Chromatin dynamics during hematopoiesis reveal discrete regulatory modules instructing differentiation

Grigorios Georgolopoulos[1,2], Mineo Iwata[1], Nikoletta Psatha[1], Andrew Nishida[1], Tannishtha Som[1], Minas Yiangou[2], John A. Stamatoyannopoulos[1,3,4], Jeff Vierstra[1]*

[1]Altius Institute for Biomedical Sciences, Seattle, WA, USA
[2]Department of Genetics, Development & Molecular Biology, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece
[3]Department of Genome Sciences, University of Washington, Seattle, Washington, USA
[4]Division of Oncology, Department of Medicine, University of Washington, Seattle, Washington, USA

***Correspondence:** (JV) jvierstra@altius.org

Abstract

Lineage commitment and differentiation is driven by the concerted action of master transcriptional regulators at their target chromatin sites. Multiple efforts have characterized the key transcription factors (TFs) that determine the various hematopoietic lineages. However, the temporal interactions between individual TFs and their chromatin targets during differentiation and how these interactions dictate lineage commitment remains poorly understood. We performed dense, daily, temporal profiling of chromatin accessibility (DNase I-seq) and gene expression changes (total RNA-seq) along *ex vivo* human erythropoiesis to comprehensively define developmentally regulated DNase I hypersensitive sites (DHSs) and transcripts. We link both distal DHSs to their target gene promoters and individual TFs to their target DHSs, revealing that the regulatory landscape is organized in distinct sequential regulatory modules that regulate lineage restriction and maturation. Finally, direct comparison of transcriptional dynamics (bulk and single-cell) and lineage potential between erythropoiesis and megakaryopoiesis illuminates the fine-scale temporal dynamics of these regulatory modules during lineage-resolution between these two fates. Collectively, these data provide novel insights into the global regulatory landscape during hematopoiesis.

Introduction

The temporal activation of stage-specific regulatory DNA instructs lineage specific gene expression programs that underpin cellular fate and potential. The establishment and maintenance of regulatory DNA is mediated by the combinatorial engagement of sequence-specific transcription factors (TFs) that bind in the place of a canonical nucleosome. Over the course of cellular differentiation programmed shifts in the global transcription factor milieu drive extensive re-organization of chromatin[1,2], where silencing of regulatory DNA associated with alternate lineage and the *de novo* activation of lineage-restricted elements result in the narrowing of the epigenetic and functional landscape[3]. However, it is unclear how and when regulatory DNA is dynamically activated and silenced during cell state transitions to establish lineage restricted gene expression programs and how these epigenetic changes relate to developmental potential.

Hematopoiesis is a prototypical system to study how genetically and epigenetically encoded programs are established during cellular differentiation[4–6]. Conventionally, hematopoiesis is depicted as a discrete hierarchical process where a multipotent hematopoietic stem and progenitor cell (HSPC) traverses a sequence of bifurcating decisions, mediated by the expression of lineage-specific TFs, with each decision resulting in an increasingly restricted fate potential. Historically, the characterization of the gene regulatory programs involved in the transition from HSPCs to terminal fates has relied on the identification of differential transcriptional programs from isolated discrete populations using defined cell surface markers[7–11]. While this approach has led to the identification of master regulatory transcription factors[10,12] that define many of the major hematopoietic cell lineages and has enabled a systematic mapping of their steady-state regulatory landscapes[9,11,13], interrogation of discretely defined populations cannot elucidate the dynamic regulatory events that mark cell-state transitions.

Recently, single-cell chromatin and transcriptional profiling assays have attempted to resolve the spatio-temporal *cis*- and *trans*- dynamics in different stages of hematopoiesis[14–17]. Typically, these studies have relied on the analysis of either bulk or immunophenotypically isolated populations of steady-state peripheral blood or bone marrow derived cells, whereby hierarchical relationships and developmental trajectories between cell states are predicted computationally. While these experimental approaches have aided in defining major subpopulations of hematopoietic cells and their respective epigenetic and transcriptional landscapes, definition of developmental trajectories within individual lineages from population snapshots is challenging due to the limited sensitivity and the resulting technical and analytical artifacts associated with single-cell genomic assays[18,19]. Furthermore, because developmental trajectories are predicted *in silico*, direct association of functional changes (i.e., lineage potential) to intermediate cellular states is not possible[20].

In order to investigate the dynamics of regulatory and functional events during differentiation, we use human erythropoiesis as a proxy for hematopoietic development. The transition from HSPCs to terminally differentiated enucleated red blood cells involves a series of morphologically, functionally, and phenotypically distinguishable states. While multiple efforts relying on the isolation of these states have characterized key transcriptional regulators[21,22] and chromatin elements implicated in erythropoiesis[9,11], our knowledge on the temporal interplay

between individual *cis*- and *trans*- elements and how these establish stage-specific transcriptional programs and lineage commitment during erythropoiesis, and hematopoiesis in general, remains only rudimentary. Furthermore, because erythrocytes share their developmental origins of with other myeloid lineages (granulocytic/monocytic and megakaryocytic), erythropoiesis represents an ideal system to study how lineage choice is genetically and epigenetically encoded.

Here, we capitalize on the *ex vivo* human differentiation scheme where dense unbiased sampling of the populations allows us to capture the dynamics of chromatin accessibility and gene expression during differentiation with a completely defined developmental trajectory. DNase I-seq and gene expression profiling (bulk and single cell) time-course during erythropoiesis coupled with lineage potential assays and morphological characterization, enabled the assignment of distal elements (alone or in combination) to target genes and individual TFs to their target DHSs which collectively comprise discrete regulatory modules associated with lineage potential. Comparing the activity patterns of the TF regulatory modules in the erythroid lineage to the closely related megakaryocytic lineage, provides insights into how these modules instruct lineage commitment. Collectively, our findings provide key insights into the organization of the functional epigenetic landscape during differentiation and its relation to lineage-potential.

Dense mapping of the temporal dynamics of *cis*- and *trans*- elements during erythropoiesis

Human erythropoiesis was induced *ex vivo* for 12 days using an established differentiation protocol[23] that faithfully recapitulates the major features of *in vivo* erythropoiesis. Starting from human adult $CD34^+$-enriched HSPCs in mobilized peripheral blood derived from 3 healthy donors we cultured the cells in defined media for 12 days (**Figure 1a** and **Methods**). Characteristic features of developing erythroblast cells were confirmed by immunophenotyping using canonical cell-surface markers of early (CD117, C-Kit) and late (CD235a, Glycophorin A) erythropoiesis as well as morphologically by hematoxylin-eosin staining of cell smears (**Supplementary Figure 1**).

To densely map both chromatin accessibility and transcriptional dynamics during the transition from HSPCs to committed erythroblasts, we subsampled a single continuous culture each day (12 days) and performed DNase I analysis and total RNA-seq (**Figure 1a,b**). Biological replicates from $CD34^+$ HSPCs from 3 donors were highly reproducible for both chromatin accessibility and gene expression profiles where the majority of the observed variability was accounted for by developmental trajectory (i.e., sampling days) (**Figure 1b,c** and **Supplementary Figure 2a**), as biological replicates were highly correlated (**Supplementary Figure 2b,c**). For many individual DHSs and genes, we observed quantitative changes in chromatin accessibility and expression over the course of differentiation highlighted by quantitative trajectories of opening or closing (**Figure 1d,e**). Notably, accessibility changes were mostly confined to compact regions of the genome (~200bp average DHS width). In many cases, we observed both opening and closing events within close proximity (**Figure 1d**), indicating focal regulation[24] of chromatin structure in contrast to previous reports that chromatin changes during differentiation occur over large domains[1,25].

To systematically identify developmentally responsive *cis*-elements, we leveraged the observed continuity of DHS signal over adjacent days (**Figure 1d**) and modelled DNase I cleavage density against differentiation time-points (**Methods**). We determined significance by comparing our full model to a reduced model (intercept-only; not accounting for developmental time) and performing a likelihood ratio test (**Methods**). Of the total 79,085 DHSs accessible in 2 or more samples/replicates, we conservatively identified 11,805 (14.9%) significantly changing DHSs (adjusted $p < 10^{-5}$ and fold-change >2), nearly evenly grouped between activated and silenced (45% and 55%, respectively) (**Supplementary Table 1**). A similar analytical approach applied to the RNA expression data identified 5,769 developmentally regulated genes (adjusted $p < 10^{-5}$ and fold-change > 2), of which 62% up-regulated and 38% down-regulated over the course of differentiation (**Supplementary Table 2**). Collectively, these data define a high-resolution and quantitative map of chromatin and gene expression dynamics during erythroid differentiation.

Stage-specific compartmentalization of the cis- and trans- landscape

PCA indicated that days 5-6 were associated with a critical developmental inflection point during *ex vivo* differentiation (**Figure 1b,c**). We next sought to characterize the relationship between temporal chromatin and gene expression dynamics with regards to the observed immunophenotypic and morphological changes present in the population of differentiating cells. We performed unsupervised clustering (*K*-means; *k*=5) on dynamically changing DHSs and developmentally responsive transcripts (**Figure 2a, b**). This analysis revealed a stark partitioning of activated and silenced genes and DHSs into non-overlapping sets that closely paralleled canonical developmental features of erythropoiesis. Particularly, DHSs rapidly silenced within the first days of differentiation (clusters E1 and E2) were found to preferentially harbor binding sequences utilized by the known HSPC regulators such as (HOXA9[26], RUNX[27] and ERG[28]) (**Supplementary Figure 3**). Similarly, immediately downregulated transcripts upon induction of differentiation (cluster G1) include these transcription factors as well as structural genes characteristic of $CD34^+$ HSPCs (**Figure 2b**). Consistent with PCA (**Figure 1b,c**), a rapid and marked turnover of chromatin and gene expression landscape is observed between days 5-7 where an early erythroid signature appears in both activated DHSs and gene expression, marked by the upregulation of *GATA1*, *KLF1*, *PPARA* and *TFRC* (cluster G4). Markers of mature erythropoiesis emerge later in the differentiation (after day 8; cluster G5) with the upregulation of hemoglobins, glycophorin A (*GYPA*) and *ALAS2* (**Figure 2b**).

In addition to canonical activation and downregulation patterns observed, we found a subset of genes exhibiting reproducible transient upregulation (clusters G2 and G3) occurring prior to establishment of the erythroid signature (**Figure 2b**). Transiently upregulated genes are expressed in myeloid lineages and correspond to several myeloid markers (e.g. *MPO*, *KIT*), including the myeloid-specific transcription factor CEBPA. Compatible with gene expression, late closing DHSs in cluster E2 and E3 were enriched in CEBPA recognition sequences (**Supplementary Figure 3**), denoting a transient emergence of myeloid-related regulatory elements. Consistent with this, the majority (~80%) of DHSs in cluster E2 and E3 were found

overlapping with DHSs active in other myeloid cell types (macrophages and monocytes) (**Supplementary Figure 4**).

Taken together these data indicate a sequence of developmentally related changes in both the *cis*- and *trans*- environment where an HSPC related chromatin and gene expression signature is succeeded by a transient activation of a broader myeloid program, prior to acquiring an erythroid-specific signature. Expectedly, activated DHSs (clusters E4-E5) were found to preferentially harbor red blood cell-related GWAS variants (1.36-fold enrichment over all detected DHSs), highlighting their functional role in regulating erythropoiesis (**Supplementary Figure 5**).

Connecting individual DHSs to genes

The overall dynamics of chromatin accessibility for individual DHSs closely mirrored that of the expression of nearby genes. To formulate this, we performed an enrichment test to investigate the DHS landscape around a gene promoter. Interestingly, we found that developmentally regulated genes are enriched for DHSs with a similar developmental profile (**Figure 2c**). For example, early closing genes (cluster G1) are significantly enriched for cluster E1 DHSs. Noteworthy, transient genes of cluster G3 are harboring DHSs belonging to both late closing DHS cluster E3 and early activated erythroid DHS cluster E4, suggesting that the transient nature of these genes is a result of the combinatorial activity of a closing and an opening chromatin landscape.

Because of fine-resolution afforded by our dense sampling approach, we sought to quantify the extent of genome-wide coactivation patterns that could potentially comprise physical regulatory links between DHSs and their target genes by correlating the temporal expression patterns of a gene to nearby (±500 kb from TSS) developmentally regulated DHSs. This analysis identified 25,624 connections at an absolute Pearson correlation coefficient (r) > 0.7 with the vast majority of gene-DHS links occurring within 50 kilobases of the transcription start site (**Figure 2d**). Overall, we connected 74.5% of the developmentally regulated genes with ≥1 DHS and 46% of changing DHSs were linked to ≥1 developmentally regulated gene. While on average 51 DHSs reside within 1 Mb of a given gene, only 5 DHSs (±4.4 SD) were found to be linked with a changing gene.

We therefore sought to functionally validate these associations by genetic perturbation of these gene-DHS links. Specifically, we disrupted two correlated DHSs (HS1 and HS2) with *CDH1* expression, situated upstream (5kb and 12kb, respectively) of the promoter of *CDH1* gene which is upregulated during differentiation (**Figure 2e**). Homozygous deletion of these DHSs as well as the promoter in the human derived erythroid progenitor cell line HUDEP-2 where these DHSs are also active, resulted in complete ablation of the CDH1 expression as determined by flow-cytometry (**Figure 2f**). These results suggest that both elements as predicted by the correlation analysis as regulators of *CDH1*, indeed drive the expression of the gene and their deletion confers effects similar to the deletion of the gene promoter.

Overall, these findings suggest that the majority of changes in transcription during development are regulated by a limited number of *cis*-regulatory inputs, situated within close proximity to the genes they regulate.

Distinct and sequential regulatory modules encode developmental stages

Clustering of dynamically changing DHSs revealed that chromatin active at different stages of hematopoiesis display differential enrichment for recognition sequences for transcription factors. indicating stage-specific transcriptional regulation of *cis*- elements. This however, does not resolve the temporal interactions between individual DHSs and individual *trans*- regulators and how this relationship shapes the developmental response of a DHS. Given the observed global correlated changes between the transcription factor expression levels and the accessibility of the DHSs containing their cognate recognition sequences (**Figure 3a**) we sought to quantify the contribution of individual TFs to the dynamic changes in DNase I density at individual regulatory *cis*-elements. We capitalized on our dense sampling approach and applied a regression strategy where the activity of an individual regulatory element (i.e. DNase I cleavage density) is modelled as a function of the gene expression profiles of developmentally regulated TFs with a compatible recognition sequence harbored within each DHS (**Figure 3b** and **Methods**). We controlled for weak and ambiguous association of TFs recognizing degenerate motifs using elastic-net regularization (**Methods**). We applied this approach to all of the 11,805 dynamically changing DHSs, identifying 11,734 (>99% of changing DHSs) with at least one explanatory TF regulator (**Methods**) where the regression coefficients broadly correspond the strength of association of a TF with an individual DHS (**Supplementary Figure 7**). Overall, 7 TFs on average (out of the total 214 tested), were positively associated with each DHS, suggesting that a small subset of TFs regulate the developmental activity of individual *cis*-elements.

We next asked to what extent the activity of DHSs with similar temporal accessibility patterns are regulated by a coherent set of TF regulators. We identified 52 TFs positively associated with at least 200 DHSs and performed unsupervised hierarchical clustering based on their regression coefficients computed for each DHS (**Figure 3c** and **Methods**). This analysis resolved the temporal associations between transcription factors and their target DHS into a sequence of five discrete and largely non-overlapping regulatory modules, reflective of developmental stages of erythropoiesis (**Figure 3d**). Module 1 consists of known HSPC transcriptional regulators (e.g. ERG[28], MEIS1[29], MYCN[30]) which are positively associated with early closing DHSs in clusters E1 and E2. In modules 2 and 3, transcription factors associated with commitment of hematopoietic progenitors to the different myeloid lineages (e.g. CEBPA[31], MYB[32], FLI1[33], RUNX1[34]) interact with DHSs in clusters E2 and E3. Modules 4 and 5 define the erythroid-specific regulatory landscape as known erythroid regulators (e.g. GATA1, KLF1[21], RXRA[35] and FOXO3[36]) positively interact with activated DHSs in clusters E4 and E5.

Plotting the fraction of DHSs in each cluster positively associated with each TF (**Figure 3d**) highlights the major drivers of chromatin accessibility in each developmental stage. Particularly, ERG appears as a major regulator of the HPSC stage as it is positively associated with ~25% of DHSs in clusters E1 and E2. Although ERG has been long implicated in HSPC regulation, it is only recently its role as a critical regulator of HSPC survival has been appreciated[37]. Interestingly, KLF12 also appears to share a significant proportion of the early chromatin landscape, although its role in HSPC regulation is not known. Overexpression of the critical HSC regulator Evi-1 in mice, resulted in maintenance of the quiescent phenotype of murine HSCs along with the more than 12-fold increase in *Klf12* expression[38]. In another

experiment, sustained expression of *Hlf* in mice also resulted in enrichment of *Klf12* in more primitive hematopoietic compartments[39], implicating KLF12 in the HSPC regulation. Apart from the canonical erythroid transcription factors, we identified MXI1 among the top regulators of the erythroid chromatin landscape. Knockdown of *Mxi1* in mice, blocks chromatin condensation and impairs enucleation of mouse erythroblasts, highlighting the role of MXI1 in erythroid maturation[40]. Additionally, we find CTCF to be positively associated with a large portion of the erythroid-specific chromatin (~25% of DHSs in clusters E4 and E5), an observation in line with the role of CTCF in the establishment of erythroid-specific functional chromatin loops[41,42].

Taken together, these findings illustrate the dynamic interaction of the *cis*- and the *trans*-regulatory landscape during erythropoiesis and their organization into well-defined and discrete regulatory modules of associated DHSs with their cognate transcription factors, reflecting distinct stages of erythroid development.

A sequence of abrupt lineage restriction events marks erythropoiesis

The organization of chromatin and transcription factors into defined regulatory modules corresponding to distinct stages of erythropoiesis indicate a functional relationship between lineage potential and module activity. To gain insight into whether these modules underpin lineage decision events we determined the lineage potential of the erythroid cultures by daily sampling a population of cells and assaying their multipotent and unipotent capacity for different myeloid lineages (**Figure 4a** and **Supplementary Figure 8a**). Total number of colonies declined with the progress of differentiation resulting in an abrupt depletion of total progenitors on day 6 of differentiation (**Supplementary Figure 8b**). After 4 days of exposure to erythroid media, the most primitive and multipotent colonies (CFU-GEMM; granulocytic, erythroid, monocytic, megakaryocytic) were no longer detected (**Figure 4b**). Day 6 marked a second event of restriction of the fate potential as all colonies derived from progenitors with unilineage capacity were no longer detected in the cultures such that frequency of erythroid progenitors (BFU-E) rapidly declined from day 5 to day 6 (**Figure 4c**). Similarly, granulocytic/monocytic progenitors (CFU-GM) were depleted by day 6 of erythroid differentiation (**Supplementary Figure 8c**). Notably, none of the changes in clonogenic capacity were associated with any changes in the growth rate of the parental erythroid cultures (which remained constant throughout the differentiation) (**Supplementary Figure 8d**).

In addition to the above lineages, we specifically tested for the ability to differentiate into megakaryocytes during erythroid development by transferring cells from the primary erythroid culture, on a daily basis, to megakaryopoiesis-inducing suspension cultures and tested for their ability to give rise to $CD41^+$ megakaryocytic populations (**Supplementary Figure 9a** and **Methods**). Consistent with the overall lineage restriction observed during colony-forming assays, erythroid cultures completely lose megakaryocytic potential on day 6 of the differentiation (**Figure 4c** and **Supplementary Figure 9b**).

The rapid changes observed in clonogenic capacity correspond to the transitions in the activity of regulatory modules (**Figure 4d**). Early depletion of primitive multipotent CFU-GEMM progenitors is concomitant with the transition from the HSPC-related modules (modules 1 and 2), while the decline of unipotent progenitors of all detectable myeloid lineages

(granulocytic/monocytic, erythroid, megakaryocytic) coincides with the transition from a program with a broader myeloid signature to erythroid specific *cis*- and *trans*- landscape. Furthermore, because these rapid lineage restriction events are not associated with other abrupt changes in morphology or cell growth (**Supplementary Figure 8d**), these data suggest that the mechanism responsible for the exit from the progenitor stage is decoupled from maturation progress.

Exit from the HSPC-related transcriptional program is shared between erythropoiesis and megakaryopoiesis

The silencing of the HPSC regulatory modules prior to lineage commitment suggested that exit from the progenitor state is necessary for erythroid commitment to proceed. We therefore asked whether this represents a canonical feature of hematopoietic development to any lineage. To investigate this, we focused on megakaryocytic differentiation, a process that shares both close common developmental origins[43] and key TF regulators with erythropoiesis[44].

We induced *ex vivo* megakaryocytic differentiation and performed dense sampling of gene expression during development (**Figure 5a** and **Methods**). Developmentally regulated genes during megakaryopoiesis exhibit largely bipartite profiles similar to those observed during erythropoiesis (**Supplementary Figure 10**). To determine whether the transcriptional changes associated with exit from HSPC state during erythropoiesis are shared with megakaryopoiesis we examined the expression profiles of erythroid developmentally regulated genes during megakaryocytic differentiation. We observed highly correlated global expression profiles for early silenced transcripts (erythroid clusters G1 and G2) between the two lineages (median Spearman's ρ=0.76 and 0.62, respectively) (**Figure 5b**), with the exception of key regulators and canonical markers of megakaryopoiesis (*MEIS1*, *FLI1, PBX1, ITGA2B,* etc.) (**Figure 5c**). In contrast, correlation for erythroid clusters G3-G5 was low (median Spearman's $\rho \leq 0.13$).

Similar to erythropoiesis, the early suppression of HSPC-related gene signature resulted in abrupt restriction of lineage potential during megakaryopoiesis. We found that cells in megakaryocytic cultures abruptly lose erythroid potential on day 4 of differentiation, coinciding with the activity of the HSPC transcriptional program (**Figure 5c** right and **Supplementary Figure 11**). However, in contrast to erythropoiesis, megakaryocytic differentiation does not exhibit a transient activation of a myeloid signature. This finding is compatible with the recently revised hematopoietic tree according to which megakaryocytes directly emerge from the primitive HSPC compartments bypassing the common myeloid progenitor[45,46].

Conclusively, these results indicate the existence of a shared mechanism between erythropoiesis and megakaryopoiesis driven by a common set of TFs which mediates the exit from HSPC state and signals the entry into lineage-restricted states.

Single-cell transcriptomics uncover discrete cell states corresponding to transcriptional programs

So far, observations based on population-level analysis suggest the existence of transcriptional programs responsible for changes in lineage potential. However, as lineage decision events occur in individual progenitor cells[47], we analyzed transcriptional changes from

~50,000 single cells sampled from frequent intervals along both the *ex vivo* erythroid and megakaryocytic differentiation (**Figure 6a**). Overall, gene expression dynamics from aggregated single-cells were highly concordant to RNA-seq performed on bulk populations (**Supplementary Figure 12**).

Given the heterogeneity between cells we sought to explore the transcriptional dynamics of individual cells during the transition from HSPC state to lineage commitment. To this end we collectively analyzed all cells sampled from both the megakaryocytic and erythroid lineages and performed principal component analysis (PCA) using 166 lineage regulating TFs detected in the dataset and 11 marker genes of mature erythropoiesis and megakaryopoiesis (**Figure 6a** and **Supplementary Table 5**). PCA readily resolved the two primary axes of differentiation where PC2 reflected HSPC to terminally committed lineages, and PC1 distinguished the erythroid and megakaryocytic lineages (**Figure 6a**). The top positive and negative loadings from each component correctly identified known key TFs that regulate distinct aspects of differentiation such as known HSPC regulators (*HOXA9* and *SOX4)*, TF regulators shared between both lineages (*GATA1* and *GFIB),* and lineage specific regulators (*KLF1, NFIA, MEIS1, PBX1*) (**Figure 6b**). Projection of the sampling time on the PCA plot, highlights day 4 as the bifurcation point where erythro-megakaryocytic lineages resolve (**Figure 6a**).

Analysis of the bulk RNA-seq and chromatin accessibility profiles revealed discrete regulatory modules associated with distinct functional states. We sought to explore whether these modules are reflected in discrete populations of individual cells with distinct transcriptional signature. To this end we performed Louvain clustering[48,49] on all cells using the top 10,000 highly variable and highly expressed genes (**Figure 6c** and **Methods**). This resulted in 9 clusters with largely distinct gene expression signature, closely associated with sampling days where cells are partitioned into primitive clusters with strong HSPC signature (clusters 0 and 5), early progenitors with myeloid signature (clusters 6, 8) and committed/mature cells for each lineage (clusters 7, 1, 2, 3 and 4) (**Supplementary Figure 13**). Compatible with earlier results, clusters that exhibit myeloid gene signature (clusters 6 and 8) are almost exclusively comprised of early erythroid cells (**Supplementary Figure 13**), while megakaryocytes are divided between early cells with HSPC signature and committed cells expressing markers of megakaryopoiesis (**Figure 6c**).

Bulk analysis and sampling time points might obscure cellular heterogeneity and true developmental time. Furthermore, as *ex vivo* erythropoiesis and megakaryopoiesis are two independent systems induced by different cytokine stimuli, the differential lineage commitment kinetics observed between the two lineages might be a result of the culture conditions and therefore not directly comparable. In order to address this and directly compare transcriptional developmental dynamics between erythroid and megakaryocytic differentiation, we sought to align the gene expression patterns of these two lineages along a common developmental coordinate system (independent of sampling time). Given the resolution of the developmental trajectories between erythropoiesis and megakaryopoiesis provided by PCA, we used the first two principal components to infer a metric of developmental pseudotime (**Supplementary Figure 14** and **Methods**). Plotting gene expression of individual cells against pseudotime, readily identifies a sequence of non-overlapping, developmentally punctuated transcriptional states punctuated for both lineages (**Figure 6e**). Downregulation of HSPC signature for both

lineages occurs over the same developmental interval, confirming that the exit from the HSPC state occurs for both lineages where the two lineages rapidly diverge after that. Erythroid development exhibits transient activation of myeloid-related transcripts prior to terminal commitment, while megakaryopoiesis rapidly acquires maturation signature. This observation becomes particularly evident when we place the Louvain clusters over developmental pseudotime. Transition from megakaryocytic progenitor cluster 0 to early committed cluster 2 occurs over a very short developmental interval. In contrast, committed erythroid cluster 7 appears much later (**Supplementary Figure 15**). Interestingly, annotating the sampling time over developmental pseudotime we observe that transcriptional inflection points coincide with the sampling time points where rapid lineage restriction was observed on population level. Day 4 of differentiation (**Figure 6e,** purple dashed line) demarcates the exit from HSPC for both lineages, while day 6 (**Figure 6e,** red dashed line) marks the switch from myeloid state to erythroid commitment.

Taken together, these findings suggest that exit from HSPC is an intrinsic mechanism, independent of the cytokine environment, that occurs during both early erythroid and megakaryocytic differentiation and it is driven by a common set of TFs. Additionally, rapid transitions in transcriptional states are a result of changes in expression of lineage regulating TFs, happening in individual cells, over short intervals of developmental time.

Discussion

Here, we systematically link individual transcription factors and their target *cis*- elements along *ex vivo* human erythropoiesis, resolving how these elements organize temporally, encoding lineage commitment and differentiation during hematopoiesis. Multiple efforts have extensively studied the individual (*cis*- and *trans*-) regulatory components involved in erythropoiesis[11] as well as other diverse hematopoietic lineages[9,50–53]. The bulk of these efforts base their findings either on immunophenotypically defined hematopoietic populations, or single-cell dissection of steady state heterogeneous sources, where developmental relationships between cells within a heterogeneous steady-state population can only be inferred[14,16,54,55]. In this work we differ by capitalizing on the continuity of the differentiating populations during *ex vivo* erythropoiesis to finely map chromatin accessibility and gene expression dynamics while we overcome the sampling biases inherent to the conventional immunophenotypic definitions of hematopoiesis[46,56]. Sequential sampling of a continuously differentiating culture allows the direct and repeated measurement of the dynamic epigenetic landscape along a defined lineage trajectory. In addition, a dense sampling approach enables the unbiased detection of transient events occurring over short intervals that would otherwise be missed by sparse sampling methodologies.

Integrative analysis of chromatin accessibility and gene expression during erythropoiesis revealed a sequence of discrete, non-overlapping regulatory modules comprising of interacting transcription factors and individual *cis*-regulatory elements, corresponding to distinct stages of erythroid development. Strikingly, the transition between the activity of these modules coincides with a sequence of experimentally validated rapid lineage restriction events. We found that the exit from the program associated with the HSPC state occurred independent of lineage outcome, as it was identified during *ex vivo* megakaryopoiesis. Furthermore, comparison of

developmental transcriptomics of single-cells along erythropoiesis and megakaryopoiesis reveals that exit from HSPC occurs over the same developmental interval for both lineages, indicating an intrinsic mechanism independent of the cytokine environment.

We found the two lineages to exhibit differential commitment kinetics after exit from HSPC state. Erythroid differentiation maintains unipotent myeloid capacity until terminal erythroid commitment, as it transiently activates a broader regulatory program involving canonical myeloid transcription factors (FLI1, C/EBPs, IRF8, etc). There are several lines of evidence from single-cell assays in both mouse and human hematopoiesis where they converge on the existence of a population early on erythroid development with a myelo-erythroid signature[14,17,57]. Notably, transgenic mice lacking a set of the C/EBP family of myeloid regulators exhibit decreased erythroid output, in addition to the granulocytic/monocytic[58]. Although all three myeloid lineages (Ery, Mk, G/M) are considered to derive from a common myeloid progenitor (CMP), we found megakaryocytic commitment to occur rapidly upon HSPC exit, bypassing the transient expression (and presumably chromatin) program. This observation is compatible with the growing evidence suggesting that megakaryocytic lineage arises directly from the primitive hematopoietic compartments[46,59–61].

The single-cell RNA-seq findings presented here, represent the most comprehensive analysis of single-cell gene expression along a closely monitored developmental system, so far. Knowledge of the actual time point where individual cells were sampled allowed us to align changes recorded from population level experiments to developmental changes in lineage regulating TF expression in single-cells. Strikingly, our single-cell based observations faithfully recapitulated the population-derived findings as the inferred developmental pseudotime aligns with the time-points where lineage restriction events were observed. This suggests that highly synchronized rapid shifts in gene expression levels of lineage regulators across individual cells, occurring over short intervals of developmental time, underpin the changes observed in bulk populations. This finding is in contrast to observations based on single-cell analyses of steady-state hematopoiesis whereby variability in the chromatin and transcriptional landscapes are interpreted as gradients of continuous regulatory states[14,15,55,62].

We present novel insights into the developmental regulatory dynamics during hematopoiesis whereby lineage regulating TFs organize in a developmental manner with their individual chromatin elements into discrete regulatory modules associated with distinct developmental stages. The transitions between these modules are demarcated by rapid shifts in lineage potential during differentiation. Furthermore, this work provides a comprehensive approach to studying the *cis*- and *trans*- regulatory element dynamics, in the context of lineage commitment and differentiation utilizing a well-defined *ex vivo* differentiation system. Although these findings might not entirely recapitulate the regulatory events occurring during *in vivo* hematopoietic development, they provide a generalizable model of how the *trans*- environment interacts with chromatin along the developmental axis to instruct fate choice and lineage commitment. Additionally, the dense sampling and the systematic linkage between distal elements and target promoters provides sufficient resolution to identify stage-specific activity of regulatory elements. Such elements can prove particularly useful in transgene-based therapies where the efficacy of these methods relies on the precise modulation of gene expression in a developmental and lineage-specific manner.

Author contributions: G.G., M.I., and N.P. performed the experiments. G.G., A.N., T.S., and J.V. performed data analysis. G.G., N.P., and J.V. designed the experiments. J.S., M.Y., and J.V. supervised the study and provided consultation. G.G. and J.V. wrote the manuscript. J.V. conceived the study. All authors approved the manuscript.

Acknowledgments: The authors would like to thank Daniel Bates, Morgan Diegel, Douglas Dunn, Fidencio Neri, Ericka Otterman, Shinny Vong, and Alister Funnell from the Altius Institute for Biomedical Sciences for help in sequencing and genome editing. We also thank John Lazar and Thalia Papayannopoulou for critical review of the manuscript.

Conflict of interest: The authors state no conflicts of interest.

References

1. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
2. Ho, L. & Crabtree, G. R. Chromatin remodelling during development. *Nature* vol. 463 474–484 (2010).
3. Stergachis, A. B. *et al.* Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell* **154**, 888–903 (2013).
4. Antoniani, C., Romano, O. & Miccio, A. Concise Review: Epigenetic Regulation of Hematopoiesis: Biological Insights and Therapeutic Applications. *STEM CELLS Translational Medicine* vol. 6 2106–2114 (2017).
5. Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
6. Cullen, S. M., Mayle, A., Rossi, L. & Goodell, M. A. Hematopoietic stem cell development: an epigenetic journey. *Curr. Top. Dev. Biol.* **107**, 39–75 (2014).
7. Goode, D. K. *et al.* Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. *Dev. Cell* **36**, 572–587 (2016).
8. González, A. J., Setty, M. & Leslie, C. S. Early enhancer establishment and regulatory locus complexity shape transcriptional programs in hematopoietic differentiation. *Nat. Genet.* **47**, 1249–1259 (2015).
9. Lara-Astiaso, D. *et al.* Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).
10. Iwasaki, H. *et al.* The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages. *Genes Dev.* **20**, 3010–3021 (2006).
11. Ludwig, L. S. *et al.* Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. *Cell Rep.* **27**, 3228–3240.e7 (2019).
12. Zhu, J. & Emerson, S. G. Hematopoietic cytokines, transcription factors and lineage commitment. *Oncogene* vol. 21 3295–3313 (2002).
13. Winter, D. R. & Amit, I. The role of chromatin dynamics in immune cell development. *Immunol. Rev.* **261**, 9–22 (2014).
14. Buenrostro, J. D. *et al.* Integrated Single-Cell Analysis Maps the Continuous Regulatory

Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535–1548.e16 (2018).
15. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).
16. Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
17. Drissen, R. *et al.* Distinct myeloid progenitor–differentiation pathways identified through single-cell RNA sequencing. *Nature Immunology* vol. 17 666–676 (2016).
18. Weinreb, C., Wolock, S., Tusi, B. K., Socolovsky, M. & Klein, A. M. Fundamental limits on dynamic inference from single-cell snapshots. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E2467–E2476 (2018).
19. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).
20. Jacobsen, S. E. W. & Nerlov, C. Haematopoiesis in the era of advanced single-cell technologies. *Nat. Cell Biol.* **21**, 2–8 (2019).
21. Cantor, A. B. & Orkin, S. H. Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene* **21**, 3368–3376 (2002).
22. Perry, C. & Soreq, H. Transcriptional regulation of erythropoiesis. Fine tuning of combinatorial multi-domain elements. *Eur. J. Biochem.* **269**, 3607–3618 (2002).
23. Giarratana, M.-C. *et al.* Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells. *Nat. Biotechnol.* **23**, 69–74 (2005).
24. Oudelaar, A. M., Beagrie, R. A., Gosden, M. & De Ornellas, S. Dissection of the 4D chromatin structure of the α-globin locus through in vivo erythroid differentiation with extreme spatial and temporal resolution. *bioRxiv* (2019).
25. Fraser, J. *et al.* Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol. Syst. Biol.* **11**, 852 (2015).
26. Lawrence, H. J. *et al.* Loss of expression of the Hoxa-9 homeobox gene impairs the proliferation and repopulating ability of hematopoietic stem cells. *Blood* **106**, 3988–3994 (2005).
27. Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. & Speck, N. A. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature* vol. 457 887–891 (2009).
28. Loughran, S. J. *et al.* The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat. Immunol.* **9**, 810–819 (2008).
29. Argiropoulos, B., Yung, E. & Humphries, R. K. Unraveling the crucial roles of Meis1 in leukemogenesis and normal hematopoiesis. *Genes Dev.* **21**, 2845–2849 (2007).
30. Laurenti, E. *et al.* Hematopoietic stem cell function and survival depend on c-Myc and N-Myc activity. *Cell Stem Cell* **3**, 611–624 (2008).
31. Friedman, A. D. C/EBPα in normal and malignant myelopoiesis. *Int. J. Hematol.* **101**, 330–341 (2015).
32. Greig, K. T., Carotta, S. & Nutt, S. L. Critical roles for c-Myb in hematopoietic progenitor cells. *Semin. Immunol.* **20**, 247–256 (2008).
33. Kawada, H. *et al.* Defective Megakaryopoiesis and Abnormal Erythroid Development inFli-1 Gene-Targeted Mice. *International Journal of Hematology* vol. 73 463–468 (2001).
34. Ichikawa, M. *et al.* AML-1 is required for megakaryocytic maturation and lymphocytic

differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. *Nat. Med.* **10**, 299–304 (2004).
35. Evans, T. Regulation of hematopoiesis by retinoid signaling. *Experimental Hematology* vol. 33 1055–1061 (2005).
36. Liang, R. *et al.* A Systems Approach Identifies Essential FOXO3 Functions at Key Steps of Terminal Erythropoiesis. *PLoS Genet.* **11**, e1005526 (2015).
37. Xie, Y. *et al.* Reduced Erg Dosage Impairs Survival of Hematopoietic Stem and Progenitor Cells. *Stem Cells* **35**, 1773–1785 (2017).
38. Kustikova, O. S. *et al.* Activation of Evi1 inhibits cell cycle progression and differentiation of hematopoietic progenitor cells. *Leukemia* vol. 27 1127–1138 (2013).
39. Wahlestedt, M. *et al.* Critical Modulation of Hematopoietic Lineage Fate by Hepatic Leukemia Factor. *Cell Rep.* **21**, 2251–2263 (2017).
40. Zhang, L., Flygare, J., Wong, P., Lim, B. & Lodish, H. F. miR-191 regulates mouse erythroblast enucleation by down-regulating Riok3 and Mxi1. *Genes Dev.* **25**, 119–124 (2011).
41. Hanssen, L. L. P. *et al.* Tissue-specific CTCF–cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nature Cell Biology* vol. 19 952–961 (2017).
42. Lee, J., Krivega, I., Dale, R. K. & Dean, A. The LDB1 Complex Co-opts CTCF for Erythroid Lineage-Specific Long-Range Enhancer Interactions. *Cell Rep.* **19**, 2490–2502 (2017).
43. Papayannopoulou, T. & Kaushansky, K. Evolving insights into the synergy between erythropoietin and thrombopoietin and the bipotent erythroid/megakaryocytic progenitor cell. *Exp. Hematol.* **44**, 664–668 (2016).
44. Wickrema, A. & Crispino, J. D. Erythroid and megakaryocytic transformation. *Oncogene* vol. 26 6803–6815 (2007).
45. Roch, A., Trachsel, V. & Lutolf, M. P. Brief Report: Single-Cell Analysis Reveals Cell Division-Independent Emergence of Megakaryocytes From Phenotypic Hematopoietic Stem Cells. *Stem Cells* **33**, 3152–3157 (2015).
46. Notta, F. *et al.* Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116 (2016).
47. Nimmo, R. A., May, G. E. & Enver, T. Primed and ready: understanding lineage commitment through single cell analysis. *Trends Cell Biol.* **25**, 459–467 (2015).
48. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* vol. 2008 P10008 (2008).
49. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports* vol. 9 (2019).
50. Tusi, B. K. *et al.* Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature* **555**, 54–60 (2018).
51. An, X. *et al.* Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood* **123**, 3466–3477 (2014).
52. Moignard, V. *et al.* Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat. Cell Biol.* **15**, 363–372 (2013).

53. Gazit, R. *et al.* Transcriptome analysis identifies regulators of hematopoietic stem and progenitor cells. *Stem Cell Reports* **1**, 266–280 (2013).
54. Pellin, D. *et al.* A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat. Commun.* **10**, 2395 (2019).
55. Macaulay, I. C. *et al.* Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep.* **14**, 966–977 (2016).
56. Georgolopoulos, G., Iwata, M., Psatha, N., Yiangou, M. & Vierstra, J. Unbiased phenotypic identification of functionally distinct hematopoietic progenitors. *J. Biol. Res.* **26**, 4 (2019).
57. Upadhaya, S. *et al.* Kinetics of adult hematopoietic stem cell differentiation in vivo. *J. Exp. Med.* **215**, 2815–2832 (2018).
58. Mancini, E. *et al.* FOG-1 and GATA-1 act sequentially to specify definitive megakaryocytic and erythroid progenitors. *EMBO J.* **31**, 351–365 (2012).
59. Woolthuis, C. M. & Park, C. Y. Hematopoietic stem/progenitor cell commitment to the megakaryocyte lineage. *Blood* **127**, 1242–1248 (2016).
60. Sanjuan-Pla, A. *et al.* Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232–236 (2013).
61. Rodriguez-Fraticelli, A. E. *et al.* Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216 (2018).
62. Karamitros, D. *et al.* Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat. Immunol.* **19**, 85–97 (2018).

Figure Legends

Figure 1. Comprehensive identification of regulatory landscape developmental dynamics. (a) Dense DNase I-seq and RNA-seq time course with daily sampling during the 12-day *ex vivo* erythroid differentiation induced from $CD34^+$ HSPCs. (b) PCA analysis using all detected DHSs (79,085 Hotspots 5% FDR) across all samples (12 time points, 3 donors). The arrow denotes the differentiation trajectory from day 0 to day 12. (c) PCA analysis using all 24,849 detected genes across all samples (13 time points, 3 donors). The arrow denotes the differentiation trajectory from day 0 to day 12. (d) Chromatin accessibility tracks for each day of differentiation with DNase Hypersensitive Sites (DHSs) harbored around the *TFRC* locus. (e) Identification of significantly changing DHS and genes with robust linear regression analysis. Scatterplots show *TFRC* expression and DNase I density for two upstream DHS. Dots represent normalized values for each of the 3 donors. Dashed line represents the fitted regression spline.

Figure 2. The *cis*- and *trans*- landscape of erythropoiesis exhibits temporal compartmentalization (a) K-means clustering of 11,805 changing DHS resulted in 5 clusters (E1-E5) with sequential activity profile for each cluster. Values are z-score of per day average normalized DHS counts from 3 donors. (b) K-means clustering of 5,792 developmentally regulated genes resulted in 5 clusters (G1-G5). Values are z-score of per day average normalized FPKM from 3 donors. (c) A matrix showing the enrichment score (log2-ratio observed over expected) for any given DHS cluster, around each developmentally regulated DHS (±50kb from TSS). Highlighted in red is cluster G3 which is enriched for both late downregulated DHS of cluster E3 and early upregulated from cluster E4. (d) Correlation density plot between developmental genes and developmental DHS ±250kb around the gene promoter. Grey shaded area highlights enrichment of correlations within ±50kb around the gene promoter. (e) DNase I accessibility track of the *CDH1* locus during erythroid differentiation, highlighting the accessibility of 3 nearby DHS correlated to *CDH1* expression. (f) DNase I accessibility track of the *CDH1* locus in HUDEP-2 cells depicting the genetic knockout of the *CDH1* promoter and two upstream DHS (-32, and -12) (above) along with the resulted ablation in CDH1 protein expression as assessed by flow cytometry (below). Asterisks denote significant positive enrichment (X^2 test p-value < 0.05).

Figure 3. Systematic modelling of *cis*- and *trans*- element temporal interactions reveals discrete regulatory modules during erythropoiesis. (a) Developmental responses of DHS accessibility and transcription factor expression levels were found to be correlated across the genome. (b) The density of a given developmental DHS is modelled after the TF binding motif composition and the expression of the binding TFs using elastic-net regression. The model returns a coefficient (β) for each pair of DHS and binding TF which denotes how strongly (positively or negatively) the TF expression is associated with the accessibility of the particular DHS. (c) Hierarchical clustering of 52 highly connected TFs based on the cosine distances of the regression coefficient from 11,734 DHS resulted in 5 clusters of developmentally regulated TFs. Transcription factors along with their positively associated DHS comprise a regulatory module (Module 1-5). (d) The fraction of DHS per cluster positively associated with a TF identifies the major drivers of chromatin accessibility during erythropoiesis.

Figure 4. Lineage restriction events during erythropoiesis reflect the sequence of regulatory programs. (a) Schematic diagram of lineage potential assays during the first 7 days of *ex vivo* erythropoiesis. Cells were sampled daily and transferred to lineage-permissive media. Multilineage capacity was determined as frequency of CFU-GEMM progenitors. Erythroid potential as frequency of BFU-Es and megakaryocytic potential as frequency of $CD41^+$ cells. (b) Frequency of multipotent CFU-GEMM over the course of erythroid differentiation. (d) Frequency of erythroid progenitors BFU-E detected in methylcellulose assay (red line) and $CD41^+$ megakaryocytic progenitors (blue line). (d) Changes in lineage potential coincide with the transitions between the regulatory modules identified earlier. Transition from modules 1 and 2 to module 3 reflects the loss of multipotency occurring between days 3 and 4, while transition from module 3 to erythroid modules 4 and 5 coincides with the depletion of unipotent progenitors and entry to erythroid maturation (days 5 to 6). Error bars denote ±1 SE of the mean from 4 replicates for colony-forming assays, 2 replicates for $CD41^+$ frequency. Asterisk denotes statistically significant difference in CFU-GEMM and BFU-E counts from day 1 (P-value < 0.05 Student's T-test). CFU-GEMM: Colony Forming Unit - Granulocytic, Erythroid, Macrophage, Megakaryocytic. BFU-E: Burst Forming Unit-Erythroid.

Figure 5. A shared transcriptional program drives the exit from HSPC state early in erythropoiesis and megakaryopoiesis. (a) Dense RNA-seq time course during *ex vivo* megakaryopoiesis induced from $CD34^+$ HSPCs. (b) Correlation of gene expression profiles between erythropoiesis and megakaryopoiesis across the erythroid gene clusters G1-G5. (c) Expression profiles during megakaryocytic development, ordered by their correlation score to their erythroid counterparts from erythroid clusters G1 and G2. (d) Lineage potential assay during *ex vivo* megakaryopoiesis whereby erythroid potential was assessed by subjecting cells to a secondary erythroid culture (left). Frequency of $CD235a^+$ erythroid cells after 12-day culture into secondary erythroid media (right).

Figure 6. Single-cell gene expression dynamics demonstrate distinct cell states during erythroid and megakaryocytic differentiation. (a) $CD34^+$ cells from a single donor were *ex vivo* differentiated towards the erythroid and the megakaryocytic lineage. Including an uncultured sample from the same donor ($CD34^+$ day 0), ~50,000 cells were totally sampled during 5 time-points from each lineage and subjected to single-cell RNA-seq. (b) PCA on a set of developmentally regulated transcription factors and mature lineage markers resolves the differentiation (PC2) and lineage (PC1) axes along erythro-megakaryocytic development. Grey line denotes where the 90% of the cells sampled prior to day 4 are located (above line). (c) The loadings of the first two principal components highlight the transcription factors that act on each direction of the differentiation and lineage axes. (d) Expression profiles of marker genes highlight the expression signature of each of the 9 Louvain clusters. Clusters were hierarchically clustered using expression profiles from 10,000 highly variable, highly expressed genes. Values shown are normalized mean TPM across cells in each cluster. (e) Expression profiles of 6 representative genes based against the erythroid (left) and megakaryocytic (right) developmental pseudotime demonstrate discrete transcriptional states along erythroid and megakaryocytic development. Fitted splines of moving average TPM values over 200-cell

windows are plotted against the pseudo-temporal distance of each lineage. The strip underneath represents individual cells from each trajectory ordered by pseudotime and are colored by the average sampling day over moving average of 200 cells.

Figure 1

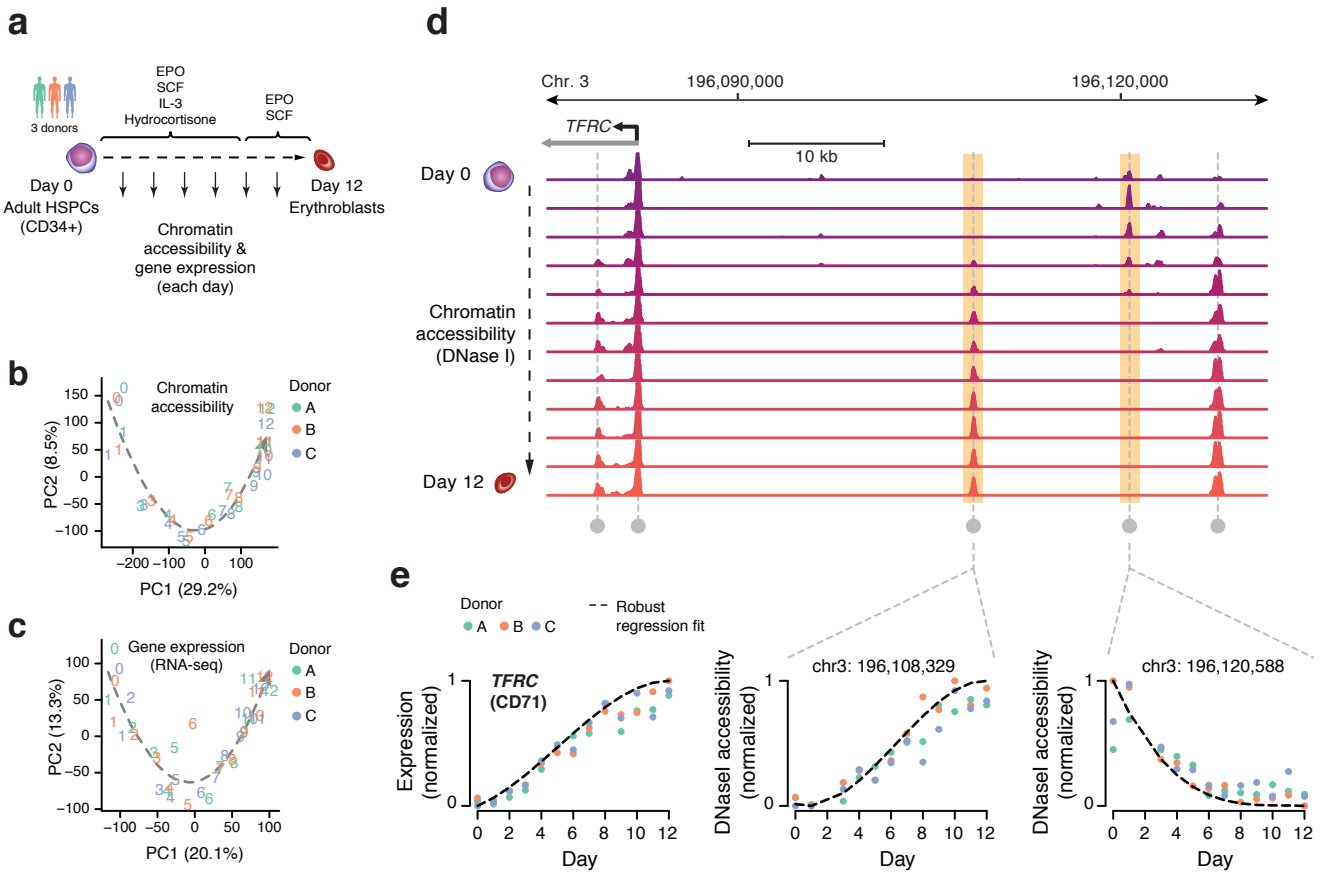

a
3 donors
EPO
SCF
IL-3
Hydrocortisone
EPO
SCF
Day 0
Adult HSPCs
(CD34+)
Day 12
Erythroblasts
Chromatin
accessibility &
gene expression
(each day)
b
Chromatin
accessibility
Donor
A
B
C
PC2 (8.5%)
PC1 (29.2%)
c
Gene expression
(RNA-seq)
Donor
A
B
C
PC2 (13.3%)
PC1 (20.1%)
d
Chr. 3
196,090,000
196,120,000
TFRC
10 kb
Day 0
Chromatin
accessibility
(DNase I)
Day 12
e
Donor
A B C
Robust
regression fit
TFRC
(CD71)
Expression
(normalized)
Day
chr3: 196,108,329
DNaseI accessibility
(normalized)
Day
chr3: 196,120,588
DNaseI accessibility
(normalized)
Day

Figure 2

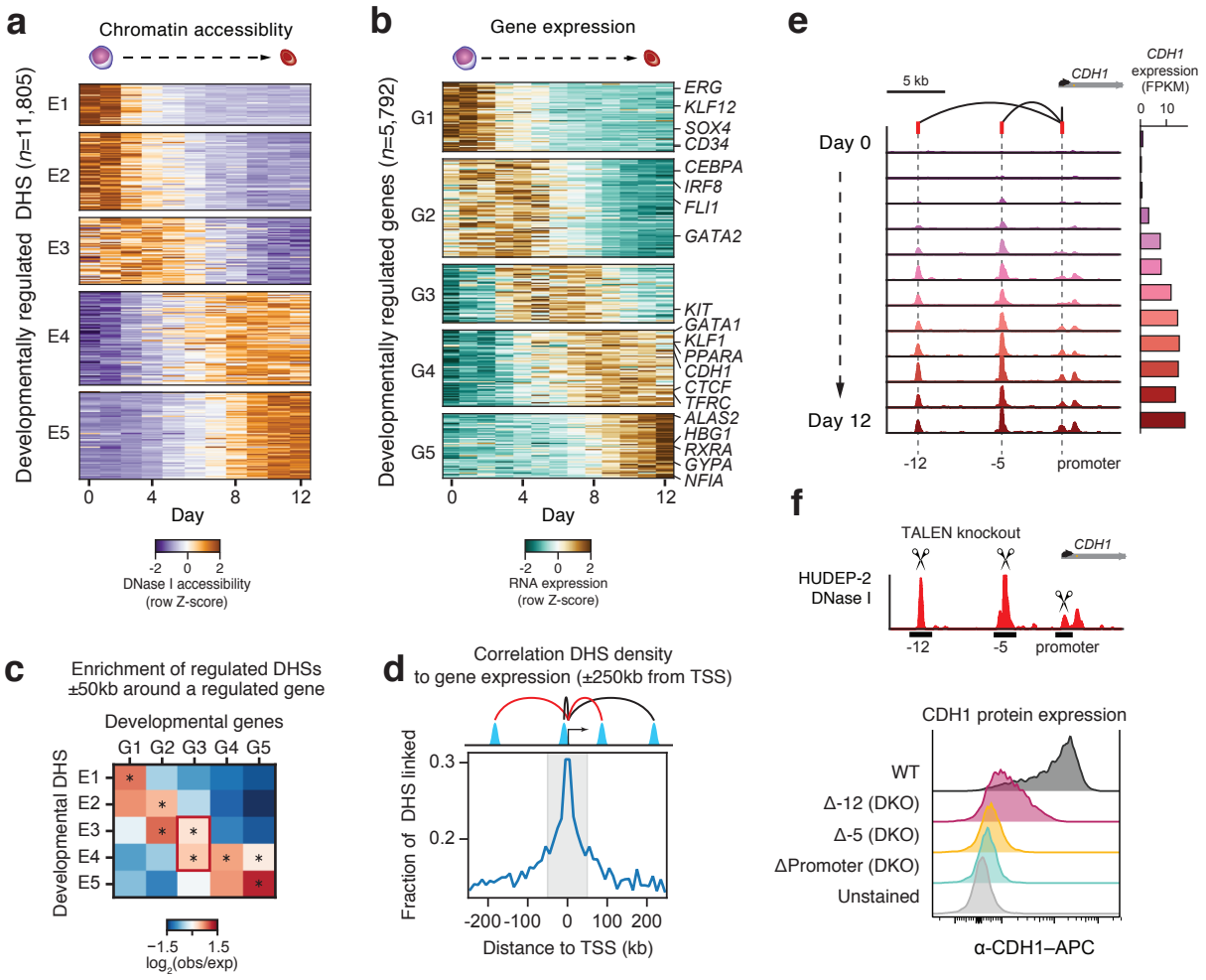

a
Chromatin accessiblity
Developmentally regulated DHS (n=11,805)
E1
E2
E3
E4
E5
0
4
8
12
Day
-2 0 2
DNase I accessibility
(row Z-score)
b
Gene expression
Developmentally regulated genes (n=5,792)
G1
G2
G3
G4
G5
ERG
KLF12
SOX4
CD34
CEBPA
IRF8
FLI1
GATA2
KIT
GATA1
KLF1
PPARA
CDH1
CTCF
TFRC
ALAS2
HBG1
RXRA
GYPA
NFIA
0
4
8
12
Day
-2 0 2
RNA expression
(row Z-score)
e
5 kb
CDH1
CDH1
expression
(FPKM)
0 10
Day 0
Day 12
-12
-5
promoter
f
TALEN knockout
CDH1
HUDEP-2
DNase I
-12
-5
promoter
CDH1 protein expression
WT
Δ-12 (DKO)
Δ-5 (DKO)
ΔPromoter (DKO)
Unstained
α-CDH1–APC
c
Enrichment of regulated DHSs
±50kb around a regulated gene
Developmental genes
G1 G2 G3 G4 G5
Developmental DHS
E1
E2
E3
E4
E5
−1.5 1.5
log2(obs/exp)
d
Correlation DHS density
to gene expression (±250kb from TSS)
Fraction of DHS linked
0.3
0.2
-200 -100 0 100 200
Distance to TSS (kb)

Figure 3

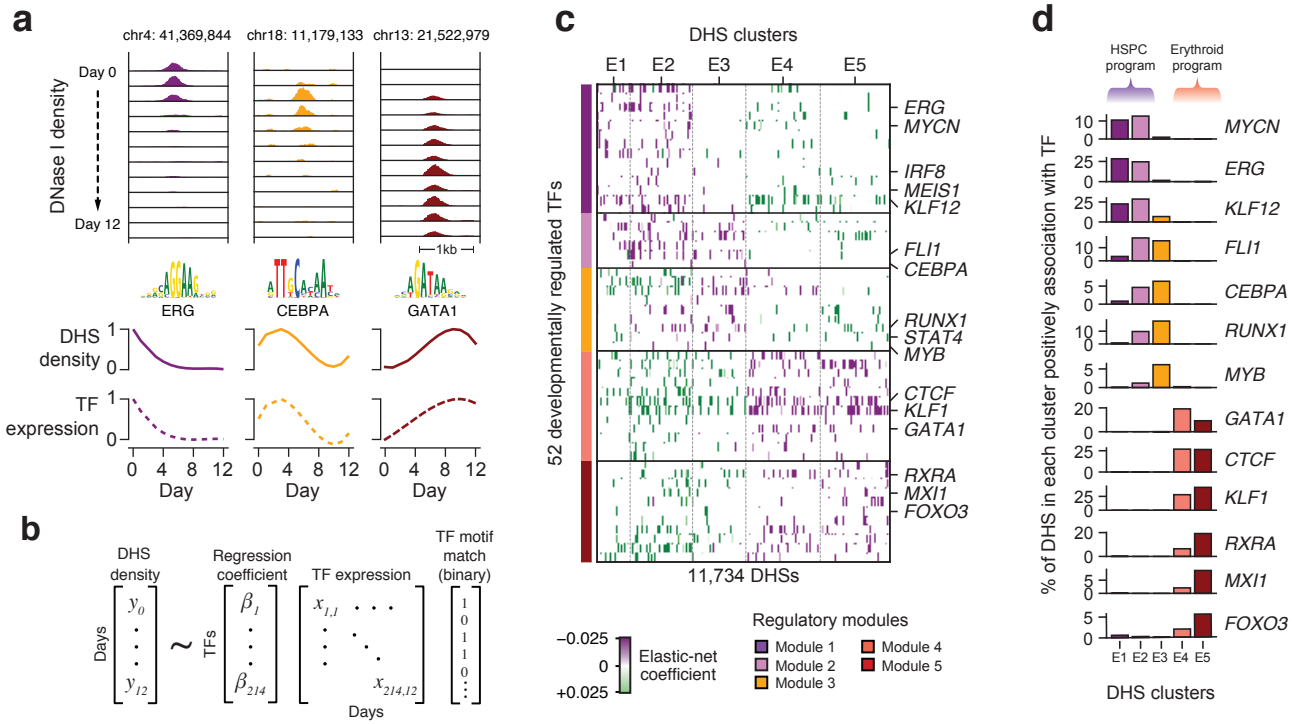

a
chr4: 41,369,844
chr18: 11,179,133
chr13: 21,522,979
Day 0
Day 12
DNase I density
1kb
ERG
CEBPA
GATA1
DHS density
TF expression
Day
b
DHS density
Regression coefficient
TF expression
TF motif match (binary)
Days
TFs
c
DHS clusters
E1
E2
E3
E4
E5
52 developmentally regulated TFs
ERG
MYCN
IRF8
MEIS1
KLF12
FLI1
CEBPA
RUNX1
STAT4
MYB
CTCF
KLF1
GATA1
RXRA
MXI1
FOXO3
11,734 DHSs
Elastic-net coefficient
Regulatory modules
Module 1
Module 2
Module 3
Module 4
Module 5
d
HSPC program
Erythroid program
% of DHS in each cluster positively association with TF
DHS clusters

Figure 4

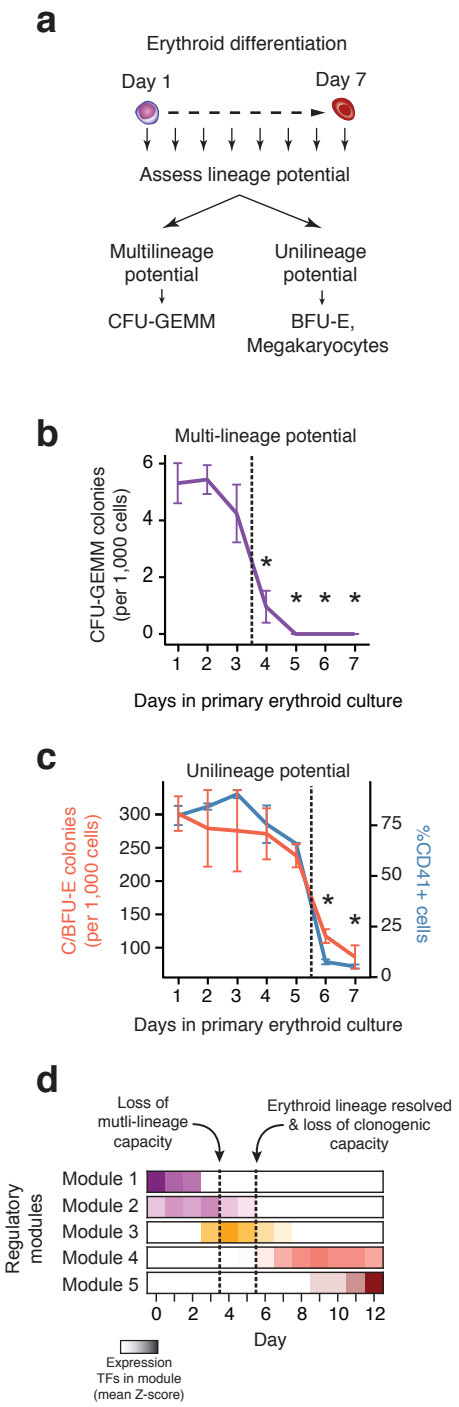

a
Erythroid differentiation
Day 1
Day 7
Assess lineage potential
Multilineage potential
CFU-GEMM
Unilineage potential
BFU-E, Megakaryocytes
b
Multi-lineage potential
CFU-GEMM colonies (per 1,000 cells)
Days in primary erythroid culture
c
Unilineage potential
C/BFU-E colonies (per 1,000 cells)
%CD41+ cells
Days in primary erythroid culture
d
Loss of mutli-lineage capacity
Erythroid lineage resolved & loss of clonogenic capacity
Regulatory modules
Module 1
Module 2
Module 3
Module 4
Module 5
Day
Expression TFs in module (mean Z-score)

Figure 5

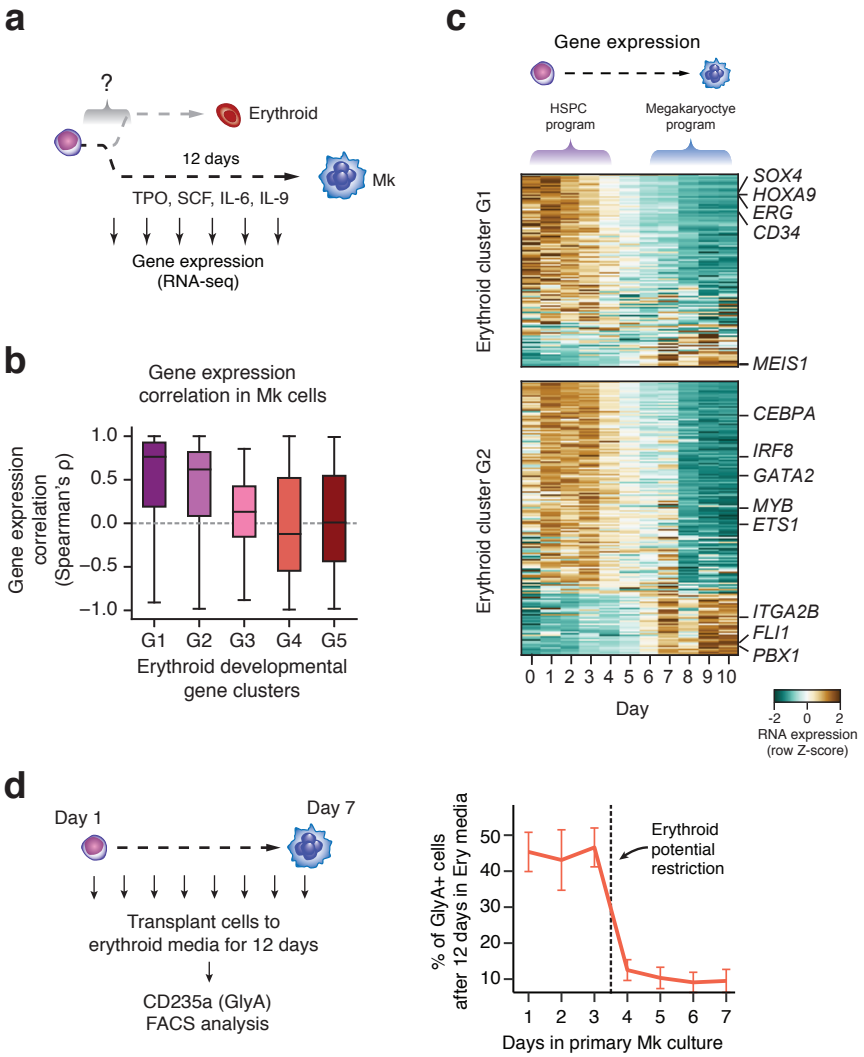
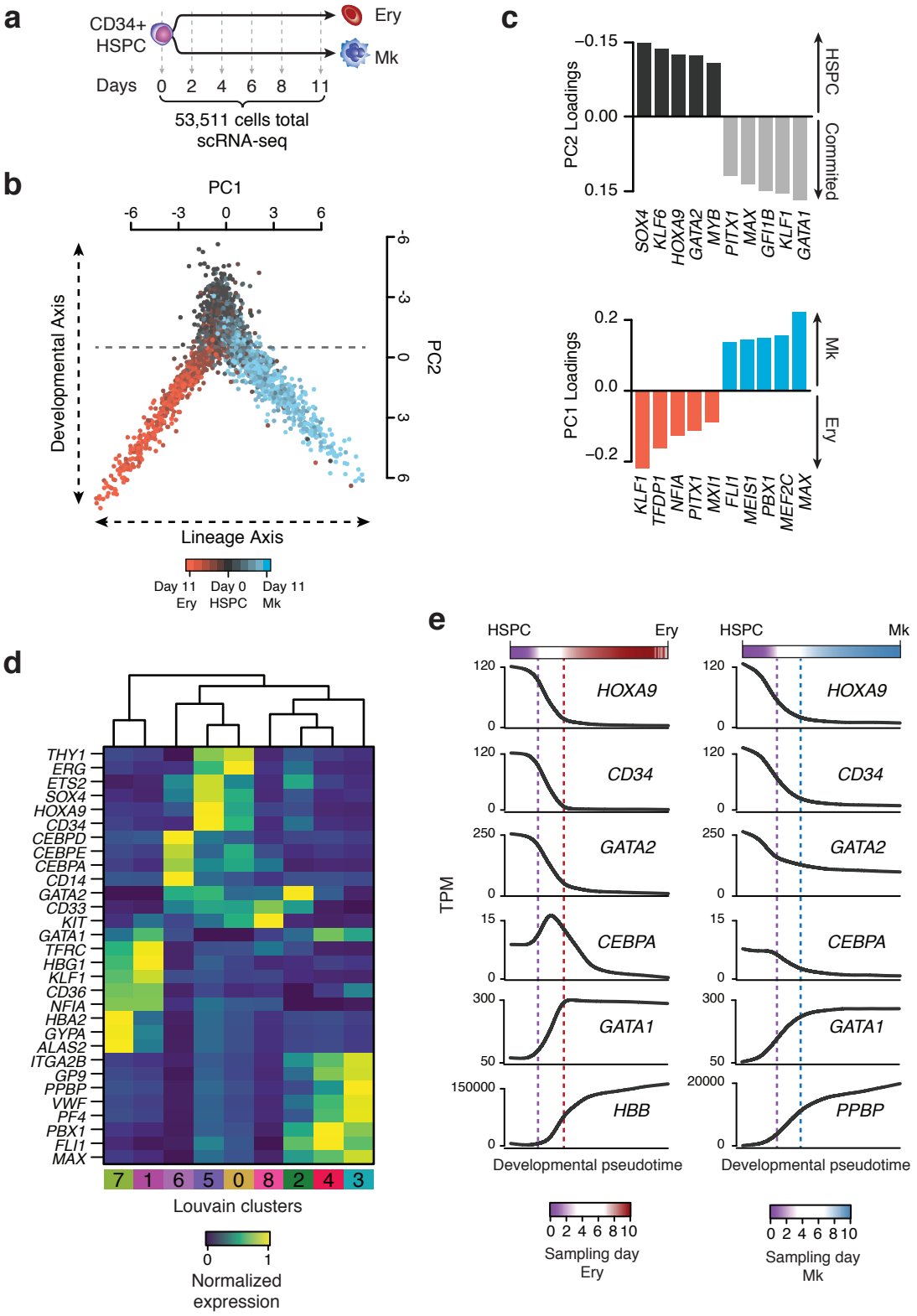

a
?
Erythroid
12 days
Mk
TPO, SCF, IL-6, IL-9
Gene expression
(RNA-seq)
b
Gene expression
correlation in Mk cells
Gene expression
correlation
(Spearman's ρ)
1.0
0.5
0.0
−0.5
−1.0
G1 G2 G3 G4 G5
Erythroid developmental
gene clusters
c
Gene expression
HSPC
program
Megakaryoctye
program
Erythroid cluster G1
Erythroid cluster G2
SOX4
HOXA9
ERG
CD34
MEIS1
CEBPA
IRF8
GATA2
MYB
ETS1
ITGA2B
FLI1
PBX1
0 1 2 3 4 5 6 7 8 9 10
Day
-2 0 2
RNA expression
(row Z-score)
d
Day 1
Day 7
Transplant cells to
erythroid media for 12 days
CD235a (GlyA)
FACS analysis
% of GlyA+ cells
after 12 days in Ery media
50
40
30
20
10
Erythroid
potential
restriction
1 2 3 4 5 6 7
Days in primary Mk culture

Figure 6


a
CD34+
HSPC
Ery
Mk
Days 0 2 4 6 8 11
53,511 cells total
scRNA-seq
b
PC1
-6 -3 0 3 6
Developmental Axis
PC2
Lineage Axis
Day 11 Ery
Day 0 HSPC
Day 11 Mk
c
PC2 Loadings
−0.15
0.00
0.15
HSPC
Commited
SOX4 KLF6 HOXA9 GATA2 MYB PITX1 MAX GFI1B KLF1 GATA1
PC1 Loadings
0.2
0.0
−0.2
Mk
Ery
KLF1 TFDP1 NFIA PITX1 MXI1 FLI1 MEIS1 PBX1 MEF2C MAX
d
THY1 ERG ETS2 SOX4 HOXA9 CD34 CEBPD CEBPE CEBPA CD14 GATA2 CD33 KIT GATA1 TFRC HBG1 KLF1 CD36 NFIA HBA2 GYPA ALAS2 ITGA2B GP9 PPBP VWF PF4 PBX1 FLI1 MAX
7 1 6 5 0 8 2 4 3
Louvain clusters
0 1
Normalized expression
e
HSPC
Ery
Mk
TPM
HOXA9
CD34
GATA2
CEBPA
GATA1
HBB
PPBP
Developmental pseudotime
0 2 4 6 8 10
Sampling day Ery
Sampling day Mk