UC Santa Cruz
UC Santa Cruz Electronic Theses and Dissertations

Title

ATAC-ing the Epigenetic Roots of Lineage Fate Choice in Hematopoiesis

Permalink

https://escholarship.org/uc/item/0w22478d

Author

Martin, Eric William

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**ATAC-ING THE EPIGENETIC ROOTS OF LINEAGE FATE CHOICE IN
HEMATOPOIESIS**

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

MOLECULAR, CELLULAR, AND DEVELOPMENTAL BIOLOGY

by

Eric William Martin

June 2022

The Dissertation of Eric W. Martin is
approved:

_____
Professor E. Camilla Forsberg, chair

_____
Professor Angela Brooks

_____
Professor Rohinton Kamakaka

_____
Professor Daniel H. Kim

_____
Peter Biel
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

List of Figures and Tables

Abstract

ATAC-ing the Epigenetic Roots of Lineage Fate Choice in Hematopoiesis

Eric William Martin

The goal of my thesis was to resolve the epigenetic mechanisms that govern hematopoietic stem cell (HSC) multipotency and differentiation. Hematopoiesis describes the essential function of constantly producing every type of blood cell daily throughout life. To accomplish this multilineage differentiation, the HSC and intermediate progenitors require precise temporal expression and control of cell-type-specific genes throughout differentiation. Cis-regulatory elements (CREs) regulate the temporal expression of these genes during hematopoiesis. The CREs themselves are regulated by altering the accessibility of those elements to transcription factor binding. The priming of CREs for transcription factor binding and gene activation without active expression allows a cell to be developmentally competent for specific cell types and act upon inductive signals to specify cell fate. Understanding the dynamics of cis-element accessibility is essential to understanding the mechanisms of stem fate decisions in hematopoiesis. My thesis project focused on characterizing the CRE accessibility dynamics in hematopoiesis to uncover regulators of fate decisions and then functionally interrogating the CREs to elucidate their function in hematopoiesis. Understanding these mechanisms during normal hematopoiesis will allow us to better understand the pathogenesis of hematological malignancies and manipulate the system to control lineage output.

Dedication

This work is dedicated to my parents and sister,

to my partner, Julianna,

and my son, Everett.

Thank you for inspiring, supporting, and believing in me.

I love you all!

Acknowledgments

My journey to my Ph.D. would not have been possible or as fun without the many colleagues, collaborators, and friends that I had the privilege of interacting with during my time here at UC Santa Cruz. I apologize if I missed recognizing you here, but I will try my best.

First and foremost, I would like to thank my mentor, Dr. E. Camilla Forsberg. I would not be here if it wasn't for her giving me the opportunity to join her lab first as an intern and then as a graduate student. Her love and excitement for science and making new discoveries is infectious. I thoroughly enjoyed our many conversations and meetings about research in our lab and the greater hematopoietic field. I am forever grateful to have had this opportunity, especially in an incredible lab and environment that Camilla has fostered as a professor at UC Santa Cruz.

Speaking of the lab, I would not have had any success if it wasn't for the support and collaboration of my fellow lab members through the years. I want to thank Anna Beaudin, Gloria Hernandez, Jessie Perez-Cunningham, Rebekah Sousae, Scott Boyer, Stephanie Smith-Berdan, and Susan Calhoun, that were friendly, welcoming, and super helpful in getting me up to speed with research in a new system and setting when I first joined. I want to thank Jana Krietsch and Smrithi Rajendiran, two incredible post-docs that I got to work directly with on several projects and who really helped me shape my own research path. To my undergrads, Sophia Graham and Sarah Tedjakusuma, you both will do amazing things. I am very proud to have been your mentor during your time in the lab. Finally, to my fellow graduate student lab mates,

I especially want to thank Oscar Cazares, Braden Larson, and Elektra Robinson for helping me navigate grad school and being great friends. I'm excited to see what amazing things you all will do in the future.

I would also like to thank all the staff who help support UCSC's research. None of the research presented in this thesis would be possible without their help and support. I would like to thank Bari Nazario for her tireless efforts advising with experimental design, setup, and keeping the flow cytometers running as well as tissue culture, and the delicious treats that gave me energy on long days and comfort on those failed experiment days. Thank you, Dr. Ben Abrams, for your help and guidance with using the microscopes and microscopy experiments, and always a friendly hello and a quick catch-up whenever passing by.

I would like to thank my friends on the unofficial UCSC STEM graduate student (and friends) softball team. The best escape and release from the stresses of grad school was going out and playing a game on a team with a wide variety of skills and comedy, learning about what fantastic research other grad students are doing, and just having a good time.

I would like to thank my family. To my mom and dad, Cathy Carson and Nick Martin, thank you for your inspiration and support throughout my entire educational career and for being a great mom and dad. I am forever grateful for all those trips to the aquarium, museums, exploratorium, academy of sciences, the random seminars at MJC, and days spent at the hospital or in the classroom. I remember not always being excited to

go, but once we were there, I always found something interesting and enjoyed learning something new each time. Thank you. To my sister, Michelle Martin, you're the bestest sister anyone could ask for. Thank you!

I would like to thank my partner, Julianna. You have been my co-pilot, and I am so grateful for your patience and support. Thank you for listening to me rant and rave about my struggles and triumphs throughout grad school. And thank you for teaching me how to better communicate my science. I am eternally grateful that I got to marry my best friend and spend all our adventures, both big and small, together. I love you.

Finally, my son, Everett. You have been my greatest joy and biggest distraction. Watching you grow up has and will continue to be my greatest inspiration. I am so very proud of you, and I hope you are proud of daddy. I love you.

Chapter 1: Lineage Priming as a Regulator of Hematopoietic Stem Cell Multipotency.

Building the Tree of Hematopoiesis

The goal of my thesis was to decode the epigenetic regulation of lineage fate choice in hematopoiesis. Hematopoiesis describes the essential function of producing every type of blood cell throughout life. With its roots stemming from the writings of Ernest Haeckel, the depiction of how a stem cell and progenitors are organized throughout differentiation is modeled after a tree (Haeckel, 1868). There has been a massive effort to shape and prune the hematopoietic tree (**Figure 1.1**), with the first isolation and characterization
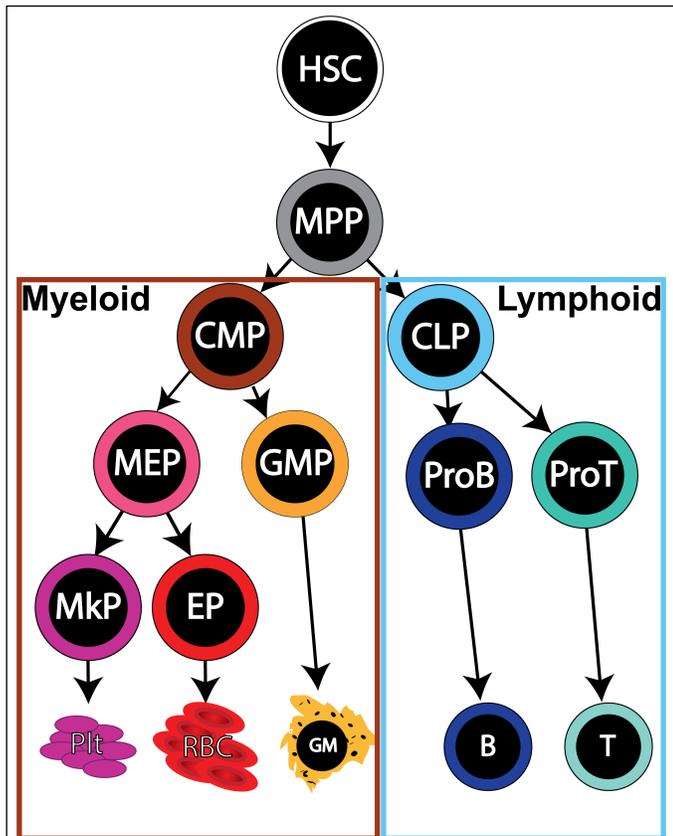
Figure 1.1: The Hematopoietic Tree.
This model of the hematopoietic tree with the hematopoietic stem cell (HSC) at the top, followed by the Multipotent Progenitor (MPP), then splits into the Myeloid and Lymphoid branches with the Common Myeloid Progenitor (CMP) and Common Lymphoid Progenitor (CLP) at the top of each branch, respectively. In the myeloid branch, the CMP differentiates into the Granulocyte/ Macrophage Progenitor (GMP) which differentiates into Granulocytes/Macrophages (GMs). Alternatively, the CMP can differentiate via Megakaryocyte/Erythrocyte Progenitors (MEPs) into either the Megakaryocyte Progenitor (MkP) which makes platelets (Plt), or the Erythroid Progenitor (EP), which makes red blood cells. The Lymphoid branch splits into B-cell Progenitors (ProB) which differentiate into B cells or T-cell Progenitors (ProT) which differentiate into T cells.

of the *bonafide* hematopoietic stem cell (HSC) by fluorescence-activated cell sorting (FACS) (Spangrude et al., 1988). With the tree's apex discovered, the pathways from stem cell to mature cell were mapped by the discovery and characterization of the intermediate progenitors in the hematopoietic tree. From the HSC, all cells differentiate through an Flk2+ Multipotent Progenitor (MPP) (Boyer et al., 2011), then the tree splits into myeloid and lymphoid lineages, with the Common Myeloid Progenitor (CMP) (Akashi et al., 2000) and the Common Lymphoid Progenitor (CLP) (Kondo et al., 1997) at the top of each branch, respectively. Further studies have led to refinement and modifications to the hematopoietic tree, driven by the advancement of flow cytometer technology and newly defined phenotypes to isolate and define the cell types in hematopoiesis (Orkin and Zon, 2008; Pronk et al., 2007).

Although the tree is modeled symmetrically, observations from steady-state and quantitative transplantation experiments indicate a heavy bias towards the erythroid and megakaryocyte output (Boyer et al., 2019; Busch et al., 2015; Rodriguez-Fraticelli et al., 2018; Yamamoto et al., 2013). Taken altogether, the multipotent ability of HSCs combined with the observed bias towards specific lineages leads us to ask how multipotency and fate decisions are regulated to continually generate the vast amount and diversity of different blood cell types that are essential for life.

Epigenetic Regulation of Lineage Fate Decisions

Developmental competence describes the ability of a cell to respond to inductive cues and is required for the proper temporal initiation of differentiation programs (Waddington, 1940). Lineage priming is hypothesized to be the mechanism of

Figure 1.2 Priming of CREs.
Cis-regulatory elements (CREs) are primed by an increase in chromatin accessibility which allows for transcription factor binding and are typically marked by H3K4me1 histone modifications. When the target gene is active, the CRE is marked by H3K27Ac and the chromatin loops to put the CRE and promoter into close contact with each other.
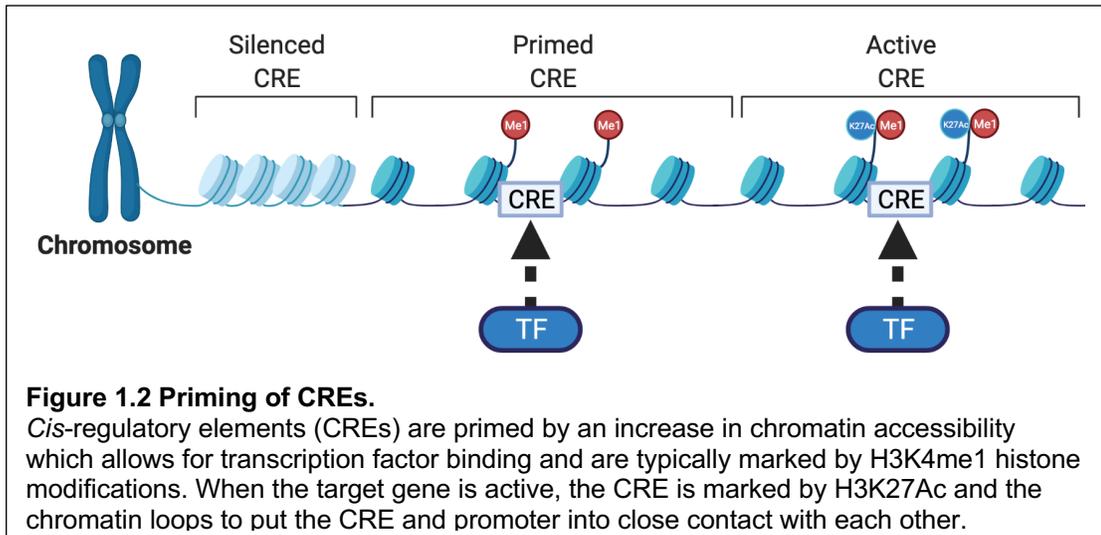
developmental competence and an essential property for multipotent stem cells. Lineage priming is defined as: Cis-regulatory elements (CREs) that drive differentiation towards a particular lineage are accessible to transcription factor binding and can respond to either an intrinsic or extrinsic inductive cue to initiate gene expression of its target gene and begin the differentiation program. Mechanistically, an increase in chromatin accessibility and the deposition of H3K4me1 histone modification marks priming. Genes become activated by subsequent chromatin looping to localize the CRE and promoter to each other and is marked by RNA Pol II binding to the promoter and the deposition of H3K27Ac at the CRE (Shlyueva et al., 2014). For genes that are important in determining lineage fate, the regulation and timing of CRE priming and subsequent target gene activation is vital to ensure proper expression at critical developmental timepoints. Multiple development studies highlight this strict control, with examples ranging from the precise regulation of the beta-globin cluster during development (Bonifer and Cockerill, 2017; Forsberg et al., 2000) to CRE priming in endoderm to pancreatic differentiation (Wang et al., 2015). This importance of proper temporal priming and subsequent activation of genes in development

warrants further study into the dynamics and mechanisms of priming. The hematopoietic system lends itself as an excellent model for studying priming dynamics throughout differentiation due to the phenotypically and functionally well-defined intermediate progenitors that allow us to capture and track priming temporally.

In hematopoiesis, we hypothesize that regulation of CRE accessibility is essential for HSC multipotency and proper differentiation. In HSCs, low levels of nascent transcription of genes important for multiple lineages (Hu et al., 1997) indicated an open chromatin state (Ugarte et al., 2015). As HSCs differentiate, genes involved in the target lineage (lineage drivers) are up-regulated in intermediate progenitors, accompanied by CRE priming and activation (marked by H3K4me1 and H3K27Ac) (Lara-Astiaso et al., 2014). In contrast, the genes important for alternative lineages are silenced into heterochromatin. Additionally, analysis of transcription factor expression and binding throughout differentiation mimics this progressive upregulation and activation of lineage drivers, as well as silencing of alternative lineage drivers by their associated transcription factors (Goode et al., 2016; Schütte et al., 2016). These studies inform us of potential mechanisms of regulating lineage priming through poising and activation of enhancer regions (Lara-Astiaso et al., 2014) and the dynamics of transcription factor networks throughout differentiation (Goode et al., 2016; Nestorowa et al., 2016; Schütte et al., 2016). However, the technical limitations of the ChIP-seq used in these studies could have miscalled identified CREs or missed identifying them. In addition, the cell lines used for (Schütte et al., 2016) have a different competence compared to wild-type HSCs. Therefore, interrogation of

4

chromatin accessibility and functional assessment of CREs in the wild-type system by complementary methods is required.

A more efficient and less subjective method to identify and characterize CREs and priming in hematopoiesis is through interrogation of chromatin accessibility by the Assay for Transposase Accessible Chromatin by High Throughput Sequencing (ATAC-seq) (Buenrostro et al., 2013). This assay allows the assessment of genome-wide chromatin accessibility at a high resolution with a low amount of input. It is ideal for hematopoiesis as many progenitors and the HSC itself are particularly rare (Orkin and Zon, 2008) and precludes them from being assayed by other methods that require orders of magnitude of input (10^6-10^7 cells per assay) higher than ATAC-seq (10^4 cells down to single-cell) (Buenrostro et al., 2013; Ma et al., 2020). Other studies have partially characterized hematopoiesis, focusing on granulocyte/macrophage development in human cells (Buenrostro et al., 2018) and erythro-megakaryopoiesis (Heuston et al., 2018). These studies were highly informative and proved the power of using ATAC-seq to identify and characterize CREs and track priming throughout differentiation.

We leveraged the sensitivity of ATAC-seq to investigate the dynamics of CRE accessibility throughout the entirety of murine hematopoiesis. Our studies interrogated the dynamics of chromatin accessibility throughout the entirety of hematopoiesis, first comparing HSCs with unipotent/mature lineage cells (Chapter 2)(Martin et al., 2021), then tracking accessibility throughout the entire continuum of hematopoiesis (Chapter 3). Our studies identified CREs that are potentially important for lineage fate

determination in hematopoiesis. These CREs are now ready for functional assessment to link each CRE to its target gene and determine the CRE's function in fate choice during differentiation.

Functional characterization of cis-regulatory elements

Thanks to the advances in immune-based probing and chromatin accessibility techniques, we now have precise and highly scalable tools to identify potential regulatory networks in cells. However, there is a large discrepancy between the discovery of regulatory elements and verifying their actual function within a cell. Initial experiments for in vivo testing of CREs utilize knock-in techniques to replace the CRE sequence region with a selection cassette or other non-coding element (McDevitt et al., 1997; Rojo et al., 2019). Although effective in testing the function of the CRE in vivo, these techniques are extremely low throughput and labor-intensive. Alternatively, with the rapid development and modifications of the CRISPR/Cas system (Adli, 2018; Anzalone et al., 2020), we can utilize a nuclease dead Cas9 fused to a Krüppel-associated Box domain (dCas9-KRAB) to introduce heterochromatin in a sequence-specific manner to target CREs (Thakore et al., 2015). In our pilot study (Chapter 4), we utilized a transgenic mouse that expressed dCas9-KRAB in all cells to validate that our system could recapitulate the functional observations from CD115 promoter and enhancer knockout experiments (Dai et al., 2002; Rojo et al., 2019), which led to a decline in Granulocyte/Macrophage differentiation.

The studies that identify CREs typically discover tens of thousands of candidate CREs which require functional validation to determine if and how they function in the cell

types of interest. Typical CRISPR/Cas9 functional experiments focus on a single gene or CRE at a time and perform one perturbation per cell. These designs vastly limit the scalability of the assay and the feasibility of assaying candidate CREs in many cell types. To overcome these limitations, Shendure and colleagues combined single-cell RNA-seq with CRISPRi perturbations at a high MOI to assay many candidate CREs and characterize their function (crisprQTL mapping) (Gasperini et al., 2019). This combination approach allowed for the capture of gene expression perturbations to determine the target gene of each CRE, and the multiple perturbations per cell increased the the ability to detect changes in gene expression, and experiment power. One caveat of this method was that detection of the guide RNA in the cell required a particular transfer plasmid that limited what could be integrated into the target cell after transduction (Hill et al., 2018). Excitingly, an improvement to the guide RNA detection in the cells was developed, which utilized a target enrichment strategy for the guide RNAs that can be detected as part of a commercial (10x Genomics) single-cell RNA-seq assay (Replogle et al., 2020). In addition, this new transfer plasmid is more customizable and amenable to loading two guide RNAs per plasmid to increase the robustness or genomic coverage of the perturbation. By combining the crisprQTL mapping technique (Gasperini et al., 2019) with the improved single-cell CRISPR screen with direct guide RNA capture (Replogle et al., 2020), we aim to functionally interrogate thousands of candidate CREs to determine their roles in lineage priming and multipotency in hematopoiesis.

Chapter 2. Chromatin accessibility maps provide evidence of multilineage gene priming in hematopoietic stem cells

[This has been adapted from publication, **Chromatin accessibility maps provide evidence of multilineage gene priming in hematopoietic stem cells** (Martin et al. 2021)]

Eric W. Martin, Jana Krietsch, Roman E. Reggiardo, Rebekah Sousae, Daniel H. Kim, E. Camilla Forsberg[1]

Institute for the Biology of Stem Cells, Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA 95064, USA

Abstract

Hematopoietic stem cells (HSCs) have the capacity to differentiate into vastly different types of mature blood cells. The epigenetic mechanisms regulating the multilineage ability, or multipotency, of HSCs are not well understood. To test the hypothesis that *cis*-regulatory elements that control fate decisions for all lineages are primed in HSCs, we used ATAC-seq to compare chromatin accessibility of HSCs with five unipotent cell types. We observed the highest similarity in accessibility profiles between megakaryocyte progenitors and HSCs, whereas B cells had the greatest number of regions with de novo gain in accessibility during differentiation. Despite these differences, we identified *cis*-regulatory elements from all lineages that displayed epigenetic priming in HSCs. These findings provide new insights into the regulation of stem cell multipotency, as well as a resource to identify functional drivers of lineage fate.

HIGHLIGHTS:

- HSCs have higher global chromatin accessibility than any unilineage progeny

- Megakaryocyte Progenitors are the most closely related unipotent cell type to HSCs

- B cell commitment involves *de novo* chromatin accessibility

- Evidence of *cis* element priming of lineage-specific genes in HSCs

9

INTRODUCTION

Multipotency is a key feature of hematopoietic stem cells (HSCs) and essential for their ability to produce all types of blood and immune cells *in situ* and upon therapeutic stem cell transplantation. The mechanistic basis of multipotency is unclear, but previous studies have shown that the regulation of differentiation programs is achieved, in large part, through epigenetic remodeling of *cis*-regulatory elements (CREs) (Forsberg et al., 2000; Shivdasani et al., 1997; Wang et al., 2015). Thus, HSC multipotency may be enabled by accessible non-promoter CREs that keep loci competent for transcription factor binding and gene activation without active expression. Such selective "CRE priming" may underlie the developmental competence of specific cell types, which is then acted upon by inductive signals to gradually specify fate (Waddington, 1940). When all CREs that drive differentiation and lineage choice are primed in stem cells, that stem cell is in a permissive state **(Figure 2.1A)** and is competent to initiate differentiation into all mature lineages.

We sought to test two models of HSC multipotency that are based on regulation of chromatin organization: the "permissive fate model" and a "*de novo* activation model" **(Figure 2.1A)**. Supporting a role for the permissive model in stem cell lineage potential are observations of bivalent histone domains that maintain key developmental genes in embryonic stem cells (ESCs) poised for activation (Bernstein et al., 2006), and an overall accessible chromatin state in both ESCs and HSCs compared to lineage-restricted progenitors and mature cells (Gaspar-Maia et al., 2009, 2011; Ugarte et al., 2015). When differentiation occurs, the genes poised for differentiation into the induced lineage are activated while CREs that would drive differentiation into

alternative lineages are silenced. This has been observed in ESCs and during differentiation of ESCs into endoderm (Wang et al., 2015; Xu et al., 2009). Our observation of global chromatin condensation and localization of H3K9me3-marked repressed domains or heterochromatin towards the nuclear periphery during HSC differentiation also support the permissive model (Ugarte et al., 2015). Inversely, in the *de novo* activation model **(Figure 2.1A)**, CREs that drive lineage fate are inaccessible in HSCs. Differentiation and lineage choice occur by "unlocking" these CREs. Transcriptional and functional analyses of hematopoietic stem and progenitor cells support this *de novo* model, where lymphoid potential is *gained* in progenitor cells rather than being a consequence of CRE priming in HSCs (Boyer et al., 2019; Cabezas-Wallscheid et al., 2014; Forsberg et al., 2005; Månsson et al., 2007).

In order to interrogate these models and how they pertain to the regulation of competence in hematopoiesis, as well as gain a better understanding of the relationships between epigenetic, transcriptomic and functional observations, we mapped global chromatin accessibility using the Assay for Transposase Accessible Chromatin by High Throughput Sequencing (ATAC-seq) (Buenrostro et al., 2013). This assay allows assessment of high resolution, genome-wide chromatin accessibility throughout differentiation programs of rare cells. The dynamics of chromatin accessibility in erythro-megakaryopoiesis (Heuston et al., 2018) and granulocyte/macrophage development (Buenrostro et al., 2018) have been highly informative. From these studies, the bulk observations gave us insight into the dynamics of lineage commitment during hematopoiesis, while single-cell analysis revealed the heterogeneity of epigenomic states and, therefore, lineage bias in

progenitors throughout hematopoiesis. Based on those studies, as well as reports of global chromatin accessibility of embryonic (Bernstein et al., 2006; Bulut-Karslioglu et al., 2018; Gaspar-Maia et al., 2011) and hematopoietic (Cabal-Hierro et al., 2020; Ugarte et al., 2015) stem cells, we hypothesized that HSCs are in a permissive chromatin state where CREs that control fate decisions are primed in HSCs. Here, we tested this hypothesis by performing in-depth ATAC-seq investigation of HSCs and 5 unipotent lineage cell populations representing the five main hematopoietic lineages **(Figure 2.1B)**, as defined by previously published phenotypes (Boyer et al., 2011, 2019).

RESULTS

Mapping of chromatin accessibility in HSCs and unipotent lineage cells identified a tight association of megakaryocyte progenitors to HSCs.

To determine the dynamics of genome accessibility throughout hematopoiesis, we sorted six primary hematopoietic cell types (**Figure 2.1B**) and performed ATAC-seq of libraries with expected fragment size distributions **(Figure 2.2)** (Buenrostro et al., 2013). We identified 70,731 peaks in HSCs, 47,363 peaks in megakaryocyte progenitors (MkPs), 38,007 in erythroid progenitors (EPs), 30,529 in granulocyte/macrophages (GMs), 70,358 in B cells, and 51,832 in T cells **(Table 2.1)**. From these peak-lists we combined and filtered the peaks using the chromVAR package to only the most significant peaks, as defined by (Schep et al., 2017a) and identified a total of 84,243 peaks, referred to as the master peak-list throughout the study **(Table 2.1)**. To assess data quality, we analyzed replicate clustering and cell type relationships of all 6 cell types using principal component analysis and

dimensionality reduction as a t-Distributed Stochastic Neighbor Embedding (tSNE) plot (Schep et al., 2017a). All biological replicate samples closely associated with each other by tSNE analysis (**Figure 2.1C**), as well as by hierarchical clustering using the chromVAR output (**Figure 2.1D**). We observed two primary clusters in **Figure 2.1D**: an HSC/MkP cluster and all other cell types. We also observed a distinct lymphoid cell subcluster containing only B and T cells, while GMs and EPs clustered independently. MkPs have the most similar accessibility to HSCs, with the ranking of the other cell types from most to least similar as EPs, GMs, Bs and then Ts. This is consistent with our tSNE analysis (**Figure 2.1C**), where HSCs and MkPs closely associated with each other, and with studies that have reported a close relationship of HSCs with the megakaryocyte lineage (Carrelha et al., 2018; Rodriguez-Fraticelli et al., 2018) and that erythropoiesis requires chromatin remodeling for differentiation to occur (Heuston et al., 2018).

Table 2.1 Peak counts and peak distribution relative to protein-coding gene promoters in each cell type

| Cell Type | ATAC peaks | Promoter Peaks (±500bp of TSS) | non-promoter peaks | | |
			coding (exons+TTS+TSS)	Introns	Intergenic
Master Peak-list	84,243	13,171	5,243	34,137	31,692
HSC	70,731	27,973	4,166	18,931	19,661
MkP	47,363	23,998	2,013	10,036	11,316
EP	38,007	23,243	2,014	7,040	5,710
GM	30,529	15,559	1,440	6,697	6,833
B	70,358	24,596	4,461	21,210	20,091
T	51,832	25,103	2,016	11,929	12,784

Figure 2.1: ATAC-seq maps of hematopoietic cell populations exhibit a high degree of reproducibility between replicates and a tight association of MkPs to HSCs. A. Two models of epigenetic regulation of HSC fate. In the "permissive fate" model, CREs of lineage-specific genes of all possible lineage outcomes are in an accessible state (green) in HSCs, keeping genes "primed" for subsequent activation. After lineage commitment occurs towards one fate, the accessibility of primed elements of the alternative fate is restricted by epigenetic remodeling (red). In contrast, the "de novo activation" model posits that CREs of lineage-specific genes are in an inaccessible state (red) in HSCs, keeping genes silenced. Lineage commitment occurs by de novo decondensing of chromatin at the appropriate CRE, allowing for subsequent activation of the differentiation program (green). The CREs of alternative lineage fates remain epigenetically repressed (red). **B.** Schematic diagram of the hematopoietic cells used in this study. Six cell populations were investigated: multipotent HSCs (Hematopoietic stem cells), unilineage MkPs (megakaryocyte progenitors) and EPs (erythroid progenitors), and mature GMs (Granulocyte/Macrophages), B cells, and T cells. **C.** tSNE analysis of the ATAC-seq peaks revealed a high concordance of biological replicates. MkPs clustered close to HSCs, while EPs, GMs, B, and T cells separated across the tSNE plot. **D**. Hierarchical clustering revealed high concordance of cell type-specific replicates. Similar to the tSNE analysis, MkPs clustered closest to HSCs. B and T cells were closely associated to each other but distant to HSCs, while GMs and EPs were contained within their own branches, closer to HSCs

Figure 2.1

A "Permissive Fate"

HSC

Gene A
Gene B

Gene B ⊣   ⊢ Gene A

Cell A   Cell B

"*De novo* activation"

HSC

Gene A
Gene B

→ Gene A   Gene B ←

Cell A   Cell B

B

HSC

MkP   EP   GM   B   T

Plt   RBC

C

tSNE dim 2

100

0

-100

-200   -100   0   100

tSNE dim 1

T

EP

B

GM

HSC

MkP

D

Bias-corrected deviations

0.6
0.4
0.2
0
-0.2

MkP
MkP
HSC
HSC
EP
EP
GM
GM
T
T
B
B

15

Figure 2.2

A   HSC   B   EP   C   MkP

D   GM   E   B cell   F   T cell
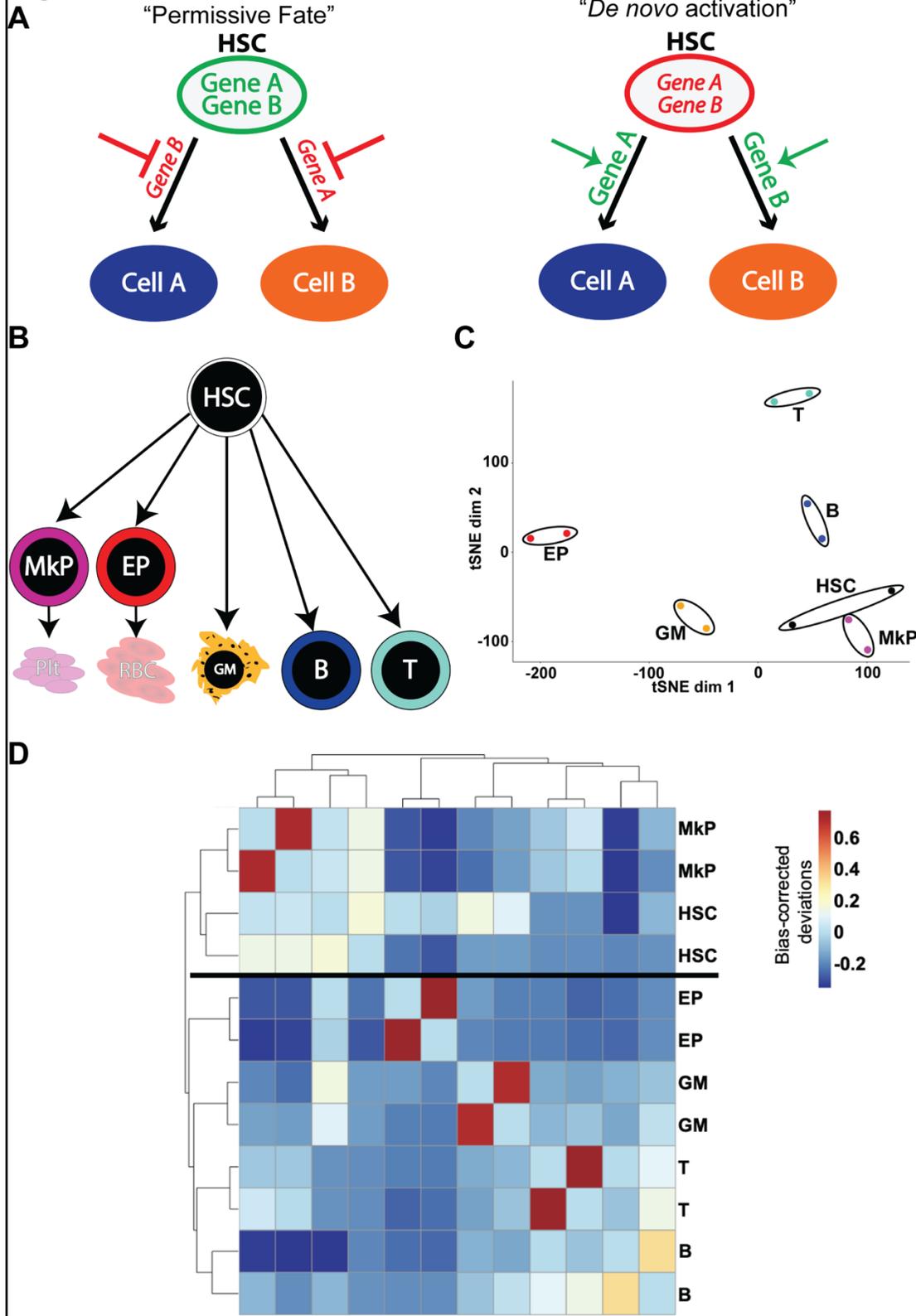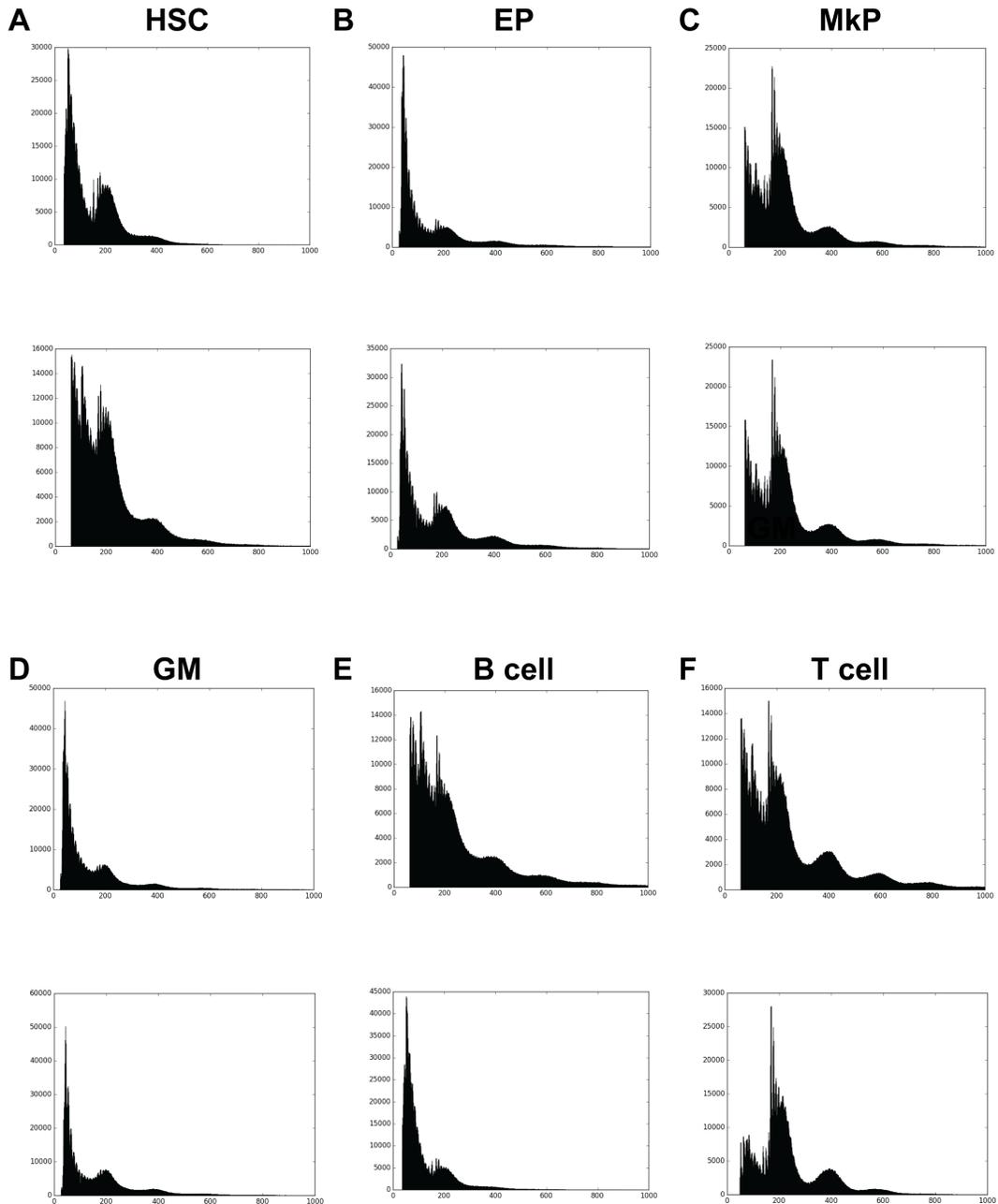
Figure 2.2: Library fragment distributions for ATAC-seq samples. The library size distribution after deep-sequencing, mapping, and filtering to unique reads is shown of both replicates for **A**) HSCs **B**) EPs **C**) MkPs **D)** GMs **E**) B cells, and **F**) T cells.

Visualization and comparison of ATAC-seq data generated in this study correlated with known expression patterns of cell type-specific genes.

As another assessment of the quality and reproducibility of our ATAC-seq data, we used the Gene Expression Commons (GEXC) expression database (Seita et al., 2012) to generate a list of genes that were expressed only in each unipotent lineage cell type (**Figure 2.3A**). From each list we calculated the normalized average signal centered at the promoter of each cell-type specific peak-list for each cell type by generating histograms using HOMER (Heinz et al., 2010) (**Figure 2.3B**). We observed the expected cell type-specific accessibility for each unipotent lineage with minimal signal from the other cell types. In addition, we visualized the ATAC-seq signals across promoters of some example genes with known cell type-specific expression patterns, plus a negative (expressed in none of the cell types) and a positive (expressed in all of the cell types) control, using the: *Gapdh* (expressed in all cell types)*, Fezf2* (not expressed in any cell type)*, Ndn* (expressed in HSCs only)*, Klf1* (EPs only)*, Gp6* (MkPs only)*, Ly6g* (GMs only), *CD19* (B cells only), and *Ccr4 (*T cells only) (**Figure 2.3C,D**). *Ly6g* was not available in GEXC but is a well-known GM-selective gene (Hestdal et al., 1991). We observed the expected accessibility peaks in each cell type, as well as a minimal signal from cell types without expression of those genes (**Figure 2.3D**). As an example of a well-characterized locus, we visualized our ATAC-seq data across the mouse b-globin cluster. As expected, we observed EP-selective accessibility of the hypersensitive sites in the locus control region (LCR) and of adult globin gene promoters *b-major* and *b-minor* (Li et al., 2002; Palstra et al., 2008) (**Figure 2.4**). The overall high level of reproducibility between independent sample replicates and clustering strategies (**Figure 2.1C,D**), as well as the expected accessibility in cell type-

specific genes (**Figure 2.3, Figure 2.4**), indicated that we had generated high-quality

chromatin accessibility maps of these 6 cell types.

Figure 2.3: Promoter accessibility correlated with known expression patterns of cell type-specific genes. A. Lineage-specific gene expression patterns used to find all genes expressed within each unipotent lineage cell type. The level of expression (red = high; blue = low/not expressed) according to the Gene Expression Commons (GEXC) database. **B.** Lineage-specific promoters had accessibility of the corresponding unipotent lineage cell types. Homer histograms of the average cumulative signal of all cell types used in this study across the lineage-specific promoter gene lists for EPs, MkPs, GMs, B cells, and T cells.
C. Lineage-specific expression of one example gene each for MkPs, EPs, B, or T cells. The level of expression (red = high; blue = low/not expressed) according to the Gene Expression Commons (GEXC) database of an example gene with cell type-specific ATAC-seq promoter peak. The probeset for the GM-specific Ly6g is not present in GEXC and therefore not displayed. **D**. Cell type-specific chromatin accessibility visualized as ATAC-seq read-counts at transcription start sites (TSS) using UCSC Genome Browser snapshots. Depiction of the six ATAC-seq libraries used in this study with example genes that had ATAC-seq signal in all samples (GAPDH; positive control), no samples (Fezf2; negative control), or in a specific cell type: HSCs (Ndn), EPs (Klf1), MkPs (Gp6), GMs (Ly6g), B cells (CD19), and T cells (Ccr4).

Figure 2.3

Figure 2.4: Erythroid-selective accessibility of the β-globin cluster. ATAC-seq signal tracks of the six cell types in this study at the β-globin cluster (chr7: 103,792,027–103,879,340; mm10). The adult globin genes β-*major (ßmaj)* and β-*minor (ßmin),* as well as the hypersensitive sites (HS1-4,6) of the Locus Control Region (LCR) that regulates expression of the genes in this locus, displayed accessibility in EPs. HS2, but not the other HSs, and the β-major promoter were also accessible in HSCs, possibly indicating a "permissive" chromatin state. Accessibility of HS2 and HS4 in MkPs may relate to a closer relationship to HSCs and/or EPs (Fig. 1). As expected, no accessibility was observed at the fetal-specific *epsilon Y globin (Ey),* β-*h1 (ßh1)*, β-*h2 (ßh2)* genes, or HS5. Likewise, GMs, B and T cells, that do not express β-globin genes, did not display accessibility of any of the regulatory elements in the locus.

HSCs have greater global accessibility and undergo more extensive chromatin remodeling upon lymphoid differentiation.

Using a number of quantitative, but non-sequence-specific assays, we previously reported that chromatin is progressively condensed upon HSC differentiation into unilineage and mature cells (Ugarte et al., 2015). To test whether the ATAC-seq data recapitulated these findings, we quantified the total number of distinct peaks, as well as the cumulative read-counts for all peaks, for each cell type. First, we took each cell type's optimal peak-list from the Irreproducible Discovery Rate (IDR) analysis (Li et al., 2011) and reported the number of peaks. We observed the highest number of peaks in HSCs (**Figure 2.5A**), closely followed by B cells. In parallel, we quantified global accessibility by calculating the normalized average signal over the master peak-list for

each cell type by generating histograms using HOMER (Heinz et al., 2010). We observed similar ordering compared to the peak number, with HSCs having the highest average signal and B cells the second highest (**Figure 2.5B**). The low signal in EPs is possibly due to widespread transcriptional silencing as the next step towards becoming highly specialized red blood cells and ejection of nuclei (An et al., 2014). Although these measurements are not completely independent, there is not a strict correlation between peak count and cumulative peak signal: for example, compared to EPs, GMs have fewer peaks (**Figure 2.5A**) but higher cumulative readcount (**Figure 2.5B**). Interestingly, HSCs displayed both the highest number of peaks and the greatest peak signal. These results are consistent with our previous findings of progressive chromatin condensation upon HSCs differentiation (Ugarte et al., 2015).

Figure 2.5: Greater overall global accessibility of HSCs. A. HSCs had the highest number of peaks of all hematopoietic cell types. The total number of individual peaks are displayed for each cell type. HSCs had the highest number of peaks followed by B cells, T cells, MkPs, EPs, then GM cells. **B**. HSCs had the highest total accessibility signal across all peaks of all hematopoietic cell types analyzed. Average cumulative signal across the master peak-list (the number of sequencing reads that fall into the detected peaks) was determined by the -hist function of HOMER annotatePeaks.pl.

Comparisons of peaks gained and lost as HSCs differentiate into unilineage cells revealed an overall gain of accessibility selectively for B cell differentiation.

To assess the number of peaks that changed upon HSCs differentiation, we took the IDR optimal peak-list for each cell type and performed pair-wise comparisons between HSCs and the five mature/unipotent cell types (**Figure 2.6A**). We quantified the number of peaks gained and lost by the unipotent progenitors/mature cells compared to HSCs (**Figure 2.6B-E**). MkPs had the lowest number of peak changes (peaks gained plus lost; **Figure 2.6B**), and therefore have the greatest proportion of peaks in common with HSCs. This was primarily driven by the low percentage of peaks gained (**Figure 2.6C**), as opposed to peaks lost (**Figure 2.6D**) upon HSC differentiation into MkPs. In contrast, EPs had the highest percentage of total peaks changed (**Figure 2.6B**) due to the greatest percentage of peaks lost (**Figure 2.6D**). This could be driven by EPs starting to shut down transcription to become highly specialized and eject their nuclei, reflected by the overall low accessibility observed (**Figure 2.5A,B**). B cells had the highest percentage of peaks gained and the lowest percentage of peaks lost compared to the other cell types (**Figure 2.6C,D**) and was the only cell type where the percentage of peaks gained was higher than peaks lost (**Figure 2.6E**). This suggests that B cell fate requires chromatin remodeling to open up sites that drive B cell lineage fate.

Figure 2.6

Figure 2.6: HSCs had more extensive chromatin remodeling upon lymphoid differentiation. A. HSCs had the highest number of peaks of all hematopoietic cell types. The total number of individual peaks are displayed for each cell type. HSCs had the highest number of peaks followed by B cells, T cells, MkPs, EPs, then GM cells. **B**. HSCs had the highest total accessibility signal across all peaks of all hematopoietic cell types analyzed. Average cumulative signal across the master peak-list (the number of sequencing reads that fall into the detected peaks) was determined by the -hist function of HOMER annotatePeaks.pl. **C–E.** Comparisons of the number of peaks gained and lost upon HSC differentiation into unipotent cells revealed that MkPs had the most similar accessibility profile to HSCs. **C**. Schematic of the pairwise comparisons made. HSC peaks were compared with one unilineage cell type at a time and those comparisons are reported in **B–E**. **B**. MkPs had the lowest percentage of altered peaks from HSCs compared to the other 4 unilineage cell types. The percentage of all non-overlapping peaks (peaks both gained and lost) calculated as the ratio of unique peaks in each cell type when compared pairwise to HSCs divided by the total number of peaks called in that cell type are displayed here. The numbers in the bars represent the total number of peaks altered (gained + lost) for each cell type. EPs had the highest percentage of peaks altered (gained + lost), followed by T cells, GMs, then B cells. **C**. B cells had the highest percentage of peaks gained from HSCs, while MkPs had the lowest. Calculations as in **B**, but only peaks gained are shown. **D**. EPs had the highest percentage of peaks lost from HSCs, while B cells had the least. Calculations as in **B**, but only peaks lost are shown. **E**. B cells were the only lineage with more peaks gained (53%) than lost (47%) upon differentiation from HSCs. In this panel, the sum of peaks gained and lost in each cell type was set to 100% and then the ratio of peaks gained and lost was displayed. T cells had the second highest proportion of peaks gained (27%), followed by MkPs (19%), EPs (16%), then GMs (15%)
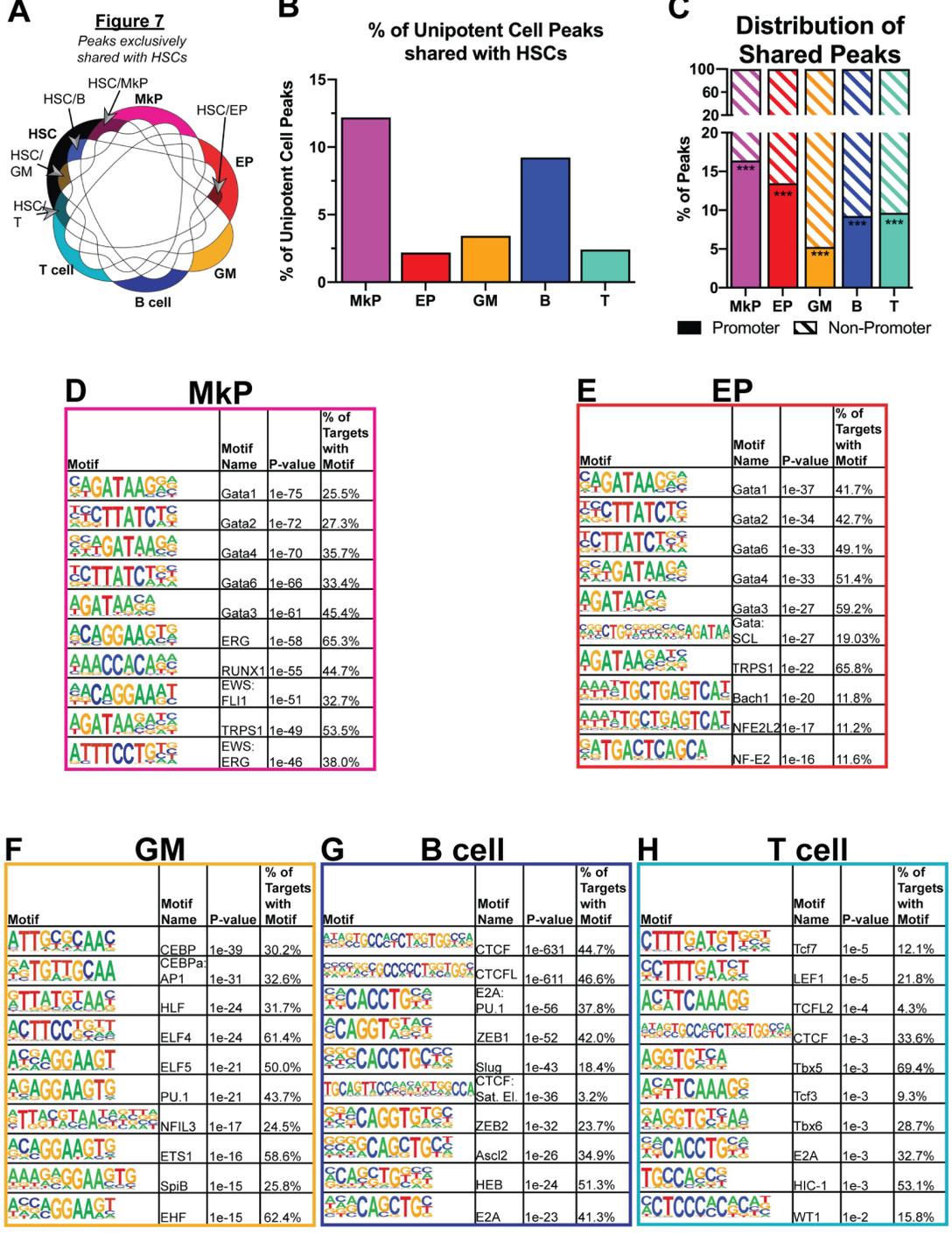
Exclusively shared peaks between HSCs and unipotent cell types are primarily non-promoter and are enriched for known cell-type specific transcription factors.

We then turned our attention from peaks that were different between HSCs and their progeny to instead focus on elements with shared accessibility. We hypothesized that peaks that are exclusively shared between HSCs and one unipotent cell type contain elements that drive lineage commitment into that cell type. We filtered the peak lists of all 6 cell types against each other using the HOMER mergePeaks.pl tool and annotated the peak lists that each of unipotent lineage cell types exclusively shared with HSCs (**Figure 2.7A**). We quantified the percentage of peaks that each unipotent cell type shared with HSCs (**Figure 2.7B**). Consistent with the clustering profiles (**Figure 2.7C-D**), MkPs had the highest percentage of peaks that were shared exclusively with HSCs. This similarity appeared to be primarily manifested in non-promoter elements: we annotated the exclusively shared peaks and categorized them as promoter or non-promoter peaks (**Figure 2.7C**) and compared the distributions to the annotated peak lists for each cell type assayed (**Table 2.1**). All of the exclusively shared peak lists had significant enrichment (p-value <0.001) of non-promoter peaks compared to the normal distribution of peaks in our dataset. Thus, non-promoter elements were shared between HSCs and their progeny significantly more frequently than promoter elements, especially with MkPs. Many, but likely not all, of these non-promoter accessible sites may serve as enhancers: about one-third of the non-promoter peaks overlapped with an enhancer catalog generated from chromatin immunoprecipitation (ChIP) experiments in blood cells (Lara-Astiaso et al., 2014) (**Figure 2.8A**). Similar levels of overlap was observed between the ATAC-accessible

peaks in our ATAC exclusively shared peak lists with H3K4me1 modifications in HSCs, while less overlap was observed for H3K27Ac, at the aggregate and cell type-specific level (**Figure 2.8B,C**).

Figure 2.7 Peaks shared between HSCs and unipotent cell types are primarily non-promoter and are enriched for known cell type-specific transcription factors. A. Schematic for how the unipotent lineage peaks exclusively intersected with HSC peaks were generated. Peaks were compared using HOMER mergePeaks.pl tool using peak-lists from the 6 cell types assayed. The resulting 5 overlapping peak-lists contained shared peaks between HSCs and only the unipotent cell type of interest (but not present in any of the other four lineages). The five exclusive pairwise comparisons (e.g., HSC/MkP only, HSC/EP, etc.) were used for panels B–H. **B**. MkPs have the highest peak overlap with HSCs. The number of unipotent lineage peaks that were uniquely intersected with HSCs was divided by the total number of peaks for each mature cell type. MkPs had the highest percentage of HSC overlap (12.2%), followed by B cells (9.2%), GMs (3.4%), T cells (2.4%), then EPs (2.2%). **C**. Peaks exclusively shared between each unipotent cell type and HSCs were significantly enriched in the non-promoter regions of the genome. The shared peak-lists described in a were annotated using HOMER annotatePeaks.pl function and filtered as promoter (< ± 500 bp from TSS), and non-promoter (< -500 bp and > + 500 bp from TSS). The number of promoter and non-promoter peaks was divided by the total number of peaks for each cell type. For all cell types, less than 20% of peaks were promoter peaks, with MkPs with the highest (16.4%) and GMs with the lowest (5.3%) percentage. This is a significant (< 0.001) difference compared to the normal distribution of promoter peaks (35-61%) for each cell type assayed. ***p-value of < .001. **D-H** Unipotent lineage peaks exclusively intersected with HSC peaks displayed enrichment of motifs for transcription factors with known roles in lineage differentiation. Motifs were found using HOMER findMotifsGenome.pl function, with a background file containing the combined peak-lists of the other 4 cell types. The top 10 results, as ranked by p-value from the known_motifs.html output, are shown. **D**. In MkP/HSC peaks, Gata family peaks made up 5 of the top 10 hits, followed by ERG, Runx1, and fusions EWS:FL1 and EWS:ERG. **E**. EP/HSC-enriched motifs also contained Gata factors, as well as the combination Gata:SCL motif and the known beta-globin locus control binder NFE2 and its paralog NFE2L2. **F**. GM/HSCs had CEBPa and PU.1 motifs as top hits, along with ETS transcription factor binding sites. **G**. B cell/HSC-enriched motifs had CTCF with CTCFL (BORIS) as the top two hits. B cells/HSC peaks also had E2A motifs enriched, as well as Ascl2, Slug, and ZEB1/2. **H**. Tcf7 motif was the top hit for T cell/HSC-shared peaks, along with CTCF and Tbx5/6. Similar to the B-cell/HSC list, the T-cell/HSC list was also enriched for E2A motifs.
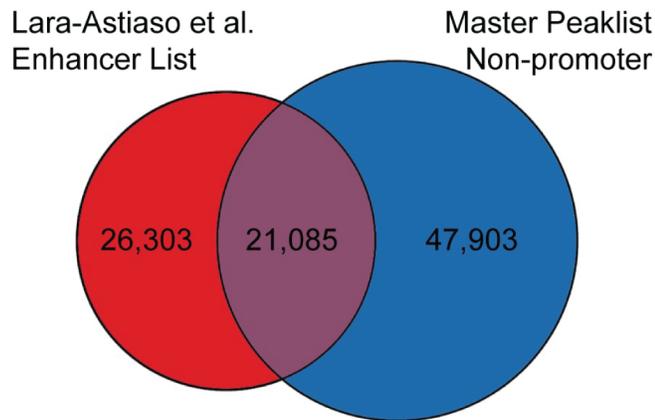
Figure 2.7

A

Figure 7
Peaks exclusively shared with HSCs

B % of Unipotent Cell Peaks shared with HSCs

C Distribution of Shared Peaks

D MkP

Motif	Motif Name	P-value	% of Targets with Motif
CAGATAAGGA	Gata1	1e-75	25.5%
TGCTTATCTG	Gata2	1e-72	27.3%
GCAGATAAGA	Gata4	1e-70	35.7%
TCTTATCTGC	Gata6	1e-66	33.4%
AGATAACA	Gata3	1e-61	45.4%
ACAGGAAGTG	ERG	1e-58	65.3%
AAACCACAGG	RUNX1	1e-55	44.7%
GACAGGAAAT	EWS:FLI1	1e-51	32.7%
AGATAAGATG	TRPS1	1e-49	53.5%
ATTTCCTGTG	EWS:ERG	1e-46	38.0%

E EP

Motif	Motif Name	P-value	% of Targets with Motif
CAGATAAGGA	Gata1	1e-37	41.7%
TGCTTATCTG	Gata2	1e-34	42.7%
TCTTATCTAX	Gata6	1e-33	49.1%
GCAGATAAGA	Gata4	1e-33	51.4%
AGATAACA	Gata3	1e-27	59.2%
CTGCCACAGATAA	Gata:SCL	1e-27	19.03%
AGATAAGATC	TRPS1	1e-22	65.8%
AAATTGCTGAGTCAT	Bach1	1e-20	11.8%
AAATTGCTGAGTCAT	NFE2L2	1e-17	11.2%
GATGACTCAGCA	NF-E2	1e-16	11.6%

F GM

Motif	Motif Name	P-value	% of Targets with Motif
ATTGCGCAAC	CEBP	1e-39	30.2%
GATGTTGCAA	CEBPa:AP1	1e-31	32.6%
GTTATGTAAC	HLF	1e-24	31.7%
ACTTCCTGTT	ELF4	1e-24	61.4%
ACCAGGAAGT	ELF5	1e-21	50.0%
AGAGGAAGTG	PU.1	1e-21	43.7%
ATTACGTAATATTA	NFIL3	1e-17	24.5%
ACAGGAAGTG	ETS1	1e-16	58.6%
AAAGAGGAAGTG	SpiB	1e-15	25.8%
ACCAGGAAGT	EHF	1e-15	62.4%

G B cell

Motif	Motif Name	P-value	% of Targets with Motif
ATAGTGCCACCCTAGTGGCCA	CTCF	1e-631	44.7%
CCCGCGCCCCCTAGTGGG	CTCFL	1e-611	46.6%
GCCACCTGGC	E2A:PU.1	1e-56	37.8%
GCAGGTGACA	ZEB1	1e-52	42.0%
GGGCACCTGCC	Slug	1e-43	18.4%
TGCAGTTCCAGAGTGGCCA	CTCF:Sat. El.	1e-36	3.2%
GGGCAGGTGTGC	ZEB2	1e-32	23.7%
GGGGCAGCTGCC	Ascl2	1e-26	34.9%
GCAGCTGCCC	HEB	1e-24	51.3%
GCAGCAGCTGCC	E2A	1e-23	41.3%

H T cell

Motif	Motif Name	P-value	% of Targets with Motif
CTTTGATGTGGT	Tcf7	1e-5	12.1%
CCTTTGATGT	LEF1	1e-5	21.8%
ACTTCAAAGG	TCFL2	1e-4	4.3%
ATAGTGCCACCCTAGTGGCCA	CTCF	1e-3	33.6%
AGGTGTCA	Tbx5	1e-3	69.4%
ACATTCAAAGG	Tcf3	1e-3	9.3%
AAGGTGTGAA	Tbx6	1e-3	28.7%
GCCACCTGCC	E2A	1e-3	32.7%
TGCCAGCG	HIC-1	1e-3	53.1%
GCTCCCACACAT	WT1	1e-2	15.8%

28

To determine what transcription factor binding sites were present within the exclusively shared peaks, we performed motif enrichment using the HOMER package and reported the top 10 results for each cell type, sorted by p-value (**Figure 2.7D-H**). The peaks that HSCs shared with MkPs (**Figure 2.7D**) or EPs (**Figure 2.7E**) were primarily enriched for Gata family transcription factors and their inhibitor TRPS1. Notably, HSC/MkP peaks also had enrichment of ERG and Runx1, which are known drivers of hematopoiesis (Growney et al., 2005; Kruse et al., 2009). For HSC/EPs, Gata1 was the most enriched motif, with the Gata:SCL combination motif and NF-E2 and NFE2L motifs also scoring in the top ten. These factors are all known to be important in red blood cell differentiation, and NF-E2 is known to regulate SCL and Gata2 (Siegwart et al., 2020). HSC/GM peaks had enrichment of known regulators of GM cell fate, such as CEBP, PU.1, and SpiB (**Figure 2.7F**). HSC/B cells primarily had CTCF and CTCFL motif enrichment (**Figure 2.7G**). These motifs could be a reason for the overall high number of peaks observed in B cells (**Figure 2.5A,B**), as 44.7% and 46.6% of the shared peaks contained CTCF or CTCFL motifs, respectively. HSC/T cell peaks were enriched for Tcf and Tbx family factors that are known to play a role in T cell development (**Figure 2.7H**). Overall, all five HSC-shared peak lists had enrichment of transcription factors that are known to be important for normal differentiation for each lineage.
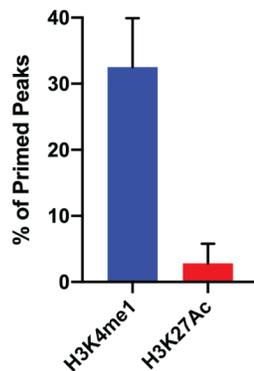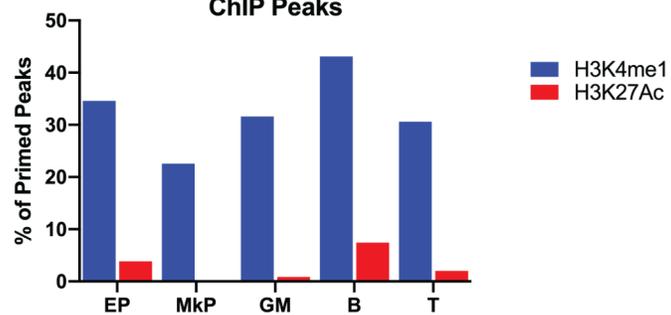
Figure 2.8: Lineage specific, HSC-primed peaks were marked by H3K4me1 and not H3K27Ac. A) About one in three (21,085 peaks out of 71,072) of our ATAC-seq non-promoter peaks in the master peak-list overlapped with peaks designated as probable enhancers based on H3K4me1 and H3K27Ac ChIP data (Lara-Astiaso et al.). **B**) About one in three (32.5%) of all HSC-primed peaks for the five unipotent lineage cell types were also marked by the histone modification H3K4me1, and 2.8% were marked by H3K27Ac. **C**) HSC-primed peaks for each unipotent lineage were primarily marked by H3K4me1 and not H3K27Ac. Results in panel B represent the aggregate of the results shown in panel C.

Evidence of *cis* element priming of lineage-specific genes in HSCs.

Previous work on understanding multipotency and developmental competence suggests a model where competence is conferred by transcriptional priming: being competent of transcription factor binding and gene expression, without active expression (Hu et al., 1997). One of the suggested regulators of transcriptional priming

are non-promoter *cis*-regulatory elements (CREs). This means that CREs that drive

lineage fate for all lineages are accessible in HSCs in our permissive fate model and

inaccessible in our *de novo* activation model. We hypothesized that CREs that are

exclusively shared between HSCs and a unipotent lineage cell are potential drivers of

that lineage. We utilized the GREAT tool (McLean et al., 2010) to annotate and predict

the target genes for each exclusively shared CRE. Here we report examples of genes

and a predicted CRE for each lineage that is primed in HSCs. In addition, we linked

the motif enrichment with the GREAT analysis by annotating the CREs using the top

10 motifs enriched by p-value **(Figure 2.7D-H)** for each exclusive HSC/unipotent cell

type. In MkPs, a predicted CRE for *Thrombin receptor like 2* (*F2rl2*) was found. This

gene is expressed only in MkPs **(Figure 2.9A)**, while the CRE is only accessible in

HSCs and MkPs **(Figure 2.9B)**. This CRE contained 9 out of the top 10 motifs, with

the Runx1 motif being the only one missing **(Figure 2.9C)**. *Pyruvate kinase liver and*

red blood cell (*Pklr)* was found to be expressed only in EPs **(Figure 2.9D)**, and a

predicted CRE was accessible only in HSCs and EPs **(Figure 2.9E).** Motifs for Gata2,

Gata3, Gata4, and TRPS1 were found within the CRE **(Figure 2.9F)**. In GMs,

Mitochondrial tumor suppressor 1 (*Mtus1*) was found to be primed in HSCs, with

expression only in GMs **(Figure 2.9G)**, accessibility of a predicted CRE only in HSCs

and GMs **(Figure 2.9H)**, and the presence of transcription factors known to play a role

in GM development, such as CEBP and PU.1 **(Figure 2.9I)**. In B cells, *Interferon*

regulatory factor 8 (*Irf8*), is only expressed in B cells **(Figure 2.9J)**, the predicted CRE

is only accessible in both B cells and HSCs **(Figure 2.9K)**, and contained 5 out of the

top 10 motifs, ZEB1/2, Slug, Ascl2, HEB, and E2A **(Figure 2.9L)**. In T cells, the gene

Inducible T cell co-stimulator (Icos) is only expressed in T cells **(Figure 2.9M)**, a

predicted linked CRE is accessible in both T cells and HSCs **(Figure 2.9N)** and contains motifs for CTCF and WT1 **(Figure 2.9O)**. Taken together, these examples represent CRE priming in HSCs, along with the corresponding transcription factors that may act on each element to guide HSC fate.

Figure 2.9 Examples of cis-element priming of lineage-specific genes in HSCs. A.
GEXC expression data reported expression of Thrombin receptor like 2 (F2lr2) selectively in
MkPs. **B**. A cis-element predicted to be associated with F2rl2 by GREAT was accessible in
both MkPs and HSCs, but not in any other unipotent cell type. **C**. The F2rl2 CRE contained
the transcription factor binding motifs for 9 out of the top 10 enriched motifs in MkPs. The
only motif not present is Runx1. **D**. GEXC expression data reported expression of Pyruvate
kinase liver and red blood cell (Pklr) in EPs, and not any other cell type. **E.** A cis-element
predicted to be associated with Pklr by GREAT was accessible in both EPs and HSCs, but
not in any other unipotent cell type. **F.** The Pklr CRE contained the binding motifs for Gata2,
Gata4, Gata3 and TRPS1. **G**. GEXC expression data reported selective expression of
Mitochondrial tumor suppressor 1 (Mtus1) in GMs and no expression in any other cell type.
H. A cis-element predicted to be associated with Mtus1 by GREAT was accessible in both
GMs and HSCs. **I**. CEBP, CEBP:AP1, HLF, PU.1, NFL3, ETS1, and EHF binding motifs were
present in the Mtus1 CRE reported in **H**. **J**. GEXC expression data reported Interferon
regulatory factor 8 (Irf8) expression only in B cells, not in the other unipotent lineage cells or
in HSCs. **K**. A cis-element predicted by GREAT to be associated with Irf8 was accessible in
both B cells and HSCs. **L**. ZEB1/2, Slug, Ascl2, HEB, and E2A binding motifs were found
within the Irf8 CRE displayed in **K**. **M**. GEXC expression data reported Inducible T cell co-
stimulator (Icos) expression only in T cells, but not in the other unipotent lineage cells or
HSCs. **N**. A cis-element predicted by GREAT to be associated with Icos was accessible in
both T cells and HSCs. **O**. CTCF and WT1 motifs were found within the Icos CRE displayed

Figure 2.9
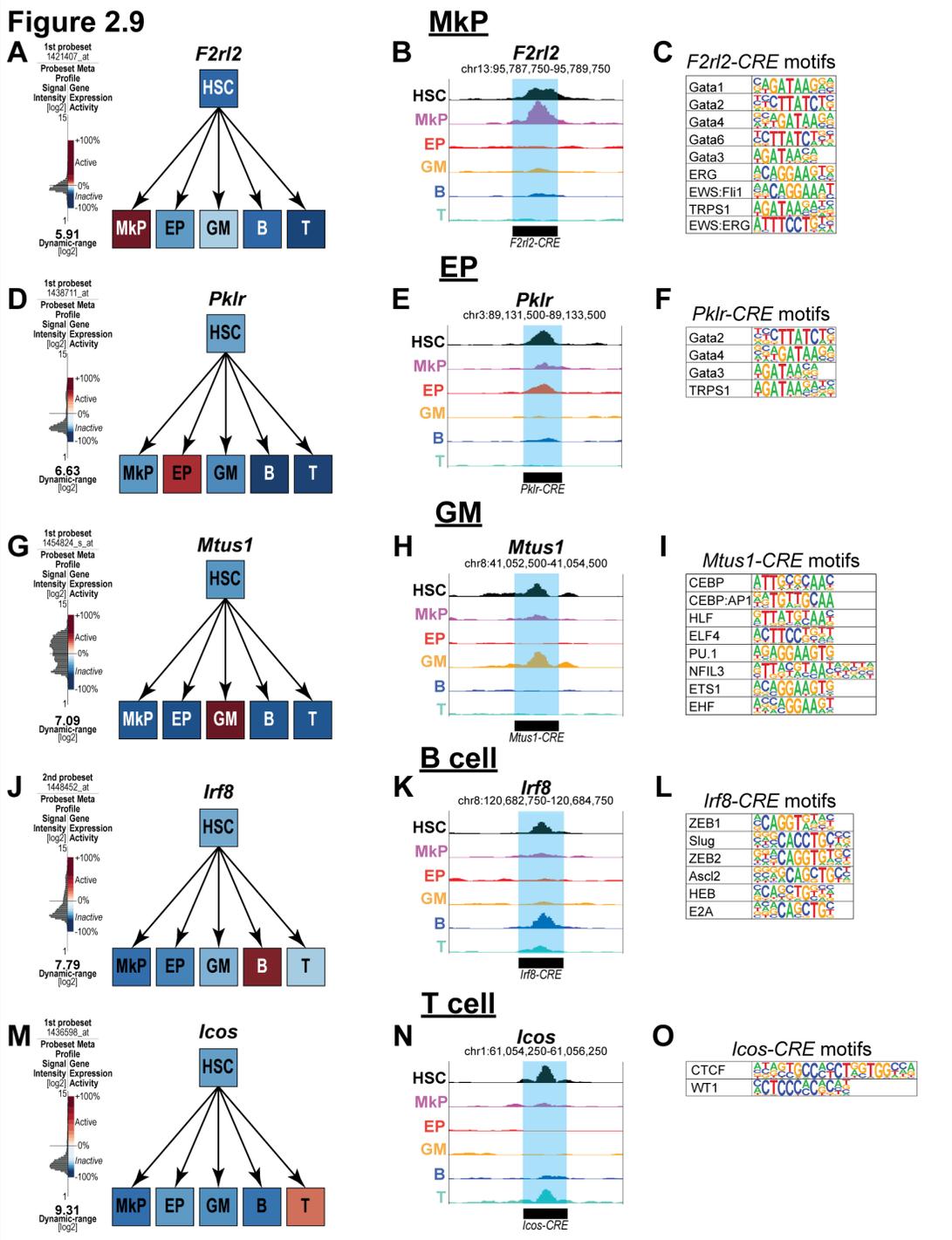
DISCUSSION

MkPs and HSCs have the most similar accessibility profile.

Here, we compared the genome-wide accessibility by ATAC-seq of the multipotent HSCs and unipotent lineage cell types (EPs, MkPs, GMs, B, and T cells). Through hierarchical clustering analysis, we observed erythromyeloid and lymphoid relationships that are consistent with the classical model of hematopoiesis **(Figure 2.1D)** (Boyer et al., 2011; Bryder et al., 2006; Laurenti and Göttgens, 2018; Seita and Weissman, 2010). By both PCA and hierarchical clustering, we observed that MkPs were the most similar to HSCs based on their accessibility profiles **(Figure 2.1)**. This relationship is reflected in a high level of overlap of peaks, as MkPs had the fewest peaks gained or lost from HSCs compared to the other cell types **(Figure 2.6)** and had the largest percentage of peaks exclusively shared with HSCs **(Figure 2.7B)**. These findings are in agreement with recent clonal studies of hematopoiesis that reported a megakaryocyte lineage bias of HSCs (Carrelha et al., 2018; Rodriguez-Fraticelli et al., 2018). According to hierarchal clustering, EPs had the second closest association to HSCs **(Figure 2.1D)** possibly supporting erythropoiesis as the default fate for hematopoiesis (Boyer et al., 2019) under conditions where chromatin remodeling silences megakaryocyte driver elements (Heuston et al., 2018). On the other end of the spectrum, the least similar cell types to HSCs were the lymphoid cell types **(Figure 2.1D)**. This greater difference was primarily due to a high proportion of peaks gained **(Figure 2.6C)** rather than lost **(Figure 2.6D)** upon differentiation from HSCs, leading to a greater ratio of peaks gained:lost for lymphoid cells than for erythromyeloid lineages **(Figure 2.6E)**.

Evidence of multilineage priming in HSCs.

The priming of genes for transcription likely initiates within CREs, which can then drive the activation of promoter targets. These enhancers can act as drivers of lineage fate (Wang et al., 2015) and their accessibility is a putative regulator of competence in stem cells. We made the assumption that peaks that are exclusively shared between HSCs and the unipotent lineage cells contain CREs that are specific for driving differentiation into that lineage. We observed that the majority of exclusively shared peaks were non-promoter peaks **(Figure 2.7C)** and were enriched for binding motifs of transcription factors known to be important for differentiation into each lineage **(Figure 2.7D-H)**. The enrichment of binding sites for known lineage-specific transcription factors suggests that many of the accessible sites may play functional roles. Additionally, about one third of the exclusively shared ATAC peaks were enriched for the H3K4me1 histone modification, which is linked to a primed enhancer state (Calo and Wysocka, 2013) indicated that a subset are likely functional enhancers (**Figure 2.8**); other ATAC-accessible elements may mark transcription start sites for non-coding genes, which are abundant and highly tissue-specific in the mouse genome (Ravasi et al., 2006). By using the GREAT tool, we made predictions for the target genes for the many ATAC-identified putative CREs that were present in the HSC/mature cell exclusive lists. The examples shown in **Figure 2.9** provide evidence that multi-lineage priming exists in HSCs.

Both permissive and *de novo* epigenetic mechanisms influence hematopoiesis.

Analogous to other stem cell systems, multipotent HSCs with the competence to differentiate into diverse cell types reside at the top of the blood cell hierarchy. We

tested two potential models of the mechanism of multipotency, the permissive fate and *de novo* activation **(Figure 2.1A)**. We found evidence for both. Supporting the permissive fate model are the observations that HSCs had the highest global accessibility **(Figure 2.5A/B)**, that peaks were lost in every unipotent cell type from HSCs **(Figure 2.6D)**, that every unipotent cell type shared some peaks exclusively with HSCs **(Figure 2.7B)**, and that evidence of multilineage priming of CREs were found in HSCs **(Figure 2.9)**. The *de novo* activation model was supported by the observation that new peaks were gained during differentiation into all five lineages (**Figure 2.6C**), and previous studies reporting progressive upregulation of lineage-specific genes as HSCs transition into progenitors (Forsberg et al., 2005; Terskikh et al., 2003). Interestingly, in the b-globin locus, HS2, the strongest enhancer of globin expression (Fiering et al., 1995; Bender et al., 2012)], was highly accessible in HSCs, whereas the other HSs were not (**Figure 2.4**). Thus, "priming" of this locus may occur in HSCs via HS2 (adhering to the permissive model of Figure 1A), followed by induced accessibility (de novo model, Figure 1A) of the other HSs and active b-globin expression upon erythroid differentiation. Thus, both permissive and de novo mechanisms likely influence hematopoietic fate decisions. Interestingly, we found evidence that the balance between the two models varies between lineages. For example, B cells, and to a lesser extent T cells, had a higher proportion of peaks gained than lost compared to erythromyeloid lineages (**Figure 2.6E**). This may indicate that the megakaryocyte/erythroid lineage is in a more primed state in HSCs, whereas lymphopoiesis requires more extensive chromatin remodeling to both prime lymphoid CREs not accessible in HSCs and simultaneously shut down the megakaryocyte/erythrocyte trajectory. The cell output and kinetics from *in vivo* lineage

tracing and reconstitution assays support these conclusions (Carrelha et al., 2018; Rodriguez-Fraticelli et al., 2018; Boyer et al., 2011, 2012, 2019; Yamamoto et al., 2013). Our identification of specific, putative regulatory CREs will enable functional testing of these elements.

EXPERIMENTAL PROCEDURES

Mice and Cells

All experiments were performed using 8- to 12-week-old C57BL/6 wild-type mice in accordance with UCSC IACUC guidelines. Hematopoietic cells were isolated from BM by crushing murine femurs, tibias, hips, and sternums as previously described (Rajendiran et al., 2020). Stem and progenitor cell fractions were enriched using CD117-coupled magnetic beads (Miltenyi). Cells were stained with unconjugated lineage rat antibodies (CD3, CD4, CD5, CD8, B220, Gr1, Mac1, and Ter119) followed by goat-α-rat PE-Cy5 (Invitrogen). Stem and progenitor cells were isolated using fluorescently labeled or biotinylated antibodies for the following antigens: cKit (2B8, Biolegend), Sca1 (D7, Biolegend), Slamf1(CD150) (TC15-12F12.2, Biolegend), CD41(MWReg30, Biolegend), and CD71(RI7217, Biolegend). Cells were sorted using a FACS Aria II (BD Bioscience). HSCs were defined as cKit$^+$ Lin$^-$ Sca1$^+$ Flk2$^-$ and Slamf1$^+$; MkPs as cKit$^+$Lin$^-$Sca1$^-$Slamf1$^-$CD41$^+$. Unipotent lineage cells were isolated by the following markers and as described previously (Cool et al., 2020; Leung et al., 2019) : EPs, Lin(CD3, CD4, CD5, CD8, B220, Gr1, and Mac1)$^-$ CD71$^+$Ter119$^{+/-}$; GMs, Lin(CD3, CD4, CD5, CD8, B220, and Ter119)$^-$ Gr1$^+$Mac1$^+$ ("GM" cells were positive for both Gr1 and Mac1); T cells, Lin(CD5, B220, Gr1, Mac1, and Ter119)$^-$ CD25$^-$ CD3$^+$CD4$^{+/-}$CD8$^{+/-}$; B cells, Lin(CD3, CD4, CD8, Gr1, Mac1, and Ter119)$^-$CD43$^-$B220$^+$.

ATAC-seq

ATAC-seq was performed as previously described (Buenrostro et al., 2013). Briefly, cells were collected after sorting into microcentrifuge tubes containing staining media (1xDPBS,1mM EDTA with 5% serum). They were centrifuged at 500xg for 5 minutes at 4˚C to pellet the cells. The supernatant was aspirated, and the cells were washed with ice-cold 1xDPBS. Cells were centrifuged and the supernatant was discarded. Cells were immediately resuspended in ice-cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2 and 0.1% IGEPAL CA-630) and centrifuged at 500xg for 10 minutes. The supernatant was aspirated, and pellets were resuspended in transposase reaction mix (25µL 2xTD Buffer, 2.5µL transposase (Illumina), and 22.5µL nuclease free water). The transposition reaction was carried out at 37˚C for 30 minutes at 600rpm in a shaking thermomixer (Eppendorf). Immediately after completion of the transposition reaction, the samples were purified using the MinElute Reaction Clean up kit (Qiagen) and eluted into 10 µL of EB. Samples were stored at -20˚C until PCR amplification step. PCR amplification was performed as previously described (Buenrostro et al., 2013) using custom Nextera primers. After initial amplification, a portion of the samples were run on qPCR (ViiA7 Applied Biosystems) to determine the additional number of cycles needed for each library. The libraries were purified using the MinElute Reaction Clean up kit (Qiagen), eluted into 20 µL EB and then size selected using AmpureXP(Beckman-Coulter) beads at a ratio of 1.8:1 beads/sample, and eluted into 40µL of nuclease-free water. Library size distribution was determined by Bioanalyzer (Agilent) capillary electrophoresis and library concentration was determined by Qubit 3 (Life Technologies). Quality of libraries were checked by shallow sequencing (1 million raw reads) on a Miseq (Illumina) at 75 x 75

paired-end sequencing. Those libraries that appeared to have size distributions similar to previous reports (**Figure 2.2**) were pooled together and deep sequenced on a HiSeq2500 (Illumina) at 100 x 100 reads at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

Data processing

Demultiplexed sequencing data was processed using the ENCODE ATAC-seq pipeline version 1.1.6 and 1.4.2 (https://github.com/ENCODE-DCC/atac-seq-pipeline) using the mm10 assembly and the default parameters. In version 1.4.2 changed: atac.multimapping=0, atac.smooth_win=150, atac.enable_idr=true, atac.idr_thresh=0.1 to be consistent with the mapping/peak calling performed with previous versions.

Peak filtering, hierarchical clustering, and tSNE plot production was performed using the chromVAR package (https://github.com/GreenleafLab/chromVAR). First, the optimal peak-list from the IDR output for each cell type was concatenated and sorted, then used as the peak input for chromVAR. The blacklist filtered bam files for reach replicate was used as input along with the sorted peak file. The fragment counts in each peak for each replicate and GC bias was calculated, and then the peaks were filtered using filterPeaks function with the default parameters and nonoverlapping=TRUE. The master peak-list was extracted at this point, which contained 84,243 peaks, and used throughout the study. The deviations were calculated using every peak, and the tSNE and correlation functions were also performed using the deviations output and the default parameters.

Annotation of peaks, generation of histogram plot, merging of peaks, and motif enrichment was performed by HOMER (http://homer.ucsd.edu/homer/). Peaks were annotated using the annotatePeaks.pl function with the mm10 assembly and default parameters. Histogram was created by first shifting the bam files using DeepTools alignmentSieve.py with the flag –ATACshift. Next, tag directories were made using the Tn5 shifted bam files using HOMER makeTagDirectory. The histogram was made using the annotatePeaks.pl function with the default settings and the flags: -size -500,500 and -hist 5. Peak lists were compared using the mergePeaks.pl function with default settings and the flags -d given, -venn, and for the unique peak lists -prefix. Motif enrichment was performed using the findMotifsGenome.pl package with default parameters using the flag -size given and custom background peaks, which consisted of the combination of all the peaklists for the cell types not being analyzed. Instances of motifs in non-promoter peaks were found by using the annotatePeaks.pl function with the -m flag, using custom made motif files for each cell type containing the top 10 enriched motifs found.

The GREAT tool (http://great.stanford.edu/public/html/) was used to annotate non-promoter peaks to target genes. The peak lists were reduced to BED4 files from the HOMER annotations output and used as input. The whole mm10 genome was used as the background regions, and the association rule settings were set as Basal plus extension, proximal window 2kb upstream, 1kb downstream, plus distal up to 1Mb and included curated regulatory domains. All genome track visualizations were made using the UCSC genome browser. Graphs were made in either Microsoft Excel or GraphPad

Prism 8. Annotations to figures was performed using Adobe Illustrator CC and Adobe Photoshop CC.

ChIP data was handled as follows: The enhancer list from (Lara-Astiaso et al., 2014) was mapped to mm10 using the liftOver tool, then compared to the master peak-list. The raw sequencing data for H3K4me1 and H3K27Ac in LT-HSCs was downloaded from GEO and mapping to mm10 and peak calling were performed using the parameters listed in the publication (Lara-Astiaso et al., 2014).

Chapter 3: Dynamics of Chromatin Accessibility during Hematopoietic Stem Cell Differentiation into Progressively Lineage-Committed Progeny

Eric W. Martin, Alessandra Rodriguez y Baena, Roman E. Reggiardo, Jana Krietsch, Daniel H. Kim, E. Camilla Forsberg

Abstract

Epigenetic mechanisms regulate the multilineage differentiation capacity of hematopoietic stem cells (HSCs) into a variety of blood and immune cells. Mapping the chromatin dynamics of functionally defined cell populations will shed mechanistic insight on two major, unanswered questions in stem cell biology: how does epigenetic identity contribute to a cell type's lineage potential, and how do cascades of chromatin remodeling dictate ensuing fate decisions? Our recent work revealed evidence of multilineage gene priming in HSCs, where open *cis*-regulatory elements (CREs) exclusively shared between HSCs and unipotent lineage cells were enriched for DNA binding motifs of known lineage-specific transcription factors. Oligopotent progenitor populations operating between the HSCs and unipotent cells play essential roles in effecting hematopoietic homeostasis. To test the hypothesis that HSC-primed lineage-specific CREs remain accessible throughout differentiation, we used ATAC-seq to map the temporal dynamics of chromatin remodeling during progenitor differentiation. We observed epigenetic-driven clustering of oligopotent and unipotent progenitors into distinct erythromyeloid and lymphoid branches, with multipotent HSCs and MPPs associating with the erythromyeloid lineage. We mapped the dynamics of lineage-primed CREs throughout hematopoiesis and identified both unique and shared CREs

as potential lineage reinforcement mechanisms at fate branch points. Additionally, quantification of genome-wide peak count and size revealed overall greater chromatin accessibility in HSCs, allowing us to identify HSC-unique peaks as putative regulators of self-renewal and multilineage potential. These findings provide insight into the regulation of stem cell multipotency and lineage commitment throughout hematopoiesis and serve as a resource to test functional drivers of hematopoietic lineage fate.

Highlights:

- HSCs displayed higher chromatin accessibility than any progeny population
- Epigenetic branchpoints were evident between CMPs and CLPs
- Lineage priming was selectively maintained throughout differentiation
- HSC-unique peaks were highly enriched for regulatory elements of erythrocyte differentiation

INTRODUCTION

Hematopoiesis is the process by which multipotent hematopoietic stem cells (HSCs) undergo orchestrated epigenetic and transcriptional changes to produce increasingly lineage-restricted progenitors. According to classical models of hematopoiesis, progressively restricting cell fate decisions allows the differentiation of HSCs into multipotent progenitors (MPPs), which further differentiate into common lymphoid progenitors (CLPs) and common myeloid progenitors (CMPs) (Boyer et al., 2011; Orkin and Zon, 2008; Pronk et al., 2007). Lymphopoiesis further results in unipotent progenitors, ProB and ProT, of B and T cells respectively. In myelopoiesis, granulocyte-macrophage progenitors (GMPs) generate primarily mature granulocytes and macrophages (GMs), while megakaryocytic-erythroid progenitors (MEPs), megakaryocyte progenitors (MkPs) and erythroid progenitors (EPs) produce primarily platelets and red cells (Krause, 2002; Orkin and Zon, 2008; Seita and Weissman, 2010; Boyer et al., 2019). This well characterized mammalian hematopoietic system is a superb model for the analysis of factors responsible for the development of functionally distinct progenitors and mature cell populations from stem cells. Lineage-specific cell fate decisions are regulated through epigenetic remodeling of cis-regulatory elements (CREs), including promoters and enhancer regions. While proximal promoter sequences can suffice to assemble the Pol II transcriptional machinery, non-promoter CREs are often necessary to confer cell type-specific transcriptional regulation. These enhancer regions can be located far upstream or downstream of the target promoter and serve as sequence-specific binding sites for lineage-determining transcription factors (TFs) that regulate the expression of genes specifying cell identity (Surani et al., 2007; Whyte et al., 2013). While TFs are important

contributors to cellular lineage specification and progressive lineage restriction, accessibility of enhancers to TFs is fundamental for spatiotemporal gene regulation during stem cell differentiation (Creyghton et al., 2010; Heintzman et al., 2009; Koch et al., 2007; Rada-Iglesias et al., 2011; Visel et al., 2009).

In hematopoietic progenitors, there is evidence that multilineage priming of CREs precedes commitment to the different cell lineages (Hu et al., 1997). "Priming", here defined as ATAC-accessibility of a putative CRE despite lack of expression of its presumed target gene, likely contributes to stem and progenitor lineage potential. As differentiation of HSCs proceeds, genes involved in the target lineage are progressively upregulated in progenitor populations while genes involved in non-target lineages are repressed (Chambers et al., 2007; Forsberg et al., 2005; Phillips et al., 2000; Terskikh et al., 2003), suggesting an essential role of epigenetic regulation in cell fate decisions. In our previous work, we showed evidence of multilineage priming in HSCs, where HSCs had increased global chromatin accessibility compared to progeny (Ugarte et al., 2015) and where open CREs exclusively shared between HSCs and unipotent lineage cells were enriched for DNA binding motifs for known lineage-specific TFs (Martin et al., 2021). These data led us to hypothesize that HSC-primed lineage-specific CREs remain accessible throughout differentiation into that specific lineage. Since CREs are often devoid of nucleosomes to allow TF binding (Gross and Garrard, 1988; Heintzman et al., 2007), we performed the Assay for Transposase Accessible Chromatin by high throughput sequencing (ATAC-seq) (Buenrostro et al., 2015, 2013) of seven, functionally well-characterized hematopoietic progenitor cell types (Boyer et al., 2019; Poscablo et al., 2021) to understand CRE

priming across hematopoiesis. Importantly, cell fate decisions, as well as lineage-selective expansion and apoptosis, appears to occur primarily in progenitor cell populations (Rodriguez-Fraticelli et al., 2018; Mohrin et al., 2010; Boyer et al., 2011; 2019). In this study, in-depth ATAC-seq investigation and comparative analysis of HSCs and 12 progeny populations of the five main hematopoietic cell lineages reveal potential multipotency, lineage-driving and/or lineage-reinforcing regulatory elements and their corresponding transcription factors that orchestrate differentiation through epigenetic remodeling.

RESULTS

HSCs had greater global chromatin accessibility compared to hematopoietic progenitor cell types

To determine the dynamics of genome accessibility of multipotent and increasingly lineage-restricted hematopoietic progenitors, we purified 7 primary hematopoietic progenitor cell types **(Figure 3.1A)** by fluorescent-activated cell sorting (FACS) and performed ATAC-seq. After careful quality control of individual and replicate samples (also see below), we tested the hypothesis that multipotency is correlated with overall chromatin openness (Gaspar-Maia et al., 2011; Martin et al., 2021; Ugarte et al., 2015); we reasoned that multipotent cell populations would have the highest level of accessibility relative to oligopotent cells, and that unipotent progenitors would have the least. Thus, we ranked the relative overall accessibility of the hematopoietic progenitors relative to HSCs from our previous report (Martin et al., 2021). We first combined the peak lists from each replicate (n=2) using the Irreproducible Discovery Rate (IDR) (Li et al., 2011) for each cell type to quantify the number of peaks. HSCs had the highest number of peaks, followed by MPPs **(Figure 3.1B, Table 3.1)**. We also quantified global accessibility by calculating the cumulative normalized average signal over the master peak-list for each cell type by generating histograms using HOMER (Heinz et al., 2010). HSCs had by far the largest peak signal of any progenitor cell type, while all the progenitors had a similar average signal **(Figure 3.1C).** Although these two measurements are not completely independent, HSCs displayed both the highest number of peaks **(Figure 3.1B)** and the cumulative greatest peak signal **(Figure 3.1C)**. Overall, these results are consistent with epigenetic stem cell priming

48

and our previous reports (Martin et al., 2021; Ugarte et al., 2015) where HSCs have the greatest chromatin accessibility compared to their progeny and differentiated cells.

Chromatin accessibility of cell type-specific genes correlated with known expression patterns in hematopoietic cells

We began the search for lineage-specific regulatory elements by using the Gene Expression Commons (GEXC) expression database (Seita et al., 2012) to generate lists of genes that were expressed specifically in each progenitor cell type (examples shown in **Figure 3.1D**). In parallel, we filtered the ATAC-seq peak lists of each progenitor cell type (HSC, MPP, CMP, GMP, MEP, CLP, ProB, ProT) against each other to generate unique peak-lists for each cell type. We then intersected the unique peak lists with the uniquely expressed genes for each progenitor. For populations that had more than ten unique promoter peaks (HSCs, MEPs, ProBs, ProTs) we used HOMER (Heinz et al., 2010) to calculate the normalized average signal centered at the promoter for peaks that overlapped with expressed genes (**Figure 3.1E**). We observed cell type-specific read-count accumulation for each progenitor cell with minimal signal from other cell types, indicating that our strategy indeed resolved lineage-specific accessibility.

Table 3.1: Peak counts and peak distribution relative to protein-coding gene promoters in each cell type.

Cell Type	ATAC peaks	Promoter Peaks (±500bp of TSS)	Sum of all non-promoter peaks	non-promoter peaks		
				coding (exons+TTS+TSS)	Introns	Intergenic
Master Peak-list	92,842	12,702	80,140	5,543	38,090	36,507
HSC	70,731	27,973	42,758	4,166	18,931	19,661
MPP	63,349	23,415	39,934	3,685	17,872	18,377
CLP	47,054	21,467	25,587	2,126	11,414	12,047
CMP	46,431	21,648	24,783	1,951	10,701	12,131
GMP	42,447	20,823	21,624	1,939	9,317	10,368
GM	30,529	15,559	14,970	1,440	6,697	6,833
MEP	50,483	26,064	24,419	2,281	10,803	11,335
EP	38,007	23,243	14,764	2,014	7,040	5,710
MkP	47,363	23,998	23,365	2,013	10,036	11,316
ProB	46,790	24,837	21,953	2,003	9,355	10,595
B	70,358	24,596	45,762	4,461	21,210	20,091
ProT	61,141	27,073	34,068	2,796	14,950	16,322
T	51,832	25,103	26,729	2,016	11,929	12,784

Figure 3.1: ATAC-seq analysis of hematopoietic progenitor cell populations revealed progressive and lineage-specific chromatin condensation.
A) Schematic diagram of the hematopoietic cells analyzed in this study. Thirteen cell populations were investigated: multipotent HSCs (Hematopoietic Stem Cells) and MPPs (Multipotent Progenitors); lineage-restricted/oligopotent CMPs (Common Myeloid Progenitors), CLPs (Common Lymphoid Progenitors), GMPs (Granulocyte Macrophage Progenitors), MEPs (Megakaryocyte Erythrocyte Progenitors); unilineage MkPs (Megakaryocyte Progenitors), EPs (Erythroid Progenitors), ProBs (B cell Progenitors), ProTs (T cell Progenitors), and mature GMs (Granulocyte/Macrophages), B cells, and T cells. ATACseq profiles for HSCs and unilineage MkPs, EPs, GMs, B and T cells were reported previously (Martin et al., 2021); data were integrated in selective analyses of the new data for intermediate progenitors for a comprehensive perspective of hematopoiesis. **B**) HSCs had the highest number of peaks of all hematopoietic progenitor cell types. The total number of Irreproducible Discovery Rate (IDR) peaks per cell type are displayed. HSCs had the highest number of peaks, followed by MPPs and then lineage-committed progenitors. **C**) HSCs had the highest average signal across all peaks. Average cumulative signal across the peak-list for each population was determined by the -hist function of HOMER annotatePeaks.pl. Multipotent HSCs and MPPs had the highest average peak signal, whereas lineage-restricted progenitors had overall lower signal. **D**) Lineage-specific gene expression patterns used to find examples of genes selectively expressed within each indicated cell type. The level of expression (red=high; blue=low/not expressed) was obtained from the Gene Expression Commons (GEXC) database. **E**) Promoter accessibility correlated with cell type-specific gene expression in the corresponding progenitor cell types. Plots depict HOMER histograms of the average cumulative signal across the cell type-specific promoters for HSCs (34 peaks), MEPs (16 peaks), ProBs (29 peaks), and ProTs (12 peaks). MPPs, CMPs, GMPs, and CLPs were not displayed as each of these populations had fewer than 10 promoter peaks of uniquely expressed genes.

Figure 3.1
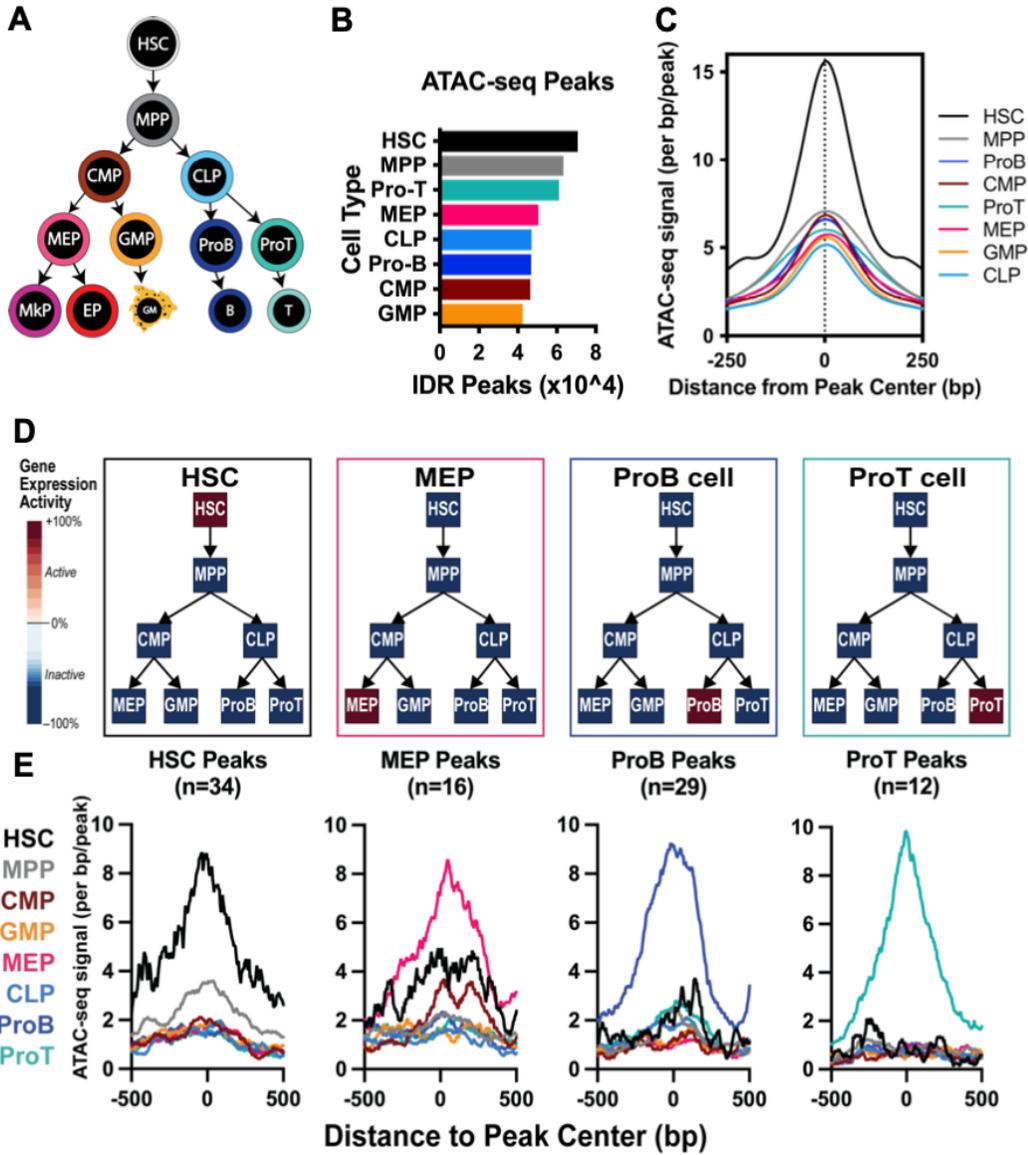
Lymphoid commitment displayed more extensive chromatin remodeling compared to myelopoiesis

Next, we sought to pinpoint epigenetic changes at a main branchpoint in hematopoiesis, where the multipotent stem and progenitor cells differentiate into either erythromyeloid- or lymphoid- committed CMPs or CLPs (Seita and Weissman, 2010). We compared the peaks gained and lost between multipotent HSCs and MPPs (combined as "KLS" peaks) and CMPs or CLPs (**Figure 3.2A**). First, we determined the number of peaks either CMPs or CLPs gained or lost from KLS. The CMP and CLP peaks were filtered against each other to focus only on peaks that were uniquely altered in either cell type. At a global level, CLPs had a larger number of peaks gained and lost from KLS cells compared to CMPs (**Figure 3.2B)**. When categorizing peaks into promoter vs non-promoter, we observed significantly more promoter peaks altered in CLPs than CMPs (**Figure 3.2C**), whereas similar numbers of non-promoter peaks were altered in both progenitors (**Figure 3.2D**). We annotated the peaks that were gained and lost using Genomic Regions Enrichment of Annotations Tool (GREAT) and reported the top 4 biological process Gene Ontology (GO) terms enriched, along with example genes in each GO term. We also performed motif enrichment by HOMER (**Figure 3.2E-H**). The peaks gained in CMPs included "Negative Regulation of B-cell Activation", and the annotated genes of all four GO terms have known roles in myeloid differentiation, such as *Prdm1* (Chang et al., 2000) and *Btk* (Schmidt et al., 2004). Gata1/2 motifs were among the highest enriched sequences (**Figure 3.2E**). The CLP peaks gained were enriched for GO terms that pertained mainly to immune response and immunity, with genes *Ikzf1, Il6,* and *Jun* present within the top 4 GO terms, and were notably enriched with IRF8 and Spi-B motifs (**Figure 2F**). Peaks lost from KLS

to CMPs were related to immune system activation and proliferation (**Figure 2G**), with known immune development genes such as *CD180, Ikzf1, and Gata3.* In addition, there were enriched motifs from ETS/ERG transcription factors as well as SpiB, a known factor in immune development (**Figure 2G**). In CLPs, peaks lost from KLS were related to immune system and activation, with example genes having known roles in erythropoiesis (*Hlx* and *Tfrc*) or HSCs self-renewal and maintenance, such as *CD44* (**Figure 2H**). The CLP-lost peaks were enriched in erythromyeloid specific Gata factors (Gata1, 2, and 6) as well as CTCF motifs (**Figure 2H**). These analyses suggest that at the first branchpoint, both myeloid and lymphoid differentiation require a combination of silencing of self-renewal and alternative lineage genes, and *de novo* activation of lineage drivers for the induced fate. Quantitatively, lymphoid differentiation appears to require more chromatin remodeling than myeloid differentiation, particularly in promoter regions.

Figure 3.2: Comparisons of peak dynamics as multipotent HSCs and MPPs differentiate into CMPs or CLPs revealed quantitatively differential gain and loss of accessibility. A) Schematic of the comparisons made between multipotent HSCs and MPPs (ckit+Lin-Sca1+; KLS) to lymphoid- or erythromyeloid-committed CLPs or CMPs. First, the peaks from HSCs and MPPs were combined using bedtools merge and then compared to CLPs or CMPs. The altered peak lists from the CMP and CLP comparisons were then intersected against each other to generate CMP- or CLP-specific peaks that were either gained or lost from KLS. **B-D)** CLPs had more peak alterations than CMPs. The number of peaks gained and lost in each cell type are displayed. Compared to CMPs, CLPs had more total number of peaks gained/lost (**B**), promoter peaks altered (**C**), and similar numbers of non-promoter peaks altered (**D**). The distribution of peaks between CMPs and CLPs was significant by Chi-square for the total number of peaks (**B**) (*** $p < 0.001$) and promoter peaks (**C**) (**** $p < 0.0001$; and not significant for non-promoter peaks (**D**) ($p = 0.42$). **E-H)** Cis-regulatory element analysis, GO term enrichment, and motif enrichment of the peaks that were altered between KLS and CLPs or CMPs, along with example target genes from each GO term. Briefly, each list of altered peaks was submitted to GREAT using the basal extension function with a parameter of 2kb upstream, 1kb downstream, and up to 1Mb extension. Example genes were extracted from the region-target association table for each GO term. The top 5 enriched known motifs from HOMER and corresponding transcription factors were also reported. **E)** GREAT analysis of CMP-gained peaks contained the GO term "Negative Regulation of B cell Activation", and were enriched for motifs of Gata transcription factors. **F)** Peaks gained by CLPs were primarily enriched in immune cell activation GO terms, with "Leukocyte Activation Involved in Immune Response" as the top hit. Peaks were enriched for motifs of ETS factor ETS1, as well as known lymphoid drivers IRF8 and SpiB. **G)** CMP peaks that were lost from KLS cells all relate to immune cell processes, and were enriched with motifs for ETS factors and SpiB, similar to the peaks gained by CLPs. **H)** CLP peaks lost from KLS contained GO terms that were immune related, such as "Regulation of Leukocyte Mediated Immunity" with Gata2 and Tlr4 as example genes. The peaks were enriched for Gata and CTCF/CTCFL transcription factor motifs. # the full title of this GO term is "Regulation of Adaptive Immune Response Based On Somatic Recombination of Immune Receptors Built from Immunoglobulin Superfamily Domains".
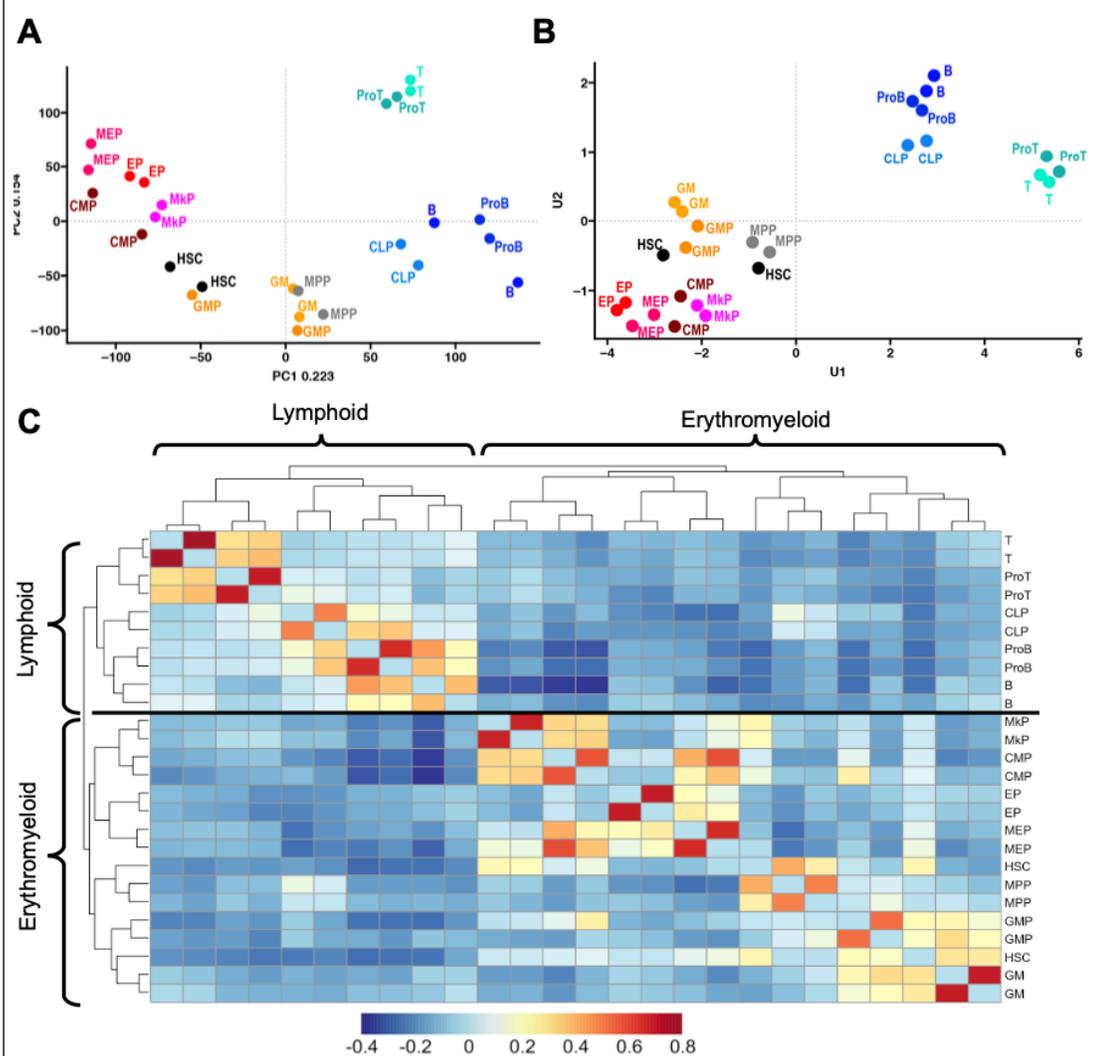
Figure 3.2
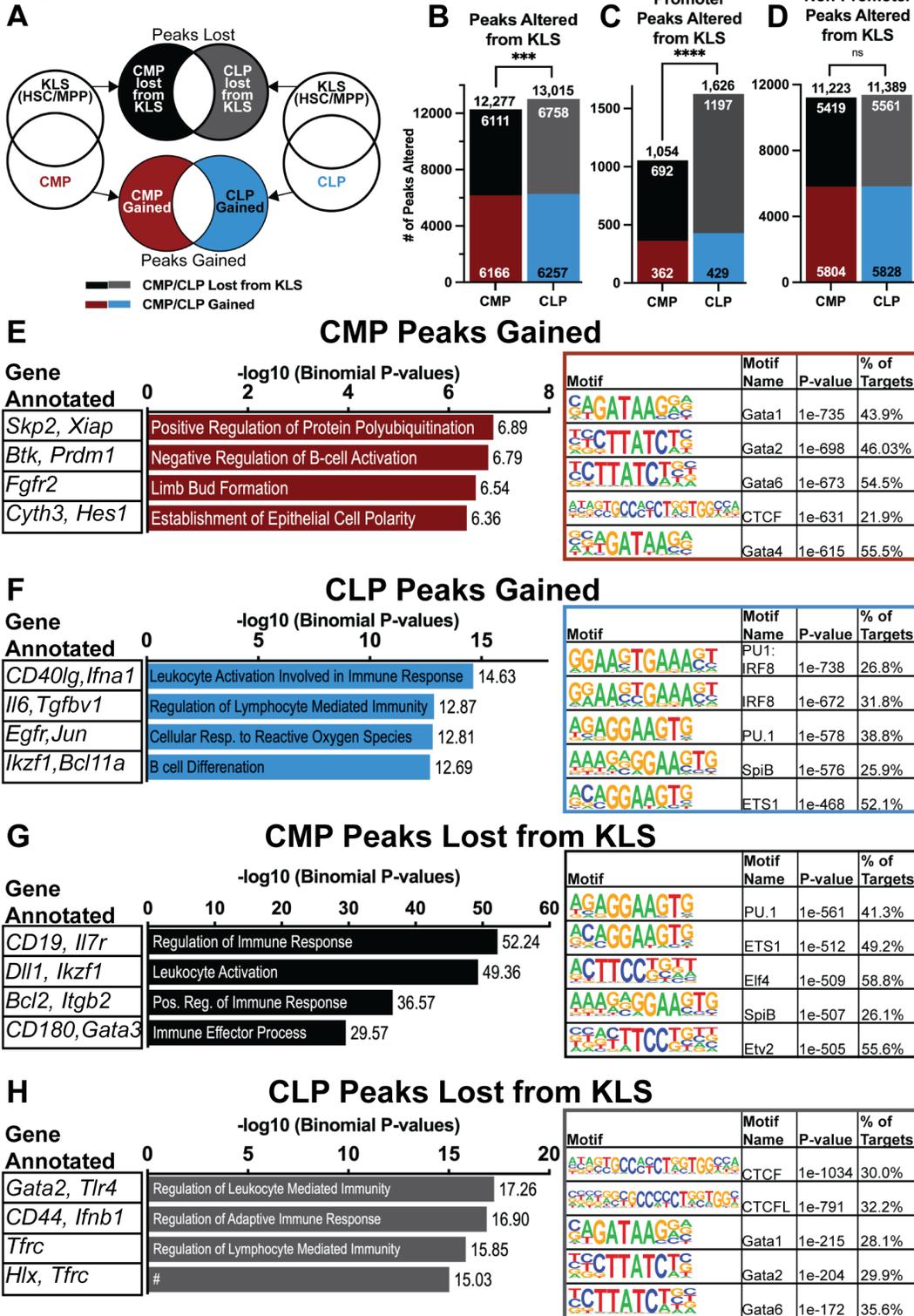
Mapping of chromatin accessibility throughout hematopoiesis identified distinct erythromyeloid and lymphoid clusters

To test our hypothesis that CREs primed in HSCs maintained accessibility throughout hematopoiesis, we needed to determine the dynamics of genome accessibility and further characterize lineage selective CREs throughout the whole continuum of hematopoiesis. To do so, we combined the ATAC-seq data from the 7 progenitors cell types with our previously reported HSCs and 5 unilineage cell types (Martin et al., 2021) **(Figure 3.1A**). A master peak-list of 92,842 peaks was produced by combining and filtering the peaks from 2 biological replicates for each of the 13 cell types using chromVAR (Martin et al., 2021; Schep et al., 2017b) **(Table 3.1)**. Principal Component Analysis (PCA) of the peak profiles of our 13 populations revealed a high concordance of replicates, as well as a distinct bifurcation of erythromyeloid and lymphoid populations, with the multipotent HSCs and MPPs landing within the erythromyeloid fraction **(Figure 3.3A)**. CMPs, MEPs, EPs, and MkPs all clustered together high on PC2, while CLPs, ProBs, and B cells clustered together, with ProTs and T cells grouped on the same PC1 scale but with higher PC2. HSCs and MPPs, together with GMPs and GMs, fell between the main myeloid and lymphoid groups. As a complement to PCA analysis, we performed Uniform Manifold Approximation and Projection (UMAP) using components derived from PCA of normalized ATAC-seq peak counts **(Figure 3.3B)**. We observed a similar bifurcation between erythromyeloid and lymphoid cell types with the multipotent HSCs and MPPs falling within the erythromyeloid quadrant. Additionally, we performed hierarchical clustering using the chromVAR output which similarly grouped the 13 populations into two distinct clusters, one erythromyeloid and one lymphoid **(Figure 3.3C)**. All biological replicates clustered

directly next to each other, except for the two HSC samples which were separated by MPPs and GMPs. We ruled out batch-effects as closely associated samples were processed independently; the separation of the HSC replicates may instead reflect the presence of primed CREs of all lineages (Martin et al., 2021; also see below). In all three clustering analyses, the multipotent HSCs and MPPs associated near each other and within the erythromyeloid cluster, indicating a similar accessibility profile of these cell types. Overall, clustering analysis confirmed a high degree of reproducibility. Regardless of the method used, we observed distinct clustering based on similar accessibility profiles of lymphoid cell types, erythromyeloid cell types, multipotent HSCs and MPPs, and of unipotent/mature cells with their presumed immediate upstream progenitor. The bifurcation of lymphoid and erythromyeloid lineages observed in the PCA **(Figure 3.3A)**, UMAP **(Figure 3.3B)** and hierarchical clustering **(Figure 3.3C)** is consistent with models of classical hematopoiesis. Of note, the observed similarity between HSCs/MPPs and erythromyeloid cells provides a potential epigenetic basis for the previously reported erythroid functional bias, where HSCs and MPPs predominantly produce red blood cells over all other cell types (Boyer et al., 2019).

Figure 3.3: ATAC-seq maps of hematopoietic cell populations revealed distinct erythromyeloid and lymphoid clusters. A) Principal Component Analysis (PCA) of chromVAR-normalized ATAC-seq peak counts revealed high concordance of replicates, and distinct erythromyeloid and lymphoid quadrants. Percent of total variance explained by each component are displayed on respective axes. **B**) Uniform Manifold Approximation and Projection (UMAP) using components derived from PCA generated distinct erythromyeloid and lymphoid clusters with the multipotent HSCs and MPPs associated with the erythromyeloid quadrant, similar to the PCA. **C**) Hierarchical clustering of all 13 cell types revealed high concordance of replicates and distinct clusters consistent with classical models of hematopoiesis (Figure 1A). Two primary associations were revealed: one erythromyeloid cluster and one lymphoid cluster. Multipotent HSCs and MPPs were designated to the erythromyeloid cluster. Additionally, there were four distinct sub-clusters: MkPs with CMPs; MEPs with EPs; ProBs with B cells and CLPs; and ProTs with T cells.

Visualization and comparison of ATAC-seq data generated in this study correlated with known expression patterns at two well characterized loci.

To determine whether our mapping could detect known CREs, we visualized our ATAC-seq data across two well-characterized loci: the mouse β-globin cluster **(Figure 3.4A)** and the mouse *Rag* locus (**Figure 3.4B/C**). At the β-globin cluster (chr7: 103,792,027-103,879,340; mm10), we observed expected EP-selective accessibility of the HS3 site in the locus control region (LCR) and *β-minor* promoter (Li et al., 2002; Palstra et al., 2008) (**Figure 3.4A**). We observed erythroid-lineage specific accessibility (HSCs, MEPs, and EPs) of the *β–major (ßmaj*) promoter as well as DNase I hypersensitive sites (HS1,2,4,6) of the LCR that are known to regulate erythroid-specific expression of the genes in this locus. This observation could indicate a "permissive" chromatin state in these erythroid-competent progenitor cells (HSCs, CMPs, MEPs, and EPs). Unexpectedly, we observed robust HS2 accessibility in GMPs, MkPs, and ProT cells, which are not currently known to have any erythroid cell potential. As expected, we did not observe any accessibility at the fetal-specific *epsilon Y globin (Ey), β-h1 (ßh1), β-h2 (ßh2)* genes, or HS5, and no accessibility was observed at any of these sites in GMs, CLPs, ProBs, and ProTs. Taken together, we observed expected accessibility in the *β–globin* locus in progenitors that give rise to cells that express *β–globin* genes, and little to no accessibility in progenitors that do not give rise to cells that express *β–globin* genes.

Similar specificity was observed for the *Rag* gene locus (chr2: 101,542,312-101,656,796; mm10) which consists of four CREs (Ep,D3,Erag,ASE) and the gene bodies for *Rag1* and *Rag2*. Both *Rag1* and *Rag2* have lymphoid-specific gene

expression patterns (**Figure 3.4B)**, and we observed lymphoid-specific accessibility of both *Rag1* and *Rag2* promoters (**Figure 3.4C**). The peak in the second intron of *Rag2*, on the opposite strand of the *Rag* genes*,* corresponds to the promoter for the *Iftap* gene (Laszkiewicz et al., 2012) (**Figure 3.4C**). The CREs Ep and Erag, which have been characterized to be enhancers in B cell lines (Hsu et al., 2003; Wei et al., 2002), exhibited CLP and B cell specific (ProBs and B cells) accessibility. D3 has been characterized to act as a lymphoid specific enhancer (Kuo and Schlissel, 2009; Wei et al., 2002) and was accessible in all lymphoid cell types, while the previously characterized anti-silencing element (ASE), important for T cell differentiation (Yannoutsos et al., 2004; Yu et al., 1999), was only accessible in ProTs (**Figure 3.4C**). In conclusion, our data demonstrated cell type-specific accessibility of unique peaks for multiple progenitors as well as lineage-specific accessibility at two well characterized erythroid and lymphoid loci, suggesting that our dataset should be sufficiently robust and accurate to also reveal novel CREs.

Figure 3.4: Accessibility correlated with known regulatory elements of well-characterized cell type-specific genes. A) Chromatin accessibility of the β-globin locus revealed expression-selective patterns at known *cis*-regulatory elements (CREs). ATAC-seq signal tracks at the β-globin cluster (chr7: 103,792,027-103,879,340; mm10) of the thirteen cell types are shown. Peaks highlighted by green boxes represent called peaks by Irreproducible Discovery Rate (IDR) at known CREs for each cell type. **B)** Lymphoid-selective expression of *Rag1* and *Rag2.* GEXC expression data reported expression of *Recombination activating gene 1 (Rag1*) and *Recombination activating gene 2 (Rag2*) in CLPs, ProBs, ProTs, B, and T cells. *Rag2* expression in non-lymphoid cell types (CMPs, GMPs, MkPs, and EPs) is due to the *Iftap* promoter on the opposite strand of the *Rag* genes in the second intron of *Rag2* (Laszkiewicz et al., 2012).**C)** Lymphoid-selective accessibility of the *Rag* locus. ATAC-seq signal tracks of the thirteen cell types in this study at the lymphoid-selective *Rag* gene locus (chr2: 101,542,312-101,656,796; mm10). The *Rag* gene locus consists of four previously characterized CREs (Ep, D3, Erag, ASE) and the gene bodies for *Rag1* and *Rag2*. The promoter for both *Rag1* and *Rag2* had accessibility only in lymphoid cell types (CLPs, ProBs, B cells, ProTs, and T cells). The lymphoid specific D3 CRE had expected lymphoid-only accessibility, and the B cell specific CREs Ep and Erag had accessibility only in CLPs, ProBs, and B cells. The T cell development specific anti-silencing element (ASE) only exhibited accessibility in ProT cells.

Figure 3.4

A LCR Beta-globin (chr7:103,792,027-103,879,340)

B

Rag1

Rag2

C Rag locus (chr2:101,542,312-101,656,796)
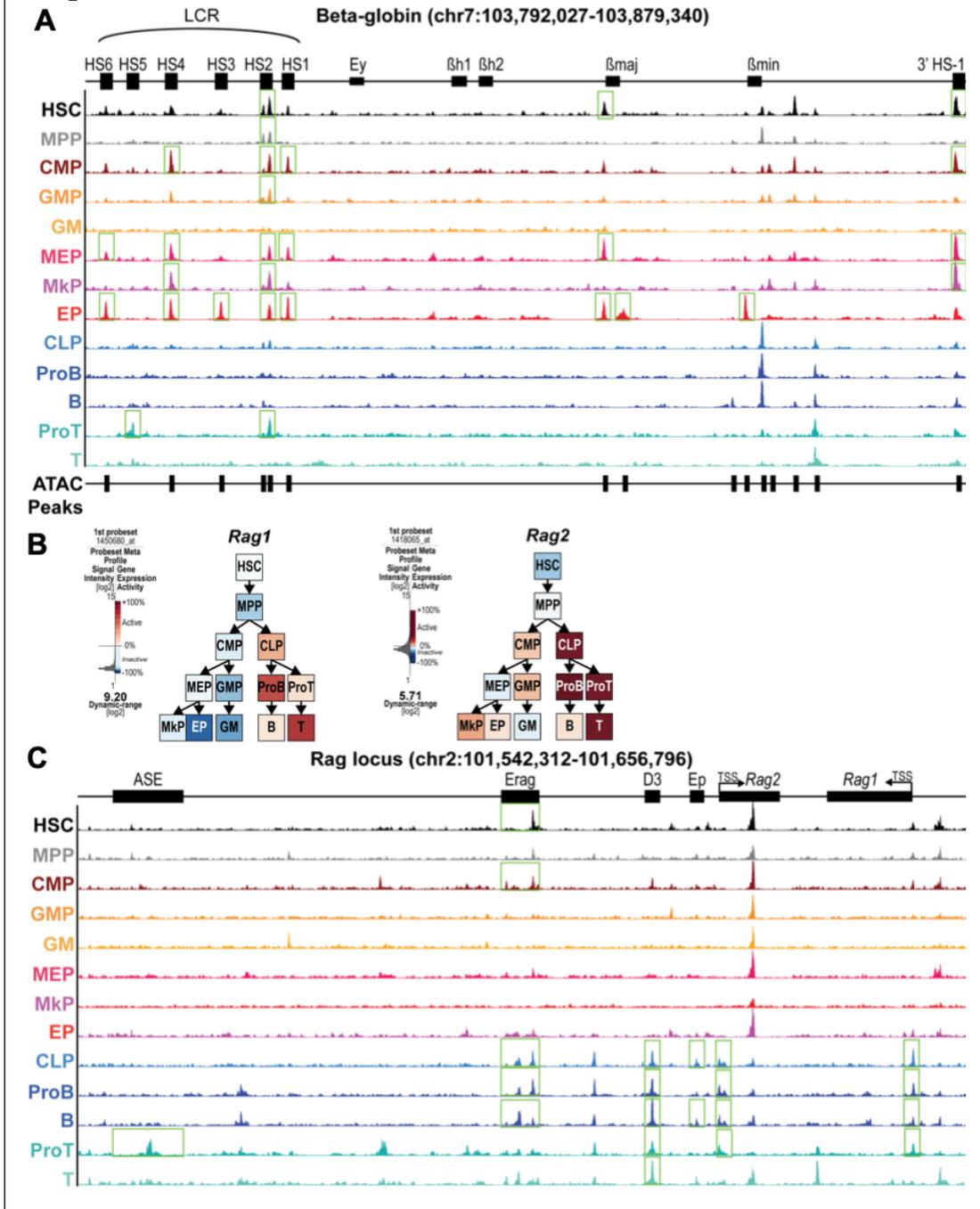
A subset of lineage-specific CREs were primed in HSCs as well as in select progenitors

Previously, we reported evidence of multilineage priming in HSCs of CREs specific for each unipotent lineage (Martin et al., 2021). We hypothesized that lineage-primed CREs are maintained throughout differentiation. To test this, we first compared the average cumulative accessibility of the lineage-specific peaks primed in HSCs to all 13 cell types (**Figure 3.5A**). As expected, we observed strong signals from HSCs and the corresponding unipotent progenitor cell type for each lineage-specific primed peak list. MPPs had a discernable peak in four out of the five primed peak lists, with a less distinct signal in EP-primed peaks. Notably, each unilineage region displayed accessibility signal in the presumed immediate upstream progenitor (MEPs in EP-primed peaks; GMPs in GM-primed peaks; ProB in B-cell peaks; and ProTs in T-primed peaks), except for MkPs, which lacked MEP signal and instead had notable accessibility in MPPs and CMPs. These observations revealed that lineage priming of a sizeable proportion of CREs persists throughout differentiation for every lineage.

To assess the distribution of the primed peaks in each progenitor population, we performed a *bedtools* intersect of the lineage-specific peaks primed in HSCs and determined the number of overlapping peaks with each progenitor. Interestingly, all progenitors from every lineage contained peaks from all 5 primed peak lists (**Figure 3.5B**). The distribution of primed peaks of all five lineages was about equally distributed at ~20% each in HSCs, with similar distribution in MPPs, and CLPs (Chi-square > 0.01). Clear lineage bias was evident in other populations: erythromyeloid progenitors (CMPs, MEPs) were significantly enriched for EP- and MkP-primed peaks,
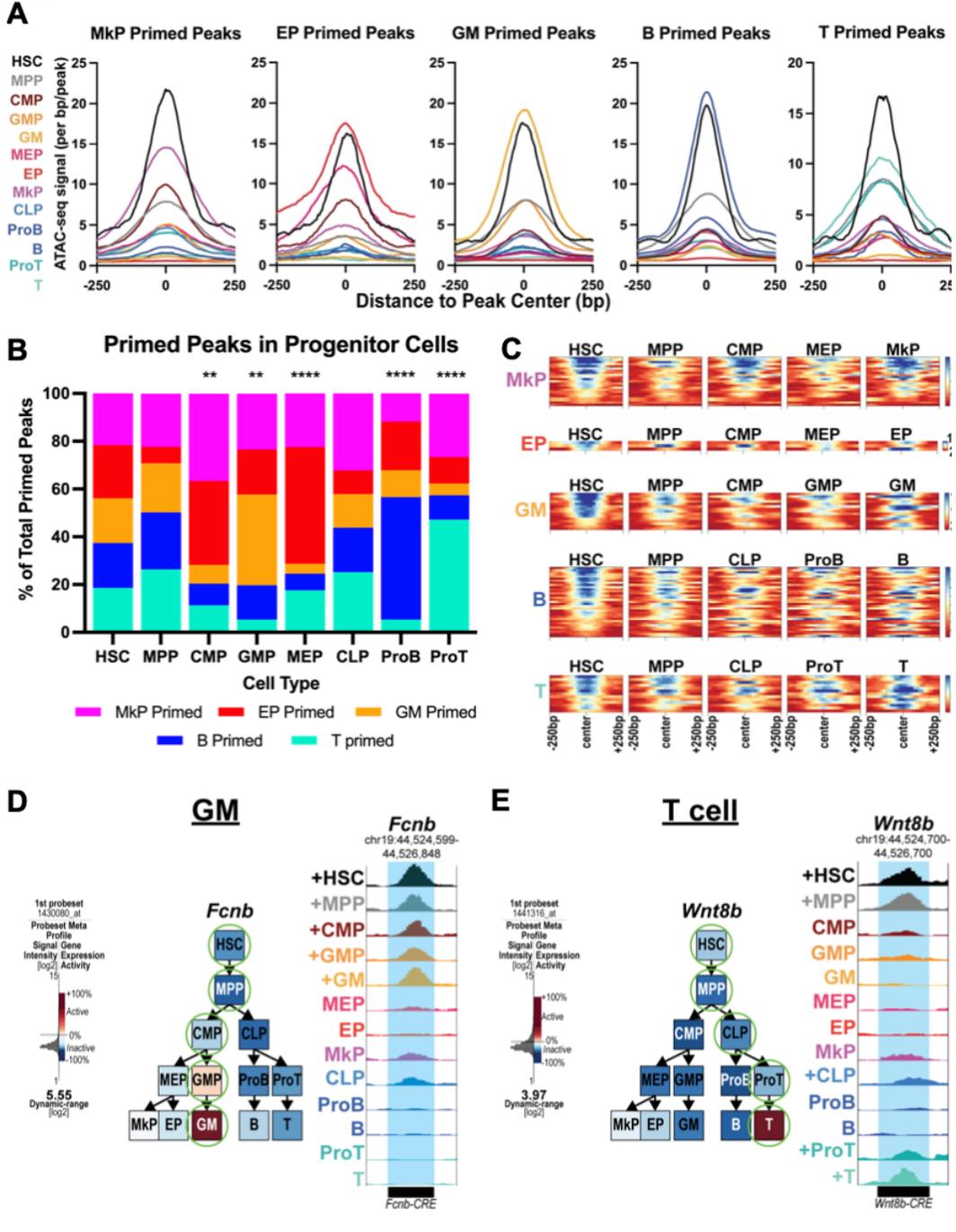
GMPs were enriched for GM-primed elements, while the unipotent lymphoid progenitors (ProBs and ProTs) were significantly enriched with peaks from their immediate downstream progeny (B and T cells).

To accomplish direct longitudinal analysis of priming through multiple differentiation stages, we intersected the HSC-primed peak lists with every assumed intermediate progenitor between HSCs and the unipotent lineage (i.e for HSC/EP shared peaks, we intersected MPPs, CMPs, and MEPs with the shared peak list, as those populations are in the HSC-to-EP lineage in Figure 1A). From those intersections, we identified and quantified the number of peaks that maintained accessibility throughout differentiation for each lineage. Surprisingly, even though perfectly primed CREs were detected for every lineage, this was far from the norm, as no lineage had more than 25% of the HSC-primed peaks maintained throughout differentiation (**Figure 3.5C**). About 10% of the persistently primed peaks were promoters, with the B cell lineage specific *BAFF-R* and T cell specific *CD28* as examples that have known functional roles in those cell types (Dodson et al., 2009; Shulga-Morskaya et al., 2004). Next, we examined two example CREs that were primed throughout differentiation for the GM (**Figure 3.5D)** and T cell (**Figure 3.5E**) lineages. The GM specific *Fcnb* gene is expressed only in GMPs and GMs, while the putative CRE associated to *Fcnb* was accessible in HSCs, MPPs, CMPs, GMPs, and GMs (**Figure 3.5D**). The T cell specific *Wnt8b* is only expressed in T cells, while the putative CRE is accessible in HSCs, MPPs, CLPs, ProTs, and T cells. (**Figure 3.5E**). These findings support that lineage priming observed in HSCs is maintained throughout differentiation for certain CREs. Unexpectedly, most of the peaks primed in HSCs did not exhibit persistent priming in

every intermediate progenitor. These observations, combined with the bias in signal and peak counts in progenitors could suggest preferred lineages at specific branchpoints. For example, EP-primed peaks had a high average signal and made up most of the overlapping peaks in MEPs which could suggest that MEPs are biased towards EPs over MkPs, or reinforce fate decisions initiated in upstream progenitors.

Figure 3.5: CREs of lineage-specific genes primed in HSCs also displayed accessibility in progenitors. A) Lineage-specific peaks primed in HSCs also displayed selective enrichment in intermediate progenitors. HOMER histograms of the average cumulative accessibility in each of the 13 cell types in each lineage-primed peak-list. MkP lineage peaks that were primed in HSCs were also enriched in MPPs and CMPs, but less so in GMPs, CLPs, ProB, and ProTs; EP peaks were selectively enriched in MEPs and CMPs; GM peaks were enriched primarily in MPPs and GMPs; B cell peaks were enriched in ProBs and MPPs, and T cell peaks were enriched in ProTs and MPPs. **B)** Peak distribution analysis revealed lineage skewing within progenitors. The distribution of lineage-primed peaks was displayed for each progenitor cell type. All progenitors contained lineage-primed peaks representing unique peaks of each of the five lineages, but at different proportions. HSCs had an almost equal distribution of peaks from all five lineages that did not deviate from an expected equal distribution (Chi-square, p = 0.97). MPPs and CLPs had similar peak distributions and were not significantly different when compared pairwise to HSCs (Chi-square, p ≥ 0.01). In contrast, pairwise comparison of the distribution of peaks between HSCs and progenitors revealed significant differences in CMPs, GMPs, MEPs, ProBs, and ProTs by Chi-square. CMPs had a relative expansion primarily of erythromyeloid (MkP, EP) peaks; GMPs had primarily GM-unique peaks; MEPs were enriched for EP-unique peaks; whereas ProBs had more B cell peaks, and ProTs had mainly T cell peaks. ** p < 0.01, **** p < 0.0001. **C)** Heatmaps of primed peaks that maintain accessibility throughout the expected differentiation trajectory for each lineage. Each line is one peak, with accessibility indicated in blue centered around the peak +/-250 bp. Less than 25% of the lineage specific primed peaks for each lineage followed the expected trajectory by maintaining accessibility throughout differentiation. **D)** A *cis* regulatory element (CRE) predicted by GREAT to be associated with *Fcnb* maintained accessibility ("priming") throughout differentiation into GMs. GEXC reported expression of *Fcnb* selectively in GMPs and GMs. Green circles indicate which cell type contained a called peak. Genome track snapshot of the *cis* regulatory element of *Fcnb* reported accessibility in HSCs, MPPs, CMPs, GMPs, and GMs. A "+" sign designated which cell type contained a called peak. **E)** A CRE predicted by GREAT to be associated with *Wnt8b* maintained accessibility throughout differentiation into T cells. GEXC reported expression of *Wnt8b* selectively in T cells only. Green circles indicate which cell type contained a called peak. Genome track snapshot of the *cis* regulatory element of *Wnt8b* reported accessibility in HSCs, MPPs, CLPs, ProTs, and T cells. A "+" sign designated which cell type contained a called peak.

Figure 3.5

A

MkP Primed Peaks | EP Primed Peaks | GM Primed Peaks | B Primed Peaks | T Primed Peaks

HSC, MPP, CMP, GMP, GM, MEP, EP, MkP, CLP, ProB, B, ProT, T

ATAC-seq signal (per bp/peak)

Distance to Peak Center (bp)

B Primed Peaks in Progenitor Cells

% of Total Primed Peaks

Cell Type: HSC, MPP, CMP, GMP, MEP, CLP, ProB, ProT

MkP Primed, EP Primed, GM Primed, B Primed, T primed

C

D GM

E T cell

Fcnb
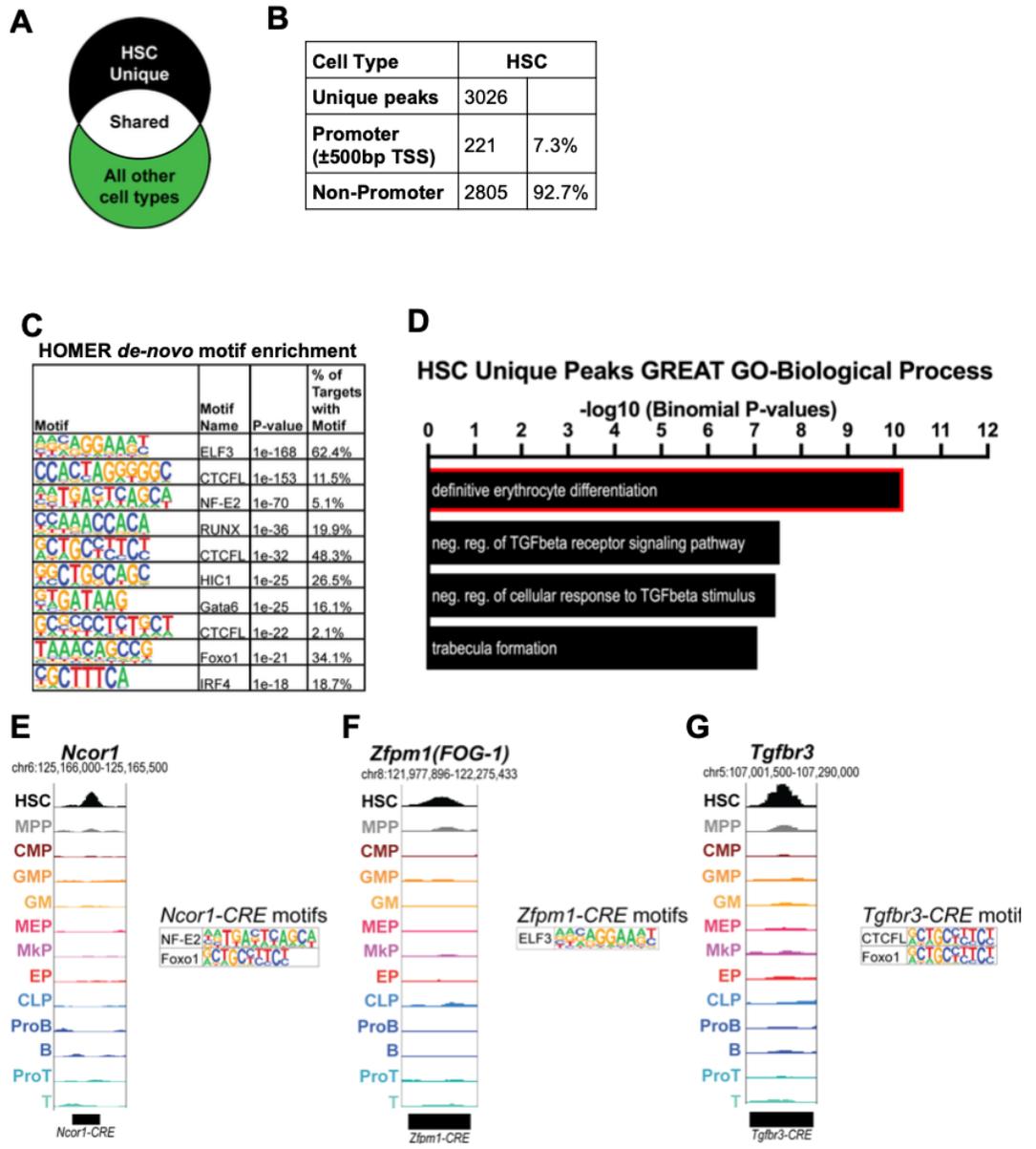chr19:44,524,599-44,526,848

Wnt8b
chr19:44,524,700-44,526,700

68

HSC-unique peaks indicated an erythropoiesis-primed chromatin state

Because HSCs are the only cell type in the hematopoietic tree that is capable of long-term reconstitution, we reasoned that HSC-unique peaks would be enriched in elements that promote self-renewal. To test this, we identified and examined HSC-unique peaks (**Figure 3.6A**). We found 3,026 HSC-unique peaks, 92.7% of which classified as non-promoter (**Figure 3.6B**). To determine what motifs were enriched within the HSC-unique peaks, we performed *de novo* motif finding and enrichment using the HOMER package and reported the top 10 results sorted by p-value (**Figure 3.6C**). ELF3 (E74 Like ETS Transcription Factor 3) was the top ranked motif, followed by CTCFL. There were 3 instances of CTCF-like motifs in the top 10 *de novo* motifs, while single instances of NF-E2, RUNX, HIC1, Gata6, Foxo1, and IRF4 rounded out the enriched motifs. Next, we annotated the CREs to nearby genes using GREAT. The top GO term was *definitive erythrocyte differentiation* (**Figure 3.6D**)*, comprised of 14 peaks linked to 4 genes: Ncor1, Tgfbr3, Zfpm1,* and *Smarca4.* All four genes have known roles in hematopoiesis, with knock-out studies presenting severe defects in erythropoiesis, or the entire hematopoietic compartment (Bultman et al., 2005; Chi et al., 2003; Jepsen et al., 2000; Stenvers et al., 2003; Tsang et al., 1998), consistent with important roles in HSCs. We then visualized the ATAC accessibility of three example peaks enriched in the *definitive erythrocyte differentiation* GO term, along with their respective linked motif enrichment (**Figure 3.6E-G**). The CRE linked to *Ncor1* contained the NF-E2 and Foxo1 motif (**Figure 3.6E**). The CRE linked to *Zfpm1* contained the motif for ELF3 (**Figure 3.6F**), while the CRE linked to *Tgfbr3* contained CTCFL and Foxo1 motifs (**Figure 3.6G**). Taken together, the unique HSC peaks are enriched for elements that prime erythroid cell fate in HSCs, such as NF-E2 binding

69

sites and the 14 peaks that are linked to *Ncor1, Tgfbr3, Zfpm1,* and *Smarca4, all of which have known roles in erythropoiesis.*

Figure 3.6: HSC-unique *cis* regulatory elements are primarily enriched for transcription factors that drive erythropoiesis. A) The HSC-unique peak-list was generated by filtering HSC peaks against the peak lists of the other 12 hematopoietic cell types. **B)** HSC-unique peaks are primarily non-promoter peaks. Table of the composition of the HSC-unique peaks and percentage of non-promoter and promoter peaks. **C)** *De novo motif* enrichment of HSC-unique peaks revealed binding sites for known hematopoietic regulators. ELF3, CTCFL, NF-E2, and Runx motifs were the top 5 enriched *de novo* motifs. **D)** "Definitive erythroid differentiation" was the top enriched GO term from GREAT annotation and analysis of the unique HSC peaks. The resulting graphs are GO Biological Process terms and the -log10 p-value for the top four terms. **E-G)** Three examples of putative CREs for target genes that were enriched in "definitive erythrocyte differentiation" and displayed unique HSC accessibility. **E)** A putative CRE for *Ncor1* was unique to HSCs and contained motifs that closely match NF-E2 and Foxo1 binding sites. **F)** A putative *Zfpm1* CRE contained the binding motif that closely matches ELF3. **G)** A putative *Tgfbr3* CRE contained DNA motifs that closely matched CTCFL and Foxo1 binding sites.

Figure 3.6

A

HSC Unique

Shared

All other cell types

B

Cell Type	HSC	
Unique peaks	3026	
Promoter (±500bp TSS)	221	7.3%
Non-Promoter	2805	92.7%

C

HOMER *de-novo* motif enrichment

Motif	Motif Name	P-value	% of Targets with Motif
	ELF3	1e-168	62.4%
	CTCFL	1e-153	11.5%
	NF-E2	1e-70	5.1%
	RUNX	1e-36	19.9%
	CTCFL	1e-32	48.3%
	HIC1	1e-25	26.5%
	Gata6	1e-25	16.1%
	CTCFL	1e-22	2.1%
	Foxo1	1e-21	34.1%
	IRF4	1e-18	18.7%

D

HSC Unique Peaks GREAT GO-Biological Process

-log10 (Binomial P-values)

definitive erythrocyte differentiation

neg. reg. of TGFbeta receptor signaling pathway

neg. reg. of cellular response to TGFbeta stimulus

trabecula formation

E *Ncor1*
chr6:125,166,000-125,165,500

F *Zfpm1(FOG-1)*
chr8:121,977,896-122,275,433

G *Tgfbr3*
chr5:107,001,500-107,290,000

Ncor1-CRE motifs

Zfpm1-CRE motifs

Tgfbr3-CRE motil

DISCUSSION

Global chromatin accessibility throughout hematopoiesis is highly dynamic

Here, we mapped accessible loci in seven hematopoietic progenitor cell types with distinct functional capacities. Integration of these new data with HSCs and mature progeny revealed epigenetic-based cell clustering into erythromyeloid and lymphoid branches (**Figure 3.3**) and robust identification of known regulatory elements (**Figure 3.4**). Consistent with previous evidence by us and others that stem cells have relatively decondensed chromatin structure (Lara-Astiaso et al., 2014; Ugarte et al., 2015; Wang et al., 2015)**,** we found that both the ATAC peak number and cumulative signal was greatest for HSCs (**Figure 3.1**). This study advances previous reports by pinpointing the location both of all putative CREs genome-wide in each population, as well as HSC-specific putative CREs and those associated with the major erythromyeloid/lymphoid branchpoint. These maps will serve as a valuable resource for functional interrogation of gene regulation, stem cell self-renewal, and fate decisions.

Lymphopoiesis requires more extensive chromatin remodeling compared to erythromyelopoiesis

Based on functional studies (Boyer et al., 2019; Rodriguez-Fraticelli et al., 2018), we hypothesized that differential epigenetic priming may be evident at major branchpoints. Indeed, we found that HSC/MPP-descendant CLPs had a significantly larger number of altered peaks compared to CMPs (**Figure 3.2**). Interestingly, this was mainly driven by promoter peak changes, possibly indicating that erythromyeloid priming, but not implementation of an erythromyeloid program, may remain present in

CLPs. From the GO term enrichment of peaks altered (**Figure 3.2E-H**), we found that erythromyeloid differentiation was accompanied by a loss of accessibility of lymphoid lineage drivers, as well as a gain of accessibility at negative regulators of lymphoid differentiation. In contrast, lymphoid differentiation was accompanied by a loss of accessibility of self-renewal genes and a gain of accessibility of lymphoid driver genes. These findings are consistent with our previous study, where we observed that mature B cells and T cells had a higher proportion of peaks gained compared to the erythromyeloid lineages (Martin et al., 2021). Similarly, a genome architecture study reported a similar magnitude of dynamic alterations during B-cell differentiation as upon somatic cell reprogramming into induced pluripotent stem cells (Vilarrasa-Blasi et al., 2021). Collectively, this suggests that HSCs are primarily primed for erythromyelopoiesis and require greater chromatin remodeling to initiate lymphopoiesis.

Lineage priming was selectively maintained throughout differentiation

Our previous study identified CREs that were exclusively shared between unipotent lineage cells and HSCs (Martin et al., 2021). Here, we found that a subset of those primed elements maintained accessibility throughout differentiation in intermediate progenitors (**Figure 3.5**). Lineage priming was also detected at the β-globin locus, where the strongest enhancer, HS2, was primed in HSCs and MPPs, with additional accessibility of HS1 and HS4 in CMPs, and then also HS3 and HS6 in MEPs and EPs **(Figure 3.4).** The global enrichment of peaks within the intermediate progenitors reflected the distinct bifurcation found in hematopoiesis, with erythromyeloid-primed peaks enriched in erythromyeloid progenitors and lymphoid-primed peaks enriched in

74

lymphoid progenitors. Surprisingly, CLPs did not significantly deviate from the distribution of peaks in HSCs and MPPs (**Figure 3.5B**), potentially indicating "inherited" priming that is not implemented *in vivo* (Schlenner et al., 2010) but can be reignited in vitro (Karsunky et al., 2003). We also tracked the accessibility of the primed peaks throughout differentiation and found that the majority of peaks do not maintain accessibility in every intermediate progenitor throughout differentiation (**Figure 3.5C,D**). Collectively, these findings provide insight into the dynamics of CRE accessibility throughout differentiation and supports a model where lineage priming in HSCs guides lineage competence during differentiation, while the gain and loss of accessibility at certain intermediate progenitors could regulate or reinforce differentiation in specific lineages.

HSC-unique peaks were highly enriched for CREs that drive erythroid differentiation

While HSCs are capable of producing all blood cell lineages, several studies have suggested lineage-specific priming within HSCs (Boyer et al., 2019; Carrelha et al., 2018; Ema et al., 2014; Yamamoto et al., 2013, 2018). From these studies we hypothesized that CREs within HSCs would uncover drivers of erythro- and/or megakaryopoiesis. Our GREAT analysis of HSC-unique peaks revealed "definitive erythrocyte differentiation" as the top GO-Biological Process hit (**Figure 3.6D**), and we observed HSC-specific accessibility in the CREs linked to genes that have known roles in erythropoiesis (**Figure 3.6E-G**). Furthermore, we observed *de novo* enrichment of transcription factor motifs in the HSC-unique peaks that are known to be key regulators of hematopoiesis, such as NF-E2 and Runx (Gasiorek et al., 2012; Shivdasani and

Orkin, 1995; Willcockson et al., 2019). This suggests that establishment of developmental competence for erythropoiesis in HSCs may occur primarily in CREs that are uniquely accessible in HSCs.

In summary, we present evidence that multilineage priming is present in HSCs and selectively maintained, or repressed, throughout differentiation. In addition, the observation that HSCs harbor the most ATAC-seq peaks of all hematopoietic cell types **(Figure 3.1)** is consistent with previous findings that linked multipotency with global epigenetic regulation and the presence of poised loci that are distal to promoters in stem cells (Lara-Astiaso et al., 2014; Ugarte et al., 2015; Wang et al., 2015). We also found that accessibility, especially of distal CREs, is highly dynamic. These results provide insight to how lineage fate is reinforced at branchpoints through the collective action of specific transcription factors at these CREs. Future investigation will determine which of these CREs are a consequence of differentiation and which elements drive differentiation into specific fates.

EXPERIMENTAL PROCEDURES

Mice and Cells

All experiments were performed using 8- to 12-week-old C57BL/6 wild-type mice in accordance with UCSC IACUC guidelines. Hematopoietic cells were isolated from BM of murine femurs, tibias, hips, and sternums. Stem and progenitor cell fractions were enriched using cKit-coupled magnetic beads (Miltenyi). Cells were stained with unconjugated lineage rat antibodies (CD3, CD4, CD5, CD8, B220, Gr1, Mac1, and Ter119) followed by goat-α-rat PE-Cy5 (Invitrogen). Stem and progenitor cells were

76

isolated using fluorescently labeled or biotinylated antibodies for the following antigens: cKit (2B8, Biolegend), Sca1 (D7, Biolegend), Slamf1(CD150) (TC15-12F12.2, Biolegend), CD34 (RAM34, ebiosciences), FcγrII (93, Biolegend), and Il7rα (A7R34, Biolegend),. Cells were sorted using a FACS Aria II (BD Bioscience). HSCs were defined as $cKit^+$ Lin^- $Sca1^+$ $Flk2^-$ and $Slamf1^+$; MPPs as $cKit^+Lin^-Sca1^+$ $Flk2^+$ $Slamf1^-$ cells. CMPs were defined as $cKit^+Lin^-Sca1^-$ $CD34^{mid}$ $FcγrII^{mid}$; GMPs as $cKit^+Lin^-Sca1^-$ $CD34^{mid}$ $FcγrII^{high}$, MEPs as $cKit^+Lin^-Sca1^-$ $CD34^{low}$ $FcγrII^{low}$. CLPs were isolated by lineage depleting BM cells through staining of unconjugated lineage rat antibodies (CD3, CD4, CD5, CD8, B220, Gr1, Mac1, and Ter119) followed by sheep-α-rat Dynabeads (Life Technologies) and separation via EasySep magnet (Stem Cell Technologies). CLPs were isolated by Lin^-Flk2^+ $Il7rα^+$ $cKit^{mid}$ $Sca1^{mid}$. Lineage restricted hematopoietic progenitor and mature cells were isolated by the following markers: EPs, Lin(CD3, CD4, CD5, CD8, B220, Gr1, and Mac1)$^-CD71^+Ter119^{+/-}$; GMs, Lin(CD3, CD4, CD5, CD8, B220, and Ter119)$^-Gr1^+Mac1^+$ ("GM" cells were positive for both Gr1 and Mac1); T-progenitors (ProT), Lin(CD5, B220, Gr1, Mac1, and Ter119)$^-$ $CD3^+CD25^+$; T cells, Lin(CD5, B220, Gr1, Mac1, and Ter119)$^-CD25^-CD3^+CD4^{+/-}CD8^{+/-}$; B-progenitors (ProB), Lin(CD3, CD4, CD8, Gr1, Mac1, and Ter119)$^-CD43^+B220^+$; B cells, Lin(CD3, CD4, CD8, Gr1, Mac1, and Ter119)$^-CD43^-B220^+$.

ATAC-seq

ATAC-seq was performed as previously described (Buenrostro et al., 2013). Briefly, cells were collected after sorting into microcentrifuge tubes. They were centrifuged at 500xg for 5 minutes at 4°C to pellet the cells. The supernatant was aspirated, and the cells were washed with ice-cold 1xDPBS. Cells were centrifuged and the supernatant

was discarded. Cells were immediately resuspended in ice-cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2 and 0.1% IGEPAL CA-630) and centrifuged at 500xg for 10 minutes. The supernatant was aspirated, and pellets were resuspended in transposase reaction mix (25µL 2xTD Buffer, 2.5µL transposase (Illumina), and 22.5µL nuclease free water). The transposition reaction was carried out at 37°C for 30 minutes at 600rpm in a shaking thermomixer (Eppendorf). Immediately after completion of the transposition reaction, the samples were purified using the MinElute Reaction Clean up kit (Qiagen) and eluted into 10 µL of EB. Samples were stored at -20°C until PCR amplification step. PCR amplification was performed as previously described (Buenrostro et al., 2013) using custom Nextera primers. After initial amplification, a portion of the samples were run on qPCR (ViiA7 Applied Biosystems) to determine the additional number of cycles needed for each library. The libraries were purified using the MinElute Reaction Clean up kit (Qiagen), eluted into 20 µL EB and then size selected using AmpureXP (Beckman-Coulter) beads at a ratio of 1.8:1 beads/sample, and eluted into 40µL of nuclease-free water. Library size distribution was determined by Bioanalyzer (Agilent) capillary electrophoresis and library concentration was determined by Qubit 3 (Life Technologies). Quality of libraries were checked by shallow sequencing (1 million raw reads) on a Miseq (Illumina) at 75 x 75 paired-end sequencing. Those libraries that appeared to have size distributions similar to previous reports were pooled together and deep sequenced on a HiSeq2500 (Illumina) at 100 x 100 reads at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

Data processing

Demultiplexed sequencing data was processed using the ENCODE ATAC-seq pipeline version 1.1.6 and 1.4.2 (https://github.com/ENCODE-DCC/atac-seq-pipeline) using the mm10 assembly and the default parameters. In version 1.4.2 changed: atac.multimapping=0, atac.smooth_win=150, atac.enable_idr=true, atac.idr_thresh=0.1 to be consistent with the mapping/peak calling performed with previous versions.

Peak filtering and hierarchical clustering was performed using the chromVAR package (https://github.com/GreenleafLab/chromVAR). First, the optimal peak-list from the IDR output for each cell type was concatenated and sorted, then used as the peak input for chromVAR. The blacklist filtered bam files for each replicate (n=2 for each cell type) was used as input along with the sorted peak file. The fragment counts in each peak for each replicate and GC bias was calculated, and then the peaks were filtered using filterPeaks function with the default parameters and nonoverlapping=TRUE. The master peak-list was extracted at this point, which contained 92,842 peaks, and used throughout the study. The deviations were calculated using every peak, and the tSNE and correlation functions were also performed using the deviations output and the default parameters.

Normalized chromVAR counts were log+1 scaled, centered, and filtered to peaks that had above-median coefficient of variance. These filtered counts were use in principal component analysis (PCA) with the R package *prcomp*. Following this, the resulting

79

components were used to calculate UMAP dimensions using the R package *uwot*. The component values were plotted using the R package *ggplot2*.

Annotation of peaks, generation of histogram plot, merging of peaks, and motif enrichment was performed by HOMER (http://homer.ucsd.edu/homer/). Peaks were annotated using the annotatePeaks.pl function with the mm10 assembly and default parameters. Histogram was created by first shifting the bam files using DeepTools alignmentSieve.py with the flag –ATACshift. Next, tag directories were made using the Tn5 shifted bam files using HOMER makeTagDirectory. The histogram was made using the annotatePeaks.pl function with the default settings and the flags: -size -500,500 and -hist 5. Peak lists were compared using the mergePeaks.pl function with default settings and the flags -d given, -venn, and for the unique peak lists -prefix. Motif enrichment was performed using the findMotifsGenome.pl package with default parameters using the flag -size given.

The GREAT tool (http://great.stanford.edu/public/html/) was used to annotate non-promoter peaks to target genes. The peak lists were reduced to BED4 files from the HOMER annotations output and used as input. The whole mm10 genome was used as the background regions, and the association rule settings were set as Basal plus extension, proximal window 2kb upstream, 1kb downstream, plus distal up to 1Mb and included curated regulatory domains. All genome track visualizations were made using the UCSC genome browser. Statistical analysis was performed using GraphPad Prism 9. Graphs were made in either Microsoft Excel or GraphPad Prism 9. Annotations to figures was performed using Adobe Illustrator CC and Adobe Photoshop CC.

Data availability

The datasets generated in the current study are available in the Gene Expression Omnibus (GEO), accession number GSE184851, reviewer token **mpinagaaxxoffyt**. Published datasets are available at GSE162949.

AUTHOR CONTRIBUTIONS

E.W.M., J.K., and E.C.F. designed the experiments. E.W.M. and J.K, isolated and sorted the primary cell types and performed ATAC-seq. E.W.M, A.R.y.B., and R.E.R. conducted data processing and analysis. E.W.M., A.R.y.B., and E.C.F. wrote the paper. E.W.M., A.R.y.B., R.E.R., J.K., D.H.K., and E.C.F. read and commented on the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interest.

Chapter 4: CRISPRi-mediated Functional Assessment of Putative *cis* Regulatory Elements on Hematopoietic Stem Cell Lineage Output.

Eric W. Martin, Alessandra Rodriguez y Baena, Atesh Worthington, Connor Mattingly, E. Camilla Forsberg

INTRODUCTION

In hematopoiesis, the regulation of self-renewal and differentiation is important to maintain homeostasis within an organism. Multiple studies have shown that hematopoietic stem cells (HSCs) are regulated intrinsically through dynamic chromatin structure and accessibility of *cis* regulatory elements (CREs) (Cullen et al., 2014; Han et al., 2019; Ludwig et al., 2019; Rodrigues et al., 2022). The gain of accessibility at CREs that allow transcription factor binding and subsequent initiation of gene expression programs that drive differentiation into a particular lineage is termed lineage priming. We and others have characterized CRE dynamics in hematopoiesis, identifying putative CREs that may regulate multipotency and lineage fate choice throughout hematopoiesis. (Heuston et al., 2018; Martin et al., 2021).

Despite the wealth of characterizations of CRE dynamics in hematopoiesis, there has been little work to determine if a CRE is truly functional in regulating multipotency and lineage production. Previous work to characterize a CD115 enhancer (Rojo et al., 2019) and a Gata1 enhancer (McDevitt et al., 1997) utilized knock-out strategies to delete CREs in a one-by-one manner. The method is very low throughput and technically challenging, especially when there are a vast number of sequences that

need to be interrogated. With the advances of CRISPR/cas and sequencing technologies, we can functionally interrogate multiple CREs in a single experiment by combining a high-throughput enhancer screening method with single-cell RNA-seq (Gasperini et al., 2019; Replogle et al., 2020). We tested the hypothesis that *silencing of lineage primed cis regulatory elements in HSCs will alter mature lineage cell production.* To test this hypothesis, we optimized and performed CRISPRi mediated knockdown of the enhancer linked to CD115 in HSCs (Rojo et al., 2019) and performed a CFU-S transplantation assay to determine the lineage output capability of the manipulated HSCs. We observed lineage specific cell type production defects when silencing the promoter and enhancer without off-target effects to other lineages. From these observations we can functionally interrogate CREs and determine those that are indispensable for normal hematopoiesis.
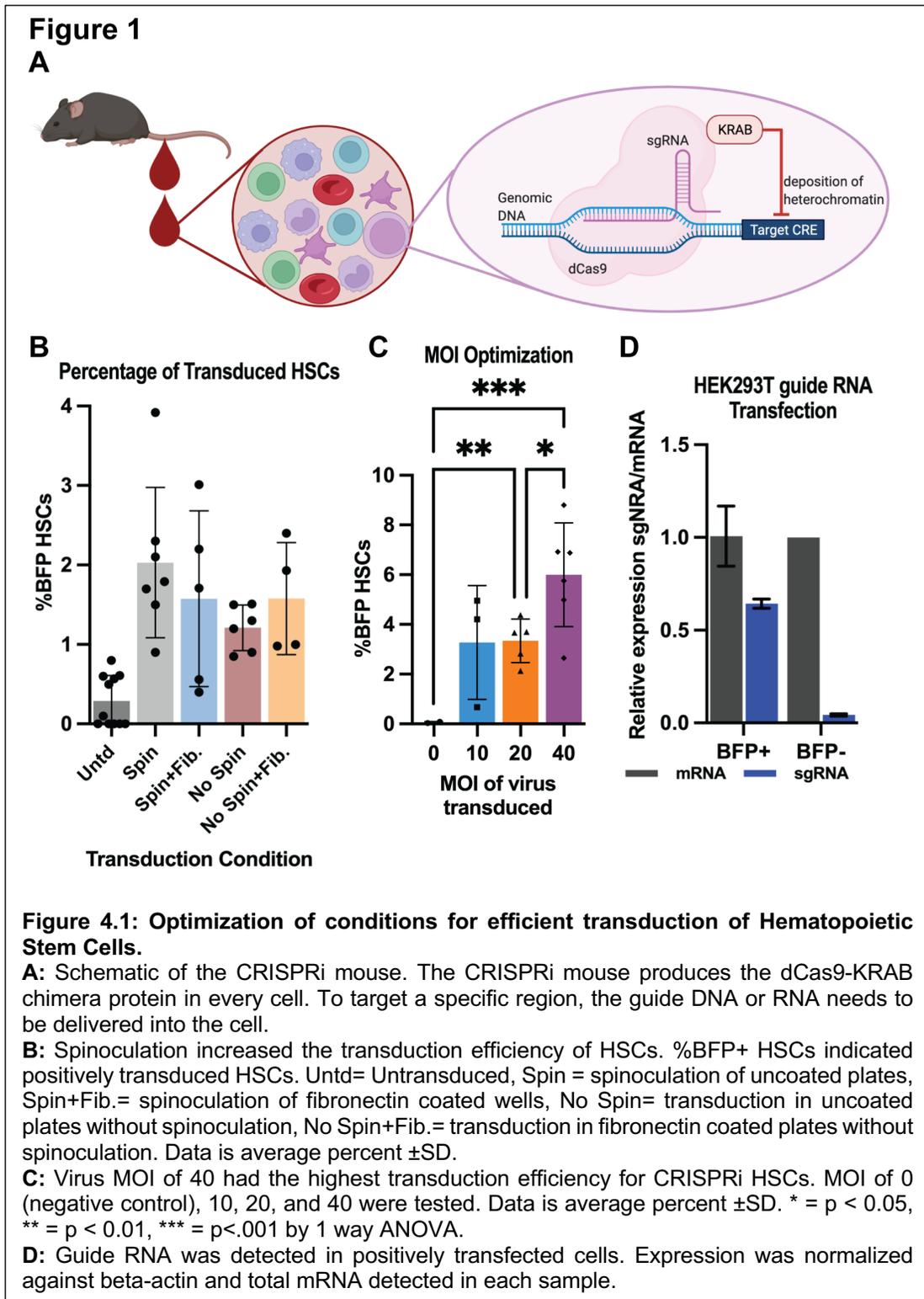
RESULTS

Optimization of transduction parameters for efficient lentiviral transduction of CRISPRi HSCs.

To be able to functionally interrogate CREs, we needed to optimize the transduction conditions for HSCs to get the highest number of transduced cells to transplant. We utilized CRISPRi mice that express the dCas9-KRAB fusion protein in every cell type (**Figure 4.1A**). This system requires the delivery of the guide RNA into the cells, so we utilized the direct capture guide RNA plasmid (Replogle et al., 2020) that is delivered by lentiviral transduction. First, we tested if fibronection and/or spinoculation increased transduction efficiency of HSCs. We used the same MOI of lentivirus across multiple conditions and assayed transduction efficiency 7 days after infection. We

observed trends of increased transduction with spinoculation and no observed difference with or without fibronectin (**Figure 4.1B**). Previous work has shown that fibronectin increased engraftment potential of HSCs that are cultured *ex vivo* (Wilkinson et al., 2019), so we concluded that spinoculation with fibronectin is the optimal condition to transduce HSCs.

Next, we determined the optimal MOI of the virus to use. We used the optimized culture conditions and transduced HSCs with a negative control virus and a guide targeting the CCR5 promoter. We observed that a MOI of 40 had highest percentage of transduced HSCs for both guides (**Figure 4.1C**), leading us to conclude that a MOI of 40 is optimal for our experiments.

In the future, we plan to use this system to perform a high-throughput screen of regulatory elements in hematopoiesis (Gasperini et al., 2020). We utilized the direct-capture guide system (Replogle et al., 2020) and wanted to validate that we could detect and identify each sgRNA in our system. We transfected HEK293T cells with a negative control guide and performed qRT-PCR using primers to detect the sgRNA capture sequence as well as cellular mRNA. We observed a clear enrichment of the sgRNA in transfected (BFP+) HEK293T cells compared to uninfected (BFP-) cells (**Figure 4.1D**), demonstrating that we can successfully detect sgRNAs and the expression of their intended gene target in our system. Taken altogether, we have established conditions for efficient transduction of HSCs and can detect the presence of guide RNA and their gene targets in lentivirally transduced cells.

Figure 4.1: Optimization of conditions for efficient transduction of Hematopoietic Stem Cells.
A: Schematic of the CRISPRi mouse. The CRISPRi mouse produces the dCas9-KRAB chimera protein in every cell. To target a specific region, the guide DNA or RNA needs to be delivered into the cell.
B: Spinoculation increased the transduction efficiency of HSCs. %BFP+ HSCs indicated positively transduced HSCs. Untd= Untransduced, Spin = spinoculation of uncoated plates, Spin+Fib.= spinoculation of fibronectin coated wells, No Spin= transduction in uncoated plates without spinoculation, No Spin+Fib.= transduction in fibronectin coated plates without spinoculation. Data is average percent ±SD.
C: Virus MOI of 40 had the highest transduction efficiency for CRISPRi HSCs. MOI of 0 (negative control), 10, 20, and 40 were tested. Data is average percent ±SD. * = p < 0.05, ** = p < 0.01, *** = p<.001 by 1 way ANOVA.
D: Guide RNA was detected in positively transfected cells. Expression was normalized against beta-actin and total mRNA detected in each sample.

Silencing of CD115 enhancer reduces GM cell numbers in reconstitution assays.

CD115 is a known regulator of macrophage development (Dai et al., 2002; Sherr et al., 1985). From previous knockout studies targeting the gene (Dai et al., 2002) and the enhancer (Rojo et al., 2019), we hypothesized that silencing the CD115 enhancer will result in reduced production of GM cells in a colony forming unit spleen (CFU-S) assay. We transduced HSCs with guides that targeted the CD115 promoter, the annotated CD115 enhancer, and a scrambled negative control guide RNA. We transplanted transduced HSCs into lethally irradiated recipient mice and isolated splenic colonies 13.5 days after transplantation. We observed a similar number of colonies per mouse (4-5) and a similar distribution of mature cell types from colonies from the promoter, enhancer, and negative control guides (**Figure 4.2A**). However, our preliminary results reported a significant reduction in the number of GM cells produced in both the enhancer and promoter knockdowns compared to the negative control (**Figure 4.2B**). This knockdown was associated with a significant reduction in surface receptor expression of CD115 in GM cells compared to negative controls (**Figure 4.2C**). From these observations we concluded that knockdown of CD115 promoter as well as the associated CRE can be functionally observed in hematopoiesis.

Figure 2

A

CD115 CFU-S Experiments

B

CD115 CRISPRi CFU-S

C

CD115 MFI of BFP+ GMs
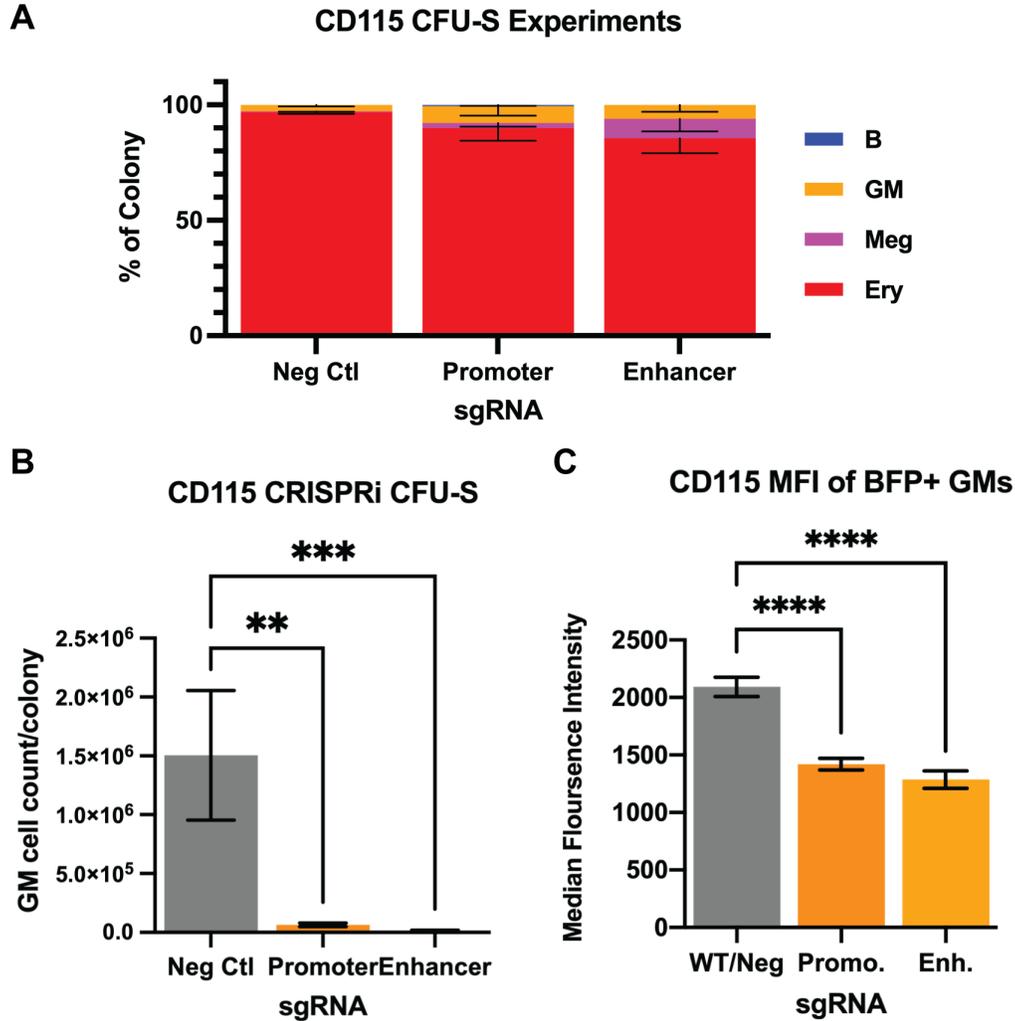
Figure 4.2: Knockdown of CD115 Enhancer perturbs GM cell production.
A: CD115 knockdown does not change the distribution of mature cells produced. Percent of colony distribution is average ±SEM of donor-derived B, GM, Megakaryocyte (Meg), and Erythroid (Ery) cells. Results not significant by 2-way ANOVA.
B: CD115 enhancer and promoter knockdown leads to loss of GM cell production. Cell count of donor-derived GM cells in negative control, CD115 promoter, and CD115 enhancer knockdown by CRISPRi. ** = p < 0.01, *** = p<.001 by 1 way ANOVA and Dunn's multiple comparison test.
C: Knockdown of CD115 enhancer results in reduction of surface expression of CD115. Median Fluorescence Intensity of CD115 on donor-derived GM cells in wild-type (WT)/negative control, CD115 Promoter (Promo), and CD115 Enhancer knockdown by CRISPRi. **** = p<.0001 by 1 way ANOVA and Dunnett's multiple comparison test.

DISCUSSION

We sought to use CRISPRi to functionally interrogate the CD115 CRE and measure changes in lineage output. We optimized the transduction of CRISPRi HSCs with lentivirus containing guide DNA along with a direct capture seed sequence and performed knockdown of the CD115 promoter and linked enhancer using CRISPRi. We performed a CFU-S assay to functionally test CD115 and its enhancer. Consistent with previous genetic deletion of the CRE, we observed similar defects in CD115 surface expression and reduced GM differentiation when targeting the CD115 CRE (Rojo et al., 2019). This study improves upon previous work interrogating CRE function in hematopoiesis (Li et al., 2020) by determining if there is a phenotypic outcome of CRE silencing instead of only reporting changes in gene expression of the target gene. This pilot study demonstrates that we can link CREs to a functional readout important to hematopoiesis. In addition, the ability to detect the sgRNA using direct capture techniques allows us to combine both CRISPRi knockdown with direct guide capture (Replogle et al., 2020) to perform the crisprQTL mapping technique (Gasperini et al., 2019) to functionally test our characterized CREs in hematopoiesis at unprecedented high throughput.

EXPERIMENTAL PROCEDURES

Mice and Cells

All experiments were performed using 8 to 12-week-old C57BL/6 wild-type or CRISPRi knock-in mice accordance with UCSC IACUC guidelines. Hematopoietic cells were isolated from BM of murine femurs, tibias, and hips. Stem and progenitor cell fractions were enriched using cKit-coupled magnetic beads (Miltenyi). Cells were stained with

unconjugated lineage rat antibodies (CD3, CD4, CD5, CD8, B220, Gr1, Mac1, and Ter119) followed by goat-α-rat PE-Cy5 (Invitrogen). Hematopoietic stem cells were isolated using fluorescently labeled or biotinylated antibodies for the following antigens: cKit (2B8, Biolegend), Sca1 (D7, Biolegend), and Slamf1(CD150) (TC15-12F12.2, Biolegend). Cells were sorted using a FACS Aria II (BD Bioscience). HSCs were defined as cKit$^+$ Lin$^-$ Sca1$^+$ Flk2$^-$ and Slamf1$^+$.

In-vitro culture and lentiviral transduction.

Sorted HSCs were plated in fibronectin-coated plates in HSC defined media from (Wilkinson et al., 2019). Briefly, flat bottom 96-well plates were incubated in a minimal volume of PBS with 5 µg/ml fibronectin (Sigma) for 1 hour. PBS was aspirated and HSCs were added with 200 µL of complete media per well and incubated at 37˚C, 95% humidity, 5% CO_2. 24 hours after initial plating, lentivirus was added and allowed to incubate for 24 hours before a complete media change. If spinoculation was performed, cells were spun at 400 xg for 1 hour at 32 ˚C in a pre-warmed centrifuge. Media changes were performed every 2-3 days. On day 7 after transduction, cells were harvested by aspiration and rinsing wells with PBS, and stained with antibodies described previously for HSCs.

CFU-S analysis

Lethally irradiated (1,000 rads) wild-type mice were transplanted with 200-300 HSCs that were purity sorted on phenotypic markers and positively transduced with lentivirus (BFP+). On day 13.5 post-transplantation, mice were sacrificed and individual splenic colonies were removed with a scalpel. Single-cell suspensions of dissected colonies

were stained with antibodies for the following antigens: Ter119, CD41, CD71, Gr1, Mac1, and B220. Cell types were defined as: EPs, CD41$^-$Ter119$^+$Gr1$^-$Mac-1$^-$B220$^-$CD71$^+$; GMs, CD41$^-$Ter119$^-$Gr1$^+$Mac-1$^+$B220$^-$("GM" cells were positive for both Gr1 and Mac1); B cells, CD41$^-$Ter119$^-$Gr1$^-$Mac-1$^-$B220$^+$; Megakaryocytes (Meg), CD41$^+$Ter119$^-$Gr1$^-$Mac-1$^-$B220$^-$.

Data processing

Flow cytometry data was analyzed by FlowJo 10. Statistical analysis was performed using GraphPad Prism 9. Graphs were made in either Microsoft Excel or GraphPad Prism 9. Annotations to figures was performed using Adobe Illustrator CC and Adobe Photoshop CC. Model figures were generated using bio render.

AUTHOR CONTRIBUTIONS

E.W.M., A.R.y.B., A.W. and E.C.F. designed the experiments. E.W.M., A.R.y.B., A.W. and C.M. performed the experiments. E.W.M., A.R.y.B., and A.W. conducted data processing and analysis. E.W.M. wrote the thesis chapter.

Chapter 5: Conclusions and Future Directions

During my PhD training, I investigated chromatin accessibility dynamics during hematopoiesis. In collaboration with other lab members, I performed ATAC-seq (Buenrostro et al., 2013) on 13 hematopoietic cell types (**Figure 1.1**) based on previously identified cell-surface markers. We identified candidate CREs that could be important for the differentiation of the five mature lineages assayed (Platelets, Red Blood Cells, Granulocytes/Macrophages, B cells, and T cells) and characterized their accessibility dynamics throughout differentiation. Finally, we established a CRISPRi-mediated knockdown technique to characterize the function of identified CREs. These findings provide insight into the epigenetic regulation of lineage fate choice in hematopoiesis.

Chapter 2 compared the genome-wide accessibility of the multipotent HSCs and unipotent lineage cell types (EPs, MkPs, GMs, B, and T cells). We observed that MkPs were most similar to HSCs through PCA and hierarchical clustering, with the lowest percentage of peaks lost from HSCs, and the greatest percentage of peaks exclusively shared with HSCs. These observations are consistent with clonal studies of hematopoiesis that reported a megakaryocyte lineage bias of HSCs (Carrelha et al., 2018; Rodriguez-Fraticelli et al., 2018). We observed evidence of multilineage priming in HSCs from the presence of exclusively shared CREs for all five lineages assayed. The exclusively shared CREs were all enriched for known lineage-specific transcription factors. About one-third were enriched for H3K4me1 histone modification, which has been linked to a primed enhancer state (Calo and Wysocka, 2013). Finally,

from the observations made throughout this study, we concluded that both permissive and de novo epigenetic mechanisms influence hematopoiesis. Evidence for permissive mechanisms included: HSCs with the highest global accessibility, most unipotent lineage cells had a higher percentage of peaks lost than gained, and the evidence of multilineage priming in HSCs. For de novo mechanisms, we observed peaks gained for all lineages compared to HSCs, consistent with previous studies that reported progressive up-regulation of lineage-specific genes during differentiation from HSCs to progenitors (Forsberg et al., 2005; Terskikh et al., 2003). Therefore, both permissive and de novo mechanisms influence hematopoietic fate decisions, and the balance between the two models likely act as another layer of regulation during hematopoietic differentiation.

The analysis and observations made in this study (Chapter 2) focused on comparisons between the top and bottom of hematopoiesis: the stem cell and five unipotent progenitors. Next, we further studied the dynamics of priming as differentiation progressed and characterized the accessibility of the exclusively primed peaks throughout the entire continuum of hematopoiesis. Chapter 3 mapped accessible loci in seven hematopoietic progenitor cell types with distinct functional capacities and compared accessibility with HSCs and the unipotent lineage cell types analyzed in chapter 2. From our ATAC-seq data, we were able to epigenetically recapitulate the classic model of hematopoiesis (**Figure 1.1**) through PCA and hierarchical clustering; and observed the multipotent HSCs and MPPs associated with the erythromyeloid cell types. Consistent with chapter 2, HSCs had the highest cumulative signal and the greatest number of peaks. Next, we characterized the first branchpoint of

hematopoiesis, the MPP to CMP/CLP transition. Consistent with our findings in chapter 2, we observed that lymphopoiesis required more extensive chromatin remodeling than erythromyelopoiesis. In addition, we observed a loss of lymphoid lineage drivers and a gain of negative regulators of lymphoid lineage differentiation during CMP differentiation. In contrast, CLP differentiation was accompanied by a loss of self-renewal elements and a gain of lymphoid lineage drivers. These findings suggest that HSCs are primarily primed for erythromyeloid differentiation, and lymphoid differentiation requires greater chromatin remodeling. Next, we tracked the dynamics of our exclusively primed peaks from chapter 2 throughout differentiation by tracking their accessibility in intermediate progenitors. We observed a subset of primed peaks that maintained accessibility throughout differentiation; however, most peaks for each lineage followed unique patterns of loss and gain of accessibility throughout differentiation, suggesting a model where lineage priming in HSCs guides lineage competence during differentiation. In contrast, the gain and loss of accessibility at intermediate progenitors could regulate or reinforce fate choice at specific branchpoints. Finally, we analyzed HSC-unique peaks and discovered that they were enriched for CREs that drive erythroid differentiation, suggesting a model where the developmental competence for erythropoiesis in HSCs may originate in CREs that are uniquely accessible in HSCs. In summary, we characterized the accessibility dynamics of the exclusively primed CREs throughout hematopoiesis which provides insight into the epigenetic regulation of lineage fate choice and reinforcement.

In Chapter 4, we established a protocol and optimized conditions for efficient transduction of HSCs to perform functional characterization of CREs in hematopoiesis.

In addition, we performed a proof-of-concept experiment knocking down CD115 expression by targeting the promoter and the annotated enhancer (Rojo et al., 2019). We observed a decrease in cell-surface expression of CD115 and a decrease in GM cell output when targeting the enhancer with CRISPRi machinery. These experiments establish our ability to perform CRISPRi mediated knockdown of CREs and detect functional changes to differentiation. We also demonstrated the ability to detect specific sgRNAs using direct capture sequences, which will allow us to perform a high-throughput screen using multiple sgRNA and detect their presence in cells while assaying their transcriptome via single-cell RNA-seq/sgRNA capture assays.

Future directions are focused on further characterization, and validation of the CREs identified in chapter 2. With the establishment of CRISPRi-mediated knockdown of CREs in the lab, we can now thoroughly interrogate our candidate lists of CREs for all five lineages through a screening approach by combining the crisprQTL mapping technique (Gasperini et al., 2019) and single-cell transcriptome with direct capture of guide sequences (Replogle et al., 2020). These screens will establish functional roles (or lack thereof) for many CREs in hematopoiesis. In addition, determining the histone modifications present on each candidate CREs by more sensitive and efficient chromatin immunoprecipitation methods like Cut 'N Run or Cut & Tag (Kaya-Okur et al., 2019; Skene and Henikoff, 2017) could be linked back to the functional characterization to identify the epigenetic signature of, and coactivators/repressors responsible for, primed and active CREs. This additional information can be used to inform annotation tools (such as GREAT (McLean et al., 2010) or ChIP-Enrich (Qin et al., 2022; Welch et al., 2014)) to better predict the target genes of CREs. The

combination of functional interrogation of CREs and greater prediction of CRE targets

will allow us in the near future to gain a comprehensive and actionable understanding

of the epigenetic mechanisms that bestow stem cells with their multilineage capacity.

Bibliography

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. Nat. Commun. *9*, 1911.

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.

An, X., Schulz, V.P., Li, J., Wu, K., Liu, J., Xue, F., Hu, J., Mohandas, N., and Gallagher, P.G. (2014). Global transcriptome analyses of human and murine terminal erythroid differentiation. Blood *123*, 3466–3477.

Anzalone, A. V, Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nat. Biotechnol. *38*, 824–844.

Bender, M.A., Ragoczy, T., Lee, J., Byron, R., Telling, A., Dean, A., and Groudine, M. (2012). The hypersensitive sites of the murine β-globin locus control region act independently to affect nuclear localization and transcriptional elongation. Blood *119*, 3820–3827.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326.

Bonifer, C., and Cockerill, P.N. (2017). Chromatin priming of genes in development: Concepts, mechanisms and consequences. Exp. Hematol. *49*, 1–8.

Boyer, S.W., Schroeder, A. V, Smith-Berdan, S., and Forsberg, E.C. (2011). All Hematopoietic Cells Develop from Hematopoietic Stem Cells through Flk2/Flt3-Positive Progenitor Cells. Cell Stem Cell *9*, 64–73.

Boyer, S.W., Beaudin, A.E., and Forsberg, E.C. (2012). Mapping differentiation pathways from hematopoietic stem cells using Flk2/Flt3 lineage tracing. Cell Cycle *11*, 3180–3188.

Boyer, S.W., Rajendiran, S., Beaudin, A.E., Smith-berdan, S., Muthuswamy, P.K., Perez-Cunningham, J., Martin, E.W., Cheung, C., Tsang, H., Landon, M., et al. (2019). Clonal and Quantitative In Vivo Assessment of Hematopoietic Stem Cell Differentiation Reveals Strong Erythroid Potential of Multipotent Cells. Stem Cell Reports *12*, 801–815.

Bryder, D., Rossi, D.J., and Weissman, I.L. (2006). Hematopoietic Stem Cells: The Paradigmatic Tissue-Specific Stem Cell. Am. J. Pathol. *169*, 338–346.

Buenrostro, J., Wu, B., Chang, H., and Greenleaf, W. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. *109*,

21.29.1-21.29.9.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Meth *10*, 1213–1218.

Buenrostro, J.D., Corces, M.R., Lareau, C.A., Wu, B., Schep, A.N., Aryee, M.J., Majeti, R., Chang, H.Y., and Greenleaf, W.J. (2018). Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. Cell *0*, 1–14.

Bultman, S.J., Gebuhr, T.C., and Magnuson, T. (2005). A Brg1 mutation that uncouples ATPase activity from chromatin remodeling reveals an essential role for SWI/SNF-related complexes in β-globin expression and erythroid development. Genes Dev. *19*, 2849–2861.

Bulut-Karslioglu, A., Macrae, T.A., Oses-Prieto, J.A., Covarrubias, S., Percharde, M., Ku, G., Diaz, A., McManus, M.T., Burlingame, A.L., and Ramalho-Santos, M. (2018). The Transcriptionally Permissive Chromatin State of Embryonic Stem Cells Is Acutely Tuned to Translational Output. Cell Stem Cell *22*, 369-383.e8.

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S.M., Reth, M., Höfer, T., Rodewald, H.-R.R., Hofer, T., et al. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature *518*, 542–546.

Cabal-Hierro, L., van Galen, P., Prado, M.A., Higby, K.J., Togami, K., Mowery, C.T., Paulo, J.A., Xie, Y., Cejas, P., Furusawa, T., et al. (2020). Chromatin accessibility promotes hematopoietic and leukemia stem cell activity. Nat. Commun. *11*, 1406.

Cabezas-Wallscheid, N., Klimmeck, D., Hansson, J., Lipka, D.B., Reyes, A., Wang, Q., Weichenhan, D., Lier, A., Von Paleske, L., Renders, S., et al. (2014). Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. Cell Stem Cell *15*, 507–522.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell *49*, 825–837.

Carrelha, J., Meng, Y., Kettyle, L.M., Luis, T.C., Norfo, R., Alcolea, V., Boukarabila, H., Grasso, F., Gambardella, A., Grover, A., et al. (2018). Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. Nature *554*, 106–111.

Chambers, S.M., Boles, N.C., Lin, K.Y.K., Tierney, M.P., Bowman, T. V., Bradfute, S.B., Chen, A.J., Merchant, A.A., Sirin, O., Weksberg, D.C., et al. (2007). Hematopoietic Fingerprints: An Expression Database of Stem Cells and Their Progeny. Cell Stem Cell *1*, 578–591.

Chang, D.H., Angelin-Duclos, C., and Calame, K. (2000). BLIMP-1: trigger for differentiation of myeloid lineage. Nat. Immunol. *1*, 169–176.

Chi, T.H., Wan, M., Lee, P.P., Akashi, K., Metzger, D., Chambon, P., Wilson, C.B., and Crabtree, G.R. (2003). Sequential Roles of Brg, the ATPase Subunit of BAF Chromatin Remodeling Complexes, in Thymocyte Development. Immunity *19*, 169–182.

Cool, T., Worthington, A., Poscablo, D., Hussaini, A., and Forsberg, E.C. (2020). Interleukin 7 receptor is required for myeloid cell homeostasis and reconstitution by hematopoietic stem cells. Exp. Hematol. *90*, 39-45.e3.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. 201016071.

Cullen, S.M., Mayle, A., Rossi, L., and Goodell, M.A. (2014). Chapter Two - Hematopoietic Stem Cell Development: An Epigenetic Journey. In Stem Cells in Development and Disease, M.B.T.-C.T. in D.B. Rendl, ed. (Academic Press), pp. 39–75.

Dai, X.-M., Ryan, G.R., Hapel, A.J., Dominguez, M.G., Russell, R.G., Kapp, S., Sylvestre, V., and Stanley, E.R. (2002). Targeted disruption of the mouse colony-stimulating factor 1 receptor gene results in osteopetrosis, mononuclear phagocyte deficiency, increased primitive progenitor cell frequencies, and reproductive defects. Blood, J. Am. Soc. Hematol. *99*, 111–120.

Ema, H., Morita, Y., and Suda, T. (2014). Heterogeneity and hierarchy of hematopoietic stem cells. Exp. Hematol. *42*, 74-82.e2.

F., D.L., S., B.J., M., D.C., D., S.D., Julia, S., L., B.T., H., R.J., and M., G.J. (2009). Targeted Knock-In Mice Expressing Mutations of CD28 Reveal an Essential Pathway for Costimulation. Mol. Cell. Biol. *29*, 3710–3721.

Forsberg, E.C., Downs, K.M., Christensen, H.M., Im, H., Nuzzi, P.A., and Bresnick, E.H. (2000). Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. Proc. Natl. Acad. Sci. *97*, 14494 LP – 14499.

Forsberg, E.C., Prohaska, S.S., Katzman, S., Heffner, G.C., Stuart, J.M., and Weissman, I.L. (2005). Differential expression of novel potential regulators in hematopoietic stem cells. PLoS Genet. *1*, e28.

Gasiorek, J.J., Nouhi, Z., and Blank, V. (2012). Abnormal differentiation of erythroid precursors in p45 NF-E2−/− mice. Exp. Hematol. *40*, 393–400.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M.J., Heidersbach, A., Ramalho-Santos, J., McManus, M.T., Plath, K., Meshorer, E., et al. (2009). Chd1

regulates open chromatin and pluripotency of embryonic stem cells. Nature *460*, 863–868.

Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. Nat. Rev. Mol. Cell Biol. *12*, 36–47.

Gasperini, M., Hill, A.J., McFaline-Figueroa, J.L., Martin, B., Kim, S., Zhang, M.D., Jackson, D., Leith, A., Schreiber, J., Noble, W.S., et al. (2019). A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. Cell *176*, 377-390.e19.

Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat. Rev. Genet. *21*, 292–310.

Goode, D.K., Obier, N., Vijayabaskar, M.S., Lie-A-Ling, M., Lilly, A.J., Hannah, R., Lichtinger, M., Batta, K., Florkowska, M., Patel, R., et al. (2016). Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. Dev. Cell *36*, 572–587.

Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. Annu. Rev. Biochem. *57*, 159–197.

Growney, J.D., Shigematsu, H., Li, Z., Lee, B.H., Adelsperger, J., Rowan, R., Curley, D.P., Kutok, J.L., Akashi, K., Williams, I.R., et al. (2005). Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. Blood *106*, 494–504.

Haeckel, E. (1868). Natürliche Schöpfungsgeschichte: gemeinverständliche wissenschaftliche Vorträge über die Entwickelungslehre im allgemeinen und diejenigen von Darwin, Goethe und Lamarck im besonderen, über die Anwendung derselben auf den Ursprung des Menschen und andere d (Reimer).

Han, L., Madan, V., Mayakonda, A., Dakle, P., Woon, T.W., Shyamsunder, P., Nordin, H.B.M., Cao, Z., Sundaresan, J., Lei, I., et al. (2019). Chromatin remodeling mediated by ARID1A is indispensable for normal hematopoiesis in mice. Leukemia *33*, 2291–2305.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

Hestdal, K., Ruscetti, F.W., Ihle, J.N., Jacobsen, S.E., Dubois, C.M., Kopp, W.C., Longo, D.L., and Keller, J.R. (1991). Characterization and regulation of RB6-8C5 antigen expression on murine bone marrow cells. J. Immunol. *147*, 22 LP – 28.

Heuston, E.F., Keller, C.A., Lichtenberg, J., Giardine, B., Anderson, S.M., Hardison, R.C., and Bodine, D.M. (2018). Establishment of regulatory elements during erythro-megakaryopoiesis identifies hematopoietic lineage-commitment points. Epigenetics and Chromatin *11*, 1–18.

Hill, A.J., McFaline-Figueroa, J.L., Starita, L.M., Gasperini, M.J., Matreyek, K.A., Packer, J., Jackson, D., Shendure, J., and Trapnell, C. (2018). On the design of CRISPR-based single-cell molecular screens. Nat. Methods *15*, 271–274.

Hsu, L.-Y., Lauring, J., Liang, H.-E., Greenbaum, S., Cado, D., Zhuang, Y., and Schlissel, M.S. (2003). A Conserved Transcriptional Enhancer Regulates RAG Gene Expression in Developing B Cells. Immunity *19*, 105–117.

Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M., Heyworth, C., and Enver, T. (1997). Multilineage gene expression precedes commitment in the hemopoietic system. Genes Dev. *11*, 774–785.

Jepsen, K., Hermanson, O., Onami, T.M., Gleiberman, A.S., Lunyak, V., McEvilly, R.J., Kurokawa, R., Kumar, V., Liu, F., Seto, E., et al. (2000). Combinatorial Roles of the Nuclear Receptor Corepressor in Transcription and Development. Cell *102*, 753–763.

Karsunky, H., Merad, M., Cozzio, A., Weissman, I.L., and Manz, M.G. (2003). Flt3 Ligand Regulates Dendritic Cell Development from Flt3+ Lymphoid and Myeloid-committed Progenitors to Flt3+ Dendritic Cells In Vivo . J. Exp. Med. *198*, 305–313.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat. Commun. *10*, 1–10.

Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaöz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P., et al. (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res. *17*, 691–707.

Kondo, M., Weissman, I.L., and Akashi, K. (1997). Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow. Cell *91*, 661–672.

Krause, D.S. (2002). Plasticity of marrow-derived stem cells. Gene Ther *9*.

Kruse, E.A., Loughran, S.J., Baldwin, T.M., Josefsson, E.C., Ellis, S., Watson, D.K., Nurden, P., Metcalf, D., Hilton, D.J., Alexander, W.S., et al. (2009). Dual requirement for the ETS transcription factors Fli-1 and Erg in hematopoietic stem cells and the megakaryocyte lineage. Proc. Natl. Acad. Sci. *106*, 13814 LP – 13819.

Kuo, T.C., and Schlissel, M.S. (2009). Mechanisms controlling expression of the RAG locus during lymphocyte development. Curr. Opin. Immunol. *21*, 173–178.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Chromatin state dynamics during blood formation. Science *55*, 1–10.

Laszkiewicz, A., Sniezewski, L., Kasztura, M., Bzdzion, L., Cebrat, M., and Kisielow, P. (2012). Bidirectional Activity of the NWC Promoter Is Responsible for RAG-2 Transcription in Non-Lymphoid Cells. PLoS One *7*, e44807.

Laurenti, E., and Göttgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. Nature *553*, 418–426.

Leung, G.A., Cool, T., Valencia, C.H., Worthington, A., Beaudin, A.E., and Camilla Forsberg, E. (2019). The lymphoid-associated interleukin 7 receptor (IL7R) regulates tissue-resident macrophage development. Dev. *146*, dev176180.

Li, K., Liu, Y., Cao, H., Zhang, Y., Gu, Z., Liu, X., Yu, A., Kaphle, P., Dickerson, K.E., Ni, M., et al. (2020). Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. Nat. Commun. *11*, 1–16.

Li, Q., Peterson, K.R., Fang, X., and Stamatoyannopoulos, G. (2002). Locus control regions. Blood *100*, 3077–3086.

Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. *5*, 1752–1779.

Ludwig, L.S., Lareau, C.A., Bao, E.L., Nandakumar, S.K., Muus, C., Ulirsch, J.C., Chowdhary, K., Buenrostro, J.D., Mohandas, N., An, X., et al. (2019). Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. Cell Rep. *27*, 3228-3240.e7.

Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T., et al. (2020). Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell *183*, 1103-1116.e20.

Månsson, R., Hultquist, A., Luc, S., Yang, L., Anderson, K., Kharazi, S., Al-Hashmi, S., Liuba, K., Thorén, L., Adolfsson, J., et al. (2007). Molecular Evidence for Hierarchical Transcriptional Lineage Priming in Fetal and Adult Stem Cells and Multipotent Progenitors. Immunity *26*, 407–419.

Martin, E.W., Krietsch, J., Reggiardo, R.E., Sousae, R., Kim, D.H., and Forsberg,

E.C. (2021). Chromatin accessibility maps provide evidence of multilineage gene priming in hematopoietic stem cells. Epigenetics and Chromatin *14*, 1–15.

McDevitt, M.A., Shivdasani, R.A., Fujiwara, Y., Yang, H., and Orkin, S.H. (1997). A "knockdown" mutation created by cis-element gene targeting reveals the dependence of erythroid cell maturation on the level of transcription factor GATA-1. Proc. Natl. Acad. Sci. *94*, 6781–6785.

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

Nestorowa, S., Hamey, F.K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N.K., Kent, D.G., Gottgens, B., Sala, B.P., et al. (2016). e- Blood A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood *128*, 20–32.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. Cell *132*, 631–644.

Palstra, R., de Laat, W., and Grosveld, F.B.T.-A. in G. (2008). Chapter 4 β-Globin Regulation and Long-Range Interactions. In Long-Range Control of Gene Expression, (Academic Press), pp. 107–142.

Phillips, R., Ernst, R.E., Brian, B., Natalia, I., Mahan, M.A., Deanehan, J.K., Moore, K.A., Christian, O.G., and Lemischka, I.R. (2000). The Genetic Program of Hematopoietic Stem Cells. Science (80-.). *288*, 1635–1640.

Poscablo, D.M., Worthington, A.K., Smith-Berdan, S., and Forsberg, E.C. (2021). Megakaryocyte progenitor cell function is enhanced upon aging despite the functional decline of aged hematopoietic stem cells. Stem Cell Reports *16*, 1598–1613.

Pronk, C.J.H., Rossi, D.J., Månsson, R., Attema, J.L., Norddahl, G.L., Chan, C.K.F., 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.

Qin, T., Lee, C., Li, S., Cavalcante, R.G., Orchard, P., Yao, H., Zhang, H., Wang, S., Patil, S., Boyle, A.P., et al. (2022). Comprehensive enhancer-target gene assignments improve gene set level interpretation of genome-wide regulatory data. Genome Biol. *23*, 105.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Rajendiran, S., Smith-Berdan, S., Kunz, L., Risolino, M., Selleri, L., Schroeder, T.,

and Forsberg, E.C. (2020). Ubiquitous overexpression of CXCL12 confers radiation protection and enhances mobilization of hematopoietic stem and progenitor cells. Stem Cells *38*, 1159–1174.

Replogle, J.M., Norman, T.M., Xu, A., Hussmann, J.A., Chen, J., Cogan, J.Z., Meer, E.J., Terry, J.M., Riordan, D.P., Srinivas, N., et al. (2020). Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. Nat. Biotechnol.

Rodrigues, C.P., Herman, J.S., Herquel, B., Valsecchi, C.I.K., Stehle, T., Grün, D., and Akhtar, A. (2022). Temporal expression of MOF acetyltransferase primes transcription factor networks for erythroid fate. Sci. Adv. *6*, eaaz4815.

Rodriguez-Fraticelli, A.E., Wolock, S.L., Weinreb, C.S., Panero, R., Patel, S.H., Jankovic, M., Sun, J., Calogero, R.A., Klein, A.M., and Camargo, F.D. (2018). Clonal analysis of lineage fate in native haematopoiesis. Nature *553*, 212–216.

Rojo, R., Raper, A., Ozdemir, D.D., Lefevre, L., Grabert, K., Wollscheid-Lengeling, E., Bradford, B., Caruso, M., Gazova, I., Sánchez, A., et al. (2019). Deletion of a Csf1r enhancer selectively impacts CSF1R expression and development of tissue macrophage populations. Nat. Commun. *10*, 1–17.

Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017a). ChromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*, 975–978.

Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017b). ChromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*, 975–978.

Schlenner, S.M., Madan, V., Busch, K., Tietz, A., Läufle, C., Costa, C., Blum, C., Fehling, H.J., and Rodewald, H.-R. (2010). Fate Mapping Reveals Separate Origins of T Cells and Myeloid Lineages in the Thymus. Immunity *32*, 426–436.

Schmidt, U., van den Akker, E., Parren-van Amelsvoort, M., Litos, G., de Bruijn, M., Gutiérrez, L., Hendriks, R.W., Ellmeier, W., Löwenberg, B., Beug, H., et al. (2004). Btk is required for an efficient response to erythropoietin and for SCF-controlled protection against TRAIL in erythroid progenitors. J. Exp. Med. *199*, 785–795.

Schütte, J., Wang, H., Antoniou, S., Jarratt, A., Wilson, N.K., Riepsaame, J., Calero-Nieto, F.J., Moignard, V., Basilico, S., Kinston, S.J., et al. (2016). An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. Elife *5*, e11469–e11469.

Seita, J., and Weissman, I.L. (2010). Hematopoietic stem cell: self-renewal versus differentiation. WIREs Syst. Biol. Med. *2*, 640–653.

Seita, J., Sahoo, D., Rossi, D.J., Bhattacharya, D., Serwold, T., Inlay, M.A., Ehrlich,

L.I.R., Fathman, J.W., Dill, D.L., and Weissman, I.L. (2012). Gene expression commons: An open platform for absolute gene expression profiling. PLoS One *7*, 1–11.

Sherr, C.J., Rettenmier, C.W., Sacca, R., Roussel, M.F., Look, A.T., and Stanley, E.R. (1985). The c-fms proto-oncogene product is related to the receptor for the mononuclear phagocyte growth factor, CSF 1. Cell *41*, 665–676.

Shivdasani, R.A., and Orkin, S.H. (1995). Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2. Proc. Natl. Acad. Sci. *92*, 8690 LP – 8694.

Shivdasani, R.A., Fujiwara, Y., McDevitt, M.A., and Orkin, S.H. (1997). A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. EMBO J. *16*, 3965–3973.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. *15*, 272–286.

Shulga-Morskaya, S., Dobles, M., Walsh, M.E., Ng, L.G., MacKay, F., Rao, S.P., Kalled, S.L., and Scott, M.L. (2004). B Cell-Activating Factor Belonging to the TNF Family Acts through Separate Receptors to Support B Cell Survival and T Cell-Independent Antibody Formation. J. Immunol. *173*, 2331 LP – 2341.

Siegwart, L.C., Schwemmers, S., Wehrle, J., Koellerer, C., Seeger, T., Gründer, A., and Pahl, H.L. (2020). The transcription factor NFE2 enhances expression of the hematopoietic master regulators SCL/TAL1 and GATA2. Exp. Hematol. *87*, 42-47.e1.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife *6*, 1–35.

Spangrude, G.J., Shelly, H., and Weissman, I. (1988). Purification and Characterization of Mouse Hematopoietic Stem Cells. Science (80-.). *241*, 58–62.

Stenvers, K.L., Tursky, M.L., Harder, K.W., Kountouri, N., Amatayakul-Chantler, S., Grail, D., Small, C., Weinberg, R.A., Sizeland, A.M., and Zhu, H.-J. (2003). Heart and Liver Defects and Reduced Transforming Growth Factor β2 Sensitivity in Transforming Growth Factor β Type III Receptor-Deficient Embryos. Mol. Cell. Biol. *23*, 4371 LP – 4385.

Surani, M.A., Hayashi, K., and Hajkova, P. (2007). Genetic and Epigenetic Regulators of Pluripotency. Cell *128*, 747–762.

Terskikh, A. V, Miyamoto, T., Chang, C., Diatchenko, L., and Weissman, I.L. (2003). Gene expression analysis of purified hematopoietic stem cells and committed progenitors. Blood *102*, 94–101.

Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Methods *12*, 1143–1149.

Tsang, A.P., Fujiwara, Y., Hom, D.B., and Orkin, S.H. (1998). Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. Genes Dev. *12*, 1176–1188.

Ugarte, F., Sousae, R., Cinquin, B., Martin, E.W., Krietsch, J., Sanchez, G., Inman, M., Tsang, H., Warr, M., Passegué, E., et al. (2015). Progressive Chromatin Condensation and H3K9 Methylation Regulate the Differentiation of Embryonic and Hematopoietic Stem Cells. Stem Cell Reports *5*, 728–740.

Vilarrasa-Blasi, R., Soler-Vila, P., Verdaguer-Dot, N., Russiñol, N., Di Stefano, M., Chapaprieta, V., Clot, G., Farabella, I., Cuscó, P., Kulis, M., et al. (2021). Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. Nat. Commun. *12*, 651.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J. a, Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854–858.

Waddington, C.H. (1940). Organisers and genes (University Press; Cambridge).

Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N.A., Muth, K., Palmer, J., Qiu, Y., Wang, J., et al. (2015). Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. Cell Stem Cell *16*, 386–399.

Wei, X.-C., Kishi, H., Jin, Z.-X., Zhao, W.-P., Kondo, S., Matsuda, T., Saito, S., and Muraguchi, A. (2002). Characterization of Chromatin Structure and Enhancer Elements for Murine Recombination Activating Gene-2. J. Immunol. *169*, 873 LP – 881.

Welch, R.P., Lee, C., Imbriano, P.M., Patil, S., Weymouth, T.E., Smith, R.A., Scott, L.J., and Sartor, M.A. (2014). ChIP-Enrich: gene set enrichment testing for ChIP-seq data. Nucleic Acids Res. *42*, e105–e105.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319.

Wilkinson, A.C., Ishida, R., Kikuchi, M., Sudo, K., Morita, M., Crisostomo, R.V., Yamamoto, R., Loh, K.M., Nakamura, Y., Watanabe, M., et al. (2019). Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation. Nature *571*, 117–121.

Willcockson, M.A., Taylor, S.J., Ghosh, S., Healton, S.E., Wheat, J.C., Wilson, T.J., Steidl, U., and Skoultchi, A.I. (2019). Runx1 promotes murine erythroid progenitor proliferation and inhibits differentiation by preventing Pu.1 downregulation. Proc. Natl. Acad. Sci. *116*, 17841 LP – 17847.

Xu, J., Watts, J.A., Pope, S.D., Gadue, P., Kamps, M., Plath, K., Zaret, K.S., and Smale, S.T. (2009). Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. Genes Dev. *23*, 2824–2838.

Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K.L., Ema, H., and Nakauchi, H. (2013). Clonal Analysis Unveils Self-Renewing Lineage-Restricted Progenitors Generated Directly from Hematopoietic Stem Cells. Cell *154*, 1112–1126.

Yamamoto, R., Wilkinson, A.C., Ooehara, J., Lan, X., Lai, C.-Y., Nakauchi, Y., Pritchard, J.K., and Nakauchi, H. (2018). Large-Scale Clonal Analysis Resolves Aging of the Mouse Hematopoietic Stem Cell Compartment. Cell Stem Cell *22*, 600-607.e4.

Yannoutsos, N., Barreto, V., Misulovin, Z., Gazumyan, A., Yu, W., Rajewsky, N., Peixoto, B.R., Eisenreich, T., and Nussenzweig, M.C. (2004). A cis element in the recombination activating gene locus regulates gene expression by counteracting a distant silencer. Nat. Immunol. *5*, 443–450.

Yu, W., Misulovin, Z., Suh, H., Hardy, R.R., Jankovic, M., Yannoutsos, N., and Nussenzweig, M.C. (1999). Coordinate Regulation of RAG1 and RAG2 by Cell Type-Specific DNA Elements 5' of RAG2. Science (80-.). *285*, 1080–1084.