Title
Deciphering the Hematopoietic Gene Regulatory Architecture

Author
Ramirez, Ricardo Noel

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Deciphering the Hematopoietic Gene Regulatory Architecture

DISSERTATION

Submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Developmental and Cell Biology

by

Ricardo Noel Ramirez

Dissertation Committee:
Assistant Professor Ali Mortazavi
Professor Peter Donovan
Professor Kyoko Yokomori
Professor Ken Cho
Professor Xiaohui Xie

2016

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my appreciation and gratitude to my advisor Assistant Professor Ali Mortazavi, for his continued support and guidance throughout my dissertation. With his support I have been able to participate and learn from several opportunities, which has furthered my academic and professional growth. With his mentorship I have exceled and become an independent scientist and without his guidance my thesis dissertation would not have been possible.

I would like to thank my committee members, Professor Peter Donovan, Professor Ken Cho, Professor Kyoko Yokomori and Professor Xiaohui Xie, who have helped guide my dissertation and demonstrated commitment in this regard.

I would like to thank Professor Peter Donovan, who has always been a positive influence in both my academic and personal life.

I would like to thank my parents, Ralph and Julia Ramirez for their encouragement and continued support. I would like to thank my sister Patricia Ramirez, who has always been a most positive and strong support for me during my studies.
I would like to thank Nicole Che El-Ali for her continued support, work on research projects, and positivity throughout my academic career.
I would like to thank my entire family for their encouragement and support.
I would like to thank my friends Joey Medina, Ashwin Jackson, Ray Rojes and Kenny Rojes for their encouragement and support over the years.

I would like to thank the ENCODE and STATegra consortiums. Specifically, David Gomez-Cabrero who has served as an esteemed colleague and friend.

I would like to thank all of my research collaborators here at UCI and across all institutes.

I would like to thank Dr. Marlene De La Cruz and Dr. Luis Mota-Bravo for their funding support and advice throughout my studies.

I thank the University of California Irvine Developmental and Cell Biology and Center for Complex Biological Systems departments. Specifically Karen Martin, who has always found a way to help a graduate student in times of great need.

I would like to thank Dr. Weihua Zeng for his advice and support on all research matters during my dissertation. I would like to thank Camden Jansen for his efforts in our co-authored manuscript. Lastly I would like to thank every member of the Mortazavi laboratory and wish all great success in his or her future endeavors.

CURRICULUM VITAE

Ricardo Noel Ramirez

2011 B.S. in Cell and Molecular Biology, California State University Long Beach

2011 California Institute for Regenerative Medicine (CIRM) training fellow

2016 PhD in Developmental and Cell Biology, University of California Irvine

FIELD OF STUDY

Gene regulation of hematopoiesis in mammalian systems

PUBLICATIONS

- *Dynamic gene regulatory networks of human myeloid differentiation*. **Ricardo N. Ramirez**, Nicole C. El-Ali, Dana Wyman, Ana Conesa, Ali Mortazavi (under review).

- *Profiling the single-cell transcriptome and chromatin landscape of pre-B cell differentiation*. **Ricardo N. Ramirez**, Camden Jansen, Ali Mortazavi. (in preparation).

- *STATegra integrative analysis of pre-B cell differentiation*. STATegra Consortium. (in preparation).

- *The landscape of global long-rang interactions in mouse and human erythroid cells mediated by YY1, GATA1, and CTCF during differentiation*. Weihua Zeng, **Ricardo N. Ramirez**, Shan Jiang, Camden Jansen, Marissa Macchietto, Ali Mortazavi. (in preparation).

- *The dynamics of NRSF/REST motif evolution favor the canonical NRSE/RE1 form*. Shan Jiang, **Ricardo N. Ramirez**, Nicole C. El-Ali, Ali Mortazavi. (under review).

- *Fully-defined generation of iPS-derived human microglia to study Neurological Diseases*. E.M. Abud[1,2,3], **R.N. Ramirez**[4,5], E.S. Martinez[1,2,3], L.M. Healy[7], C.H.H. Nguyen[1,2,3], V.M. Scarfone[2], S.E. Marsh[1,2,3], A. Madany[8], C. Fimbres[2,3], C. Caraway[2,3], M.D. Torres[1,2,3], A. Park[2,3], K. Rakez[6], A., K.H. Gylys[7], A. Mortazavi[4,5], J.P. Antel[7], M.J. Carson[2], W.W. Poon[2,3], M. Blurton-Jones[1,2,3,*]**.** (in preparation).

- *DNMT3a-dependent specific hypermethylation is associated with a tolerogenic phenotype in myeloid-derived suppressor cells*. Javier Rodriguez-Ubreva, Francesc Catala-Moll, Natasa Obermajer, Damiana Alvarez-Errico, **Ricardo N. Ramirez**, Carlos Company, Roser Vento-Tormo, Ali Mortazavi, Pawel Kalinski, Esteban Ballestar. (under review).

- *TCF7L1 is a BMP4 target and opposes mesendoderm differentiation to stabilize human embryonic stem cell pluripotency*. Robert A. Sierra, Nathan P. Hoverter, **Ricardo N. Ramirez**, Ali Mortazavi, Bradley J. Merrill, Marian L. Waterman, Peter J. Donovan. (in preparation).

- *Fat cells regenerate from myofibroblasts during wound healing*. Maksim V. Plikus[1,2*], Christian Fernando Guerrero-Juarez[2], Mayumi Ito[3], Yun Rose Li[4], Priya H. Dedhia[5], Olga Shestova[5], Denise L. Gay[1,6], Raul Ramos[2], Tsai-Ching Hsi[2], Ji Won Oh[2], Xiaojie Wang[2], **Ricardo Ramirez**, Rabi Murad, Amanda Ramirez[2], Sara E. Konopelski[2], Arijh Elzein[2], Anne Wang[1], Rarinthip June Supapannachart[1], Zaixin Yang[1], Ying Zheng[1], Arben

Nace[1], Amy Guo[1], Elsa Treffeisen[1], Thomas Andl[7], Stefan Offermanns[8], Daniel Metzger[9], Pierre Chambon[9], Ali Mortazavi, Rana K. Gupta[10], Bruce A. Hamilton[11], Sarah E. Millar[1], Patrick Seale[5, 12], Warren S. Pear[5], Mitchell A. Lazar[4, 13], George Cotsarelis[1*] (under review).

- *Apolipoprotein E LDL receptor-binding domain-containing high-density lipoprotein: A nanovehicle to transport curcumin, an antioxidant and anti-amyloid bioflavonoid*. Panupon Khumsupan, **Ricardo N. Ramirez** , Darin Khumsupan ,Vasanthy Narayanaswami. Biochim Biophys Acta. 2011 Jan;1808(1):352-9. doi: 10.1016/j.bbamem.2010.09.007. Epub 2010 Sep 17.

AWARDS and HONORS

- Keystone Symposia scholarship (National Institute of Allergy and Infectious Diseases, Grant #1R13AI120448-01). 01/10/2016.
- Developmental and Cell Biology Poster competition award (1st Place), UCI Developmental and Cell Biology Retreat. 04/23/14.
- Broad Institute Visiting Computational Scientist Award. 01/04/14-02/07/14.
- Predoctoral Training Grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NIH/NICHD) "Training Program in the Systems Biology of Development". Grant Number: 1 T32 HD060555-05: 07/01/13-06/30/14.
- AAAS Student Poster Competition Winner in Cellular and Molecular Biology. 03/25/13.
- Developmental and Cell Biology Poster competition award (1st Place), UCI Developmental and Cell Biology Retreat. 02/28/13.
- Minority Biomedical Research Support (MBRS-IMSD) Grant GM055246. (NIH). 06/01/11-06/01/13.
- UCI Center for Complex Biological Systems Opportunity Research Award. 09/01/12.

PRESENTATIONS

- Gene expression & signaling in the immune system. CSHL, New York USA. 01/11/16. Poster.
- Keystone Symposia, Systems Immunology: From Molecular Networks to Human Biology, Montana. 01/11/16. **Speaker**.
- Exploring Complex Systems: From Human Biology to Exercise Science, University of California Irvine. 12/09/15. **Speaker.**
- Fluidigm Single-cell Users Meeting, City of Hope, CA. 10/15/15. **Speaker.**
- Statistical Methods for Omics Data Integration and Analysis (SMODIA), Centro de Investigacion Principe Felipe (CIPF), Valencia, Spain. 09/15/15. **Speaker.**
- STATegra Summer School in Omics Data Integration. Benicassim, Spain. 09/09/15. **Speaker.**
- STATegra Summer School in Omics Data Integration. Benicassim, Spain. 09/10/15. **Speaker.**
- STATegra Consortium meeting, LMU Munich Germany, 05/07/15. **Speaker.**
- NHGRI NIH, Bethesda, Maryland, 08/11/14. Poster,
- Developmental and Cell Biology Poster session, University of California, Irvine**,** 04/23/14. Poster.
- **STATegra** Next Generation Sequencing Workshop, University of Amsterdam, Amsterdam, Netherlands, 03/24/14. **Speaker.**
- ENCODE Consortium Meeting, Stanford University, 06/08/13. Poster.
- Center for Complex Biological Systems Department Retreat, 03/23/13. **Speaker.**
- AAAS Conference, Boston, MA. 02/15/13. Poster.
- Biology Seminar Series, California State University Long Beach. 11/08/12. **Speaker.**
- International Society for Computational Biology (ISMB) Hit-Seq Conference, Long Beach, California.07/13/12. Poster.
- Center for Complex Biological Systems Annual Retreat, University of California, Irvine. 03/30/12. Poster.
- Cold Spring Harbor Laboratory; "Systems Biology: Global Regulation of Gene Expression", One Bungtown Road Cold Spring Harbor, NY. 03/20/12. Poster.
- Department of Developmental and Cell Biology Retreat, University of California, Irvine CA. 11/11/11. Poster.

- Minority Biomedical Research Support/NIH Program Symposium, University of California Irvine, CA. 09/20/11. Poster.

PROFESSIONAL DEVELOPMENT AND TEACHING EXPERIENCE

- STATegra Summer School in Omics Data Integration Instructor, Benicassim Spain. 09/07/15-09/11/15.
- NHGRI NIH Genomics course. 08/05/14-08/11/14.
- Broad Institute CEGS visiting computational scientist. 01/04/14-02/07/14.
- DNase-seq training course instructor, Imperial College London, United Kingdom. 06/08/13-06/15/13.
- NIH Minority in Biomedical Research Science (MBRS) Undergradute/Graduate Instructor. 06/2013-06/2016.
- University of California Irvine, Sue and Bill Gross CIRM Stem Cell Training Course. 08/2011.

ABSTRACT OF THE DISSERTATION

Deciphering the Hematopoietic Gene Regulatory Architecture

By

Ricardo Noel Ramirez

Doctor of Philosophy in Developmental and Cell Biology

University of California, Irvine, 2016

Assistant Professor Ali Mortazavi

The immune system is a complex and interactive network of diverse cell types, with a myriad of functional properties that are fundamental to maintaining an immunological-responsive balance within an organism. Thus, coordinated organization of cellular differentiation is established early and throughout the development of an organism, resulting in the generation of the interacting innate and adaptive immune systems. The temporal component is critical for capturing transient regulatory events during cellular differentiation. Thus, high-resolution analyses are necessary to derive immune regulatory dynamics. To address this, I focused my graduate studies on the genome-wide analysis of cellular differentiation with a primary focus on: (i) myeloid transcriptional and cis-regulatory modules; (ii) dynamic myeloid gene regulatory networks; and (iii) analysis of the expression and accessible chromatin landscape in single-cells. Utilizing high-resolution time-series analyses, I have identified temporal modules of cell-type specific cis-regulatory and transcriptional regulation during myeloid and pre-B cell differentiation. Such genome-wide modules were then utilized to infer novel regulatory interactions and generation of dynamic gene regulatory networks during myeloid differentiation. Secondly, both an independent and integrative analysis of single-cell chromatin maps and

expression patterns of pre-B cell differentiation, revealing complex relationships between the chromatin landscape and TF regulators in differentiating single-cells. Take together; these results depict the temporal dynamics mediated across the cis-regulatory landscape and transcriptional regulators that ultimately shape differentiating mammalian cells across the innate and adaptive immune systems.

Theme of thesis

The 'instructions', or regulatory architecture, for development in all living organisms is ultimately encoded in their DNA. While every cell in an organism contains the same DNA molecules, these 'instructions' are read and functionally applied in a cell autonomous manner. The central theme of my thesis is to decipher the DNA regulatory architecture for cells undergoing cell fate determination. Focusing on the hematopoietic system, for which a wealth of information is known, I employed next-generation sequencing and computational approaches to construct time-course driven gene regulatory networks. Chapter 2 of my thesis discusses the rapid changes in gene expression and chromatin accessibility in the differentiation of promyelocytic cells into macrophages, neutrophils, monocytes, and monocyte-derived macrophages, demonstrating that each cell-type response varies during differentiation. Additionally I detail the temporal transcriptional modules and chromatin accessible patterns, revealing several candidate regulators that are integral to cell specification during myeloid differentiation. I also examine the immediate immune response by LPS-stimulation and find relatively few immediate changes in chromatin accessibility compared to the more significant gene expression changes mediated by LPS stimulation. In chapter 3, I describe the power of integrating gene expression and leverage the deep sequencing of chromatin accessibility data to perform open chromatin footprinting and to build gene regulatory networks. I mapped the regulatory interactions for 23 dynamically expressed immune transcription factors, recovering both well-known interactions for key regulators such as PU.1, GFI1 and EGR, also identifying novel regulatory circuits not previously described in human myeloid cells. This work is currently under review in a peer-reviewed journal.

Following the theme of deciphering transcriptional regulation in immune cells, chapter 4 focuses on the transcriptome and chromatin dynamics during mouse pre-B cell differentiation. As part of my contribution to the STATegra consortium, I generated 36 DNase-seq, 324 single-cell RNA-seq and 227 single-cell ATAC-seq datasets, totaling 587 experiments of the 787 experimental datasets generated by STATegra. Here, I mapped the pseudo-time trajectory of scRNA-seq using Monocle to reveal heterogeneity and dynamic expression of transcription factors and signaling pathways in our pre-B model system. Secondly, I utilize regular DNase-seq and scATAC-seq to investigate the accessibility of differentiating pre-B cells, revealing dynamics in accessibility patterns for cis-regulatory elements of *Igll1, Vpreb1* and *Rag* genes. A main STATegra consortium paper is currently under preparation, which will include my DNase-seq and scRNA-seq results. A second co-first author manuscript in collaboration with Camden Jansen is currently under preparation that develops a computational method for integrating and studying time-course scRNA-seq and scATAC-seq data.

INTRODUCTION

Eukaryotic development depends on the orchestrated transcriptional regulation of gene expression with acute spatial-temporal precision. The gene regulatory networks that control development and differentiation are encoded in the genome, shared by all cells in the organism (Peters and Davidson, 2015). However, only a subset of genes are expressed in any given cell type through epigenetic control of DNA accessibility at promoters and enhancers for each gene. Genes that are active have "open", accessible chromatin while genes that are shut off lose that accessibility or are bound by repressors. This is ultimately controlled by transcription factors and long non-coding RNAs recruiting the chromatin remodeling at the precise genomic locations and times (Guttman et al., 2011). The dynamics and developmental commitment states within the cell rely on this remodeling of chromatin accessibility of promoter and enhancer Cis-Regulatory Modules (CRMs) across the genome. The identification of which CRMs have potential for activity in a cell is critical for understanding the complexities of gene regulatory circuits controlling cell fate and how development is encoded in the genome.

I. From the innate to adaptive system: Cellular characterization and the use of cancer cell lines of the immune system

The immune system is an integral component of all life. In vertebrates, an intricate defense system has evolved from invertebrates, which employ a minimalist, 'innate-like' program (Stuart and Ezekowitz, 2008). Importantly, the innate (first response) and adaptive (second response) immune systems allow for a layered defense with increased specificity to bacterial and viral pathogens. In humans, 1 microliter of blood contains roughly 4-6 million red blood cells, 4,000-11,000 leukocytes (white blood cells), and 150,000-450,000 thrombocytes (Ganong, 2003), all of

which are derived from 0.7 to 1.5 self-renewing hematopoietic stem cells /10^8 nucleated marrow cells or roughly 1 in every 10,000 bone marrow cell (Abkowitz et al., 2002; TILL and McCullouch, 1961). White blood cells play a primary role in protecting the body from infection. Remarkably, hematopoiesis generates over a million new hematopoietic cells a second (Sackmann, 1995) and must do so precisely to avoid both leukemia and leukopenia. Given the voluminous literature that exists on all aspects of immune system development, which has been extensively reviewed before (Amulic et al., 2012; Auffray et al., 2009; Chow et al., 2011; Varol et al., 2015), this brief review chapter focuses on the specific myeloid and lymphoid cells used during my graduate studies.

With respect to the innate immune system, hematopoietic progenitors diverge into the myeloid and lymphoid lineages of the blood system. The myeloid lineages that are primarily responsible for initial immune responses such as the neutralizing and clearing of bacteria, are cells known as phagocytes (Kaufmann and Dorhoi, 2016). Phagocytes were first described as immune cells that neutralize bacterial pathogens and are shown to be integral in eliciting antimicrobial responses (Metschnikoff, 1884). Extensive study of phagocytes has provided a comprehensive understanding of the inflammatory signaling mediated by bacterial pathogens *Listeria* and *Staphylococcus* (Kaufmann and Dorhoi, 2016). The myeloid lineage phagocytes include neutrophil, monocyte, macrophage, mast, basophil and eosinophil cells (Chow et al., 2011; Hume, 2006; Hume et al., 2002).

II. Innate system : Neutrophils

Neutrophils are the most abundant phagocytes in the blood, representing 50-75% of white blood cells in circulation, with a half-life in circulation of 1.5 and 8 hours in mice and humans

respectively (Kolaczkowska and Kubes, 2013). As an initial responder in bacterial response, the neutrophil has been shown to be important in neutralization (Kaufmann and Dorhoi, 2016), cell-to-cell communication (Scapini and Cassatella, 2014), and inflammatory disease (Amulic et al., 2012). A distinct, multi-lobed polymorphonuclear body is the primary morphological characteristic of mature neutrophils. While it was proposed initially that neutrophils in circulation were predominantly homogenous, experimental evidence supports the existence of several neutrophil sub-types in circulation (Kolaczkowska and Kubes, 2013; Scapini et al., 2001). Interestingly, neutrophil swarming has been shown to initiate novel repair mechanisms by facilitating cellular recruitment and tissue restructuring at sites of damage (Lämmermann et al., 2013). In brief, this relatively short-lived phagocyte is a critical component of innate bacterial defense for stabilizing immediate responses and relaying information to all cells of the immune system.

III. Innate system : Monocytes

Monocytes are phagocytic cell that directly interact with neutrophils during immune challenges (Scapini and Cassatella, 2014) that are characterized by a bean-shaped nucleus. Monocytes are conserved across all vertebrates (Ginhoux and Jung, 2014), with evidence of a parallel cell in fly haemolymph (Williams, 2007). Monocytes comprise of 4% and 10% of white blood cells in mice and humans respectively (Ginhoux and Jung, 2014) have a half-life of 20 hours (van Furth and Cohn, 1968). Similarly to neutrophils, the concept of monocyte heterogeneity was first shown through the diverse expression of markers such as CD14 and CD16, which ultimately determine two distinct monocyte populations (Passlick et al., 1989). As monocytes exit the bone marrow and enter circulation, the expression of several markers (CD43,

CCR2, LY6C, CD14, CD16) ultimately delineate monocyte functions in either: (i) the recruitment towards sites of inflammation through cellular differentiation; (ii) and the patrolling of the endothelium and recruitment of neutrophils in circulation (Ginhoux and Jung, 2014). In both human and mouse studies, monocytes have been shown to differentiate into macrophages (Auffray et al., 2007) and dendritic cells (Randolph et al., 1999), which are important for replenishing these cells throughout the organism. Interestingly, in addition to differentiation and neutrophil recruitment, it has been shown that monocytes may also contribute to vascular inflammation (Yang et al., 2014) and are a major component in propagation of atherosclerotic plaques (Zhang et al., 2012a). Thus, monocytes are not only crucial during an immune response and inflammation, but are important in replenishing macrophage and dendritic cells in the blood.

IV. Innate system : Macrophages

In contrast to both monocytes and neutrophils, macrophages are unique in that they are both a component of the circulatory system, but are uniquely specified in tissues across the body. They are also unique in that they are an inherently adherent cell, unlike monocyte and neutrophils. While the precise origin of embryonic macrophages is still highly debated (Lavin et al., 2015; Varol et al., 2015), two hypothesis have been proposed. The first states that all embryonic macrophages are derived from embryonic erythromyeloid progenitors (EMP) in the yolk sac at E7.5 during development, ultimately giving rise to all macrophages in the adult. A second hypothesis suggests that the yolk sac-derived EMPs arise in two waves that differentially contribute to adult microglia and other tissue-resident macrophages that are separate from the circulating cells that will generate adult macrophages. Initially, an early wave of yolk sac-derived EMPs that appear around E7.5 in the yolk sac colonizes the brain. Additional fetal

tissues around E9 propagate all tissue macrophages, and a secondary wave of yolk sac-derived EMPs that colonize and expand in the fetal liver, ultimately giving rise to fetal liver monocytes. These fetal monocytes differentiate into macrophages that are sustained and replenished during adulthood. Tissue-specific macrophages enable a robust immune response across the entire organism, whereby macrophage cells can react immediately to bacterial or viral response from the vital tissues where they reside. This in turn sustains organ homeostasis during immune challenges. This remarkable plasticity provides macrophages the potential to react to multiple stimuli. Thus, the concept of macrophage polarization and plasticity initially led to the notion that macrophage cells can adopt two states: (i) M1 inflammatory ; (ii) M2 reparative or alternatively activated (Mackaness, 1962; Nathan et al., 1983; Stein et al., 1992). While it serves as an attractive and long-held model, work studying many transcriptomes and epigenetic profiles during macrophage polarization using various stimuli have led to the identification of additional polarization states in macrophages (Xue et al., 2014). The observed diversity of macrophage polarization led to a proposal for a new macrophage nomenclature to address the deficiencies of the simple bipolar scheme (Murray et al., 2014). While multiple studies have compared the transcriptome and epigenetic profiles of macrophages across tissues and in the bone marrow in both populations and single-cells (Lara-Astiaso et al., 2014; Lavin et al., 2014; Paul et al., 2015; Xue et al., 2014), the underlying gene regulatory networks during immune responses and cellular differentiation still remain to be elucidated for myeloid cells.

V. Adaptive system : Pre-B cells

The adaptive immune system is the secondary response primed by the innate response resulting in the activation of B and T cell lineages to neutralize bacteria, viral pathogens, and

retain memory of immune challenges for faster future responses. The first evidence of the function of cells that we now refer to as B cells was initially discovered through the study of circulating antitoxins in immunity to diphtheria and tetanus (Behring and Kitasato, 1890), though it was not discovered until 1947 that plasma B cells produce antibodies (Fagraeus, 1947), which was later verified using immunofluorescence in 1955 (Coons, 1955).

Similar to myeloid cells, B cells originate from hematopoietic progenitors in the bone marrow where they reside as pre-B cells, until differentiation events drive maturation and progression into secondary lymphoid organs. The pre-B cell contains a well-studied pre-BCR (B cell receptor) which includes the μ-heavy chain, VpreB, Igα and Igβ receptors (Cooper, 2015; Melchers, 2005). Pre-B differentiation requires that cells undergo light-chain rearrangement, subsequent loss of the pre-BCR (VpreB, Igα and Igβ receptors) and eventual development of a mature BCR that is capable of binding antigens (Cambier et al., 2007; Melchers, 2005). This is mediated in part by the RAG1 and RAG2 genes which encode enzymes that are critical for V(D)J recombination events (Oettinger et al., 1990; Schatz et al., 1989). Failure of proper V(D)J recombination events result in apoptosis of pre-B cells. While we do not yet have a complete understanding of pre-B cell differentiation, the focus on key transcriptional regulators and signaling pathways integral to pre-B cell differentiation have illuminated several components that influence this process (Clark et al., 2014; Ferreir´os-Vidal et al., 2013; Johnson et al., 2008; Merkenschlager, 2010; Thompson et al., 2007).

VI. Origin and use of immune cancer cell lines

Human cancer-derived cell lines have been fundamental for the study of various cancer etiologies, useful in technology development, and for the testing of therapeutic efficacy for

developed anti-cancer agents. The HeLa cell line was the first human cultured cancer cell, derived from Henrietta Lacks in 1952 (Scherer and Syverton, 1952). The use of HeLa cells has contributed to the understanding of a wide range of fundamental biological mechanisms and more than 60,000 research publications as of 2013 (statistic from EMBL.de). One of the earliest uses of HeLa cells was in the development of the vaccine against the polio virus. The work on HeLa cells has resulted in two Nobel prizes, awarded for (i) the link between human papillomavirus and cervical cancer (Harald zur Hausen in 2008) and (ii) the role of telomerase in preventing chromosome degradation (Elizabeth Blackburn, Carol Greider, and Jack Szostak in 2011). While the use of HeLa cells in research is now controversial, it has served as a fundamental resource in science (Hyman A. & Simons K, 2011; Masters, 2002). There are now over 3,400 cancer cell lines available at the ATCC repository and they have contributed invaluable insights into both the mechanisms of cancer as well as normal development.

Compound screen studies using transplantable murine neoplasms for solid tumors were yielding poor success in the late 1980s. To address this issue, the National Cancer Institute (NCI) launched an initiative in 1990 to generate a panel of cancer cell lines called NCI-60, which was composed of 60 cancer cell lines representing nine different cancer types (Shoemaker, 2006) to standardize studies on compound efficacy. This initiative was followed with several efforts in generating cell lines across various cancers. One prime example of how cancer cell lines have shaped the understanding of both cancer and normal biology of hematopoietic cells is in the study of acute myelogenous leukemia (AML). Murine erythroleukemia cells were initially derived and used to study AML in non-human cells (Friend et al., 1971). While murine erythroleukemia cells served as an important resource, the etiology of AML in rodents does not parallel that of the human disease. In 1975, the K562 cell line was derived from the pleural fluid

of a patient with chronic myeloid leukemia in blast crisis (Lozzio and Lozzio, 1975) and widely adopted. Shortly after, a human myelogenous leukemia cell designated HL-60 was derived from the blood of a woman with AML (Collins et al., 1977) and a third myeloid cell line, known as KG-I, was derived from the bone marrow of a man with erythroleukemia (Koeffler and Golde, 1978). The use of these cell lines continues today with studies from the ENCODE consortium, whereby the use of K562 and HL-60 cells have been used extensively to catalog functional DNA elements genome-wide (Bernstein et al., 2012).

The HL-60 promyelocytic cell line has been used extensively to study the properties of granulocyte proliferation, myeloid differentiation, and malignant function since its initial characterization (Gallagher et al., 1979a), including its aberrant mutations and proliferative capability (Collins, 1987; Harris and Ralph, 1985). One of its most powerful properties, is the ability to terminally differentiate upon stimulation with chemical agents into monocyte (Mangelsdorf et al., 1984), macrophage (Murao et al., 1983), and neutrophil (Breitman et al., 1980) lineages. While its aberrant proliferative capacity is not comparable to any normal promyelocytic cell, it shares a required transcriptional program that mediates proper cellular commitment, that is also inherent in normal granulocyte differentiation (Collins, 1987).

Cell lines have served in allowing scientists' to explore several areas of biological research that span from cancer to development that would otherwise be impossible, including large scale studies of gene expression (Eisen et al., 1998). A first study comparing diffuse large B-cell lymphoma across several patients developed a gene expression microarray platform ('Lymphochip') to interrogate 17,856 cDNA clones (Alizadeh et al., 2000). Importantly, this study detailed the great diversity of gene expression among tumors of DLBCL patients, tumor proliferation rate, and differentiation state of the tumor. The term 'transcriptome' was coined for

the study of transcript expression genome-wide comparing tissues and cell lines from both mouse and human using microarrays (Su et al., 2002). Since the advent of microarrays, several other functional genomic technologies have been developed using cell lines, such as ChIP-seq (Johnson et al., 2007a), RNA-seq (Mortazavi et al., 2008) and ChIA-PET (Fullwood et al., 2009).

VII. Key transcriptional regulators of hematopoiesis

The control of cellular identity is encoded by groups of transcription factors that function in regulating downstream target genes. Within each cell of the organism, specific key transcription factors mediate this process efficiently and robustly. For example, in embryonic stem cells it has been shown that the factors OCT4, SOX2 and NANOG predominantly function as a core transcriptional regulatory triad that maintains self-renewal and pluripotency (Boyer et al., 2005). It was further shown that the core regulators above plus the transcription factors c-MYC and KLF4 were able to induce the reprogramming of fibroblasts into pluripotent stem cells (iPSC) in both mouse and human (Takahashi and Yamanaka, 2006; Takahashi et al., 2007). The ability to reprogram cell identity has revolutionized our interpretation of the 'Waddington landscape' paradigm, now emphasizing that cells exhibit dramatic plasticity during lineage commitment (Noble, 2015). Thus, a brief description of specific key transcriptional regulators and cis-regulatory function will be discussed to address plasticity in the context of the hematopoietic system.

VIII. PU.1/SPI1 transcriptional regulation

A variant of the simian virus SV40 that lacked a 72 bp enhancer region revealed that a duplicate region containing a purine-rich sequence called the PU box (5'-GAGGAA-3') (Petterson and Schaffner, 1987). This box was bound by nuclear transcription factors present only in lymphocytes and the binding was correlated with the growth of the SV40 variant. (Petterson and Schaffner, 1987). A cDNA clone for the PU.1 tissue-specific DNA binding protein was isolated (Klemsz et al., 1990). PU.1 recognizes the purine-rich sequence, 5'-GAGGAA-3' (PU box) based on both binding site analysis and DNaseI protection assays (Klemz, 1999). PU.1 expression was identified in both macrophages and B cells and the factor was shown to function as a transcriptional activator. The amino acid sequence in the binding domain of PU.1 exhibits a shared homology with the *ets* oncogene family. PU.1 has been shown to be important in the regulation of multipotent fate determination in several contexts. PU.1 directs tissue specific expression through the regulation of macrophage colony stimulating factor (M-CSF) (Zhang et al., 1994), functions as a key transcriptional toggle between erythroid and myeloid specification (Nerlov et al., 2000), and is fundamental for determining granulocyte cell commitment (Dahl et al., 2003). Importantly, the complete block on the specification of the myeloid lineage in PU.1$^{-/-}$ mice demonstrates the importance of PU.1 for the proper development of a mature hematopoietic system (Scott et al., 1994).

PU.1 binding sites have been identified in almost all myeloid-specific promoters. In addition to the regulation of M-CSF and G-CSF by PU.1, the CD11b promoter has also been shown to be co-occupied by both PU.1 and SP1 transcription factors (Chen et al., 1993). CD11b (or macrophage-1 antigen; MAC-1) leukocyte integrin subunit is expressed on the surface of human granulocytes and monocytes/macrophages coupled with the CD18 subunit. Likewise, the CD18

promoter is also occupied by PU.1 (Bottinger et al., 1994), and it was shown that cells lacking PU.1 do not exhibit expression for CD11b/CD18 in granulocytes (Olson et al., 1995).

Primary granule enzymes are a group of serine proteases expressed in human and mouse granulocytic cells. PU.1-deficient myeloid cells exhibit low expression of primary granule enzymes, which include myeloperoxidase (MPO), neutrophil elastase (NE), proteinase 3 (PR3) and lysozyme in granulocyte cells (Iwama et al., 1998). An enhancer in the second intron of the PR3 gene with a consensus ETS sequence was shown to be regulated by PU.1, functioning as an active regulatory element that mediates neutrophil elastase expression (Nuchprayoon et al., 1999). Moreover, it was shown that PU.1 cooperates with CEBPα and c-MYB to directly regulate the neutrophil elastase promoter (Oelgeschläger et al., 1996). Additionally, PU.1-deficient neutrophils do not express secondary granule components, including collagenase and lysozyme, limiting the ability to extinguish bacteria (Anderson et al., 1998). PU.1 has also been show to function as a key regulator in maintaining T helper cells (Chang et al., 2010) and a key developmental regulator in sustaining differentiation of T-cells in both human and mouse (Georgescu et al., 2008; Rothenberg et al., 2016). The ability for PU.1 to transcriptionally regulate its downstream targets in myeloid cells reflects the importance for maintaining a competent innate immune response.

In addition to PU.1 functioning as a transcriptional activator, the first evidence supporting its role as a transcriptional suppressor was observed in the regulation of the major histocompatibility complex I-Ab gene (Borras et al., 1995). PU.1 mediated repression occurs through the binding of PU.1 to its consensus sequence located at the transcription start site of I-Ab. The c-MYC promoter is a direct PU.1 target, subsequently leading to down regulation as measured by luciferase assays and a restricted region of PU.1 is responsible for this

transcriptional repression, with PU.1 also repressing the promoters of PCNA, c-FOS, and SV-40 (Kihara-Negishi et al., 2005). Interestingly, PU.1 has been found to directly interact with methyl-CpG-binding protein (MeCP) 2 via its C-terminal *ets* domain (Suzuki et al., 2006). Suzuki et al also revealed an interaction between PU.1 and DNA methyltransferase (DNMT) 3a and DNMT3b via the *ets* domain, demonstrating that PU.1 overexpression can lead to methylation of CpG sites in the p16 (INK4A) promoter in NIH3T3 cells. Similarly, it was demonstrated that MeCP2 and PU.1 occupy CpG-rich regions of the MT-1 and VIM promoters. Bisuflite sequencing analysis showed a significant methylation CpG sites for PU.1 bound regions within these promoters, also correlating with PU.1 expression levels (Imoto et al., 2010). In summary, the myeloid regulator PU.1 has been shown to function as both a transcriptional activator and repressor, controls cell fate determination in myeloid cells, and is important for maintaining innate responses, thus serving as a core regulator in all aspects of hematopoiesis.

IX. Regulation of PU.1 expression during hematopoiesis

As described above, the expression of PU.1 is critical for ensuring proper development of myeloid cells and in the propagation of several lineages within the hematopoietic system. It has been demonstrated that PU.1 can auto-regulate itself through binding of its own promoter (Chen et al., 1995). Additionally, PU.1 promoter activity can also be directed in myeloid cells by the binding of transcription factor CEBPα and AP1 complex (Cai et al., 2008). However, it was shown that the PU.1 promoter alone could not drive reporter expression, which led to further studies defining the cis-regulatory function and control of PU.1 (Cai et al., 2008).

The PU.1 upstream regulatory element (URE) approximately -14 kb was initially described to function as a myeloid-specific enhancer, that enhanced promoter activity in myeloid cells (Li, 2001). Moreover, it was shown that Runx1/AML1 bind to sites within the URE,

regulating PU.1 expression in both embryonic and adult stages of hematopoietic development (Huang et al., 2008). Work from the Rothenberg lab revealed an additional novel cis-regulatory element (-10 kb), that amplifies PU.1 expression in myeloid cells (Zarnegar et al., 2010). Lastly, it was shown that CEBPα, which binds to the URE and induces activation of an enhancer -12 kb, potentiates a secondary autoregulatory feedback loop which elicits binding of additional PU.1 transcription factors, such as early growth response 2 (EGR2) (Leddin et al., 2011).

While the regulation of PU.1 has been shown to occur via transcriptional mechanisms, evidence of post-transcriptional regulation has also been demonstrated in myeloid cells. For example, noncoding antisense RNA's were first identified as overlapping elements in the PU.1 gene coding regions (Ebralidze et al., 2008). These antisense RNAs can functionally inhibit PU.1 expression through modification of RNA translation in cells and are regulated via evolutionarily conserved cis-regulatory elements. Moreover, PU.1 is a direct target of microRNA-155 (miR-155) in activated B cells (Vigorito et al., 2007). Evidence supporting miRNA inhibition was additionally shown through the ectopic expression of miR-155, resulting in the decreased protein levels of PU.1 (Hu et al., 2010). Thus, the regulation of PU.1 is mediated via transcriptional and post-transcriptional mechanisms.

X. GFI1 and EGR regulation in myeloid cells

Growth Factor Independence I (GFI1)

Growth factor independence 1 (GFI1) was first identified as the target gene for the insertion locus of the Moloney murine leukemia virus in a leukemia T-cell line (Grimes et al., 1996a). The human GFI1 gene is encoded as a 422 amino-acid transcription factor. A small conserved SNAG (Snail/GFI1) domain was found at the N terminus, which is also present in

other transcriptional repressors (Grimes et al., 1996a, 1996b). GFI is able to bind DNA but also recruit proteins that modify DNA or DNA-bound histones, altering the accessible chromatin landscape (Zhang and Reinberg, 2001). Specifically, GFI1 bound to DNA can suppress gene expression through the recruitment of histone demethylases and histone deacetylases via a series of corepressors (McGhee et al., 2003; Montoya-Durango et al., 2008). While acting predominantly as a repressor, it has been reported that GFI1 can also activate gene expression (Duan et al., 2007; Laurent et al., 2009).

Mouse studies have shown that GFI1 is implicated in the development and function of hematopoietic stem cells (HSCs), B and T cells, dendritic cells, granulocytes and macrophages. Initially it was shown that GFI1 mediates hematopoietic stem cell quiescence and shown to be directly regulated by the tumor suppressor gene p53 (Liu et al., 2009; Zeng et al., 2004). Importantly, GFI1 has been implicated in the proper development and function of neutrophils and myeloid leukemia. In humans, rare mutations in GFI1 have been shown to cause severe congenital neutropenia (SCN) and non-immune chronic idiopathic neutropenia of adults (Person et al., 2003; Xia et al., 2009). Additional evidence supports an inability to produce granules in neutrophils when GFI1 expression is diminished (Khanna-Gupta et al., 2007). Hence, GFI1 is an important regulator in maintaining the normal development of circulating neutrophils.

In myeloid cells, GFI1 is a critical factor in the regulatory networks that specifies lineage fate decisions between monocyte/macrophage and granulocytic development. It has been shown that GFI1 and PU.1 antagonize each other's function through competitive DNA binding in promoters of target genes (Spooner et al., 2009) and protein-protein interactions (Dahl et al., 2007). GFI1 has also been shown to inhibit PU.1 by direct repression of the PU.1 promoter (Spooner et al., 2009), whereas PU.1 can enhance the expression of transcription factors that

inhibit GFI1 gene expression directly (Laslo et al., 2006). While GFI1 plays an important developmental role in myeloid progenitor cells by controlling their proliferation and differentiation, GFI1 is not required for the differentiation of myeloid progenitors towards the monocytic lineage. Although expressed in $CD14^{+}$ monocytes, GFI1 was shown to regulate genes required for granulocytic differentiation, suggesting that GFI1 likely represses the granulocytic programs in monocytes (Marteijn et al., 2007). *In vitro* differentiation experiments show that GFI1-null progenitors are incapable of developing into myeloid dendritic cells, but by default develop into macrophages (Rathinam et al., 2005). Additionally, mature macrophages show modest levels of GFI1 expression and have been shown to inhibit the production of proinflammatory cytokines (Sharif-Askari et al., 2010). In conclusion, the control of proliferation, differentiation and survival of hematopoietic cells is mediated in part by GFI1's ability to repress target genes, recruit histone modifying factors, and transcriptionally antagonize transcription factors like PU.1.

Early Growth Response (EGR) transcription factors

Early Growth Response (EGR) genes constitute a family of C2H2 zinc finger transcription factors induced under mitogenic signals in a variety of cell types such as fibroblasts and granulocytes. Family members EGR 1 through 4 have function as the immediate response factors between cell surface receptor signaling and gene expression regulation (Decker et al., 2003; Safford et al., 2005). In terms of structure, their zinc fingers exhibit extensive conservation in their respective DNA-binding domains with 90% homology, which suggests common DNA target sequences among the members of this family (Joseph et al., 1988).

Specifically, the granulocyte gene regulatory network is finely tuned by EGR, enabling control of cell fate determination between lineages (Laslo et al., 2006). Moreover, the depletion of only one EGR factor does not impair cell specification, as other expressed EGR family members were able to compensate transcriptionally (Laslo et al., 2006). Interestingly, ectopic expression of EGR1 alone in murine myeloblastic cells leads to the up-regulation AP1 transcriptional complex and IL-6 mediated inflammatory response (Krishnaraju et al., 1998). Additionally, early work in peritoneal macrophages has shown that GM-CSF promotes a 12-fold up-regulation of EGR1 mRNA within 30 minutes (Liu et al., 1991). This rapid induction in expression highlights the regulatory importance and sensitivity of the EGR-mediated response. Rapid up-regulation in EGR expression has also been observed in mouse bone-marrow derived macrophages treated with lipid A, in which EGR expression was induced dramatically within 5 minutes of stimulation (Tong et al., 2016). While EGR is has been studied for its role in determining cell fate and immune responsiveness, recent evidence highlights the importance of EGR regulation in systemic autoimmunity (Sumitomo et al., 2013). The immediate responsiveness and redundancy of these transcription factors is likely to play a key role in the robustness of myeloid gene regulatory networks.

Chapter 1 : Transcriptional and chromatin landscape dynamics during human myeloid differentiation

1.1 Abstract

The dynamics of vertebrate development are encoded in the binding of transcription factors to regulatory elements of target genes in each cell. To address the transcriptional complexity of cellular commitment during myeloid differentiation, we collected a 5-day time-series of human HL-60 cells differentiating into macrophages, monocytes, monocyte-derived macrophages, and neutrophils to measure changes in chromatin accessibility and gene expression. We detect 69,658 open chromatin elements, of which 8,919 are differentially accessible across 96 ATAC-seq datasets. We observe temporal modules of transcription factor expression and chromatin patterns during differentiation. We measure the greatest LPS-mediated changes in gene expression patterns than across the chromatin landscape. Interestingly, we find the greatest change of chromatin accessible elements in neutrophils than in other myeloid cell types. Altogether, our results highlight and emphasize the complex chromatin and expression patterns during human myeloid differentiation.

1.2 Introduction

The immune system is a complex and interactive network of diverse cell types, with a myriad of functional properties that are fundamental to maintaining an immunological-responsive balance within an organism. The coordinated organization of cellular differentiation starting from a hematopoietic stem cell is established early and maintained throughout the development of an organism, resulting in the generation of the interacting innate and adaptive immune systems. Much is known about the vast heterogeneity of surface marker expression that exists throughout hematopoietic cellular differentiation and maturation, demonstrating marker and

cellular plasticity across the adaptive (Lee et al., 2009; Zhu and Paul, 2010) and innate immune systems (Auffray et al., 2009; Ginhoux and Jung, 2014; Glass and Natoli, 2015; Lawrence and Natoli, 2011). Due to the difficulty in differentiating primary immune cells *ex vivo*, recent studies have extracted dynamics of cellular commitment from comparisons of terminally differentiated primary cells (Chen et al., 2014), still limited by sorting of cell populations.

The HL-60 promyelocytic cell line (Gallagher et al., 1979b) has been used extensively to study the properties of granulocyte proliferation, myeloid differentiation, and malignant function (Collins, 1987). HL-60 cells' aberrant mutations in the MYC locus and their proliferative capability have been compared to normal cells (Harris and Ralph, 1985). Moreover, HL-60 cells have been extensively used over the years to catalog gene expression, chromatin elements and transcription factor binding in projects such as the ENCODE Project Consortium (The ENCODE Project Consortium, 2012). Furthermore, a considerable amount of recent single-cell genomic data is also available for HL-60 (Buenrostro et al., 2015; Cusanovich et al., 2015; Farlik et al., 2015). One of this cell line's most interesting properties is its ability to terminally differentiate through the induction of chemical agents into monocyte (Mangelsdorf et al., 1984), macrophage (Murao et al., 1983), and neutrophil (Breitman et al., 1980) lineages. While its aberrant proliferative capacity is not comparable to any normal promyelocytic cell, it shares a required transcriptional program that mediates proper cellular commitment inherent in normal granulocyte differentiation (Collins, 1987).

We use the HL-60 model system to address the transcriptional complexity and the dynamics of cis-regulatory control in myeloid differentiation using a 5-day time-series analysis of genome-wide transcriptome and chromatin accessibility dynamics by probing the earliest (3 hours) to the latest (up to 168 hours depending on the cell) stages of cellular differentiation.

Additionally, we measure immediate (3 hours post-stimulus) transcriptome and chromatin accessibility immune responses mediated by LPS stimulation in cells at various stages during myeloid differentiation. Collectively, our results provide a comprehensive and high-resolution view of the dynamic immune responses, transcriptome, and chromatin landscape during differentiation.

1.3 Results

A transcriptional time course of myeloid differentiation

We induced the differentiation of HL-60 cells into three distinct lineages to measure the transcriptional and open chromatin dynamics that drive cellular commitment into macrophage, monocyte, and neutrophil lineages. We characterized the morphology of each lineage during induction of differentiation using Giemsa staining (Figure R1.1A) and also measured immune cell function using phagocytosis assays. We observed changes in cellular morphology as early as 3 hours post-differentiation, with all lineages reaching terminal differentiation by 120 hours. Furthermore, to characterize the relationship of how HL-60 derived macrophages differ epigenetically from monocyte-derived macrophages, we induced the differentiation of monocytes to macrophages (monocyte-derived macrophage) for 48 hours and measured changes in chromatin accessibility, gene expression, and function (Figure R1.1B). Moreover we included immune effects induced by LPS-stimulation at various time-points of differentiation (Figure R1.1B). All measurements were done in triplicate, resulting in 96 RNA-seq and 96 ATAC-seq datasets. This allowed us to reproducibly measure 69,658 accessible chromatin elements across human myeloid differentiation. We observed genome-wide similarities of chromatin accessibility and gene expression for closely timed-samples (3, 6 and 12 hr ; 24 and 48 hr ; 96 and 120 hr) in monocyte, neutrophil, and macrophage time-series (Figure R1.C-D, black boxes). Thus, in order

to compare cell stages during differentiation across all myeloid cell-types, we grouped time-points into early (3 to 12 hrs), intermediate (12 to 48 hrs), and late (48 to 120 hrs) temporal stages for both chromatin accessibility and gene expression (Figure R1.1D). We used these pooled stages below for the purposes of calling transcription factor footprints from the ATAC-seq data in order to build a dynamic gene regulatory network of myeloid differentiation into macrophages, monocytes, and neutrophils.

Cell-specific trajectories during differentiation reflect the transcriptional differences between macrophage sub-types

We performed a principal component analysis on the RNA-seq and ATAC-seq time-series respectively (Figure R1.2A-B). We found that the first principal component for the RNA-seq time-series explains the differences between each myeloid cell type (Figure R1.2A). Conversely, the first component for the ATAC-seq data reflects the temporal attribute of our differentiation system (Figure R1.2B). Additionally, the principal component analysis of our RNA-seq and ATAC-seq datasets for the different myeloid lineages revealed that while both directly derived macrophage and monocyte-derived macrophages look morphologically identical, they have dramatically different gene expression and open chromatin profiles. This is particularly striking given the profile similarity of early directly-derived macrophages (3hr) and late monocytes (120hr, right before differentiation into macrophages). In order to identify the candidate regulators that exhibit either shared or specific expression profiles between macrophage and monocyte-derived macrophage cells, we hierarchically clustered 901 expressed transcription factors in our system (> 1 FPKM) (Figure R1.2C). Several regulators were highly expressed in monocyte-derived macrophages compared to macrophages, which include the

nuclear orphan receptor factors (NR4A1, NR4A2, NR4A3), EGR3, FOSB, HES1 and PRDM8 genes (Figure R1.2D). A second cluster shows comparable expression in both macrophage sub-types for EGR1, EGR2, MAFB, POU2F2, RELB, NFKB2 and ATF3 regulators (Figure R1.2E). These regulators are important in the pro-inflammatory function, maturation, and differentiation of macrophages (Aziz et al., 2009; Ginhoux and Jung, 2014; Lawrence and Natoli, 2011). Recently, rapid up-regulation of EGR1, EGR2 and NR4A1 was observed in mouse macrophages within minutes of lipid A stimulation (Tong et al., 2016), which suggests a similar role in function during both stimulation and differentiation of macrophages. Interestingly, we detect expression of MAF and TEAD2 only in the latest stages of macrophage specification (Figure R1.2F), reflecting the specific differences in regulator expression between macrophage and monocyte-derived macrophages in our system.

Measuring the earliest regulatory transitions during cellular commitment

Having identified rapid changes in cellular morphology upon differentiation in our Giemsa staining assays, we focused our attention on characterizing the earliest expression and chromatin dynamics during the first 24 hours of differentiation. We first focused on genes that are up-regulated as early as 3 hours. Transcriptional regulators such as the key myeloid regulator PU.1, the macrophage-specific factor MAFB, and members of the EGR and STAT families show an increase in expression after 3 hours of macrophage induction (Figure R1.5A). We found a higher number of genes to be up regulated in macrophages (321 genes, Figure R1.5D-E) when compared to neutrophils (71 genes, Figure 5F) and monocytes (64, genes Figure R1.5G) at our earliest 3-hour time-point. The number of differentially expressed genes at 6 hours post-differentiation in neutrophils (146 genes, Figure R1.5B) and monocytes (306 genes, Figure

R1.5C) are more comparable to the number of genes up regulated after 3 hours post-macrophage differentiation (Figure R1.5E). Known markers, such as the monocyte markers CD14 and transcription factor CEBPE show increased expression in monocytes while CCR6 and B2M reflect an appropriate neutrophil-specific differentiation response. Thus, the macrophage commitment program changes the expression of a larger number of genes sooner than neutrophil and monocyte terminal differentiation does in our system.

While changes in gene expression can rapidly occur as a result of cell-signaling induced by stress or immune stimulus, chromatin accessibility dynamics depend on both transcription factor occupancy and nucleosome remodeling, which may be rate limiting. To this end, we analyzed accessibility dynamics of regulatory chromatin elements over the first 24 hours and identified a total of 893 differentially accessible chromatin regions, 112 of which lose their chromatin accessibility over the first 24 hours. The largest fraction of these differentially accessible regions occurs during macrophage specification (465 chromatin elements, Figure R1.5I), with a majority of chromatin elements becoming accessible by the 24-hour mark in all lineages. A similar enrichment of differential accessible elements 24 hours post-differentiation were also observed for neutrophil (152 chromatin elements, Figure R1.5J) and monocyte (276 chromatin elements, Figure R1.5K) lineages. These numbers represent a fraction of the overall changes in open chromatin detected in our time course, which suggests that changes in chromatin accessibility are more important for controlling the middle and later parts of myeloid differentiation programs.

Time-dependent modules of transcriptional regulation define myeloid differentiation programs

We identified 2,854 genes that were differentially expressed across our time-series and partitioned them into 13 clusters, which display both temporal and lineage-specific profiles (Figure R1.3A). We explored co-expression of genes across all clusters and report representative genes that are relevant to cell-specific differentiation programs (Figure R1.3B). Interestingly, we found that cluster 3 (324 genes, Figure R1.3B) exhibits maximal expression at 12 hours in macrophages and equivalent 132/144 hour time-point in monocyte-derived macrophage cells, while the expression profiles in neutrophils and monocytes were relatively static. This identifies a set of genes expressed with similar expression patterns, but with different relative magnitudes during the differentiation of our two macrophage sub-types, within the same time-window of differentiation. We also report cell-specific clusters that encompass markers of neutrophil (cluster 7: CFP, CLL2, S100A9), monocyte (cluster 13: CD14, GNLY), and macrophage (cluster 6: FCRLB, MERTK) lineages (Figure R1.3B). From our analysis, we identify several clusters of gene expression that capture both temporal and cell-specific patterns of cellular commitment.

From our previous analysis, we had identified several transcriptional regulators that exhibited clear temporal and lineage dynamics. This prompted us to focus on transcription factor expression in our time-series. We partitioned 232 differentially expressed regulators into 7 clusters (Figure R1.3C), and identified time-dependent modules of transcriptional regulator expression. Such modules have also been described as ‘waves’ of transcriptional regulation in a previous study of Th17 differentiation (Yosef et al., 2013). We identify immediate, intermediate, and late modules of transcription factor expression during differentiation, and assign transcriptional classes to each cluster and lineage respectively (Figure R1.3D-G). Regulators that do not exhibit a distinct transcriptional class are considered static. Transcriptional regulators in cluster 4 (for example MYC, MYB, MAZ) are uniformly assigned to the immediate class across

all lineages, as the expression of these regulators is highest in the undifferentiated state and gradually decreases across time (Figure R1.3C). The undifferentiated HL-60 cells have been shown to exhibit high-levels of amplified MYC mRNA as characterized by the presence of extra copies on a double-minute chromosome (Schwartsmann et al., 1987). As cells are induced to commit to a differentiated program, we see a down regulation of MYC and additional cell cycle genes, which suggests that differentiating cells are exiting from cellular division. Interestingly, intermediate regulators identified in cluster 6 (for example FOS, EGR, JUN) exhibit similar expression profiles in monocyte-derived macrophage and macrophage lineages during differentiation. Conversely, regulators in cluster 1 such as MAF, SOX4, and IRF2, which are in a late expression module, are expressed highest solely in the macrophage lineage. Our expression analysis found distinct time-dependent modules of transcriptional regulation that account for the distinct temporal role of key regulators in the differentiation program of our model of myeloid commitment.

Assessing chromatin landscape dynamics during differentiation and enrichment of transcriptional regulators

We addressed the genome-wide dynamics of chromatin accessibility using the same methods applied in deriving our expression results and identify 8,907 differentially accessible chromatin elements distributed in 13 clusters (Figure R1.4A). We observe a loss of differential chromatin accessibility in 2,732/8,907 (31%) chromatin elements (clusters; 7 Figure R1.4B, 9, 11) as cells were induced to differentiate, while 6,175/8,907 (69%) sites show a gain in chromatin accessibility that are shared or lineage-specific during myeloid differentiation. For example, elements from cluster 6 (n=778) show increased chromatin accessibility in both

macrophage and monocyte-derived macrophages whereas cluster 5 elements (n=429) are primarily accessible in neutrophils (Figure R1.4A-B). Interestingly, we observe a cell-specific preference for chromatin element accessibility in macrophage and monocyte-derived macrophage cells (clusters; 2 & 6 Figure 4D). The FC receptor genes (FCRLA, FCRLB), which are enriched for accessibility in macrophages and monocyte-derived macrophages, exhibit shared and cell-specific preferences for chromatin element accessibility. Importantly, the similarity in accessibility between macrophage and monocyte-derived macrophage cells shows a clear demarcation from the chromatin landscape of both differentiated monocytes and neutrophils. Lastly, we detect a subset (cluster 8, n=275) of shared chromatin elements that are accessible in all terminally differentiated cells across the time-series (Figure R1.4D).

We performed a *de novo* motif analysis on each accessible element across all 13 clusters to identify the transcriptional regulators enriched in our differentially accessible chromatin elements. We identified 21 transcription factor motifs *de novo* (significant; q-value < 0.05, highly significant; q-value < 5.0×10^{-4}) enriched in our chromatin clusters (Figure R1.4C). Motifs for MYC and E2F1 were enriched in chromatin clusters 7 and 11, which exhibit a decrease in accessibility during myeloid differentiation. Since MYC and E2F1 were identified in clusters assigned to the immediate transcriptional class in our expression analysis (Figure R1.3C), it is likely that a depletion of MYC and E2F1 occupancy occurs at these elements during cellular commitment. Additionally, we observe the PU.1 motif in 12 of 13 chromatin clusters, EGR (11 of 13), STAT (4 of 13), and IRF (8 of 13), among many other transcription factor binding site motif enrichments. Here, an initial analysis of transcription factor motif enrichment in our chromatin accessible clusters, combined with gene expression information provide a series of potential candidates to further understanding immune cell regulation.

Transcriptome and chromatin dynamics in response to LPS stimulation

We also measured immediate changes in the transcriptome and chromatin landscape of our differentiated cells after 3 hours in response to LPS treatment for macrophage (48 and 120 hours), neutrophil (120 hours), monocyte (120 hours) and monocyte-derived macrophage (168 hours) cells. We found 25 genes commonly up-regulated in monocyte, macrophage and neutrophil cells during LPS treatment (Figure R1.6D) including CXCL2, IL23A, CD44, and IL8, which have been shown previously to elicit an LPS-response in the immune system (De Filippo et al., 2013; Fujishima et al., 1993; Gee et al., 2002; Ohno et al., 1997; Sheikh et al., 2011). Gene ontology analysis demonstrates significant term enrichment for immune response (p-value < $1x10^{-7}$) and programmed cell death (p-value $5x10^{-3}$). We measured differential expression of 48 and 120 hours macrophages under LPS treatment, observing up-regulation of 115 and 210 genes respectively (Figure R1.6A, R1.7A-B) such as known LPS-responsive genes IL8, SOD2, TLR2, as well as transcriptional regulators such as MAFB and FOSL1. We then compared the differentiation states of macrophages at 48 and 120 hours treated with LPS. We find that the majority of the up-regulated genes during this comparison are markers of cellular maturation in macrophages (CD68, CSF1, FCGRT), with few up-regulated genes during LPS stimulation (Figure R1.7B). Based on the assumption that the timing of cellular differentiation for monocyte-derived macrophage cells (168 hours) and macrophages (48 hours) are equivalent, we sought to address the similarity of expression profiles based on the cell-state and LPS response. While LPS-responsive genes (SOD2, PTGS2, IL8) showed no change differential change in expression during stimulation, several up-regulated genes demonstrate changes in expression for markers defining cellular identity (Figure R1.S2C; CSF1: macrophage, CD14: monocyte). We

furthermore found 80 genes that are up-regulated in monocytes treated with LPS (Figure R1.7C), while we detect 687 genes differentially expressed in neutrophils (Figure R1.6B). Interestingly, our analysis identifies the differential expression of IL6 in neutrophils after LPS treatment, which has been widely debated (Bazzoni et al., 1991; Cicco et al., 1990; Palma et al., 1992; Zimmermann et al., 2015), while PU.1 expression is unchanged. Our results demonstrate gene expression dynamics mediated by LPS treatment that are shared and cell-specific during cellular differentiation.

We measured the effects of LPS treatment on the chromatin landscape in each myeloid lineage. We found a total 352 chromatin elements that are differentially accessible in our time-series after 3 hours of LPS treatment. A considerable fraction of differentially accessible chromatin elements are dynamic specifically in neutrophils (266/352, 75%), while only minor changes in the chromatin landscape are significant in macrophages (19/352, 5%), monocytes (6/352, ~2%) and monocyte-derived macrophages (32/352, 9%) (Figure R1.6F). We also identify a subset of elements that are shared in neutrophils and monocyte-derived macrophages that become accessible upon LPS treatment (31/352, ~9%). Because neutrophils demonstrate a significant change in the chromatin landscape during LPS treatment, we performed an analysis to integrate gene expression and chromatin accessibility profiles (Experimental methods and Figure R1.6E). We identified three clusters with dynamic profiles of chromatin accessibility during neutrophil differentiation (Figure R1.6G). Each differentially accessible chromatin element was associated to the nearest gene TSS (Experimental methods). We identified several LPS-mediated gene-element associations in neutrophil cells (Figure R1.6G). We show two examples upon LPS treatment in neutrophils; positively correlated expression and chromatin dynamics (Figure R1.S2H; NFKBIA), dynamic gene expression with a static chromatin landscape (Figure

R1.S2H', SOD2), and unchanged accessibility and gene expression patterns (Figure R1.7D, S100A9). Our analysis of LPS-treated myeloid cells demonstrates that a measured gene expression response occurs quite rapidly, while the reorganization of the chromatin landscape is both cell-specific and time-sensitive.

1.4 Discussion

We used the transcriptome and open chromatin changes in differentiating HL-60 promyelocytes to map the gene regulatory networks of human myeloid differentiation into macrophages, monocytes, and neutrophils. Our time-series of gene expression allowed us to carefully define temporal profiles of expression in connection with the timing of myeloid cell differentiation. Interestingly, we found greater changes in gene expression earlier (~3 hours) during macrophage differentiation than in monocytes and neutrophils. Immediately up-regulated genes include key transcriptional regulators such as PU.1, MAFB, and EGR that are known to drive macrophage maturation and differentiation. We observed greater cell-specific changes in expression of neutrophil and monocyte genes 6 hours post-differentiation. Conversely, while we detect substantial changes in gene expression early (3-6 hours) during differentiation, we find relatively few differential changes in the chromatin landscape until 24 hours post-differentiation in all three lineages. Instead, the majority of changes observed across the accessible chromatin landscape occurred during the middle to late temporal stages of differentiation. Our analysis of the most immediate transcriptome and accessible chromatin dynamics highlights that the earliest changes in gene expression are independent of chromatin remodeling, while intermediate and late changes in gene expression are likely dependent on chromatin organization, enzymatic histone modifications, and the exact levels of transcription factors during cell differentiation.

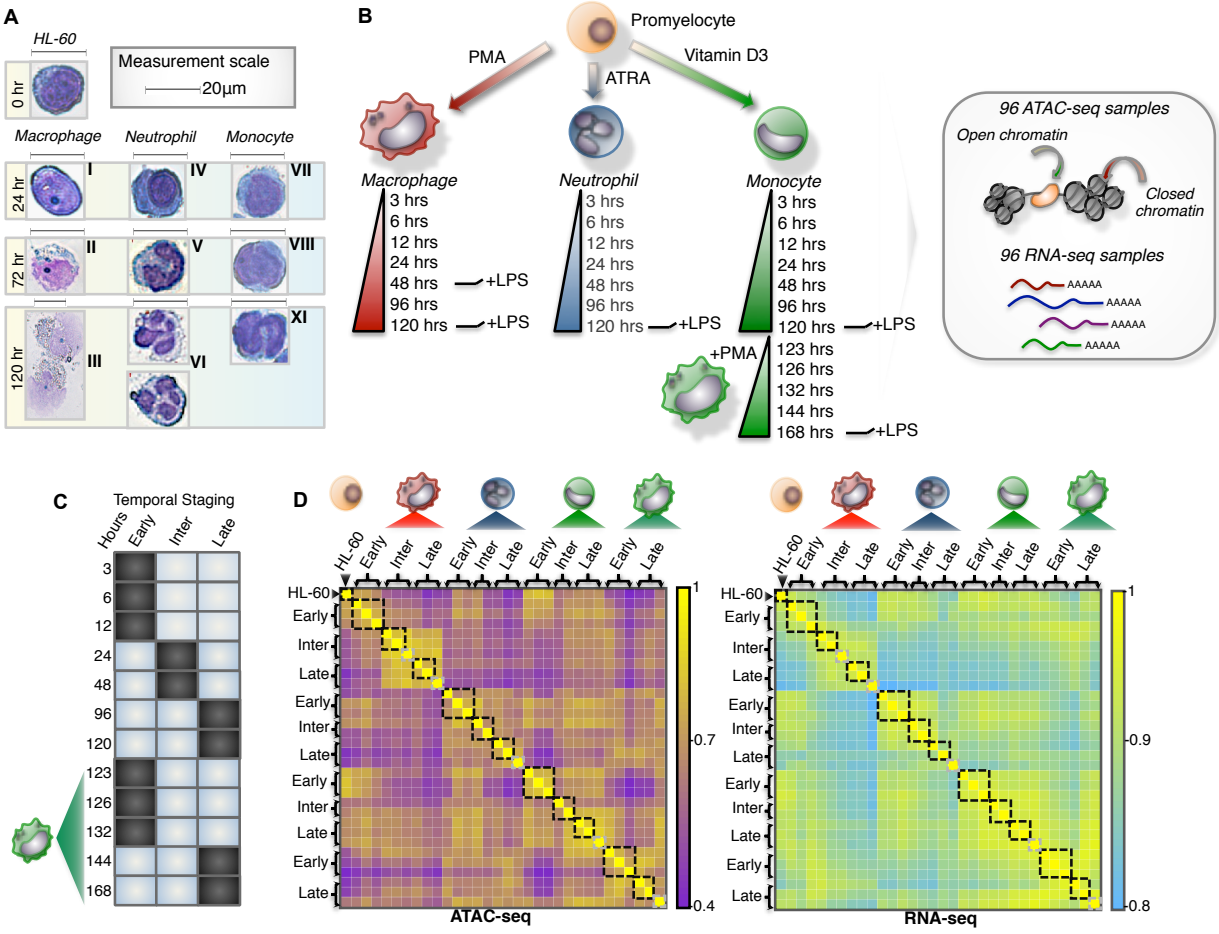
Previously, blood monocytes were believed to predominantly give rise to all resident macrophages in adult tissues (van Furth and Cohn, 1968). Recent evidence now challenges this paradigm and offers insight about macrophage differentiation in the context of both tissue maintenance and development (Ginhoux and Jung, 2014). While monocytes contribute most to the tissue macrophage compartments during inflammation and in the steady state, adult tissue macrophages are additionally derived from embryonic precursors before birth (Ginhoux and Jung, 2014). We leveraged our human myeloid system to study and biologically simulate the derivation of macrophages in the context of adult specification and embryonic development; a monocyte-derived macrophage and an HL-60 derived macrophage. Our analysis identified clusters of genes and chromatin elements shared between monocyte-derived macrophages and macrophages (Figure R1.3, R1.4, R1.6), demonstrating similarities in regulatory components that are distinct from monocytes and neutrophils. However, our PCA analysis demonstrates that the cell trajectories measured during differentiation are quite distinct for the two macrophage subtypes with respect to the transcriptome and chromatin landscape, even though both differentiation events were triggered by the same stimulus (PMA), resulting in subtype-specific expression for key transcriptional regulators (Figure R1.2). Furthermore, an analysis comparing LPS-mediated gene expression changes for equivalent time-points of macrophage (48 hours) and monocyte-derived macrophages (168 hours) showed that most of the differentially expressed genes were mediators of macrophage or monocyte specification. While the monocyte-derived macrophage cells exhibit phagocytosis functionality and cellular morphology consistent with macrophages (Figure R1.1), it is still possible that these cells may not have matured fully or reached a terminally differentiated state, as they still show expression for CD14. While the time

required for differentiation of adult macrophages is not precisely known, our analysis provides a temporal paradigm of the earliest events specified during macrophage differentiation.

The induction of LPS on each myeloid cell type ultimately revealed greater changes in gene expression than across the chromatin landscape. Interestingly, we observed the greatest change in chromatin accessibility for neutrophils 3-hours post LPS stimulation. Our analysis of coordinated changes in expression with changes of accessibility of nearest regulatory elements revealed several LPS responsive regulatory elements specified in neutrophils. The structure of neutrophil nuclei as segmented lobes offers an interesting paradigm in understanding how the chromatin landscape is organized in such a complex and convoluted nuclear assembly.

1.5 Figures

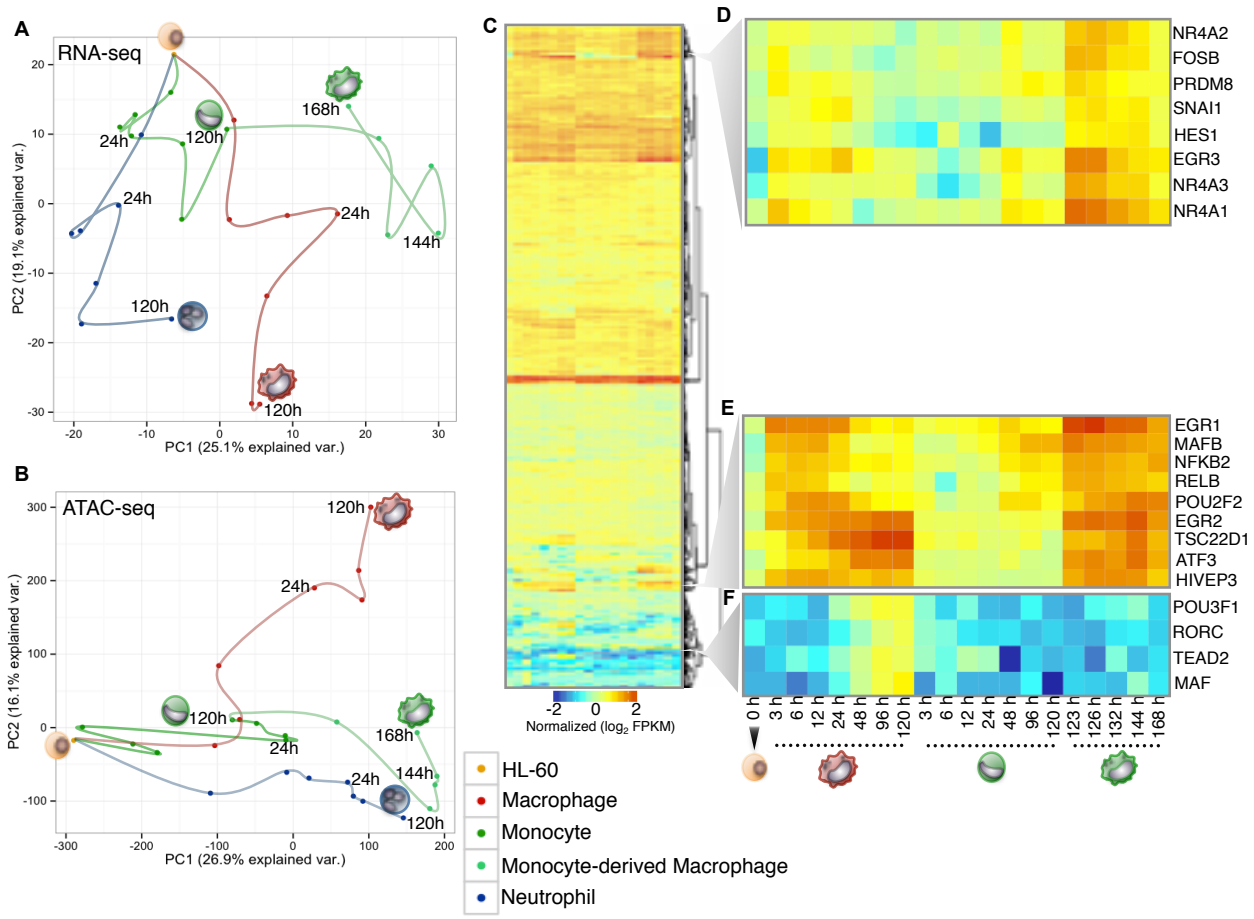
Figure R1.1 A dynamic model of human myeloid differentiation using HL-60s



(A) HL-60 directed differentiation of neutrophil, monocyte and macrophages show cell intermediates at varying time-points using Giemsa staining. Intermediate progenitors were characterized based on morphology during the course of differentiation and categorized as: (I) immature macrophage, (II) macrophage < 20µm, (III) macrophage > 20µm, (IV) myelocyte, (V) banded neutrophil, (VI) segmented neutrophil, (VII) monoblast, (VIII) promonocyte, (IX) monocyte. Scale indicates 20µm. (B) Schematic outline of study design. Colored cell identifier denotes myeloid cell-types. (C) Temporal staging of time-points for each cell-type was based on the clustering of RNA-seq and ATAC-seq data. (D) Genome-wide clustering reveals inherent

structure of the transcriptome and chromatin landscape during myeloid differentiation (Pearson correlation). Black boxes mark samples grouped as temporal stages. Grey boxes indicate LPS induced time-points.

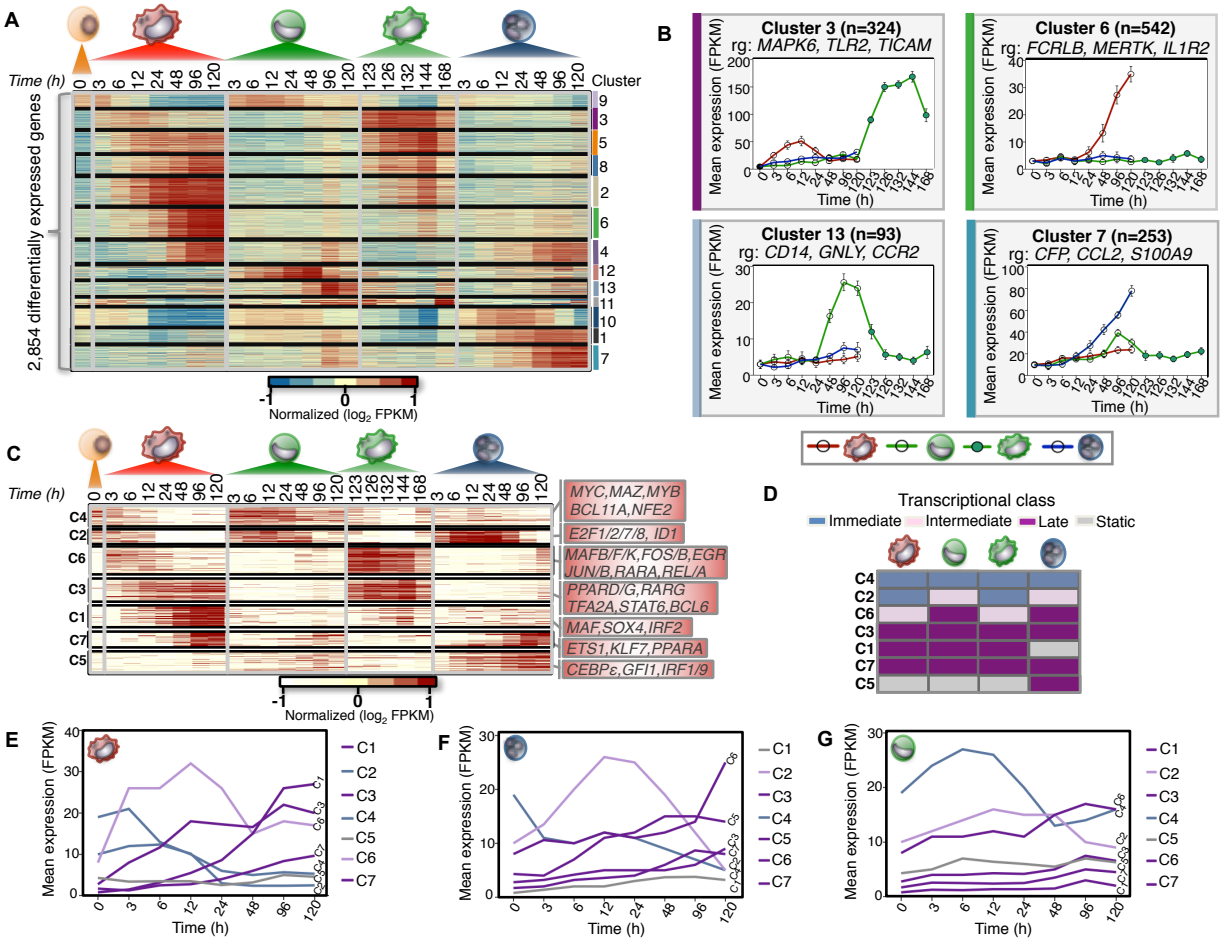
Figure R1.2 Distinct cells during myeloid differentiation



(A and B) Principal component analysis of RNA-seq and ATAC-seq time-series. Time-points were connected serially to illustrate cell-specific trajectories. Cell-types are labeled with distinct colored points. (C) Heatmap of 902 expressed transcription factors during macrophage, monocyte, and monocyte-derived macrophage differentiation. Each column represents the expression for a time-point, whereas each row represents a transcription factor. RNA-seq data is row-mean normalized and row clustered using Euclidean distance. (D) Representative cluster of transcription factors showing the highest expression in monocyte-derived macrophages. (E) cluster of transcription factors that demonstrate similar expression patterns between macrophage

and monocyte-derived macrophage time-points. (F) representative cluster of transcription factors expressed specifically in macrophage cells.

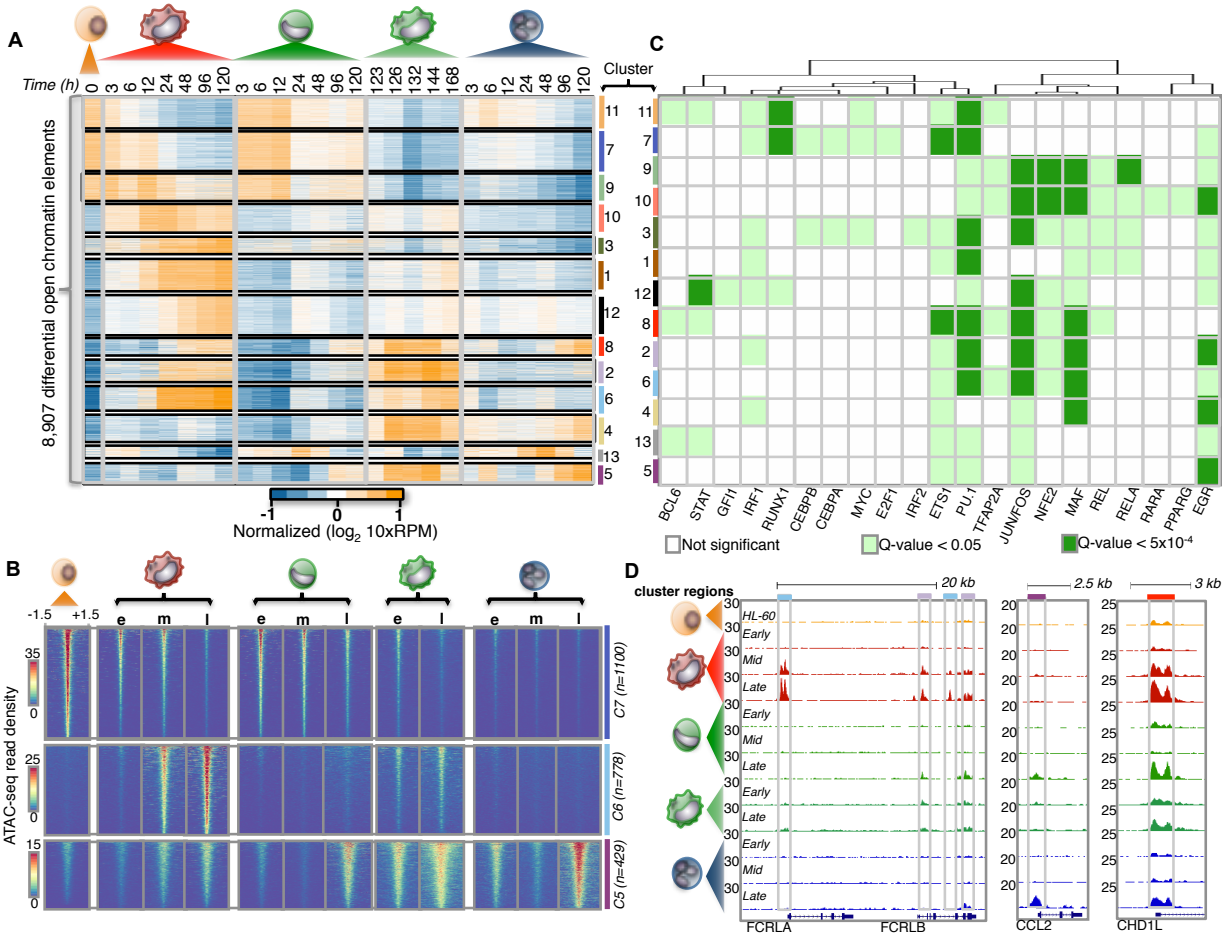
Figure R1.3 Temporal modules of transcriptional regulator expression



(A) Heatmap of 2,854 differentially expressed genes during myeloid differentiation (α=0.05, FDR < 1%, p-val < 0.01). 13 expression clusters were derived using k-means and denoted by both color and number for all differentially expressed genes. FPKM values log transformed and row-mean normalized for all genes. (B) Gene expression profiles for clusters (3, 6, 7, and 13). Mean expression and standard deviation (error bars) are plotted across all biological replicates and for all genes in each corresponding cluster. Cluster size (n) and representative genes (rg) are denoted. (C) Heatmap of 232 differentially expressed transcriptional regulators (α=0.05, FDR < 1%, p-val < 0.01). Seven regulator clusters were derived using k-means (C1-C7). Representative

regulators from each cluster are shown respectively. FPKM values are log transformed and row-mean normalized for all regulators. (D) Schematic of transcriptional regulator classification. Transcriptional classes are denoted as immediate (blue), intermediate (pink), late (Purple) and static (Gray). (E, F and G) Transcriptional profiles for each cluster based on classification for macrophage, neutrophil and monocyte cells respectively. Mean FPKM expression is shown for each cluster.

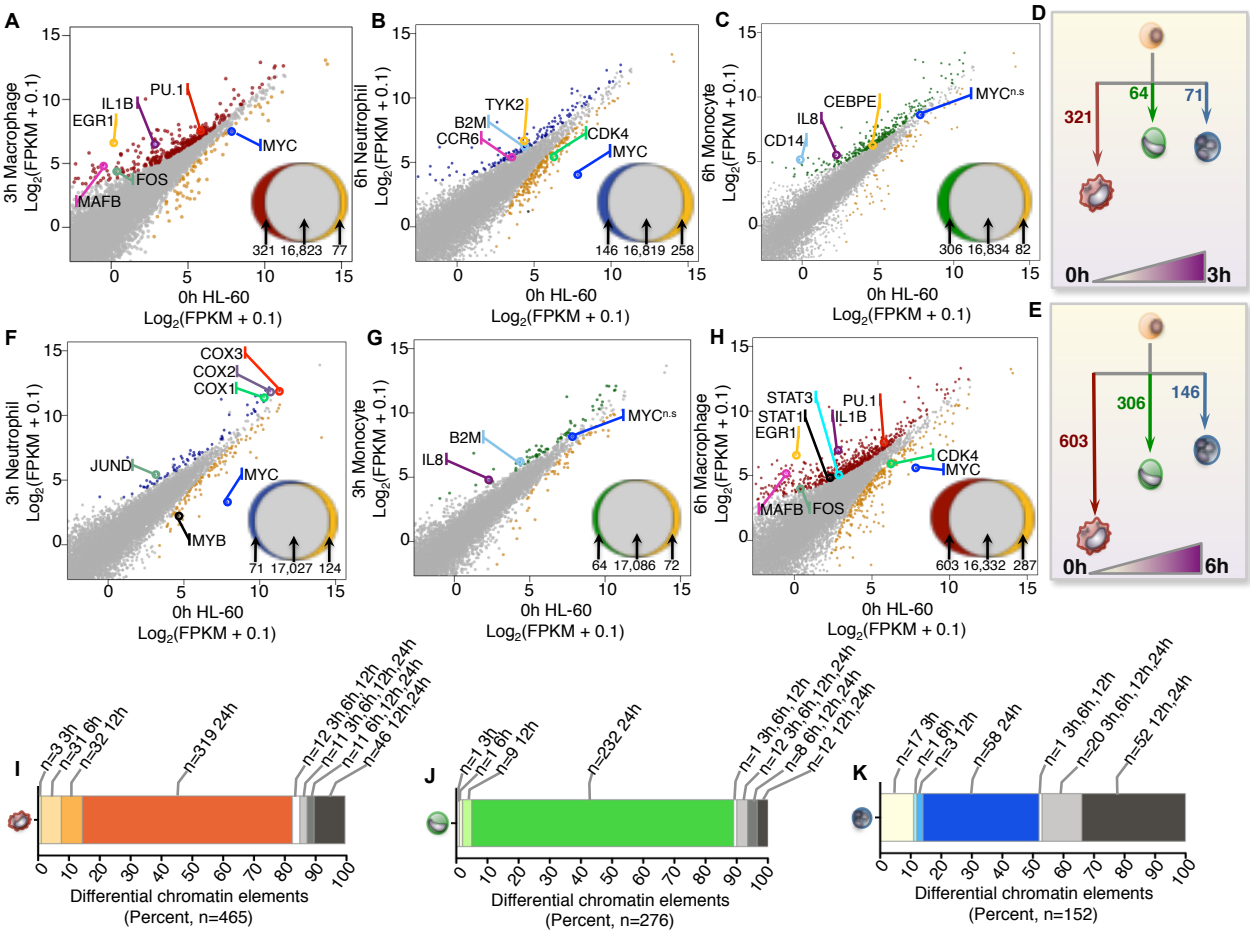
Figure R1.4 Temporal modules of differential chromatin accessibility



(A) Heatmap of 8,919 differential accessible chromatin elements during myeloid differentiation (α=0.05, FDR < 1%, p-val < 0.01). 13 expression clusters were derived using k-means and denoted by both color and number for all chromatin elements. 10xRPM (Reads Per Million) values are log transformed and row-mean normalized. (B) Heatmaps showing ATAC-seq read density of chromatin elements from three clusters (5,6, and 7). Mean ATAC-seq read density was derived using temporal stages (e:Early, m: Middle, l: Late). ATAC-seq signal is shown for a window of +/-1.5 kb from the chromatin element center and ranked from strongest to weakest for all comparisons. Cluster size (n) is denoted. (C) Heatmap of *de novo* motif transcription factor enrichment. Rows indicate cluster of chromatin elements mined for motifs, while columns indicate transcription factor motif of interest. Transcription factor motifs were hierarchically

clustered based on significance using a Euclidean distance. Non-significant motifs are represented as white boxes. Motif significance is shown for a q-val < 0.05 and q-val < $5x10^{-4}$ denoted by light or dark green boxes respectively. (D) Examples of chromatin element clusters specified during differentiation. Browser tracks of ATAC-seq data for all cell-types are normalized by read density. Chromatin elements from two differing cluster profiles reflect the complex regulatory diversity (left browser panel) during myeloid differentiation. Cell-specific chromatin accessibility is strongly enriched in neutrophils (middle panel), while temporal changes in chromatin element accessibility can be observed across all cell-types (last panel). Colored boxes identify with chromatin cluster.

Figure R1.5 Characterizing immediate changes in expression and chromatin accessibility

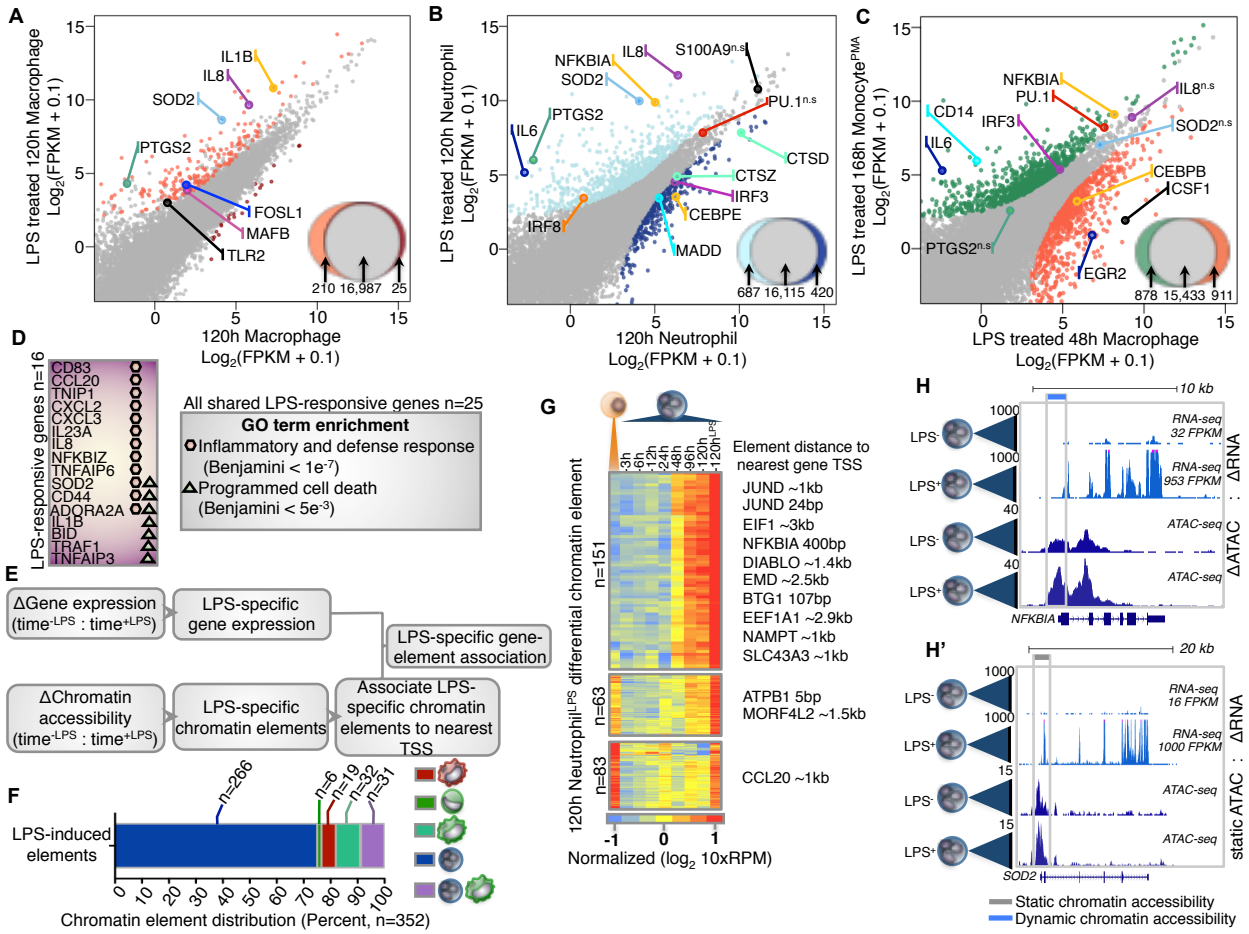


(A) Differential gene expression changes between undifferentiated HL-60 and 3 hour macrophages (321 genes), FDR < 1%, p-value < 0.01. (B and C) Differential gene expression of undifferentiated HL-60 and 6 hour neutrophils (146 genes) and monocytes (306 genes), FDR < 1%, p-val < 0.01.Representative genes are denoted by color.

(D and E) Schematic of immediate changes in gene expression. Branch lengths in diagrams represent the number of differentially expressed genes up regulated for each cell-type at 3 and 6 hours post-differentiation. (F) Differential gene expression changes between undifferentiated HL-60 and 3 hour neutrophil (71 genes), FDR < 1%, p-value < 0.01.(G) Differential gene

expression changes between undifferentiated HL-60 and 3 hour monocyte (64 genes), FDR < 1%, p-value < 0.01. (H) Differential gene expression changes between undifferentiated HL-60 and 6 hour macrophage (603 genes), FDR < 1%, p-value < 0.01. (I, J and K) Differential chromatin accessibility of early (0-24 hours) macrophage (n=465 chromatin elements), neutrophil (n=152 chromatin elements), and monocyte (n=276 chromatin elements) differentiation (p-val < 0.05). Chromatin accessibility dynamics are observed across several time-points and denoted respectively.

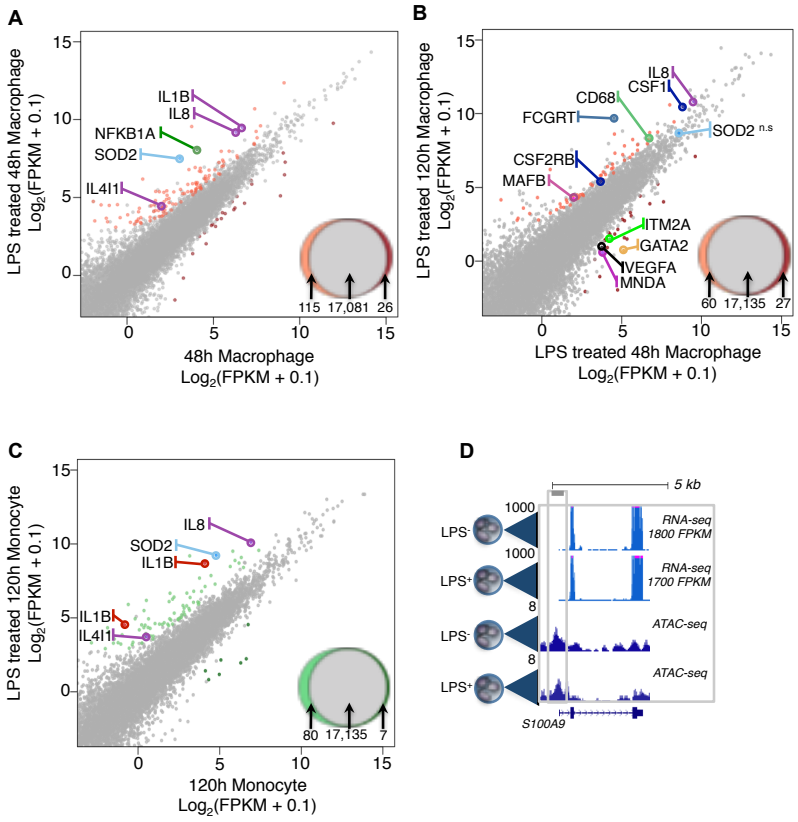
Figure R1.6 Immediate effects of LPS-stimulation are dynamic across the transcriptome and chromatin landscape during myeloid differentiation



(A and B) Differential gene expression changes between 120-hour macrophages/neutrophils and 120-hour macrophages^{+LPS} (210 genes)/neutrophils^{+LPS} (687 genes), FDR < 1%, p-value < 0.01. Representative genes are colored and labeled for identification. FPKM values were log transformed. (C) Differential gene expression changes between 48-hour macrophages^{+LPS} (911 genes) and monocyte-derived macrophages^{+LPS} (878 genes), reflecting a majority cell-specific expression changes relative to LPS-mediated differences (FDR < 1%, p-value < 0.01). FPKM values were log transformed and n.s denotes not significant. (D) Diagram of LPS-mediated genes

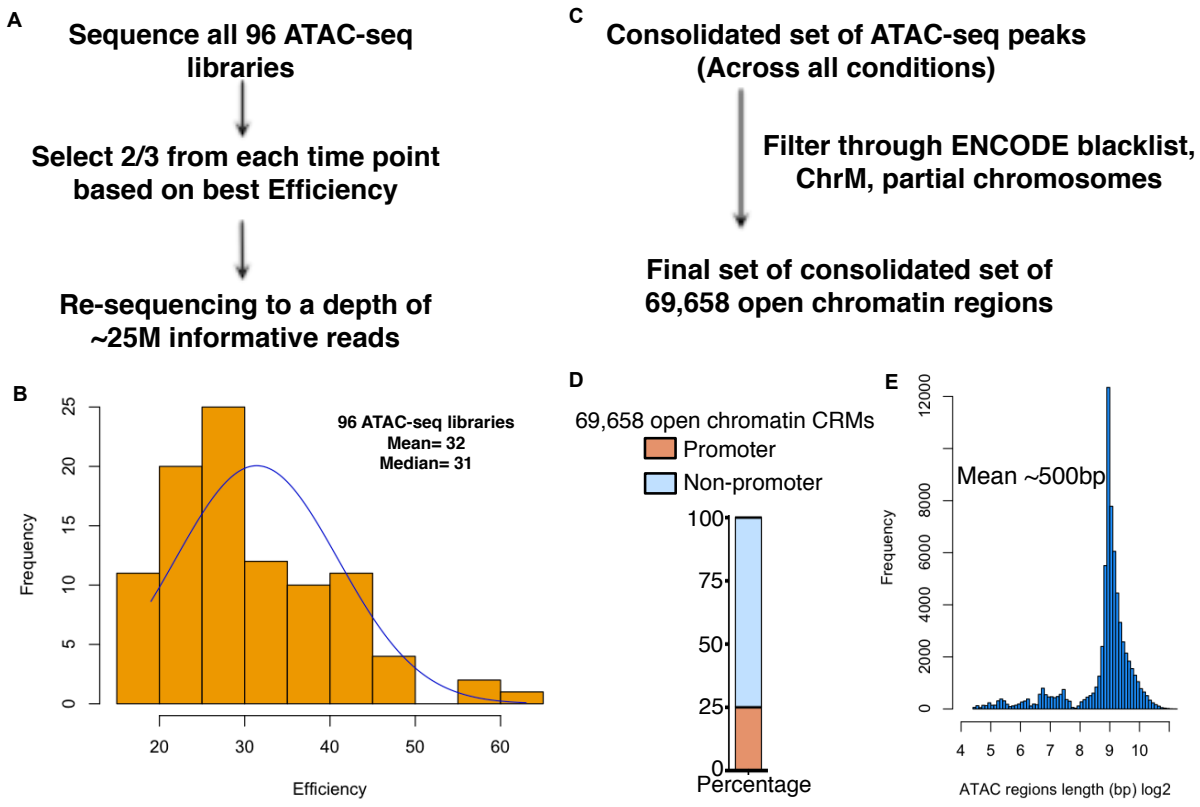
up regulated and shared across all LPS treated time-points. Gene ontology analysis shows significant enrichment in terms for 16/25 genes. Red hexagon denotes enrichment in inflammatory and defense response (Benjamini < $1x10^{-7}$). Green triangle denotes enrichment in programmed cell death (Benjamini < $5x10^{-3}$). (E) Schematic design for integrative analysis of LPS-mediated changes in gene expression and chromatin accessibility. See experimental methods. (F) Cell-specific changes in differential chromatin accessibility across all myeloid cell-types, FDR < 1%, p-value < 0.01 (n=352 chromatin elements). Cluster size is denoted as (n) and colored boxes reflect cell-specific or combinatorial representation. (G) Heatmap of 237 differentially accessible LPS-mediated chromatin elements in neutrophils. Three chromatin clusters were derived using k-means. 10xRPM (Reads Per Million) values are log transformed and row-mean normalized for all data. Differential chromatin elements were associated to nearest genes (schematic, E) and shown with approximate chromatin element distance to gene TSS. Cluster size (n) is denoted. (H and H') Examples of LPS-mediated changes in chromatin accessibility and gene expression in neutrophils. Browser tracks of ATAC-seq data for neutrophils are normalized by read density. RNA-seq FPKM values are indicated for each gene and treatment. Colored boxes denote chromatin accessibility as static or dynamic. The change in gene expression and chromatin accessibility are well correlated for gene NFKBIA during LPS stimulation (top panel). Significant change in gene expression is observed for SOD2 without a significant change in promoter accessibility (bottom panel).

Figure R1.7 LPS-mediated gene and open chromatin changes



(A) Differential gene expression changes between 48-hour macrophages and 48-hour macrophages^{+LPS} (115 genes), FDR < 1%, p-value < 0.01. (B) Differential gene expression changes between 48-hour macrophages^{+LPS}(27 genes) and 120-hour macrophages^{+LPS} (60 genes), FDR < 1%, p-value < 0.01. (C) Differential gene expression changes between 120-hour monocytes and 120-hour monocytes^{+LPS} (80 genes), FDR < 1%, p-value < 0.01. Representative genes are colored and labeled for identification. FPKM values were log transformed. (D) UCSC browser screenshot for S100A9 showing no change in gene expression and chromatin accessibility during LPS stimulation. ATAC-seq data for neutrophils are normalized by read density. RNA-seq FPKM values are indicated for each gene and treatment.

Figure R1.8 ATAC-seq data quality and peak calling strategies



(A) Sequencing strategy for 96 ATAC-seq libraries. Deep sequencing was performed on samples with best efficiency measurement (fraction of mapped reads in peaks/ total mapped reads). (B) Histogram of ATAC-seq library efficiencies. Calculated mean is 32%, median is 31%. (C) Pipeline for removing blacklist regions and generating consolidated peak list. (D) Distribution of ATAC-seq peaks that overlap promoter and non-promoter elements genome-wide. (E) Histogram of ATAC-seq peak lengths for consolidated peaks (n=69,658).

1.6 Methods

Experimental design

HL-60 cells (ATCC) were grown in Modified Dulbecco's Medium in a final concentration of 20% FBS with penicillin antibiotics (1%). Cells were routinely cultured at a density of $1x10^6$ cells/ml. Differentiation of HL-60 cells into macrophage (Murao et al., 1983), monocytes (Mangelsdorf et al., 1984), and neutrophils (Breitman et al., 1980) was performed as previously described. Additionally, monocytes (120 hours post-differentiation) were stimulated with PMA to induce differentiation into monocyte-derived macrophages. LPS stimulation at 48, 120, and 168 hours for specific cell-types was induced at a final concentration of 100ng/ml for ~3h, immediately followed by expression and chromatin analysis. Cells were induced to differentiate from three biological HL-60 cell culture growths over the course of 5-7 days. For each time-point, differentiating cells were collected for both RNA-seq (~2-3 million cells) and ATAC-seq (~50,000 cells) from the same treated cells. All biological replicates were collected for the same time-point simultaneously and libraries were generated from each single biological replicate were processed together for both RNA-seq and ATAC-seq measurements. We assayed a total of 13 cell-specific time-points upon myeloid differentiation. This resulted in the generation of 96 ATAC-seq and 96 RNA-seq datasets.

Morphological and functional characterization of myeloid cells

Cell morphology was profiled during myeloid differentiation using an optimized Giemsa staining procedure. Cells were imaged using the Zeiss Observer at 40X magnification. To test for cell phagocytosis, HL-60 and differentiated cells were tested with the Cayman Phagocytosis Assay kit (IgG FITC) at various time-points of differentiation. Phagocytosis was tracked through

interval imaging (5-10 seconds) for a total of 30-60 minutes using the EVOS FL. Videos were compiled from images and assessed for active cell phagocytosis and movement.

Tandem gene expression and chromatin profiling

Approximately ~2 million HL-60 and differentiated cells were collected for RNA-seq and 50,000 cells were harvested for ATAC-seq for each replicate of each time point. Briefly, cells were assessed for cell viability, counted, and washed with PBS. RNA-seq was performed as previously described (Mortazavi et al., 2008). ATAC-seq was performed as previously described (Buenrostro et al., 2013)with the addition of a DNA size selection step after library generation to enrich for accessible chromatin ranging from 100-400bp. RNA-seq libraries were sequenced as single-end 86bp and ATAC-seq libraries as paired-end 43bp reads on the Nextseq 500 Illumina platform. Approximately 1 billion RNA-seq and 2 billion ATAC-seq reads were generated.

Gene expression analysis of myeloid differentiation

RNA-seq reads were mapped to the hg38 reference genome using STAR (Dobin et al., 2013) aligner and mapped to Gencode version 20 gene annotations using Cufflinks (Trapnell et al., 2010). Batch effects due to the generation of libraries were considered and corrected for using Combat (Johnson et al., 2007b). Batch-corrected data was normalized using TMM function in EdgeR (Robinson et al., 2010). maSigPro (Nueda et al., 2014) allows for a two-step regression modeling strategy, which was used in identifying gene expression dynamics across differentiation of all lineages. An alpha of 0.05 for multiple hypothesis testing and a false discovery control of 1% were used, in both gene and transcription factor analysis. A k-cluster of 13 was selected based on previous analysis using hierarchical clustering and k-means clustering on the entire dataset. Gene ontology enrichments were determined for each cluster using DAVID

(Huang et al., 2007). Gene expression heatmaps were generated using Tree View 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm) and using R software. PCA analysis was performed using R.

ATAC-seq data processing and analysis

ATAC-seq reads were mapped to the hg38 reference genome using bowtie (Langmead et al., 2009). HOMER was used to call open chromatin regions across all replicates. A dual selection strategy was employed for calling open chromatin regions using HOMER. Briefly, HOMER was run on a size setting to enrich for 'narrow' regions (120-150bp) then sequentially run on size selecting 'broad' regions (500bp) at an FDR 1%. 'Narrow' and 'broad' regions were then merged into a single bed file for each replicate. This allowed for an enrichment of diverse accessible chromatin elements. A region identified in all three biological replicates was then considered as a biologically reproducible specific open chromatin region. A master set of open chromatin regions was generated by consolidation of all peaks identified across all time-points. ENCODE 'blacklist' regions and open chromatin regions mapping to ChrM were discarded from our analysis. ATAC-seq data quality was measured using a sample efficiency measurement: fraction of uniquely mapped reads in peaks divided by all uniquely mapping reads. The mean efficiency for all ATAC-seq data was 32%, ranging from 25%-65% for all samples. Read coverage was estimated for each open chromatin region and normalized by sample size and efficiency to detect changes in open chromatin dynamics across all time-points. Data was then corrected for batch effects using Combat and normalized using the TMM function in EdgeR. maSigPro was used to identify open chromatin dynamics across the time-series (alpha=0.05, FDR=1%). Chromatin heatmaps were generated using Deeptools software (Ramírez et al., 2014). PCA analysis was performed with ATAC-seq counts using R functions.

Transcription factor motif enrichment

Chromatin clusters were mined for *de novo* motifs using EXTREME (Quang and Xie, 2014). Chromatin elements were converted to fasta format using a masked hg38 genome reference. Fasta dinucleotide shuffling was performed on masked data (fasta-dinucleotide-shuffle.py) and k-mer search (GappedKmerSearch.py). PWM's were generated (Consensus2PWM.py) and used to identify motifs *de novo* (EXTREME.py). Motifs similarities were quantified using TOMTOM (Gupta et al., 2007) and recovered based on significant q-values.

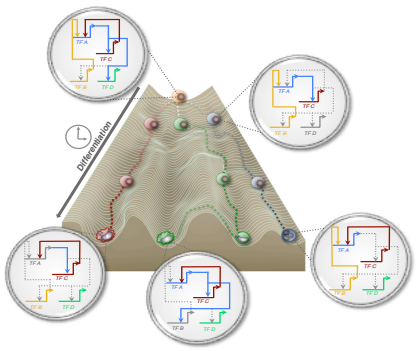
LPS-stimulation analysis

Corresponding time-points that included LPS stimulation were analyzed for both differential chromatin and gene expression comparisons. EdgeR was used to call differentially expressed and accessible chromatin elements with an FDR <1%, p-value < 0.05 and fold change > 1.5. Hierarchical clustering was performed using Euclidean distance on neutrophil chromatin elements. Association of LPS-specific accessible chromatin elements to LPS-specific identified genes was performed using GREAT (McLean et al., 2010). Gene-element associations were then filtered to a distance of 3kb from element to TSS.

Chapter 2 : Dynamic gene regulatory networks of human myeloid differentiation

2.1 Abstract

Hematopoiesis requires the coordinated regulation of several transcription factors that define the regulatory circuits driving cellular specification. Here we focus on the transcriptional and chromatin landscape of human myeloid differentiation using a 5-day time-series of promyelocyte differentiation into macrophage, neutrophil, monocyte, and monocyte-derived macrophages to derive dynamic gene regulatory networks. We identify temporal modules of transcriptional regulator expression and integrate with open chromatin footprinting to build dynamic gene regulatory networks of 23 key immune transcription factors. We find from our gene regulatory networks that immune transcription factors like PU.1 only regulate a subset of their targets at a specified time, via a cell-dependent manner during differentiation. Interestingly, we derive subnetworks of immune regulatory interactions for transcription factors EGR and GFI1, revealing that their complex regulatory circuits are configured in a cell-dependent manner. Additionally, we observe diverse VDR cis-regulatory architecture in macrophages and monocytes, which are differentially specified by key regulators during cellular differentiation. Altogether, our simple model reveals the underlying complexity of regulatory circuits, transcriptional dynamics, and cell intrinsic programs required for human cells of the myeloid branch.



2.2 Introduction

Vertebrate developmental commitments are implemented within cells through remodeling of chromatin accessibility of promoter and enhancer cis-regulatory modules (CRMs) across the genome to allow for transcription factor binding. The identification of CRMs is critical to understanding the complexities of gene regulatory circuits within the genome in a variety of organisms (Hardison and Taylor, 2012; Peters and Davidson, 2015; Wray, 2007). A powerful application of open chromatin assays such as ATAC-seq and DNase-seq is the derivation of transcription factor footprints. DNaseI footprinting has been used to identify transcription factor occupancy (Hesselberth et al., 2009; Neph et al., 2012a) and to derive transcriptional networks in many biological contexts (Neph et al., 2012b; Stergachis et al., 2014; Sullivan et al., 2014). Recently, ATAC-seq was also applied to characterizing transcription factor regulation in the mammalian brain (Mo et al., 2015) and identifying variation in primary T cells (Qu et al., 2015). There has been relatively less work in incorporating open chromatin directly in a dynamic gene regulatory network (GRN). Sullivan et al. characterized the light/dark time-specific dynamics through the generation of chromatin interaction networks in *A. thaliana* (Sullivan et al., 2014). Yet all GRNs are by their very nature dynamic and should ideally capture the many steps of differentiation that have been described in well-defined systems such as T-cell development (Zhang et al., 2012b).

Several studies have looked at genome-wide chromatin accessibility dynamics in myeloid cells such as tissue macrophage populations (Lavin et al., 2014), profiling of terminally differentiated immune cells (Lara-Astiaso et al., 2014), chromatin dynamics of macrophages and monocytes (Gosselin et al., 2014; Saeed et al., 2014), immune-responsive late enhancers in macrophages upon stimulation (Ostuni et al., 2013), and immediate responses to lipid A

stimulation in mouse macrophages (Tong et al., 2016). Another study measured the chromatin accessibility landscape of neutrophil populations (Wong et al., 1999). While studies have investigated gene networks in several immune cell-types (Clark et al., 2014; Georgescu et al., 2008; Spooner et al., 2009; Yosef et al., 2013; Zhang et al., 2012b), dynamic gene regulatory networks during myeloid differentiation have not been previously described. Studies have focused on changes in histone modifications and transcription factor binding by comparing terminally differentiated cells, but a comprehensive view of how chromatin elements gain or lose accessibility throughout myeloid differentiation has not been published and would by default not capture any intermediate stages for which appropriate markers have not yet been identified. An unbiased and appropriately controlled model system is needed to probe changes in gene expression and the chromatin landscape for the purpose of mapping the regulatory networks that act during myeloid differentiation.

We use the HL-60 model system to address the transcriptional complexity and the dynamics of cis-regulatory control in myeloid differentiation using a 5-day time-series analysis of genome-wide transcriptome and chromatin accessibility dynamics by probing the earliest (3 hours) to the latest (up to 168 hours depending on the cell) stages of cellular differentiation. We use gene expression to select dynamic regulators and chromatin footprinting in our 192 datasets to generate draft gene regulatory networks of macrophage, monocyte and neutrophil differentiation. We identify both previously described transcription factor interactions as well as novel immune regulatory circuits. Collectively, our results provide a comprehensive and high-resolution view of the dynamic transcriptome, chromatin landscape, and complex immune gene regulatory network driving human myeloid differentiation.

2.3 Results

Building myeloid gene regulatory networks using chromatin footprinting data

Based on the transcription factor modules and *de novo* motif analysis of our differentiated myeloid cells, we had observed several key immune regulators to display distinct cell and temporal patterns of expression. We focused our analysis on 23 transcriptional regulators which displayed dynamic expression patterns and whose DNA motifs were significantly enriched across myeloid differentiation. Gene expression differences for the 23 regulators seemed to vary across early, intermediate, and late stages of differentiation (Figures R2.1-4). For example, the regulator MAF was expressed in the later stages of the promyelocyte derived macrophages (Figure R2.2), whereas GFI1 expression was most differential in neutrophils (Figure R2.3). Interestingly, we observed dramatic and dynamic kinetics in expression for EGR transcription factors (Figures R2.1-4).

We leveraged our ATAC-seq datasets by merging biological replicates and time-points based on temporal staging to achieve ≥ 200 million reads for chromatin footprinting analysis. We identified an average of 85k footprints per time-point, resulting in more than 1.02 million chromatin footprints across our time-series. To infer transcriptional interactions using chromatin footprints, we generated a gene regulatory network of footprints in the promoter and proximal enhancers of key differentially regulated transcription factors (Experimental methods), allowing us to identify temporal and cell-specific regulatory dynamics during cellular differentiation (Figure R2.5A). Our strategy captured well-known myeloid regulators such as PU.1, EGR, GFI1, and CEBPα among others. A ‘genome-view’ (Peters and Davidson, 2015) of all identified interactions derived from our networks, consisting of 23 regulators with 158 interactions is illustrated as a circuit diagram (Figure R2.5B). One key difference between our draft GRNs and

classical, perturbation-based GRNs is that our footprinting alone does not allow us to assign activating or repressing activity to footprints. We evaluated the precision of our networks in identifying PU.1 interactions with published HL-60 PU.1 ChIP-seq data. Approximately ~90% of our PU.1 derived footprints are supported by PU.1 ChIP-seq data in undifferentiated HL-60 cells (Figure R2.6A-B). Furthermore, all five PU.1 edges in our HL-60 gene regulatory network were also supported by PU.1 ChIP-seq data. We show gene regulatory networks for macrophage, monocyte, neutrophil, and monocyte-derived stages of differentiation for all 23 transcriptional regulators (R2.7-11). Several other linkages in our GRNs were previously identified in myeloid cells (Figure R2.12), which suggest that our networks capture known regulatory interactions along with many new candidate interactions.

Having observed time-dependent modules of expression for many transcriptional regulators, we investigated the changes in our draft, footprinting-based GRNs during early, intermediate, and late time points of differentiation in macrophage, neutrophil, monocyte, and monocyte-derived macrophage differentiation based on the grouping of time-points into temporal stages. One example out of many dynamics in our footprinting GRN is the behavior of the vitamin D receptor (VDR) gene regulatory elements. VDR has been shown to elicit and modulate immune responses in several cell types including macrophage and monocyte cell-types (Aranow, 2011; Baeke et al., 2010; Mora et al., 2008). We observed several dynamic regulatory inputs at the VDR promoter element in both monocyte and macrophage gene regulatory networks. Interestingly, we detect early regulation of VDR by MYC, CEBPα, PPARG, and PU.1 during monocyte differentiation, followed by a subsequent loss of CEBPα and MYC regulation in the intermediate time-points and maintained regulation for both PU.1 and PPARG in the late stages of differentiation (R2.13B). This dramatically contrasts the level of regulatory control

during macrophage differentiation, whereby the VDR promoter is regulated in the late stages of macrophage differentiation, occupied by PU.1, STAT, NFE2, MAF, MAFB, STAT and JUND regulators (Figure R2.13A). We also observe cell-specific expression patterns for the immune regulators MAF and MAFB, which are important for monocyte and macrophage differentiation and self-renewal (Aziz et al., 2009; Kelly et al., 2000). While it has been shown that the VDR promoter contains potential binding sites for regulators JUND (AP-1), PU.1, and MYC (Jehan and DeLuca, 1997), our analysis recovers these previously known potential regulatory interactions but also emphasizes the dynamic and cell-specific regulation of VDR during both monocyte and macrophage differentiation. Interestingly, a gain in regulatory inputs at the VDR promoter correlates with its general increase in gene expression during late macrophage specification (Figure R2.13B,D), while the expression of VDR in monocytes is sustained throughout the time-series, suggesting cell-specific regulatory control.

Myeloid cell specification network analysis

The PU.1 locus has been extensively characterized for transcription factor regulatory interactions in the context of hematopoietic development (Hoogenkamp et al., 2007; Leddin et al., 2011). To study the PU.1 regulatory interactions recovered in our myeloid time courses, we generated PU.1-specific subnetworks. A total of 23 interactions were observed, of which 11 imply targeted regulation of PU.1 and 12 suggest PU.1-mediated interactions (Figure R2.14B). As a key regulator of myeloid cells, we observe a general trend for increased expression of PU.1 across our time-series, but with observably different kinetics between each cell-type (Figure R2.14A). It has been previously shown that RUNX1 specifically regulates both the upstream regulatory elements (Huang et al., 2008) and promoter of PU.1 (Koh et al., 2013) in myeloid cells. Our PU.1 network recovers RUNX1 regulatory interactions for undifferentiated cells, but

also for differentiated macrophage, monocyte, monocyte-derived macrophage and neutrophil subnetworks (Figures R2.14C-G). Similarly, CEBPα has also been shown to regulate PU.1 through regulation of its promoter and upstream regulatory elements in hematopoietic cells (Kummalue and Friedman, 2003; Yeamans et al., 2007). We observe CEBPα regulating PU.1 specifically in differentiated macrophage and neutrophil cells (Figures R2.14D,G). Our subnetworks also illustrate the PU.1 auto-regulatory feedback loop, maintaining a constant transcriptional burst of expression across differentiation. Interestingly, our networks also identify PU.1 occupancy at several additional regulators such as MAFB, EGR, STAT, and VDR. Our chromatin analysis identifies likely candidates regulated by PU.1 during granulopoiesis.

Because PU.1 acts as potent regulator of myeloid fate specification, we focused on PU.1-specific regulation of EGR family members in terminally differentiated cells. Neutrophil and macrophage cell fate specification is mediated through concerted regulation of PU.1, EGR (1,2), GFI1, and CEBPα (Laslo et al., 2006). Our PU.1 subnetwork identifies regulation of PU.1 of EGR family members during macrophage and neutrophil differentiation (Figures R2.14D,G). To this end, we generated both EGR (Figure R2.15) and GFI1 (Figure R2.6) subnetworks to illustrate their regulatory interactions in myeloid cells. EGR regulators can directly repress GFI1 promoter activity in NIH3T3 cells in luciferase assays (Laslo et al., 2006), and our EGR subnetwork captures this regulatory interaction using footprinting. We also identify EGR-specific regulation of lineage and temporal mediated regulators such as MAFB, which is specified during macrophage commitment, and JUND, which is specified in both monocyte and macrophage differentiation (Figure R2.15D,E). Our analysis emphasizes the dynamic expression patterns for each differentially expressed EGR (1,2,3) member during myeloid differentiation (Figure R2.15G). We observe the most dramatic change (~10-100 fold) in EGR expression in

macrophage cells 3 hours post-differentiation. Interestingly, the monocyte-derived macrophage cells also display the same sharp and dramatic increase in EGR expression. We identify auto-regulatory feedback interactions for each EGR members across the different cell-types (Figure R2.15C-F), a likely explanation for dramatic and sustained EGR expression during differentiation. While the role of EGRs in specifying neutrophil and macrophage fates has been previously described (Laslo et al., 2006, 2008), our analysis provides additional novel chromatin interactions of EGR regulation during myeloid differentiation.

Similarly, the transcriptional repressor GFI1 has been previously shown to function as an integral regulator in multi-lineage blood cell development through the regulation of gene expression (Van der Meer et al., 2010; Möröy and Khandanpour, 2011). We identify dynamic changes in both the temporal expression (Figure R2.6C) and GFI1 occupancy for several transcription factors in our gene regulatory networks (Figure SD-H). GFI1 can promote neutrophil differentiation by antagonizing PU.1 and EGR activity via direct protein-protein interactions (Dahl et al., 2007) and through direct transcriptional repression (Kubosaki et al., 2009; Laslo et al., 2006; Spooner et al., 2009). Our analysis identifies this well-studied regulation of PU.1 by GFI1 in our HL-60 and neutrophil networks (Figures R2.6D,E). We recover the GFI1 auto-regulatory feedback loop in our neutrophil network (Figure R2.6E), which has been studied in previous work on GFI1 auto-regulation (Yücel et al., 2004), as well as novel GFI1 regulatory interactions for STAT6 and MAFB in neutrophil and monocyte cell-types (Figures R2.6F,G). Our network analysis recovers many previously identified regulatory interactions fundamental in hematopoietic cells, in addition to many new candidate interactions. It also demonstrates the power of chromatin networks in uncovering the dynamic regulatory circuitry specified during cellular differentiation.

2.4 Discussion

We derived clusters of differentially expressed transcription factors as well as chromatin accessible elements genome-wide and analyzed the difference in expression of several transcriptional regulators and their corresponding motif enrichment across the chromatin clusters. We built a myeloid differentiation gene regulatory network using open chromatin footprinting using methods previously used to study transcriptional regulatory interactions in several biological contexts (Neph et al., 2012b; Sullivan et al., 2014). In our case, we integrated transcription factor expression and chromatin footprints from ATAC-seq to build gene regulatory networks of myeloid differentiation. Our networks include 23 differentially expressed regulators interconnected through 158 footprints. We recovered regulatory interactions that were derived experimentally from previous studies in hematopoietic cells, thus demonstrating the accuracy in our analysis framework. For example, our analysis recovered the PU.1 and GFI1 auto-regulatory feedback loops, which have been implicated in the controlled expression of both transcription factors in hematopoietic cells (Chen et al., 1995; Okuno et al., 2005; Yücel et al., 2004). Interestingly, we detect PU.1 footprints in distinctly different target genes in each lineage, despite PU.1 expression increasing to similar end-levels across all cell types in our time-series. These lineage-specific differences could come from a combination of regulation of chromatin opening at these target genes by other TFs or simply from combinatorial interactions with co-factors. In contrast to PU.1, the expression of GFI showed highest expression in terminally differentiated neutrophils, was present at lower levels in macrophage and monocytes, and was not significantly detected in monocyte-derived macrophages. Unlike PU.1, the subnetworks for GFI1 mirror our observations from its expression profile in that the regulatory interactions in neutrophils were unique, were common between macrophage and monocytes, and absent in

monocyte-derived macrophage cells. Thus, network analysis highlights the complexity of the PU.1 subnetwork compared to other TFs such as GFI1.

We observed disparate EGR expression between differing cell-types and amongst the EGR family members across differentiation. Our EGR subnetwork analysis revealed a complex series of regulatory interactions for each EGR member that were distinct to each lineage. Earlier investigations have demonstrated that while expression of the EGR factors is lineage specific, EGR1/2 regulators can transcriptionally compensate for one another (Laslo et al., 2006). Interestingly, auto-regulatory feedback loops at different EGR family members were prominent features that were detected and for each lineage, likely driving and/or tightly regulating their own expression. While we cannot specify which EGR member is directly interacting in our network due to the similarity of their DNA-binding motifs, we rely on the expression for each member as an indication of the most likely candidate, and thus specify inferred interactions from all EGR1/2/3 members in our networks. Interestingly, the cooperative regulatory interactions detected for the EGR regulators were highly complex. For example, EGR1 was regulated by E2F1/8, MYC, EGR, and PPARG in macrophages, while in neutrophils EGR1 was regulated by PU.1, E2F1/8, CEBPα, EGR, and PPARG. Strikingly, 4 out of 5 and 4 out of 6 regulatory interactions were shared upstream of EGR1 between macrophage and neutrophil cells, respectively. Our network analysis of the EGR sub-circuitry expands on the current knowledge of known regulatory linkages and elucidates the novel interactions encoded in the gene regulatory networks of myeloid cells.

Our experimental data and analysis present the first example of gene regulatory networks built using footprinting during human myeloid differentiation. Our networks encompass early, intermediate, and late stages of differentiation. While we observed many interesting changes in

footprint-based regulatory interactions during differentiation, we focused on the regulatory control of VDR, because of its relatively understudied role in macrophages, monocytes, and the innate immune system. While HL-60 cells require the stimulus of 1,25-dihydroxyvitamin D3 to differentiate into monocytes in our system, HL-60-derived macrophage cells do not. We observed VDR up-regulation in late macrophages, but observed a sustained expression of VDR throughout monocyte differentiation. Strikingly, the VDR sub-circuits we identified in macrophage and monocyte cells revealed a dramatic difference in their respective regulatory circuitry. While the interactions for VDR in monocytes were sustained primarily by PU.1 and PPARG throughout, we observed a dramatic increase in footprints for several transcription factors in the late stages of macrophage differentiation. Several studies have found interactions between PPARG and VDR exist in the context of melanoma (Sertznig et al., 2009), the innate immune system, and across several immunopathologies (Kiss et al., 2013), but their protein-DNA interactions have not been extensively characterized during cellular differentiation. Our results demonstrate that PPARG and PU.1-mediated interactions account for the observed stable expression of VDR during monocyte differentiation, while the combinatorial regulatory interactions from several transcription factors presumably drive the late up-regulation in VDR expression observed in maturing macrophages. Thus, the dynamics of gene regulatory circuits we identify within different immune cells illustrate different strategies for controlling the regulation of VDR in a cell-specific manner. Interestingly, the deletion of VDR specifically in macrophage cells leads to both atherosclerosis and insulin resistance in mice (Oh et al., 2015), which suggests that VDR plays an important role in controlling the inflammatory properties of macrophages and many other immune cell types (Wöbke et al., 2014).

In summary, our results highlight the complex immune regulatory circuitry of human neutrophil, monocyte, and macrophage differentiation. We demonstrate the power of combined time course open chromatin and gene expression analysis, allowing us to construct dynamic draft gene regulatory networks that recover both previously established as well as novel interactions in human myeloid cells. These networks will serve in furthering the understanding of transcriptional regulation in the context of cellular differentiation and etiology of human immunopathologies.

2.5 Limitations on chromatin footprinting analysis

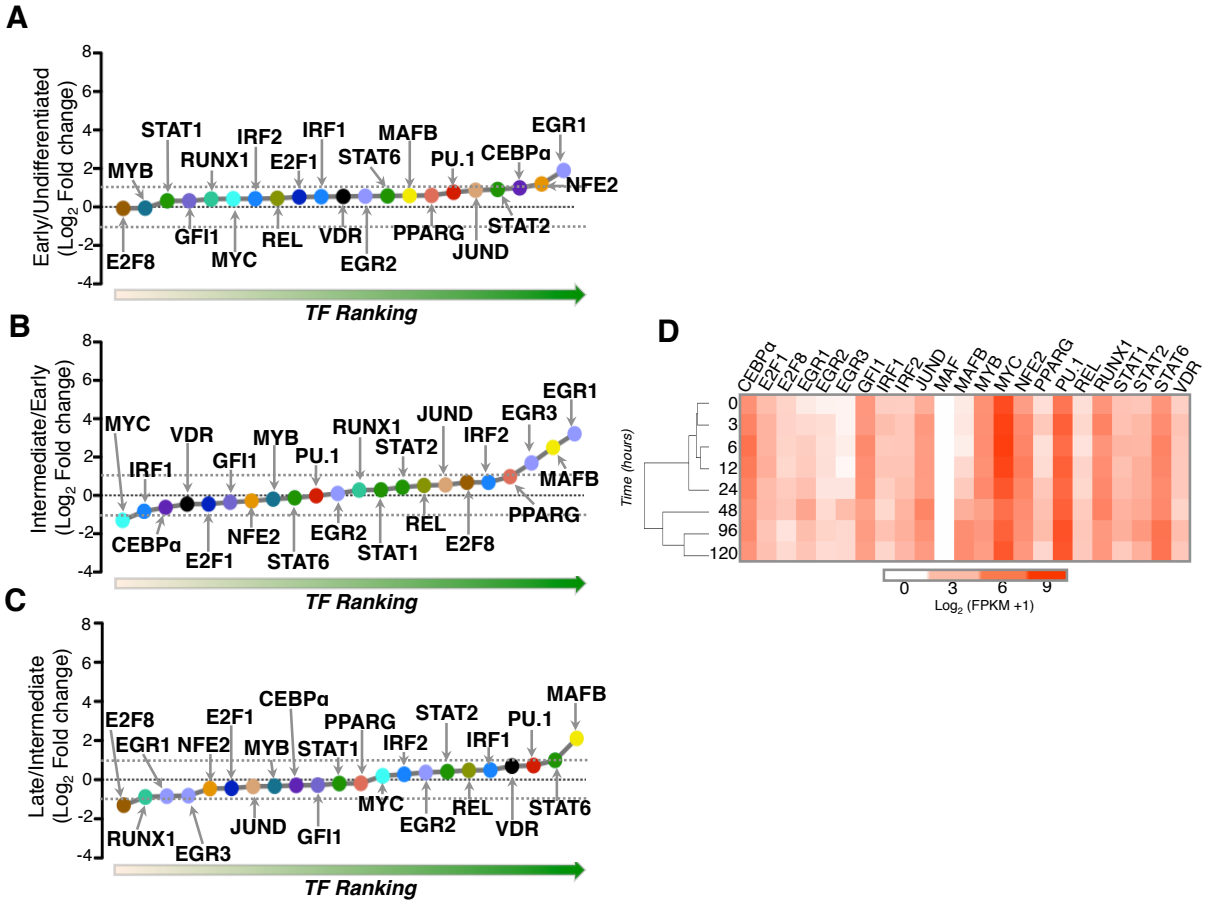
Genomic footprinting offers the ability to derive protein-DNA interactions from genome-wide open chromatin assays to understand regulatory interactions across several biological systems (Mo et al., 2015; Neph et al., 2012b; Qu et al., 2015; Stergachis et al., 2014; Sullivan et al., 2014). While genomic footprinting offers the great potential to infer regulatory interactions, several technical considerations and challenges should be addressed when applying this method. A detailed description of both the prospects and challeneges of genomic footprinting are discussed in the following perspectives (Sung et al., 2016; Viestra and Stamatoyannopoulos, 2016), we will focus on key issues our study has addressed. An important aspect of genomic footprinting is sequencing depth of biological samples. The sequencing depth of samples have been important when analyzing assays such as ChIP-seq and RNA-seq (Pepke et al., 2009) and open chromatin assays such as DNase-seq (Thurman et al., 2012). It has been shown that increasing the sequencing depth allows for higher-resolution and recovery of TF footprints when performing genomic footprinting (Viestra and Stamatoyannopoulos, 2016). To address this issue, we maintained a sequencing depth of 200 million reads for all time-points in which we derived

genomic footprints. While increasing the number of reads in our study could potentially recover additional information, our novel draft gene regulatory networks highlight the most sensitive regulatory interactions detected during myeloid differentiation.

Several computational methods for both *de novo* footprint detection and transcription factor recognition sequence occupancy have been developed to identify genomic footprints from open chromatin data (Viestra and Stamatoyannopoulos, 2016). Our strategy was to employ a two-step footprinting analysis that first used the Wellington (Piper et al., 2013) algorithm, which optimizes a windowing strategy incorporating the DNA strand information, and scores a cleavage rate using flanking windows of a potential footprint. The *de novo* derived footprints were then processed through an unsupervised approach CENTIPEDE (Pique-Regi et al., 2011), that classifies and models cleavage profiles simultaneously. This allowed us to derive the genomic footprints of 23 transcriptional regulators that showed significant and dynamic gene expression profiles, have a well-described role in immune regulation, and contain a DNA binding motif. In summary, a detailed and carefully curated analysis of genomic foorptining is critical for inferring both meaningful and informative regulatory networks across biological systems.

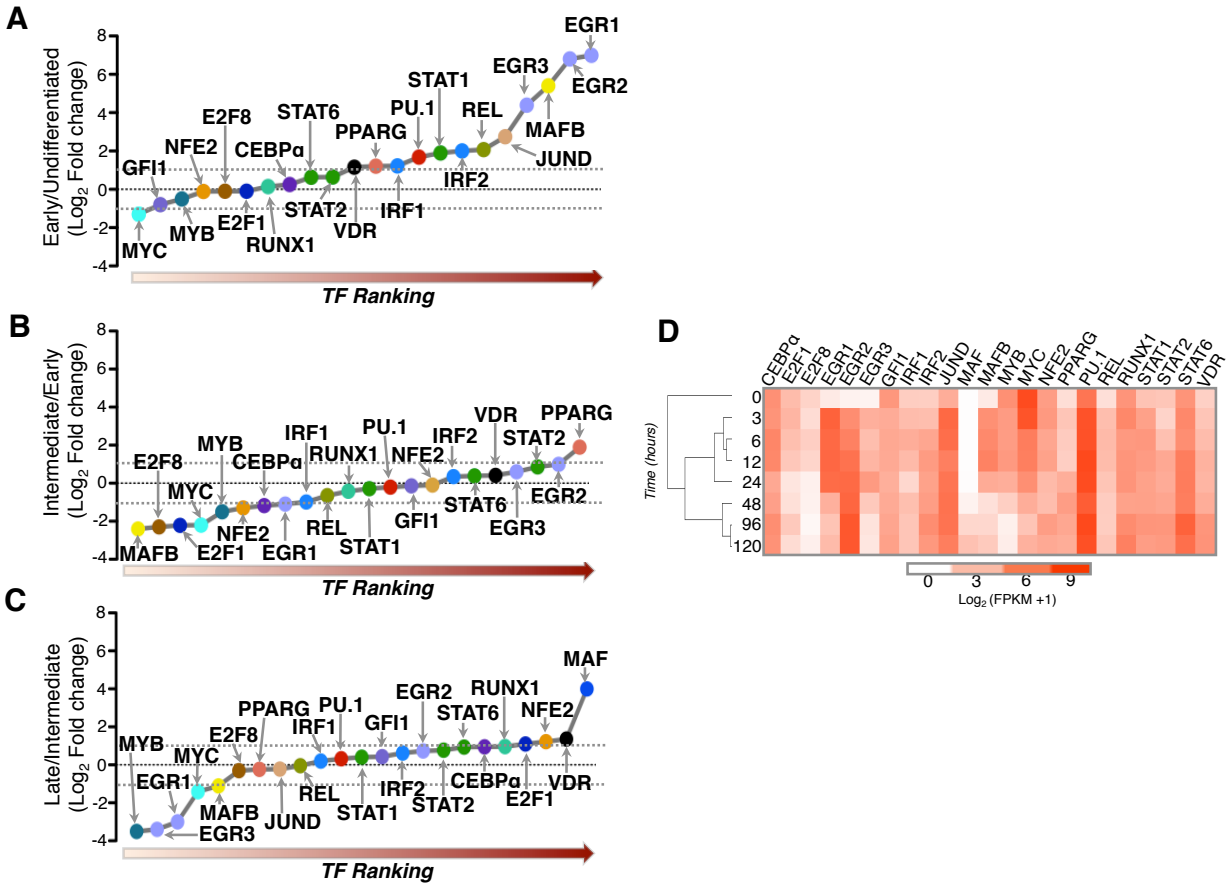
2.6 Figures

Figure R2.1 Ranking of transcriptional regulators during Monocyte differentiation



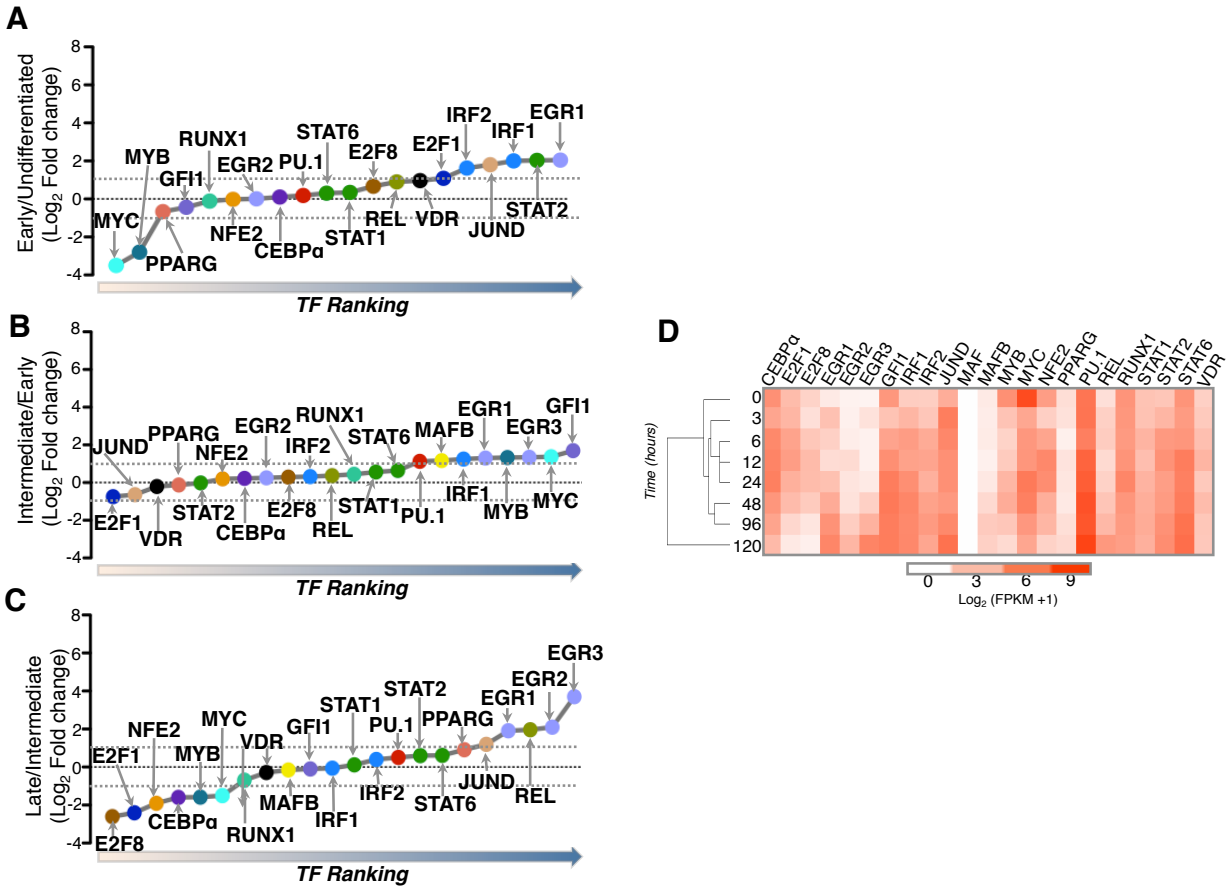
(A,B,C) Fold change in expression of 23 transcription factors comparing undifferentiated with early, early with intermediate, and intermediate with late. Fold change of 2 is indicated by horizontal dashed line. (D) Heatmap of regulator expression alphabetically listed. Log2 transform of FPKM +1.

Figure R2.2 Ranking of transcriptional regulators during Macrophage differentiation



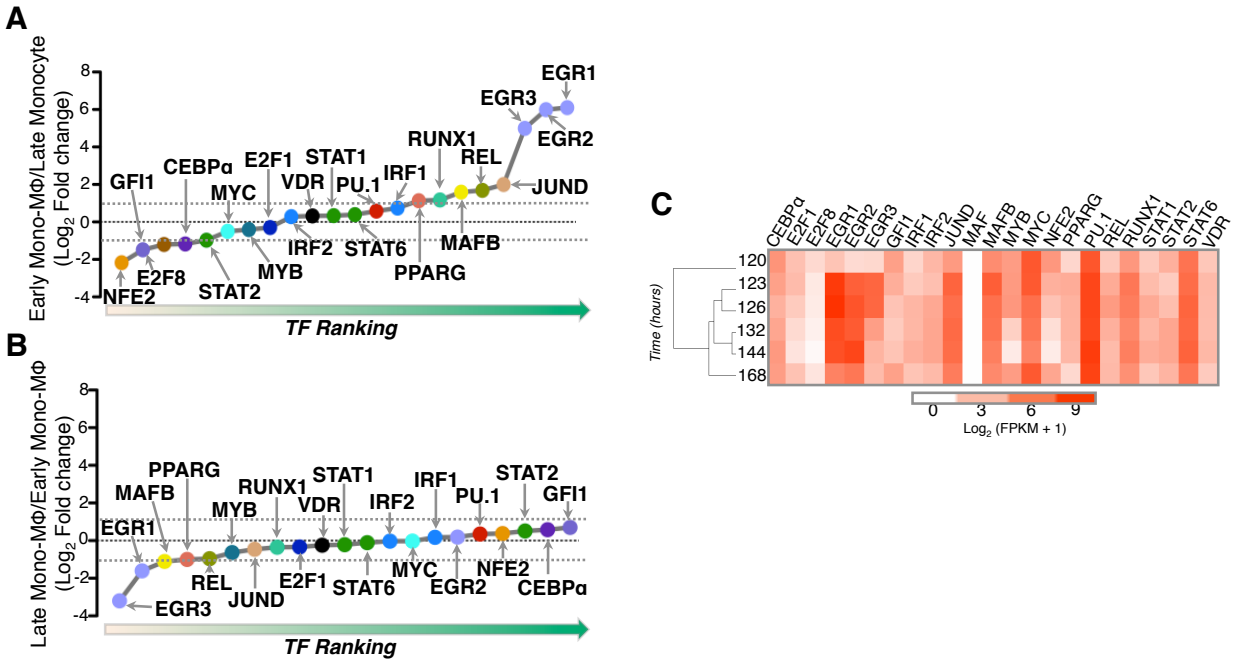
(A,B,C) Fold change in expression of 23 transcription factors comparing undifferentiated with early, early with intermediate, and intermediate with late. Fold change of 2 is indicated by horizontal dashed line. (D) Heatmap of regulator expression alphabetically listed. Log2 transform of FPKM +1.

Figure R2.3 Ranking of transcriptional regulators during Neutrophil differentiation



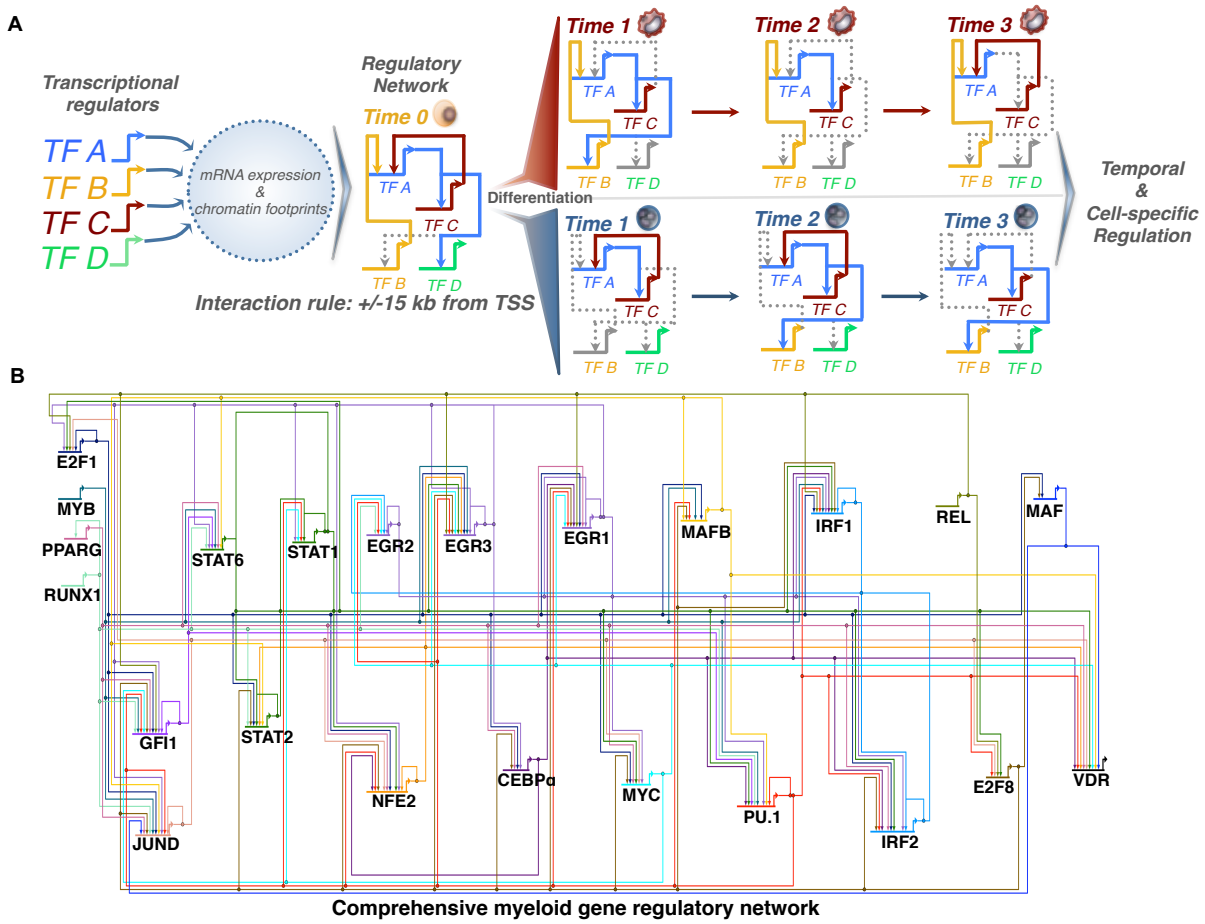
(A,B,C) Fold change in expression of 23 transcription factors comparing undifferentiated with early, early with intermediate, and intermediate with late. Fold change of 2 is indicated by horizontal dashed line. (D) Heatmap of regulator expression alphabetically listed. Log2 transform of FPKM +1.

Figure R2.4 Ranking of transcriptional regulators during Monocyte-derived Macrophage differentiation



(A,B) Fold change in expression of 23 transcription factors comparing undifferentiated with early, early with intermediate, and intermediate with late. Fold change of 2 is indicated by horizontal dashed line. (C) Heatmap of regulator expression alphabetically listed. Log2 transform of FPKM +1.

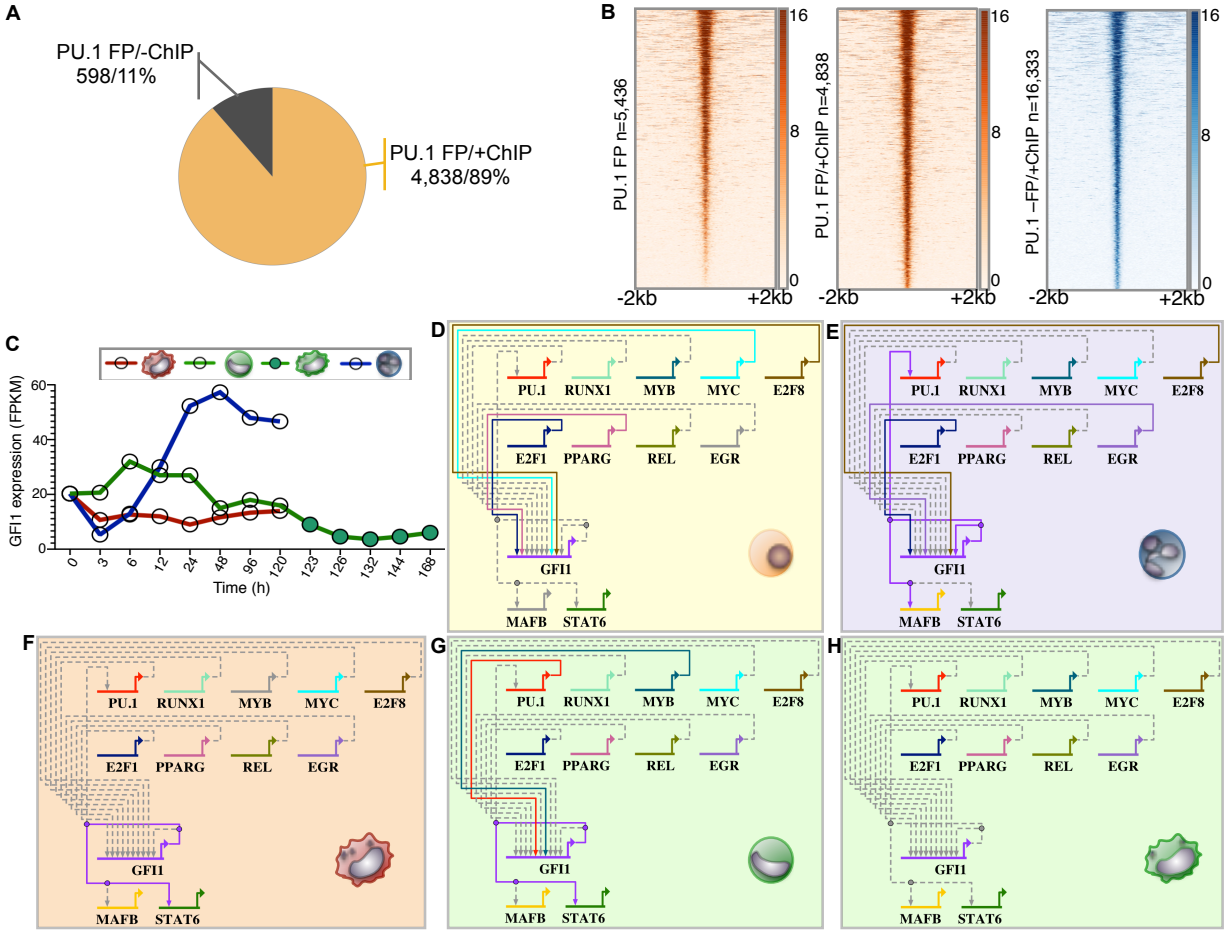
Figure R2.5 Human myeloid gene regulatory network architecture



(A) Overall schematic design for constructing gene regulatory circuits. An interaction rule was defined whereby an edge was drawn when a transcriptional regulator footprint (TF A) was identified within a +/-15kb window of another transcriptional regulator TSS (TF C and TF D). Colored edges indicate that a regulatory interaction was observed for the respective time-point. Dashed grey edges indicate regulatory interactions observed at other time-points or cell-types respectively. Colored grey gene arrows indicate no mRNA detected at a given time-point. (B) A comprehensive 'genome-view' gene regulatory network snapshot of 23 transcriptional regulators and 158 inferred regulatory interactions were generated from ATAC-seq footprinting (Experimental methods). The 'genome-view' network representation shows all inferred

interactions during human myeloid differentiation. Each TF regulator is assigned a unique color identifier to track changes in regulatory interactions during differentiation.

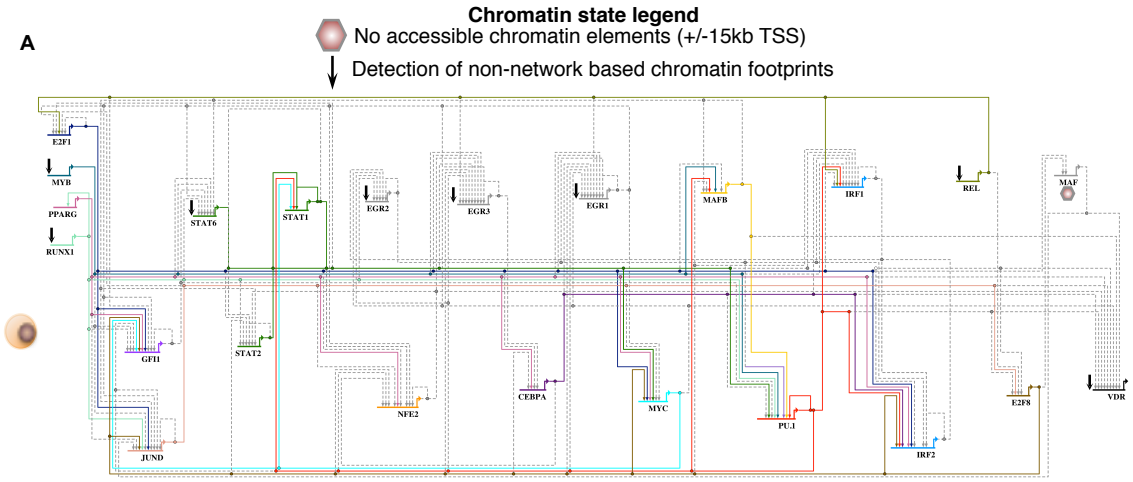
Figure R2.6 GFI1 gene regulatory subnetwork



(A) Distribution of undifferentiated HL-60 PU.1 footprints identified by ATAC-seq that are also detected by PU.1 ChIP-seq. 4,838 (~89%) of PU.1 ATAC-seq footprints are also detected by PU.1 ChIP-seq. (B) Heatmaps of ATAC-seq read density is shown for undifferentiated HL-60 inferred PU.1 ATAC-seq footprints (Left; n=5436), PU.1 ATAC-seq footprints with an overlapping ChIP-seq PU.1 peak (Center; n=4,838), and PU.1 ChIP-seq read density for ChIP-seq PU.1 peaks with no ATAC-seq footprint detected (Right Blue heatmap; n=16,333). 10xRPM (Reads Per Million) values are log transformed and row-mean normalized for ATAC-seq and ChIP-seq data. ATAC-seq and ChIP-seq signal is shown for a window of +/-2 kb from the footprint center and ranked from strongest to weakest for all comparisons. (C) GFI1 expression

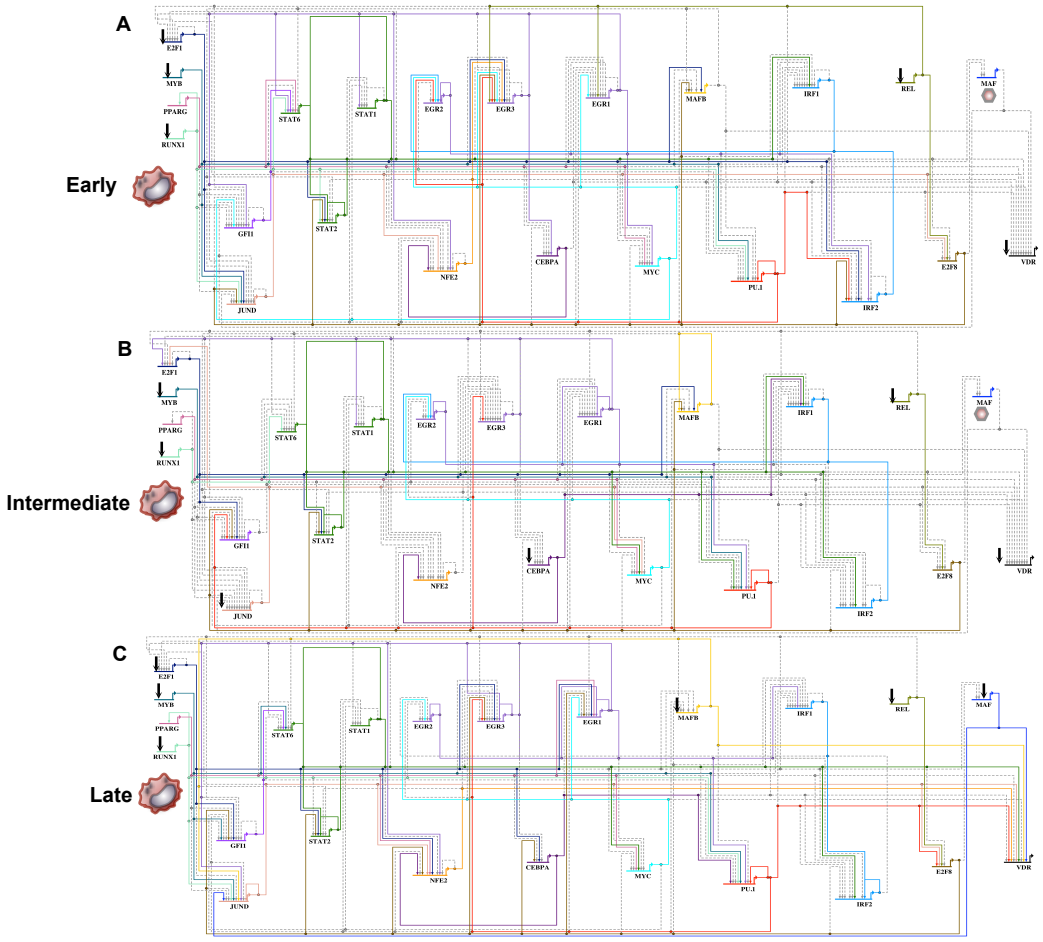
demonstrates a cell-specific pattern during myeloid differentiation. Maximal expression is observed in neutrophils, with minimal to no expression detected in monocyte-derived macrophages. Mean FPKM values for each time-series are shown. (D-H) GFI1 sub-circuits are shown for (D) undifferentiated HL-60, (E) neutrophil, (F) macrophage, (G) monocyte, and (H) monocyte-derived macrophages. The GFI1-MAFB interaction and GFI1 auto-regulatory feedback loop is specific to neutrophil cells. Colored edges indicate that a regulatory interaction was observed for the respective time-point. Dashed grey edges indicate regulatory interactions observed at other time-points or cell-types respectively. Colored grey gene arrows indicate no mRNA detected at a given time-point.

Figure R2.7 HL-60 genome-view network



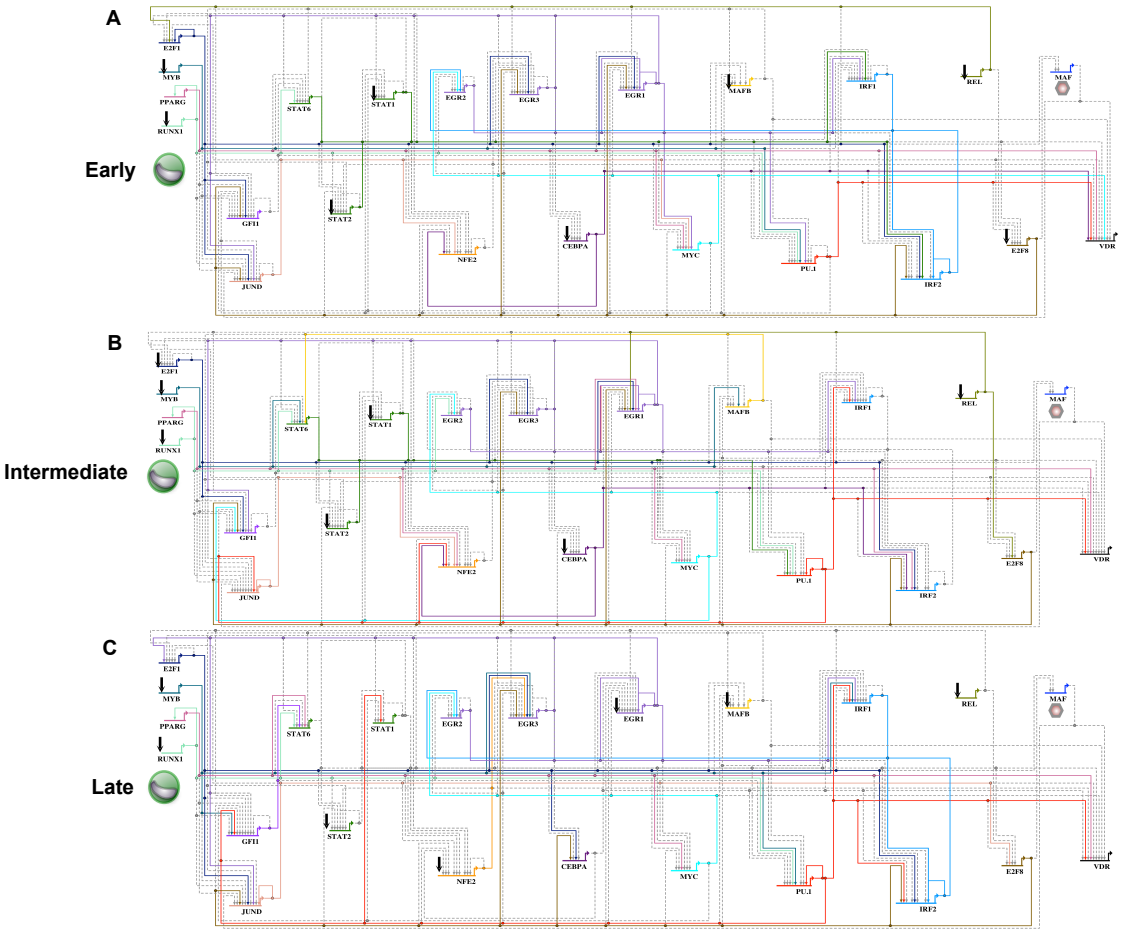
(A) Genome-view network representation of 23 regulators for HL-60 cells. Colored edges indicate that a regulatory interaction was observed for the respective time-point. Dashed grey edges indicate regulatory interactions observed at other time-points or cell-types respectively. Colored grey gene arrows indicate no mRNA detected at a given time-point.

Figure R2.8 Macrophage genome-view networks



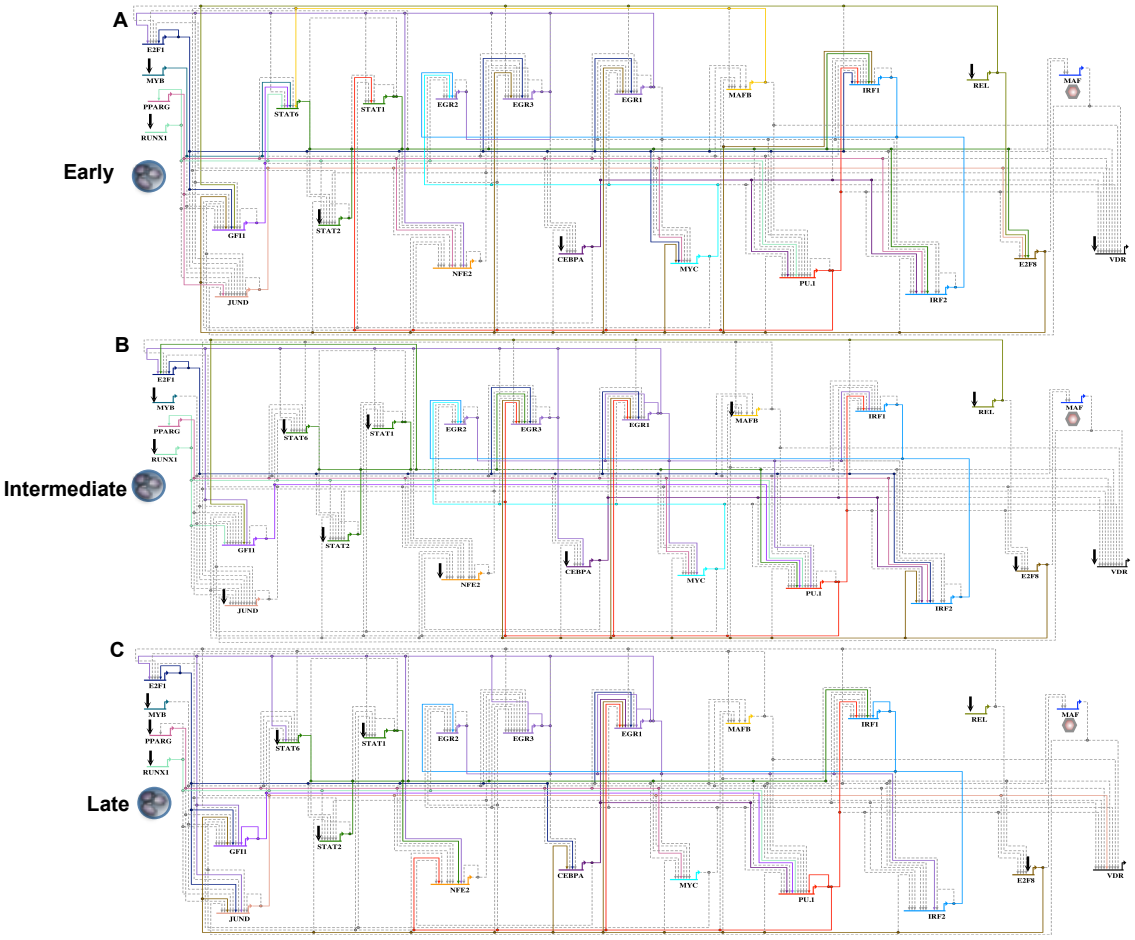
(A-C) Genome-view network representation of 23 regulators of early (A), intermediate (B) and late (C) stages of macrophage differentiation. Colored edges indicate that a regulatory interaction was observed for the respective time-point. Dashed grey edges indicate regulatory interactions observed at other time-points or cell-types respectively. Colored grey gene arrows indicate no mRNA detected at a given time-point.

Figure R2.9 Monocyte genome-view networks



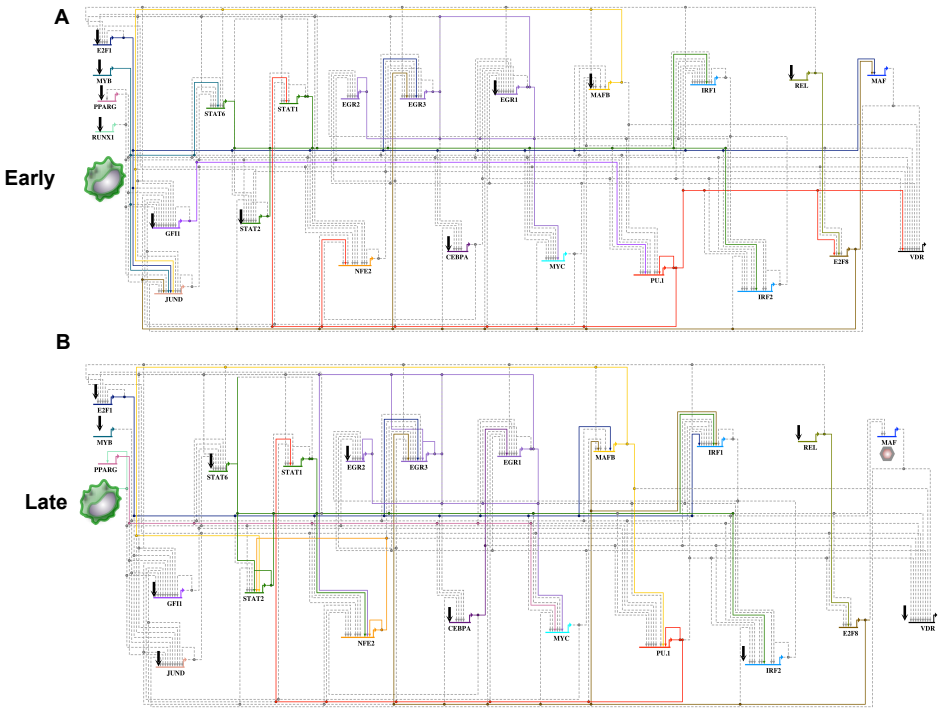
(A-C) Genome-view network representation of 23 regulators of early (A), intermediate (B) and late (C) stages of monocyte differentiation. Colored edges indicate that a regulatory interaction was observed for the respective time-point. Dashed grey edges indicate regulatory interactions observed at other time-points or cell-types respectively. Colored grey gene arrows indicate no mRNA detected at a given time-point.

Figure R2.10 Neutrophil genome-view networks



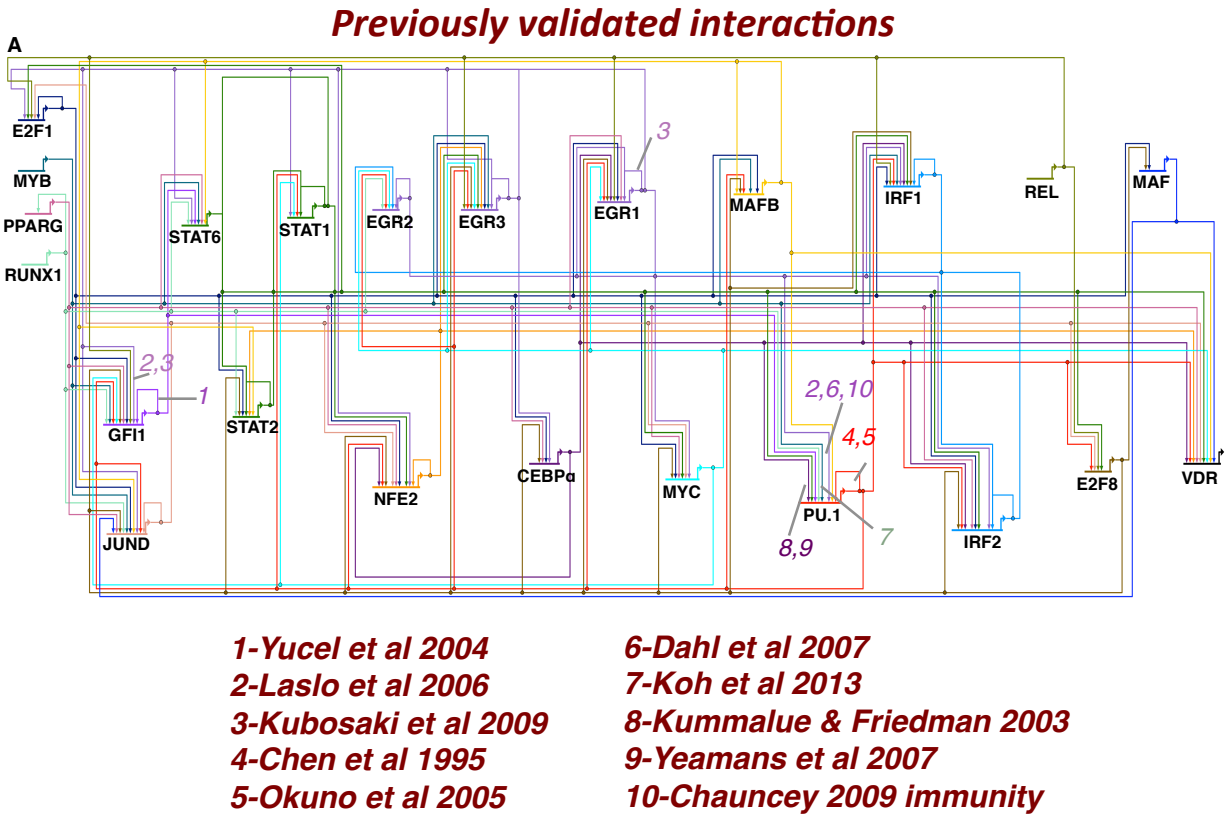
(A-C) Genome-view network representation of 23 regulators of early (A), intermediate (B) and late (C) stages of neutrophil differentiation. Colored edges indicate that a regulatory interaction was observed for the respective time-point. Dashed grey edges indicate regulatory interactions observed at other time-points or cell-types respectively. Colored grey gene arrows indicate no mRNA detected at a given time-point.

Figure R2.11 Monocyte-derived macrophage genome-view networks



(A-B) Genome-view network representation of 23 regulators of early (A) and late (B) stages of monocyte-derived macrophage differentiation. Colored edges indicate that a regulatory interaction was observed for the respective time-point. Dashed grey edges indicate regulatory interactions observed at other time-points or cell-types respectively. Colored grey gene arrows indicate no mRNA detected at a given time-point.

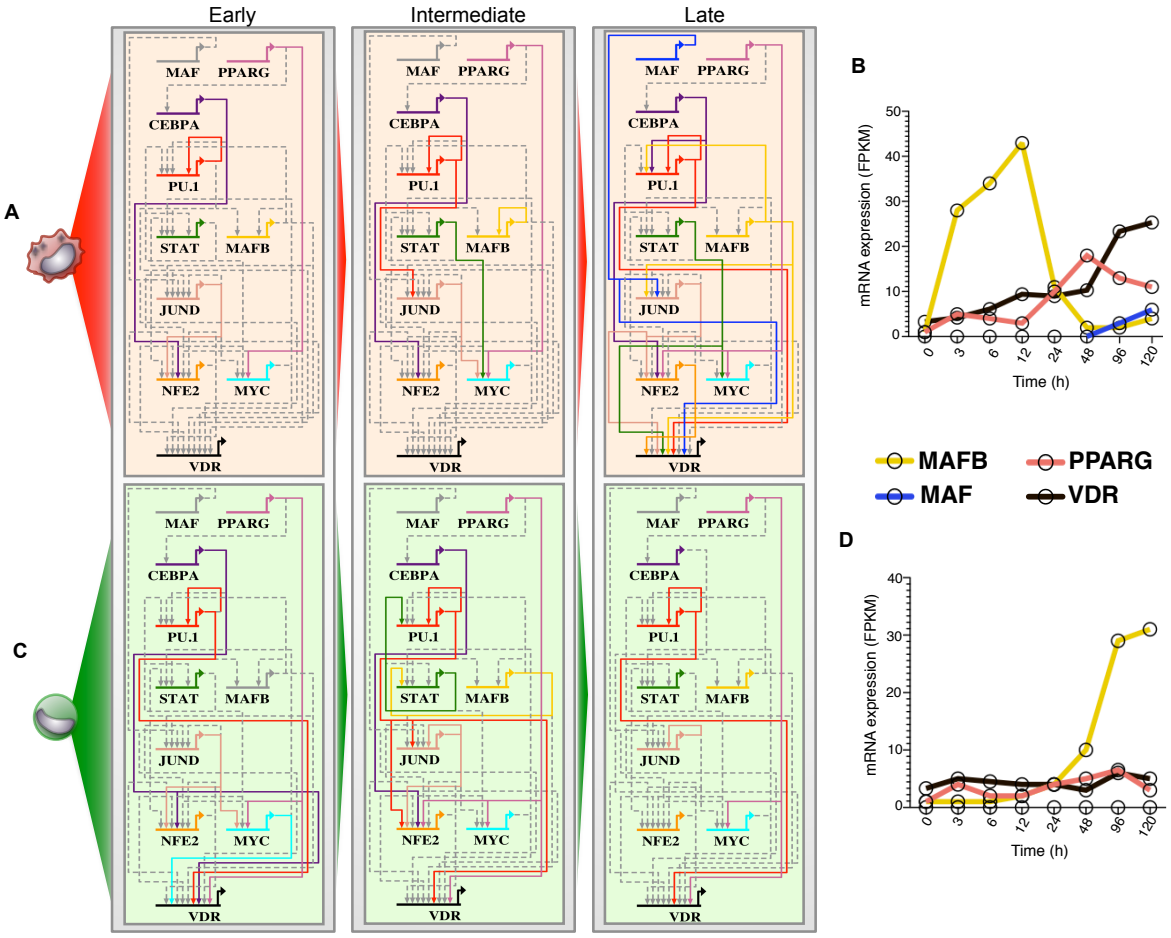
Figure R2.12 Previously validated hematopoietic regulatory interactions



(A) Genome-view of myeloid regulatory network interactions in differentiating HL-60 cells.

Regulatory interactions are indicated by numbers and associated citations.

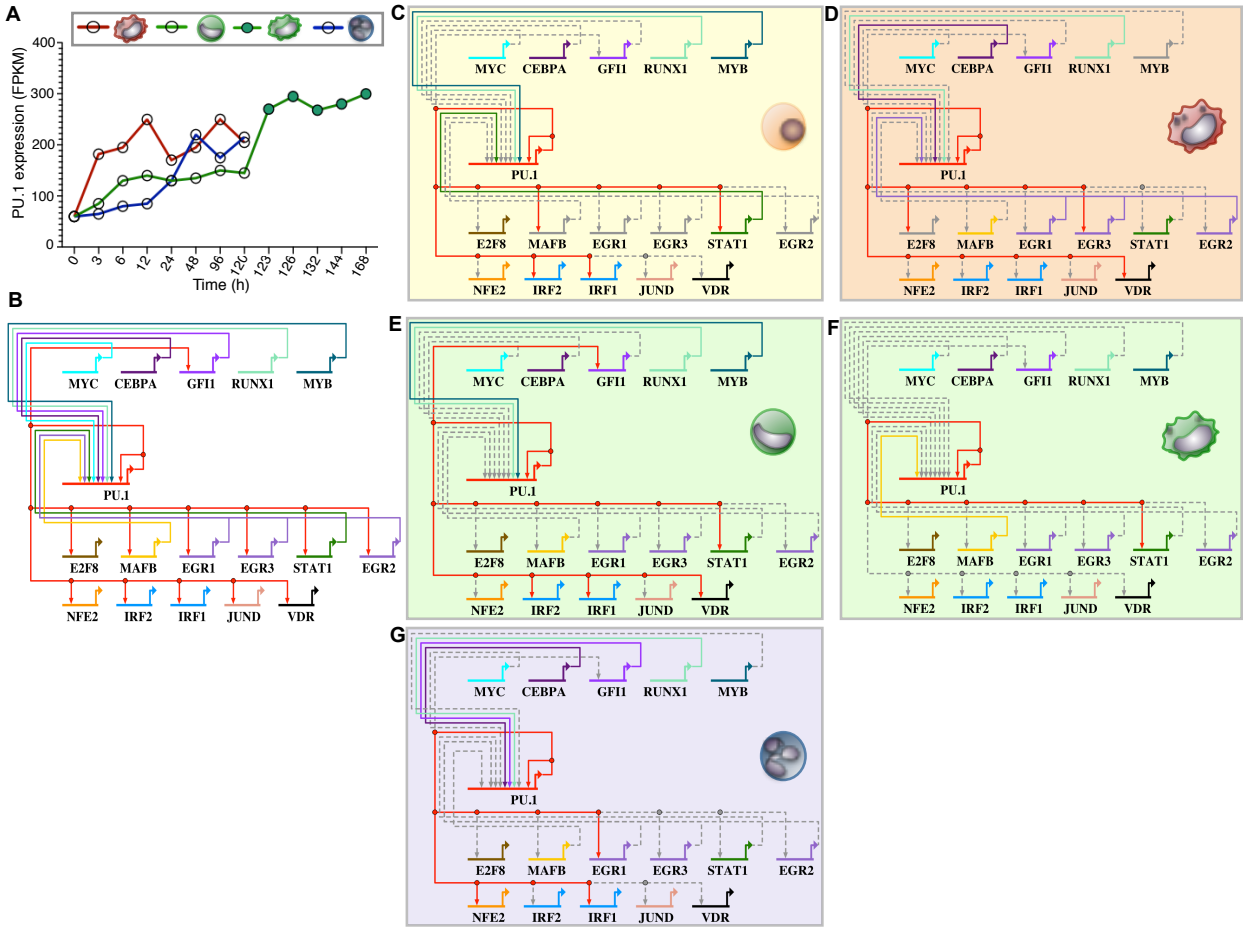
Figure R2.13 Cell-specific cis-regulation of VDR during myeloid differentiation



(A and B) Temporal gene regulatory sub-circuits of VDR regulation in macrophage (top panel) and monocyte (bottom panel) cell-types demonstrate the dramatic differences in regulatory interactions during cellular differentiation. Early is denoted as 3-12 hours, intermediate as 24-48 hours, and late as 96-120 hours of differentiation (Experimental methods). Colored edges indicate that a regulatory interaction was observed for the respective time-point. Dashed grey edges indicate regulatory interactions observed at other time-points or other cell-types, respectively. Colored grey gene arrows indicate no mRNA detected at a given time-point. (A' and B') Gene expression for regulators MAFB, MAF, PPARG, and VDR during macrophage

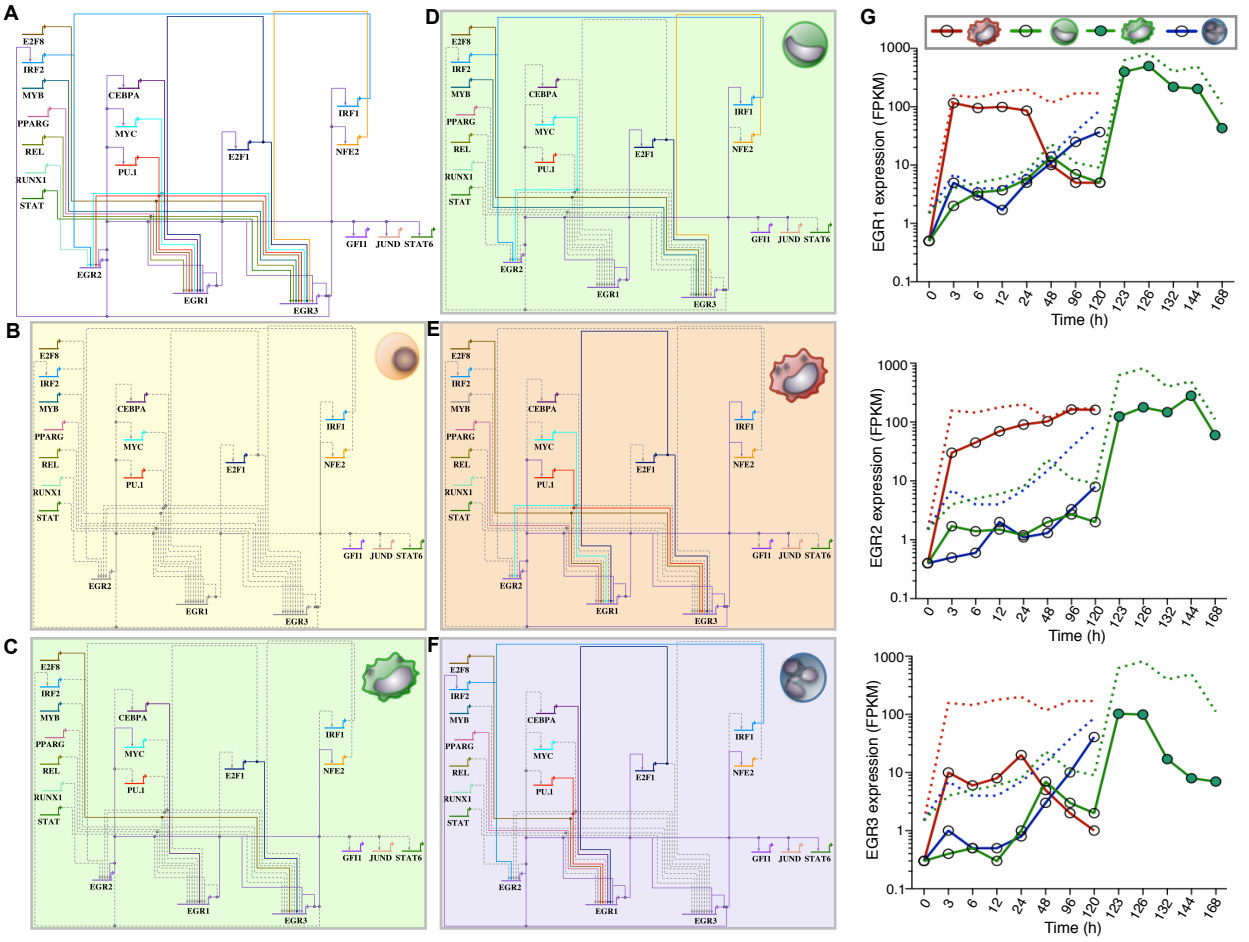
(top panel) and monocyte (bottom) differentiation, respectively. Mean FPKM values for each regulator are shown.

Figure R2.14 PU.1–regulated targets change in differentiated human myeloid cell types



(A) PU.1 gene expression during myeloid differentiation generally increases across all cell-types. The mean FPKM values for each time-series are shown. (B) Genome-view of the PU.1 sub-circuitry in myeloid cells. We inferred 12 PU.1-mediated regulatory interactions and 11 that regulate PU.1. (C-G) PU.1 sub-circuits of (C) HL-60, (D) 96-120 hour macrophage, (E) 96-120 hour monocyte, 144-168 hour monocyte-derived macrophage, and 120-hour neutrophil cells. Colored edges indicate that a regulatory interaction was observed for the respective time-point. Dashed grey edges indicate regulatory interactions observed at other time-points or cell-types respectively. Colored grey gene arrows indicate no mRNA detected at a given time-point.

Figure R2.15 EGR gene regulatory subnetworks



(A) Genome-view of the EGR transcription factor sub-circuitry in differentiated myeloid cells. EGR motifs are generally indistinguishable between protein family members, thus regulatory interactions are inferred from all three EGR factors (EGR1, EGR2, EGR3). (B-F) Sub-circuits of EGR regulation for (B) undifferentiated HL-60, (C) monocyte-derived macrophage, (D) monocyte, (E) macrophage, and (F) neutrophil. Our network analysis infers regulatory interactions specific to both myeloid cells and EGR members. Colored edges indicate that a regulatory interaction was observed for the respective time-point. Dashed grey edges indicate regulatory interactions observed at other time-points or cell-types respectively. Colored grey gene arrows indicate no mRNA detected at a given time-point. (G) Dynamic kinetics of EGR1

(top), EGR2 (middle), EGR3 (bottom) gene expression profiles during myeloid differentiation. We observe a rapid change in EGR expression in macrophage cells with a general increase in expression across all cells. Mean FPKM values for each time-series are shown. Colored dash lines indicate the summed expression of all EGR members for each respective time-point and cell-type.

2.7 Methods

Chromatin footprinting and construction of myeloid gene regulatory networks

ATAC-seq data partitioned by early, intermediate and late stages were merged to achieve ≥ 200 million reads for footprinting analysis. Reads were also shifted as previously described in our subsequent analysis (Buenrostro et al., 2013). Chromatin footprints were determined using the Wellington algorithm (Piper et al., 2013) with the following parameters (-fp 6,31,1 -sh 7,36,4 -fdrlimit -2), restricting our analysis to footprints with an of FDR 1%. We then scanned chromatin footprints for motifs using FIMO (Grant et al., 2011; Neph et al., 2012b) to identify transcription factor motifs identified from the most recent JASPAR database (Mathelier et al., 2014). Lastly, we determined the quality of identified TF-footprints using Centipede (Pique-Regi et al., 2011) and compiled a final set of factor-specific footprints. To build myeloid gene regulatory networks, we focused on a subset of transcriptional regulators that were differentially expressed in our time-series. This allowed us to investigate networks of transcription factors that were demonstrated temporal and cell-specific expression dynamics. We employed a similar approach in constructing networks as shown in previous studies (Mo et al., 2015; Neph et al., 2012b; Sullivan et al., 2014). We padded the TSS for each transcription factor in using a +/-15kb window for identifying regulatory interactions that would include promoter and enhancer interactions. Directed edges were drawn from the first gene node to another gene node when a TF-motif potentially bound by the first gene was identified within a 15 kb distance of the second gene's TSS. Networks were generated for undifferentiated, early, intermediate, and late grouped time-points across all cell-types. Edges not identified for a corresponding time-point were drawn and colored as grey dashed lines. All networks were generated using Biotapestry software (Longabaugh et al., 2005, 2009).

Chapter 3: Genome-wide transcriptome and chromatin survey during pre-B cell differentiation

3.1 Abstract

The international STATegra consortium used several 'omic' platforms to profile the the transition from the B-cell progenitor and self-renewal pre-B(I) cell state to the growth arrested and differentiated pre-B(II) cell state that is mediated Ikaros (Ikzf1) upregulation. Specifically, the consortium used the pre-B cell line, B3 with a tamoxifen inducible version of Ikaros. We use DNase-seq to identify changes in the open chromatin landscape during this differentiation process in a 24-hour time course. We observe two general trends in changes of chromatin accessibility that ultimately define lymphocyte activation and cell cycle arrest programs. A pseudo-time trajectory during pre-B cell differentiation using single-cell RNA-seq reveals the underlying heterogeneity of cellular differentiation in our system. Additionally, we survey the chromatin landscape and measure the accessibility of single-cells before and after differentiation using single-cell ATAC-seq. We observe agreement in chromatin accessibility changes between time-points when averaging signal from all single-cells at a given time point while also showing the existence of distinct subgroups within each population. Overall our analysis reveals the complexity of even "single-step" developmental decisions at the single-cell level.

3.2 Introduction

The development of B cells is a complex multistage process whereby self-renewing hematopoietic stem cells are guided towards a terminal cell fate associated with the successful expression of a functional B cell receptor. The developmental stages of B-cell progenitors are characterized by known transition states that facilitate somatic rearrangement of the working B-cell receptor. One critical cell transition state involves the differentiation of a pre-B cell to a mature B cells. This involves changes in genes that define the B-cell receptor. The pre-B cell receptor includes the μ-heavy chain, VpreB, Igα and Igβ receptors (Cooper, 2015; Melchers, 2005). Additionally, pre-B cellular differentiation requires that cells undergo light-chain rearrangement, subsequent loss of the pre-BCR (VpreB, Igα and Igβ receptors) and eventual development of a mature BCR that is capable of binding antigens (Cambier et al., 2007; Melchers, 2005). Failure of proper V(D)J recombination events result in apoptosis of pre-B cells. Importantly, the proliferative state of cycling pre-B cells is dependent on both pre-BCR and IL-7 signaling (Clark et al., 2014; Ochiai et al., 2012; Rickert, 2013). IL-7R mediated-signaling limits the ability of cells to enter cell cycle arrest and transition from a cycling to resting state in pre-B cells. A fine-tuning of signaling and transcriptional control is critical for maintaining proper pre-B cell development.

A screen for transcriptional regulators that control T cell fate commitment isolated cDNA for the transcription factor Ikaros (*Ikzf1*) (Georgopoulos et al., 1992). Ikaros is considered one of the master hematopoietic regulators of myeloid and lymphoid development (Dovat, 2011). Ikaros null mice lack proper development of B cells, peripheral lymph nodes, natural killer cells and fetal T cells (Georgopoulos, 2002; Georgopoulos et al., 1994). Additional work demonstrates that loss of Ikaros activity promotes transformation of pre-B cells into a malignant state (Joshi et

al., 2014). Ikaros binds to DNA through its N-terminal domain comprising of four C2H2 zinc fingers (F1 to F4) (John and Ward, 2011). The C-terminal zinc-finger domain, allows for heterodimerization and multimerization with various protein partners (Bottardi et al., 2013). Ikaros can bind to its consensus DNA sequence (5'-TGGGAA/T-3') at both proximal and distal to gene regulatory elements, thereby contributing to the assembly of multiprotein complexes that can either repress or activate gene expression (Merkenschlager, 2010; Sabbattini et al., 2001; Yoshida et al., 2010).

Ikaros is primarily associated with the NuRD complex in all hematopoietic cells (Sridharan and Smale, 2007). The NuRD complex includes histone deacetylases HDAC1 and HADAC2, which are frequently referred to as transcriptional repressors (Bottardi et al., 2015). However, the NuRD complex can also be recruited to transcriptionally active genes targeted by Ikaros (Bottardi et al., 2014; Zhang et al., 2011). Furthermore, because NuRD interacts with complexes such as polycomb repressive complex 2 (PRC2) (Ross et al., 2012), direct chromatin re-organization may be mediated through interaction with Ikaros. In hematopoietic progenitors as well as in erythroid and lymphoid cells, Ikaros has been shown to interact with the hematopoietic GATA factors (GATA1, GATA2, and GATA3) (Ross et al., 2012). GATA factors are zinc-finger proteins capable of promoting both transcription activation and repression (Bresnick et al., 2010; Ko and Engel, 1993). Ikaros facilitates GATA1 binding to its DNA consensus sequence in vitro and to chromatin regulatory regions with both GATA1 and cofactor Friend of GATA1 (FOG1) in regulating long-range interactions at the human β-globin locus (Keys et al., 2008). Therefore Ikaros likely functions as a key regulator in defining cell identity in hematopoietic cells through the control of chromatin organization via recruitment of transcriptional regulators and co factors that include the NuRD complex.

The STATegra consortium is a European FP7 funded project that aims to develop statistical methods and software for the integration of Next Generation Sequencing and Omics data such as Transcriptomics, Proteomics, and Metabolomics. STATegra selected the mouse B3 cell line differentiation system as a model for the transition of cells from the pre-B(I) stage to the pre-B(II) stage in order to test the power of their novel statistical methods. B3 differentiation models the transition of self-renewing pre-B cells to a differentiated state in which cells have undergone growth arrest, which is accomplished by the induction of an Ikaros construct (Ikaros-ERt2) in these cells (Ferreir´ os-Vidal et al., 2013). This model system is not only relevant in the context of basic B cell biology, but has clinical interest. Ikaros mutations or deletions result in acute lympoblastic leukemias (Dovat, 2011). STATegra's goals are to develop new integrative methods using different analysis strategies, thus leveraging the expertise of the different partners in the consortium. Using ikaros-inducible B3 cells, I investigated the chromatin accessible landscape and transcriptome in differentiating pre-B cells. To this end, I generated 36 DNase-seq datasets for both control and Ikaros time-points, 324 single-cell RNA-seq and 227 single-cell ATAC-seq datasets to understand the regulation in differentiating pre-B cells.

3.3 Results

Genome-wide changes in DNaseI Hypersensitive elements during *Ikzf1* induction

A description of all 787 datasets generated and collected by the STATegra consortium is detailed in Figure R3.1 for 0, 2, 6, 12, 18, and 24-hour time-points and experimental procedures. Briefly, data was generated in biological triplicates from both control and *Ikzf1*-induced time-series for all experimental methods with the exception of single-cell RNA-seq and single-cell ATAC-seq. Induction of the Ikaros-ERt2 vector in pre-B cells is accomplished by treating cells with tamoxifen over the course of 24 hours. We performed DNase-seq to measure changes in open chromatin accessibility across pre-B cell differentiation in *Ikzf1*-induced and control time-series (Figure R3.2A). We initially optimized DNaseI concentrations in differentiating cells using qPCR for both positive and negative control targets to reduce over-digestion of native DNA. Initially, we surveyed genomic targets in genes with functions such as metabolism, cell cycle, lymphocyte biology, and transcription factors previously known to have accessible promoter elements in pre-B cells (Figures R3.2B-D). We observe changes in DNaseI accessibility for various targets during pre-B cell differentiation that we expected to occur in *Ikzf1*-induced cells. Libraries with optimal DNaseI signal to noise ratio as compared to control samples were selected for sequencing. To measure changes in accessibility genome-wide from our DNase-seq data, we performed an analysis comparing Hotspot (John et al., 2011) and HOMER (Heinz et al., 2010) peak callers (Figure R3.3A). We observed that a higher number of DHS peaks were estimated using Hotspot, but upon further inspection found that these DHS were often 'noisy' peaks. In contrast, peaks called by HOMER demonstrate a more accurate method for calling DHS from our data compared to Hotspot (Figure R3.3B). We employed a two-step peak calling strategy to capture narrow (<150bp) and broad (>150bp) DHS peaks using

HOMER, allowing us to capture diverse DHS from our differentiation time-series (Figure R3.3C).

To understand the changes in accessibility during pre-B cell differentiation, we generated a consolidated list of DHS reproducibly detected across biological replicates in our time-series. In total we identified 52,788 consensus DHS in our time-series (Figure R3.4A). DNase-seq datasets were both normalized and batch corrected for all subsequent analysis (Experimental methods). Principal component analysis of *Ikzf1*-induced time-points shows a high variance (82.9%) in the first principal component for our time-series, which corresponds to the progression of the *Ikzf1* time-series from 0 to 24 hours upon induction (Figure R3.4B). To investigate the dynamics of DHS during *Ikzf1* induction, we performed hierarchical clustering on the 7018 strongest DHS peaks. We identify two distinct profiles in chromatin accessibility in our time-series. 3,587 DHS elements show a loss in chromatin accessibility during pre-B cell differentiation. Moreover, gene ontology (GO) analysis of genes closest to the 3,587 DHS peaks shows enrichment in cell cycle, B-cell differentiation, and cell activation terms (Figure R3.4C). Interestingly, 3,431 DHS elements show an increase in DNaseI accessibility, with GO term enrichment for lymphocyte differentiation, immune response and B-cell differentiation. Additionally, DNaseI footprints were also derived for all DHS data using HOMER consolidated peaks (Figure R3.5). These results highlight that cis-regulatory elements undergo changes in DNaseI accessibility and that dynamic DHS changes mediate cell cycle control and lymphocyte activation during *Ikzf1* induction.

Single-cell RNA-seq analysis of pre-B cell differentiation

To understand the dynamic changes in gene expression during *Ikzf1* induction in B3 cells, we performed 8 single-cell experiments using the C1 system from Fluidigm (Experimental

methods). Pre-B cells were captured and visualized to estimate cell size and single-cell capture efficiencies across all time-points (Figure R3.6A). In total we analyzed 324 single-cell RNA-seq (scRNA-seq) libraries during the first 24 hours of pre-B differentiation (Figure R3.6B). The distribution of mapped reads for our scRNA-seq experiments range from 450,000 to 2.8 million mapped reads with a mean of 1.25 million mapped reads (Figure R3.7A). Additionally, we observed between 2000 and 6000 genes detected (> 1 FPKM, median=4,500) per single-cell across all data (Figure R3.7B) and find that only 375 genes are expressed in all 324 single-cells (Figure R3.7C). Lastly, we observe a strong correlation between individual single-cells (r=0.88, Figure R3.8A) and between pooled single-cell for each Fluidigm C1 run where a second experiment was performed for 18 and 24 hours time-points (r=0.96, 0.98, Figures R3.8B-C).

To identify distinct gene expression profiles in our time-series, we first performed a weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008). We identified 19 expression modules enriched across our single-cell data (Figure R3.9A). We observed significant enrichment for genes in module 1 (ME1, red box), which contains the regulator *Foxp1*, *Cd40,* and *Wnt4* ligand. A second module was also identified, which included the key regulators *Ikzf1, Ebf1,* and *Myc (*ME0, orange box). The eigengene profiles for modules 0 and 1 across our time-series show correlated genes that both increase and decrease during differentiation (Figure R3.9B). Furthermore, we observe cellular heterogeneity within these modules during pre-B cell differentiation, as we note that not all 24-hour pre-B cells are fully differentiated. While the analysis provides distinct modules of expression in our time-series, we found it difficult to interpret the heterogeneity of cells undergoing differentiation in our time-series from this analysis alone since WGCNA does not explicitly take the time course into account.

In order to understand the dynamics of our single-cells time course, we employed Monocle (Trapnell et al., 2014) to both recover differentially expressed genes and map the pseudo-temporal ordering of our 324 single-cells (Experimental methods). Monocle is designed to accommodate individual cells that are ahead or behind in differentiation compared to their "expected" position in the time course and it reorders these cells in a pseudo-temporal ordering. We identified 2,611 differentially expressed genes (FDR < 0.01) across our time-series. We pseudo-ordered 324 single-cells based on 2,611 differentially expressed genes to identify an unbiased pre-B cell trajectory (Figure R3.10A). As expected, we observed heterogeneity in cells differentiating from 0 to 24 hours in our time-series using the pseudo-temporal ordering that revealed that some cells collected at later time points looked like less-differentiated cells from earlier time points. The ordering of 324 single-cells robustly and accurately displays the expression of the pre-B cell receptor *Igll1* in single-cells during differentiation (Figures R3.10B-C). Additionally, we observe interesting single-cell expression kinetics for the pre-B cell receptor gene *Vpreb1,* regulator *Myc,* metabolic enzyme *Hk2,* and *Rag (1/2)* genes (Figures R3.11-3.12). Here, we leverage the pseudo-temporal ordering of 324 single-cells, where we observe cellular heterogeneity and dynamic gene expression during pre-B cell commitment.

We evaluated the differences in gene expression by comparing regular RNA-seq generated by STATegra consortium collaborators at the Karolinska Institute and scRNA-seq measurements during *Ikzf1* induction. Briefly, differential expression was performed on regular RNA-seq using limma and maSigPro (Nueda et al., 2014). Differential expression from regular and scRNA-seq for the *Ikzf1* time-courses were compared (Figure R3.13A). We found that 1,509 genes were differentially expressed in regular and single-cell measurements during pre-B cell differentiation (Figure R3.13B). We performed a correlation analysis of regular and single-cell

measurements using 1,509 differentially expressed genes and observe technical differences between regular and single-cell measurements (Figure R3.13C). We find stronger correlations between neighboring time-points (0/2 hour, 18/24 hour) during *Ikzf1* induction, both preserved in the regular and scRNA-seq data. Furthermore, we performed a hierarchical clustering on 1,509 differentially expressed genes for both regular and scRNA-seq data and observe two distinct profiles. 1,187 differentially expressed genes show a down regulation (red), whereas 322 genes show an up regulation (green) during pre-B cell differentiation in both bulk and single-cell data (Figure R3.13D). Moreover, gene ontology analysis of the shared 1,509 differentially expressed genes show a significant enrichment for lymphocyte activation (p-value < $1.2x10^{-5}$), cell cycle (p-value < $5x10^{-5}$), and metabolic process (p-value <$5.6x10^{-39}$) terms (Figure R3.13E). Our analysis combines regular and scRNA-seq time-course data to identify key biological pathways and transcriptional regulators that drive and maintain the transition states of pre-B cells.

We had identified enrichment of transcription factors from our analysis of differentially expressed genes between regular and scRNA-seq. We identified 266 transcription factors to be differentially expressed from our regular RNA-seq data. Using the single-cell data, we determined the frequency and proportion of transcription factor (266 TFs) expression in all 324 single-cells (Figure R3.14A). We find that 105 transcription factors are expressed (>1 FPKM) in less than 50 cells, while 161 are expressed in more than 50 cells. We also observe *Ikzf1* to be the only transcription factor expressed in all single-cells of our time-series. From our single-cell data, we find that 52 of 266 transcription factors are differentially expressed by our pseudo-temporal analysis and 38 of 52 transcription factors are differentially expressed in both regular and scRNA-seq. A heatmap of all 52 transcription factors shows the diverse expression across all 324 single-cells (Figure R3.14B). We find transcription factors *Foxo1/3, Foxp1,* and *Irf8* to be

differentially regulated in our time-series. Additionally, immune responsive regulators like *Rel, Nfia* and *Nfe2l1* show differential expression in both regular and scRNA-seq. Overall our analysis highlights transcription factors that exhibit differential expression supported by regular and scRNA-seq during *Ikzf1* induction.

Technical analysis of single-cell chromatin accessibility during differentiation

Our time-series analysis of the scRNA-seq revealed cellular heterogeneity and differential expression of transcriptional regulators in our *Ikzf1* inducible time-series. Additionally, from our initial DNase-seq analysis, we observed that most changes in chromatin accessibility occurred between 0 and 24 hours in our time-series. To measure changes in the chromatin accessible landscape of single-cells induced with *Ikzf1*, we performed 4 single-cell ATAC-seq (Buenrostro et al., 2015) Fluidigm C1 experiments, focusing on the 0 and 24 hour time-points. We generated of 227 single-cell ATAC-seq (scATAC-seq) profiles (Figure R3.15A) after data processing (Experimental methods). We observe good concordance of open chromatin signal between our bulk DNase-seq (~25 million cells), bulk ATAC-seq (50,000 cells) and scATAC-seq (average of 95 cells) profiles (Figures R3.15B). scATAC-seq libraries from individual C1 runs were first analyzed to ensure proper reproducibility between experiments. We considered successful library criteria to include more than 5,000 unique mapped fragments and observed that between 0 hour and 24 hour C1 experiments the mean library size was similar across all single-cells (Figures R3.16A-G). Additionally, we observe significant correlation between C1 experiments for 0 and 24-hour time-points (Figures R3.17A-B). These results highlight the reproducibility and robustness between scATAC-seq experiments using the Fluidigm C1 system.

Having observed that pre-B receptor genes *Igll1* and *Vpreb1* were down regulated in both regular and scRNA-seq data, we investigated the chromatin accessibility of these two genes in differentiating single-cells. We observe a general agreement in the signal intensity from DNase-seq, regular ATAC-seq and averaged single-cell ATAC-seq profiles for 0 hour pre-B cells (Figure R3.18A). Interestingly, we observe a great diversity in single-cell accessibility for *Igll1, Vpreb1, Top3b* promoters and the upstream enhancer of *Igll1* (Figure R3.18B). We highlight possible clusters of single-cells based on combinatorial accessibility of these elements (colored boxes). We observe that 24-hour cells show less accessibility for *Igll1* and *Vpreb1* promoter elements in DNase-seq, regular ATAC-seq and averaged single-cell ATAC-seq profiles (Figure R3.18C). This result is as expected, as 24 hour induced cells should no longer express pre-B cell receptors as they enter a resting state. Moreover, our 24-hour scATAC-seq data show that very few cells are accessible for pre BCR promoters (Figure R3.18D). This result is also consistent with our scRNA-seq observations, where few cells at 24 hours still express pre BCR genes, reflecting the heterogeneity of cellular differentiation in our B3 model system.

We performed an analysis to assess the technical characteristics and limitations of our scATAC-seq. We utilized the peaks called in the DNase-seq for our subsequent technical analysis. We estimate an average of 3,483 sites/cell are accessible, with a mean of 4 cells/site and median of 2 cells/site (Figure R3.19A). Moreover, we observe that only ~5,000 sites are accessible in 10% of all cells (n=4.8/48), with only 100 sites accessible in 50% of cells (n=24/48) (Figure R3.19B). We evaluated the extent of ATAC-seq signal from these 5,000 accessible sites in both regular ATAC-seq and single-cell ATAC-seq. We observe similar signal for sites ranked in regular and averaged single-cells (Figure R3.19C). We mapped chromatin accessibility in three single-cells with differing mapping rates and efficiencies (fragments in

peaks/ mapped fragments). Although we observe that the library for single-cell A1 has 4x lower mapped fragments and 3x less fragments in sites in comparison to library A5, the relative efficiency is slightly higher (1.7%) between the two cells (Figure R3.19C). This result shows that the number of mapped fragments for a given sample may not be the most accurate measure for assessing data quality of highly sparse scATAC-seq data.

We initially observed that few cells by 24-hours post-induction, were expressing *Rag1/2* from our scRNA-seq analysis. Similarly, we observe increased chromatin accessibility of promoter and upstream regulatory elements of the *Rag1/2* genes in bulk and scATAC-seq for 24-hour single-cells (Figure R3.20A-B). To identify patterns of accessibility we analyzed the top 6,000 accessible elements and performed hierarchical clustering. We observe distinct profiles in chromatin accessibility from our scATAC-seq profiles. Moreover, regulatory elements for *Igll1* and *Vpreb1* show maximal accessibility in 0-hour cells (Figures R3.21A-B). Overall our technical analysis of scATAC-seq provides high-resolution dynamics of chromatin accessibility that are supported by scRNA-seq during pre-B cell differentiation.

3.4 Discussion

The aims of the STATegra consortium are to develop methods for integrating multi-omics data to understand the process of B cell development. Here, we leverage a model of pre-B cell differentiation to study regulatory changes in DNaseI hypersensitivity from pooled samples and single-cell profiles of gene expression and chromatin accessibility. From our DNaseI and scATAC-seq analysis we find two distinct profiles that detail a loss and gain in accessibility over 24-hours of differentiation. These profiles were enriched for terms in B-cell activation, lymphocyte differentiation, and cell cycle.

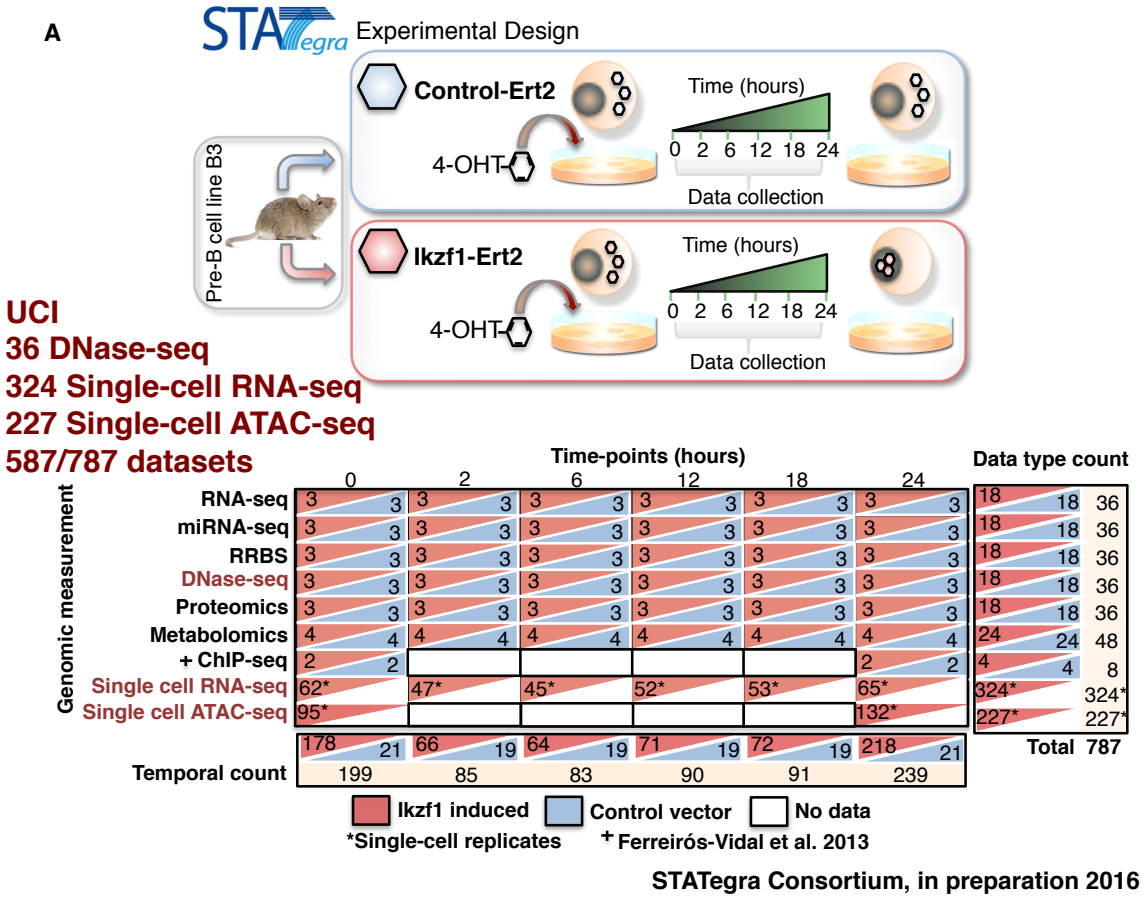
Additionally, our analysis of single-cell gene expression using monocle provided a strategy for mapping cellular differentiation trajectories using an unbiased pseudo-time. Although we observed considerable cellular heterogeneity in our model system, we show an enrichment of genes and pathways that are consistent in both regular and scRNA-seq analysis. We observe previously reported changes in gene expression for pre BCR genes during pre-B cell differentiation (Clark et al., 2014; Ferreir´ os-Vidal et al., 2013). In addition to changes in lymphocyte specific gene expression, we observe changes in metabolic enzymes. While it has been shown that transcriptional and metabolic processes are necessary for pre-B cell differentiation (Zouali, 2014), we observe these changes in differentiating single-cells. We also observed single-cell dynamic profiles of transcriptional regulators which further emphasize the role of *Foxo* and *Foxp* transcription factors during pre-B cell commitment (Ferreir´ os-Vidal et al., 2013).

The ability to profile chromatin accessibility in single-cells has shown to be a powerful application for studying cell-to-cell differences (Buenrostro et al., 2015). To this date, no studies have described the changes in chromatin accessibility in single-cells using a time-series approach. We leverage our B cell model system to analyze both regular and single-cell chromatin accessibility measurements during differentiation. Unlike scRNA-seq were several molecules of RNA can be recovered and quantified, scATAC-seq is constrained by the amount of DNA material in a single-cell. In short, scATAC-seq measurements are quantified to near binary site estimates. Thus, scATAC-seq is limited to surveying a small subset of all open chromatin elements in a single-cell genome-wide (Maurano and Stamatoyannopoulos, 2015; Pott and Lieb, 2015). Despite this technical issue, we observed distinct profiles in chromatin accessibility in cells of 0 and 24 hours-post differentiation, supported by both DNaseI and

regular ATAC-seq measurements. Moreover, we observe sub-population like chromatin accessibility for regulatory elements in single-cells (Figure R3.18). Although we cannot confirm that these observations reveal true cell sub-populations based on single-cell accessibility alone, this offers a potential application for understanding cis-regulatory element use across single-cells. Further studies employing droplet based technologies (Klein et al., 2015; Macosko et al., 2015) for scATAC and scDNase-seq (Jin et al., 2015) will allow for profiling thousands of cells in parallel and provide robust detection of the cis-regulatory landscape dynamics across many biological systems.

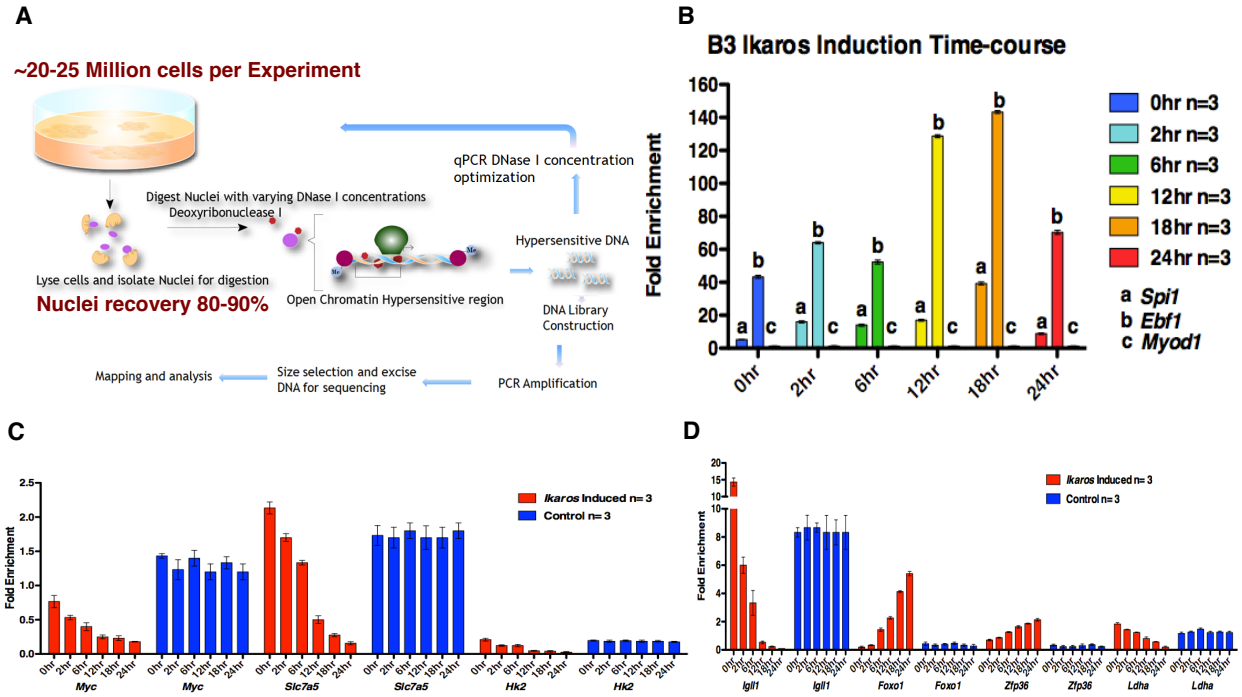
3.5 Figures

Figure R3.1 STATegra multi-omic time-course of pre-B cell differentiation



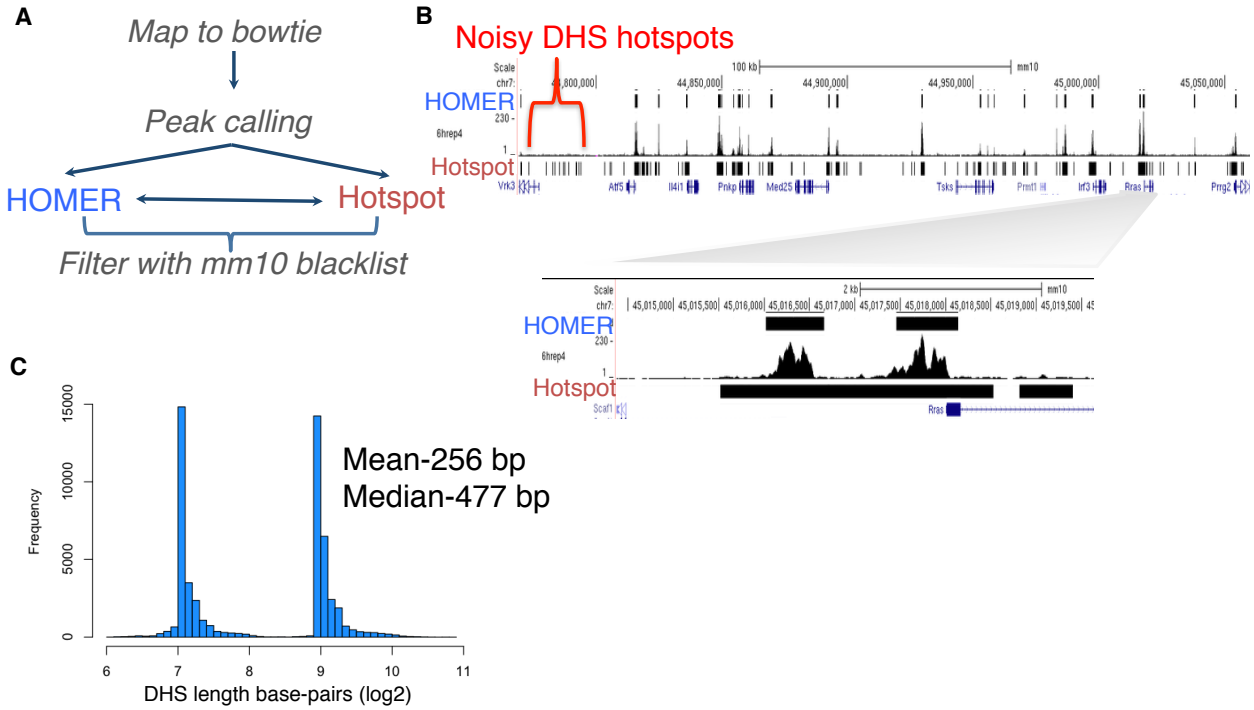
(A) STATegra consortium multi-omics experimental data collection during pre-B cell differentiation. Genomic assays are shown with number of biological replicates for each condition. In total 787 datasets are integrated and used to study pre-B cell differentiation from many genomic assays, of which 587 were generated at UCI.

Figure R3.2 DNaseI qPCR enrichment optimization and analysis



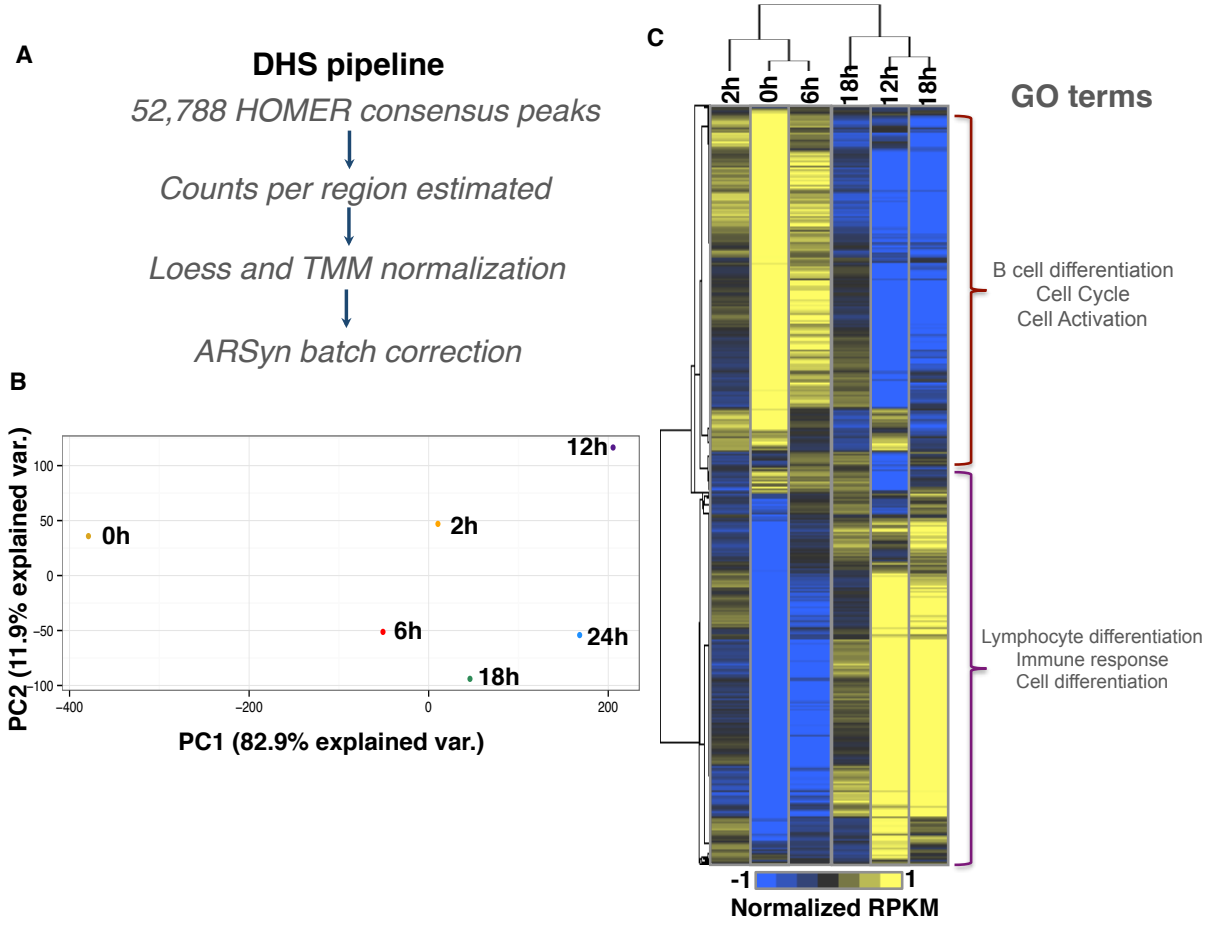
(A) DNase-seq experimental pipeline for both *Ikzf1*-induced and control time-series. (B) DNaseI qPCR fold enrichment for *Ikzf1* time-series for accessible (*Spi1, Ebf1*) and inaccessible promoters (*Myod*). (C) DNaseI qPCR fold enrichment of control and Ikaros data for promoters of several genes previously shown to change in gene expression (Ferreir´os-Vidal et al., 2013) during *Ikzf1* induction.

Figure R3.3 DNase-seq peak calling strategy



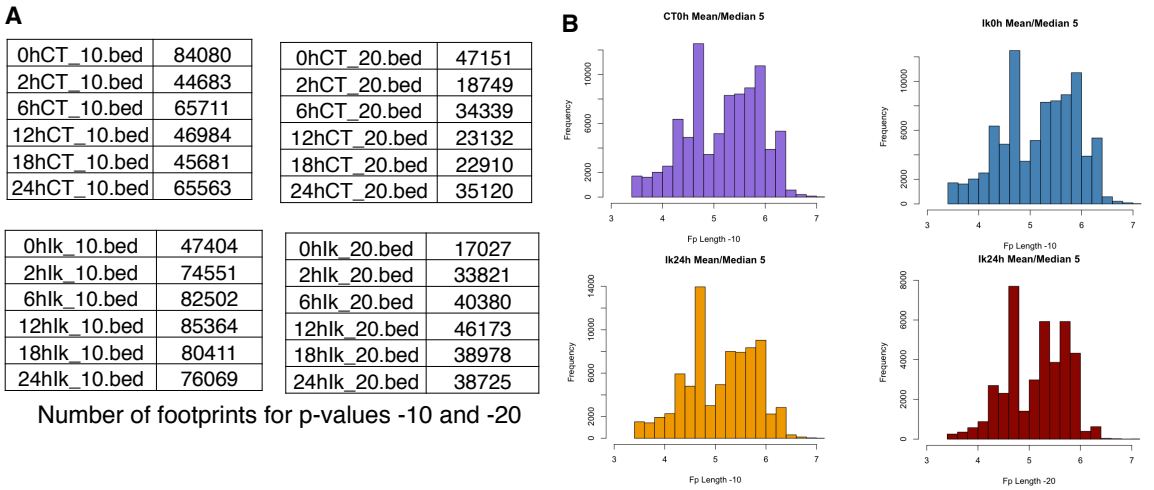
(A) DNase-seq peak calling strategy comparing HOMER and Hotspot. (B) UCSC browser shot of HOMER and Hotspot peaks. DNase-seq bigWig track is shown in black. Noisy DHS peaks called by Hotspot are shown. Zoomed-in view of *Rras* promoter showing more precise peak calling by HOMER. (C) Distribution of HOMER DHS peak size (bp) after merging narrow and broad peaks (n=52,788).

Figure R3.4 DNaseI hypersensitivity dynamics of *Ikzf1*-induction during pre-B cell differentiation.



(A) DNase-seq data processing and analysis pipeline. (B) Principal component analysis of *Ikzf1* time-series. We observe that PC1 corresponds to the time-series after Ikaros induction. (C) Hierarchical clustering of top DNaseI hypersensitive sites (7,018). Two distinct clusters of DHS loss (top) and gain (bottom) in accessibility reflect specific GO term enrichments.

Figure R3.5 DNaseI footprinting estimates for Ikaros time-series



A

0hCT_10.bed	84080
2hCT_10.bed	44683
6hCT_10.bed	65711
12hCT_10.bed	46984
18hCT_10.bed	45681
24hCT_10.bed	65563

0hCT_20.bed	47151
2hCT_20.bed	18749
6hCT_20.bed	34339
12hCT_20.bed	23132
18hCT_20.bed	22910
24hCT_20.bed	35120

0hIk_10.bed	47404
2hIk_10.bed	74551
6hIk_10.bed	82502
12hIk_10.bed	85364
18hIk_10.bed	80411
24hIk_10.bed	76069

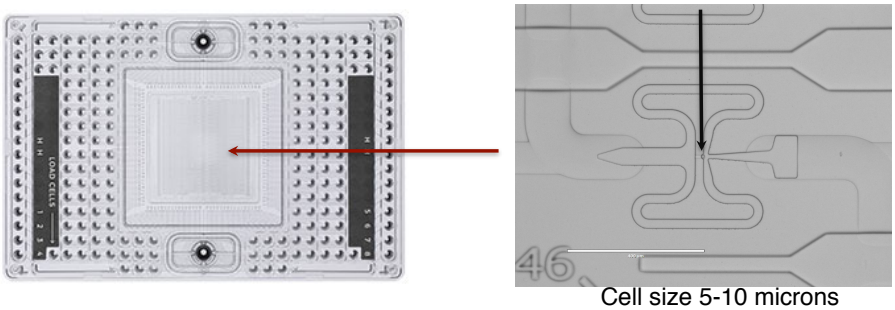
0hIk_20.bed	17027
2hIk_20.bed	33821
6hIk_20.bed	40380
12hIk_20.bed	46173
18hIk_20.bed	38978
24hIk_20.bed	38725

Number of footprints for p-values -10 and -20

(A) Tables for number of DNaseI footprints in control and Ikaros time-courses. P-values < -10 and -20 are shown. (B) Distribution of DNaseI footprint lengths are shown for control and Ikaros 0 hour, Ikaros 24 hour (p-value < -10) and Ikaros 24 hour (p-value <-20)

Figure R3.6 Fluidigm C1 integrated fluidic circuits and single-cell capture estimates

A



B

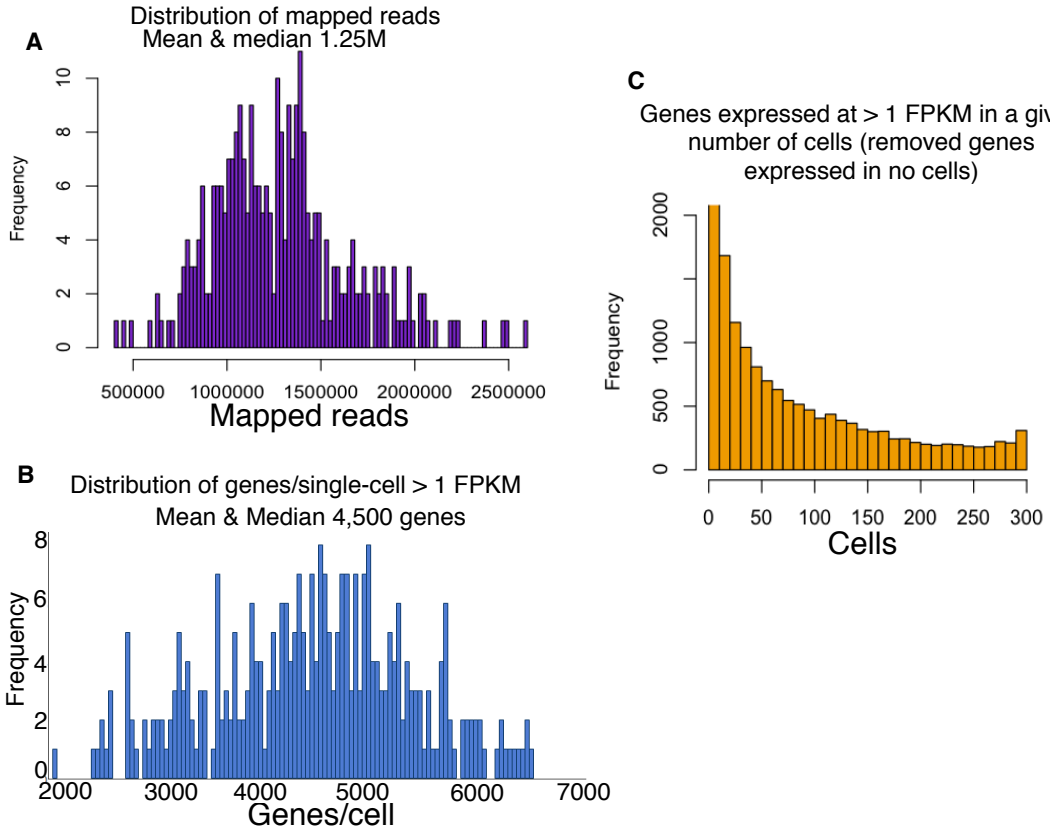
Condition	Cell libraries sequenced	Single cell libraries pass filter	Read lengths
0h Ikaros	82	62	2x75
2h Ikaros	67	47	2x75
6h Ikaros	68	45	2x75
12h Ikaros	64	52	2x75
18h Ikaros	137	53	2x75
24h Ikaros	142	65	2x75

324 single-cell libraries > 50% and/or > 500k mapped reads

(A) Fluidigm IFC microfluidic chip designed for capturing of up to 96 single-cells. Capture site shown with an Ikaros-induced B3 cell captured. B3 cell size is estimated between 5 and 10 microns. (B) Distribution of cell libraries sequenced and libraries used for downstream analysis (n=324).

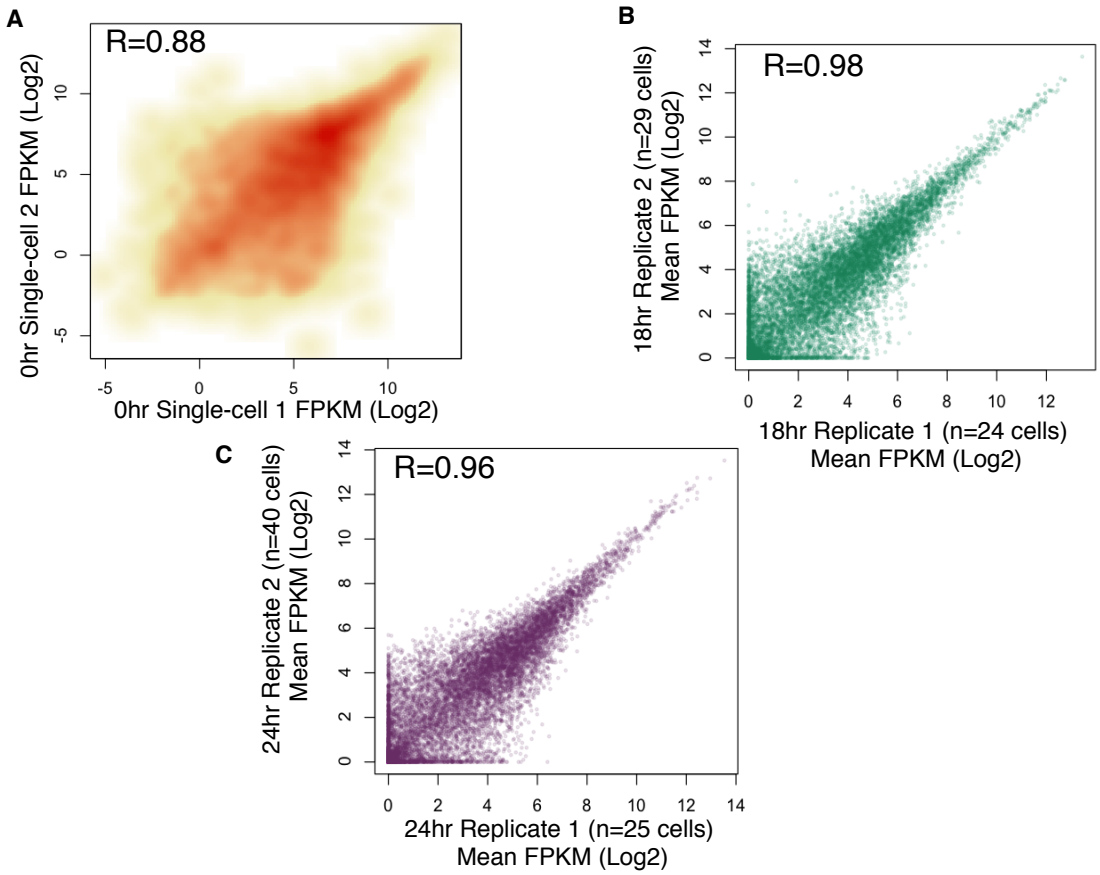
Figure R3.7 scRNA-seq gene statistics

R3.7



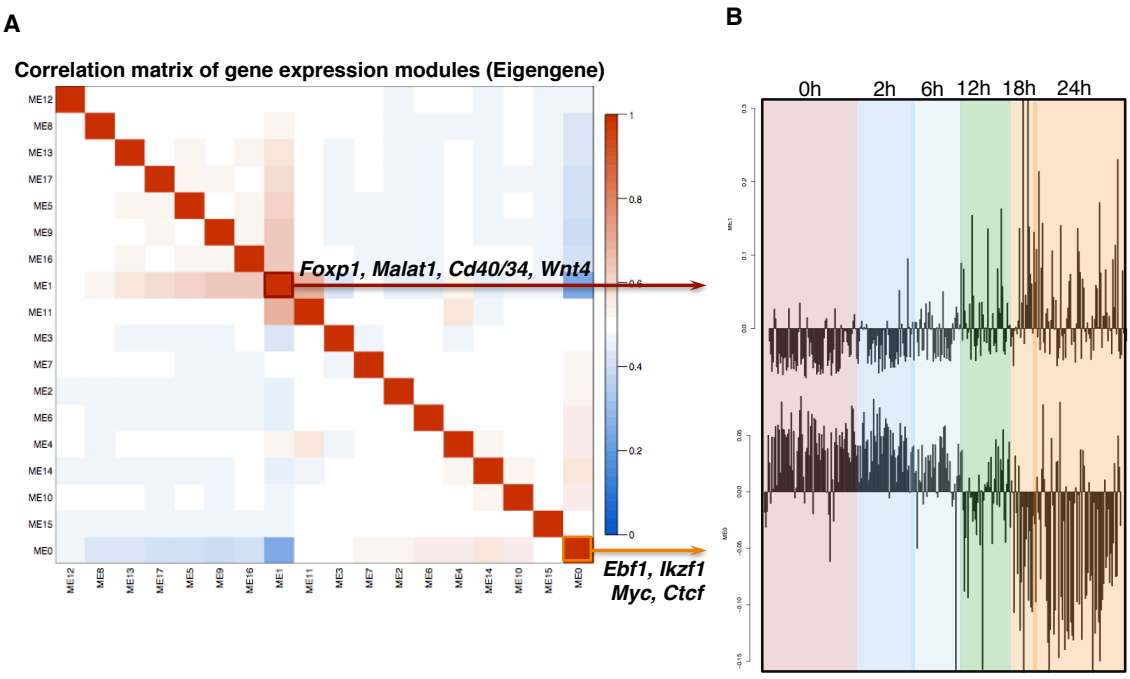
(A) Distribution of number of mapped reads for scRNA-seq libraries (n=324). Mean and median are 1.25 million mapped reads. (B) Distribution of the number of genes expressed in a single-cell (> 1 FPKM). Mean and median are 4,500 genes per single-cell. (C) Distribution for number of genes expressed in respective single-cell libraries.

Figure R3.8 Single-cell RNA-seq reproducibility



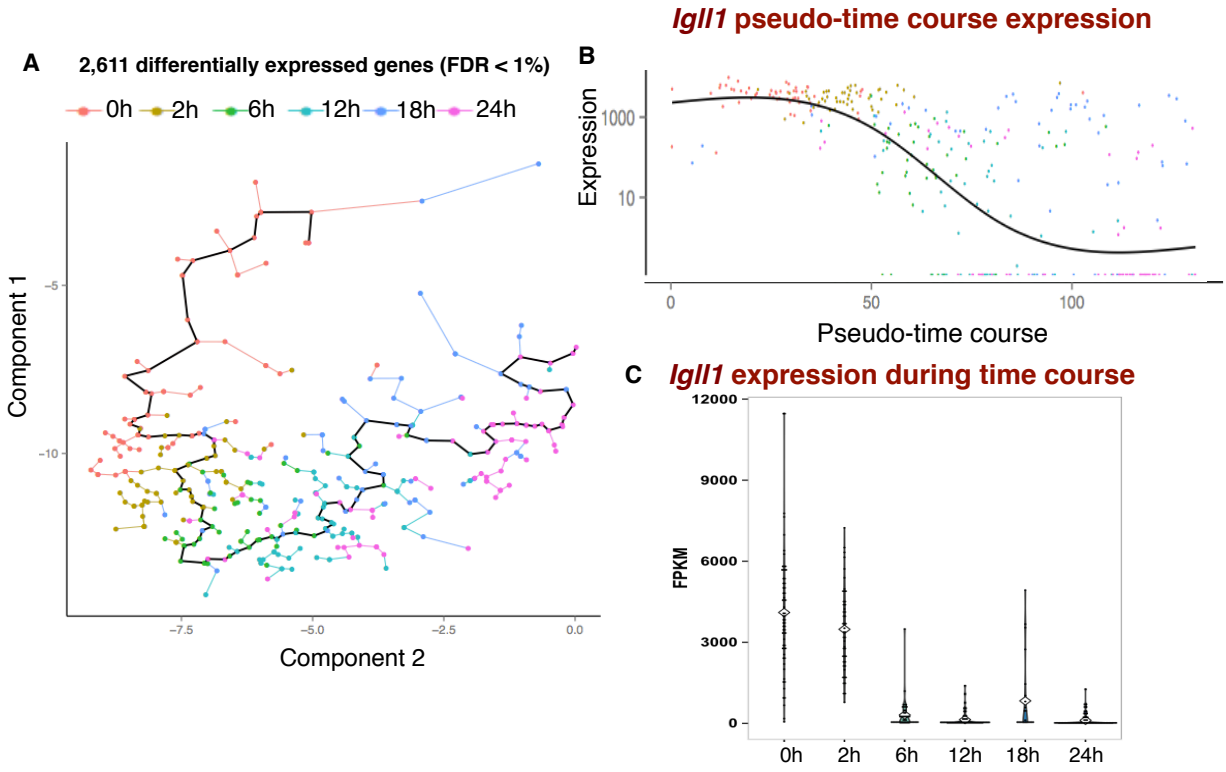
(A) Single-cell correlation for two 0h Ikaros single-cell libraries. Correlation coefficient is 0.88 for these two libraries. (B-C) Correlation for 18 and 24 hour Ikaros C1 experimental runs. Correlation coefficient is 0.98 and 0.96 respectively for 18 and 24 hours data.

Figure R3.9 Weighted gene co-expression network analysis of pre-B differentiation



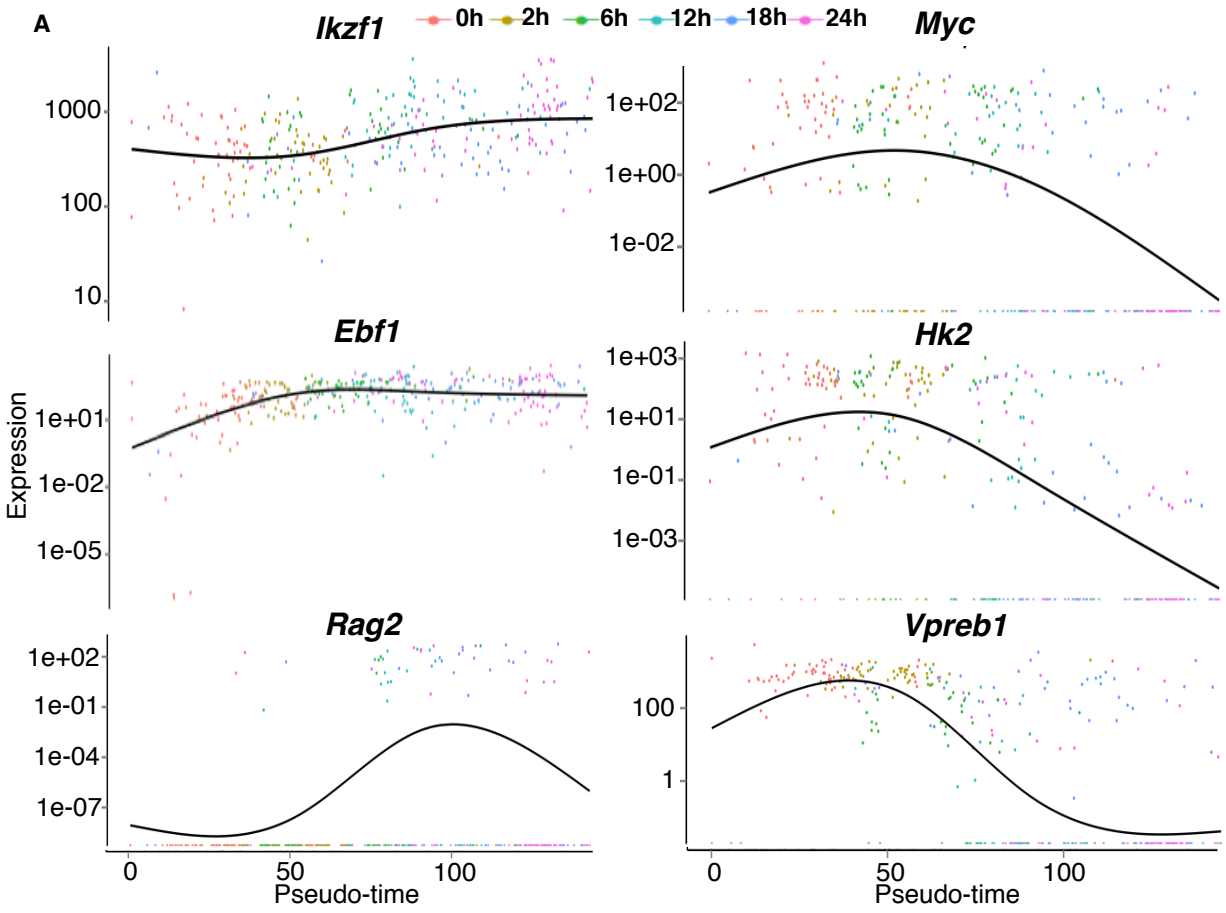
(A) Correlation matrix for all eigengene modules identified with WGCNA. Red indicates strong eigengene expression. Genes are highlighted from modules 0 and 1. (B) Fraction of *Ikzf1* cells that are enriched in modules 1 (top) and 0 (bottom) during pre-B cell differentiation.

Figure R3.10 Pseudo-temporal ordering of gene expression in pre-B cell differentiation



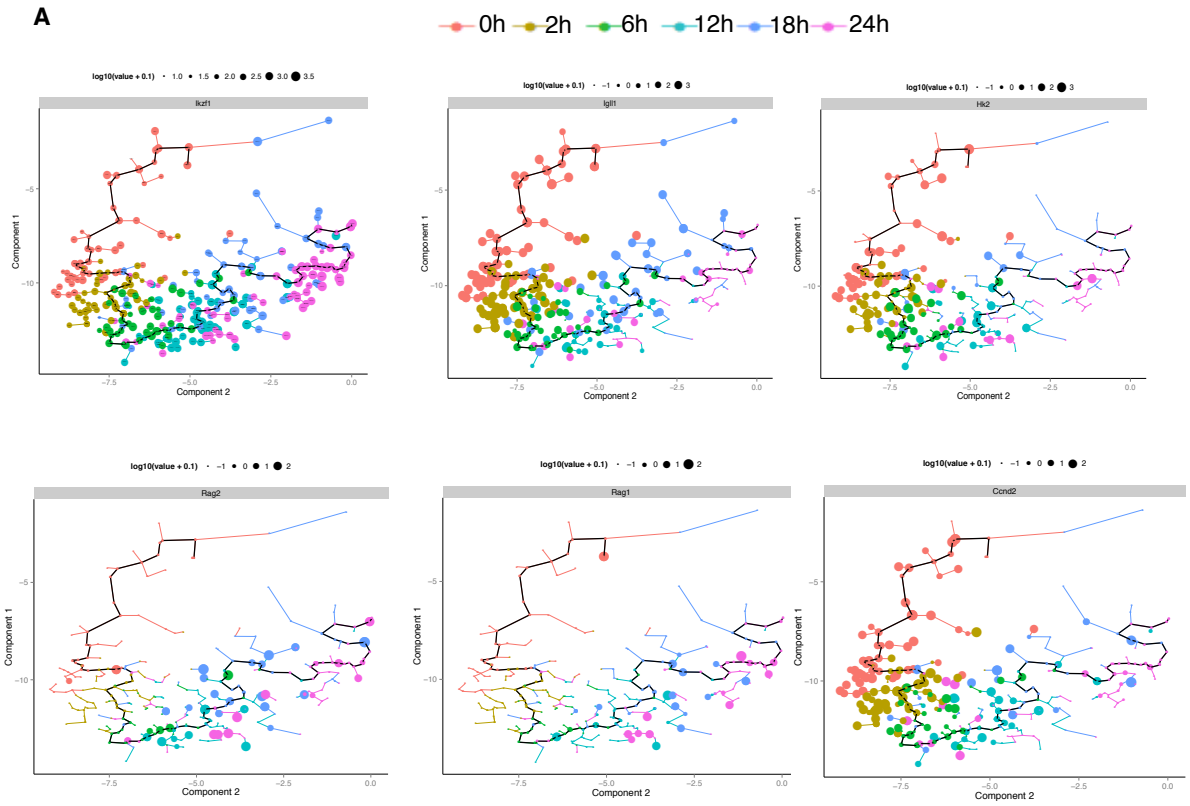
(A) Pseudo-temporal ordering of 324 single-cells using independent component analysis using Monocle. A minimum spanning tree was derived from differentially expressed genes and is shown as a black line for the trajectory of the *Ikzf1* time-series. Single-cells corresponding to time-points are colored. (B-C) *Igll1* expression kinetics are consistent for both pseudo-temporal ordering and non-ordering of single-cells.

Figure R3.11 Dynamic expression of key genes in differentiating pre-B single-cells



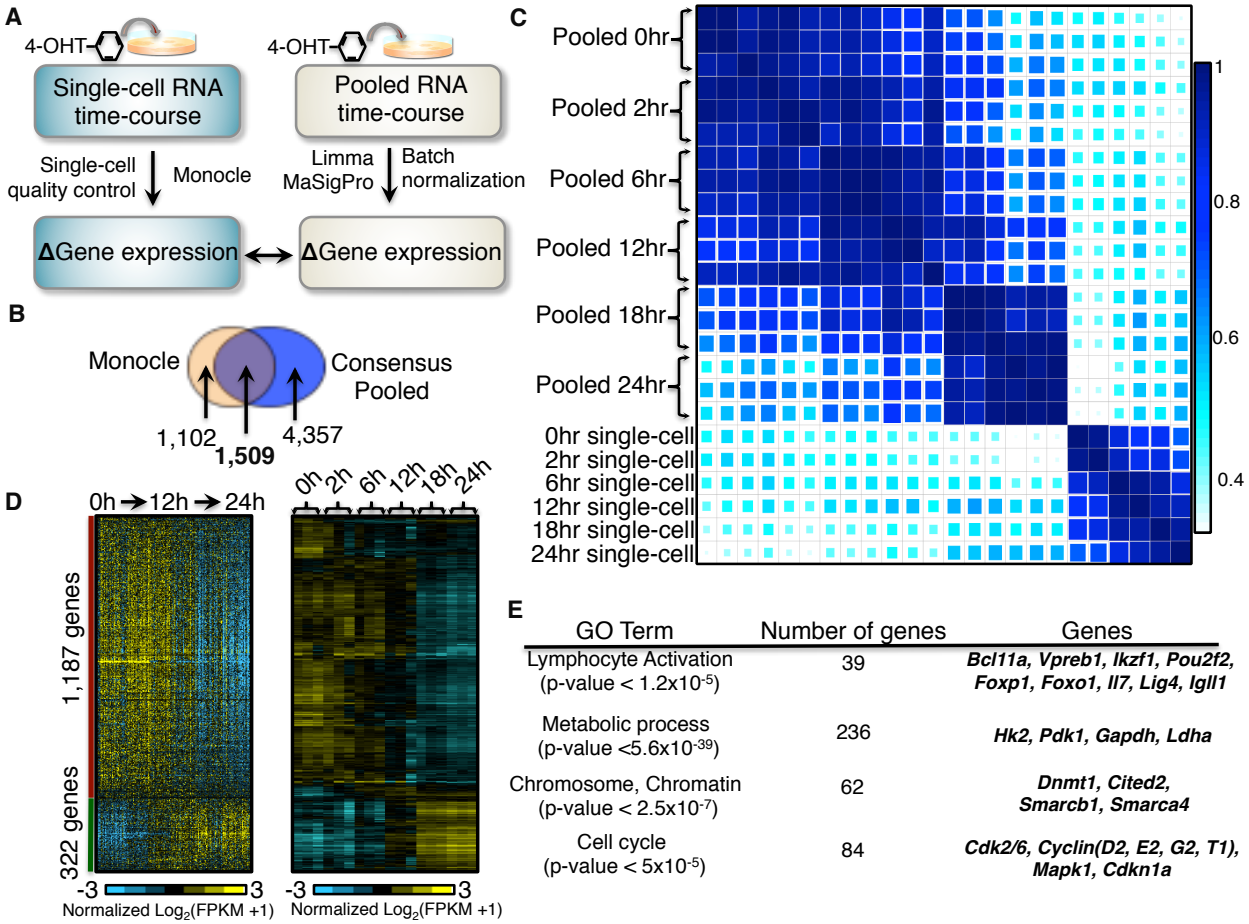
(A) Pseudo-temporal gene expression profiles for *Ikzf1, Myc, Ebf1, Hk2, Rag2*, and *Vpreb1*.

Figure R3.12 Dynamic gene expression of pseudo-temporal ordered cells compared to specific genes



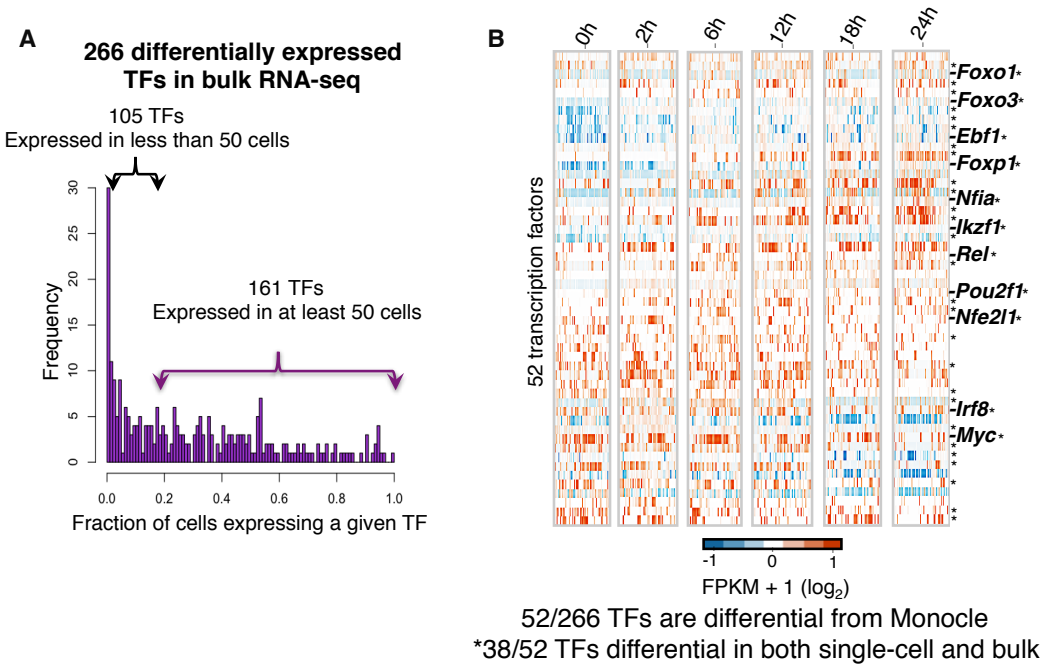
(A) Pseudo-temporal ordering of 324 single-cells using independent component analysis. A minimum spanning tree was derived from differentially expressed genes and is shown as a black line for the trajectory of the *Ikzf1* time-series. Size of circle (single-cell) indicates relative level of expression for *Ikzf1, Igll1, Hk2, Rag1/2,* and *Ccnd2*. Single-cells corresponding to time-points are colored.

Figure R3.13 Regular and single-cell RNA-seq comparative analysis



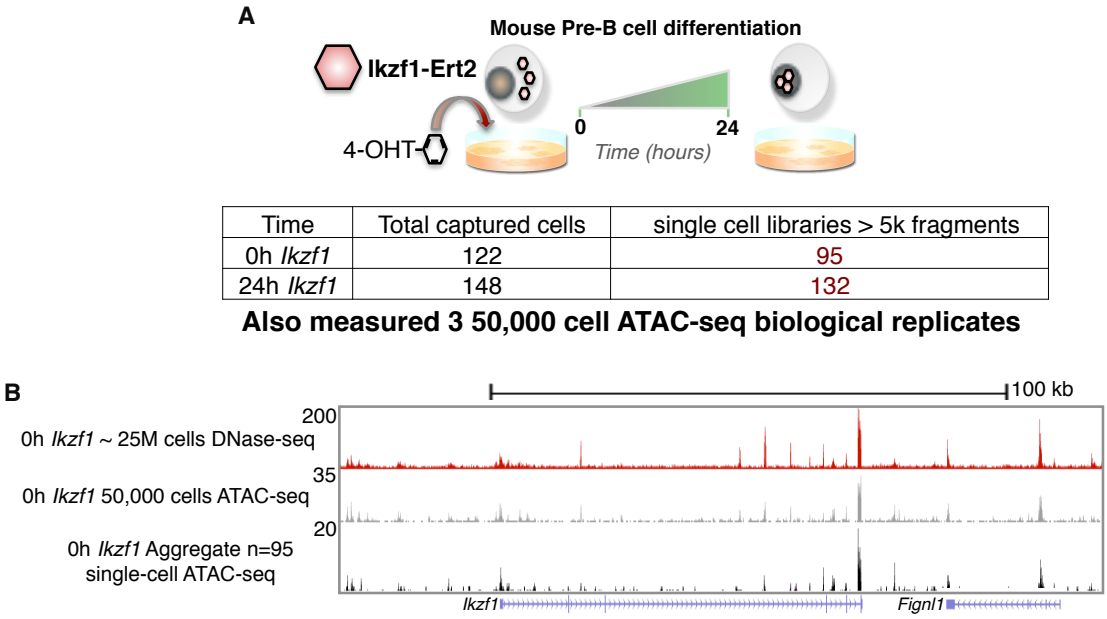
(A) Schematic of regular and scRNA-seq comparative analysis. (B) Venn diagram of shared differentially expressed genes (1,509) from determined by both regular and single cell RNA-seq. (C) Pearson coefficients between bulk and scRNA-seq. Single-cell data was averaged by time-points. (D) Heatmap of 1,509 differentially expressed genes by regular and scRNA-seq. 1,187 genes show down regulation, while 322 genes show up regulation during *Ikzf1* induction. (E) Gene ontology analysis of differentially expressed genes (1,509). Representative genes for each enriched term are shown.

Figure R3.14 Robust transcription factor expression in regular and scRNA-seq



(A) Distibution of single-cells expressing a differentially expressed transcription factor (n=266). 161 TFs are expressed in at least 50 of 324 single-cells. (B) Heatmap of 52 TFs differentially expressed in single-cells. Asterisk (*) indicates genes that are differential in both regular and scRNA-seq (n=38).

Figure R3.15 Profiling the accessible chromatin landscape in single-cells

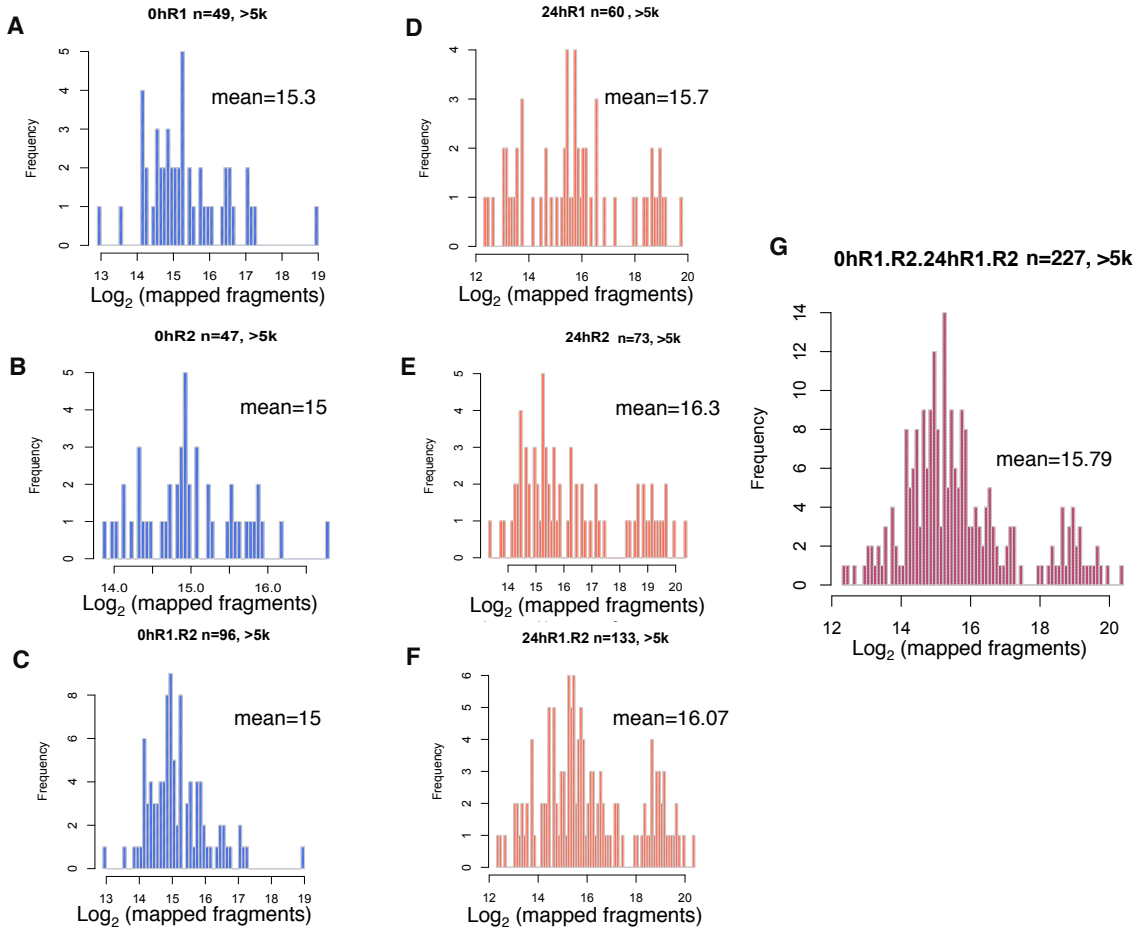


Time	Total captured cells	single cell libraries > 5k fragments
0h *Ikzf1*	122	95
24h *Ikzf1*	148	132

Also measured 3 50,000 cell ATAC-seq biological replicates

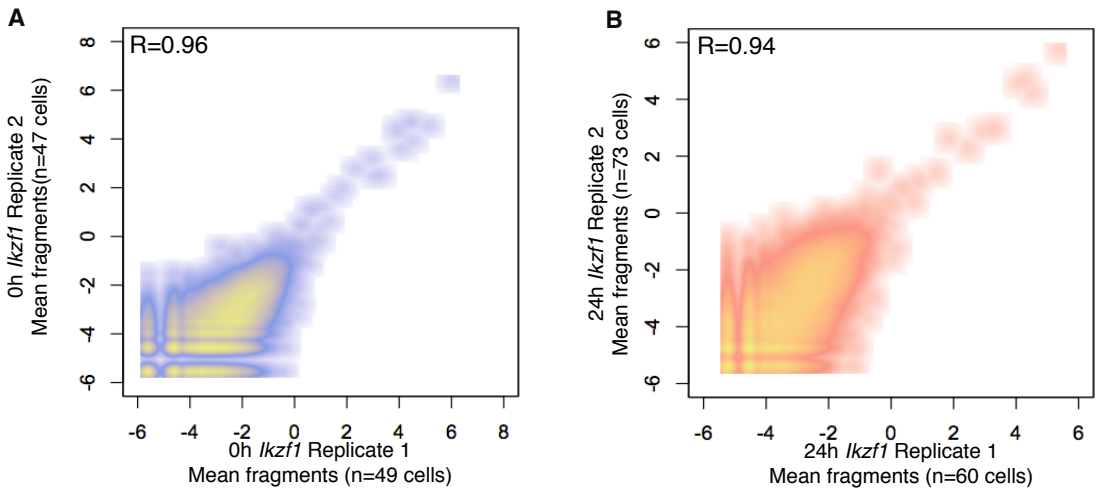
(A) scATAC-seq experimental design. 0 and 24 hour *Ikzf1* cells are profiled using the C1 Fluidigm. 227/270 single-cells passed data quality threshold for downstream analysis. (B-C) UCSC genome browser view of the *Ikzf1* loci for bulk DNaseI (red), ATAC-seq (grey), and scATAC (black, n=95) chromatin accessibility. Single-cells for the 0 hour time-point were averaged.

Figure R3.16 Mapped fragment distribution between Fluidigm C1 experiments



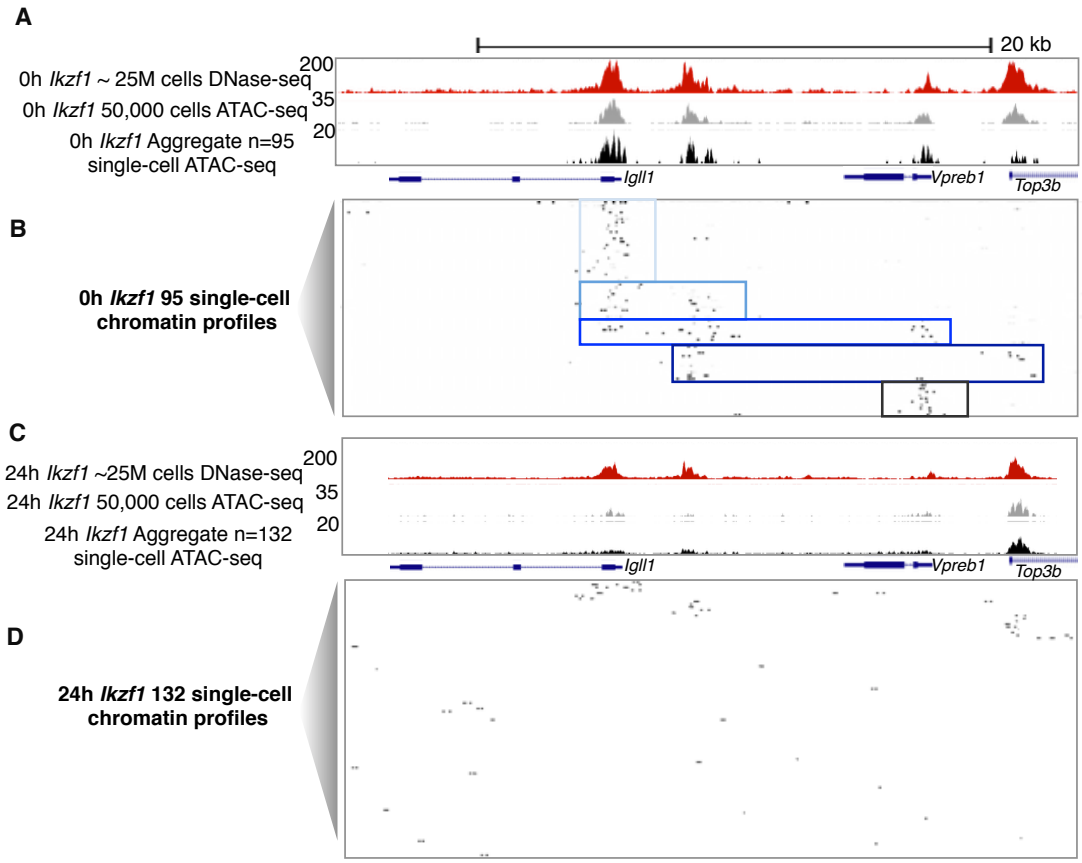
(A-C) Distribution for the number of mapped fragments per 0 hour single-cell in C1 experiment 1 (A), C1 experiment 2 (B) or combined (C). (D-F) Distribution for the number of mapped fragments per 24 hour single-cell in C1 experiment 1 (D), C1 experiment 2 (E) or combined (F). (G) Distribution for the number of mapped fragments for all single-cell ATAC-seq libraries (n=227).

Figure R3.17 Mapped fragment reproducibility between Fluidigm C1 experiments



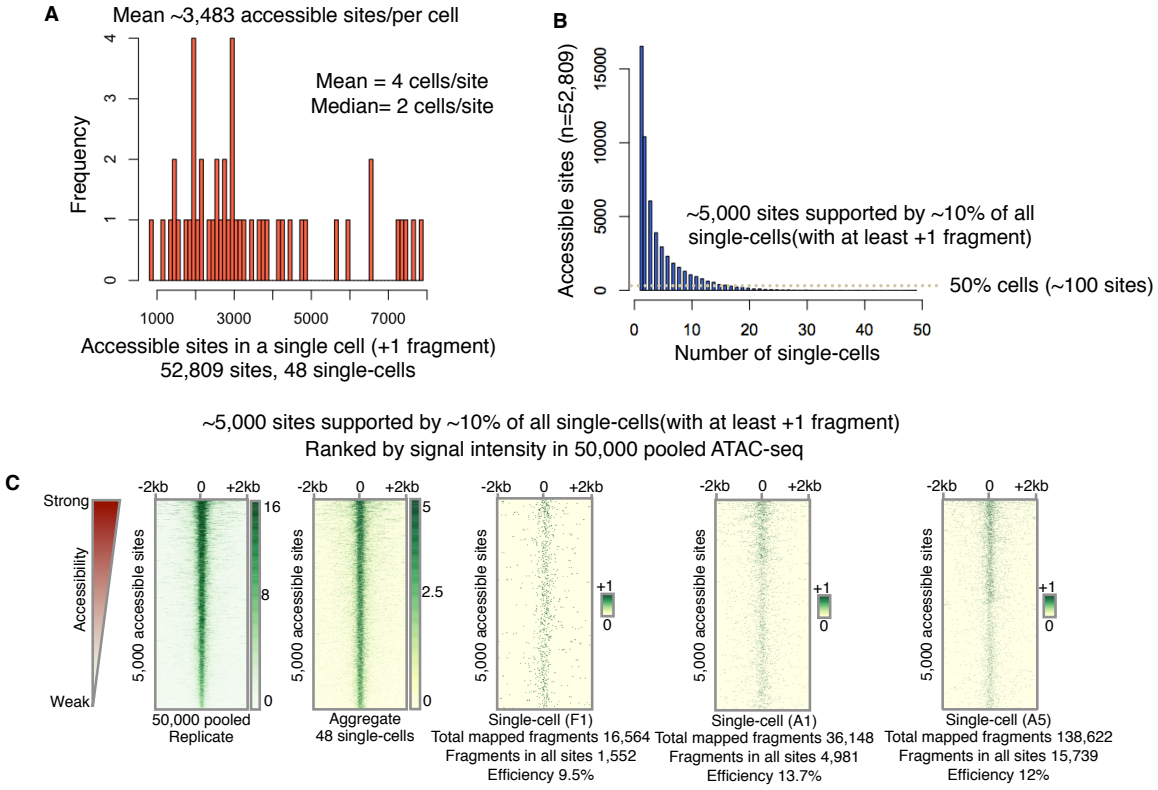
(A-B) Correlation plots of 0 hour C1 Fluidigm experiments (A, R=0.96) and 24 hour C1 Fluidigm experiments (B, R=0.94).

Figure R3.18 Heterogeneous chromatin accessibility in pre-B single-cells



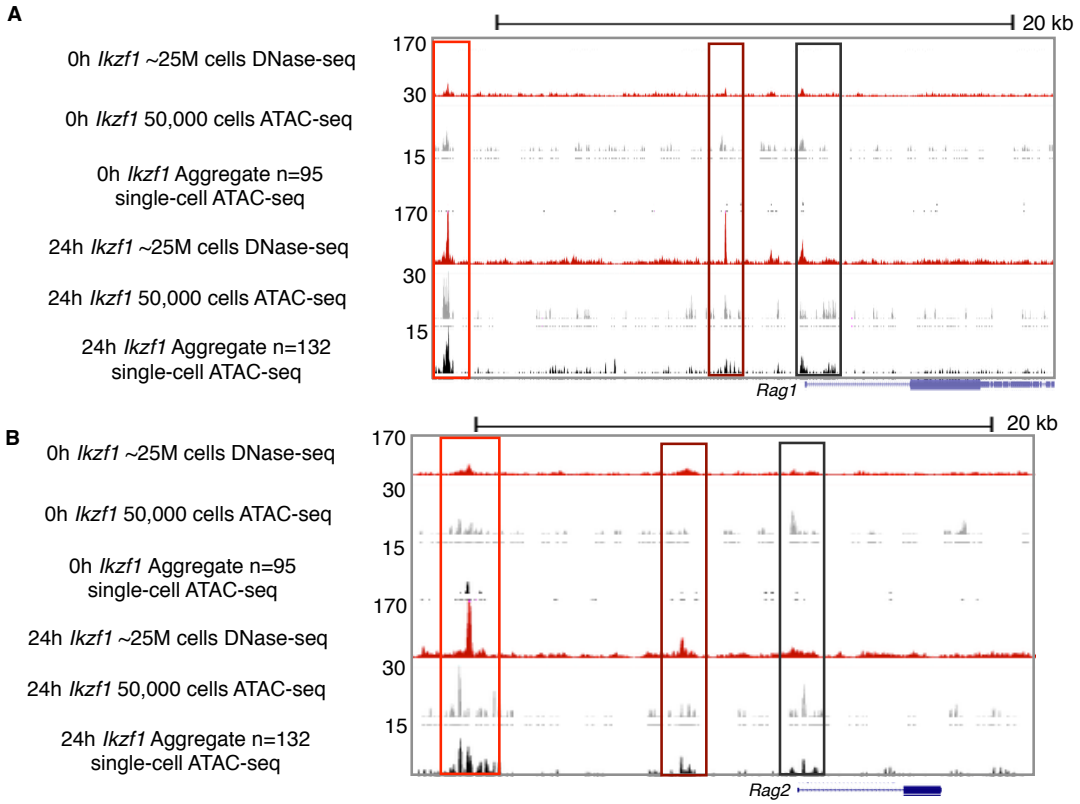
A) UCSC genome browser view of the *Igll1* locus for regular DNaseI (red), ATAC-seq (grey), and scATAC (black, n=95 single-cells) chromatin accessibility. Single-cells for the 0 hour time-point were averaged. (B) 95 single-cell chromatin profiles (0 hour) for the respective locus. (C) UCSC genome browser view of the *Igll1* locus for bulk DNaseI (red), ATAC-seq (grey), and scATAC (black, n=132 single-cells) chromatin accessibility. Single-cells for the 24-hour time-point were averaged. (D) 132 single-cell chromatin profiles (24 hour) for the respective locus.

Figure R3.19 Technical characteristics of scATAC-seq libraries



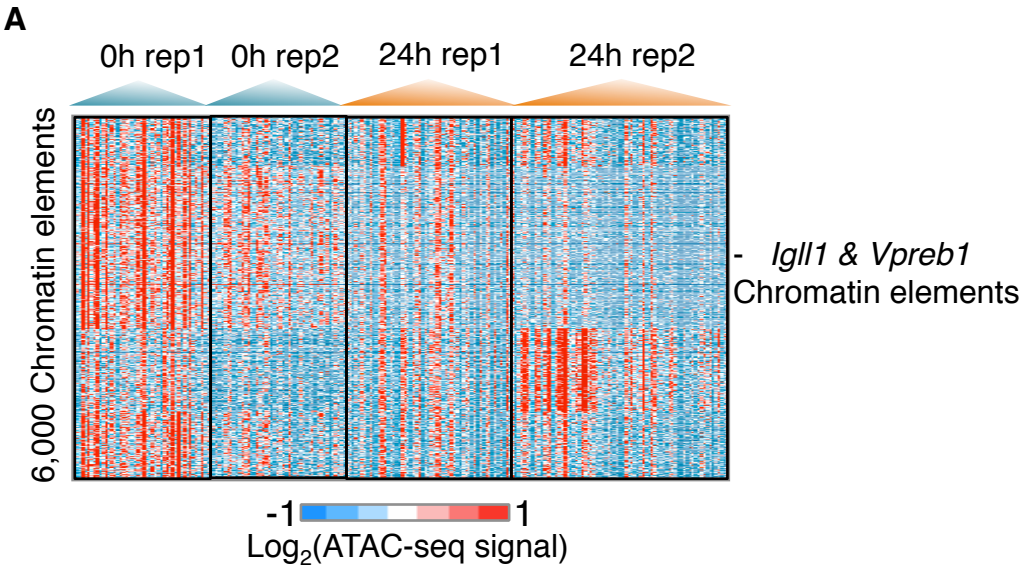
(A) Distribution for the number of accessible sites per single-cell (n=48, mean 3,483 sites/cell). On average we observe 4 single-cells per site, median 2 cells per site. (B) Plot of accessible sites in a given single-cell (n=48). 100 sites are accessible in 50% of cells (n=24). 5,000 sites are supported in 10% of all single-cells (n=5 cells). (C) Chromatin accessible heatmaps (5,000 sites) for regular (50,000 cells), pooled single-cells (n=48) and three individual single-cell libraries. Mapped fragment estimates, fragments in accessible sites, and efficiencies are shown for each single-cell library.

Figure R3.20 Dynamic cis-regulatory landscape of Rag1/2 during pre-B differentiation



(A-B) UCSC genome browser view of the *Rag1* (A) and *Rag2* (B) loci for regular 0h/24h DNaseI (red), 0h/24h ATAC-seq (grey), and 0h/24h scATAC (black, n=95 and n=132) chromatin accessibility. Single-cells for the 0 and 24 hour time-points were averaged. Promoters (black) and upstream regulatory elements (red) are boxed respectively.

Figure R3.21 Single-cell cis-regulatory landscape during pre-B differentiation



(A) Heatmap 6,000 chromatin elements for 0 and 24 hour scATAC-seq data. C1 experimental replicates are indicated for both time-points.

3.6 Methods

DNase-seq

DNase-seq was performed on ~20-25 million cells with 3 biological replicates for each time-points (0-24 hours) and conditions (Ikaros-inducible and control). Briefly, cells were harvested and washed with cold 1X PBS, prior to nuclei lysis. Lysing conditions were optimized to ensure >90% recovery of intact nuclei. DNaseI concentrations were titrated on Ikaros-inducible and control cells using qPCR against known positive DNaseI hypersensitive promoters (*Ap2a1, Ikzf1, Igll1*) and negative inaccessible hypersensitive promoters (*Myog, Myod*) in our biological system, thereby reducing excessive digestion of DNA. Enrichment of DNaseI hypersensitive fragments (0-500bp) was performed using a low-melt gel size selection protocol and built into libraries, which were sequenced as 43bp paired-end NextSeq 500 Illumina reads at a minimum depth of 20 million reads per each biological replicate. For DNaseI footprinting analysis, libraries were further sequenced and merged to achieve a minimum of 200 million mapped reads.

DNase-seq mapping, peak and footprint calling

DNase-seq reads were trimmed to 36bp and paired-end mapped to the mm10 reference genome using Bowtie (Langmead et al., 2009) with options: -v 2 -k 1 -m 1 --best –strata. DNase-seq peaks were called for each replicate using the HOMER findPeaks function. We employed a specific peak-calling strategy to capture several features of our DNaseI hypersensitive sites. Our strategy was to include both ‘narrow’ and ‘broad’ DHS peaks in our analysis. This captured a comprehensive set of sites with a wide DHS dynamic range. Initially, we used HOMER to determine narrow DHS peaks using a default size parameter (120-150bp) with a minimum peak

distance of 50bp between DHS and an FDR of 1%. We then included a second round of peak calling, restricting to a peak size of 500bp with a minimum peak distance of 50bp between DHS peaks and an FDR of 1%. We then merged the two peak sets for each replicate. We required a minimum 1bp overlap of peaks across all three biological replicates for each time-point respectively and generated a consensus DHS peak list across all time-points.

DNase-seq quantification & normalization

The consensus DHS peaks (53,624) were filtered to remove peaks on the mitochondrial DNA, chromosome fragments, and mouse ENCODE blacklist regions. Counts were estimated for each consensus DHS using the Bedtools coverageBed function (Quinlan and Hall, 2010). Additionally, no DHS were considered with less than 10 reads (~1 RPM) in all time-points. We used 52,788 consensus DHS peaks in our biological system for all subsequent analysis. Counts were estimated for each DHS regions on each dataset and normalized using loess and TMM. ARSyn was used to batch correct data and converted to RPKM values. Consensus DHS peaks were subsequently scanned for chromatin footprints using the Wellington algorithm with the following parameters (Piper et al., 2013).

Single-cell RNA-seq

Single cells were isolated on the Fluidigm C1 System using the smallest IFC (5-10 um) based on the estimated size of B3 cells. Briefly, cells were collected for each time-point at a concentration of 400 cells/μl in a total of 50 μl. To optimize cell capture rates on the C1, buoyancy estimates were optimized prior to each run. Our C1 single-cell capture efficiency was ~75-90% across 8 C1 runs. Each individual C1 capture site was visually inspected to ensure

single-cell capture and cell viability. After visualization, the IFC was loaded with Clontech SMARTer kit lysis, RT, and PCR amplification reagents. After harvesting, cDNA was normalized across all libraries from 0.1-0.3 ng/μl and libraries were constructed using Illumina's Nextera XT library prep kit per Fluidigm's protocol. Constructed libraries were multiplexed and purified using AMPure beads. The final multiplexed single-cell libraries were analyzed on an Agilent 2100 Bioanalyzer for fragment distribution and quantified using Kapa Biosystem's universal library quantification kit. The pooled libraries were normalized to 2 nM and sequenced as 75bp paired-end dual indexed reads using Illumina's NextSeq 500 system at a depth of ~1.0-2.0 million reads per cell. Each Ikaros time-point was performed once, with the exception of 18 and 24 hour time-points, in which two C1 runs were required in order to achieve approximately ~50 single-cells per each time-point.

Single-cell RNA-seq mapping, quantification, and pseudo-time analysis

A total of 560 single-cell RNA-seq libraries were mapped with Tophat (Trapnell et al., 2009) to the mouse Ensembl gene annotations and mm10 reference genome. Single-cell libraries with a mapping rate less than 50% and less than 450,000 mapped reads were excluded from any downstream analysis, resulting in 324 single-cells for all subsequent analysis. Cufflinks (Trapnell et al., 2010) version 2.2.1 was used to quantify expression from single-cell libraries using Cuffquant. Gene expression measurements for each single-cell library were merged and normalized into a single data matrix using Cuffnorm. Monocle (Trapnell et al., 2014) was used to determine the pseudo-time trajectory of Ikaros-induced cells across our time-series. See supplemental code Monocle parameters.

ATAC-seq

ATAC-seq was performed as previously described (Buenrostro et al., 2013) with the addition of a DNA size selection step after library generation to enrich for accessible chromatin ranging from 100-400bp. ATAC-seq libraries were sequenced as paired-end 43bp reads on the Nextseq 500 Illumina platform. Approximately 20-25 million ATAC-seq reads were generated per library.

ATAC-seq and scATAC-seq data processing

Single-cell libraries were mapped with Bowtie (Langmead et al., 2009) to the mm10 reference genome using the parameters -S -p 2 --trim3 10 -X 2000. Duplicate fragments were removed using Picard (http://picard.sourceforge.net) as previously performed (Buenrostro et al., 2015). We only analyzed single-cell libraries with more than 5000 fragments after mapping and duplication removal. Bulk ATAC-seq replicates were mapped to the mm10 reference genome using the parameters -S --trim3 10 -p 32 -m 3 -k 1 -v 2 --best -X 2000. Peak calling was performed on bulk replicates using HOMER with the command findPeaks <tags> -o <output> -localSize 50000 -size 150 -minDist 50 –fragLength 0. The intersection of peaks in three biological replicates were retained. A consolidated list of peaks was generated from the union of peaks from 0 and 24 hour time-points. Fragment counts for single-cell and bulk samples were estimated from the consolidated peak and compiled into a single experiment data matrix.

Chapter 4: Discussion and perspective

One of the most complete gene regulatory networks constructed to date describes the first 6-30 hours of sea urchin development (http://sugp.caltech.edu/endomes/). While these networks were primarily constructed using gene expression information, each linkage was systematically validated using perturbation-based experiments. The arduous task of perturbing each gene and measuring its individual effect on the entire gene network, has provided a deep and mechanistic understanding of sea urchin development (Howard and Davidson, 2004; Longabaugh et al., 2005; Peters and Davidson, 2015). This approach systematically validates a regulatory role for each linkage defined. These linkages can then be evaluated to assess their importance for maintaining the gene regulatory circuitry. Perturbation experiments were critical to constructing the sea urchin gene regulatory network, and similar approaches are also necessary to validating mammalian GRNs, which is complicated by the multiplicity of paralogs that would need to be analyzed. A clear next step would be the siRNA-mediated perturbation experiments for several key regulators during HL-60 differentiation to validate the importance of several regulatory linkages identified in our myeloid-derived networks. The information from these gene knockdown experiments would feed back into refining our gene regulatory networks. We foresee a long program of iterative dissection using perturbation-based experiments to validate the key linkages in our dynamic myeloid gene regulatory circuits that would allow us to move beyond the "draft" designation to a "validated" state similar to the sea urchin GRN.

The ability to study individual single-cells has revealed the complex biological nature of cellular heterogeneity often lost in pooled data at both the transcriptional and epigenetic level (Elowitz et al., 2002; Lacramioara Bintu, 2016). Moreover, functional applications utilizing single-cell technologies have expanded the current view of the dynamic hematopoietic system

through identification of sub-populations of granulocytic progenitors and diverse transcriptional programs (Lara-Astiaso et al., 2014; Paul et al., 2015). Using single-cell RNA-seq, we mapped the pseudo-temporal trajectory of 324 differentiating pre-B cells as part of the STATegra project. Our analysis revealed the dynamic gene expression profiles and heterogeneous behavior of our pre-B model system over the first 24 hours of *Ikzf1* induction. Additionally, we observed significant enrichment for signaling pathways and key transcriptional regulators that show time-specific expression dynamics in single-cells. To gain insight into the cis-regulatory landscape of differentiating pre-B single-cells, we performed single-cell ATAC-seq (Buenrostro et al., 2015) on 0 and 24 hour *Ikzf1* induced cells. This analysis of 227 single-cell chromatin profiles was challenging due to the sparse nature of scATAC-seq data. We observed predominant changes in accessibility for elements of *Igll1, Vpreb1* and *Rag1/2* genes, which correlated with the changes in single-cell gene expression. Interestingly, we observed heterogeneity for site accessibility in single-cells. It is possible that this observed heterogeneity for site accessibility is a result of technical or biological variation. From our technical analysis we observed that only about 100 sites were accessible in more than 50% of cells analyzed and show that only 5,000 sites we accessible in 10% of the cells in our scATAC-seq datasets. Because single-cell accessibility estimates are near binary, increasing the number of single-cells analyzed will allow for a more reliable assessment of whether the observation of variations in single sites accessibility is due to technical or true biological difference between subsets of single-cells.

Conrad Waddington's 'epigenetic landscape' concept initially described the genetic robustness of a developing organism (Waddington, 1957). The 'epigenetic landscape' is now generally applied to explain the cellular commitment decisions and reprogramming potential of individual cells (Ladewig et al., 2013; Nicol-Benoit et al., 2013). While the power of cellular

reprogramming using induced pluripotent cells (iPSC) (Takahashi and Yamanaka, 2006; Takahashi et al., 2007) has transformed the biomedical field in developing novel applications and therapeutics for the study of human disease (Soldner and Jaenisch, 2012a, 2012b), much of the transcriptional control that mediates this process is still under investigation. Less understood is the concept of cellular transdifferentiation. This phenomenon was first described in the observation of cuticle producing cells transdifferentiating into salt-secreting cells in silk moths undergoing morphogenesis (Selman and Kafatos, 1974). Moreover, the first observation of transdifferentiation in mammalian cells was described for conversion of mouse embryonic fibroblast into myoblasts following induced expression of the transcriptional regulator MyoD alone (Davis et al., 1987). Additionally, Th17 transdifferentiation into T regulatory cells and pre-B cell transdifferentiation into macrophages has also been described (Gagliani et al., 2015; Van Oevelen et al., 2015). Because HL-60 cells can be terminally differentiated into macrophage, monocyte and neutrophil cells, it would be interesting to explore the transdifferentiation potential of these myeloid progenitors into non-myeloid cells. We could use CRISPR dCas9-VP64 (Maeder et al., 2013) and several guide RNAs targeting the activation of the MYOD promoter to see whether HL-60 cells induced into myoblasts. Robust and rationale reprogramming of one cell state into a completely different cell state is a proof of principle that we understand the key regulatory mechanisms controlling the identity of the starting and ending cell states.

The ability to generate high-throughput single-cell measurements for several biochemical assays has become a powerful tool for studying complex biological systems. There is therefore a growing need for new computational approaches and tools for analyzing single-cell results from multiple assays. Existing computational approaches include self-organizing maps (SOM), pseudo-temporal ordering, diffusion mapping and dimensionality reduction have all been used

for analyzing single-cell RNA-seq data (Haghverdi et al., 2015; Kim et al., 2015; Satija et al., 2015; Setty et al., 2016; Trapnell et al., 2014). However, the development and implementation of these approaches for integrating multiple genomic platforms remains unresolved for single-cell analysis.

Self-organizing maps (SOM) are an artificial neural network that are trained using unsupervised learning to reduce high-dimensional data into a low-dimensional representation . The representation of the SOM can be viewed as a two-dimensional map, in which each node is weighted and has the same dimension as the average of input data vectors associated with that node. Moreover, the position of the node within the two-dimensional map represents similarity between data represented in each node. Previously, SOMs have been used to analyze large-scale datasets comparing multiple cell types with ChIP-seq and DNase-seq data both independently and in some cases integratively (ENCODE Consortium Project., 2012; Mortazavi et al., 2013; Sheffield et al., 2013; Thurman et al., 2012; Yue et al., 2014). For example, Mortazavi et al. used self-organizing maps to identify candidate cell-type-specific enhancers from multiple ENCODE cell types. A single-cell experiment is capable of producing up to thousands of single-cell data sets (Klein et al., 2015; Macosko et al., 2015) at once, which can be computationally challenging to analyze. SOMs have also been used to study the dynamics of cellular reprogramming in single-cells using scRNA-seq, which revealed that the dynamic activation of lncRNAs occurs during cellular reprogramming and that activated lncRNAs are also capable of repressing lineage-specific genes (Kim et al, 2015). Additionally, the SOM analysis revealed dynamic single-cell profiles that highlight the heterogeneity and transition of each cell during reprogramming (Kim et al., 2015). The application of SOMs to the STATegra scRNA-seq and scATAC-seq time-course data should allow us to view the dynamic transition of gene expression

and chromatin accessibility in single-cells during pre-B cell differentiation. If we could fuse those SOMs and incorporate dynamic scRNA-seq and scATAC-seq data, we could mine the relationships between the different subsets present in our one-datatype single-cell data in a way that is currently impossible to do from pooled data.

Longer term, the ability to reconstruct gene regulatory networks accurately from high-throughput functional genomics methods such as RNA-seq and ATAC-seq in the course of cell differentiation or embryonic development will be key to understanding how development is encoded in the genome. Even now with the ability to perform single-cell genomic assays that measure gene expression, chromatin accessibility, methylation state, and chromatin architecture in single-cells, utilizing this information in a dynamic system remains a challenge. Future studies of human myeloid differentiation in a time-specific manner in single-cells should reveal the underlying transcriptional circuits of a single-cell. While a time-series of differentiating cells would offer much insight, this would only provide temporal snapshots of individual cell profiles since individual captured cells can no longer be traced and profiled at subsequent time points. The ability to probe biochemical measurements from a single-cell without its destruction would reveal the individual dynamic circuits of any respective cell, which is currently being done in a low throughput manner using live imaging. Additionally, few technologies have begun to profile multiple biochemical assays that include the DNA methylation state, gene expression profiles, and genome sequence from the same individual cell (Angermueller et al., 2016; Hou et al., 2016; Macaulay et al., 2015). Expanding on the notion of simultaneous profiling, measuring chromatin accessibility and gene expression would yield a powerful approach for linking the cis-regulatory landscape and corresponding expression profile of a single-cell. As the technologies to probe the nature of single-cells become central, so does the requirement for proper analytical methods to

analyze ever more complicated datasets. In summary, approaches that integrate multiple genomic assays and incorporate single-cell data into gene regulatory network reconstruction will be critical to gain further insights into the biological logic of development.

References

Abkowitz, J.L., Catlin, S.N., McCallie, M.T., and Guttorp, P. (2002). Evidence that the number of hematopoietic stem cells per animal is conserved in mammals. Blood *100*, 2665–2667.

Alizadeh, a a, Alizadeh, a a, Eisen, M.B., Eisen, M.B., Davis, R.E., Davis, R.E., Ma, C., Ma, C., Lossos, I.S., Lossos, I.S., et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature *403*, 503–511.

Amulic, B., Cazalet, C., Hayes, G.L., Metzler, K.D., and Zychlinsky, A. (2012). Neutrophil Function: From Mechanisms to Disease. Annual Review of Immunology *30*, 459–489.

Anderson, K.L., Smith, K. a, Pio, F., Torbett, B.E., and Maki, R. a (1998). Neutrophils deficient in PU.1 do not terminally differentiate or become functionally competent. Blood *92*, 1576–1585.

Angermueller, C., Clark, S.J., Lee, H.J., Macaulay, I.C., Teng, M.J., Hu, T.X., Krueger, F., Smallwood, S.A., Ponting, C.P., Voet, T., et al. (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nature Methods *13*, 229–232.

Aranow, C. (2011). Vitamin D and the immune system. Journal of Investigative Medicine : The Official Publication of the American Federation for Clinical Research *59*, 881–886.

Auffray, C., Fogg, D., Garfa, M., Elain, G., Join-Lambert, O., Kayal, S., Sarnacki, S., Cumano, A., Lauvau, G., and Geissmann, F. (2007). Monitoring of Blood Vessels and\nTissues by a Population of Monocytes\nwith Patrolling Behavior\n. Science (New York, N.Y.) *317*, 666–670.

Auffray, C., Sieweke, M.H., and Geissmann, F. (2009). Blood monocytes: development, heterogeneity, and relationship with dendritic cells. Annual Review of Immunology *27*, 669–692.

Aziz, A., Soucie, E., Sarrazin, S., and Sieweke, M.H. (2009). MafB/c-Maf deficiency enables self-renewal of differentiated functional macrophages. Science (New York, N.Y.) *326*, 867–871.

Baeke, F., Takiishi, T., Korf, H., Gysemans, C., and Mathieu, C. (2010). Vitamin D: Modulator of the immune system. Current Opinion in Pharmacology *10*, 482–496.

Bazzoni, F., Cassatella, M.A., Laudanna, C., and Rossi, F. (1991). Phagocytosis of opsonized yeast induces tumor necrosis factor-alpha mRNA accumulation and protein release by human polymorphonuclear leukocytes. Journal of Leukocyte Biology *50*, 223–228.

Behring, E. von, and Kitasato, B.S. (1890). Ueber das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren. Deutsche Medizinische Wochenschrift *49*, 1–7.

Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
Borras, F.E., Lloberas, J., Maki, R.A., and Celada, A. (1995). Repression of I-A beta gene expression by the transcription factor PU.1. The Journal of Biological Chemistry *270*, 24385–24391.
Bottardi, S., Mavoungou, L., Bourgoin, V., Mashtalir, N., Affar, E.B., and Milot, E. (2013). Direct Protein Interactions Are Responsible for Ikaros-GATA and Ikaros-Cdk9 Cooperativeness in Hematopoietic Cells. Molecular and Cellular Biology *33*, 3064–3076.
Bottardi, S., Mavoungou, L., Pak, H., Daou, S., Bourgoin, V., Lakehal, Y.A., Affar, E.B., and Milot, E. (2014). The IKAROS Interaction with a Complex Including Chromatin Remodeling and Transcription Elongation Activities Is Required for Hematopoiesis. PLoS Genetics *10*.
Bottardi, S., Mavoungou, L., and Milot, E. (2015). IKAROS: A multifunctional regulator of the polymerase II transcription cycle. Trends in Genetics *31*, 500–508.
Bottinger, E.P., Shelley, C.S., Farokhzad, O.C., and Arnaout, M.A. (1994). The human beta 2 integrin CD18 promoter consists of two inverted Ets cis elements. Mol Cell Biol *14*, 2604–2615.
Boyer, L. a, Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. Cell *122*, 947–956.
Breitman, T.R., Selonick, S.E., and Collins, S.J. (1980). Induction of differentiation of the human promyelocytic leukemia cell line (HL-60) by retinoic acid. Proceedings of the National Academy of Sciences of the United States of America *77*, 2936–2940.
Bresnick, E.H., Lee, H.Y., Fujiwara, T., Johnson, K.D., and Keles, S. (2010). GATA switches as developmental drivers. Journal of Biological Chemistry *285*, 31087–31093.
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. Nature Methods *10*, 1213–1218.
Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature.
Cai, D.H., Wang, D., Keefer, J., Yeamans, C., Hensley, K., and Friedman, a D. (2008). C/EBP alpha:AP-1 leucine zipper heterodimers bind novel DNA elements, activate the PU.1 promoter

and direct monocyte lineage commitment more potently than C/EBP alpha homodimers or AP-1. Oncogene *27*, 2772–2779.

Cambier, J.C., Gauld, S.B., Merrell, K.T., and Vilen, B.J. (2007). B-cell anergy: from transgenic models to naturally occurring anergic B cells? Nature Reviews. Immunology *7*, 633–643.

Chang, H.-C., Sehra, S., Goswami, R., Yao, W., Yu, Q., Stritesky, G.L., Jabeen, R., McKinley, C., Ahyi, A.-N., Han, L., et al. (2010). The transcription factor PU.1 is required for the development of IL-9-producing T cells and allergic inflammation. Nature Immunology *11*, 527–534.

Chen, H., Ray-Gallet, D., Zhang, P., Hetherington, C.J., Gonzalez, D.A., Zhang, D.E., Moreau-Gachelin, F., and Tenen, D.G. (1995). PU.1 (Spi-1) autoregulates its expression in myeloid cells. Oncogene *11*, 1549–1560.

Chen, H.M., Pahl, H.L., Scheibe, R.J., Zhang, D.E., and Tenen, D.G. (1993). The Sp1 transcription factor binds the CD11b promoter specifically in myeloid cells in vivo and is essential for myeloid-specific promoter activity. Journal of Biological Chemistry *268*, 8230–8239.

Chen, L., Kostadima, M., Martens, J.H. a, Canu, G., Garcia, S.P., Turro, E., Downes, K., Macaulay, I.C., Bielczyk-Maczynska, E., Coe, S., et al. (2014). Transcriptional diversity during lineage commitment of human blood progenitors . Science *Accepted Pending Minor Review *345*.

Chow, A., Brown, B.D., and Merad, M. (2011). Studying the mononuclear phagocyte system in the molecular age. Nature Reviews. Immunology *11*, 788–798.

Cicco, N.A., Lindemann, A., Content, J., Vandenbussche, P., Lübbert, M., Gauss, J., Mertelsmann, R., and Herrmann, F. (1990). Inducible production of interleukin-6 by human polymorphonuclear neutrophils: role of granulocyte-macrophage colony-stimulating factor and tumor necrosis factor-alpha. Blood *75*, 2049–2052.

Clark, M.R., Mandal, M., Ochiai, K., and Singh, H. (2014). Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling. Nature Reviews. Immunology *14*, 69–80.

Collins, S.J. (1987). The HL-60 promyelocytic leukemia cell line: proliferation, differentiation, and cellular oncogene expression. Blood *70*, 1233–1244.

Collins, S.J., Gallo, R.C., and Gallagher, R.E. (1977). Continuous growth and differentiation of

human myeloid leukaemic cells in suspension culture. Nature *270*, 347–349.
Coons, A.H. (1955). Studies on Antibody Production: I. a Method for the Histochemical Demonstration of Specific Antibody and Its Application To a Study of the Hyperimmune Rabbit. Journal of Experimental Medicine *102*, 49–60.
Cooper, M.D. (2015). The early history of B cells. Nature Reviews Immunology *15*, 191–197.
Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *348*.
Dahl, R., Walsh, J.C., Lancki, D., Laslo, P., Iyer, S.R., Singh, H., and Simon, M.C. (2003). Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBPalpha ratio and granulocyte colony-stimulating factor. Nature Immunology *4*, 1029–1036.
Dahl, R., Iyer, S.R., Owens, K.S., Cuylear, D.D., and Simon, M.C. (2007). The transcriptional repressor GFI-1 antagonizes PU.1 activity through protein-protein interaction. The Journal of Biological Chemistry *282*, 6473–6483.
Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.
Decker, E.L., Nehmann, N., Kampen, E., Eibel, H., Zipfel, P.F., and Skerka, C. (2003). Early growth responce proteins (EGR) and nuclear factors of activated T cells (NFAT) form heterodimers and regulate proinflammatory cytokine gene expression. Nucleic Acids Research *31*, 911–921.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
Dovat, S. (2011). Ikaros in hematopoiesis and leukemia. World Journal of Biological Chemistry *2*, 105–107.
Duan, Z., Person, R.E., Lee, H.-H., Huang, S., Donadieu, J., Badolato, R., Grimes, H.L., Papayannopoulou, T., and Horwitz, M.S. (2007). Epigenetic regulation of protein-coding and microRNA genes by the Gfi1-interacting tumor suppressor PRDM5. Molecular and Cellular Biology *27*, 6889–6902.
Ebralidze, A.K., Guibal, F.C., Steidl, U., Zhang, P., Lee, S., Bartholdy, B., Jorda, M.A., Petkova, V., Rosenbauer, F., Huang, G., et al. (2008). PU.1 expression is modulated by the balance of

functional sense and antisense RNAs regulated by a shared cis-regulatory element. Genes and Development *22*, 2085–2092.
Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences of the United States of America *95*, 14863–14868.
Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science (New York, N.Y.) *297*, 1183–1186.
Fagraeus, a (1947). Plasma cellular reaction and its relation to the formation of antibodies in vitro. Nature *159*, 499.
Farlik, M., Sheffield, N.C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J., and Bock, C. (2015). Single-Cell DNA Methylome Sequencing and Bioinformatic Inference of Epigenomic Cell-State Dynamics. Cell Reports 1386–1397.
Ferreir´ os-Vidal, I., Carroll, T., Taylor, B., Terry, A., Liang, Z., Bruno, L., Dharmalingam, G., Khadayate, S., Cobb, B.S., Smale, S.T., et al. (2013). Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B-cell differentiation. Blood *121*, 1769–1782.
De Filippo, K., Dudeck, A., Hasenberg, M., Nye, E., van Rooijen, N., Hartmann, K., Gunzer, M., Roers, A., and Hogg, N. (2013). Mast cell and macrophage chemokines CXCL1/CXCL2 control the early stage of neutrophil recruitment during tissue inflammation. Blood *121*, 4930–4937.
Friend, C., Scher, W., Holland, J.G., and Sato, T. (1971). Hemoglobin synthesis in murine virus-induced leukemic cells in vitro: stimulation of erythroid differentiation by dimethyl sulfoxide. Proceedings of the National Academy of Sciences of the United States of America *68*, 378–382.
Fujishima, S., Hoffman, A.R., Vu, T., Kim, K.J., Zheng, H., Daniel, D., Kim, Y., Wallace, E.F., Larrick, J.W., and Raffin, T.A. (1993). Regulation of neutrophil interleukin 8 gene expression and protein secretion by LPS, TNF-alpha, and IL-1 beta. Journal of Cellular Physiology *154*, 478–485.
Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58–64.
van Furth, R., and Cohn, Z.A. (1968). The origin and kinetics of mononuclear phagocytes. The Journal of Experimental Medicine *128*, 415–435.

Gagliani, N., Vesely, M.C.A., Iseppon, A., Brockmann, L., Xu, H., Palm, N.W., de Zoete, M.R., Licona-Limón, P., Paiva, R.S., Ching, T., et al. (2015). Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation. Nature *523*, 221–225.

Gallagher, R., Collins, S., Trujillo, J., McCredie, K., Ahearn, M., Tsai, S., Metzgar, R., Aulakh, G., Ting, R., Ruscetti, F., et al. (1979a). Characterization of the continuous, differentiating myeloid cell line (HL-60) from a patient with acute promyelocytic leukemia. Blood *54*, 713–733.

Gallagher, R., Collins, S., Trujillo, J., McCredie, K., Ahearn, M., Tsai, S., Metzgar, R., Aulakh, G., Ting, R., Ruscetti, F., et al. (1979b). Characterization of the continuous, differentiating myeloid cell line (HL-60) from a patient with acute promyelocytic leukemia.

Ganong, W.F. (2003). Review of medical physiology 21st Edition.

Gee, K., Lim, W., Ma, W., Nandan, D., Diaz-Mitoma, F., Kozlowski, M., and Kumar, A. (2002). Differential regulation of CD44 expression by lipopolysaccharide (LPS) and TNF-alpha in human monocytic cells: distinct involvement of c-Jun N-terminal kinase in LPS-induced CD44 expression. Journal of Immunology (Baltimore, Md. : 1950) *169*, 5660–5672.

Georgescu, C., Longabaugh, W.J.R., Scripture-Adams, D.D., David-Fung, E.-S., Yui, M. a, Zarnegar, M. a, Bolouri, H., and Rothenberg, E. V (2008). A gene regulatory network armature for T lymphocyte specification. Proceedings of the National Academy of Sciences of the United States of America *105*, 20100–20105.

Georgopoulos, K. (2002). Haematopoietic cell-fate decisions, chromatin regulation and ikaros. Nature Reviews. Immunology *2*, 162–174.

Georgopoulos, K., Moore, D.D., and Derfler, B. (1992). Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. Science (New York, N.Y.) *258*, 808–812.

Georgopoulos, K., Bigby, M., Wang, J.H., Molnar, A., Wu, P., Winandy, S., and Sharpe, A. (1994). The ikaros gene is required for the development of all lymphoid lineages. Cell *79*, 143–156.

Ginhoux, F., and Jung, S. (2014). Monocytes and macrophages: developmental pathways and tissue homeostasis. Nature Reviews. Immunology *14*, 392–404.

Glass, C.K., and Natoli, G. (2015). Molecular control of activation and priming in macrophages. Nature Immunology *17*, 26–33.

Gosselin, D., Link, V.M., Romanoski, C.E., Fonseca, G.J., Eichenfield, D.Z., Spann, N.J.,

Stender, J.D., Chun, H.B., Garner, H., Geissmann, F., et al. (2014). Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. Cell *159*, 1327–1340.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics (Oxford, England) *27*, 1017–1018.

Grimes, H.L., Chan, T.O., Zweidler-McKay, P. a, Tong, B., and Tsichlis, P.N. (1996a). The Gfi-1 proto-oncoprotein contains a novel transcriptional repressor domain, SNAG, and inhibits G1 arrest induced by interleukin-2 withdrawal. Molecular and Cellular Biology *16*, 6263–6272.

Grimes, H.L., Gilks, C.B., Chan, T.O., Porter, S., and Tsichlis, P.N. (1996b). The Gfi-1 protooncoprotein represses Bax expression and inhibits T-cell death. Proceedings of the National Academy of Sciences of the United States of America *93*, 14569–14573.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biology *8*, R24.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295–300.

Haghverdi, L., Buettner, F., and Theis, F.J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics *31*, 2989–2998.

Hardison, R.C., and Taylor, J. (2012). Genomic approaches towards finding cis-regulatory modules in animals. Nature Reviews Genetics *13*, 469–483.

Harris, P., and Ralph, P. (1985). Human leukemic models of myelomonocytic development: a review of the HL-60 and U937 cell lines. Journal of Leukocyte Biology *37*, 407–422.

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. Molecular Cell *38*, 576–589.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., et al. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nature Methods *6*, 283–289.

Hoogenkamp, M., Krysinska, H., Ingram, R., Huang, G., Barlow, R., Clarke, D., Ebralidze, A., Zhang, P., Tagoh, H., Cockerill, P.N., et al. (2007). The Pu.1 locus is differentially regulated at

the level of chromatin structure and noncoding transcription by alternate mechanisms at distinct developmental stages of hematopoiesis. Molecular and Cellular Biology *27*, 7425–7438.
Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., Wu, X., Wen, L., Tang, F., Huang, Y., et al. (2016). Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Research *26*, 304–319.
Howard, M.L., and Davidson, E.H. (2004). cis-Regulatory control circuits in development. Developmental Biology *271*, 109–118.
Hu, Y.L., Fong, S., Largman, C., and Shen, W.F. (2010). HOXA9 regulates miR-155 in hematopoietic cells. Nucleic Acids Research *38*, 5472–5478.
Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2007). The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biology *8*, R183.
Huang, G., Zhang, P., Hirai, H., Elf, S., Yan, X., Chen, Z., Koschmieder, S., Okuno, Y., Dayaram, T., Growney, J.D., et al. (2008). PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. Nature Genetics *40*, 51–60.
Hume, D.A. (2006). The mononuclear phagocyte system. Current Opinion in Immunology *18*, 49–53.
Hume, D.A., Ross, I.L., Himes, S.R., Sasmono, R.T., Wells, C.A., and Ravasi, T. (2002). The mononuclear phagocyte system revisited. Journal of Leukocyte Biology *72*, 621–627.
Hyman A. & Simons K (2011). Beyond HeLa cells. Nature 2011.
Imoto, A., Okada, M., Okazaki, T., Kitasato, H., Harigae, H., and Takahashi, S. (2010). Metallothionein-1 isoforms and vimentin are direct PU.1 downstream target genes in leukemia cells. Journal of Biological Chemistry *285*, 10300–10309.
Iwama, A., Zhang, P., Darlington, G.J., McKercher, S.R., Maki, R., and Tenen, D.G. (1998). Use of RDA analysis of knockout mice to identify myeloid genes regulated in vivo by PU.1 and C/EBPα. Nucleic Acids Research *26*, 3034–3043.
Jehan, F., and DeLuca, H.F. (1997). Cloning and characterization of the mouse vitamin D receptor promoter. Proceedings of the National Academy of Sciences *94*, 10138–10143.
Jin, W., Tang, Q., Wan, M., Cui, K., Zhang, Y., Ren, G., Ni, B., Sklar, J., Przytycka, T.M., Childs, R., et al. (2015). Genome-wide detection of DNase I hypersensitive sites in single cells

and FFPE tissue samples. Nature *528*, 142–146.
John, L.B., and Ward, A.C. (2011). The Ikaros gene family: Transcriptional regulators of hematopoiesis and immunity. Molecular Immunology *48*, 1272–1278.
John, S., Sabo, P.J., Thurman, R.E., Sung, M.-H., Biddie, S.C., Johnson, T. a, Hager, G.L., and Stamatoyannopoulos, J. a (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nature Genetics *43*, 264–268.
Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007a). Genome-Wide Mapping of in Vivo. Science *316*, 1497–1502.
Johnson, K., Hashimshony, T., Sawai, C.M., Pongubala, J.M.R., Skok, J.A., Aifantis, I., and Singh, H. (2008). Regulation of Immunoglobulin Light-Chain Recombination by the Transcription Factor IRF-4 and??the Attenuation of Interleukin-7 Signaling. Immunity *28*, 335–345.
Johnson, W.E., Li, C., and Rabinovic, A. (2007b). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics (Oxford, England) *8*, 118–127.
Joseph, L.J., Le Beau, M.M., Jamieson, G. a, Acharya, S., Shows, T.B., Rowley, J.D., and Sukhatme, V.P. (1988). Molecular cloning, sequencing, and mapping of EGR2, a human early growth response gene encoding a protein with “zinc-binding finger” structure. Proceedings of the National Academy of Sciences of the United States of America *85*, 7164–7168.
Joshi, I., Yoshida, T., Jena, N., Qi, X., Zhang, J., Van Etten, R. a, and Georgopoulos, K. (2014). Loss of Ikaros DNA-binding function confers integrin-dependent survival on pre-B cells and progression to acute lymphoblastic leukemia. Nature Immunology *15*, 294–304.
Kaufmann, S.H.E., and Dorhoi, A. (2016). Molecular Determinants in Phagocyte-Bacteria Interactions. Immunity *44*, 476–491.
Kelly, L.M., Englmeier, U., Lafon, I., Sieweke, M.H., and Graf, T. (2000). MafB is an inducer of monocytic differentiation. The EMBO Journal *19*, 1987–1997.
Keys, J.R., Tallack, M.R., Zhan, Y., Papathanasiou, P., Goodnow, C.C., Gaensler, K.M., Crossley, M., Dekker, J., and Perkins, A.C. (2008). A mechanism for Ikaros regulation of human globin gene switching. In British Journal of Haematology, pp. 398–406.
Khanna-Gupta, A., Sun, H., Zibello, T., Han, M.L., Dahl, R., Boxer, L.A., and Berliner, N. (2007). Growth factor independence-1 (Gfi-1) plays a role in mediating specific granule deficiency (SGD) in a patient lacking a gene-inactivating mutation in the C/EBP?? gene. Blood

109, 4181–4190.

Kihara-Negishi, F., Suzuki, M., Yamada, T., Sakurai, T., and Oikawa, T. (2005). Impaired repressor activity and biological functions of PU.1 in MEL cells induced by mutations in the acetylation motifs within the ETS domain. Biochemical and Biophysical Research Communications *335*, 477–484.

Kim, D.H., Marinov, G.K., Pepke, S., Singer, Z.S., He, P., Williams, B., Schroth, G.P., Elowitz, M.B., and Wold, B.J. (2015). Single-Cell Transcriptome Analysis Reveals Dynamic Changes in lncRNA Expression during Reprogramming. Cell Stem Cell *16*, 88–101.

Kiss, M., Czimmerer, Z., and Nagy, L. (2013). The role of lipid-activated nuclear receptors in shaping macrophage and dendritic cell function: From physiology to pathology. Journal of Allergy and Clinical Immunology *132*, 264–286.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187–1201.

KLEMSZ, L.A. and M.J. (1999). DNA Methylation and Chromatin Structure Regulate PU.1 Expression. DNA AND CELL BIOLOGY *18*, 875–884.

Klemsz, M.J., McKercher, S.R., Celada, A., Van Beveren, C., and Maki, R.A. (1990). The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. Cell *61*, 113–124.

Ko, L.J., and Engel, J.D. (1993). DNA-binding specificities of the GATA transcription factor family. Molecular and Cellular Biology *13*, 4011–4022.

Koeffler, H.P., and Golde, D.W. (1978). Acute myelogenous leukemia: a human cell line responsive to colony-stimulating activity. Science (New York, N.Y.) *200*, 1153–1154.

Koh, C.P., Wang, C.Q., Ng, C.E.L., Ito, Y., Araki, M., Tergaonkar, V., Huang, G., and Osato, M. (2013). RUNX1 meets MLL: epigenetic regulation of hematopoiesis by two leukemia genes. Leukemia *27*, 1793–1802.

Kolaczkowska, E., and Kubes, P. (2013). Neutrophil recruitment and function in health and inflammation. Nature Reviews. Immunology *13*, 159–175.

Krishnaraju, K., Hoffman, B., and Liebermann, D. a (1998). The zinc finger transcription factor Egr-1 activates macrophage differentiation in M1 myeloblastic leukemia cells. Blood *92*, 1957–1966.

Kubosaki, A., Tomaru, Y., Tagami, M., Arner, E., Miura, H., Suzuki, T., Suzuki, M., Suzuki, H., and Hayashizaki, Y. (2009). Genome-wide investigation of in vivo EGR-1 binding sites in monocytic differentiation. Genome Biology *10*, R41.

Kummalue, T., and Friedman, A.D. (2003). Cross-talk between regulators of myeloid development: C/EBPalpha binds and activates the promoter of the PU.1 gene. Journal of Leukocyte Biology *74*, 464–470.

Lacramioara Bintu, M.B.E. (2016). Dynamics of epigenetic regulation at the single-cell level. Science.

Ladewig, J., Koch, P., and Brüstle, O. (2013). Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. Nature Reviews. Molecular Cell Biology *14*, 225–236.

Lämmermann, T., Afonso, P. V, Angermann, B.R., Wang, J.M., Kastenmüller, W., Parent, C. a, and Germain, R.N. (2013). Neutrophil swarms require LTB4 and integrins at sites of cell death in vivo. Nature *498*, 371–375.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Langmead, B., Langmead, B., Trapnell, C., Trapnell, C., Pop, M., Pop, M., Salzberg, S.L., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

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 (New York, N.Y.) 1–10.

Laslo, P., Spooner, C.J., Warmflash, A., Lancki, D.W., Lee, H.J., Sciammas, R., Gantner, B.N., Dinner, A.R., and Singh, H. (2006). Multilineage Transcriptional Priming and Determination of Alternate Hematopoietic Cell Fates. Cell *126*, 755–766.

Laurent, B., Randrianarison-Huetz, V., Kadri, Z., Rom??o, P.H., Porteu, F., and Dum??nil, D. (2009). Gfi-1B promoter remains associated with active chromatin marks throughout erythroid differentiation of human primary progenitor cells. Stem Cells *27*, 2153–2162.

Lavin, Y., Winter, D., Blecher-gonen, R., David, E., Keren-shaul, H., and Merad, M. (2014). Tissue-Resident Macrophage Enhancer Landscapes Are Shaped by the Local Microenvironment. Cell *159*, 1312–1326.

Lavin, Y., Mortha, A., Rahman, A., and Merad, M. (2015). Regulation of macrophage development and function in peripheral tissues. Nature Reviews. Immunology *15*, 731–744.
Lawrence, T., and Natoli, G. (2011). Transcriptional regulation of macrophage polarization: enabling diversity with identity. Nature Reviews Immunology *11*, 750–761.
Leddin, M., Perrod, C., Hoogenkamp, M., Ghani, S., Assi, S., Heinz, S., Wilson, N.K., Follows, G., Schönheit, J., Vockentanz, L., et al. (2011). Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells. Blood *117*, 2827–2838.
Lee, Y.K., Turner, H., Maynard, C.L., Oliver, J.R., Chen, D., Elson, C.O., and Weaver, C.T. (2009). Late Developmental Plasticity in the T Helper 17 Lineage. Immunity *30*, 92–107.
Li, Y. (2001). Regulation of the PU.1 gene by distal elements. Blood *98*, 2958–2965.
Liu, J.W., Lacy, J., Sukhatme, V.P., and Coleman, D.L. (1991). Granulocyte-macrophage colony-stimulating factor induces transcriptional activation of Egr-1 in murine peritoneal macrophages. The Journal of Biological Chemistry *266*, 5929–5933.
Liu, Y., Elf, S.E., Miyata, Y., Sashida, G., Liu, Y., Huang, G., Di Giandomenico, S., Lee, J.M., Deblasio, A., Menendez, S., et al. (2009). p53 Regulates Hematopoietic Stem Cell Quiescence. Cell Stem Cell *4*, 37–48.
Longabaugh, W.J.R., Davidson, E.H., and Bolouri, H. (2005). Computational representation of developmental genetic regulatory networks. Developmental Biology *283*, 1–16.
Longabaugh, W.J.R., Davidson, E.H., and Bolouri, H. (2009). Visualization, documentation, analysis, and communication of large-scale gene regulatory networks. Biochimica et Biophysica Acta *1789*, 363–374.
Lozzio, C.B., and Lozzio, B.B. (1975). Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. Blood *45*, 321–334.
Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.I., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., et al. (2015). G{&}T-seq : parallel sequencing of single- cell genomes and transcriptomes. Nature Methods *12*, 1–7.
Mackaness, G.B. (1962). Cellular resistance to infection. The Journal of Experimental Medicine *116*, 381–406.
Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA-guided activation of endogenous human genes. Nat Methods *10*, 977–979.

Mangelsdorf, D.J., Koeffler, H.P., Donaldson, C.A., Pike, J.W., and Haussler, M.R. (1984). 1,25-Dihydroxyvitamin D3-induced differentiation in a human promyelocytic leukemia cell line (HL-60): Receptor-mediated maturation to macrophage-like cells. Journal of Cell Biology *98*, 391–398.

Marteijn, J.A.F., Van Der Meer, L.T., Van Emst, L., De Witte, T., Jansen, J.H., and Van Der Reijden, B.A. (2007). Diminished proteasomal degradation results in accumulation of Gfi1 protein in monocytes. Blood *109*, 100–108.

Masters, J.R. (2002). HeLa cells 50 years on: the good, the bad and the ugly. Nature Reviews Cancer *2*, 315–319.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C., Chou, A., Ienasescu, H., et al. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Research *42*, D142–D147.

Maurano, M.T., and Stamatoyannopoulos, J.A. (2015). Previews Taking Stock of Regulatory Variation. Cell Systems *1*, 18–21.

McGhee, L., Bryan, J., Elliott, L., Grimes, H.L., Kazanjian, A., Davis, J.N., and Meyers, S. (2003). Gfi-1 attaches to the nuclear matrix, associates with ETO (MTG8) and histone deacetylase proteins, and represses transcription using a TSA-sensitive mechanism. Journal of Cellular Biochemistry *89*, 1005–1018.

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. Nature Biotechnology *28*, 495–501.

Van der Meer, L.T., Jansen, J.H., and van der Reijden, B. a (2010). Gfi1 and Gfi1b: key regulators of hematopoiesis. Leukemia *24*, 1834–1843.

Melchers, F. (2005). The pre-B-cell receptor: selector of fitting immunoglobulin heavy chains for the B-cell repertoire. Nature Reviews. Immunology *5*, 578–584.

Merkenschlager, M. (2010). Ikaros in immune receptor signaling, lymphocyte differentiation, and function. FEBS Letters *584*, 4910–4914.

Metschnikoff, E. (1884). Eine neue Entzu¨ ndungstheorie (Entzu¨ ndung und intracellula¨re

Verdauung). Allgemeine Wiener medizinische. Zeitung 27/29, 307–332.
Mo, A., Mukamel, E.A., Davis, F.P., Luo, C., Henry, G.L., Picard, S., Urich, M.A., Nery, J.R., Sejnowski, T.J., Lister, R., et al. (2015). Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. Neuron *86*, 1369–1384.
Montoya-Durango, D.E., Velu, C.S., Kazanjian, A., Rojas, M.E.B., Jay, C.M., Longmore, G.D., and Grimes, H.L. (2008). Ajuba functions as a histone deacetylase-dependent co-repressor for autoregulation of the growth factor-independent-1 transcription factor. Journal of Biological Chemistry *283*, 32056–32065.
Mora, J.R., Iwata, M., and von Andrian, U.H. (2008). Vitamin effects on the immune system: vitamins A and D take centre stage. Nature Reviews. Immunology *8*, 685–698.
Möröy, T., and Khandanpour, C. (2011). Growth factor independence 1 (Gfi1) as a regulator of lymphocyte development and activation. Seminars in Immunology *23*, 368–378.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods *5*, 621–628.
Mortazavi, A., Pepke, S., Jansen, C., Marinov, G.K., Ernst, J., Kellis, M., Hardison, R.C., Myers, R.M., and Wold, B.J. (2013). Integrating and mining the chromatin landscape of cell-type specificity using self-organizing maps. Genome Research *23*, 2136–2148.
Murao, S., Gemmell, M.A., Callaham, M.F., Anderson, N.L., and Huberman, E. (1983). Control of macrophage cell differentiation in human promyelocytic HL-60 leukemia cells by 1,25-dihydroxyvitamin D3 and phorbol-12-myristate-13-acetate. Cancer Research *43*, 4989–4996.
Murray, P.J., Allen, J.E., Biswas, S.K., Fisher, E.A., Gilroy, D.W., Goerdt, S., Gordon, S., Hamilton, J.A., Ivashkiv, L.B., Lawrence, T., et al. (2014). Macrophage Activation and Polarization: Nomenclature and Experimental Guidelines. Immunity *41*, 14–20.
Nathan, C.F., Murray, H.W., Wiebe, M.E., and Rubin, B.Y. (1983). Identification of interferon-gamma as the lymphokine that activates human macrophage oxidative metabolism and antimicrobial activity. The Journal of Experimental Medicine *158*, 670–689.
Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012a). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature *489*, 83–90.
Neph, S., Stergachis, A.B., Reynolds, A., Sandstrom, R., Borenstein, E., and Stamatoyannopoulos, J.A. (2012b). Circuitry and dynamics of human transcription factor

regulatory networks. Cell *150*, 1274–1286.
Nerlov, C., Querfurth, E., Kulessa, H., and Graf, T. (2000). GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription. Blood *95*, 2543–2551.
Nicol-Benoit, F., Le Goff, P., and Michel, D. (2013). Drawing a Waddington landscape to capture dynamic epigenetics. Biology of the Cell *105*, 576–584.
Noble, D. (2015). Conrad Waddington and the origin of epigenetics. The Journal of Experimental Biology *218*, 816–818.
Nuchprayoon, I., Shang, J., Simkevich, C.P., Luo, M., Rosmarin, A.G., and Friedman, A.D. (1999). An enhancer located between the neutrophil elastase and proteinase 3 promoters is activated by Sp1 and an Ets factor. Journal of Biological Chemistry *274*, 1085–1091.
Nueda, M.J., Tarazona, S., and Conesa, A. (2014). Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. Bioinformatics (Oxford, England) 1–5.
Ochiai, K., Maienschein-Cline, M., Mandal, M., Triggs, J.R., Bertolino, E., Sciammas, R., Dinner, A.R., Clark, M.R., and Singh, H. (2012). A self-reinforcing regulatory network triggered by limiting IL-7 activates pre-BCR signaling and differentiation. Nature Immunology *13*, 300–307.
Oelgeschläger, M., Nuchprayoon, I., Lüscher, B., and Friedman, a D. (1996). C/EBP, c-Myb, and PU.1 cooperate to regulate the neutrophil elastase promoter. Molecular and Cellular Biology *16*, 4717–4725.
Oettinger, M.A., Schatz, D.G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science (New York, N.Y.) *248*, 1517–1523.
Van Oevelen, C., Collombet, S., Vicent, G., Hoogenkamp, M., Lepoivre, C., Badeaux, A., Bussmann, L., Sardina, J.L., Thieffry, D., Beato, M., et al. (2015). C/EBP?? Activates Pre-existing and de Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. Stem Cell Reports *5*, 232–247.
Oh, J., Riek, A.E., Darwech, I., Funai, K., Shao, J., Chin, K., Sierra, O.L., Carmeliet, G., Ostlund, R.E., and Bernal-Mizrachi, C. (2015). Deletion of Macrophage Vitamin D Receptor Promotes Insulin Resistance and Monocyte Cholesterol Transport to Accelerate Atherosclerosis in Mice. Cell Reports *10*, 1872–1886.
Ohno, Y., Lee, J., Fusunyan, R.D., MacDermott, R.P., and Sanderson, I.R. (1997). Macrophage

inflammatory protein-2: chromosomal regulation in rat small intestinal epithelial cells. Proc Natl Acad Sci U S A *94*, 10279–10284.

Olson, M.C., Scott, E.W., Hack, A.A., Su, G.H., Tenen, D.G., Singh, H., and Simon, M.C. (1995). PU. 1 is not essential for early myeloid gene expression but is required for terminal myeloid differentiation. Immunity *3*, 703–714.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. Cell *152*, 157–171.

Palma, C., Cassone, a., Serbousek, D., Pearson, C. a., and Djeu, J.Y. (1992). Lactoferrin release and interleukin-1, interleukin-6, and tumor necrosis factor production by human polymorphonuclear cells stimulated by various lipopolysaccharides: Relationship to growth inhibition of Candida albicans. Infection and Immunity *60*, 4604–4611.

Passlick, B., Flieger, D., and Ziegler-Heitbrock, H.W. (1989). Identification and characterization of a novel monocyte subpopulation in human peripheral blood. Blood *74*, 2527–2534.

Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell *163*, 1663–1677.

Pepke, S., Wold, B., and Mortazavi, A. (2009). Computation for ChIP-seq and RNA-seq studies. Nat Methods *6*, S22–S32.

Person, R.E., Li, F.-Q., Duan, Z., Benson, K.F., Wechsler, J., Papadaki, H. a, Eliopoulos, G., Kaufman, C., Bertolone, S.J., Nakamoto, B., et al. (2003). Mutations in proto-oncogene GFI1 cause human neutropenia and target ELA2. Nature Genetics *34*, 308–312.

Peters, I., and Davidson, E.H. (2015). Genomic Control Process.

Petterson, M., and Schaffner, W. (1987). A purine-rich DNA sequence motif present in SV40 and lymphotropic papovavirus binds a lymphoid-specific factor and contributes to enhancer activity in lymphoid cells. Genes & Development *1*, 962–972.

Piper, J., Elze, M.C., Cauchy, P., Cockerill, P.N., Bonifer, C., and Ott, S. (2013). Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Research *41*, e201.

Pique-Regi, R., Degner, J.F., Pai, A. a, Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). Accurate inference of transcription factor binding from DNA sequence and chromatin

accessibility data. Genome Research *21*, 447–455.

Pott, S., and Lieb, J.D. (2015). Single-cell ATAC-seq: strength in numbers. Genome Biology *16*, 172.

Qu, K., Zaba, L.C., Giresi, P.G., Li, R., Longmire, M., Kim, Y.H., Greenleaf, W.J., and Chang, H.Y. (2015). Individuality and Variation of Personal Regulomes in Primary Human T Cells. Cell Systems *1*, 51–61.

Quang, D., and Xie, X. (2014). EXTREME: An online em algorithm for motif discovery. Bioinformatics *30*, 1667–1673.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Research *42*, W187–W191.

Randolph, G.J., Inaba, K., Robbiani, D.F., Steinman, R.M., and Muller, W. a (1999). Differentiation of phagocytic monocytes into lymph node dendritic cells in vivo. Immunity *11*, 753–761.

Rathinam, C., Geffers, R., Y??cel, R., Buer, J., Welte, K., M??r??y, T., and Klein, C. (2005). The transcriptional repressor Gfi1 controls STAT3-dependent dendritic cell development and function. Immunity *22*, 717–728.

Rickert, R.C. (2013). New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. Nature Reviews. Immunology *13*, 578–591.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England) *26*, 139–140.

Ross, J., Mavoungou, L., Bresnick, E.H., and Milot, E. (2012). GATA-1 Utilizes Ikaros and Polycomb Repressive Complex 2 To Suppress Hes1 and To Promote Erythropoiesis. Molecular and Cellular Biology *32*, 3624–3638.

Rothenberg, E. V., Kueh, H.Y., Yui, M.A., and Zhang, J.A. (2016). Hematopoiesis and T-cell specification as a model developmental system. Immunological Reviews *271*, 72–97.

Sabbattini, P., Lundgren, M., Georgiou, A., Chow, C.M., Warnes, G., and Dillon, N. (2001). Binding of Ikaros to the ??5 promoter silences transcription through a mechanism that does not require heterochromatin formation. EMBO Journal *20*, 2812–2822.

Sackmann, E. (1995). Biological Membranes Architecture and Function., Handbook of Biological Physics.

Saeed, S., Quintin, J., Kerstens, H.H.D., Rao, N. a., Aghajanirefah, a., Matarese, F., Cheng, S.-C., Ratter, J., Berentsen, K., van der Ent, M. a., et al. (2014). Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. Science.

Safford, M., Collins, S., Lutz, M.A., Allen, A., Huang, C.T., Kowalski, J., Blackford, A., Horton, M.R., Drake, C., Schwartz, R.H., et al. (2005). Egr-2 and Egr-3 are negative regulators of T cell activation. Nat Immunol *6*, 472–480.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature Biotechnology *33*, 495–502.

Scapini, P., and Cassatella, M.A. (2014). Social networking of human neutrophils within the immune system. Blood *124*, 710–719.

Scapini, P., Laudanna, C., Pinardi, C., Allavena, P., Mantovani, a, Sozzani, S., and Cassatella, M. a (2001). Neutrophils produce biologically active macrophage inflammatory protein-3alpha (MIP-3alpha)/CCL20 and MIP-3beta/CCL19. European Journal of Immunology *31*, 1981–1988.

Schatz, D.G., Oettinger, M.A., and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. Cell *59*, 1035–1048.

Scherer, W.F., and Syverton, J.T. (1952). Studies on the propagation in vitro of poliomyelitis viruses. III. The propagation of poliomyelitis viruses in tissue cultures devoid of nerve cells. The Journal of Experimental Medicine *96*, 389–400.

Schwartsmann, G., Madan, K., Heins, Y., Pinedo, H.M., and Leyva, A. (1987). Double-minute chromatin bodies in HL-60 leukemia cells sensitive and resistant to differentiation inducing agents. Cell Biology International Reports *11*, 651–655.

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

Selman, K., and Kafatos, F.C. (1974). Transdifferentiation in the labial gland of silk moths: is DNA required for cellular metamorphosis? Cell Differentiation *3*, 81–94.

Sertznig, P., Dunlop, T., Seifert, M., Tilgen, W., and Reichrath, J. (2009). Cross-talk between vitamin D receptor (VDR)- and peroxisome proliferator-activated receptor (PPAR)-signaling in melanoma cells. Anticancer Research *29*, 3647–3658.

Setty, M., Tadmor, M.D., Reich-Zeliger, S., Angel, O., Salame, T.M., Kathail, P., Choi, K.,

Bendall, S., Friedman, N., and Pe'er, D. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. Nature Biotechnology 1–14.

Sharif-Askari, E., Vassen, L., Kosan, C., Khandanpour, C., Gaudreau, M.-C., Heyd, F., Okayama, T., Jin, J., Rojas, M.E.B., Grimes, H.L., et al. (2010). Zinc finger protein Gfi1 controls the endotoxin-mediated Toll-like receptor inflammatory response by antagonizing NF-kappaB p65. Molecular and Cellular Biology *30*, 3929–3942.

Sheffield, N.C., Thurman, R.E., Song, L., Safi, A., Stamatoyannopoulos, J.A., Lenhard, B., Crawford, G.E., and Furey, T.S. (2013). Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. Genome Research *23*, 777–788.

Sheikh, S.Z., Kobayashi, T., Matsuoka, K., Onyiah, J.C., and Plevy, S.E. (2011). Characterization of an interferon-stimulated response element (ISRE) in the Il23a promoter. The Journal of Biological Chemistry *286*, 1174–1180.

Shoemaker, R.H. (2006). The NCI60 human tumour cell line anticancer drug screen. Nature Reviews. Cancer *6*, 813–823.

Soldner, F., and Jaenisch, R. (2012a). iPSC Disease Modeling. Science *338*, 1155–1156.

Soldner, F., and Jaenisch, R. (2012b). Medicine. iPSC disease modeling. Science (New York, N.Y.) *338*, 1155–1156.

Spooner, C.J., Cheng, J.X., Pujadas, E., Laslo, P., and Singh, H. (2009). A Recurrent Network Involving the Transcription Factors PU.1 and Gfi1 Orchestrates Innate and Adaptive Immune Cell Fates. Immunity *31*, 576–586.

Sridharan, R., and Smale, S.T. (2007). Predominant interaction of both Ikaros and Helios with the NuRD complex in immature thymocytes. Journal of Biological Chemistry *282*, 30227–30238.

Stein, M., Keshav, S., Harris, N., and Gordon, S. (1992). Interleukin 4 potently enhances murine macrophage mannose receptor activity: a marker of alternative immunologic macrophage activation. The Journal of Experimental Medicine *176*, 287–292.

Stergachis, A.B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A.P., Zhang, M., Byron, R., Canfield, T., Stelhing-Sun, S., Lee, K., et al. (2014). Conservation of trans-acting circuitry during mammalian regulatory evolution. Nature *515*, 365–370.

Stuart, L.M., and Ezekowitz, R.A. (2008). Phagocytosis and comparative innate immunity:

learning on the fly. Nature Reviews. Immunology *8*, 131–141.
Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A., et al. (2002). Large-scale analysis of the human and mouse transcriptomes. Proceedings of the National Academy of Sciences of the United States of America *99*, 4465–4470.
Sullivan, A.M., Arsovski, A.A., Lempe, J., Bubb, K.L., Weirauch, M.T., Sabo, P.J., Sandstrom, R., Thurman, R.E., Neph, S., Reynolds, A.P., et al. (2014). Mapping and Dynamics of Regulatory DNA and Transcription Factor Networks in A. thaliana. Cell Reports *8*, 2015–2030.
Sumitomo, S., Fujio, K., Okamura, T., and Yamamoto, K. (2013). Egr2 and Egr3 are the unique regulators for systemic autoimmunity. Jak-Stat *2*, e23952.
Sung, M.-H., Baek, S., and Hager, G.L. (2016). Genome-wide footprinting: ready for prime time? Nat Meth *13*, 222–228.
Suzuki, M., Yamada, T., Kihara-Negishi, F., Sakurai, T., Hara, E., Tenen, D.G., Hozumi, N., and Oikawa, T. (2006). Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. Oncogene *25*, 2477–2488.
Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676.
Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell *131*, 861–872.
Thompson, E.C., Cobb, B.S., Sabbattini, P., Meixlsperger, S., Parelho, V., Liberg, D., Taylor, B., Dillon, N., Georgopoulos, K., Jumaa, H., et al. (2007). Ikaros DNA-Binding Proteins as Integral Components of B Cell Developmental-Stage-Specific Regulatory Circuits. Immunity *26*, 335–344.
Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.
TILL, J.E., and McCullouch, E.A. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiation Research *14*, 213–222.
Tong, A.-J., Liu, X., Thomas, B.J., Lissner, M.M., Baker, M.R., Senagolage, M.D., Allred, A.L., Barish, G.D., and Smale, S.T. (2016). A Stringent Systems Approach Uncovers Gene-Specific

Mechanisms Regulating Inflammation. Cell *165*, 165–179.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B. a, Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology *28*, 511–515.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature Biotechnology *32*, 381–386.

Varol, C., Mildner, A., and Jung, S. (2015). Macrophages: development and tissue specialization.

Vierstra, J., and Stamatoyannopoulos, J.A. (2016). Genomic footprinting. *13*.

Vigorito, E., Perks, K.L., Abreu-Goodger, C., Bunting, S., Xiang, Z., Kohlhaas, S., Das, P.P., Miska, E.A., Rodriguez, A., Bradley, A., et al. (2007). microRNA-155 Regulates the Generation of Immunoglobulin Class-Switched Plasma Cells. Immunity *27*, 847–859.

Waddington, C.H. (1957). The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. The Strategy of the Genes A Discussion of Some … ix + – 262 .

Williams, M.J. (2007). Drosophila hemopoiesis and cellular immunity. Journal of Immunology (Baltimore, Md. : 1950) *178*, 4711–4716.

Wöbke, T.K., Sorg, B.L., and Steinhilber, D. (2014). Vitamin D in inflammatory diseases. Frontiers in Physiology *5 JUL*.

Wong, E.T., Jenne, D.E., Zimmer, M., Porter, S.D., and Gilks, C.B. (1999). Changes in chromatin organization at the neutrophil elastase locus associated with myeloid cell differentiation. Blood *94*, 3730–3736.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nature Reviews. Genetics *8*, 206–216.

Xia, J., Bolyard, A.A., Rodger, E., Stein, S., Aprikyan, A.A., Dale, D.C., and Link, D.C. (2009). Prevalence of mutations in ELANE, GFI1, HAX1, SBDS, WAS and G6PC3 in patients with severe congenital neutropenia. British Journal of Haematology *147*, 535–542.

Xue, J., Schmidt, S. V., Sander, J., Draffehn, A., Krebs, W., Quester, I., DeNardo, D., Gohel,

T.D., Emde, M., Schmidleithner, L., et al. (2014). Transcriptome-Based Network Analysis Reveals a Spectrum Model of Human Macrophage Activation. Immunity *40*, 274–288.

Yang, J., Zhang, L., Yu, C., Yang, X.-F., and Wang, H. (2014). Monocyte and macrophage differentiation: circulation inflammatory monocyte as biomarker for inflammatory diseases. Biomarker Research *2*, 1.

Yeamans, C., Wang, D., Paz-Priel, I., Torbett, B.E., Tenen, D.G., and Friedman, A.D. (2007). C/EBPα binds and activates the PU.1 distal enhancer to induce monocyte lineage commitment. Blood *110*, 3136–3142.

Yosef, N., Shalek, A.K., Gaublomme, J.T., Jin, H., Lee, Y., Awasthi, A., Wu, C., Karwacz, K., Xiao, S., Jorgolli, M., et al. (2013). Dynamic regulatory network controlling TH17 cell differentiation. Nature *496*, 461–468.

Yoshida, T., Ng, S.Y.M., and Georgopoulos, K. (2010). Awakening lineage potential by Ikaros-mediated transcriptional priming. Current Opinion in Immunology *22*, 154–160.

Yücel, R., Kosan, C., Heyd, F., and Möröy, T. (2004). Gfi1: Green fluorescent protein knock-in mutant reveals differential expression and autoregulation of the growth factor independence 1 (Gfi1) gene during lymphocyte development. The Journal of Biological Chemistry *279*, 40906–40917.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355–364.

Zarnegar, M. a, Chen, J., and Rothenberg, E. V (2010). Cell-type-specific activation and repression of PU.1 by a complex of discrete, functionally specialized cis-regulatory elements. Molecular and Cellular Biology *30*, 4922–4939.

Zeng, H., Yücel, R., Kosan, C., Klein-Hitpass, L., and Möröy, T. (2004). Transcription factor Gfi1 regulates self-renewal and engraftment of hematopoietic stem cells. The EMBO Journal *23*, 4116–4125.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails. Genes and Development *15*, 2343–2360.

Zhang, D., Fang, P., Jiang, X., Nelson, J., Moore, J.K., Kruger, W.D., Berretta, R.M., Houser, S.R., Yang, X., and Wang, H. (2012a). Severe hyperhomocysteinemia promotes bone marrow-

derived and resident inflammatory monocyte differentiation and atherosclerosis in LDLr/CBS-deficient mice. Circulation Research *111*, 37–49.

Zhang, D.E., Hetherington, C.J., Chen, H.M., and Tenen, D.G. (1994). The macrophage transcription factor PU.1 directs tissue-specific expression of the macrophage colony-stimulating factor receptor. Molecular and Cellular Biology *14*, 373–381.

Zhang, J., Jackson, A.F., Naito, T., Dose, M., Seavitt, J., Liu, F., Heller, E.J., Kashiwagi, M., Yoshida, T., Gounari, F., et al. (2011). Harnessing of the nucleosome-remodeling-deacetylase complex controls lymphocyte development and prevents leukemogenesis. Nature Immunology *13*, 86–94.

Zhang, J. a., Mortazavi, A., Williams, B. a., Wold, B.J., and Rothenberg, E. V. (2012b). Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell *149*, 467–482.

Zhu, J., and Paul, W.E. (2010). Heterogeneity and plasticity of T helper cells. Cell Research *20*, 4–12.

Zimmermann, M., Aguilera, F.B., Castellucci, M., Rossato, M., Costa, S., Lunardi, C., Ostuni, R., Girolomoni, G., Natoli, G., Bazzoni, F., et al. (2015). Chromatin remodelling and autocrine TNFα are required for optimal interleukin-6 expression in activated human neutrophils. Nature Communications *6*, 6061.

Zouali, M. (2014). Transcriptional and metabolic pre-B cell receptor-mediated checkpoints: Implications for autoimmune diseases. Molecular Immunology *62*, 315–320.