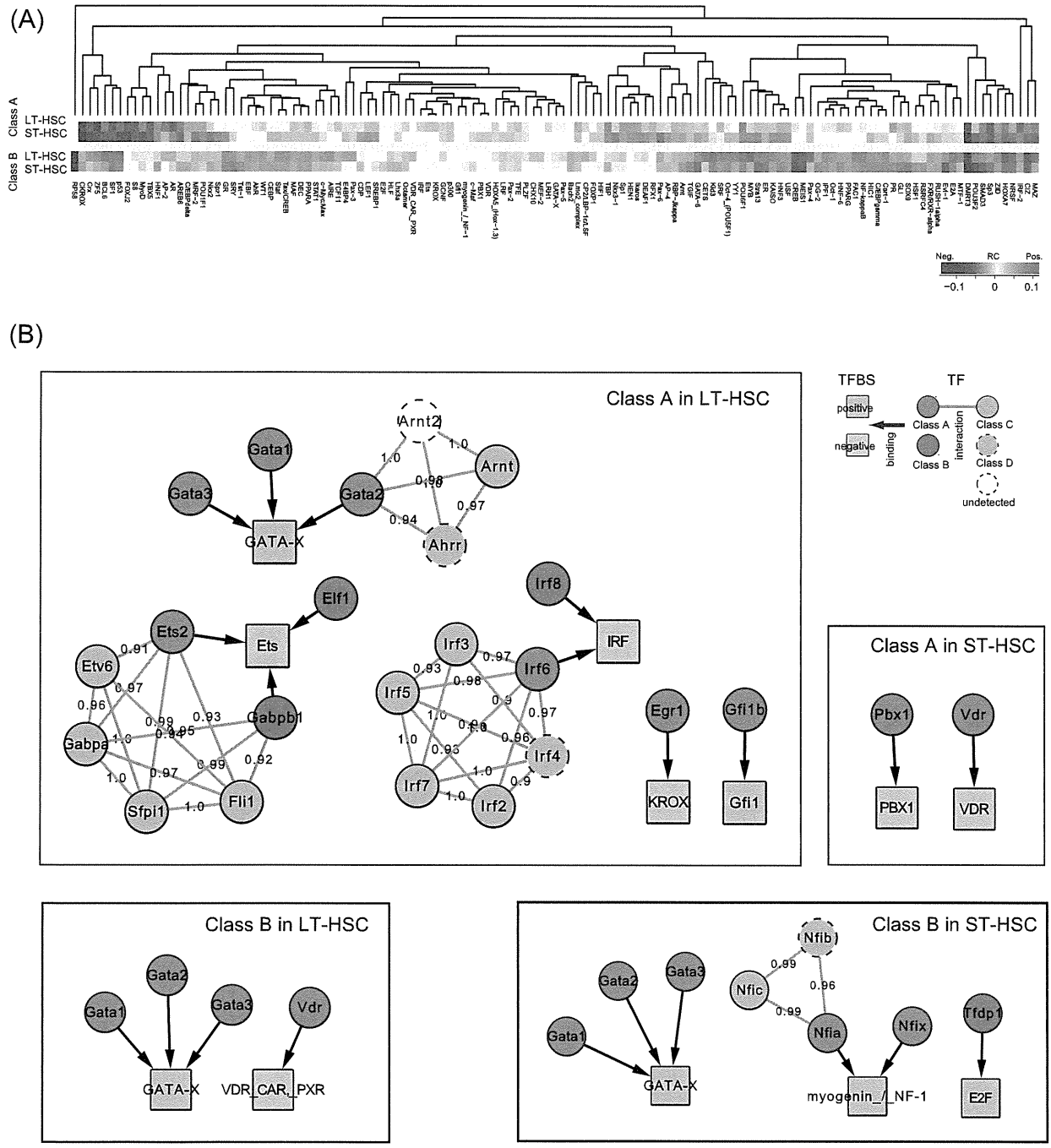

Figure 6. Alternative regulators potentially important in the presence of dysfunctional TFBSs that are targeted by differentially expressed TFs. (A) Heat map showing the regression coefficients (RCs) of 129 potentially important TFBSs ($p < 0.05$) that were identified after the removal of the TFBSs in Figures 4B and S4. The overall propensity of TFBS activities were not different from those shown in Figure 3A. (B) This removal test identified subnetworks that involve alternative TFBSs targeted by differentially expressed TFs. These included GATA-X, Ets, and IRF, which are related to erythroid/megakaryocytic lineage commitment; 6 TFBSs were targeted by 11 TFs in LT-HSCs, and 5 TFBSs were targeted by 8 TFs in ST-HSCs.

doi:10.1371/journal.pone.0093853.g006

overall propensity of the activities was not different from those shown in Figure 3A (Figure 6A).

Interestingly, specific TFBSs (e.g., GATA-X, Ets, and IRF) that were targeted by differentially expressed TFs were determined (Figure 6B). The most remarkable change was that GATA-X acquired positive activities in LT-HSCs. It is well known that GATA and AP-1 frequently co-occupy chromatin sites and that they play critical roles in cell fate decisions to commit to erythroid

vs. myeloid lineages [57,58]. More recent studies have shown that epigenetic marks control the interactions among Gata factors and other hematopoietic TFs [55], and that the DNA methylation patterns of the GATA and AP-1 motifs are mutually exclusive during early hematopoiesis [56].

Overall, our results suggest that the 24 TFs that target 21 TFBSs (Figure 4B) are key regulators of HSCs. The ST-HSCs used here exhibited lymphoid-priming features [8] with preferentially repressive potential megakaryocyte/erythroid genes (Table S7). Therefore, these regulators may be related to lymphoid-lineage development. Our model showed that dysfunctions in these regulators led to alternative regulators related to erythroid/megakaryocytic lineage development competence. This supports the recent remarkable finding of a novel lineage commitment pathway [4].

Discussion

HSC fate is controlled tightly by extrinsic and intrinsic factors [1,2,10-12,36]. The identification and characterization of these factors may lead to more effective clinical therapies for acquired and congenital blood disorders. Owing to recent advances in experimental and computational techniques, many recent studies [3,4,25] have begun to move beyond the traditional beliefs regarding hematopoiesis. However, the determination of the upstream regulatory elements that are responsible for the development of the hematopoietic system remains far from adequate and requires the application of various approaches. In the present study, we established novel transcriptome profiles from mouse LT- and ST-HSCs using an RNA-seq assay and developed a computational method for exploring the potential modes of transcriptional regulation based on these profiles.

Our RNA-seq assay confirmed the transcriptionally active state of ST-HSCs [6,7,15] with markedly high numbers of DEGs. These DEGs included 77 cell-surface molecules and 57 TFs (Tables 1 and S2–S5), which indicates that specific extrinsic and intrinsic regulators respond actively during the transition between LT- and ST-HSCs. During this transition, we observed that many previously annotated lineage-specific genes [8] were up- and downregulated (Table S7). In particular, lymphoid potential genes that preferentially undergo upregulation in ST-HSCs and potential megakaryocyte/erythroid genes had opposite patterns, suggesting that lymphoid priming occurs during this early stage.

To investigate the regulatory activities of known factors, we conducted a preliminary study using our previous method [29] and ChIP-seq data for 10 major hematopoietic regulators [17]; however, we were unable to obtain any significant results ($R<0.3$). This failure prompted us to extend our approach in the following manner (Figure 2). To approximate TFBS activities, we employed cis- and trans-regulatory information from TRANSFAC [39]. Furthermore, to consider the combinatorial regulation of TFs, we incorporated the probabilities of the conditional TF–TF interactions inferred by LLM [31]. Thus, our approach systematically inferred the regulatory activities of TFBSs, and suggested potential synergistic TF modules. Consequently, we found that motif similarity, the positional distribution of motifs, and expression changes in TFs were the most informative features for the promoter modeling of DEGs. Using LLM, we quantified the TFBS activities on the basis of the fine-tuned explanations of DEGs (TGAS V in Table 2).

Many hematopoietic TFs [6,17] were included among the transcriptional steady-state gene set (Class C), the low-level expression gene set (Class D), or the genes expressed at undetectable levels. Throughout this study, we found that the regulatory effects of these TFs and their target sites are essential to explain the regulation of DEGs. This may explain, in part, the observation that our preliminary model using 10 major hematopoietic TFs was not well fitted. We further supported this finding by performing a transplantation assay of LT-HSCs cultured with activated *Pparg* (Figure 5). Furthermore, we found that these TFs modulated differentially expressed TFs that are likely to be important during commitment to specific lineages (Figures 4B and 6B). However, LLM inferred low probabilities for interactions between known co-operative TF pairs (Tables S12 and S13), e.g., *Gata2* and *Erg* ($Pr=0.23$ in Classes A and B) and *Gata2* and *Tal1* ($Pr=0.32$ in Class A, $Pr=0.4$ in Class B), which suggests that their co-operation regulates specific gene sets.

We identified 142 TFBSs that contributed significantly to the regression models ($p<0.05$). Among these, 71 TFBSs (Class A) and 58 TFBSs (Class B) exhibited a considerable gain or loss of their activities during cell differentiation ($p<0.001$). As illustrated in Figure 4A, the effects of TFBS activities represented by plus or minus signs of RCs were mostly unchanged between cells but were inverted between DEGs. The strengths of TFBS activities increased markedly in ST-HSCs compared with LT-HSCs. We applied our method to 2 public RNA-seq datasets that represented sequential cell development (MII oocytes and two-cell embryos) and lineage commitment (megakaryocyte/erythroid precursors and megakaryocytes) (Figure S5). This analysis showed that the results of cell-lineage commitment agreed with the propensity of the regulatory activities detected in HSCs, rather than with that of sequential cell development. Therefore, regulators that play similar or different roles in accordance with cellular contexts might be general features that underlie cell fate decisions.

Overall, our results suggest that HSCs exhibit flexible and rapid responses to local needs by controlling TFs that are expressed at steady-state or low levels via a highly complex regulatory network. Further studies should consider the implications of these regulatory modes based on instructive and/or stochastic models of stem cell fate decisions. In the present study, we demonstrated that specific lineage-affiliated TFs formed a resultant set of transcriptional regulation, i.e., 24 differentially expressed TFs that contributed significantly to the model were modulated by other TFs that were not differentially expressed. These TFs include immediate early genes (e.g., *Fos*, *Jun*, and *Egr1*) that induce an early genomic response related to HSC biology [50,54]. If they become dysfunctional, LT-HSCs may be primed to an erythroid/megakaryocytic lineage via pathways that are controlled by other TFs (e.g., Gata factors, ETS family, and IRF family).

In summary, we obtained novel transcriptome data and developed a computational method for promoter modeling. Our method can be applied easily to other biological systems. Using these approaches, we identified transcriptional regulation modes that provide insights into how HSCs determine their phenotype. Future works that overcome the limitations of the present study, such as the inclusion of enhancer activities that appear to be important in hematopoiesis [17,42] and the analysis of the influence of transcriptional heterogeneity at the single-cell level [4,10,34], which can be assayed using promising techniques [59–61], would refine our findings and advance our understanding of the kinetic and regulatory aspects of stem cell biology.

Materials and Methods

Animals

All experimental protocols were reviewed and approved by the Institutional Animal Care and Use Committee of Tokyo Women's Medical University (approval ID: 13-99-2-B). Mice were purchased from Sankyo Labo Service.

Cell collection

$CD34^-KSL$ ($c\text{-}kit^+Sca1^+Lin^-$) LT-HSCs or $CD34^+KSL$ ST-HSCs were sorted, as described previously [36]. In brief, we isolated bone marrow cells from 8- to 10-week-old C57BL/6 mice and stained them with antibodies for CD34 (RAM34, eBiosciences, San Diego, CA), Sca-1 (E13-161.7, BD Biosciences Pharmingen, San Jose, CA), c-kit (2B8, BD Biosciences Pharmingen), and a lineage marker (Lineage Detection Kit, Miltenyi Biotec Inc., Bergisch Gladbach, Germany). Subsequently, we analyzed the stained cells using a MoFlo XDP cell sorter system (Beckman Coulter, Fullerton, CA).

RNA sequencing and real-time PCR

After obtaining total RNA extracts from 5000 LT- or ST-HSCs using Isogen (Nippon Gene, Tokyo, Japan) in triplicate, we synthesized cDNA using a SMARTer Pico cDNA amplification kit (Clonetech, Mountain View, CA) and amplified them with 20 cycles of PCR. Using the standard protocols for the SOLiD system, we sequenced the amplified cDNA using a SOLiD sequencer (Life Technologies, Carlsbad, CA), as described previously [36]. In the RT-PCR assay, total RNA was obtained from the sorted cells and cDNA was synthesized as described above. We performed RT-PCR using a TaqMan Gene Expression Assay (Life Technologies) for the genes indicated with the BioMark HD system (Fludigm, South San Francisco, CA).

Read mapping and quantification

We used the TopHat (v1.4.1)/Cufflinks (v.2.0.2) pipeline [33] with the sequenced reads (quality score, >15). The pipeline was coupled to Bowtie (v.0.12.7) [62]. We employed the recursive read mapping method, as described previously [32]. In brief, we applied TopHat by truncating the $3'$ ends of unmapped reads and by realigning the reads using more stringent parameters. We set the parameters empirically, which were used sequentially, as the read length, "-initial-read-mismatches", "-segment-mismatches", and "-segment-length": (50, 3, 2, 25), (46, 3, 2, 23), (42, 3, 2, 21), (38, 2, 0, 19), and (34, 2, 0, 17).

The pipeline, which quantifies RNA abundance as fragments per kilobase of exon per million mapped reads (FPKM), mapped sequenced reads to the mouse genome (mm9), and then assembled transcripts with uniquely mapped reads (uni-reads) for each replicate. We used Cuffcompare to merge all the transcript assemblies; 14,728 and 14,128 RefSeq-annotated genes in LT- and ST-HSCs, respectively. Using the merged transcript assembly, we performed Cuffdiff, which calculates FPKMs across all replicates and detects DEGs via two-group t-tests coupled to a Benjamini–Hochberg false discovery rate (FDR) procedure. We further used transcripts that satisfied the following conditions: successful deconvolution, FDR of <0.05, complete match of intron chain, and FPKM of >0.001. The mouse genome and RefSeq annotation were downloaded from http://genome.ucsc.edu/.

Long-term competitive reconstitution assay

We cultured $CD34^-KSL$ HSCs derived from C57BL/6-Ly5.1 congenic mice for 5 days with or without 20μM GW1929 (Sigma-Aldrich, St. Louis, MO) in S-Clone SF-03 medium (Sanko-Junyaku Co., Tokyo, Japan) supplemented with 0.5% bovine serum albumin (Sigma, St. Louis, MO) and 50 ng/ml mouse stem cell factor and 50 ng/ml mouse TPO (all from R&D systems, Minneapolis, MN). Next, we performed a long-term competitive reconstitution assay by transplanting cultured cells with 5×10^5 whole bone marrow competitor cells derived from C57BL/6-Ly5.2 Wt mice into lethally irradiated (9.5 Gy) C57BL/6-Ly5.2 Wt mice.

Log-linear model (LLM)

Suppose that we consider binary-stated (absence or presence) TFs $\{A, B, C\}$. The observed counts fall into 2^3-dimensional contingency table by cross-classifying the TF states. The full model (FM), which contains all the possible interactions, gives the logarithms of probabilities as follows:

$$\log p_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}, \quad (1)$$

where i, j and k are the state indices of $\{A, B, C\}$, λs are unknown parameters, λ_{ij}^{AB}, λ_{ik}^{AC} and λ_{jk}^{BC} represent the interaction effects among the indexed variables. If an instance of A is independent of B, FM can be reduced to a reduced model (RM) with respect to the hierarchy [31], which is given as follows:

$$\log p_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}. \quad (2)$$

This model can be reformulated as

$$p_{ijk} = (p_{i+k} \cdot p_{+jk})/p_{++k}, \quad (3)$$

where "+" denotes the summation over the corresponding index. This formula is equivalent to $\Pr(A=i,B=j|C=k) = \Pr(A=i|C=k).)\Pr(B=j|C=k)$, which means that A and B are independent in the conditional distribution given C ($A \perp\!\!\!\perp B|C$).

To find the most parsimonious RM, we remove an interaction term from the current model and measure two p-values for the asymptotic χ^2 test of a likelihood ratio G^2 statistic [31]. The p-values comprise p_FM, which is the difference between FM and RM, and p_RM, which is the difference between the current model and RM. We accept a removal if it yields the largest p_RM (≥ 0.01), and we terminate if any removal test yields <0.01 for either p_RM or p_FM.

Iterative random sampling for LLM

A large number of TFs can easily yield a vast dimensional contingency table. To find a near optimal parsimonious model even in such higher-dimensional space, we designed an iterative sampling scheme that allowed us to calculate interaction probability Pr as follows.

Let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is an undirected graph, where $\mathcal{V}$ is a finite set of vertices (TFs) and $\mathcal{E}$ is a set of edges, which represent the interactions between vertex pairs. The scheme is as follows.

1. 1. $\mathcal{S} = \{s_1, \ldots, s_k\}$, a nonredundant combination of TFs, is selected randomly from all TFs ($k = 10$ in the present study).
2. For all possible vertex pairs (s_i, s_j), the trial number $ntry_{ij}$ of an edge between s_i and s_j is counted (i.e., FM of k variables).

3. LLM infers the best model $\mathcal{G}' = (\mathcal{S}, \mathcal{E}')$, where $\mathcal{E}'$ is a set of edges that represents TF–TF interactions.
4. For all possible vertex pairs (s_i, s_j), if an edge in $\mathcal{E}'$ links a pair (s_i, s_j), the observed edge frequency $nobs_{ij}$ for this pair is counted.
5. For all possible vertex pairs (s_i, s_j), the interaction probability Pr for a pair (s_i, s_j) is updated using $nobs_{ij}/ntry_{ij}$.
6. If $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E}$ is a set of edges ($Pr = 1.0$), is not changed with a large number of samplings ($= 100{,}000$); therefore, this procedure is terminated. Otherwise, steps 1–5 are repeated.

Linear regression model

We used a multivariate regression model

$$\log Y_i = \sum_j w_j X_{ij} + e_i, \tag{4}$$

$$X_{ij} = \sum_k x_k, \tag{5}$$

where Y_i is the expression of gene i, X_{ij} is TGAS of the jth TFBS in the promoter region of gene i, w_j is RC of the jth TFBS, and e_i is the error term. TGAS is the sum of scores x_k, where k represents the position of the jth TFBS in promoter i. We tested the following forms of x_k.

- I: matrix similarity s of TFBS j scored using MATCH[43] ($x_k = s_k$).
- II: TGAS I modified by a location-dependent weight L,

$$x_k = s_k \times L_k. \tag{6}$$

- III: TGAS II weighted by the expression fold change (F) of TFs,

$$x_k = s_k \times L_k \times \sum_{k'} F_{k'}, \tag{7}$$

where k' is the index of TFs binding to TFBS j. If FPKM for TF is ≤ 3, we use $F = 1$.

- IV: the same as TGAS III, but we removed TFBSs where none of the TFs had FPKM of > 3.
- V: TGAS III weighted using both Fs of interactive TFs and the interaction probability Pr estimated by LLM,

$$x_k = s_k \times L_k \times \left(\sum_{k'} F_{k'} + I_{k'}\right) \tag{8}$$

$$I_{k'} = \sum_{l'=1}^{k'} \sum_{j'>i'}^{k'} F_{i'} F_{j'} Pr_{i'j'}. \tag{9}$$

We used a published method to calculate L [40]. First, we calculated the distribution of TFBS j in bins ($= 500$ bp) of promoter regions and created a histogram H_{real}. Next, we randomized the positions of TFBS j and created a histogram H_{rand}. L for the kth TFBS j is given by the following:

$$L_k = \begin{cases} 0, & if\ H_{real}(m) < H_{rand}(m) \\ \dfrac{H_{real}(m) - H_{rand}(m)}{H_{real}(m)}, & if\ H_{real}(m) \geq H_{rand}(m), \end{cases} \tag{10}$$

where m represents the index of bin that corresponds to the position of the kth TFBS j. This location-dependent weight takes a value between 0 and 1, where a higher weight implies nonrandom occurrence.

Stepwise selection of the regression model

We built a regression model with the explanatory variable X and then reduced the model using AIC. Let the reduced model be Y' with X'. $X - X' = \{x_1, x_2, \ldots\}$ is the variables removed on the basis of AIC. V is the set of all pairwise terms of $x_i x_j$ ($i \neq j$). We searched any elements of V that improve Pearson's correlation coefficient r of 5-fold CV on testing datasets.

1. Randomly select v_i ($\in V$) and add it to X', which yields X''.
2. Perform 5-fold CV with X'' and calculate the averaged r on testing datasets.
3. If the r has been improved, update X'' to X'.
4. Repeat step 1–3 until all v_i have been tested.
5. Calculate Pearson's correlation coefficient R between observed and predicted FPKMs of all genes by using the final model.

We run this procedure 100 times using different random seeds. The final R is referred to as a model quality in this study.

Bioinformatics analysis

We obtained array-based gene expression profiles [8,9] from BloodExpress [63], RNA-seq data for megakaryocyte/erythroid precursors and megakaryocytes from http://genome.ucsc.edu/encode/, and RNA-seq data for MII oocytes and two-cell embryos from DDBJ DRA001066. The public RNA-seq datasets were analyzed using the pipeline mentioned above. To search putative TFBSs and TFs in TRANSFAC professional (released in January 2013) [39], we prepared ± 5kb DNA sequences from transcription start sites (TSSs) annotated in RefSeq (http://www.ncbi.nlm.nih.gov/refseq/), and applied the MATCH tool in the minimize false-positive mode [43].

To analyze the enriched GO terms, we used the DAVID Bioinformatics Resources [35]. Significant terms detected by DAVID (EASE score, a modified Fisher's exact p-value, < 0.01) were grouped into representative ancestor terms in the dataset GO Slim2 using CateGOrizer [64]. We used the R programming language (http://www.r-project.org/) for regression modeling and to perform statistical tests. Although all p-values were adjusted by Bonferroni correction (Tables S6 and S8–S11), we used uncorrected p-values throughout this study to avoid too conservative interpretation that would reduce biologically meaningful findings.

Data access

The RNA-seq data generated in this study have been deposited in the DDBJ (DNA Data Bank of Japan) Sequence Read Archive (DRA) under accession number DRA001213. The online version of LLM is available at http://dbtmee.hgc.jp/tools/.

Supporting Information

Figure S1 Correlation analysis of gene expression levels measured using RNA-seq assays. (A) Reproducibility based on triplicate analyses of LT- and ST-HSCs. (B) Comparison of the gene expression correlations in the present study to those reported by Karlsson et al. [15], who purified HSCs using $CD48^-$, $CD150^+$, $CD34^-$, $CD9^{high}$ KSL for LT-HSCs and $CD48^-$, $CD150^+$, $CD9^{low}$ KSL for ST-HSCs.
(EPS)

Figure S2 Contribution of higher-order TF interaction scores estimated by LLM. (A) Statistical differences of 2 regression coefficient (RC) ensembles of a TFBS found commonly by TGAS III and V (two-sample *t*-test). (B) Distribution of the TF interaction score I_k in Equation 9.
(EPS)

Figure S3 Box plots of RCs estimated by 100 iterations of regression modeling with TGAS V. Pos and Neg represent the positive (red) and negative (blue) mean values of RCs (red line), respectively.
(EPS)

Figure S4 Subnetworks involved in ST-HSC regulation. Although the majority of TF-coding genes found in ST-HSCs (Figure 4A) were not differentially expressed, 26 differentially expressed TFs that putatively bind to 21 TFBSs were present among DEGs (Class A and Class B).
(EPS)

Figure S5 Propensity of the TFBS activities inferred from public RNA-seq datasets. We applied our method to public RNA-seq datasets related to sequential cell development (A) and lineage commitment (C). Our procedure evaluates the averaged R of 5-fold CV on testing datasets (blue line). If a model improved R in testing, the model was accepted and its R value between the observed and predicted gene expression of all genes was measured (red line). (B) Of 147 TFBSs ($p<0.05$), 67 TFBSs (Class A; upregulated in Oo) and 80 TFBSs (Class B; upregulated in 2C) exhibited significant gains and losses of activity ($p<0.001$). In addition, 73% (49/67) of Class A and 52.5% (42/80) of Class B genes exhibited no changes in the effects of their TFBS activities between cells, i.e., positive (negative) in Oo was still positive (negative) in 2C. We found that 16% (8/49) of Class A and 83% (35/42) of Class B genes had increased activities in 2C compared with Oo. (D) Among 150 TFBSs ($p<0.05$), 98 TFBSs (Class A, upregulated in MEP) and 114 TFBSs (Class B, upregulated in Mk) exhibited significant gains and losses of activity ($p<0.001$). We also found that 83% (81/98) of Class A and 76% (87/114) of Class B genes exhibited no changes in the effects of their TFBS activities. All of the TFBSs in both classes exhibited increases in the strengths of their activities in Mk compared with MEP. R, Pearson's correlation coefficient; Oo, MII oocytes; 2C, 2-cell embryo; MEP, megakaryocyte/erythroid precursor; Mk, megakaryocyte.
(EPS)

Table S1 RNA-seq mapping statistics.
(XLSX)

Table S2 Differentially expressed cell-surface molecules.
(XLSX)

Table S3 Differentially expressed transcription factors.
(XLSX)

Table S4 Transcription factors categorized into Class C.
(XLSX)

Table S5 Low expressed transcription factors (Class D).
(XLSX)

Table S6 Average regression coefficient of 142 TFBSs.
(XLSX)

Table S7 Classification of MkE, GM, and Lymphoid-associated genes.
(XLSX)

Table S8 TFBSs significantly different in the regression coefficient between LT- and ST-HSCs (Class A).
(XLSX)

Table S9 TFBSs significantly different in the regression coefficient between LT- and ST-HSCs (Class B).
(XLSX)

Table S10 Enriched GO terms in Class A.
(XLSX)

Table S11 Enriched GO terms in Class B.
(XLSX)

Table S12 Result of log-linear model in Class A.
(XLSX)

Table S13 Result of log-linear model in Class B.
(XLSX)

Acknowledgments

We thank Drs S. Mitani and T. Furukawa for their great help with RNA sequencing. Computational resources were provided by the supercomputer system at Human Genome Center, the Institute of Medical Science, the University of Tokyo.

Author Contributions

Conceived and designed the experiments: SJP KN. Performed the experiments: TU YS MY. Analyzed the data: SJP MSA. Contributed reagents/materials/analysis tools: TU SJP. Wrote the paper: SJP.

References

1. Hoang T (2004) The origin of hematopoietic cell type diversity. Oncogene 23: 7188–98.
2. Forsberg EC, Bhattacharya D, Weissman IL (2006) Hematopoietic stem cells: expression profiling and beyond. Stem Cell Rev 2: 23–30.
3. Sanjuan-Pla A, Macaulay IC, Jensen CT, Woll PS, Luis TC, et al. (2013) Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. Nature 502: 232–6.
4. Yamamoto R, Morita Y, Ooehara J, Hamanaka S, Onodera M, et al. (2013) Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell 154: 1112–26.
5. Passegue E, Wagers AJ, Giuriato S, Anderson WC, Weissman IL (2005) Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. J Exp Med 202: 1599–611.
6. Forsberg EC, Prohaska SS, Katzman S, Heffner GC, Stuart JM, et al. (2005) Differential expression of novel potential regulators in hematopoietic stem cells. PLoS Genet 1: e28.
7. Zhong JF, Zhao Y, Sutton S, Su A, Zhan Y, et al. (2005) Gene expression profile of murine longterm reconstituting vs. short-term reconstituting hematopoietic stem cells. Proc Natl Acad Sci U S A 102: 2448–53.
8. Mansson R, Hultquist A, Luc S, Yang L, Anderson K, et al. (2007) Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. Immunity 26: 407–19.
9. Ficara F, Murphy MJ, Lin M, Cleary ML (2008) Pbx1 regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence. Cell Stem Cell 2: 484–96.

10. Kent DG, Copley MR, Benz C, Wohrer S, Dykstra BJ, et al. (2009) Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. Blood 113: 6342–50.
11. Chotinantakul K, Leeanansaksiri W (2012) Hematopoietic stem cell development, niches, and signaling pathways. Bone Marrow Res 2012: 270425.
12. Kunisaki Y, Bruns I, Scheiermann C, Ahmed J, Pinho S, et al. (2013) Arteriolar niches maintain haematopoietic stem cell quiescence. Nature 502: 637–43.
13. Liu P, Barb J, Woodhouse K, Taylor JG, Munson PJ, et al. (2011) Transcriptome profiling and sequencing of differentiated human hematopoietic stem cells reveal lineage-specific expression and alternative splicing of genes. Physiol Genomics 43: 1117–34.
14. Lu R, Neff NF, Quake SR, Weissman IL (2011) Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nat Biotechnol 29: 928–33.
15. Karlsson G, Rorby E, Pina C, Soneji S, Reckzeh K, et al. (2013) The tetraspanin cd9 affords high-purity capture of all murine hematopoietic stem cells. Cell Rep 4: 642–8.
16. Weishaupt H, Sigvardsson M, Attema JL (2010) Epigenetic chromatin states uniquely define the developmental plasticity of murine hematopoietic stem cells. Blood 115: 247–56.
17. Wilson NK, Foster SD, Wang X, Knezevic K, Schutte J, et al. (2010) Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell 7: 532–44.
18. Bissels U, Bosio A, Wagner W (2012) Micrornas are shaping the hematopoietic landscape. Haematologica 97: 160–7.
19. Whichard ZL, Sarkar CA, Kimmel M, Corey SJ (2010) Hematopoiesis and its disorders: a systems biology approach. Blood 115: 2339–47.
20. Bonzanni N, Garg A, Feenstra KA, Schutte J, Kinston S, et al. (2013) Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. Bioinformatics 29: i80–8.
21. Hannah R, Joshi A, Wilson NK, Kinston S, Gottgens B (2011) A compendium of genome-wide hematopoietic transcription factor maps supports the identification of gene regulatory control mechanisms. Exp Hematol 39: 531–41.
22. Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, et al. (2011) Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell 144: 296–309.
23. Will B, Vogler TO, Bartholdy B, Garrett-Bakelman F, Mayer J, et al. (2013) Satb1 regulates the self-renewal of hematopoietic stem cells by promoting quiescence and repressing differentiation commitment. Nat Immunol 14: 437–45.
24. Mirshekar-Syahkal B, Haak E, Kimber GM, van Leusden K, Harvey K, et al. (2013) Dlk1 is a negative regulator of emerging hematopoietic stem and progenitor cells. Haematologica 98: 163–71.
25. Gazit R, Garrison B, Rao T, Shay T, Costello J, et al. (2013) Transcriptome analysis identifies regulators of hematopoietic stem and progenitor cells. Stem Cell Reports 1: 266–280.
26. Osawa M, Hanada K, Hamada H, Nakauchi H (1996) Long-term lymphohematopoietic reconstitution by a single cd34-low/negative hematopoietic stem cell. Science 273: 242–5.
27. Ema H, Morita Y, Yamazaki S, Matsubara A, Seita J, et al. (2006) Adult mouse hematopoietic stem cells: purification and single-cell assays. Nat Protoc 1: 2979–87.
28. Bussemaker HJ, Foat BC, Ward LD (2007) Predictive modeling of genome-wide mrna expression: from modules to molecules. Annu Rev Biophys Biomol Struct 36: 329–47.
29. Park SJ, Nakai K (2011) A regression analysis of gene expression in es cells reveals two gene classes that are significantly different in epigenetic patterns. BMC Bioinformatics 12 Suppl 1: S50.
30. Irie T, Park SJ, Yamashita R, Seki M, Yada T, et al. (2011) Predicting promoter activities of primary human dna sequences. Nucleic Acids Res 39: e75.
31. Lauritzen S (1996) Graphical Models. New York: Oxford University Press.
32. Park SJ, Komata M, Inoue F, Yamada K, Nakai K, et al. (2013) Inferring the choreography of parental genomes during fertilization from ultralarge-scale whole-transcriptome analysis. Genes Dev 27: 2736–48.
33. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, et al. (2012) Differential gene and transcript expression analysis of rna-seq experiments with tophat and cufflinks. Nat Protoc 7: 562–78.
34. Benz C, Copley MR, Kent DG, Wohrer S, Cortes A, et al. (2012) Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. Cell Stem Cell 10: 273–83.
35. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using david bioinformatics resources. Nat Protoc 4: 44–57.
36. Umemoto T, Yamato M, Ishihara J, Shiratsuchi Y, Utsumi M, et al. (2012) Integrin-alphavbeta3 regulates thrombopoietin-mediated maintenance of hematopoietic stem cells. Blood 119: 83–94.
37. Domen J, Cheshier SH,Weissman IL (2000) The role of apoptosis in the regulation of hematopoietic stem cells: Overexpression of bcl-2 increases both their number and repopulation potential. J Exp Med 191: 253–64.
38. Peng C, Chen Y, Shan Y, Zhang H, Guo Z, et al. (2012) Lsk derived lsk- cells have a high apoptotic rate related to survival regulation of hematopoietic and leukemic stem cells. PLoS One 7: e38614.
39. Wingender E, Chen X, Hehl R, Karas H, Liebich I, et al. (2000) Transfac: an integrated system for gene expression regulation. Nucleic Acids Res 28: 316–9.
40. Chen X, Xu H, Yuan P, Fang F, Huss M, et al. (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133: 1106–17.
41. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, et al. (2008) Genome-scale dna methylation maps of pluripotent and differentiated cells. Nature 454: 766–70.
42. Moignard V, Macaulay IC, Swiers G, Buettner F, Schutte J, et al. (2013) Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. Nat Cell Biol 15: 363–72.
43. Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, et al. (2003) Match: A tool for searching transcription factor binding sites in dna sequences. Nucleic Acids Res 31: 3576–9.
44. Boyle P, Despres C (2010) Dual-function transcription factors and their entourage: unique and unifying themes governing two pathogenesis-related genes. Plant Signal Behav 5: 629–34.
45. Whitfield TW, Wang J, Collins PJ, Partridge EC, Aldred SF, et al. (2012) Functional analysis of transcription factor binding sites in human promoters. Genome Biol 13: R50.
46. Diffner E, Beck D, Gudgin E, Thoms JA, Knezevic K, et al. (2013) Activity of a heptad of transcription factors is associated with stem cell programs and clinical outcome in acute myeloid leukemia. Blood 121: 2289–300.
47. Chute JP, Ross JR, McDonnell DP (2010) Minireview: Nuclear receptors, hematopoiesis, and stem cells. Mol Endocrinol 24: 1–10.
48. Henke BR, Blanchard SG, Brackeen MF, Brown KK, Cobb JE, et al. (1998) N-(2-benzoylphenyl)-l-tyrosine ppargamma agonists. 1. discovery of a novel series of potent antihyperglycemic and antihyperlipidemic agents. J Med Chem 41: 5020–36.
49. Brown KK, Henke BR, Blanchard SG, Cobb JE, Mook R, et al. (1999) A novel n-aryl tyrosine activator of peroxisome proliferator-activated receptor-gamma reverses the diabetic phenotype of the zucker diabetic fatty rat. Diabetes 48: 1415–24.
50. Liebermann DA, Gregory B, Hoffman B (1998) Ap-1 (fos/jun) transcription factors in hematopoietic differentiation and apoptosis. Int J Oncol 12: 685–700.
51. Shen LJ, Chen FY, Zhang Y, Cao LF, Kuang Y, et al. (2013) Myen transgenic zebrafish model with the characterization of acute myeloid leukemia and altered hematopoiesis. PLoS One 8: e59070.
52. Yu S, Jing X, Colgan JD, Zhao DM, Xue HH (2012) Targeting tetramer-forming gabpbeta isoforms impairs self-renewal of hematopoietic and leukemic stem cells. Cell Stem Cell 11: 207–19.
53. Ghiaur G, Yegnasubramanian S, Perkins B, Gucwa JL, Gerber JM, et al. (2013) Regulation of human hematopoietic stem cell self-renewal by the microenvironment's control of retinoic acid signaling. Proc Natl Acad Sci U S A.
54. Okada S, Fukuda T, Inada K, Tokuhisa T (1999) Prolonged expression of c-fos suppresses cell cycle entry of dormant hematopoietic stem cells. Blood 93: 816–25.
55. Wozniak RJ, Keles S, Lugus JJ, Young KH, Boyer ME, et al. (2008) Molecular hallmarks of endogenous chromatin complexes containing master regulators of hematopoiesis. Mol Cell Biol 28: 6681–94.
56. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, et al. (2013) Charting a dynamic dna methylation landscape of the human genome. Nature 500: 477–81.
57. Kawana M, Lee ME, Quertermous EE, Quertermous T (1995) Cooperative interaction of gata-2 and ap1 regulates transcription of the endothelin-1 gene. Mol Cell Biol 15: 4225–31.
58. Zhang P, Behre G, Pan J, Iwama A, Wara-Aswapati N, et al. (1999) Negative cross-talk between hematopoietic regulators: Gata proteins repress pu.1. Proc Natl Acad Sci U S A 96: 8705–10.
59. Peixoto A, Monteiro M, Rocha B, Veiga-Fernandes H (2004) Quantification of multiple gene expression in individual cells. Genome Res 14: 1938–47.
60. Warren L, Bryder D, Weissman IL, Quake SR (2006) Transcription factor profiling in individual hematopoietic progenitors by digital rt-pcr. Proc Natl Acad Sci U S A 103: 17807–12.
61. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, et al. (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 498: 236–40.
62. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short dna sequences to the human genome. Genome Biol 10: R25.
63. Miranda-Saavedra D, De S, Trotter MW, Teichmann SA, Gottgens B (2009) Bloodexpress: a database of gene expression in mouse haematopoiesis. Nucleic Acids Res 37: D873–9.
64. Hu ZL, Bao J, Reecy J (2008) Categorizer: A web-based program to batch analyze gene ontology classification categories. Online J Bioinform 9: 108–12.