Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis

Noa Novershtern,[1,2,3,11] Aravind Subramanian,[1,11] Lee N. Lawton,[4] Raymond H. Mak,[1] W. Nicholas Haining,[5]
Marie E. McConkey,[6] Naomi Habib,[3] Nir Yosef,[1] Cindy Y. Chang,[1,6] Tal Shay,[1] Garrett M. Frampton,[2,4] Adam C.B. Drake,[2,7]
Ilya Leskov,[2,7] Bjorn Nilsson,[1,6] Fred Preffer,[8] David Dombkowski,[8] John W. Evans,[5] Ted Liefeld,[1] John S. Smutko,[9]
Jianzhu Chen,[2,7] Nir Friedman,[3] Richard A. Young,[2,4] Todd R. Golub,[1,5,10] Aviv Regev,[1,2,10,12,*]
and Benjamin L. Ebert[1,5,6,12,*]
[1]Broad Institute, 7 Cambridge Center, Cambridge MA, 02142, USA
[2]Department of Biology, Massachusetts Institute of Technology, Cambridge MA, 02140, USA
[3]School of Computer Science, Hebrew University, Jerusalem 91904, Israel
[4]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA
[5]Dana-Farber Cancer Institute, Boston, MA 02115, USA
[6]Brigham and Women's Hospital, Boston, MA 02115, USA
[7]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139
[8]Massachusetts General Hospital, Boston, MA 02114, USA
[9]Nugen Technologies, San Carlos, CA 94070, USA
[10]Howard Hughes Medical Institute, Chevy Chase, MD 20815-6789, USA
[11]These authors contributed equally to this work
[12]These authors contributed equally to this work
*Correspondence: aregev@broad.mit.edu (A.R.), bebert@partners.org (B.L.E.)
DOI 10.1016/j.cell.2011.01.004

SUMMARY

Though many individual transcription factors are known to regulate hematopoietic differentiation, major aspects of the global architecture of hematopoiesis remain unknown. Here, we profiled gene expression in 38 distinct purified populations of human hematopoietic cells and used probabilistic models of gene expression and analysis of *cis*-elements in gene promoters to decipher the general organization of their regulatory circuitry. We identified modules of highly coexpressed genes, some of which are restricted to a single lineage but most of which are expressed at variable levels across multiple lineages. We found densely interconnected *cis*-regulatory circuits and a large number of transcription factors that are differentially expressed across hematopoietic states. These findings suggest a more complex regulatory system for hematopoiesis than previously assumed.

INTRODUCTION

Hematopoiesis is an ideal model for the study of multilineage differentiation in humans. More than 2×10^{11} hematopoietic cells from at least 11 lineages are produced daily in humans from a small pool of self-renewing adult stem cells (Quesenberry and Colvin, 2005). Production of each cell type is highly regulated and responsive to environmental stimuli. Mutations or aberrant expression of regulatory proteins cause both benign and malignant hematologic disorders.

The hematopoietic system is also well suited for an analysis of the global architecture of the molecular circuits controlling human cellular differentiation. Hematopoietic stem cells, progenitor cells, and terminally differentiated cells can be isolated using flow cytometry. Moreover, many aspects of hematopoietic differentiation can be recapitulated in vitro. Finally, high-speed multiparameter flow cytometry and cDNA amplification procedures allow us to purify and profile gene expression from rare subpopulations (Ebert and Golub, 2004).

A dominant model of hematopoiesis posits that it is controlled by a hierarchy of a relatively small number of critical transcription factors (TFs) that are sequentially expressed, are largely restricted to a specific lineage, and can interact directly to mediate and reinforce cell fate decisions (Iwasaki and Akashi, 2007). Genetically engineered mice have been used to map the maturation stage at which key TFs are essential (Orkin and Zon, 2008).

Recent genome-wide studies suggest a more complex architecture in regulatory circuits involving larger numbers of TFs that control different combinations of modules of coexpressed genes (Amit et al., 2009; Suzuki et al., 2009). Complex circuits with a larger number of TFs than previously assumed, each with a major regulatory effect, are emerging from studies in immune cell types (Amit et al., 2009; Suzuki et al., 2009), stem cell populations (Müller et al., 2008), and cell differentiation in invertebrates (Davidson, 2001).

These two views leave open several key questions in understanding the regulatory architecture of human hematopoiesis. (1) Are distinct hematopoietic cell states characterized mostly
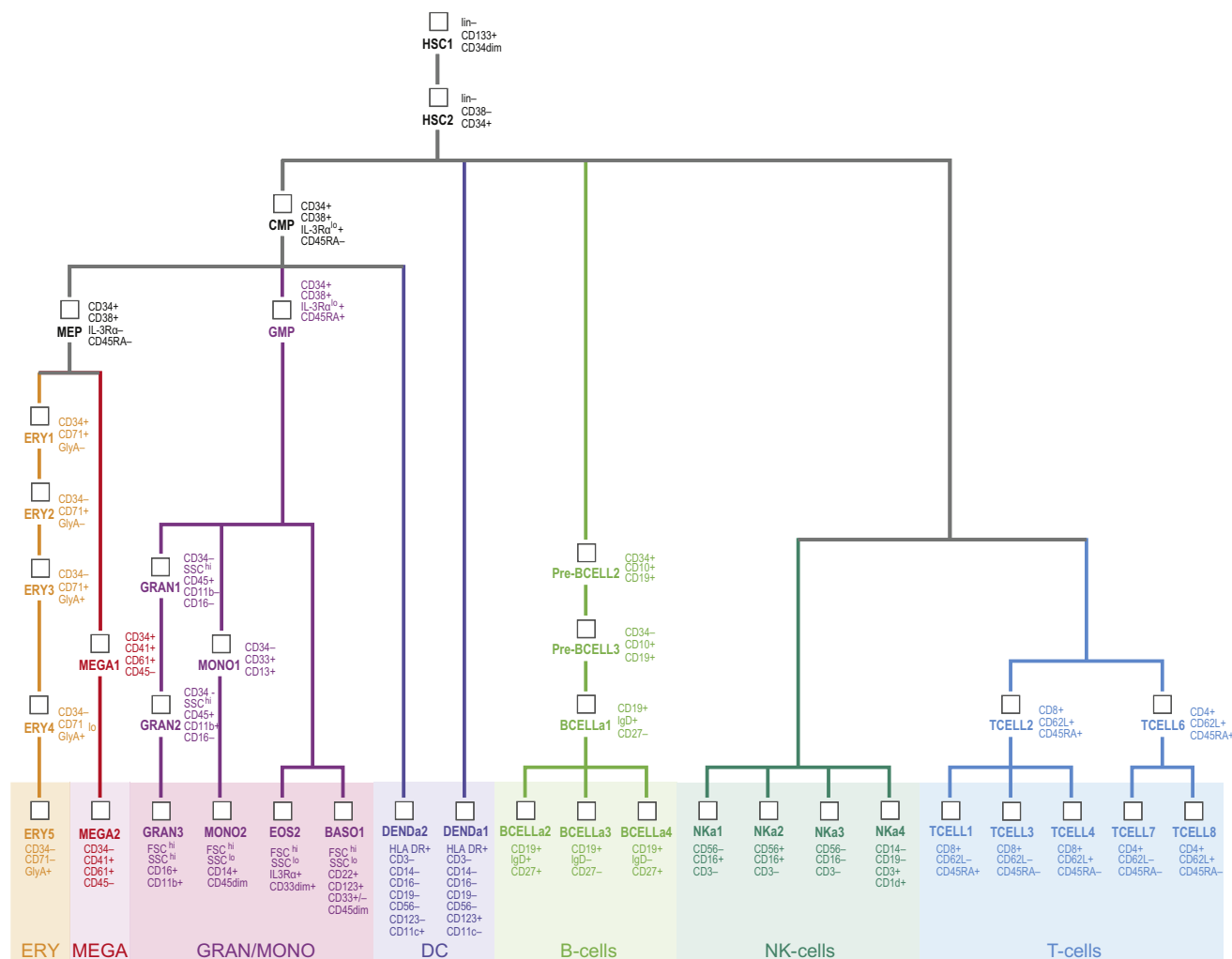
Figure 1. Hematopoietic Differentiation

The 38 hematopoietic cell populations purified by flow sorting and analyzed by gene expression profiling are illustrated in their respective positions in hematopoiesis. (Gray) Hematopoietic stem cell (HSC1,2), common myeloid progenitor (CMP), megakaryocyte/erythroid progenitor (MEP). (Orange) Erythroid cells (ERY1–5). (Red) CFU-MK (MEGA1) and megakaryocyte (MEGA2). (Purple) Granulocyte/monocyte progenitor (GMP), CFU-G (GRAN1), neutrophilic metamyelocyte (GRAN2), neutrophil (GRAN3), CFU-M (MONO1), monocytes (MONO2), eosinophil (EOS), and basophil (BASO). (Blue) Myeloid dendritic cell (DENDa2) and plasmacytoid dendritic cell (DENDa1). (Light green) Early B cell (Pre-BCELL2), pro-B cell (Pre-BCELL3), naive B cell (BCELLa1), mature B cell, class able to switch (BCELLa2), mature B cell (BCELLa3), and mature B cell, class switched (BCELLa4). (Dark green) Mature NK cell (NK1–4). (Turquoise) Naive CD8+ T cell (TCELL2), CD8+ effector memory RA (TCELL1), CD8+ effector memory (TCELL3), CD8+ central memory (TCELL4), naive CD4+ T cell (TCELL6), CD4+ effector memory (TCELL7), and CD4+ central memory (TCELL8). See Table S1 for markers information.

by induction of lineage-specific genes or by a unique combination of modules, wherein the distinct capacities of each cell type are largely determined through the reuse of modules? (2) Is hematopoiesis determined solely by a few master regulators, or does it involve a more complex network with a larger number of factors? (3) What are the regulatory mechanisms that maintain cell state in the hematopoietic system, and how do they change as cells differentiate?

Here, we measured mRNA profiles in 38 prospectively purified cell populations, from hematopoietic stem cells, through multiple progenitor and intermediate maturation states, to 12 terminally differentiated cell types (Figure 1). We found distinct, tightly integrated, regulatory circuits in hematopoietic stem cells and

differentiated cells, implicated dozens of new regulators in hematopoiesis, and demonstrated a substantial reuse of gene modules and their regulatory programs in distinct lineages. We validated our findings by experimentally determining the binding sites of four TFs in hematopoietic stem cells, by examining the expression of a set of 33 TFs in erythroid and myelomonocytic differentiation in vitro, and by investigating the function of 17 of these TFs using RNA interference. Our data provide strong evidence for the role of complex interconnected circuits in hematopoiesis and for "anticipatory binding" to the promoters of their target genes in hematopoietic stem cells. Our data set and analyses will serve as a comprehensive resource for the study of gene regulation in hematopoiesis and differentiation.
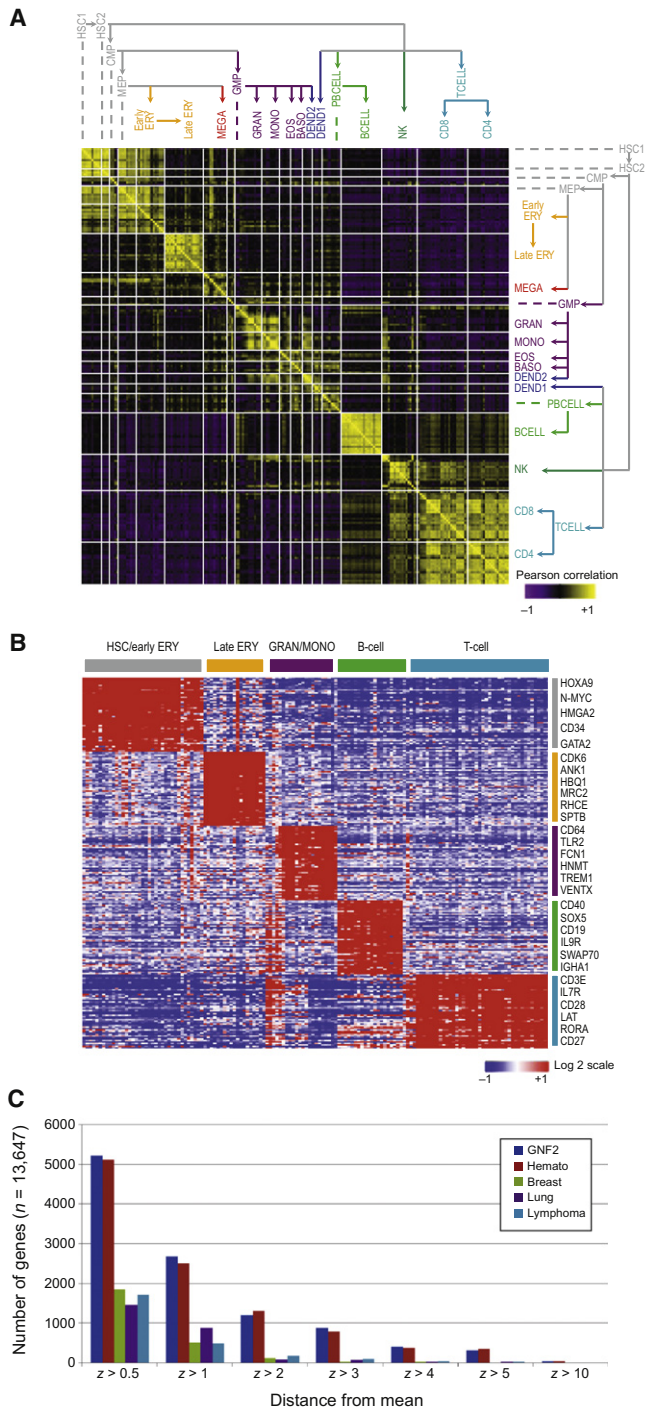
RESULTS

An Expression Map of Hematopoiesis Reveals Cell State-Specific Profiles

We defined 38 distinct cell states based on cell surface marker expression, representing hematopoietic stem and progenitor cells, terminally differentiated cells, and intermediate states (Figure 1 and Table S1 available online). For each state, we purified samples separately from four to seven independent donors by multiparameter flow cytometry (Experimental Procedures), yielding 211 samples. Cells from all stem and progenitor populations were purified from umbilical cord blood. Terminally differentiated lymphocyte populations were purified from peripheral blood, as terminal differentiation is completed in these cells upon exposure to antigens after birth (Table S1). In all cases, cells were harvested fresh and were processed and sorted immediately. We isolated mRNA from each cell type and measured expression profiles using Affymetrix microarrays (Experimental Procedures).

The global transcriptional profiles are consistent with the established topology of hematopoietic differentiation. Replicate samples from a single state but different donors and samples from multiple states within a lineage are highly correlated with each other, and profiles from related lineages are also similar (Figure 2A). Of note, hematopoietic stem cell (HSC) samples do not form a separate cluster but are highly similar to early progenitors in the megakaryocyte/erythrocytes lineage (MEGA/ERY), suggesting that their transcriptional state is largely maintained in some of the early progenitors. These findings are also apparent in a systematic unsupervised analysis using nonnegative matrix factorization (NMF) (Brunet et al., 2004) (Figure S1A) and hierarchical clustering (Figure S1B). We further validated our data set by confirming that previously published lineage-specific gene signatures are significantly enriched in the expected lineage compared to other lineages (FDR < 0.25; Figure S1C and Extended Experimental Procedures).

Unique and Complex Gene Signatures Characterize Distinct Hematopoietic Lineages

In a supervised analysis, we found that each of the five dominant states in our data set—HSPCs, differentiated erythroid cells, granulocytes/monocytes B cells, and T cells—is distinguished by a set of significantly differentially expressed genes specific to each lineage as compared to the others (Figure 2B and Table S2). Some of these genes are expressed more than 100-fold higher in one cell type (e.g., granzyme genes in certain T cell and NK cell populations, PROM1 [CD133 antigen], and HOXA9 in stem and progenitor cells).

Figure 2. A Transcriptional Map of Hematopoietic Differentiation Identifies Lineage-Specific Transcription

(A) Similarity in global expression profiles between proximate differentiation states. The heat map shows the pairwise Pearson correlation coefficients between all 211 samples ordered according to the differentiation tree (right and top). A positive correlation is portrayed in yellow and a negative correlation in purple.

(B) Signature genes characterizing the five main lineages. Expression levels are shown for the top 50 marker genes (rows) for each of four major lineages plus hematopoietic stem and progenitor cells. High relative expression is

shown in red and low relative expression in blue; the expression of each gene is normalized to a mean expression of zero across all the samples; labels as in Figure 1. Genes were selected by high expression in one lineage compared to the others (t test).

(C) The number of genes that are differentially expressed, according to an outlier statistic, was calculated for all hematopoietic cell states profiled (red); a compendium of 79 tissues in the GNF atlas (Su et al., 2004) (blue); and data sets of lymphomas (Monti et al., 2005) (turquoise), lung cancers (Bhattacharjee et al., 2001) (purple), and breast cancers (Chin et al., 2006) (green).
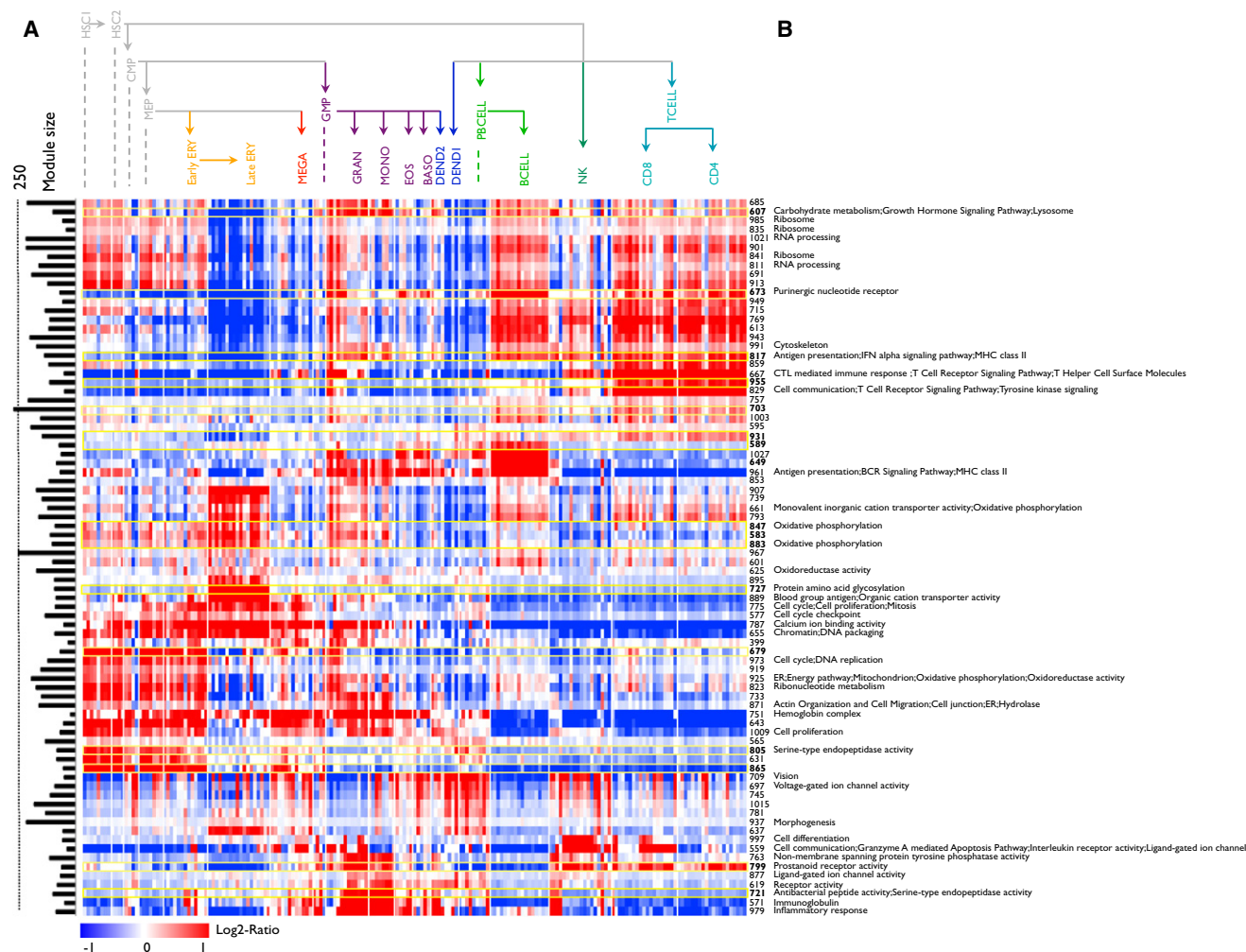
See also Figure S1.

Figure 3. Expression Pattern and Functional Enrichment of 80 Transcriptional Modules

(A) Average expression levels of 80 gene modules. Shown is the average expression pattern of the gene members in each of the 80 modules (rows) across all 211 samples (columns). Colors and normalization as in Figure 2B. The samples are organized according to the differentiation tree topology (top) with abbreviations as in Figure 1. The number of genes in each module is shown in the bar graph (left). The expression profiles of a few example modules discussed in the text are highlighted by vertical yellow lines. The expression of individual genes in each module is shown in Figure S2.

(B) Functional enrichment in gene modules. Functional categories with enriched representation (FDR < 5%) in at least one module are portrayed. Categories were selected for broad representation. The complete list appears in Table S3.

See also Figure S2 and Figure S7.

The signature genes are enriched for molecular functions and biological processes consistent with the functional differences between lineages (Figure S1D and Table S2). Of note, a set of 16 genes comprised of the 5′ partners of known translocations in leukemias (Mitelman et al., 2010) is enriched in the HSPC population (p < 0.013). This suggests that the 5′ partners of leukemia-causing translocations, containing the promoters of the fusion genes, tend to be selectively expressed in stem and progenitor cell populations.

The diversity of gene expression across hematopoietic lineages is comparable to the diversity in gene expression observed across a host of human tissue types. The number of genes that are differentially expressed throughout our hematopoiesis data set (outlier analysis) (Tibshirani and Hastie, 2007) (Extended

Experimental Procedures) is comparable to that determined for an atlas of 79 different human tissues (Su et al., 2004) and far higher than in lymphomas (Monti et al., 2005), lung cancers (Bhattacharjee et al., 2001), or breast cancers (Chin et al., 2006) (Figure 2C).

Coherent Functional Modules of Coexpressed Genes Are Reused across Lineages

To dissect the architecture of the gene expression program, we used the Module Networks (Segal et al., 2003) algorithm (Experimental Procedures) to find modules of strongly coexpressed genes and associate them with candidate regulatory programs that (computationally) predict their expression pattern. We identified 80 gene modules (Figure 3A; modules are numbered

arbitrarily by the algorithm) covering the 8968 genes that are expressed in the majority of the samples of at least one cell population. The genes in each of the modules are tightly coexpressed (Figure S2), and the 80 modules have largely distinct expression patterns (Figure 3A and Figure S2) and are enriched for genes with distinct biological functions (Figure 3B and Table S3).

A small number of modules are expressed in very specific cell states and reflect the unique functional capacities of a single lineage. For example, module 889 is expressed in terminal erythroid differentiation and is enriched for genes encoding blood group antigens and organic cation transporters; module 691 is expressed in B lymphocytes and is enriched for genes encoding immunoglobulins and BCR-signaling pathway components; and module 721 is expressed in granulocytes and monocytes and includes genes encoding enzymes and cytokine receptors that are essential for inflammatory responses.

Conversely, most modules are expressed at varying levels across multiple lineages, suggesting reuse of their genes in multiple hematopoietic contexts. These include modules expressed in both HSC and progenitor populations (e.g., numbers 865, 679, and 805), in both B and T cells (e.g., 673 and 703), in both granulocyte/monocyte populations and lymphocytes (e.g., 817, 799, and 649), and across all myeloid (e.g., 583) or all lymphoid cells (e.g., 931).

Reuse of modules reflects the differential functional requirements for specific biochemical programs in the various cell states. For example, mitochondrial and oxidative phosphorylation modules (e.g., 847, 583, and 883) are induced in erythroid progenitors that produce high levels of heme and are affected most by mitochondrial mutations (Chen et al., 2009; Fontenay et al., 2006), as well as in granulocytes and monocytes, which are capable of a respiratory burst following phagocytosis.

Module States Persist through Multiple Differentiation Steps

To delineate the relation between gene expression and differentiation, we projected each module's expression pattern onto the known topology of the differentiation tree (Figure 4 and Figure S4). For example, consider module 865 (Figure 4A and Figure S3), which is strongly induced in hematopoietic stem and progenitor cells and contains genes encoding key HSPC cell surface markers (CD34 and CD117) and transcriptional regulators (GATA2, HOXA9, HOXA10, MEIS1, and N-MYC). By projecting the module on the differentiation tree, we observe that its induced state in HSCs persists through several consecutive differentiation steps and is repressed at three main points (Figure 4A, arrowheads): (1) after the granulocyte/monocyte progenitor, (2) after erythroid progenitors, and (3) in the differentiation of HSCs toward the lymphocyte lineage.

We identified a host of such differentiation-associated patterns in gene regulation. One major pattern (31 modules) is HSC-persistent states: such modules are active in the HSC state and persist in an active state in several progenitor populations on the erythroid/myeloid branch (Figures 4A and 4E), the lymphoid branch (Figure S4A), or both (Figures S4B and S4H). The HSC state changes gradually at different points in different modules. Indeed, only module 631 (Figure S4C) is primarily HSC specific and includes the known stem cell-specific TFs NANOG and

SMAD1 (Xu et al., 2008). In other patterns, modules have low or inactive expression in HSCs but are activated in a single lineage (10 modules) on either the erythroid/myeloid branch (Figures 4B and 4C and Figure S4D) or the lymphoid branch (Figure 4D). In most cases (39 modules), modules are inactive in HSPCs but are activated in multiple independent lineages (Figure 4F and Figure S4F).

A Sequence-Based Model of the Regulatory Code

The high degree of coexpression of genes within modules suggests that they may be coregulated by common transcriptional circuits. We therefore examined each module for enrichment of known and candidate *cis*-regulatory elements in their promoters (Extended Experimental Procedures). We used six motif-finding methods and a motif-clustering pipeline to identify a nonredundant library of enriched elements. We scored each module for the enrichment of each of the candidate sites or of known elements or binding events (Sandelin et al., 2004; Subramanian et al., 2005) (Extended Experimental Procedures). We identified 156 sequence motifs and 28 binding profiles of 12 TFs (measured by ChIP) that were enriched in the promoters of at least one module (data available on http://www.broadinstitute.org/dmap/). Of these, 66 are previously unannotated motifs, and 118 are associated with 72 TFs (Table S4).

Of these 72 TFs, 11 are known hematopoietic factors (Table S4), and their sites are often enriched in modules consistent with their known functions. For example, the site for the erythroid TF GATA1 (Pevny et al., 1991) is enriched in the late erythroid module 889, and sites for the lymphocyte regulators Helios and NFATC (Aramburu et al., 1995; Hahm et al., 1998) are enriched in the T and NK module 559. We also found significant enrichments for TFs with roles in other differentiation processes, which were not previously implicated in hematopoiesis, such as HNF4 α (in the HSPC Module 865) and HNF6 (in the lymphoid modules 859 and 961).

Tightly Integrated *cis*-Regulatory Circuits Govern Differentiation States

To explore how these *cis*-regulatory associations can give rise to stable cell states, we assembled the regulatory circuits connecting the 276 TFs whose binding sites were enriched in any gene set with each other (Figure 5). We connected an edge from each factor with a known motif to all of the factors that harbor this motif in their gene promoters (Extended Experimental Procedures) and focused only on those factors that were expressed in a given cell state. For example, the circuit of HSC-expressed TFs with known binding sites (Figure 5A) includes many major known regulators of the HSC state (Orkin and Zon, 2008), which are densely interconnected through autoregulatory (12 of 23 active factors), feedback (15 and 39 loops of size 2 and 3), and feedforward (206 loops of size 3) loops. Abnormal expression of many of the circuit's TFs is known to cause hematologic malignancies (Look, 1997). This integrated circuitry can give rise to a robust transcriptional network in terminally differentiated cells and HSCs. Of note, because the sequence of the binding site for most TFs is unknown, including 66 of the putative enriched binding sites, the density of regulation is likely even greater than we observed.
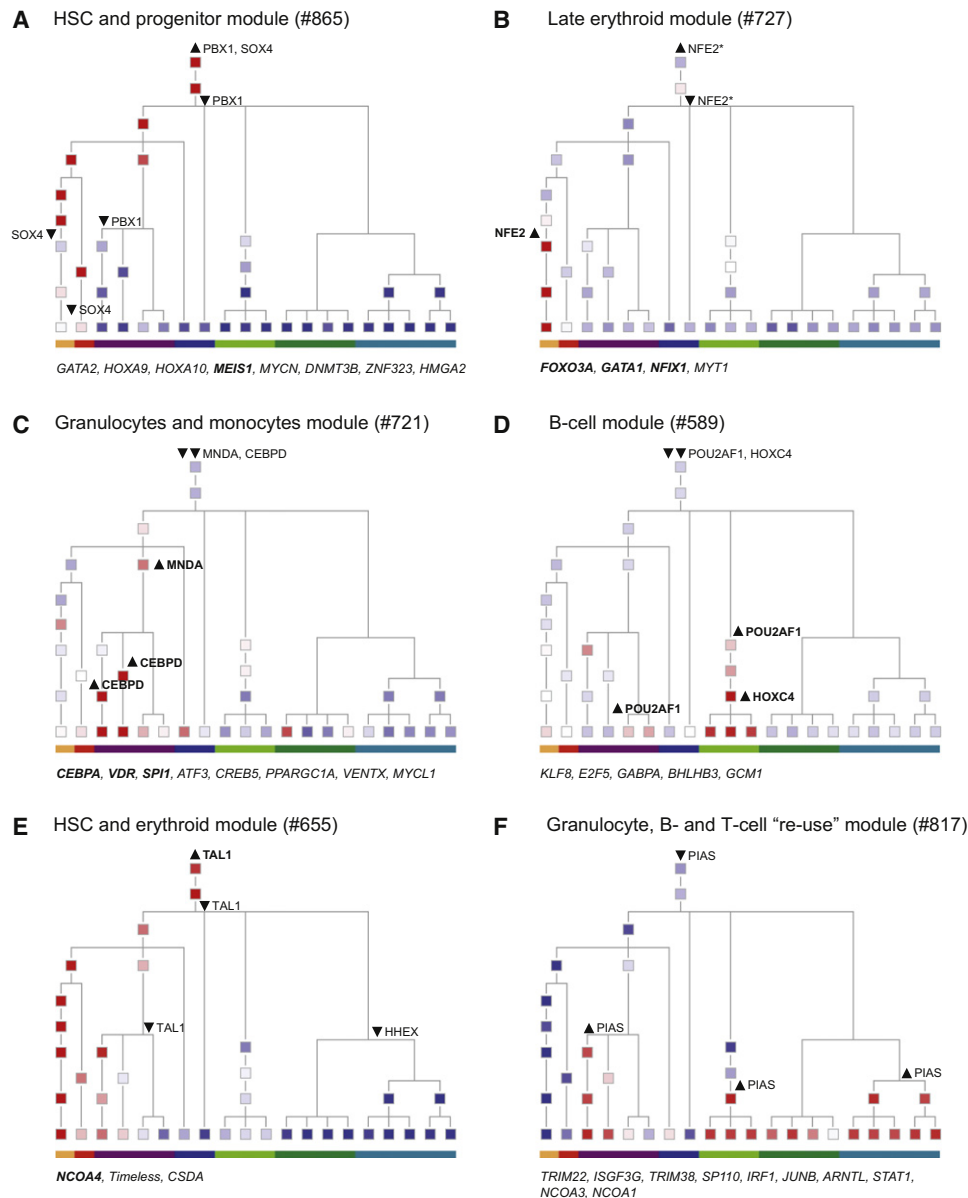
A HSC and progenitor module (#865)

▲ PBX1, SOX4
▼ PBX1
SOX4 ▼
▼ PBX1
▼ SOX4

*GATA2, HOXA9, HOXA10, **MEIS1**, MYCN, DNMT3B, ZNF323, HMGA2*

B Late erythroid module (#727)

▲ NFE2*
▼ NFE2*
NFE2 ▲

***FOXO3A**, **GATA1**, **NFIX1**, MYT1*

C Granulocytes and monocytes module (#721)

▼▼ MNDA, CEBPD
▲ MNDA
▲ CEBPD
▲ CEBPD

***CEBPA**, **VDR**, **SPI1**, ATF3, CREB5, PPARGC1A, VENTX, MYCL1*

D B-cell module (#589)

▼▼ POU2AF1, HOXC4
▲ POU2AF1
▲ POU2AF1
▲ HOXC4

KLF8, E2F5, GABPA, BHLHB3, GCM1

E HSC and erythroid module (#655)

▲ TAL1
▼ TAL1
▼ TAL1
▼ HHEX

***NCOA4**, Timeless, CSDA*

F Granulocyte, B- and T-cell "re-use" module (#817)

▼ PIAS
▲ PIAS
▲ PIAS
▲ PIAS

TRIM22, ISGF3G, TRIM38, SP110, IRF1, JUNB, ARNTL, STAT1, NCOA3, NCOA1

Figure 4. Propagation and Transitions in Modules' Expression along Hematopoiesis

Shown are the mean expression levels of the module's genes in each cell state (colored squares) and selected changes in the predicted regulators, as highlighted in the text (upward arrowhead, regulator induced; downward arrowhead, regulator repressed). Member genes (rather than regulators) in each module encoding TFs are noted below each module, as these may reflect alternative regulators at the same differentiation points. TFs that were validated as regulators of erythroid or granulocyte/monocyte differentiation in a functional assay (Figure 7) are highlighted in bold. The color bar at the bottom of each tree denotes the key lineages, as in Figure 1.

(A) HSC and progenitor expression in module 865.
(B) Lineage-specific induction in late erythrocytes in module 727.
(C) Lineage-specific induction in granulocytes and monocytes in module 721.
(D) Lineage-specific induction in B cells in module 589.
(E) One-sided propagation of induced state from HSC to the erythroid lineage in module 655.
(F) Reuse of module 817, which is inactive in HSCs and independently induced in both lymphoid cells and granulocytes.
See also Figure S3 and Figure S4.

During the course of differentiation, the HSC circuit gradually disappears along multiple lineages due to loss of expression of the relevant TFs (Figure 5A and data available on http://www. broadinstitute.org/dmap/). Conversely, in terminally differentiated cells, other dense circuits emerge through the induction of other TFs. For example, the 14 factors in the erythroid circuit

A HSC network
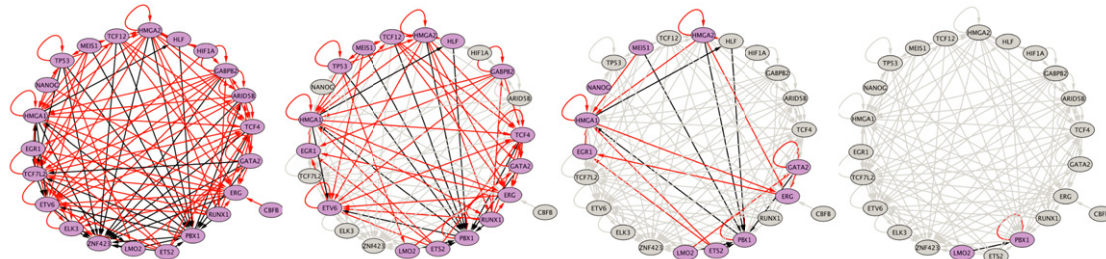
B Late erythrocyte network

HSC ⟶ MEP ⟶ Early ERY ⟶ Late ERY

⟶ Correlated   ⬭ TF
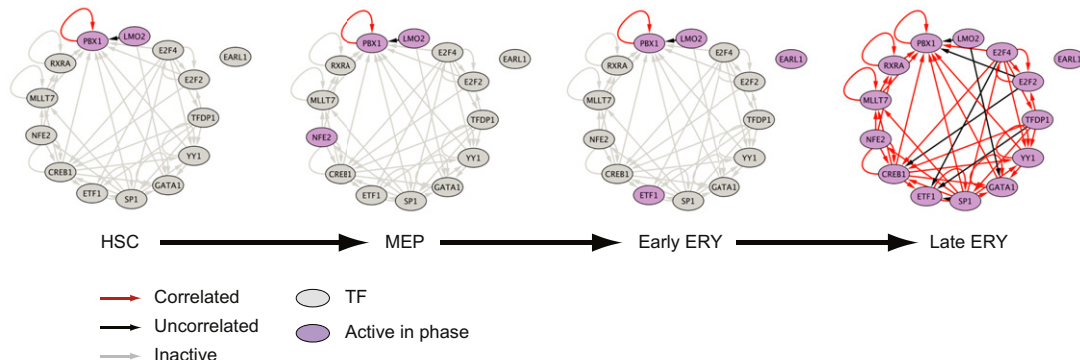⟶ Uncorrelated ⬭ Active in phase
⟶ Inactive

Figure 5. Dynamic Organization of Tightly Integrated *cis*-Regulatory Circuits in HSCs and Erythroid Cells
(A and B) Shown are *cis*-regulatory networks between TFs (nodes) that are enriched in at least one gene set and are expressed (fold change > 1.5) in (A) HSCs or (B) late erythroid cells. Nodes represent TFs that are expressed (purple) or not (gray) in each of the four phases of the erythroid lineage (HSC, MEP, early ERY, and late ERY). An edge from node a to node b indicates that the promoter of the gene in node b has a binding site for the TF encoded by the gene in node a. Edge colors indicate the Pearson correlation between the expression profiles of the TFs in the connected nodes: red, positive correlation (coefficient > 0.4); black, no correlation (absolute Pearson ≤ 0.4); gray, nonactive edge (at least one of the two connected nodes was not expressed in that phase). See Table S4 for enriched motif information.

include many of the known major regulators of erythroid differentiation (Cantor and Orkin, 2002), including GATA1, LMO2, FOXO4, NFE2, and RXRA (Figure 5B). We find similarly distinct networks in the granulocyte lineage, T cells, and B cells.

Hundreds of Transcription Factors Are Differentially Expressed across Lineages in Coherent Modules
The dense regulatory circuits between TFs in our sequence-based model suggest that the expression of TF genes is likely to be highly regulated in hematopoiesis. Indeed, supervised analysis finds that many TF genes are strongly differentially expressed in each primary lineage (Figure 6A and Figure S5A) and that the diversity of TF gene expression is comparable between hematopoiesis and the tissue compendium (Su et al., 2004) (Figure S5B).

Some TFs are expressed predominantly in a single lineage, including well-studied TFs that are known to be essential for differentiation in HSCs or a particular lineage (Figure S6). However, the expression of those factors often increases gradually along differentiation (Figures S6D, S6H, and S6I), similar to the gradations observed in gene modules (Figure 4 and Figure S4).

Many other TFs are "reused" across lineages either through persistent expression from a single progenitor population or by independent activation in multiple lineages (Figure 4 and Figure S4). For example, module 793 (Figure S4F), which is

induced in both B cells and late erythroid cells, includes several TFs and chromatin regulators. Among these, KLF3 has a reported role in erythroid cells (Funnell et al., 2007), whereas NFAT5 has a demonstrated function in B cells (Kino et al., 2009).

Many TFs—not previously associated with these lineages—are expressed similarly to known factors and belong to the same modules, suggesting that the transcriptional circuit consists of a greater number of TFs than previously assumed. For example, the late erythroid module 727 (Figure 4B) contains four TFs: two are known erythroid TFs (GATA1 and FOXO3A) (Bakker et al., 2007), whereas the others (NFIX1, MYT1) were not previously linked to erythropoiesis. Similarly, the granulocytes/monocytes module 721 (Figure 4C) contains eight TFs, only two with known roles in the lineage (CEBPA and PU.1/SPI1).

An Expression-Based Model of the Regulatory Code of Hematopoiesis Identifies Putative Regulators Controlling Changes in Differentiation
To identify the potential regulatory role of differentially expressed TFs, we examined the combinations of TFs (regulatory program), which the Module Networks algorithm (Segal et al., 2003) used in order to "explain" the expression of each of the 80 modules (Experimental Procedures). For example, the algorithm associated module 865 (Figure S3, bottom) with five regulators, most prominently PBX1 ("top regulator") and SOX4 ("2nd level
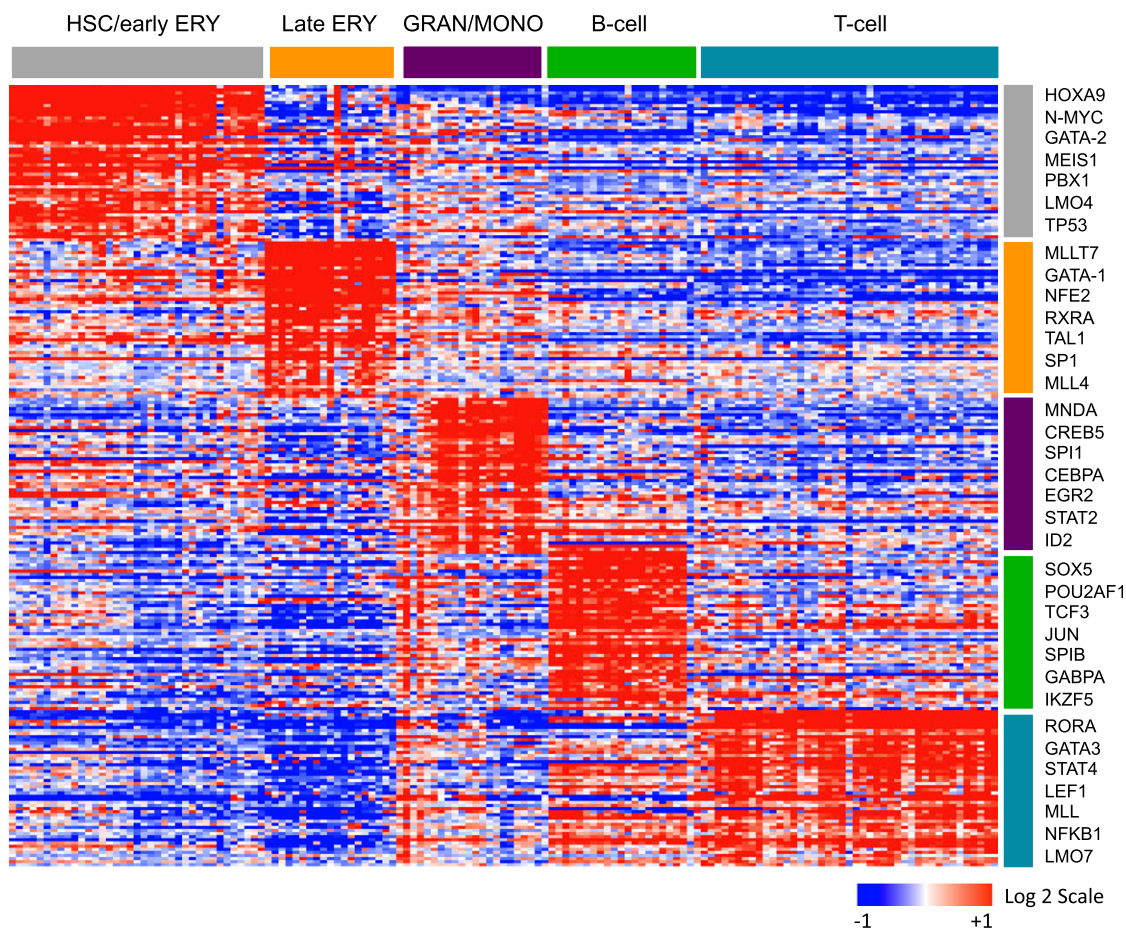
Figure 6. Lineage-Specific Regulation of TF Expression

Signature TF genes with lineage-specific expression in the five main lineages. Shown are the expression levels of the top 50 marker TF genes (rows) selected for each of four major lineages plus hematopoietic stem and progenitor cells (labels as in Figure 1). Genes were selected by high expression in one lineage compared to the others (t test). High expression is shown in red and low expression in blue; the expression of each gene is normalized to a mean expression of zero across all the samples. See also Figure S5 and Figure S6.

regulator") (Figure S3, top). It predicts that, when both PBX1 and SOX4 are induced (in HSCs, CMPs, MEPs, GMPs, early ERY, and early MEGA cells), the module's genes are induced too. PBX1 is an established regulator of HSPCs, and SOX4 has recently been shown to be a direct target of HOXB4, a known HSC regulator (Lee et al., 2010), supporting the algorithm's result. The regulators were chosen by their expression alone, and though the model chooses one combination of "representative" regulators, there may be several highly similar TFs that could fulfill the role.

We next interpreted these regulatory connections within the context of the lineage tree. We associated each regulator with the tree positions (Figure 4 and Figure S4, arrowheads), in which a change in the regulator's expression is associated with a change in the module's expression. For example, there are four such positions for PBX1 and SOX4 in module 865 (Figure 4A, arrowheads), such as the association between the repression of PBX1 and the repression of the module in differentiation toward lymphoid lineages (Figure 4A, downward arrows, labeled PBX1). In this way, we predict the roles of distinct TFs at distinct

differentiation points, such as MNDA at the granulocyte/monocytes progenitor (Figure 4C and Figure S4G) or NCOA4 and KLF1 at late erythrocytes (Figure S4D).

Overall, the algorithm associated 220 TFs (Table S3) with at least one regulatory program and 63 TFs as top regulators (e.g., Figure S3, top) of at least one module. These include 15 TFs previously associated with hematopoiesis (e.g., TAL1, KLF1, BCL11b, LMO2, and MYB) and 7 associated with differentiation in other systems (e.g., CREG1, MEF2A, and NHLH2). For example, we correctly found HOXA9 associated with HSPCs and early erythroid induction (module 679); NFE2, RXRA, KLF1, and FOXO3 associated with late erythroid induction (modules 727, 895, 889, and 739) (Figure 4B and Figure S5D); HIVEP2 and BCL11b associated with T cell induction (modules 859, 949, and 667); and HOXC4 and POU2AF1 associated with B cell induction (module 589) (Figure 4D). In addition, the algorithm predicted a regulatory role for proteins that were not previously associated with regulating hematopoietic differentiation (e.g., MNDA and NCOA4). The selected regulators are enriched for TFs that are known to participate as 3′ partners in fusions in hematologic
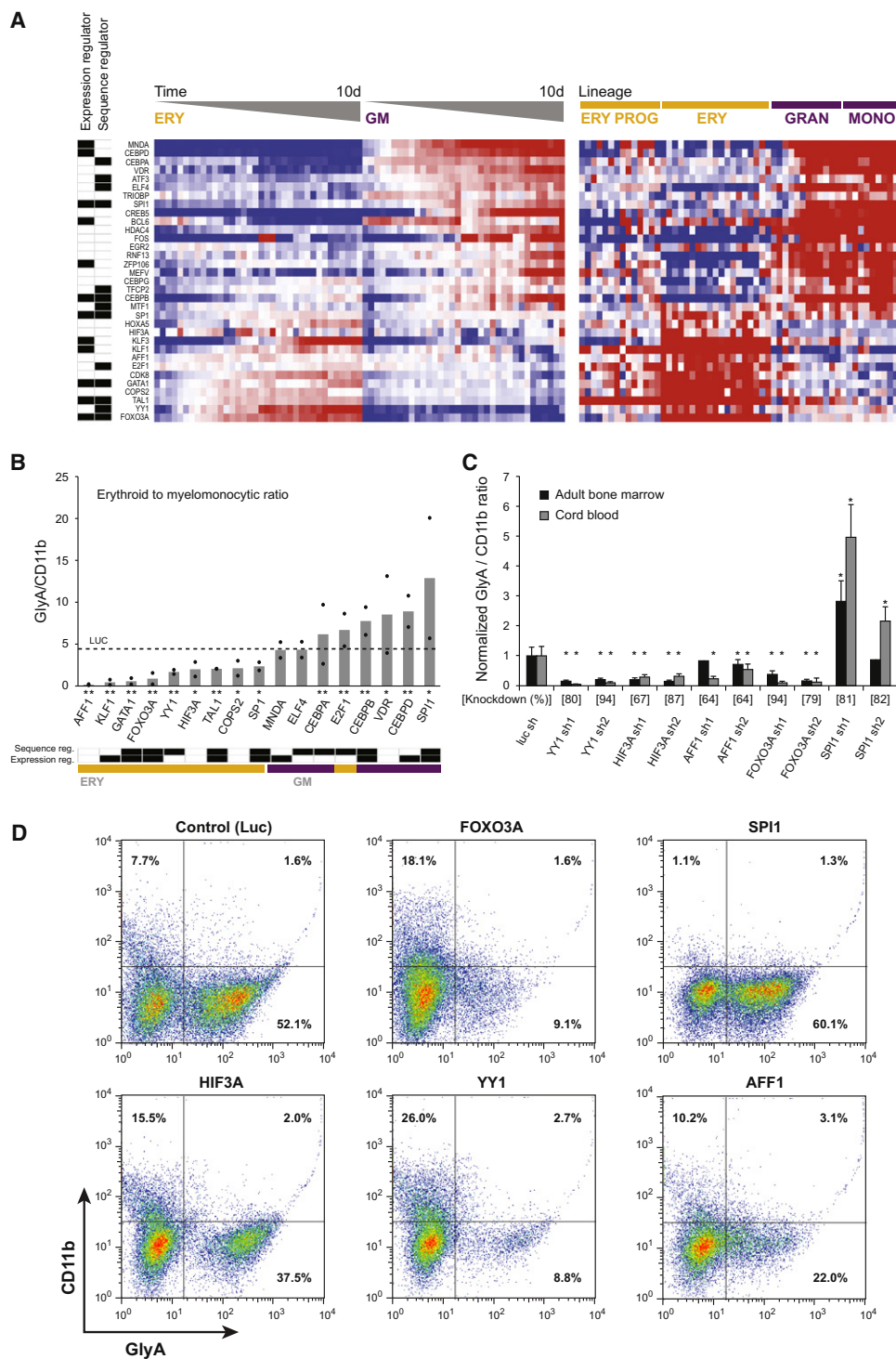
Figure 7. Experimental Validation of 33 TFs

(A) The expression of 33 TFs was detected in primary human bone marrow CD34+ progenitor cells undergoing differentiation in vitro, harvested at 12 time points between days 3 to 10 of differentiation, and detected by a multiplexed assay using LMA followed by fluorescent bead-based detection (left heat map). In the heat map in the right panel, the expression of the same TFs in the original Affymetrix data set is illustrated. The labels at the far left indicate whether the TF was chosen as a regulator in the expression-based model or in the sequence-based model.

(B) Differentiation following TF silencing with shRNA. Human bone marrow CD34+ cells expressing shRNAs targeting TFs were induced to differentiate in vitro for 10 days, and the ratio of erythroid (glycophorin A-positive) and myelomonocytic (CD11b-positive) cells was measured by flow cytometry. Each black dot represents an individual shRNA (mean of three replicates), and bars indicate their average. The effect of a control shRNA targeting the luciferase gene, which is not

cancers (Mitelman et al., 2010) (25 of the regulators; p < 0.028), consistent with a regulatory role in hematopoiesis.

Finally, we compared the predictions of the expression- and sequence-based models. The two models were different due to two reasons. First, 85% of the TFs chosen as regulators in the expression model (187 of 220) do not have a characterized binding motif in current databases and cannot be identified in the sequence model. Second, 29 of 41 TFs (70%) whose known sites are incorporated in the sequence model and appear in the expression model show little or no correlation in expression (absolute Pearson < 0.4) to the module with which they are associated in the sequence model (data available on http://www.broadinstitute.org/dmap/). Thus, the two models are likely complementary, each capturing a substantial but distinct number of known regulators in the relevant states. To gain confidence in their predictions, we next pursued experimental approaches.

Direct Targets of MEIS1, TAL1, IKAROS, and PU.1 in HSPCs Reveal Dense Circuits and Anticipatory Binding

To validate and further investigate the gene modules and *cis*-circuits, we examined the direct binding of TFs across the genome using chromatin immunoprecipitation followed by sequencing (ChIP-Seq) in HSPCs. We analyzed the binding of MEIS1, TAL1, PU.1/SPI1, and IKAROS/IKZF1, four key regulators of the specification, maintenance, or differentiation of HSCs (Argiropoulos et al., 2007; Lécuyer and Hoang, 2004; Ng et al., 2007; Singh et al., 1999), in two replicates, often in independently expanded populations of primary human HSPCs (Extended Experimental Procedures). We scored each experiment for statistically significant binding (Extended Experimental Procedures and Table S5) and tested each of our expression modules for enrichment in binding events (Table S5).

In modules whose genes are highly induced in terminal differentiation, we found enrichment of binding by corresponding lineage specific factors in HSPCs, suggesting anticipatory regulation. For example, module 727 (Figure 4B), expressed in terminally differentiated erythroid cells, was enriched with target genes bound in HSPCs by TAL1, an erythroid transcription factor (Table S5). Similarly, genes in the granulocyte/monocyte module 763 were enriched for targets bound by PU.1 in HSPCs (Table S5), and genes in the lymphoid module 949 were enriched for target genes bound by IKAROS in HSPCs (Table S5). In many (but not all) cases, expression of the target module is already moderate in HSCs and increases with differentiation. This strongly supports an anticipatory regulation in which relevant differentiation TFs are bound at target promoters in HSPCs, resulting in mild expression of targets that persists and further increases upon differentiation.

Some of our expression-based model's predictions for HSPCs are supported by the ChIP-Seq data. For example, the two modules that are induced in HSPCs and are associated in our model with either MEIS1 (module 961) or its known binding partner PBX1 (module 865, Figure 4A) are enriched in target genes bound by MEIS1. MEIS1 and HOXA9 are members of module 865, consistent with MEIS1's autoregulatory binding (Table S5). The ChIP-Seq data also support module reuse. For example, several of the modules enriched with PU.1 are reused in granulocytes and B lymphoid cells (e.g., modules 853, 649, 979, 769, and 817), consistent with an established role for PU.1 in both lineages. In other cases, module reuse may be mediated by combinatorial binding of two factors (e.g., by both PU.1 and IKAROS in module 607, which is expressed in granulocytes, monocytes, and some lymphoid cells).

The individual binding events in our profiles also support the overall organization observed in the *cis*-circuits in the sequence model. First, three of the factors bind their own promoter (IKAROS and MEIS1) or enhancer (PU.1), forming autoregulatory loops, as observed for many known master regulators (Boyer et al., 2005) and in our sequence model. Second, PU.1, IKAROS, and MEIS1 are integrated in a feed-forward loop. Third, there is a significant overlap between the targets of any pair of factors (Table S5). Finally, in aggregate, the factors bind 13 of the 23 other TFs in our HSC circuit, further increasing its density.

Differential Expression of Candidate Transcription Factors during In Vitro Differentiation

We confirmed the lineage-specific expression of 33 TFs in primary human hematopoietic progenitor cells induced to differentiate in vitro. We focused on the erythroid and myelomonocytic lineages, as differentiation of primary human hematopoietic progenitor cells can be faithfully recapitulated and genetically manipulated along these lineages in vitro. We selected a set of 33 TFs identified in either the sequence or gene expression-based models as candidate regulators of these two lineages.

We developed a quantitative, multiplexed assay to detect the expression of the signature genes in a single well using ligation-mediated amplification (LMA) followed by amplicon detection on fluorescent beads (Peck et al., 2006). We cultured primary human CD34+ cells from adult bone marrow in vitro in cytokine conditions promoting either erythroid or myelomonocytic differentiation. We harvested cells at 12 time points between days 3 and 10 of erythroid and myelomonocytic differentiation and determined TF gene expression using the multiplexed bead-based assay. We confirmed that the 33 TFs are differentially expressed between the two lineages, providing a robust expression signature that can distinguish between the two states independent of profiling platform in cells derived from adult bone marrow or umbilical cord blood and in cells that differentiated in vivo or in vitro (Figure 7A).

expressed in human cells, is indicated with a dashed line. Below the shRNA labels, * or ** indicates p < 0.05 for one or both shRNAs, respectively. (Bottom) Classification of the TFs according to their roles in the expression-based and sequence-based models and to their induction pattern in the LMA profiling.

(C) The effects of additional shRNAs targeting candidate TFs expressed in CD34+ cells derived from both umbilical cord blood and adult bone marrow and assayed as in (B) (*p < 0.01).

(D) Representative flow cytometry scatter plots from shRNAs expressed in umbilical cord blood.

See additional information in Table S5, Table S6, and Table S7.

Changes in Expression Levels in Transcription Factor Circuits Functionally Modulate Differentiation In Vitro

We next tested whether acute loss of expression of each TF using RNA interference can functionally affect erythroid and myelomonocytic differentiation. We used our multiplexed bead-based assay to identify short hairpin RNAs (shRNAs) that effectively knock down each TF and found 17 genes with at least two different effective shRNAs. Next, we infected primary human adult bone marrow CD34+ cells with the validated lentiviral shRNAs, cultured the cells in cytokine conditions supporting both erythroid and myelomonocytic differentiation, and assessed the number of erythroid (glycophorin A-positive) cells relative to myelomonocytic (CD11b-positive) cells by flow cytometry (Figure 7B). In most cases, the shRNA perturbation dramatically altered differentiation, with the ratio of erythroid to myeloid cells ranging from less than 1:10 to more than 10:1 with different shRNAs.

The perturbations associated with the lowest fraction of erythroid cells in culture corresponded to the samples expressing shRNAs targeting nine TFs expressed at higher levels in the erythroid lineage (Table S6). Consistent with our models, six were regulators in either the expression or the sequence model, and the other three were members of erythrocyte-induced modules (Figure 7B, bottom). These include GATA-1 and KLF1, TFs with well-established roles in erythroid differentiation (Funnell et al., 2007; Pevny et al., 1991), and TAL1 and FOXO3A, which have been implicated in erythroid differentiation (Aplan et al., 1992; Bakker et al., 2007). The TF YY1 was identified in our sequence-based models, has higher expression in erythroid cells, and was functionally validated by our shRNA screen. A physical association between YY1 and GATA-1 was reported in the chicken α-globin enhancer (Rincón-Arano et al., 2005). Finally, we validated a new role for HIF3A and AFF1 (AF4) in erythroid differentiation based on module membership and perturbation. Of note, AFF1 is a common translocation partner with the MLL gene in leukemia (Li et al., 1998).

Conversely, eight perturbations resulted in the lowest fraction of myelomonocytic cells and corresponded to samples expressing shRNAs targeting seven TFs induced in granulocyte/monocyte cells and one (E2F1) with higher expression in erythroid cells. Four TFs were predicted by the expression model to regulate modules induced in granulocytes/monocytes, and five were predicted in the sequence network (Figure 7B, bottom). These include the well-established granulocyte/monocyte TFs, PU.1/SPI1 and C/EBP family members (Hirai et al., 2006; Scott et al., 1994), and VDR, a gene that has been implicated in myeloid differentiation (Liu et al., 1996).

We further validated three TFs that had not previously been associated with erythroid differentiation (AFF1, HIF3A, and YY1) alongside a known erythroid regulator (FOXO3A) and a known granulocyte regulator (PU.1/SPI1) (Figures 7C and 7D). We tested additional shRNAs for each gene by quantitative PCR and identified two shRNAs per gene that decrease expression of their target genes by 63% to 95% in human CD34+ cells derived from both adult bone marrow and umbilical cord blood. Using flow cytometry for lineage-specific markers following 10 days of differentiation, we validated our initial findings that AFF1, HIF3A, and YY1 decrease the relative production of erythroid lineage cells. These results are further supported by profiling mRNA levels following knockdown of these five TFs at 4 days following lentiviral infection. Compared to a control shRNA, gene expression in cells expressing AFF1, HIF3A, and FOXO3A were anticorrelated with erythroid profiles and positively correlated with granulocytes (Figure S5C), and knockdown of PU.1/SPI1 had the inverse pattern, as expected. Knockdown of YY1 caused a transcriptional profile more similar to HSCs, indicating a more substantial block in terminal differentiation. Taken together, our findings indicate that modulating the expression of TF genes can powerfully alter hematopoietic differentiation.

Web-Based Portal as a Research Resource

To facilitate interrogation of our hematopoietic gene expression database by the broader scientific community, we have created a Web-based portal (http://www.broadinstitute.org/dmap) to provide access to the primary data, sample information, processed results from both models, and a suite of analytic tools.

DISCUSSION

General Principles of Transcriptional Circuits in Differentiation

The changes in gene expression over the course of hematopoietic differentiation are profound. The number of differentially expressed genes is similar within hematopoiesis and across human tissues, suggesting comparable complexity. Our findings reveal several major principles about the organization of this transcriptional program.

Gene expression in hematopoiesis can be decomposed into modules of tightly coexpressed genes, some of which are restricted to specific lineages, whereas most are reused in multiple lineages. Furthermore, a module's transcriptional state persists through multiple differentiation steps. For example, the transcriptional state of HSCs is not switched off immediately but instead persists with gradually decreasing expression in progenitor cells.

Many of the TFs with known binding sites can be assembled into densely interconnected circuits. These can provide a mechanism for robust gene regulation in both terminally differentiated cells and HSCs. Because the binding sites for many factors remain unknown, we therefore expect that the circuit's density and complexity is even higher.

A large number of TFs are differentially expressed across hematopoiesis, often in tightly coregulated modules, and at comparable complexity to that of the other (nonregulatory) genes. Leveraging this correspondence, we associated TFs to the modules and differentiation states that they may regulate. We automatically rediscovered (without prior knowledge) many of the key known TFs and predict regulatory functions for numerous additional TFs.

By monitoring the binding of four major TFs in HSPCs, we found that anticipatory regulation may be a major feature of these circuits. In such cases, TFs that direct lineage-specific differentiation bind a significant portion of their target genes in HSPCs. These target genes are often moderately expressed in the stem and progenitor cells, with substantial further induction as differentiation progresses in the relevant lineage. This is

consistent with the concept of "lineage priming" in HSCs (Akashi, 2005), providing flexibility in cell fate commitments.

Discovering and Validating Transcriptional Regulators in Hematopoiesis

Our examination of the global architecture of hematopoietic differentiation offers a complementary strategy to studies of individual genes in murine models. In this approach, gene expression and sequence-based analyses nominate a host of candidate regulators and point to groups of factors that may act together and hence introduce redundancies. The two computational approaches complement each other: the expression model may identify factors whose binding specificity is unknown, the sequence model may help detect those factors whose mRNA levels do not change or do not correspond to changes in targets (Lu et al., 2009).

We used a perturbation-based approach to validate TFs derived from the sequence and expression models in an in vitro differentiation system. Modulating expression of candidate TFs with RNA interference altered differentiation of hematopoietic progenitor cells in vitro in the direction predicted by our models. We reconfirmed the role of several known factors and identified several new ones (e.g., YY1, AFF1, and HIF3A). In vitro manipulation can be more sensitive than genetic ablation experiments in vivo, wherein perturbations may be corrected by homeostatic mechanisms, such as cytokine or transcriptional feedback loops.

A Transcriptional Roadmap for Hematologic Malignancies

Balanced translocations involving TFs play a major role in the pathogenesis of human leukemias. Of 200 known translocations in AML (Mitelman et al., 2010), there are 53 in which at least one translocation partner is a TF, 16 as a 5′ partner, and 43 as a 3′ partner (6 in both). Twenty-five of these 43 known 3′ partners are among the 220 regulators in the expression model (p < 0.028), and 5 of the known 3′ partners are among the 72 known TFs in the sequence model. Furthermore, 5′ partners are enriched in genes expressed in HSPCs. These results support the role of chosen regulators in differentiation and are consistent with a broader paradigm in which lineage-specific promoters can dysregulate key TFs to disrupt differentiation (Rosenbauer and Tenen, 2007; Tomlins et al., 2005).

Impaired or blocked hematopoietic differentiation is a defining characteristic of leukemia, and the gene expression profiles of leukemias cluster strongly into subgroups that correspond to specific molecular subgroups (Bullinger et al., 2004; Tamayo et al., 2007; Valk et al., 2004). Gene signatures induced in various leukemias significantly overlap those induced in normal hematopoiesis (Figures S7A and S7B). In most cases, there is a coherent overlap between the leukemia subtype and the cell type from which it is known to arise. However, human leukemias often express more complex combinations of modules that are not observed in normal samples, including HSPC modules, as has been reported in murine models of leukemia (Krivtsov et al., 2006).

Toward a Programming "Code" of Hematopoietic Differentiation

A more complete understanding of hematopoietic differentiation will likely require an integration of gene expression data with other genomic data, including epigenetic analyses, genome-wide ChIP-Seq studies, proteomics, and systematic functional studies. Given the ability to produce high-quality measurements from small numbers of cells, gene expression data provides a first draft of the transcriptional program controlling hematopoiesis, opening the way to manipulate and reprogram these circuits, through perturbation and manipulation of each regulatory factor. These can highlight avenues for therapeutic intervention, including "reprogramming" of cells to more desired states. Though deriving mechanistic models from mammalian gene expression profiling data has been challenging, hematopoiesis provides a paradigm for the testing of more advanced algorithms (Kim et al., 2009). Our data set and analyses provide a resource for further inquiries into normal and pathologic hematopoietic differentiation in humans.

EXPERIMENTAL PROCEDURES

Further details for data analysis, chromatin immunoprecipitation, and functional validation experiments are described in the Extended Experimental Procedures.

Subjects and Samples

Human umbilical cord blood was harvested from postpartum placentas at Brigham and Women's Hospital under an Institution Review Board (IRB)-approved protocol. Peripheral blood samples were obtained from healthy volunteers at the Dana-Farber Cancer Institute with informed consent under an IRB-approved protocol.

The majority of cells were purified from umbilical cord blood, an enriched source of undifferentiated populations. However, terminally differentiated lymphocyte populations, including T cells (TCELL1-8), B cells (BCELL1-4), natural killer cells (NKa1, NKa2, NKa3, and NKT), and dendritic cells (DENDa1 and DENDa2), were purified from adult peripheral blood because terminal differentiation in these populations requires exposure to antigens after birth. For each cell population, we purified samples from four to seven distinct donors. All blood samples were harvested fresh and immediately processed for flow sorting.

Cell Sorting Strategy and Flow Cytometry

First, mononuclear cells were isolated by Ficoll-Hypaque sedimentation. For relatively rare populations, including hematopoietic stem cell populations (HSC1 and HSC2), progenitor populations such as common myeloid progenitor (CMP), megakaryocyte/erythroid progenitor (MEP), granulocyte-monocyte progenitor (GMP), and the erythroid lineage populations (ERY1–5), lineage depletion was performed using antibodies against CD2, CD3, CD4, CD5, CD8, CD11b, CD14C, CD19, and CD56 with a magnetic column (Miltenyi Biotec, Auburn, CA). Positive selection was then performed using flow cytometry for labeled antibodies to the markers described in Table S1. For the more common or terminally differentiated populations, including neutrophil populations (GRAN1-3), basophils (BASO1), monocytes (MONO1–2), eosinophils (EOS2), megakaryocytes (MEGA1–2), B-lymphoid progenitor (PRE_BCELL1), pro and early B lymphocytes (PRE_BCELL2 and PRE_BCELL3), dendritic cells (DENDa1 and DENDa2), mature T cells (TCELL1–8), mature B cells (BCELL1–4), and natural killer cells (NKa1–3, NKT), cells were positively selected using flow scatter properties and antibodies based on the immunophenotypes described in Table S7. The gene expression profiles for a subset of the lymphoid populations has been analyzed previously (Haining et al., 2008).

Sorting was performed with Vantage SE. Diva or FACSAria flow cytometers (Becton Dickinson, San Jose, CA). Cell populations of interest were collected into tubes containing PBS in a collection unit at 4°C. The > 95% purity of populations was confirmed by performing FACS analysis of the sorted cells. Sorted cells were spun down, immediately resuspended in TriZol (Invitrogen, San Diego, CA), and stored at −70°C.

Microarray Data Acquisition

Total RNA was isolated from TriZol. The concentration of RNA was quantified using the RiboGreen RNA Quantitation Kit (Invitrogen, San Diego, CA). Ten nanograms of total RNA were amplified using the Ovation Biotin RNA Amplification and Labeling System (NuGEN, San Carlos, CA). The cDNA was fragmented, labeled, and hybridized to Affymetrix HG_U133AAofAv2 microarrays (Affymetrix, Santa Clara, CA), which contain 22,944 probes.

Expression-Based Module Networks Model

The modules and their regulation programs were automatically learned using the Module Networks algorithm (Segal et al., 2003). This method detects modules of coexpressed genes and their shared regulation programs. The regulation program is a small set of genes whose expression is predictive of the expression level of the module genes using a decision (regression) tree structure. Given the expression values and a pool of candidate regulator genes, a set of modules and their associated regulation programs are automatically inferred by an iterative procedure. This procedure searches for the best gene partition into modules and for the regulation program of each module while optimizing a target function. The target function is the Bayesian score derived from the posterior probability of the model (see Segal et al., 2005 for a detailed description of the algorithm).

ACCESSION NUMBERS

Data set is available on http://www.ncbi.nlm.nih.gov/geo/, GSE24759.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and seven tables and can be found with this article online at doi:10.1016/j.cell.2011.01.004.

REFERENCES

Akashi, K. (2005). Lineage promiscuity and plasticity in hematopoietic development. Ann. N Y Acad. Sci. *1044*, 125–131.

Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O., et al. (2009). Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. Science *326*, 257–263.

Aplan, P.D., Nakahara, K., Orkin, S.H., and Kirsch, I.R. (1992). The SCL gene product: a positive regulator of erythroid differentiation. EMBO J. *11*, 4073–4081.

Aramburu, J., Azzoni, L., Rao, A., and Perussia, B. (1995). Activation and expression of the nuclear factors of activated T cells, NFATp and NFATc, in human natural killer cells: regulation upon CD16 ligand binding. J. Exp. Med. *182*, 801–810.

Argiropoulos, B., Yung, E., and Humphries, R.K. (2007). Unraveling the crucial roles of Meis1 in leukemogenesis and normal hematopoiesis. Genes Dev. *21*, 2845–2849.

Bakker, W.J., van Dijk, T.B., Parren-van Amelsvoort, M., Kolbus, A., Yamamoto, K., Steinlein, P., Verhaak, R.G., Mak, T.W., Beug, H., Löwenberg, B., and von Lindern, M. (2007). Differential regulation of Foxo3a target genes in erythropoiesis. Mol. Cell. Biol. *27*, 3839–3854.

Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., et al. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc. Natl. Acad. Sci. USA *98*, 13790–13795.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Brunet, J.-P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proc. Natl. Acad. Sci. USA *101*, 4164–4169.

Bullinger, L., Döhner, K., Bair, E., Fröhling, S., Schlenk, R.F., Tibshirani, R., Döhner, H., and Pollack, J.R. (2004). Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N. Engl. J. Med. *350*, 1605–1616.

Cantor, A.B., and Orkin, S.H. (2002). Transcriptional regulation of erythropoiesis: an affair involving multiple partners. Oncogene *21*, 3368–3376.

Chen, M.L., Logan, T.D., Hochberg, M.L., Shelat, S.G., Yu, X., Wilding, G.E., Tan, W., Kujoth, G.C., Prolla, T.A., Selak, M.A., et al. (2009). Erythroid dysplasia, megaloblastic anemia, and impaired lymphopoiesis arising from mitochondrial dysfunction. Blood *114*, 4045–4053.

Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T., et al. (2006). Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell *10*, 529–541.

Davidson, E.H. (2001). Genomic Regulatory Systems: In Development and Evolution (Burlington, MA: Academic Press).

Ebert, B.L., and Golub, T.R. (2004). Genomic approaches to hematologic malignancies. Blood *104*, 923–932.

Fontenay, M., Cathelin, S., Amiot, M., Gyan, E., and Solary, E. (2006). Mitochondria in hematopoiesis and hematological diseases. Oncogene *25*, 4757–4767.

Funnell, A.P., Maloney, C.A., Thompson, L.J., Keys, J., Tallack, M., Perkins, A.C., and Crossley, M. (2007). Erythroid Krüppel-like factor directly activates the basic Krüppel-like factor gene in erythroid cells. Mol. Cell. Biol. *27*, 2777–2790.

Hahm, K., Cobb, B.S., McCarty, A.S., Brown, K.E., Klug, C.A., Lee, R., Akashi, K., Weissman, I.L., Fisher, A.G., and Smale, S.T. (1998). Helios, a T cell-restricted Ikaros family member that quantitatively associates with Ikaros at centromeric heterochromatin. Genes Dev. *12*, 782–796.

Haining, W.N., Ebert, B.L., Subrmanian, A., Wherry, E.J., Eichbaum, Q., Evans, J.W., Mak, R., Rivoli, S., Pretz, J., Angelosanto, J., et al. (2008). Identification of an evolutionarily conserved transcriptional signature of CD8 memory differentiation that is shared by T and B cells. J. Immunol. *181*, 1859–1868.

Hirai, H., Zhang, P., Dayaram, T., Hetherington, C.J., Mizuno, S., Imanishi, J., Akashi, K., and Tenen, D.G. (2006). C/EBPbeta is required for 'emergency' granulopoiesis. Nat. Immunol. *7*, 732–739.

Iwasaki, H., and Akashi, K. (2007). Hematopoietic developmental pathways: on cellular basis. Oncogene *26*, 6687–6696.

Kim, H.D., Shay, T., O'Shea, E.K., and Regev, A. (2009). Transcriptional regulatory circuits: predicting numbers from alphabets. Science *325*, 429–432.

Kino, T., Takatori, H., Manoli, I., Wang, Y., Tiulpakov, A., Blackman, M.R., Su, Y.A., Chrousos, G.P., DeCherney, A.H., and Segars, J.H. (2009). Brx mediates the response of lymphocytes to osmotic stress through the activation of NFAT5. Sci. Signal. *2*, ra5.

Krivtsov, A.V., Twomey, D., Feng, Z., Stubbs, M.C., Wang, Y., Faber, J., Levine, J.E., Wang, J., Hahn, W.C., Gilliland, D.G., et al. (2006). Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. Nature 442, 818–822.

Lécuyer, E., and Hoang, T. (2004). SCL: from the origin of hematopoiesis to stem cells and leukemia. Exp. Hematol. 32, 11–24.

Lee, H.M., Zhang, H., Schulz, V., Tuck, D.P., and Forget, B.G. (2010). Downstream targets of HOXB4 in a cell line model of primitive hematopoietic progenitor cells. Blood 116, 720–730.

Li, Q., Frestedt, J.L., and Kersey, J.H. (1998). AF4 encodes a ubiquitous protein that in both native and MLL-AF4 fusion types localizes to subnuclear compartments. Blood 92, 3841–3847.

Liu, M., Lee, M.H., Cohen, M., Bommakanti, M., and Freedman, L.P. (1996). Transcriptional activation of the Cdk inhibitor p21 by vitamin D3 leads to the induced differentiation of the myelomonocytic cell line U937. Genes Dev. 10, 142–153.

Look, A.T. (1997). Oncogenic transcription factors in the human acute leukemias. Science 278, 1059–1064.

Lu, R., Markowetz, F., Unwin, R.D., Leek, J.T., Airoldi, E.M., MacArthur, B.D., Lachmann, A., Rozov, R., Ma'ayan, A., Boyer, L.A., et al. (2009). Systems-level dynamic analyses of fate change in murine embryonic stem cells. Nature 462, 358–362.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28, 495–501.

Mitelman, F., Johansson, B., and Mertens, F. (2010). Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (http://cgap.nci.nih.gov/Chromosomes/Mitelman).

Monti, S., Savage, K.J., Kutok, J.L., Feuerhake, F., Kurtin, P., Mihm, M., Wu, B., Pasqualucci, L., Neuberg, D., Aguiar, R.C., et al. (2005). Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. Blood 105, 1851–1861.

Müller, F.J., Laurent, L.C., Kostka, D., Ulitsky, I., Williams, R., Lu, C., Park, I.H., Rao, M.S., Shamir, R., Schwartz, P.H., et al. (2008). Regulatory networks define phenotypic classes of human stem cell lines. Nature 455, 401–405.

Ng, S.Y., Yoshida, T., and Georgopoulos, K. (2007). Ikaros and chromatin regulation in early hematopoiesis. Curr. Opin. Immunol. 19, 116–122.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell 132, 631–644.

Peck, D., Crawford, E.D., Ross, K.N., Stegmaier, K., Golub, T.R., and Lamb, J. (2006). A method for high-throughput gene expression signature analysis. Genome Biol. 7, R61.

Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature 349, 257–260.

Quesenberry, P.J., and Colvin, G.A. (2005). Hematopoietic Stem Cells, Progenitor Cells, and Cytokines. In Williams Hematology, E. Beutler, M.A. Lichtman, B.S. Coller, T.J. Kipps, and U. Seligsohn, eds. (New York: McGraw-Hill), pp. 153–174.

Rincón-Arano, H., Valadez-Graham, V., Guerrero, G., Escamilla-Del-Arenal, M., and Recillas-Targa, F. (2005). YY1 and GATA-1 interaction modulate the chicken 3'-side alpha-globin enhancer activity. J. Mol. Biol. 349, 961–975.

Rosenbauer, F., and Tenen, D.G. (2007). Transcription factors in myeloid development: balancing differentiation with transformation. Nat. Rev. Immunol. 7, 105–117.

Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. 32(Database issue), D91–D94.

Scott, E.W., Simon, M.C., Anastasi, J., and Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science 265, 1573–1577.

Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., and Friedman, N. (2003). Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat. Genet. 34, 166–176.

Segal, E., Pe'er, D., Regev, A., Koller, D., and Friedman, N. (2005). Learning module networks. J. Mach. Learn. Res. 6, 557–588.

Singh, H., DeKoter, R.P., and Walsh, J.C. (1999). PU.1, a shared transcriptional regulator of lymphoid and myeloid cell fates. Cold Spring Harb. Symp. Quant. Biol. 64, 13–20.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. USA 101, 6062–6067.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550.

Suzuki, H., Forrest, A.R., van Nimwegen, E., Daub, C.O., Balwierz, P.J., Irvine, K.M., Lassmann, T., Ravasi, T., Hasegawa, Y., de Hoon, M.J., et al; FANTOM Consortium; Riken Omics Science Center. (2009). The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. Nat. Genet. 41, 553–562.

Tamayo, P., Scanfeld, D., Ebert, B.L., Gillette, M.A., Roberts, C.W., and Mesirov, J.P. (2007). Metagene projection for cross-platform, cross-species characterization of global transcriptional states. Proc. Natl. Acad. Sci. USA 104, 5959–5964.

Tibshirani, R., and Hastie, T. (2007). Outlier sums for differential gene expression analysis. Biostatistics 8, 2–8.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 310, 644–648.

Valk, P.J., Verhaak, R.G., Beijen, M.A., Erpelinck, C.A., Barjesteh van Waalwijk van Doorn-Khosrovani, S., Boer, J.M., Beverloo, H.B., Moorhouse, M.J., van der Spek, P.J., Löwenberg, B., and Delwel, R. (2004). Prognostically useful gene-expression profiles in acute myeloid leukemia. N. Engl. J. Med. 350, 1617–1628.

Valouev, A., Johnson, D.S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., Myers, R.M., and Sidow, A. (2008). Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat. Methods 5, 829–834.

Xu, R.H., Sampsell-Barron, T.L., Gu, F., Root, S., Peck, R.M., Pan, G., Yu, J., Antosiewicz-Bourget, J., Tian, S., Stewart, R., and Thomson, J.A. (2008). NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. Cell Stem Cell 3, 196–206.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137.