insights. 2020;14:1177932219899051. 10.1177/1177932219899051.

8. Meng Q, Mäkinen V-P, Luk H, Yang X. Systems Biology Approaches and Applications in Obesity, Diabetes, and Cardiovascular Diseases. Curr Cardiovasc Risk Rep. 2013;7(1):73–83. 10.1007/sl2170-012-0280-y. [PubMed: 23326608]

9. Arneson D, Shu L, Tsai B, Barrere-Cain R, Sun C, Yang X. Multidimensional Integrative Genomics Approaches to Dissecting Cardiovascular Disease. Front Cardiovasc Med. 2017;4. [Online]. Available: https://www.frontiersin.org/articles/ https://doi.org/10.3389/fcvm.2017.00008. Accessed 16 Sep 2023.

10. Blencowe M, Ameson D, Ding J, Chen Y-W, Saleem Z, Yang X. Network modeling of single-cell omics data: challenges, opportunities, and progresses. Emerg Top Life Sci. 2019;3(4):379–98. 10.1042/etls20180176. [PubMed: 32270049]

11. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017;169(7): 1177–86. 10.1016/j.cell.2017.05.038. [PubMed: 28622505]

12. Erdmann J, Kessler T, Munoz Venegas L, Schunkert H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. Cardiovasc Res. 2018;114(9):1241–57. 10.1093/cvr/cvy084. [PubMed: 29617720]

13. Koyama S, et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. Nat Genet. 2020;52(11). 10.1038/S41588-020-0705-3.

14. Akers K, Murali TM. Gene regulatory network inference in single-cell biology. Curr Opin Syst Biol. 2021;26:87–97. 10.1016/j.coisb.2021.04.007.

15. Armingol E, Officer A, Harismendy O, Lewis NE. Deciphering cell–cell interactions and communication from gene expression. Nat Rev Genet. 2021 ;22(2). 10.1038/s41576-020-00292-x.

16. Badia-I-Mompel P, et al. Gene regulatory network inference in the era of single-cell multi-omics. Nat Rev Genet. 2023. 10.1038/s41576-023-00618-5.

17. Blencowe M, Karunanayake T, Wier J, Hsu N, Yang X. Network Modeling Approaches and Applications to Unravelling Non-Alcoholic Fatty Liver Disease. Genes. 2019; 10(12). 10.3390/genesl0120966.

18. Chai LE, Loh SK, Low ST, Mohamad MS, Deris S, Zakaria Z. A review on the computational approaches for gene regulatory network construction. Comput Biol Med. 2014;48:55–65. 10.1016/j.compbiomed.2014.02.011. [PubMed: 24637147]

19. Zuo Y, Wei D, Zhu C, Naveed O, Hong W, Yang X. Unveiling the Pathogenesis of Psychiatric Disorders Using Network Models. Genes. 2021;12(7). 10.3390/genesl2071101.

20. 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(1):232. 10.1186/sl2859-018-2217-z. [PubMed: 29914350]

21. Liu Z, Sun D, Wang C. Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information. Genome Biol. 2022;23(1):218. 10.1186/sl3059-022-02783-y. [PubMed: 36253792]

22•. Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat Methods. 2020;17(2):147–54. 10.1038/s41592-019-0690-6. [PubMed: 31907445] This gene regulatory network method

benchmarking study for single cell data uses multiple accuracy metrics on real and simulated data to highlight the strengths and weaknesses of different approaches.

23. Woodhouse S, Piterman N, Wintersteiger CM, Göttgens B, Fisher J. SCNS: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data. BMC Syst Biol. 2018;12(1):59. 10.1186/sl2918-018-0581-y. [PubMed: 29801503]
24. Matsumoto H, et al. SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. Bioinformatics. 2017;33(15):2314–21. 10.1093/bioinformatics/btx194. [PubMed: 28379368]
25. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. PLoS One. 2010;5(9). 10.1371/journal.pone.0012776.
26. Bravo González-Blas C, et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Nat Methods. 2023. 10.1038/s41592-023-01938-4.
27. Aibar S, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. 2017;14(11). 10.1038/nmeth.4463.
28. Moerman T, et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. Bioinformatics. 2019;35(12):2159–61. 10.1093/bioinformatics/bty916. [PubMed: 30445495]
29. Littman R, Cheng M, Wang N, Peng C, Yang X. SCING: Inference of robust, interpretable gene regulatory networks from single cell and spatial transcriptomics. IScience. 2023;26(7):107124. 10.1016/j.isci.2023.107124. [PubMed: 37434694]
30. Jiang J, et al. IReNA: Integrated regulatory network analysis of single-cell transcriptomes and chromatin accessibility profiles. iScience. 2022;25(11): 105359. 10.1016/j.isci.2022.105359. [PubMed: 36325073]
31. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020;17(2):159–62. 10.1038/S41592-019-0667-5. [PubMed: 31819264]
32. Sanchez-Castillo M, Blanco D, Tienda-Luna IM, Carrion MC, Huang Y. A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. Bioinformatics. 2018;34(6):964–70. 10.1093/bioinformatics/btx605. [PubMed: 29028984]
33. Zhu J, et al. Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. PLoS Comput Biol. 2007;3(4):e69. 10.1371/joumal.pcbi.0030069. [PubMed: 17432931]
34. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012;13(9). 10.1038/nrg3207.
35. Castro-Mondragon JA, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2022;50(D1):D165–73. 10.1093/nar/gkab1113. [PubMed: 34850907]
36. Papili Gao N, Ud-Dean SMM, Gandrillon O, Gunawan R. SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. Bioinformatics. 2018;34(2):258–66. 10.1093/bioinformatics/btx575. [PubMed: 28968704]
37. Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA. Dissecting cell identity via network inference and in silico gene perturbation. Nature. 2023;614(7949). 10.1038/s41586-022-05688-9.
38. Martin EW, Sung M-H. Challenges of Decoding Transcription Factor Dynamics in Terms of Gene Regulation. Cells. 2018;7(9). 10.3390/cens7090132.
39. Raj A, van Oudenaarden A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell. 2008;135(2):216–26. 10.1016/j.cell.2008.09.050. [PubMed: 18957198]
40. Statello L, Guo C-J, Chen L-L, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021;22(2). 10.1038/S41580-020-00315-9.
41. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. CeU Res. 2011;21(3). 10.1038/cr.2011.22.
42. Jin S, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021 ;12(1). 10.1038/s41467-021-21246-9.

43. Shao X, et al. Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. Nat Commun. 2022; 13(1). 10.1038/S41467-022-32111-8.

44. Cang Z, Nie Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. Nat Commun. 2020;11(1):2084. 10.1038/s41467-020-15968-5. [PubMed: 32350282]

45. Wang Y-P, Lei Q-Y. Metabolite sensing and signaling in cell metabolism. Signal Transduct Target Ther. 2018;3(1). 10.1038/s41392-018-0024-7.

46. Mäkinen V-P, et al. Integrative Genomics Reveals Novel Molecular Pathways and Gene Networks for Coronary Artery Disease. PLOS Genet. 2014;10(7):el004502. 10.1371/joumal.pgen.1004502.

47. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinforma. 2008;(1):559, [Online]. Available: https://bmcbioinformatics.biomedcentral.com/articles/ https://doi.org/10.1186/1471-2105-9-559.

48••. Koplev S, et al. A mechanistic framework for cardiometabolic and coronary artery diseases. Nat Cardiovasc Res. 2022;1(1). 10.1038/s44161-021-00009-1. Findings from this study that employs both within- and cross-tissue gene regulatory networks from metabolically relevant tissues highlight the ability of gene regulatory networks to predict coronary artery disease heritability beyond what has been identified by GWAS risk loci.

49. Zeng L, et al. Contribution of Gene Regulatory Networks to Heritability of Coronary Artery Disease. J Am Coll Cardiol. 2019;73(23):2946–57. 10.1016/j.jacc.2019.03.520. [PubMed: 31196451]

50•. Ma WF, et al. Enhanced single-cell RNA-seq workflow reveals coronary artery disease cellular cross-talk and candidate drug targets. Atherosclerosis. 2022;340:12–22. 10.1016/j.atherosclerosis.2021.11.025. [PubMed: 34871816] This study proves the ability of an existing cell-cell communication method to leverage the granularity of single cell RNA-sequendng data and further provides a publicly available and interactive single cell database.

51. Bennett BJ, et al. A high-resolution association mapping panel for the dissection of complex traits in mice. Genome Res. 2010;20(2):281–90. 10.1101/gr.099234.109. [PubMed: 20054062]

52. Lusis AJ, et al. The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. J Lipid Res. 2016;57(6):925–42. 10.1194/jlr.R066944. [PubMed: 27099397]

53. Bennett BJ, et al. Genetic Architecture of Atherosclerosis in Mice: A Systems Genetics Analysis of Common Inbred Strains. PLOS Genet. 2015;11(12):el005711. 10.1371/joumal.pgen.1005711.

54. Hui ST, et al. The genetic architecture of NAFLD among inbred strains of mice. eLife. 2015;4:e05607. 10.7554/eLife.05607. [PubMed: 26067236]

55. Parks BW, et al. Genetic Control of Obesity and Gut Microbiota Composition in Response to High-Fat, High-Sucrose Diet in Mice. Cell Metab. 2013;17(1): 141–52. 10.1016/j.cmet.2012.12.007. [PubMed: 23312289]

56. Rau CD, et al. Mapping Genetic Contributions to Cardiac Pathology Induced by Beta-Adrenergic Stimulation in Mice. Circ Cardiovasc Genet. 2015;8(1):40–9. 10.1161/CIRCGENETICS.113.000732. [PubMed: 25480693]

57. Franzén O, et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. Science. 2016;353(6301):827–30. 10.1126/science.aad6970. [PubMed: 27540175]

58. Hägg S, et al. Multi-Organ Expression Profiling Uncovers a Gene Module in Coronary Artery Disease Involving Transendothelial Migration of Leukocytes and LIM Domain Binding 2: The Stockholm Atherosclerosis Gene Expression (STAGE) Study. PLOS Genet. 2009;5(12):el000754. 10.1371/joumal.pgen.1000754.

59. Andersson C, Johnson AD, Benjamin EJ, Levy D, Vasan RS. 70-year legacy of the Framingham Heart Study. Nat Rev Cardiol. 2019;16(11). 10.1038/S41569-019-0202-5.

60. Taliun D, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature. 2021;590(7845). 10.1038/s41586-021-03205-y.

61. Ding J, et al. Mergeomics 2.0: a web server for multi-omics data integration to elucidate disease networks and predict therapeutics. Nucleic Acids Res. 2021;49(Wl):W375–87. 10.1093/nar/gkab405. [PubMed: 34048577]

62. Shu L, et al. Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. BMC Genomics. 2016;17(1):874. 10.1186/S12864-016-3198-9. [PubMed: 27814671]

63. Hartman RJG, et al. Sex-Stratified Gene Regulatory Networks Reveal Female Key Driver Genes of Atherosclerosis Involved in Smooth Muscle Cell Phenotype Switching. Circulation. 2021;143(7):713–26. 10.1161/CIRCULATIONAHA.120.051231. [PubMed: 33499648]

64. Zhao Y, et al. Network-Based Identification and Prioritization of Key Regulators of Coronary Artery Disease Loci. Arterioscler Thromb Vase Biol. 2016;36(5):928–41. 10.1161/ATVBAHA.115.306725.

65. von Scheldt M, et al. Transcription Factor MAFF (MAF Basic Leucine Zipper Transcription Factor F) Regulates an Atherosclerosis Relevant Network Connecting Inflammation and Cholesterol Metabolism. Circulation. 2021;143(18):1809–23. 10.1161/CIRCULATIONAHA.120.050186. [PubMed: 33626882]

66. Kurt Z, et al. Shared and distinct pathways and networks genetically linked to coronary artery disease between human and mouse. eLife. 2023;12. 10.7554/eLife.88266.

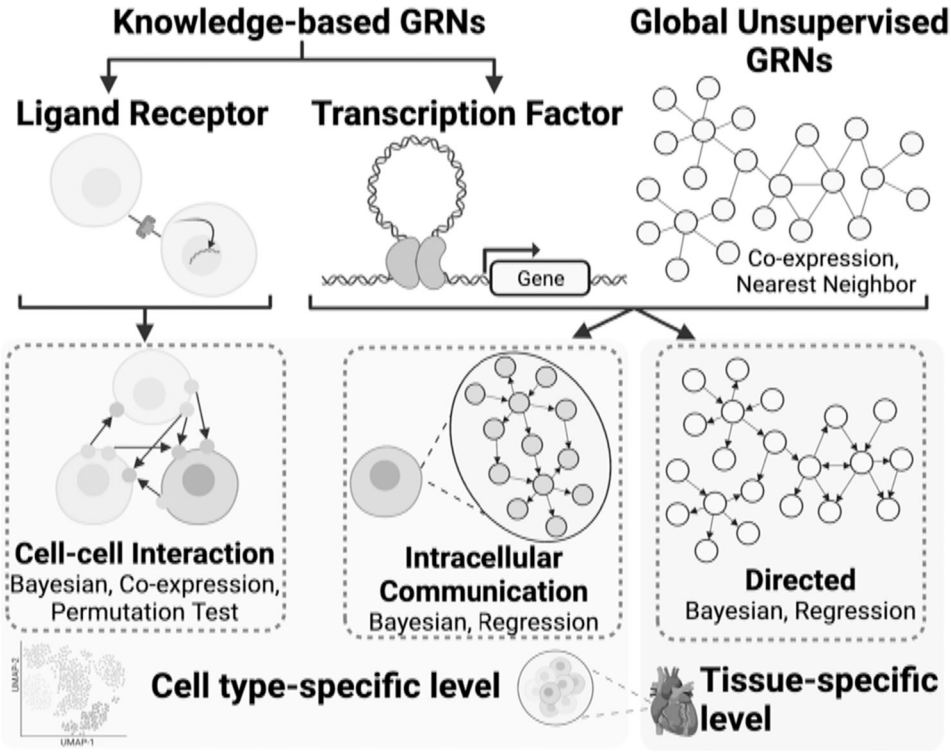

Fig. 1.

Overview of GRNs. GRN methods describe various forms of gene regulation. Knowledge-based methods supply known biological regulator and target information to orient the network interactions. TF-based GRNs focus on TF-target gene regulatory pathways while ligand-receptor based networks model ligand-receptor mediated cell-cell interaction. Global unsupervised networks calculate and refine data-driven gene expression patterns to capture broad regulatory landscapes. Bulk RNAseq GRNs explain tissue-level gene regulation, while scRNAseq GRNs can explain within and between cell-type regulation

Curr Atheroscler Rep. Author manuscript; available in PMC 2024 October 10.

Table 1

Description of select GRN methods

Category	Method Name	Algorithm	Data Type	Advantage	Limitation
Transcription Factor Based	GENIE3 [25]	Random forest	Bulk RNAseq or scRNAseq	Simple model to build complex nonlinear gene relationships	Performs as a random network if it does not assume prior knowledge of TFs[a]
	GRNBoost [28]	Gradient boosting	scRNAseq	Early stop monitoring and parallel computing improves run time	Dense network Prone to bi-directional edges with-out prior TF assumptions
	SCODE [24]	Linear ODE[b]	scRNAseq, pseudotime	Reduced run time	Assumes all cells have the same trajectory
	SINCERITIES [36]	Granger causality, ridge regression, partial correlation	scRNAseq, experimental time point	Low computational complexity	Requires data from multiple timepoints
	SCNS [23]	Boolean	scRNAseq, pseudotime	Models combinatorial effects on genes	Requires binarization, Limited to Boolean relationships
	SCENIC [27]	Random forest (GENIE3)	scRNAseq	Prunes TF edges with motif enrichment	Limited to TF-based regulation
	SCENIC+ [26]	Gradient boosting, motif enrichment	scRNAseq, scATACseq	Refine edges with chromatin accessibility	Limited to TF-based regulation
	IReNA [30]	Random forest (GENIE3), correlation, motif enrichment	scRNAseq, bulk/ scATACseq (optional)	Pseudotime-based smoothed expression mitigates sparsity	Subject to GENIE3 limitations
	CellOracle [37]	Ridge regression, signal propagation, motif enrichment	scRNAseq, scATACseq	Uses regression coefficients to refine edges	Constrained to linear regression in order to perform TF signal propagation
Ligand Receptor Based	NicheNet [31]	Bayesian parameter optimization, random forest	scRNAseq	Uses literature-based ligand-receptor and ligand-target gene network priors	Does not account for sparsity, cannot identify new regulations outside of databases
	CellChat [42]	Law of mass action, permutation test	scRNAseq	Extracts contextual latent patterns	Cannot capture low to moderate signaling due to sparsity constraints
	SpaTalk [43]	Nonnegative linear model, per-mutation test, knowledge graph random walk	Spatial transcriptomics	Utilize distance and ligand-receptor expression to prioritize cell-pairs	Coexpression binarizes ligand receptor expression, Requires scRNAseq reference for spot deconvolution
	SpaOTsc [44]	Optimal transport, partial infor-mation decomposition	scRNAseq, spatial transcriptom-ics	Utilize distance and ligand-receptor expression to prioritize cell-pairs	Intractable for large datasets
Global Unbiased Network	RIMBANET [33]	Bayesian, MCMCC[c]	Bulk RNAseq	Uses eQTL[d] to generate network prior	Computationally intensive, cannot model feedback loops
	SCING [29]	Gradient boosting	scRNAseq, spatial transcriptomics	Supercells mitigate gene sparsity, KNN[e] feature selection reduces run time, bagging removes noise edges	Many manually tuned hyperpa-rameters
	GRNVBEM [32]	First Order Autoregressive Model, Variational Bayes Expectation Maximization	scRNAseq, pseudotime	Accounts for noisy data	Uses fold change instead of actual expression values

Abbreviations

[a] TF: Transcription Factor

[b] ODE: Ordinary Differential Equations

[c] MCMC: Markov Chain Monte Carlo

[d] eQTL: Expression Quantitative Trait Loci

[e] KNN: K-Nearest Neighbors