IReNA: integrated regulatory network analysis of single-cell transcriptomes and chromatin accessibility profiles

Junyao Jiang[1], Pin Lyu[3], Jinlian Li[1], Sunan Huang[1], Seth Blackshaw[4], Jiang Qian[3], Jie Wang[1,2*]

[1] CAS Key Laboratory of Regenerative Biology, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, 510530, China

[2] China-New Zealand Joint Laboratory on Biomedicine and Health, Guangzhou, 510530, China

[3] Department of Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA

[4] Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA

[*]Corresponding author. Email: wang_jie01@gibh.ac.cn

Abstract

Although single-cell RNA sequencing (scRNA-seq) and single-cell assay for transposase accessible chromatin using sequencing (scATAC-seq) have been widely used, few methods can reliably integrate these data to perform regulatory network analysis. Here, we developed IReNA (Integrated Regulatory Network Analysis) for network inference through integrated analysis of scRNA-seq and scATAC-seq data, network modularization, transcription factor enrichment, and construction of simplified intermodular regulatory networks. Using public datasets, we showed that integrated network analysis of scRNA-seq and scATAC-seq data using IReNA outperformed currently available methods in identifying known regulators. IReNA facilitates the systems-level understanding of biological regulatory mechanisms, and is available at https://github.com/jiang-junyao/IReNA.

Keywords: gene regulatory networks, single-cell RNA sequencing, single-cell chromatin accessibility

Background

Dynamic changes of *trans*-regulators (e.g., transcription factors) and *cis*-regulatory elements (e.g., promoters) control gene expression in biological systems [1]. This fact makes it possible to infer gene regulatory networks using transcriptomic and epigenomic profiles. Recent advances in single-cell sequencing technologies provide new opportunities to reconstruct cell type-specific regulatory networks [2]. Single-cell transcriptomes have been widely detected through single-cell RNA sequencing (scRNA-seq). Currently, dozens of methods have used scRNA-seq data to infer regulatory networks, including top performing methods GENIE3 and PIDC [3–5]. Complementary to scRNA-seq, epigenomic profiling technique assay for transposase accessible chromatin using sequencing (ATAC-seq) measures accessible states of *cis*-regulatory elements to *trans*-regulators facilitating regulatory network inference [6]. Using scRNA-seq and bulk ATAC-seq data, we have developed a method to infer regulatory networks controlling retinal regeneration [7]. However, with the advent of the latest technology single-cell ATAC-seq (scATAC-seq), a new method for network inference is needed to integrate scRNA-seq and scATAC-seq data.

Besides network inference, dissecting regulatory networks to detect modules and to identify key regulators is another major challenge in systems biology. Clustering and decomposition are two frequently used methods for module detection [8]. Weighted correlation network analysis (WGCNA) performs hierarchical clustering of expression profiles to detect gene modules after correlation-based network inference [9]. Using single-cell transcriptomes, the SCENIC software combined GENIE3 and Rcistarget separately for network inference and identification of key transcription factors [10]. Although current methods can infer regulatory networks to identify gene modules and key individual regulatory genes, they do not construct a simple and statistically robust regulatory network among modules to provide new biological insights.

Here, we developed IReNA to perform regulatory network analysis through integrating scRNA-seq and scATAC-seq data. Since only scRNA-seq data are available for many biological samples, IReNA can also provide network analysis for using only scRNA-seq data. Using IReNA, we analyzed published single-cell RNA-seq and ATAC-seq profiles, reconstructed regulatory networks, identified key regulators, and revealed simplified regulatory networks among modules. In comparison with Rcistarget, one of the most

frequently used existing methods, IReNA shows better overall performance at identifying known regulators.

Result

The framework of IReNA

To perform regulatory network analysis using single-cell RNA-seq data or integrating with ATAC-seq data, we developed IReNA which consists of two components: network inference and network decoding (Figure 1 and Figure S1). Two pipelines of network inference were developed separately for analyzing scRNA-seq data alone or integrating scRNA-seq data with bulk or single-cell ATAC-seq data. Using scRNA-seq data, potential regulatory relationships were firstly inferred by the tree-based ensemble method GENIE3. Then, transcription factor binding motifs were used to refine regulatory relationships. If bulk or single-cell ATAC-seq data is available, both transcription factor binding motifs and footprints were used to refine regulatory relationships. For scATAC-seq data, peaks were linked to genes and used to identify transcription factor binding motifs and footprints. Network decoding in IReNA includes network modularization, identification of enriched transcription factors and a unique function for the construction of simplified regulatory networks among modules.

To illustrate the features and application of IReNA, we used public scRNA-seq and scATAC-seq data of hepatocytes from a mouse model of hepatectomy in the study of liver regeneration [11]. Although scRNA-seq and scATAC-seq data were obtained, the original study didn't perform regulatory network analysis.

Network analysis only using scRNA-seq data

Here, IReNA was applied to infer regulatory networks using scRNA-seq data alone. We firstly analyzed scRNA-seq data to identify genes used for network inference. Single-cell expression profiles of 2815 hepatocytes from the control and 48h liver tissues after hepatectomy were used to construct the trajectory of hepatocytes (Figure 2A). In the trajectory, we observed three branches, including the rest, activation and proliferation. Based on the trajectory, we calculated the pseudotime and identified 4014 differentially expressed

genes (DEGs) including 165 transcription factors along the pseudotime. Meanwhile, we identified 45 transcription factors which are expressed in more than 5% hepatocytes but not statistically differential during pseudotime. Next, all 4059 DEGs and expressed transcription factors were modularized into five modules through K-means clustering of the smoothed expression profiles (Figure 2B). Genes in each module showed specific expression profiles. For instance, genes in the first and fifth modules were specifically expressed in the rest and proliferating hepatocytes, respectively. Function enrichment analysis revealed relevant biological functions enriched in each module of genes, including fatty acid metabolism, organic acid catabolism, cytoplasmic translation, autophagy and cell cycle (Figure 2C).

We then applied GENIE3 to infer regulatory relationships of all DEGs and expressed transcription factors in hepatocytes. We identified 434,218 potential regulatory relationships, each of which has > 0.0001 weight and contains at least one transcription factor. For each regulatory relationship, Pearson's correlation was calculated to determine the positive or negative regulation. To refine 434,218 potential regulatory relationships, we further analyzed transcription factor binding motifs in the promoter regions of genes. Totally, 180,322 regulatory relationships with binding motifs were used to reconstruct regulatory networks (Figure 2D). Meanwhile, regulatory networks were modularized according to five modules of genes identified through K-means clustering in Figure 2B. Statistically analyzing modular regulatory networks, we identified 115 transcription factors which significantly regulated each module of genes (Table S1). We reconstructed regulatory networks of these 115 enriched transcription factors, which were also divided into five modules (Figure 2E).

To obtain a simple regulatory network among modules for providing a clearer picture of biological regulatory relationships, we statistically analyzed modular regulatory networks of enriched transcription factors. Fifteen significant regulatory relationships among modules were identified and used to establish simplified regulatory networks among modules (Figure 2F). In simplified intermodular regulatory networks, we observed that transcription factors related to fatty acid metabolism, organic acid catabolism and autophagy significantly activated each other. In return, these factors significantly repressed transcription factors controlling cell cycle regulation. These indicate that inhibition of transcription factors related to fatty acid metabolism, organic acid catabolism and autophagy would activate cell cycle progression. Transcription factors relevant to ribosome repressed transcription factors of fatty

acid metabolism, suggesting transcription factors regulating ribosome have synergistic effects on regulators of cell cycle.

Network analysis through integrating scRNA-seq and scATAC-seq data

Next, IReNA was used to infer regulatory networks through integrating scRNA-seq and scATAC-seq data from the study of liver regeneration. We analyzed scATAC-seq data of 7004 hepatocytes after hepatectomy using ArchR [12]. We identified 94,595 significant peak-to-gene links, each of which had high peak-to-gene correlation, e.g., *Rora* and *Mlx* (Figure 3A). We further uncovered transcription factors binding to peaks, and identified 386,597 regulatory relationships of transcription factors to genes. Among 386,597 regulatory relationships, 154,601 regulatory relationships had high footprint occupancy scores. Overlapping 154,601 regulatory relationships with 434,218 potential regulatory relationships inferred by GENIE3, we refined 47,721 regulatory relationships to reconstruct regulatory networks which consisted of 3185 genes.

Analyzing modular regulatory networks of 3185 genes, we identified 47 transcription factors which significantly regulate gene expression in each module. We then reconstructed modular regulatory networks of 47 enriched transcription factors (Figure 3B). Six significant regulation relationships among modules were identified to reconstruct simplified regulatory networks among modules (FDR < 0.05, Figure 3C). We observed that intermodular regulatory networks from the integrated analysis of scRNA-seq and scATAC-seq data are consistent with intermodular regulatory networks obtained from analyzing scRNA-seq data alone. In liver regeneration, the module of transcription factors related to organic acid catabolism of hepatocytes, e.g., *Hnf4a*, significantly repressed cell cycle. This suggests that the inhibition of hepatocyte metabolism may promote liver regeneration.

Performance of IReNA on identifying reported transcription factors

To directly compare IReNA with other methods on network analysis, we analyzed scRNA-seq data to reconstruct regulatory networks using the Rcistarget package from the SCENIC software [10]. Using Rcistarget, we identified 108 transcription factors whose binding motifs were overrepresented in the promoter regions of 4059 DEGs and expressed transcription factors identified by scRNA-seq data analysis. We obtained 1716 significant regulatory relationships for 108 enriched transcription factors, and then reconstructed modular regulatory networks of enriched transcription factors (Figure 3D).

We then compared regulatory networks of enriched transcription factors inferred through three different approaches described above. Among 47 transcription factors identified by IReNA using integrated analysis of scRNA-seq and scATAC-seq data, 36 (76.60%) transcription factors were also present in regulatory networks inferred using IReNA analysis of scRNA-seq data alone (Figure 3E). By comparing gene regulatory networks obtained by analyzing only scRNA-seq data separately using IReNA and Rcistarget, we observed an overlap of 66.96% (77 in 115) of all enriched transcription factors. These indicated that a large fraction of transcription factors was identified by two methods.

To assess the significance of transcription factors identified using IReNA or Rcistarget, we manually examined whether these factors had been previously reported in the literature related to tissue regeneration. To compare different methods, we ranked the enriched transcription factors according to the significance in statistics (FDR for IReNA or normalized enrichment score for Rcistarget). We found that regulatory networks from the integrated analysis of scRNA-seq and ATAC-seq data have the best performance, 55.00% in top 20, 53.33% in top 30 and 55.32% in all enriched transcription factors were separately reported in regeneration-related literature (Figure 3F and Table S1). In regulatory networks inferred from analyzing scRNA-seq data alone, 50.00% in top 20, 50.00% in top 30 and 53.91% in all transcription factors were reported in regeneration-related literature. In regulatory networks inferred by Rcistarget, there were 35.00% in top 20, 46.67% in top 30 and 51.85% in all transcription factors reported in regeneration-related literature. Among three types of regulatory networks, the highest fraction of transcription factors was reported for regulatory networks from the integrated analysis of scRNA-seq and scATAC-seq data. These results indicate that integrated network analysis of scRNA-seq and scATAC-seq data using IReNA improved the precision of identifying known transcription factors. Moreover, IReNA shows better performance at identifying known regulators than the Rcistarget method.

To further demonstrate the performance of IReNA, we performed regulatory network analysis on another two datasets from heart regeneration and NASH. Prior to inferring gene regulatory networks controlling heart regeneration, we reconstructed the trajectory of 4884 cardiomyocytes from neonatal heart tissues (Figure S2A). We found that cardiomyocytes formed two distinct branches (named the activation branch and the proliferative branch) following myocardial infarction. We further identified and divided 4340 DEGs and expressed transcription factors into 4 modules (Figure S2B). Genes in the activation branch (module 2

and 3) were related to oxidative phosphorylation and muscle cell differentiation, whereas genes in the proliferative branch (module 4) are enriched for cell cycle (Figure S2C). Then, regulatory networks were inferred using the same three methods described above (Figure 4A-4B and Figure S2D-S2E). In comparison with regulatory networks inferred by Rcistarget, regulatory networks reconstructed by IReNA contained a higher fraction of transcription factors previously reported in the literature related to regeneration (Figure 4C and Table S1). The precision of identifying known regulators among top 20 enriched transcription factors were 80.00%, 70.00% and 45.00% separately for scRNA-seq + scATAC-seq + IReNA, scRNA-seq + IReNA and scRNA-seq + Rcistarget. We also observed that scRNA-seq + scATAC-seq + IReNA identified the highest fraction of known regulators for both top 30 and all enriched transcription factors.

For the study of NASH, we used 2748 Kupffer cells to construct the trajectory (Figure S3A). 2742 DEGs and expressed transcription factors were identified and divided into three modules, which were separately enriched for myeloid cell differentiation, ribosome and oxidative phosphorylation (Figure S3B-S3C). In NASH, only bulk ATAC-seq data was available and used to refine regulatory relationships inferred from scRNA-seq data analysis (Figure S3D). We reconstructed three types of regulatory networks and a simplified regulatory network among modules (Figure 4D-4E and Figure S3E-S3F). Regulatory network comparison of NASH study has a similar trend with studies in liver regeneration and heart regeneration. IReNA analysis using scRNA-seq and bulk ATAC-seq data identified the most transcription factors (65.00% in top 20, 53.33% in top 30 and 59.26% in all enriched transcription factors) which are reported to associate with NASH, followed by IReNA analysis using scRNA-seq data alone (60.00% in top 20, 53.33% in top 30 and 49.33% in all), then is Rcistarget analysis using scRNA-seq data alone (30.00% in top 20, 33.33% in top 30 and 36.36% in all) (Figure 4F and Table S1). The integrated analysis of single-cell or bulk ATAC-seq data with scRNA-seq data overall substantially improved the reconstruction of gene regulatory networks, and had higher precision of identifying known transcription factors. In addition, regulatory networks reconstructed from scRNA-seq data alone using IReNA showed improved accuracy relative to those identified using Rcistarget.

Discussion

In the study, we developed IReNA to perform regulatory network analysis including network inference and network decoding. In IReNA, gene regulatory networks are inferred through analyzing either scRNA-seq data alone or integrating scRNA-seq and scATAC-seq data. The regulatory relationships between transcription factors and target genes are firstly inferred according to the weights calculated by GENIE3 using scRNA-seq data. Then, transcription factor binding motifs are identified to refine regulatory relationships if only scRNA-seq data are available. When both scRNA-seq and ATAC-seq data are used, transcription factor binding motifs and footprints are applied to further refine transcriptional regulatory relationships. IReNA also provides functions to decode inferred regulatory networks, including the modularization of regulatory networks, the enrichment of transcription factors, and the construction of simplified regulatory networks among modules.

Several of these functions of IReNA are new, and not performed by other analytic tools. First, unlike current methods using scRNA-seq data for network analysis, IReNA integrates scRNA-seq data with scATAC-seq data to reconstruct regulatory networks. The results indicate that the integrated analysis of scRNA-seq and scATAC-seq data could more precisely identify known regulators than network analysis using scRNA-seq data alone. Second, IReNA could provide cell state-specific regulatory networks through modularizing regulatory networks. Regulatory networks are modularized according to the clustering results of gene expression profiles which represent different cell states and specific biological functions. Third, IReNA statistically analyzes modular regulatory networks and identifies reliable transcription factors including known regulators. Applied to multiple datasets, the method used in IReNA showed a consistently better performance on identification of known regulators than Rcistarget, which conducts transcription factor enrichment analysis based on the rank of all genes for each motif [13]. Fourth, we created a unique function in IReNA to construct simplified regulatory networks among modules which reveal key regulatory modules and factors, facilitating the interpretation of dynamic biological regulations.

IReNA can be used to infer regulatory networks for the same cell type, or for several cell types which are related by lineage during the development or disease progression. Correspondingly, cell type-specific or lineage-specific transcriptomes and epigenomes must be extracted prior to network inference. Transcriptomes could be measured by bulk RNA-seq or scRNA-seq, whereas epigenomes may be detected using either bulk ATAC-seq or

scATAC-seq. Transcriptomes and epigenomes used in IReNA should be matched for the same cell type or the same condition. Currently, IReNA has used scRNA-seq and scATAC-seq data of unpaired cells from the same condition to perform network analysis. However, parallel scRNA-seq and scATAC-seq profiles from the same cells are emerging, and updates to IReNA will accommodate these new data.

In previous studies, we demonstrated that IReNA can be used to integrate scRNA-seq and ATAC-seq to reconstruct gene regulatory networks controlling retinal regeneration and retinal development [7,14]. Through network analysis using IReNA, we identified modular gene regulatory networks and key transcription factors regulating different cell states in retinal Müller glia in zebrafish and mice. Using genetic loss of function analysis, we confirmed that several transcription factors identified by IReNA are critical for retinal regeneration, including *hmga1a* and *yap1* in zebrafish and *Nfia/b/x* in mice [7]. These results indicate that IReNA can provide reliable regulatory networks and reveal key regulators during biological processes including retinal regeneration.

Using public scRNA-seq and ATAC-seq data from three studies, we further performed regulatory network analysis through IReNA, which provides meaningful biological insights. According to simplified intermodular regulatory networks in liver regeneration, organic acid catabolism-related modules and transcription factors are needed to be repressed to activate cell cycle progression in hepatocytes, e.g., hepatocyte nuclear factor 4 alpha (*Hnf4a*) (Figure 2F, 3C). The study in mouse liver regeneration demonstrated that deletion of *Hnf4a* leads to sustained proliferation [15]. In heart regeneration, simplified modular regulatory networks imply that regulatory modules and transcription factors controlling cardiac muscle contraction promote cell cycle of cardiomyocytes (Figure 4B). It has been reported that one of such factors *Gata4* could activate heart regeneration in zebrafish and mice [16,17]. In regulatory networks of NASH, we identified *Fos* and *Klf6*, which regulate TH17/Treg cells in NASH patients and are up-regulated in the context of steatohepatitis, respectively [18,19]. The results indicate that IReNA could be applied to identify key modules and transcription factors that regulate a range of different biological processes, although further functional validation of putative key regulators is required.

Conclusion

In summary, we have developed IReNA to perform regulatory network analysis, including network inference, network modularization, transcription factor enrichment and construction of simplified regulatory networks among modules. IReNA showed a consistently better performance at identifying known regulators when integrating scRNA-seq data with scATAC-seq data to reconstruct gene regulatory networks. IReNA also outperformed the existing method Rcistarget on identifying known regulators. Key transcription factors and regulatory relationships identified by IReNA are potential targets controlling tissue regeneration, diseases, and other dynamic biological processes. Through the construction of modular regulatory networks and simplified regulatory networks among modules, IReNA facilitates the understanding of regulatory mechanisms and provides meaningful biological insights.

Materials and methods

Description of scRNA-seq, bulk ATAC-seq and scATAC-seq data

To demonstrate analysis flow of IReNA, we used public scRNA-seq and ATAC-seq data from three studies, which analyzed datasets obtained from models of liver regeneration, heart regeneration, and nonalcoholic steatohepatitis (NASH), respectively [11,20,21].

For liver regeneration, partial hepatectomy (PHx) was performed in adult mice [20]. The study conducted both scRNA-seq and scATAC-seq on liver tissues at 0 and 48 hours after PHx (accession number GSE158866 and GSE158873, available at Gene Expression Omnibus database https://www.ncbi.nlm.nih.gov/geo/). According to the original annotation of cell types in the study, there are 2815 and 7004 hepatocytes measured separately for scRNA-seq and scATAC-seq.

In the study of heart regeneration, scRNA-seq profiles were measured on 4884 cardiomyocytes at 1 day, 3 days after myocardial infarction, and 1 day after sham surgery in neonatal mice (accession number GSE130699) [22]. Meanwhile, scATAC-seq was performed on 755 cardiomyocytes at 3 days after myocardial infarction on postnatal day one (accession number GSE142365).

The scRNA-seq and bulk ATAC-seq data from the study of NASH are available through the accession numbers GSE128334 and GSE128335 [11]. In this study, scRNA-seq profiles were measured on 6184 non-parenchymal cells, including 2748 Kupffer cells, from liver tissues of healthy and NASH mice. Bulk ATAC-seq was conducted on Kupffer cells from two healthy and two NASH samples.

Regulatory network inference

Prior to network inference, we identified differentially expressed genes (DEGs) and the expressed transcription factors as the potential genes in network analysis. Monocle (version 2.1.8) was used to analyze scRNA-seq data and to construct the trajectory [23]. Pseudotime of individual cells was further inferred and used to identify DEGs (q-value < 0.005, fraction of expressed cells > 10% and single-cell expression difference > 0.1). Single-cell expression difference was defined as previously described [7]. Given that some key transcription factors may not be DEGs, we included all expressed transcription factors which expressed in > 5% cells for network analysis.

According to the previous evaluation of the methods for network inference, we selected top performing method GENIE3 as the default method for network inference in IReNA [3,4]. GENIE3 infers regulatory relationships of transcription factors to target genes based on random forest regression [4]. GENIE3 (version1.16) was used to analyze scRNA-seq data and to calculate the weight of regulation for each gene pair. We selected gene pairs which have > 0.0001 weight. Meanwhile, only gene pairs which contain at least one transcription factor from the TRANSFAC database (version 2018.3) were chosen as potential regulatory relationships.

To determine activating and repressive regulatory relationships, we calculated Pearson's correlation of gene pairs using smoothed expression profiles. When Pearson's correlation > 0, the regulation type of gene pair is defined as activation. When Pearson's correlation < 0, the regulation type is defined as repression.

Identify transcription factor binding motifs in regulatory regions of targeted genes

If bulk or single-cell ATAC-seq data is not available, gene regulatory relationships inferred from scRNA-seq data analysis were further refined according to transcription factor binding motifs present in the promoter regions of genes. Fimo was used to identify

transcription factor binding motifs in the promoter regions (ranging from 1000bp upstream to 500bp downstream of the transcription start sites) of the genes [24]. Position weight matrices of binding motifs were from TRANSFAC database (version 2018.3). Regulatory relationships were selected for further network analysis if the binding motif of transcription factor occurs in the promoter region of the target gene.

Analyze bulk or single-cell ATAC-seq data to refine regulatory relationships

If bulk ATAC-seq data is available, we use the following six steps to preprocess raw data in fastq format. (I) Remove adaptors of pair-end raw reads using fastp software (version 0.21.0) [25]. (II) Align reads the GRCm38/mm10 genome using bowtie2 (version 2.4.1) with default parameters [26]. (III) Filter low-mapping-quality reads (MAPQ < 10) and exclude duplicated reads separately using Samtools (version 1.3.1) and Picard (http://broadinstitute.github.io/picard/) [27]. (IV) Call peaks through MACS2 (version 2.1.0) with the parameter extsize = 200 and shift = 100 [28]. (V) Use HTseq (version 0.12.4) to calculate the count number of each peak [29]. (VI) Combine the peaks across all samples to obtain the union peaks and identify differentially accessible peaks using EdgeR [30].

Different from bulk ATAC-seq data, scATAC-seq data was preprocessed through following steps. (I) Map raw sequencing data in fastq format to the reference genome (GRCm38/mm10) with cellranger (version 2.0.0) [31]. (II) Use ArchR (version 1.0.1) to integrate scATAC-seq and scRNA-seq data with unconstrained integration methods [12]. Identify peak-to-gene links by calculating the correlation between peak accessibility and gene expression across individual cells, and retain peak-to-gene links with absolute value of correlation > 0.2, FDR < 1E-6, varCutOffATAC (variance of peak accessibility) > 0.7 and varCutOffRNA (variance of gene expression) > 0.3.

After processing bulk or single-cell ATAC-seq data, the following steps were used to identify footprints and to refine regulatory relationships. (I) Identify the footprints of peaks through HINT (version 0.13.2) and select high-quality footprints (tag-count score > 80th percentile) for downstream analysis [32]. (II) Select footprints which are covered by differentially accessible peaks. (III) Run Fimo to find binding motifs in the footprints according to the position weight matrices of motifs from TRANSFAC database [24]. (IV) Identify footprint-related genes. For bulk ATAC-seq data, ChIPseeker (version 1.26.2) is used to annotate footprint regions. Genes related to footprint regions are considered as

footprint-related genes [33]. For scATAC-seq data, genes linked by ArchR to peaks are considered as footprint-related genes. (V) Use Rsamtools (version 2.6.0) to obtain the sequencing depth of the mapped reads which is used to calculate the number of insertions at each position of footprints (https://bioconductor.org/packages/Rsamtools). (VI) Use the number of insertions to calculate footprint occupancy score (FOS), and then select regulatory relationships which have high FOS (FOS > 0.1) to reconstruct regulatory networks. FOS was calculated using the formula defined as previously described [7].

$$FOS = min\left(-log_2\frac{N_C+1}{N_L+1}, -log_2\frac{N_C+1}{N_R+1}\right) \qquad (1)$$

where N_L, N_C and N_R are numbers of insertions separately in the left, center and right regions of the motif.

Network modularization based on gene co-expression

Given the sparsity of scRNA-seq data, we used the smoothed expression profiles to perform gene co-expression analysis. The smoothed expression profiles were calculated according to pseudotime and branches on the trajectory. If there is a branch in the trajectory, pseudotime in each branch was divided into 20 equal intervals. Otherwise, pseudotime was divided into 50 equal intervals. Then, we calculated the average expression profile of single cells in each interval and obtained the smoothed expression profiles. DEGs and the expressed transcription factors were divided into different modules using the K-means clustering of the smoothed expression profiles. The optimal number of modules was determined by the silhouette coefficient calculated by R package 'cluster'. Based on the modules of genes, the inferred regulatory networks were modularized. For each module, ClusterProfile (version 3.18.1) was used to perform functional enrichment analysis [34].

Transcription factor enrichment for modular regulatory networks

Refined regulatory relationships were used to reconstruct modular regulatory networks of DEGs and expressed transcription factors. Cytoscape was used to display regulatory networks [35]. We performed the hypergeometric test to calculate the probability P that an individual transcription factor regulates a module, and then adjusted P values to false discovery rate (FDR) values. Enriched transcription factors which significantly regulate each module of genes were used to reconstruct regulatory networks. For the hypergeometric test, the probability was calculated as follows.

$$P(x = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \qquad (2)$$

Here, $\binom{a}{b}$ is a binomial coefficient. For identifying the significant transcription factor regulating module A, *N* and *n* represent numbers of all regulations and regulations targeting module A, respectively. *K* and *k* separately indicate the number of regulations from transcription factor and the number of regulations targeting module A from transcription factor.

Construct simplified regulatory networks among modules

Based on modular regulatory networks of enriched transcription factors, we carried out another hypergeometric test to determine significant regulatory relationships among modules and reconstructed simplified regulatory networks among modules.

The same formula as (2) was used to calculate the probability of the regulation of module A to module B. In the formula, *N* and *n* represent numbers of all regulations and regulations from module A, respectively. *K* and *k* separately indicate the number of regulations from module B and the number of regulations from module A to module B.

The P values of regulations among modules were adjusted to FDR values. Regulatory relationships with FDR < 0.05 were regarded as significant regulations, and used to construct simplified regulatory networks among modules.

Comparison of regulatory networks

To compare regulatory networks inferred from the integrated analysis of both scRNA-seq and ATAC-seq data, and from scRNA-seq data alone, we examined whether top and all enriched transcription factors in network analysis had been previously reported in the literature. Enriched transcription factors were ranked by FDR. For the study of nonalcoholic steatohepatitis, we used the biological terms 'nonalcoholic steatohepatitis, steatosis or fatty liver disease' to search the Google Scholar and PubMed databases. Given a limited number of studies reporting organ-specific regeneration, the common biological term 'regeneration' was used for each study of liver regeneration, and heart regeneration. We confirmed if the gene symbol and/or common gene name for individual transcription factors were present. If

the gene symbol/gene name and biological term were both present in the title or the same sentence in the abstract, the biological function of the enriched transcription factor was regarded to have been reported in the literature.

To compare IReNA with existing methods for transcription factor enrichment, we used Rcistarget to identify key regulators analyzing the same scRNA-seq data [10]. Rcistarget calculates the normalized enrichment score (NES) which measures the enrichment of transcription factor binding motifs in the promoter regions of DEGs. We reconstructed regulatory networks for transcription factors which have > 3 NES. Then, we checked whether top (ranked by NES) and all transcription factors identified through Rcistarget were reported in the literature.

Supplementary data

Supplementary data are available online.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [32170849], and the Guangdong Province Science and Technology Program [2020B1212060052].

Availability of data and materials

IReNA is an open software and available online at https://github.com/jiang-junyao/IReNA.

Author contributions

J.W., and J.Q. conceived the project. J.W., J.Q., and S.B. supervised the research. J.Y., P.L., and J.W. developed IReNA and performed sequencing data analysis to construct regulatory networks. J.Y., J.L., and S.H. checked transcription factors reported in the literature.

Competing interests

The authors declare that they have no competing interests.

References

1. Thompson D, Regev A, Roy S. Comparative analysis of gene regulatory networks: from network reconstruction to evolution. Annu Rev Cell Dev Biol. 2015;31:399–428.

2. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015;161:1202–14.

3. Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat Methods. 2020;17:147–54.

4. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. PLoS One [Internet]. 2010;5. Available from: http://dx.doi.org/10.1371/journal.pone.0012776

5. Chan TE, Stumpf MPH, Babtie AC. Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. Cell Syst. 2017;5:251–67.e3.

6. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10:1213–8.

7. Hoang T, Wang J, Boyd P, Wang F, Santiago C, Jiang L, et al. Gene regulatory networks controlling vertebrate retinal regeneration. Science [Internet]. 2020;370. Available from: http://dx.doi.org/10.1126/science.abb8598

8. Saelens W, Cannoodt R, Saeys Y. A comprehensive evaluation of module detection methods for gene expression data. Nat Commun. 2018;9:1090.

9. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

10. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. 2017;14:1083–6.

11. Seidman JS, Troutman TD, Sakai M, Gola A, Spann NJ, Bennett H, et al. Niche-Specific Reprogramming of Epigenetic Landscapes Drives Myeloid Cell Diversity in Nonalcoholic Steatohepatitis. Immunity. 2020;52:1057–74.e7.

12. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet. 2021;53:403–11.

13. Imrichová H, Hulselmans G, Atak ZK, Potier D, Aerts S. i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. Nucleic Acids Res. 2015;43:W57–64.

14. Lyu P, Hoang T, Santiago CP, Thomas ED, Timms AE, Appel H, et al. Gene regulatory networks controlling temporal patterning, neurogenesis, and cell-fate specification in mammalian retina. Cell Rep. 2021;37:109994.

15. Huck I, Gunewardena S, Espanol-Suner R, Willenbring H, Apte U. Hepatocyte Nuclear Factor 4 Alpha Activation Is Essential for Termination of Liver Regeneration in Mice. Hepatology. 2019;70:666–81.

16. Kikuchi K, Holdway JE, Werdich AA, Anderson RM, Fang Y, Egnaczyk GF, et al. Primary contribution to zebrafish heart regeneration by gata4(+) cardiomyocytes. Nature. 2010;464:601–5.

17. Malek Mohammadi M, Kattih B, Grund A. The transcription factor GATA 4 promotes myocardial regeneration in neonatal mice. molecular medicine [Internet]. embopress.org; 2017; Available from: https://www.embopress.org/doi/abs/10.15252/emmm.201606602

18. Cai C, Chen D-Z, Tu H-X, Chen W-K, Ge L-C, Fu T-T, et al. MicroRNA-29c acting on FOS plays a significant role in nonalcoholic steatohepatitis through the interleukin-17 signaling pathway. Front Physiol. Frontiers Media SA; 2021;12:597449.

19. Stärkel P, Sempoux C, Leclercq I, Herin M, Deby C, Desager J-P, et al. Oxidative stress, KLF6 and transforming growth factor-β up-regulation differentiate non-alcoholic steatohepatitis progressing to fibrosis from uncomplicated steatosis in rats. J Hepatol. 2003;39:538–46.

20. Chen T, Oh S, Gregory S, Shen X, Diehl AM. Single-cell omics analysis reveals functional diversification of hepatocytes during liver regeneration. JCI Insight [Internet]. 2020;5. Available from: http://dx.doi.org/10.1172/jci.insight.141024

21. Guo J, Grow EJ, Yi C, Mlcochova H, Maher GJ, Lindskog C, et al. Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development. Cell Stem Cell. 2017;21:533–46.e6.

22. Cui M, Wang Z, Chen K, Shah AM, Tan W, Duan L, et al. Dynamic Transcriptional Responses to Injury of Regenerative and Non-regenerative Cardiomyocytes Revealed by Single-Nucleus RNA Sequencing. Dev Cell. 2020;55:665–7.

23. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017;14:979–82.

24. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015;43:W39–49.

25. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90.

26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

28. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137.

29. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. Oxford University Press; 2015;31:166–9.

30. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.

31. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8:14049.

32. Li Z, Schulz MH, Look T, Begemann M, Zenke M, Costa IG. Identification of transcription factor binding sites using ATAC-seq. Genome Biol. 2019;20:45.

33. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015;31:2382–3.

34. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.

35. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

Figure and table legends

Figure 1. Flowchart of IReNA. IReNA consists of two components: network inference and network decoding. Network inference is performed according to weights calculated by GENIE3. FIMO is then used to identify binding motifs of transcription factors and refine regulatory relationships in cases where only gene expression profiles (single-cell or bulk RNA-seq data) are used. If bulk or single-cell ATAC-seq data is available, binding motifs and footprint occupancy score (FOS) are used to refine regulatory relationships inferred from expression analysis. Especially for single-cell ATAC-seq data, ArchR is used to link peaks and genes. Next, inferred networks are modularized and used to identify enriched transcription factors (TFs). Regulatory networks of enriched TFs are further decoded by establishing simplified regulatory networks among modules.

Figure 2. Regulatory network analysis of hepatocytes from liver regeneration. (A) Trajectory of 2815 hepatocytes from liver tissues at 0 and 48 hours after partial hepatectomy. Two trajectories are separately colored by pseudotime and samples. (B) Heatmap of scRNA-seq expression profiles of 4059 differentially expressed genes (DEGs) and expressed transcription factors. K-means clustering was used to divide genes into five modules. Rows represent genes, and columns indicate intervals of pseudotime separated by three branches. (C) Enriched functions of five modules of genes. (D) Fraction of regulatory gene pairs inferred by GENIE3 which contain transcription factor binding motifs in the promoter regions of target genes. (E) Regulatory networks of 41 enriched transcription factors obtained from analyzing scRNA-seq data alone. Color of each circle indicates the module. Grey edge represents negative regulation, and the yellow edge represents positive regulation. (F) Simplified regulatory networks among five modules obtained from analyzing scRNA-seq data alone. Representative biological functions and transcription factors were selected from Figure 2C and 2E to label each module.

Figure 3. Comparison of regulatory networks related to liver regeneration. (A) Heatmap of peak-to-gene correlations (top panel) calculated by ArchR and weights of gene pairs (bottom panel) calculated by GENIE3. In the top panel, the gene in the column represents a transcription factor which has the binding motif in the peak. Gene pairs with both high correlations and high weights are used to infer regulatory networks of integrating scRNA-seq and scATAC-seq data. (B) Regulatory networks of 47 enriched transcription factors obtained through the integrated analysis of scRNA-seq and scATAC-seq data. Color of the circle indicates the module. Grey edge represents negative regulation, and the yellow edge represents positive regulation. (C) Simplified regulatory networks among modules obtained through the integrated analysis of scRNA-seq and scATAC-seq data. (D) Regulatory networks for 108 enriched transcription factors from scRNA-seq data analysis using Rcistarget. (E) Venn diagram of enriched transcription factors from three types of regulatory networks, one from the integrated analysis of scRNA-seq and ATAC-seq data using IReNA (ATAC-seq + scRNA-seq + IReNA), one from scRNA-seq data analysis using IReNA (scRNA-seq + IReNA) and one from scRNA-seq data analysis using Rcistarget (scRNA-seq + Rcistarget). (F) Fraction of enriched transcription factors reported in the literature related to

tissue regeneration. Top 20, top 30 and all enriched transcription factors were compared for all three types of regulatory networks. Enriched transcription factors were ranked according to FDR value or normalized enrichment score.

Figure 4. Regulatory network analysis for two studies of heart regeneration and NASH. (A) Regulatory networks of 54 enriched transcription factors obtained by integrating scRNA-seq and scATAC-seq data from heart regeneration. Modules are represented by the color of the circles. Grey edge represents negative regulation, and the yellow edge represents positive regulation. (B) Simplified regulatory networks among modules obtained by integrating of scRNA-seq and scATAC-seq data from heart regeneration. (C) Fraction of enriched transcription factors reported in the literature related to tissue regeneration. (D) Regulatory networks of 75 enriched transcription factors obtained by integrating scRNA-seq and bulk ATAC-seq data from nonalcoholic steatohepatitis (NASH). (E) Simplified regulatory networks among modules obtained by integrating scRNA-seq and ATAC-seq data from NASH. (F) Fraction of transcription factors reported in the literature related to NASH.

Figure S1. Workflow of IReNA. Differentially expressed genes (DEGs) and expressed transcription factors (TFs) are identified from single-cell or bulk RNA-seq expression profiles. Regulatory relationships among genes are inferred according to weights calculated by GENIE3. If ATAC-seq data is not available, binding motifs are further identified to refine regulatory relationships. Otherwise, both binding motifs and footprint occupancy score (FOS) are computed to refine regulatory relationships. If scATAC-seq is used, peaks are linked to genes through ArchR. After network inference, networks are decoded through modularizing regulatory networks, identifying enriched TFs, and establishing simplified regulatory networks among modules.

Figure S2. Regulatory network analysis of heart regeneration. (A) Trajectory of 4884 cardiomyocytes from neonatal heart tissues at day 1, day 3 after myocardial infarction, and day 1 after sham surgery. (B) 4340 DEGs and expressed transcription factors were divided into 4 modules according to K-means clustering of scRNA-seq expression profiles. Rows represent genes, and columns indicate intervals of pseudotime divided by branches. (C) Enriched functions of four modules of genes. Colors represent modules. (D) Regulatory networks of 11enriched transcription obtained through integrated analysis of scRNA-seq and scATAC-seq data. Grey edge represents negative regulation, and the yellow edge represents positive regulation. (E) Regulatory networks of 90 enriched transcription factors obtained by Rcistarget analysis of scRNA-seq data.

Figure S3. Regulatory network analysis of Kupffer cells from healthy and nonalcoholic steatohepatitis (NASH) liver tissues. (A) Trajectory of 2748 Kupffer cells separately colored by pseudotime and samples. (B) Heatmap of scRNA-seq expression profiles of 2742 differentially expressed genes and expressed transcription factors. K-means clustering was used to divide genes into three modules. Each row represents one gene, and each column indicates one interval of pseudotime. (C) Enriched functions of three modules of DEGs. (D) Genome-wide chromatin accessibility of healthy and NASH Kupffer cells measured by bulk

ATAC-seq. (E) Regulatory networks of 27 enriched transcription factors obtained through integrated analysis of scRNA-seq and bulk ATAC-seq data. Color of each circle indicates the module. Grey edge represents negative regulation, and the yellow edge represents positive regulation. (F) Regulatory networks of 119 enriched transcription factors obtained by Rcistarget analysis of scRNA-seq data.

Table S1. The list of transcription factors enriched in liver regeneration, heart regeneration and nonalcoholic steatohepatitis. Transcription factors were enriched by IReNA analysis of scRNA-seq alone (scRNA-seq + IReNA), the integrated analysis of scRNA-seq and scATAC-seq data (scRNA-seq + scATAC-seq + IReNA) or Rcistarget analysis of scRNA-seq alone (scRNA-seq + Rcistarget). The related literature and sentences are also listed.

Figure 1

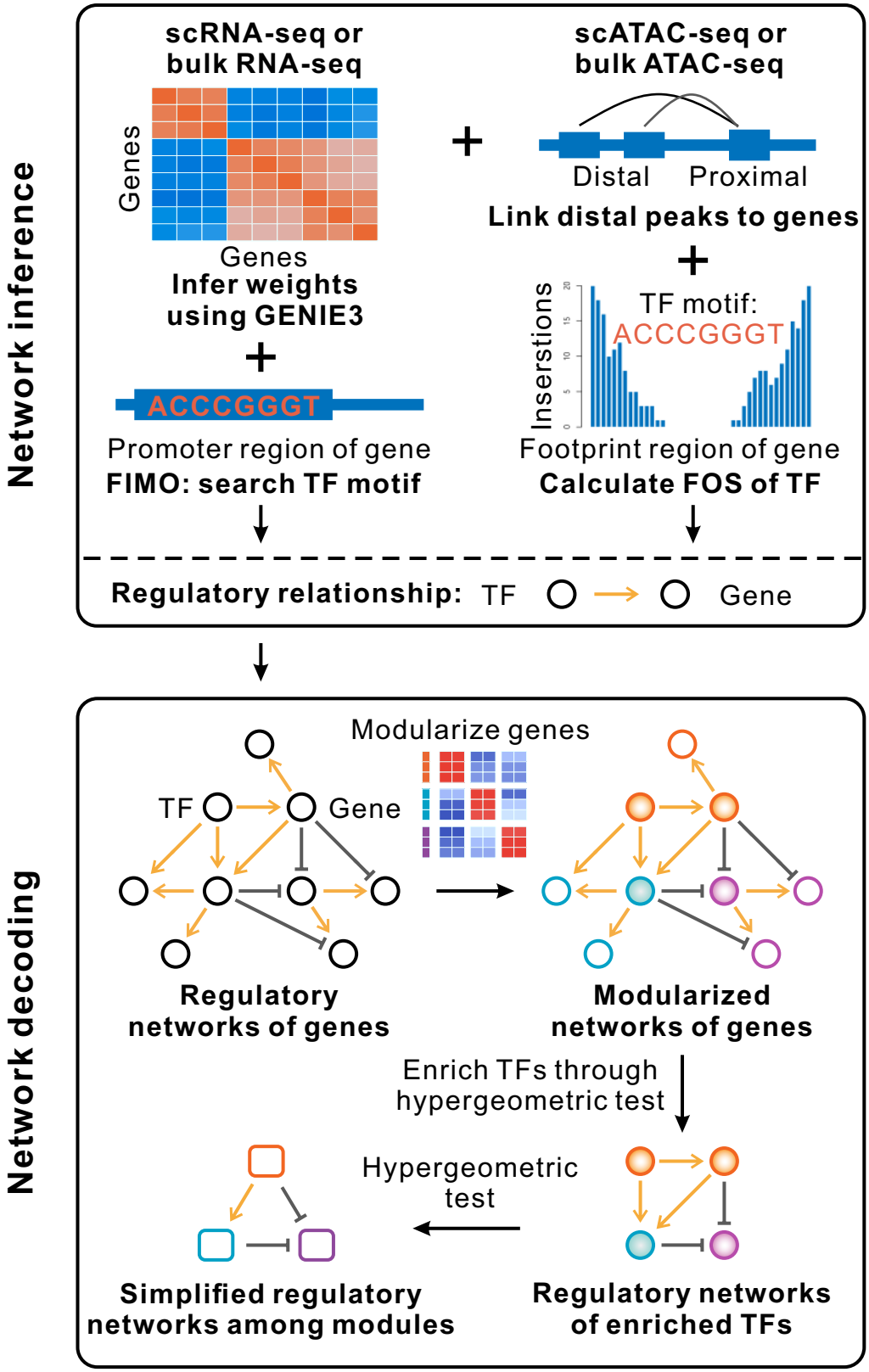

Network inference
scRNA-seq or bulk RNA-seq
Genes
Genes
Infer weights using GENIE3
ACCCGGGT
Promoter region of gene
FIMO: search TF motif
scATAC-seq or bulk ATAC-seq
Distal
Proximal
Link distal peaks to genes
Inserstions
TF motif:
ACCCGGGT
Footprint region of gene
Calculate FOS of TF
Regulatory relationship: TF Gene
Network decoding
Modularize genes
TF
Gene
Regulatory networks of genes
Modularized networks of genes
Enrich TFs through hypergeometric test
Hypergeometric test
Simplified regulatory networks among modules
Regulatory networks of enriched TFs

Figure 2

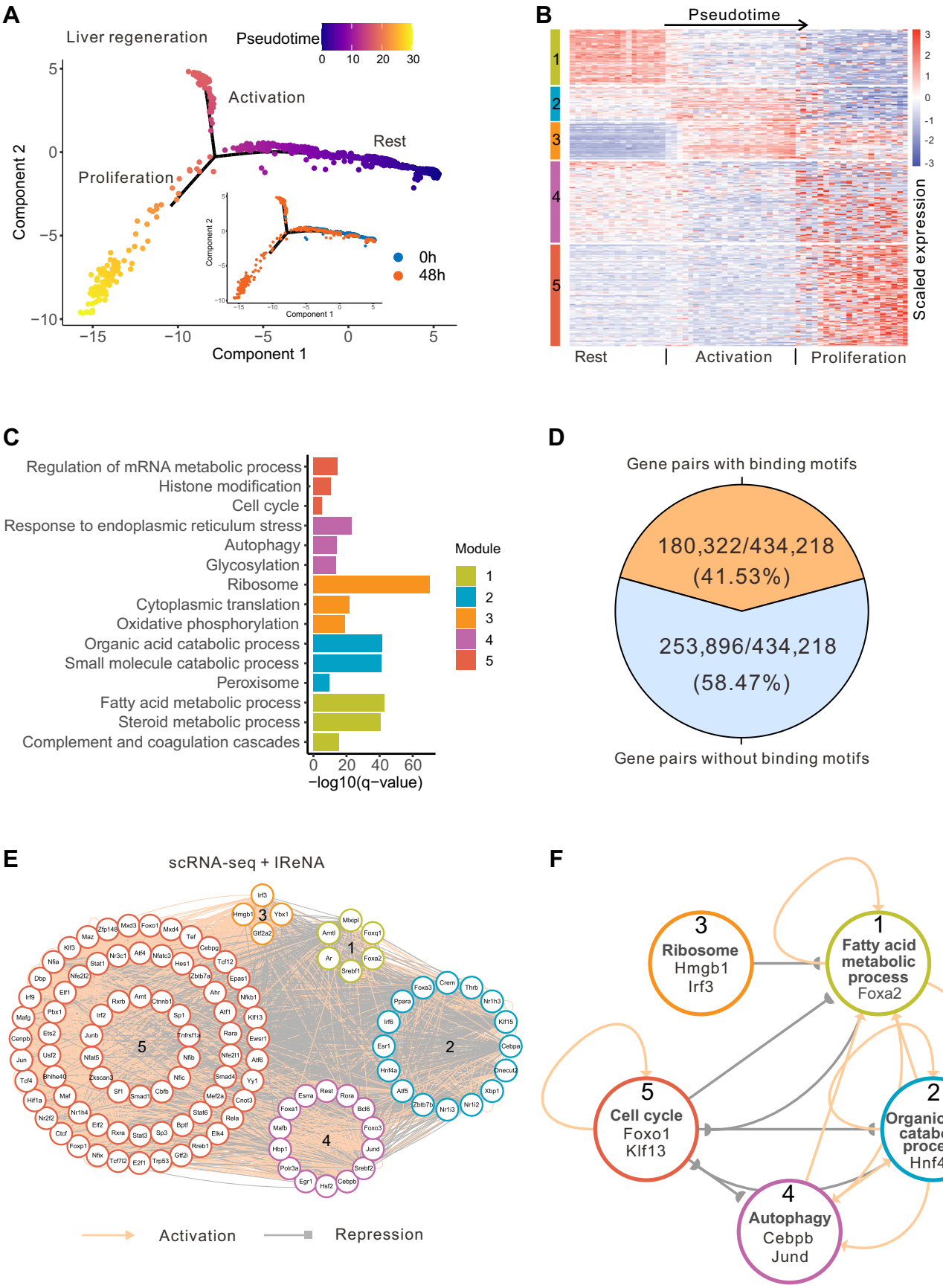

A
Liver regeneration
Pseudotime
0 10 20 30
Activation
Rest
Proliferation
Component 2
Component 1
0h
48h
B
Pseudotime
Rest
Activation
Proliferation
Scaled expression
C
Regulation of mRNA metabolic process
Histone modification
Cell cycle
Response to endoplasmic reticulum stress
Autophagy
Glycosylation
Ribosome
Cytoplasmic translation
Oxidative phosphorylation
Organic acid catabolic process
Small molecule catabolic process
Peroxisome
Fatty acid metabolic process
Steroid metabolic process
Complement and coagulation cascades
−log10(q−value)
Module
D
Gene pairs with binding motifs
180,322/434,218
(41.53%)
253,896/434,218
(58.47%)
Gene pairs without binding motifs
E
scRNA-seq + IReNA
Activation
Repression
F
3
Ribosome
Hmgb1
Irf3
1
Fatty acid metabolic process
Foxa2
5
Cell cycle
Foxo1
Klf13
2
Organic acid catabolic process
Hnf4a
4
Autophagy
Cebpb
Jund

Figure 3

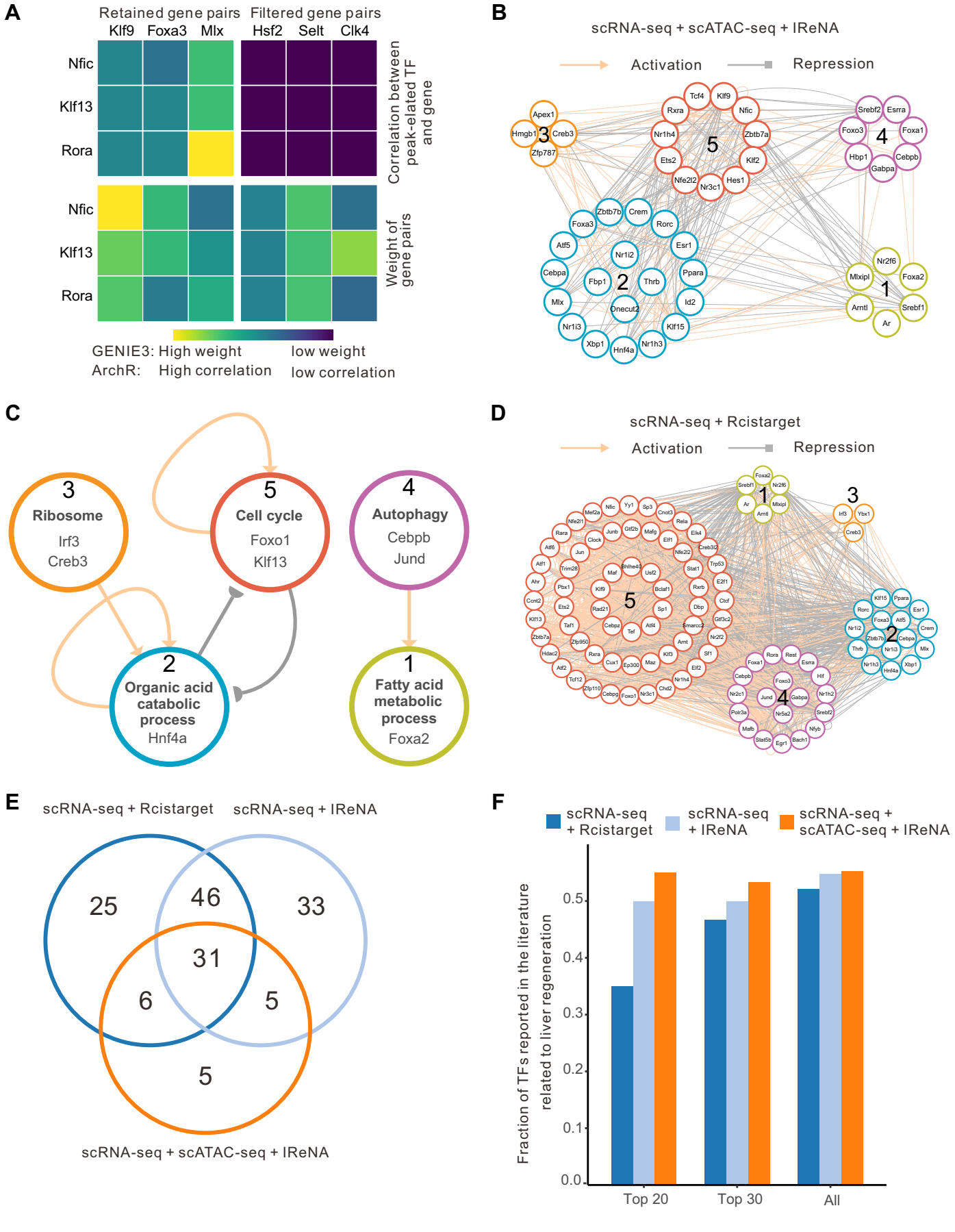

A
Retained gene pairs
Filtered gene pairs
Klf9 Foxa3 Mlx
Hsf2 Selt Clk4
Nfic
Klf13
Rora
Correlation between peak-elated TF and gene
Weight of gene pairs
GENIE3: High weight low weight
ArchR: High correlation low correlation
B
scRNA-seq + scATAC-seq + IReNA
Activation
Repression
C
3 Ribosome Irf3 Creb3
5 Cell cycle Foxo1 Klf13
4 Autophagy Cebpb Jund
2 Organic acid catabolic process Hnf4a
1 Fatty acid metabolic process Foxa2
D
scRNA-seq + Rcistarget
Activation
Repression
E
scRNA-seq + Rcistarget
scRNA-seq + IReNA
25
46
33
31
6
5
5
scRNA-seq + scATAC-seq + IReNA
F
scRNA-seq + Rcistarget
scRNA-seq + IReNA
scRNA-seq + scATAC-seq + IReNA
Fraction of TFs reported in the literature related to liver regeneration
0.0
0.1
0.2
0.3
0.4
0.5
Top 20
Top 30
All

Figure 4

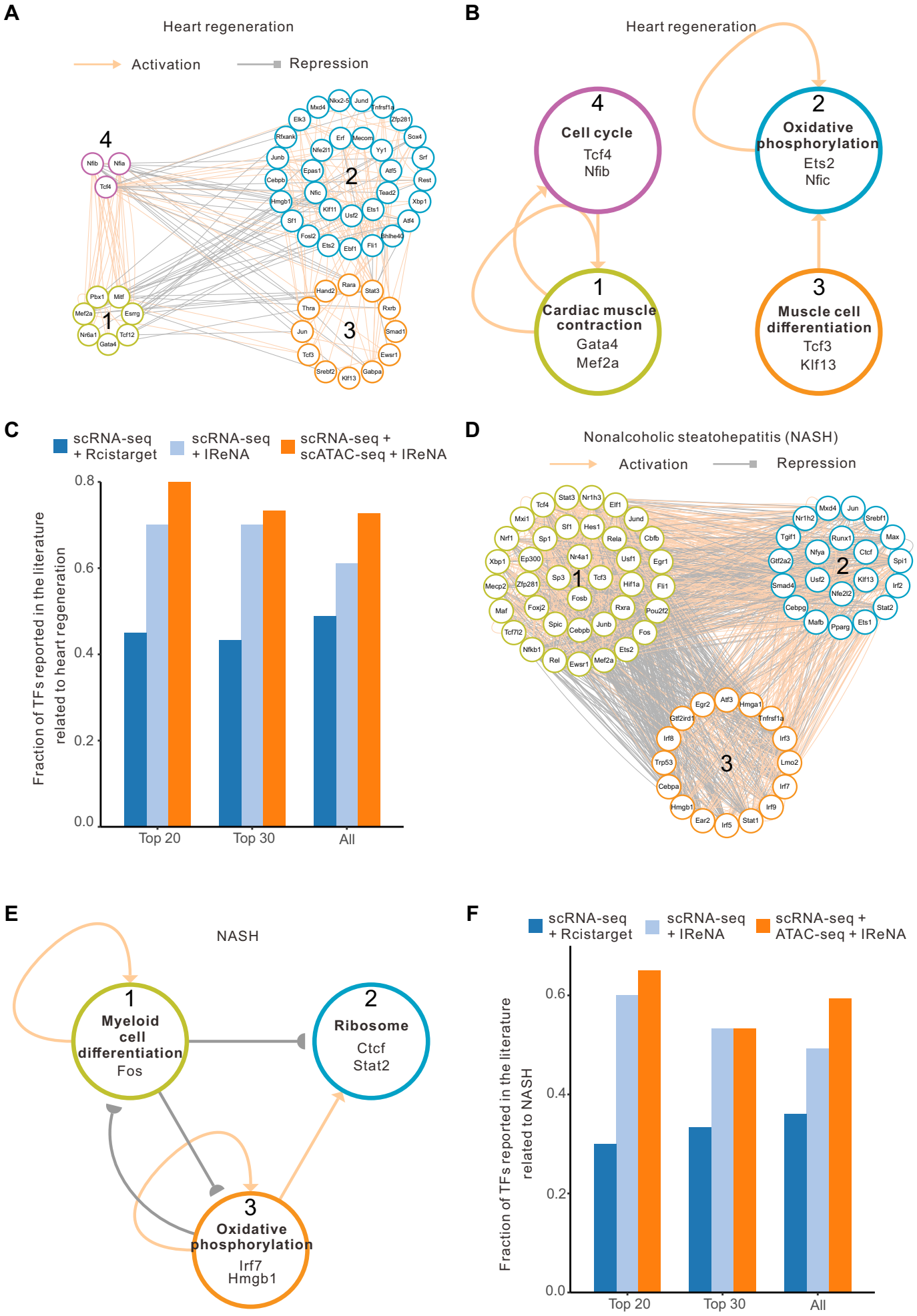

A
Heart regeneration
Activation
Repression
4
Nfib
Nfia
Tcf4
1
Pbx1
Mitf
Mef2a
Esrrg
Nr6a1
Tcf12
Gata4
2
Nkx2-5
Jund
Tnfrsf1a
Mxd4
Elk3
Zfp281
Rfxank
Erf
Mecom
Sox4
Nfe2l1
Yy1
Junb
Srf
Epas1
Atf5
Cebpb
Rest
Nfic
Tead2
Hmgb1
Xbp1
Klf11
Ets1
Usf2
Sf1
Atf4
Fosl2
Bhlhe40
Ets2
Ebf1
Fli1
3
Rara
Hand2
Stat3
Thra
Rxrb
Jun
Smad1
Tcf3
Ewsr1
Srebf2
Gabpa
Klf13
B
Heart regeneration
4
Cell cycle
Tcf4
Nfib
2
Oxidative phosphorylation
Ets2
Nfic
1
Cardiac muscle contraction
Gata4
Mef2a
3
Muscle cell differentiation
Tcf3
Klf13
C
scRNA-seq + Rcistarget
scRNA-seq + IReNA
scRNA-seq + scATAC-seq + IReNA
Fraction of TFs reported in the literature related to heart regeneration
0.0
0.2
0.4
0.6
0.8
Top 20
Top 30
All
D
Nonalcoholic steatohepatitis (NASH)
Activation
Repression
1
Stat3
Nr1h3
Tcf4
Elf1
Mxi1
Jund
Sf1
Hes1
Nrf1
Sp1
Rela
Cbfb
Xbp1
Ep300
Nr4a1
Usf1
Egr1
Mecp2
Zfp281
Sp3
Tcf3
Hif1a
Fli1
Maf
Foxj2
Fosb
Rxra
Pou2f2
Tcf7l2
Spic
Cebpb
Junb
Fos
Nfkb1
Rel
Ewsr1
Mef2a
Ets2
2
Mxd4
Jun
Nr1h2
Srebf1
Tgif1
Runx1
Max
Nfya
Ctcf
Gtf2a2
Spi1
Usf2
Klf13
Smad4
Irf2
Nfe2l2
Cebpg
Stat2
Mafb
Pparg
Ets1
3
Egr2
Atf3
Hmga1
Gtf2ird1
Tnfrsf1a
Irf8
Irf3
Trp53
Lmo2
Cebpa
Irf7
Hmgb1
Irf9
Ear2
Irf5
Stat1
E
NASH
1
Myeloid cell differentiation
Fos
2
Ribosome
Ctcf
Stat2
3
Oxidative phosphorylation
Irf7
Hmgb1
F
scRNA-seq + Rcistarget
scRNA-seq + IReNA
scRNA-seq + ATAC-seq + IReNA
Fraction of TFs reported in the literature related to NASH
0.0
0.2
0.4
0.6
Top 20
Top 30
All

Figure S1

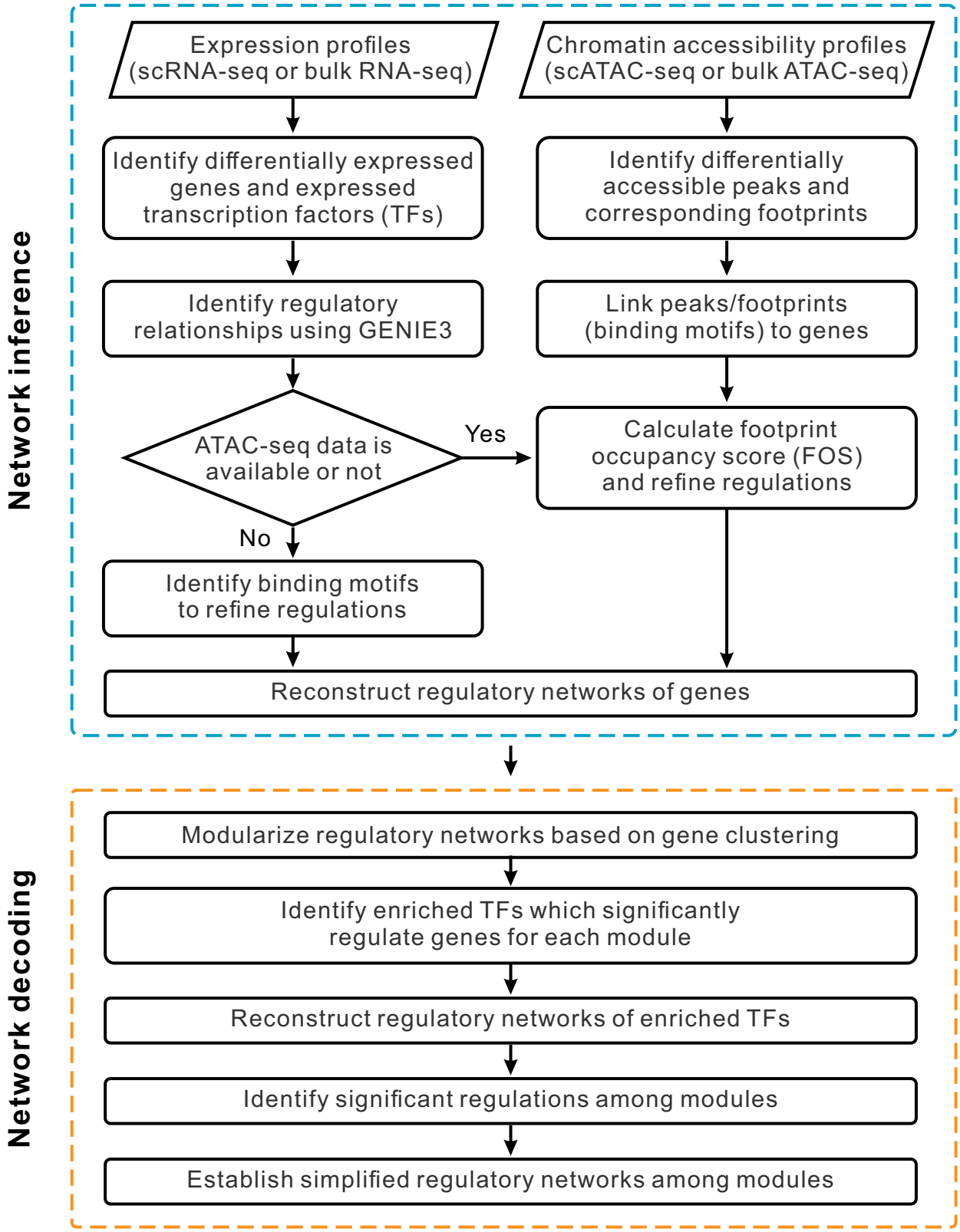

Network inference
Expression profiles (scRNA-seq or bulk RNA-seq)
Chromatin accessibility profiles (scATAC-seq or bulk ATAC-seq)
Identify differentially expressed genes and expressed transcription factors (TFs)
Identify differentially accessible peaks and corresponding footprints
Identify regulatory relationships using GENIE3
Link peaks/footprints (binding motifs) to genes
ATAC-seq data is available or not
Yes
No
Calculate footprint occupancy score (FOS) and refine regulations
Identify binding motifs to refine regulations
Reconstruct regulatory networks of genes
Network decoding
Modularize regulatory networks based on gene clustering
Identify enriched TFs which significantly regulate genes for each module
Reconstruct regulatory networks of enriched TFs
Identify significant regulations among modules
Establish simplified regulatory networks among modules

Figure S2

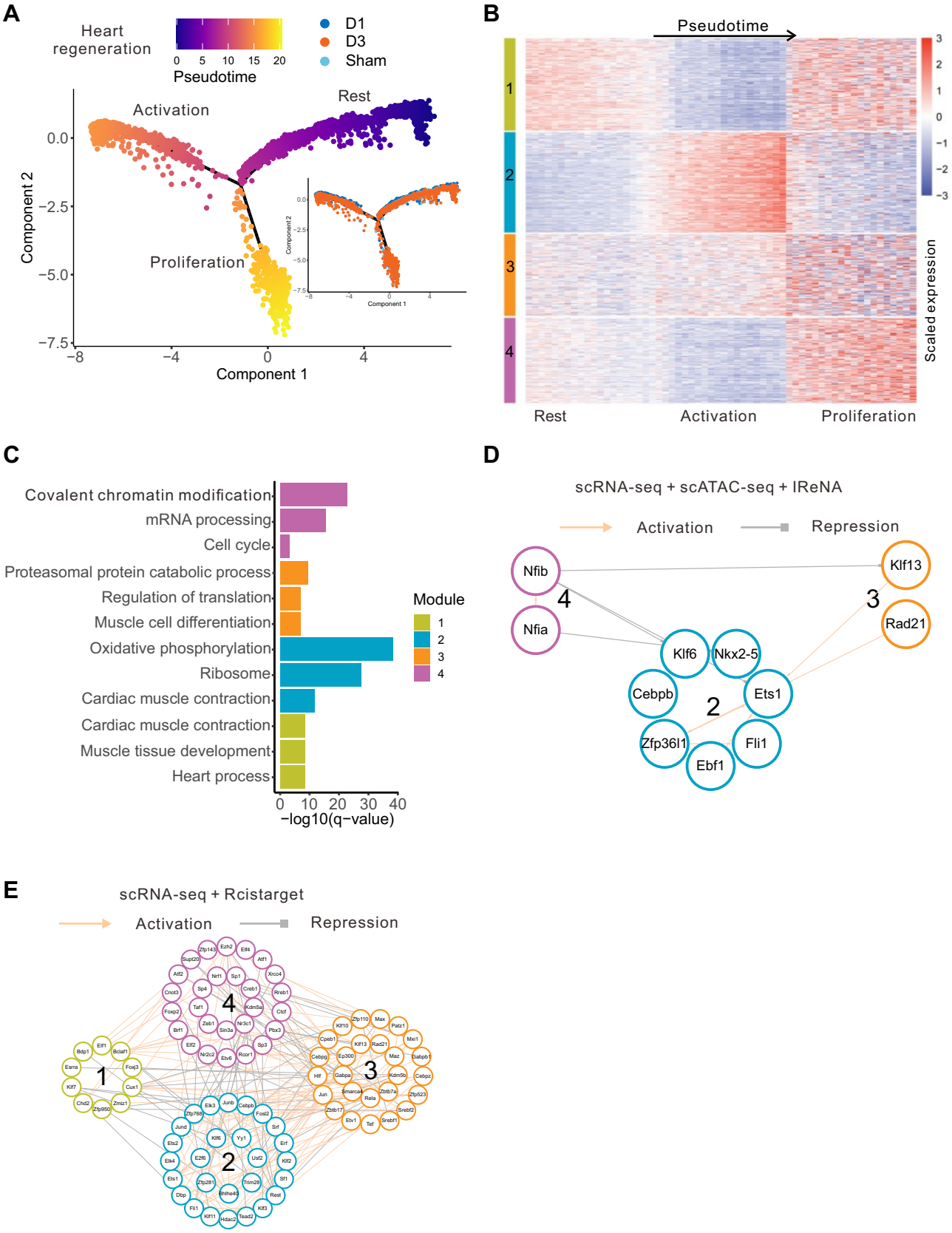
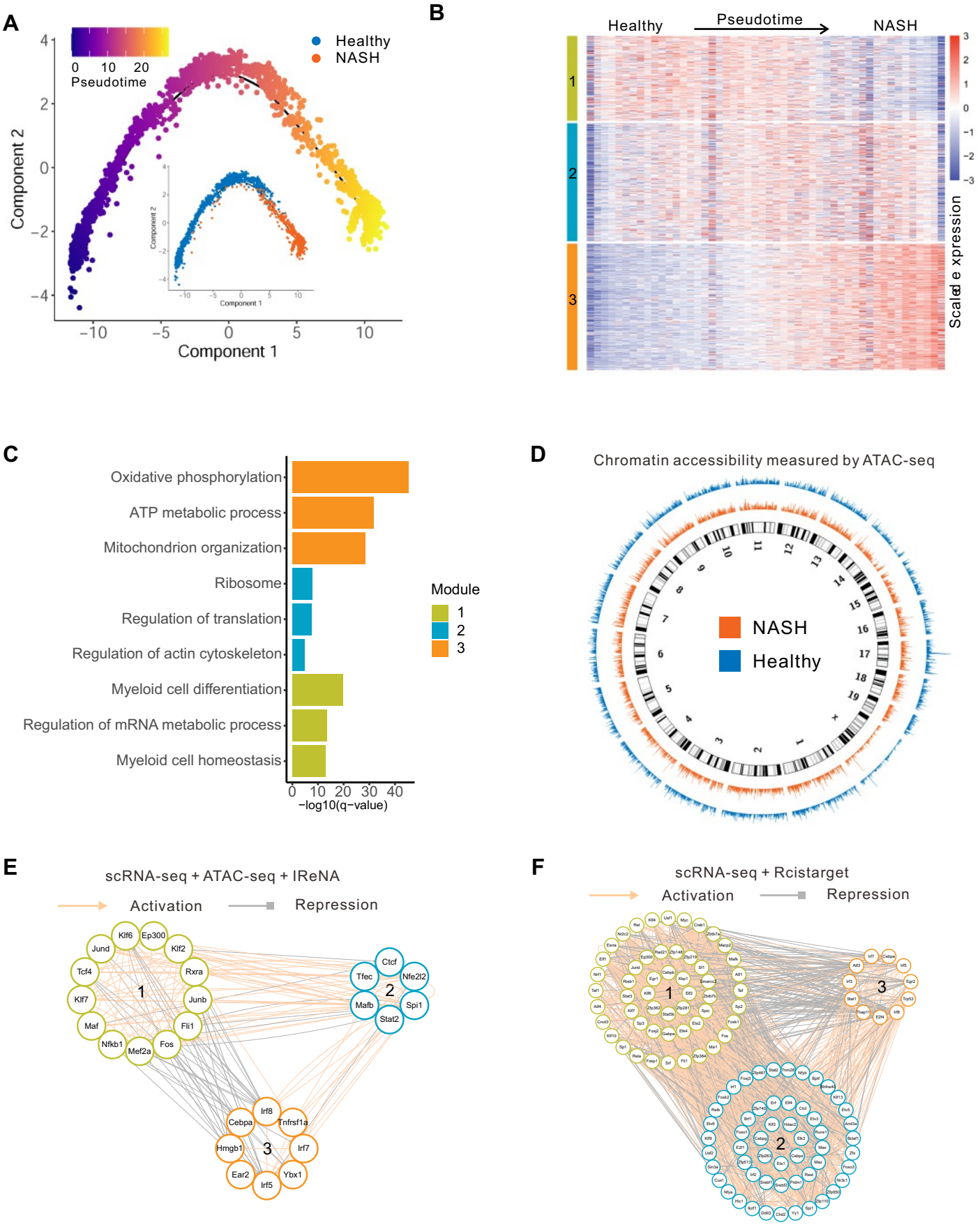

A
Heart regeneration
0 5 10 15 20
Pseudotime
D1
D3
Sham
Activation
Rest
Proliferation
Component 2
Component 1
B
Pseudotime
Scaled expression
Rest
Activation
Proliferation
C
Covalent chromatin modification
mRNA processing
Cell cycle
Proteasomal protein catabolic process
Regulation of translation
Muscle cell differentiation
Oxidative phosphorylation
Ribosome
Cardiac muscle contraction
Cardiac muscle contraction
Muscle tissue development
Heart process
Module
−log10(q−value)
D
scRNA-seq + scATAC-seq + IReNA
Activation
Repression
Nfib
Nfia
Klf13
Rad21
Klf6
Nkx2-5
Cebpb
Ets1
Zfp36l1
Fli1
Ebf1
E
scRNA-seq + Rcistarget
Activation
Repression

Figure S3


A
0 10 20
Pseudotime
Healthy
NASH
Component 2
Component 1
B
Healthy
Pseudotime
NASH
Scaled expression
C
Oxidative phosphorylation
ATP metabolic process
Mitochondrion organization
Ribosome
Regulation of translation
Regulation of actin cytoskeleton
Myeloid cell differentiation
Regulation of mRNA metabolic process
Myeloid cell homeostasis
−log10(q−value)
Module
1
2
3
D
Chromatin accessibility measured by ATAC-seq
NASH
Healthy
E
scRNA-seq + ATAC-seq + IReNA
Activation
Repression
Klf6
Ep300
Jund
Klf2
Tcf4
Rxra
Klf7
Junb
Maf
Fli1
Nfkb1
Mef2a
Fos
Ctcf
Tfec
Nfe2l2
Mafb
Spi1
Stat2
Irf8
Cebpa
Tnfrsf1a
Hmgb1
Irf7
Ear2
Ybx1
Irf5
F
scRNA-seq + Rcistarget
Activation
Repression