OXFORD

Briefings in Bioinformatics, 2023, **24(1)**, 1–13
https://doi.org/10.1093/bib/bbac508
Problem Solving Protocol

Complexity of enhancer networks predicts cell identity and disease genes revealed by single-cell multi-omics analysis

Danni Hong†, Hongli Lin†, Lifang Liu, Muya Shu, Jianwu Dai, Falong Lu, Mengsha Tong and Jialiang Huang

Corresponding author. Jialiang Huang, State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Faculty of Medicine and Life Sciences, Xiamen University, Xiamen, Fujian 361102, China. E-mail: jhuang@xmu.edu.cn

†Danni Hong and Hongli Lin contributed equally.

Abstract

Many enhancers exist as clusters in the genome and control cell identity and disease genes; however, the underlying mechanism remains largely unknown. Here, we introduce an algorithm, eNet, to build enhancer networks by integrating single-cell chromatin accessibility and gene expression profiles. The complexity of enhancer networks is assessed by two metrics: the number of enhancers and the frequency of predicted enhancer interactions (PEIs) based on chromatin co-accessibility. We apply eNet algorithm to a human blood dataset and find cell identity and disease genes tend to be regulated by complex enhancer networks. The network hub enhancers (enhancers with frequent PEIs) are the most functionally important. Compared with super-enhancers, enhancer networks show better performance in predicting cell identity and disease genes. eNet is robust and widely applicable in various human or mouse tissues datasets. Thus, we propose a model of enhancer networks containing three modes: Simple, Multiple and Complex, which are distinguished by their complexity in regulating gene expression. Taken together, our work provides an unsupervised approach to simultaneously identify key cell identity and disease genes and explore the underlying regulatory relationships among enhancers in single cells.

Keywords: enhancer, single-cell multi-omics, complexity of enhancer network, super-enhancers, gene regulation

Introduction

Enhancers play a central role in orchestrating spatiotemporal gene expression programs during development and diseases [1–4]. Many enhancers exist as clusters in the genome to control the expression of the same target gene, termed enhancer clusters or super-enhancers (SEs) [5, 6]. While enhancer clusters are remarkably widespread features in the genome and provide an effective regulatory buffer for phenotypic robustness during development [7, 8], the underlying mechanisms remain poorly understood.

Genome editing using the CRISPR/Cas9 system offers an opportunity for functionally dissecting enhancer clusters [9]. Several groups, including ours, have utilized genome editing assays to functionally dissect individual constituent elements of a couple of SEs [10–18]. These studies suggest the diversity of enhancer cluster regulatory mechanisms, where the individual components may act additively, redundantly, synergistically or temporally. Meanwhile, proximity ligation-based chromatin interactions have been used to investigate the relationship among the individual components of enhancer clusters and their effects on target gene expression [19–24]. We and other groups uncover hub enhancers, the enhancers with frequent chromatin interactions, play distinct roles in chromatin organization and gene activation [25–28]. However, it is infeasible to scale up these approaches to rigorously test a wide range of enhancers due to technical limitations, such as the resolution in proximity ligation-based methods and the scalability in genome editing-based methods.

With the rapid development of single cell RNA sequencing (scRNA-seq) and single-cell Assay for Transposase-Accessible Chromatin using sequencing (scATAC-seq) [29–34], a large number of single cell chromatin accessibility and gene expression profiles have been generated in various biological systems [35–40]. However, these existing studies have largely focused on connecting enhancers with their target genes [41], but rarely on the regulatory relationship between enhancers. There remains a lack of method development to quantitatively assess how individual elements work together to regulate gene expression.

In this study, we developed a computational method termed eNet to build enhancer networks based on single-cell chromatin accessibility and gene expression data. Applying eNet on various biological systems, we found that the complexity of enhancer networks can predict cell identity and disease genes. Altogether,

Danni Hong is a PhD student at the School of Life Sciences, Xiamen University.
Hongli Lin is a postgraduate student at the School of Life Sciences, Xiamen University.
Lifang Liu is a postgraduate student at the School of Life Sciences, Xiamen University.
Muya Shu is a Postdoctoral Fellow at the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences.
Jianwu Dai is a Professor at the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences.
Falong Lu is a Professor at the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences.
Mengsha Tong is an Assistant Professor at the School of Life Sciences, Xiamen University.
Jialiang Huang is a Professor at the School of Life Sciences, Xiamen University.
Received: August 17, 2022. **Revised:** October 21, 2022. **Accepted:** October 24, 2022

we proposed the concept of complexity of enhancer networks and established its functional links with cell identity or disease.

Results

eNet builds enhancer networks based on single cell multi-omics data

Many enhancers exist as clusters in the genome; however, the underlying mechanism through which the clustered enhancers work together to regulate the same target gene remains largely unknown. To this end, we developed an algorithm eNet to build an enhancer network for each gene to quantitatively assess how multiple enhancers work together to regulate gene expression based on scATAC-seq and scRNA-seq data (Methods). The enhancer network we proposed is a gene regulation model that not only delineates the mapping between enhancers and target genes but also quantifies the underlying regulatory relationships among enhancers, which differs from previous studies [5–7, 31]. First, given the scATAC-seq and scRNA-seq profiles, the enhancer accessibility and gene expression matrix of single cells were prepared as the input of eNet (Figure 1A). Second, a set of enhancers were identified, termed a putative enhancer cluster hereafter, which putatively regulate a specific target gene within a ±100 kb window based on the correlation between gene expression and enhancer accessibility in single-cell data (Figure 1B). Third, we evaluated the enhancer interaction potential based on their chromatin co-accessibility calculated by Cicero [41] and determined the enhancer pairs with significantly high co-accessibility as the predicted enhancer interactions (PEIs) (Figure 1C). Fourth, an enhancer network was built to delineate how multiple enhancers interact with each other to regulate gene expression, where nodes represent enhancers and edges represent the PEIs between enhancers (Figure 1D). Fifth, the complexity of the enhancer network was evaluated by two metrics: (i) the number of enhancers, termed the network size (*x*-axis), and (ii) the frequency of PEIs, termed the network connectivity (*y*-axis), quantified by the average degree of network [42] (Figure 1E). Lastly, based on the network size and network connectivity, we classified the enhancer networks into several modes: Simple, Multiple, Complex and others (but will not be discussed due to limited cases) (Figure 1E and F). Intuitively, the complexity of the enhancer network increased from Simple mode to Multiple mode by involving more enhancers and further to Complex mode by increasing the interactions between enhancers. Altogether, eNet builds enhancer networks to clarify how a putative enhancer cluster regulates gene expression based on scATAC-seq and scRNA-seq data.

Cell identity and disease genes tend to be regulated by complex enhancer networks during human hematopoiesis

We first applied eNet to build enhancer networks during human hematopoiesis using a human blood dataset [35], including the single cell chromatin accessibility and transcriptional landscapes in human bone marrow and peripheral blood mononuclear cells (Figure 2A). In total, we built 11 438 enhancer networks during human hematopoiesis (Figure 2B, Supplementary Figure S1A available online at http://bib.oxfordjournals.org/). We noticed several blood-related cell identity or disease genes, such as *BCL11B*, *ETS1*, *CCR7* and *IL7R* displayed obviously large network size and high-network connectivity (Figure 2B). This inspired the question that whether cell identity genes tend to be regulated by complex enhancer networks. To test this hypothesis, we classified these enhancer networks into three modes: Simple (controlled by one or few enhancers), Multiple (multiple enhancers but limited PEIs) and Complex (multiple enhancers and frequent PEIs). It resulted in 6894 Simple, 2992 Multiple and 1552 Complex enhancer networks (Figure 2B; Supplementary Table S2 available online at http://bib.oxfordjournals.org/; Methods). For example, the *CD3E* gene, encoding a subunit of the T-cell receptor-CD3 complex, was controlled by an enhancer network consisting of 14 PEIs among 9 enhancers. In contrast, the *SERPINE2* gene, encoding a member of the serpin family of proteins that inhibit serine proteases, was controlled by an enhancer network containing the same number of enhancers but only two PEIs (Figure 2C). Interestingly, the *CD3E* enhancer network showed significant higher chromatin co-accessibility than *SERPINE2*, irrespective of their indistinguishable chromatin accessibility and similar enhancer number (Supplementary Figure S1B and C available online at http://bib.oxfordjournals.org/).

Next, we curated a list of known cell identity genes in the blood system (Supplementary Table S3 available online at http://bib.oxfordjournals.org/; Methods) and calculated their enrichment in the genes regulated by three enhancer network modes (Figure 2E). We observed that genes regulated by Multiple mode showed higher enrichment in cell identity genes than those by Simple mode, which is consistent with previous reports that developmentally expressed genes are commonly associated with multiple enhancers [7, 31, 43]. In addition, we found that genes regulated by Complex mode exhibited the highest enrichment in cell identity genes, 8.7-fold using the whole genome as the background (Figure 2E). Similarly, genes regulated by Complex mode displayed a higher enrichment in blood-related disease genes curated from DisGeNET [44] than those by Multiple mode (4.8-fold versus 2.4-fold, $P = 3.4E{-}20$, binomial test, Figure 2F). Notably, these observations were robust to various threshold values of network size, network connectivity and chromatin accessibility (Supplementary Figures S2–S4 available online at http://bib.oxfordjournals.org/). These results suggested that cell identity and disease genes tend to be regulated by complex enhancer networks.

Complexity of enhancer networks predicts cell identity and disease genes

To systematically evaluate the performance of the complexity of enhancer networks in predicting cell identity and disease genes, we ranked enhancer networks by network size, network connectivity and overall chromatin accessibility. We then calculated the enrichment of cell identity and disease genes in the list of top ranked enhancer networks related genes, using the whole genome as the background (Figure 2G and H). We found that the genes controlled by enhancer networks with more enhancers were overall preferentially more enriched for cell identity genes (Figure 2G), which concurs with previous studies [6, 7, 31]. Importantly, network connectivity displayed better performance in predicting cell identity genes than the network size. For example, the top 50 genes ranked by network connectivity were 77.7-fold enriched in cell identity genes, compared with those by network size (29.9-fold). Both the network connectivity and network size showed remarkably better performance in predicting cell identity genes than the chromatin accessibility of enhancers in the network. Similarly, network connectivity displayed the best performance in predicting blood-related disease genes (6.8-fold in the top 50 genes, Figure 2H). Therefore, these analyses suggest that the complexity of enhancer networks can predict cell identity and disease genes.

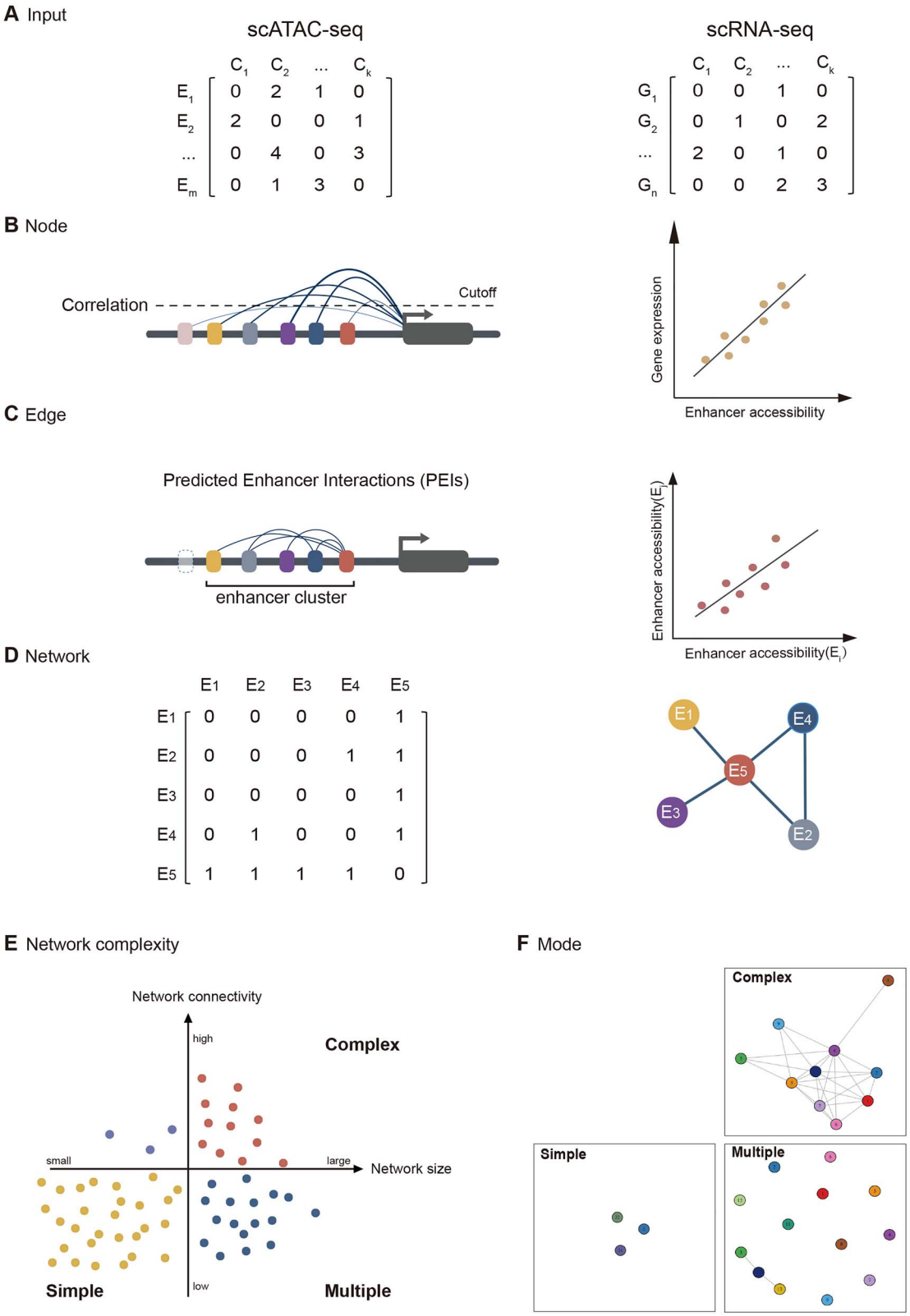

Figure 1. eNet, an algorithm to build enhancer networks based on scATAC-seq and scRNA-seq data. (**A**) Input: Preparation of the enhancer accessibility and gene expression matrix from scATAC-seq and scRNA-seq data. Each row represents an enhancer or a gene, while each column represents a cell. (**B**) Node: Identification of putative enhancer clusters regulating a specific target gene based on the correlation between gene expression and enhancer accessibility. (**C**) Edge: Determination of the PEIs, the enhancer pairs with significantly high co-accessibility calculated using Cicero. (**D**) Network: Construct enhancer network to represent the PEIs among enhancers in a putative enhancer cluster, where nodes represent enhancers and edges represent PEIs. (**E**) Network complexity: Calculation of the network complexity by (i) network size, the number of enhancers (x-axis); and (ii) network connectivity, the PEIs frequency, quantified by the average degree of network (y-axis). (**F**) Mode: Classification of the enhancer networks into three modes based on network complexity: Complex, Multiple and Simple, with representative examples shown in the cartoon.

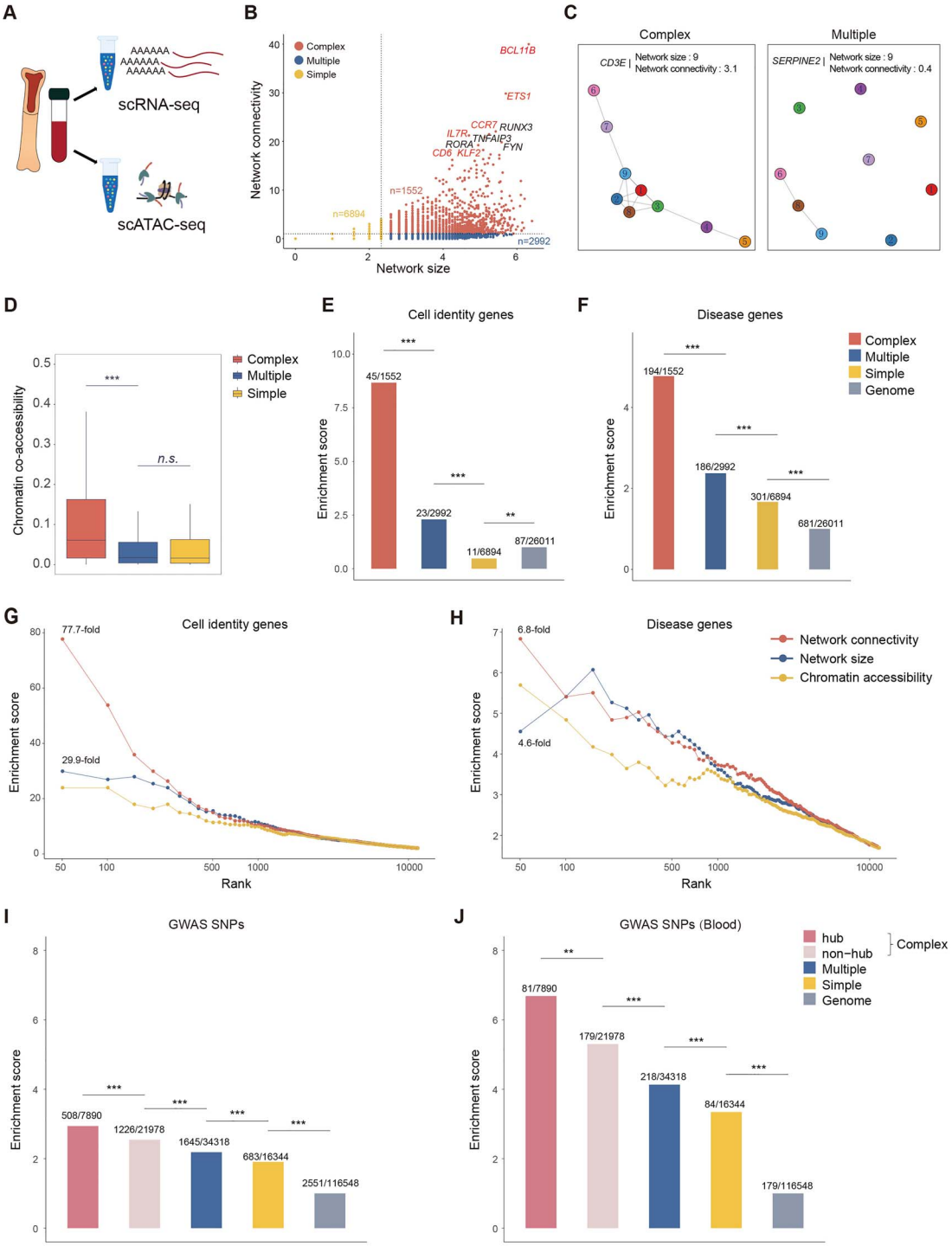

Figure 2. Enhancer networks during human hematopoiesis. (**A**) The human blood dataset. (**B**) Scatter plot of the enhancer networks during hematopoiesis, where the x-axis represents the network size (log2-scaled) and the y-axis represents network connectivity. Top 10 genes ranked by network connectivity are labelled, where known blood-related cell identity or disease genes are red-highlighted. (**C**) Representative enhancer networks in Complex or Multiple mode. (**D**) Chromatin co-accessibility of PEIs calculated using Cicero in Complex, Multiple and Simple modes. *P*-values were calculated using Student's *t*-test. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; *n.s.*, not significant. (**E**) and (**F**) Enrichment of cell identity (**E**) and disease genes (**F**) in genes in Complex, Multiple and Simple modes, using the whole genome as the background. The number of cell identity or disease genes and total genes in each group are labelled on each bar. *P*-values were calculated using the binomial test. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; *n.s.*, not significant. (**G**) and (**H**) Enrichment of cell identity (**G**) and disease genes (**H**) (y-axis) is plotted for top genes (x-axis) ranked by different properties of enhancer networks, including network connectivity (the frequency of PEIs in this study), network size (equivalent to the enhancer number in multiple enhancers [7], DORCs [31]) or overall chromatin accessibility of enhancers (similar to the sum of the individual constituent enhancers in SEs [6]). (**I**) Enrichment of the diseases/traits-related SNPs curated in the GWAS catalog for enhancers in Complex (hub and non-hub), Multiple and Simple modes, using randomly selected genomic regions as the control. *P*-values were calculated using the binomial test. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; *n.s.*, not significant. (**J**) Enrichment of blood-related GWAS SNPs. *P*-values were calculated using the binomial test. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; *n.s.*, not significant.

Network hub enhancers are functionally important

Enhancer networks provide an opportunity to study how individual elements work and then how they interact with each other to control gene expression. Toward this end, we focused on the enhancers with frequent PEIs in enhancer networks in Complex mode, termed network hub enhancers (Methods). We first found that network hub enhancers displayed significantly higher level of sequence phastCons conservation [45] than non-hub enhancers ($P = 3.8E{-}8$, Student's t-test, Supplementary Figure S1D available online at http://bib.oxfordjournals.org/), suggesting that network hub enhancers might be more functionally important. Next, we assessed the enrichment of single-nucleotide polymorphisms (SNPs) linked to diverse phenotypic traits and diseases in the genome-wide association study (GWAS) catalog [46], in enhancers in Complex (hub and non-hub), Multiple and Simple modes. We observed significantly higher enrichment of blood-associated GWAS SNPs in enhancers in Multiple mode than those in Simple mode ($P = 2.8E{-}4$, binomial test, Figure 2J), which is consistent with previous studies [6, 7]. Additionally, the enhancers in Complex mode (hub and non-hub) showed higher enrichment in GWAS SNPs associated with blood traits than those in Multiple mode. In particular, in Complex mode, hub enhancers displayed higher enrichment of GWAS SNPs associated with blood traits than non-hub enhancers (6.7-fold versus 5.3-fold, $P = 5.8E{-}3$, binomial test, Figure 2J). These results suggest that compared with Multiple and Simple modes, enhancers in Complex mode might be more important in diseases, where hub enhancers are major functional constituents.

Enhancer network outperforms SE in predicting cell identity and disease genes

We next compared the performance of predicting cell identity and disease genes by enhancer networks and SEs [6]. To this end, we downloaded a list of SEs associated with hematopoiesis-related cell types from the dbSUPER database [47] and curated a catalog of hematopoiesis-related SEs containing 2306 SEs (Supplementary Figure S5D and E, Supplementary Table S4 available online at http://bib.oxfordjournals.org/). We identified 2159 potential target genes regulated by these SEs using ROSE algorithm [6]. Comparing the genes regulated by SEs or by enhancer networks in Complex mode, we separated them into three groups: Complex-only (836), SE-only (1443) and Complex SE (716) (Figure 3A). The constituent enhancers in these two groups (SE-only versus Complex SE) showed significantly different chromatin co-accessibility, but indistinguishable chromatin accessibility (Figure 3B and C). It might explain the diverse and heterogeneous mechanisms of SEs, such as cooperative, redundant and hierarchical revealed by CRISPR/Cas9 genome editing assays. Strikingly, genes in Complex-only group displayed significantly higher enrichment in cell identity and disease genes than those in SE-only group, while genes in Complex SE group showed the highest enrichment (Figure 3D and E). Similar patterns that enhancer networks precede SEs in predicting cell identity and disease genes were observed in GM12878 cell line (Supplementary Figure S5A–C available online at http://bib.oxfordjournals.org/). We further ranked genes by network connectivity, network size, chromatin accessibility or SE ranks based on H3K27ac signal and found that network connectivity showed the best performance in predicting both cell identity and disease genes (Figure 3F and G). These results suggested that the enhancer networks precede SEs in predicting cell identity and disease genes.

Enhancer networks based on PEIs remedy the resolution limitations in Hi-C chromatin interactions

The proximity ligation-based methods to capture genome-wide chromatin interactions at high-resolution for the analysis of enhancer interactions remains difficult and costly [21, 48, 49]. We wonder to what extent the PEIs in eNet analysis resolve the resolution limitations in Hi-C data. To this end, we compared enhancer networks based on PEIs and Hi-C data in GM12878 cell line, where scATAC-seq [31], H3K27ac ChIP-seq [3] and high-resolution Hi-C data [19] are available. We observed the high co-accessible enhancer pairs (PEIs) that showed significant enrichment of Hi-C chromatin interactions (Figure 4A), indicating the overall concordance between co-accessible pairs and proximity ligation-based chromatin interactions [41]. For example, at the locus controlling *CCR7*, a gene expressed in various lymphocytes, we predicted 20 PEIs based on scATAC-seq data, while only 10 chromatin interactions were detected via Hi-C probably due to the limited resolution at 5 kb level (Figure 4B and, C). We systematically compared the enhancer networks based on scATAC-seq and Hi-C data by replacing PEIs with Hi-C interactions and re-built enhancer networks. We observed a significant overlap between the genes controlled by the complex enhancer networks based on PEIs and Hi-C data (Figure 4D, $P < 2.2E{-}16$, Fisher's exact test). Interestingly, PEI-only group showed significant higher enrichment of cell identity and disease genes than HiC-only group, where PEI-with-HiC showed the highest enrichment (Figure 4E and F). Moreover, we found that the network hub enhancers derived from PEIs showed significant higher enrichment of GWAS SNPs than those from Hi-C data (Supplementary Figure S5F–H available online at http://bib.oxfordjournals.org/). Taken together, these results suggested that enhancer networks based on PEIs remedy the resolution limitations of chromatin interactions in Hi-C data.

Dynamics of *PAX5* enhancer network drives gene expression during B cell lineage commitment

Enhancer networks were built based on single cell multi-omics data, providing an opportunity to investigate the dynamic role of enhancer networks in determining gene expression during cell differentiation. To this end, we focused on B cell differentiation, from hematopoietic stem cell (HSC), lymphoid-primed multipotent progenitor (LMPP), common lymphoid progenitor (CLP), pre-B, to B cells (Figure 5A; Methods). The *PAX5* gene, a known key regulator for B cell differentiation, specifically expressed in pre-B and B cells, was controlled by a putative enhancer cluster consisting of 24 enhancers (Figure 5B and C). Then, we built cell-type-specific enhancer networks by constructing the enhancer networks for each cell type independently (Methods). Comparing the enhancer networks specific for each cell type, we observed the sequential changes in the constituent enhancers during B cell differentiation, in terms of both chromatin accessibility and network interactions (Figure 5D). Within the *PAX5* enhancer network, we noticed that enhancer E14, constitutively accessible from HSC to B cells, functions as a network hub enhancer to coordinate enhancer network interactions to establish the enhancer network gradually during B cell differentiation (Figure 5D). Interestingly, we found that the *PAX5* enhancer network was almost fully established in the CLP and pre-B stages, which preceded the gene expression of *PAX5* in pre-B and B cells (Figure 5C and D). It suggests that the establishment of an enhancer network may drive gene expression during lineage commitment.

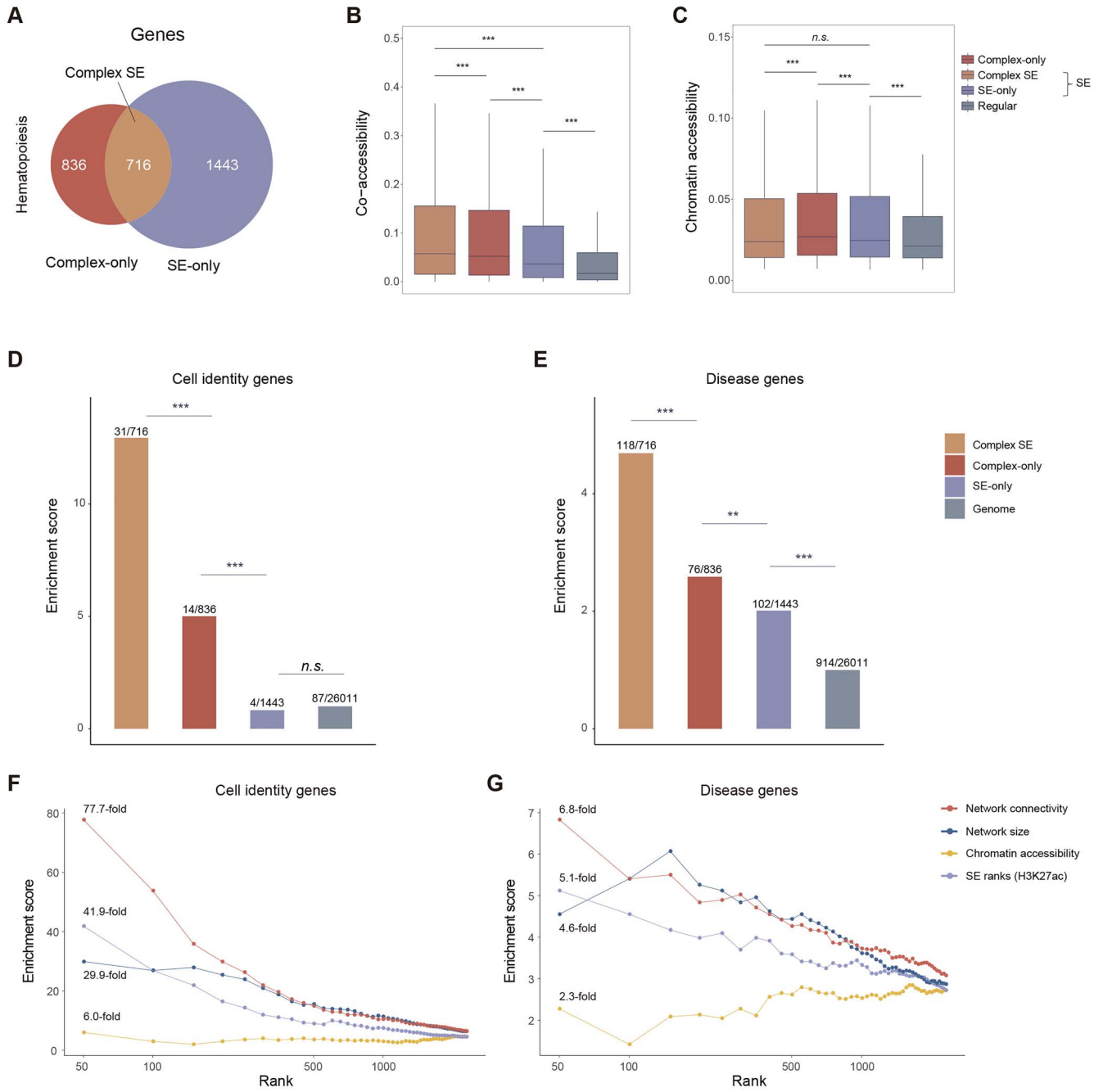

Figure 3. Enhancer network outperforms SE in predicting cell identity and disease genes. (**A**) Venn diagram showing the overlap between genes in Complex mode in Figure 2 in blood dataset and hematopoiesis-related SEs, resulting in three groups, Complex-only, Complex SE (SEs with network structure) and SE-only (SEs without network structure). (**B**) and (**C**) Co-accessibility (**B**) and chromatin accessibility (**C**) of the constituent enhancers in three groups, using regular enhancers as control. *P*-values were calculated using Student's *t*-test. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; *n.s.*, not significant. (**D**) and (**E**) Enrichment of cell identity (**D**) and disease genes (**E**) in genes in three groups, using the whole genome as the background. *P*-values were calculated using the binomial test. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; *n.s.*, not significant. (**F**) and (**G**) Enrichment of cell identity (**F**) and disease genes (**G**) (*y*-axis) for the genes in (**A**), ranked by the network complexity (*x*-axis), measured by (i) network connectivity, as well as the overall enhancer activity, (ii) network size as well as enhancer number, (iii) chromatin accessibility and (iv) SE ranks based on H3K27ac signals calculated by ROSE [6].

eNet is robust and broadly applicable

To investigate the broad applicability of eNet, we applied it to various datasets in human or mouse tissues across different single-cell platforms, including SHARE-seq mouse skin dataset [31], SNARE-seq mouse cerebral cortex dataset [34] and sci-ATAC-seq3 human fetal kidney and heart datasets [50]. Similar to the above findings, we found cell identity and disease genes tended to be regulated by complex enhancer networks (Supplementary Figure S6A, C, E and G available online at http://bib.oxfordjournals.org/). The network connectivity showed the best performance in predicting cell identity genes and disease genes (Figure 6A, C, E and G; Supplementary Figure S6B, D, F and H available online at http://bib.oxfordjournals.org/). Hub enhancers in Complex mode displayed the highest enrichment of tissue-related GWAS SNPs (Figure 6B, D, F and H). These analyses in various human or mouse tissues datasets (Figures 2 and 6;

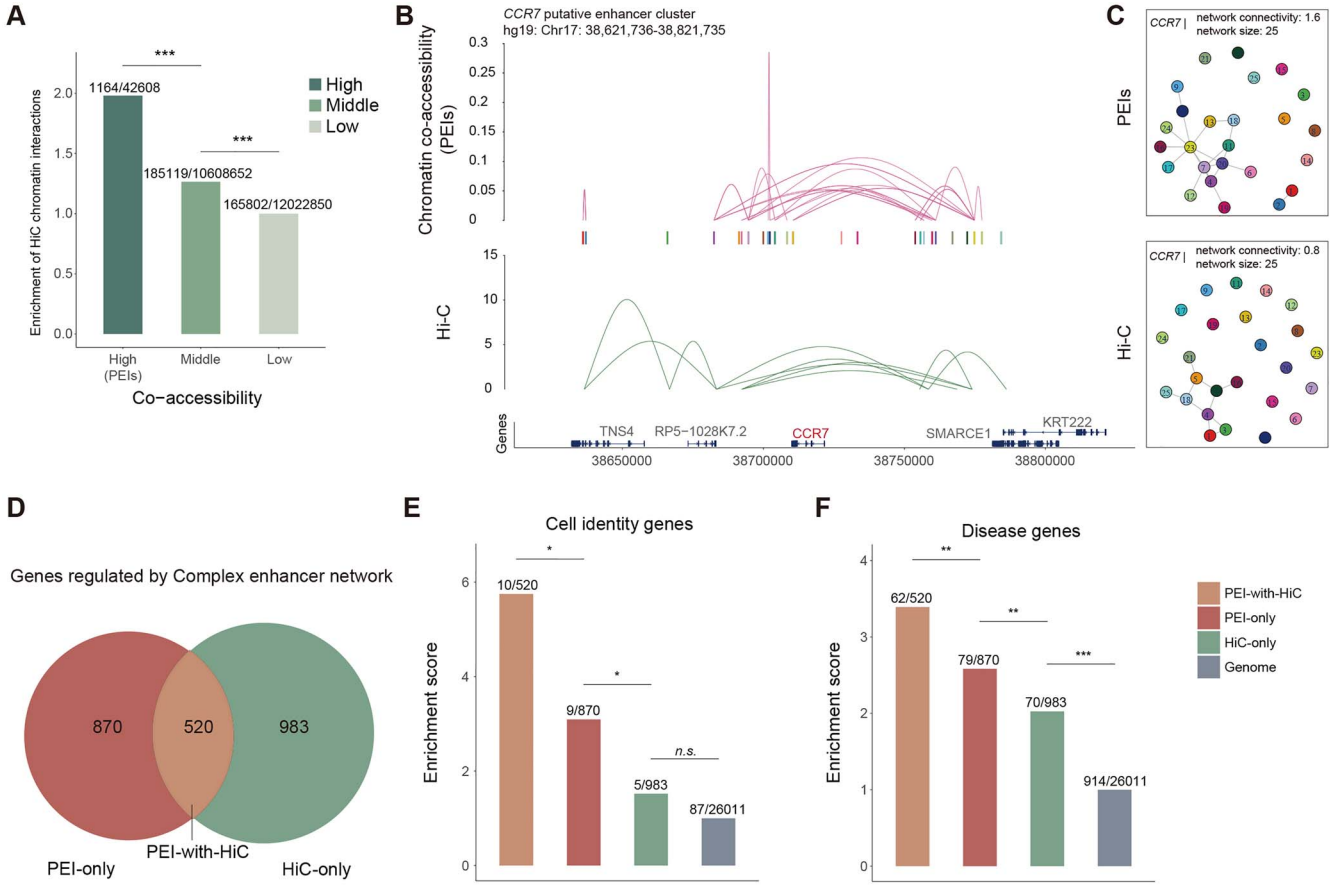

Figure 4. Comparison of enhancer networks based on single cell data and Hi-C data in GM12878 cell line. (**A**) Enrichment of Hi-C chromatin interactions in three groups of enhancer pairs divided by chromatin coaccessibility:High (PEIs), Middle and Low, using the group Low as the background. *P*-values were calculated using the binomial test. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *n.s.*, not significant. (**B**) The PEIs and Hi-C chromatin interactions in the *CCR7* locus. Y-axis indicated the chromatin coaccessibility of PEIs calculated by Cicero (top) or the fragments frequency of Hi-C chromatin interaction (bottom). (**C**) *CCR7* enhancer networks based on PEIs (top) or Hi-C chromatin interactions (bottom). (**D**) Venn diagram showing the overlap of genes regulated by the Complex enhancer networks defined based on PEIs and Hi-C data, resulting in three groups: PEIs-with-HiC, PEIs-only and HiC-only. (**E**) and (**F**) Enrichment of cell identity (**E**) and disease genes (**F**) in three groups, using the whole genome as the background. *P*-values were calculated using the binomial test. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *n.s.*, not significant.

Supplementary Figure S6 available online at http://bib.oxfordjournals.org/) support the conclusion that eNet is robust and broadly applicable in various biological systems and different single-cell platforms.

Model of enhancer networks in gene regulation

Our analysis revealed three modes of enhancer networks in regulating gene expression according to their network complexity: Complex, Multiple and Simple. We found that genes regulated by the Simple mode were primarily enriched in housekeeping functions, such as RNA modification and DNA repair (Figure 7A). In contrast, genes regulated by the Complex mode were enriched in key genes related to cell fate commitment, such as the regulation of leukocyte differentiation in human blood, skin development in mouse skin and cerebellar cortex formation in mouse cerebral datasets [51]. In addition, Complex mode preferentially regulated upstream regulators, such as transcription factors [52] (Supplementary Figure S7A available online at http://bib.oxfordjournals.org/).

Therefore, we proposed a model of enhancer networks containing three modes according to their network complexity: Simple, Multiple and Complex (Figure 7B). By definition, in Simple mode, gene regulation was controlled simply by one or a limited number of enhancers; we speculated that it provided a quick response to control a large number of regular genes, such as housekeeping genes, at a low cost. Meanwhile, in Multiple mode, gene regulation was controlled by multiple enhancers but limited PEIs; this might increase the strength of regulation and redundancy of gene expression at the cost of involving more enhancers. Lastly, gene regulation was controlled by multiple enhancers and frequent PEIs in Complex mode, perhaps the most robust to random failures of individual enhancers (transcriptional noise or genetic mutation), at the cost of connecting enhancers and primarily controls key cell identity genes. Enhancer networks are established gradually during lineage commitment and drive the expression of cell identity genes, where network hub enhancers play central roles to coordinate the network system.

Discussion

Enhancer networks have been reported in previous studies through integrated analysis of cell type-specific epigenomic data or CRISPR-based technique [53–57]. Here, we reported eNet to build enhancer network based on the rich source of single-cell multi-omics data and quantified its complexity by two metrics: network size and network connectivity. The first metric, network size (the number of enhancers) is equivalent or similar to the

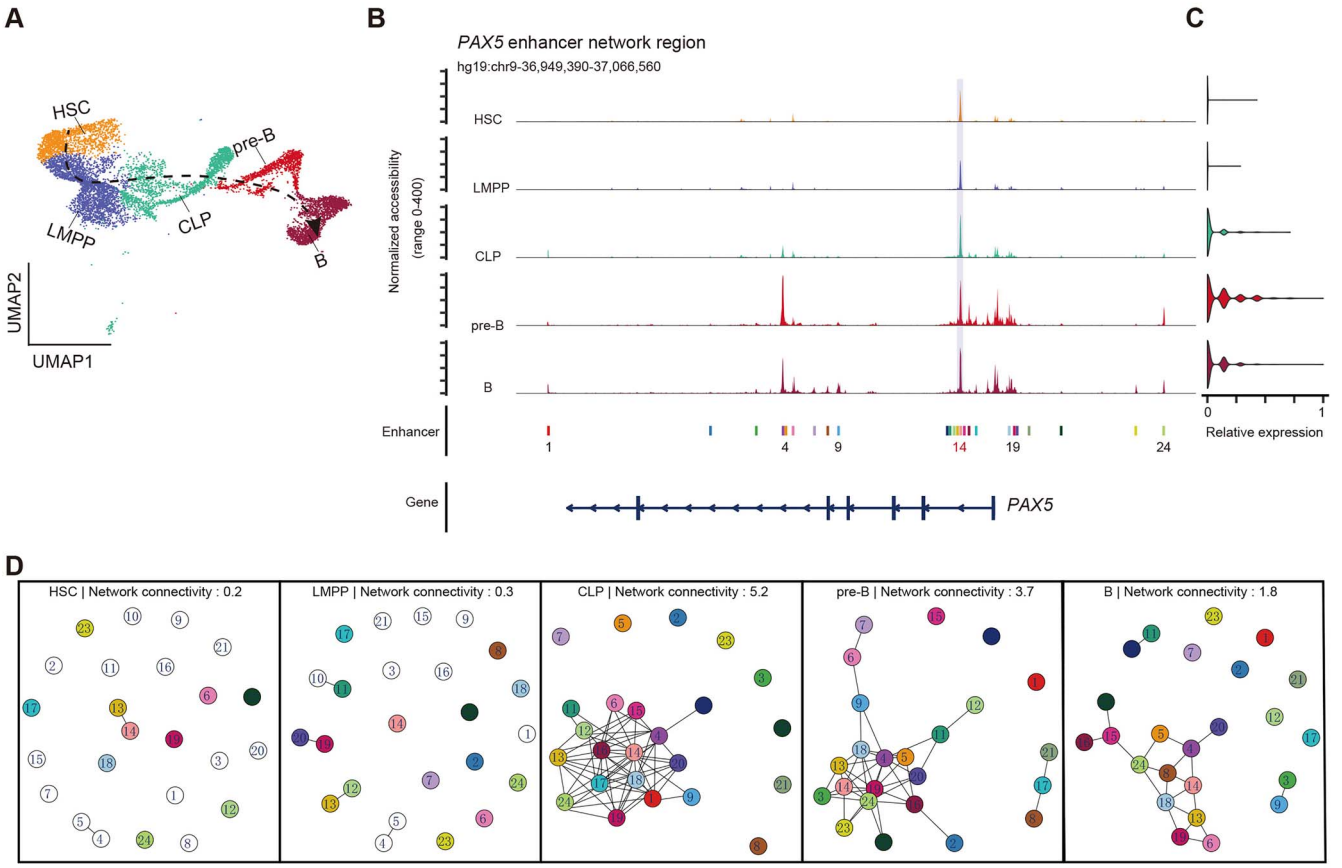

Figure 5. Dynamics of enhancer networks during B cell differentiation. (**A**) UMAP of B cell differentiation colored by cell-type annotation, the dash-line indicates the pseudotime during B cell differentiation inferred based on scATAC-seq data. (**B**) Genome browser track of *PAX5* putative enhancer cluster, involving 24 enhancers, that are accessible at any of HSC, LMPP, CLP, pre-B and B cell types. (**C**) Violin plot showing *PAX5* relative expression. (**D**) The *PAX5* enhancer networks in HSC, LMPP, CLP, pre-B and B cells, where the colored nodes represent accessible enhancers, while the empty nodes represent closed enhancers. The edges represent PEIs.

sum of the individual constituent enhancers in previous studies, such as multiple enhancers [7], domains of regulatory chromatin (DORCs) [31], regulatory locus complexity [58] or SEs [6]. However, the second metric, network connectivity (the frequency of PEIs), measuring the potential enhancer interactions, differs from these existing studies. Network connectivity had the best performance in predicting cell identity and disease genes, where the network hub enhancers are the most functionally important. Interestingly, SEs with or without network structure displayed different network connectivity between the constituent enhancers, irrespective of their indistinguishable chromatin accessibility, which might explain their diverse and heterogeneous mechanisms [13–16]. Taken together, to our knowledge, we for the first time propose the concept of 'complexity of enhancer network' and establish its functional links with cell identity or disease, which greatly extended these previous findings in understanding the biological relevance and implications of enhancer network.

The development of single-cell technologies generated a large number of single cell multi-omics profiles in various biological systems [35–38]. However, these studies have largely focused on connecting distal enhancers with their target genes. eNet allows us to explore how individual elements interact with each other to control gene expression during lineage commitment at single-cell resolution. Next, we discovered a simple rule that the complexity of enhancer network is effective in predicting cell identity genes, disease and phenotype-associated genes, which outperforms the existing models such as SE, enhancer cluster and gene expression variance (Figures 2 and 3, Supplementary Figures S1 and S6 available online at http://bib.oxfordjournals.org/). We also mapped each enhancer to multiple target genes and build enhancer networks, and observed similar patterns (Supplementary Figure S1J and K available online at http://bib.oxfordjournals.org/). Importantly, it is not necessary to know the cell identity in advance from primary samples or conduct challenging experimental steps, such as cell subpopulation isolation and ChIP-seq. Thus, our method and findings made it possible to re-visit and make full use of such rapidly growing single cell multi-omics data.

This study had several limitations. First, eNet builds enhancer networks based on the assumption that the Cicero-detected co-accessible pairs [41], the PEIs in this study, are overall concordant with proximity ligation-based chromatin interactions. However, it is important to systematically compare the coherence of the enhancer networks from scATAC-seq with those from proximity ligation-based chromatin interactions at higher resolution, if available in the future. Second, in a parallel study, Shu *et al.* dissected *Atoh1* enhancer network revealed by eNet analysis through CRISPR/Cas9-mediated perturbation and confirmed the central role of network hub enhancers during spinal cord development [59]. However, fully determining the functional regulatory roles of

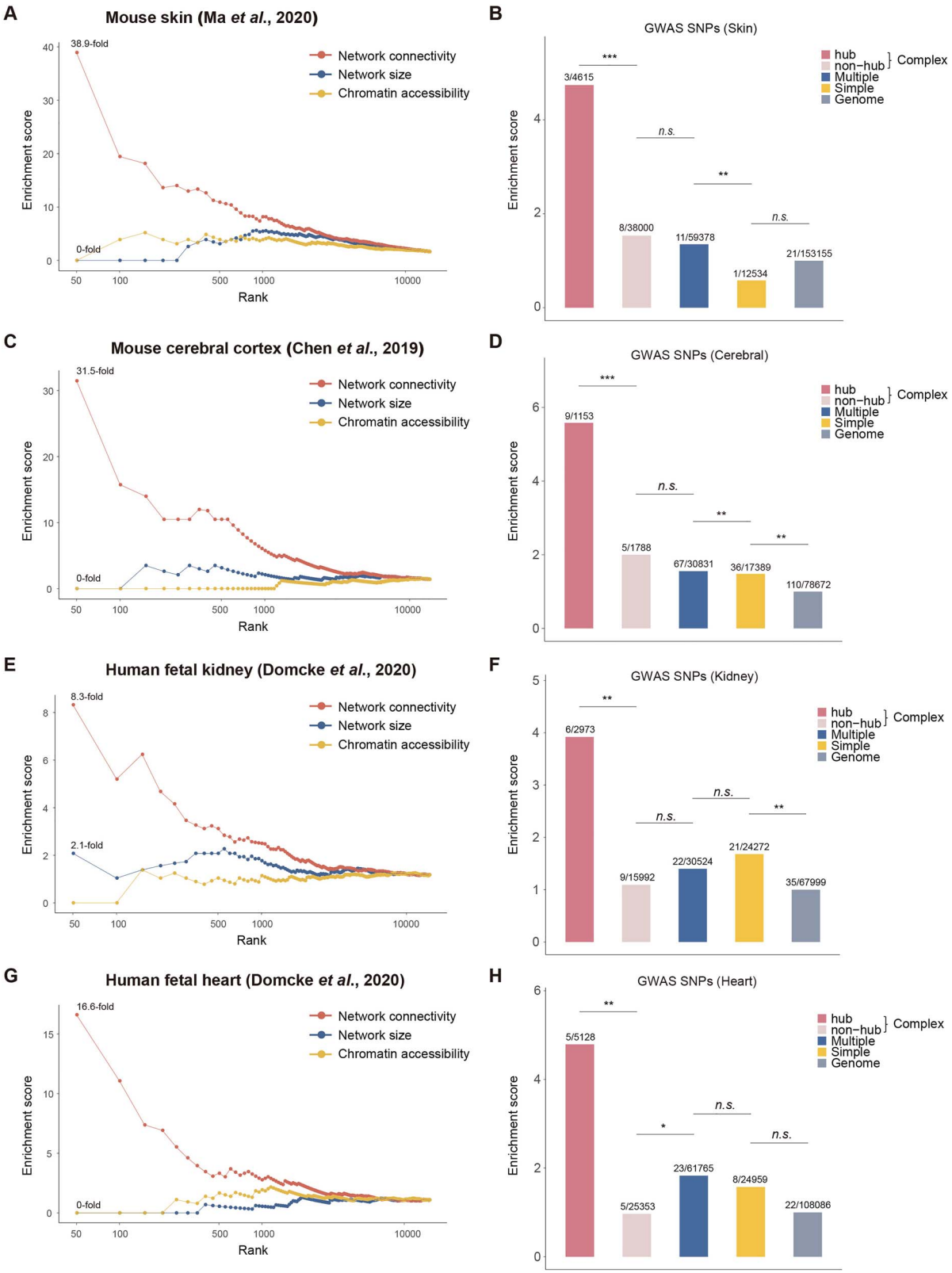

Figure 6. Enhancer networks in various human or mouse tissues across different single-cell platforms. (**A**), (**C**), (**E**), (**G**) Enrichment of cell identity genes (*y*-axis) is plotted for top genes ranked by various scoring methods (x-axis) in different tissues and approaches. (**A**) mouse skin dataset (SHARE-seq) [31], (**C**) mouse cerebral cortex dataset (SNARE-seq) [34], (**E**) human fetal kidney dataset (sci-ATAC-seq3) [50] and (**G**) human fetal heart dataset (sci-ATAC-seq3) [50]. (**B**), (**D**), (**F**), (**H**) Enrichment of tissue-related diseases/traits SNPs curated in GWAS catalog in enhancers in Complex (hub and non-hub), Multiple and Simple modes, using randomly selected genomic regions as the control. (**B**) mouse skin dataset, (**D**) mouse cerebral cortex dataset, (**F**) human fetal kidney dataset and (**H**) human fetal heart dataset. *P*-values were calculated using the binomial test. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; *n.s.*, not significant.

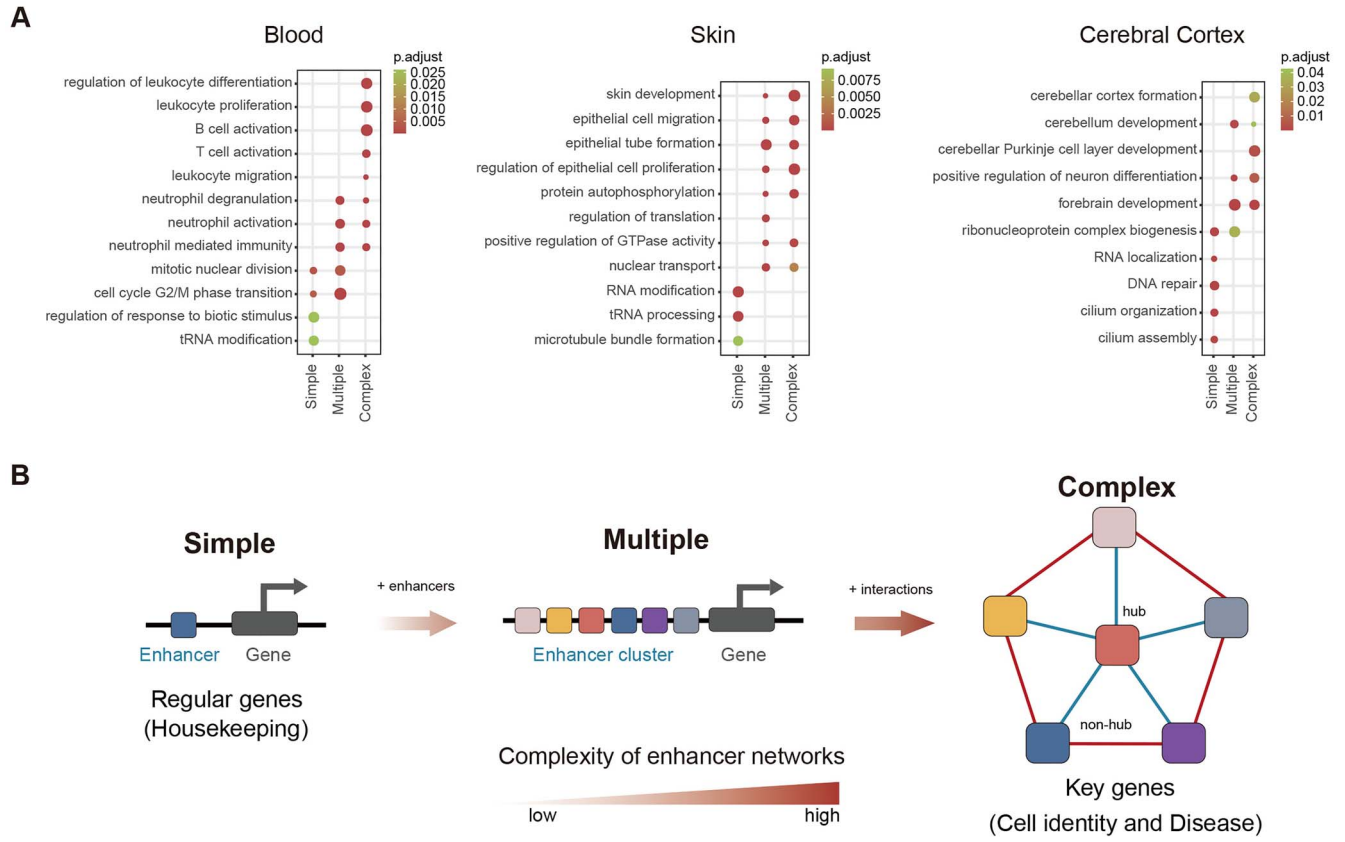

Figure 7. Model of enhancer networks in gene regulation. (**A**) Functional enrichment of genes regulated by enhancer networks in Simple, Multiple and Complex modes in human blood, mouse skin and mouse cerebral cortex datasets. (**B**) Three modes of enhancer networks. Simple mode, involving one or very few enhancers, provides quick response to control a large number of regular genes, such as housekeeping genes, at low cost; Multiple mode, involving multiple enhancers but limited PEIs, increases regulation strength as well as redundancy at the cost of the number of enhancers (nodes); Complex mode, involving multiple enhancers and frequent PEIs, provides robustness of gene regulation for key genes, such as cell identity and disease genes, at the cost of edges, where hub enhancers are functionally important.

enhancer networks in various biological systems requires more comprehensive investigations in future.

Methods

Overview of eNet

eNet is an algorithm to build enhancer networks for clustered enhancers controlling the same gene based on scATAC-seq and scRNA-seq datasets. Briefly, it contains the following six steps.

Step 1. Preparing input matrix (input)

In this study, the processed single cell chromatin accessibility and gene expression matrix data were downloaded directly from public literatures and used as the input for eNet.

Step 2. Identifying the putative enhancer cluster (node)

We identified putative enhancer cluster by adapting the method previously described [31, 40], with some modifications. Briefly, given a gene, we first selected the enhancers located within a ±2∼100 kb window around each annotated transcriptional start site as enhancer candidates. For each gene-enhancer pair, we then calculated the Spearman correlation between enhancer chromatin accessibility and gene expression. The Spearman correlations were z-score normalized using genome-wide gene-enhancer pairs as the background. Lastly, the enhancers with a significant z-score (P-value < 0.01, one-sided Student's t-test), were defined as a putative enhancer cluster regulating the specific target gene, as the nodes in the network.

Step 3. Identifying the PEIs (edge)

The chromatin co-accessibility of enhancer pairs across various cells was calculated using Cicero [41], a method that predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. By applying a threshold value of the co-accessibility calculated, we determined the significant co-accessible enhancer pairs, termed as the PEIs, as the edges of the network.

Step 4. Building enhancer networks (network)

We built a binary adjacency matrix to represent the PEIs for each putative enhancer cluster. Thus, the adjacency matrix can be visualized as an enhancer network, where nodes represent enhancers and the edges represent PEIs.

Step 5. Calculating network complexity (network complexity)

Network size was quantified by the quantity of nodes in the network. Network connectivity was quantified by the average degree [42], which was calculated as 2-fold of the number of edges and divided by the number of nodes.

Step 6. Classification of enhancer networks (mode)

We built the enhancer network for each gene genome-wide by repeating steps 1–5. Then, by applying a threshold value of network size and connectivity, we can classify the enhancer networks into several groups: Complex, Multiple, Simple and others.

Defining network hub enhancers

In Complex mode, we calculated the node degree for each enhancer and normalized them by the total number of edges in network, termed as normalized node degree. By applying a threshold value of the normalized node degree, we divided the enhancers into two groups, termed as network hub enhancers and non-hub enhancers, where network hub enhancers are those with high frequency of PEIs.

Robustness analysis of eNet

Building weighted enhancer network in Step 4

In addition to the binary adjacency matrix in Step 4, we also built the weighted co-accessibility enhancer networks and evaluated the performance of the complexity of weighted network connectivity in predicting cell identity and disease genes. It resulted in not obvious difference between two methods (Supplementary Figure S1G–I available online at http://bib.oxfordjournals.org/).

Thresholds to classify enhancer networks in Step 6

To test the robustness of thresholds of network size and network connectivity in defining Complex, Multiple and Simple mode, we set different thresholds and calculated the enrichment of cell identity and disease genes (Supplementary Figure S2 available online at http://bib.oxfordjournals.org/).

Effect of network size or chromatin accessibility on network connectivity

To decouple the effect of network size on network connectivity, we ranked the enhancer networks based on the network size and separated them into 5 groups from high to low, which resulted in similar network size level within each group. Then, we compared the network connectivity and cell identity/disease genes enrichment of the Complex and Multiple networks in each group (Supplementary Figure S3A–D available online at http://bib.oxfordjournals.org/). Similar analyses were performed to evaluate the Effect of network size or chromatin accessibility (Supplementary Figure S4A–D available online at http://bib.oxfordjournals.org/).

Retrieval of cell identity, disease and phenotype-associated genes

The cell identity genes were retrieved from related previous studies and the website (https://www.biolegend.com/cell_markers) [60]. The disease genes were from MalaCards (https://www.malacards.org), OMIM (https://omim.org) and DisGeNET [44]. The phenotype-associated genes were from MGI (http://www.informatics.jax.org/) [61,62]. All these cell identity and disease genes are provided in Supplementary Table S3 available online at http://bib.oxfordjournals.org/.

Enrichment analysis of cell identity and disease genes

Briefly, given a gene group, the enrichment score was calculated as the fold enrichment relative to the genome background. The computing method was determined as

$$(m/n) / (M/N),$$

where m and M represent the number of cell identity genes within the group and genome-wide, respectively, and n and N represent the number of genes within the group and genome-wide, respectively.

Performance evaluation in predicting cell identity and disease genes

To evaluate the performance of enhancer networks in predicting the cell identity and disease genes, we ranked all genes by various scoring methods, including network connectivity, network size, overall chromatin accessibility and gene expression variance. We then calculated the fold-enrichment of cell identity or disease genes in top-ranked genes with a moving window of 50, using the whole genome as the background.

Enrichment analysis of GWAS SNPs

The GWAS Catalog SNPs [46] were downloaded through the UCSC Table Browser (http://genome.ucsc.edu/). We curated a list of cell-type-related GWAS SNPs using a semi-automatic text mining method (Supplementary Table S5 available online at http://bib.oxfordjournals.org/). The overlap between loci and GWAS SNPs was performed using bedtools intersect [63]. For enhancers in each group, the enrichment score was calculated as the fold enrichment relative to the genome background. The computing method was listed as following:

$$(m/n) / (M/N),$$

where m and M represent the number of SNPs within the group and genome-wide, respectively, and n and N represent the number of loci within the group and genome-wide, respectively. The genome-wide background is generated from a list of loci obtained by randomly shuffling the list of regular enhancers.

Sequence conservation score

phastCons 100-way vertebrate conservation scores were downloaded from the UCSC Genome Browser [45]. We calculated the mean phastCons score for each enhancer as previously described [36].

Comparison of PEIs and Hi-C chromatin interactions

High-resolution Hi-C data in GM12878 cell were obtained from the literature [19] and processed as previously described [25]. We compared the enrichment of chromatin interactions detected by Hi-C in enhancer pairs with different co-accessibility (Figure 4A).

Comparison of enhancer networks based on PEIs and Hi-C chromatin interactions

We mapped Hi-C chromatin interactions to the enhancer clusters defined by single cell GM12878 data to replace the PEIs by using bedtools map, then built enhancer networks, evaluated the complexity of enhancer networks and defined network hub enhancers following the workflow in eNet analysis.

Cell-type-specific enhancer networks

To build cell-type-specific enhancer networks (Figure 5), we used the enhancer accessibility and gene expression matrix from a specific cell type as the input for eNet algorithm. The gene expression and chromatin accessibility of cell-type-specific enhancer network were represented by their average across all cells per cell type, followed by min-max normalization.

Blood-related SEs

The blood-related SEs were downloaded from the dbSUPER database [47] and merged into an SE list using bedtools [63] (Supplementary Table S4 available online at http://bib.oxfordjournals.org/).

Key Points

- We develop eNet, a computational method to build enhancer network based on scATAC-seq and scRNA-seq data.
- Cell identity and disease genes tend to be regulated by complex enhancer networks, where network hub enhancers are functionally important.
- Enhancer network outperforms the existing models in predicting cell identity and disease genes, such as super-enhancer and enhancer cluster.
- We propose the concept of complexity of enhancer networks and establish its functional links with cell identity or disease.

Authors' contributions

D.H. and J.H. conceived and designed the study. D.H., H.L., L.L. and M.T. performed the computational analysis. D.H., H.L., L.L., M.S., J.D., F.L. and J.H. wrote the manuscript. J.H. supervised the study.

Supplementary Data

Supplementary data are available online at https://academic.oup.com/bib.

Acknowledgement

We thank the useful comments from Drs Guo-Cheng Yuan, Jian Xu and the members in J.H. lab.

Funding

The National Natural Science Foundation of China (31871317 and 32070635 to J.H., 82002529 to M.T., 81891000 and 81891002 to J.D. and F.L.); Natural Science Foundation of Fujian Province of China (2020J01028 to J.H. and 2020J05012 to M.T.); the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16040000 to J.D. and F.L.).

Data and code availability

All datasets analyzed in this study were published previously [31,34,35,50,64] (Supplementary Table S1 available online at http://bib.oxfordjournals.org/). The code of eNet was available via GitHub https://github.com/xmuhuanglab/eNet.

References

1. Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 1981;**27**:299–308.
2. Long HK, Prescott SL, Wysocka J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* 2016;**167**:1170–87.
3. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57–74.
4. Maurano MT, Humbert R, Rynes E, *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012;**337**:1190–5.
5. Blobel GA, Higgs DR, Mitchell JA, *et al.* Testing the super-enhancer concept. *Nat Rev Genet* 2021;**22**:749–55.
6. Hnisz D, Abraham BJ, Lee TI, *et al.* Super-enhancers in the control of cell identity and disease. *Cell* 2013;**155**:934–47.
7. Osterwalder M, Barozzi I, Tissieres V, *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 2018;**554**:239–43.
8. Perry MW, Boettiger AN, Levine M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc Natl Acad Sci U S A* 2011;**108**:13570–5.
9. Jinek M, Chylinski K, Fonfara I, *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;**337**:816–21.
10. Bahr C, von Paleske L, Uslu VV, *et al.* A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* 2018;**553**:515–20.
11. Fulco CP, Munschauer M, Anyoha R, *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* 2016;**354**:769–73.
12. Canver MC, Smith EC, Sher F, *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* 2015;**527**:192–7.
13. Huang J, Liu X, Li D, *et al.* Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. *Dev Cell* 2016;**36**:9–23.
14. Shin HY, Willi M, HyunYoo K, *et al.* Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat Genet* 2016;**48**:904–11.
15. Hay D, Hughes JR, Babbs C, *et al.* Genetic dissection of the alpha-globin super-enhancer in vivo. *Nat Genet* 2016;**48**:895–903.
16. Kai Y, Li BE, Zhu M, *et al.* Mapping the evolving landscape of super-enhancers during cell differentiation. *Genome Biol* 2021;**22**:269.
17. Cai W, Huang J, Zhu Q, *et al.* Enhancer dependence of cell-type-specific gene expression increases with developmental age. *Proc Natl Acad Sci U S A* 2020;**117**:21450–8.
18. Thomas HF, Kotova E, Jayaram S, *et al.* Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Mol Cell* 2021;**81**:969, e913–82.
19. Rao SS, Huntley MH, Durand NC, *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;**159**:1665–80.
20. Dixon JR, Selvaraj S, Yue F, *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012;**485**:376–80.
21. Lieberman-Aiden E, van Berkum NL, Williams L, *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009;**326**:289–93.
22. Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* 2019;**20**:437–55.
23. Song M, Pebworth MP, Yang X, *et al.* Cell-type-specific 3D epigenomes in the developing human cortex. *Nature* 2020;**587**:644–9.
24. Liu X, Zhang Y, Chen Y, *et al.* In situ capture of chromatin interactions by biotinylated dCas9. *Cell* 2017;**170**:1028, e1019–43.
25. Huang J, Li K, Cai W, *et al.* Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat Commun* 2018;**9**:943.
26. Schmitt AD, Hu M, Jung I, *et al.* A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep* 2016;**17**:2042–59.

27. Huang J, Marco E, Pinello L, *et al.* Predicting chromatin organization using histone marks. *Genome Biol* 2015;**16**:162.
28. Liu X, Chen Y, Zhang Y, *et al.* Multiplexed capture of spatial configuration and temporal dynamics of locus-specific 3D chromatin by biotinylated dCas9. *Genome Biol* 2020;**21**:59.
29. Tang F, Barbacioru C, Wang Y, *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 2009;**6**: 377–82.
30. Buenrostro JD, Wu B, Litzenburger UM, *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 2015;**523**:486–90.
31. Ma S, Zhang B, LaFave LM, *et al.* Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* 2020;**183**:1103, e1120–16.
32. Cao J, Cusanovich DA, Ramani V, *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 2018;**361**:1380–5.
33. Zhu C, Yu M, Huang H, *et al.* An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nat Struct Mol Biol* 2019;**26**:1063–70.
34. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol* 2019;**37**:1452–7.
35. Granja JM, Klemm S, McGinnis LM, *et al.* Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol* 2019;**37**:1458–65.
36. Sarropoulos I, Sepp M, Fromel R, *et al.* Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells. *Science* 2021;**373**(6558):eabg4696.
37. Argelaguet R, Clark SJ, Mohammed H, *et al.* Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* 2019;**576**: 487–91.
38. Trevino AE, Muller F, Andersen J, *et al.* Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell* 2021;**184**(19):5053–69.
39. Ziffra RS, Kim CN, Ross JM, *et al.* Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* 2021;**598**: 205–13.
40. Li YE, Preissl S, Hou X, *et al.* An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* 2021;**598**:129–36.
41. Pliner HA, Packer JS, McFaline-Figueroa JL, *et al.* Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol Cell* 2018;**71**:858, e858–71.
42. Barabasi A-L. *Network Science*. Cambridge, England, Cambridge University Press, 2016.
43. Tsai A, Alves MR, Crocker J. Multi-enhancer transcriptional hubs confer phenotypic robustness. *Elife* 2019;**8**:e45325.
44. Pinero J, Bravo A, Queralt-Rosinach N, *et al.* DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* 2017;**45**: D833–9.
45. Siepel A, Bejerano G, Pedersen JS, *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;**15**:1034–50.
46. Welter D, MacArthur J, Morales J, *et al.* The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014;**42**:D1001–6.
47. Khan A, Zhang X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res* 2016;**44**:D164–71.
48. Tang Z, Luo OJ, Li X, *et al.* CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* 2015;**163**:1611–27.
49. Mumbach MR, Satpathy AT, Boyle EA, *et al.* Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet* 2017;**49**: 1602–12.
50. Domcke S, Hill AJ, Daza RM, *et al.* A human cell atlas of fetal chromatin accessibility. *Science* 2020;**370**(6518):eaba7612.
51. Yu G, Wang LG, Han Y, *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;**16**:284–7.
52. Lambert SA, Jolma A, Campitelli LF, *et al.* The human transcription factors. *Cell* 2018;**172**:650–65.
53. Malin J, Aniba MR, Hannenhalli S. Enhancer networks revealed by correlated DNAse hypersensitivity states of enhancers. *Nucleic Acids Res* 2013;**41**:6828–38.
54. Chen X, Zhou J, Zhang R, *et al.* Tissue-specific enhancer functional networks for associating distal regulatory regions to disease. *Cell Syst* 2021;**12**:353, e356–62.
55. Xu Q, Georgiou G, Frolich S, *et al.* ANANSE: an enhancer network-based computational approach for predicting key transcription factors in cell fate determination. *Nucleic Acids Res* 2021;**49**: 7966–85.
56. Carleton JB, Berrett KC, Gertz J. Multiplex enhancer interference reveals collaborative control of gene regulation by estrogen receptor alpha-bound enhancers. *Cell Syst* 2017;**5**:333, e335–44.
57. Lin X, Liu Y, Liu S, *et al.* Nested epistasis enhancer networks for robust genome regulation. *Science* 2022;**377**(6610):1077–85.
58. Gonzalez AJ, Setty M, Leslie CS. Early enhancer establishment and regulatory locus complexity shape transcriptional programs in hematopoietic differentiation. *Nat Genet* 2015;**47**: 1249–59.
59. Shu M, Hong D, Lin H, *et al.* Single-cellchromatin accessibility reveals enhancer networks driving gene expression during spinal cord development. *Dev Cell* (accepted).
60. Ranzoni AM, Tangherloni A, Berest I, *et al.* Integrative single-cell RNA-seq and ATAC-seq analysis of human developmental hematopoiesis. *Cell Stem Cell* 2021;**28**:472, e477–87.
61. Blake JA, Baldarelli R, Kadin JA, *et al.* Mouse Genome Database (MGD): knowledgebase for mouse-human comparative biology. *Nucleic Acids Res* 2021;**49**:D981–7.
62. Krupke DM, Begley DA, Sundberg JP, *et al.* The mouse tumor biology database: a comprehensive resource for mouse models of human cancer. *Cancer Res* 2017;**77**:e67–70.
63. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–2.
64. Cao J, O'Day DR, Pliner HA, *et al.* A human cell atlas of fetal gene expression. *Science* 2020;**370**(6518):eaba7721.