

# UC Santa Cruz

## UC Santa Cruz Previously Published Works

### Title

Prioritizing transcriptional factors in gene regulatory networks with PageRank

### Permalink

<https://escholarship.org/uc/item/8s05m2p4>

### Journal

iScience, 24(1)

### ISSN

2589-0042

### Authors

Ding, Hongxu  
Yang, Ying  
Xue, Yuanqing  
et al.

### Publication Date

2021

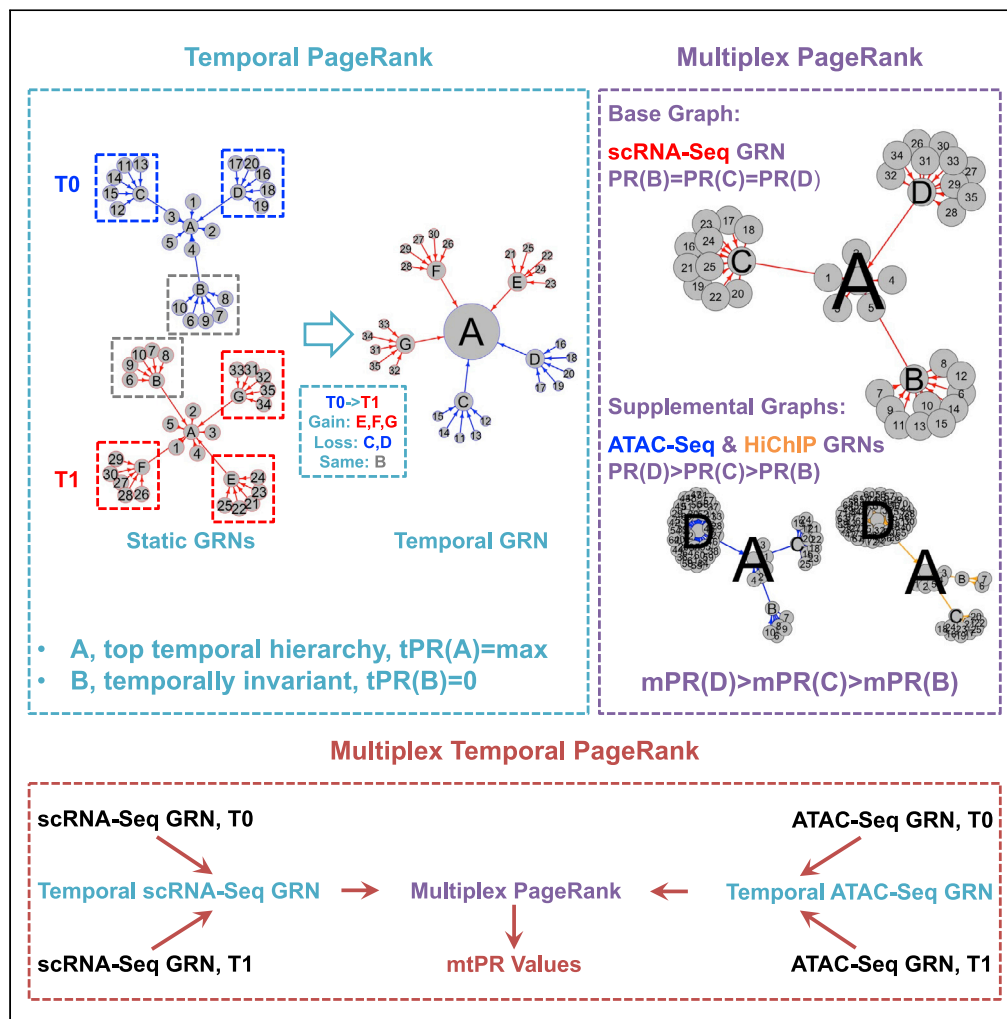
### DOI

10.1016/j.isci.2020.102017

Peer reviewed

Article

# Prioritizing transcriptional factors in gene regulatory networks with PageRank



Hongxu Ding,  
Ying Yang,  
Yuanqing Xue, ...,  
Rojin Safavi,  
Andrea Califano,  
Joshua M. Stuart

hdng16@ucsc.edu (H.D.)  
jstuart@ucsc.edu (J.M.S.)

**HIGHLIGHTS**

Temporal PageRank  
prioritizes TFs controlling  
cellular state dynamics

Multiplex PageRank  
prioritizes TFs by  
integrating multi-omics  
GRNs

Temporal and multiplex  
PageRank can be used in  
combination for TF  
prioritization



## Article

Prioritizing transcriptional factors  
in gene regulatory  
networks with PageRank

Hongxu Ding,<sup>1,2,4,5,\*</sup> Ying Yang,<sup>1,3,4</sup> Yuanqing Xue,<sup>1,4</sup> Lucas Seninge,<sup>1</sup> Henry Gong,<sup>1</sup> Rojin Safavi,<sup>1</sup> Andrea Califano,<sup>2</sup> and Joshua M. Stuart<sup>1,\*</sup>

## SUMMARY

**Biological states are controlled by orchestrated transcriptional factors (TFs) within gene regulatory networks. Here we show TFs responsible for the dynamic changes of biological states can be prioritized with temporal PageRank. We further show such TF prioritization can be extended by integrating gene regulatory networks reverse engineered from multi-omics profiles, e.g. gene expression, chromatin accessibility, and chromosome conformation assays, using multiplex PageRank.**

## INTRODUCTION

Biological processes are primarily executed via gene regulatory networks (GRNs), which are controlled by key transcriptional factors (TFs) (Levine and Davidson, 2005; Califano and Alvarez, 2017). Such key TFs usually occupy the top of the gene regulatory hierarchy (Chan and Kyba, 2013). The regulatory hierarchy of a specific TF depends on the number and hierarchy of corresponding transcriptional targets, thus can be quantified by PageRank centrality (Brin and Page, 1998; Page et al., 1999). Originally proposed for ranking search results of World Wide Web (WWW) snapshots, PageRank and related algorithms has been successfully applied to the analysis of single static biological networks (Morrison et al., 2005; Koschützki and Schreiber, 2008; Tarca et al., 2009; Iván and Grolmusz, 2011). The advent of high-throughput sequencing technologies provide unprecedented temporal and multi-dimensional biological information for understanding transcriptional regulation. For instance, transcriptional regulatory dynamics among consecutive biological states can be characterized with single cell RNA sequencing (scRNA-Seq) (Kolodziejczyk et al., 2015) and trajectory analysis (Herring et al., 2018). Meanwhile, epigenetic regulation of gene transcription can be illustrated using, e.g. chromatin accessibility (Klemm et al., 2019) and chromosome conformation (Sati and Cavalli, 2017) assays. Here, we show within such temporal and multiplex GRNs, TFs can be prioritized with temporal (Rozenshtein and Gionis, 2016) and multiplex (Halu et al., 2013) PageRank.

As the extension of original steady-state PageRank in temporal networks, temporal PageRank ranks nodes based on their connections that change over time (Rozenshtein and Gionis, 2016). In temporal GRNs, important TFs are those connected with more time-related targets and other important TFs. Such TFs will then be considered at the top of the temporal gene regulatory hierarchy and prioritized (Figure 1A, see Transparent methods). Multiplex PageRank, on the other hand, extends PageRank analysis to multiplex networks. In such networks, the same nodes might interact with one another in different layers. Multiplex PageRank is then calculated according to the topology of a predefined base network, with regular PageRank of other supplemental networks as edge weights and personalization vector (Halu et al., 2013). Therefore, GRNs reverse engineered from multi-omics assays can be integrated for TF prioritization (Figure 1B, see Transparent methods).

## RESULTS

## Interpreting regulatory dynamics using temporal PageRank

We first demonstrated TFs controlling cellular state transitions can be prioritized with temporal PageRank. Specifically, we analyzed the human myoblast-muscle cell differentiation process, during which single cells were harvested and profiled every 24 hr from T0 to T72 (Trapnell et al., 2014). To provide intuitions for the rationale of temporal PageRank analysis, following the schematic diagram in Figure 1A, we visualized static GRNs at neighboring timepoints, as well as the yielding temporal GRNs. Considering the sizes of such static/temporal GRNs, for clear visualization, we only highlighted key regulatory modules by filtering out less confident

<sup>1</sup>Department of Biomolecular Engineering and Genomics Institute, University of California, Santa Cruz, Santa Cruz, CA, USA

<sup>2</sup>Department of Systems Biology, Columbia University, New York, NY, USA

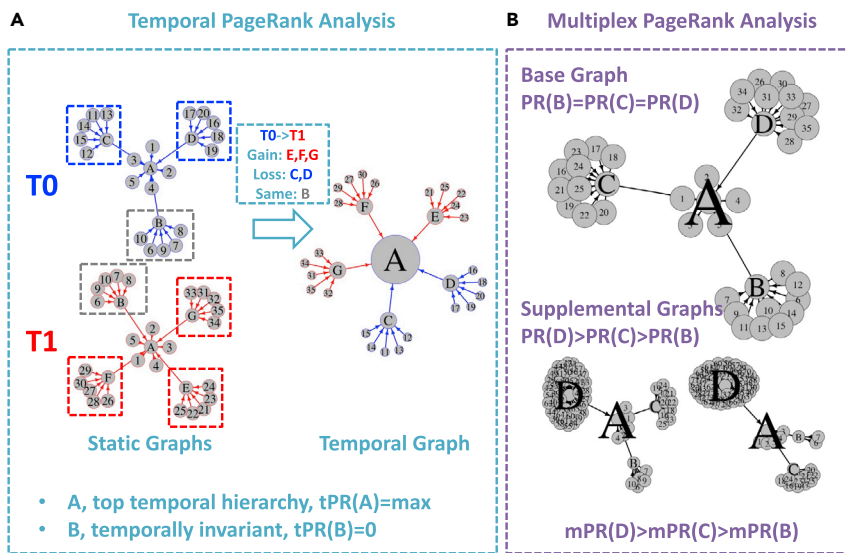
<sup>3</sup>Department of Dermatology, Stanford University, Stanford, CA, USA

<sup>4</sup>These authors contributed equally

<sup>5</sup>Lead Contact

\*Correspondence: hding16@ucsc.edu (H.D.), jstuart@ucsc.edu (J.M.S.)  
<https://doi.org/10.1016/j.isci.2020.102017>





**Figure 1. Graphic overview**

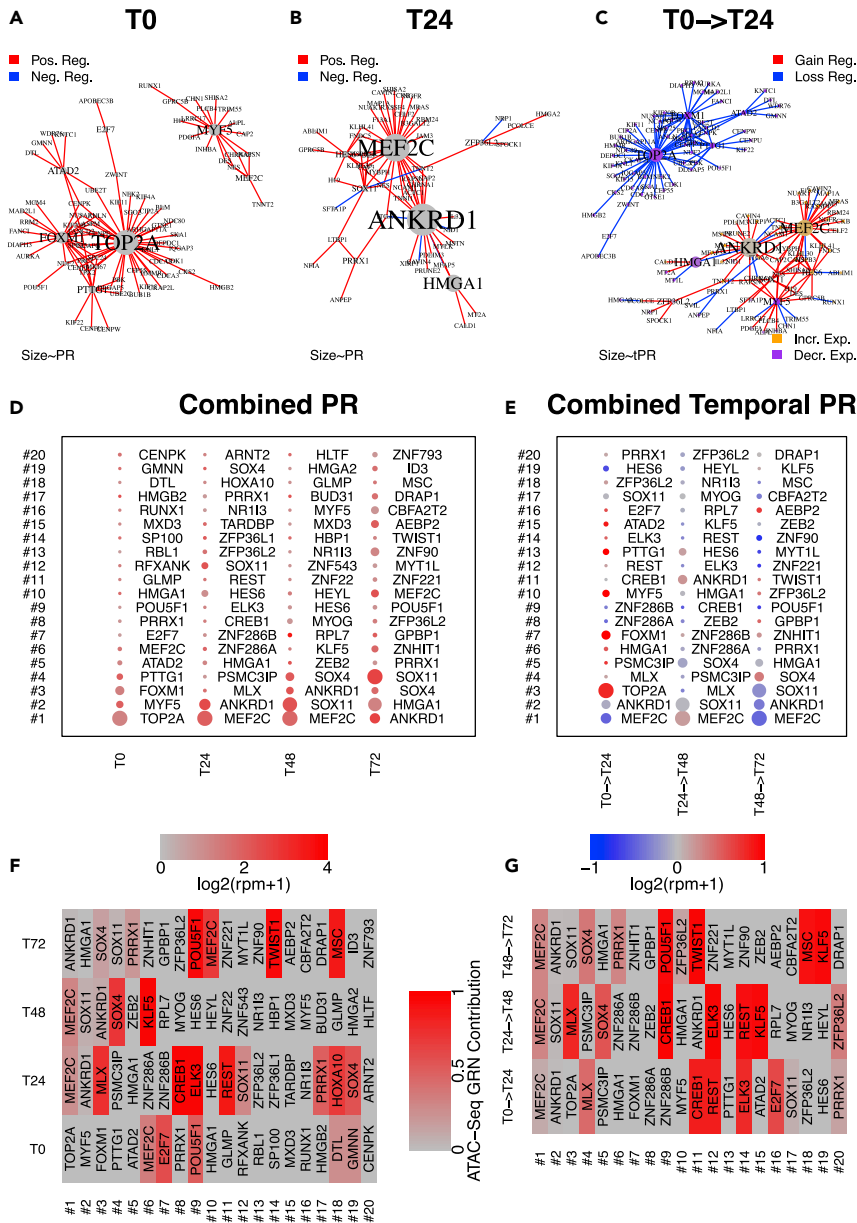
Schematic diagrams of temporal and multiplex PageRank analysis are shown in (A) and (B), respectively.

interactions (see [Transparent methods](#)). As shown in [Figure 2A](#), two independent regulatory modules were identified at T0. The first module was controlled by cell cycle TFs *TOP2A* ([Mjelle et al., 2015](#)) and *FOXM1* ([Wierstra and Alves, 2007](#)), in concordance with the active proliferation of myoblasts. The second module, which is responsible for the lineage identity of myoblast, was marked by the lineage-specific TF *MYF5* ([Blais et al., 2005](#)). As for T24 ([Figure 2B](#)) and the following timepoints ([Figure S1](#)), the corresponding GRNs were majorly controlled by a single regulatory module. Such a module was composed of muscle cell-specific TFs, including muscle cell lineage markers *MEF2C* and *ANKRD1* ([Blais et al., 2005](#)), as well as epigenetic modifier *HMGAI* ([Brocher et al., 2010](#)). Thus, the differentiation of myoblast can be described as the sequential interplay of key TFs. We further applied temporal PageRank on the differential GRNs derived from the corresponding adjacent static counterparts. As shown in [Figures 2C and S1](#), the regulatory dynamics of myoblast-muscle cell differentiation was recapitulated, by discovering all the above-mentioned key TFs. We also analyzed the 33 major lineages during mouse organogenesis reported in the MOCA data sets ([Cao et al., 2019](#)) as the additional proof-of-concept ([Figure S2](#)).

### Integrating multi-omics GRNs using multiplex PageRank

We then demonstrated GRNs reverse engineered from multi-omics assays can be integrated through multiplex PageRank for TF prioritization. Specifically, we included matching ATAC-Seq profiles ([Pliner et al., 2018](#)) of the above-mentioned differentiation process. We then constructed static and temporal GRNs, and performed corresponding PageRank analysis, following the workflow described in [Transparent methods](#). Although the scRNA-Seq and ATAC-Seq GRNs were topologically different ([Figure S1](#)), muscle cell signature TF *MEF2C* was identified with both GRN types across the entire differentiation process ([Figure S3](#)). Meanwhile, additional muscle cell TFs, e.g. *KLF5* ([Hayashi et al., 2016](#)) and *REST* ([Iannotti et al., 2013](#)) were recapitulated by analyzing ATAC-Seq GRNs ([Figure S3](#)). Such results suggested GRNs reverse engineered from multi-omics profiles agreed on the general principle, while each provided unique insights into the gene regulatory machinery. Aiming at prioritizing TFs by combining scRNA-Seq and ATAC-Seq GRNs, we performed multiplex PageRank analysis. The contributions of scRNA-Seq and ATAC-Seq GRNs were quantified in [Figures 2F and 2G](#). Noticeably, multiplex PageRank can be applied to integrate GRNs under both static and temporal scenarios. As shown in [Figures 2D, 2E and S4](#), key TFs elucidated from scRNA-Seq and ATAC-Seq GRNs were together recapitulated.

As an additional proof-of-concept, we analyzed the human hematopoiesis process, including the linear lineage progression of hematopoietic stem cell, multi-potent progenitor, and CMP (common myeloid progenitor), as well as the bifurcation from CMP to granulocyte-macrophage progenitor and megakaryocyte-erythroid progenitor, with multiplex PageRank. Following the same pipeline as the previous analysis, GRNs assembled from matching scRNA-Seq ([Pellin et al., 2019](#)) and ATAC-Seq ([Corces et al., 2016](#)) data sets were analyzed ([Figure S5](#)).



**Figure 2. PageRank analysis on the myoblast-muscle cell differentiation process**

(A and B) Static scRNA-Seq GRNs for T0 and T24. Vertex and label sizes correspond to static PageRank values. Red and blue edges correspond to positive and negative interactions, respectively.

(C) Temporal scRNA-Seq GRN between T0 and T24. Vertex and label sizes correspond to temporal PageRank values. Orange and purple vertices correspond to increased and decreased gene expression from T0 to T24, respectively. Red and blue edges correspond to gained and lost interactions from T0 to T24, respectively.

(D and E) Bubble plots showing the top 20 combined static and temporal PageRank candidates by analyzing scRNA-Seq and ATAC-Seq GRNs. The size of bubbles correspond to the degree values (number of connecting interactions). The color of bubbles correspond to the gene expression quantified by  $\log_2(\text{rpm}+1)$ , where rpm stands for reads per million. For static PageRank, absolute gene expression was quantified, with red and gray corresponding to high and low gene expression, respectively. For temporal PageRank, differential gene expression was quantified, with red and blue corresponding to increased and decreased gene expression, respectively.

(F and G) Heatmaps showing the contribution of ATAC-Seq GRNs in TF prioritization. The contributions of scRNA-Seq and ATAC-Seq GRNs were normalized to 1. F and G describes static and temporal PageRank analysis, respectively.

We further expanded multiplex PageRank to integrate gene expression, chromatin accessibility and chromosome conformation GRNs, by analyzing scRNA-Seq, ATAC-Seq, and HiChIP (Mumbach et al., 2017) profiles of human T-cells. We used scRNA-Seq GRN as the base network for the integration. As shown in Figure S6, several known crucial TFs responsible for T cell homeostasis, e.g. *FOXP1* (Feng et al., 2011) and functionalities, e.g. *LEF1* (Travis et al., 1991) were recapitulated among the top 20 identified TFs. Moreover, a systematic survey of the top 20 TFs was performed with GO analysis (<http://geneontology.org/>). As shown in Table S1, a significant amount of T-cell-related biological processes were recapitulated. Noticeably, the three included GRNs complement each other by providing unique insights into gene regulatory machinery. For instance, the prioritization of *LEF1* and *FOXP1* was majorly contributed by the HiChIP and ATAC-Seq GRNs, respectively.

## DISCUSSION

Taken together, by analyzing diverse biological questions, we demonstrated that key TFs responsible for biological processes can be prioritized by analyzing GRNs using PageRank. Specifically, we showed that the crucial TFs controlling the dynamic transition of biological states can be prioritized with temporal PageRank. Further, we showed GRNs reverse engineered from multi-omics profiles can be integrated for TF prioritization with multiplex PageRank.

PageRank quantifies the importance of TFs during biological processes by performing comprehensive surveys on GRN hierarchies, therefore extremely suitable for TF prioritization. Specifically, PageRank analysis can prioritize TFs even if their expression patterns are obscure. For instance, as shown in Figure 2E, although no strong differential expression was observed during the transition from T0 to T24, muscle cell-specific TF *ANKRD1* was ranked #2 with temporal PageRank analysis. Meanwhile, PageRank analysis prioritizes TFs with the entire GRN hierarchies, rather than based on “flattened” architectures which only consider the direct targets. For instance, as shown in Figure S2, during the mouse embryo development of inhibitory interneuron lineage from stage E10.5 to stage E11.5, the degree centrality of *Sox6* was insignificant, while ranked #3 by temporal PageRank analysis. We further performed a systematic comparison between our PageRank analysis with a state-of-the-art TF prioritization algorithm, VIPER (Alvarez et al., 2016; Ding et al., 2018) (see Figure S11 for details).

## Limitations of the study

Taken together, we anticipate the PageRank analysis would provide novel and comprehensive insights for the understanding of transcriptional regulation, by identifying regulators that potentially reside at the top of the regulatory hierarchy. One thing to be noticed for temporal PageRank analysis is that, we would not recommend applying it on distinct networks. Consider an extreme case, where (1) the number of nodes in network A and B are the same, while A has 10 times more interactions than B, and (2) the two networks have no overlapping interactions, for example. The differential network between A and B will include all interactions in A and B. If temporal PageRank analysis is performed on such a differential network, the yielded top ranks will be dominated by A nodes, considering the 10 times more interactions. B nodes, on the other hand, are under-appreciated even though they also convey important information describing differences between A and B. As for analyzing GRNs, similar biases could happen. For instance, the epigenetic landscape of zygotes is less restricted, thus more regulatory interactions are expected. In contrast, the more restricted epigenetic landscapes reduce possible regulatory interactions in terminally differentiated cells such as T-cells. Thus, the temporal PageRank analysis between zygote and T cell GRNs is highly likely to ignore functional TFs in T-cells. We thus would suggest only applying temporal PageRank analysis between temporally adjacent networks.

As for multiplex PageRank analysis, one limitation is that the results might vary according to the choice of the base network. For instance, during the myoblast-muscle cell differentiation process, although known myoblast-specific TFs *SP1/3* (Parakati and DiMario, 2002) were prioritized by PageRank analyses based on only ATAC-Seq GRNs (Figure S3), they failed to be captured when using scRNA-Seq GRNs as base networks for multiplex PageRank analyses (Figures 2D and 2E). On the other hand, the effect of base network choice on the final multiplex PageRank integration was minor when analyzing T-cells (Figure S7). Thus, one possible future direction might be developing “reciprocal” multiplex PageRank analysis for GRNs, in which feedbacks among multiplex networks are considered as reported in (Tu et al., 2018).

## Resource availability

### Lead contact

Hongxu Ding, [hding16@ucsc.edu](mailto:hding16@ucsc.edu).

### Material availability

This study did not generate any new material.

### Data and code availability

scRNA-Seq profiles of myoblast-muscle cell differentiation were downloaded from Gene Expression Omnibus (GEO) under accession number GSE52529. MOCA scRNA-Seq profiles were downloaded from GEO under accession number GSE119945. Hematopoiesis scRNA-Seq profiles were downloaded from GEO under accession number GSE117498. Healthy PBMC T cell scRNA-Seq profiles were downloaded from: [https://support.10xgenomics.com/single-cell-gene-expression/datasets/3.0.0/pbmc\\_10k\\_v3](https://support.10xgenomics.com/single-cell-gene-expression/datasets/3.0.0/pbmc_10k_v3). ATAC-Seq profiles of myoblast-muscle cell differentiation were downloaded from GEO under accession number GSE109828. Hematopoiesis ATAC-Seq profiles were downloaded from GEO under accession number GSE74912. T cell ATAC-Seq and Hi-ChIP profiles were downloaded from GEO under accession number GSE101498.

The pageRank R package is available on Bioconductor, and GitHub repository <https://github.com/hd2326/pageRank>. Custom scripts used to reproduce the results and figures are also available at <https://github.com/hd2326/pageRank>.

## METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.102017>.

## ACKNOWLEDGMENTS

J.M.S. was supported by a grant 5R01GM109031 from the NIGMS. J.M.S. and H.D. were supported by a grant from the Chan Zuckerberg Initiative's Human Cell Atlas portals project. H.D. was supported by a gift from Seagate Technology. J.M.S. was supported by grant GC1R-06673-C from the California Institute for Regenerative Medicine's Center of Excellence for Stem Cell Genomics.

## AUTHOR CONTRIBUTIONS

H.D. conceived the idea. H.D. and A.C. performed ARACNe analysis. H.D. and Y.Y. performed epigenetic network analysis. H.D., Y.Y., Y.X., H.G., and R.S. performed the PageRank analysis. H.D., Y.Y., Y.X., and L.S. collected and pre-processed the data. H.D. and J.M.S. supervised the project. H.D., Y.Y. and J.M.S. wrote the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 6, 2020

Revised: November 6, 2020

Accepted: December 28, 2020

Published: January 22, 2021

## REFERENCES

- Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* 48, 838–847.
- Blais, A., Tsikitis, M., Acosta-Alvarez, D., Sharan, R., Kluger, Y., and Dynlacht, B.D. (2005). An initial blueprint for myogenic differentiation. *Genes Dev.* 19, 553–569.
- Brin, S., and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Comput. Netw.* 30, 107–117.
- Brocher, J., Vogel, B., and Hock, R. (2010). HMGA1 down-regulation is crucial for chromatin composition and a gene expression profile permitting myogenic differentiation. *BMC Cell Biol.* 11, 1–12.
- Califano, A., and Alvarez, M.J. (2017). The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nat. Rev. Cancer* 17, 116.
- Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., and Trapnell, C. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566, 496–502.
- Chan, S.S.K., and Kyba, M. (2013). What is a master regulator? *J. Stem Cell Res. Ther.* 3, <https://doi.org/10.4172/2157-7633.1000e114>.
- Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., and Majeti, R. (2016). Lineage-specific and single-cell chromatin accessibility charts human



- hematopoiesis and leukemia evolution. *Nat. Genet.* 48, 1193–1203.
- Ding, H., Douglass, E.F., Sonabend, A.M., Mela, A., Bose, S., Gonzalez, C., Canoll, P.D., Sims, P.A., Alvarez, M.J., and Califano, A. (2018). Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nat. Commun.* 9, 1–10.
- Feng, X., Wang, H., Takata, H., Day, T.J., Willen, J., and Hu, H. (2011). Transcription factor Foxp1 exerts essential cell-intrinsic regulation of the quiescence of naive T cells. *Nat. Immunol.* 12, 544–550.
- Halu, A., Mondragón, R.J., Panzarasa, P., and Bianconi, G. (2013). Multiplex pagerank. *PLoS One* 8, e78293.
- Hayashi, S., Manabe, I., Suzuki, Y., Relaix, F., and Oishi, Y. (2016). Klf5 regulates muscle differentiation by directly targeting muscle-specific genes in cooperation with MyoD in mice. *Elife* 5, e17462.
- Herring, C.A., Chen, B., McKinley, E.T., and Lau, K.S. (2018). Single-cell computational strategies for lineage reconstruction in tissue systems. *Cell Mol. Gastroenterol. Hepatol.* 5, 539–548.
- Iannotti, F.A., Barrese, V., Formisano, L., Miceli, F., and Tagliatela, M. (2013). Specification of skeletal muscle differentiation by repressor element-1 silencing transcription factor (REST)-regulated Kv7.4 potassium channels. *Mol. Biol. Cell* 24, 274–284.
- Iván, G., and Grolmusz, V. (2011). When the Web meets the cell: using personalized PageRank for analyzing protein interaction networks. *Bioinformatics* 27, 405–407.
- Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 20, 207–220.
- Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The technology and biology of single-cell RNA sequencing. *Mol. Cell* 58, 610–620.
- Koschützki, D., and Schreiber, F. (2008). Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regul. Syst. Biol.* 2, GRSB-S702.
- Levine, M., and Davidson, E.H. (2005). Gene regulatory networks for development. *Proc. Natl. Acad. Sci. U S A* 102, 4936–4942.
- Mjelle, R., Hegre, S.A., Aas, P.A., Slupphaug, G., Drabløs, F., Sætrum, P., and Krokan, H.E. (2015). Cell cycle regulation of human DNA repair and chromatin remodeling genes. *DNA Repair* 30, 53–67.
- Morrison, J.L., Breitling, R., Higham, D.J., and Gilbert, D.R. (2005). GeneRank: using search engine technology for the analysis of microarray experiments. *BMC Bioinformatics* 6, 233.
- Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., and Wei, Y. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* 49, 1602.
- Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The PageRank Citation Ranking: Bringing Order to the Web (Stanford InfoLab).
- Parakati, R., and DiMario, J.X. (2002). Sp1-and Sp3-mediated transcriptional regulation of the fibroblast growth factor receptor 1 gene in chicken skeletal muscle cells. *J. Biol. Chem.* 277, 9278–9285.
- Pellin, D., Loperfido, M., Baricordi, C., Wolock, S.L., Montepeloso, A., Weinberg, O.K., Biffi, A., Klein, A.M., and Biasco, L. (2019). A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat. Commun.* 10, 1–15.
- Pliner, H.A., Packer, J.S., McFaline-Figueroa, J.L., Cusanovich, D.A., Daza, R.M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A., and Adey, A.C. (2018). Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* 71, 858–871.
- Rozenshtein, P., and Gionis, A. (2016). September. Temporal pagerank. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (Springer)*, pp. 674–689.
- Sati, S., and Cavalli, G. (2017). Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma* 126, 33–44.
- Tarca, A.L., Draghici, S., Khatri, P., Hassan, S.S., Mittal, P., Kim, J.S., Kim, C.J., Kusanovic, J.P., and Romero, R. (2009). A novel signaling pathway impact analysis. *Bioinformatics* 25, 75–82.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381.
- Travis, A., Amsterdam, A., Belanger, C., and Grosschedl, R. (1991). LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function [corrected]. *Genes Dev.* 5, 880–894.
- Tu, X., Jiang, G.P., Song, Y., and Zhang, X. (2018). Novel multiplex PageRank in multilayer networks. *IEEE Access* 6, 12530–12538.
- Wierstra, I., and Alves, J. (2007). FOXM1, a typical proliferation-associated transcription factor. *Biol. Chem.* 388, 1257–1274.



**iScience, Volume 24**

## **Supplemental Information**

### **Prioritizing transcriptional factors in gene regulatory networks with PageRank**

**Hongxu Ding, Ying Yang, Yuanqing Xue, Lucas Seninge, Henry Gong, Rojin Safavi, Andrea Califano, and Joshua M. Stuart**

## Supplemental Data Items

**Figure S1. [Graph visualization of myoblast-muscle cell differentiation GRNs], related to Figure 2.** (A) Static scRNA-Seq GRNs. Bigger vertices corresponded to higher regular PageRank values. Red/blue edges corresponded to positive/negative interactions, respectively. (B) Static ATAC-Seq GRNs. Bigger vertices corresponded to higher regular PageRank values. (C, D) Temporal scRNA-Seq and ATAC-Seq GRNs. Bigger vertices corresponded to higher temporal PageRank values. Orange/purple vertices corresponded to increased/decreased gene expression, respectively. Red/blue edges corresponded to gained/lost interactions, respectively. T0, T24, T48 and T72 corresponded to sampling timepoints alongside the differentiation process as reported in [1].

**Figure S2. [Static PageRank analysis on MOCA lineages], related to Figure 2.** We analyzed mouse organogenesis lineages reported in the MOCA datasets [2]. For every lineage, the top 20 static and temporal PageRank (the left and right panels of every lineage) candidates were shown in bubble plots. The size of bubbles corresponded to the degree: the more connecting interactions the bigger the bubble. The color of bubbles corresponded to the gene expression. For static PageRank panels, red/grey corresponded to high/low gene expression, respectively. For temporal PageRank panels, red/blue corresponded to increased/decreased gene expression, respectively. E9.5 etc, embryonic stages.

**Figure S3. [Static and temporal PageRank analysis on myoblast-muscle cell differentiation], related to Figure 2.** The top 20 static and temporal PageRank candidates by analyzing scRNA-Seq [1] and ATAC-Seq [3] GRNs were shown in bubble plots. The size of bubbles corresponded to the degree: the more connecting interactions the bigger the bubble. The color of bubbles corresponded to the gene expression. For static PageRank analysis, red/grey corresponded to high/low gene expression, respectively. For temporal PageRank analysis, red/blue corresponded to increased/decreased gene expression, respectively. PR, static PageRank; tPR, temporal PageRank; T0 etc, differentiation timepoints.

**Figure S4. [Multiplex PageRank analysis on myoblast-muscle cell differentiation], related to Figure 2.** The top 20 multiplex (additive, multiplicative, combined and neutral, see METHODS for details) PageRank candidates by combining scRNA-Seq [1] and ATAC-Seq [3] GRNs were shown in bubble plots. The size of bubbles corresponded to the degree: the more connecting interactions the bigger the bubble. The color of bubbles corresponded to the gene expression. For static PageRank analysis, red/grey corresponded to high/low gene expression, respectively. For temporal PageRank analysis, red/blue corresponded to increased/decreased gene expression, respectively. PR, static PageRank; tPR, temporal PageRank; T0 etc, differentiation timepoints.

**Figure S5. [Multiplex PageRank analysis on hematopoiesis], related to Figure 2.** The top 20 multiplex (additive, multiplicative, combined and neutral, see METHODS for details) PageRank candidates by combining scRNA-Seq [4] and ATAC-Seq [5] GRNs were shown in bubble plots. The size of bubbles corresponded to the degree: the more connecting interactions the bigger the bubble. The color of bubbles corresponded to the gene expression. For static PageRank analysis, red/grey corresponded to high/low gene expression, respectively. For temporal PageRank analysis, red/blue corresponded to increased/decreased gene expression, respectively. PR, static PageRank; tPR, temporal PageRank; T0 etc, differentiation timepoints.

**Figure S6. [PageRank analysis on T-cells], related to Figure 2.** (A) Static PageRank values calculated from scRNA-Seq, ATAC-Seq and HiChIP GRNs reverse engineered from healthy human T-cells were analyzed [6]. The size of points correspond to the combined static PageRank values, and the color of points correspond to whether they are TFs or targets. (B) Top 20 combined static PageRank candidates. The contributions of each GRN were represented as normalized fractions. Red green and blue correspond to scRNA-Seq, ATAC-Seq and HiChIP GRNs, respectively.

**Figure S7. [Effect of base network on multiplex PageRank], related to Figure 2.** The scRNA-Seq, ATAC-Seq and HiChIP static GRNs reverse engineered from healthy human T-cells were analyzed [6]. (A) Multiplex PageRank values calculated from different base GRNs. (B) Scatter plot of regular versus multiplex (additive, multiplicative, combined and neutral, see METHODS for details) PageRank values calculated from different base GRNs.

**Figure S8. [Overview of the analytical workflow], related to Figure 2.**

**Figure S9. [Overview of epigenetic GRN construction], related to Figure 2.**

**Figure S10. [Complexity analysis], related to Figure 2.** We generated 10 random Erdos-Renyi graphs, represented as ER( $n, p$ ), per  $n$ - $p$  combination to gain enough statistical power for complexity analysis. Specifically,  $\log_2(n)$  ranges from 5 to 14, while  $p$  equals 0.1. For both multiplex and temporal PageRank analyses, the complexity scales exponentially with the network size. Noticeably, even the largest ER graphs, which contain  $2^{14}$  (16384) nodes, can be processed within 2 hours, suggesting the decent efficiency of our temporal and multiplex PageRank implementations.

**Figure S11. [Systematic comparison with state-of-the-art TF prioritization algorithm VIPER], related to Figure 2.** Enrichment analysis-based algorithms, e.g. VIPER [7] and successor metaVIPER [8], as well as SCENIC [9], emerged as powerful tools for prioritizing TFs from GRNs. Such algorithms evaluate the activity of TFs by enriching the gene expression signatures against the corresponding transcriptional targets. If a majority of positive/negative targets of a specific TF are over/under-expressed, then such a TF can be considered as activated, and vice versa. With the quantified enrichment score, TFs can be prioritized. Specifically, we chose the VIPER algorithm for comparison study. VIPER has successfully prioritized TFs in various biological conditions, thus can be considered as the “ground truth” algorithm. For systematic comparison, we analyzed all the 33 MOCA lineages [2] using the VIPER algorithm [8,9]. We further compared the VIPER-inferred TF activities with temporal PageRank values, as described in Figure S2. The x-axis denotes the normalized enrichment score (nES, TF activity) quantified by VIPER, and the y-axis denotes the temporal PageRank values. The colored dots represent TFs during different developmental transitions. As shown, in general PageRank values agree with VIPER inferences, suggesting the recapitulation of key TFs. However, there were cases where discrepancies between the PageRank values and nES scores. VIPER evaluates TF activities by only visiting their direct transcriptional targets, thus cannot fully appreciate the hierarchy of gene regulatory networks. PageRank, on the other hand, performs a comprehensive survey on the network topology for systematic evaluation of the importance of TFs. For instance, during the E12.5-E13.5 development of mouse radial glia, *Whsc1* regulates *Tcf12* and *Pax3* (purple). As shown in panel “Radial-glia”, the moderate differential expression of *Tcf12* and *Pax3* gave less significant VIPER nES. In contrast, the PageRank value of *Tcf12* ranked #1 among all TFs, thus significantly promoted the importance of *Whsc1*. Such a dissection of the *Whsc1* sub-network provides a good example of the advantage of PageRank analysis.

**Table S1. [Result of the T-cell GO analysis], related to Figure 2.** Results of online GO analysis (<http://geneontology.org/>) with the top 20 identified TFs reported in Figure S6.

## Transparent Methods

### Overview of the analytical workflow.

As shown in Figure S8, cell type-specific TF-target networks will first be constructed from corresponding multi-omics profiles (see subsections “Gene expression-based GRNs” and “Epigenetic GRNs” for details). TF-target interactions will then be filtered based on confidence scores derived from the expression of TFs and corresponding targets. The confidence scores were calculated based on marginal and joint expression distribution, using either their differences or mutual information. Based on the null model calculated from random TF-target pairs, interactions with p-value smaller than  $10^{-3}$  were kept for downstream analysis. Such filtered GRNs can then be used to generate temporal GRNs, followed by temporal PageRank analysis (see section “temporal PageRank” for details). Such steady-state and temporal GRNs can be cleaned by removing certain nodes, further adjusted by user-provided edge weights, personalization vectors and damping factors. Finally, multiplex PageRank can be calculated among such filtered and adjusted GRNs (see section “multiplex PageRank” for details).

### Gene expression-based GRNs.

GRNs are conventionally reverse engineered from expression profiles [10]. Several well-established algorithms have been proposed to construct expression-based tissue-specific GRNs, as reviewed in [11]. Here, we specifically chose the ARACNe algorithm, considering the superior accuracy [12-14]. ARACNe was run on scRNA-Seq profiles with 200 bootstrap iterations using 1813 transcription factors (genes annotated in gene ontology molecular function database, as GO:0003700, “transcription factor activity”, or as GO:0003677, “DNA binding”, and GO:0030528, “transcription regulator activity”, or as GO:00034677 and GO: 0045449, “regulation of transcription”). Parameters were set to zero DPI (Data Processing Inequality) tolerance and MI (Mutual Information) p-value (using MI computed by permuting the original dataset as null model) threshold of  $10^{-8}$ . The ARACNe algorithm is available at <https://github.com/califano-lab/ARACNe-AP>.

### Epigenetic GRNs.

In this study we focused on chromatin accessibility and chromosome conformation epigenetic assays. Chromatin accessibility assays, e.g. DNase-Seq [15,16], MNase-Seq [17], NOMe-Seq [18], ATAC-Seq [19] and derivatives are designed to identify physically accessible DNA regions, which are likely to be involved in transcriptional regulation. As shown in Figure S9, motif searching will be performed within called peaks (accessible regions) to identify putative TFs. Meanwhile, TSSs (Transcription Start Sets) will be searched within certain upstream and downstream genomic ranges of peaks to identify potential target genes. Thus, TF-target can be corresponded for GRNs assembly.

Chromosome conformation assays, e.g. ChIA-PET [20], 3C [21], 4C [22], 5C [23], Hi-C [24], HiChIP [6] and derivatives are designed to map spatial organization of chromosomes. Such spatial organization usually results in transcriptional regulation. As shown in Figure S9, chemical crosslink will preserve long-range interactions of genomic regions. Within such regions, TF binding motifs and TSSs will be searched for GRNs assembly, as described before.

Specifically, for TSS searching, genes will first be retrieved from Bioconductor R TxDb annotations [25] using the function `genes` in Bioconductor R package `GenomicFeatures` [26]. Then the corresponding promoter regions will be determined using the function `promoters` in Bioconductor R package `GenomicRanges` [26]. For motif searching, the most updated JASPAR database (available in Bioconductor R package `JASPAR2018`) [27] will be searched using the function `getMatrixSet` in Bioconductor R package `TFBSTools` [28]. Then functions `matchMotifs` and `motifMatches` in Bioconductor R package `motifmatchr` [29] will be used to match motifs to TFs.

### Temporal PageRank.

Temporal PageRank extends the original static PageRank by only considering temporal edges instead of all edges. Temporal edges denote edges that change at certain stages during the lifetime of the networks [30]. Here, we simplify the definition of temporal PageRank by examining GRN pairs describing consecutive biological states, e.g. T0 and T24 of myoblast-muscle differentiation. Thus, we represented such GRN pairs using adjacency matrices, then took the subtraction as representation of temporal differential networks, further for temporal PageRank calculation. We implemented temporal PageRank as the function `diff_graph` in R package `pageRank`, based on CRAN R package `igraph` [31]. As shown in Figure S10, the complexity of our temporal PageRank implementation scales exponentially with the network size.

### Multiplex PageRank.

Multiplex PageRank extends PageRank analysis to multiplex networks, in which the same nodes might connect differently across layers. Multiplex PageRank integrates network information from supplementary layers to base layer, as defined by

$$X_{Bi} = \alpha_B \sum_j X_{Aj}^\beta B_{ij} \frac{X_{Bj}}{G_j} + (1 - \alpha_B) \frac{X_{Ai}^\gamma}{N \langle X_A^\gamma \rangle}$$

, where

$$G_j = \sum_r B_{rj} X_{Ar}^\beta + \delta(0, \sum_r B_{rj} X_{Ar}^\beta)$$

with  $\delta$  as the Kronecker delta and  $\langle \dots \rangle$  as mean value.

In the equation,  $A$  and  $B$  denote the adjacency matrices of supplementary and base layers. For such adjacency matrices, the first/second subscripts represent the row/column indices.  $X_{Ai}$  represent the PageRank value of  $i$ th node calculated from  $B$ . The same representation rule applies for  $X_{Bj}$  and  $X_{Ar}$ .  $N$  is the total number of nodes in the base layer.  $\alpha_B$  is the damping factor used for.  $\beta$  and  $\gamma$  are user-defined parameters. As shown in the equation, multiplex PageRank is calculated according to the topology of the base layer, with regular PageRank of supplemental layer as outgoing weights and personalization vector. For networks with multiple supplementary layers, we can simply multiply their PageRank to fit into the above equation. Based on parameters  $\beta$  and  $\gamma$ , four special multiplex PageRank forms were defined, including additive ( $\beta = 0, \gamma = 1$ ), multiplicative ( $\beta = 1, \gamma = 0$ ), combined ( $\beta = 1, \gamma = 1$ ) and neutral ( $\beta = 0, \gamma = 0$ ) [32]. We implemented the four special forms of multiplex PageRank as the function `multiplex_page_rank` in R package `pageRank`, based on CRAN R package `igraph` [31]. As shown in Figure S10, the complexity of our multiplex PageRank implementation scales exponentially with the network size.

We assume that regulatory machinery can be directly reflected on expression profiles, thus gene expression-based GRNs were used as base networks in this study. In contrast, epigenetic profiles are usually considered as indirect evidence of TF-target regulation. For instance ATAC-Seq and HiChIP GRNs were reverse engineered through motif-matching and TSS-searching, thus significant amounts of false positives should be expected. Also, TFs might not necessarily bind to the regions to be searched, e.g. ATAC-Seq peaks. In the case of ATAC-Seq, TF might bind to enhancer regions, which are usually distant from the corresponding promoters, thus cannot be captured. Therefore, considerable amounts of false negatives should also be expected.

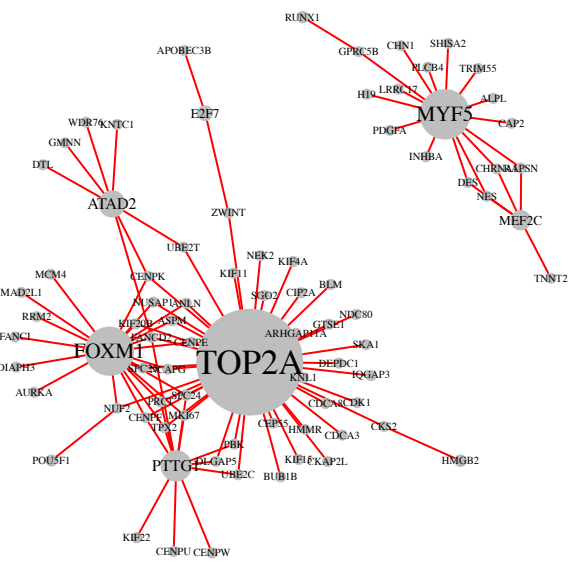
## Supplemental References

1. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S. and Rinn, J.L., 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology*, 32(4), p.381.
2. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J. and Trapnell, C., 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 566(7745), pp.496-502.
3. Pliner, H.A., Packer, J.S., McFaline-Figueroa, J.L., Cusanovich, D.A., Daza, R.M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A. and Adey, A.C., 2018. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Molecular cell*, 71(5), pp.858-871.
4. Pellin, D., Loperfido, M., Baricordi, C., Wolock, S.L., Montepeloso, A., Weinberg, O.K., Biffi, A., Klein, A.M. and Biasco, L., 2019. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nature communications*, 10(1), pp.1-15.
5. Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J. and Majeti, R., 2016. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature genetics*, 48(10), pp.1193-1203.
6. Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R. and Wei, Y., 2017. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nature genetics*, 49(11), p.1602.
7. Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H. and Califano, A., 2016. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nature genetics*, 48(8), pp.838-847.
8. Ding, H., Douglass, E.F., Sonabend, A.M., Mela, A., Bose, S., Gonzalez, C., Canoll, P.D., Sims, P.A., Alvarez, M.J. and Califano, A., 2018. Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nature communications*, 9(1), pp.1-10.
9. Aibar, S., González-Blas, C.B., Moerman, T., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., van den Oord, J. and Atak, Z.K., 2017. SCENIC: single-cell regulatory network inference and clustering. *Nature methods*, 14(11), pp.1083-1086.
10. Emmert-Streib, F., Glazko, G. and De Matos Simoes, R., 2012. Statistical inference and reverse engineering of gene regulatory networks from observational expression data. *Frontiers in genetics*, 3, p.8.
11. Mercatelli, D., Scalambra, L., Triboli, L., Ray, F. and Giorgi, F.M., 2020. Gene regulatory network inference resources: A practical overview. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1863(6), p.194430.
12. Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R. and Califano, A., 2005. Reverse engineering of regulatory networks in human B cells. *Nature genetics*, 37(4), pp.382-390.
13. Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R. and Califano, A., 2006, March. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. In *BMC bioinformatics* (Vol. 7, No. S1, p. S7). BioMed Central.
14. Lachmann, A., Giorgi, F.M., Lopez, G. and Califano, A., 2016. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics*, 32(14), pp.2233-2235.
15. Crawford, G.E., Davis, S., Scacheri, P.C., Renaud, G., Halawi, M.J., Erdos, M.R., Green, R., Meltzer, P.S., Wolfsberg, T.G. and Collins, F.S., 2006. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nature methods*, 3(7), pp.503-509.
16. Sabo, P.J., Kuehn, M.S., Thurman, R., Johnson, B.E., Johnson, E.M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A. and Weaver, M., 2006. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nature methods*, 3(7), pp.511-518.

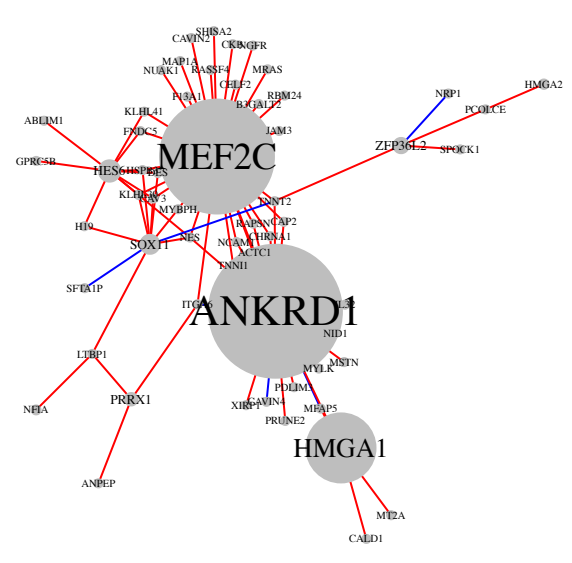
17. Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G. and Zhao, K., 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell*, 132(5), pp.887-898.
18. Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L. and Schübeler, D., 2017. Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. *Molecular cell*, 67(3), pp.411-422.
19. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. and Greenleaf, W.J., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods*, 10(12), p.1213.
20. Fullwood, M.J. and Ruan, Y., 2009. ChIP-based methods for the identification of long-range chromatin interactions. *Journal of cellular biochemistry*, 107(1), pp.30-39.
21. Dekker, J., Rippe, K., Dekker, M. and Kleckner, N., 2002. Capturing chromosome conformation. *science*, 295(5558), pp.1306-1311.
22. Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., De Wit, E., Van Steensel, B. and De Laat, W., 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture—on-chip (4C). *Nature genetics*, 38(11), pp.1348-1354.
23. Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C. and Green, R.D., 2006. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome research*, 16(10), pp.1299-1309.
24. Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. and Sandstrom, R., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326(5950), pp.289-293.
25. Carlson, M.R., Pagès, H., Arora, S., Obenchain, V. and Morgan, M., 2016. Genomic annotation resources in R/Bioconductor. In *Statistical Genomics* (pp. 67-90). Humana Press, New York, NY.
26. Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T. and Carey, V.J., 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol*, 9(8), p.e1003118.
27. Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., Van Der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G. and Baranasic, D., 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic acids research*, 46(D1), pp.D260-D266.
28. Tan, G. and Lenhard, B., 2016. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics*, 32(10), pp.1555-1556.
29. Schep, A.N., Wu, B., Buenrostro, J.D. and Greenleaf, W.J., 2017. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nature methods*, 14(10), pp.975-978.
30. Rozenshtein, P. and Gionis, A., 2016, September. Temporal pagerank. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 674-689). Springer, Cham.
31. Csardi, G. and Nepusz, T., 2006. The igraph software package for complex network research. *InterJournal, complex systems*, 1695(5), pp.1-9.
32. Halu, A., Mondragón, R.J., Panzarasa, P. and Bianconi, G., 2013. Multiplex pagerank. *PLoS one*, 8(10), p.e78293.



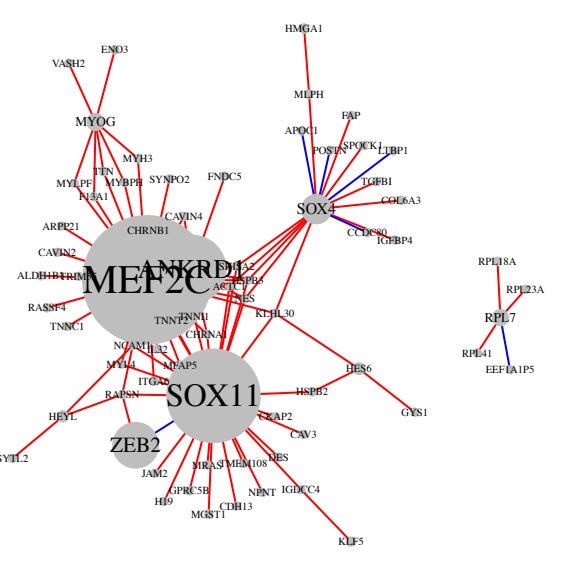
scRNA-Seq T0



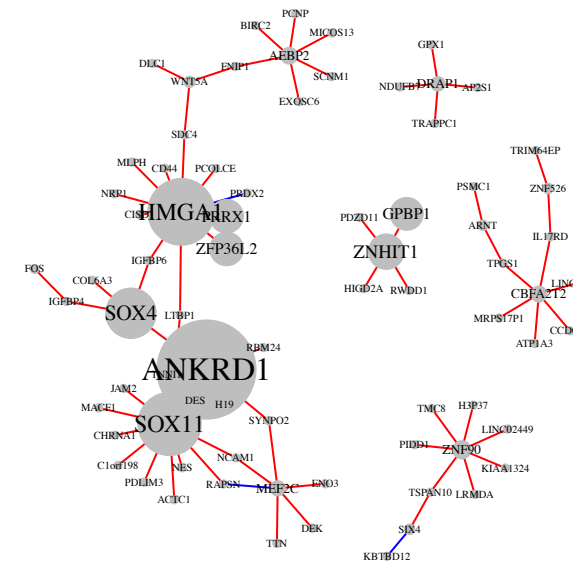
scRNA-Seq T24



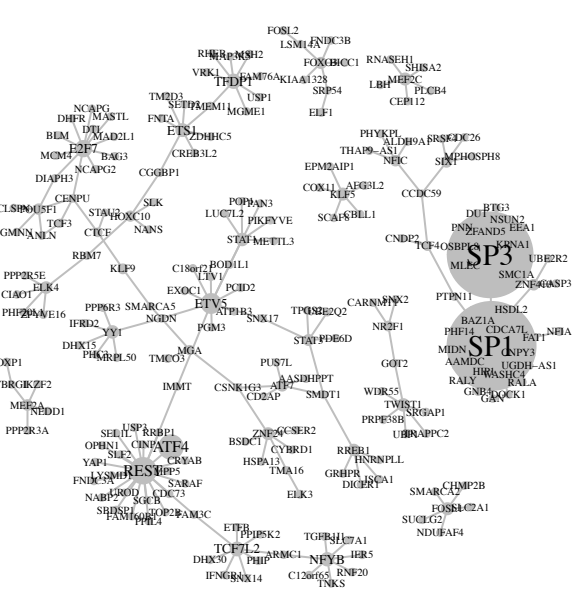
scRNA-Seq T48



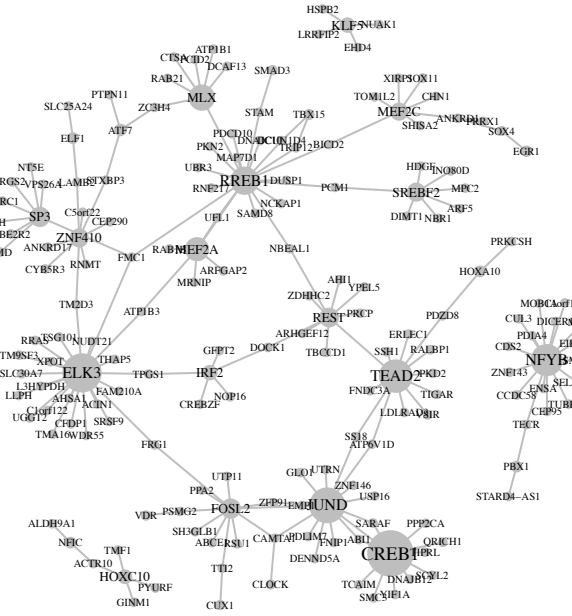
scRNA-Seq T72



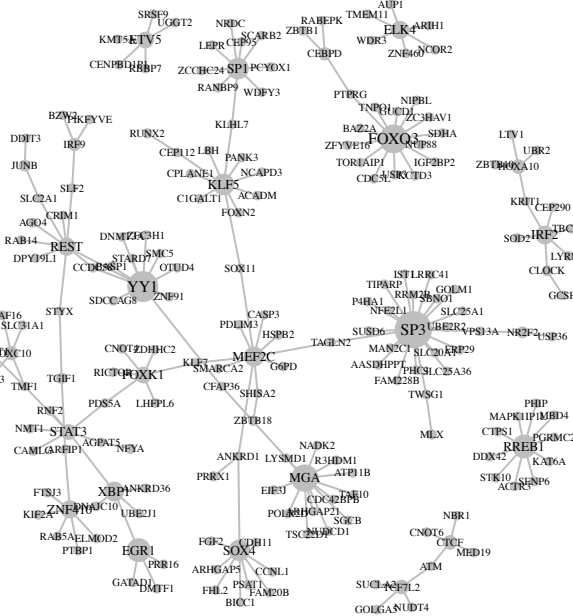
ATAC-Seq T0



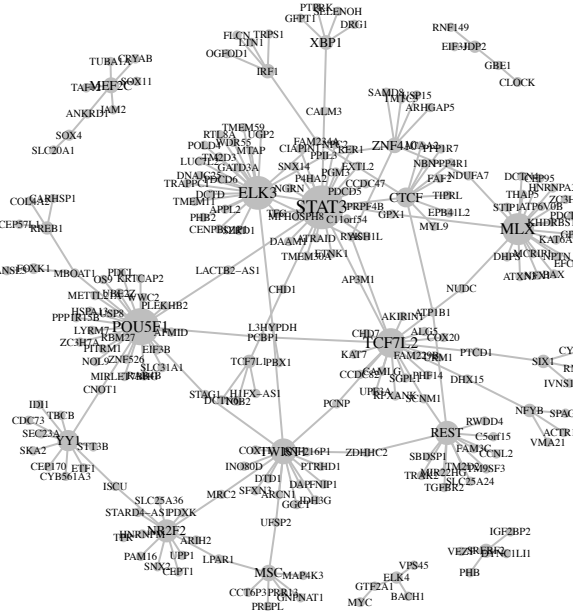
ATAC-Seq T24



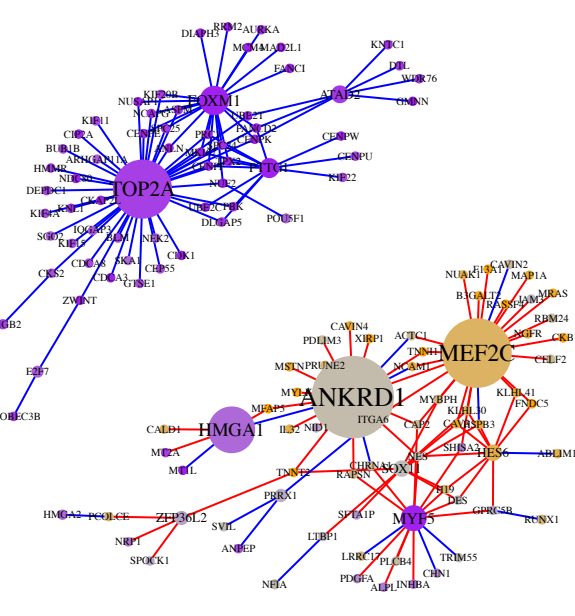
ATAC-Seq T48



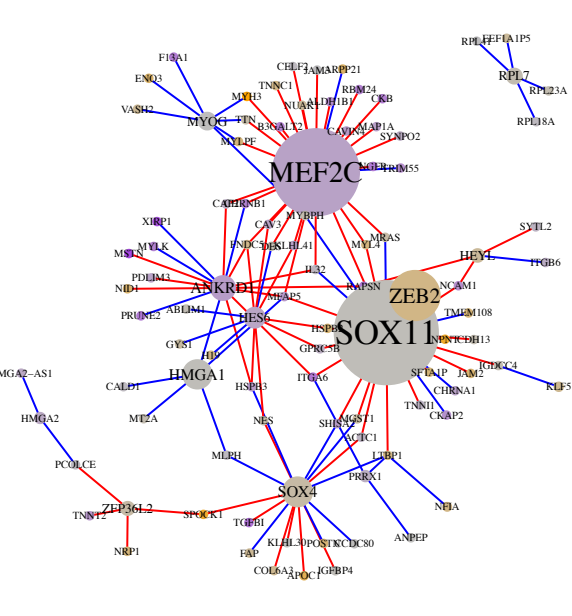
ATAC-Seq T72



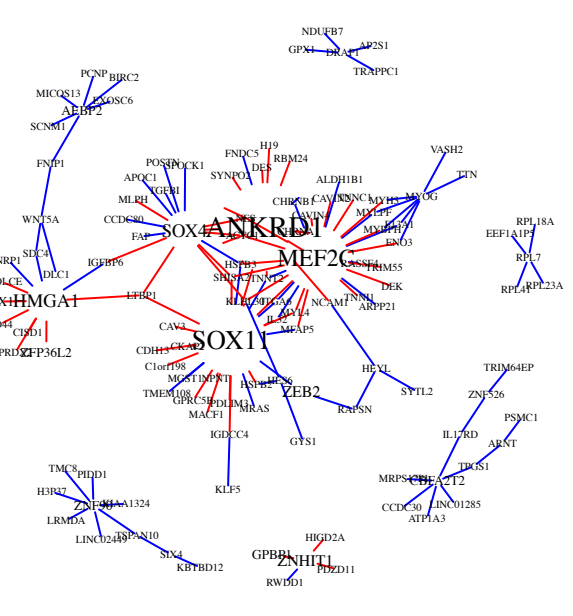
scRNA-Seq T0-&gt;T24



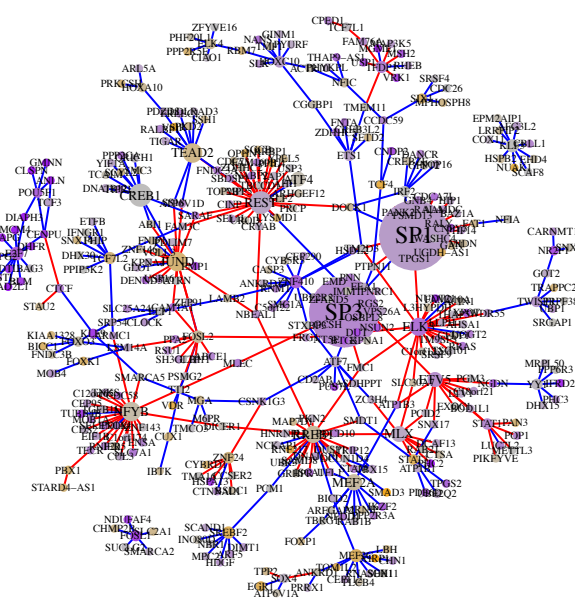
scRNA-Seq T24-&gt;T48



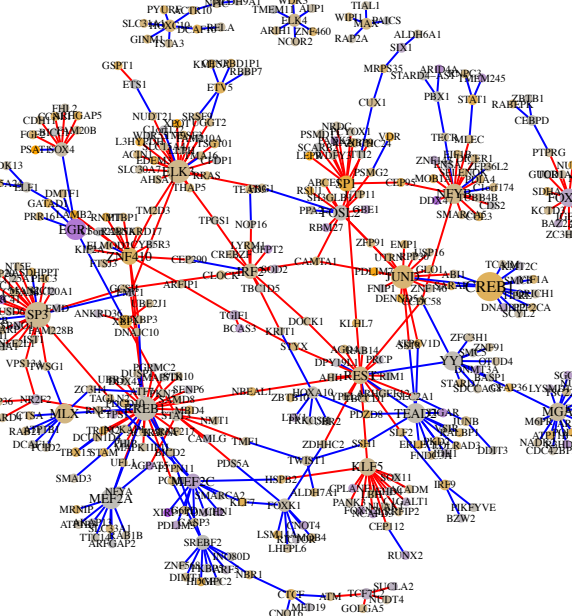
scRNA-Seq T48-&gt;T72



ATAC-Seq T0-&gt;T24



ATAC-Seq T24-&gt;T48



ATAC-Seq T48-&gt;T72

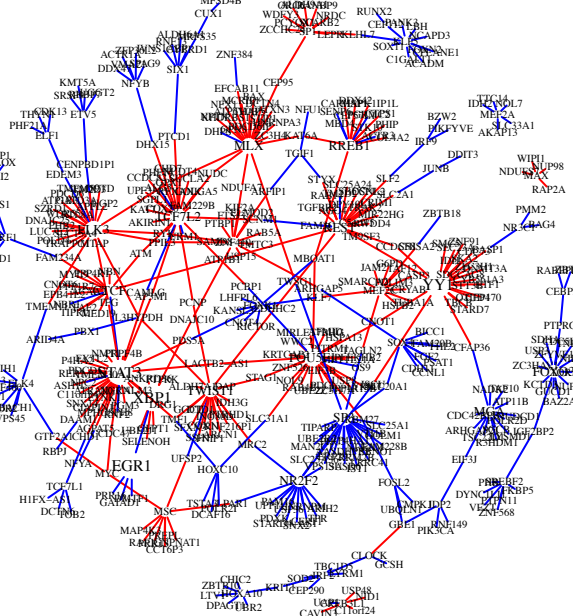


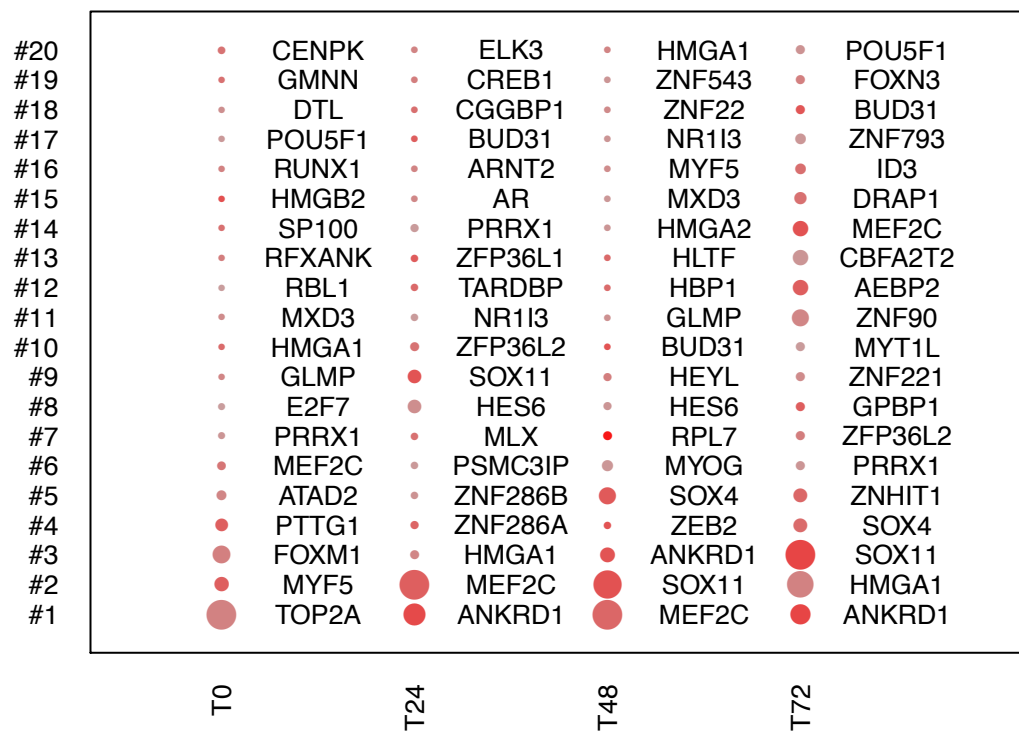




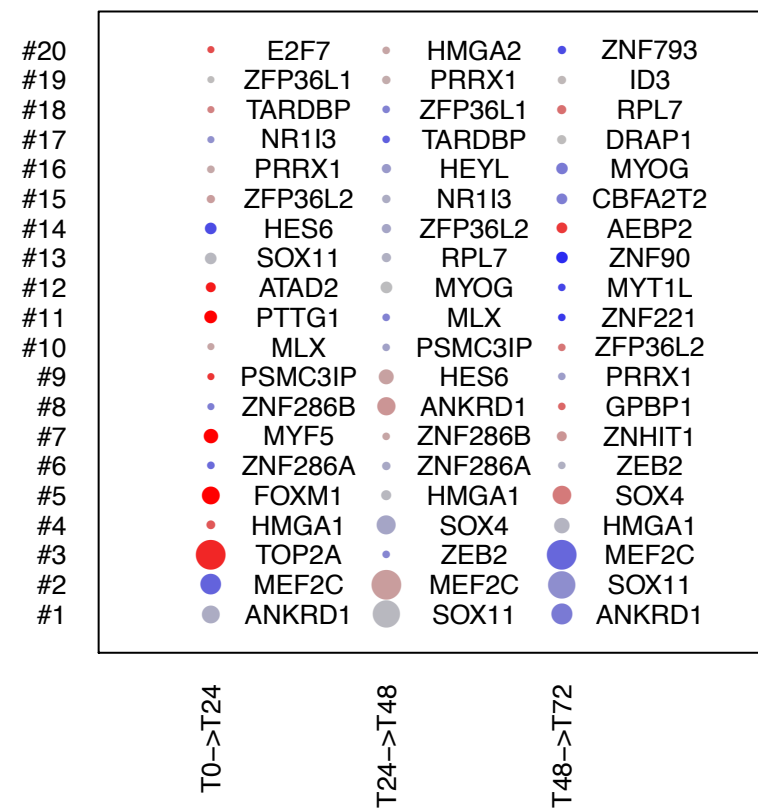
Figure S2



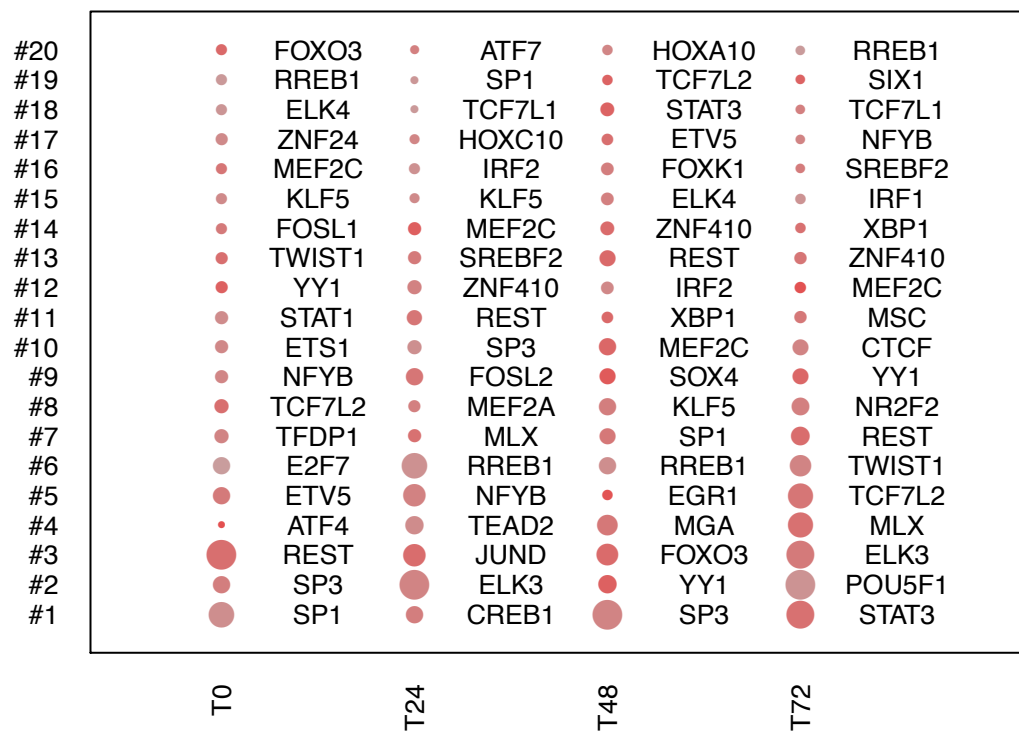
## scRNA-Seq PR



## scRNA-Seq tPR



## ATAC-Seq PR



## ATAC-Seq tPR

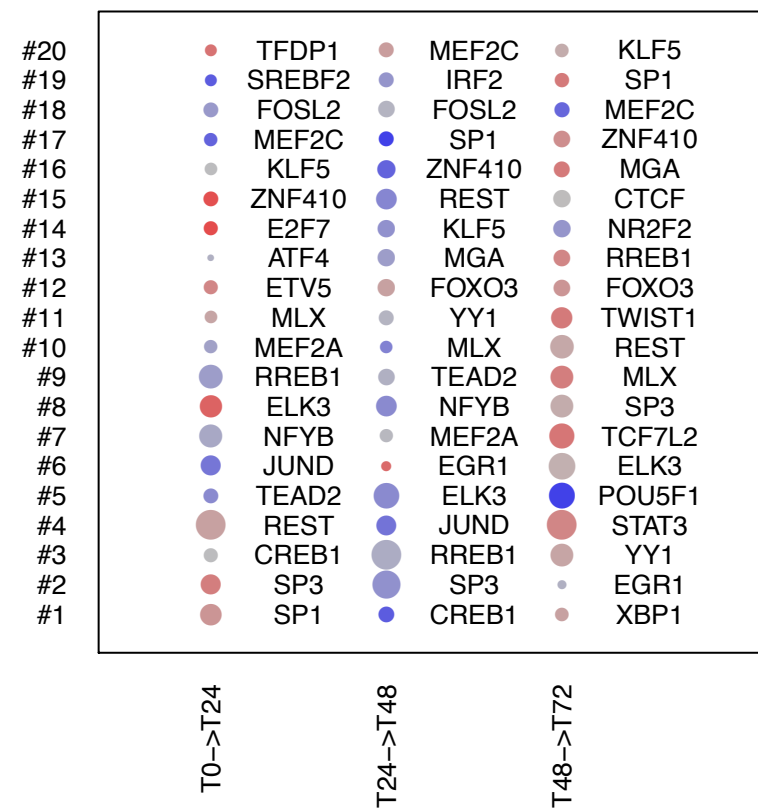
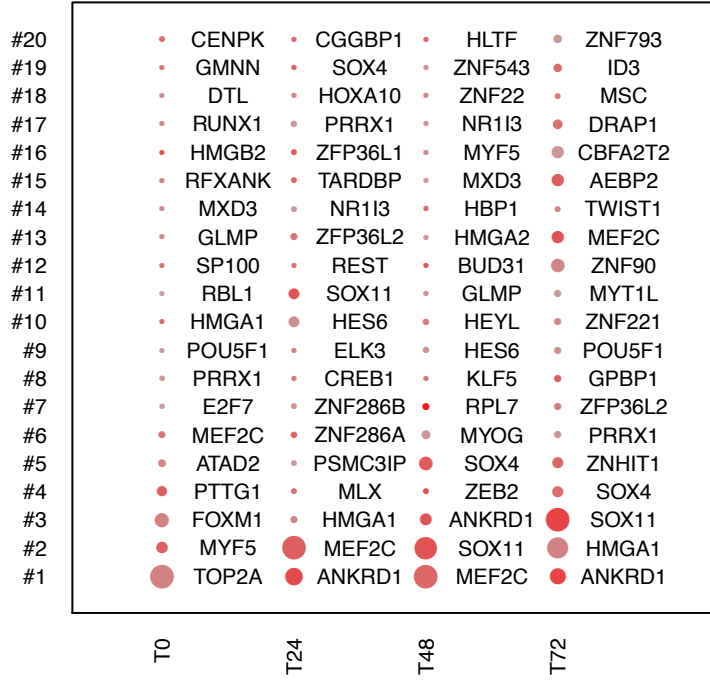
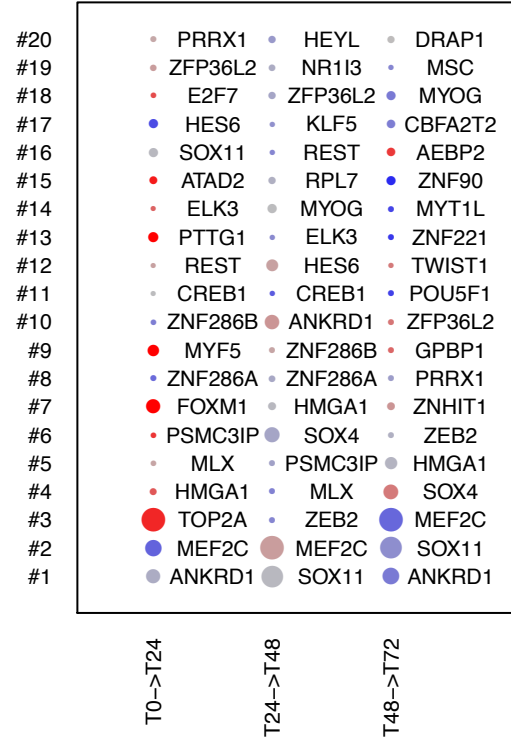


Figure S3

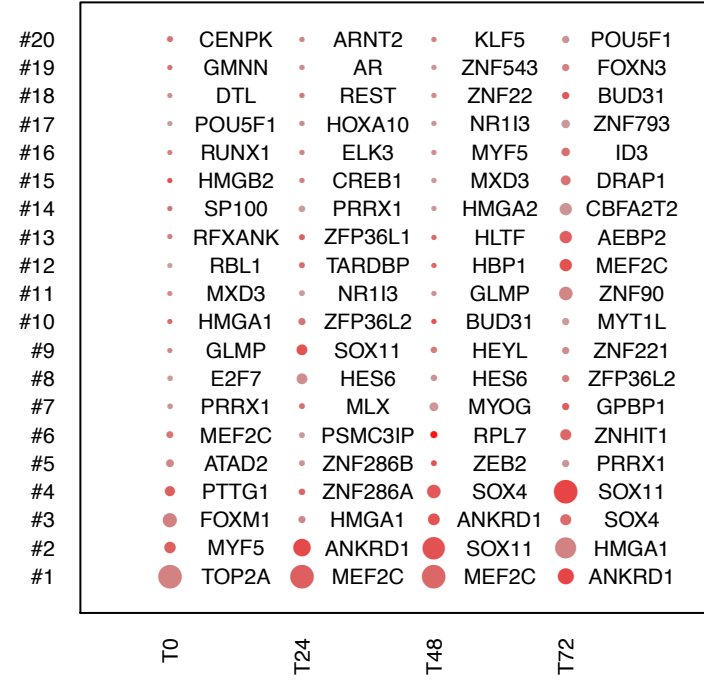
### additive PR



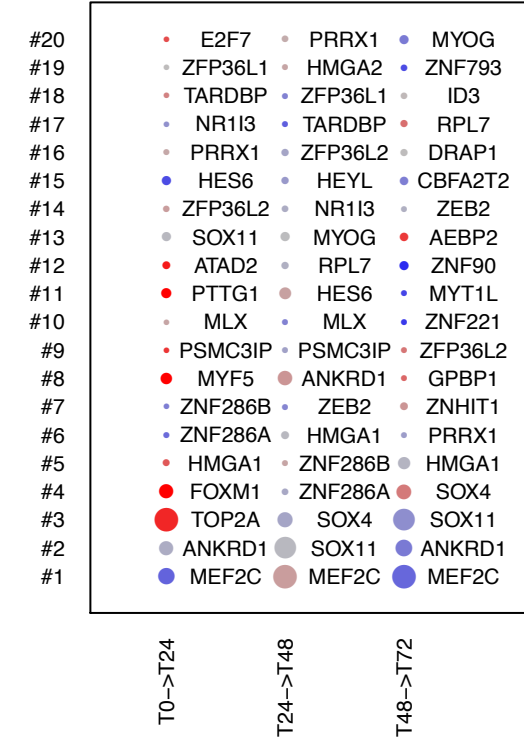
### additive tPR



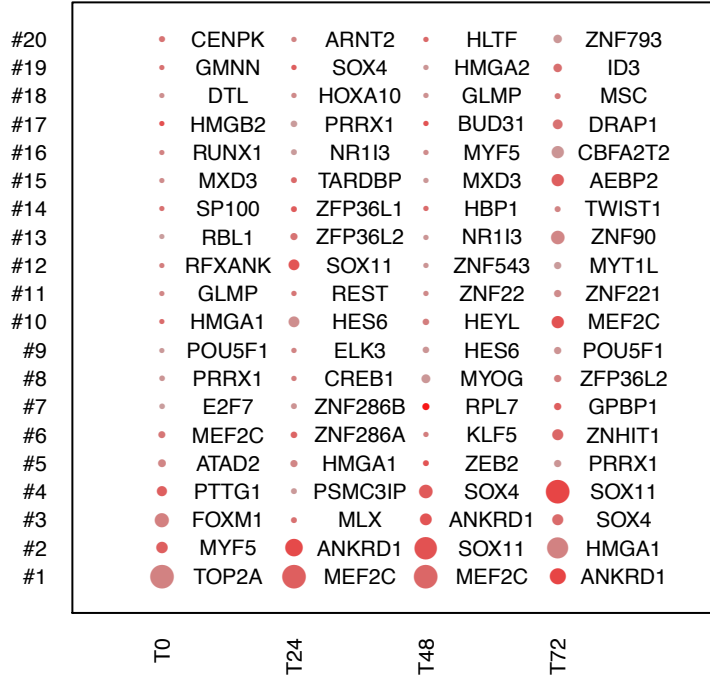
### multiplicative PR



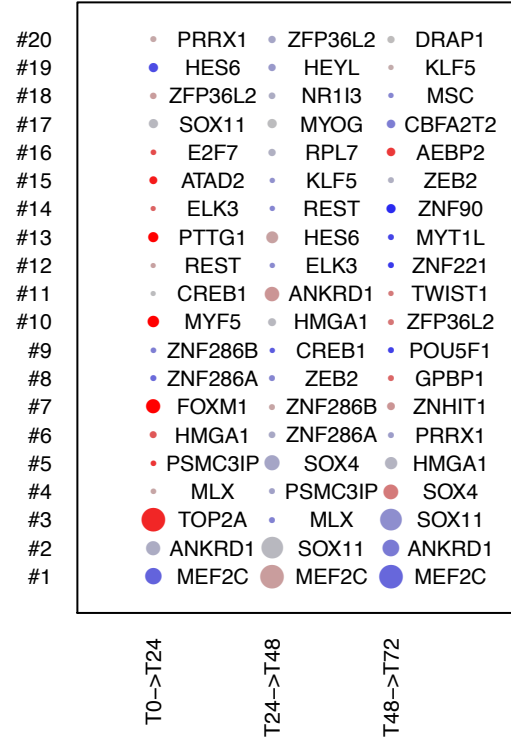
### multiplicative tPR



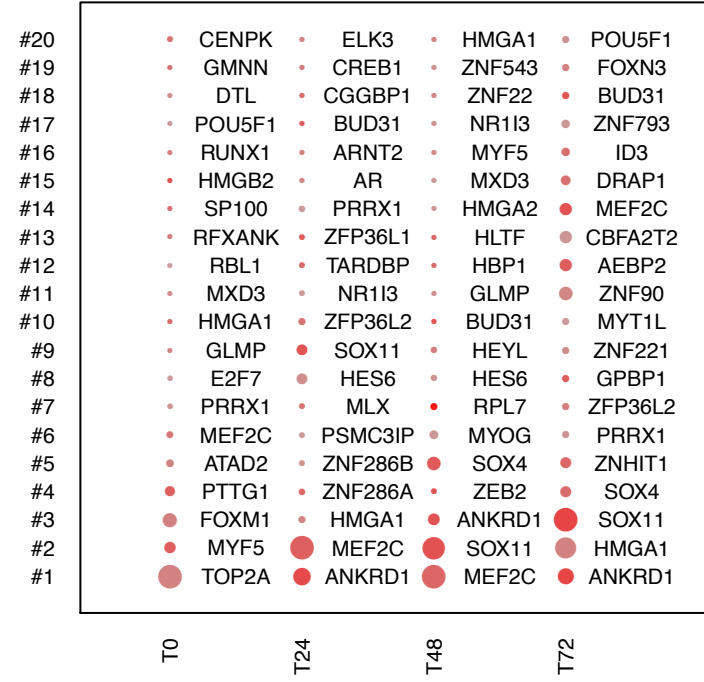
### combined PR



### combined tPR



### neutral PR



### neutral tPR

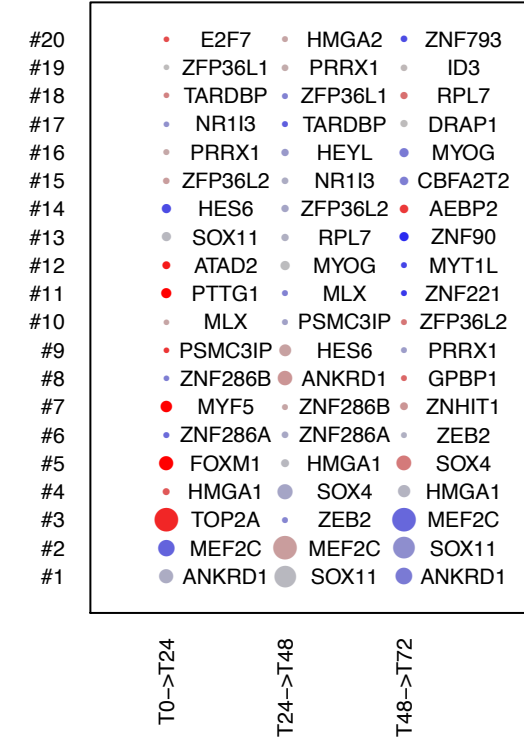
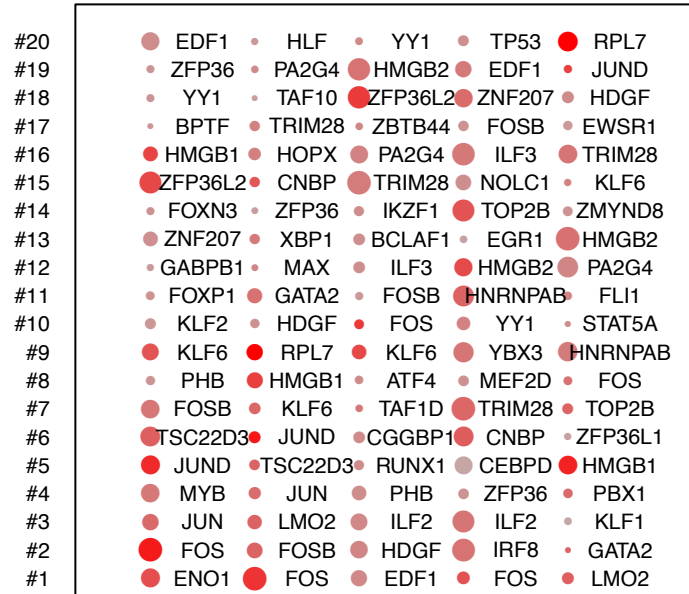


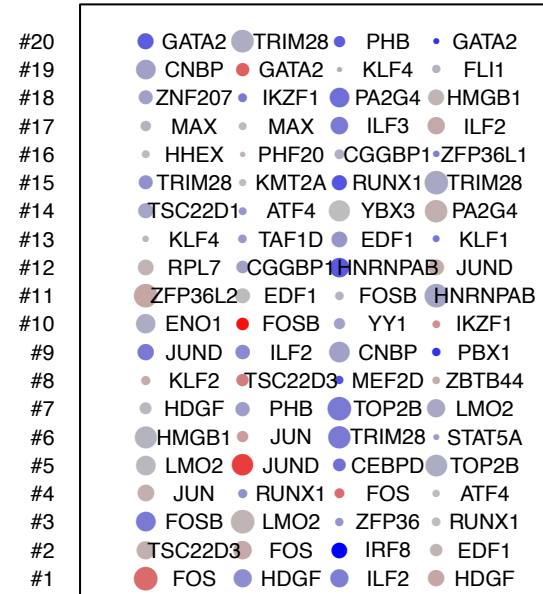
Figure S4

### additive PR



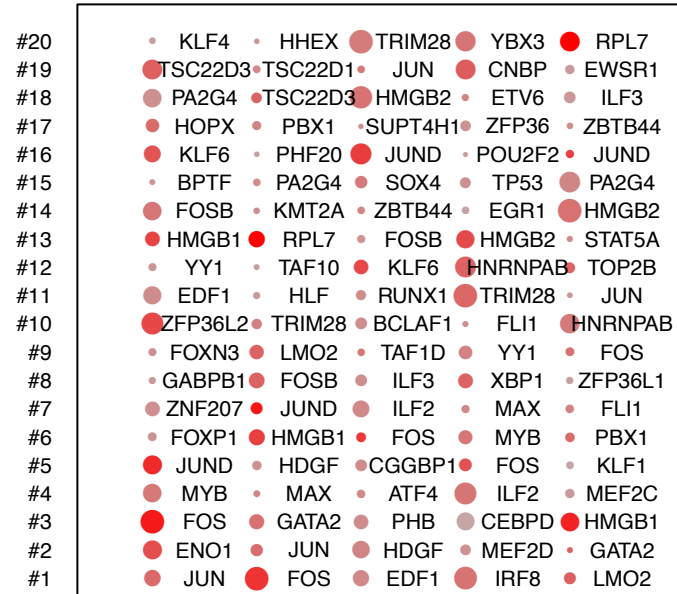
HSC MPP CMP GMP MEP

### additive tPR



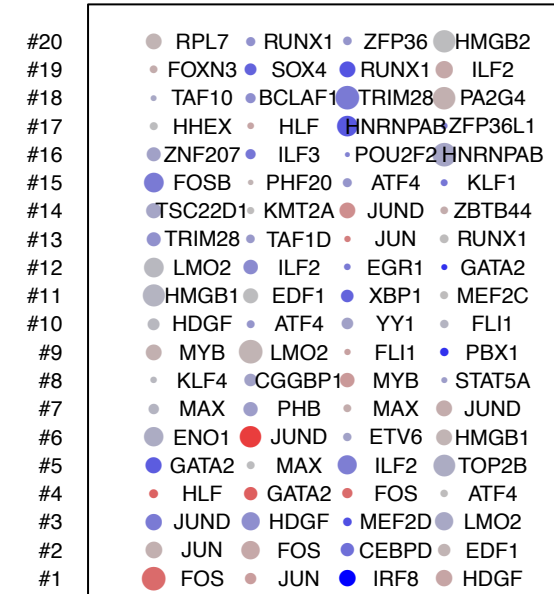
HSC->MPP MPP->CMP CMP->GMP CMP->MEP

### multiplicative PR



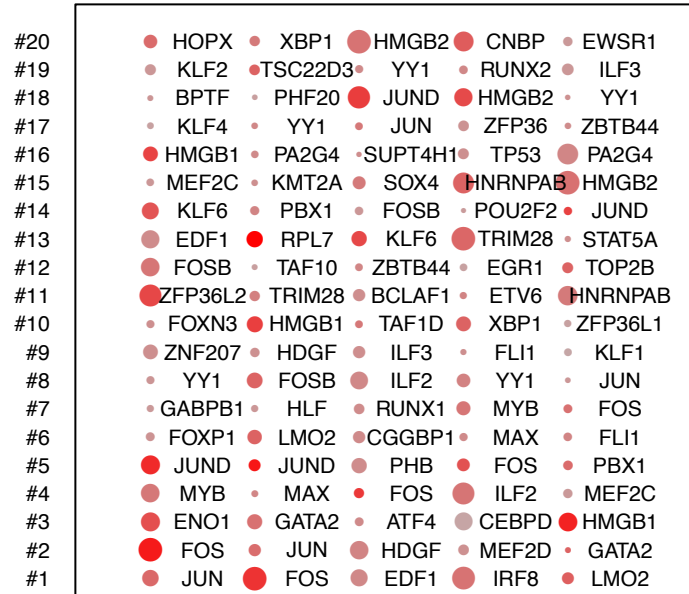
HSC MPP CMP GMP MEP

### multiplicative tPR



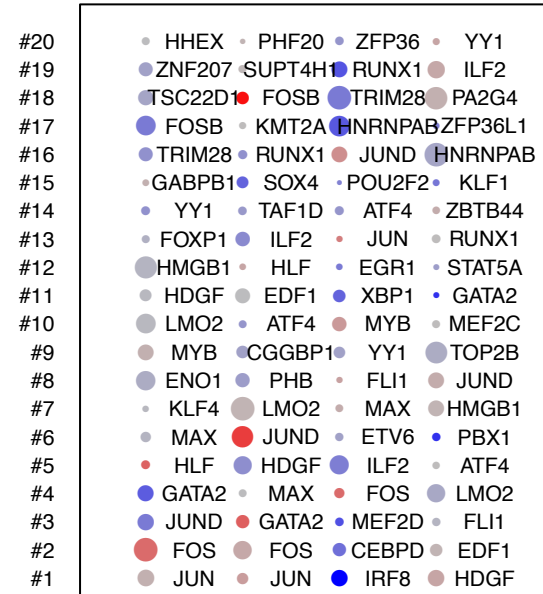
HSC->MPP MPP->CMP CMP->GMP CMP->MEP

### combined PR



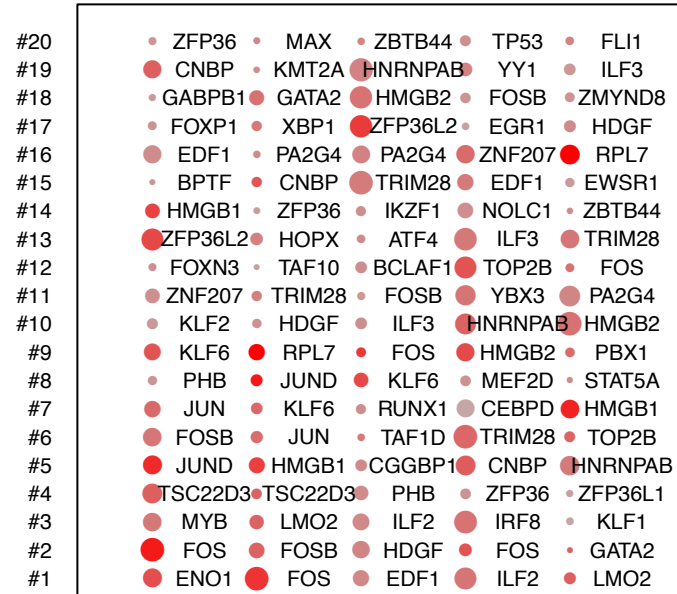
HSC MPP CMP GMP MEP

### combined tPR



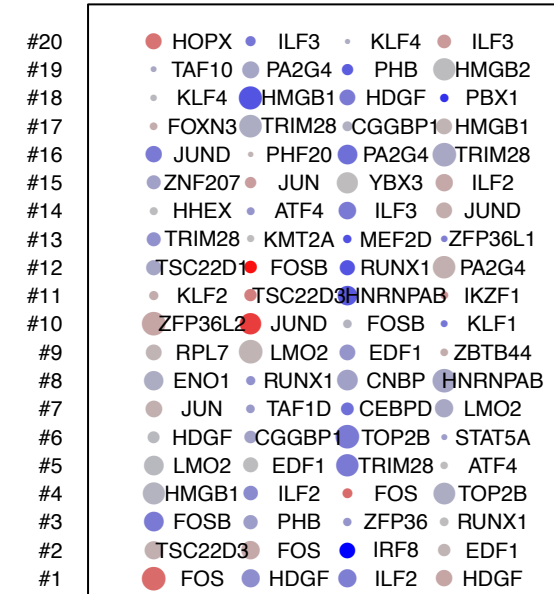
HSC->MPP MPP->CMP CMP->GMP CMP->MEP

### neutral PR



HSC MPP CMP GMP MEP

### neutral tPR



HSC->MPP MPP->CMP CMP->GMP CMP->MEP

Figure S5

# Combined PR

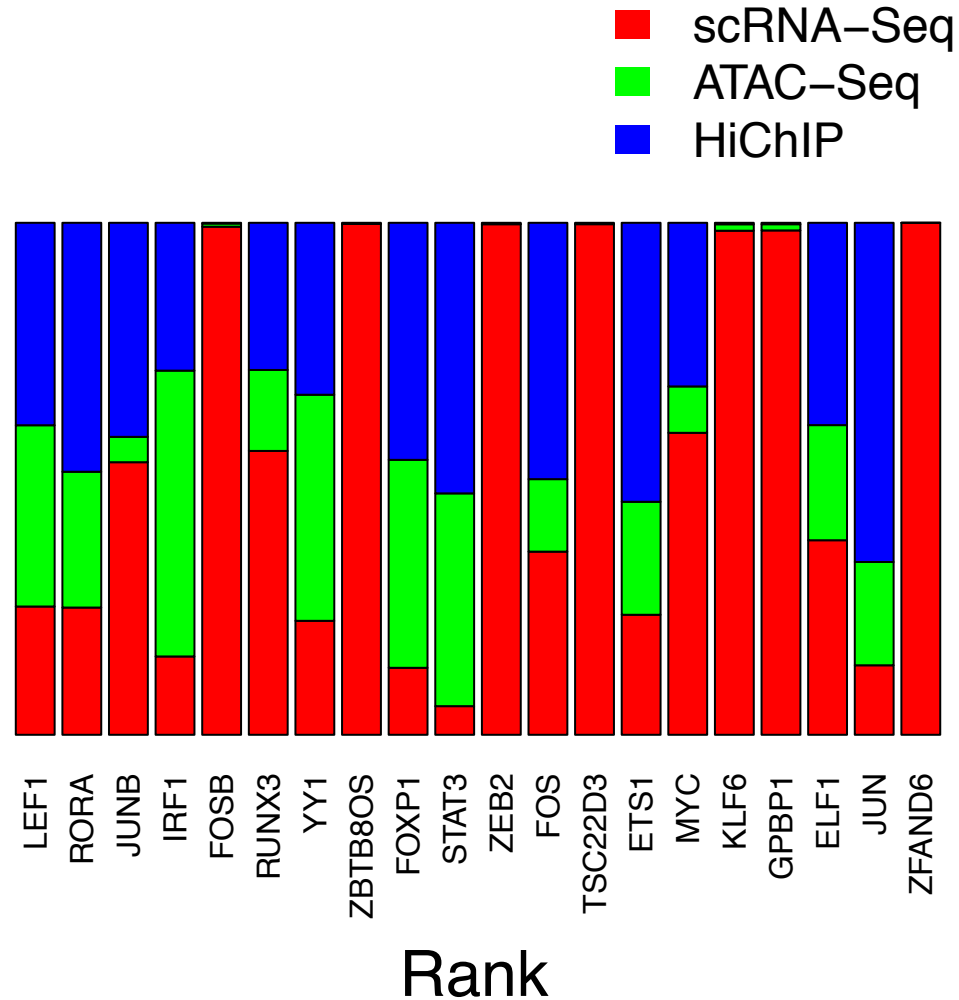
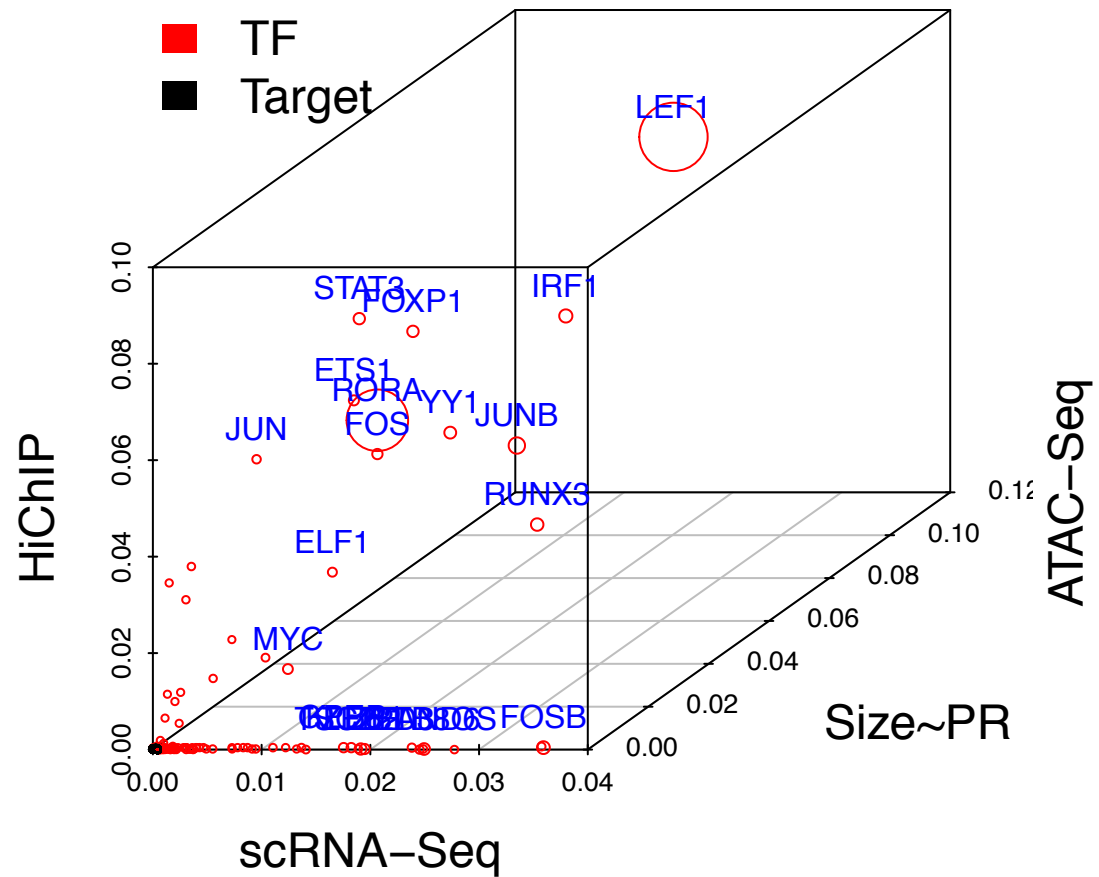
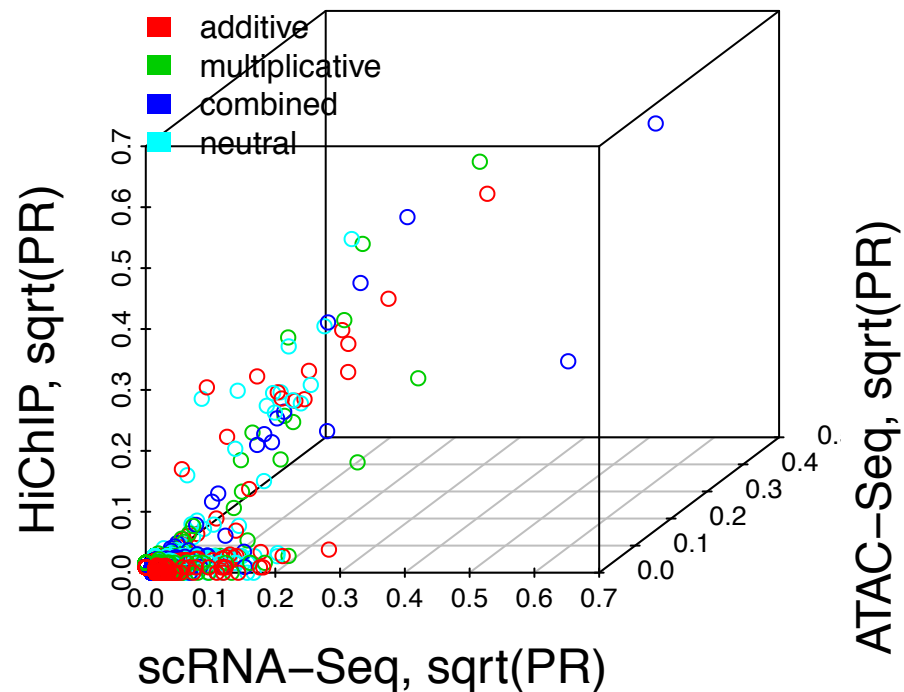


Figure S6

# A Base Network



# B PageRank Method

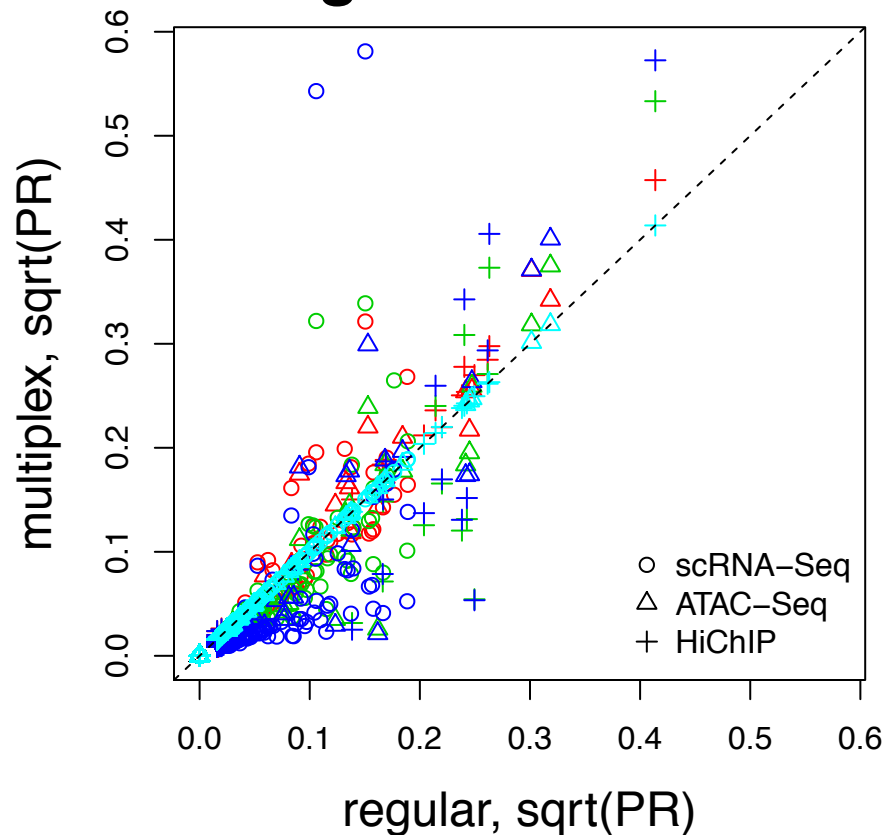
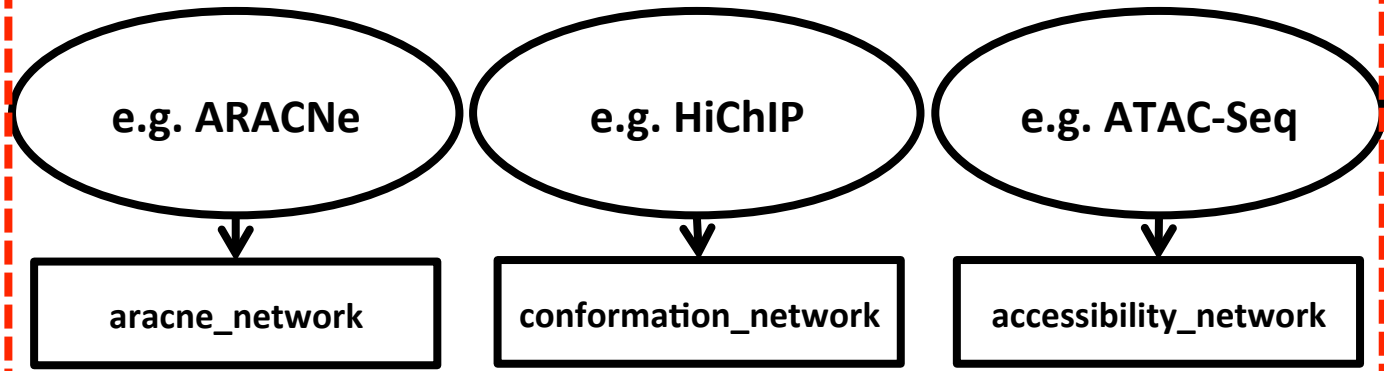
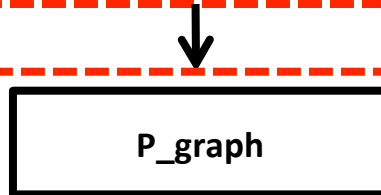
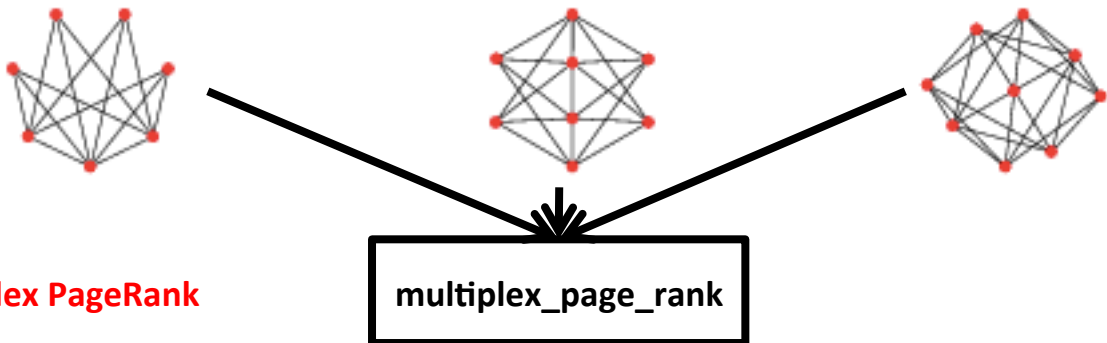


Figure S7



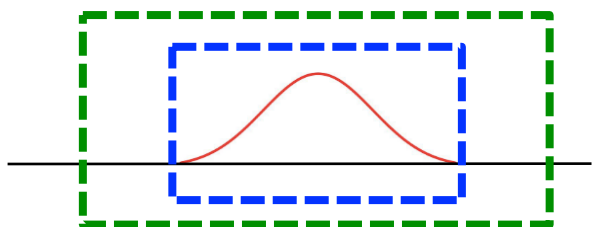
**GRN Construction****GRN Filtering (Optional)****GRN Processing (Optional)****Multiplex PageRank**

- `aracne_network`: re-format ARACNe regulon R object.
- `accessibility_network`: generate genome accessibility-based GRNs from called peaks.
- `conformation_network`: generate chromosome conformation-based GRNs from records.
- `P_graph`: filter GRNs by probability-based approaches, with interaction confidence and mode as edge attributes, and regular PageRank as vertex attribute.
- `clean_graph`: remove vertex by corresponding sub-graph size, name and PageRank, with updated regular PageRank as vertex attribute.
- `adjust_graph`: re-calculate PageRank with weight, personalization and damping.
- `diff_graph`: calculate differential graph, with mode of changing as edge attribute, and regular PageRank as vertex attribute.
- `multiplex_page_rank`: calculate multiplex PageRank from different networks.

# accessibility\_network



Motif-Searching (TF)

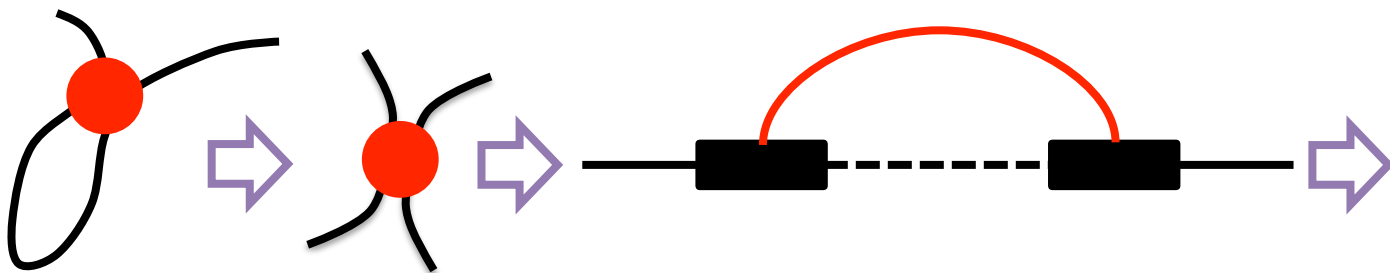


TSS-Searching (Target)

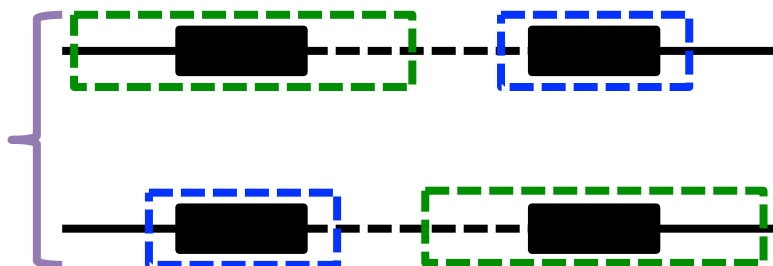


TF-Target Network

# conformation\_network



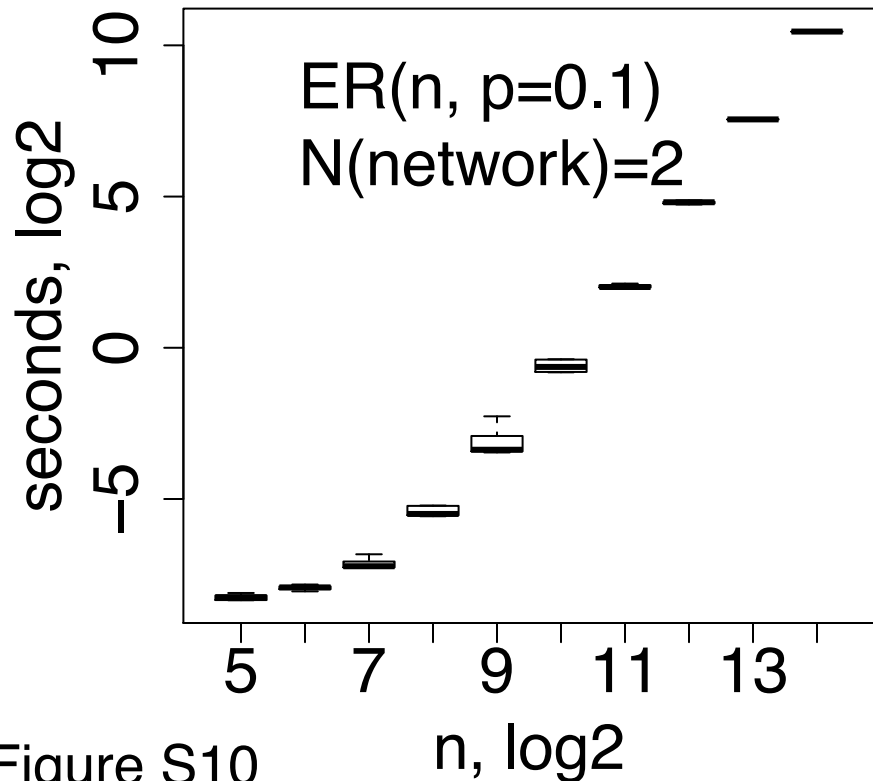
TSS-Searching (Target) Motif-Searching (TF)



TF-Target Network

Figure S9

## multiplex PageRank



## temporal PageRank

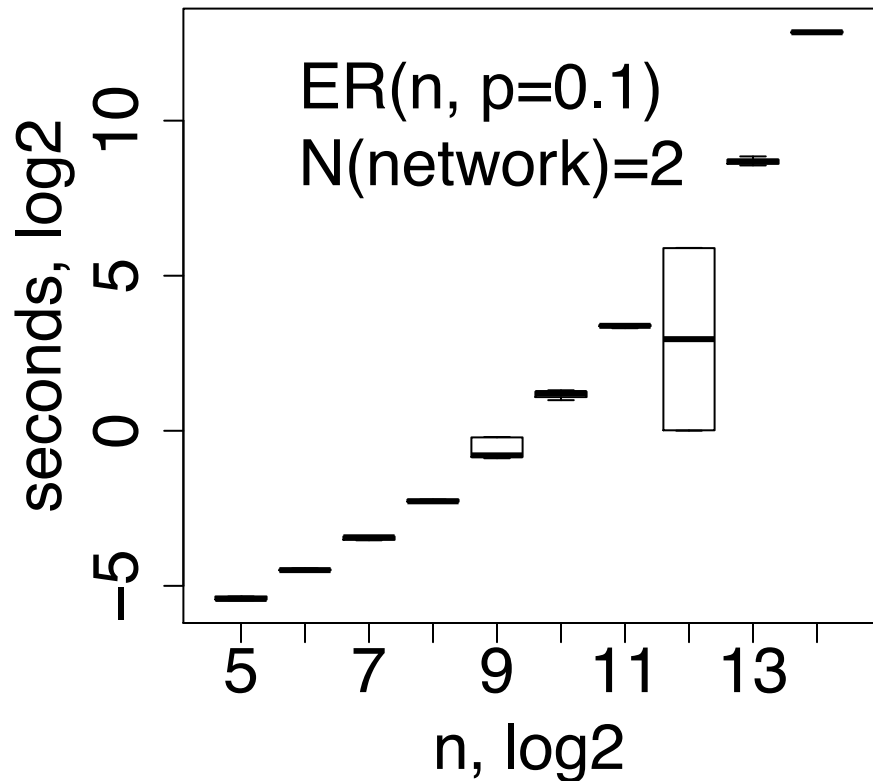


Figure S10



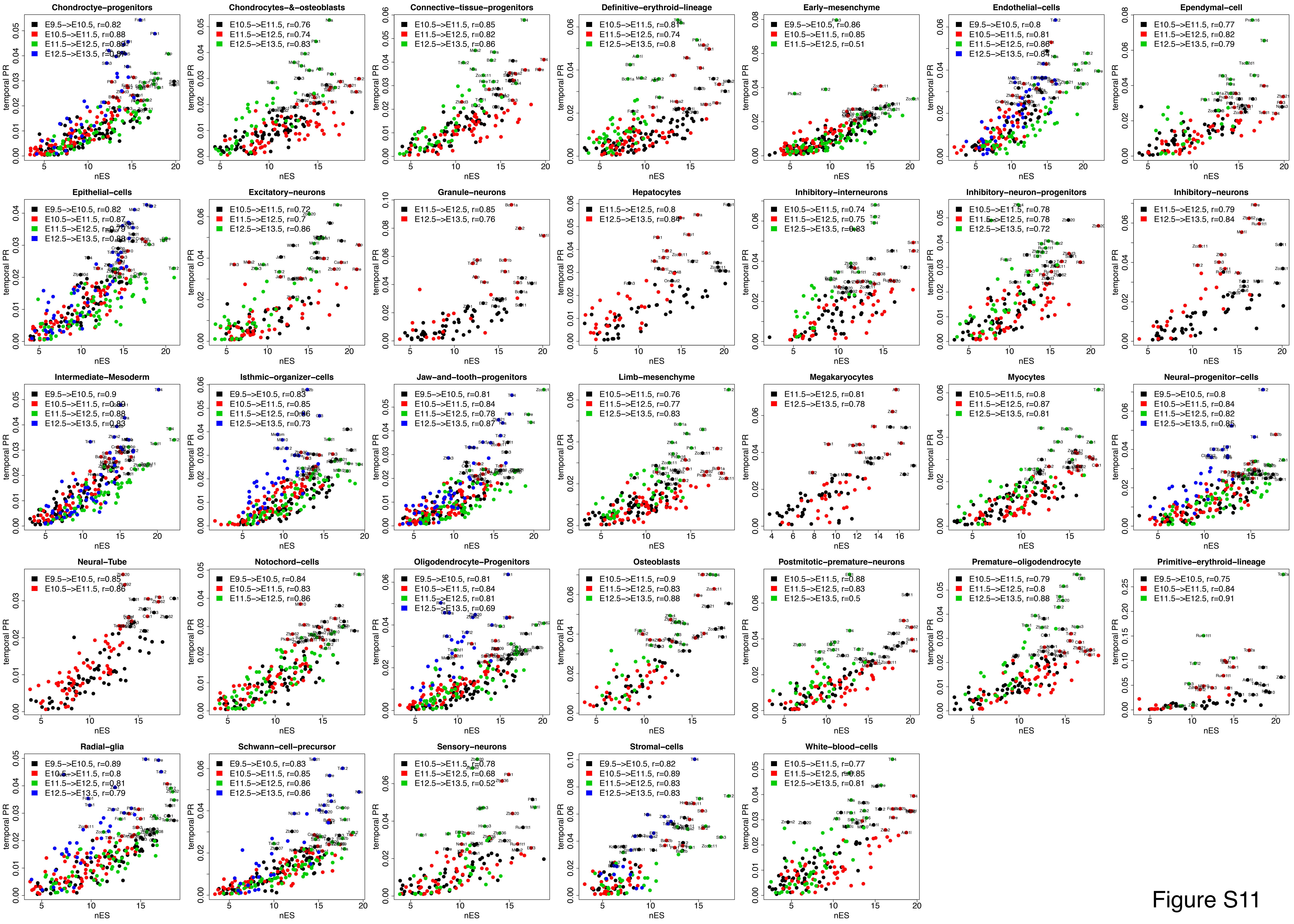


Figure S11

Table S1

Analysis Type: **PMMER Overrepresentation Test (Released 20200628)**  
 Annotation Version and Release Date: **GO Ontology database 002: 10.5281/zenodo.4033854 Released 2020-09-10**  
 Analyzed List: **upload\_1 (Homo sapiens)**  
 Reference List: **Homo sapiens (all genes in database)**  
 Test Type: **FISHER**  
 Correction: **None**

GO biological process complete	Homo sapiens - REFLIST (20051)	upload_1 (20)	upload_1 (expected)	upload_1 (over/under)	upload_1 (fold enrichment)	upload_1 (raw P-value)	upload_1 (FDR)
T-helper 17 cell differentiation (GO:0072538)	2	.01	> 100	4.77E-05	5.86E-03		
T-helper 17 type immune response (GO:0072388)	10	.01	> 100	5.72E-05	6.86E-03		
radial glial cell differentiation (GO:0060019)	13	.01	> 100	9.05E-05	6.76E-03		
regulation of monoacylglycerol biosynthesis (GO:0045655)	20	.02	> 100	2.12E-04	2.41E-04		
CD4-positive, alpha-beta T cell differentiation involved in immune response (GO:0002294)	29	.03	> 100	4.46E-08	9.33E-06		
positive regulation of ATP biosynthetic process (GO:2001171)	15	.01	> 100	1.18E-04	1.09E-02		
alpha-beta T cell differentiation involved in immune response (GO:0002293)	30	.04	> 100	2.79E-08	9.85E-06		
alpha-beta T cell activation involved in immune response (GO:0002292)	30	.04	> 100	2.79E-08	9.85E-06		
T cell differentiation involved in immune response (GO:0002292)	34	.04	> 100	4.43E-08	1.36E-05		
positive regulation of viral transcription (GO:0005023)	17	.02	> 100	1.40E-04	1.33E-02		
regulation of CD8-positive, alpha-beta T cell activation (GO:0011185)	18	.02	> 100	1.64E-04	1.42E-02		
CD4-positive, alpha-beta T cell differentiation (GO:0043267)	39	.04	> 100	7.29E-08	2.06E-05		
regulation of ATP biosynthetic process (GO:2001169)	20	.02	> 100	1.99E-04	1.68E-02		
positive regulation of pri-miRNA transcription by RNA polymerase II (GO:1902895)	41	.04	> 100	8.91E-08	2.32E-05		
response to corticosterone (GO:004127)	2	.02	> 100	1.81E-02	1.81E-02		
regulation of pri-miRNA transcription by RNA polymerase II (GO:1902893)	53	.05	> 100	98.35	2.09E-09	1.01E-06	
hypothalamus development (GO:0002154)	22	.02	> 100	3.38E-04	1.26E-06		
alpha-beta T cell differentiation (GO:0046632)	56	.05	> 100	2.70E-09	1.26E-06		
CD4-positive, alpha-beta T cell activation (GO:0035370)	45	.04	> 100	92.67	1.26E-07	3.04E-05	
cellular response to growth hormone stimulus (GO:0071378)	23	.02	> 100	5.86E-04	2.09E-02		
response to muscle stretch (GO:0035994)	24	.02	> 100	2.86E-04	2.28E-02		
positive regulation of purine nucleotide biosynthetic process (GO:0003733)	25	.02	> 100	83.40	3.02E-04	2.34E-02	
positive regulation of nucleotide biosynthetic process (GO:0030810)	25	.02	> 100	83.40	3.02E-04	2.33E-02	
regulation of cell receptor signaling pathway (GO:0005055)	2	.02	> 100	60.28	3.25E-04	2.44E-02	
beta-catenin-TCF complex assembly (GO:0046311)	69	.07	> 100	7.26E-09	3.04E-06		
negative regulation of lymphocyte apoptotic process (GO:0078229)	31	.02	> 100	67.26	4.52E-04	3.21E-02	
positive regulation of erythrocyte differentiation (GO:0045648)	32	.03	> 100	65.10	4.80E-04	3.37E-02	
response to mineralocorticoid (GO:0051385)	33	.03	> 100	5.09E-06	2.54E-02		
T cell activation involved in immune response (GO:0002286)	67	.04	> 100	62.24	5.72E-07	1.15E-04	
cellular response to growth hormone stimulus (GO:0000416)	2	.03	> 100	57.92	6.01E-04	4.10E-02	
positive regulation of myeloid leukocyte differentiation (GO:0002763)	58	.06	> 100	33.46	8.06E-04	53.93	2.61E-05
cellular response to calcium ion (GO:0042766)	39	.04	> 100	4.02E-04	2.75E-02		
positive regulation of myeloid cell differentiation (GO:0045639)	100	.05	> 100	52.13	4.27E-08	1.33E-03	
cellular response to calcium ion (GO:0071248)	85	.06	> 100	49.06	1.02E-06	2.57E-04	
regulation of myeloid leukocyte differentiation (GO:0002761)	123	.05	> 100	62.38	1.16E-07	2.83E-05	
neural precursor cell proliferation (GO:0061351)	75	.03	> 100	41.70	5.47E-05	5.65E-03	
dimethylsulfoniopropionate biosynthesis (GO:0002877)	3	.07	> 100	5.90E-05	6.86E-04		
lymphocyte activation involved in immune response (GO:0002285)	112	.04	> 100	4.20E-04	37.23	4.10E-06	6.39E-04
limbic system development (GO:0002761)	11	.11	> 100	3.90	1.24E-04	1.40E-04	
T cell differentiation (GO:0030217)	147	.14	> 100	35.46	2.72E-07	5.85E-05	
positive regulation of smooth muscle cell proliferation (GO:0048661)	90	.03	> 100	33.90	4.87E-06	6.73E-04	34.75
myeloid leukocyte differentiation (GO:0002733)	12	.12	> 100	31.91	1.18E-04	1.89E-02	
response to cAMP (GO:0051591)	98	.03	> 100	31.91	1.18E-04	1.89E-02	
positive regulation of hemopoiesis (GO:0037081)	6	.09	> 100	31.12	3.06E-08	1.01E-05	
regulation of myeloid cell differentiation (GO:0045637)	238	.07	> 100	30.66	1.92E-09	9.54E-07	
negative regulation of leukocyte differentiation (GO:1906246)	104	.04	> 100	27.44	1.33E-05	1.78E-03	
response to calcium ion (GO:0051592)	152	.15	> 100	27.44	1.33E-05	1.78E-03	
modulation of process of other organism involved in symbiotic interaction				116	.33	.11	26.96
positive regulation of leukocyte differentiation (GO:0005187)	289	.07	> 100	25.25	7.12E-08	3.80E-06	
regulation of leukocyte differentiation (GO:1902105)	289	.07	> 100	25.25	7.12E-08	3.80E-06	
lymphocyte differentiation (GO:0005086)	24	.24	> 100	24.06	2.66E-04	2.20E-02	
regulation of DNA binding (GO:0051101)	130	.12	> 100	23.52	2.84E-04	2.22E-02	
cellular response to reactive oxygen species (GO:0045614)	122	.12	> 100	23.52	2.84E-04	2.22E-02	
leukocyte differentiation (GO:0002521)	354	.34	> 100	23.52	2.84E-04	2.22E-02	
modulation of process of other organism (GO:0055521)	133	.13	> 100	23.52	2.84E-04	2.22E-02	
response to mechanical stimulus (GO:0002833)	223	.21	> 100	23.38	3.04E-08	5.71E-03	
regulation of smooth muscle cell proliferation (GO:0048660)	139	.13	> 100	23.38	3.04E-08	5.71E-03	
response to organophosphate (GO:0002833)	139	.13	> 100	23.38	3.04E-08	5.71E-03	
female pregnancy (GO:0007565)	488	.48	> 100	22.18	3.00E-05	3.53E-03	
positive regulation of hemopoiesis (GO:0037077)	142	.14	> 100	22.18	3.00E-05	3.53E-03	
response to estradiol (GO:0032355)	146	.14	> 100	22.18	3.00E-05	3.53E-03	
cellular response to metal ion (GO:0071248)	195	.19	> 100	21.39	3.45E-05	9.92E-03	
T cell activation (GO:0002110)	246	.24	> 100	21.19	3.45E-05	9.92E-03	
positive regulation of epithelial cell migration (GO:0018634)	148	.14	> 100	21.19	3.45E-05	9.92E-03	
response to interleukin-1 (GO:0008706)	19	.19	> 100	20.75	3.87E-05	3.87E-04	2.83E-02
regulation of T cell differentiation (GO:0045580)	151	.15	> 100	20.71	4.10E-04	2.96E-02	
response to purine-containing compound (GO:0014074)	155	.15	> 100	20.18	4.10E-04	3.15E-02	
regulation of hemopoiesis (GO:0037081)	9	.45	> 100	19.88	4.28E-04	3.15E-02	
response to transforming growth factor beta (GO:0015599)	221	.22	> 100	19.88	4.28E-04	3.15E-02	
multi-organismal organismal process (GO:0044706)	221	.22	> 100	19.88	4.28E-04	3.15E-02	
cellular response to inorganic substance (GO:0071241)	223	.22	> 100	18.70	5.76E-05	5.80E-03	
transcription by RNA polymerase II (GO:0006297)	409	.41	> 100	18.70	5.76E-05	5.80E-03	
cellular response to interleukin-1 (GO:0071347)	176	.17	> 100	17.77	6.35E-04	4.32E-02	
myeloid cell differentiation (GO:0008929)	235	.23	> 100	17.75	6.35E-04	4.32E-02	
regulation of lymphocyte differentiation (GO:0045619)	181	.18	> 100	17.28	6.82E-04	4.64E-02	
negative regulation of cell-cell adhesion (GO:0022408)	185	.18	> 100	16.91	7.32E-04	4.83E-02	
response to alcohol (GO:0009738)	24	.24	> 100	16.42	8.46E-05	8.43E-03	
lymphocyte activation (GO:0046648)	394	.39	> 100	15.88	1.47E-06	1.60E-04	
hemopoiesis (GO:0002761)	57	.57	> 100	15.62E-09	1.60E-04		
transcription, DNA-templated (GO:0006351)	558	.54	> 100	14.95	2.63E-08	9.51E-06	
nucleic acid-templated transcription (GO:0007650)	54	.54	> 100	14.95	2.63E-08	9.51E-06	
hematopoietic or lymphoid organ development (GO:0040534)	640	.64	> 100	14.58	3.10E-08	1.40E-09	1.45E-06
BNA biosynthetic process (GO:0032774)	572	.57	> 100	14.58	3.10E-08	1.40E-09	1.45E-06
cellular response to chemical stress (GO:0062127)	292	.29	> 100	14.58	3.10E-08	1.40E-09	1.45E-06
immune system development (GO:0002520)	675	.67	> 100	13.90	3.83E-09	2.22E-06	
response to drug (GO:0042423)	402	.40	> 100	12.97	5.32E-05	5.32E-05	
regulation of leukocyte cell-cell adhesion (GO:1903037)	322	.32	> 100	12.97	5.32E-05	5.32E-05	
angiogenesis (GO:0001525)	322	.32	> 100	12.97	5.32E-05	5.32E-05	
cell population proliferation (GO:0002833)	486	.48	> 100	12.56	3.89E-05	4.21E-03	
blood vessel morphogenesis (GO:0048514)	415	.41	> 100	12.56	3.89E-05	4.21E-03	
positive regulation of transcription by RNA polymerase II (GO:0004504)	422	.42	> 100	12.56	3.89E-05	4.21E-03	
response to lipid (GO:0033993)	857	.85	> 100	12.17	1.84E-09	9.44E-07	
regulation of epithelial cell proliferation (GO:0050678)	346	.34	> 100	12.17	1.84E-09	9.44E-07	
positive regulation of cell migration (GO:0030335)	528	.52	> 100	11.85	7.79E-06	1.13E-03	
response to inorganic substance (GO:0018035)	540	.54	> 100	11.41	9.61E-06	1.34E-03	
positive regulation of cell motility (GO:2000147)	51	.51	> 100	11.41	9.61E-06	1.34E-03	
response to metal ion (GO:0001038)	375	.37	> 100	11.41	9.61E-06	1.34E-03	
positive regulation of transcription, DNA-templated (GO:0006297)	1691	.16	> 100	11.41	9.61E-06	1.34E-03	
positive regulation of locomotion (GO:0040017)	567	.56	> 100	11.03	1.16E-05	1.53E-03	
positive regulation of cellular component morphogenesis (GO:0001272)	569	.56	> 100	10.92	1.39E-06	3.37E-04	
tube morphogenesis (GO:0032329)	668	.66	> 100	10.92	1.39E-06	3.37E-04	
regulation of nucleic acid-templated transcription (GO:1903508)	1886	.18	> 100	10.51	1.62E-02	2.69E-12	2.09E-12
positive regulation of RNA biosynthetic process (GO:1902608)	168	.16	> 100	10.51	1.62E-02	2.69E-12	2.09E-12
forebrain development (GO:0030980)	484	.48	> 100	10.32	5.44E-04	3.70E-02	
blood vessel development (GO:0001568)	595	.59	> 100	10.32	5.44E-04	3.70E-02	
response to cytokine (GO:0034097)	1131	1.13	> 100	10.32	5.44E-04	3.70E-02	
positive regulation of RNA metabolic process (GO:0051254)	1780	.17	> 100	10.32	5.44E-04	3.70E-02	
gland development (GO:0048732)	422	.42	> 100	9.88	6.40E-04	4.33E-02	
vasculature development (GO:0001944)	528	.52	> 100	9.88	6.40E-04	4.33E-02	
regulation of cytokine production (GO:0001817)	741	.74	> 100	9.88	6.40E-04	4.33E-02	
response to growth factor (GO:0008048)	533	.53	> 100	9.78	1.20E-04	1.40E-02	
cellular response to hormone stimulus (GO:0032329)	537	.53	> 100	9.78	1.20E-04	1.40E-02	
negative regulation of transcription by RNA polymerase II (GO:0000122)	978	.97	> 100	9.93	9.93E-05	9.67	1.12E-07
regulation of cell-cell adhesion (GO:0002761)	42	.42	> 100	9.65	9.93E-05	9.67	2.77E-05
positive regulation of cell differentiation (GO:0045597)	1012	.97	> 100	9.22	1.02E-08	1.09E-04	
response to organic cyclic compound (GO:0014078)	905	.87	> 100	9.22	1.02E-08	1.09E-04	
positive regulation of macroscopic biosynthetic process (GO:0001957)	1930	.19	> 100	9.22	1.02E-08	1.09E-04	
positive regulation of nucleobase-containing compound metabolic process (GO:0004593)	17	.17	> 100	9.22	1.02E-08	1.09E-04	
cellular response to chemical stress (GO:0062127)	292	.29	> 100	8.93	1.02E-08	1.09E-04	



positive regulation of macromolecule metabolic process (GO:0010684)	3630	17	3.40	+	4.88	8.52E-11	5.89E-08
cellular response to organic substance (GO:0071318)	2368	11	2.27	+	4.84	2.59E-06	4.34E-04
response to organic substance (GO:0010033)	3832	14	4.81	+	3.14E-08	1.02E-05	
regulation of cellular macromolecule biosynthetic process (GO:2000112)	3924	18	3.76	+	4.78	1.16E-11	1.23E-08
regulation of apoptotic process (GO:0042981)	1566	7	4.66	+	4.40E-04	3.15E-02	
regulation of macromolecule biosynthetic process (GO:0010556)	4832	18	3.87	+	4.65	1.87E-11	1.65E-08
regulation of nucleobase-containing compound metabolic process (GO:0019219)	4078	18	4.56	+	3.91	4.68	2.28E-11 1.91E-08
regulation of programmed cell death (GO:0043067)	1588	7	1.52	+	3.68	3.37E-02	
generation of neurons (GO:0048091)	1559	7	4.56	+	4.99E-04	3.48E-02	
positive regulation of metabolic process (GO:0009093)	3937	17	3.78	+	4.50	3.21E-10	2.04E-07
regulation of cellular biosynthetic process (GO:0031326)	4184	18	4.81	+	4.49	3.58E-11	2.84E-08
animal organ development (GO:0048513)	3277	14	3.14	+	4.45	8.61E-08	2.28E-05
regulation of response to stress (GO:0008134)	1641	7	1.57	+	4.45	5.84E-04	4.02E-02
regulation of biosynthetic process (GO:0009889)	4270	18	4.10	+	4.39	5.10E-11	3.69E-08
neurogenesis (GO:0022008)	1703	7	4.29	+	7.30E-04	4.84E-02	
cellular response to chemical stimulus (GO:0070037)	2841	12	2.82	+	4.25	2.62E-06	4.34E-04
anatomical structure morphogenesis (GO:0009653)	2286	9	2.12	+	4.25	9.47E-05	8.96E-03
nucleobase-containing compound metabolic process (GO:0000139)	2740	11	2.63	+	4.19	1.09E-05	1.46E-03
negative regulation of nitrogen compound metabolic process (GO:0051172)	2497	18	2.40	+	4.18	3.61E-05	4.08E-03
regulation of multicellular organismal process (GO:0051239)	3248	13	3.12	+	4.17	8.45E-07	1.60E-04
cell differentiation (GO:0030154)	3769	15	3.63	+	4.13	4.84E-08	1.43E-05
cellular developmental process (GO:0048869)	3845	15	3.69	+	4.07	5.96E-08	1.72E-05
immune system process (GO:0002761)	2953	11	2.74	+	4.02	1.62E-05	2.04E-03
heterocycle metabolic process (GO:0046483)	2936	11	2.82	+	3.91	2.13E-05	2.65E-03
negative regulation of cellular metabolic process (GO:0051324)	2780	10	2.58	+	3.86	7.14E-05	7.01E-03
nervous system development (GO:0007399)	2437	9	2.34	+	3.85	2.05E-04	1.71E-02
cellular aromatic compound metabolic process (GO:0006725)	2996	11	2.86	+	3.84	2.51E-05	3.02E-03
regulation of gene expression (GO:0010468)	4013	18	4.71	+	3.82	5.89E-10	3.40E-07
organic cyclic compound metabolic process (GO:1901360)	3218	11	3.09	+	3.56	5.14E-05	5.34E-03
negative regulation of macromolecule metabolic process (GO:0010665)	2951	10	2.83	+	3.53	1.53E-04	1.34E-02
cellular nitrogen compound metabolic process (GO:0034641)	3481	11	3.26	+	3.37	8.67E-05	8.41E-03
response to chemical (GO:0042212)	4417	14	4.24	+	3.30	3.88E-06	5.14E-04
regulation of signal transduction (GO:0009966)	3174	10	3.84	+	3.28	2.84E-04	2.23E-02
negative regulation of metabolic process (GO:0009092)	3285	10	3.07	+	3.25	3.08E-04	2.35E-02
system development (GO:0046731)	4525	14	4.24	+	3.22E-06	1.91E-04	
regulation of nitrogen compound metabolic process (GO:0051171)	5920	18	5.68	+	3.17	1.49E-08	5.92E-06
negative regulation of cellular process (GO:0048523)	4081	15	4.78	+	3.14	2.09E-06	3.58E-04
positive regulation of cellular process (GO:0048522)	5742	17	5.51	+	3.09	1.42E-07	3.27E-05
regulation of response to stimulus (GO:0048583)	4404	13	4.22	+	3.08	2.85E-05	3.39E-03
regulation of primary metabolic process (GO:0008090)	6118	18	5.87	+	3.07	2.62E-08	9.70E-06
regulation of cellular metabolic process (GO:0031223)	6329	18	6.87	+	2.97	4.89E-08	1.41E-06
regulation of macromolecule metabolic process (GO:0008253)	6510	18	6.24	+	2.88	7.01E-08	2.09E-05
multicellular organism development (GO:0007275)	5106	14	4.90	+	2.86	2.32E-05	2.84E-03
positive regulation of biological process (GO:0048518)	6347	17	6.09	+	2.79	6.93E-07	1.36E-04
negative regulation of biological process (GO:0048519)	5643	15	5.41	+	2.77	1.12E-05	1.48E-03
anatomical structure development (GO:0048565)	5489	14	5.26	+	2.66	5.58E-05	4.98E-03
regulation of metabolic process (GO:0019222)	7860	18	6.77	+	2.66	3.04E-07	6.46E-05
developmental process (GO:0032582)	5941	15	5.70	+	2.21E-05	2.72E-03	
multicellular organismal process (GO:0032981)	7148	15	6.86	+	2.19	2.41E-04	1.95E-02
regulation of cellular process (GO:0050794)	11398	19	10.93	+	1.74	1.39E-04	1.25E-02
regulation of biological process (GO:0050789)	11955	19	11.47	+	1.66	3.58E-04	2.65E-02