

Joint reconstruction of cis-regulatory interaction networks across multiple tissues using single-cell chromatin accessibility data

Kangning Dong and Shihua Zhang

Corresponding author: Shihua Zhang, Tel./Fax: +86 01 82541360; E-mail: zsh@amss.ac.cn

Abstract

The rapid accumulation of single-cell chromatin accessibility data offers a unique opportunity to investigate common and specific regulatory mechanisms across different cell types. However, existing methods for cis-regulatory network reconstruction using single-cell chromatin accessibility data were only designed for cells belonging to one cell type, and resulting networks may be incomparable directly due to diverse cell numbers of different cell types. Here, we adopt a computational method to jointly reconstruct cis-regulatory interaction maps (JRIM) of multiple cell populations based on patterns of co-accessibility in single-cell data. We applied JRIM to explore common and specific regulatory interactions across multiple tissues from single-cell ATAC-seq dataset containing ~80 000 cells across 13 mouse tissues. Reconstructed common interactions among 13 tissues indeed relate to basic biological functions, and individual cis-regulatory networks show strong tissue specificity and functional relevance. More importantly, tissue-specific regulatory interactions are mediated by coordination of histone modifications and tissue-related TFs, and many of them may reveal novel regulatory mechanisms.

Key words: cis-regulatory interaction networks; single-cell ATAC-seq; Gaussian graphical LASSO

Introduction

Cis-regulatory elements (CREs) are a key class of regulatory DNA sequences and typically regulate the transcription of target genes by binding to transcription factors (TFs) [1, 2]. Lots of efforts have been made to characterize combinatorial patterns and systematically define CREs [3–6]. Furthermore, interactions between CREs are vital components of genetic regulatory networks, controlling cell type-specific biological processes [7–9]. However, linking CREs to their target genes is still a challenging problem due to their distal distance (in some cases hundreds of kilobases) and complicated regulatory mechanisms. Moreover,

direct DNA contacts could be inferred using computational methods (e.g. CHiCAGO [10]) from data of the chromosome conformation capture (3C) technique and its variants such as ChIA-PET [11], Hi-C [12] and capture Hi-C [13].

But thus far, these types of data are only available for a few cell types. As a result, several computational methods have been developed to estimate cis-regulatory DNA interactions using epigenetic data [14–18]. For example, Corces *et al.* [14] proposed a computational strategy based on the correlation of transcriptional expression and chromatin accessibility to identify promoter-enhancer interactions. Cao *et al.* [15]

Kangning Dong is a PhD candidate of Academy of Mathematics and Systems Science, Chinese Academy of Sciences. His research interests include bioinformatics, single-cell omics, epigenomics and machine learning.

Shihua Zhang is a Professor at NCMIS, CEMS, RCSDS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences. His research interests are mainly in bioinformatics, single-cell omics, cancer genomics, epigenomics and machine learning.

Submitted: 20 March 2020; Received (in revised form): 16 May 2020

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

adopted a random forest model to reconstruct enhancer-target networks using both histone modification data and chromatin accessibility data. However, the above methods were all designed to infer CREs interactions from bulk sequencing data. The limited number of samples per cell type makes it hard to generate robust genome-wide *cis*-regulatory interaction maps.

Fortunately, the advent of single-cell ATAC-seq technology has enabled the genome-wide profiling of chromatin accessibility at single-cell resolution [19, 20]. Chromatin accessibility is a hallmark of active DNA regulatory elements, which delineates the *in vivo* availability of binding sites to TFs [21, 22]. These large-scale datasets ($10^4 \sim 10^5$ cells) provide a unique opportunity to detect cell type-specific CREs, assess the gene regulatory landscape, model the transcription factor regulatory grammar and reconstruct CREs interaction maps robustly using statistical models [23–26]. For example, Cicero [26] employs a graphical lasso model to detect linkages between CREs based on co-accessibility patterns in single-cell data. Co-accessible DNA elements have been illustrated to be functionally related. Specifically, they exhibit physical proximity, are mediated by interacting TFs and undergo coordinated changes in histone modifications.

When working with single-cell data, identifying the differences between different cell types or the dynamic changes during cell development is a vital problem. Therefore, one might want to compare *cis*-regulatory interaction networks estimated from different cell types to investigate the common and specific regulatory patterns. The strategy adopting Cicero to reconstruct regulatory networks for each cell type, respectively, has several deficiencies: (1) resulting networks may be incomparable directly due to diverse numbers of cells as well as data sparsity in different cell types; (2) estimating separate networks does not exploit the similarity between multiple cell types, which will cause inaccurate results for identifying common patterns; (3) when we explore the cell type-specific *cis*-regulatory interactions, the uncorrelated technology noises of single-cell data from different cell types, which are caused by amplification process and batch effects [27], could lead to many false positives. Consequently, more accurate estimations are expected by considering multiple cell types simultaneously and alleviating uncorrelated technology noises.

To this end, we adopt a statistical method to jointly reconstruct *cis*-regulatory interaction maps (JRIM) of multiple cell populations based on patterns of co-accessibility in single-cell chromatin accessibility data. JRIM employs a group lasso penalty to encourage a similar pattern of sparsity across all the regulatory networks and alleviate uncorrelated technology noises. We applied JRIM to explore common and specific regulatory patterns across *cis*-regulatory interaction networks of different tissues using single-cell ATAC-seq dataset generated in [23], which profiled $\sim 80,000$ cells across 13 adult mouse tissues. Results show that common interactions of all tissues are significantly enriched in housekeeping gene regions, and common interactions of similar functional tissues are involved in regulations of tissue-shared biological processes. Furthermore, tissue-specific differential activity genes perform specific biological functions and have higher transcriptional expressions in the corresponding tissues. More interestingly, tissue-specific functional peaks are enriched for histone modification marks and motifs of TFs that are known to be important for tissue-specific functions. Last but not least, the reconstructed *cis*-regulatory interaction networks reveal distinct regulatory mechanisms of the sodium channel gene *Scn5a* and identify a verified liver-specific promoter-enhancer loop of clock-controlled gene *Gys2*.

Materials and methods

Materials

Chromatin accessibility profiles at single-cell resolution in 13 adult mouse tissues have been generated using single-cell ATAC-seq technology in a recent study [23]. We downloaded this single-cell ATAC-seq data from GSE111586. These 13 tissues include bone marrow, cerebellum, large intestine (including cecum and colon), heart, kidney, liver, lung, prefrontal cortex, small intestine (including duodenum, jejunum, and ileum), spleen, testes, thymus and whole brain. To avoid batch effects, we only selected cells from one sample for each tissue. The processed single-cell ATAC-seq dataset contains $\sim 80,000$ cells and has called $\sim 400,000$ differentially accessible peaks (potential CREs) using MACS2 [28] as described in [23]. The total number of cells profiled per tissue ranges widely from 2278 for cerebellum to 8991 for heart (Supplementary Figure S1A).

We collected available single-cell RNA-seq datasets for 11 tissues (except prefrontal cortex and whole brain). The single-cell RNA-seq data of cerebellum were collected from Dropviz database [29]. Those of large intestine and heart were collected from Tabula Muris database [30]. Those of other tissues were obtained from [31]. We also collected available chromatin modification ChIP-seq datasets for 10 tissues, including H3K4me1, H3K4me3, H3K27ac and CTCF (except large intestine, prefrontal cortex and whole brain) from the ENCODE portal (<https://www.encodeproject.org/>) [32]. And the genome-wide occupancy profiles of TBX3 in heart measured by ChIP-seq were collected from [33]. In addition, we obtained mouse housekeeping gene set and 27 consistently expressed genes from [34] and collected tissue-specific differential expression genes of 12 tissues (except prefrontal cortex) from two mouse tissue-specific gene datasets [34, 35].

JRIM

Let X^k be a $M \times N_k$ matrix representing the binary accessibility values of the k -th tissue ($k = 1, \dots, K$), where X_{ij}^k is 1 if one or more reads are observed to overlap the accessible peak i in cell j of tissue k , and 0 otherwise. JRIM takes the binary matrices of all K tissues as input (Figure 1A) and consists of three main steps as follows.

Step 1. Grouping similar cells per tissue. Sparsity of single-cell epigenetic data is intrinsic due to the limited DNA copy number, and the percentage of 0 in used single-cell ATAC-seq data is $>98\%$ (Supplementary Figure S1B). The exceeding sparsity may cause several computational issues. For example, sample covariance matrix may be singular. To reduce the sparsity, control the technical variation and improve the stability of estimation, JRIM adopts a similar grouping procedure used as Cicero. This procedure can be regarded as a type of ‘bagging’ [36].

JRIM first maps cells in each tissue into low dimensional space using a t-SNE map [37] and constructs a k -nearest neighbor graph based on the two-dimensional t-SNE embedding via the FNN package (<https://CRAN.R-project.org/package=FNN>) (Supplementary Figure S2). Then, JRIM samples cells randomly and clusters their k nearest neighbors (default $k=50$). Binary accessibility profiles of cells in a group are aggregated into integer counts to construct the grouped matrix if a group does not overlap any existing one with more than 80% cells (a tunable parameter directly related to the number of groups). This step will be performed for each tissue separately and generates grouped matrices A^1, A^2, \dots, A^K . Note that a cell will sometimes belong to more than one group. The different

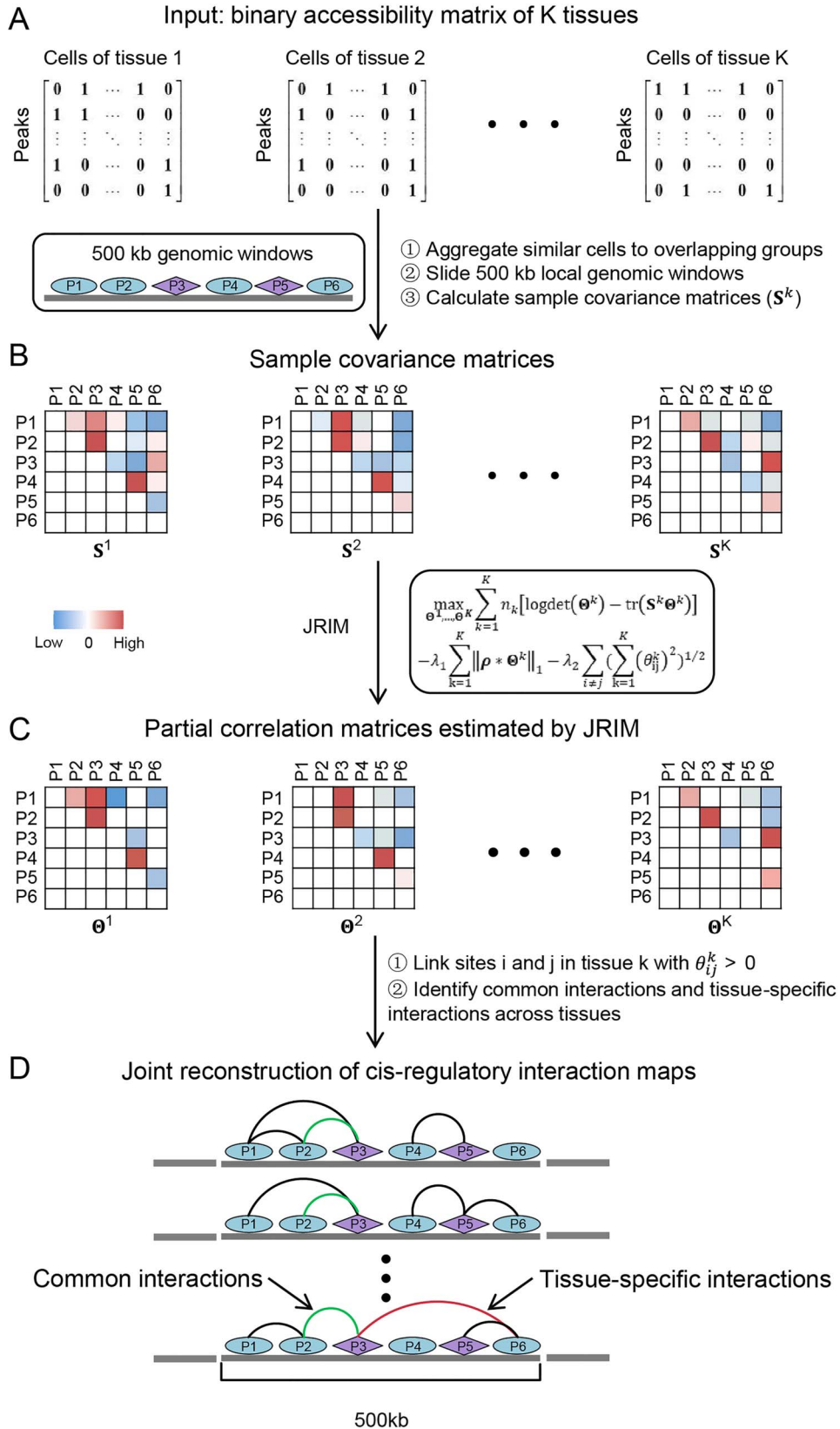


Figure 1. Illustration of the workflow of JRIM. (A) The input is binary single-cell chromatin accessibility data of multiple tissues. (B) JRIM aggregates similar cells to overlapping groups per tissue to overcome sparsity of single-cell data and calculates sample covariance matrices for local genomic windows. The purple diamond represents peaks overlapped with gene promoter and the ellipse represents remaining peaks. (C) JRIM jointly estimates local partial correlation matrices of K tissues and identifies co-accessible DNA element pairs, that is, cis-regulatory interactions. (D) Reconcile local regulatory interaction networks to achieve the reconstruction of genome-wide cis-regulatory interaction networks of all K tissues.

initializations of random selection will slightly influence the results (Supplementary Methods).

We further normalize these grouped matrices $\mathbf{A}^1, \mathbf{A}^2, \dots, \mathbf{A}^K$ to alleviate the impact of heterogeneous sequencing depths. Explicitly, each matrix \mathbf{A}^k is divided by a group-wise scaling factor computed by `estimateSizeFactors()` function of a R package `Monocle2` and then performs log-transformation with `pseudo-count 1`.

Step 2: Adopting joint graphical lasso to compute co-accessibility scores between accessible peaks in local genomic windows. JRIM jointly estimates sparse inverse covariance matrices, which encode partial correlations between variables, to capture the co-accessibility structure of accessible peaks via a joint graphical lasso model [38] (Figure 1C). Formally, JRIM employs the joint graphical lasso to maximize the following objective function:

$$\max_{\Theta^1, \dots, \Theta^K} \sum_{k=1}^K n_k \left[\log \left\{ \det \left(\Theta^k \right) \right\} - \text{tr} \left(\mathbf{S}^k \Theta^k \right) \right] - \lambda_1 \sum_{k=1}^K \left\| \rho * \Theta^k \right\|_1 - \lambda_2 \sum_{i \neq j} \left(\sum_{k=1}^K \left(\theta_{ij}^k \right)^2 \right)^{1/2},$$

where Θ^k is the inverse covariance matrix of the aggregated count matrix \mathbf{A}^k , which captures the conditional dependence structure of accessible peaks in tissue k , and \mathbf{S}^k is the sample covariance matrix of \mathbf{A}^k . n_k is the number of cell groups in tissue k , λ_1 and λ_2 are two nonnegative tuning parameters, the matrix ρ encodes the distance penalty between accessible peaks and $*$ denotes component-wise multiplication.

The first term of this formula is the maximum likelihood of estimating K graphical lasso models [39], respectively. To extract meaningful networks, the joint graphical lasso model expects only a small fraction of pairs to be partially correlated, i.e. to be nonzero in the inverse covariance matrices. Therefore, the second term is a l_1 -norm penalty to increase sparsity of the resulting inverse covariance matrices. The matrix ρ is determined by the following equation to induce distance penalty such that long-distance pairs have a relatively high penalty factor:

$$\rho_{ij} = 1 - d_{ij}^{-s},$$

where d_{ij} is the distance in the genome (in kilobases) between sites i and j , and s is a constant (s is set to 0.75 by default as suggested in Cicero).

The third term of this equation is the group lasso penalty [40] across all K inverse covariance matrices. This group lasso penalty encourages a similar pattern of sparsity across all the inverse covariance matrices—i.e. an interaction is encouraged to be 0s in all the K estimated inverse covariance matrices if it only shows nonsignificant co-accessibility in a few tissues. This penalty helps to remove uncorrelated technology noises and exploit the similarity between multiple cell types.

Furthermore, as promoter capture Hi-C data suggested, >75% of three-dimensional promoter-based interactions occur within a 500 kb distance [41]. Therefore, JRIM pays more attention to local cis-regulatory interactions and estimates co-accessibility structure within overlapping 500 kb genomic windows (windows are spaced by 250 kb such that any accessible site is covered by two genomic windows).

Step 3: Constructing genome-wide cis-regulatory interaction maps. JRIM calculates partial correlations between local accessible peaks for every 500 kb genomic window. Then we

will reconcile local interaction networks to construct genome-wide cis-regulatory interaction networks (Figure 1D). Note that many pairs of peaks are covered by two local co-accessibility networks and have two co-accessibility scores. In principle, these two co-accessibility scores have the same sign (>95% in our experiment). If so, the mean score of these two pairs is assigned to the interaction in the resulting genome-wide cis-regulatory interaction networks. If not, such pairs are considered undetermined, and their co-accessibility values are set as zero. Finally, JRIM reconstructs genome-wide cis-regulatory networks where nodes are the accessible peaks, and edges link two co-accessible DNA elements which have co-accessibility scores above a user-defined threshold (set as 0 in this paper to avoid the impact of different distribution of co-accessibility scores across tissues).

Selection of tuning parameters

The tuning parameters λ_1, λ_2 could control the sparsity and similarity of regulatory networks across all K tissues. However, both tuning parameters contribute to sparsity: λ_1 encourages an individual network to be sparse and λ_2 drives interaction intensity to be 0s in all K networks. Therefore, as joint graphical lasso [38] recommended, we reparametrize λ_1, λ_2 as follows:

$$\omega_1 = \lambda_1 + \frac{1}{\sqrt{2}} \lambda_2,$$

$$\omega_2 = \frac{1}{\sqrt{2}} \lambda_2 / \left(\lambda_1 + \frac{1}{\sqrt{2}} \lambda_2 \right),$$

where ω_1 and ω_2 , respectively, reflect the levels of sparsity and similarity. To determine values of ω_1 and ω_2 , JRIM first randomly selects 50 random 500 kb genomic windows and calculates the minimum ω_1 with a fixed ω_2 (a relatively small value) such that nonzero entries in $\Theta^1, \dots, \Theta^K$ are less than 10% of all entries and no more than 5% long-distance pairs (pairs of sites at a distance greater than 250 kb) are nonzeros. As for parameter ω_2 , we defined a similarity measurement as below:

$$S = \left(\sum_{i \neq j} c_{ij} \right) / (2 \times C_K^2 \times M),$$

where c_{ij} denotes the number of common nonzero pairs in tissues i and j , K is the number of tissues, C_K^2 is the number of permutations and M is the mean number of nonzero pairs in all tissues. Then in the same 50 genomic windows, we fix ω_1 as the mean of calculated values of ω_1 and determine the minimum ω_2 value such that measurement S is more than 0.6. Finally, the mean values of the calculated parameters are used for setting the penalties for each of 500 kb genomic window sub-problem.

Mapping cis-regulatory interactions to various genomic features

We obtained the genomic features including promoter regions (200 bp upstream TSS and 100 bp downstream TSS), 5' UTR, 3' UTR, exons, introns and intergenic regions, which are located further than 1 kb from any RefSeq gene, based on mm9 mouse reference genome from the UCSC Genome Browser [42]. Firstly, peaks were categorized into the six feature categories if they overlap with any one genomic feature region. Then, cis-regulatory interactions were assigned into the same six feature categories according to accessible peaks they were associated. Note that one peak may be mapped to more than one features and so do the interactions. We also linked the regulatory

regions to their target genes based on gene promoter-related interactions.

Definition of tissue-specific differential activity genes (DAGs)

We depicted the activity of genes in a given tissue by the number of gene promoter-related interactions in the reconstructed cis-regulatory interaction network. For gene l , we denoted the gene activity score vector as $GA_l = \{GA_l^1, GA_l^2, \dots, GA_l^{13}\}$, where GA_l^k represents the number of promoter-related interactions of gene l in the k th tissue. Then we normalized gene activity scores through dividing them by the total number of interacting co-accessible pairs in the corresponding tissue and calculated the z-scores of normalized gene activity scores \hat{GA}_l :

$$ZGA_l = \{z_l^1, z_l^2, \dots, z_l^{13}\}, \text{ where } z_l^k = (\hat{GA}_l^k - \mu(\hat{GA}_l)) / \sigma(\hat{GA}_l),$$

The gene l is considered as the k th tissue-specific differential activity one if $z_l^k > 2$.

Exploring transcription and chromatin modification levels of tissue-specific DAGs

To test the transcription regulation functions of estimated cis-regulatory interactions, we investigated transcription levels and transcription specificity of DAGs using single-cell RNA-seq data. We first normalized the single-cell RNA-seq data of 11 tissues using the R package *Seurat* [43]. For each tissue, genes were divided into two groups: 'tissue-specific DAGs' and 'other genes'. We first calculated the mean expression of these two groups in their corresponding tissue. To characterize transcription specificity of DAGs, we also calculated the mean expression of 'tissue-specific DAGs' in other tissues.

We also investigated chromatin modification marks level around TSSs of DAGs to access the regulatory mechanisms. We calculated the mean signal of collected chromatin modification marks (H3K4me1, H3K4me3, H3K27ac and CTCF) in the promoter region of all genes and normalized these signals to have the same mean value across all tissues. Then we grouped genes into two groups as in the analysis of gene expression data and compared their chromatin modification level.

Gene ontology enrichment analysis

To assess the functional relevance of reconstructed interaction networks, we performed gene ontology (GO) enrichment analysis for common interaction-related genes of similar functional tissues and tissue-specific differential activity genes via R package *clusterProfile* [44]. Common interactions of similar functional tissues are defined as interactions that only detected in immune tissues (spleen, thymus, bone marrow and lung) or nervous tissues (cerebellum, prefrontal cortex and whole brain). We only selected genes whose promoters are associated with more than one common interaction of similar functional tissues to perform GO enrichment analysis. In the enrichment analysis of DAGs, we selected top 200 differential activity ones for each tissue (gene number of kidney, prefrontal cortex and large intestine are less than 200). The enrichment significance is corrected by the Benjamini-Hochberg procedure.

Motif analysis in tissue-specific interaction-related regions

We defined the peaks related to tissue-specific promoter-related interactions as 'tissue-specific functional ones'. To investigate

the regulatory mechanisms of tissue-specific interactions, we employed the HOMER software (<http://homer.salk.edu/homer/>) [45] to detect transcription factor motifs that are enriched in 'tissue-specific functional peaks' (Table 1 and Supplementary Table S1). The background sequences were generated from all accessible peaks. The brief description of TFs is obtained from UniProtKB database (<https://www.uniprot.org/>).

Results

JRIM reconstructs comparable cis-regulatory networks of multiple tissues

We applied JRIM to jointly reconstruct cis-regulatory interaction maps of multiple mouse tissues from single-cell ATAC-seq data generated in [23]. For the 436 206 accessible peaks, we identified 2.28 million regulatory interactions in each tissue on average (ranging from 1.28 million in small intestine to 3.38 million in prefrontal cortex) (Supplementary Figure S3A). Compared with Cicero [26], JRIM generates tissue-specific networks with relatively more comparable and balanced interactions (Supplementary Figure S3B). Proportions of tissue-specific interactions (only detected in one tissue) by JRIM have significantly decreased compared to those by Cicero, indicating JRIM characterize the shared interactions well among multiple tissues. The interactions frequency decreases with the increasing of genomic distance (Supplementary Figure S4).









Note that the number of peak-associated interactions is higher in transcription-related genomic regions (especially in promoters and 5' UTR), indicating their transcriptional regulation roles, and lower in intron as well as intergenic regions (Figure 2A and Supplementary Figure S5). Moreover, we observed the same order when comparing conservation of interactions within different genomic regions, i.e. interactions associated with 5' UTRs and promoters showed the highest conservation across tissues (Figure 2B and Supplementary Figure S6), which is consistent with previous studies that DNA regulatory elements far from TSS exhibited a greater tissue specificity and wide dynamic range of activity [14, 26, 46].

The cis-regulatory networks reveal common interaction patterns among tissues

The comparable cis-regulatory interaction networks generated by JRIM enable direct identification of common or tissue-shared interactions among tissues. Note that the genes whose promoters are associated with common interactions of all 13 tissues are significantly enriched in housekeeping gene set (Figure 2C; $P < 1e-65$, Fisher's exact test), suggesting that these common interactions are involved in regulations of fundamental biological functions. We also observed that highly and constantly expression genes are regulated by more regulatory elements than remaining genes in all 13 tissues (Supplementary Figure S7). Both observations suggest that the common patterns among all tissues are related to basic biological functions.

We observed that similar functional tissues are clustered together based on the overlapping ratio of their common interactions, including nervous tissues (cerebellum, prefrontal cortex and whole brain) and immune tissues (spleen, thymus, bone marrow and lung) (Figure 2D). We also found that tissues within these two clusters have more common interactions than random selected ones (Figure 2E). GO enrichment analysis of genes related to common interactions of similar functional tissues

Table 1. Examples of motifs enriched in the ‘tissue-specific functional peaks’. Examples are selected from the top two enriched motifs of each tissue

Tissue	Motif	P-value	TF	Brief description	Ref.
Spleen		1e-136	ELF4	ELF4 plays a role in the development and function of NK and NK T-cells.	(47)
		1e-123	ETS1	ETS1 controls the differentiation, survival and proliferation of lymphoid cells.	(48)
Kidney		1e-282	PPARA	PPARA is a key regulator of lipid metabolism and peroxisomal β -oxidation pathway of fatty acids.	(49)
		1e-277	HNF4A	HNF4A binds fatty acids and may be essential for development of the liver, kidney and intestine.	(50)
Prefrontal cortex		1e-143	NEUROD1	NEUROD1 associates with chromatin to enhancer regulatory elements in neurogenesis regulation genes.	(51)
Cerebellum		1e-100	NEUROD1	The same as above.	(51)
Whole brain		1e-164	NEUROD1	The same as above.	(51)
Heart		1e-153	TBX5	TBX5 regulates the transcription of ion channel genes and is essential for heart development.	(52)

(Materials and Methods) reveals distinct common biological processes of these tissues (Figure 2F). For example, genes of common interactions for the nervous tissues are strongly enriched in ‘synapse organization’, ‘cognition’ and ‘learning or memory’ functional terms, and those of immune system are significantly enriched in ‘T cell activation’, ‘myeloid leukocyte differentiation’ and so on, indicating that the common interactions among tissues play key roles in regulations of their sharing biological functions.

The reconstructed interaction networks show strong tissue specificity and functional relevance

Similar sparsity of cis-regulatory interaction networks enables quantification of the changes of gene activity in different tissues. Tissue-specific differential expression genes from two datasets [34, 35] both show higher activities in the corresponding tissue except tissue-specific genes of small intestine as well as the spleen-specific genes in TiSCeD dataset (Figure 3A; $P < 0.001$ by one-sample Wilcoxon tests). Interestingly, the overlapping parts of two tissue-specific gene datasets tend to exhibit higher activity specificity compared with tissue-specific genes from an individual dataset. And the similar functional tissues (nervous tissues and immune tissues) are also clustered together in terms of gene activity scores (Supplementary Figure S8). These results shed light on the underlying transcription regulation functions of cis-regulatory interactions and imply that tissue-specific gene expression is regulated by promoter-enhancer interactions.

Functional enrichment analysis for tissue-specific differential activity genes (DAGs) (Materials and Methods) reveals that, except large intestine and small intestine, overrepresented GO

terms were indeed relevant to the distinct tissue-specific biological functions (Figure 3B). For example, tissue-specific DAGs of heart are enriched in cardiac muscle development-related GO terms, such as ‘cardiac muscle cell development’, ‘cardiac muscle tissue development’ and so on. Those of testis are closely related to ‘spermatid development’ and ‘germ cell development’. Genes related to metabolism terms such as ‘fatty acid metabolic process’ and ‘alcohol metabolic process’ are strongly enriched in liver. As for nervous tissues, overrepresented GO terms are both related to neuron development or cognition functions, such as ‘cognition’ and ‘learning’ terms in both cerebellum and whole brain. These results highlight the underlying role of regulatory interactions in performing tissue-specific biological functions.

Next, we investigated the transcriptional behavior of tissue-specific DAGs and observed that tissue-specific DAGs have significantly higher transcriptional levels than other genes in all 11 tissues (Figure 4A; $P < 0.001$, two sample Wilcoxon tests), highlighting the potential tissue-specific functions of DAGs. Moreover, tissue-specific DAGs from all tissues except those of small intestine exhibit strong transcriptional specificity (Figure 4A; $P < 0.01$ for thymus and $P < 0.001$ for others, two sample Wilcoxon tests). Thus, the reconstructed promoter-enhancer interactions are involved in gene activation to perform tissue-specific biological functions.

We also found that chromatin modification marks around TSSs of tissue-specific DAGs in the corresponding tissue exhibit strong difference compared with those in other tissues (Figure 4B and Supplementary Figure S9). Specifically, the active promoter mark H3K4me3 is strongly enriched in regions around TSSs of DAGs in most tissues (except heart and lung) (Figure 4B). And H3K4me3 also shows tissue-specific enrichment around

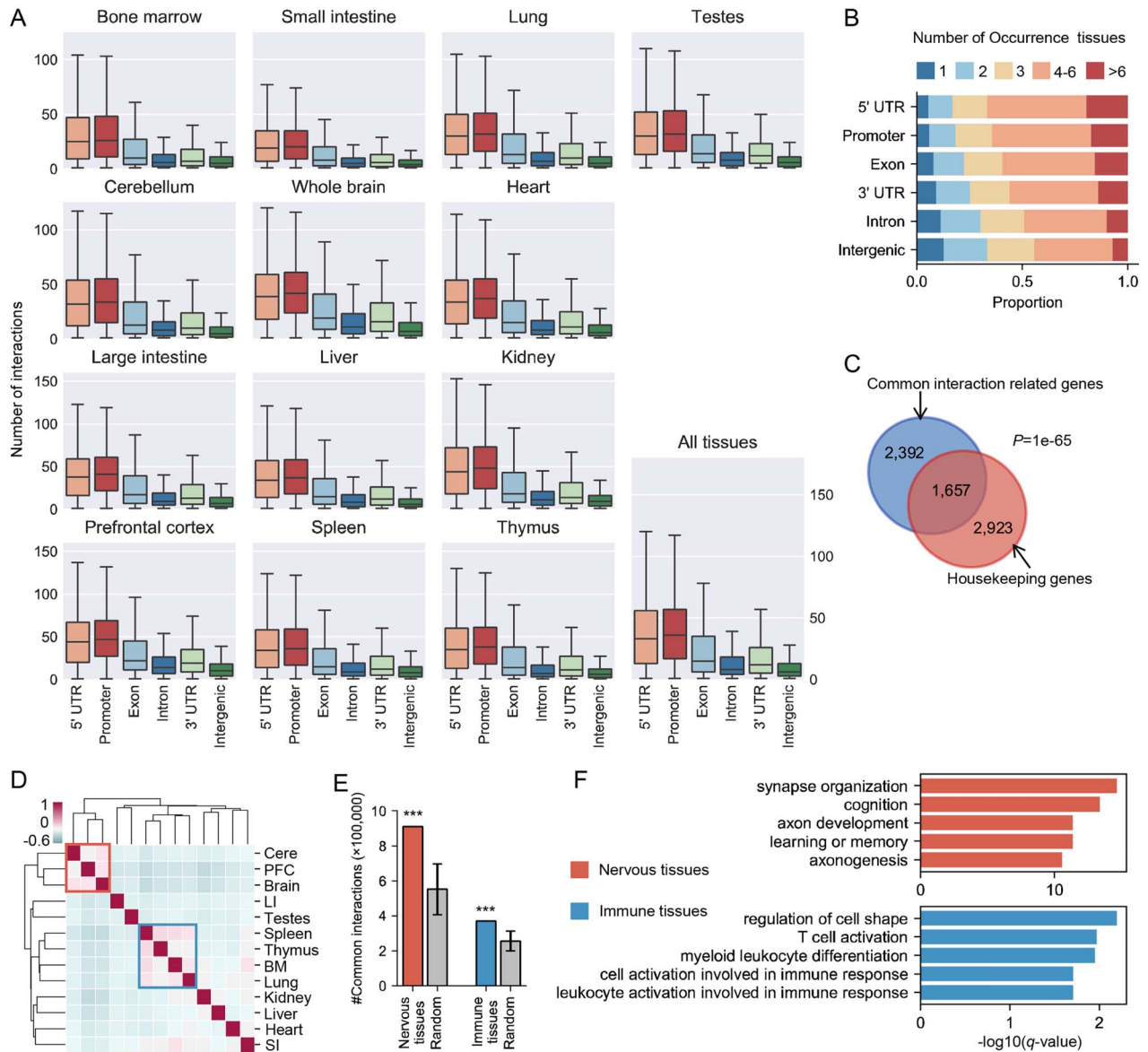


Figure 2. Basic characteristics of the reconstructed cis-regulatory interaction networks and the common interaction patterns across tissues. **(A)** Distribution of cis-regulatory interactions in six different genomic regions, including 5' UTR, promoter, exon, intron, 3' UTR and intergenic regions. **(B)** Distribution of conservation of genomic region-related interactions. Color reflects the number of tissues detecting corresponding regulatory interactions. **(C)** Venn diagram of the overlapping between common interaction-related genes and housekeeping gene set. The statistical significance was evaluated with Fisher's exact test. **(D)** Hierarchical clustering of 13 tissues in terms of overlapping ratio of their cis-regulatory interactions. The two similar functional tissue clusters are marked by blue and red boxes. The values of heatmap represent Pearson's correlation coefficients about the overlap ratio of cis-regulatory interactions. Cere: cerebellum; PFC: prefrontal cortex; LI: large intestine; BM: bone marrow; SI: small intestine. **(E)** Comparison of the number of common interactions of similar functional tissue clusters (nervous tissues and immune tissues) and those of random selection tissues. The error bars indicate the standard deviation and *** indicates $P < 0.001$ using one-sample Wilcoxon test. **(F)** GO terms enrichment in common interaction-related genes of similar functional tissues is associated with tissue sharing biological processes. Top 5 enriched GO terms of each cluster are shown.

TSSs of tissue-specific DAGs, except those of small intestine (Figure 4B), supporting the activity specificity of DAGs identified in the reconstructed networks by JRIM. Meanwhile, the signals of the architectural protein CTCF around TSSs of DAGs are also significantly higher in the corresponding tissue than those in other ones (Supplementary Figure S9A), indicating that the promoter of tissue-specific DAGs tend to be related to chromatin looping in the given tissues. Moreover, the H3K27ac mark, known to be associated with the activation of transcription, is also strongly enriched in promoter regions of tissue-specific DAGs (Supplementary Figure S9B). These observations indicate that

epigenetic features may contribute to the regulation functions of promoter-enhancer interactions.

Tissue-specific cis-regulatory interactions are mediated by coordination of tissue-related TFs and histone modifications

JRIM enables the identification of tissue-specific interactions and investigation of tissue-specific regulatory mechanisms. The tissue-specific functional peaks associated with tissue-specific

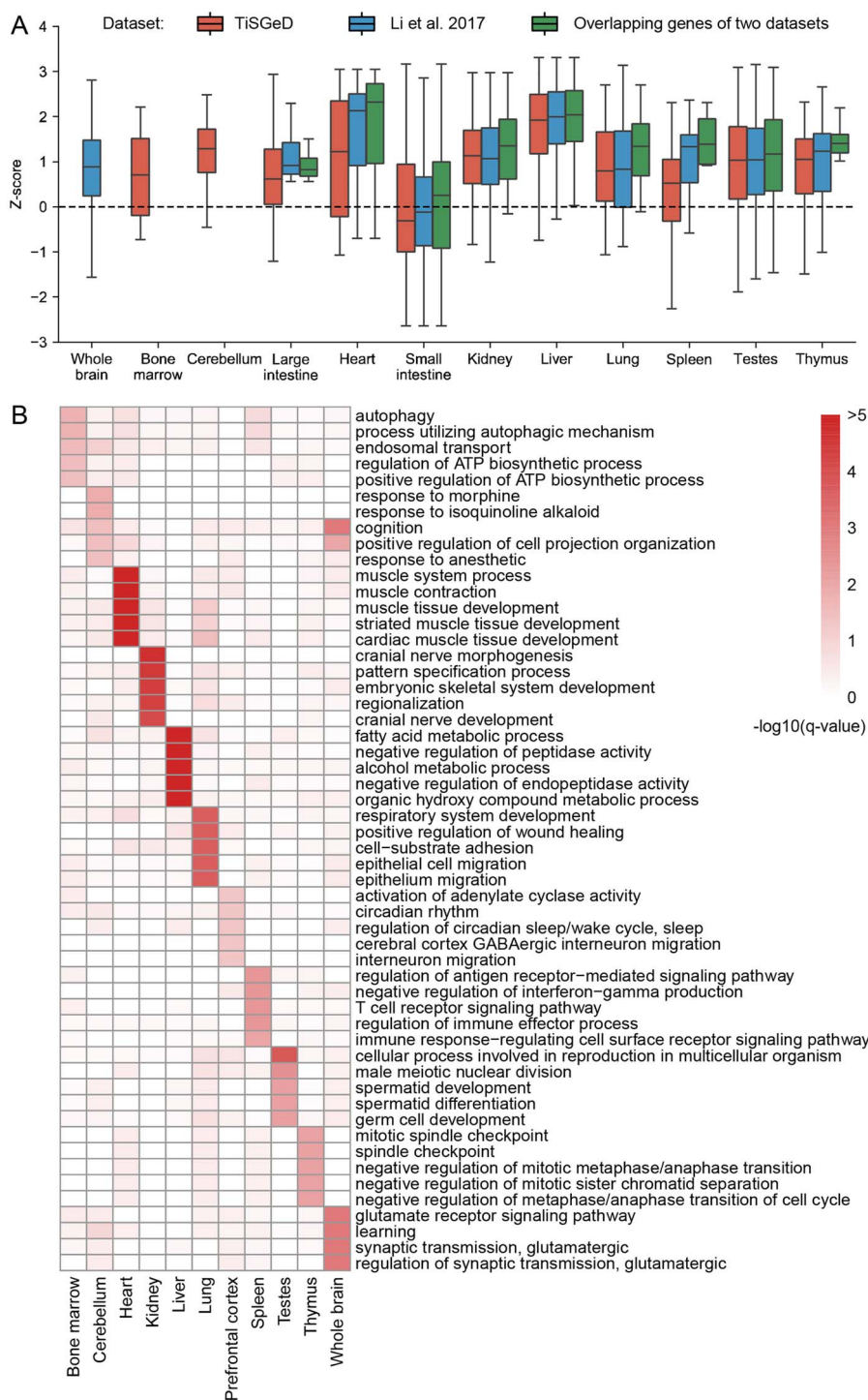


Figure 3. Tissue specificity and functional relevance of the reconstructed *cis*-regulatory interaction networks. (A) Boxplots of z-scores of two tissue-specific differential expression gene datasets and their overlapping genes. The z-scores are calculated by the number of promoter-associated interactions in the reconstructed interaction networks. The dashed line is the theoretical mean value of z-scores for randomly selected genes. (B) Differential activity genes for each tissue are enriched in tissue-specific biological processes. Top 5 enriched GO terms with q -value < 0.05 for each tissue are selected. $-\log_{10}(q$ -value) is used to plot this heatmap.

interactions are enriched for motifs of TFs that are functionally associated with tissue functions (Materials and Methods; Table 1 and Supplementary Table S1). For instance, motifs of ELF4 and ETS1 were significantly overrepresented in spleen-specific functional peaks ($P = 1e-136$ and $P = 1e-123$, respectively). Note that ELF4 plays a role in development of NK and NK T-cells [47],

and ETS1 controls the differentiation, survival and proliferation of lymphoid cells [48]. In other examples, motifs of PPARA and HNF4A are significantly enriched in kidney-specific functional peaks ($P = 1e-282$ and $P = 1e-277$, respectively). PPARA has been reported to be a key regulator of lipid metabolism [49], and HNF4A is essential for development of kidney [50]. Moreover,

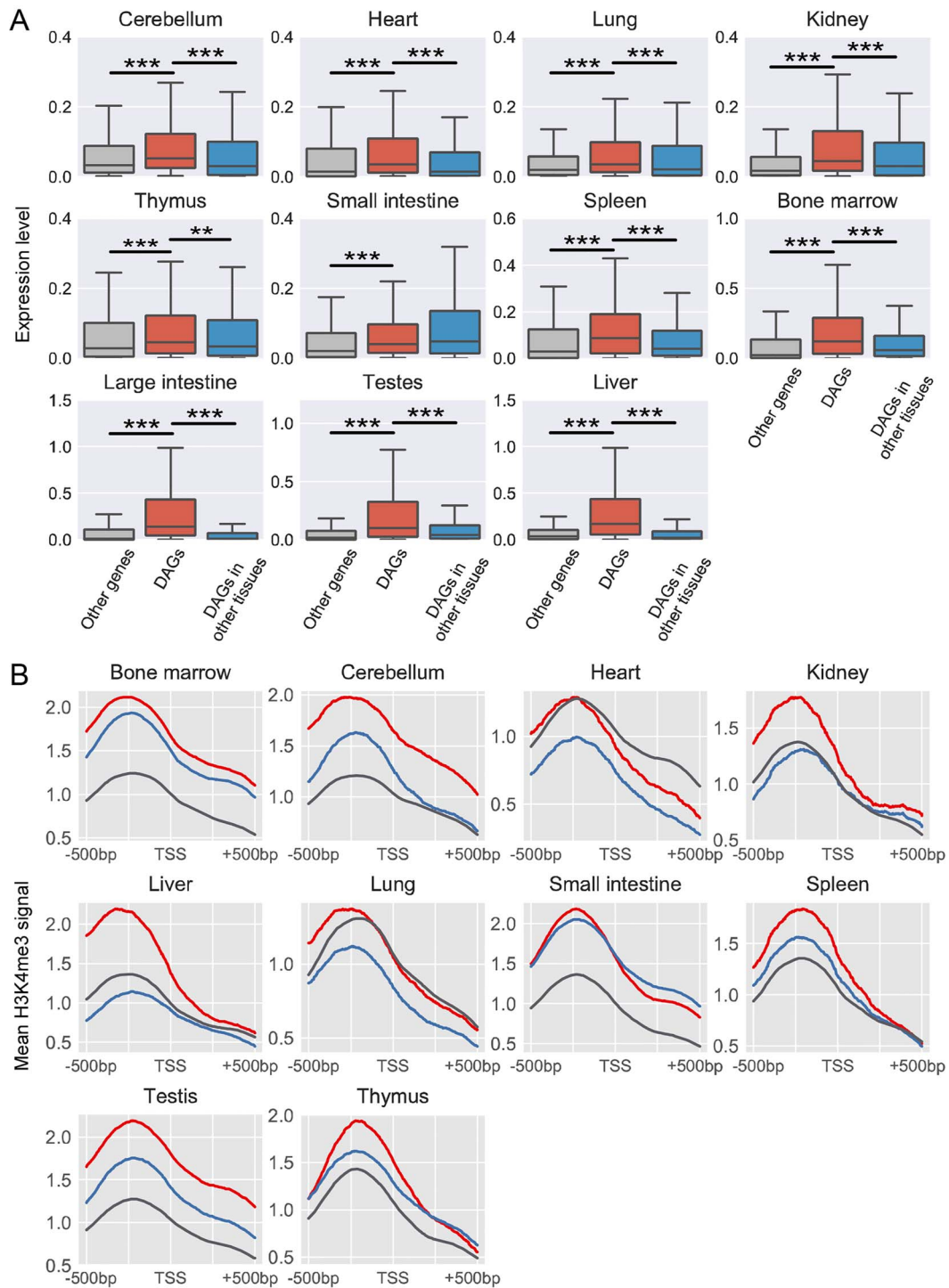


Figure 4. Transcription and histone modification levels of tissue-specific differential activity genes. **(A)** Boxplots of transcription levels of differential activity genes (DAGs) and other genes in the single-cell RNA-seq dataset. The middle red box represents the RNA expression level of DAGs in the corresponding tissues. The left gray box represents expressions of other genes. The right blue box represents these DAGs expression in other tissues. Each middle box was compared with both left and right boxes using two-sample Wilcoxon tests. ** and *** indicate $0.001 < P < 0.01$ and $P < 0.001$, respectively. **(B)** Enrichment analysis of H3K4me3 mark around TSSs of DAGs in ten tissues. The signals around DAGs TSSs and TSSs of remaining genes in the corresponding tissue are labeled with red and gray colors, respectively. The blue line is the mean signal around TSSs of DAGs in other tissues.

NEUROD1 was shown to be associated with neurogenesis [51], and motifs of NEUROD1 are highly enriched in prefrontal cortex, cerebellum and whole brain ($P = 1e-143$, $P = 1e-104$ and $P = 1e-164$, respectively). We also investigated histone modification markers in tissue-specific functional peaks. Indeed, the enhancer marks

(H3K27ac and H3K4me1) are significantly enriched in tissue-specific functional peaks (Figure 5A and Supplementary Figure S10A), highlighting the potential regulation function of these regions. Moreover, the CTCF marker is also enriched in tissue-specific functional peaks (Supplementary Figure S10B), which

might be related with tissue-specific chromatin looping. Overall, the reconstructed cis-regulatory networks by JRIM could reveal tissue-specific interactions, which are mediated by coordination of tissue-related TFs and histone modifications.

The cis-regulatory networks reveal distinct regulatory mechanisms of the sodium channel gene *Scn5a*

The cis-regulatory interaction network of heart around *Scn10a-Scn5a* locus provides a strong illustration of cooperative regulatory roles about TF binding and histone modification. Motifs of TBX5 are significantly enriched in heart-specific functional peaks (Table 1, $P=1e-153$). Note TBX5 is known to activate ion channel genes and is essential for heart development [52]. In addition, TBX3 is also critical for the development of heart, which competes with TBX5 for the same binding sites, but represses the expression of ion channels genes, such as the sodium channel gene *Scn5a* [53]. In the reconstructed cis-regulatory interaction network of heart, the occupancy profile of TBX3 shows distinctly higher enrichment in the peaks linked to *Scn5a* promoter A (accessible peak: chr9: 119,487,317-119,490,317) (Figure 5B), implying that *Scn5a*-related regulatory interactions in heart are mediated by bindings of TBX3 and TBX5. Meanwhile, the enhancer marks H3K27ac and H3K4me1 are also strongly enriched in *Scn5a* promoter-related regions (Figure 5B). More interestingly, with the increase of co-accessibility scores, the signals of TBX3 and H3K27ac also enhance a lot, but the signals of H3K4me1 are relatively stable. H3K27ac has been reported to be used to find active enhancers by subtracting H3K4me1 [54]. This suggests that some histone modifications contribute to the binding of TFs, and the binding of TFs in the active enhancer region has a greater impact on gene regulations.

The reconstructed cis-regulatory interactions-related to *Scn5a* promoters are associated with TBX3 enriched regions (Figure 5C). Two of these TBX3 enriched regions have been well-documented to modulate the expression of *Scn5a* [33, 55, 56] (a region located in *Scn10a* gene region labeled as enhancer A: chr9:119 540 800–119 544 032 and an intergenic region labeled as enhancer B: chr9:119 378 051–119 379 479). Specifically, Boogaard et al. [56] have leveraged high-resolution 4C-seq to investigate the chromosome structure between enhancer A, enhancer B and promoters of *Scn5a*, finding that these two enhancers and *Scn5a* promoters are in close contact and form a complex. They also engineered deletions of enhancer A and B and found that removal of these two enhancers significantly abrogates the ventricular conduction system and the expression of *Scn5a*, indicating the necessity of enhancer A and B for *in vivo* *Scn5a* expression regulation.

The validated interactions between enhancer B and *Scn5a* promoters were identified with high co-accessibility scores in the cis-regulatory network of heart, supporting the transcription regulation role of enhancer B (Figure 5C). However, as for enhancer A, we only observed the regulatory contacts between enhancer A and enhancer B in heart (Figure 5C). And the H3K4me1 signal in enhancer A region is relatively high, but the H3K27ac signal is low. We guessed that although the physical proximity between enhancer A and *Scn5a* promoters and the bindings of TBX3 and TBX5 in enhancer A region, enhancer A regulates the expression of *Scn5a* through indirect interactions (enhancer A-enhancer B-promoters of *Scn5a*) due to its low enhancer activity. In short, the regulatory network of heart around *Scn10a-Scn5a* locus reveals distinct regulatory mechanisms of *Scn5a*: regions enriched in both TBX3 bindings and H3K27ac are likely to directly regulate the expression of

Scn5a (region A, C-F and enhancer B in Figure 5C) and regions only enriched in TBX3 bindings tend to not interact with *Scn5a* promoters (region B and enhancer A in Figure 5C). Moreover, the regulatory mechanisms of enhancer A demonstrate that physical proximity does not definitively mean regulatory interaction. Therefore, reconstructing cis-regulatory networks using chromatin accessibility data could help to better understand regulatory mechanisms and guide downstream analysis (see another example in Supplementary Figure S11).

The cis-regulatory networks identify a liver-specific promoter–enhancer loop of clock-controlled gene *Gys2*

We further investigated the regulatory interaction networks around Glycogen Synthase 2 (*Gys2*) gene, which is a liver-specific clock-controlled gene, catalyzing the rate-limiting step in hepatic glycogen synthesis [57]. The transcript of *Gys2* is rhythmically expressed in liver during a day. In contrast, the accumulation of *Gys2* mRNA is constant and low in kidney [58]. We compared the *Gys2* promoter related interactions in all 13 tissues and found a liver-specific contact with a region of 12 to 21 kb downstream from *Gys2* TSS (Figure 6A). Two of three peaks located in this intragenic region exhibit higher co-accessibility with *Gys2* promoter in liver comparing to those in kidney (Figure 6B). A recent study investigated the clock-dependent chromatin topology of *Gys2* in kidney and liver [58]. They measured the interaction frequencies of *Gys2* promoter at different time points in liver and kidney using 4C-seq technology and found that the promoter of *Gys2* contacts with an intragenic region, 21 kb downstream from the TSS of *Gys2*, more frequently at the day versus at the night only in liver (Figure 6C and Supplementary Figure S12A). To verify this liver-specific promoter-enhancer loop, they cultured clock-deficient *Bmal1* knockout mice through hybridization and found that rhythms of this liver-specific interaction are abolished (Supplementary Figure S12B). This putative liver-specific promoter-enhancer loop is highly consistent with our observation in the reconstructed cis-regulatory networks.

Discussion

Exploring tissue-specific cis-regulatory interaction patterns across multiple cell types could help to investigate cell type-specific regulation. Here, we adopt JRIM to jointly reconstruct cis-regulatory interaction maps of multiple cell populations from single-cell chromatin accessibility data. JRIM controls the sparsity of resulting cis-regulatory interaction networks simultaneously and alleviates uncorrelated technological noises across different cell populations. These cis-regulatory interaction networks could advance the understanding of regulatory mechanisms and guide downstream analysis such as CRISPR-mediated (epi)-genome editing.

We believe that our analysis could also help to identify novel tissue-specific interactions. For example, comparing the cis-regulatory interaction networks of kidney and liver around *Gys2*, JRIM detects a long-range interaction within a region of 280–320 kb downstream from *Gys2* TSS only in kidney (Figure 6A). This region is suggested to be involved in a kidney-specific promoter-enhancer loop. As expected, the comparison of liver and kidney 4C-seq data showed significant increased contact frequency with the *Gys2* promoter within this region in kidney (Figure 6C). We employed *FourCSeq* to estimate the spatial interactions of *Gys2* promoter in liver and kidney using 4C-seq data [59], finding a long-range kidney-specific chromatin loop

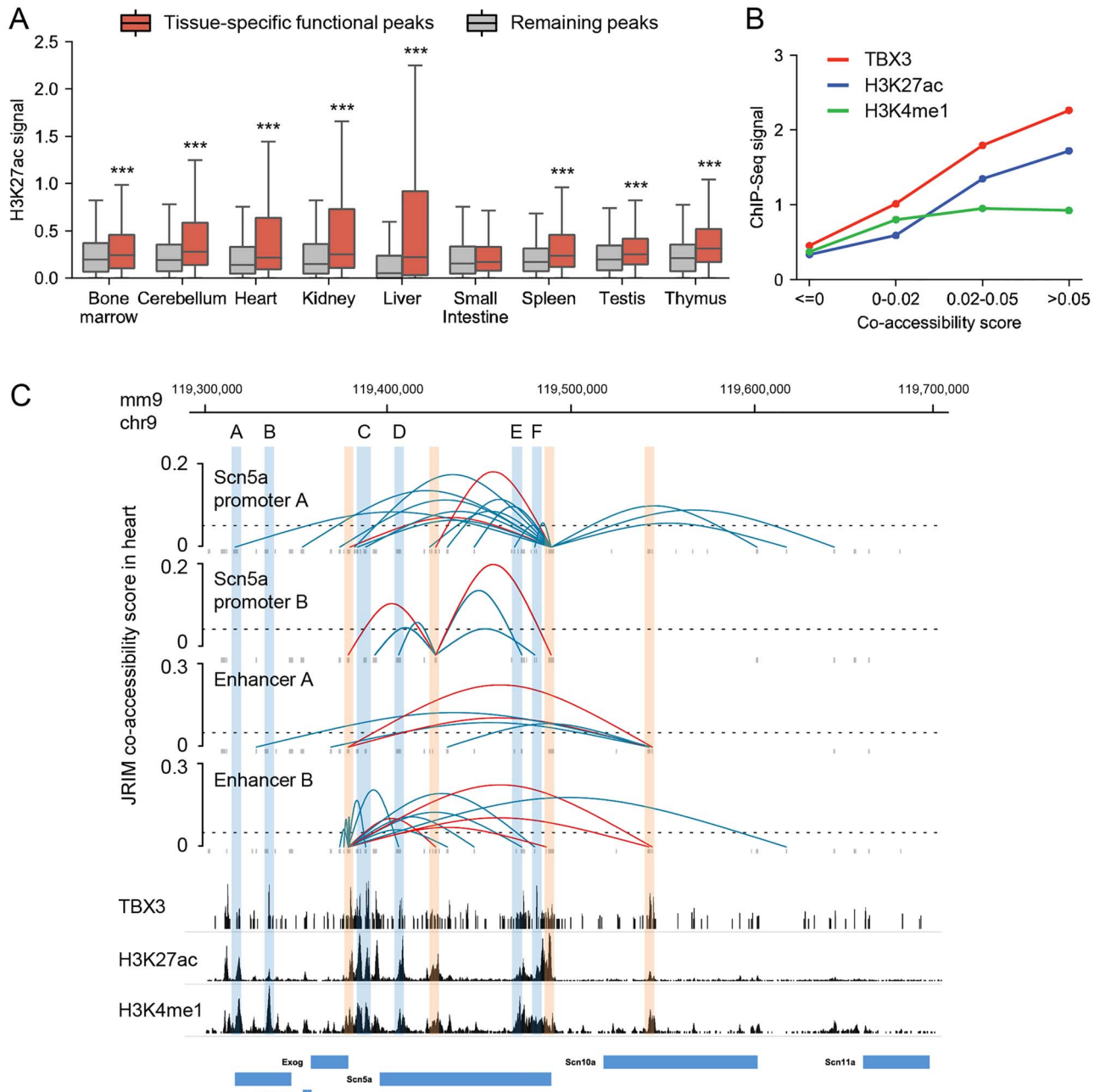


Figure 5. Illustration of the distinct regulatory mechanisms of *Scn5a*. (A) Boxplots of the H3K27ac signal of tissue-specific functional peaks compared to those of other peaks. (B) The occupancy profiles of TBX3, H3K4me1 signal and H3K27ac signal of *Scn5a* promoter-related peaks (co-accessibility score > 0) and other peaks. (C) The reconstructed cis-regulatory interaction network around the *Scn10a-Scn5a* locus (chr9: 119 300 000–119 700 000) in heart. The regions of *Scn5a* promoters and enhancer A, B are labeled with red boxes. Blue boxes indicate locations of TBX3 enriched regions A-F. The bottom is ChIP-seq data around *Scn5a* gene. The image is drawn on basis of WashU epigenome browser with RefSeq gene annotations.

within this region (Figure 6D and Supplementary Figure S13). Furthermore, we compared the histone modification marks in this region and observed that the enhancer marks (H3K4me1, H3K27ac) of peaks within this region in kidney are far greater than those in liver (Figure 6E). The physical proximity and the higher histone modification level in kidney indicated that this region modulates expression of *Gys2* through a long-range promoter-enhancer loop in kidney.

Although we focused on tissue level and did not emphasize the cell heterogeneity within tissues in this study, the benefits of considering cell heterogeneity can be demonstrated through

the poor functional relevance of small intestine-specific differential activity genes. Unlike other tissues, the profiled cells of small intestine were dissected from different regions including duodenum, jejunum, and ileum. The detected co-accessibility patterns (number of interactions) in small intestine are much fewer than other tissues (Supplementary Figure S3A), which is probably caused by cell heterogeneity. JRIM could be applied to chromatin accessibility data with respect to cell populations defined ahead to obtain accurate regulatory networks.

Comparing to Cicero, JRIM exploits the similarity between tissues and generates cis-regulatory networks that can be

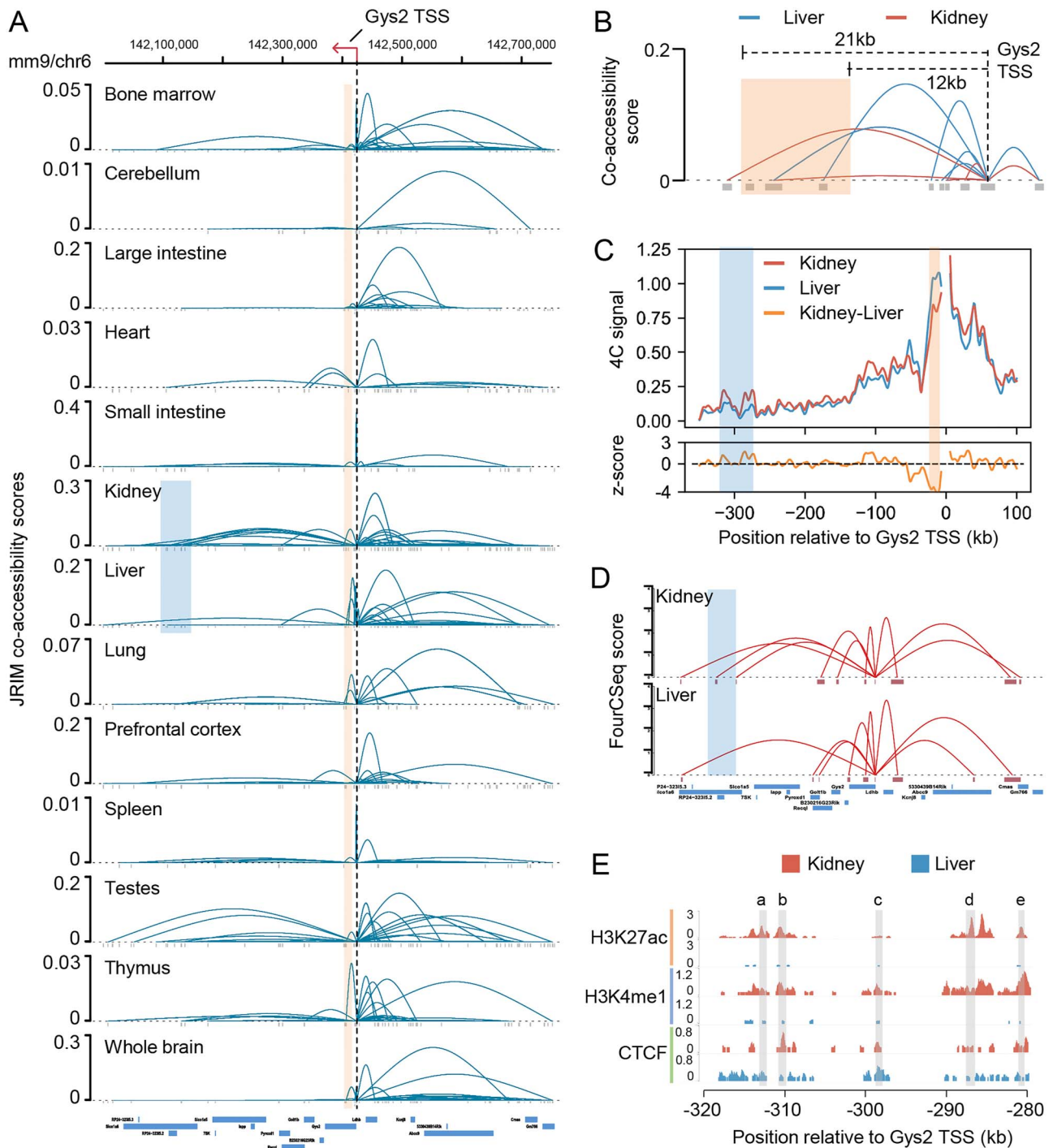


Figure 6. Illustration of two tissue-specific enhancer-promoter interactions revealed by the reconstructed networks. (A) Visualization of *Gys2* promoter-related interactions within region (chr6:142 000 000–142 750 000) in all 13 mouse tissues. The black dashed line indicates the location of *Gys2* TSS. The red box indicates the genomic location of validated liver-specific enhancers. The blue box indicates a region located from 280 to 320 kb downstream of *Gys2* promoter. (B) Comparison of regulatory networks of kidney and liver in the region from 25 kb downstream to 10 kb upstream of *Gys2* promoter. The red box indicates the same genomic location as in (A). (C) The 4C-seq signal in a 450 kb genomic region surrounding *Gys2* in kidney and liver. The 4C-seq data generated in [58] depict the interaction frequencies of *Gys2* promoter. The red box and blue box are the same as in (A). (D) Spatial interaction networks around *Gys2* TSS in kidney and liver estimated by FourC-seq using 4C-seq data. The blue box indicates the same region as in (A). (E) Signals of three chromatin modification marks across the region located from 280 to 320 kb downstream of *Gys2* promoter. Grey boxes indicate the location of accessible peaks a–e.

compared directly. In a simple experiment, we found that although the two scaling parameters of JRIM λ_1, λ_2 both contribute to the sparsity of networks, the combined parameters ω_1, ω_2 separately control the sparsity level and similarity among

reconstructed networks (Supplementary Figure S14). Especially when fix ω_1 , enlarging ω_2 brings almost no change in sparsity, but the similarity among tissues is increasing. We regarded it as an important merit relative to Cicero, and JRIM could generate

biologically plausible results based on the data-driven parameter selection process (Materials and Methods).

Currently, most tissue-specific genes are identified on the basis of differential expression or tissue diseases-related coding-region mutations. In this paper, we identified tissue-specific differential activity gene sets based on the changes of promoter-associated interactions across tissues. We have shown that tissue-specific differential activity genes tend to be highly expressed in corresponding tissues (Figure 4A). Moreover, the mutations within some DAG regions have been reported to be associated with tissue-related complex diseases on mouse or human. Taking heart as an example, the mouse embryos harboring a cardiomyocyte-restricted mutation in the *Myocd* gene exhibit myocardial hypoplasia and defective atrial [60], recessive *Ttn* truncating mutations in human are known to be associated with core myopathies [61] and *Myh6* mutations affect myofibril formation and are associated with congenital heart defects [62]. The methodologies developed here for identifying tissue-specific genes may provide new biomarkers for the illustration of mutation and disease-related regulatory mechanisms.

In this work, we mainly focused on the direct transcriptional regulations, i.e. protein coding gene-related promoter-enhancer interactions. However, as shown in the case study of *Scn5a*, indirect interactions also play a critical role in regulations of gene expression. Moreover, long intergenic noncoding RNAs (lincRNAs) are known for important regulators of gene expression [63], and aberrant epigenetic patterns in lincRNAs might be associated with cancer development [64]. Therefore, characterizing the regulatory functions of indirect interactions and noncoding regions will be an intriguing problem. Deciphering modular structure of the reconstructed cis-regulatory networks can be used to distinguish biological function units and shed light on the understanding of noncoding genes as well as noncoding mutations. Furthermore, a very interesting question in single-cell biology is modeling the dynamic changes during cell development. JRIM could detect dynamic changes of cis-regulatory interaction networks during cell development by analyzing time-series single-cell ATAC-seq data [24, 65, 66] or coupling with 'pseudotime' algorithms such as Monocle 2 [67].

Key Points

- We introduce JRIM to jointly reconstruct cis-regulatory interaction maps of multiple cell populations based on patterns of co-accessibility in single-cell data.
- JRIM was applied to explore common and specific regulatory interactions from single-cell ATAC-seq dataset containing ~80 000 cells across 13 mouse tissues.
- Reconstructed common interactions indeed relate to basic biological functions, and individual cis-regulatory networks show strong tissue specificity and functional relevance.
- Tissue-specific regulatory interactions are mediated by coordination of histone modifications and tissue-related TFs.

Availability

The package of JRIM is available at <http://page.amss.ac.cn/shihua.zhang/software.html>. All datasets in this manuscript are public datasets.

Supplementary Data

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

Funding

This work has been supported by the National Key R&D Program of China [2019YFA0709501]; National Natural Science Foundation of China [11661141019, 61621003]; National Ten Thousand Talent Program for Young Top-notch Talents; CAS Frontier Science Research Key Project for Top Young Scientist [QYZDB-SSW-SYS008].

References

1. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* 2012;13:59–69.
2. Zhang S, Tian D, Tran NH, et al. Profiling the transcription factor regulatory networks of human cell types. *Nucleic Acids Res* 2014;42:12380–7.
3. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 2012;9:215–6.
4. Chen C, Zhang S, Zhang X-S. Discovery of cell-type specific regulatory elements in the human genome using differential chromatin modification analysis. *Nucleic Acids Res* 2013;41:9230–42.
5. Yang X, Shao X, Gao L, et al. Systematic DNA methylation analysis of multiple cell lines reveals common and specific patterns within and across tissues of origin. *Hum Mol Genet* 2015;24:4374–84.
6. Wang C, Zhang S. Large-scale determination and characterization of cell type-specific regulatory elements in the human genome. *J Mol Cell Biol* 2017;9:463–76.
7. Butler JE, Kadonaga JT. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev* 2002;16:2583–92.
8. Leung KK, Ng LJ, Ho KK, et al. Different cis-regulatory DNA elements mediate developmental stage- and tissue-specific expression of the human COL2A1 gene in transgenic mice. *J Cell Biol* 1998;141:1291–300.
9. Ye Y, Gao L, Zhang S. MSTD: an efficient method for detecting multi-scale topological domains from symmetric and asymmetric 3D genomic maps. *Nucleic Acids Res* 2019;47:e65–5.
10. Cairns J, Freire-Pritchett P, Wingett SW, et al. CHiCAGO: robust detection of DNA looping interactions in capture hi-C data. *Genome Biol* 2016;17:127.
11. Fullwood MJ, Liu MH, Pan YF, et al. An oestrogen-receptor- α -bound human chromatin interactome. *Nature* 2009;462:58–64.
12. 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.
13. Schoenfelder S, Furlan-Magaril M, Mifsud B, et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res* 2015;25:582–97.
14. Corces MR, Granja JM, Shams S, et al. The chromatin accessibility landscape of primary human cancers. *Science* 2018;362:eaav1898.

15. Cao Q, Anyansi C, Hu X, et al. Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. *Nat Genet* 2017;**49**:1428–36.
16. Roy S, Siahpirani AF, Chasman D, et al. A predictive modeling approach for cell line-specific long-range regulatory interactions. *Nucleic Acids Res* 2015;**43**:8694–712.
17. Whalen S, Truty RM, Pollard KS. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet* 2016;**48**:488–96.
18. Zhu Y, Chen Z, Zhang K, et al. Constructing 3D interaction maps from 1D epigenomes. *Nat Commun* 2016;**7**:10812.
19. Buenrostro JD, Wu B, Litzgenberger UM, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 2015;**523**:486–90.
20. Cusanovich DA, Daza R, Adey A, et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 2015;**348**:910–4.
21. Felsenfeld G, Boyes J, Chung J, et al. Chromatin structure and gene expression. *Proc Natl Acad Sci U S A* 1996;**93**:9384–8.
22. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* 2019;**20**:207–20.
23. Cusanovich DA, Hill AJ, Aghamirzaie D, et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* 2018;**174**:1309–24.
24. Cusanovich DA, Reddington JP, Garfield DA, et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* 2018;**555**:538–42.
25. Clark SJ, Lee HJ, Smallwood SA, et al. Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol* 2016;**17**:72.
26. 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**:858–71.
27. Tung P-Y, Blischak JD, Hsiao CJ, et al. Batch effects and the effective design of single-cell gene expression studies. *Sci Rep* 2017;**7**:39921.
28. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;**9**:R137.
29. Saunders A, Macosko EZ, Wysoker A, et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* 2018;**174**:1015–30.
30. Consortium, T.M. Single-cell transcriptomics of 20 mouse organs creates a tabula Muris. *Nature* 2018;**562**:367–72.
31. Han X, Wang R, Zhou Y, et al. Mapping the mouse cell atlas by microwell-seq. *Cell* 2018;**172**:1091–107.
32. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57–74.
33. van den Boogaard M, Wong LE, Tessadori F, et al. Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. *J Clin Invest* 2012;**122**:2519–30.
34. Li B, Qing T, Zhu J, et al. A comprehensive mouse transcriptomic BodyMap across 17 tissues by RNA-seq. *Sci Rep* 2017;**7**:4200.
35. Xiao S-J, Zhang C, Zou Q, et al. TiSGeD: a database for tissue-specific genes. *Bioinformatics* 2010;**26**:1273–5.
36. Breiman L. Bagging predictors. *Mach Learn* 1996;**24**:123–40.
37. Maaten L.v.d, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;**9**:2579–605.
38. Danaher P, Wang P, Witten DM. The joint graphical lasso for inverse covariance estimation across multiple classes. *J R Stat Soc Ser B-Stat Methodol* 2014;**76**:373–97.
39. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 2008;**9**:432–41.
40. Yuan M, Lin Y. Model selection and estimation in the Gaussian graphical model. *Biometrika* 2007;**94**:19–35.
41. Javierre BM, Burren OS, Wilder SP, et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* 2016;**167**:1369–84.
42. Haeussler M, Zweig AS, Tyner C, et al. The UCSC genome browser database: 2019 update. *Nucleic Acids Res* 2018;**47**:D853–8.
43. Stuart T, Butler A, Hoffman P, et al. Comprehensive integration of single-cell data. *Cell* 2019;**177**:1888–902.
44. Yu G, Wang L-G, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;**16**:284–7.
45. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;**38**:576–89.
46. Yoshida H, Lareau CA, Ramirez RN, et al. The cis-regulatory atlas of the mouse immune system. *Cell* 2019;**176**:897–912.
47. Lacorazza HD, Miyazaki Y, Di Cristofano A, et al. The ETS protein MEF plays a critical role in perforin gene expression and the development of natural killer and NK-T cells. *Immunity* 2002;**17**:437–49.
48. Wasyluk C, Schlumberger SE, Criqui-Filipe P, et al. Sp100 interacts with ETS-1 and stimulates its transcriptional activity. *Mol Cell Biol* 2002;**22**:2687–702.
49. Fu J, Gaetani S, Oveisi F, et al. Oleyethanolamide regulates feeding and body weight through activation of the nuclear receptor PPAR- α . *Nature* 2003;**425**:90–3.
50. Yue F, Cheng Y, Breschi A, et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 2014;**515**:355–64.
51. Seo S, Lim JW, Yellajoshiyula D, et al. Neurogenin and NeuroD direct transcriptional targets and their regulatory enhancers. *EMBO J* 2007;**26**:5093–108.
52. Li QY, Newbury-Ecob RA, Terrett JA, et al. Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet* 1997;**15**:21–9.
53. Hoogaars WM, Engel A, Brons JF, et al. Tbx3 controls the sinoatrial node gene program and imposes pacemaker function on the atria. *Genes Dev* 2007;**21**:1098–112.
54. Creighton MP, Cheng AW, Welstead GG, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* 2010;**107**:21931–6.
55. Arnolds DE, Liu F, Fahrenbach JP, et al. TBX5 drives Scn5a expression to regulate cardiac conduction system function. *J Clin Invest* 2012;**122**:2509–18.
56. van den Boogaard M, Smemo S, Burnicka-Turek O, et al. A common genetic variant within SCN10A modulates cardiac SCN5A expression. *J Clin Invest* 2014;**124**:1844–52.
57. Doi R, Oishi K, Ishida N. CLOCK regulates circadian rhythms of hepatic glycogen synthesis through transcriptional activation of Gys2. *J Biol Chem* 2010;**285**:22114–21.
58. Mermet J, Yeung J, Hurni C, et al. Clock-dependent chromatin topology modulates circadian transcription and behavior. *Genes Dev* 2018;**32**:347–58.
59. Klein FA, Pakozdi T, Anders S, et al. FourCSeq: analysis of 4C sequencing data. *Bioinformatics* 2015;**31**:3085–91.
60. Huang J, Elicker J, Bowens N, et al. Myocardin regulates BMP10 expression and is required for heart development. *J Clin Invest* 2012;**122**:3678–91.

61. Chauveau C, Bonnemann CG, Julien C, et al. Recessive TTN truncating mutations define novel forms of core myopathy with heart disease. *Hum Mol Genet* 2013;**23**: 980–91.
62. Granados-Riveron JT, Ghosh TK, Pope M, et al. α -Cardiac myosin heavy chain (MYH6) mutations affecting myofibril formation are associated with congenital heart defects. *Hum Mol Genet* 2010;**19**:4007–16.
63. Yoon J-H, Abdelmohsen K, Srikantan S, et al. LincRNA-p21 suppresses target mRNA translation. *Mol Cell* 2012;**47**: 648–55.
64. Zhi H, Ning S, Li X, et al. A novel reannotation strategy for dissecting DNA methylation patterns of human long intergenic non-coding RNAs in cancers. *Nucleic Acids Res* 2014;**42**:8258–70.
65. Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* 2016;**48**:1193–203.
66. Preissl S, Fang R, Huang H, et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat Neurosci* 2018;**21**:432–9.
67. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;**32**:381–6.