Cell Stem Cell
Resource

Mapping Cellular Hierarchy by Single-Cell Analysis of the Cell Surface Repertoire

Guoji Guo,[1] Sidinh Luc,[1] Eugenio Marco,[3] Ta-Wei Lin,[4] Cong Peng,[1] Marc A. Kerenyi,[1] Semir Beyaz,[1] Woojin Kim,[1] Jian Xu,[1] Partha Pratim Das,[1] Tobias Neff,[5] Keyong Zou,[6] Guo-Cheng Yuan,[3] and Stuart H. Orkin[1,2,*]
[1]Division of Pediatric Hematology/Oncology, Boston Children's Hospital and Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Harvard Medical School, Boston, MA 02115, USA
[2]Howard Hughes Medical Institute, Boston, MA 02115, USA
[3]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston, MA 02115, USA
[4]Molecular Genetics Core Facility, Children's Hospital Boston, Boston, MA 02115, USA
[5]Pediatric Hematology/Oncology/BMT, University of Colorado, Aurora, CO 80045, USA
[6]Boston Open Labs, Cambridge, MA 02138, USA
*Correspondence: stuart_orkin@dfci.harvard.edu
http://dx.doi.org/10.1016/j.stem.2013.07.017

SUMMARY

Stem cell differentiation pathways are most often studied at the population level, whereas critical decisions are executed at the level of single cells. We have established a highly multiplexed, quantitative PCR assay to profile in an unbiased manner a panel of all commonly used cell surface markers (280 genes) from individual cells. With this method, we analyzed over 1,500 single cells throughout the mouse hematopoietic system and illustrate its utility for revealing important biological insights. The comprehensive single cell data set permits mapping of the mouse hematopoietic stem cell differentiation hierarchy by computational lineage progression analysis. Further profiling of 180 intracellular regulators enabled construction of a genetic network to assign the earliest differentiation event during hematopoietic lineage specification. Analysis of acute myeloid leukemia elicited by MLL-AF9 uncovered a distinct cellular hierarchy containing two independent self-renewing lineages with different clonal activities. The strategy has broad applicability in other cellular systems.

INTRODUCTION

Cellular differentiation is commonly depicted as a sequential binary commitment process through multiple intermediate states. Using combinations of markers, different types of stem and progenitor cells have been identified in various systems. Further enrichment and analysis of these populations has aided appreciation of stepwise lineage specification. However, the choice of a small number of markers for enrichment of cell populations often masks potential heterogeneity and may bias an understanding of the cellular hierarchy.

Extensive cellular and molecular studies have contributed to the characterization of vertebrate hematopoietic differentiation pathways (Orkin and Zon, 2008). The prospective identification of mouse hematopoietic stem and progenitor cells (Muller-Sieburg et al., 1986; Visser et al., 1984) and further separation of hematopoietic stem cells (HSCs) from multipotent progenitors (MPPs) (Kiel et al., 2005; Morrison et al., 1997; Morrison and Weissman, 1994; Osawa et al., 1996) suggested a cellular hierarchy, whereby self-renewing HSCs produce transiently amplifying MPP. Subsequent identification of common lymphoid (CLPs) and myeloid progenitors (CMPs) (Akashi et al., 2000; Kondo et al., 1997) led to the conventional model in which lineage specification first takes place as a lymphoid (CLP) versus myeloid (CMP) bifurcation event. Several findings, however, challenge this simple view. They describe heterogeneity of early progenitor populations and posit that lymphomyeloid lineage commitment may occur upstream of the separation of CLP and CMP (Adolfsson et al., 2005; Arinobu et al., 2007; Pronk et al., 2007). Different marker panels and fluorescence-activated cell sorting (FACS) purification schemes have prevented resolution of these alternative models.

Cells within leukemias are also believed to form a hierarchy, yet descriptions of leukemia stem cells (LSCs) are often seemingly contradictory. Original support for the existence of LSCs rested on the observation that only a rare subset of human acute myeloid leukemia (AML) cells, characterized by a surface phenotype similar to that of hematopoietic stem/progenitor cells, was competent to reinitiate disease upon transplantation in immunodeficient mice (Bonnet and Dick, 1997). More recent findings derived from a mouse model of AML driven by MLL-AF9 suggest that LSCs display a granulocyte/monocyte progenitor (GMP)-like phenotype and stand at the top of the leukemia hierarchy (Krivtsov et al., 2006). Other reports argue that leukemia cells with immunophenotypes of lineage cells may perform as functional LSCs in mouse AML (Gibbs et al., 2012; Somervaille and Cleary, 2006), adding to the complexity of the leukemia hierarchy.

Single-cell gene expression analysis offers potential to resolve these issues. Recently, several hallmark technical advances

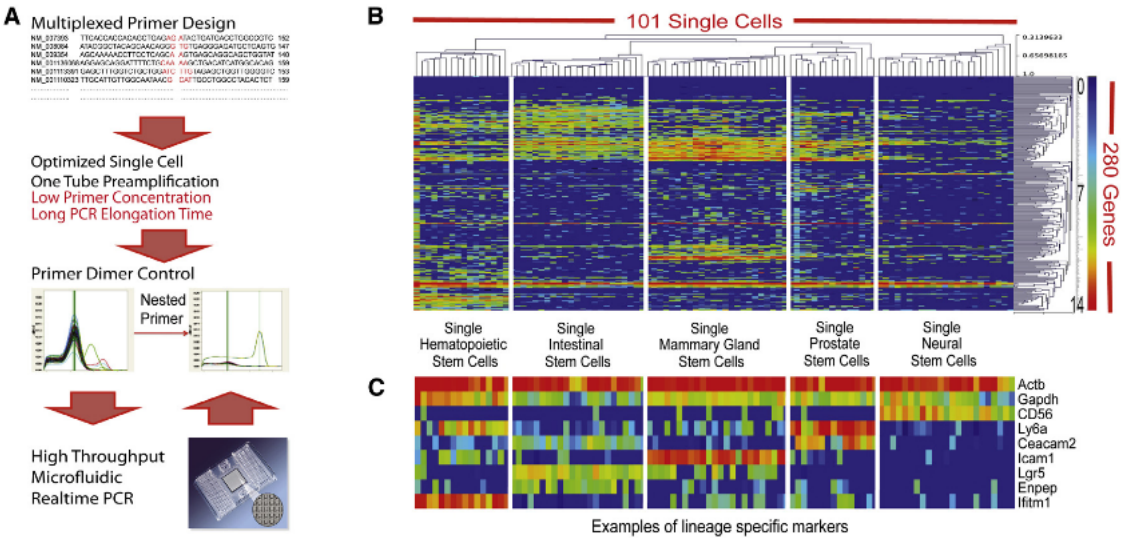


Figure 1. Single-Cell Gene Expression Analysis of the Cell Surface Repertoire

(A) Flow chart of single-cell assay development.

(B) A heatmap showing that the unbiased hierarchical clustering well separates single-cell gene expression signatures from different types of adult stem cells. Each row corresponds to a specific gene; each column corresponds to a particular single cell. Red to yellow suggest high to middle expression, whereas green to blue suggest low to no expression.

(C) A heatmap highlighting examples of lineage-specific markers from Figure 1B. The color scale and sample layout are the same as in Figure 1B.

See also Figure S1 and Tables S1 and S7.

have been achieved. Single-cell messenger RNA (mRNA) sequencing strategies enable whole transcriptome analysis from individual cells (Islam et al., 2012; Ramsköld et al., 2012; Tang et al., 2009, 2010). Alternatively, single-cell mass cytometry constitutes a powerful system for multiplexed gene expression analysis at the protein level (Bendall et al., 2011). When both sample size and assayed gene number are taken into consideration, high-throughput single-cell quantitative PCR (qPCR) represents a favorable option (Buganim et al., 2012; Dalerba et al., 2011; Guo et al., 2010; Moignard et al., 2013). The qPCR approach is highly sensitive in detecting quantitative differences at mRNA level (Guo et al., 2010).

Here, we sought to improve the utility and value of current single-cell qPCR technology by increasing its throughput so as to assess expression of nearly all commonly used cell surface markers. We illustrate how this enhanced approach provides biological insights into normal and leukemic hematopoiesis. The approach we describe should be applicable to other developmental systems and allow for cross-tissue and cross-experiment comparisons. The method allows dissection of heterogeneous populations and the identification of cellular states at single-cell resolution.

RESULTS

Single-Cell Gene Expression Analysis of the Cell Surface Repertoire

By introducing algorithm-based primer design and optimizing the cycling conditions for highly multiplexed PCR, we have increased the capacity of single-cell mRNA sequence-specific preamplification (Figure 1A). In addition, the use of EvaGreen real-time PCR chemistry (Biotium) and melting curve analysis allows for nonspecific signal control during gene-specific qPCR on the BioMark real-time PCR system (Fluidigm). Finally, inclusion of nested primers filters out primer dimer signals (Figure 1A). In the highly multiplexed PCR preamplification, the chance of forming a dimer between a given primer pair of the same gene is actually very low. The subsequent gene-specific qPCR will select and enrich target amplicons, even from extremely low starting materials (Figures S1A and S1B available online). We have designed and optimized a panel of assays to cover all commonly used cell surface markers (Lai et al., 1998) with a total of 280 genes (a few important transcription factors are also included) in establishing an analysis platform for all mouse cell types. After 280 multiplexed, single-cell preamplification, individual gene expression is quantified on the BioMark real-time PCR system (Fluidigm) using three 96.96 dynamic arrays.

To assess the ability of the assay to discriminate different cell types at single-cell level, we used flow cytometry to sort stem cell populations from a broad range of tissues, including neural, prostate, mammary gland, intestinal, and hematopoietic stem cells, according to published protocols (Table S1A), and applied the single-cell assay for gene expression profiling. As shown in Figure 1B, hierarchical clustering of the single cell data faithfully groups cells of the same origin together. The clustering also reveals lineage-specific markers, such as *CD56* for neural stem cells, *Ceacam2* for prostate stem cells, *Icam1* for mammary gland stem cells, *Lgr5* for intestinal stem cells, and *Ifitm1* for hematopoietic stem cells (Figure 1C). False positive signal from a no-cell preamplification control is extremely rare and weak (Figure S1C). These results provide initial evidence on behalf of the robustness of the single-cell approach.

Comprehensive Single-Cell Analysis of the Hematopoietic System

We utilized the single-cell assay for a systematic analysis of the mouse hematopoietic system. To enrich stem cell and progenitor cell populations and represent all possible cellular transitional states during differentiation, we used FACS to sort the principal hematopoietic compartments of the bone marrow by use of the cell surface markers Kit and Sca1, as well as a lineage (Lin) cocktail that recognizes mature cells of the major hematopoietic cell lineages, including T lymphocytes, B lymphocytes, monocytes/macrophages, granulocytes, and erythrocytes. Sorted populations include Lin+, Lin-Sca1+Kit+ (LSK), Lin-Sca1+Kit□ (LSK□), Lin-Sca1□Kit+ (LS□K), and Lin-Sca1□Kit□ (LS□K□) populations (Figure 2A). In addition, we sorted conventionally defined stem and progenitor cell types (including HSCs, MPP, CMP, CLP, common dendritic cell progenitor [CDP] [Onai et al., 2007], megakaryocyte/erythroid progenitor [MEP], and GMP) as well as a set of differentiated cell types (Figure S2; Table S1). A similar strategy was used to sort CD4+ T cells, CD8+ T cells, CD4+CD8+ double positive T cells, earliest thymic progenitors (ETP), CD4□CD8□ double-negative (DN) 2, DN3, and DN4 stage thymocyte progenitors (Figures 2A and S2; Table S1). An average of around 50 single cells are analyzed for each sorted population (Table S1). We analyzed more than 1,500 single cells throughout the mouse hematopoietic system and quantified all 280 genes for each individual cell (Table S2).

Unsupervised hierarchical clustering of the single cell data set reveals high correlation of gene expression clusters with cell-type clusters (Figure 2B). As highlighted by white boxes, *CD11c*, *CD3*, *Blnk*, *Kit*, *CD11b*, and *Gypa* clusters correspond to dendritic, T, B, stem and progenitor, myeloid, and megakaryocytic and erythroid (MegE) lineage cells, respectively. The principal lineage-specific gene clusters are summarized in Table S3. Subclusters also exist within these main clusters. In addition, the quantified mRNA level differences correlate with different FACS sorting schemes; *Actb* and *Gapdh* expression levels are relatively consistent (Figure S3A). The clustering data suggest that differential global gene expression signatures at the single level are reproducible in both progenitor and differentiated cells types. The clustering pattern may then be used to identify novel markers and populations.

To visualize the overall pattern of gene expression (280 parameters) at the single-cell level, we used the Gene Expression Dynamics Inspector (GEDI) program (Chang et al., 2008) to generate individual expression maps (Figure 2C). The color of each pixel on the map indicates the centroid value of the gene expression level for each minigene cluster generated by the software. Representative single-cell maps from different populations illustrate how the method can be used to identify and classify virtually all cell types. As an example, we show that an incompletely defined LSK□ population is very heterogeneous, as revealed by the clustering (Figure 2C). According to the single-cell gene expression signature, this population contains not only CLP-like progenitors and B cell progenitors but also plasmacytoid dendritic cells (PDC) (Onai et al., 2007) and nuocytes (Neill et al., 2010). Changes during cellular differentiation may be visualized from the maps. The gradual transition of the GEDI map from sorted MEP to CD71+ erythroid progenitors and then to Ter119+ cells provides an example. Interestingly, we have identified bone marrow MPP-like cells in the spleen (MPP-SP) and thymus (MPP-TH), consistent with the circulation of hematopoietic progenitor cells throughout the body.

Heterogeneity of Hematopoietic Progenitor Cell Types

Having established a robust methodology for single-cell analysis, we proceeded to examine the classically defined hematopoietic progenitor cell types (Figures 3A–3E). Each of these progenitor types reveals marked heterogeneity. For example, we profiled 47 single CMP cells, originally defined by the Lin-IL7R-Kit+Sca1-CD34+CD16/CD32lo profile, and ranked 280 genes by their standard deviation across all CMP samples. The top four most variable genes were *CD53*, *Sell*, *CD55*, and *Flt3* (Figure 3A). Hierarchical clustering of these variable genes reveals two principal populations with different gene expression patterns. To address whether these gene expression differences reflect stochastic noise (Chang et al., 2008), we applied violin plot analysis to visualize the distribution of gene expression levels. In this plot, the Y- and X-axes correspond to the gene expression level and distribution frequency, respectively. Theoretically, expression noise should exhibit unimodal distribution around a reference level, whereas a multimodal distribution should indicate quantitative differences. As expected, the distributions of *Actb* and *Gapdh* levels are unimodal, with a very narrow peak indicative of low variation (Figures 3A and S3B). In contrast, the top four most variable genes within the CMP population show clear bimodal distribution. To confirm mRNA level differences at the protein level, we used available antibodies to the surface marker CD55 to further analyze the CMP compartment. Flow cytometry validated the heterogeneous nature of the CMP population detected by single-cell qPCR (Figure 3A). We then continued to dissect heterogeneity further in the CD55-CMP population and revealed Csf1r as one of the most differentially expressed markers (Figure 3B). Comparable analyses were performed for GMP, CLP, MEP, ETP, and CDP. We observed discrete heterogeneity within all populations (Figures 3C–3E). The analysis also reveals dynamic changes in LSK heterogeneity during the aging process (Figure S3C) and permits assessment of the purity of HSCs from different enrichment protocols (Figure S3D). The bimodal distribution of mRNA transcripts is present in all the cell types that we have purified, suggesting extensive unknown heterogeneities. Although the mRNA level expression is not always reflective of protein level expression, we argue that it should be indicative of a cell's transcriptional state and functional potential.

Mapping Hematopoietic Hierarchy by Computational Lineage Progression Analysis

We hypothesized that the similarity of different single-cell signatures and continuity of transitional states during differentiation could form the foundation of an in silico strategy to organize high-dimensional data into ordered, stepwise cell fate commitment pathways. To accomplish this, we first removed redundancy by extracting the average value of 40 distinct gene expression clusters from the entire data set (Table S3) and then used spanning-tree progression analysis of density-normalized events (SPADE) (Bendall et al., 2011; Qiu et al., 2011) analysis to distill 40 dimensional single-cell data down to a single interconnected cluster of transitional cell populations.

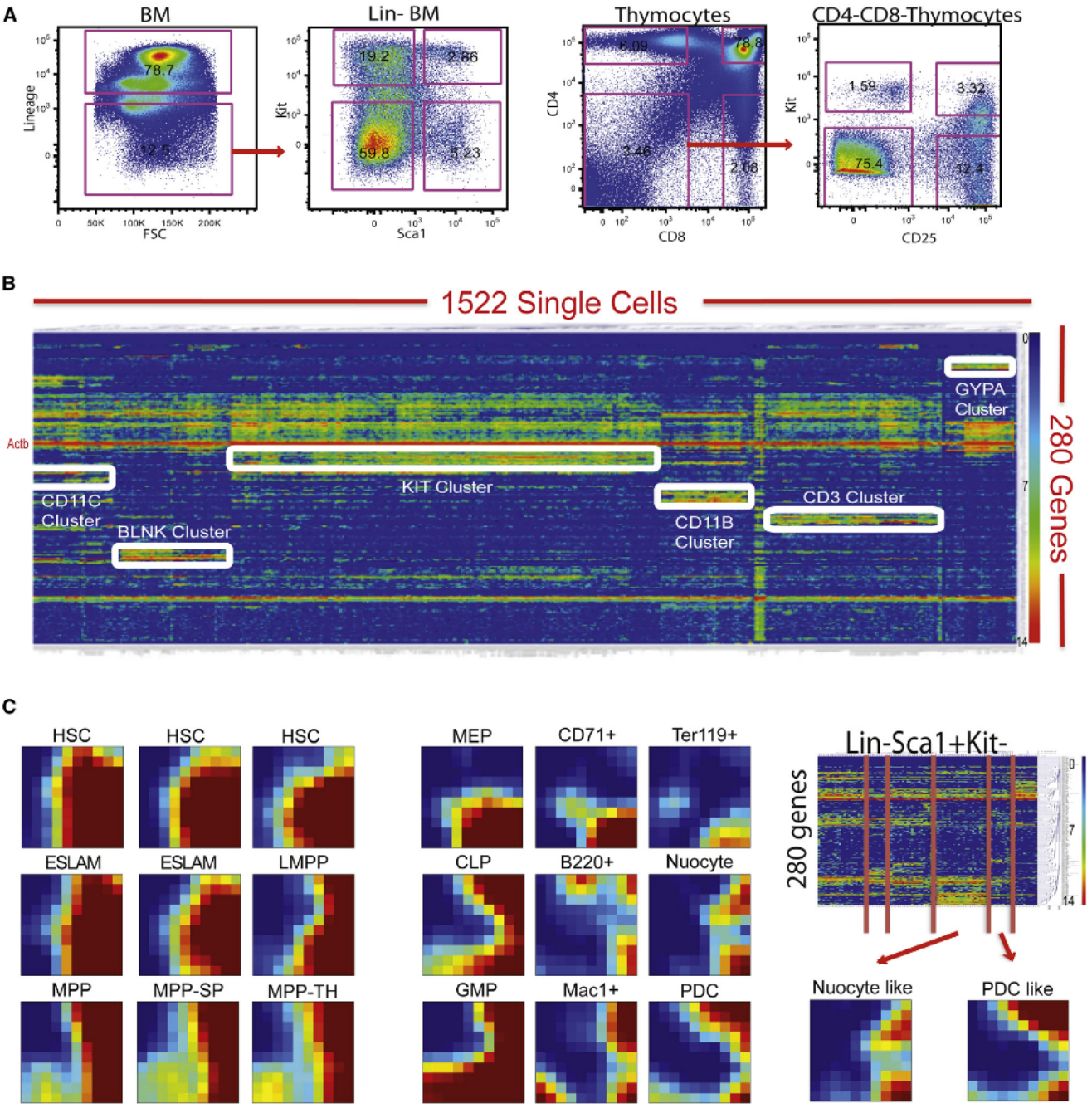


Figure 2. Comprehensive Single-Cell Analysis of the Mouse Hematopoietic System

(A) Single-cell sorting strategy to enrich stem and progenitor cells but to cover all possible populations.

(B) A master heatmap showing the hierarchical clustering of gene expression signatures from 1,500 single cells throughout the hematopoietic system. Each row corresponds to a specific gene; each column corresponds to a particular single cell. Strong correlation between gene and cell clusters are highlighted by white boxes and labeled by cell type-specific clusters. Red to yellow suggest high to middle expression, whereas green to blue suggest low to no expression.

(C) GEDI plot allows for visualization of single-cell global signatures. Examples of single-cell GEDI map from different cell types are presented. Color scale is as described in Figure 1B. The lower right corner, which is always red, corresponds to endogenous control genes that are highly expressed in all single-cell samples. From the Lin-Sca1+Kit□ population, there are clusters of single cells (the red lines separate different clusters in the heatmap of Lin-Sca1+Kit□ single-cell data) with nuocyte signature and PDC signature.

See also Figure S2 and Tables S2 and S7.

The unsupervised computationally constructed hierarchy shows high resemblance to the hematopoietic differentiation lineage tree (Figure 4A). Different cell lineages are readily separated into distinct branches, as revealed by the overlaid expression level of different gene clusters. Branches expressing *Kit* cluster and *Gypa* cluster genes correspond to stem and progenitor

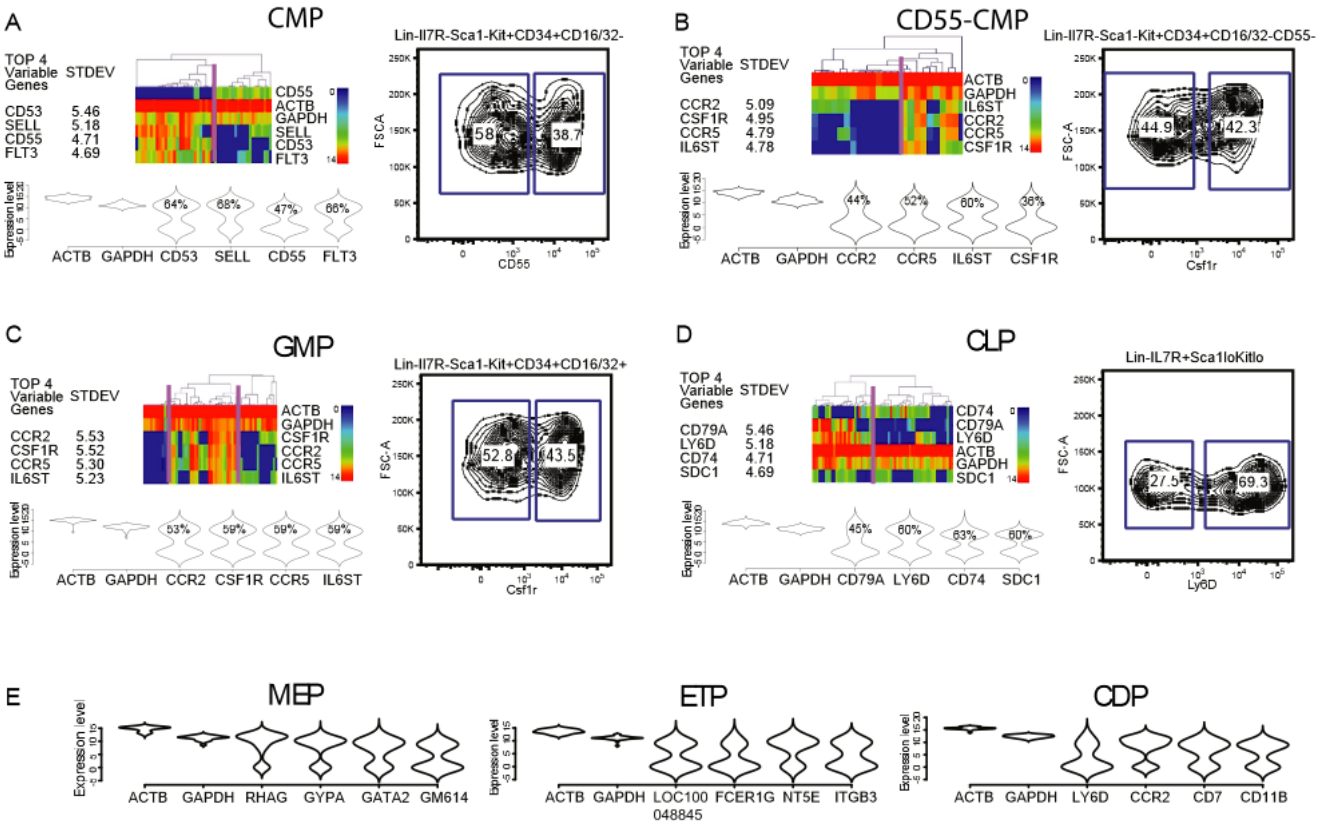


Figure 3. Dissection of Heterogeneity within Classical Progenitor Types

(A–D) Top four most variable genes are listed according to their standard deviation value within a particular progenitor cell type. The hierarchical clustering heatmap and violin density plot reveal the heterogeneity in the population. The percentages of cells with positive expression levels are marked on the violin plot. Color scale is as described in Figure 1B. FACS analysis confirms gene expression differences at protein level.

(E) Violin plots showing the expression pattern of top four most variable genes in MEP, ETP, and CDP progenitor populations.

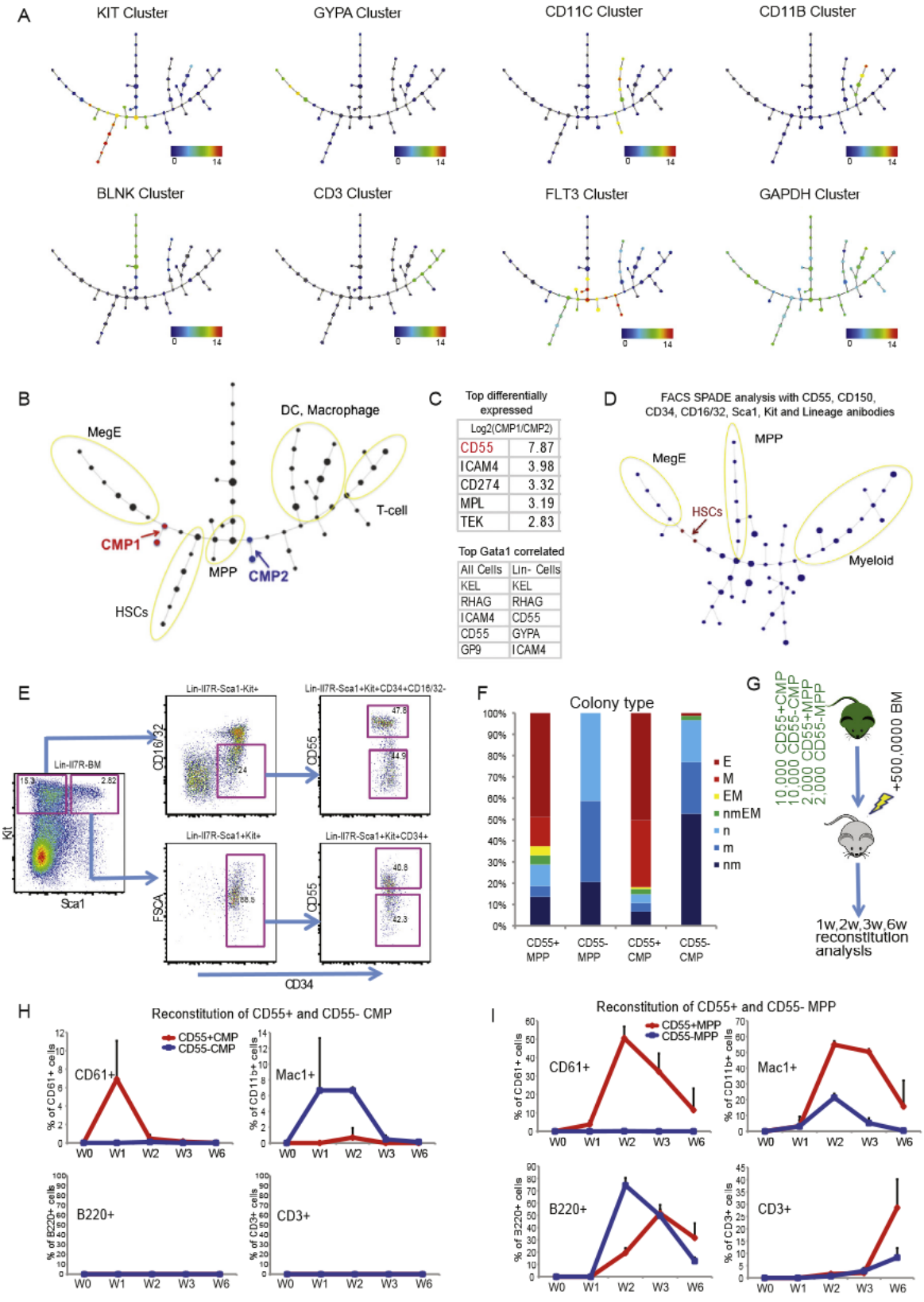
See also Figure S3 and Table S7.

and to MegE lineage cells, respectively. The dendritic, macrophage, B cell, and T cell branches, as well as lymphomyeloid progenitor cells, are marked by expression of *CD11c*, *CD11b*, *Blnk*, *CD3*, and *Flt3* clusters, respectively. The *Gapdh* endogenous control cluster is expressed broadly.

In the hierarchy generated from single-cell expression data, the MegE lineage branch is closely connected to the long-term repopulating HSC branch. These data suggest that the MegE lineage separates very early from lymphomyeloid lineage cells. Upon inspection of the composition of different nodes, we found that phenotypic CMP cells are located on two separate differentiation pathways, with half merged to the MegE lineage and half merged to the lymphomyeloid lineage (Figure 4B). This pattern is inconsistent with the conventionally portrayed, classical differentiation scheme that positions the MegE progenitor after the bifurcation of CMP and CLP and is reminiscent of an alternative model (Adolfsson et al., 2005; Pronk et al., 2007).

To validate this alternate scheme functionally, we sought to predict an early MegE lineage-specific marker from our data resource. We compared gene expression differences between the two separated CMP compartments (CMP1 and CMP2) and identified *CD55* as the most differentially expressed MegE marker (Figure 4C). In addition, we found that *CD55* expression strongly correlated with the *Gata1* transcription factor (Figure 4C and S4A), a master regulator of MegE lineage specification (Arinobu et al., 2007; Fujiwara et al., 1996; Iwasaki et al., 2003). FACS analyses indicate that Lin-Kit+Sca1 cells can be separated by CD55 into two main compartments (Figure S4B). To overcome the limitation of traditional two-dimensional gating strategy, we used SPADE analysis to analyze multidimensional FACS data from mouse bone marrow stained with CD55, CD150, CD34, CD16/CD32, Sca1, Kit, and lineage antibodies. We focused on Lin Kit+ data points and generated a simplified lineage tree with seven dimensional single-cell profiles. Consistent with our qPCR expression findings, the MegE lineage branch is closely connected with the HSC containing cell cluster nodes (Figures 4D and S4C), confirming early MegE specification.

We next separated CMP (Lin-IL7R-Sca1+Kit+CD34+CD16/CD32lo) and MPP (Lin-Sca1+Kit+CD34+) compartments into CD55+ and CD55 subpopulations (Figure 4E) and tested their function using in vitro colony-forming assays. Both CD55+ MPP and CD55+ CMP produce predominantly erythroid and megakaryocytic colonies, whereas few MegE colonies arise from CD55 MPP or CD55 CMP, revealing a functional difference in these early progenitor compartments (Figures 4F and



(legend on next page)

S4D). In order to confirm the early MegE separation in vivo, we used Actb-GFP mice for transplantation studies (Figure 4G). CD55+ CMPs transiently give rise to CD61+ platelets, whereas CD55− CMPs produce mainly myeloid cells (Figure 4H). CD55+ MPPs achieved more than 50% platelet reconstitution, whereas there was no reproducible contribution of CD55− MPPs to CD61+ platelets (Figures 4I and S4E). Importantly, CD55− CMPs and CD55− MPPs failed to produce platelets in vivo, whereas CD150− progenitors exhibited robust MegE potential (Pronk et al., 2007), suggesting that CD55 is an improved marker for separating early MegE progenitors. In conclusion, by computational analysis of single-cell data, we have predicted and validated CD55 as a marker to establish a functional separation between early MegE and lymphomyeloid differentiation at both CMP and MPP stages.

Genetic Network Construction by Single-Cell Analysis

To explore potential molecular mechanisms underlying early hematopoietic lineage specification, we designed primers to assay expression of an additional 180 genes, including lineage-specific transcription factors, epigenetic modifiers, and cell-cycle regulators. We assayed single cells from HSCs (CD48−CD34−CD150+LSK), MPP (CD34+LSK), CMP, MEP, GMP, and CLP populations (Table S4) and calculated gene expression covariance across the data set to uncover hidden regulatory links. We then used Cytoscape software to integrate expression correlations with published chromatin immunoprecipitation sequencing (ChIP-seq) binding data sets for ten major stem cell transcriptional regulators (including Scl/Tal1, Lyl1, Lmo2, Gata2, Runx1, Meis1, PU.1, Erg, Fli1, and Gfi1b) from HPC-7 cell line (Wilson et al., 2010). The network (Figure S5A) only depicts links in which the covariance was above 0.1 for correlated genes (green edges) or below −0.1 for anticorrelated genes (red edges). The network contains 76 nodes, connected through 71 edges between correlated genes and 74 edges between anticorrelated genes. Figure 5A highlights the transcription factor components of the complete network. As revealed, Gata2, a central hematopoietic stem cell regulator (Tsai et al., 1994; Wilson et al., 2010), lies at the core of the lineage specification pathway (Figure 5A), positively correlates with a MegE lineage module (characterized by *Gata1*, *Gfi1b*, *Nfia*, and *Klf1*), and negatively correlates with a lymphomyeloid module (characterized by *Flt3*, *Sell*, *Cebpa*, and *Notch1*). This is also depicted on a gene-to-gene correlation heatmap in Figure 5B. Time-course, single-cell tracing experiments suggest that upregulation or downregulation of Gata2 marks the first molecular event during colony formation (Figure S5B). The correlation in expression level between *Gata2* and *Gata1* is maintained during both in vivo and in vitro differentiation (Figure S5C). As revealed in Figure 5C, Gata2, Runx1, Meis1, Scl, Lyl1, and Lmo2 co-occupy at *Gata1* and *Gfi1b* regulatory regions (Figure 5C). The stem cell transcription factor Gata2 occupies regulatory elements of multiple MegE lineage-related genes, as well as HSC-enriched genes (Figure S5D).

As a functional test of this predicted genetic network, we examined the consequences of perturbation of the level of *Gata2*. Because *Gata2* −/− embryos die early due to hematopoietic failure (Tsai et al., 1994), we analyzed gene expression changes in viable *Gata2*+/− mice at single-cell resolution. Consistent with a previous report (Rodrigues et al., 2005), we observed a reduction in the size of the LSK population in Gata2 heterozygous mice as compared with wild-type. Single-cell gene expression analysis of *Gata2* +/+ and *Gata2* +/− LSK reveals that haploinsufficiency is associated with an altered regulatory network during early lineage differentiation (Figures 5D and S5E), revealing sensitivity of the network to modest quantitative changes in *Gata2* expression. Haploinsufficiency for *Gata2* leads to downregulation of the MegE marker *Gfi1b* and *Gata1* and upregulation of lymphomyeloid markers, including *Flt3*, *Sell*, *CD34*, *CD53*, and *Cebpa* in hematopoietic stem and progenitor cell populations (Figures 5D and S5E). Taken together, single-cell-level gene expression in combination with functional studies validates a genetic network underlying early differentiation of MegE lineage from the lymphomyeloid lineage.

MegE Priming in the Most Primitive HSCs

In single-cell, in vitro tracing experiments (Figure S5B), we noticed that pure megakaryocytic colonies are the first to emerge in cultures of HSCs. These results encouraged us to investigate the heterogeneity and existing MegE network within the most primitive HSC (CD48−CD34−CD150+LSK) population. Remarkably, a MegE module, characterized by expression of

Figure 4. Mapping Cellular Hierarchy by Lineage Progression Analysis

(A) Spanning-tree progression analysis of density-normalized events from single-cell expression pattern of 40 gene clusters. The information regarding each cluster is listed in Table S3. Overlaid expression pattern of different gene clusters helped to define distinct cell lineages. Color scale is as described in Figure 1B.

(B) Single-cell SPADE hierarchy suggests early separation of MegE lineage and lymphomyeloid lineage.

(C) CD55 is the most differentially expressed MegE cell surface marker between the CMP1 and CMP2. In addition, *CD55* is highly correlated with *Gata1* expression across the single-cell data set.

(D) SPADE analysis of FACS data from mouse bone marrow stained with CD55, CD150, CD34, CD16/CD32, Sca1, Kit, and lineage antibodies. Only Lin-Kit+ (as defined by lineage signal < 1,000, kit signal > 1,000) data points are included in the analysis to reduce complexity. The two cell cluster nodes that contain most of the HSCs (as defined by CD150 signal > 1,000, CD48 signal < 1,000, CD34 signal < 1,000, Sca1 signal > 1,000) containing nodes are labeled in red. The two nodes are closely related with MegE branch, as defined by overlaid expression in Figure S4C.

(E) CD55 can be used to separate both CMP and MPP progenitors.

(F) In vitro colony-forming assays using erythropoietin (EPO)-containing methylcellulose suggest that CD55 divides both CMP and MPP into functionally different subpopulations. One hundred fifty cells from each population were plated in 1.5 ml of Methocult M3434 (Stem Cell Technologies) in duplicates. E, erythrocytes; M, megakaryocytes; m, monocytes; n, neutrophils.

(G–I) Reconstitution experiment using Actb GFP mice validated the early separation of MegE lineage potential and lymphomyeloid potential in vivo. X-axes correspond to sampling time. Y-axes correspond to GFP% of the total reconstituted cells. Mice are irradiated by two doses of 5 Gy with a 4 hr interval. Data are represented as mean of four biological replicates for each group. The error bars represent SD.

See also Figure S4 and Tables S3 and S7.

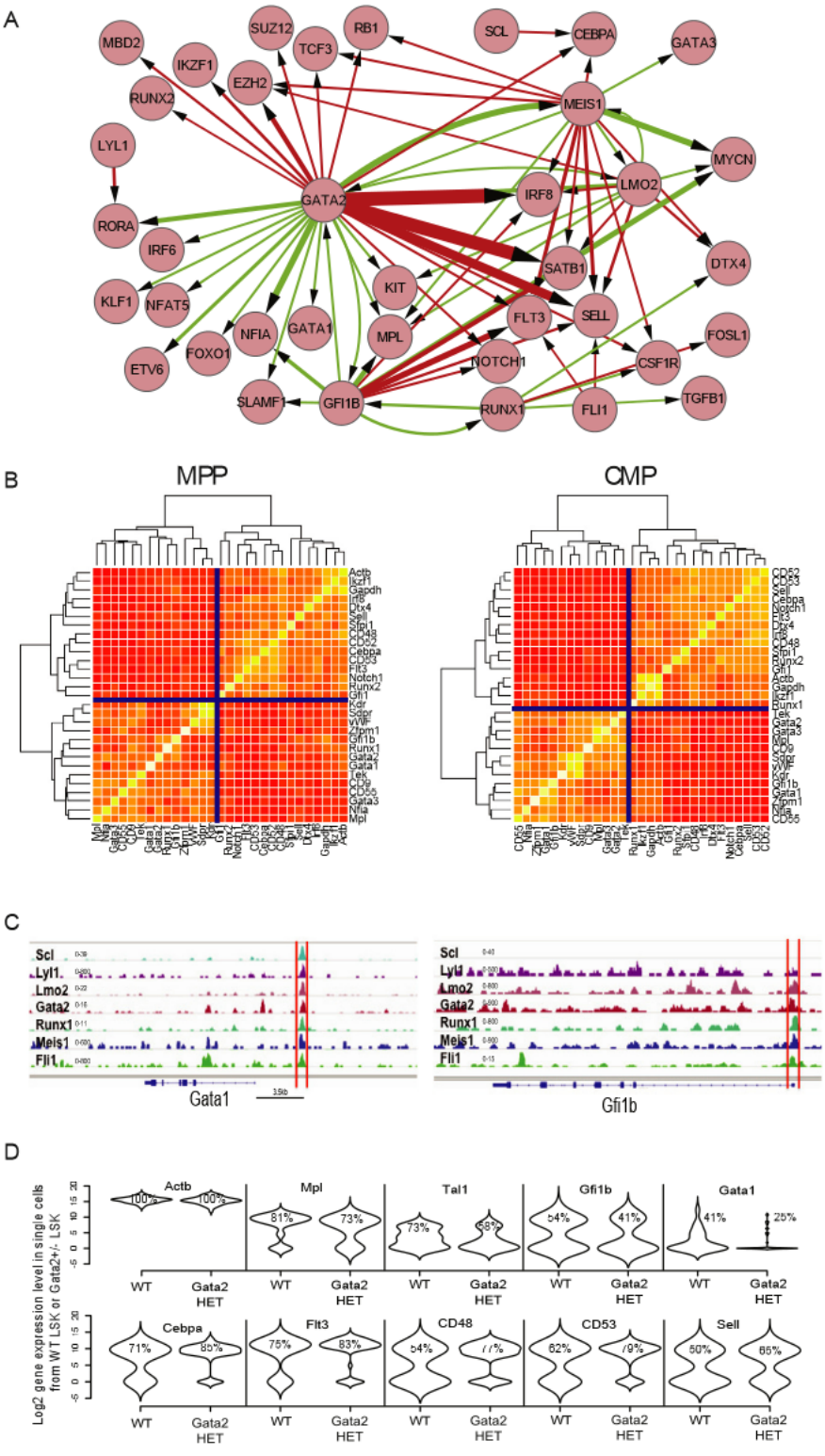


Figure 5. Genetic Regulation during HSC Differentiation

(A) A genetic network constructed by Cytoscape using transcription factor ChIP-seq binding information and single cell-level gene expression correlation data. It highlights transcription factors components within the complete network in Figure S5A. Green arrow corresponds to positive correlation, whereas red arrow corresponds to anticorrelation. The width of the line corresponds to absolute value of the covariance between two linked gene nodes.

(B) Gene-to-gene correlation heatmaps containing HSC, MegE, myeloid, and lymphoid modules in MPP and CMP.

(C) HSC module transcription factors co-occupy *Gata1* and *Gfi1b* upstream region.

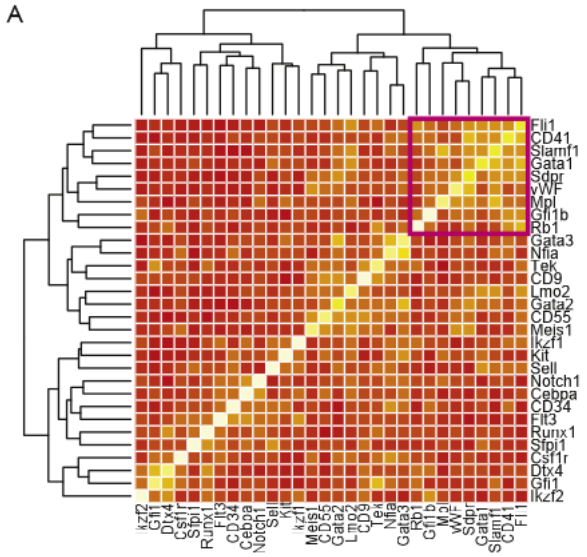
(D) Gene expression level distribution of LSK single cells from wild-type and *Gata2* heterozygous mice are presented with violin density plots. The percentages of cells with positive expression levels are marked on the violin plots. Note the decrease in MegE-primed cells (*Gfi1b*+ or *Gata1*+) cells and increase in lymphomyeloid primed cells (*Cebpa*+, *Flt3*+, *CD53*+, or *Sell*+) in the *Gata2* +/□ LSK population.

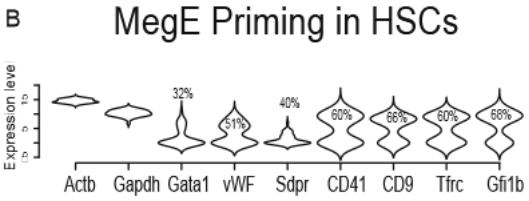
See also Figure S5 and Tables S4 and S7.

scriptional priming does not appear to be stochastic but rather controlled by an intertwined HSC regulatory network. We ranked single HSCs by the expression level of *Gata2* and compared gene expression between *Gata2*high HSCs (top 50%) and *Gata2*int HSCs (bottom 50%). *CD150* emerged as a candidate marker for separating HSCs according to different levels of *Gata2* expression (Figure 6C), as well as different degrees of MegE priming (Figure S6C). To confirm these differences, we FACS-sorted HSCs into CD150high and CD150int compartments for gene expression analysis (Figure 6D). Indeed, CD150high HSCs express higher levels of *Gata2*, *Gata1*, and *CD61*, as well as other MegE lineage-related genes (Figure 6E). In colony-forming assays, CD150high HSCs generate greater numbers of MegE lineage-containing colonies than CD150int HSCs (Figure 6F). Similar biased differentiation readouts were also seen in HSCs that were separated by relative expression levels of CD55, CD41, or CD9 (Figure S6D). These results suggest that MegE differentiation bias is already established at the HSC level.

Fli1, *CD41*, *CD150*, *Gata1*, *vWF*, *Mpl*, and *Gfi1b*, maintains high correlation in single HSCs (Figure 6A). MegE lineage-specific gene expression is detected in HSCs purified by different enrichment protocols and further confirmed by single-cell NanoString technology (Figures 6B, S6A, and S6B; Table S5). Such tran-

The positive correlation of *Gata2* with the MegE priming expression suggests that the regulatory network within HSCs is intrinsically unstable. As such, higher levels of *Gata2* in

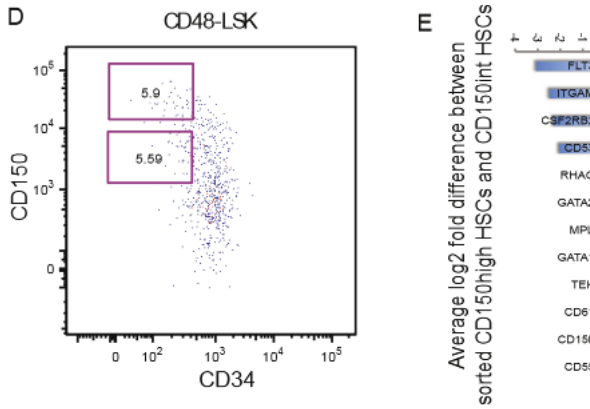


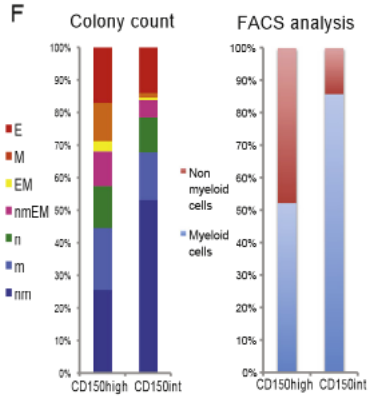


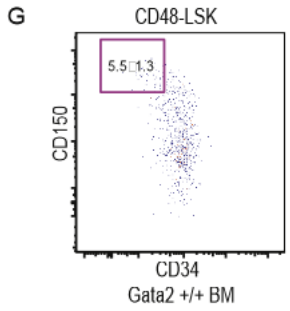
C

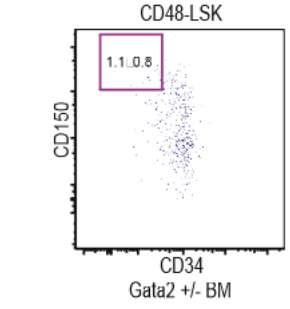
Top 5 differentially expressed genes between Gata2high HSCs and Gata2int HSCs
Average Log2(Gata2high/Gata2int)

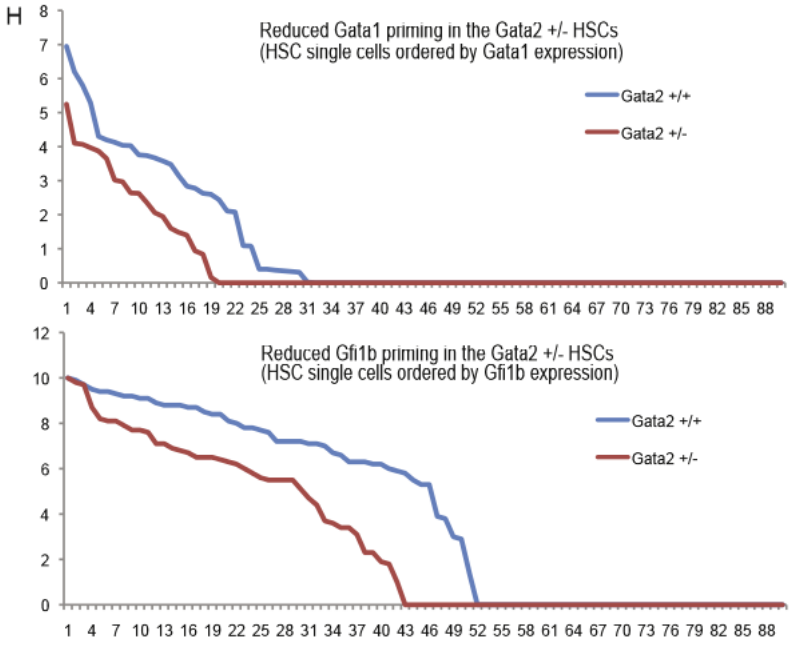
Top		Bottom	
CD150	3.33	FLT3	-3.01
CD41	2.34	CD34	-2.72
GATA1	2.24	CD11B	-2.13
MYCN	2.14	CEBPA	-1.91
GATA2	1.98	LY6A	-1.81











(legend on next page)

HSCs may activate MegE lineage expression and promote MegE lineage skewing. When stained with the full panel of HSC markers, we observed a reduced number of CD150high HSCs in the *Gata2* haploinsufficient state (Figure 6G). In addition, in the most primitive HSCs of *Gata2* +/− mice, we observed a reduction in the number of *Gata1*+ or *Gfi1b*+ HSCs, as well as the average level of MegE priming (Figure 6H). Consistent with these findings, overexpression of Gata2 has been reported to promote MegE differentiation (Huang et al., 2009; Kitajima et al., 2006). The characterization of MegE priming in HSCs supports the cellular hierarchy and genetic network derived from single-cell expression data and illustrates the power of single-cell analysis in detecting the earliest regulatory events during stem cell differentiation.

Single-Cell Analysis of the AML Cellular Hierarchy

Having obtained a comprehensive data set in the wild-type hematopoietic system, we next applied the single-cell expression approach to characterization of LSCs in MLL-AF9-driven AML, a clinically relevant model of hematopoietic malignancy (Krivtsov et al., 2006; Neff et al., 2012). In this model, LSCs resemble GMPs and are hence described as LGMPs. Others, however, have described alternative cellular hierarchies of AML (Gibbs et al., 2012; Somervaille and Cleary, 2006). We generated MLL-AF9 primary leukemia in mice (Neff et al., 2012) and profiled single cells of the originally defined LGMP LSC population (Lin-Il7r-Kit+Sca1−CD34+CD16/CD32+), as well as the leukemic Lin+ (LLin+) population from bone marrow (Figure 7A and Table S1). As shown in Figure 7B, hierarchical clustering of gene expression data from leukemia cells and the wild-type myeloid cells reveals clear separation of the two groups. The LLin+ clusters closely with a group of LGMP cells, suggesting that lineage marker expression does not define a clear hierarchy in the leukemia. Two strong gene clusters are observed in the leukemia cells: a *Csf1r*, *Ccr2, Ccr5* cluster and a *CD24*, *Vcam1*, *CD133* cluster (Figure 7B). We adapted SPADE to analyze the data (Figure 7C). To allow for comparison of the wild-type and leukemia lineages, we extracted 40 clusters from the combined data sets of LGMP, LLin+, GMP, and Lin+ single cells (Table S6). We then used these clusters to infer lineage hierarchy for both cellular systems. From the overlaid expression level of different gene clusters, we observed clear separation of the *CD24*+ lineage branch and the *Csf1r*+ lineage branch within the tested leukemia cells (Figure 7C). By comparing the two main leukemia cell type signatures with other hematopoietic cell types, we find that MLL-AF9 leukemia cells display a unique signature with high expression of *Lamp1*, *Lamp2*, *Ifngr1*, *CD47*, and *CD33* (Figure 7D). Notably, the leukemia cellular state differs from other hematopoietic cellular states both at the single cell and population levels.

In the previously defined LGMP population, in which LSCs are highly enriched (Krivtsov et al., 2006), we observed clear heterogeneity. Guided from single-cell data, we separated the LGMP into two populations using CD24 antibody (Figure 7E). To assess potential functional difference of these two compartments, we transplanted each into sublethally irradiated secondary recipients. Both CD24− LGMP and CD24+ LGMP are capable of initiating AML (Figure 7F). However, mice transplanted with CD24+ LGMPs exhibited a marked delay in disease progression. Analysis of the bone marrow from secondary leukemia mice indicated that CD24− leukemia cells and CD24+ leukemia cells maintain their respective signatures and fail to reconstitute each other during clonal expansion (Figure 7G). Thus, CD24 marks two distinct, self-renewing clones within MLL-AF9-driven AML. Further profiling of additional intracellular regulators reveals different genetic programs used by CD24− LGMP and CD24+ LGMP (Figure S7A). Interestingly, Ezh2, a core polycomb repressive complex 2 (PRC2) component, is overexpressed in CD24− LGMPs (Figure S7A). Our analysis also reveals high variation of *Ezh2* at the single-cell level, which strongly correlates with *Ccna2*, *Ccnb1*, and *Ccnb2* expression (Figure S7B). Such correlation may account in part for the more aggressive behavior of the CD24− Ezh2high leukemia clone, as compared with the CD24+ Ezh2low leukemia clone. In microarray data of synchronized HeLa cells (Whitfield et al., 2002), *Ezh2* expression is lowest in G1 and peaks at S phase (Figure S7C). In addition, many cell-cycle regulators are direct targets of PRC2, as assessed from PRC2 chromatin occupancy data (Figures S7D–S7G). Moreover, inhibition of Ezh2 function with the specific inhibitor GSK126 (McCabe et al., 2012) leads to an increase in G1 phase cells and a decrease in S phase cells in MLL-AF9 cultures (Figure S7H). Our findings are in general agreement with the observation that EZH2 overexpression correlates with poor prognosis in several tumor types (Cavalli, 2012; McCabe et al., 2012).

DISCUSSION

Single-cell analysis technologies provide a powerful approach to the study of rare cell types and cell heterogeneity. For both genome analysis and transcriptome analysis of single cells, amplification of small amounts of material is required and presents technical challenges. For assessment of gene expression,

Figure 6. MegE Lineage Priming in HSCs

(A) Gene-to-gene correlation heatmaps reveal correlation of MegE lineage markers in single cells from HSCs (CD48−CD34−CD150+LSK).

(B) Violin plot suggests significant MegE lineage priming in HSCs (CD48−CD34−CD150+LSK).

(C) CD150 stands out as the top differentially expressed gene between Gata2high HSCs and Gata2int HSCs.

(D) FACS of CD150high and CD150int HSCs.

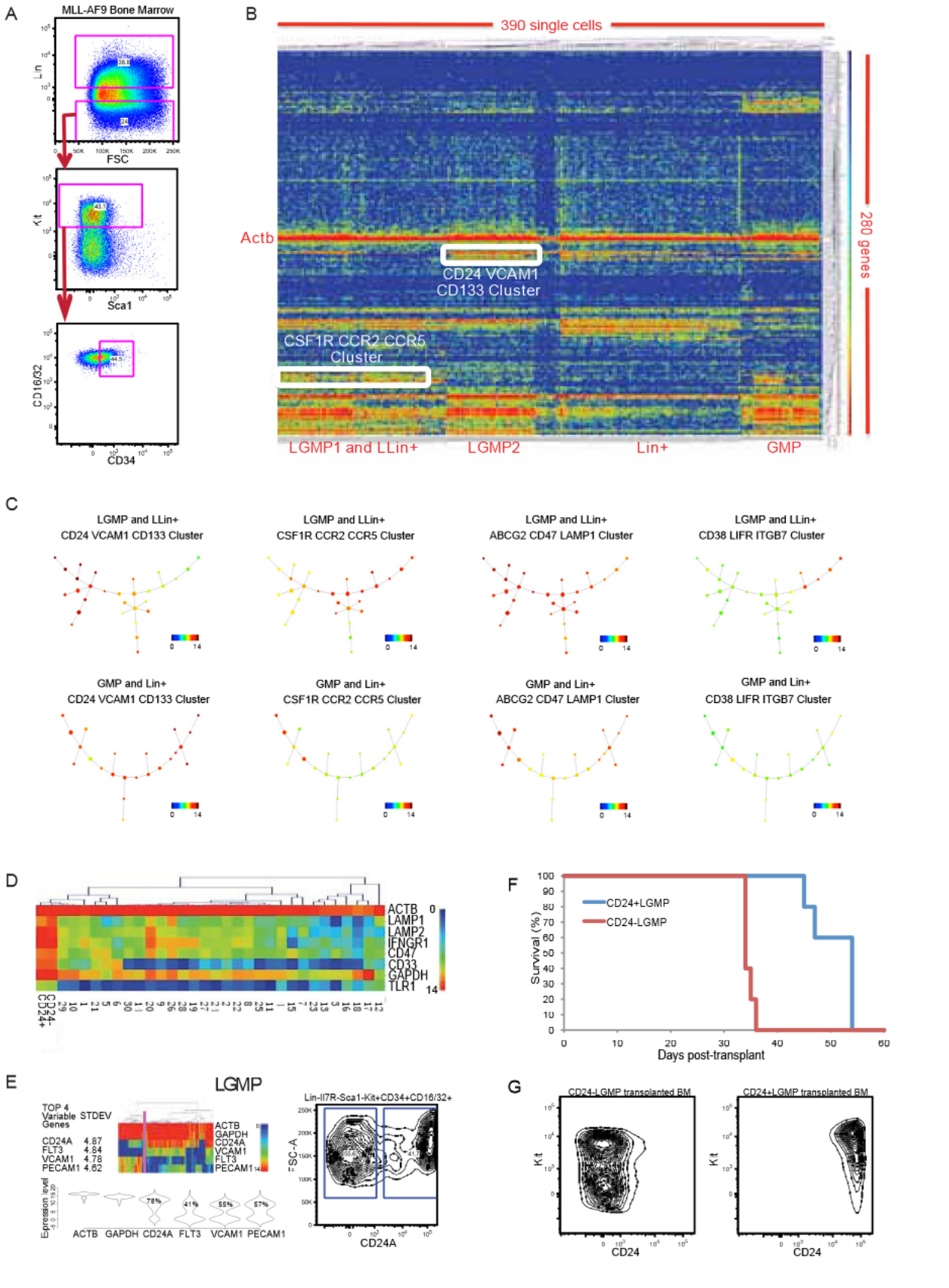
(E) Gene expression difference between the sorted CD150high versus CD150int HSCs.

(F) In vitro colony-forming assays using Methocult M3434 (Stem Cell Technologies) suggest that CD150high HSCs produce more MegE lineage-containing colonies than the CD150int HSCs. FACS analysis of day 7 methylcellulose cultures also suggests a decreased percentage of CD11b+ or Gr1+ myeloid cells generated from the CD150high HSCs when compared to CD150int HSCs. CD11b− and Gr1− cells were defined as nonmyeloid cells.

(G) Gata2 haploinsufficiency results in a reduction of CD150high HSCs. Three animals were analyzed for each genotype; results are shown as mean ± SD.

(H) Gata2 haploinsufficiency results in a reduction of *Gata1* and *Gfi1b* priming in the HSC compartment. A total of 87 single cells were analyzed for each genotype. Single cells are ordered by *Gata1* or *Gfi1b* expression.

See also Figure S6 and Tables S5 and S7.



(legend on next page)

single-cell, high-throughput qPCR has several advantages. First, it utilizes one-tube one-step single-cell sequence-specific preamplification, which involves minimal sample handling time and allows for high-throughput complementary DNA (cDNA) library generation. Second, the targeted PCR approach enables specific amplification of lowly expressed genes. Finally, Biomark (Fluidigm) microfluidic qPCR permits well-controlled, parallel analysis of 96 single-cell samples. The system minimizes technical variation, allowing for comparison of different samples without normalization. A major challenge relates to primer dimer formation during multiplexed, sequence-specific preamplification step, which generates false-positive signals (Guo et al., 2010). To overcome this obstacle, we have introduced multiplexed primer design, lowered the preamplification primer concentration, and included nested primers to avoid primer dimer signals. These optimizations significantly increased the throughput of single-cell qPCR technology and permitted analysis of the cell surface repertoire in many single cells from a broad range of tissue types. We show with functional validation that such data sets can be used for classification of cell type, dissection of heterogeneity, mapping of cellular hierarchy, and computational construction of genetic networks.

We have applied our analysis to examine cellular lineages of the mouse hematopoietic system. Computational lineage progression analysis provided an unbiased view of cellular state transition during differentiation from HSCs. Our findings independently support an alternative hematopoietic hierarchy first proposed by Jacobsen and colleagues (Adolfsson et al., 2005) and provide a molecular model for early MegE lineage separation. The sensitivity of the assay allowed detection of coordinated MegE transcriptional priming within the most highly enriched HSCs. We then extended the analysis to a robust, clinically relevant model of AML. At the level of single cells, we showed that leukemia cells are intrinsically distinct from any of the wild-type hematopoietic lineages. Interestingly, the most significant heterogeneity within the leukemia corresponded to two independent, disease-initiating clones. Moreover, we found that *Ezh2* is overexpressed in the more highly proliferative leukemia cells and uncovered a link with cell-cycle progression.

No two cells are identical, a concept evident in our data set. Even the most closely related cells, which correspond to two erythroid progenitor cells from our complete data-clustering heatmap (Figure 2B), exhibit differences in gene expression patterns. Variation detected in single-cell gene expression data may reflect biological and/or technical noise or may correspond to function. Distinguishing these two types of variation is very important. Here, we assayed many single-cell samples and searched for correlated gene clusters rather than variation in expression of individual genes. Such correlated gene expression behavior is more likely to represent genetic network function rather than biological noise. By this approach, we correlated MegE priming in the HSCs and correlated *Ezh2* dynamics with cell cycle, which were then both functionally validated. Single-cell gene expression data is extremely valuable for extracting such correlated gene expression clusters, because single cells represent the fundamental unit of genetic network regulation.

The mammalian epigenetic landscape contains numerous transitional cellular states within lineage differentiation pathways. Comprehensive mapping of this landscape requires single-cell gene expression analysis in order to represent all possible states. Such an assay needs to be both quantitative and thorough, so that the data are experimentally robust and reflects all major cell types. We suggest that the strategy described here satisfies these requisites. In order to validate biological differences in cell populations, we have relied extensively on study of cell surface markers, as available antibodies can then be used for prospective cell isolation. The approach is readily applicable to other biological contexts. Its use should facilitate identification of new surface markers for functional assessment of stem and progenitor cells and the construction of cellular hierarchies in other organ systems. The strategy is suitable for deconvoluting cellular heterogeneity within different types of cancers. Further accumulation of data sets from diverse contexts should eventually allow for the mapping of all nonredundant cellular states on the mammalian differentiation hierarchy.

EXPERIMENTAL PROCEDURES

Multiplexed Primer Design for Single-Cell Analysis

Gene symbol list for commonly used surface markers is summarized from two resources: a comprehensive mouse cell surface antigens review paper (Lai et al., 1998) and the eBioscience website mouse cellular antigen charts (http://www.ebioscience.com/resources/mouse-cd-chart.htm). Gene symbols are then converted to mRNA refseq ID by DAVID tools (http://david.abcc.ncifcrf.gov/). mRNA sequences for each gene are retrieved from University of California Santa Cruz table browser; only common regions are used for genes with different isoforms. Multiplexed, gene-specific primers are designed using a Primer3-based (http://primer3.wi.mit.edu/) algorithm to ensure that each primer within the designed pool has a maximum complimentary sequence of 7 bp to all the other primers. All primers (Table S7) are synthesized and provided by Boston Open Labs (http://bolresearch.com/).

Figure 7. Single-Cell Analysis of the AML Cellular Hierarchy

(A) Single-cell sorting strategy for different leukemia compartments in the MLL-AF9 AML mouse model.

(B) A heatmap showing hierarchical clustering of gene expression signatures from 390 single cells from wild-type myeloid cells (Lin-Il7r-Kit+Sca1□CD34+CD16/CD32+ or Lin+ bone marrow) and MLL-AF9 primary leukemia cells (Lin-Il7r-Kit+Sca1□CD34+CD16/CD32+ or Lin+ bone marrow). Each row corresponds to a specific gene; each column corresponds to a particular single cell. White boxes highlight strongly correlated gene and cell clusters. Color scale is as described in Figure 1B.

(C) SPADE analysis of the wild-type myeloid hierarchy and leukemia hierarchy using the high-dimensional, single-cell data. Overlaid expression pattern of different gene clusters helped to define distinct cell lineages in the MLL-AF9 leukemia system.

(D) Gene expression clustering heatmap of gene expression from the main leukemia cell clusters and all the main hematopoietic cell clusters reveals distinct leukemia-specific expression.

(E) Dissection of heterogeneity in the LGMP (Lin-Il7r-Kit+Sca1□CD34+CD16/CD32+) according to described method in Figure 3.

(F) Survival of secondary recipient mice receiving 800 CD24+ LGMP or CD24□ LGMP cells.

(G) Reconstitution of two leukemia lineages in the secondary recipient bone marrow.

See also Figure S7 and Tables S6 and S7.

FACS and Single-Cell Collection

Seven- to twelve-week-old C57Bl/6 mice or Actb-GFP C57Bl/6 transgenic mice were used throughout this study (except for fetal liver and aged mice). Bone marrow cells were isolated by crushing iliac crest bones, femurae, and tibiae in PBS containing 5% fetal calf serum and 2 mM EDTA. After red blood cell lysis, the remaining cells were stained with monoclonal antibodies, analyzed, and sorted on the BD FACSAria II (BD Bioscience). Individual cells were sorted directly into 96 well PCR plates loaded with PCR buffer under single-cell mode. Monoclonal antibodies and conjugations used in this study are found in Table S7. All data were analyzed with FlowJo (Tree Star).

One-Tube Single-Cell Sequence-Specific Preamplification

Individual primer sets (total of 300) were pooled to a final concentration of 0.1 mM for each primer. Individual cells were sorted directly into 96 well PCR plates loaded with 5 ml RT-PCR master mix (2.5 ml CellsDirect reaction mix, Invitrogen; 0.5 ml primer pool; 0.1 ml RT/Taq enzyme, Invitrogen; 1.9 ml nuclease free water) in each well. Sorted plates were immediately frozen on dry ice. After brief centrifugation at 4°C, the plates were immediately placed on PCR machine. Cell lyses and sequence-specific reverse transcription were performed at 50°C for 60 min. Then, reverse transcriptase inactivation and Taq polymerase activation was achieved by heating to 95°C for 3 min. Subsequently, in the same tube, cDNA went through 20 cycles of sequence-specific amplification by denaturing at 95°C for 15 s, annealing, and elongation at 60°C for 15 min. After preamplification, PCR plates were stored at −80°C to avoid evaporation.

High-Throughput Microfluidic Real-Time PCR

Preamplified products were diluted 5-fold prior to analysis. Amplified single-cell samples were analyzed with Universal PCR Master Mix (Applied Biosystems), EvaGreen Binding Dye (Biotium), and individual qPCR primers using 96.96 Dynamic Arrays on a BioMark System (Fluidigm). Three Dynamic Arrays loaded with different primer sets were used for each sample plate. Threshold crossing (Ct) values were calculated using the BioMark Real-Time PCR Analysis software (Fluidigm).

Single-Cell NanoString

Reporter probes are designed and synthesized by NanoString R&D team. Target sequences are amplified from single cells using one-tube single-cell sequence-specific preamplification as described before. Twenty-five percent of the amplified cDNA are subject to gene expression quantification using the GEN2 Digital Analyzer. Raw counts are compiled, normalized, and analyzed using nSolver. The data are then subtracted with the background signal and transformed to Log2 scale before analysis.

Computational Processing of Single-Cell Data

A background Ct of 28 was used for all real-time signals. Samples with low *Actb* expression level (Ct higher than 18) are outliers of normal distribution and are excluded from the analysis. These samples had low or no expression for all the other genes, suggesting that they correspond to empty wells or bad single-cell samples. Hierarchical clustering was done with MultiExperiment Viewer program. For all hierarchical clustering heatmaps, the rainbow scheme color scale is set from 0 to 14, corresponding to Log2 gene expression above background of 28. GEDI plots are generated using the gene expression dynamics inspector. Each pixel on the 10.10 GEDI map corresponds to a particular minigene cluster generated by the software. Violin plot, box plot, and correlation heatmap were generated with R software. SPADE analysis was performed with Matlab. Lineage specific gene lists for the 180 intracellular regulator assay set and for Figure S5D are generated from the Immgen website analysis tool. ChIP-seq peak visualization was done with Integrative Genomics Viewer program. The genetic networks in Figures 5A and S5A were constructed using Cytoscape 3 software.

SUPPLEMENTAL INFORMATION

Supplemental Information for this article includes Supplemental Experimental Procedures, seven figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2013.07.017.

ACKNOWLEDGMENTS

We thank H. Skaletsky from Whitehead Institute for extensive help on the multiplexed primer design; Y. Fujiwara, E. Baena, O. Yilmaz, M. Nguyen, X. Han, V. Bragt, D. Linn, and J. Buchman for help with different parts of the sample preparations; and H. Huang, Z. Li, D. Scadden, H. Xie, and H. Hock for insightful discussions on the project. This work was supported by funding from the National Institutes of Health and the Harvard Stem Cell Institute (S.H.O). S.H.O. is an investigator of the Howard Hughes Medical Institute.

Received: May 24, 2013
Revised: July 4, 2013
Accepted: July 22, 2013
Published: September 12, 2013

REFERENCES

Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.J., Thoren, L.A., et al. (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell *121*, 295–306.

Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature *404*, 193–197.

Arinobu, Y., Mizuno, S., Chong, Y., Shigematsu, H., Iino, T., Iwasaki, H., Graf, T., Mayfield, R., Chan, S., Kastner, P., and Akashi, K. (2007). Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. Cell Stem Cell *1*, 416–427.

Bendall, S.C., Simonds, E.F., Qiu, P., Amir-el, A.D., Krutzik, P.O., Finck, R., Bruggner, R.V., Melamed, R., Trejo, A., Ornatsky, O.I., et al. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science *332*, 687–696.

Bonnet, D., and Dick, J.E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat. Med. *3*, 730–737.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209–1222.

Cavalli, G. (2012). Molecular biology. EZH2 goes solo. Science *338*, 1430–1431.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544–547.

Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P.S., Rothenberg, M.E., Leyrat, A.A., Sim, S., Okamoto, J., Johnston, D.M., Qian, D., et al. (2011). Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat. Biotechnol. *29*, 1120–1127.

Fujiwara, Y., Browne, C.P., Cunniff, K., Goff, S.C., and Orkin, S.H. (1996). Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. Proc. Natl. Acad. Sci. USA *93*, 12355–12358.

Gibbs, K.D., Jr., Jager, A., Crespo, O., Goltsev, Y., Trejo, A., Richard, C.E., and Nolan, G.P. (2012). Decoupling of tumor-initiating activity from stable immunophenotype in HoxA9-Meis1-driven AML. Cell Stem Cell *10*, 210–217.

Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Dev. Cell *18*, 675–685.

Huang, Z., Dore, L.C., Li, Z., Orkin, S.H., Feng, G., Lin, S., and Crispino, J.D. (2009). GATA-2 reinforces megakaryocyte development in the absence of GATA-1. Mol. Cell. Biol. *29*, 5168–5180.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.B., Lönnerberg, P., and Linnarsson, S. (2012). Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. Nat. Protoc. *7*, 813–828.

Iwasaki, H., Mizuno, S., Wells, R.A., Cantor, A.B., Watanabe, S., and Akashi, K. (2003). GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. Immunity *19*, 451–462.

Kiel, M.J., Yilmaz, O.H., Iwashita, T., Yilmaz, O.H., Terhorst, C., and Morrison, S.J. (2005). SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. Cell *121*, 1109–1121.

Kitajima, K., Tanaka, M., Zheng, J., Yen, H., Sato, A., Sugiyama, D., Umehara, H., Sakai, E., and Nakano, T. (2006). Redirecting differentiation of hematopoietic progenitors by a transcription factor, GATA-2. Blood *107*, 1857–1863.

Kondo, M., Weissman, I.L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell *91*, 661–672.

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.

Lai, L., Alaverdi, N., Maltais, L., and Morse, H.C., 3rd. (1998). Mouse cell surface antigens: nomenclature and immunophenotyping. J. Immunol. *160*, 3861–3868.

McCabe, M.T., Ott, H.M., Ganji, G., Korenchuk, S., Thompson, C., Van Aller, G.S., Liu, Y., Graves, A.P., Della Pietra, A., 3rd, Diaz, E., et al. (2012). EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature *492*, 108–112.

Moignard, V., Macaulay, I.C., Swiers, G., Buettner, F., Schütte, J., Calero-Nieto, F.J., Kinston, S., Joshi, A., Hannah, R., Theis, F.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*, 363–372.

Morrison, S.J., and Weissman, I.L. (1994). The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. Immunity *1*, 661–673.

Morrison, S.J., Wandycz, A.M., Hemmati, H.D., Wright, D.E., and Weissman, I.L. (1997). Identification of a lineage of multipotent hematopoietic progenitors. Development *124*, 1929–1939.

Muller-Sieburg, C.E., Whitlock, C.A., and Weissman, I.L. (1986). Isolation of two early B lymphocyte progenitors from mouse marrow: a committed pre-pre-B cell and a clonogenic Thy-1-lo hematopoietic stem cell. Cell *44*, 653–662.

Neff, T., Sinha, A.U., Kluk, M.J., Zhu, N., Khattab, M.H., Stein, L., Xie, H., Orkin, S.H., and Armstrong, S.A. (2012). Polycomb repressive complex 2 is required for MLL-AF9 leukemia. Proc. Natl. Acad. Sci. USA *109*, 5028–5033.

Neill, D.R., Wong, S.H., Bellosi, A., Flynn, R.J., Daly, M., Langford, T.K., Bucks, C., Kane, C.M., Fallon, P.G., Pannell, R., et al. (2010). Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. Nature *464*, 1367–1370.

Onai, N., Obata-Onai, A., Schmid, M.A., Ohteki, T., Jarrossay, D., and Manz, M.G. (2007). Identification of clonogenic common Flt3+M-CSFR+ plasmacytoid and conventional dendritic cell progenitors in mouse bone marrow. Nat. Immunol. *8*, 1207–1216.

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

Osawa, M., Hanada, K., Hamada, H., and Nakauchi, H. (1996). Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. Science *273*, 242–245.

Pronk, C.J., Rossi, D.J., Månsson, R., Attema, J.L., Norddahl, G.L., Chan, C.K., Sigvardsson, M., Weissman, I.L., and Bryder, D. (2007). Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. Cell Stem Cell *1*, 428–442.

Qiu, P., Simonds, E.F., Bendall, S.C., Gibbs, K.D., Jr., Bruggner, R.V., Linderman, M.D., Sachs, K., Nolan, G.P., and Plevritis, S.K. (2011). Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. Nat. Biotechnol. *29*, 886–891.

Ramsköld, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. *30*, 777–782.

Rodrigues, N.P., Janzen, V., Forkert, R., Dombkowski, D.M., Boyd, A.S., Orkin, S.H., Enver, T., Vyas, P., and Scadden, D.T. (2005). Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. Blood *106*, 477–484.

Somervaille, T.C., and Cleary, M.L. (2006). Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. Cancer Cell *10*, 257–268.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382.

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., and Surani, M.A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. Cell Stem Cell *6*, 468–478.

Tsai, F.Y., Keller, G., Kuo, F.C., Weiss, M., Chen, J., Rosenblatt, M., Alt, F.W., and Orkin, S.H. (1994). An early haematopoietic defect in mice lacking the transcription factor GATA-2. Nature *371*, 221–226.

Visser, J.W., Bauman, J.G., Mulder, A.H., Eliason, J.F., and de Leeuw, A.M. (1984). Isolation of murine pluripotent hemopoietic stem cells. J. Exp. Med. *159*, 1576–1590.

Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O., and Botstein, D. (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol. Biol. Cell *13*, 1977–2000.

Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E., et al. (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell *7*, 532–544.