

Review

Elucidation of Biological Networks across Complex Diseases Using Single-Cell Omics

Yang Li ^{1,4}, Anjun Ma ^{1,4}, Ewy A. Mathé ^{2,@}, Lang Li,¹ Bingqiang Liu ^{3,*} and Qin Ma ^{1,4,*,@}

Single-cell multimodal omics (scMulti-omics) technologies have made it possible to trace cellular lineages during differentiation and to identify new cell types in heterogeneous cell populations. The derived information is especially promising for computing cell-type-specific biological networks encoded in complex diseases and improving our understanding of the underlying gene regulatory mechanisms. The integration of these networks could, therefore, give rise to a heterogeneous regulatory landscape (HRL) in support of disease diagnosis and drug therapeutics. In this review, we provide an overview of this field and pay particular attention to how diverse biological networks can be inferred in a specific cell type based on integrative methods. Then, we discuss how HRL can advance our understanding of regulatory mechanisms underlying complex diseases and aid in the prediction of prognosis and therapeutic responses. Finally, we outline challenges and future trends that will be central to bringing the field of HRL in complex diseases forward.

Significance in Elucidating the Gene Regulatory Mechanism of Complex Diseases

Fundamental biological questions include how individual cells differentiate into various tissues/cell types, how cellular activities are operated in a coordinated manner, and what gene regulatory mechanisms support these processes [1,2]. Gene regulatory mechanisms involve diverse procedures carried out by cells to increase/decrease the production of specific cell products (metabolites or RNA). They mainly involve regulation of the rate of transcription and translation, processing of RNA molecules (e.g., alternative splicing), control of mRNA molecule stability, and modulation of the pathway behavior and interactions [3].

Disorders in the above regulatory activities usually relate to the occurrence and development of complex diseases, which refer to diseases caused by a combination of genetic, environmental, and lifestyle factors [4–9]. Examples of complex diseases include asthma [10], epilepsy [11], hypertension [12], Alzheimer's disease (AD) [13], manic depression [14], schizophrenia [15], cancer [16], diabetes [17], and heart diseases [18]. Early detection of these diseases relies on the identification of predictive biomarkers, which plays an important role in improving prognosis and identifying appropriate therapeutics. Although some predictive diagnostic tests, for example, MammaPrint and Oncotype DX, have been identified as biomarkers and applied to clinical practice for breast cancer prognosis, it remains a challenge to reverse-engineer the gene regulatory mechanisms underlying these biomarkers [19]. Furthermore, gene regulatory mechanisms disrupted by genotype aberrations may cause altered expression of genes and lead to pathogenic phenotypes [20,21]. For example, genetic aberrations that disrupt topologically associated domains (TADs) were found to rewire promoter–enhancer interactions and lead to limb malformations [20]. Accurately translating genotype to pathological phenotype requires effective approaches to unravel the gene regulatory mechanisms.

Highlights

Advances in single-cell sequencing technologies open a window to understanding the heterogeneous regulatory landscape (HRL) encoded in complex diseases, by inferring various biological networks.

The development of single-cell multimodal omics (scMulti-omics) technologies combined with scMulti-omics integration tools provides multimodal measurements of HRL.

Application of HRL to complex diseases presents opportunities and poses challenges.

Among the remaining challenges, establishing a robust benchmarking pipeline is paramount.

In support of the integration of diverse single-cell modalities, bulk and single-cell omics, the deep learning and artificial intelligence (AI) omics become the major trends.

¹Department of Biomedical Informatics, The Ohio State University, Columbus, OH, 43210, USA

²Division of Preclinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health (NIH), Rockville, MD, 20892, USA

³School of Mathematics, Shandong University, Jinan, Shandong, 250100, China

⁴<https://u.osu.edu/bmb/>

*Correspondence:

bingqiang@sdu.edu.cn (B. Liu) and qin.ma@osumc.edu (Q. Ma).

[@]Twitter: @QinMaBMBL (Q. Ma) and @EwyMathe (E.A. Mathe).

Meanwhile, discovering the pathways that are perturbed and dysregulated in each disease type is critical for elucidating the gene regulatory mechanisms relating genotypes to phenotypes and may direct efforts to develop treatments. Pathway perturbation encompasses functional alterations of a pathway induced by internal or external factors, and pathway dysregulation refers to metabolic disorders and cellular homeostasis disruption [22]. From a systems biology viewpoint, perturbed or dysregulated pathways foster the development or deterioration of complex diseases [5]. Therefore, interpretation of pathway perturbations and dysregulations could shed light on improving diagnosis and personalized treatments [5].

Heterogeneous Regulatory Landscape Enabled by Single-Cell Omics Technologies

Although bulk omics approaches for interrogating genetic variations that drive the regulation of pathways have determined genetic aetiologies for thousands of human diseases, they are typically limited to studies at the ecosystem or organism level [23,24]. Practically, the diversity of individual cells within an ecosystem is greater than we can measure by studying a mixture of cells of the same organism, and the genomes within the cells of an individual multicellular organism are not always the same [23]. In these settings, omics assays at the single-cell level represent powerful, high-resolution tools for probing specific molecular mechanisms and pathways and uncovering the nature of cell heterogeneity [22,25].

The fast development of single-cell omics technologies has enabled comprehensive profiling of genetic, epigenetic, spatial, proteomic, and lineage information [23,26,27]. These technologies have provided exciting opportunities for systematic investigation of rare cell types, cellular heterogeneity [28], evolution [28], and cell-to-cell interactions [29] in a wide range of tissues and cell populations [30–32]. The generated multimodal information from individual cells has enabled the elucidation of the cellular reprogramming, developmental dynamics, and communication networks in disease development, and identification of unique malfunctions of individual cells [1,2,26,30,31,33–35]. Interrogation of the contributions of individual cells to pathogenesis and disease progression [23], especially the immune cells [35], enables aetiology interpretation and therapeutics identification in complex diseases [30].

A central question that arises now is how we can discover the network models that control cellular differentiation and drive transitions from one cell type to another. Specifically, a **gene coexpression network (GCN)** (see Glossary) is a comprehensive compilation of coexpression linkages among genes, which can be inferred from single-cell RNA-sequencing (scRNA-Seq) [36–41]. From scRNA-Seq data, **gene regulatory networks (GRNs)** can be constructed to model the regulatory interactions between transcription factors (TFs)/noncoding RNAs and target genes [25,42–54]. Computed from single-cell ATAC-sequencing (scATAC-Seq), **cis-coaccessibility networks (CCANs)** in single cells indicate how accessible *cis*-regulatory elements (CREs) orchestrate gene regulatory mechanisms dynamically [55–57]. **Methylation-associated gene regulatory networks (MGRNs)** could be inferred from single-cell DNA methylation sequencing (scMethyl-Seq) to capture the impacts of epigenetic factors on gene regulatory mechanisms [58,59]. To quantify the interplays between chromatin loci in a 3D space, **chromatin interaction networks (CINs)** in individual cells can be identified from single-cell Hi-C (scHi-C) data [60]. In single cells, **CRE–gene interaction networks (CGNs)** can be constructed through the integration of scRNA-Seq and scATAC-Seq, which provide a detailed view of how CREs influence gene expression in single cells [61–65]. Similarly, **TF–CRE interaction networks (TCNs)** in individual cells are mainly inferred by integrating scRNA-Seq and scATAC-Seq [61–65]. Although each network provides a different glimpse into the entire cell heterogeneity [66], they can jointly build up a **heterogeneous regulatory landscape (HRL)**

Glossary

Artificial intelligence (AI): also known as machine intelligence. Intelligence demonstrated by machines simulating natural intelligence.

cis-coaccessibility network (CCAN): an undirected graph, where each node represents a CRE and a pair of CREs is connected by an edge if there is a significant coaccessibility relationship between them based on available scATAC-Seq data.

Chromatin interaction network (CIN): an undirected graph, where each node represents a chromatin locus, and a pair of regions are connected by an edge if there is a significant interaction between them based on available scHi-C data.

CRE–gene interaction network (CGN): a directed graph with nodes representing CREs and genes, and edges corresponding to the *cis*-regulatory relationships between the two kinds of nodes.

Convolutional neural network (CNN): a kind of neural network that performs convolution operations in at least one of their layers.

Deep learning (DL): a type of ML algorithm that uses multiple layers to progressively extract higher-level features from the raw input.

Deep neural network (DNN): a type of artificial neural network with multiple layers between the input and output layers.

Federated learning (FL): also known as collaborative learning. An ML technique that trains an algorithm across multiple decentralized edge devices or servers holding local data samples, without exchanging their data samples.

Gene coexpression network (GCN): an undirected graph, where each node corresponds to a gene, and a pair of nodes is connected with an edge if there is a significant coexpression relationship between them.

Graph-convolutional neural network: a neural network designed to work directly on graphs and leverage their structural information.

Gene regulatory network (GRN): a directed graph with nodes representing transcriptional regulators and genes, and edges corresponding to the regulatory relationships between these two kinds of nodes.

Heterogeneous regulatory landscape (HRL): a large genomic region of heterogeneous cell types

that characterizes cell-type-specific genetic and epigenetic regulatory relationships. All the above networks are thus called HRL-associated networks, and they are outlined in Figure 1, where their applications to complex diseases are also summarized.

Moreover, **single-cell multimodal omics (scMulti-omics)** opens up new frontiers by simultaneously measuring multiple modalities. In scMulti-omics, the information of one modality can be leveraged to improve the interpretation of another. Currently, at most four types of single-cell omics can be measured simultaneously, leading to 13 combinations [66]. Specifically, there are nine double-modality sequencing techniques (e.g., sci-CAR-Seq combines RNA expression and chromatin accessibility), three triple-modality sequencing techniques (e.g., scNMT-Seq combines RNA expression, chromatin accessibility, and DNA methylation), and one quad-modality sequencing technique (ECCITE-Seq combines RNA expression, protein expression, T cell receptor, and perturbation). Hence, scMulti-omics has brought about new resources for understanding HRL [66].

Overall, HRL underlies the multilevel genotype and phenotype observations across diverse cell types. Changes of HRL alongside the cell developmental trajectory reflect the pathological

containing several long-range-acting regulatory sequences that control target genes in a coordinated manner.

Methylation associated gene regulatory network (MGRN): a GRN where regulator-gene regulatory relationships are influenced by methylation status.

Machine learning (ML): an application of AI that improves automatically through experience.

Perturb-Seq: also known as Crisp-Seq or CROP-Seq. A high-throughput method of performing scRNA-Seq on pooled genetic perturbation screens.

Perturbing strategy: a category of ML interpretation strategies that measure how changes in the input data impact model predictions.

Probing strategy: a category of ML interpretation strategies that involve directly inspecting the structure and parameters in a trained model.

Recurrent neural network (RNN): a type of artificial neural network where connections between nodes form a directed graph along a temporal sequence.

Single-cell multimodal omics (scMulti-omics): a category of technologies measuring multiple types of molecules from the same individual cell.

Surrogate strategy: a category of ML interpretation strategies that involve training an inherently interpretable model (e.g., a linear model) to approximate the black-box model using the same data.

TF-CRE interaction network (TCN): a directed graph, where nodes denote TFs and CREs, and edges represent the binding activity of TFs to CREs.

Transfer learning (TL): an ML technique that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem.

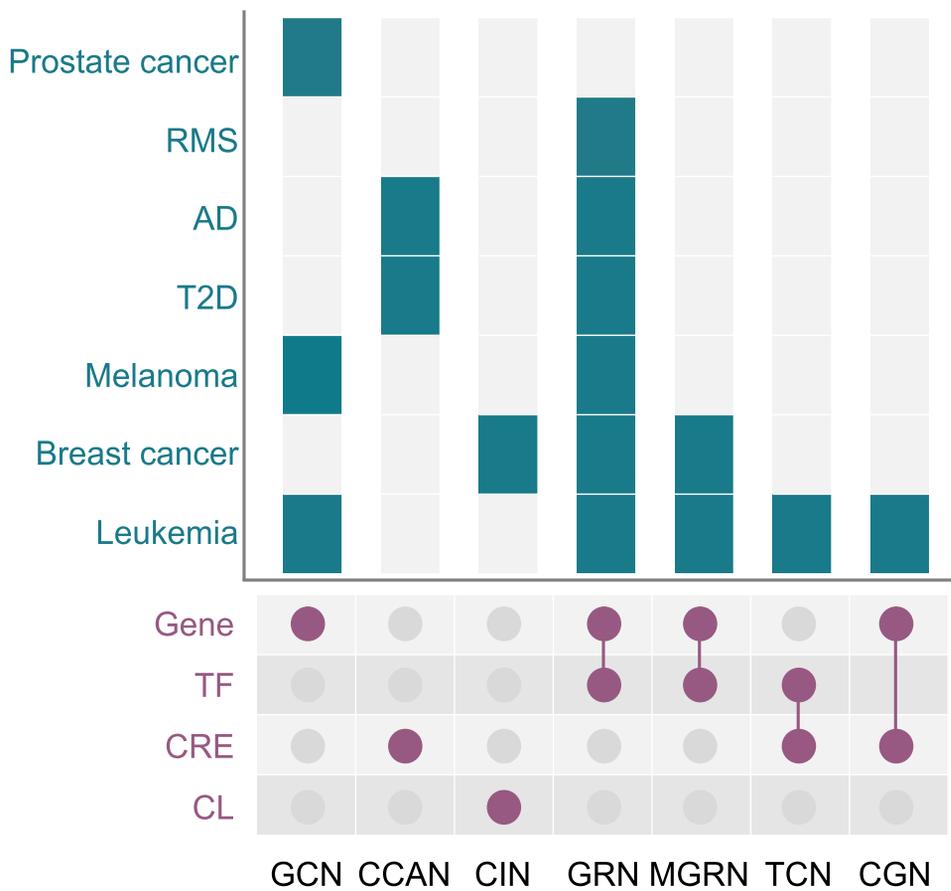


Figure 1. Graphical Elucidations of Different Networks Using Single-Cell Data and Applications to Complex Diseases. Abbreviations: AD, Alzheimer’s disease; CCAN, cis-coaccessibility network; CGN, CRE–gene interaction network; CIN, chromatin interaction network; CL, chromatin locus; CRE, cis-regulatory element; GCN, gene coexpression network; GRN, gene regulatory network; MGRN, methylation-associated gene regulatory network; RMS, rhabdomyosarcoma; T2D, type 2 diabetes; TCN, TF–CRE interaction network; TF, transcription factor.

process in complex diseases. By comparing HRLs across different cell populations, the discovery of the causal genetic patterns driving the cell developmental trajectories can assist in the prediction of prognosis and treatment responses for complex diseases. Besides, HRL diversity between cell populations can be exploited to guide the identification of some rare cell populations associated with complex diseases [67].

HRL-Associated Network Inference from Single-Cell Omics Data in Complex Diseases

All aspects of life activities or complex diseases in living cells are mediated by various regulatory mechanisms that are jointly characterized by an HRL. HRL-associated networks can then be built to delineate the interplays between molecules within the HRL. These networks, including GCNs [68], GRNs [42], CCANs [69], MGRNs [70], CINs [71], CGNs [72], and TCNs [73], were originally defined at tissue level and then extended to individual cells in single-cell data. Some CCANs seek to exploit the coaccessibility between promoters and are called promoter–promoter coaccessibility networks [74]. CINs can be further classified into promoter–promoter interaction networks and promoter–enhancer interaction networks [75]. Rapid advances in single-cell sequencing technologies have provided exciting opportunities for construction of HRL-associated networks and applications to complex diseases. Taking leukemia as an example, five HRL-associated networks have been applied, namely, GCNs, GRNs, MGRNs, CGNs, and TCNs (Figure 1). Specifically, the GCN of acute myeloid leukemia (AML) discovered aberrant coexpression between stemness and myeloid priming genes in primitive AML cells [76]. SINCERITIES was utilized to infer the GRN driving the differentiation of monocytic THP-1 human myeloid leukemia cell into macrophages [47]. The MGRN constructed in chronic lymphocytic leukemia (CLL) suggests that conserved CpG sites are protected from DNA methylation by TF binding through direct exclusion of methylases or negative selection caused by disrupted CLL regulatory code [58]. CGNs were leveraged to detect putative *cis*-regulatory linkages of genes specific to mixed-phenotype acute leukemia (MPAL) [64]. Meanwhile, TCNs were inferred to identify TFs that regulate the leukemia genes [64]. Besides, CGNs were also applied to CLL and discovered several cell clusters enriched for enhancers specific to CLL cells [65], and TCNs were constructed to identify a consistent regulatory program in CLL cells [65]. In the following subsections, we provide an overview of the available tools for HRL-associated network inference. Around 21 examples were collected to elucidate seven HRL-associated networks in Table 1, covering the details such as networks, diseases, sequencing methods, number of cells, and accession IDs.

GCN Inference in Prostate Cancer, Melanoma, and AML from scRNA-Seq

The inference of GCNs (Figure 2A) is advancing rapidly and is generating many new insights into the prediction of prognosis and treatment responses of prostate cancer [77], AML [76], and melanoma [78]. For example, Chen and colleagues used GCNs to identify key biological processes regulated by the androgen receptor, which is essential to the development of normal and cancer prostate gland [77]. Galen and collaborators discovered aberrant coexpression between stemness and myeloid priming genes in primitive AML cells, unveiling dysregulated transcriptional programs in malignant cells [76]. Peng and colleagues constructed a localized GCN using a subset of the cells from a melanoma patient-derived scRNA-Seq dataset and then used it to predict survival rates [78].

Until now, a few tools have been developed to infer GCNs, among which LEAP, PPCOR, and SCIMITAR are three that are widely used [39–41]. LEAP is an R package that utilizes the pseudotime of the cells to find gene coexpression involving time delay from a scRNA-Seq dataset [39]. Harly and collaborators applied LEAP to identify the coexpressed genes alongside the development of innate lymphoid cells [79], an important immune cell in diseases such as asthma,

Table 1. Examples of HRL-Associated Network Inference and Applications in Complex Diseases

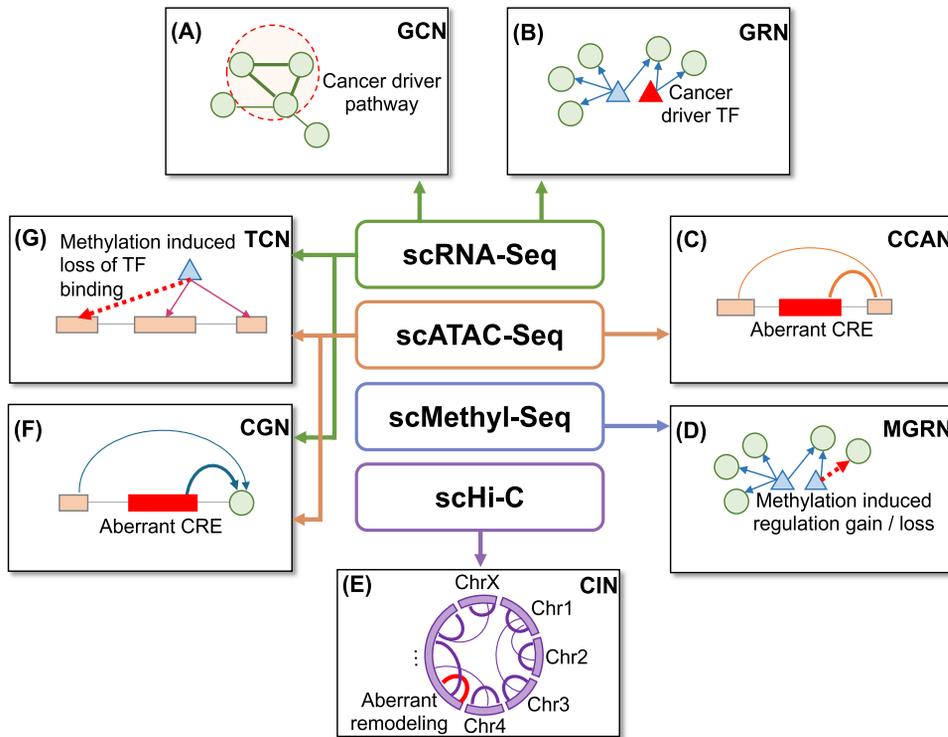
Network	Disease	Seq method	Cell no.	Tool	Main contribution	Accession	Refs
Applications to complex diseases							
GCN	Prostate cancer	scRNA-Seq	144	WGCNA	Identify key biological processes regulated by the androgen receptor, which is essential to the development of normal and cancer prostate glands.	GSE99795	[77]
	AML	scRNA-Seq	38 410	—	Discovered aberrant coexpression between stemness and myeloid priming genes in primitive AML cells, indicating dysregulated transcriptional programs in malignant cells.	GSE116256	[76]
	Melanoma	scRNA-Seq	4645	COAC	Constructed a localized GCN to predict the survival rate of melanoma patients.	https://github.com/ChengF-Lab/COAC	[78]
GRN	RMS	scRNA-Seq	9970	—	Identified driver TFs, e.g., SOX8, underlying childhood RMS by dissecting the GRN.	GSE116344	[80]
	T2D	scRNA-Seq	2491	—	Discovered the altered GRN architecture in the pancreas from T2D patients.	http://sandberg.cmb.ki.se/pancreas/	[52]
	AD	scRNA-Seq	22 951	—	Identified rewiring of microglia gene regulation of AD by comparing the characteristics of GRNs in AD and controls.	GSE98971	[52]
	Melanoma	scRNA-Seq	4645	ACTION	Use cell-type-specific GRN to identify several novel biomarkers (e.g., CCND1 and SNAI2) and TF-gene linkages (e.g., MITF-CTSK), which may be associated to melanoma.	Upon reasonable request	[81]
	Melanoma	scRNA-Seq	674	GENIE3	Infer GRN from melanoma scRNA-Seq data and identified TFs such as SOX10 and RXRG as central hubs governing the neural crest stem cell states.	GSE116237	[82]
	Luminal breast cancer	scRNA-Seq	9979	PIDC	Leveraged PIDC to identify the regulatory relationships between genes from the scRNA-Seq profiles of CD44 ^{high} and CD44 ^{low} cells in luminal breast cancer.	GSE122743	[83]
	Uveal melanoma	scRNA-Seq	8598	GRNBoost2	Utilized GRNBoost2 to infer GRN from scRNA-Seq profiles of uveal melanoma tumor cells, and identified gene modules together with their regulatory motifs.	GSE139829	[84]
	Myeloid leukemia	scRNA-Seq	960	SINCERITIES	Used SINCERITIES to infer the GRN driving the differentiation of monocytic THP-1 human myeloid leukemia cell into macrophages.	https://cabsel.ethz.ch/tools/sincerities.html	[47]
CCAN	T2D	scATAC-Seq	1456	Cicero	Utilized Cicero to nominate the cell-specific target genes at 104 noncoding T2D genome-wide association studies signals.	dbGap: phs001188.v2.p1	[85]
	AD	scATAC-Seq	~100 000	Cusanovich2018	Identify cell-type-specific enrichments of the heritability signal for hundreds of complex traits and diseases, such as heritability for AD in microglia.	GSE111586	[56]

(continued on next page)

Table 1. (continued)

Network	Disease	Seq method	Cell no.	Tool	Main contribution	Accession	Refs
	Heart diseases	scATAC-Seq	~80 000	JRIM	Reported several mutations associated with tissue-related heart diseases such as myocardial hypoplasia, defective atrial, and congenital heart defects.	GSE111586	[57]
MGRN	CLL	scMethyl-Seq	1821	—	Suggested that conserved CpG sites are protected from DNA methylation by TF binding through direct exclusion of methylases or negative selection caused by disrupted CLL regulatory code.	GSE109085	[58]
	Breast cancer	scMethyl-Seq	—	iRegulon	Found that stem- and proliferation-related TFs are transcriptionally active in circulating tumor cells, which correlates with a poor prognosis in breast cancer.	ENA: PRJEB25101	[59]
CIN	Breast cancer	scHi-C	—	—	Utilized CIN to identify 110 putative target genes, (e.g., PEX14 and DLX2), from interaction peaks within risk loci.	PRJEB23968	[87]
CGN/TCN	CLL	scRNA-Seq, ATAC-Seq	43 049	—	Identified a consistent regulatory program starting with a sharp decrease of NF- κ B binding in CLL cells, which is followed by reduced activity of lineage-defining transcription factors, erosion of CLL cell identity, and acquisition of a quiescence-like gene signature.	GSE111015	[65]
	MPAL	CITE-Seq, scATAC-Seq	35 882	Seurat, Cicero	Identified 91 601 peak-to-gene links, and inferred RUNX1 as a potential oncogene of MPAL.	GSE139369	[64]
Other applications using single-cell data							
GCN	—	scRNA-Seq	466	SCIMITAR	Revealed 92 progression-associated genes and uncovered three coregulatory states from human fetal neurons.	GSE67835	[40]
GRN	—	scRNA-Seq	376	GRNVBEM	Identified 491 cell type marker genes for each of the four sample groups associated with low and high expression of CD41 in zebrafish, and detected two hub genes, slc2a6 and csf3r.	ArrayExpress: E-MTAB-3947	[48]
	—	scRNA-Seq	148–5069, 49–8522	IRIS3	Demonstrated superior performance compared with SCENIC on 19 scRNA-Seq datasets as well as reproducibility and robustness.	https://bmbi.bmi.osumc.edu/iris3/	[54]
CIN	—	scHi-C	10 696, 180	scHiCluster	Identified TADs within single cells after imputation, and allowed single-cell clustering.	GSE84920, GSE80006	[60]

autoimmune diseases, allergic rhinitis, and inflammatory bowel disease. PPCOR is an R package used for computing the partial/semipartial correlations between gene expressions relative to the other genes [41]. SCIMITAR is a Python package to infer GCN dynamics throughout the biological progression from single-cell transcriptomes [40]. Although the application value of PPCOR and SCIMITAR to complex diseases remains unexplored, they possess potential for the future.



Trends In Genetics

Figure 2. Illustration of Each Type of HRL-Associated Network in the Context of Evaluating Cancer Heterogeneity.

(A) A GCN is an undirected graph where each node represents a gene, and a pair of genes are connected by an edge if there is a significant coexpression relationship between them based on available scRNA-Seq data. (B) In a GRN, each edge is directed from the regulator to the target gene corresponding to activation or inhibition. (C) A CCAN is an undirected graph, where each node represents a CRE, and a pair of CREs are connected by an edge if there is a significant coaccessibility relationship between them based on available scATAC-Seq data. (D) An MGRN is a GRN where regulator-gene regulatory relationships are influenced by methylation states. (E) A CIN is an undirected graph, where each node represents a CL, and a pair of loci are connected by an edge if a significant interaction exists between them based on scHi-C data. (F) A CGN is a directed graph with nodes representing CREs and genes and edges corresponding to the *cis*-regulatory relationships between the two kinds of nodes. (G) A TCN is a directed graph, where nodes denote TFs and CREs, and edges represent the binding activities of TFs to CREs. Abbreviations: CCAN, *cis*-coaccessibility network; CGN, CRE-gene interaction network; CIN, chromatin interaction network; CL, chromatin locus; CRE, *cis*-regulatory element; GCN, gene coexpression network; GRN, gene regulatory network; HRL, heterogeneous regulatory landscape; MGRN, methylation-associated gene regulatory network; scATAC-Seq, single-cell ATAC sequencing; scHi-C, single-cell Hi-C; scRNA-Seq, single-cell RNA sequencing; TCN, TF-CRE interaction network; TF, transcription factor.

GRN Inference in Rhabdomyosarcoma, Luminal Breast Cancer, Myeloid Leukemia, Melanoma, Type 2 Diabetes, and AD from scRNA-Seq

The applications of GRNs (Figure 2B) mainly aim at rhabdomyosarcoma (RMS) [80], melanoma [81,82], type 2 diabetes (T2D) [52], luminal breast cancer [83], myeloid leukemia [47], and AD [52]. For instance, by dissecting the GRN, Gryder and collaborators identified the driver TFs (e.g., SOX8) underlying childhood RMS [80]. Iacono *et al.* discovered the altered GRN architecture in the pancreas from T2D patients [52]. In addition, they also identified the rewiring of microglia gene regulation of AD by comparing the characteristics of GRNs in AD and controls [52]. Mohammadi *et al.* reconstructed a cell-type-specific GRN for the newly discovered melanoma subtype from MITF-associated melanoma patients [81]. The GRN highlighted several novel biomarkers (e.g., CCND1 and SNAI2) and TF-gene linkages (e.g., MITF-CTSK), potentially associated with melanoma.

Presently, many GRN inference efforts have been performed via developing tools such as CSN [51], PIDC [50], GENIE3 [42], GRNBoost2 [84], Jump3 [44], SCODE [45], SINCERTIES [47], GRNVBEM [48], SCNS [49], and IRIS3 [54]. GENIE3 is a widely-used Bioconductor package for GRN inference [42], which has been integrated into several scRNA-Seq analysis pipelines, for example, SCENIC and Arboreto [43,46]. Rambow *et al.* used GENIE3 to infer GRN from melanoma scRNA-Seq data and identified TFs such as SOX10 and RXRG as central hubs governing the neural crest stem cell states [82]. PIDC is a Julia package developed based on multivariate information theory [50]. Hong *et al.* leveraged PIDC to identify the regulatory relationships between genes from the scRNA-Seq profiles of CD44^{high} and CD44^{low} cells in luminal breast cancer [83]. GRNBoost2 is an efficient GRN inference tool integrated into the Python package, Arboreto [43]. Using GRNBoost2, Durante and colleagues inferred GRN from scRNA-Seq profiles of uveal melanoma tumor cells and identified gene modules together with their regulatory motifs [84]. SINCERTIES was designed based on linear regression and temporally dynamic changes in expression and implemented using R and MATLAB, respectively [47]. The efficacy of SINCERTIES was demonstrated by identifying GRNs driving the differentiation of monocytic THP-1 human myeloid leukemia cells into macrophages [47]. IRIS3 is an integrated web server for cell-type-specific regulon (CTSR) prediction from human or mouse scRNA-Seq data [54]. Based on the reasoning that a CTSR can reliably characterize and distinguish the cell types, these CTSRs can aid in the elucidation of regulatory mechanisms. Albeit efforts to apply the other GRN inference tools, such as CSN, Jump3, SCODE, GRNVBEM, SCNS, and IRIS3, to complex diseases are still in their infancy, they will be essential for a deeper understanding of complex diseases.

CCAN Inference in AD and T2D Based on scATAC-Seq

Lately, applications of CCANs (Figure 2C) aim to provide insight into the regulatory mechanisms underlying AD [56] and T2D [85]. To date, several tools have been proposed to infer CCANs, including but not limited to, Cicero [55], Cusanovich2018 [56], and JRIM [57]. Cicero is an R package to predict *cis*-regulatory interactions from scATAC-Seq data [55]. Rai *et al.* utilized Cicero to nominate the cell-specific target genes at 104 noncoding, T2D genome-wide association studies signals [85]. Cusanovich 2018 focuses on identifying cell types, defining candidate tissue-specific enhancers, modeling TF regulatory grammar that specifies each cell type, and linking distal CREs to promoters of target genes [56]. Cusanovich2018 was used to identify cell-type-specific enrichments of the heritability signal for hundreds of complex traits and diseases, such as heritability for AD in microglia [56]. JRIM is an R package based on Cicero, aiming to investigate common and specific regulatory mechanisms across different cell types [57]. Based on the changes in promoter-associated interactions predicted by JRIM, several mutations were reported to be associated with tissue-related heart diseases such as myocardial hypoplasia, defective atrial, and congenital heart defects [57].

MGRN Inference in Breast Cancer and CLL Based on scMethyl-Seq

DNA methylation is involved in various cellular processes, including X chromosome inactivation, genomic imprinting, and silencing of transposable elements [86], whereas aberrant DNA methylation relates to several diseases, for example, breast cancer [59] and CLL [58]. MGRNs (Figure 2D) have been applied to the prediction of prognosis and treatment responses of breast cancer and CLL [58,59]. For example, Gaiti *et al.* utilized MGRNs to explore the impact of methylation on transcriptional activity and cellular phenotypes in CLL [58]. Experimental results suggested that conserved CpG sites are protected from DNA methylation by TF binding through the direct exclusion of methylases or negative selection caused by disrupted CLL regulatory code. Using MGRNs, Gkoutela and collaborators found that stem- and proliferation-related TFs are transcriptionally active in circulating breast cancer cells [59]. Though many single-cell

scMethyl-Seq techniques are developed, they all suffer from low sequence coverage and robustness, as well as high costs and handling time [86]. Hence, these MGRN inference tools have not been applied to complex diseases, yet it is possible that advancing of sequencing technologies will bring wide application space for these tools.

CIN Inference in Breast Cancer from Single-Cell Hi-C

Clinical applications of CINs (Figure 2E) focus mainly on breast cancer [87]. For example, Baxter and colleagues utilized CIN to identify 110 putative target genes, for example, PEX14 and DLX2, from interaction peaks within risk loci [87]. Presently, several tools have been proposed to infer CINs, for example, scHiCluster [60]. scHiCluster is an R package for clustering on Hi-C contact matrices based on imputations using convolution and random walk. After imputation using scHiCluster, TAD can be identified in single cells [60]. Although there is no study that applied scHiCluster to complex diseases, scHiCluster lays a basis for future advances.

Meanwhile, two HRL-associated networks, CGNs (Figure 2F) and TCNs (Figure 2G), are inferred from multiple single-cell modalities rather than from single modalities alone. Applications of CGNs and TCNs mainly concentrate on leukemia (Table 1) [64,65]. A CGN was used to identify 91 601 putative *cis*-regulatory linkages of the leukemia-specific genes, such as marker gene CD69, in MPAL [64]. Besides, several TFs (e.g., RUNX1) that regulate the leukemia genes were also discovered through the inference of TCNs [64]. Additionally, CGNs identified four clusters in CLL, which are enriched for enhancers specific to CLL cells and/or B cells, B cells and T cells, NF- κ B binding sites, and transcribed regions marked by H3K36me3 in hematopoietic cells, respectively [65]. Meanwhile, TCNs discovered a consistent regulatory program starting with a sharp decrease of NF- κ B binding in CLL cells, which is followed by reduced activity of lineage-defining TFs, erosion of CLL cell identity, and acquisition of a quiescence-like gene signature [65]. More examples of network inference from scMulti-omics are discussed in the next section and shown in Table 2.

Generally, network inference methods relying on correlation calculation (e.g., semipartial correlation and covariance) act as the main forces in networks including GCNs, CCANs, and MGRNs. The GCN inference methods, WGCNA [77], COAC [78], SCIMITAR [40], LEAP [39], and PPCOR [41], are all developed based on correlations. Similarly, the three CCAN inference methods, Cicero [55], JRIM [57], and Cusanovich2018 [56], also depend on covariance, a kind of correlation metric. GRN inference encompasses a variety of methods, such as regression models (GENIE3, GRNBoost2, and SINCERITIES) [42,43,47], information theory (PIDC) [50], and Bayesian models (GRNVBEM) [48]. Furthermore, SCENIC and IRIS3 reconstruct GRNs from scRNA-Seq data by integrating TFs to coexpressed gene modules [46,54]. Likewise, iRegulon infers MGRN via associating TFs to a gene list [59]. Chromatin interactions in CINs can be captured directly by scHi-C [60].

Integration of scMulti-Omics Data for HRL Construction

Although these HRL-associated networks yield informative insights into cellular diversity and disease development, they measure only particular aspects of cellular identity, motivating the need to jointly build an overall HRL based on these HRL-associated networks (Figure 3, Key Figure).

Indeed, scMulti-omics technologies encompass the diverse characterization of a single-cell's DNA methylation, chromatin accessibility, RNA expression, protein abundance, gene perturbation, and even spatial information [66]. Correspondingly, a few studies have focused on the integration of different single-cell technologies, allowing for a deeper understanding of cellular reprogramming in complex diseases. For example, Vickovic *et al.* developed a new single-cell technology by combining

Table 2. Examples of scMulti-Omics Integration Tools and Applications in Complex Diseases

Seq methods	Input ^a	Disease	Cell no.	Tool	Main contribution	Accession	Refs
Applications to complex diseases							
HDST	RE+SI	Breast cancer	50 000	—	Distinguished cell types and niches from a breast cancer resection.	Single cell portal	[88]
CITE-Seq	RE+ES	Salivary gland squamous cell carcinoma	~23 000	—	Identified two subpopulations of cancer stem and basal cells, and four macrophage subpopulations.	GSE124425	[89]
Other applications using single-cell data							
scRNA-Seq, STARmap	RE+SI	—	71 000 + 2500	LIGER	Spatially located fine subtypes of cells present in the mouse frontal cortex and predicted even complex spatial expression patterns across many individual genes.	GSE126836	[63]
scNMT-Seq	RE+CA+DM	—	1828	MOFA+	Suggested that independent cell lineages are characterized by different epigenetic variations, such as global changes in ExE endoderm, and local patterns in embryonic endoderm and mesoderm.	GSE121708	[91]
scRNA-Seq, scATAC-Seq	RE+CA	—	464 + 415	CoupledNMF	Demonstrated that peaks specific to various mESC cell clusters are enriched in different tissues such as heart, forebrain, and midbrain.	GSE115968, GSE115970	[62]
scRNA-Seq, scATAC-Seq, HiChIP	RE+CA+CI	—	464 + 96	DC3	Simultaneously identify cell subpopulations and deconvolve the bulk data into subpopulation-specific data. Moreover, the accessibility, expression, and loop profiles inferred by DC3 lay a basis for further analyses of the regulatory systems, such as constructing subpopulation-specific gene regulatory networks.	GSE115968, GSE107651, GSE127807	[61]

^aInput data-type combination examples using the corresponding tools: CA, chromatin accessibility; CI, chromatin interactions; DM, DNA methylation; ES, epitope sequencing; PE, protein expression; RE, RNA expression; SI, spatial information.

transcriptome and spatial information and demonstrated its clinical potential by distinguishing cell types and niches in a breast cancer resection [88]. Praktijnjo and collaborators utilized single-cell transcriptome and epitope profiling to identify two subpopulations of cancer stem and basal cells, and four macrophage subpopulations [89]. By combining flow cytometry, scRNA-seq, and ATAC-Seq, Rendeiro and colleagues identified a consistent regulatory program starting with a sharp decrease of NF-κB binding in CLL cells, which is followed by reduced activity of lineage-defining TFs, erosion of CLL cell identity, and acquisition of a quiescence-like gene signature [65].

Several integration tools, such as Seurat [90], MOFA+ [91], LIGER [63], CoupledNMF [62], and DC3 [61], have been developed and even applied to scMulti-omics data of complex diseases. Seurat is a multifunctional R package for scRNA-Seq and scMulti-omics analysis based on canonical correlation analysis [90]. Granja and colleagues integrated scRNA-Seq and scATAC-Seq data using Seurat and inferred RUNX1 as a potential oncogene of MPAL [64]. LIGER is an R platform allowing for the integration of gene expression, epigenetic regulation, and spatial relationships across single-cell datasets [63]. Experiments on an unmatched scRNA-Seq and DNA methylation dataset showcased its ability to identify methylation regions anticorrelated with Arx expression [66]. MOFA+ is an R/Python package for comprehensive and scalable integration of single-cell multimodal data [91]. MOFA+ was applied to a scNMT-Seq dataset in which RNA expression, chromatin accessibility, and DNA methylation were assayed [91]. Experimental results

Key Figure

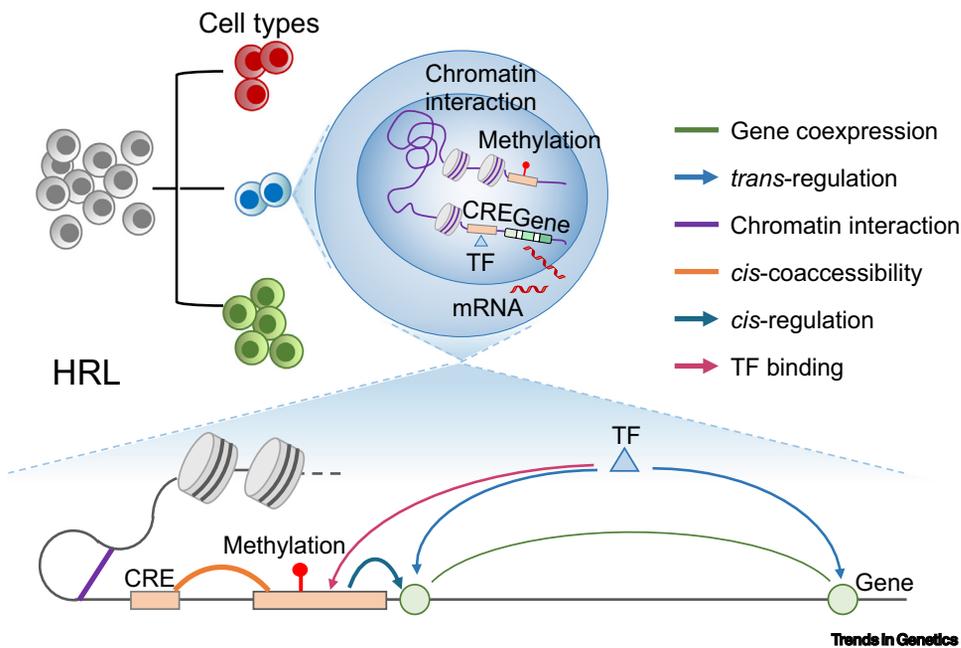
Towards HRL Construction from the Integration of **scMulti-omics**

Figure 3. Abbreviations: CRE, cis-regulatory element; HRL, heterogeneous regulatory landscape; scMulti-omics, single-cell multimodal omics; TF, transcription factor.

demonstrated that independent cell fate commitment events undergo different modes of epigenetic variations, such as global changes in ExE endoderm and local patterns in embryonic endoderm and mesoderm. CoupledNMF is a Python tool to integrate gene expression and chromatin accessibility that are not measured on the same cell [62]. CoupledNMF enables a systematic mapping of CREs to genes and TFs to CREs, informative for downstream analyses such as inferring cell-cluster-specific regulatory networks at the single-cell level. The cell-cluster-specific gene expression profiles and chromatin accessibility profiles computed by CoupledNMF derived useful insights into the constituent subpopulations of mouse embryonic stem cells (mESCs) [62]. The results showed that peaks specific to various subpopulations are enriched in different tissues such as heart, forebrain, and midbrain. DC3 is a Python package for deconvolution and coupled clustering from bulk and single-cell genomics data (e.g., RNA-Seq, ATAC-Seq, and HiChIP) [61]. Similar to CoupledNMF, DC3 can also construct cell-cluster-specific regulatory networks based on the identified TF–CRE–gene triplets [61]. The experiment on a mixture of scRNA-Seq, scATAC-Seq, and bulk HiChIP from mESCs demonstrated that DC3 can deconvolve bulk profiles into subpopulation-specific profiles, and meanwhile the subpopulation-specific profiles, in turn, leads to improved coupled clustering results of single-cell data [61]. Details of applications of scMulti-omics to complex diseases can be found in Table 2, including sequencing methods, diseases, number of cells, contributions, and accession IDs.

Remaining Challenges

Development of single-cell sequencing technologies and their applications in complex diseases has been astonishing in the past decade, but many challenges still exist and much remains to be

explored. As new techniques arise to measure distinct cellular modalities, the paramount challenge is to integrate these datasets to better understand cellular identity and regulatory mechanisms. Because each technology measures only partial aspects of cellular identity, and noise may be introduced during cell isolation, most analyses merely focus on the clustering and subtyping of cells. Therefore, there is a pressing need to leverage information in one dataset to improve the interpretation of another as examples in bulk data [72,92,93]. PECA is a statistical approach to infer transcriptional regulatory networks from paired expression and chromatin accessibility data across diverse cellular contexts [92]. The networks inferred by PECA provide a detailed view of how *trans*- and *cis*-regulatory elements work together to affect gene expression in a context-specific manner. TimeReg is a framework for time-course regulatory analysis based on paired expression and chromatin accessibility data from the time course, which can be used to prioritize CREs, to extract core regulatory modules, to identify key regulators driving changes of cellular states, and to casually connect the modules across different time points [72]. One solution to this challenge is to transfer the pretrained models at the tissue level to the single-cell level. Another solution is to resort to **artificial intelligence (AI)**, for example, multimodal learning models [94].

Moreover, single-cell omics data typically suffer from technical issues including sparsity, biased cell subpopulations, and false-positive variants [23,34,95]. Sparsity mainly originates from dropout, low-throughput, and low coverage, and cell subpopulation biases are introduced during single-cell isolation, and false-positive variants result from the amplification or sequencing. These issues may lead to missing nodes or edges in constructed networks. One solution to deal with sparsity is to use bulk omics data to impute single-cell omics data. Typical tools for scRNA-Seq imputation include MAGIC [96], SCRABBLE [97], and SAVER-X [98]. Comparing the variant alleles discovered in the single-cells with the bulk population could ensure there were no biases in cell subpopulations. Besides, using the bulk sample as a reference can reduce false-positive variants (e.g., SAVER-X).

Furthermore, batch effects exist when integrating single-cell omics data across samples, experiments, and modalities. These effects result from varying resolution levels, the uncertainty of any measurements, and scaling of single-cell methodology to more cells and more features measured at once. Batch effects may thus influence the stability of inferred networks. One way is to develop techniques to measure different modalities within the same cells. Another way is to design tools to harmonize the modalities measured in different cells into a single reference, for example, Seurat [90].

Finally, the construction of networks depends on many factors such as data type, network complexity and topology, and tool settings. Evaluating the reliability of the predicted networks remains a challenge. One common solution is to crossvalidate the linkages in the network with known interactions from literature. However, this method may be inaccurate in evaluating negative interactions and global interactions, as their information is difficult to obtain. **Perturb-Seq**, a sequencing method combining CRISPR-mediated genes with scRNA-Seq, can be performed to curate TF–gene relationships, allowing for validation of the stability and plasticity of networks [99]. In addition, the functional homogeneity of network modules could be estimated by enrichment analyses against gene ontology terms, pathways, and motifs [40,46,48,54,56]. However, Perturb-seq and scMulti-omics crossvalidation requires special sequencing techniques that are not suitable for networks predicted from a single dataset.

Concluding Remarks

Exploring single-cell gene regulatory mechanisms provides the opportunity to uncover the unique malfunctions associated with complex diseases in individual cells. To elucidate the roles of TFs, genes, CREs, and chromatin loci in gene regulation, HRL-associated networks (GCNs, GRNs,

Outstanding Questions

How can we bridge different HRL-associated networks and cell–cell networks through data integration and modeling approaches?

Will it be feasible to bridge basic cell/molecular biology models to clinically predictive models in complex diseases?

Are there any data-driven ways to guide sample collection to model dynamic time-series networks?

How can we build networks based on all the different modalities?

How can we extend systems biology concepts to the tissue level?

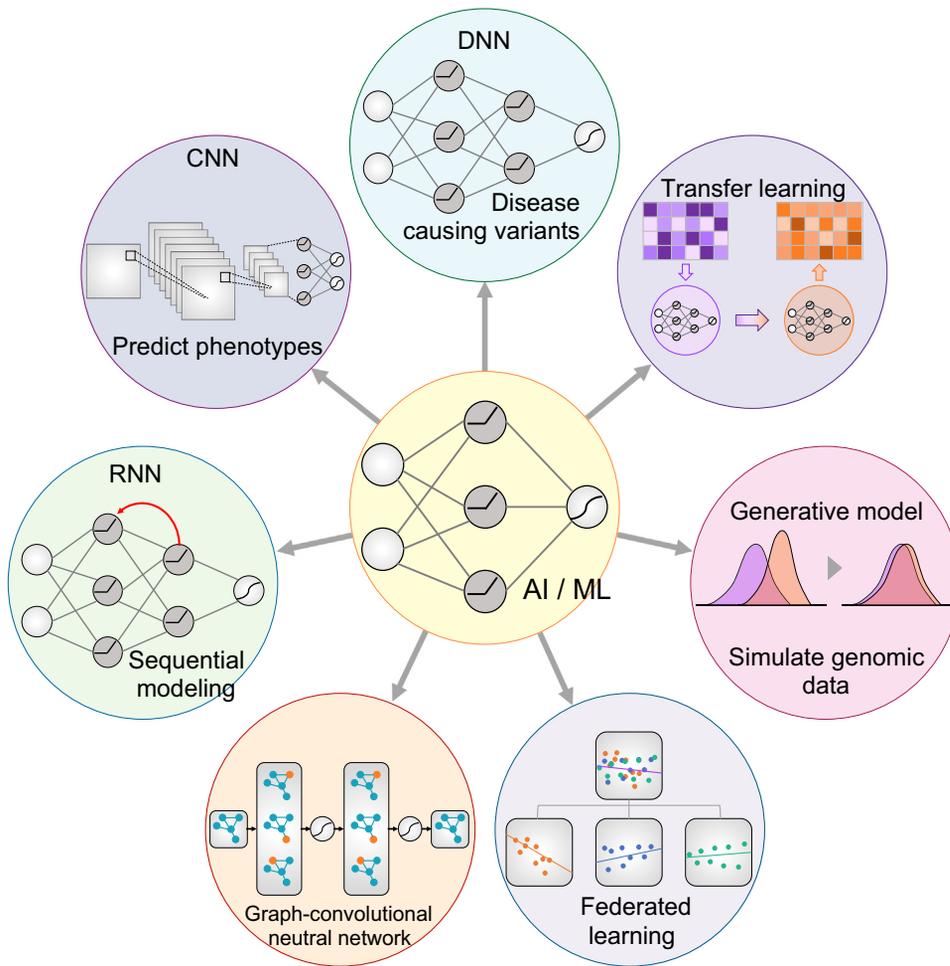
What other information can be used to interpret the networks in addition to single-cell omics data?

How can high throughput perturbation experiments (e.g., CRISPR, single cell) be used to inform and validate network and system information?

CCANs, MGRNs, CINs, CGNs, and TCNs) are constructed, which jointly build an HRL to characterize the multilevel genotype and phenotype observations across diverse cell types. We raise several questions to be answered regarding HRL in complex diseases using single-cell data (see Outstanding Questions). Meanwhile, several future trends in HRL in complex diseases are discussed below.

One clear trend is the increasing significance of HRL in complex diseases. HRL is informative in identifying the TFs, genes, CREs, DNA methylation states, and chromatin architectures that possibly link to complex diseases. First, HRL construction will be further refined by integrating other information, such as spatial, time-series, and treatment, leading to dynamic HRL alongside time-series. Besides, HRL reconstruction could promote the delineation of cell–cell differences via building cell–cell interaction networks. Uncovering the HRL by inferring cell-type-specific networks will provide biological insights into the development, evolution, and pathology of organisms, and contribute to the prognosis and treatment of diseases. Lastly, the integration of bulk omics and single-cell omics data will be an important strategy in support of HRL elucidation, for example, imputing gene expression from scRNA-Seq using bulk RNA-Seq data [66]. This topic will attract considerable attention to the advances in multimodal omics and the fast development of AI.

Another trend is the increasing application of AI and machine learning (ML) methods in single-cell omics analyses (Figure 4), especially deep learning (DL) models [94]. The capability to extract new insights from the exponentially increasing volume of genomics data requires more expressive ML models. By effectively leveraging large datasets, DL can capture dependencies in sequential and graph-structured data using various models such as deep neural networks (DNNs), recurrent neural networks (RNNs), convolutional neural networks (CNNs), and graph-convolutional neural networks [94]. DNNs are mainly used to model complex dependencies, which include predicting the percentage of spliced exons, discovering TF binding motifs, and prioritizing potential disease-causing genetic variants [94]. Besides, DNNs can be easily extended to take multiple modalities as inputs in order to leverage complementary information from between them [94]. Applications of CNNs include discovering local patterns in sequential data, for example, predicting molecular phenotypes, classifying TF binding sites, denoising ChIP-Seq, and enhancing Hi-C data resolution [94]. The CNN model, CNNC, has been proposed for supervised gene relationship inference, which improves upon prior methods in tasks ranging from predicting TF targets to identifying disease-related genes to causality inference [100]. RNNs have been used to model long-range dependencies in sequences, for example, predicting TF binding sites or RNA binding proteins, detecting methylation states, and discovering miRNA targets [94]. Tasks that graph-convolutional neural networks can be trained for include modeling dependencies in graph-structured data, such as deriving features from protein–protein interaction networks, modeling polypharmacy side effects, and predicting binarized gene expressions [94]. In the scenario where data are scarce, training a model from scratch might be infeasible. Transfer learning (TL) can be used to initiate the model with the majority of parameters from another model trained on a similar task [94]. The utility of TL has been demonstrated for sequence-based predictive models of chromatin accessibility by training models in other cell types [94]. In federated learning (FL), model instances are deployed on distinct sites and trained on local data to optimize a global model [94]. By avoiding data transfer among devices, FL could attain high efficiency and facilitate the respect of medical data privacy. FL has achieved advantageous performance in predicting hospitalization due to heart diseases within a calendar year based on patients' electronic health records prior to that year [101]. Moreover, generative models (e.g., generative adversarial network) could be used to simulate clinical data that can be analyzed by others without privacy violation [94]. Finally, to make the inner working of ML models understandable to humans, three strategies for interpretable ML have been proposed [102]: (i) measuring how changes in



Trends in Genetics

Figure 4. The power of AI and ML, Especially Graph-Convolutional Neural Networks, in the Prediction of Network Models. Abbreviations: AI, artificial intelligence; CNN, convolutional neural network; DNN, deep neural network; ML, machine learning; RNN, recurrent neural network.

the input data impact model predictions (perturbing strategy); (ii) training an interpretable model using the same data to approximate the predictions of the to-be-interpreted model (surrogate strategy); and (iii) inspecting the structure and parameters in a trained model (probing strategy).

Although DL models have showcased outstanding performances in this field, potential limitations remain. For example, the performance of DL models strongly depends on the choice of models and hyperparameters [94]. The increase in the number of parameters might pose an overfitting challenge. One solution is to develop an effective method for hyperparameter optimization. Besides, some models, for example, RNNs, are time-consuming and cannot be easily parallelized [94]. This limitation might be overcome by designing more efficient DL models.

Acknowledgments

This work was supported by R01-131399, P30-CA016058, and U24-CA252977 from the NIH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation and the NIH.

References

- Tanay, A. and Regev, A. (2017) Scaling single-cell genomics from phenomenology to mechanism. *Nature* 541, 331–338
- Packer, J. and Trapnell, C. (2018) Single-cell multi-omics: an engine for new quantitative models of gene regulation. *Trends Genet.* 34, 653–665
- Li, R. *et al.* (2015) A gene regulatory program in human breast cancer. *Genetics* 201, 1341–1348
- Tommasini-Gheffi, S. *et al.* (2019) Cancer-associated mutation and beyond: The emerging biology of isocitrate dehydrogenases in human disease. *Sci. Adv.* 5, eaaw4543
- Buschur, K.L. *et al.* (2019) Causal network perturbations for instance-specific analysis of single cell and disease samples. *Bioinformatics* 36, 2515–2521
- Malik, B. *et al.* (2019) Gene expression analysis reveals early dysregulation of disease pathways and links Chmp7 to pathogenesis of spinal and bulbar muscular atrophy. *Sci. Rep.* 9, 3539
- Jadiya, P. *et al.* (2019) Impaired mitochondrial calcium efflux contributes to disease progression in models of Alzheimer's disease. *Nat. Commun.* 10, 3885
- Needham, E.J. *et al.* (2019) Illuminating the dark phosphoproteome. *Sci. Signal* 12, eaau8645
- Kaneshwaran, K. *et al.* (2019) Sleep fragmentation, microglial aging, and cognitive impairment in adults with and without Alzheimer's dementia. *Sci. Adv.* 5, eaax7331
- Cardenas, A. *et al.* (2019) The nasal methylome as a biomarker of asthma and airway inflammation in children. *Nat. Commun.* 10, 3095
- Vezzani, A. *et al.* (2019) Neuroinflammatory pathways as treatment targets and biomarkers in epilepsy. *Nat. Rev. Neurol.* 15, 459–472
- Xu, Z. *et al.* (2017) Characterization of serum miRNAs as molecular biomarkers for acute Stanford type A aortic dissection diagnosis. *Sci. Rep.* 7, 13659
- Zetterberg, H. and Bendlin, B.B. (2020) Biomarkers for Alzheimer's disease-preparing for a new era of disease-modifying therapies. *Mol. Psychiatry* Published online April 6, 2020. <https://doi.org/10.1038/s41380-020-0721-9>
- Fitzgerald, P.J. and Watson, B.O. (2018) Gamma oscillations as a biomarker for major depression: an emerging topic. *Transl. Psychiatry* 8, 177
- Li, A. *et al.* (2020) A neuroimaging biomarker for striatal dysfunction in schizophrenia. *Nat. Med.* 26, 558–565
- Sveen, A. *et al.* (2020) Biomarker-guided therapy for colorectal cancer: strength in complexity. *Nat. Rev. Clin. Oncol.* 17, 11–32
- Koppe, L. and Poutou, V. (2016) CMPF: a biomarker for type 2 diabetes mellitus progression? *Trends Endocrinol. Metab.* 27, 439–440
- Nguyen, M.N. *et al.* (2018) Mechanisms responsible for increased circulating levels of galectin-3 in cardiomyopathy and heart failure. *Sci. Rep.* 8, 8213
- Ahsen, M.E. *et al.* (2019) NeTFactor, a framework for identifying transcriptional regulators of gene expression-based biomarkers. *Sci. Rep.* 9, 12970
- Hekselman, I. and Yeger-Lotem, E. (2020) Mechanisms of tissue and cell-type specificity in heritable traits and diseases. *Nat. Rev. Genet.* 21, 137–150
- Hanson, C. *et al.* (2018) Principled multi-omic analysis reveals gene regulatory mechanisms of phenotype variation. *Genome Res.* 28, 1207–1216
- Macaulay, I.C. *et al.* (2017) Single-cell multiomics: multiple measurements from single cells. *Trends Genet.* 33, 155–168
- Gawad, C. *et al.* (2016) Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* 17, 175–188
- Hasin, Y. *et al.* (2017) Multi-omics approaches to disease. *Genome Biol.* 18, 83
- Pratapa, A. *et al.* (2020) Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* 17, 147–154
- Stuart, T. and Satija, R. (2019) Integrative single-cell analysis. *Nat. Rev. Genet.* 20, 257–272
- McGranahan, N. and Swanton, C. (2017) Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* 168, 613–628
- Baslan, T. and Hicks, J. (2017) Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nat. Rev. Cancer* 17, 557–569
- Tirosh, I. *et al.* (2016) Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196
- Cheung, P. *et al.* (2019) Single-cell technologies - studying rheumatic diseases one cell at a time. *Nat. Rev. Rheumatol.* 15, 340–354
- Liu, L. *et al.* (2019) Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. *Nat. Commun.* 10, 470
- Wen, L. and Tang, F. (2018) Boosting the power of single-cell analysis. *Nat. Biotechnol.* 36, 408–409
- Efremova, M. and Teichmann, S.A. (2020) Computational methods for single-cell omics across modalities. *Nat. Methods* 17, 14–17
- Ren, X. *et al.* (2018) Understanding tumor ecosystems by single-cell sequencing: promises and limitations. *Genome Biol.* 19, 211
- Papalexi, E. and Satija, R. (2018) Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* 18, 35–45
- Crow, M. and Gillis, J. (2018) Co-expression in single-cell analysis: saving grace or original sin? *Trends Genet.* 34, 823–831
- Bartlett, T.E. *et al.* (2017) Single-cell co-expression subnetwork analysis. *Sci. Rep.* 7, 15066
- Lamere, A.T. and Li, J. (2019) Inference of gene co-expression networks from single-cell RNA-sequencing data. *Methods Mol. Biol.* 1935, 141–153
- Specht, A.T. and Li, J. (2017) LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering. *Bioinformatics* 33, 764–766
- Cordero, P. and Stuart, J.M. (2017) Tracing co-regulatory network dynamics in noisy, single-cell transcriptome trajectories. *Pac. Symp. Biocomput.* 22, 576–587
- Kim, S. (2015) ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. Methods* 22, 665–674
- Huynh-Thu, V.A. *et al.* (2010) Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 5
- Moerman, T. *et al.* (2019) GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* 35, 2159–2161
- Huynh-Thu, V.A. and Sanguinetti, G. (2015) Combining tree-based and dynamical systems for the inference of gene regulatory networks. *Bioinformatics* 31, 1614–1622
- Matsumoto, H. *et al.* (2017) SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics* 33, 2314–2321
- Aibar, S. *et al.* (2017) SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* 14, 1083–1086
- Papili Gao, N. *et al.* (2018) SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. *Bioinformatics* 34, 258–266
- Sanchez-Castillo, M. *et al.* (2018) A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. *Bioinformatics* 34, 964–970
- Woodhouse, S. *et al.* (2018) SCNS: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data. *BMC Syst. Biol.* 12, 59
- Chan, T.E. *et al.* (2017) Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.* 5, 251–267 e3
- Dai, H. *et al.* (2019) Cell-specific network constructed by single-cell RNA sequencing data. *Nucleic Acids Res.* 47, e62
- Iacono, G. *et al.* (2019) Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol.* 20, 110
- Moignard, V. *et al.* (2015) Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* 33, 269–276
- Ma, A. *et al.* (2020) IRIS3: integrated cell-type-specific regulon inference server from single-cell RNA-Seq. *Nucleic Acids Res.* 48, W275–W286

55. Pliner, H.A. *et al.* (2018) Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* 71, 858–871 e8
56. Cusanovich, D.A. *et al.* (2018) A single-cell atlas of *in vivo* mammalian chromatin accessibility. *Cell* 174, 1309–1324 e18
57. Dong, K. and Zhang, S. (2020) Joint reconstruction of cis-regulatory interaction networks across multiple tissues using single-cell chromatin accessibility data. *Brief. Bioinform.* Published online July 31, 2019. <https://doi.org/10.1101/721290>
58. Gaiti, F. *et al.* (2019) Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature* 569, 576–580
59. Gkoutela, S. *et al.* (2019) Circulating tumor cell clustering shapes DNA methylation to enable metastasis seeding. *Cell* 176, 98–112 e14
60. Zhou, J. *et al.* (2019) Robust single-cell Hi-C clustering by convolution- and random-walk-based imputation. *Proc. Natl. Acad. Sci. U. S. A.* 116, 14011–14018
61. Zeng, W. *et al.* (2019) DC3 is a method for deconvolution and coupled clustering from bulk and single-cell genomics data. *Nat. Commun.* 10, 4613
62. Duren, Z. *et al.* (2018) Integrative analysis of single-cell genomics data by coupled nonnegative matrix factorizations. *Proc. Natl. Acad. Sci. U. S. A.* 115, 7723–7728
63. Welch, J.D. *et al.* (2019) Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* 177, 1873–1887 e17
64. Granja, J.M. *et al.* (2019) Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* 37, 1458–1465
65. Rendeiro, A.F. *et al.* (2020) Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL. *Nat. Commun.* 11, 577
66. Ma, A. *et al.* (2020) Integrative methods and practical challenges for single-cell multi-omics. *Trends Biotechnol.* March 26, 2020. <https://doi.org/10.1016/j.tibtech.2020.02.013>
67. Hainer, S.J. *et al.* (2019) Profiling of pluripotency factors in single cells and early embryos. *Cell* 177, 1319–1329 e11
68. Voineagu, I. *et al.* (2011) Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 474, 380–384
69. Ott, C.J. *et al.* (2018) Enhancer architecture and essential core regulatory circuitry of chronic lymphocytic leukemia. *Cancer Cell* 34, 982–995 e7
70. Wu, J. *et al.* (2018) Characterization of DNA methylation associated gene regulatory networks during stomach cancer progression. *Front. Genet.* 9, 711
71. Song, M. *et al.* (2019) Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat. Genet.* 51, 1252–1262
72. Duren, Z. *et al.* (2020) Time course regulatory analysis based on paired expression and chromatin accessibility data. *Genome Res.* 30, 622–634
73. Zeng, W. *et al.* (2020) Integrating distal and proximal information to predict gene expression via a densely connected convolutional neural network. *Bioinformatics* 36, 496–503
74. Li, W. *et al.* (2017) Gene co-opening network deciphers gene functional relationships. *Mol. BioSyst.* 13, 2428–2439
75. Li, W. *et al.* (2019) DeepTACT: predicting 3D chromatin contacts via bootstrapping deep learning. *Nucleic Acids Res.* 47, e60
76. van Galen, P. *et al.* (2019) Single-cell RNA-Seq reveals AML hierarchies relevant to disease progression and immunity. *Cell* 176, 1265–1281 e24
77. Chen, X. *et al.* (2019) Single cell gene co-expression network reveals FECH/CROT signature as a prognostic marker. *Cells* 8, 698
78. Peng, H. *et al.* (2019) A component overlapping attribute clustering (COAC) algorithm for single-cell RNA sequencing data analysis and potential pathobiological implications. *PLoS Comput. Biol.* 15, e1006772
79. Harly, C. *et al.* (2019) The transcription factor TCF-1 enforces commitment to the innate lymphoid cell lineage. *Nat. Immunol.* 20, 1150–1160
80. Gryder, B.E. *et al.* (2019) Histone hyperacetylation disrupts core gene regulatory architecture in rhabdomyosarcoma. *Nat. Genet.* 51, 1714–1722
81. Mohammadi, S. *et al.* (2018) A geometric approach to characterize the functional identity of single cells. *Nat. Commun.* 9, 1516
82. Rambow, F. *et al.* (2018) Toward minimal residual disease-directed therapy in melanoma. *Cell* 174, 843–855 e19
83. Hong, S.P. *et al.* (2019) Single-cell transcriptomics reveals multi-step adaptations to endocrine therapy. *Nat. Commun.* 10, 3840
84. Durante, M.A. *et al.* (2020) Single-cell analysis reveals new evolutionary complexity in uveal melanoma. *Nat. Commun.* 11, 496
85. Rai, V. *et al.* (2020) Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures. *Mol. Metab.* 32, 109–121
86. Karemaker, I.D. and Vermeulen, M. (2018) Single-cell DNA methylation profiling: technologies and biological applications. *Trends Biotechnol.* 36, 952–965
87. Baxter, J.S. *et al.* (2018) Capture Hi-C identifies putative target genes at 33 breast cancer risk loci. *Nat. Commun.* 9, 1028
88. Vickovic, S. *et al.* (2019) High-definition spatial transcriptomics for *in situ* tissue profiling. *Nat. Methods* 16, 987–990
89. Praktijn, S.D. *et al.* (2020) Tracing tumorigenesis in a solid tumor model at single-cell resolution. *Nat. Commun.* 11, 991
90. Stuart, T. *et al.* (2019) Comprehensive integration of single-cell data. *Cell* 177, 1888–1902 e21
91. Argelaguet, R. *et al.* (2020) MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* 21, 111
92. Duren, Z. *et al.* (2017) Modeling gene regulation from paired expression and chromatin accessibility data. *Proc. Natl. Acad. Sci. U. S. A.* 114, E4914–E4923
93. Wang, Y. *et al.* (2016) Modeling the causal regulatory network by integrating chromatin accessibility and transcriptome data. *Natl. Sci. Rev.* 3, 240–251
94. Eraslan, G. *et al.* (2019) Deep learning: new computational modelling techniques for genomics. *Nat. Rev. Genet.* 20, 389–403
95. Dueck, H. *et al.* (2015) Deep sequencing reveals cell-type-specific patterns of single-cell transcriptome variation. *Genome Biol.* 16, 122
96. van Dijk, D. *et al.* (2018) Recovering gene interactions from single-cell data using data diffusion. *Cell* 174, 716–729 e27
97. Peng, T. *et al.* (2019) SCRABBLE: single-cell RNA-seq imputation constrained by bulk RNA-seq data. *Genome Biol.* 20, 88
98. Wang, J. *et al.* (2019) Data denoising with transfer learning in single-cell transcriptomics. *Nat. Methods* 16, 875–878
99. Dixit, A. *et al.* (2016) Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167, 1853–1866 e17
100. Yuan, Y. and Bar-Joseph, Z. (2019) Deep learning for inferring gene relationships from single-cell expression data. *Proc. Natl. Acad. Sci. U. S. A.* 116, 27151–27158
101. Brisimi, T.S. *et al.* (2018) Federated learning of predictive models from federated Electronic Health Records. *Int. J. Med. Inform.* 112, 59–67
102. Azodi, C.B. *et al.* (2020) Opening the black box: interpretable machine learning for geneticists. *Trends Genet.* 36, 442–455