Structural changes in transcriptional regulatory networks for cell-type-specific gene expression during hematopoiesis.

Jun Nakabayashi

Department of Mathematics
Tokyo Medical Dental University
Ichikawa, Japan
junakabayashi@gmail.com

Abstract—Hematopoiesis is an extensively studied model system for cell differentiation. Cell-type-specific gene expression patterns are observed during hematopoiesis. Gene expression is governed by regulatory networks composed of cell-type-specific transcription factors. Resolving the transcriptional regulatory network for cell-type-specific gene expression provides a promising means of understanding the mechanisms underlying cell fate decisions. In this study, transcriptional regulatory networks in hematopoietic stem and progenitor cells were predicted based on gene expression profiles and distributions of transcription factor binding motifs in the promoter regions of cell-type-specific transcription factors. In particular, structural changes that occur when pluripotent stem cells progress to lineage-committed progenitors were evaluated. Marked changes in the regulatory circuit of transcription throughout the differentiation process could be elucidated by network analysis. Modular structures were a frequently described feature of biological networks observed in estimated networks. Within a module, most transcription factors were found to be regulated by a small number of regulators acting as downstream targets. Certain regulators within these modules coincide with known key regulators of hematopoietic cell differentiation. In addition to the modular structure, a two-layered structure was clearly observed in progenitor regulatory networks. Transcription factors could be distinctly divided into regulators within the regulatory layer and into targets in the output layer according to their degree of distribution. The restriction of mutual regulation between transcription factors was remarkable in that it allowed for alterations in network structures between hematopoietic stem cells and progenitors. Thus, using this approach, the relationships among transcription factors could be revealed by a reduction in mutual regulation to form a modular structure within the regulatory network.

Index Terms—hematopoiesis, transcriptional regulatory network, cell-type-specific gene expression,

I. Introduction

In multicellular organisms, a variety of cell types exist owing to diverse gene expression patterns governed by distinct regulatory networks composed of cell-type-specific transcription factors (TFs). It is difficult to explain how regulatory networks for cell-type-specific gene expression are developed throughout the differentiation process. Hematopoiesis has been well studied as a model system of cell differentiation. More than 10 types of hematopoietic cells are differentiated from hematopoietic stem cells (HSCs) via the formation of various progenitors committed to specific lineages. The differentiation process is controlled by the orchestration of intrinsic gene regulatory circuitry and external signals by various cytokines, chemokines, and growth factors [1]. Cell-type-specific gene expression patterns, including surface markers, are observed throughout hematopoiesis. Cell-type-specific gene expression is governed by regulatory circuitry that is composed of cell-type-specific TFs. Binding of TFs to specific sequences on genomic DNA is a primary step for gene expression. Then, TFs mutually activate or inhibit their expression to facilitate gene regulatory programs [2]. Cascades of TFs that range from upstream regulators to downstream targets ultimately form a complicated network.

Numerous efforts have been made to resolve the genetic programs of cell-type decisions. Cell type is defined by its cell-type-specific gene expression. Gene expression is governed by transcriptional regulatory networks (TRNs). Comprehensive analyses of TRNs have been performed, yielding large-scale datasets obtained from next-generation sequencing (NGS)-based technologies to estimate various large networks present within different organisms [3]–[6], [8]–[13]. TRNs are observed not only in hematopoietic cells but also in many other tissues, and these networks allow us to better understand the mechanisms underlying cell fate decisions [13]–[19]. Estimated networks are becoming more accurate and useful owing to findings obtained by advanced technologies. For example, the hierarchy and heterogeneity of the differentiation trajectory from HSCs to progenitors has been revealed by single-cell gene expression data [20]–[23]. Despite such sustained efforts, it remains challenging to resolve TRNs in the context of the entire system. One of the major difficulties in estimating the TRNs underlying cell-type-specific gene expression is that functional interactions between TFs must be correctly identified. A portion of TFs functionally regulate each other, whereas others are merely co-expressed in a particular cell type. A number of TF binding events onto genomic DNA are also functionally silent [24]. Combining gene expression profiles and sequence-based network modeling provides a promising means to model TRNs.

Network analysis provides an effective tool for analysis of

biological systems [25], [26]. Modules can be established as structures of TRNs [27] and various biological and social networks [28]–[31]. Numerous modules are known to occur within regulatory networks at the global level of genetic interactions. In addition to modules, various network motifs, such as positive and negative feedback circuitry in TRNs, play crucial roles in cell fate decisions [32], [33]. Structural changes in TRNs that occur during cell differentiation are currently an important research topic. One of the difficulties of network analysis is determining the means to evaluate these structural differences among networks. It is necessary to consider differences in the number of nodes or the probability of edges among networks. Various TRNs that function within numerous cell types throughout the process of differentiation must be considered. HSCs can be functionally distinguished from progenitors by their pluripotency and self-renewal capability. Structural differences between the TRNs of HSCs and those of progenitors may reflect these functional difference between HSCs and progenitors. Thus, it may be useful to consider the lineage commitment of progenitors.

In general, when a regulatory network is estimated from gene expression profiles, it is important that truly regulated genes are selected among the co-expressed genes. Among co-expressed genes, indirect regulation exists. To address this problem, researchers have developed the database for iRegulon cytoscape APP [34], curated by chromatin immunoprecipitation sequencing (ChIP-seq) data, which can indicate direct binding of TFs to genomic DNA in addition to sequence motifs.

In this study, TRNs from hematopoietic stem and progenitor cells (HSPCs) were estimated from gene expression profiles of 14 hematopoietic cells and the distributions of TF binding motifs within the promoter regions of cell-type-specific TFs. Critical changes between HSCs and progenitors were clarified in the context of network analysis.

II. METHOD

A. Detection of differentially expressed transcription factors

RNA-seq data for 14 types of hematopoietic cells were obtained from the GEO database (accession number GSE60101). This dataset was used because it includes the comprehensive expression profiles of 14 cell types from hematopoietic stem cells (HSCs) to terminally differentiated cells. Cell types and their abbreviations are summarized in Table**??**. Fastq formatted sequence tags were mapped onto mouse reference genome version mm10 using the mapping software tophat, and differentially expressed genes (DEGs) were detected by cuffdiff [51]. DEGs were selected as having false discovery rates of less than 0.05. Expression profiles were processed and visualized using cummeRbund R package [52]. Among DEGs, TFs deposited onto RIKEN Transcription Factor Database [53] were selected for subsequent analysis. Expression levels of TFs in HSCs were compared with those in multipotent progenitors (MPPs) to identify HSC-specific TFs whose expression levels in HSCs were significantly higher than those in MPPs. Inversely, the expression levels of MPP-specific TFs were

Cell Name	Abbreviation	Pluripotency	Self-renewal	SRA Accession number
Hematopoietic stem cell	HSC	pluripotent	+	SRR1536380 SRR1536381
Multi-potent progenitor	MPP	multipotent	+	SRR1536383 SRR1536384
Common lymphoid progenitor	CLP	olygopotent	+	SRR1536385 SRR1536386
Common myeloid progenitor	CMP	olygopotent	+	SRR1536389 SRR1536390
Megakatyo erythroid progenitor	MEP	olygopotent	+	SRR1536423 SRR1536424
Granulocyte-monocyte progenitor	GMP	olygopotent	+	SRR1536393 SRR1536394
Erythroid A	Ery A	differentiated	-	SRR1536427 SRR1536427
Monocyte	Mono	differentiated	-	SRR1536407 SRR1536408
Granulocyte	GN	differentiated	-	SRR1536401 SRR1536402
Macrophage	MF	differentiated	-	SRR1536397 SRR1536398
Natural killer cell	NK	differentiated	-	SRR1536421 SRR1536422
B cell	B	differentiated	-	SRR1536411 SRR1536412
CD4(+) T cell	CD4T	differentiated	-	SRR1536413 SRR1536414
CD8(+) T cell	CD8T	differentiated	-	SRR1536417 SRR1536418

TABLE I: Cell name abbreviations: Names and abbreviations of hematopoietic cells. Cells were divided into four groups: pluripotent, multipotent, olygopotent, and differentiated, according to the pluripotency status. HSPCs and differentiated cells were clearly divided by the self-renewal capability.

significantly higher than those in HSCs. Common lymphoid progenitor (CLP)- and common myeloid progenitor (CMP)-specific TFs were identified as having higher TF expression levels in CLPs or CMPs than those in MPPs. These lists of TFs were imported into the iRegulon plugin of cytoscape to detect TF binding motifs within the promoter regions of these TFs [34].

B. Construction of TRNs

TF expression is regulated by TF binding motifs that exist within the promoter regions of target TFs. A representative case with three TFs is shown in Fig1. In this scenario, the binding motif of TF B was detected in the promoter region of TF A. This indicated that TF A was regulated by TF B. The binding motif of TF A was detected in the promoter regions of TF B and C. Thus, TF A regulated TFs B and C. Not only binding motif of of TF A but also biding motif of TF C was detected in promoter region of TF C. Thus, TF C was regulated by TF A and C. Based on these findings, a TRN was obtained as a directional graph. The node and edge indicated the TF and the relationships among TFs, respectively. Arrows were drawn from the regulating TF to the regulated TF.

C. Network analysis

Network analysis was performed using the igraph R package [54]. Reciprocity, transitivity, and closeness were calculated using functions available in this package.

For dyad census, z-scores of mutual, asymmetrical, and null dyads were calculated as follows. Briefly, 1000 random graphs containing the same degree and probability of edges as HSC, MPP, CLP, and CMP networks were generated using the random.graph.game function in igraph package. Means and standard deviations (SDs) of mutual, asymmetrical, and null dyads of 1000 random graphs were calculated. Numbers
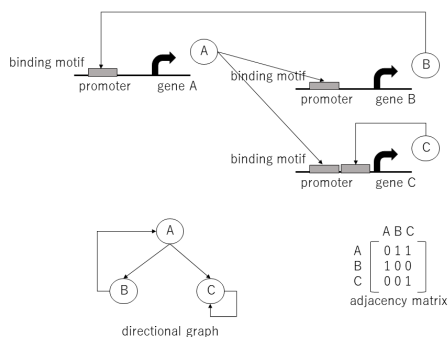
Fig. 1: Network construction. TF expression was assumed to be regulated by TF binding motifs enriched within promoter regions. Relationships among TFs are presented in a directional graph. Circles represent TFs, and relationships among TFs are indicated by arrows from regulators to downstream targets. Adjacency matrix is obtained. In this case TF A regulates B and C. B regulates A. C is regulated by A and C.

of mutual, asymmetrical, and null dyads were subtracted from the mean and divided by the SD to obtain the z-score. All codes for the analysis in this study are summarized in the Supplementary Methods.

III. RESULTS

A. Structural changes in TRNs from HSCs to progenitors

1) Identification of cell-type-specific TFs in HSPCs: Expression profiles of 14 types of hematopoietic cells originally deposited by Lara-Astiaso et al. were re-analyzed [35]. According to gene expression profiles, hematopoietic stem and progenitor cells (HSCs, MPPs, CLPs, and CMPs) were clustered into one group by hierarchical clustering. Similar to hierarchical clustering, HSPCs were expected to exist in proximity in a two-dimensional surface by multidimensional scaling. The classification of hematopoietic cells by gene expression profiles was nearly identical to that of the conventional theory of hematopoiesis. These results indicated that gene expression patterns of HSPCs were distinguished from those of terminally differentiated cells. Thus, expression patterns of HSPCs were common to some extent. Among HSPCs, HSCs are distinguished from progenitors by their pluripotency and self-renewal capability. To clarify the cause of the differences observed between HSCs and progenitors, the structures of TRNs of specific gene expression in HSPCs were investigated.

First, TFs significantly expressed in specific cells throughout hematopoiesis were identified to select TFs that functionally contributed to hematopoietic cell differentiation. Differentially expressed genes were identified based on whether their expression levels increased or decreased significantly in at least one cell type using 14 types of hematopoietic cells assessed by the cuffdiff program under the condition that the q-value was less

than 0.05. TFs were isolated from these DEGs as candidates of cell-type-specific TFs. These results indicated that a total of 823 TFs satisfied these criteria. HSC-, CLP-, and CMP-specific TFs were defined based on having expression levels higher than those found in MPPs. Conversely, MPP-specific TFs were defined based on having expression levels that were higher than those found in HSCs. From these results, 414 HSC-specific, 396 MPP-specific, 370 CLP-specific, and 343 CMP-specific TFs were identified. Lists of these TFs used to detect the enriched TF binding motifs in their promoter regions were uploaded to the iRegulon database [34].

B. TRNs in HSPCs

1) Modular structures of TRNs in HSPCs: TRNs of HSPCs presented as directional graphs are shown in Fig2(a-d).

(a)HSC (b)MPP

(c)CLP (d)CMP
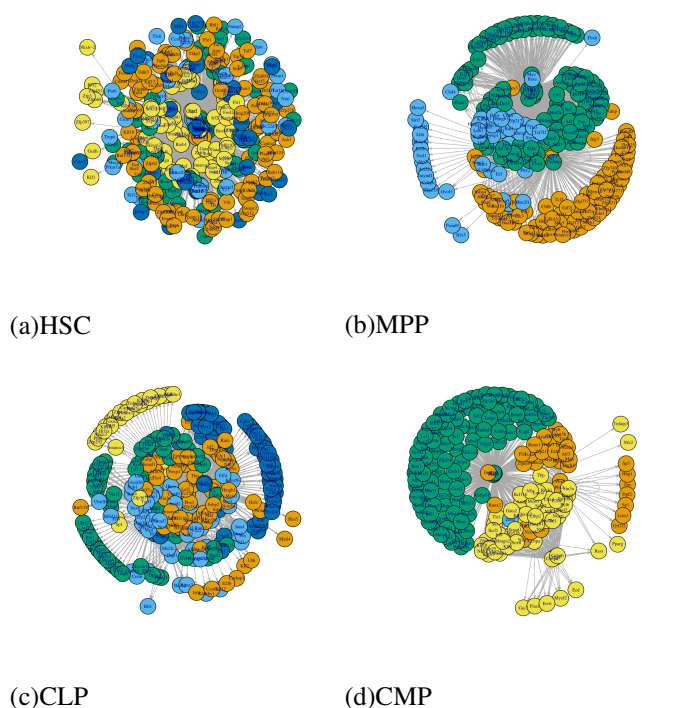
Fig. 2: TRNs in HSPCs. (a) TRNs in HSCs, (b) MPPs, (c) CLPs, and (d) CMPs TRNs are presented as directional graphs. Circles in the graph represent TFs. Labels in the vertex designate gene symbols of TFs. Arrow indicate the relationships among TFs from regulators to their downstream targets. TFs in these networks are subdivided into groups using the spinglass method. The groups are indicated by colors. Vertexes in the graphs are arranged by the Kamada-Kawai layout algorithm. Borders between groups in HSC network are ambiguous compared with TRNs in MPPs, CLPs, and CMPs.

2) Clustering TFs using the spinglass algorithm: TFs in TRNs were heuristically clustered into certain groups by maximizing the ratio of network density within a group to that between groups using the spinglass algorithm [36]. In

Fig2, colors indicates groups obtained by spinglass algorithm. As shown in Fig2, most TFs were clustered by their regulators. The results indicate modules within the TRNs. In a random graph having the same sizes of TRNs as observed in HSPCs, these modules were not observed (Data not shown). Modules in TRNs have been described in detail in previous studies [27], [37] as an important feature of TRNs. Some of the modules in the TRNs estimated in this study coincided with previously reported modules. For example, a module including C/EBP family members in CMPs (Fig2) coincided with the granulocyte and monocyte modules in human hematopoiesis, as described by Novershtern. Moreover, CEBPA and CEBPD have been reported to be included in this module. A module including Hbp1, Klf3, and Irf2, as indicated in the CLP network in Fig2, is equivalent to the erythroid, B-cell, and T-cell "re-use" module described by Novershtern [27]. Additionally, HBP1 , KLF3, and IRF2 are included in this Module, consistent with previously reported modules, supporting the validity of the TRNs estimated in this study.

TFs known as master regulators of differentiation for specific cell type are also included in such modules. In this study, TFs that act as key regulators of hematopoietic cell differentiation were included in the modules of TRNs estimated in this study. For example, a module including C/EBP family members (C/EBPα, β, δ, and ϵ), which are master regulators of myeloid cell differentiation [38]–[40], was observed in the CMP network. Lef1 and Tcf7 are master regulators of lymphoid cell differentiation [41] and were included in a module in the CLP network. Transcription factors within these modules contribute to the differentiation of specific cell types under the control of master regulators.

The formation of modules occurs prior to gene expression, and the average expression level of genes within the module gradually increases during progression from a progenitor to a terminally differentiated cell. The average expression level of TFs in a module was monitored to test whether the modular structure occurred prior to gene expression. Similar to Noversthtern's granulocyte/monocyte module, the average expression level of the C/EBP module gradually increased from CMPs to terminally differentiated myeloid cells. Moreover, the average expression level of TFs in the Lef1 module increased from CLPs to B- and T-cell lineages and erythroid and myeloid lineages, similar to the erythroid B- and T-cell "re-use" module detailed by Novershtern. These results indicated that module formation in TRNs preceded gene expression.

3) Two-layered structure: The basic statistics of these networks are summarized in TableII. The average path lengths derived from 1000 random graphs generated by the random.graph.game function in igrpah package possessing the same number of nodes and probabilities of edges for HSCs, MPPs, CLPs, and CMPs were 2.599165, 3.07034, 3.273771, and 3.64743, respectively. The average path lengths of TRNs from HSPCs were shorter than those of random graphs. These results, combined with the modular structures described in the previous subsection, suggested that both short path length and high modularity may be characteristics of TRNs of HSPCs.

Next, differences among TRNs in HSPCs were considered. The numbers of nodes and edges in the HSC network were 305 and 3539, respectively. The nodes and edges in the HSC network were both larger than those in progenitors. The average degree of the HSC network was larger than that of progenitors. The diameter and average path length of the HSC network were similar to those of progenitors. As described above, the HSC network appeared dense compared with the TRNs of progenitors. The borders between groups in the HSC network were ambiguous, but became clear in the TRNs of progenitors, as shown in Fig2 (a-d). Thus, the structure of the HSC network was different from that of progenitors.

To fully clarify the network structure, degree distribution was investigated. The probability of degree was plotted against the degree, as shown in Fig3(a-d). The degree distribution of the HSC network was wide and continuous. In contrast, the degree distribution of the progenitors was wide but discontinuous. TFs in the TRNs of hematopoietic progenitors were clearly subdivided into two groups according to their degree distribution. The first group included TFs exhibiting a high degree distribution, and the second group contained TFs exhibiting a low degree distribution, reflecting two-layered structure of TRNs. TFs with a high degree distribution that occupied the central regions of TRNs mutually regulated their expression, suggesting the presence of a regulatory layer. In contrast, most TFs with a low degree distribution surrounding TRNs were regulated by a small number of TFs, suggestive of an output layer. This two-layered structure, including regulatory and output layers, was clearly observed using the Kamada-Kawai layout algorithm [42] in Fig2(a-d). Assortativities of the HSC, MPP, CLP, and CMP networks calculated using the assortatibity.degree function in igraph package were -0.2937269, -0.1255377, -0.3657887, and -0.5141146, respectively, as shown in TableIII. These values indicated that the degree distribution was negatively correlated. TFs having a high degree distribution in the TRNs of HSPCs tended to regulate TFs having a low degree distribution. The observed assortativity also supported the two-layered structure of TRNs, in which TFs were clearly divided into regulators and targets.

TABLE II: Basic statistics of TRNs in HSPCs

	HSCs	MPPs	CLPs	CMPs
Number of nodes	305	234	244	196
Number of edges	3539	1567	1449	881
Average degree	23.20656	13.39316	11.87705	8.989796
Diameter	5	3	5	4
Average path length	1.892139	1.46323	1.924054	1.846971

TABLE III: Assortativity.

	HSC	MPP	CLP	CMP
assortativity	-0.2937269	-0.1255377	-0.3657887	-0.5141146

4) Bow-tie decomposition: Within various social and biological networks, the bow-tie structure is frequently observed [43]. Accordingly, the bow-tie structure in TRNs of HSPCs
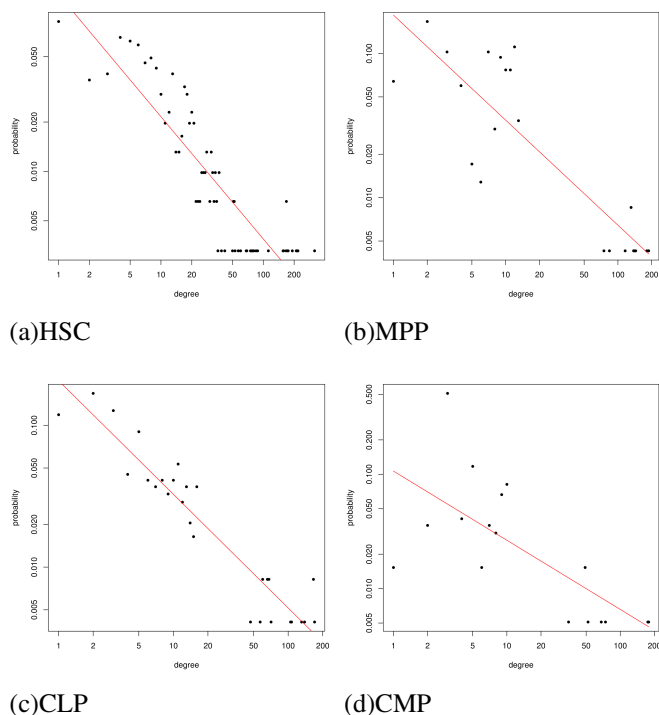
(a)HSC   (b)MPP

(c)CLP   (d)CMP

Fig. 3: Degree distributions of TRNs in HSPCs. (a) TRNs in HSCs, (b) MPPs, (c) CLPs, and (d) CMPs. Graphs indicate the degree distributions of TRNs in HSPCs. Vertical and horizontal axes indicate the degree numbers and distributions, respectively. Both axes are presented in log scale. The degree distribution of the HSC network was wide and continuous. There was a gap in the degree distributions of MPPs, CLPs, and CMPs. TFs having an intermediate degree were absent in the TRNs of these progenitors.

(a)HSC   (b)MPP

(c)CLP   (d)CMP

Fig. 4: Bow-tie decomposition. Input, regulatory, and output layers are indicated by green, orange, and light blue, respectively. The ratio of the output layer to the input and regulatory layers was extremely large in TRNs in HSPCs. The two-layered structure, rather than a bow-tie structure, was shown by these networks.

was investigated according to Yang's protocol. As shown in Fig4(a-d), a small fraction of TFs in the TRNs was classified into input and output layers; these were considered output-oriented bow-ties, as described previously [19]. TRNs in HSPCs showed a two-layered structure rather than a bow-tie structure. These data supported the idea that TRNs in HSPCs exhibited a two-layered structure composed of regulators and targets.

5) K-core decomposition: Next, a k-core decomposition of the graph was performed. Maximal subgraphs in which TFs possessed at least degree k were detected. Degree k is indicated in the node and reflected as the diameter of the node, as shown in Fig5(a-d). In the HSC network, TFs with various degree k values overlapped each other, particularly in the central region of the network. In contrast, TFs with the same degree were clearly clustered in the TRNs of progenitors. In these networks, TFs with a small degree k surrounding the network and TFs with a large degree k in the central region of the network were both clustered depending on the degree k. The structure of the TRNs in progenitors, but not HSCs, was arranged properly.

C. Mutuality of TRNs of HSPCs

1) Dyad census: The modular structures within the TRNs of progenitors were clearer than those in the HSC network. Changes in TRNs that contributed to module formation within the network were considered. Mutual regulation between TFs in the regulatory layer of the HSC network appeared complicated compared with that observed in the TRNs of progenitors. Thus, mutual regulation among TFs was restricted in the TRNs of progenitors compared with that of HSCs. To confirm the changes in mutual regulation between TFs from HSCs to progenitors, a dyad census was performed. The results of the dyad census are summarized in TableIV. The number of mutual dyad values in the HSC network was larger than that in the TRNs of progenitors. Z-scores in mutual, asymmetrical, and null dyads were calculated. Although the mutual dyad of the HSC network showed greater deviations than the means of the 1000 random graphs having the same node and probability degree in the HSC network, the mutual dyad in the TRNs of progenitors showed smaller deviation than the means of the 1000 random graphs. Conversely, the asymmetrical dyad in the TRNs in HSCs showed deviations larger than the means of 1000 random graphs. Reciprocity (proportion of mutual connection) was also smaller in the TRNs in progenitors compared with those in HSCs. The mean

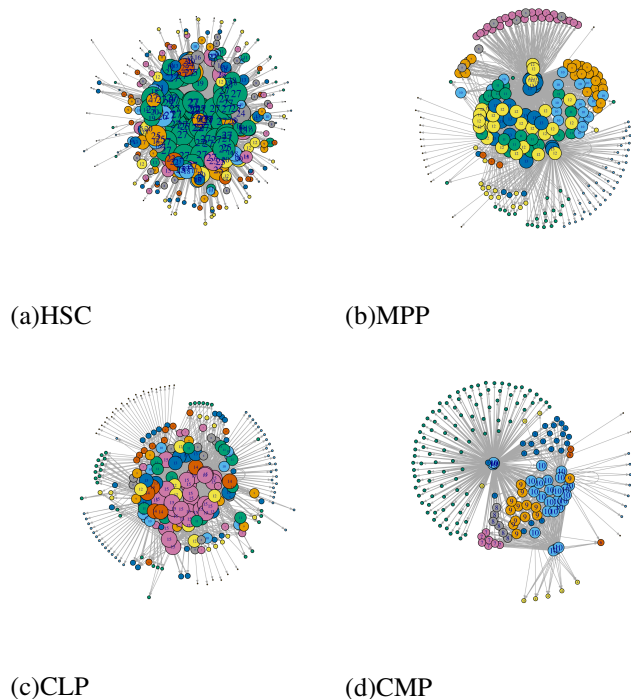(a)HSC

(b)MPP

(c)CLP

(d)CMP

Fig. 5: K-core decomposition. Maximal subgraphs in which TFs exhibit at least degree k were detected. Degree k is indicated by the numbers in the circles in the TRNs of the HSPCs. TFs possessing the same degree k values are indicated by the same colors. TFs with various k values overlapped with each other in the HSC network. TFs were arranged properly in the TRNs of MPPs, CLPs, and CMPs.

reciprocities of 1000 random graphs with the same nodes and probability of edges for HSC, MPP, CLP, and CMP networks were 0.03815487, 0.02887527, 0.02413445, and 0.02312244, respectively. In particular, the reciprocity of the CMP network was smaller than those of the HSC, MPP, and CLP networks and smaller than that of the mean of 1000 random graphs using the same nodes and probability of edges as the CMP network. These results indicated that mutual regulation among TFs in HSCs was dramatically restricted in progenitors.

TABLE IV: Dyad census. Mutual and asymmetric interactions between pairs of TFs in TRNs in HSPCs were identified by the dyad.census function of igrpah package. To compare the mutuality among TRNs in HSPCs, deviations from random graphs possessing the same number of nodes and probabilities were estimated. Z-scores were obtained as described in the Methods.

		HSC	MPP	CLP	CMP
mutual	number	82	27	21	3
	z-score	1.755567	0.9826579	0.7565645	-2.192493
asymmetrical	number	3374	1513	1407	875
	z-score	-0.5751004	-0.2531842	-0.1600614	0.5568425
null	number	42904	25721	28218	18232
	z-score	0.3076151	0.1298958	0.07200748	-0.2939236
reciprocity		0.04256527	0.02825947	0.02359473	0.004550626
	z-score	0.9731325	-0.1059991	-0.09469029	-2.687319

2) Triad census: Next, a triad census was performed. Similar to the dyad census, a reduction in the mutual regulation among TFs was observed. The transitivity values of HSCs, MPPs, CLPs, and CMPs were 0.210478, 0.1347141, 0.1381967, and 0.05728737, respectively as shown in TableV. The ratio of triangles to connected triples in the network reduced during progression from HSCs to progenitors. This result also indicated that mutual regulation among TFs was reduced during progression from HSCs to progenitors.

TABLE V: Triad census. Mutual and asymmetric interactions among three TFs in TRNs in HSPCs were identified using the triad.census function of igrpah.

	HSC	MPP	CLP	CMP
null	3925220	1866759	2130787	1122660
012	502644	135840	179471	57193
102	16886	4684	1289	19
021D	195467	92755	71377	53202
021U	8704	757	1784	1067
021C	6376	1147	1210	275
111D	347	30	21	5
111U	7254	1234	1447	249
030T	12871	2750	2749	996
030C	13	0	0	0
201	71	0	4	0
120D	266	51	30	0
120U	5745	2122	1227	108
120C	153	5	12	5
210	283	31	28	1
300	60	19	8	0
transitivity	0.210473	0.134714	0.138197	0.057287

D. Evaluation of key regulators in modules based on network structures

1) Closeness centrality: In the previous section, the results showed that a reduction in mutual regulation was responsible for structural differences in TRNs in HSCs and progenitors. Most TFs were unidirectionally regulated by a small number of regulators. Imbalance between incoming and outgoing ties was the key feature of TRNs in HSPCs and was apparent as cells progressed from HSCs to progenitors. To further evaluate the network structure of TRNs, closeness centrality was calculated to reflect the features of the network structure. As shown in Fig6(a-d), calculated closeness was reflected by the diameter and color density of TFs. Certain TFs exhibiting large closeness coincided with known key regulators. For example, C/EBP family members in the CMP network possessed large closeness values. These results indicated that indexes such as centrality could be used to evaluate these key regulators of hematopoietic cell differentiation as important players in TRNs. The top 20 TFs exhibiting large closeness centrality are summarized in TableVI. These TFs may play critical roles as key regulators determining cell-type-specific gene expression patterns.

E. Prediction of new key regulators

As described above, key regulators appeared in TRNs and contributed to specific lineage determination. Such key

(b)MPP

(a)HSC

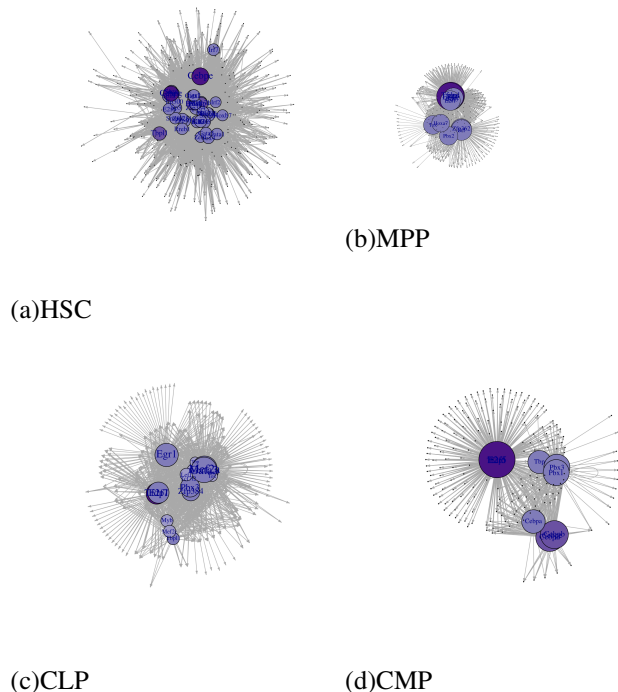(c)CLP   (d)CMP

Fig. 6: Closeness centrality. Closeness centrality was reflected by the diameters of TFs in TRNs in HSPCs. (a) TRNs in HSCs, (b) MPPs, (c) CLPs, and (d) CMPs.

TABLE VI: Top 20 TFs exhibiting large closeness centrality.

	HSC	MPP	CLP	CMP
1	Cebpe	E2f4	Tfdp1	E2f5
2	Cebpg	Tfdp1	Egr1	E2f6
3	Tbpl1	Srf	Gata3	Tfdp1
4	Pbx4	Zfp362	Mef2a	Cebpb
5	Tfap4	E2f1	E2f7	Cebpd
6	Sirt6	E2f2	Pbx3	Cebpe
7	Klf5	E2f3	Tbpl1	Pbx3
8	Klf16	E2f5	Zfp384	Pbx1
9	Klf6	E2f7	Tox4	Cebpa
10	Klf7	Tox	Tox	Tbp
11	Mtf1	Pbx2	Tcf7	Eed
12	Zbtb7b	Hoxa7	Lef1	Etv6
13	Pbx3	Tbx6	Tcf7l2	Pias1
14	Pbx1	Hoxa5	Tbp	Gas7
15	Zfp384	Sirt7	Mef2c	Myef2
16	Nfe2l1	Psmd9	Myb	Sp3
17	Smad4	Rfx5	Hif1a	Hbp1
18	Irf7	Mcm2	Zbtb2	Lmo2
19	Egr1	Mcm7	Phf8	Rest
20	Zfp281	Mcm5	Zdhhc1	Sp1

regulators may be correctly evaluated by indexes representing the features of the network. Finally, other key regulators that play crucial roles in specific hematopoietic cell differentiation patterns were predicted based on network structure. In the CMP network, many TFs were mainly regulated by Pbx1 and Pbx3, as illustrated in Fig7a. The average expression levels of TFs in this Pbx module were assessed. The results showed that the average expression levels of these Pbx module members specifically increased from CMPs to Ery-A, as indicated in Fig7b. This result suggested that Pbx1 and Pbx3 regulate many TFs in the context of erythropoiesis. Importantly, Pbx family members are known to be important TFs that play critical roles in maintaining the pluripotency of HSCs rather than in erythrocyte differentiation.

IV. DISCUSSION

A. Validation of network estimation

In this study, using comprehensive datasets of expression profiles from 14 types of hematopoietic cells, differentially expressed genes that likely contributed to cell fate decisions through hematopoiesis were properly selected. The results showed that expression profiles from 14 types of hematopoietic cells were sufficient to select appropriate genes contributing to differentiation. Among the co-expressed genes, indirect regulation exists. Information regarding TF binding motifs in the promoter regions was suitable for selection of direct interactions between TFs [44]. Analysis of the database for iRegulon cytoscape APP [34] curated by ChIP-seq data indicated direct binding of TFs to genomic DNA in addition to sequence motifs. Notably, genome-wide TF binding maps are useful for detecting the key regulators of mast cell differentiation [5], and binding sites detected by iRegulon are sufficiently probable.

TRNs of cell-type-specific gene expression in HSPCs were estimated from gene expression profiles and the distributions of TF binding motifs in the promoter regions of cell-type-specific TFs. Gene expression of TFs was assumed to be regulated by TFs, as their binding motifs were enriched in the promoter regions of target TFs. Regulator TFs could positively or negatively regulate the expression of their targets. When cell-type-specific TFs were identified, TF expression was increased in specific cell types, and these TFs were selected for further analysis. Using these criteria, activation was a prerequisite, and negative regulation was ignored. Although positive and negative regulation of transcription occur complementarily, it was likely that activating the genetic switch may play a critical role in cell fate decisions. For cell fate decisions, signature genes characterizing a particular cell type must be expressed.

In estimated networks, modular structures are observed. TFs in a particular module regulate numerous targets, such as the C/EBP family in the CMP network and Lef1 in the CLP network. Certain modules in TRNs in HSPCs coincide with previously reported modules. This result supported the validity of the modeling method and estimated TRNs evaluated in this study. TFs that truly contribute to cell fate

decisions through hematopoiesis were precisely identified as differentially expressed genes using 14 types of hematopoietic cells. Indirectly co-expressed genes were eliminated from the TRNs by a combination of TF binding motif analysis within promoter regions. Overall, the findings of this study overcame a major problem with predicting genome-wide networks.

B. Structure of TRNs with cell-type-specific gene expression in HSPCs

1) Two-layered structure: The importance of modular structures in regulatory networks has been established in studies of biological sciences. The TFs in the TRNs estimated in this study were clustered into various groups using the spinglass algorithm. Similar to previous studies, a modular structure in TRNs was observed. In such modules, most TFs are specifically regulated by a small number of regulators to form the functional ensemble of TFs.

In addition to the modular structure, a two-layered structure was observed in the TRNs of HSPCs, and this was particularly true in progenitors. TFs in the TRNs of progenitors were clearly divided into regulators and targets according to the degree distribution. Moreover, regulators were identified in the central region of the TRNs through the use of the Kamada-Kawai layout algorithm, representing the regulatory layer. Additionally, TRNs were surrounded by targets, representing the output layer. In previous reports, various biological networks have been shown to exhibit a bow-tie structure [45]; however, the TRNs in HSPCs exhibited output-oriented bow-tie structures wherein the ratio of the output layer to the input and regulatory layers was extremely large compared with that in the typical three-layered bow-tie structure. Notably, the two-layered structure was observed in TRNs in HSPCs.

2) Structural differences between the HSC network and TRNs in progenitors: HSPCs were clustered into one group according to their gene expression profiles. HSCs, however, are functionally distinguished from lineage-committed progenitors according to their pluripotency and self-renewal capabilities. In the HSC network estimated in this study, Klf family members known to contribute to self-renewal and pluripotency were included [46]–[48]. These factors were found to possess large closeness centrality values. Additionally, these factors have been shown to play critical roles in maintenance of the self-renewal and pluripotency of HSCs. In addition to investigating individual TFs, the structure of the entire regulatory network was also assessed. In this study, differences between HSCs and lineage-committed progenitors are clearly shown in the context of network analysis.

Differences between HSCs and progenitors can be determined from the structure of the TRNs. The HSC networks appeared different from those of progenitors. When performing integrative analysis of regulatory networks of hematopoiesis [6], [27], [49], it is important to clearly depict structural differences in regulatory networks between HSCs and progenitors. The borders between modules in the HSC network were ambiguous compared with those in the TRNs of progenitors. Moreover, the HSC network appeared similar to a random

(a)

(b)

Fig. 7: Pbx module in the CMP network. TFs clustered into a group with Pbx1 and Pbx3 in the CMP network were identified. The average expression levels of these TFs are reflected by the diagram of hematopoiesis. These TFs were specifically expressed in the erythroid lineage.

graph, and the modular structure was developed during the differentiation process from HSCs to progenitors.

According to an analysis of the basic statistics of TRNs in HSPCs, the HSC network was dense compared with the other networks. The average degree of the HSC network was larger than that of the progenitor network. Additionally, the degree distribution of the HSC network was different from that of the progenitor network. The degree distribution of the HSC network was wide and continuous, whereas that of the progenitor network was wide but not continuous. TFs in TRNs in progenitors could be clearly divided into regulators and targets. Furthermore, a two-layered structure was clearly observed in TRNs in progenitors but not in TRNs of the HSC network. In particular, the regulatory layer of the HSC network was complicated. When the TFs in the TRNs were clustered into various groups using the spinglass method, TFs clustered into various groups overlapped. In the TRNs of progenitors, the TFs were all properly arranged.

Next, the causes underlying the structural differences between the HSC network and the TRNs in progenitors were considered. The dyad census revealed that a reduction in mutual regulation between TFs occurred as cells progressed from HSCs to progenitors. Most TFs possessing low degree distributions were unidirectionally regulated by a small number of key regulators and were clustered to form a modular structure. Relationships among TFs were arranged according to this reduction in mutual regulation among TFs. According to Waddington's landscape model, cell-fate can be induced by activation of a positive feedback system [50]. Once key regulators for a specific cell lineage, such as the C/EBP family for myeloid cells or Lef1 for lymphoid cells, are activated by a positive loop within mutual regulatory networks in TRNs, pluripotency may be lost. As a result, many targets are unidirectionally regulated by a small number of regulators within a particular module in lineage-committed progenitors. This provides a potential scenario for the differentiation process that involves structural changes in TRNs as cells progress from HSCs to progenitors.

C. Prediction of new key regulators of hematopoietic cell differentiation

The TRNs estimated in this study not only coincided with previously reported networks but were also useful for predicting new critical regulators of hematopoietic cell differentiation. Pbx1 and Pbx3 regulate many TFs as their targets in the CMP network. Although Pbx family members are known as important regulators of the maintenance of pluripotency in HSCs, their roles in the context of terminal differentiation are poorly understood. The average expression levels of TFs regulated by Pbx1 and Pbx3 in this Pbx module were found to be particularly increased in erythrocytes, indicating that Pbx1 and Pbx3 function as key regulators of erythrocyte differentiation. It is useful to find new key regulators for understanding about both normal hematopoiesis and diseases caused by abnormal hematopoiesis. The prediction of these proteins as key regulators can be confirmed by experiments

that perturb these TFs in TRNs. The findings of this study reinforced the concept that network analysis is an important tool for studying biological regulatory networks.

ETHICS

In this study, data obtained from humans or human tissues were not used.

DATA ACCESSIBILITY

Datasets (accession number GSE60101) were obtained from the GEO database in this study.

AUTHOR CONTRIBUTIONS

J.N. designed the study and analyzed the data. J.N. interpreted the results and wrote the manuscript. The author approved this work for publication.

COMPETING INTERESTS

The author declares no competing interests.

FUNDING

REFERENCES

[1] X. Chen, H. Xu, P. Yuan, F. Fang, M. Huss, VB. Vega, et al. "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells," Cell, vol. 133(6), pp.1106–1117, Jun 2008.

[2] NE. Buchler, U. Gerland, and T Hwa. "On schemes of combinatorial transcription logic," Proceedings of the National Academy of Sciences," vol. 100(9), pp. 5136–5141, Apr 2003.

[3] I. Amit, A. Regev, and N. Hacohen. "Strategies to discover regulatory circuits of the mammalian immune system," Nature Reviews Immunology, vol. 11(12), pp. 873–880, Nov 2011.

[4] D. Beck, JA. Thoms, D. Perera, J. Schütte, A. Unnikrishnan, K. Knezevic, et al. "Genome-wide analysis of transcriptional regulators in human hspcs reveals a densely interconnected network of coding and noncoding genes," Blood, vol. 122(14) pp. e12–e22, Oct 2013.

[5] FJ. Calero-Nieto, FS. Ng, NK. Wilson, R. Hannah, V. Moignard, AI. Leal-Cervantes, et al. "Key regulators control distinct transcriptional programmes in blood progenitor and mast cells," The EMBO journal, vol. 33(11), pp. 1212–1226, Jun 2014.

[6] DK. Goode, N. Obier, MS. Vijayabaskar, M. Lie-A-Ling, AJ. Lilly, R. Hannah, et al. "Dynamic gene regulatory networks drive hematopoietic specication and differentiation," Developmental cell, vol. 36(5), pp. 572–587, Mar 2016.

[7] B. Göttgens. "Regulatory network control of blood stem cells," Blood, vol. 125(17), pp. 2614–2620, Apr 2015.

[8] M. Ichikawa, T. Asai, S. Chiba, M. Kurokawa, and S. Ogawa. "Runx1/aml-1 ranks as a master regulator of adult hematopoiesis," Cell Cycle, vol. 3(6), pp. 722–724, Jun 2004.

[9] V. Jojic, T. Shay, K. Sylvia, O. Zuk, X. Sun, J. Kang, et al., "Identification of transcriptional regulators in the mouse immune system," Nature immunology, vol. 14(6), pp. 633–643, Jun2013.

[10] P. Laslo, JM. Pongubala, DW. Lancki, and H. Singh. "Gene regulatory networks directing myeloid and lymphoid cell fates within the immune system," Semi immunology, vol. 20(4), pp. 228–235, Aug 2008.

[11] E. Laurenti, S. Doulatov, S. Zandi, I. Plumb, J. Chen, C. April, et al. "The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment," Nature immunology, vol. 14(7), pp. 756–763, Jul 2013.

[12] Y. You, Cuevas-Diaz Duran R, L. Jiang, X. Dong, S. Zong, M. Snyder, and JQ. Wu. "An integrated global regulatory network of hematopoietic precursor cell self-renewal and differentiation," Integr Biology, vol. 16;10(7), pp. 390-405, Jul 2018.

[13] SF. Boj, M. Párrizas, M. A Maestro, and Jorge Ferrer. "A transcription factor regulatory circuit in differentiated pancreatic cells," Proc Natl Aca Sci, vol. 98(25), pp. 14481–14486, Dec 2001.

[14] LA. Boyer, TI. Lee, MF. Cole, SE. Johnstone, SS. Levine, JP. Zucker, et al. "Core transcriptional regulatory circuitry in human embryonic stem cells," Cell, vol. 122(6), pp. 947–956, Sep 2005.

[15] V. Chickarmane, C. Troein, UA. Nuber, HM. Sauro, and C. Peterson. "Transcriptional dynamics of the embryonic stem cell switch," PLoS Comput Biol, vol. 2(9), pp. e123, Sep 2006.

[16] V. Kashyap, NC. Rezende, KB. Scotland, SM. Shafier, JL. Persson, LJ. Gudas, and NP. Mongan. "Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the nanog, oct4, and sox2 pluripotency transcription factors with polycomb repressive complexes and stem cell micrornas," Stem Cells Dev, vol. 18(7), pp. 1093–1108, Sep 2009.

[17] J. Kim, J. Chu, X. Shen, J. Wang, and SH. Orkin. "An extended transcriptional network for pluripotency of embryonic stem cells," Cell, vol. 132(6), pp. 1049–1061, Mar 2008.

[18] J. Kyrmizi, P. Hatzis, N. Katrakili, FT.ronche, FJ. Gonzalez, and I. Talianidis. "Plasticity and expanding complexity of the hepatic transcription factor network during liver development. Genes Dev, vol. 20(16), pp. 2293–2305, Aug 2006.

[19] J. Li, X. Hua, M. Haubrock, J. Wang, and E. Wingender. "The architecture of the gene regulatory networks of different tissues. Bioinformatics, vol. 28(18), pp. i509–i514, Sep 2012.

[20] H. Ema, Y. Morita, and T. Suda. Heterogeneity and hierarchy of hematopoietic stem cells. Exp Hematol, vol. 42(2), pp. 74–82, Feb 2014.

[21] FK. Hamey, S. Nestorowa, S. J Kinston, DG. Kent, NK. Wilson, and B. Göttgens. "Reconstructing blood stem cell regulatory network models from single-cell molecular profiles," Proc Natl Aca Sci, vol. 6;114(23), pp. 5822-5829, Jun 2017.

[22] V. Moignard, S. Woodhouse, L. Haghverdi, AJ. Lilly, Y. Tanaka, AC. Wilkinson, et al. "Decoding the regulatory network of early blood development from single-cell gene expression measurements," Nat Biotechnol, vol. 33(3), pp. 269–276, Mar 2015.

[23] S. Watcham, I. Kucinski, and Berthold Gottegens. "New insights into hematopoietic dierentiation landscapes from single-cell rna sequencing," Blood, vol. 133(13), pp. 1415–1426, Mar 2019.

[24] MB. Gerstein, A. Kundaje, M. Hariharan, SG. Landt, KK. Yan, C. Cheng, et al. "Architecture of the human regulatory network derived from encode data," Nature, vol. 489(7414), pp. 91–100, Sep 2012.

[25] S. Fortunato. "Community detection in graphs," Physics reports, vol. 486(3-5), pp. 75–174, Feb 2010.

[26] M. Girvan and MEJ. Newman. "Community structure in social and biological networks. "Proc Natl Acad Sci, vol. 99(12), pp. 7821–7826, Jun 2002.

[27] N. Novershtern, A. Subramanian, LN. Lawton, RH. Mak, W. Haining, ME. McConkey, et al. "Densely interconnected transcriptional circuits control cell states in human hematopoiesis," Cell, vol. 144(2), pp. 296–309, Jan 2011.

[28] LH. Hartwell, J.J Hopfield, S. Leibler, and AW. Murray. "From molecular to modular cell biology," Nature, vol. 402, pp. C47–52, Dec 1999.

[29] MEJ. Newman and M. Girvan. "Finding and evaluating community structure in networks," Phys Rev E Stat Nonlin Soft Matter Phys, vol. 69(2 Pt 2), Nov 2004.

[30] MEJ. Newman. "Modularity and community structure in networks. Proc Natl Acad Sci, vol. 103(23), pp. 8577–8582, Jun 2006.

[31] D. Vella, S. Marini, F. Vitali, D. Silvestre, G. Mauri, and R. Bellazzi. "MTGO: PPI network analysis via topological and functional module identification. Sci Rep, vol. 8(1), pp. 5499, Apr 2018.

[32] U. Alon. "Network motifs: theory and experimental approaches," Nat Rev Genet, vol. 8(6), pp. 450–461, Jun 2007.

[33] A. Becskei, B. Séraphin, and L. Serrano. "Positive feedback in eukaryotic gene networks: cell dierentiation by graded to binary response conversion." EMBO J, vol. 20(10), pp. 2528–2535, May 2001.

[34] A. Verfaillie, H. Imrichová, B. Van de Sande, L. Standaert, V. Christiaens, et al. "iRegulon: from a gene list to a gene regulatory network using large motif and track collections," PLoS Computl Biol, vol. 10(7), pp. e1003731, Jul 2014.

[35] D. Lara-Astiaso, A. Weiner, E. Lorenzo-Vivas, I. Zaretsky, DA. Jaitin, E. David, et al. "Chromatin state dynamics during blood formation," Science, vol. 345(6199), pp. 943–949, Aug 2014.

[36] J. Reichardt and S. Bornholdt. "Statistical mechanics of community detection. Phys Rev E Stat Nonlin Soft Matter Phys, vol. 74(1 pt 2), pp. 016110, Jul 2006.

[37] E. Segal, M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller, and N. Friedman. "Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data," Nature Genet, vol. 34(2), pp. 166-176, Jun 2003.

[38] R. Avellino, M. Havermans, C. Erpelinck, MA. Sanders, R. Hoogenboezem, HJG. van de Werken, E. Rombouts, et al. "An autonomous cebpa enhancer specific for myeloid-lineage priming and neutrophilic differentiation," Blood, vol. 16;127(24), pp. 2991-3003, Jun 2016.

[39] B. Di Stefano, JL. Sardina, C. van Oevelen, S. Collombet, EM. Kallin, GP. Vicent, J. Lu, et al. "C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells," Nature, vol. 506(7487), pp. 235–239, Feb 2014.

[40] A. Itoh-Nakadai, M. Matsumoto, H. Kato, J. Sasaki, Y. Uehara, Y. Sato, et al. "A Bach2-CEBP gene regulatory network for the commitment of multipotent hematopoietic progenitors. Cell Rep, vol. 18(10), pp. 2401–2414, Mar 2017.

[41] E. Montecino-Rodriguez, D. Casero, M. Fice, J. Le, and K. Dorshkind. "Differential expression of PU. 1 and key T lineage transcription factors distinguishes fetal and adult t cell development. J Immunol, vol. 200(6), pp. 2046–2056, Mar 2018.

[42] T. Kamada, S. Kawai, et al. "An algorithm for drawing general undirected graphs," Information Processing Letters, vol. 31(1), pp.7–15, Apr 1989.

[43] R. Yang, L. Zhuhadar, and O. Nasraoui. "Bow-tie decomposition in directed graphs," 14th International Conference on Information Fusion, pp. 1–5, Jul 2011.

[44] KL. MacQuarrie, AP. Fong, RH. Morse, and SJ. Tapscott. "Genome-wide transcription factor binding: beyond direct target regulation," Trends in Genet, vol. 27(4), pp. 141–148, Apr 2011.

[45] P. Csermely, A. London, LY Wu, and B. Uzzi. "Structure and dynamics of core/periphery networks," J Complex Networks, vol. 1(2), pp.93–123, Sep 2013.

[46] H. Jeon, T. Waku, T. Azami, J. Yanagisawa, S. Takahashi, Masatsugu Ema, et al. "Comprehensive identification of krüppellike factor family members contributing to the self-renewal of mouse embryonic stem cells and cellular reprogramming," PLoS One, vol. 11(3), pp. e0150715, Mar 2016.

[47] R. Pearson, J. Fleetwood, S. Eaton, M. Crossley, and S. Bao. "Krüppel-like transcription factors: a functional family," Int J Biochem Cell Biol, vol. 40(10), pp. 1996–2001, Aug 2008.

[48] M. Yamane, S. Ohtsuka, K. Matsuura, A. Nakamura, and H. Niwa. "Overlapping functions of krüppel-like factor family members: targeting multiple transcription factors to maintain the naïve pluripotency of mouse embryonic stem cells," Development, vol. 145(10), pp. dev162404, May 2018.

[49] G. May, S. Soneji, AJ. Tipping, J. Teles, SJ. McGowan, M. Wu, et al. "Dynamic analysis of gene expression and genome-wide transcription factor binding during lineage specification of multipotent progenitors," Cell stem cell, vol. 13(6), pp.754–768, Dec 2013.

[50] JE. Ferrell Jr. "Bistability, bifurcations, and waddington's epigenetic landscape," Curr Biol, vol. 22(11), pp. R458–R466, Jun 2012.

[51] C. Trapnell, A. Roberts, L. Goff, G. Pertea, DKR. Kelley, H. Pimentel, et al. "Differential gene and transcript expression analysis of RNA-seq experiments with tophat and cufflinks. Nat Protoc, vol. 7(3), pp. 562, Mar 2012.

[52] LA. Goff, C. Trapnell, and D. Kelley. "CummeRbund: visualization and exploration of cufflinks high-throughput sequencing data," R package version 2.26.0, 2019.

[53] M. Kanamori, H. Konno, N. Osato, J. Kawai, Y. Hayashizaki, H. Suzuki. "A genome-wide and nonredundant mouse transcription factor database," Biochem Biopphys Res Commun. Vol. 322(3), pp. 787-793, Sep 2004

[54] G. Csardi and T. Nepusz. "The igraph software package for complex network research," Int J Comp Sys. vol. 1695(5), pp. 1-9, 2006

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)