Global Analysis of Dynamic Epigenetic Marking and Transcriptional Regulation Underlying T-cell Lineage Commitment

Thesis by

Jingli A. Zhang

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2012

(Defended September 29, 2011)

Acknowledgements

It has been a long journey for me to finally reach this point — the final section of my PhD thesis. There are far too many people I would like to acknowledge who have guided, collaborated with, helped and rescued, and sometimes endured with me.

Firstly and most importantly, I would like to thank my advisor, Dr. Ellen Rothenberg, for providing inspiration, supervision, and unreserved support during my PhD studies and researches. Her acute scientific insights, her breadth of knowledge, and her ever-encouraging way of providing guidance, new ideas, and discussions have been more than indispensable during my PhD career. Dr. Rothenberg has also shown me that a career devoted to science can be satisfying, rewarding, and above all, intriguing. Following her footsteps, I witnessed that Biology, in addition to having huge practical values to the human race, is also astonishingly beautiful.

Working in Rothenberg lab has always been very stimulating and fulfilling due to the highly collaborative and supportive culture. I am deeply thankful to our lab manager Rochelle (Shelley) Diamond. Shelley makes sure that everything in and around the lab runs smoothly, and she was always there when I was lost and needed be rescued. More than being a "super mom" figure to the whole lab, I shall not forget the countless times she lent her rich knowledge in flow cytometry to help me solve FACS analysis problems.

My experience and growth in Rothenberg lab also is inseparable from all the collaboration, comradeship, and companionship, with all of my lab members and peers. Here I would take the opportunity to acknowledge a few of them. I would like to thank Dr. Long Li, with whom I am currently collaborating on a very exciting Bcl11b project. I am deeply impressed by his work and research attitude, intelligence and insights, and

collaboration spirits. I would also like to thank Dr. Hao Yuan Kueh, for providing numerous critical advice on selecting the right statistical tools in data analysis, and helping review my manuscript as well as this thesis; Marissa Morales, for always being there for discussing, helping, cheering, and comforting each other; Dr. Mary Yui and Dr. Ameya Champhekar, for all the teachings and tips I received in various and numerous technologies for doing successful experiments; Ni Feng and Rob Butler, for being such super techs in the lab, and who can always be counted on to address all sorts of preparation and experimental deeds.

I would like to acknowledge one of my chief collaborating researchers, Dr. Ali Mortazavi, for teaching me how to use his powerful ChIP-seq and RNA-seq analysis software, starting with how to open the terminal window all the way to the successful data analyses; for enduring my endless questions on statistics, software, and a lot of his other expertise; and for sharing many exciting discovers and advances in his own research. I also want to thank Ali's former adviser Dr. Barbara Wold for being supportive and providing valuable advice on our genomic project. In fact, it was in part because of Dr. Wold's original suggestion and encouragement that our genomics project was triggered and grew.

I have been constantly asking for help from Dr. Brian Williams and Georgi Marinov of Wold lab. I would like to thank Brian for helping me improve the ChIP-seq and RNA-seq protocols, and for preparing RNA-seq libraries of our genomic project; and Georgi for intellectual discussions on statistics models and different ChIP-seq and RNA-seq analysis programs.

I also got lots of support from people in Caltech facilities. My paper cannot be

done without purified cell populations. Diana Perez in Flow Cytometry Facility spent hundreds of hours, often overtime, on FACS cell sorting for me without any complaint. Igor Antoshechkin, Lorian Schaeffer, and Vijaya Kumar in Millard and Muriel Jacobs Genetics and Genomics Laboratory helped me prepare over 100 ChIP- or RNA-seq libraries, and ran the sequencing, which is pivotal to the success of this genomics project. I am sincerely grateful to these wonderful people.

I would also like to thank all my committee members, Dr. Marianne Bronner, Dr. Sarkis Mazmanian, Dr. Angela Stathopoulos, and Dr. Paul Sternberg. As accomplished scholars and researchers with super busy schedules, each and every one of them dived into my project, offered diverse perspective and insightful inputs during my research committee meetings and many other discussions.

Last but not least, I would like to thank my parents for always believing and understanding me even when they really needed me but I was absent, and my husband Tian for his support and faith that kept me continuing. In addition, Tian has lent his expertise in software and programming to help address various computer obstacles especially when I was working at home, which is invaluable for my project.

Dedicated to Tian and Anthony Ling

Abstract

T-cell lineage specification and commitment success depends on precise temporal induction of T-lineage specific genes, as well as repression of lineage-inappropriate programs. After entry into the thymus, T-cell progenitors still retain inherited lineage plasticity, reflected by the mixed-lineage pattern of gene expression and the abilities to give rise to alternative lineages. Although Notch-Delta signaling is an essential force to trigger and sustain T-lineage differentiation, it does not appear to be the only requirement for this process. Successful commitment also depends on additional transcription factors, which often cooperatively interact with Notch-Delta signaling. However, the molecular mechanism by which pro-T cells are advanced to become committed T cells, in particular how the alternative lineage potentials are eliminated, is not fully understood. Using the genome-wide high-throughput sequencing, we track global shifts in gene expression pattern and transcriptional activity associated histone modifications in five successive stages of T-cell differentiation that span the commitment process. Our results show that T-lineage commitment is defined by the surprisingly complex downregulation of progenitor- and/or alternative lineage-associated programs, with relatively few regulatory genes are substantially upregulated. Rather than being silenced by a single global repression event, progenitor- and/or alternative lineage-associated genes are regulated by individual gene-specific mechanisms, indicated by the unsynchronized epigenetic transformations at discrete cis-elements of genes loci linked to progenitor and/or alternative lineage programs. We also investigate the genome-wide occupancies of PU.1 and GATA-3, two regulatory factors that have critical but complementary roles in early T-cell development. Binding sites choices of these two factors imply that transcriptional

regulation by one particular factor is developmental context as well as dosage dependent. Furthermore, We combine this genome-wide approach with gene perturbation to study the function of Bcl11b, a transcription factor required for the completion of T-cell lineage commitment. Our analyses reveal that, in part through directly or indirectly regulation of Notch1 and GATA-3, Bcl11b mediates the modulation of T-cell lineage specification and commitment.

Table of Contents

Chapter 1

Introduction

T-cell lineage specification and commitment occur in thymus. Nevertheless, committed T cells ultimately derive from hematopoietic stem cells (HSCs), which reside in the bone marrow. Rather than through a single step or cell division, this is an irreversible journey including multiple distinct successive stages and various fate choice decisions a cell has to make. HSCs are rare self-renewing cells, and sit atop a hierarchy of all adult blood cell lineage progenitors that become progressively lineage restricted and eventually commit to a single blood lineage with specific characteristics (Lai and Kondo, 2008; Luc et al., 2008; Orkin and Zon, 2008) (Figure 1). T-cell development from multipotent self-renewing HSCs is an intertwining process comprising gradual loss of self-renewal and disparate lineage potentials, and stepwise activation of T-lineage specification program that prepares for the generation of diverse functional T cells in periphery. It is achieved without altering DNA sequence, but through coordinative activation and repression of subsets of genes that collectively favor T-cell fate over others. This process is regulated by various transcription factors that act synergistically or antagonistically in response to environmental cues at different stages. The combinatorial interactions of these transcription factors are intimately connected with the functions of chromatin-modifying enzymes and chromatin-remodeling complexes, which govern the accessibility of lineage-specific gene loci and ultimately provide control over the gene expression program.

Any understanding of the molecular mechanism by which T-lineage program is installed must take into account the fact that thymic settling progenitors inherit a mixed-lineage pattern of gene expression and epigenetic structure from previous stages, and

therefore the intrathymic development, modulated by a distinct set of regulators, has to attain epigenetic switches that eventually confer durable accessibilities at T-cell gene loci, as well as permanently silent structures at progenitor-specific gene loci to make commitment irreversible.

Developmental Origins of Hematopoietic Stem Cells

Establishment and maintenance of blood system relies on the extensive self-renewal ability and all blood cell lineage differentiation potentials of HSCs. In vertebrate embryos, HSCs have been found to arise mainly in two independent anatomic sites, the yolk sac (or the ventral blood islands in amphibians), which transiently produces blood cell during embryonic stages, and the intraembryonic aorta-gonad mesonephros (AMG) region, where fully competent adult-type HSCs are originated (Dzierzak, 1999; Durand and Dzierzak, 2005). In addition, large numbers of HSCs have been found in the mouse placenta parallel to the emergence of HSCs in AMG (Gekas et al., 2005; Ottersbach and Dzierzak, 2005). Subsequently, circulating HSCs migrate to and expand in some secondary tissues, such as fetal liver, thymus, spleen, and eventually the bone marrow, where they remain throughout adult life. None of these secondary tissues are believed to be the founding sites of HSCs, but instead, their niches provide support for long-term maintenance of HSCs (Durand and Dzierzak, 2005; Orkin and Zon, 2008).

From HSCs to Lineage-Restricted Progenitors: A Hierarchical Model of Hematopoiesis

At least eight morphologically and functionally distinct mature blood cell types arise from HSCs (Figure 1). Lineage-restricted differentiation from HSCs is a sequential and irreversible process, in which alternative lineage choices become increasingly limited

(Dzierzak, 1999; Lai and Kondo, 2008; Luc et al., 2008) (Figure 1).

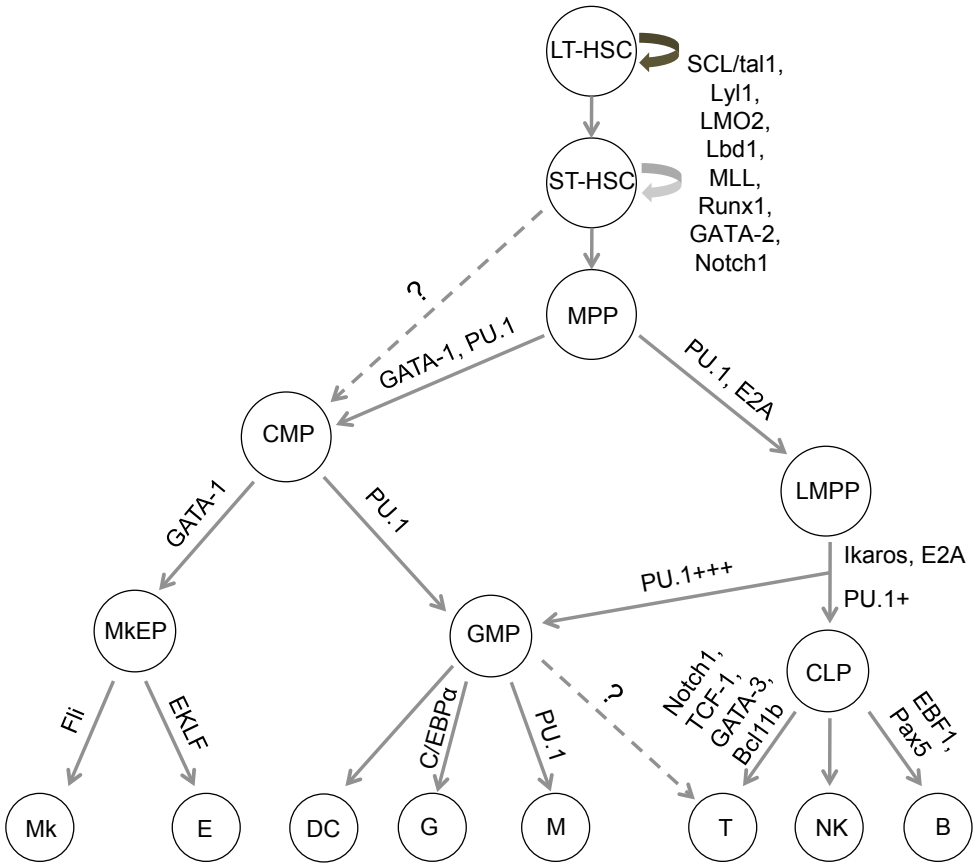


Figure 1. Hierarchical diagram of adult hematopoiesis.

LT-HSC, long-term hematopoietic stem cell; ST-HSC, long-term hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; LMPP, lymphoid primed multipotent progenitor; MkEP, megakaryocyte/erythroid progenitor; GMP, granulocyte/monocyte progenitor; CLP, common lymphoid progenitors; Mk, megakaryocyte; E, erythroid; G, granulocyte; M, monocyte (or macrophage); DC, dendritic cell; T, T cell; NK, natural killer cell; B, B cell. Curved arrows highlight the indefinite or limited self-renewal ability of LT-HSCs (darker arrow) and ST-HSCs (lighter arrow), respectively. Some known HSC regulators and lineage-restricting factors

(but not limited to) are indicated. Question mark with dashed arrow line indicates the relationship remains to be established.

Upon receiving differentiation signals, the first biological response in HSCs is the gradual loss of self-renewal ability. Based on the various cell surface markers and the ability of maintaining long-term self-renewal characteristics, HSCs can be categorized as long-term (LT-) HSCs ($CD34^{-}Flt3^{-}Lin^{-}Sca^{+}c\text{-}Kit^{+}$ (LSK)), which have a lifetime self-renewal capacity; and short-term (ST-) HSCs ($CD34^{+}Flt3^{-}$LSK), which are more limited in self-renewal and repopulation (Lai and Kondo, 2008; Luc et al., 2008). Upregulation of FMS-like tyrosine kinase 3 (Flt3) expression within the LSK compartment marks the gradual loss of the self-renewal ability and the transition to multipotent progenitors (MPPs, $CD34^{+}Flt3^{lo}VCAM\text{-}1^{+}$LSK), which are able to further differentiate into megakaryocyte/erythroid (MkE), granulocyte/monocyte (GM) and lymphoid lineage. The expression of Flt3 regulates the earliest lymphoid commitment step from HSCs (Sitnicka et al., 2002). Compelling evidences show that, under different intrinsic and extrinsic cues, MPP (or ST-HSCs) can give rise to either GM-lymphoid restricted lymphoid primed multipotent progenitors (LMPPs, $CD34^{+}Flt3^{hi}VCAM\text{-}1^{+}$LSK), which lose the MkE potential (Adolfsson et al., 2005), or common myeloid progenitors (CMPs, $CD34^{+}Flt3^{-}VCAM\text{-}1^{+}$LSK), which have potential for both GM (through granulocyte/monocyte progenitors, GMPs) and MkE (through megakaryocyte/erythroid progenitors, MkEPs) lineages but not lymphoid lineage (Lai and Kondo, 2008; Luc et al., 2008).

Within the LMPP compartment, GM-potential decreases with increased lymphoid potential, reflected by marked loss of surface expression of VCAM-1. While the most

primitive $Flt3^{+}VCAM\text{-}1^{+}$ LSK cells can differentiate into both GM and lymphoid lineages, more developmental advanced $Flt3^{+}VCAM\text{-}1^{-}$ LSK cells only give rise to common lymphoid progenitors (CLPs), precursors for all lymphoid cells including T, B, and NK cells (Lai et al., 2005; Lai and Kondo, 2006).

Temporal- and Stage-Specific Transcription Factors for Hematopoiesis

Hematopoiesis has long been used as a model system for studying transcriptional regulation in lineage differentiation. As ultimate determinants governing cell type-specific gene expression, transcription factors involved in hematopoiesis are of special interest in understanding how HSCs are emerged and maintained, how lineage restriction process is programmed, and how genes reflective of diverse hematopoietic lineages are regulated (Luc et al., 2008; Orkin and Zon, 2008).

Unlike embryonic stem cells, HSCs do not have a defined set of core regulators. Various HSC specific and general transcription factors have been shown to be required in HSCs' formation, maintenance and function (Figure 1). These factors include the basic helix-loop-helix (bHLH) factors SCL/tal1 and Lyl1, the LIM domain-containing protein LMO2, the LIM domain-binding protein Lbd1, the SET-domain containing histone methyltransferase MLL, runt-domain protein Runx1, zinc finger transcription factor GATA-2, Notch1, and several others (Godin and Cumano, 2002; Orkin and Zon, 2008). While Mice deficient in any one of SCL/tal1, its associated non-DNA binding partner LMO2, or MLL lack both primary (yolk sac) and definitive (AGM or Adult) hematopoiesis (Robb et al., 1995; Porcher et al., 1996; Robb et al., 1996; Yamada et al., 1998; Ernst et al., 2004; Jude et al., 2007), GATA-2, Notch1 and it downstream factor Runx1 are each individually important for the specification and expansion of HSCs

within AGM, but have less pronounced or no effect on primary hematopoiesis (Tsai et al., 1994; Kumano et al., 2003; Burns et al., 2005). The differences in molecular regulation imply the possible independent origins between primary and definitive hematopoiesis (Godin and Cumano, 2002). On the other hand, the requirement for particular transcription factor(s) can be stage specific. For example, although being required for the ontogeny of HSCs, SCL/tal1 is no longer necessary for the subsequent maintenance or self-renewal of HSCs and multipotent progenitors, and instead becomes essential for proper differentiation of erythroid and megakaryocytic precursors (Mikkola et al., 2003) (and see below). A closely related bHLH factor Lyl1 is believed to compensate the loss of SCL/tal1 in maintaining HSC function (Souroullas et al., 2009).

Lineage specification and commitment involves fundamental changes of the cell's gene expression program. These changes include induction of genes that confer lineage-specific function, and repression of genes antagonizing its differentiation. This process can be accomplished by the concerted action of general and lineage-restricting transcription factors (Figure 1). GATA-1, another GATA family member that is highly expressed in megakaryocyte/erythroid progenitors (MEPs) and red blood cell precursors, is one of the most critical transcription factors regulating proper erythroid and megakaryocyte development. Loss of GATA-1 blocks erythroid development at proerythroblast stage, and also leads to defects in megakaryocyte maturation and platelet formation, while myeloid and lymphoid development is not affected (Weiss and Orkin, 1995; Fujiwara et al., 1996; Shivdasani et al., 1997; Vyas et al., 1999). Furthermore, a modest decrease in GATA-1 level caused by interruption of an upstream cis-regulatory element results in lower efficiency of erythroid maturation (McDevitt et al., 1997).

Conversely, PU.1, an ETS family transcription factor, is a multilevel regulator that is involved in multiple decision checkpoints in both myeloid lineage and lymphoid lineage pathways. PU.1 promotes myelopoiesis at the expense of erythroid/megakaryocyte/ eosinophil differentiation in zebrafish (Galloway et al., 2005; Rhodes et al., 2005). Conditional deletion of PU.1 in BM leads to a complete loss of CMPs, GMPs, and CLPs, but normal or slightly increased MkEPs (Iwasaki and Akashi, 2007). In addition, fate choice between macrophage vs. B cell is regulated by the graded expression of PU.1. While high levels of PU.1 expression promote macrophage differentiation, low levels of PU.1 favor B-cell formation (DeKoter and Singh, 2000). Besides being involved in directing macrophage and B-cell differentiation, PU.1 also plays a functionally distinct role in early T-cell development (see Chapter 2). It is required for the survival of uncommitted pro-T cell, but has to be silent otherwise becomes antagonistic, during or immediately after the T-lineage commitment (see below).

In addition to GATA-1 and PU.1, a variety of lineage-specific transcription factors are involved in regulating the specification and commitment of different hematopoietic lineages. Acting cooperatively with GATA-1, EKLF (Erythroid Krüppel-like factor) facilities erythroid lineage commitment and directly upregulates erythroid gene expression program (Siatecka and Bieker, 2011), whereas GATA-2 and ETS family transcription factor Fli help maintain megakaryocyte identity (Starck et al., 2003; Huang et al., 2009). Likewise, along the opposite branch of hematopoiesis, distinct sets of transcription factors are required in myeloid and lymphoid lineage pathways. CCAAT enhancer-binding protein C/EBPs, including C/EBPα and C/EBPβ, cooperate with PU.1 in granulocyte/monocyte development in part by directly regulating the expression of

many myeloid genes (Friedman, 2002). C/EBPα null mice have normal numbers of mature MkE cells and lymphoid, but completely lack granulocytes, a defect that can be rescued by inserting C/EBPβ into C/EBPα locus, indicating the essential roles for C/EBPs in granulopoiesis (Iwasaki and Akashi, 2007). Moreover, C/EBPα also promotes differentiation by limiting HSC self-renewal and repopulating capacity through the inhibition of polycomb gene Bmi1(Zhang et al., 2004).

With respect to lymphoid development, two transcription factors, the Krüppel-type zinc finger DNA binding protein Ikaros and the bHLH E protein E2A, have been shown to be the common regulators for both B- and T-lineage differentiation. While E2A is required for LMPP generation from HSCs or MPPs (Dias et al., 2008), LMPPs can arise from Ikaros-deficient HSCs or MPPs; however, Ikaros-deficient LMPPs are unable to produce B cells, and display reduced T-cell potential (Yoshida et al., 2006; Nutt and Kee, 2007). Further restrictions into either B- or T-lineage pathways are achieved by the inputs from additional B- or T-lineage specific factors.

Under B-cell differentiation conditions, E2A first induces the COE family transcription factor EBF1 (Early B cell Factor 1). EBF1 and E2A cooperatively promote B-lineage specification through upregulation of most B-cell signature genes (Nutt and Kee, 2007). One EBF1 target is Pax5, an essential factor for completion of B-lineage commitment process. Pax5 not only activates B-lineage specific genes, but, more importantly, also represses alternate lineage programs (Cobaleda et al., 2007b). In the absence of Pax5, B-cell development is blocked at pro-B cell stage when D-J_H rearrangement occurs (Nutt et al., 1997). These mutant pro-B cells can grow indefinitely *in vitro* in the presence of IL-7 and stroma, express cell surface receptors MCSF-R

(macrophage colony stimulating factor receptor) and Notch, and are capable of differentiating into other hematopoietic lineages, such as macrophage and T cell (Nutt et al., 1999; Rolink et al., 1999). In addition, Pax5 also upregulates E2A and EBF1, which helps amplify B-cell program and conceal its lineage fate (Nutt and Kee, 2007).

T-lineage specification and commitment, on the other hand, are regulated by Notch-Delta signaling and various T-cell-specific or non-T-specific factors, including (but not limited to) the high-mobility group (HMG) transcription factor TCF-1, GATA-3, Bcl11b, PU.1 and E2A (see below).

Combinatorial Interactions between Key Regulators Define Hematopoeisis

As discussed so far, hematopoeisis is regulated by a complex hierarchical system involving a considerable number of transcription factors. Orchestration of lineage commitment is often modulated by synergistic or antagonistic (competitive or inhibitive) protein-protein interactions among different regulators, or by directly transcriptional regulation of each other. In HSCs, SCL/tal1, Lmo2 and GATA-2 simultaneously interact with a ubiquitously expressed nuclear adaptor Ldb1 to form a transactivating multiprotein complex that cooperatively promotes the HSC-specific program (Kim and Bresnick, 2007; Li et al., 2011). GATA-1, as a central mediator of erythropoiesis, can partner with a group of functionally distinct transcription factors including SCL/tal1, Lmo2, FOG1, E2A and HEB, EKLF, CBP/p300 and Sp1 to keep erythrocyte development on track (Cantor and Orkin, 2002). In particular, SCL/tal1 and its heterodimeric partners E2A and HEB have been shown to preferentially cooccupy with GATA-1 at the activated gene loci, revealing a molecular mechanism by which GATA-1 distinguishes targets for transcriptional activation vs. repression (Yu et al., 2009). Likewise, recent genome-wide

binding studies in macrophages and B cells have highlighted the coordinated DNA binding between PU.1, a common regulator of macrophage and B-cell differentiation, and subsets of lineage-specific transcription factors (for instance, C/EBPs in macrophage or E2A and Oct-2 in B cell; see below). Such lineage-specific coordinated binding allows PU.1 and its colocalized partners to selectively regulate cell type-specific genes (Heinz et al., 2010). Similar context-dependent DNA binding mechanisms have been suggested for the stages-specific E2A functions during B-cell development (Lin et al., 2010).

Besides collaboratively promoting their own lineage differentiation, key lineage specific regulators often concurrently antagonize factors associated with disparate lineages in a context-dependent manner (Orkin and Zon, 2008). The antagonism can be achieved via direct protein-protein inhibition, competition for limiting common cofactors to form mutually exclusive protein complex (Siatecka and Bieker, 2011), or transcriptional repression. PU.1 and GATA-1 (as well as its close relative GATA-2) inhibit each other's function by direct physical interaction during the myeloid vs. erythroid/megakaryocytic fate decision. Enforced expression of PU.1 or GATA-1 in multipotent progenitors induce myeloid or erythroid/megakaryocytic lineage commitment, respectively. It has been shown that the N-terminal transactivation domain of PU.1 can specifically block GATA-1/2 DNA binding activity by interacting with the C-terminal zinc-finger of GATA-1/2, which in turn can interfere with PU.1 C-terminal ETS domain (Rekhtman et al., 1999; Zhang et al., 1999; Nerlov et al., 2000; Zhang et al., 2000). GATA-1 also prevents PU.1 from binding to its essential coactivator, c-Jun (Zhang et al., 1999). In addition, GATA-1/2 can inhibit PU.1 by transcriptional repression through direct binding to the promoter and a -18 kb cis-regulatory element of

Sfpi1 (gene that encodes PU.1) (Chou et al., 2009). Additional examples of cross-antagonism have been described at various bifurcation points of hematopoiesis, including EKLF and Fli for erythroid vs. megakaryocyte (Starck et al., 2003), C/EBPα and Pax5 for macrophage vs. B cell (Xie et al., 2004; Hsu et al., 2006), and Notch1 and Pax5 for T vs. B fate decision (Pui et al., 1999; Cobaleda et al., 2007a). With respect to myeloid development, although both PU.1 and C/EBPα are required for myeloid lineage differentiation, and both are highly expressed within myeloid progenitors and synergistically regulate myeloid gene expression, the cell fate choices for granulocyte and macrophage are in fact determined by the relative level of C/EBPα and PU.1 in granulocyte-macrophage progenitors (Dahl et al., 2003; Laslo et al., 2006). Germline deletion of C/EBPα leads to complete abolishment of granulocytes but not macrophage, whereas PU.1 deficiency in granulocyte-macrophage progenitors results in excess granulocyte production (Zhang et al., 1997; Dakic et al., 2005). The imbalance between PU.1 and C/EBPα is amplified through cross-antagonism among secondary cell fate determinants comprised of Egr-1,2/Nab-2 and Gfi-1. PU.1 activates Egr-1,2/Nab-2, which in turn promote the macrophage program while selectively repressing the granulocyte gene expression. Reciprocally, C/EBPα upregulates the granulocyte program and represses macrophage genes through the action of Gfi-1(Laslo et al., 2006). This type of secondary regulatory circuit has been further elucidated in the context of innate vs. adaptive immune cell fates to explain the graded manner of PU.1 in directing macrophage and B-cell lineage differentiation (Spooner et al., 2009).

Specification and Commitment of T-lineage Program

Distinct Stages of Early T-cell Developmental

Unlike most other hematopoietic lineages, T-lineage specification and commitment occurs outside of BM in the thymus. The thymus requires perpetual supply of hematopoietic progenitors to maintain T-cell development (Foss et al., 2001). Although various different BM progenitors are able to give rise to thymocytes if they are transferred intrathymically or exposed to permissive conditions *in vitro*, the precise identity of the thymic settling progenitor (TSP) remains exclusive (Zlotoff and Bhandoola, 2011). It has been shown the coordinated actions of several membrane homing and adhesion molecules are needed for circling progenitors to enter the thymus. They include the interaction between P-selectin (expressed by the thymic endothelium) and P-selection glycoprotein ligand-1 (PSGL-1; found on subsets of LSKs and CLPs), and between thymic-expressed chemokines CCL19/21/25 and two chemokine-receptors CCR7/CCR9 (found on subsets of LSKs and CLPs) (Zlotoff and Bhandoola, 2011). It is likely that the candidate TSPs express at least one of such surface molecules (PSGL-1, CCR7 and CCR9).

Early T-cell development can be separated into a series of distinct developmental stages defined by unique constellations of cell surface markers and/or developmental potentials, as well as TCR (T cell receptor) gene recombination (Figure 2). The least mature intrathymic progenitors are $Lin^{-}c\text{-}Kit^{hi}CD44^{+}CD25^{-}$ (Lin: CD3, CD4, CD8, TCRβ, TCRγδ, CD11b, CD11c, CD19 and NK1.1), and they are termed as ETPs (early T-cell precursors), or DN1 (double negative 1; $CD4^{-}CD8^{-}$) cells. ETPs are very efficient in proliferation and can give rise to $Lin^{-}c\text{-}Kit^{+}CD44^{+}CD25^{+}$ DN2 cells. Based on the

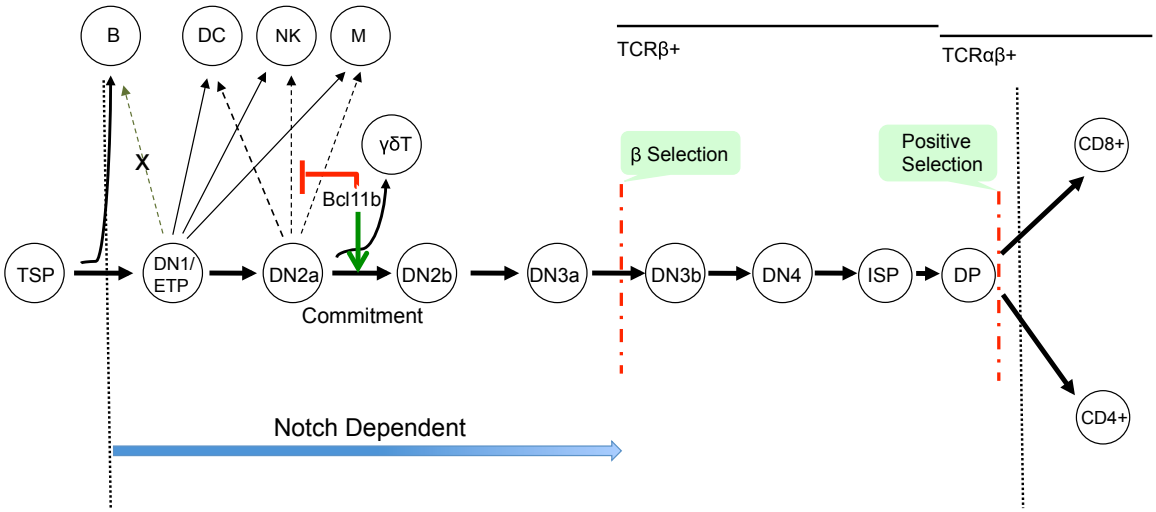


Figure 2. Landmarks for early T-cell development.

TSP, thymic-settling progenitor; ETP, early T-cell progenitor; DN, double negative (CD4-CD8-); ISP, intermediate single positive; DP, double positive (CD4+CD8+). CD4+, CD4 T cell; CD8+, CD8 T cell; M, macrophage; DC, dendritic cell; NK, natural killer cell; B, B cell. Space between two black dashed lines represents thymus. Differentiation from thymic-settling progenitor to DN3a including β-selection is Notch-dependent. Commitment occurs during the progression from DN2a to DN2b. Bcl11b is required for the completion of T-lineage commitment process, during which progenitor- and/or alternative lineage-associated gene programs are permanently silent.

surface expression level of c-Kit, DN2 can be subdivided into earlier c-Kit^{hi} DN2a cells and later c-Kit^{int} DN2b cells (Yui et al., 2010). With the exception of a rare subset of Flt3^{+}CCR9^{+} ETPs (presumably the most immature intrathymic T-cell precursors) that are capable of giving rise to B cell in the absence of Notch signaling (and under B-cell permissive condition), majority of ETPs have lost B-cell potential (Rothenberg et al.,

2008). Nevertheless, both ETP and DN2a cells still retain developmental potentials of NK, dendritic cells (DC) and myeloid lineages, although these alternative lineage potentials are gradually reduced in DN2a (Bell and Bhandoola, 2008; Rothenberg et al., 2008; Wada et al., 2008). Lineage commitment occurs as DN2a cells differentiate into DN2b cells. This is the first critical checkpoint for T-cell development inside thymus, when non-T-fate choices and self-renewal potential are terminated and before extensive TCR recombination takes place (Masuda et al., 2007; Yui et al., 2010). DN2b cells further downregulate c-Kit and CD44, and become c-Kit^{lo}CD44^{-}CD25^{+} DN3 cells. Based on cell surface CD27 expression level and cell size, DN3 cells are further subdivided into those pre- and immediately post-β-selection (or γδ-selection; DN3a and DN3b, respectively). β-selection is another major developmental checkpoint where DN3 cell express pre-TCRβ after a successful TCRβ VDJ recombination (Taghon et al., 2006). Passing β-selection, cells begin to downregulate CD25 expression and eventually become DN4 cells (c-Kit$^{lo/-}$CD44^{-}CD25^{-}). After DN4 stage, cells start to turn on cell surface expression of both CD4 and CD8, and become DP (double positive; CD4^{+}CD8^{+}) cells. DP cells ultimately acquire TCRαβ complex, which prepares cells for the positive selection and negative selection to generate mature CD4 or CD8 SP (single positive) T cells (Rothenberg et al., 2008).

Notch-Delta Signaling Regulates Early T-cell Development

Thymic epithelial cells supply various ligands, growth factors and cytokines that are required for T-cell differentiation, proliferation, and survival. Among them, arguably the most important one is Delta-like 4, the ligand for the T-cell surface receptor Notch1. For normal T-cell development to occur, potential T-cell progenitors must be able to

response to Notch signaling after entering the thymus (Sambandam et al., 2005). Notch-Delta interaction initiates a cascade of proteolytic reactions that liberate Notch intracellular domain (NICD). The NICD then translocates into the nucleus and heteodimerizes with transcription factor CSL (RBP-J). This binding converts CSL from a repressor to a activator, and subsequently activates the expression of target genes (Yashiro-Ohtani et al., 2010). Notch-Delta signaling is an essential driving force during early stages of T-cell development up to and including β-selection (or γδ-selection). Upon successful β-selection, though being continuously active, Notch-Delta signaling becomes dispensable for the rest of maturation process (Rothenberg et al., 2008; Yashiro-Ohtani et al., 2010). Notch signaling promotes T-lineage specification and inhibits the alternative lineage potentials in pro-T cells. It has been shown that overexpression or constitutive activation of Notch signaling in bone marrow progenitors promotes thymic-independent T-cell formation at the expense of B-cell development, while deletion or inhibition of Notch in progenitors or DN1 cells abolishes T-cell development and results in vastly increased intrathymic B cells and dendritic cells (Feyerabend et al., 2009; Yashiro-Ohtani et al., 2010). At the moment, the mechanism by which the B-lineage potential is repressed by Notch1 remains unclear. A closely related question is whether additional factors other than Notch1 are also required to keep pro-T cells on track from converting into B cells, since B cells can no longer arise from $CD25^{+}$ DN2a or DN2b cells even if Notch signaling is inactivated (Schmitt et al., 2004; Sambandam et al., 2005; Feyerabend et al., 2009).

Additional Essential Regulators for Early T-cell Development

Notch1 is not a T-lineage specific regulator; rather, it is a multilevel factor involved in various aspects of embryonic development. In order to launch a T-specific program, Notch signaling has to collaborate with many other regulators (Rothenberg et al., 2008). One such factor is E2A. E2A is expressed at constantly high level throughout early T-cell development (see Chapter 2). As discussed earlier, E2A is a common regulator for both B- and T-lineage development. E2A deficiency causes a partial arrest at the DN1 stage and leads to T-cell lymphomas at very young age (Bain et al., 1997). Retroviral infection of E2A-deficient fetal liver precursors with NICD recues the developmental block (Ikawa et al., 2006). The presence of CSL-preferred binding sites and neighboring bHLH A protein binding sites has been suggested to be the minimal DNA code necessary and sufficient to drive Notch signaling mediated gene activation (Cave et al., 2005), which may in turn help E2A selectively target T-cell genes instead of B-cell genes. Further characterizations uncovered that E47 (a splicing isoform of E2A gene) is directly involved in activating a group of genes known to be Notch1 targets, including *Hes1*, *Hes5*, *Ptcra* and *Notch3* (Ikawa et al., 2006). Taken together, these findings suggest that E2A acts synergistically with Notch signaling to modulate T-lineage differentiation.

Another multilineage regulator that is required for early T-cell development is PU.1. PU.1 is highly expressed in prethymic mulitpotent progenitors as well as DN1 cells, and is slightly reduced but still at considerable level in DN2a cells. It is rapidly downregulated during commitment, and essentially undetectable before β-selection (Nutt et al., 2005; Tydell et al., 2007) (see Chapter 2). PU.1 is required for generating thymic

settling progenitors, and PU.1 deficiency leads to a developmental block at DN1-like stage (Scott et al., 1994; Spain et al., 1999; Dakic et al., 2005). We have shown that PU.1 global occupancy pattern in DN1 (and DN2a) cells is significantly distinct from that in B cell or macrophage, implying PU.1 DNA binding is lineage specific. Furthermore, a substantial percentage of PU.1 targets are either upregulated or stably expressed in pro-T cells, including many T-cell identity genes (see Chapter 2). These observations collectively suggest that PU.1 may positively regulate T-cell specification during early stages of development. On the other hand, PU.1 downregulation is a prerequisite for the T-lineage commitment; PU.1 antagonizes T-cell development if expressed constitutively in DN2b or DN3 cells (Anderson et al., 2002). Enforced expression of PU.1 in pro-T cells reduces *Hes1* and *c-Myb* mRNA level, suggesting an antagonistic function of PU.1 imposed on Notch-Delta signaling (Anderson et al., 2002). Moreover, exogenous PU.1, either by itself or in collaboration with exogenous C/EBPα, is capable of reprograming committed DN3 cells into myeloid fate, and this process can be blocked, at least partially, by Notch-Delta signaling (Franco et al., 2006; Laiosa et al., 2006). Thus, during early stages of T-cell development PU.1 may "negotiate" with Notch1 and other factors in establishing a balance between promoting T-lineage specification and at the same time preserving some non-T-cell potentials (Rothenberg, 2007).

Among the core group of T-lineage regulatory factors, TCF-1, GATA-3 and Bcl11b are the few factors that are expressed in a seemingly T-lineage-specific manner (Rothenberg et al., 2010). Besides to Notch1, TCF-1 is the only other identified factor that can specifically direct bone marrow progenitors into the T-lineage pathway (Weber et al., 2011). TCF-1 is turned on at a substantial level in the earliest DN1 cells. Its

expression moderately increases from DN1 to DN2a, and stays steady afterward (see Chapter 2). TCF-1 is indispensable for T-lineage specification. Recently, it has been shown to be a downstream mediator of Notch signaling, and, if ectopically expressed in bone marrow LMPP, be able to promote T-cell development in the absence of Notch signaling (Weber et al., 2011). TCF-1 directly upregulates a group of T-cell identity genes involved in TCR complex assembly and signaling components, including *Lck*, *Cd3g/d/e*, *Lat* and *Rag2*. However, without Notch-Delta signaling, TCF-1-expression precursors only give rise to DN2/3-like cell, and fail to induce *Dtx1* and *Ptcra*, two known Notch1 targets, the latter one of which is an essential component of pre-TCRβ complex (Weber et al., 2011). Thus, a TCF-1-indenpendent Notch pathway is also required for the completion of T-cell specification.

Similar to TCF-1, GATA-3 is among the earliest factors induced by Notch signaling in T-cell precursors (Taghon et al., 2005). It is the only GATA factor that is T-lineage specific and is required recurrently during many stages of T-cell development. It is essential for the whole T-lineage commitment process from DN1 stage and onward. GATA-3 is also the main driving force in generating peripheral T helper 2 (Th2) cell and maintaining its polarized immune response (Paul and Zhu, 2010). Compared to that in Th2 cell, GATA-3 expression in early stages (i.e., DN1 to DN3) of T-cell development is relatively low and rather stable (within 2~4x change, see Chapter 2). Genome-wide DNA binding studies reveal that GATA-3 binds to a number of T-cell identity genes, implicating its role in T-lineage specification (Wei et al., 2011) (see Chapter 2). GATA-3 function in early T-cell development has been shown to be context- and dosage-dependent. GATA-3-deficient embryonic stem cells can contribute to the development of

mature erythroid, myeloid and B-lineages, but fail to give rise to T-lineage (Ting et al., 1996). Paradoxically, rather than promoting developmental progression, overexpression of GATA-3 in the DN2a stage either blocks cell survival when Notch-Delta signaling is activated, or diverts differentiation program into mast cell pathway under non-T-cell permissive condition (Taghon et al., 2007).

In addition to the activation of a specification program for T-cell identity genes and transcription factors that regulate these genes, to become a committed T cell, T-cell precursors must concurrently repress genes associated with progenitor or/and alternative lineages, including abovementioned PU.1. Compelling evidence has shown that these two seemingly related processes are regulated by separate mechanisms (Rothenberg et al., 2010). One regulator functionally implicated to play an essential role in the latter process is Bcl11b. Bcl11b is a Krüppel-like C2H2 type zinc finger transcription factor, and is exclusively expressed in T-lineage of all hematopoietic lineages. It is absent in ETPs, and first turns on at relatively low level in DN2a cells. Bcl11b is among very few transcription factors that are strongly upregulated (>50×) during the commitment from DN2a to DN2b, and is then maintained at similar level throughout T-cell development (Li et al., 2010b) (see Chapter 2). Conditional deletion of Bcl11b in T-lineage progenitors arrests cells at DN2a-like (c-Kit^{hi}CD44^{+}CD25^{+}) stage with increased alternative lineage potentials, in particular NK potential, that are less sensitive to the Notch1 inhibition (Ikawa et al., 2010; Li et al., 2010a; Li et al., 2010b). Unlike GATA-3, Bcl11b is not required for the survival or proliferation of DN2a cells; Bcl11b deficient DN2a-like cells can sustain growth with Notch-Delta signaling and cytokines for weeks *in vitro*. And in contrast to the failure of making αβ T cells, Bcl11b deficient cells are able to give rise to

γδ T cells (Wakabayashi et al., 2003; Li et al., 2010a). Gene expression profiling indicates that Bcl11b-deficient DN2a-like cells upregulate NK-lineage signature factors, including Id2, Zbtb16, Nfil3 and Il2rβ even before these cells actually differentiate into NK-like cells (Li et al., 2010a; Li et al., 2010b) (see Chapter 3). Since some of these genes, such as *Il2rb*, are normally kept silent even before Bcl11b is induced, they may not directly repressed by Bcl11b or repressed by other mechanisms when Bcl11b is absent (Rothenberg et al., 2010). In addition to NK lineage, Bcl11b-deficient cells also fail to completely silence some myeloid and DC lineage as well as progenitor specific genes (Li et al., 2010a).

In summary, as a branch of hematopoeisis, T-cell development in thymus is defined by a progression of phenotypically well-characterized stages. And this process is achieved by deploying a battery of T-lineage specific and non-specific transcription factors to stepwise induce T-cell identity genes and gradually eliminate alterative lineage potentials.

Multilineage Priming and Lineage Plasticity

Using single-cell expression profiling and genetically marked lineage-restricted transcription reporters, such as GATA-1-GFP, PU.1-GFP, Ikaros-GFP, Lysozyme-YFP and Rag1-GFP, researchers have been able to trace the onset and dynamics of lineage-restricted genes during lineage commitment process (Luc et al., 2008; Orkin and Zon, 2008). These studies showed that although majority of lineage-restricting transcription factors or lineage-determinants display lineage-specific expression patterns that mirror the hierarchy diagram of hematopoiesis, such as GATA-1 in megakaryocyte/erythroid progenitors (MEPs) and red blood cell precursors, and PU.1 and C/EBPα in myeloid

lineage, many of them are also "promiscuously" coexpressed, albeit at low level, within HSCs and MPPs prior to differentiation — a phenomenon termed lineage-priming (Hu et al., 1997; Miyamoto et al., 2002; Ye et al., 2003; Adolfsson et al., 2005; Nutt et al., 2005; Arinobu et al., 2007; Mansson et al., 2007; Ng et al., 2009). Lineage-priming has been suggested to reflect the coexisting developmental potentials in HSCs and MPPs, and provide a framework for "crosstalk(s)" between individual lineage potentials in response to different environmental cues (Orkin and Zon, 2008). In HSCs or MPPs, the antagonistic interplays between diverse lineage-affiliated transcription factors are balanced, keeping opposing factors at levels that are too low to elicit lineage differentiation. During differentiation, triggered by extrinsic or intrinsic signals, this balance tilts toward favoring one particular lineage pathway, which resolves the mixed-lineage pattern of gene expression and subsequently propagates the lineage-specific program (Laslo et al., 2006; Ng et al., 2009).

Mixed-lineage patterns of gene expression are also common phenomena in uncommitted precursors, revealing the inherited developmental plasticity before alternative lineage potentials are completely lost (Yoshida et al., 2010) (see Chapter 2). Lineage commitment has long been considered irreversible under physiological conditions. However, instead of being "hit-and-run" factors, several lineage-committing regulators (with the loose exception of Notch1), such as GATA-1, Pax5, C/EBPα and Bcl11b, are continuously present after commitment or sometimes even in terminally differentiated cells. Several studies have found that alteration of a single or a few regulatory factors can "de-commit" lineage committed cells back to uncommitted state, or reprogram committed cells to another cell fate (Kulessa et al., 1995; Xie et al., 2004;

Franco et al., 2006; Laiosa et al., 2006; Cobaleda et al., 2007a; Li et al., 2010b). Taken together, these studies suggest that commitment may actually be more flexible than previously thought. It also suggests that, even in lineage committed cells, the regulatory pathways promoting disparate lineages have to be permanently maintained in silent state by specific lineage-committing factors or by epigenetic modifications (Orkin, 2003).

Epigenetic Modifications and Lineage Differentiation

Through the collective functions of DNA binding regulators and chromatin-modifying enzymes and chromatin-remodeling complexes, external signaling is converted into epigenetic information that often has profound influence on gene expression and cell fate determination. Changes in epigenetic modifications, including histone modifications and DNA methylation, may alter DNA accessibility of transcription factors' target sites, therefore, together with the availability of lineage/stage specific transcription factors control cell type-specific transcription programme (Schones and Zhao, 2008).

While majority of CpG dinucleotides are cytosine methylated, CpG islands, often localized at the promoter regions of many genes, are more resistant to methylation. In HSCs or MPPs, CpG island methylation largely occurs at lineage-restricted gene loci to maintain their long-term stable repression and prevent premature differentiation. Upon lineage differentiation, these loci undergo lineage-restricted demethylation, while, reciprocally, DNA methylation starts to build up at the CpG islands of some stem-cell self-renewal- and/or disparate lineage-associated genes loci (Borgel et al., 2010; Hawkins et al., 2010; Ji et al., 2010; Cedar and Bergman, 2011). In contrast, histone proteins are subject to numerous types of reversible covalent posttranslational modifications,

including acetylation, methylation, phosphorylation and ubiquitination, primarily at the N-terminal tails. Lysine acetylation and methylation are the most studied and better-understood modifications. While lysine acetylation, such as H3K(9,14)Ac, H3K27Ac and H4K16Ac are considered as hallmarks of active or accessible chromatin and transcriptional activity, lysine methylation can have different, sometimes completely opposite, biological effects depending on which residue is modified. Methylation of H3K36 is associated with transcribed chromatin, and methylation of H3K4 is linked to active or accessible chromatin (H3K4me3/2), or poised chromatin state (H3K4me2/1). On the other hand, methylation of H3K9, H3K27 and H4K20 has been implicated to associate with heterochromatic domains and transcriptional repression (Schones and Zhao, 2008). Many DNA-sequence specific transcription regulators also contain functional domains that recognize different modified histones or methylated CpG (Martin and Zhang, 2005; Cedar and Bergman, 2011). Thus, besides regulating the accessibility of sequence-specific DNA target sites, modified (or de-modified) histones and/or methylated CpG islands can act as docking sites for some transcription factors, thereby allowing them to engage chromatinized targets, or in some cases preventing other DNA binding factors from binding to their target sites and causing unregulated gene activation (Martin and Zhang, 2005; Cedar and Bergman, 2011).

Many hematopoietic regulators have been implicated to associate with or possibly even direct chromatin-modifying enzymes and chromatin-remodeling complexes in establishing stem cell or lineage-specific epigenetic landscapes. For instance, in HSCs, members of PRC1 (polycomb repression complex 1), such as Bmi1, Mel18 and Rae28, appear to specifically silence genes that are involved in apoptosis, senescence, and

differentiation, thereby regulating stem cell self-renewal and maintenance (Akala and Clarke, 2006). (However, since it is the PRC2 component histone methyltransferase EZH2 that catalyzes the trimethylation of H3K27, which in turn recruits PRC1, it remains unclear how EZH2 recognizes specific targets in the first place (Cedar and Bergman, 2011)) As discussed earlier, GATA-1 physically interacts with CBP/p300 to recruit HAT (histone acetyltransferases) activity to erythroid-specific gene loci during erythroid development, and thereby facilitating their expression (Cantor and Orkin, 2002). FOG-1, another factor associated with GATA-1 (as well as GATA-2), can bind to CtBP (corepressors C-terminal-binding protein) and the NuRD (nucleosome remodeling and deacetylase) complex, which may selectively mediate GATA-1 linked repression and chromatin remodeling (Kim and Bresnick, 2007). Ikaros, a key regulator for lymphoid development, has been shown to directly associate with two chromatin remodeling complexes, the NuRD and the SWI-SNF complex in both T cells and erythroid progenitors (Ng et al., 2007). It is likely that by selectively altering the accessibility of target gene loci, Ikaros restricts erythroid and/or myeloid-specific programs while promoting lymphoid development (Ng et al., 2007). Through the conserved SNAG domain, Zinc-finger repressors Gfi-1 and Gfi-1b, which respectively play important roles in myelolymphoid development or erythroid and megakaryocytic development, recruits HDACs (histone deacetylases) 1 and 2, CoREST (REST corepressor), and the histone demethlyase LSD1 to their binding sites, leading to the silencing of relevant target genes (Saleque et al., 2007). Interestingly, depending on its partners, LSD1 can act as a repressor or activator by demethylating methylated H3K4 or H3K9, respectively (Metzger et al., 2005; Garcia-Bassets et al., 2007; Wang et al., 2007). Within Gfi-1/1b

mediated multiprotein complex, LSD1 apparently only affects methylated H3K4, suggesting the repression guided by the Gfi proteins is context-dependent (Saleque et al., 2007). In T cells, prior to Notch activation, the DNA binding transcription factor CSL acts as transcriptional repressor through interaction with a corepressor complex containing HDAC-1, SMRT (silencing mediator of retinoid and thyroid hormone receptor) and MINT (Msx2-interacting nuclear target protein), which can in turn recruit CtBP corepressor (Kao et al., 1998; Oswald et al., 2005). Ligand activation of Notch induces the liberation and subsequent nuclear translocation of NICD. NICD displaces corepressors and heterodimerizes with CSL, therefore converting the whole complex into a transcriptional activator. Once bound to CSL, NICD recruits MAML1 (Mastermind-like 1), which in turn interacts with CBP/p300 and lysine acetyltransferases (HATs) PCAF and GCN5 to modulate transcription of specific targets (Kurooka and Honjo, 2000; Wallberg et al., 2002). Finally, Bcl11b has been biochemically implicated to directly interact with the NuRD complex through its N-terminal domain to repress target sites, consistent with its largely repressive function during T-cell development (Cismasiu et al., 2005; Ikawa et al., 2010; Li et al., 2010a; Li et al., 2010b) (see Chapter 3).

Thus, by modifying cis-regulatory elements, epigenetic modifications stabilize the effects of developmentally controlled trans-activators (Cedar and Bergman, 2011). Recent genome-wide analyses have demonstrated that temporal and spatial variations of histone modifications are closely linked with developmentally regulated genes (Bernstein et al., 2005; Mikkelsen et al., 2007; Orford et al., 2008; Araki et al., 2009; Wei et al., 2009; Hawkins et al., 2010; Weishaupt et al., 2010; Koche et al., 2011). The coordination of lineage-specific expression during development is often preceded by the programmed

changes in epigenetic marking. Several lineage-determining transcription factors, including (not limited to) GATA-1, PU.1, E2A and EBF1, have been suggested to be able to induce chromatin remodeling (often at modest level) through the recruitment of methyltransferases. Methyltransferases mono- or di-methylate H3K4 of genomic regions flanking the respective transcription factor binding sites, which often contain a mixture of closely spaced DNA binding sites that are specific to small sets of other lineage-determining transcription factors (Cheng et al., 2009; Heinz et al., 2010; Lin et al., 2010; Treiber et al., 2010). Mono- or di-methylated H3K4 (H3K4me or H3K4me2) makes chromatins accessible to those secondary lineage-determining factors, which in turn facilitate more H3K4 methylation and/or H3/H4 acetylation to eventually entail a lineage-permissive epigenetic landscape available to more lineage-specific or general transcription factors. In addition, collaborative DNA binding of different transcription factors within a short distance may stabilize their binding. Therefore, through epigenetic marking a small set of lineage-determining transcription factors acting in concert initiate and propagate lineage-specific gene program (Heinz et al., 2010; Lin et al., 2010).

In HSC or progenitors, opposing histone modifications (e.g. H3K27m3 and H3K4me3) can be found to coexist “bivalently” at the same gene locus. Upon differentiation, such “bivalency” often resolves in subsequent stages with “bivalent” sites either to become activated by shedding H3K27me3, or reciprocally to be repressed by losing H3K4me3 and/or adding more H3K27me3 (Bernstein et al., 2005; Barski et al., 2007; Mikkelsen et al., 2007; Cui et al., 2009). In addition, it has been shown that developmentally poised genes are more likely enriched with H3K4me2 (Attema et al., 2007; Maes et al., 2008; Orford et al., 2008; Koche et al., 2011) (see Chapter 2). These

observations suggest that H3K4me2/3 marking may serve as a flexible platform for future lineage specification programming (Cedar and Bergman, 2011). One possible effect of lineage-priming (or mixed-lineage gene expression) is to keep chromatins accessible for a broad spectrum of lineage-determining factors through H3K4me2 marking. This permissive chromatin structure becomes progressively lineage-restricted as HSCs or MPPs differentiates (Attema et al., 2007; Maes et al., 2008). For example, although both *Gata1* and *Gata3* are expressed at low or background level in HSCs and MPPs, they are both enriched with H3K4me2 at their promoters. *Gata1* promoter gains more H3K4me2 and H3K4me3 in the MEPs, where *Gata1* expression is upregulated, and becomes demethylated in the CLPs, in which *Gata1* is repressed (marked by H3K27me3). The opposite is true for *Gata3* (Attema et al., 2007). The existing of permissive chromatin structure can be further confirmed by the finding that ectopically expressed EBF1 is able to bind to some B cell-specific gene loci without inducing expression in a non-B hematopoietic cellular context (committed T cells), but fails to do so in a non-hematopoietic context (NIH 3T3 cells) (Treiber et al., 2010). Furthermore, EBF1 is unable to bind to know adipocyte- or neuronal-specific target genes in pro-B cells (Treiber et al., 2010). One possible explanation is that T cells (or B cells) inherit the hematopoietic specific permissive chromatin structure from progenitors (epigenetic memory). However, since B- and T-lineage share several common factors, such as E2A, coordinate EBF1 and E2A (or others) binding may occur when EBF1 is ectopically expressed in T cells (Treiber et al., 2010). It would be informative to investigate whether EBF1 can bind to B cell-specific gene loci in another hematopoietic context, such as committed erythroid cells, which are developmentally more remote than T cells.

Nevertheless, all abovementioned studies suggest that the epigenetic marking plays an important role in modulating lineage differentiation program.

Theme of Thesis

As discussed earlier, multiple T-lineage specific and non-T-lineage specific transcription factors are involved in the regulation of T-cell lineage specification and commitment. The roles that individual factors play during early T-cell development have been the subjects of intensive studies. However, one largely unsolved problem is how T-specific and non-T-specific factors act cooperatively to construct a gene regulatory network that is T-lineage specific. The major roadblock to answer this question is that very few T-lineage specific cis-regulatory elements have been identified.

In Chapter 2, I present that, using genome-wide approaches, we mapped developmentally dynamic histone modification sites, and their linked gene expression. Our study provides a new way to locate potential cis-regulatory elements that translate and stabilize distinct trans-regulatory inputs. I also show that many genes associated with lineage plasticity in pro-T cell are actively repressed by individual gene-specific mechanisms. Furthermore, I show that two essential regulators, PU.1 and GATA-3, exercise their functions in lineage/stage specific manners. In Chapter 3, I summarize an ongoing project about Bcl11b's role during T-cell lineage commitment, which I collaborate on with Dr. Long Li. Using a gene perturbation approach combined with the genome-wide analyses, we profiled global gene expression pattern and GATA-3 occupancy in Bcl11b-deficient cells. Our finding suggests that Bcl11b connects T-lineage specification and commitment partially through properly regulating *Gata3* and *Notch1*. Finally, Chapter 4 concludes with a brief discussion of some implicates from the studies

described herein.

Reference

Adolfsson, J., Mansson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.J., Thoren, L.A., Anderson, K., Sitnicka, E., Sasaki, Y., Sigvardsson, M., and Jacobsen, S.E. (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell *121*, 295-306.

Akala, O.O., and Clarke, M.F. (2006). Hematopoietic stem cell self-renewal. Curr Opin Genet Dev *16*, 496-501.

Anderson, M.K., Hernandez-Hoyos, G., Dionne, C.J., Arias, A.M., Chen, D., and Rothenberg, E.V. (2002). Definition of regulatory network elements for T cell development by perturbation analysis with PU.1 and GATA-3. Dev Biol *246*, 103-121.

Araki, Y., Wang, Z., Zang, C., Wood, W.H., 3rd, Schones, D., Cui, K., Roh, T.Y., Lhotsky, B., Wersto, R.P., Peng, W., Becker, K.G., Zhao, K., and Weng, N.P. (2009). Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory CD8+ T cells. Immunity *30*, 912-925.

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

Attema, J.L., Papathanasiou, P., Forsberg, E.C., Xu, J., Smale, S.T., and Weissman, I.L. (2007). Epigenetic characterization of hematopoietic stem cell differentiation using miniChIP and bisulfite sequencing analysis. Proc Natl Acad Sci U S A *104*, 12371-12376.

Bain, G., Engel, I., Robanus Maandag, E.C., te Riele, H.P., Voland, J.R., Sharp, L.L., Chun, J., Huey, B., Pinkel, D., and Murre, C. (1997). E2A deficiency leads to abnormalities in alphabeta T-cell development and to rapid development of T-cell lymphomas. Mol Cell Biol *17*, 4782-4791.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bell, J.J., and Bhandoola, A. (2008). The earliest thymic progenitors for T cells possess myeloid lineage potential. Nature *452*, 764-767.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., Schreiber, S.L., and Lander, E.S. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet *42*, 1093-1100.

Burns, C.E., Traver, D., Mayhall, E., Shepard, J.L., and Zon, L.I. (2005). Hematopoietic stem cell fate is established by the Notch-Runx pathway. Genes Dev *19*, 2331-2342.

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

Cave, J.W., Loh, F., Surpris, J.W., Xia, L., and Caudy, M.A. (2005). A DNA transcription code for cell-specific gene activation by notch signaling. Curr Biol *15*, 94-104.

Cedar, H., and Bergman, Y. (2011). Epigenetics of haematopoietic cell development. Nat Rev Immunol *11*, 478-488.

Cheng, Y., Wu, W., Kumar, S.A., Yu, D., Deng, W., Tripic, T., King, D.C., Chen, K.B., Zhang, Y., Drautz, D., Giardine, B., Schuster, S.C., Miller, W., Chiaromonte, F., Blobel, G.A., Weiss, M.J., and Hardison, R.C. (2009). Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. Genome Res *19*, 2172-2184.

Chou, S.T., Khandros, E., Bailey, L.C., Nichols, K.E., Vakoc, C.R., Yao, Y., Huang, Z., Crispino, J.D., Hardison, R.C., Blobel, G.A., and Weiss, M.J. (2009). Graded repression of PU.1/Sfpi1 gene transcription by GATA factors regulates hematopoietic cell fate. Blood *114*, 983-994.

Cismasiu, V.B., Adamo, K., Gecewicz, J., Duque, J., Lin, Q., and Avram, D. (2005). BCL11B functionally associates with the NuRD complex in T lymphocytes to repress targeted promoter. Oncogene *24*, 6753-6764.

Cobaleda, C., Jochum, W., and Busslinger, M. (2007a). Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. Nature *449*, 473-477.

Cobaleda, C., Schebesta, A., Delogu, A., and Busslinger, M. (2007b). Pax5: the guardian of B cell identity and function. Nat Immunol *8*, 463-470.

Cui, K., Zang, C., Roh, T.Y., Schones, D.E., Childs, R.W., Peng, W., and Zhao, K. (2009). Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell *4*, 80-93.

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. Nat Immunol *4*, 1029-1036.

Dakic, A., Metcalf, D., Di Rago, L., Mifsud, S., Wu, L., and Nutt, S.L. (2005). PU.1 regulates the commitment of adult hematopoietic progenitors and restricts granulopoiesis. J Exp Med *201*, 1487-1502.

DeKoter, R.P., and Singh, H. (2000). Regulation of B lymphocyte and macrophage development by graded expression of PU.1. Science *288*, 1439-1441.

Dias, S., Mansson, R., Gurbuxani, S., Sigvardsson, M., and Kee, B.L. (2008). E2A proteins promote development of lymphoid-primed multipotent progenitors. Immunity *29*, 217-227.

Durand, C., and Dzierzak, E. (2005). Embryonic beginnings of adult hematopoietic stem cells. haematologica *90*, 100-108.

Dzierzak, E. (1999). Embryonic beginnings of definitive hematopoietic stem cells. Ann N Y Acad Sci *872*, 256-262; discussion 262-254.

Ernst, P., Fisher, J.K., Avery, W., Wade, S., Foy, D., and Korsmeyer, S.J. (2004). Definitive hematopoiesis requires the mixed-lineage leukemia gene. Dev Cell *6*, 437-443.

Feyerabend, T.B., Terszowski, G., Tietz, A., Blum, C., Luche, H., Gossler, A., Gale, N.W., Radtke, F., Fehling, H.J., and Rodewald, H.R. (2009). Deletion of Notch1 converts pro-T cells to dendritic cells and promotes thymic B cells by cell-extrinsic and cell-intrinsic mechanisms. Immunity *30*, 67-79.

Foss, D.L., Donskoy, E., and Goldschneider, I. (2001). The importation of hematogenous precursors by the thymus is a gated phenomenon in normal adult mice. J Exp Med *193*, 365-374.

Franco, C.B., Scripture-Adams, D.D., Proekt, I., Taghon, T., Weiss, A.H., Yui, M.A., Adams, S.L., Diamond, R.A., and Rothenberg, E.V. (2006). Notch/Delta signaling constrains reengineering of pro-T cells by PU.1. Proc Natl Acad Sci U S A *103*, 11993-11998.

Friedman, A.D. (2002). Transcriptional regulation of granulocyte and monocyte development. Oncogene *21*, 3377-3390.

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

Galloway, J.L., Wingert, R.A., Thisse, C., Thisse, B., and Zon, L.I. (2005). Loss of gata1 but not gata2 converts erythropoiesis to myelopoiesis in zebrafish embryos. Dev Cell *8*, 109-116.

Garcia-Bassets, I., Kwon, Y.S., Telese, F., Prefontaine, G.G., Hutt, K.R., Cheng, C.S., Ju, B.G., Ohgi, K.A., Wang, J., Escoubet-Lozach, L., Rose, D.W., Glass, C.K., Fu, X.D., and Rosenfeld, M.G. (2007). Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. Cell *128*, 505-518.

Gekas, C., Dieterlen-Lievre, F., Orkin, S.H., and Mikkola, H.K. (2005). The placenta is a niche for hematopoietic stem cells. Dev Cell *8*, 365-375.

Godin, I., and Cumano, A. (2002). The hare and the tortoise: an embryonic haematopoietic race. Nat Rev Immunol *2*, 593-604.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., Antosiewicz-Bourget, J., Ye, Z., Espinoza, C., Agarwahl, S., Shen, L., Ruotti, V., Wang, W., Stewart, R., Thomson, J.A., Ecker, J.R., and Ren, B. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell *6*, 479-491.

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.

Hsu, C.L., King-Fleischman, A.G., Lai, A.Y., Matsumoto, Y., Weissman, I.L., and Kondo, M. (2006). Antagonistic effect of CCAAT enhancer-binding protein-alpha and Pax5 in myeloid or lymphoid lineage choice in common lymphoid progenitors. Proc Natl Acad Sci U S A *103*, 672-677.

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

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

Ikawa, T., Hirose, S., Masuda, K., Kakugawa, K., Satoh, R., Shibano-Satoh, A., Kominami, R., Katsura, Y., and Kawamoto, H. (2010). An essential developmental checkpoint for production of the T cell lineage. Science *329*, 93-96.

Ikawa, T., Kawamoto, H., Goldrath, A.W., and Murre, C. (2006). E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment. J Exp Med *203*, 1329-1342.

Iwasaki, H., and Akashi, K. (2007). Myeloid lineage commitment from the hematopoietic stem cell. Immunity *26*, 726-740.

Ji, H., Ehrlich, L.I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K., Rossi, D.J., Inlay, M.A., Serwold, T., Karsunky, H., Ho, L.,

Daley, G.Q., Weissman, I.L., and Feinberg, A.P. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature *467*, 338-342.

Jude, C.D., Climer, L., Xu, D., Artinger, E., Fisher, J.K., and Ernst, P. (2007). Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. Cell Stem Cell *1*, 324-337.

Kao, H.Y., Ordentlich, P., Koyano-Nakagawa, N., Tang, Z., Downes, M., Kintner, C.R., Evans, R.M., and Kadesch, T. (1998). A histone deacetylase corepressor complex regulates the Notch signal transduction pathway. Genes Dev *12*, 2269-2277.

Kim, S.I., and Bresnick, E.H. (2007). Transcriptional control of erythropoiesis: emerging mechanisms and principles. Oncogene *26*, 6777-6794.

Koche, R.P., Smith, Z.D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B.E., and Meissner, A. (2011). Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell *8*, 96-105.

Kulessa, H., Frampton, J., and Graf, T. (1995). GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. Genes Dev *9*, 1250-1262.

Kumano, K., Chiba, S., Kunisato, A., Sata, M., Saito, T., Nakagami-Yamaguchi, E., Yamaguchi, T., Masuda, S., Shimizu, K., Takahashi, T., Ogawa, S., Hamada, Y., and Hirai, H. (2003). Notch1 but not Notch2 is essential for generating hematopoietic stem cells from endothelial cells. Immunity *18*, 699-711.

Kurooka, H., and Honjo, T. (2000). Functional interaction between the mouse notch1 intracellular region and histone acetyltransferases PCAF and GCN5. J Biol Chem *275*, 17211-17220.

Lai, A.Y., and Kondo, M. (2006). Asymmetrical lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. J Exp Med *203*, 1867-1873.

Lai, A.Y., and Kondo, M. (2008). T and B lymphocyte differentiation from hematopoietic stem cell. Seminars in Immunology *20*, 207-212

Lai, A.Y., Lin, S.M., and Kondo, M. (2005). Heterogeneity of Flt3-expressing multipotent progenitors in mouse bone marrow. J Immunol *175*, 5016-5023.

Laiosa, C., Stadtfeld, M., Xie, H., de Andres-Aguayo, L., and Graf, T. (2006). Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. Immunity *25*, 731-744.

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.

Li, L., Jothi, R., Cui, K., Lee, J.Y., Cohen, T., Gorivodsky, M., Tzchori, I., Zhao, Y., Hayes, S.M., Bresnick, E.H., Zhao, K., Westphal, H., and Love, P.E. (2011). Nuclear adaptor Ldb1 regulates a transcriptional program essential for the maintenance of hematopoietic stem cells. Nat Immunol *12*, 129-136.

Li, L., Leid, M., and Rothenberg, E.V. (2010a). An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. Science *329*, 89-93.

Li, P., Burke, S., Wang, J., Chen, X., Ortiz, M., Lee, S.C., Lu, D., Campos, L., Goulding, D., Ng, B.L., Dougan, G., Huntly, B., Gottgens, B., Jenkins, N.A., Copeland, N.G., Colucci, F., and Liu, P. (2010b). Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. Science *329*, 85-89.

Lin, Y.C., Jhunjhunwala, S., Benner, C., Heinz, S., Welinder, E., Mansson, R., Sigvardsson, M., Hagman, J., Espinoza, C.A., Dutkowski, J., Ideker, T., Glass, C.K., and Murre, C. (2010). A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. Nat Immunol *11*, 635-643.

Luc, S., Buza-Vidas, N., and Jacobsen, S.E. (2008). Delineating the cellular pathways of hematopoietic lineage commitment. Semin Immunol *20*, 213-220.

Maes, J., Maleszewska, M., Guillemin, C., Pflumio, F., Six, E., Andre-Schmutz, I., Cavazzana-Calvo, M., Charron, D., Francastel, C., and Goodhardt, M. (2008). Lymphoid-affiliated genes are associated with active histone modifications in human hematopoietic stem cells. Blood *112*, 2722-2729.

Mansson, R., Hultquist, A., Luc, S., Yang, L., Anderson, K., Kharazi, S., Al-Hashmi, S., Liuba, K., Thoren, L., Adolfsson, J., Buza-Vidas, N., Qian, H., Soneji, S., Enver, T., Sigvardsson, M., and Jacobsen, S.E. (2007). Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. Immunity *26*, 407-419.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol *6*, 838-849.

Masuda, K., Kakugawa, K., Nakayama, T., Minato, N., Katsura, Y., and Kawamoto, H. (2007). T cell lineage determination precedes the initiation of TCR beta gene rearrangement. J Immunol *179*, 3699-3706.

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

Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T., Buettner, R., and Schule, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437*, 436-439.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E.S., and Bernstein, B.E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Mikkola, H.K., Klintman, J., Yang, H., Hock, H., Schlaeger, T.M., Fujiwara, Y., and Orkin, S.H. (2003). Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. Nature *421*, 547-551.

Miyamoto, T., Iwasaki, H., Reizis, B., Ye, M., Graf, T., Weissman, I.L., and Akashi, K. (2002). Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. Dev Cell *3*, 137-147.

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.

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

Ng, S.Y., Yoshida, T., Zhang, J., and Georgopoulos, K. (2009). Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. Immunity *30*, 493-507.

Nutt, S.L., Heavey, B., Rolink, A.G., and Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature *401*, 556-562.

Nutt, S.L., and Kee, B.L. (2007). The transcriptional regulation of B cell lineage commitment. Immunity *26*, 715-725.

Nutt, S.L., Metcalf, D., D'Amico, A., Polli, M., and Wu, L. (2005). Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors. J Exp Med *201*, 221-231.

Nutt, S.L., Urbanek, P., Rolink, A., and Busslinger, M. (1997). Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B

lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. Genes Dev *11*, 476-491.

Orford, K., Kharchenko, P., Lai, W., Dao, M.C., Worhunsky, D.J., Ferro, A., Janzen, V., Park, P.J., and Scadden, D.T. (2008). Differential H3K4 methylation identifies developmentally poised hematopoietic genes. Dev Cell *14*, 798-809.

Orkin, S.H. (2003). Priming the hematopoietic pump. Immunity *19*, 633-634.

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

Oswald, F., Winkler, M., Cao, Y., Astrahantseff, K., Bourteele, S., Knochel, W., and Borggrefe, T. (2005). RBP-Jkappa/SHARP recruits CtIP/CtBP corepressors to silence Notch target genes. Mol Cell Biol *25*, 10379-10390.

Ottersbach, K., and Dzierzak, E. (2005). The murine placenta contains hematopoietic stem cells within the vascular labyrinth region. Dev Cell *8*, 377-387.

Paul, W.E., and Zhu, J. (2010). How are T(H)2-type immune responses initiated and amplified? Nat Rev Immunol *10*, 225-235.

Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F.W., and Orkin, S.H. (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. Cell *86*, 47-57.

Pui, J.C., Allman, D., Xu, L., DeRocco, S., Karnell, F.G., Bakkour, S., Lee, J.Y., Kadesch, T., Hardy, R.R., Aster, J.C., and Pear, W.S. (1999). Notch1 expression in early lymphopoiesis influences B versus T lineage determination. Immunity *11*, 299-308.

Rekhtman, N., Radparvar, F., Evans, T., and Skoultchi, A.I. (1999). Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. Genes Dev *13*, 1398-1411.

Rhodes, J., Hagen, A., Hsu, K., Deng, M., Liu, T.X., Look, A.T., and Kanki, J.P. (2005). Interplay of pu.1 and gata1 determines myelo-erythroid progenitor cell fate in zebrafish. Dev Cell *8*, 97-108.

Robb, L., Drinkwater, C.C., Metcalf, D., Li, R., Kontgen, F., Nicola, N.A., and Begley, C.G. (1995). Hematopoietic and lung abnormalities in mice with a null mutation of the common beta subunit of the receptors for granulocyte-macrophage colony-stimulating factor and interleukins 3 and 5. Proc Natl Acad Sci U S A *92*, 9565-9569.

Robb, L., Elwood, N.J., Elefanty, A.G., Kontgen, F., Li, R., Barnett, L.D., and Begley, C.G. (1996). The scl gene product is required for the generation of all hematopoietic lineages in the adult mouse. Embo J *15*, 4123-4129.

Rolink, A.G., Nutt, S.L., Melchers, F., and Busslinger, M. (1999). Long-term in vivo reconstitution of T-cell development by Pax5-deficient B-cell progenitors. Nature *401*, 603-606.

Rothenberg, E.V. (2007). Negotiation of the T lineage fate decision by transcription-factor interplay and microenvironmental signals. Immunity *26*, 690-702.

Rothenberg, E.V., Moore, J.E., and Yui, M.A. (2008). Launching the T-cell-lineage developmental programme. Nat Rev Immunol *8*, 9-21.

Rothenberg, E.V., Zhang, J., and Li, L. (2010). Multilayered specification of the T-cell lineage fate. Immunol Rev *238*, 150-168.

Saleque, S., Kim, J., Rooke, H.M., and Orkin, S.H. (2007). Epigenetic regulation of hematopoietic differentiation by Gfi-1 and Gfi-1b is mediated by the cofactors CoREST and LSD1. Mol Cell *27*, 562-572.

Sambandam, A., Maillard, I., Zediak, V.P., Xu, L., Gerstein, R.M., Aster, J.C., Pear, W.S., and Bhandoola, A. (2005). Notch signaling controls the generation and differentiation of early T lineage progenitors. Nat Immunol *6*, 663-670.

Schmitt, T.M., Ciofani, M., Petrie, H.T., and Zuniga-Pflucker, J.C. (2004). Maintenance of T cell specification and differentiation requires recurrent notch receptor-ligand interactions. J Exp Med *200*, 469-479.

Schones, D.E., and Zhao, K. (2008). Genome-wide approaches to studying chromatin modifications. Nat Rev Genet *9*, 179-191.

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.

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

Siatecka, M., and Bieker, J.J. (2011). The multifunctional role of EKLF/KLF1 during erythropoiesis. Blood *25*, 25.

Sitnicka, E., Bryder, D., Theilgaard-Monch, K., Buza-Vidas, N., Adolfsson, J., and Jacobsen, S.E. (2002). Key role of flt3 ligand in regulation of the common lymphoid progenitor but not in maintenance of the hematopoietic stem cell pool. Immunity *17*, 463-472.

Souroullas, G.P., Salmon, J.M., Sablitzky, F., Curtis, D.J., and Goodell, M.A. (2009). Adult hematopoietic stem and progenitor cells require either Lyl1 or Scl for survival. Cell Stem Cell *4*, 180-186.

Spain, L.M., Guerriero, A., Kunjibettu, S., and Scott, E.W. (1999). T cell development in PU.1-deficient mice. J Immunol *163*, 2681-2687.

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.

Starck, J., Cohet, N., Gonnet, C., Sarrazin, S., Doubeikovskaia, Z., Doubeikovski, A., Verger, A., Duterque-Coquillaud, M., and Morle, F. (2003). Functional cross-antagonism between transcription factors FLI-1 and EKLF. Mol Cell Biol *23*, 1390-1402.

Taghon, T., Yui, M.A., Pant, R., Diamond, R.A., and Rothenberg, E.V. (2006). Developmental and molecular characterization of emerging beta- and gammadelta-selected pre-T cells in the adult mouse thymus. Immunity *24*, 53-64.

Taghon, T., Yui, M.A., and Rothenberg, E.V. (2007). Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. Nat Immunol *8*, 845-855.

Taghon, T.N., David, E.S., Zuniga-Pflucker, J.C., and Rothenberg, E.V. (2005). Delayed, asynchronous, and reversible T-lineage specification induced by Notch/Delta signaling. Genes Dev *19*, 965-978.

Ting, C.N., Olson, M.C., Barton, K.P., and Leiden, J.M. (1996). Transcription factor GATA-3 is required for development of the T-cell lineage. Nature *384*, 474-478.

Treiber, T., Mandel, E.M., Pott, S., Gyory, I., Firner, S., Liu, E.T., and Grosschedl, R. (2010). Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. Immunity *32*, 714-725.

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

Tydell, C.C., David-Fung, E.S., Moore, J.E., Rowen, L., Taghon, T., and Rothenberg, E.V. (2007). Molecular dissection of prethymic progenitor entry into the T lymphocyte developmental pathway. J Immunol *179*, 421-438.

Vyas, P., Ault, K., Jackson, C.W., Orkin, S.H., and Shivdasani, R.A. (1999). Consequences of GATA-1 deficiency in megakaryocytes and platelets. Blood *93*, 2867-2875.

Wada, H., Masuda, K., Satoh, R., Kakugawa, K., Ikawa, T., Katsura, Y., and Kawamoto, H. (2008). Adult T-cell progenitors retain myeloid potential. Nature *452*, 768-772.

Wakabayashi, Y., Watanabe, H., Inoue, J., Takeda, N., Sakata, J., Mishima, Y., Hitomi, J., Yamamoto, T., Utsuyama, M., Niwa, O., Aizawa, S., and Kominami, R. (2003). Bcl11b is required for differentiation and survival of alphabeta T lymphocytes. Nat Immunol *4*, 533-539.

Wallberg, A.E., Pedersen, K., Lendahl, U., and Roeder, R.G. (2002). p300 and PCAF act cooperatively to mediate transcriptional activation from chromatin templates by notch intracellular domains in vitro. Mol Cell Biol *22*, 7812-7819.

Wang, J., Scully, K., Zhu, X., Cai, L., Zhang, J., Prefontaine, G.G., Krones, A., Ohgi, K.A., Zhu, P., Garcia-Bassets, I., Liu, F., Taylor, H., Lozach, J., Jayes, F.L., Korach, K.S., Glass, C.K., Fu, X.D., and Rosenfeld, M.G. (2007). Opposing LSD1 complexes function in developmental gene activation and repression programmes. Nature *446*, 882-887.

Weber, B.N., Chi, A.W., Chavez, A., Yashiro-Ohtani, Y., Yang, Q., Shestova, O., and Bhandoola, A. (2011). A critical role for TCF-1 in T-lineage specification and differentiation. Nature *476*, 63-68.

Wei, G., Abraham, B.J., Yagi, R., Jothi, R., Cui, K., Sharma, S., Narlikar, L., Northrup, D.L., Tang, Q., Paul, W.E., Zhu, J., and Zhao, K. (2011). Genome-wide Analyses of Transcription Factor GATA3-Mediated Gene Regulation in Distinct T Cell Types. Immunity *35*, 299-311.

Wei, G., Wei, L., Zhu, J., Zang, C., Hu-Li, J., Yao, Z., Cui, K., Kanno, Y., Roh, T.Y., Watford, W.T., Schones, D.E., Peng, W., Sun, H.W., Paul, W.E., O'Shea, J.J., and Zhao, K. (2009). Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity *30*, 155-167.

Weishaupt, H., Sigvardsson, M., and Attema, J.L. (2010). Epigenetic chromatin states uniquely define the developmental plasticity of murine hematopoietic stem cells. Blood *115*, 247-256.

Weiss, M.J., and Orkin, S.H. (1995). Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. Proc Natl Acad Sci U S A *92*, 9623-9627.

Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663-676.

Yamada, Y., Warren, A.J., Dobson, C., Forster, A., Pannell, R., and Rabbitts, T.H. (1998). The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. Proc Natl Acad Sci U S A *95*, 3890-3895.

Yashiro-Ohtani, Y., Ohtani, T., and Pear, W.S. (2010). Notch regulation of early thymocyte development. Semin Immunol *22*, 261-269.

Ye, M., Iwasaki, H., Laiosa, C.V., Stadtfeld, M., Xie, H., Heck, S., Clausen, B., Akashi, K., and Graf, T. (2003). Hematopoietic stem cells expressing the myeloid lysozyme gene retain long-term, multilineage repopulation potential. Immunity *19*, 689-699.

Yoshida, T., Ng, S.Y., and Georgopoulos, K. (2010). Awakening lineage potential by Ikaros-mediated transcriptional priming. Curr Opin Immunol *22*, 154-160.

Yoshida, T., Ng, S.Y., Zuniga-Pflucker, J.C., and Georgopoulos, K. (2006). Early hematopoietic lineage restrictions directed by Ikaros. Nat Immunol *7*, 382-391.

Yu, M., Riva, L., Xie, H., Schindler, Y., Moran, T.B., Cheng, Y., Yu, D., Hardison, R., Weiss, M.J., Orkin, S.H., Bernstein, B.E., Fraenkel, E., and Cantor, A.B. (2009). Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. Mol Cell *36*, 682-695.

Yui, M.A., Feng, N., and Rothenberg, E.V. (2010). Fine-scale staging of T cell lineage commitment in adult mouse thymus. J Immunol *185*, 284-293.

Zhang, D.E., Zhang, P., Wang, N.D., Hetherington, C.J., Darlington, G.J., and Tenen, D.G. (1997). Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. Proc Natl Acad Sci U S A *94*, 569-574.

Zhang, P., Behre, G., Pan, J., Iwama, A., Wara-Aswapati, N., Radomska, H.S., Auron, P.E., Tenen, D.G., and Sun, Z. (1999). Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. Proc Natl Acad Sci U S A *96*, 8705-8710.

Zhang, P., Iwasaki-Arai, J., Iwasaki, H., Fenyus, M.L., Dayaram, T., Owens, B.M., Shigematsu, H., Levantini, E., Huettner, C.S., Lekstrom-Himes, J.A., Akashi, K., and Tenen, D.G. (2004). Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. Immunity *21*, 853-863.

Zhang, P., Zhang, X., Iwama, A., Yu, C., Smith, K.A., Mueller, B.U., Narravula, S., Torbett, B.E., Orkin, S.H., and Tenen, D.G. (2000). PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. Blood *96*, 2641-2648.

Zlotoff, D.A., and Bhandoola, A. (2011). Hematopoietic progenitor migration to the adult thymus. Ann N Y Acad Sci *1217*, 122-138.

Chapter 2

Dynamic Transformations of Epigenetic Marking and Genome-wide Transcriptional Regulation Establish T-cell Identity

Jingli A. Zhang, Ali Mortazavi, Brian A. Williams, Barbara J. Wold, and Ellen V. Rothenberg

Abstract

T-cell development comprises a stepwise process of commitment from a multipotent precursor. To define molecular mechanisms controlling this progression, we probed five stages spanning the commitment process using RNA-seq and ChIP-seq to track genome-wide shifts in transcription, cohorts of active transcription factor genes, histone modifications at diverse classes of cis-regulatory elements, and binding repertoire of GATA-3 and PU.1, transcription factors with complementary roles in T-cell development. The results highlight potential promoter-distal cis-regulatory elements in play and reveal both activation sites and diverse mechanisms of repression that silence genes used in alternative lineages. Histone marking is dynamic and reversible, and while permissive marks anticipate, repressive marks often lag behind changes in transcription. *In vivo* binding of PU.1 and GATA-3 relative to epigenetic marking reveals distinctive, factor-specific rules for recruitment of these crucial transcription factors to different subsets of their potential sites, dependent on dose and developmental context.

INTRODUCTION

T-lymphocyte development illuminates the sequential process of cell fate choice for descendants of multipotent stem cells. Notch pathway signaling in the thymus causes hematopoietic precursors to become committed to the T-cell fate, while mobilizing a T-cell gene expression program that prepares the cells for T-cell antigen receptor (TCR) expression, TCR-based repertoire selection, and long, versatile careers as immune effectors. Sequential events that exclude alternative lineages occur at phenotypically well-defined stages within the thymus, providing a revealing model for the kinds of events needed to channel multipotent stem cells into a single developmental path (Rothenberg, 2011; Yang et al., 2010). However, major questions about the molecular mechanisms involved in this process have remained.

One question is how commitment works. Most of the alternatives to T-cell differentiation that are initially open to the precursors depend on the action of dominant transcriptional regulators, such as PU.1 (and C/EBPα) for myeloid and dendritic cells, and EBF1 and Pax5 for B cells. Although these genes are either expressed or inducible in the precursors entering the thymus, they end up not only repressed but irreversibly silenced as a result of commitment. The mechanisms responsible for these regulatory changes have been unknown.

Another question has been how the T-cell program is deployed. Notch signaling, triggered by ligands in the thymic microenvironment, is crucial to initiate and sustain differentiation. T-cell development also depends on additional transcription factors, including basic helix-loop-helix E protein dimers (E2A and HEB), the high mobility-group factors TCF-1 and LEF-1, GATA-3, Myb, Runx1, Ikaros, and Gfi1 (Rothenberg et

al., 2008). However, it is not clear if these are all the factors involved. Exactly what their roles are in establishing T-cell identity remains particularly obscure because so few T-cell specific cis-regulatory elements have been identified. Consequently, almost none have been functionally dissected in enough detail to explain the expression of the genes they control in terms of their specific and complete regulatory inputs.

For other hematopoietic cell types where known lineage-specific factors positively confer cell identity, key cis-regulatory sequences of developmental genes have been identified through the collaborative binding of these factors. For example, combined binding sites of E2A, EBF1, and/or Pax5 are target cis-regulatory elements used in the B-cell program (Lin et al., 2010; Schebesta et al., 2007). PU.1 binding, with various partner factors, has been a useful probe to detect myeloid cis-elements (Ghisletti et al., 2010; Heinz et al., 2010). In contrast, for T cells, nearly all the transcription factors known to be required are also required by other hematopoietic lineages (Rothenberg and Scripture-Adams, 2008). Thus, no formula known *a priori* is useful to define T-lineage specific cis-regulatory elements. However, it is possible to work in the opposite direction. If the cis-regulatory elements that are "in play" at crucial transitions of T-cell development can be defined globally, then the motifs enriched in these elements could be matched with the cognate transcription factors that also change at those stages, thus narrowing the search for the key factors controlling commitment (Novershtern et al., 2011).

Here we identify the dynamic transformations in transcription and epigenetic marking that occur across the genome through five successive stages of T-cell differentiation that span lineage commitment. The results provide a genome-wide view of a lineage choice process in unusually fine resolution. To test the functional relevance of

the histone marking patterns at potential proximal and distal cis-regulatory elements, we also track *in vivo* binding of GATA-3 and PU.1, two transcription factors previously shown to have complementary, dose-dependent and context dependent roles in these stages of T-cell development (Rothenberg and Scripture-Adams, 2008). Site selection rules for these two factors are revealed to be context-dependent but differently affected by dose. The results provide evidence for an actively sustained regulatory phase dominated by stem/ progenitor-cell regulatory genes that initially persists under Notch signaling, then is dismantled to establish T-cell identity.

RESULTS

Strategy for Analysis of T-lineage Commitment

Our goals were first, to map comprehensively the genes that undergo transcriptional change during T-lineage choice, especially genes encoding transcription factors; and second, to locate likely cis-regulatory sites mediating these gene expression changes by defining regions where histone marks are altered at each step of the process. RNA analysis by deep sequencing (RNA-seq) was used to measure changes in global transcriptional activity (Mortazavi et al., 2008), while chromatin immune precipitation and deep sequencing (ChIP-seq) monitored the changing distributions of histone marks across the genome in parallel (Wold and Myers, 2008).

In mice, the earliest stages of T-cell differentiation represent only a minuscule fraction of thymocytes present at steady state (Petrie and Zuniga-Pflucker, 2007). Cells from the first major stage, “early T-cell precursors” or Kit^{high} DN1 cells (Kit^{++} $CD44^{+}$

CD25$^-$ and CD4$^-$ CD8$^-$ TCR$^-$"double negative"), proliferate as they pass through the "DN2a" stage (Kit^{++} CD44$^+$ CD25$^+$) to the "DN2b" stage (Kit$^+$ CD44$^+$ CD25$^+$), when they undergo T-lineage commitment. Post-commitment, they accumulate in the "DN3" stage (Kit^{low} CD44^{low} CD25$^+$), and slow their cycling to allow rearrangement of the TCR genes and create a checkpoint for TCR rearrangement success. Only cells that successfully express TCR proteins proliferate again, differentiating to the "DP" stage (CD25$^-$ CD4$^+$ CD8$^+$). Cells are selected after this based on their TCR recognition specificity, and further differentiation refines their mature immunological roles.

However, all pre-commitment cells together represent only 0.1% of the cells in a thymus. To obtain large numbers of the earliest cells, we therefore used an *in vitro* differentiation system that generates copious yields of early T-cell precursors from fetal liver (FL)-derived hematopoietic progenitor cells. These precursors are co-cultured with lymphoid-permissive cytokines and OP9 stromal cells expressing a Delta-like Notch ligand (OP9-DL1). In these conditions a cohort of Kit^{high} DN1 (FLDN1) and FLDN2a cells is generated by day 4.5 of culture, mostly progressing to DN2b cells by day 8.5 (FLDN2b, Figure S1A). As an *in vivo* counterpart, we also purified slightly more advanced DN3 stage cells from freshly isolated adult mouse thymus, and to evaluate changes that follow after commitment, DP thymocytes were also purified (ThyDN3, ThyDP) (Figure S1B, C, see Supplemental Methods). *In vitro* differentiated FL-derived DN1 and DN2 cells showed gene expression well matched to that of normal *in vivo* thymocyte counterparts, based on previous analyses of adult thymocyte gene expression (Yui et al., 2010; David-Fung et al., 2009; Kawazu et al., 2007; Tabrizifard et al., 2004) (http://www.immgen.org (Heng et al., 2008)). Their lineage commitment status was also

in good agreement with that of *in vivo* counterparts (Table S1). FLDN1 cells after 4.5 days of culture with OP9-DL1 cells could still generate NK, myeloid or dendritic cells, and even some B cells if shifted to non-T conditions (OP9-control co-culture), whereas FLDN2a cells mostly generated NK cells under these conditions and FLDN2b cells could not generate any alternative cell types (Table S1). Some details of lineage exclusion timing were slightly shifted from those in corresponding steady-state adult thymocytes (Rothenberg, 2011) (see Table S1 legend), but the discontinuity between these DN2a and DN2b subsets was as clear as for populations from normal adult thymus (Yui et al., 2010).

Global Gene expression Analysis: Selective Changes during Early T-cell Development

To identify when major changes in gene expression occurred along the pathway from early T-cell precursor to DP stages, RNA-seq was performed on 2~3 independent biological replicates each of FLDN1, FLDN2a, FLDN2b, ThyDN3, and ThyDP cells. At least 10^7 mappable 32 bp reads were obtained for each sample, and independent biological replicates were well correlated ($r > 0.97$) (Figure S1D). Using the programs ERANGE and DEGseq to compare expression in different stages, ~10,000 of the 20,861 Refseq annotated genes were detectably expressed (RNA-seq > 1 RPKM) in each population; of these, ~50% statistically significantly changed in expression ($p < 0.001$) between at least one pair of stages and ~ 40% changed overall from FLDN1 to ThyDP (Figure 1A).

Figure 1B shows the hierarchical clustering of the expression patterns of the 3,697 genes that change expression at least twofold between any stages. Genes undergoing

induction between DN1 and DN2b included key T-cell specific genes involved in pre-TCR expression and function, i.e., genes encoding TCR complex components *Cd3g*, *Cd3d*, *Cd3e*, *Cd3z* (*Cd247*), T-cell specific signaling components *Itk* and *Lat*, recombinase *Rag1*, mutagenic DNA polymerase *Dntt*, and the surrogate a chain (pTα) *Ptcra*. Transcripts from all of these increased strongly from a low or undetectable background level in the FLDN1 stage (Table S2). In addition, a conspicuous group of genes were repressed or silenced during these transitions. They included the progenitor-cell specific growth factor receptor genes *Kit*, *Flt3*, and *Csf2rb,* and also a notable cluster of transcription factor genes described below.

The DN1 to DN2 transition is the first definitive sign of T-lineage entry in response to Notch signaling. However, a much larger difference was seen between pre-commitment FLDN2a and post-commitment FLDN2b cells (2,429 genes different, Figure 1A, B) than between FLDN1 and FLDN2a (<900 genes different), as supported by hierarchical clustering. Conversely, despite their different origins and manipulation, the newly-committed FLDN2b and ThyDN3 populations were more similar to each other as well (Figure 1A, B). Thus, the major genome-wide transcriptomic changes leading to T-lineage identity do not occur in the DN1 to DN2a transition. Instead, the cells make a larger break via transition into the DN2b or DN3 stages, associated with commitment.

Transcription Factor Expression Dynamics in T-lineage Commitment

Genes likely to encode transcriptional regulators (Table 3A, see Supplemental Methods) accounted for 379 of the genes that changed expression ≥2x between stages (Table S3B). Hierarchical clustering of their patterns of expression (Figure 1C) confirmed the FLDN1 pattern to resemble the FLDN2a pattern, while the FLDN2b

resembled the ThyDN3 expression pattern, much more than the pre-commitment FLDN2a cells resembled the newly committed FLDN2b. Another major discontinuity occurred between the DN2b or DN3 cells and the ThyDP cells (Figure 1C). Thus, the major transitions in regulatory gene expression occur at commitment and during the TCR-dependent selection event (" β-selection") that separates DN3 cells from DP cells.

From FLDN1 to FDN2b, the most strongly upregulated "regulatory" loci in the whole genome were found to be *Lef1* and *Bcl11b* (>75x increased). *Pou6f1, SpiB, Ikzf3,* and *Ets1* among others also increased >8x, with weaker increases for known T-cell regulators *Id3, Tcf12, Gfi1, Tcf7, Hes1*, *Gata3* (Table S3B). However, many regulatory genes sharply decreased in expression between FLDN1 and FLDN2b, including genes with known, important functions in hematopoietic progenitors, e.g., *Gfi1b, Lmo2, Mef2c, Hoxa9, Sfpi1 (PU.1), Gata2, Mycn (NMyc), Cebpb, Bcl11a, Hhex, Nfe2, Lyl1,* and several *Irf* factors. A major regulatory shift, with broad repression of progenitor-cell transcription factor genes, thus accompanies T-lineage commitment.

Dynamic Histone Modification Changes Identify Developmentally Regulated Promoters and Distal Cis-elements

The specific cis-regulatory elements affected by changing transcription factor action during commitment should be sites of developmentally changing histone modifications (Natoli, 2010; Kouzarides, 2007). These are of greatest interest where implicated in changes in expression of linked genes (Koche et al., 2011; Weishaupt et al., 2010; Hawkins et al., 2010; He et al., 2010; Araki et al., 2009). To detect sites of positive and negative regulation with high sequence resolution we used ChIP-seq (Johnson et al., 2007) to enrich DNA associated with three H3 modifications: H3K(9,14)Ac, H3K4me2,

and H3K27me3. Histone H3K(9, 14) acetylation (H3Ac) is functionally linked to activation at transcriptional start sites (TSS), H3K27me3 is used in one mechanism for transcriptional silencing, and H3K4me2 is associated with activation, poising for activation or repression, or repression (Wang et al., 2008; Orford et al., 2008; Barski et al., 2007; Johnson et al., 2007; Heintzman et al., 2007). For many enhancers, H3K4me2 provides more precise localization than H3K4me1 (Koche et al., 2011). These marks were monitored in two independent biological replicates of each cell type except FLDN2a (one sample), with excellent correlation between replicates (Figure S1E).

The relationships among these marks and RNA expression generally agreed with previous studies. Nearly 42,000 "regions" (≥200 bp, Figure S2A) were enriched for marks in at least one stage (see Supplemental Methods), of which ~35% were within 1 kb of a TSS ("proximal") and the others in intergenic regions or within transcription units ("distal"). The marks were distributed at annotated TSS and non-TSS sites of expressed and silent genes as shown in Figure 2A and Figure S2B. Tables S2A (TSS) and S2B (non-TSS) summarize modifications at all marked genomic elements at each stage as correlated with RNA expression of the nearest genes. Figure S3 presents these comprehensive results for TSS and non-TSS sites at all Refseq loci in hierarchically clustered heat maps.

At any given stage, the ~10,000 genes that were detectably expressed were overwhelmingly distinguished by H3Ac (85 ± 2%) and H3K4me2 (>91%) at their annotated promoters. Silent genes fell into two classes, ~35% with H3K27me3 at their promoters and many with unmarked promoters. However, >25% of the silent genes, including many with H3K27me3, were still marked with H3K4me2 (Figures S2C, S3),

suggesting at least three kinds of repressed states (Filion et al., 2010). Interestingly, silent genes encoding regulatory factors were more likely than silent genes overall to be marked with H3K27me3 (Figure 2A, B right panels).

The cumulative probability plots in Figure 2C show that of all genes, those with H3K4me2 but not H3Ac at their promoters in a given stage (red tracks), with or without H3K27me3, were most likely to be newly repressed or poised for incipient transcriptional activation in the next stage. Thus, H3K4me2 alone or bivalent with H3K27me3 marked the most developmentally labile promoters in these cells as described elsewhere (Koche et al., 2011; Orford et al., 2008). Histone modification at TSS sites was relatively stable across development, more than levels of corresponding RNAs (Figure S3A). However, distal elements were more dynamically marked: >1/3 of all regions marked with H3K4me2 or H3K27me3 in one stage lacked those marks in at least one other stage (Figure S3B). Thus the regulatory shifts occurring in development most sensitively affect histone marking at non-promoter elements (Heinz et al., 2010; Lin et al., 2010). H3K4me2-marked distal elements in fact included a number of previously noted regulatory elements (Figure S4A, B): the DP-specific *Rag1/2* gene antisilencer 71-75 kb 5' of the *Rag2* gene (Yannoutsos et al., 2004), and the DN2b/3-specific E1a promoter for the *Notch1* gene (Gomezdel Arco et al., 2010). Both of these discrete cis-elements acquired H3K4me2 specifically at stages when they contribute to gene regulation (Figure S4A, B). Other non-TSS regions with developmentally dynamic marking may thus locate stage-specific cis-regulatory elements as well.

Timing of TSS Epigenetic Changes Relative to Transcriptional Changes

To relate the timing of changes in TSS marks with changes in RNA expression

during T-cell commitment, we focused on 3,697 differentially regulated genes. First, these were subdivided by K-means clustering into 25 clusters based on expression pattern (Figure S5; genes listed in Table S4A). Figure 2D tracks the histone marks from stage to stage at the TSS's of genes undergoing upregulation (clusters 1, 2, and 6), downregulation (clusters 7, 9 and 23), and transient decreases (cluster 12) or increases in expression (clusters 17 and 19). H3Ac (Figure 2D, first group of columns) was tightly coordinated with transcriptional activity (last columns), but H3K4me2 was often present before and after expression (second columns, e.g., clusters 1 & 6). Though H3K27me3 was inversely correlated with expression, only a fraction of the repressed genes ever acquired this mark (Figure 2D, third columns). These patterns were confirmed at promoters genome-wide (Figure S3A).

Most relevant to the regulatory decisions that underlie T-lineage commitment (Rothenberg, 2011; Yang et al., 2010) are the effects on genes involved in diverse hematopoietic cell fates, options that are foreclosed in an ordered sequence during T-cell commitment. We identified 389 key hematopoietic genes by Gene Ontology (Table S4B), including "signature" regulators of erythroid cells (*Gata1, Nfe2, Epor*), myeloid cells (*Sfpi1, Cebpa, Cebpe, Csf1r*), B cells (*Pax5, Ebf1*), NK cells (*Eomes, Il2rb*), and stem cells (*Gata2, Tal1, Lmo2*). The expression levels of these genes were tracked in T-cell precursors ("DN1"-"DP"), in parallel with the changing status of the three histone marks at their promoters (Figure 3). Full results are reported in Table S4B and shown in the Figure 3 master panel, while labeled subgroups of distinctive expression types are also shown in the zoom-in panels so that individual genes can be identified. Again, H3Ac modification at promoters was tightly correlated with transcription, while H3K4me2

marking also preceded and persisted after transcription.

The central importance of alternative-lineage exclusion in T-cell commitment raised the question of whether common or diverse mechanisms of silencing of non-T regulatory genes were used. H3K27me3 use at these functionally relevant loci was both variable and dynamic. Some genes were kept silent throughout T-cell specification, and many had strong H3K27me3 marks at the promoter, some apparently together with H3K4me2 (e.g., Group e, *Epor, Irf4, Ebf1*, and *Eomes*) and others without (e.g., Group e, Pax5). Other genes were turned off during development, and while these lost H3Ac they often gained H3K27me3 (Group d). Some genes poised for early repression already had some H3K27me3 at the TSS from FLDN1 stage (e.g., *Cebpa* in Group e*, Gata2, Lmo1, Tal1* in Group d), suggesting repression already underway in at least part of the population. Yet H3K27me3 did not mandate future silencing, for some T-cell genes like *Lef1* were strongly activated during commitment despite initially strong H3K27me3 marking (Figure 3, Group b). Furthermore, other genes stayed silent from FLDN1 to ThyDP without any H3K27me3 at the TSS (e.g., *Cebpe, Cx3cr1, Zbtb32, Cd79a,* and *VpreB1;* Group c). Unexpectedly, these variations in H3K27me3 marking cut across myeloid, erythroid, NK cell, and B-cell program boundaries.

Most Epigenetic Change of Loci Affected by T-cell Development Occurs between DN1 and DP

Both the foreshadowing of future expression by H3K4me2 marking of promoters and the ability of some genes to be repressed without appearance of H3K27me3 (Figures 2D, 3) raised the question of whether critical changes in promoter status occurred during the developmental transitions to DN1, or after DP stage. We therefore compared our

results with H3Ac, H3K4me2, and H3K27me3 ChIP-seq data for a prethymic lymphoid precursor population, "PPB" ($EBF^{-/-}$ pre-pro B cells (Lin et al., 2010; Heinz et al., 2010)) and with H3K27me3 data for post-thymic naïve CD4 T cells, "CD4" (Wei et al., 2009), shown in flanking columns in Figures 2D and 3. This comparison showed that the prethymic lymphoid precursor data in general concurred well with the FLDN1 patterns for all three histone marks. Furthermore, repressed genes that lacked H3K27me3 marks by the DP stage in our samples also remained silent without H3K27me3 marks in the peripheral T cells (the uniquely regulated *Rag* genes were an exception). The FLDN1 to ThyDP interval thus encompasses the crucial epigenetic changes for the great majority of genes affected by T-cell specification.

Distinct Mechanisms for Control of Key Developmental Genes

Changes in modification at distal sites (compiled in Table S2B) as well as TSS sites (Table S2A) often appeared implicated in gene regulation, as shown for the key genes highlighted in Figure 4.

Figure 4A, B illustrate two highly T-cell-specific loci activated in parallel from DN2a to DN2b, the *Cd3gde* gene cluster and *Bcl11b,* respectively. These genes initially lack RNA transcripts (Figure 4, black tracks) and H3Ac marks (blue tracks) in the FLDN1 cells, but then are strongly upregulated and kept on thereafter. For the *Cd3* genes, there was no H3K4me2 (red tracks) at the promoters and light H3K27me3 marking across the locus (Figure 4A, green tracks) during the initial silence, but the classic enhancer elements at the 3' ends of *Cd3e* and *Cd3d* (Georgopoulos et al., 1988; van de Wetering et al., 1991) were already marked by focal H3K4me2. Note that these enhancers were already accessible to transcription factor binding even in the FLDN1

stage, as shown by ChIP-seq evidence for binding by the factor GATA-3 (Figure S4C; see below). These H3K4me2 sites then intensified while both H3Ac and H3K4me2 were recruited to the promoters of the genes during the DN2a/2b stages, when transcription began. A similar pattern for early activation without initial promoter marking was seen for *Il2ra* (Figure S4D).

In contrast to the *Cd3gde* cluster, the *Bcl11b* gene (Figure 4B) began with substantial H3K27me3 (green tracks) over its promoter and across the whole gene body at FLDN1 stage. However, its TSS also had a cryptic positive cis-regulatory element marked by H3K4me2. *Bcl11b* then was activated from FLDN2a to FLDN2b stage through a process that swept back the H3K27me3 repressive marks off the promoter, while expanding the H3K4me2 marks into the first intron and creating a new H3K4me2 marked region in the third intron.

The changes in histone marks at these loci contrast sharply with the precisely positioned but virtually unchanging H3K27me3, H3Ac, and H3K4me2 marks that characterized the *Gata3* gene (Figure 4C). Despite a block of H3K27me3 close to the major promoter, this gene was already activated by the time of the FLDN1 stage and underwent only a few-fold increase in expression after that.

Repression of essential B-cell regulatory factors, myeloid-cell regulatory factors, and stem or progenitor-cell regulatory factors is a crucial aspect of T-lineage commitment. This clearly required a variety of distinct mechanisms (Figure 4D-H). The *Pax5* gene, crucial for the B-cell program, had no H3Ac modified regions at any stages (Figure 4D; compared with neighboring Zcchc7 promoter mark). Four small peaks of H3K4me2 modification were seen in intronic regions, one of them corresponding to a

known hematopoietic enhancer (Decker et al., 2009). However, the gene was completely buried in H3K27me3 at all stages. *Ebf1* (Figure S4E) was also repressed from FLDN1 on despite H3K4me2 at several sites.

Hhex and *Bcl11a* (Figure 4E, F), in contrast, were expressed strongly in FLDN1 cells but then downregulated sharply by the FLDN2b/ThyDN3 stages, showing evidence of distinct modulating roles for distal and TSS elements. For *Hhex* (Figure 4E), the TSS and two H3K4me2-marked distal regions lost activation marks as expression decreased, while H3K27me3 appeared focally at the TSS and then spread. A similar pattern was seen for *Flt3* and the *Zbtb7b* (Thpok) gene, which were active in FLDN1 and then repressed (Figure S4F, I). For *Bcl11a* (Figure 4F), H3Ac persisted at the promoter while RNA expression declined during commitment, reflecting a tail of low-level expression through DN3. Here, H3K27me3 marks only appeared at the last stage of silencing at the DP stage. However, the H3K4me2 modification of an element just downstream of the last exon decreased sharply between FLDN2a and FLDN2b, in parallel with RNA expression, suggesting a potential regulatory role.

The myeloid and progenitor-cell transcription factor gene *Sfpi1* (encoding PU.1), silenced in parallel with *Hhex* and never re-expressed in most T-cell lineages, used a different mechanism of repression (Figure 4G). H3Ac disappeared from the promoter while H3K4me2 marks in the upstream cis-regulatory elements of the gene (Hoogenkamp et al., 2007; Rosenbauer et al., 2006; Zarnegar et al., 2010) became narrowed as transcription declined (Figure 4G). Yet minimal H3K27me3 was deposited.

Not only were H3K27me3 marks dispensable for repression; they were also labile. Figure 4H shows that dense H3K27me3 marks on *Mpzl2* (same as *Eva1*)

diminished during transient induction of RNA expression in the DN2b and DN3 stages, but then returned during re-silencing in DP stage. Conversely, despite potent silencing, *Zbtb7b* (Figure S4I) is later reactivated for $CD4^+$ cell positive selection.

Early T-cell Specific Sites for PU.1: A Positive Role

The functional interplay between histone modifications and transcription factors could be probed by monitoring the binding patterns of known transcription factors, to evaluate site accessibility and the order of transcription factor binding relative to histone marking. We focused on two well-studied factors with contrasting roles, GATA-3 and PU.1 (encoded by *Sfpi1*), which are both needed for early T-cell development, playing dose-dependent, stage-specific roles in collaboration with Notch signals (Rothenberg and Scripture-Adams, 2008; Hosoya et al., 2010). PU.1 is a key representative of the progenitor-associated cohort of transcription factors in early pro-T cells. As PU.1 is even more critical for B, dendritic and myeloid cell development, an issue is whether it has distinct T-lineage target genes or simply delays or interferes with commitment.

PU.1 bound to about 34,000 sites in double negative T cells, comparable to B and myeloid cells (Heinz et al., 2010). Although PU.1 RNA and protein levels decline sharply during T lineage commitment (Figure 5A) (Yui et al., 2010), PU.1 site binding preferences were highly consistent from stage to stage. We compared FLDN1 and FLDN2a cells; FLDN2b cells, where PU.1 is at least 4-5x downregulated; and DP cells, where PU.1 is absent. Although PU.1 binding intensity per site was reduced 4-5x in the FLDN2b cells, its site choices remained highly correlated with those in the earlier stages ($r = 0.65$-0.66) (Figure 5B). Thus, within T-cell precursors, a major determinant of PU.1 binding at most sites may simply be PU.1's own availability.

Nevertheless, the PU.1 binding sites in FLDN1 and FLDN2a cells were not the same as those reported for PU.1 binding in B cells, macrophages or $E2A^{-/-}$ pre-pro B cells (representing prethymic lymphoid progenitors (Heinz et al., 2010)) (Figure 5C). Although most closely related to the sites bound in multipotent $E2A^{-/-}$ pre-pro B cells overall, some known PU.1 target sites bound by PU.1 in pre-pro B cells were not accessible to PU.1 binding in the FLDN1 cells, e.g., the intronic enhancer of *Pax5* (Figure 5D). *De novo* motif analysis showed that PU.1 target sites in FLDN1 cells also included a different hierarchy of preferred sequences than in the pre-pro B cells (Figure 5E). Thus, the consistent site choices of PU.1 from DN1 stage through commitment include a distinct T lineage-specific component.

PU.1 is required to generate T-cell precursors, but at high levels it inhibits expression of many T-cell specific genes, particularly when Notch signaling is interrupted (Franco et al., 2006). To test whether the special T-lineage specific roles of endogenous PU.1 may be to delay T-cell gene activation, we asked whether the PU.1 sites specific to early T cells were most often linked to genes that are active or repressed, as compared with sites that were only bound by PU.1 in non-T, $E2A^{-/-}$ pre-pro B cells. Results showed that sites actually bound by PU.1 in FLDN1 cells, including FLDN1-specific sites, were broadly associated with "positive" marks (H3Ac and/or H3K4me2) and completely uncorrelated with H3K27me3 marking, unlike sites bound by PU.1 only in $E2A^{-/-}$ pre-pro B cells (Figure 5F). In the aggregate, genes with sites actually bound by PU.1 in FLDN1 cells were also much more likely to show strong expression than those with potential PU.1 sites unoccupied, with higher expression the more sites bound (Figure 5G). This was true of FLDN1-specific sites as well as of sites occupied in both

FLDN1 and pre-pro B cells. Thus, PU.1 binding globally correlates with positive target gene regulation in FLDN1 cells.

PU.1 Binding Dynamics, Histone Marking, and Temporal Control of Target Gene Expression

Experimental perturbation analyses have shown many specific genes in pro-T cells that are activated or repressed by manipulations of PU.1 level (Franco et al., 2006) (A. Champhekar, M. M. Del Real, and E. V. R., unpublished data). However, with so many binding sites for PU.1, binding alone clearly could not define genes that depend on PU.1 for positive or negative regulation. Most PU.1 binding sites in fact were linked to genes expressed stably in all stages whether PU.1 is present or not, like the majority of genes expressed in T-cell development overall. PU.1 may thus be recruited to sites at many active genes where it has no required role. The challenge was to filter the binding sites identified genome-wide to enrich for functionality, to consider properties of likely functional sites in global terms.

Whereas PU.1 sites linked to stably expressed genes may include many opportunistic “passengers”, functionally important PU.1 sites should be enriched near genes which themselves change in RNA expression, up or down, during development as PU.1 binding to their regulatory sites declines. To assess broadly whether PU.1 participates in positive or negative regulation at a given set of loci, we compared changes in local PU.1 occupancy from FLDN1 to FLDN2b with the direction and magnitude of changes in RNA expression of the genes linked to those sites.

First, we considered a pool of all genes up or downregulated from DN1 to DN2b (Figure 6B) and reclassified them purely according to whether their linked sites lost PU.1

occupancy faster (Figure 6B-C, blue) or slower (red) than the global ~4x average (green = genes with both kinds of sites). We then assessed whether these changes in PU.1 occupancy predicted the direction of changes in mRNA expression across the same interval (Figure 6C), i.e., whether early PU.1 loss was more linked to genes turning off like PU.1 itself, or with upregulating genes that PU.1 might initially have repressed. These cumulative probability plots could reveal either group-wide trends or subgroups with different expression trends within each group. Second, we classified individual PU.1 sites according to whether their linked genes were upregulated, downregulated, stably expressed, or silent across the DN1 to DN2b interval. We then asked whether expression categorized sites tended to lose PU.1 faster or slower than sites linked to stably expressed genes, a standard for likely full epigenetic accessibility (Figure S6A).

The results implied that PU.1 binding is rarely if ever directly repressive, whether the genes are expressed in the same pattern as PU.1 or not. Figure 6C shows that genes which lost PU.1 occupancy most rapidly (blue curve) were more likely downregulated ($\log_2$(FLDN2b/FLDN1) < -1) and less upregulated than those with mixed sites. Conversely, almost 80% of genes with sites that retained PU.1 best (red curve) increased their expression from FLDN1 to FLDN2b. In accord, the changes in PU.1 occupancy at individual binding sites were significantly different for downregulated, upregulated, and stably expressed genes (Figure S6A), such that sites linked to the upregulated genes retained PU.1 even better than fully “accessible”, stably expressed ones.

PU.1 binding can recruit histone methyltransferases and create locally “accessible” chromatin states (Ghisletti et al., 2010; Heinz et al., 2010), and in early T-lineage cells as in non-T cells, PU.1 occupancy was dynamically linked with local

H3K4me2 modification. Due to the developmental stability of H3K4me2 modification at TSS overall, this was most evident at distal PU.1 binding sites, where H3K4me2 modification usually melted away as PU.1 binding decreased (Figures S6B, bottom & S6D). In contrast to H3K4me2, PU.1 binding had little overlap with H3K27me3, even at silent genes (Figure S6C1, 2; S6D). PU.1 binding-linked H3K4me2 was not simply an effect of general "accessibility" or expression level of the linked gene (Figure S6D, "silent" vs. "E2A-/-" sites). Thus, PU.1 occupancy-linked changes in H3K4me2 could be used to screen candidate distal cis-regulatory sites with PU.1-dependent activity.

At candidate target genes both identified by co-regulation with PU.1 and by PU.1 activity perturbation effects, PU.1 typically occupied multiple sites, implying that full PU.1 regulatory function is commonly mediated through combinations of binding complexes. At *Tal1*, both PU.1 binding and local H3K4me2 were lost jointly from three sites, as transcription also declined (Figure 6D). At the TSS regions and intragenic sites, PU.1 loss appeared to open the way for H3K27me3 deposition. Similar patterns were seen at the TSS regions of the known PU.1 target *Flt3*, and at a downstream element and a known intronic enhancer of *Hhex* (Donaldson et al., 2005) (Figure S4F, G). Other genes with binding sites that lose PU.1 early include *Lmo1* and *Bcl11a* as well as *Itgam*, which decrease naturally from FLDN1 to FLDN2b; all are sharply downregulated in FLDN2 cells if *Sfpi1* is deleted (not shown; A. Champhekar, M. M. Del Real, S. Carotta, S. Nutt, and E. V. R., unpublished). Many sites most sensitive to loss of PU.1 may thus mediate rate limiting positive regulatory function. At the other extreme, persistent PU.1 binding occurred at the *Il7r* locus, which is upregulated from FLDN1 to FLDN2b. Despite the decreasing level of PU.1 protein, PU.1 occupancy at the *Il7r* gene remained essentially

unchanged by FLDN2b stage, both at a known TSS positive regulatory site (DeKoter et al., 2002; Xue et al., 2004), and at another putative cis-element within a silent neighboring gene, *Capsl* (Figure 6E). Here too PU.1 occupancy sites have a positive link to expression, even for this gene integral for the T-cell program.

Globally, with or without distal binding sites, PU.1 binding near the TSS appeared most consistent with a positive role (Figure 6F). Genes of diverse expression pattern clusters (defined in Figures 2D and S5) could all harbor PU.1 binding either within the body of the gene or in flanking regions, but differed sharply in frequencies of genes with PU.1 binding at the TSS regions (Figure 6G; Table S6). Genes coregulated with PU.1 itself (clusters 7; also 3, 9 & 23; blue bars) were more likely to have PU.1 binding at the TSS regions than genes regulated divergently from it (clusters 1, 2, 6, and other clusters; red bars). Progenitor-specific genes with TSS binding included *Bcl11a*, *Gfi1b*, and *Irf5* (Table S6B) as well as *Tal1* and *Hhex*. In contrast, many T-cell genes that can be downregulated by high-level PU.1 (Franco et al., 2006) either had no PU.1 binding in early T cells or had binding only in the body or flanking regions of the genes. Genes with particularly low expression in DN1 stage were most impoverished for PU.1 sites at the TSS regions (Table S6A, χ^2 test $p < 0.0001$). Thus, PU.1 binding at the promoter may provide or indicate specific antisilencing functions that maintain key stem- and progenitor-cell genes in early FLDN1 and FLDN2a stages.

Developmentally Plastic Deployment of GATA-3 Binding

GATA-3 is needed repeatedly in T-cell stages from ETP/DN1 onward and is crucial for T-lineage commitment, but capable of paradoxical effects at high doses (Taghon et al., 2007). Unlike PU.1, it is expressed almost stably across all the stages

analyzed (Figure 4C; and unpublished results). We therefore asked whether it acts through the same sites in the distinct regulatory states of FLDN1, FLDN2b, and ThyDP cells.

GATA-3 detectably bound only ~1,500 sites (Table S7). In accord with its recurrent T-cell roles, these GATA-3 sites were prominently enriched for cis-elements of T-lineage genes including *Cd3d*, *Tcf7, Zbtb7b,* and the DP-specific *Rag1-Rag2* distal enhancer (Figure S4C, H, I, B). Yet progenitor-specific genes like *Lyl1* and *Erg* (Figure 7A, B) as well as later-expressed T-cell genes like *Ets2* and *Itk* (Figure 7C, D) harbored GATA-3 sites. Occupancy patterns in our ThyDP samples were broadly consistent with those in DP $CD3^{lo}$ samples published elsewhere (Wei et al., 2011) ($r = 0.60$; Figure S7A).

Changes in GATA-3 binding were positively correlated with expression trajectories of linked genes both from DN1 to DN2b and from DN2b to DP (Figure 7E), and these were also correlated with H3K4me2 modification changes (Figure 7F). GATA-3 binding was even more likely to be a site of H3K4me2 enrichment than PU.1 binding (Figure S7B). Yet GATA-3 also bound sites linked to silent and active genes alike at early stages, and dramatically differed from PU.1 in its ability to engage sites with H3K27me3 marks (Figure S7C). For example, it remained bound to *Zbtb7b* even as it became silenced with H3K27me3 (Figure S4I). Intriguingly, GATA-3 occupancy also preceded full cis-element activation for *Cd3d* and the *Rag* enhancer (Figure S4B, C), suggesting a possible "pioneering" role.

However, the distribution of GATA-3 occupancies among different sites was strikingly different in FLDN1, FLDN2b, and ThyDP. This was despite nearly constant protein availability: global occupancy levels and peak heights at stably occupied sites

such as the *Tcf3* (*Tcfe2a*) promoter (Figure 7D) and the *Tcrb* 3' enhancer (not shown) were similar in all stages. Also, the most common motifs at sites of occupancy were a classic GATA site and an Ets family-like site (Figure 7G), in FLDN1, FLDN2b, and ThyDP alike. But from FLDN1 to ThyDP, GATA-3 occupancy increased sharply at some sites (e.g., *Ets2* in DN2b, *Itk* promoter in DP; Figure 7B, C), while disappearing from others entirely (e.g., *Lyl1, Erg*, *Itk* introns*;* Figure 7A-C). Overall, whereas sites occupied in FLDN1 and FLDN2b stages were moderately well correlated (r = 0.61), sites in FLDN2b and ThyDP were poorly correlated (r = 0.22) and those in FLDN1 and ThyDP entirely uncorrelated (r = -0.0064, Figure 7H).

These results locate elements in T and non-T genes where the crucial T-cell factor GATA-3 can be contributing to regulation, from the FLDN1 stage on (Table S7). Nevertheless, they also reveal that a target gene for GATA-3 at one stage of T-cell development may not normally receive input from GATA-3 at another stage, despite similar GATA-3 availability. In contrast to PU.1, GATA-3's physiological deployment at any given stage depends not only on its own availability but also on a specific developmental regulatory context.

DISCUSSION

Our results provide a resource for T cell development, a new reference case for regulatory epigenomics, and potentially powerful new explanatory elements for a complex developmental process.

For T-cell development, the results here provide a global, base-resolution time-course of the chromatin landscape and transcriptome during the finely defined stages of

T-cell specification, which substantially extends previous knowledge. The transcriptome data not only quantify RNA levels but also provide detailed information about promoter and exon choice that may affect gene regulation as well as function. We also note that these data reveal noncoding transcripts that maybe important in various regulatory roles. Here, we focused on two elements of greatest causal importance to explain T-cell identity: the ~400 regulatory gene loci that themselves are developmentally regulated during this process, and also the particular subset of candidate cis-regulatory genomic sites that undergo developmental changes in histone modifications in response to changes in local regulatory inputs. These are the most likely trans and cis-components of nodes in the gene network that gate successive steps of the T-cell program. For example, the mapping of specific GATA-3 binding sites at a core set of important T-cell genes will clearly speed resolution of the role of this crucial factor during early T-cell specification.

Most powerfully, the results reveal the subcomponent processes out of which T-cell specification is built. Our results confirm that relatively few regulatory genes undergo strong upregulation during lineage commitment itself. Instead, a major feature integrally linked to commitment is specific and marked downregulation of progenitor-cell genes, which is revealed to be unexpectedly complex. Important non-T hematopoietic regulatory genes are still expressed in the precursors through 4 days of consistent Notch pathway signaling, and many persist even into the DN2a stage before they are shut off. However, their repression during commitment is not due to a single switching mechanism, nor is due to the decay kinetics of old mRNA retained long after transcription has ceased. Importantly, the histone marking status of the promoters and linked cis-elements provides an independent line of evidence about the timing of

regulatory changes, supplementing measurements of population mRNA levels. Thus, the diverse histone mark transformations that are applied to different repression targets imply that a variety of biochemically and temporally distinct silencing mechanisms must be used. Notably, this rules out Notch signaling itself as a common mechanism of repression and implies that the T-lineage program deploys multiple waves of repressor functions needed to establish T-cell identity.

Dynamically regulated transcriptional repression during this process is often separable from "epigenetic silencing". At finest scale, our results show that deposition of H3K27me3 histone marks is more probably an effect and stabilizer of repression than an initial cause of repression. *De novo* H3K27me3 marking accumulates at newly-repressed loci in two distinct, major patterns. One appears to be lateral invasion from a neighboring patch of pre-existing "closed" chromatin, often after a positive regulator (e.g., PU.1) is removed, but this can be forestalled indefinitely as in the case of the *Gata3* locus. Another is by tight focal deposition at a previously active TSS or enhancer site, followed by spreading. At many downregulated genes in our survey, decreased RNA levels clearly precede deposition of H3K27me3. In other cases repression does not involve H3K27me3 at all, including *Sfpi1* and *Cd4,* two key genes known to be repressed by Runx factors in DN3 cells. Even when H3K27me3 is used, it is readily and precisely reversible. Repression via DNA methylation was not studied here but is also reversible, as shown recently by the cell type-specific demethylation of CpGs in DN2-DN3 cells at loci that include *Tcf7* and *Bcl11b* (Ji et al., 2010) (http://charm.jhmi.edu/hsc/). These examples underline the need for transcriptional repressors to act first to trigger chromatin closing, and the power of transcriptional activators to undo it.

Our results also address a long-standing question in T-cell development, namely explaining how the positive drivers of the T-cell program work. Essential roles for GATA-3, Notch1 and TCF-1 were established early, but validation of their specific cis-regulatory targets has been slow. GATA-3 effects in early T cells have been especially difficult to dissect, in part due to the profound loss of viability when GATA-3 dose is reduced (Hosoya et al., 2009), and in part due to lineage-inappropriate effects of GATA-3 in gain of function experiments (Taghon et al., 2007). The identification of a battery of cis-regulatory elements activated *de novo* from DN1 to DN2b is an important new resource for clarifying these links. Identification of sites for potential GATA-3 regulatory inputs into *Tcf7* as well as *Tcfe2a* suggests a new level of regulatory interlinkage, which could explain the acuteness of the GATA-3 requirement. At least in DP cells, data from (Wei et al., 2011) do suggest that the GATA-3 binding we see positively regulates *Tcf7*, *Cd3d*, and possibly *Zfpm1*, and negatively regulates *Tcfe2a*. Our results may also help to explain GATA-3's lineage infidelity in gain of function experiments by revealing how conditionally this factor normally provides its inputs at legitimate target sites. Even at a fixed level of expression, GATA-3's action is stage specific. Altered dosages could thus override the mechanisms that must provide appropriate targeting specificity.

T-lineage commitment is an ordered process in which different alternative hematopoietic fates are relinquished sequentially. We show that the repressive events through which alternative lineages are excluded are mechanistically separable. Thus the B-cell regulatory genes *Pax5* and *Ebf1* are silenced by H3K27me3 and rendered inaccessible to PU.1 binding from the start, whereas the myeloid regulatory gene *Cebpa* is bivalently marked. The myeloid and progenitor regulatory gene *Sfpi1* (PU.1), initially

fully activated, plays a surprisingly effective regulatory role even into the DN2b stage, and is silenced only when T-cell gene expression is under way. However, there is no simple mapping of developmental lineage exclusion order with a particular molecular class of repression mechanism. Drivers of the most "distant" fate in developmental terms, the erythroid genes, can be repressed via H3K27me3 (*EpoR*), or without it (*Gata1*), as can genes associated with the "closest", NK-cell fate (*Eomes*, *Il2rb* respectively). Like *Sfpi1*, multipotent progenitor-cell regulatory genes are strongly expressed initially and shut off during the commitment process. Many of these genes may be sustained by a common progenitor-cell positive regulator, *Lmo2* (McCormack et al., 2010), and we show many are also direct binding targets of PU.1 itself. Each of these genes responds to its own combination of positive and regulatory inputs, as a distinct node in the T-lineage specification network. Indeed, the complexity and diversity of mechanisms revealed in this initial global study argue that evolution has drawn liberally on combinations of factors and pathways to regulate a T-cell developmental progression that integrates multiple external and internal inputs.

Our multistage analysis shows that many mouse hematopoietic genes are each likely controlled by different constellations of cis-regulatory elements at one stage of development versus another, even within the same cell lineage. In this light, the quest for single, minimal sufficient regulator elements for such genes seems naïve, as it would *a priori* sacrifice the full range of developmental control. The roles of the candidate cis-elements and their rules for engagement with promoters should be greatly clarified by future extensions of this analysis, to detect specific chromatin looping events, enhancer activation states mapped by association with p300 and H3K4me3, and latent enhancers

using H3K4me1 at transcription factor binding sites. Mechanisms of repression could be clarified when effects on a broader range of non-activating cis-elements are mapped based on DNase hypersensitivity and DNA methylation. The mapping of developmentally dynamic histone modification sites provides a new way to locate the sites in cis-regulatory DNA that process distinct inputs for crucial regulatory genes. In this collection of regulatory domains lie the answers to how cells are driven to T-lineage commitment.

METHODS

Full materials and experimental procedures are given in the Supplement. **Accession codes.** GEO:GSE31235.

Acknowledgments

We thank Lorian Schaeffer and Vijaya Kumar for library preparation and sequencing; Henry Amrhein and Diane Trout for data curation; Igor Antoshechkin for sequencing facility management; Diana Perez, Josh Verceles, and Rochelle Diamond for cell sorting and advice; Tian Ling, Georgi Marinov, and Hao Yuan Kueh for statistical advice and programming; Rothenberg group members for sharing advice and unpublished data; and Robert Butler, Lorena Sandoval and Scott Washburn for care of the mice. Support was from NIH grants R33HL089123, R01CA090233, and RC2CA148278, the Beckman Institute, the Millard and Muriel Jacobs Genetics and Genomics Center, the L. A. Garfinkle Memorial Laboratory Fund, the Al Sherman Foundation, the Bren Professorship (B.J.W.), and the A. B. Ruddock Professorship (E.V.R.).

References

Araki,Y., Wang,Z., Zang,C., Wood,W.H., III, Schones,D., Cui,K., Roh,T.Y., Lhotsky,B., Wersto,R.P., Peng,W., Becker,K.G., Zhao,K., and Weng,N.P. (2009). Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory $CD8^+$ T cells. Immunity. *30*, 912-925.

Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Schones,D.E., Wang,Z., Wei,G., Chepelev,I., and Zhao,K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

David-Fung,E.-S., Butler,R., Buzi,G., Yui,M.A., Diamond,R.A., Anderson,M.K., Rowen,L., and Rothenberg,E.V. (2009). Transcription factor expression dynamics of early T-lymphocyte specification and commitment. Dev. Biol. *325*, 444-467.

Decker,T., Pasca,d.M., McManus,S., Sun,Q., Bonifer,C., Tagoh,H., and Busslinger,M. (2009). Stepwise activation of enhancer and promoter regions of the B cell commitment gene Pax5 in early lymphopoiesis. Immunity. *30*, 508-520.

DeKoter,R.P., Lee,H.-J., and Singh,H. (2002). PU.1 regulates expression of the Interleukin-7 receptor in lymphoid progenitors. Immunity. *16*, 297-309.

Donaldson,I.J., Chapman,M., Kinston,S., Landry,J.R., Knezevic,K., Piltz,S., Buckley,N., Green,A.R., and Gottgens,B. (2005). Genome-wide identification of cis-regulatory sequences controlling blood and endothelial development. Hum. Mol. Genet. *14*, 595-601.

Filion,G.J., van Bemmel,J.G., Braunschweig,U., Talhout,W., Kind,J., Ward,L.D., Brugman,W., de Castro,I.J., Kerkhoven,R.M., Bussemaker,H.J., and van Steensel,B. (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212-224.

Franco,C.B., Scripture-Adams,D.D., Proekt,I., Taghon,T., Weiss,A.H., Yui,M.A., Adams,S.L., Diamond,R.A., and Rothenberg,E.V. (2006). Notch/Delta signaling constrains re-engineering of pro-T cells by PU.1. Proc. Natl. Acad. Sci. U. S. A *103*, 11993-11998.

Georgopoulos,K., van den,E.P., Bier,E., Maxam,A., and Terhorst,C. (1988). A T cell-specific enhancer is located in a DNase I-hypersensitive area at the 3' end of the CD3-d gene. EMBO J. *7*, 2401-2407.

Ghisletti,S., Barozzi,I., Mietton,F., Polletti,S., De Santa,F., Venturini,E., Gregory,L., Lonie,L., Chew,A., Wei,C.L., Ragoussis,J., and Natoli,G. (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity. *32*, 317-328.

Gomez-del Arco,P., Kashiwagi,M., Jackson,A.F., Naito,T., Zhang,J., Liu,F., Kee,B., Vooijs,M., Radtke,F., Redondo,J.M., and Georgopoulos,K. (2010). Alternative promoter usage at the Notch1 locus supports ligand-independent signaling in T cell development and leukemogenesis. Immunity. *33*, 685-698.

Hawkins,R.D., Hon,G.C., Lee,L.K., Ngo,Q., Lister,R., Pelizzola,M., Edsall,L.E., Kuan,S., Luu,Y., Klugman,S., Antosiewicz-Bourget,J., Ye,Z., Espinoza,C., Agarwahl,S.,

Shen,L., Ruotti,V., Wang,W., Stewart,R., Thomson,J.A., Ecker,J.R., and Ren,B. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell *6*, 479-491.

He,H.H., Meyer,C.A., Shin,H., Bailey,S.T., Wei,G., Wang,Q., Zhang,Y., Xu,K., Ni,M., Lupien,M., Mieczkowski,P., Lieb,J.D., Zhao,K., Brown,M., and Liu,X.S. (2010). Nucleosome dynamics define transcriptional enhancers. Nat. Genet. *42*, 343-347.

Heintzman,N.D., Stuart,R.K., Hon,G., Fu,Y., Ching,C.W., Hawkins,R.D., Barrera,L.O., Van Calcar,S., Qu,C., Ching,K.A., Wang,W., Weng,Z., Green,R.D., Crawford,G.E., and Ren,B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311-318.

Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H., and Glass,C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576-589.

Heng,T.S.P., Painter,M.W., and Consortium,T.I.G.P. (2008). The Immunological Genome Project: networks of gene expression in immune cells. Nat. Immunol. *9*, 1091-1094.

Hoogenkamp,M., Krysinska,H., Ingram,R., Huang,G., Barlow,R., Clarke,D., Ebralidze,A., Zhang,P., Tagoh,H., Cockerill,P.N., Tenen,D.G., and Bonifer,C. (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. Mol Cell Biol *27*, 7425-7438.

Hosoya,T., Kuroha,T., Moriguchi,T., Cummings,D., Maillard,I., Lim,K.C., and Engel,J.D. (2009). GATA-3 is required for early T lineage progenitor development. J. Exp. Med. *206*, 2987-3000.

Hosoya,T., Maillard,I., and Engel,J.D. (2010). From the cradle to the grave: activities of GATA-3 throughout T-cell development and differentiation. Immunol. Rev. *238*, 110-125.

Ji,H., Ehrlich,L.I., Seita,J., Murakami,P., Doi,A., Lindau,P., Lee,H., Aryee,M.J., Irizarry,R.A., Kim,K., Rossi,D.J., Inlay,M.A., Serwold,T., Karsunky,H., Ho,L., Daley,G.Q., Weissman,I.L., and Feinberg,A.P. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature *467*, 338-342.

Johnson,D.S., Mortazavi,A., Myers,R.M., and Wold,B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497-1502.

Kawazu,M., Yamamoto,G., Yoshimi,M., Yamamoto,K., Asai,T., Ichikawa,M., Seo,S., Nakagawa,M., Chiba,S., Kurokawa,M., and Ogawa,S. (2007). Expression profiling of immature thymocytes revealed a novel homeobox gene that regulates double-negative thymocyte development. J. Immunol. *179*, 5335-5345.

Koche,R.P., Smith,Z.D., Adli,M., Gu,H., Ku,M., Gnirke,A., Bernstein,B.E., and Meissner,A. (2011a). Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell *8*, 96-105.

Kouzarides,T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Lin,Y.C., Jhunjhunwala,S., Benner,C., Heinz,S., Welinder,E., Mansson,R., Sigvardsson,M., Hagman,J., Espinoza,C.A., Dutkowski,J., Ideker,T., Glass,C.K., and Murre,C. (2010). A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. Nat. Immunol. *11*, 635-643.

McCormack,M.P., Young,L.F., Vasudevan,S., de Graaf,C.A., Codrington,R., Rabbitts,T.H., Jane,S.M., and Curtis,D.J. (2010). The Lmo2 oncogene initiates leukemia in mice by inducing thymocyte self-renewal. Science *327*, 879-883.

Mortazavi,A., Williams,B.A., McCue,K., Schaeffer,L., and Wold,B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.

Natoli,G. (2010). Maintaining cell identity through global control of genomic organization. Immunity. *33*, 12-24.

Novershtern,N., Subramanian,A., Lawton,L.N., Mak,R.H., Haining,W.N., McConkey,M.E., Habib,N., Yosef,N., Chang,C.Y., Shay,T., Frampton,G.M., Drake,A.C., Leskov,I., Nilsson,B., Preffer,F., Dombkowski,D., Evans,J.W., Liefeld,T., Smutko,J.S., Chen,J., Friedman,N., Young,R.A., Golub,T.R., Regev,A., and Ebert,B.L. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell *144*, 296-309.

Orford,K., Kharchenko,P., Lai,W., Dao,M.C., Worhunsky,D.J., Ferro,A., Janzen,V., Park,P.J., and Scadden,D.T. (2008). Differential H3K4 methylation identifies developmentally poised hematopoietic genes. Dev Cell *14*, 798-809.

Petrie,H.T. and Zuniga-Pflucker,J.C. (2007). Zoned out: functional mapping of stromal signaling microenvironments in the thymus. Annu. Rev. Immunol. *25*, 649-679.

Rosenbauer,F., Owens,B.M., Yu,L., Tumang,J.R., Steidl,U., Kutok,J.L., Clayton,L.K., Wagner,K., Scheller,M., Iwasaki,H., Liu,C., Hackanson,B., Akashi,K., Leutz,A., Rothstein,T.L., Plass,C., and Tenen,D.G. (2006). Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU.1. Nat Genet. *38*, 27-37.

Rothenberg,E.V. (2011). T cell lineage commitment: identity and renunciation. J. Immunol. *186*, 6649-6655.

Rothenberg,E.V., Moore,J.E., and Yui,M.A. (2008). Launching the T-cell-lineage developmental programme. Nat. Rev. Immunol. *8*, 9-21.

Rothenberg,E.V. and Scripture-Adams,D.D. (2008). Competition and collaboration: GATA-3, PU.1, and Notch signaling in early T-cell fate determination. Semin. Immunol *20*, 236-246.

Schebesta,A., McManus,S., Salvagiotto,G., Delogu,A., Busslinger,G.A., and Busslinger,M. (2007). Transcription factor Pax5 activates the chromatin of key genes involved in B cell signaling, adhesion, migration, and immune function. Immunity. *27*, 49-63.

Tabrizifard,S., Olaru,A., Plotkin,J., Fallahi-Sichani,M., Livak,F., and Petrie,H.T. (2004). Analysis of transcription factor expression during discrete stages of postnatal thymocyte differentiation. J. Immunol. *173*, 1094-1102.

Taghon,T., Yui,M.A., and Rothenberg,E.V. (2007). Mast cell lineage diversion of T lineage precursors by the essential T-cell transcription factor GATA-3. Nat. Immunol. *8*, 845-855.

van de Wetering,M., Oosterwegel,M., Dooijes,D., and Clevers,H. (1991). Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. EMBO J. *10*, 123-132.

Wang,Z., Zang,C., Rosenfeld,J.A., Schones,D.E., Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Peng,W., Zhang,M.Q., and Zhao,K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. *40*, 897-903.

Wei,G., Abraham,B.J., Yagi,R., Jothi,R., Cui,K., Sharma,S., Narlikar,L., Northrup,D.L., Tang,Q., Paul,W.E., Zhu,J., and Zhao,K. (2011). Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. Immunity. *35*, 299-311.

Wei,G., Wei,L., Zhu,J., Zang,C., Hu-Li,J., Yao,Z., Cui,K., Kanno,Y., Roh,T.Y., Watford,W.T., Schones,D.E., Peng,W., Sun,H.W., Paul,W.E., O'Shea,J.J., and Zhao,K. (2009). Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity. *30*, 155-167.

Weishaupt,H., Sigvardsson,M., and Attema,J.L. (2010). Epigenetic chromatin states uniquely define the developmental plasticity of murine hematopoietic stem cells. Blood *115*, 247-256.

Wold,B. and Myers,R.M. (2008). Sequence census methods for functional genomics. Nat Methods *5*, 19-21.

Xue,H.-H., Bollenbacher,J., Rovella,V., Tripuraneni,R., Du,Y.-B., Liu,C.-Y., Williams,A., McCoy,J.P., and Leonard,W.J. (2004). GA binding protein regulates interleukin 7 receptor a-chain gene expression in T cells. Nat. Immunol. *5*, 1036-1044.

Yang,Q., Bell,J.J., and Bhandoola,A. (2010). T-cell lineage determination. Immunol. Rev. *238*, 12-22.

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

Yui,M.A., Feng,N., and Rothenberg,E.V. (2010). Fine-scale staging of T cell lineage commitment in adult mouse thymus. J. Immunol. *185*, 284-293.

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. Mol. Cell Biol. *30*, 4922-4939.

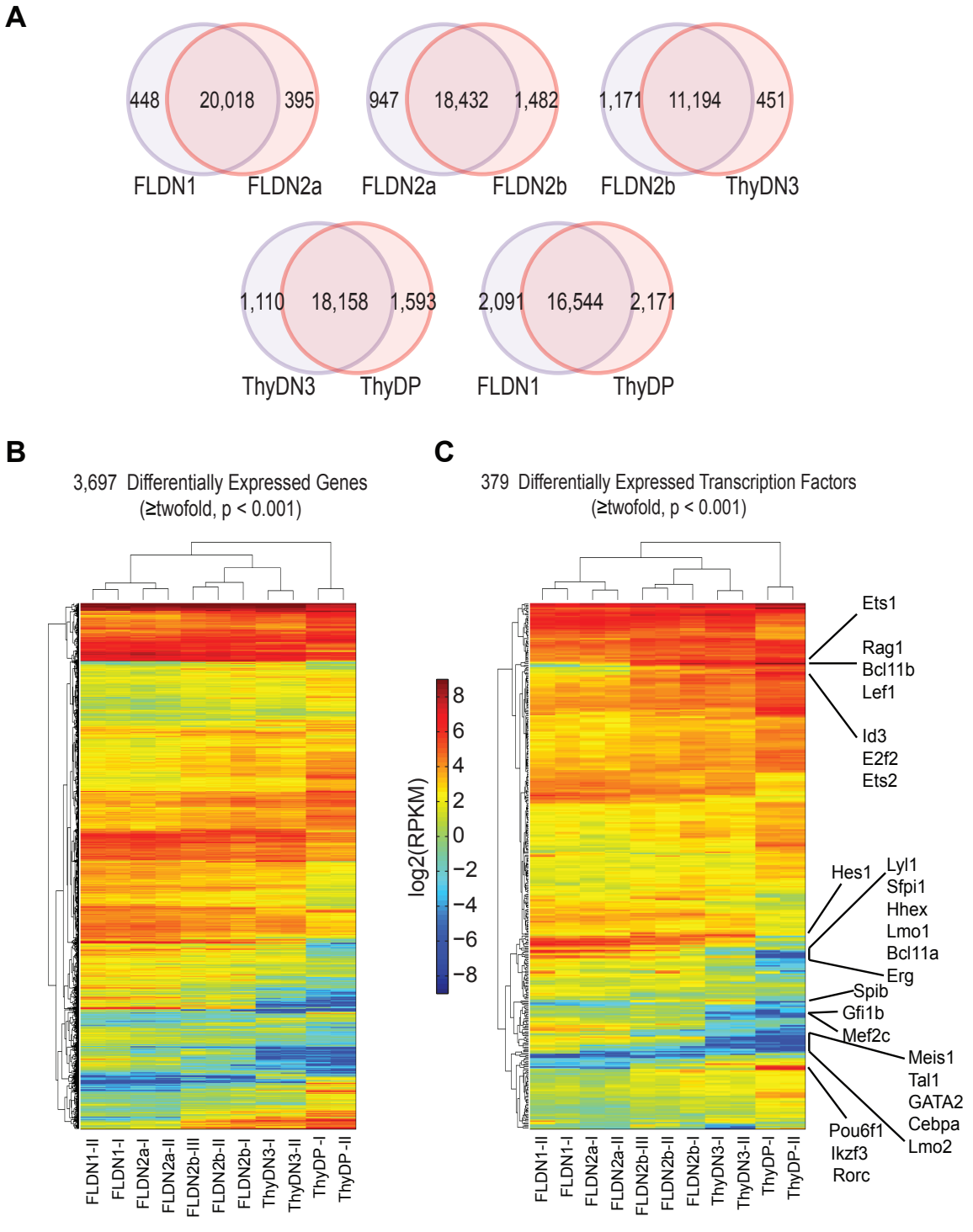


Figure 1. Global comparisons of gene expression among five developmentally related immature T cell populations.

A. Pair-wise comparisons in gene expression between successive populations as well as initial FLDN1 vs. final ThyDP stages. Statistically changed genes were defined by DEGseq ($p < 0.001$).

B. Hierarchical clustering of expression patterns of all differentially expressed genes. Expression levels of all DEGseq positive genes with ≥2x difference in expression are hierarchically clustered along both sample and gene dimensions and displayed as a heatmap.

C. Hierarchical clustering of expression patterns of differentially expressed transcription factors. Several known early T cell development related transcription factors are indicated.

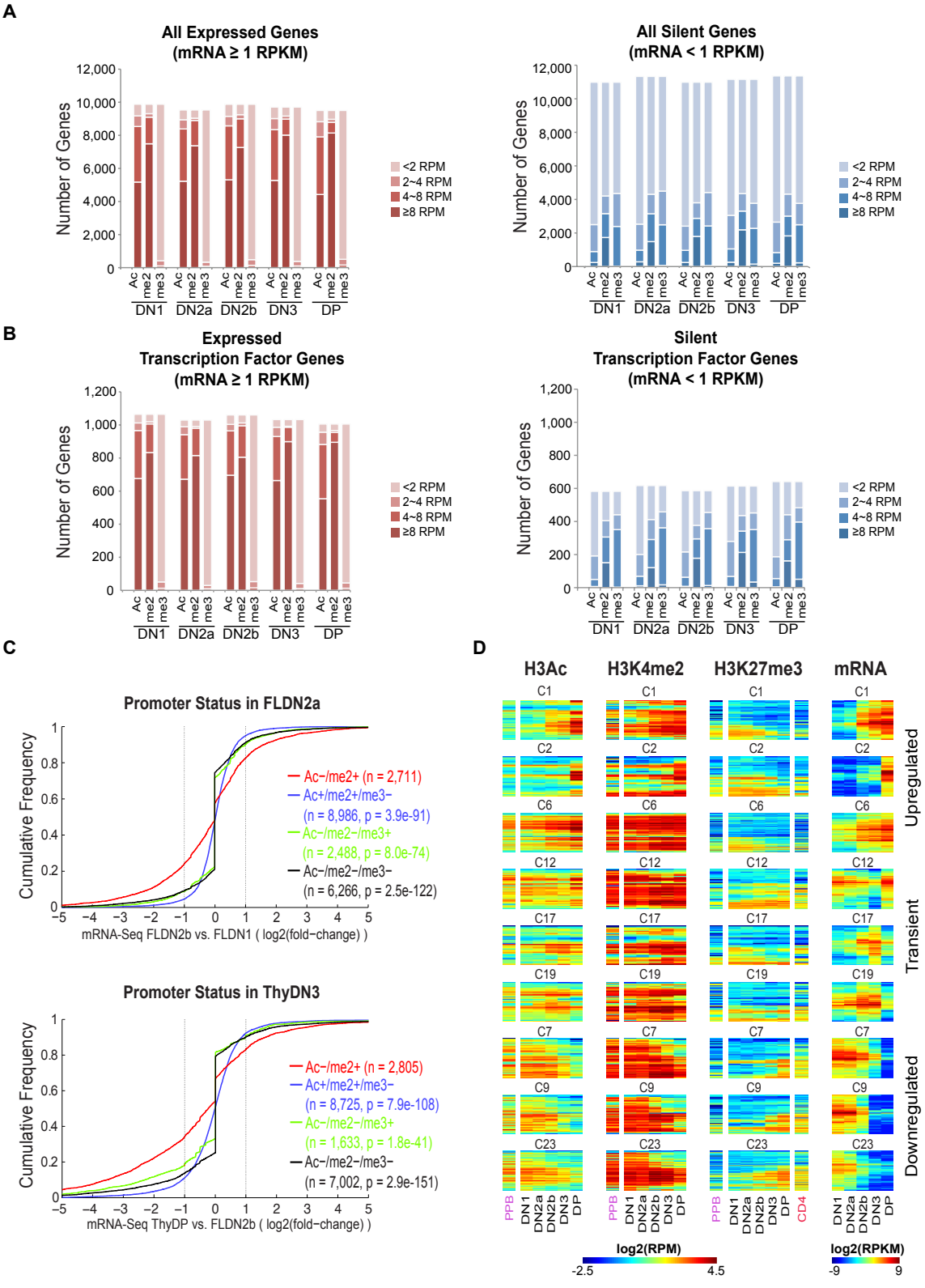

A
All Expressed Genes
(mRNA ≥ 1 RPKM)
All Silent Genes
(mRNA < 1 RPKM)
Number of Genes
<2 RPM
2~4 RPM
4~8 RPM
≥8 RPM
Ac
me2
me3
DN1
DN2a
DN2b
DN3
DP
B
Expressed
Transcription Factor Genes
(mRNA ≥ 1 RPKM)
Silent
Transcription Factor Genes
(mRNA < 1 RPKM)
C
Promoter Status in FLDN2a
Cumulative Frequency
Ac−/me2+ (n = 2,711)
Ac+/me2+/me3−
(n = 8,986, p = 3.9e-91)
Ac−/me2−/me3+
(n = 2,488, p = 8.0e-74)
Ac−/me2−/me3−
(n = 6,266, p = 2.5e-122)
mRNA-Seq FLDN2b vs. FLDN1 (log2(fold−change))
Promoter Status in ThyDN3
Ac−/me2+ (n = 2,805)
Ac+/me2+/me3−
(n = 8,725, p = 7.9e-108)
Ac−/me2−/me3+
(n = 1,633, p = 1.8e-41)
Ac−/me2−/me3−
(n = 7,002, p = 2.9e-151)
mRNA-Seq ThyDP vs. FLDN2b (log2(fold−change))
D
H3Ac
H3K4me2
H3K27me3
mRNA
C1
C2
C6
C12
C17
C19
C7
C9
C23
Upregulated
Transient
Downregulated
PPB
CD4
log2(RPM)
-2.5
4.5
log2(RPKM)
-9
9

Figure 2. Distinct gene expression patterns are associated with characteristic histone modifications.

A. Association of gene expression and histone modifications at the promoters of 20,861 UCSC annotated genes. Expressed (≥1 RPKM) and silent (<1 RPKM) genes in each stage were distinguished by RNA-seq. Color shading denotes the level of the given histone modification at promoters, defined as ±1 kb of the TSS. Ac: H3Ac, me2: H3K4me2, me3: H3K27me3.

B. Association of gene expression and histone modifications at the promoters of 1,646 genes encoding DNA-binding proteins or transcription factors.

C. Association of promoter-linked histone modifications with developmental change in expression. Genes are grouped based on histone marks of their TSS in the FLDN2a (top) or ThyDN3 (bottom) stages: H3Ac- H3K4me2+ (H3K27me3+ or -) in red, H3Ac+ H3K4me2+ H3K27me3- in blue, H3Ac- H3K4me2- H3K27me3+ in green, and H3Ac- H3K4me2- H3K27me3- in black. Plots show the cumulative distributions of genes in each group with various degrees of expression change from the previous stage to the following stage, i.e., from FLDN1 to FLDN2b (top), and from FLDN2b to ThyDP (bottom). Genes showing different changes in expression ($\log_2$ ratio of RNA-seq levels) are separated on the x-axis, with downregulated genes to the left and upregulated genes to the right, and their fractional representations within the group are accumulated on the y-axis. The group of genes that were H3Ac-/H3K4me2+, with or without H3K27me3, are most developmentally labile, with >20% that are downregulated more than 2x and >20% that are upregulated more than 2x (vertical lines: boundaries for >2x change), and this difference is highly significant (Kolmogorov-Smirnov [K-S] (two-sided) test for

comparison between H3Ac-/H3K4me2+ and each of the other three groups; p value for each comparison in parentheses).

D. Heatmaps correlating TSS histone modifications with various patterns of developmentally regulated gene expression for 9 representative clusters (see Figure S5). Clusters 1, 2, and 6: upregulated from DN1 to DP stage. Clusters 12, 17 and 19: genes transiently changed in mRNA expression from DN2a to DN3. Clusters 7, 9, and 23 downregulated from DN1 to DN2b stage. Normalized signal densities of histone modifications at the promoter region of each gene were aligned with normalized mRNA data, and hierarchical clustering (one-dimensional, along genes) was performed for individual clusters using Ward linkage and Euclidean distance. As comparisons, histone modification data from $EBF^{-/-}$ pre-pro B cells (H3Ac, H3K4me2 and H3K27me3; "PPB") and CD4 naïve T cells (H3K27me3 only, "CD4") were then added to the heatmaps after the hierarchical clustering had been generated.

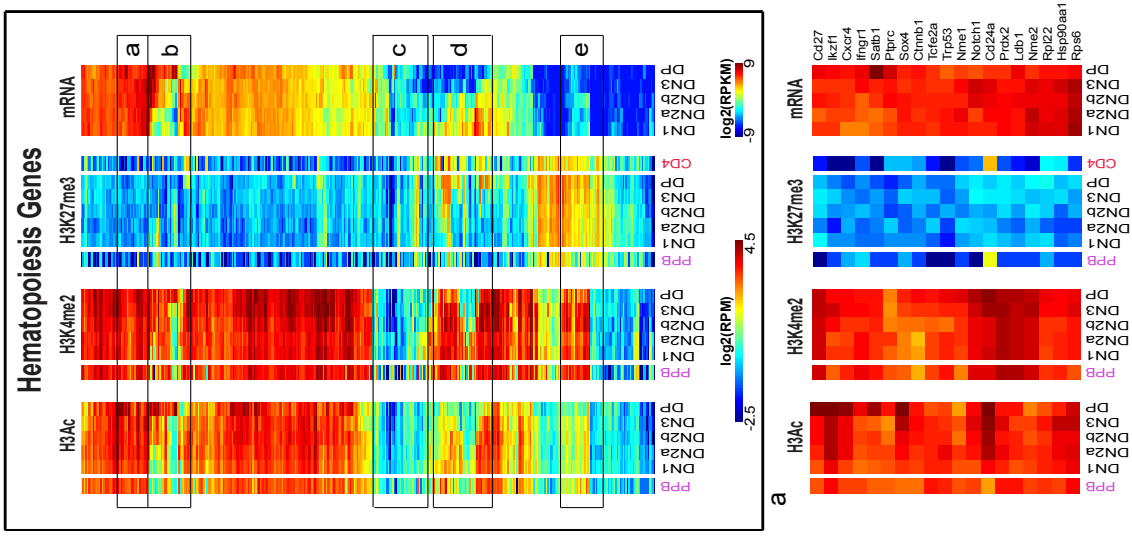

Hematopoiesis Genes
mRNA
H3K27me3
H3K4me2
H3Ac
DN1
DN2a
DN2b
DN3
DP
CD4
PPB
log2(RPKM)
9
-9
4.5
-2.5
a
b
c
d
e

Figure 3. Histone modifications and gene expression profiles of genes characterizing hematopoiesis.

Results for 379 “hematopoietic” genes are processed and displayed as in Figure 2D. Master panel shows results for all 379 genes. Signature cluster regions (a-e) are zoomed in to allow individual genes to be seen.

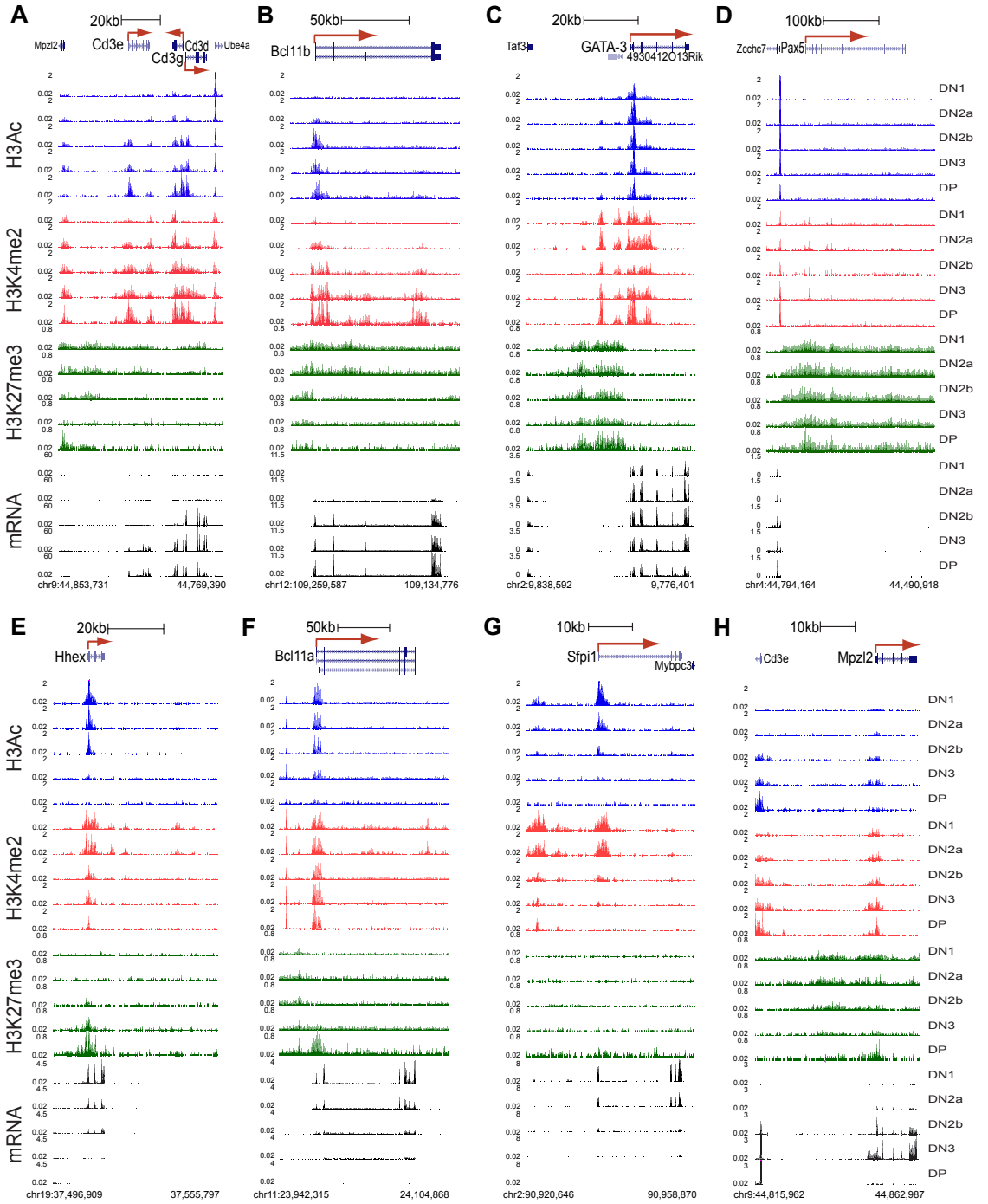


Figure 4. Portraits of key T-lineage and alternative-lineage genes.

A – H. UCSC browser track images depicting distinct epigenetic marking and gene expression patterns at eight different loci: *Bcl11b* (A), *Cd3e/d/g* cluster (B), *Gata3* (C),

Pax5 (D), *Hhex* (E), *Bcl11a* (F), *Sfpi1* (G) and *Mpzl2* (H), in all five immature T-populations (top to bottom, DN1, DN2a, DN2b, DN3 and DP). Red arrow: TSS and direction of transcription. H3Ac: blue, H3K4me2: red, H3K27me3: green, and RNA-seq: black. Uniform scales are used for histone marks in all panels, and mRNA scales are uniform within each panel, as shown by y-axis labels (units in RPM for ChIP-seq, RPKM for RNA-seq). Chromosomal coordinates are indicated below each panel.

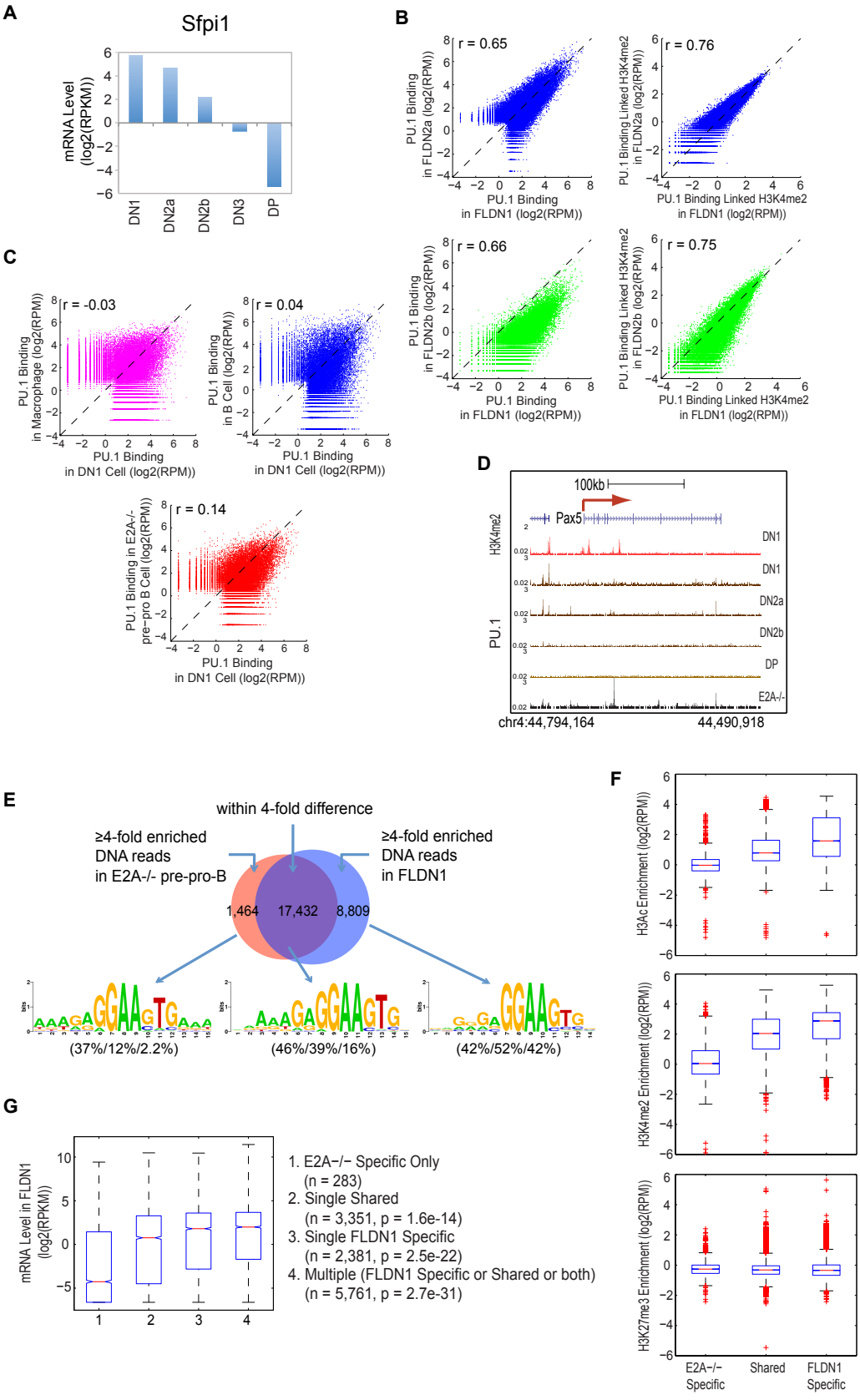

A
Sfpi1
mRNA Level (log2(RPKM))
DN1
DN2a
DN2b
DN3
DP
B
r = 0.65
PU.1 Binding in FLDN2a (log2(RPM))
PU.1 Binding in FLDN1 (log2(RPM))
r = 0.76
PU.1 Binding Linked H3K4me2 in FLDN2a (log2(RPM))
PU.1 Binding Linked H3K4me2 in FLDN1 (log2(RPM))
r = 0.66
PU.1 Binding in FLDN2b (log2(RPM))
r = 0.75
PU.1 Binding Linked H3K4me2 in FLDN2b (log2(RPM))
C
r = -0.03
PU.1 Binding in Macrophage (log2(RPM))
PU.1 Binding in DN1 Cell (log2(RPM))
r = 0.04
PU.1 Binding in B Cell (log2(RPM))
r = 0.14
PU.1 Binding in E2A-/- pre-pro B Cell (log2(RPM))
D
100kb
Pax5
H3K4me2
PU.1
DN1
DN1
DN2a
DN2b
DP
E2A-/-
chr4:44,794,164
44,490,918
E
within 4-fold difference
≥4-fold enriched DNA reads in E2A-/- pre-pro-B
≥4-fold enriched DNA reads in FLDN1
1,464
17,432
8,809
(37%/12%/2.2%)
(46%/39%/16%)
(42%/52%/42%)
F
H3Ac Enrichment (log2(RPM))
H3K4me2 Enrichment (log2(RPM))
H3K27me3 Enrichment (log2(RPM))
E2A−/− Specific
Shared
FLDN1 Specific
G
mRNA Level in FLDN1 (log2(RPKM))
1. E2A−/− Specific Only (n = 283)
2. Single Shared (n = 3,351, p = 1.6e-14)
3. Single FLDN1 Specific (n = 2,381, p = 2.5e-22)
4. Multiple (FLDN1 Specific or Shared or both) (n = 5,761, p = 2.7e-31)

Figure 5. Lineage specific PU.1 DNA binding is associated with lineage specific histone modifications and gene expression.

A. Mean RNA-seq level of PU.1 (*Sfpi1*) at each stage of early T cell development.

B. (Left) Comparisons of PU.1 DNA binding site distributions between FLDN1 and FLDN2a or FLDN2b. Pearson's correlation coefficient for each comparison (r) is indicated. (Right) Comparisons of PU.1 DNA binding-associated H3K4me2 enrichment between FLDN1 and FLDN2a or FLDN2b. H3K4me2 signal densities were calculated within ±1 kb of the summit of a given PU.1 bound region (as in Figure 5B).

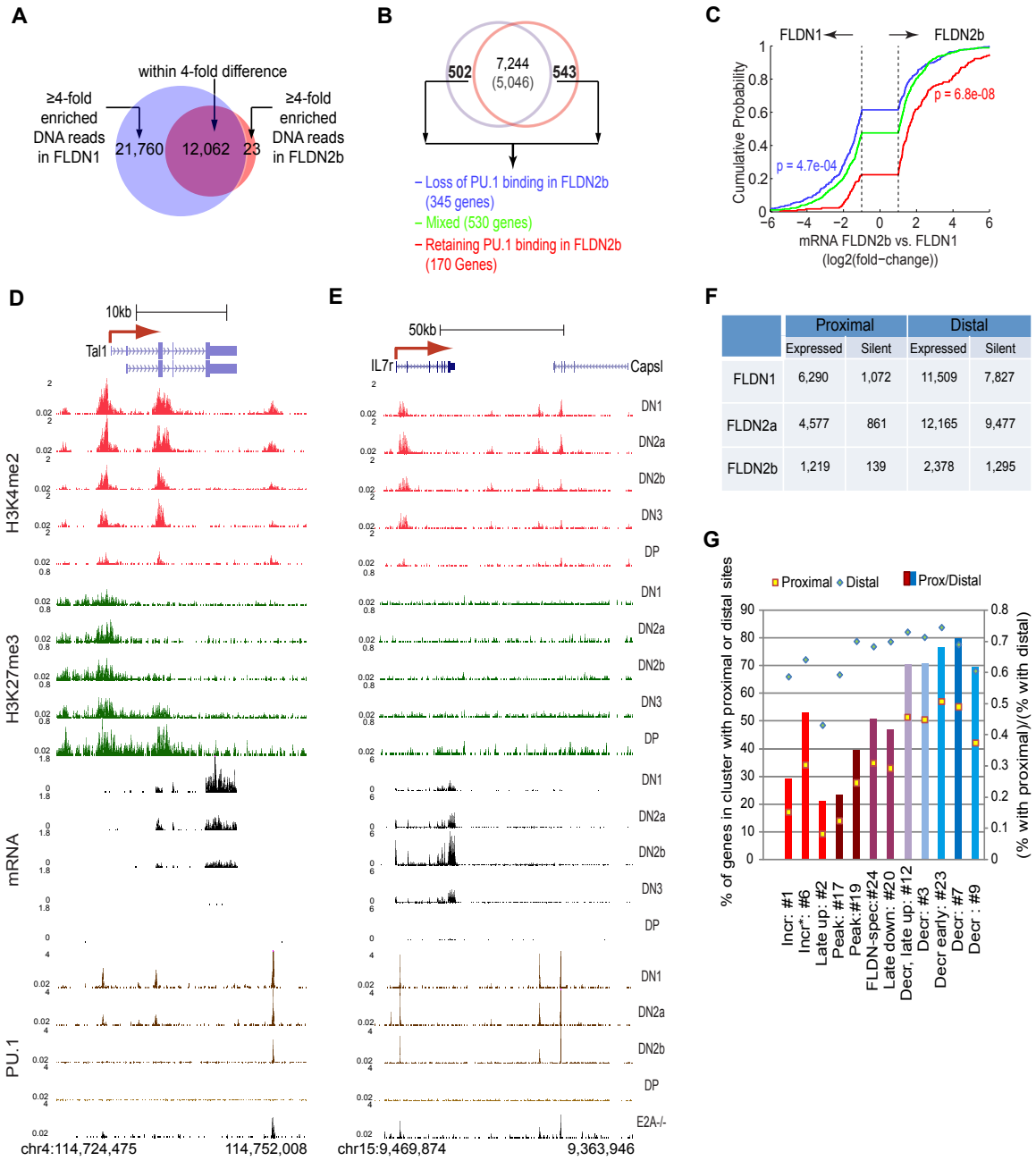
C. Comparisons of PU.1 DNA binding between FLDN1 and $E2A^{-/-}$ pre-pro B, macrophage or mature B cells.

D. Lineage-specific PU.1 binding at the *Pax5* locus. B cell specific Pax5 intronic enhancer is bound by PU.1 (black arrow) in $E2A^{-/-}$ pre-pro pre-pro B cells (Heinz et al., 2010) (black track), but not in DN cells (brown tracks). PU.1 ChIP-seq in ThyDP is used as a negative control. For orientation, H3K4me2 pattern in DN1 stage is included (red track). Note low level of H3K4me2 modification at the non-T specific PU.1 binding site, within a known Pax5 enhancer (Decker et al., 2009).

E. Lineage specific and shared PU.1 binding sites between FLDN1 and $E2A^{-/-}$ pre-pro B cells. Lineage specific binding regions: more than 4x difference in PU.1 occupancy between the two populations. Sequence logos are shown for the most highly enriched sequence motif for each occupancy subgroup. Percentages of regions from the three subgroups containing at least one instance of each motif are indicated in parentheses beneath each sequence logo ($E2A^{-/-}$ specific/Shared/FLDN1 specific).

F. Distribution of the enrichment of specified histone modification over genomic regions within ±1 kb of lineage specific and shared PU.1 binding sites in FLDN1 cells.

G. Correlation of mRNA expression levels in FLDN1 with presence of lineage-specific or shared PU.1 sites. Distribution of mRNA value in FLDN1 for subgroups of genes that are linked to either $E2A^{-/-}$ pre-pro B specific, FLDN1 specific, or shared PU.1 binding sites, and genes linked to more than one PU.1 site occupied in FLDN1 cells (Multiple). K-S test for mRNA levels was performed between $E2A^{-/-}$ Specific Only and each of the other three subgroups (number of genes in each group and p values in parentheses).



	Proximal		Distal	
	Expressed	Silent	Expressed	Silent
FLDN1	6,290	1,072	11,509	7,827
FLDN2a	4,577	861	12,165	9,477
FLDN2b	1,219	139	2,378	1,295

Figure 6. Functional and stage-dependent PU.1 binding in early T cell development.

A. Stage specific and non-stage-specific (shared) PU.1 binding sites: stage specific binding defined by ≥4x difference in signal densities between FLDN1 and FLDN2b.

B. Differential expression of PU.1 binding linked genes. Among 13,335 PU.1 binding linked genes in DN cells, 1,045 genes were differentially expressed (≥2x change in expression), with 502 genes downregulated and 543 upregulated from FLDN1 to FLDN2b. The rest of genes include 7,244 stably expressed genes (<2x change in expression) and 5,046 silent genes (<1 RPKM in both stages). Based on the type(s) of associated PU.1 binding sites (as in 6A), differentially expressed genes were divided into three subgroups: genes linked only to FLDN1 specific sites (Loss of PU.1 binding in FLDN2b, blue), those with either shared or FLDN2b specific sites or both (Retaining of PU.1 binding in FLDN2b, red), and genes that rapidly lose PU.1 binding in some sites while retain binding in other sites (Mixed, green).

C. Relationship between PU.1 occupancy patterns and changes in mRNA expression between FLDN2b and FLDN1. Plots are cumulative distributions of expression changes for three groups of genes depicted in B. The number of genes in each group and p values (K-S tests for comparisons with "Mixed") are indicated next to the plots.

D & E. Developmentally distinct PU.1 binding patterns at the *Tal1* (D) and *Il7ra* (E) loci in FLDN1, FLDN2a, FLDN2b and $E2A^{-/-}$ pre-pro B cells, compared with H3K4me2, H3K27me3 and mRNA in all five immature T-populations. Note several PU.1 bound sites with an occupancy pattern similar to *Il7r* sites in the silent *Capsl* downstream.

F. Distribution of PU.1 sites in individual stages. In each stage, PU.1 binding sites were divided into four subgroups based on the location (promoter-proximal and -distal) and expression level of binding sites associated genes (expressed and silent).

G. Location of PU.1 sites in potential target genes according to expression pattern. Genes with PU.1 binding sites from clusters with different developmental expression patterns

(Figure S5) were classified according to the location of PU.1 binding, within 1 kb of the TSS (proximal) or in non-TSS sites (distal). Point symbols (left axis) show the percentages of genes in a cluster with PU.1 binding to each region type. Bar graphs show the ratio of the number of genes in a cluster with TSS sites to the number of genes in the cluster with distal sites, to correct for accessibility (right axis). Colors of bars denote similarity of expression pattern of each cluster to endogenous PU.1 expression (most similar: blue; inverse: red). See Figures 2D, S5, &Table S6.

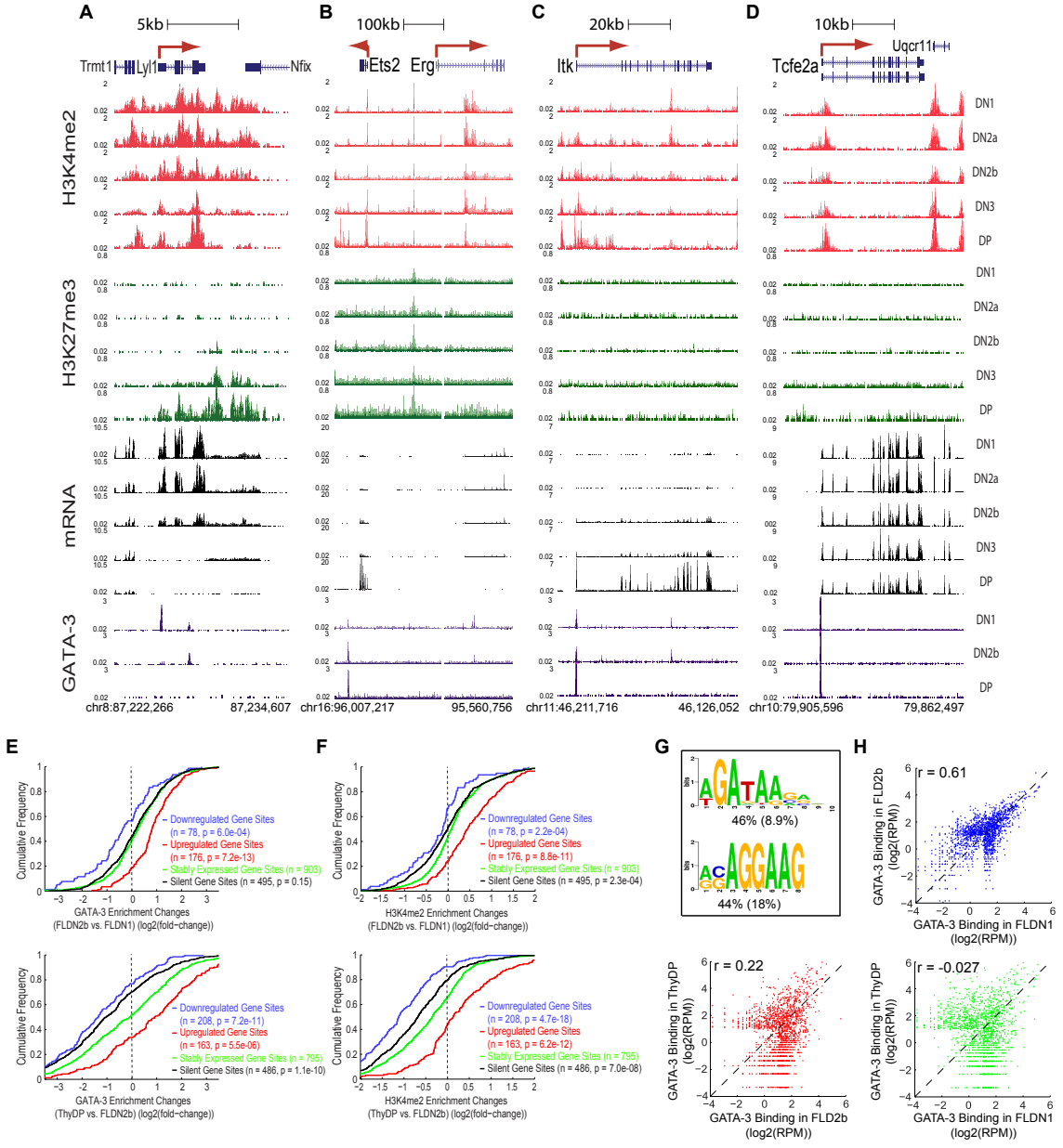


Figure 7. Developmental plasticity of GATA-3 DNA binding and associated epigenetic marking.

A – D. Stage-specific GATA-3 binding (brown) in *Lyl1*, *Ets2-Erg*, *Itk* and *Tcfe2a* loci of FLDN1, FLDN2b and ThyDP cells, shown with binding associated H3K4me2 (red) and

H3K27me3 (green) enrichment and mRNA (black) expression in all five immature T-populations.

E. Cumulative distributions of changes in GATA-3 occupancy between FLDN2b and FLDN1 (top) and between ThyDP and FLDN2b (bottom), for genes differentially regulated across the same intervals. GATA-3 binding sites (see panel H) were divided into four subgroups, based on linkage to downregulated genes (blue), upregulated genes (red), stably expressed genes (<2x change in expression, green) and silent gene sites (<1 RPKM in both stages, black). P values from K-S tests between stably expressed gene sites and each of the other three subgroups and the number of sites in each group are shown.

F. Cumulative distributions of changes in H3K4me2 marks associated with GATA-3 binding between FLDN2b and FLDN1 (top) and between ThyDP and FLDN2b (bottom) stages. H3K4me2 signal densities were calculated within ±1 kb of the summit of a given GATA-3 bound region (depicted in Figure 7H). K-S test was performed between stably expressed gene sites and each of the other three subgroups as described in 7E. The number of sites in each group and p value for each comparison are indicated in parentheses.

G. Most highly enriched sequence motifs in GATA-3 binding regions (see panel H). The percentages of regions containing at least one instance of each motif are indicated beneath each sequence logo, with the expected frequency of the motif in random regions in parentheses.

H. Scatter plots depicting the comparisons in GATA-3 binding between FLDN1, FLDN2b and ThyDP. Pearson correlation coefficients are shown for each comparison.

Supplementary Material Index

Materials and Experimental Procedures

Cell Culture

Fetal liver (FL) cells from embryonic day 13.5 to 14 (E13.5-E14) C57BL/6 mouse embryos were first depleted of Gr-1$^+$, F4/80$^+$, Ter119$^+$, and CD19$^+$ ("Lin cocktail 1") cells using streptavidin-coupled magnetic microbeads (Miltenyi Biotec) against biotin-conjugated antibodies. Lin$^-$c-Kit$^+$CD27$^+$ multi-lineage precursors were then sorted from lineage-depleted FL cells by FACS and co-cultured with OP9-DL1 stromal cells as described previously (Taghon et al., 2005). 50~100 × 10^3 Lin$^-$c-Kit$^+$CD27$^+$ FL cells were plated on OP9-DL1 monolayers in 10 cm plates in the presence of 5 ng/mL Flt3-L and 5 ng/mL IL-7 (both from Peprotech). After 4.5 d of culture, half of the cells were harvested and sorted to isolate DN1 (FLDN1, Lin$^-$c-Kit^{hi}CD45$^+$CD44$^+$CD25$^-$) and DN2a cells (FLDN2a, Lin$^-$c-Kit^{hi}CD45$^+$CD44$^+$CD25$^+$) (using "Lin cocktail 2" = antibodies to Ter-119, CD19, F4/80, Gr-1, NK1.1, CD122, CD11c, TCRγδ, TCRβ, CD3ε, CD8α). After 8.5 d of culture, the rest of the cells were harvested and sorted for FLDN2a and DN2b (FLDN2b, Lin$^-$c-Kit^{int}CD45$^+$CD44^{int}CD25$^+$). The FLDN2a samples used for analysis were each pools of day 4.5 and day 8.5 DN2a cells in approximately 2:1 ratio. For subsets from adult (4-6 weeks old) thymus, wild-type C57BL/6 mouse thymi were first depleted with antibodies to Ter-119, CD19, F4/80, Gr-1, NK1.1, CD122, CD11b, CD11c, TCRγδ, TCRβ, CD3ε, CD4 and CD8α ("Lin cocktail 3"). Thymic DN3 (ThyDN3) cells were then sorted from lineage-depleted thymocytes as Lin$^-$c-Kit$^-$CD44$^-$CD25$^+$. Finally, to prepare ThyDP populations free of contaminating cells in early stages of TCR-dependent positive selection, while maintaining viability of these fragile cells, ThyDP cells were collected from TCRα$^{-/-}$ thymi (4-6 weeks old, The Jackson Laboratory, B6.129S2-

Tcra^{tm1Mom}/J) by simple streptavidin-coupled magnetic microbead depletion with biotinylated antibodies against Ter-119, CD19, F4/80, Gr-1, NK1.1, CD122, CD11b, CD11c, c-Kit, CD44 and CD25. After this depletion, 90~95% of cells were $CD4^+CD8^+$.

Antibodies used were from eBioscience and Biolegend, including anti-CD4 (GK1.5; biotin), anti-CD8α (53-6.7; biotin), anti-CD11b (M1/70; biotin), anti-CD11c (N418; biotin), anti-CD19 (eBio1D3; biotin), anti-CD122 (5H4; biotin), anti-Gr1 (RB6-8C5; biotin), anti-F4/80 (BM8; biotin), anti-TCRβ (H57-597; biotin), anti-TCRγδ (eBioGL3; biotin), anti-NK1.1 (PK136; biotin), anti-Ter119 (Ter-119; biotin), anti-c-Kit (2B8; PE, APC, biotin), anti-CD27 (LG.7F9; APC), anti-CD25 (PC61.5; APC-Alexa 750, APC-Alexa 780, biotin), anti-CD44 (1M7; Pacific Blue, eFluor 450, biotin), anti-CD45 (30-F11; APC), anti-CD3ε (145-2c11, PerCp-cy5.5). For detection of biotinylated antibodies, streptavidin–PerCp-Cy5.5 was used.

Lineage Commitment Assay

Samples of 25 FLDN1, FLDN2a or FLDN2b cells were each sorted into 96 well plates coated with either OP9-DL1 or OP9-Mig (control) monolayers in the presence of 5 ng/mL Flt3-L, 5 ng/mL IL-7, 5 ng/mL SCF, and either 200 units/mL IL2 or 5 ng/mL MCSF. After 7 days of co-culture, cells were harvested and subjected to FACS analysis, using NK1.1 and/or Dx5 as a marker for NK progeny, CD19 as a marker for B-cell progeny, and CD11b/CD11c as markers for myeloid and dendritic progeny. Results are shown in Table S1.

Chromatin Immunoprecipitation

Each histone modification ChIP was generated using 5 million cells and 20 μg of each of the following antibodies: H3K(9,14)Ac (Millipore 06-599), H3K4me2 (Millipore

07-030) and H3K27me3 (Millipore 07-449). PU.1 ChIP was generated using 7.5~10 million cells and 5μg anti-PU.1 (Santa Cruz sc-352). GATA-3 ChIP was generated using 15~20 million cells and 2.5 μg anti-GATA-3 (Santa Cruz sc-268). In addition to FLDN1, FLDN2a and 2b samples, we performed PU.1 ChIP on a ThyDP sample, which naturally lacks PU.1 expression, as a negative control. ChIP was carried out essentially following the manufacturer's protocol (http://www.millipore.com/userguides/tech1/mcproto407), with the exceptions that protein A or G agarose beads were replaced by anti-rabbit or anti-mouse secondary antibody-coupled magnetic beads (Dynabeads, pre-washed with 1xPBS/0.5%BSA, 1 μg Ig/10 μL beads), and the pre-clear step was omitted. Independent biological replicates were generated for histone modification ChIP of FLDN1, FLDN2b, ThyDN3 and ThyDP. Different batches of histone modification antibodies (Millipore 06-599, Millipore 07-030, Millipore 07-449) were used among biological replicates. Purified ChIP DNA was subjected to end repairing, adaptor ligation, PCR amplification, size selection by gel electrophoresis (200~300 bp, insert plus adaptor and PCR primer sequences) and a second round of PCR amplification to generate each ChIP DNA library as described (Johnson et al. 2007) (Illumina ChIP-seq sample preparation kit #IP-102-1001).

mRNA Purification and cDNA Library Building

Total RNA was extracted from 2.5~20 million cells using Trizol (Invitrogen), and then subjected to two rounds of selection using Oligo-dT coupled magnetic beads (Dynabeads) according to the manufacturer's protocol. About 100 ng polyadenylated mRNA per sample was obtained after double selection. Independent biological replicates were generated for all five populations (triplicates for FLDN2b). cDNA library building

was performed as described (Mortazavi et al., 2008). Briefly, RNA was fragmented to an average length of 200 bp by Mg^{2+}-catalyzed hydrolysis and then converted into cDNA by random priming. cDNA was then subjected to end repairing, adaptor ligation ligation (using Illumina ChIP-seq sample preparation kit #IP-102-1001), size selection and one round of PCR amplification.

High-Throughput Sequencing

Each ChIP DNA library or cDNA library was sequenced with the Illumina Genome Analyzer II and IIX following the manufacturer's protocols (http://www.illumina.com) (Johnson et al., 2007, Mortazavi et al., 2008).

RNA-seq Data Analysis

Sequence reads from each cDNA library (38 bp, single-read) were trimmed to 32 bp long and mapped onto the mouse genome build NCBI37/mm9 using Bowtie (bowtie-0.12.1, http://bowtie-bio.sourceforge.net/index.shtml) with setting '-v 2 -k 11 -m 10 -t --best --strata'. The mappable data were then processed by the ERANGE v. 3.3 RNA-seq analysis program (Mortazavi et al, 2008). Assuming total transcriptional activity is comparable between different cell types, the obtained data (data units in RPKM, reads per kilobase exon model per million mapped reads) were first $\log_2$ transformed and linearly normalized between individual samples, then averaged among biological replicates or triplicates. At the same time, in order to find genes that were changed in expression between two populations to a statistically significant degree, ERANGE processed data were analyzed by the Bioconductor DEGseq program ((Wang et al., 2010) http://www.bioconductor.org/packages/2.6/bioc/html/DEGseq.html) (data units in RPM, reads per million mapped reads, method = "MARS", $p < 0.001$) (Figure 1A). This

analysis yielded 3,697 DEGseq positive genes that had more than a twofold change in RNA-seq reads (after normalization and averaging), either between any two successive stages or between FLDN1 and ThyDP, and these were defined as differentially expressed genes. To identify differentially regulated transcription factors, we did Gene Ontology analysis of this set with key term "DNA-dependent regulation of transcription" (GO:0006350), and the resulting list was then hand curated to remove cell surface receptors, cytokines, and other genes of questionable categorization. The final list used for alignment against our DEGseq set is presented as Table S3 part A.

Hierarchical clustering: To determine the overall tendencies of change in gene expression and the connection between different populations, we hierarchically clustered RNA-seq data of these 3,697 selected genes from all 11 samples (using normalized data, biological replicates and triplicates were treated independently, Figure 1B). Hierarchical clustering was performed along both dimensions with sample similarities clustered first, and then genes. Euclidean distance and complete linkage were used (MATLAB 7.10.0). Separately, two-dimensional hierarchical clustering was also performed on 379 differentially expressed transcription factors (Figure 1C).

K-means clustering: To profile and categorize the behavior of clusters of similarly regulated genes during early T cell development, we first normalized individual mRNA data for the 3,697 selected genes by the corresponding geometric mean of five stages, and then performed K-means clustering analysis on the results after $\log_2$ transformation (Figure S5). K was set at 25 and squared Euclidean distance was used (MATLAB 7.10.0).

ChIP-seq Data Analysis

Histone Modification ChIP-seq: DNA sequence reads from each ChIP-seq library (single-read) were trimmed and mapped onto NCBI37/mm9 using the same setting as for RNA-seq data, and uniquely mapped reads were used for further analysis. The data were processed by the ERANGE v. 3.3 findall peak finder (Johnson et al. 2007) to identify enriched genomic regions. We used a stringent setting of '-spacing 100 -minimum 4 -ratio 4 -minPeak 0.5 -shift learn' for H3Ac and H3K4me2 ChIP-seq data, and a relatively less stringent setting of '-notrim -nodirectionality -spacing 100 -minimum 2 -ratio 4 -minPeak 0.25 -shift learn' for H3K27me3. The sequence data of the input DNA from the same cell type were used as background control. Since on average the total amount of mappable DNA reads of each H3K27me3 ChIP-seq data was about two times of that of each H3Ac and H3K4me2 ChIP-seq data, the minimum total DNA reads for called regions were comparable for all three histone modifications (that is, about minimum 60 to 80 enriched DNA reads per region).

All called regions (from all 27 samples) were pooled and merged if overlapping. Only resulting regions of at least 200 bp were considered for further analysis. This conservative approach treats any local change in peak height or spreading of histone modification as effects on a single region, thus providing a minimum estimate of the number of centers of regulatory change. Thus, for example, the change in shape factor of H3K4me2 commonly observed at active promoters is not considered to change the number of marked regions. We considered the positive regions overlapping ±1 kb from the TSSs of UCSC known genes (mm9, NCBI v.37) as promoter-proximal regions, and the rest as promoter-distal regions. Individual regions were then assigned to the nearest

genes using ERANGE (200 kb as the maximum radius). Signal densities (number of DNA reads) were calculated using ERANGE v. 3.3 regionCounts, for each region of every histone modification dataset. For global histone modification status of promoter regions, we expanded every transcriptional starting site (TSS) of UCSC known genes to a window of ±1 kb, and calculated signal densities of each TSS regions using ERANGE. Assuming that total DNA enrichment of the same histone marker is comparable among different cell types, we linearly normalized the read number (after $\log_2$ transformation) between samples from the same histone marker (i.e., based on slopes of correlation plots in Figure S1B). The mean for biological replicates was used for analysis. Since our RNA-seq data cannot accurately distinguish among isoforms, for genes that have multiple alternative promoters we selected one promoter that had the highest H3K4me2 level (or H3Ac if all had the same level of H3K4me2). Regions (both distal and promoter regions) that had more than 4 RPM in either H3Ac or H3K4me2, or more than 2 RPM in H3K27me3 were considered as positive for the particular histone modification(s) (Figures 2, S2). The processed data were plotted and visualized in MATLAB.

All RNA-seq and ChIP-seq sequencing tracks were generated in WIG file format and uploaded onto the UCSC genome browser for visualization. Publicly available data used in this study (Lin et al., 2010; Heinz et al., 2010; Wei et al., 2009) were downloaded as raw sequence data (http://www.ncbi.nlm.nih.gov/geo) and remapped onto NCBI37/mm9 using the same settings.

PU.1 and GATA-3 ChIP-seq: Since PU.1 ChIP enriched genomic regions were in general narrower than histone modification enriched regions, we used a setting of "-spacing 50 -minimum 2 -ratio 4 -minPeak 0.5 -shift learn -listPeak" for the ERANGE

findall peak finder. The sequence data of the input DNA from the same cell type was used as background control. Publicly available PU.1 ChIP-seq and input data from $E2A^{-/-}$ pre-pro B cells, mature B cells and macrophages (Heinz et al., 2010) were downloaded as raw sequence data (http://www.ncbi.nlm.nih.gov/geo) and remapped using the same setting.

Called regions were pooled and merged if overlapping from each pair-wise or three-way comparison (from $E2A^{-/-}$ pre-pro B vs. FLDN1, B cell vs. FLDN1, macrophage vs. FLDN1, or FLDN1 vs. FLDN2a vs. FLDN2b). Individual regions were calculated for PU.1 enriched signal densities and then assigned to the nearest genes (200 kb as the radius) using ERANGE. We next aligned the summits of all positive regions and calculated histone modification signal densities in a window of ±1 kb. All histone modification data were linearly normalized (using the parameters generated from global histone modification analysis). The mean for biological replicates was used for analysis. Scatter plots were generated and visualized in MATLAB.

To compare differential PU.1 binding with associated differential gene expression and H3K4me2 enrichment during early T cell development, we divided PU.1 binding linked genes into four subgroups: upregulated and downregulated genes (selected from the differentially expressed genes group and having more than 2-fold change in expression from FLDN1 to FLDN2b; see "RNA-seq Data Analysis"), stably expressed genes (less than 2-fold change in expression between FLDN1 and FLDN2b), and silent genes (<1 RPKM in both stages). The changes in PU.1 occupancy and in H3K4me2 enrichment (within ±1 kb of binding summits) between FLDN2b and FLDN1 were calculated and plotted separately as cumulative distribution for each group (Figure S6

A&B). To determine whether PU.1 binding sites linked to upregulated or downregulated genes were more likely differentially bound by PU.1 and enriched by H3K4me2 compared to sites linked to stably expressed genes, two-sample Kolmogorov-Smirnov test was performed between stably expressed gene sites and each of the other three subgroups (Figure S6 A&B).

To visualize histone modifications and degree of PU.1 occupancy surrounding the summits, we further expanded positive regions to a window of ±2 kb, and divided each window into 50 bins (80 bp each). Histone modification and PU.1 enrichment were calculated for each bin using the same method mentioned above. The data obtained were aligned with RNA-seq data of associated genes, and then hierarchically clustered (one dimensional clustering of binding regions; using Euclidean distance and Ward linkage) and visualized as heat maps in MATLAB as shown in Figure S6C-D.

GATA-3 ChIP-seq data was processed similarly to PU.1 ChIP-seq data. To compare our findings with published results, raw sequence data for "CD3lo DP" cell samples (Wei et al., 2011) were downloaded from Gene Expression Omnibus and remapped using the same settings as used for our data, as described for comparing PU.1 results (Heinz et al., 2010) above.

De Novo Motif Analysis

We selected the top 1,000 PU.1 enriched peaks from each of the three subgroups ($E2A^{-/-}$ pre-pro B cells high, shared, and FLDN1 high), and performed MEME analysis on regions ±50 bp from the peaks by ERANGE v. 3.3 using the default setting to generate the position specific frequency matrix (PSFM) representation of the motifs. The PSFMs

were mapped separately back to the three enriched regions subgroups at 85% match (Johnson et al. 2007).

All 1,652 enriched GATA-3 regions (pooled from FLDN1, FLDN2b and ThyDP) were subjected to MEME analysis. Since the consensus sequence motifs of GATA-3 binding sites were shorter than the ones of PU.1 binding sites, the PSFMs were mapped back to the 1,652 enriched GATA-3 regions and 1,652 random genomic regions at 90% match instead. Random genomic regions were comparable to the GATA-3 binding regions in both length and chromosomal distribution.

Supplemental References

David-Fung, E.S., Yui, M.A., Morales, M., Wang, H., Taghon, T., Diamond, R.A., and Rothenberg, E.V. (2006). Progression of regulatory gene expression states in fetal and adult pro-T cell development. Immunol Rev. *209*, 212-236.

Heinzel, K., Benz, C., Martins, V.C., Haidl, I.D., and Bleul, C.C. (2007). Bone marrow-derived hemopoietic precursors commit to the T cell lineage only after arrival in the thymic microenvironment. J. Immunol. *178*, 858-868.

Masuda, K., Kakugawa, K., Nakayama, T., Minato, M., Katsura, Y., and Kawamoto, H. (2007). T cell lineage determination precedes the initiation of TCRb rearrangement. J. Immunol. *179*, 3699-3706.

Wang, L., Feng, Z., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics *26*, 136-138.

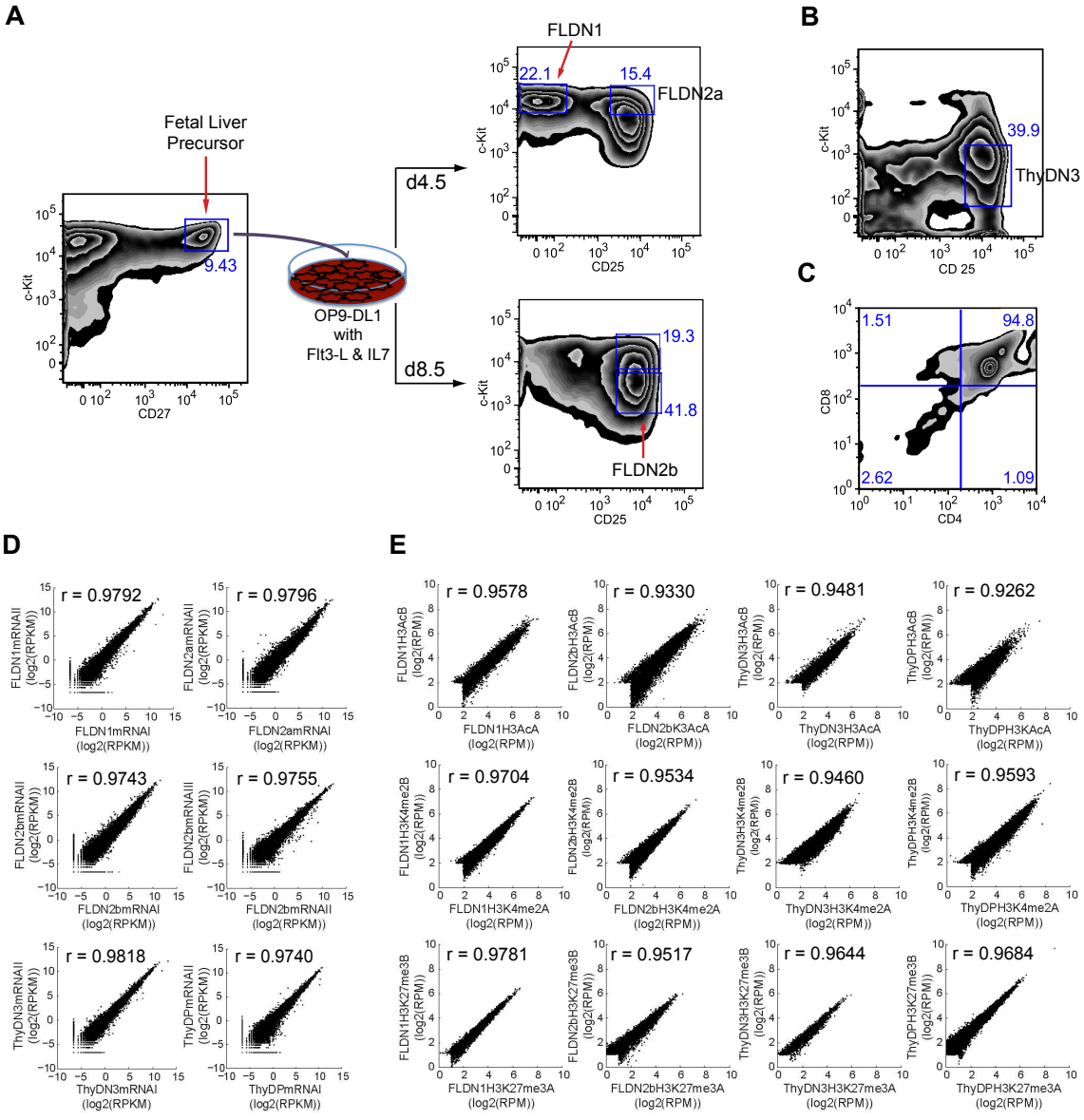


Figure S1. Cell purification and biological reproducibility.

A. Lin$^-$c-Kit$^+$CD27$^+$ fetal liver precursors were sorted from E13.5 to 14 C57BL/6 mouse embryos and co-cultured with OP9-DL1 stromal cells in the presence of IL-7 and Flt3-L. At day 4.5, FLDN1 cells were sorted from the co-culture as Lin$^-$CD45$^+$cKit^{hi}Cd44$^+$ CD25$^-$, FLDN2a as Lin$^-$CD45$^+$cKit^{hi}Cd44$^+$CD25$^+$. At day 8.5, FLDN2b cells were sorted

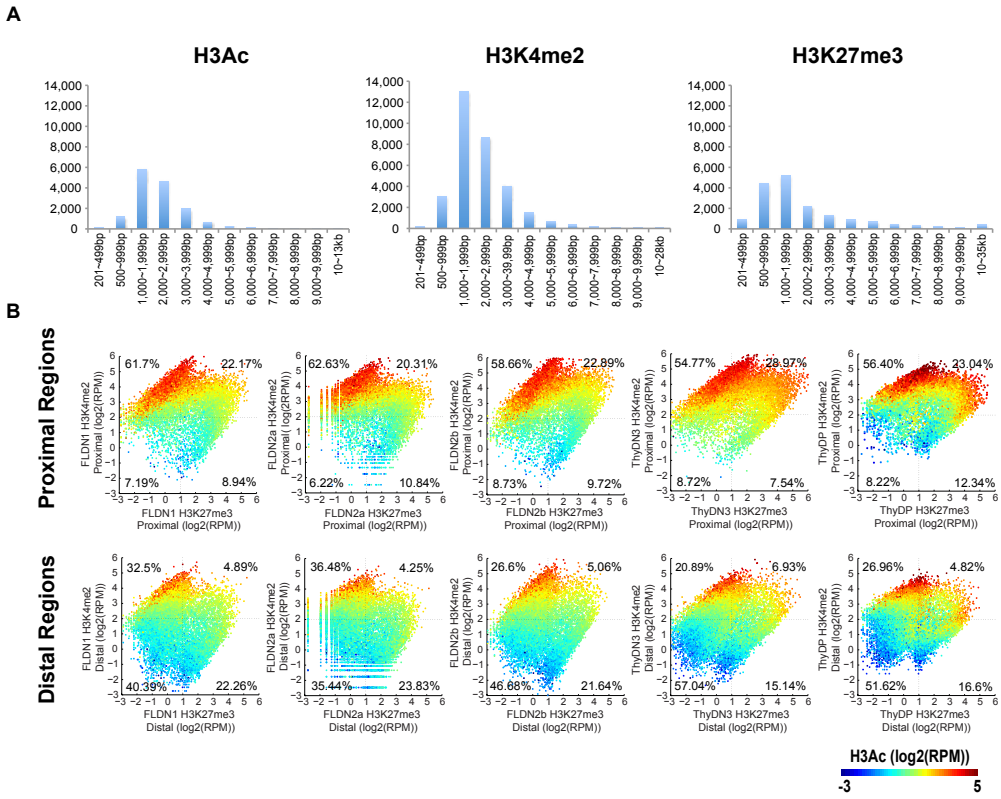
as $Lin^-CD45^+cKit^{lo}Cd44^-CD25^+$, and FLDN2a cells were also collected (see experimental procedure).

B. ThyDN3 cells were sorted from lineage-depleted thymocytes as $Lin^-c\text{-}Kit^-CD44^-CD25^+$ (see experimental procedure).

C. ThyDP cells were collected from B6 background $TCR\alpha^{-/-}$ thymi by depletion with Ter-119, CD19, F4/80, Gr-1, NK1.1, CD11c, c-Kit and CD44. After depletion, 90~95% of cells were $CD4^+CD8^+$(see experimental procedure). Cells lacking TCRα were used in order to enable DP cells to be generated normally, a process dependent only on TCRβ, but leaving them without the capacity to undergo positive selection to any later stages of T-cell development.

D. Global comparisons of RNA-seq between biological duplicates or triplicates (log_2 transformed) for individual cell populations (including FLDN1 sample I vs. II, FLDN2a sample I vs. II, FLDN2b sample I vs. II, FLDN2b sample II vs. III, ThyDN3 sample I vs. II, and ThyDP sample I vs. II). The Pearson's correlation coefficient for each comparison is indicated in the insert.

E. Global comparisons of histone modifications ChIP-seq experiments between biological duplicates (after log_2 transformed) for individual cell populations (including FLDN1 experiment A vs. B, FLDN2b experiment A vs. B, ThyDN3 experiment A vs. B, and ThyDP experiment A vs. B). From top row to bottom are H3Ac, H3K4me2 and H3K27me3. The Pearson's correlation coefficient for each comparison is indicated in the insert.



C

	FLDN1 (9,878*)	FLDN2a (9,531*)	FLDN2b (9,875*)	ThyDN3 (9,705*)	ThyDP (9,501*)
H3KAc+ (≥4 RPM)	8,469 (86%)	8,406 (88%)	8,603 (87%)	8,357 (86%)	7,948 (84%)
H3K4me2+ (≥4 RPM)	9,104 (92%)	8,913 (94%)	9,002 (91%)	8,963 (92%)	8,758 (92%)
H3K27me3+ (≥2 RPM)	366 (3.7%)	296 (3.1%)	437 (4.4%)	371 (3.8%)	494 (5.2%)
H3Ac-/ H3K4me2+/ H3K27me3+	137 (1.4%)	105 (1.1%)	106 (1.1%)	92 (0.95%)	121 (1.3%)
H3Ac+/ H3K4me2+/ H3K27me3-	8,261 (84%)	8,234 (86%)	8,274 (84%)	8,094 (83%)	7,612 (80%)
H3Ac+/ H3K4me2+/ H3K27me3+	179 (1.8%)	159 (1.7%)	274 (2.8%)	256 (2.6%)	332 (3.5%)
H3Ac-/ H3K4me2-/ H3K27me3-	695 (7.0%)	573 (6.0%)	761 (7.7%)	712 (7.3%)	699 (7.4%)

* Total number of expressed genes (mRNA ≥ 1 RPKM) in each stage

	FLDN1 (10,983*)	FLDN2a (11,330*)	FLDN2b (10,986*)	ThyDN3 (11,156*)	ThyDP (11,360*)
H3Ac+ (≥4 RPM)	850 (7.7%)	990 (8.7%)	1,006 (9.2%)	1,064 (9.5%)	841 (7.4%)
H3K4me2+ (≥4 RPM)	3,180 (29%)	3,177 (28%)	2,887 (26%)	3,256 (29%)	2,844 (25%)
H3K27me3+ (≥2 RPM)	4,134 (38%)	4,320 (38%)	4,199 (38%)	3,733 (33%)	3,734 (33%)
H3Ac-/ H3K4me2+/ H3K27me3+	1,829 (17%)	1,630 (14%)	1,521 (14%)	1,690 (15%)	1,495 (13%)
H3Ac+/ H3K4me2+/ H3K27me3-	647 (5.9%)	752 (6.6%)	607 (5.5%)	631 (5.7%)	685 (6.0%)
H3Ac+/ H3K4me2+/ H3K27me3+	200 (1.8%)	234 (2.1%)	399 (3.6%)	433 (3.9%)	299 (2.6%)
H3Ac-/ H3K4me2-/ H3K27me3-	5,695 (52%)	5,693 (50%)	5,820 (53%)	6,290 (56%)	6,574 (58%)

* Total number of silent genes (mRNA < 1 RPKM) in each stage

Figure S2. Global distribution of three histone modifications.

A. Length distribution of discrete genomic regions enriched with the specified histone modification in at least one cell population. First, more regions were enriched with H3K4me2 in at least one cell population than H3Ac or H3K27me3. Second, while the majority of enriched regions were within 500~5,000 bp (>95% for H3Ac and H3K4me2;

81.5% for H3K27me3), H3K27me3 enriched regions were relatively more diverse in length, including some regions of much greater extent.

B. Signal density plots of histone modifications at individual promoter-proximal or distal regions in each population. Signal densities were calculated for every discrete genomic region enriched with at least one histone modification in at least one population, then $\log_2$ transformed and normalized (see experimental procedure). X-axis indicates H3K27me3 value and y-axis indicates H3K4me2 value, while the color coding specifies the signal density of H3Ac in each element. In each stage, total regions were separated into four subgroups, based on the enrichment of H3K4me2 (≥4 RPM as positive) and H3K27me3 (≥2 RPM as positive). The percentage of regions for each group is indicated in the insert.

C. Distribution of histone modifications at the promoters with respect to transcriptional activity. Based on the transcriptional activity, total Refseq annotated genes were divided into subgroups of expressed genes (in red table, normalized RNA-seq value ≥ 1 RPKM) and silent genes (in blue table, normalized RNA-seq value < 1 RPKM). Each number indicts total amount of promoters enriched with the specified histone modification or combination of histone modifications. The percentage of promoters enriched with the specified histone modification or combination of histone modifications within the respective subgroup is given in the parenthesis.

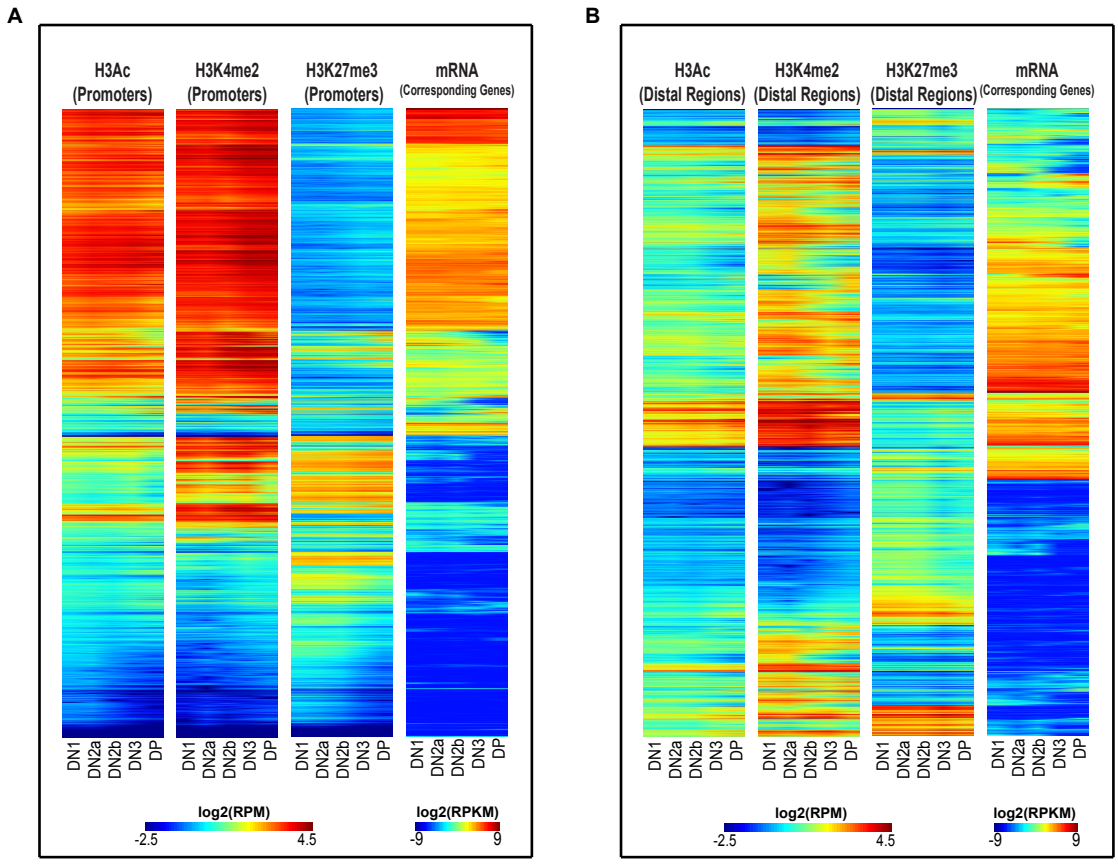


Figure S3. Global profiles of histone modifications and gene expression at both promoter regions and distal regions.

A. The normalized signal densities of histone modifications at the promoter region were aligned with the normalized mRNA data for each of 20,861 Refseq annotated genes, and one-dimensional (along genes) hierarchical clustering was performed using Ward linkage and Euclidean distance.

B. Each distal region was assigned to the closest gene, and one-dimensional hierarchical clustering, along genes, was performed as in Figure S3A.

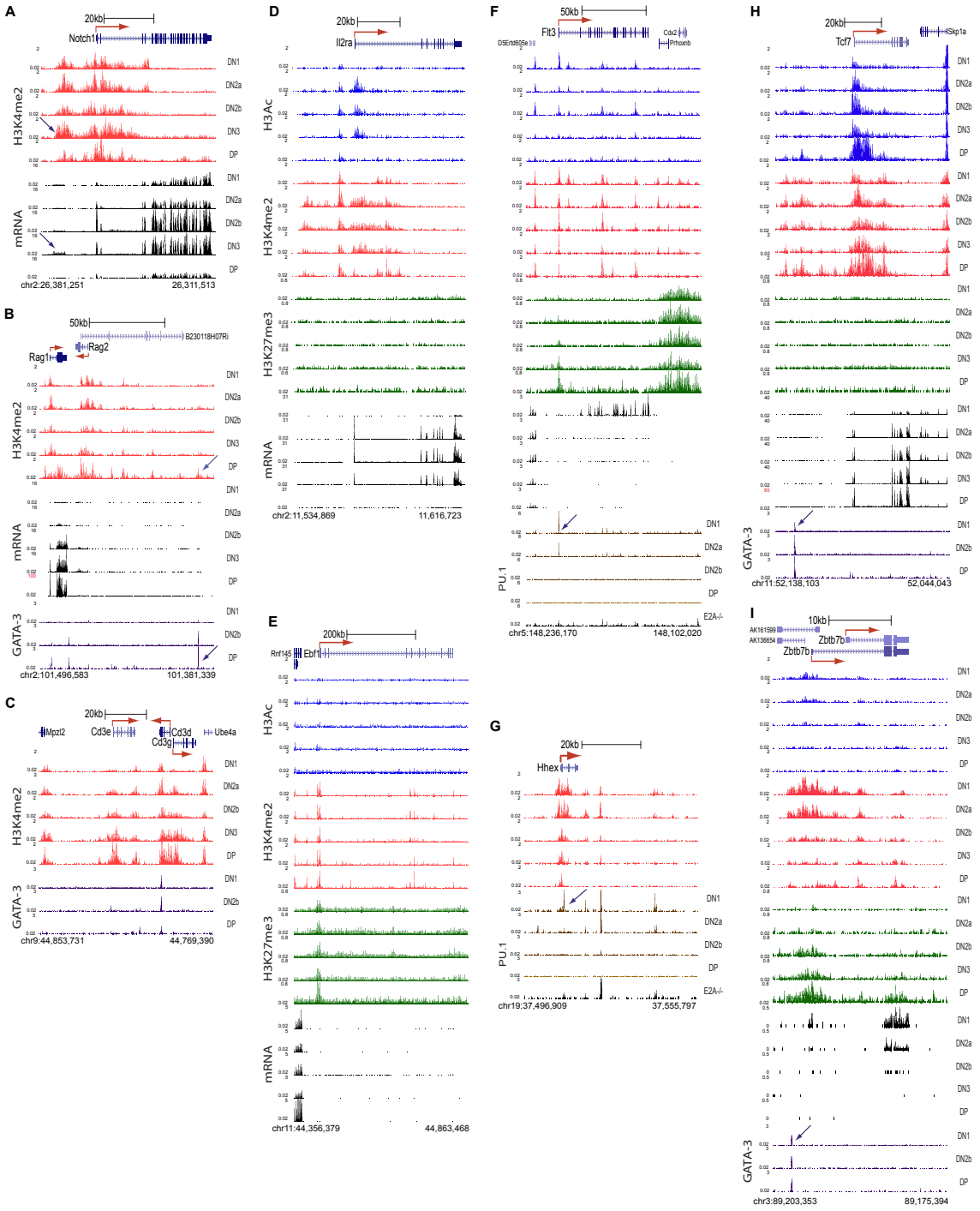


Figure S4. Dynamic histone modifications and transcription factor binding are linked to differentially regulated gene expression.

A. UCSC genome browser tracks depicting H3K4me2 and gene expression at *Notch1* locus. A *Notch1* distal alternative promoter or cis-regulatory element (black arrow, most 5' component of H3K4me2 peak) activity was upregulated at DN2b and DN3 stages.

B. UCSC genome browser tracks depicting GATA-3 binding, together with binding associated H3K4me2 and gene expression, at *Rag1/2* loci. Black arrows indicate the DP specific Rag1/2 enhancer, where GATA-3 occupancy was observed at a low level in DN2b stage and sharply increased in DP stages, as did binding associated H3K4me2. Due to the strong upregulation of gene expression of *Rag1* and *Rag2* from DN3 to DP, note that in this panel the range of RNA-seq signal densities of Rag1 and Rag2 for DP (0.02 to 100 in red) is different from the one for other stages (0.02 to 16).

C. UCSC genome browser tracks showing GATA-3 binding, together with binding associated H3K4me2 at *Cd3e/d/g* loci.

D&E. UCSC genome browser tracks depicting histone modifications and gene expression at *Il2ra* (D) and *Ebf1* (E) loci.

F&G. UCSC genome browser tracks depicting PU.1 binding, together with binding associated histone modification(s) and gene expression (of *Flt3*), at *Flt3* (F) and *Hhex* (G) loci. PU.1 occupancies, as well as binding associated H3K4me2, at the promoters of *Flt3* and *Hhex* (black arrows) decreased from DN1 to DN2a stages, and completely disappeared in DN2b, in parallel with the gene expression pattern of two genes (*Hhex* expression and histone modification patterns are depicted in Figure 4E).

H&I. UCSC genome browser tracks depicting GATA-3 binding, together with binding associated histone modifications and gene expression, at *Tcf7* (H) and *Zbtb7b* (I) loci. GATA-3 occupancy at an upstream distal region of *Tcf7* (black arrow) increased from

DN1 to DN2b, as gene expression of *Tcf7* was upregulated from DN1 to DN2b. Although repressed by H3K27me3 in DN2b and DP (pre-positive selection), *Zbtb7b* upstream distal region was constantly bound by GATA-3 (black arrow) in both stages. Note that the range of RNA-seq signal densities of Tcf7 for DP (0.02 to 80 in red) is different from the one for other stages (0.02 to 40).

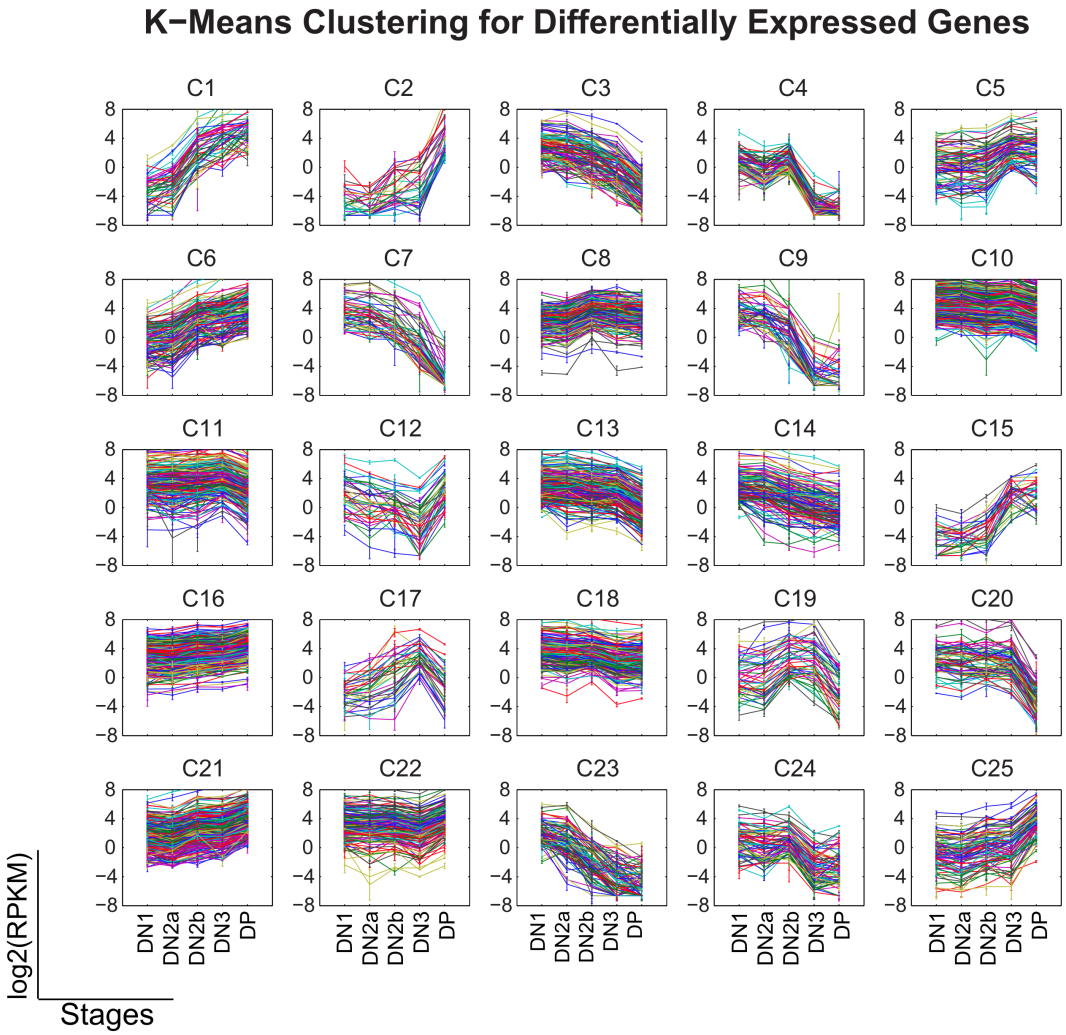


Figure S5. K−means clustering for differentially expressed genes.

3,697 differentially expressed genes were subjected to K-means clustering analysis that inferred to 25 differentially expressed patterns (see Experimental Procedures). Error bar represents the standard deviation of biological replicates or triplicates of individual genes at each stage. The genes in each cluster are listed in order in Table S4A.

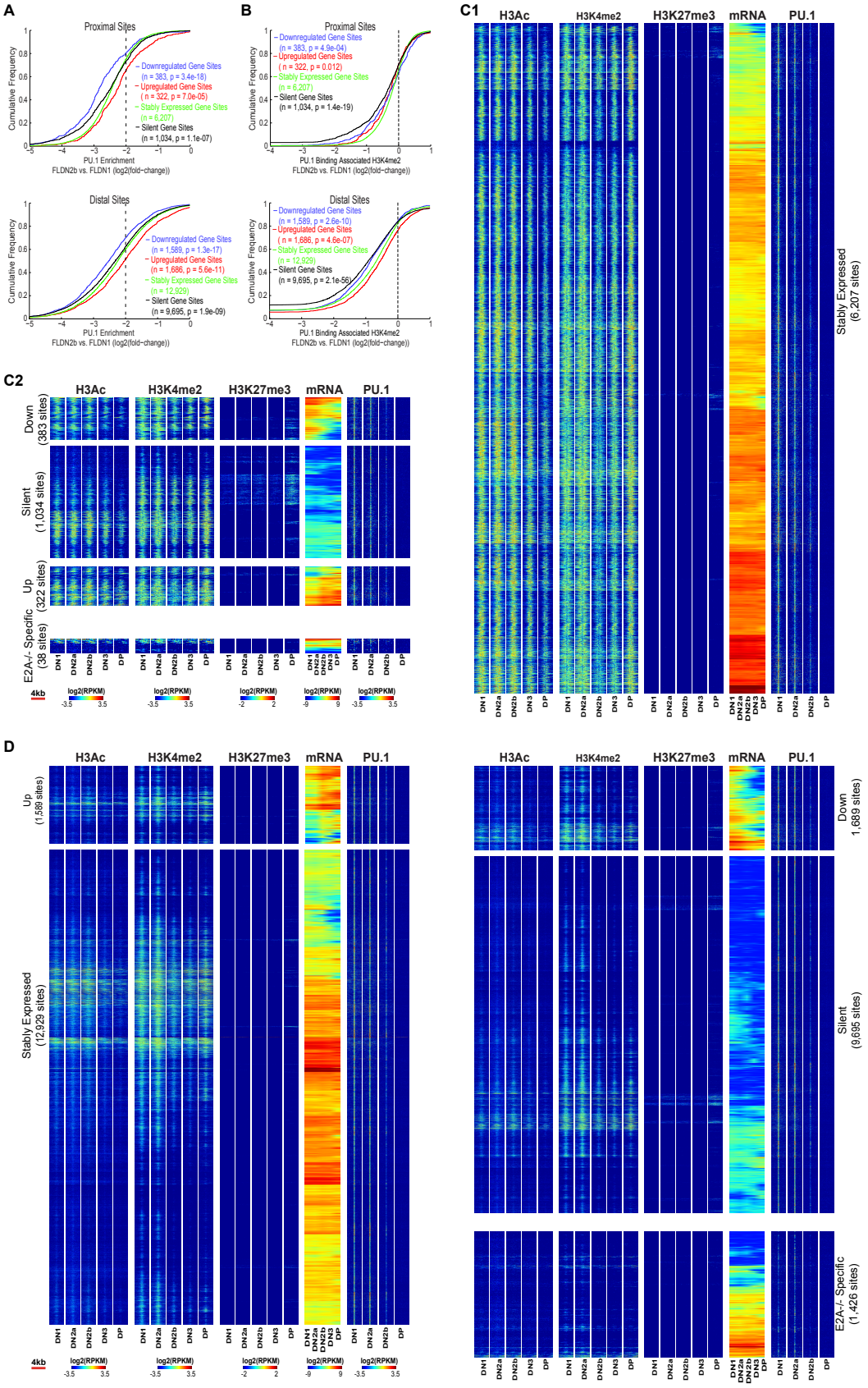

A
Proximal Sites
Cumulative Frequency
Downregulated Gene Sites (n = 383, p = 3.4e-18)
Upregulated Gene Sites (n = 322, p = 7.0e-05)
Stably Expressed Gene Sites (n = 6,207)
Silent Gene Sites (n = 1,034, p = 1.1e-07)
PU.1 Enrichment FLDN2b vs. FLDN1 (log2(fold-change))
Distal Sites
Downregulated Gene Sites (n = 1,589, p = 1.3e-17)
Upregulated Gene Sites (n = 1,686, p = 5.6e-11)
Stably Expressed Gene Sites (n = 12,929)
Silent Gene Sites (n = 9,695, p = 1.9e-09)
B
Proximal Sites
Downregulated Gene Sites (n = 383, p = 4.9e-04)
Upregulated Gene Sites (n = 322, p = 0.012)
Stably Expressed Gene Sites (n = 6,207)
Silent Gene Sites (n = 1,034, p = 1.4e-19)
PU.1 Binding Associated H3K4me2 FLDN2b vs. FLDN1 (log2(fold-change))
Distal Sites
Downregulated Gene Sites (n = 1,589, p = 2.6e-10)
Upregulated Gene Sites (n = 1,686, p = 4.6e-07)
Stably Expressed Gene Sites (n = 12,929)
Silent Gene Sites (n = 9,695, p = 2.1e-56)
C1
H3Ac
H3K4me2
H3K27me3
mRNA
PU.1
Stably Expressed (6,207 sites)
C2
Down (383 sites)
Silent (1,034 sites)
Up (322 sites)
E2A-/- Specific (38 sites)
DN1
DN2a
DN2b
DN3
DP
4kb
log2(RPKM)
D
Up (1,588 sites)
Stably Expressed (12,929 sites)
Down 1,689 sites)
Silent (9,695 sites)
E2A-/- Specific (1,426 sites)

Figure S6. PU.1 occupancy associated with epigenetic modifications and gene expression during early T cell development.

A. Cumulative distributions of changes in PU.1 occupancy between FLDN2b and FLDN1 among promoter-proximal sites (top) and promoter-distal sites (bottom). The total PU.1 binding sites in DN cells (as depicted in Figure 5B) were separated in two groups, promoter-proximal sites (≤1 kb from nearby TSS) and promoter-distal sites (>1 kb from nearby TSS). Based on the expression patterns of binding linked genes from FLDN1 to FLDN2b, each group was then divided into 4 subgroups: downregulated gene sites (linked to genes ≥2x downregulated, blue), upregulated gene sites (linked to genes ≥2x upregulated, red), stably expressed gene sites (linked to genes with <2x change in expression, green) and silent gene sites (linked to genes with <1 RPKM in both stages, black). Binding sites linked to downreulated genes, both distal and proximal, tend to lose PU.1 occupancy more rapidly than other groups of sites. K-S test was performed between stably expressed gene sites and each of the other three subgroups. The number of sites in each group and p value for each comparison are indicated in parentheses.

B. Cumulative distributions of changes in PU.1 binding associated H3K4me2 between FLDN2b and FLDN1 among promoter-proximal sites (top) and promoter-distal sites (bottom). H3K4me2 signal densities were calculated within ±1 kb of the summit of a given PU.1 bound region (as depicted in Figure 5B). K-S test was performed between stably expressed gene sites and each of the other three subgroups as described in S6A. The number of sites in each group and p value for each comparison are indicated in parentheses.

C1&2. Heat maps of PU.1 occupancy and distribution of H3Ac, H3K4me2 and H3K27me3 surrounding ±2 kb of the binding summits for promoter-proximal regions (see experimental procedures). The PU.1 binding sites in DN cells were divided into four subgroups based on linked gene expression patterns as described in S6A. As comparison, a separate group of heat maps for promoter-proximal regions that were selected for much greater PU.1 binding in $E2A^{-/-}$ cells than in early T-lineage cells (as in Figure 5) are included.

D. Heat maps of PU.1 occupancy at promoter-distal regions are shown as in C1&2, correlated with distribution of H3Ac, H3K4me2 and H3K27me3 surrounding ±2 kb of the binding summits (see Experimental Procedures). RNA expression heat maps refer to the nearest linked gene and a single gene can be represented by more than one PU.1-bound distal region. The PU.1 binding sites in DN cells were divided into four subgroups as described in panel A. As comparison, a separate group of heat maps for promoter-distal regions that were specific for PU.1 binding in $E2A^{-/-}$ cells (as in Figure 5) are included.

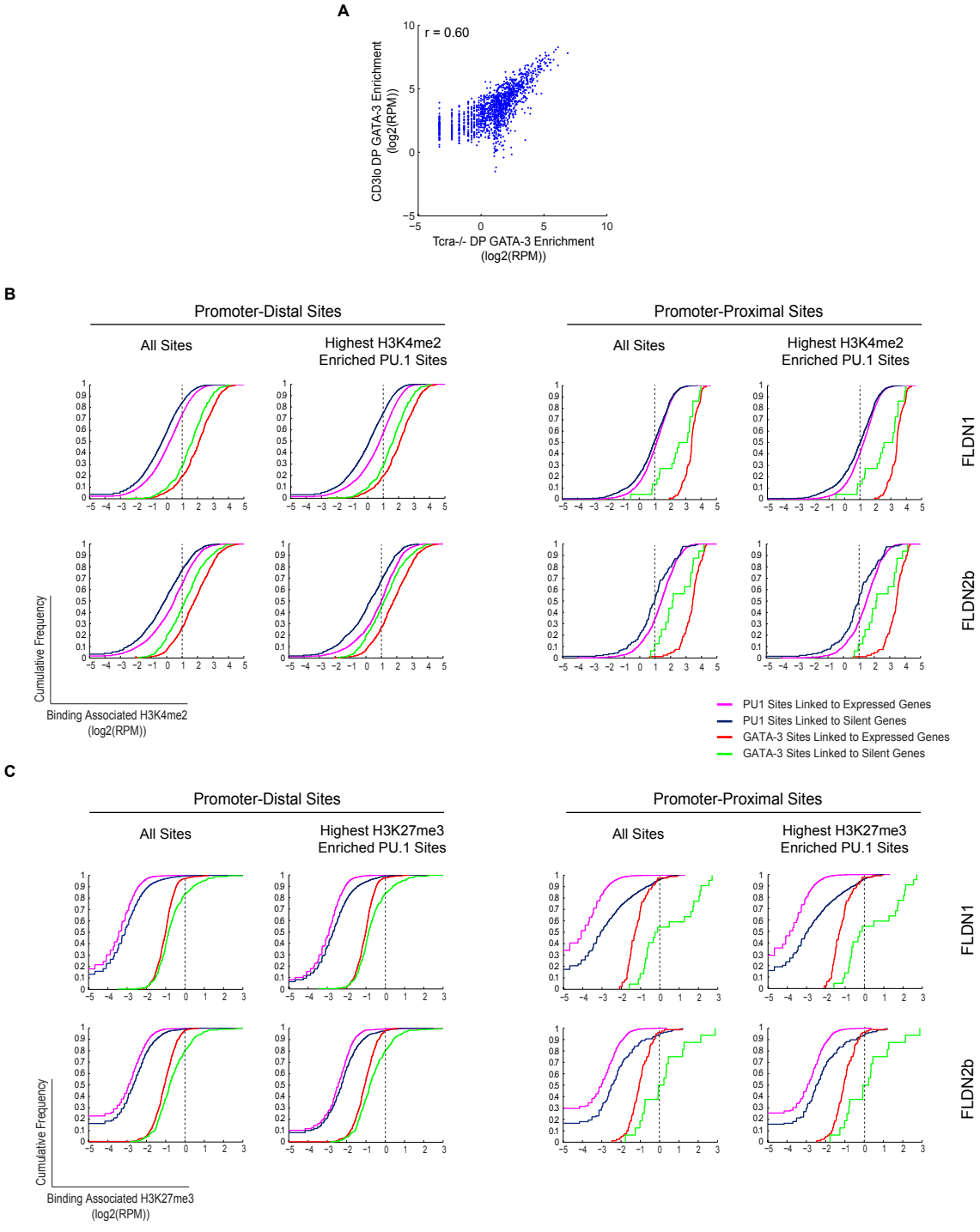


Figure S7. Characterization of sites of GATA-3 binding in early developing T cells.

A. Comparisons of GATA-3 DNA binding site distributions in ThyDP (Tcrα-/- DP, using Santa Cruz sc-268 antibody) and $CD3^{lo}$ DP (using BD biosciences #558686 antibody) (Wei et al. 2011). Pearson's correlation coefficient (r) is indicated.

B. Cumulative distributions of H3K4me2 enrichment over genomic regions within ±1 kb of PU.1 binding sites and GATA-3 binding sites in FLDN1 (top panels) and FLDN2b (bottom panels) cells. In each stage, positive binding sites (≥2 RPM of PU.1 or GATA-3 enrichment, see table S5 or S7 respectively) were divided into two groups, promoter-distal sites and promoter-proximal sites. Each group of binding sites was further divided into two subgroups based on the expression level of binding associated genes (expressed and silent). Since PU.1 tends to bind at multiple sites of a single gene locus, it is possible that binding sites with low or no histone modifications of a particular gene locus are nonfunctional. PU.1 binding sites with the highest H3K4me2 enrichment at each gene locus were selected (sites in promoter-distal regions and promoter-proximal regions were selected separately), and plotted accordingly.

C. Cumulative distributions of H3K27me3 enrichment over genomic regions within ±1 kb of PU.1 binding sites and GATA-3 binding sites in FLDN1 (top panels) and FLDN2b (bottom panels) cells. In each stage, positive binding sites were divided similarly as in Figure S7B.

Supplementary Tables 1-7

Table S1. Lineage commitment status measurements for FLDN cells differentiated *in vitro* (see supplementary files).

Table S2. Genome-wide quantitation of histone modifications and gene expression (see supplementary files).

Table S3. Dynamic expression pattern of transcription factors during T-lineage commitment (see supplementary files).

Table S4. K-means clustering for differentially expressed genes and histone marking and transcriptome profiles of hematopoiesis genes (see supplementary files).

Table S5. Genome-wide PU.1 binding from FLDN1 to FLDN2b (see supplementary files).

Table S6: Distribution of PU.1 binding sites among differentially expressed gene clusters (see supplementary files).

Table S7. Genome-wide GATA-3 binding in FLDN1, FLDN2b and ThyDP (see supplementary files).

Chapter 3

A Gene Network Involving Bcl11b, Notch1 and GATA-3

Regulates T-cell Lineage Specification and Commitment

Abstract

Bcl11b is crucial for the proper T-cell lineage commitment. Bcl11b deletion leads to a developmental block at uncommitted DN2a stage with enhanced NK potential and retained stem cell/progenitor-linked properties. Using global gene expression profiling, we show that Bcl11b-deficient pro-T cells are able to initiate T-cell specification but fail to complete the whole process, and Bcl11b is essential to silence progenitor- and alternative lineage-associated gene program. Our data reveal that Notch1 and GATA-3 are dysregulated in the absence of Bcl11b, with Notch1 being downregulated and GATA-3 being upregulated. Furthermore, several NK gene loci that are upregulated specifically in Bcl11b-deficient DN2a cell are bound by GATA-3. Taken together, our data suggest cooperative interplay between Bcl11b, Notch1 and GATA-3 is required for T-lineage commitment.

INTRODUCTION

In mammalian, multipotent progenitors in bone marrow migrate into thymus and give rise to T cells. The thymus is the major organ where T-cell differentiation is initiated and nourished. It supplies many T-lineage promoting cytokines and ligands, including Notch1 ligand Delta-like 4. The interaction between Notch1 expressed on thymocytes and Delta-like 4 expressed on the thymic epithelium is a pivotal driving force in supporting pro-T cell survival, proliferation, and differentiation during early stages of T-lineage development up to and including the β-selection (Sambandam et al., 2005; Yashiro-Ohtani et al., 2010). Ectopic expression of Notch intracellular domain (NICD) promotes thymic-independent T-cell development (Yashiro-Ohtani et al., 2010). However, Notch-Delta signaling by itself is not sufficient; additional factors, including Myb, Runx1, TCF-1, E protein E2A and HEB, Gfi1, Ikaros, GATA-3 and Bcl11b, are required to coordinate with Notch-Delta signaling in directing T-lineage commitment (Ikawa et al., 2010; Rothenberg et al., 2010; Weber et al., 2011).

Among the aforementioned factors, Bcl11b is one of a few factors that are exclusively expressed in T-lineage of all hematopoietic lineages. It is absent in ETP (DN1), and first turns on at relatively low level in DN2a cells. Bcl11b is one of the very few transcriptional regulators that are strongly upregulated (>50x) during the commitment from DN2a to DN2b, and its expression sustained at fairly stable level onward throughout T-cell development (Li et al., 2010b) (see Chapter 2). The unique expression pattern suggests that Bcl11b is not required for the initiation of T-lineage specification, but, rather, a major player during the commitment process. T-lineage commitment is coupled with eliminating alternative lineage potentials as well as

progenitor-associated stem-ness characteristics. The DN2a to DN2b transition is the first critical checkpoint for T-cell development, when non-T-fate choices and self-renewal potential are terminated (Rothenberg et al., 2010). Studies from our lab and others have shown that conditional deletion of Bcl11b in uncommitted pro-T cells *in vitro* arrests cells at DN2a-like stage with increased alternative lineage potentials — in particular NK potential — that are less sensitive to the Notch1 inhibition (Ikawa et al., 2010; Li et al., 2010a; Li et al., 2010b). Intriguingly, Bcl11b-deficient cells also fail to completely silence stem cell/progenitor-associated program, which, as result, possibly hinders the ability of Bcl11b-deficient cells to further differentiate along αβ T-lineage pathway despite the fact that some T-lineage specification genes appear expressed at compatible levels with those in wild-type DN2b cells (Li et al., 2010a).

To assess the global defect in T-lineage differentiation program caused by Bcl11b deletion, we performed RNA-seq analysis on *in vitro* derived Bcl11b-deficient DN2a-like cells. Consistent with the c-Kit^{hi}CD25^{+} DN2a-like cell surface phenotype, the global gene expression profiles indicated Bcl11b-deficient DN2a cells was more closely related to wild-type DN2a cells than DN2b cells, implying Bcl11b is absolutely required for T-cell precursors to successfully pass the commitment checkpoint. Additionally, we identified two transcription factors essential for T-lineage development — *Notch1* and *Gata3* — that were abnormally regulated in the absence of Bcl11b. Loss of Bcl11b led to upregulation of GATA-3 and downregulation of Notch1. Our data suggest that Bcl11b is an essential connecting node in the regulatory network that controls T-lineage commitment, and removal of Bcl11b causes the collapse of the topology of the network.

RESULTS

Global Expression Profile of Bcl11b-Deficient DN2a-Like Cells

$Bcl11b^{fl/fl}$ prethymic precursors was deleted for Bcl11b using retroviral transduced Cre recombinase, and grown *in vitro* on OP9 stromal cells expressing Notch ligands (DL1 or DL4) in the presence of IL-7 and Flt3-L (Li et al., 2010a). As expected, after 7 to 9 days of culture, while the majority of wild-type control cells (>60%) differentiated into DN2b stage ($Lin^-c\text{-}Kit^{lo/-}CD25^+$), most Bcl11b-deficient cells stayed at DN1-like ($Lin^-c\text{-}Kit^{hi}CD25^-$, ~18%) or DN2a-like ($Lin^-c\text{-}Kit^{hi}CD25^+$, ~69%) stages (Figure 1A). We noticed that while c-Kit expression was linearly correlated with CD44 surface expression in wild-type Lin^- DN cells, CD44 level was no longer coupled with c-Kit surface expression in Bcl11b-deficient Lin^- DN cells; more than half of $c\text{-}Kit^{hi}$ Bcl11b-deficient Lin^- DN cells were $CD44^-$ (Figure 1A). And in agreement with previous studies, deletion of Bcl11b gave the cells survival or growth advantages. Single cell analysis confirmed that, starting with the same number of precursors and under the comparable culture conditions, Bcl11b-deficient cells gave rise to more DN2a-like cells than wild-type controls (Li et al., 2010a) (data not shown).

To identify major changes in gene expression directly or indirectly caused by Bcl11b deletion, RNA-seq was performed for Bcl11b-deficient $Lin^-c\text{-}Kit^{hi}CD25^+$ DN2a-like cells and the wild-type $Lin^-c\text{-}Kit^{lo}CD25^+$ DN2b cells. As a comparison, RNA-seq data from fetal liver precursor derived DN2a cells was also included (see Chapter 2). We used RNA-seq programs ERANGE and DEGseq to evaluate differentially expressed genes between Bcl11b-deficient DN2a cells with either wild-type DN2a or DN2b cells. The analysis generated total 1,746 genes that were statistically significantly changed in

expression between at least one pair of cells, with 900 genes between Bcl11b-deficient DN2a and wild-type DN2a and 1,146 genes between Bcl11b-deficient DN2a and wild-type DN2b ($p < 0.001$ & ≥2x change in expression, Figure 1B). As expected many T-lineage specification regulators that are expressed rather constantly during the commitment in wild-type cells, such as *Tcf7, Runx1/2/3, Ikaros, Myb* and *Tcfe2a*, were not affected by Bcl11b deletion (Li et al., 2010a). And, more genes were upregulated than downregulated in the absence of Bcl11b, consistent with the repressive function of Bcl11b suggested by biochemical studies (Cismasiu et al., 2005). On the other hand, 202 genes, or 10% of differentially expressed genes, were statistically significantly upregulated in Bcl11b-deficient DN2a cells against those in both wild-type DN2a and DN2b cells, whereas only 46 genes were downregulated in Bcl11b-deficient DN2a against both wild-type counterparts (Figure 1B). It is possible that some differences between wild-type DN2a cells and its knockout counterparts were due to the low level of Bcl11b protein present in wild-type DN2a cells. Nevertheless, among the genes upregulated in Bcl11b-deficient cells, some are normally kept silence or expressed at background level throughout early T-cell development; the elevation of these genes cannot be attributed solely to Bcl11b deletion, and additional mechanism(s) are involved (Rothenberg et al., 2010).

While surface markers suggested that Bcl11b-deficient cells were blocked at DN2a stage, the global gene expression profiling showed that, in the absence of Bcl11b, a considerable number of genes departed from their normal expression levels at the DN2a stage, and some were actually expressed at the levels that are considered as DN2b-like, implying that a part of T-lineage specification and commitment program is Bcl11b-

independent (Li et al., 2010a) (see below). Nevertheless, the gene profile of Bcl11b-deficient DN2a-arrested cells resembled to the profile of wild-type DN2a more than that of DN2b, as indicated by the difference distribution plot and the hierarchical clustering (Figure 1C & D). Expression levels of the 1,746 differentially expressed genes in Bcl11b-deficient cells were in general deviate less from those in control DN2a than those in DN2b (Figure 1C; Kolmogorov-Smirnov two-sided test; p = 4.18e-22). Furthermore, the hierarchical clustering also suggests Bcl11b-deficient DN2a-like cells are more closely related to wild-type DN2a than DN2b (Figure 1D).

Distinct Expression Patterns of Dysregulated Genes in Bcl11b-Deficient Cells

In previous studies from our lab, gene expression profiles of Bcl11b-deficient DN2a-like cells were analyzed by quantitative PCR with emphasis on a group of genes related to T-lineage specification, and alternative lineages and stem cell/progenitor-associated genes. In Bcl11b-deficient cells, many genes known to regulate T-lineage specification were expressed at the comparable levels as those in the wild-type counterparts. Some alternative lineage-associated genes, specifically NK-promoting genes, however, were abnormally augmented (Li et al., 2010a). Furthermore, Bcl11b-deficient cells failed to silence a subset of genes related to stem-cell self-renewal potential (Li et al., 2010a). To determine how genes are dysregulated due to the absence of Bcl11b, we aligned the three populations in the order of DN2a, Bcl11b-deficient DN2a-like and DN2b, and performed K-means clustering for the 1,746 differentially expressed genes (Figure 2A and Table S1). Based on the result, we divided genes into four distinct subgroups: (1) “DN2a-like” genes, which were expressed at levels more like those in wild-type DN2a than those in DN2b cells (in C7 and part of C3; Figure 2A); (2)

"DN2b-like" genes, genes whose expression were closer to DN2b than DN2a cells (in C1 and C5; Figure 2A); (3) "intermediate" genes, genes whose expression levels lay somewhere in between those of DN2a and of DN2b (in part of C3 and C8; Figure 2A); and (4) "Bcl11b-deficiency-specific upregulated" genes, genes that were substantially upregulated in Bcl11b-deficient cells compared to those in both DN2a and DN2b (in C2, C6 and part of C4; Figure 2A). Although as mentioned earlier 46 genes were specifically downregulated in Bcl11b-deficient cells against those in both wild-type DN2a and DN2b (Figure 1B), majority were more decreased compared to those in DN2b than those in DN2a and were included in cluster C7 or C8 (Figure 2A).

T-lineage commitment encompasses gradual exclusion of alternative lineage potentials and stepwise gain of T-cell identity. This process is reflected by the broad repression of progenitor- and alternative lineage-associated genes, as well as the rapid induction of T-cell identity genes (see Chapter 2). In wild-type cells, genes associated with progenitor-cells and alternative lineages are expressed exclusively in DN1 and DN2a stages, whereas T-cell signature genes are either sharply upregulated from DN1/2a to DN2b, or stably expressed throughout early T-cell development. Interestingly, Genes from these two categories were found in all four aforementioned groups in Bcl11b-deficient DN2a cells (Figure 2B). For example, "DN2a-like" genes include genes from both categories. The first one includes *Kit, Bmi1*, *Mpo* and *Il7r* (in C3) — genes that are functionally implicated in the maintenance of the progenitors or the development of alternative lineages, and may help Bcl11b-deficient DN2a survive and proliferate. The latter one includes *Ptcra*, *Spib*, *Lck*, *Gfi1*, *Notch3,* and *Eva1* (in C7) — genes that are involved in T-cell specification or provide essential signaling for cell development. These

T-cell genes were not upregulated in Bcl11b-deficient DN2a cells compared to wild-type DN2b cells. During early T-cell development, *Il7r* expression increases from DN1 to DN2a/b, and then quickly drops after DN2b stage (see Chapter 2). IL-7 signaling is required for pro-T cell survival and proliferation. However, enhanced or persisted IL-7 signaling has been shown to arrest early T-cell development at pre-committed stage through a Bcl11b-dependent mechanism (Ikawa et al., 2010). Another cell surface receptor gene *Eva1,* encoding Epithelial V-like Antigen 1, has been suggested to play an important role in T-lineage commitment (Iacovelli et al., 2009). Normally, *Eva1* is actively repressed in DN1 and DN2a stages, and transiently upregulated in DN2b and DN3 stages, and then repressed again afterward (see Chapter 2). In the absence of Bcl11b, both *Il7r* and *Eva1* stayed as "DN2a-like", which might contribute to the developmental block caused by Bcl11b deletion. On the other hand, despite being provided with Notch-Delta signaling for the same length of time as wild-type DN2b cells, Bcl11b-deficient cells failed to fully induce *Ptcra*, a direct target of Notch, as wild-type DN2b cells normally do. This result suggests that Notch-Delta signaling was compromised in the absence of Bcl11b.

Likewise, "DN2b-like" genes also include both progenitor/disparate lineage-linked genes, such as *Mycn* and *Erg* (in C1), and T-lineage specification genes, such as *Ccr9*, *TcrgC*, *Itk*, *Cd3g*, *Rag1, Lef1* and *Ets1* (in C5) (Figure 2B). The DN2b-like expression of *TcrgC* and *Rag1* is consistent with the less pronounced γδT potential defect of Bcl11b-deleted pro-T cells observed both *in vivo* and *in vitro* (Wakabayashi et al., 2003; Li et al., 2010a). To make things more complicated, one of aforementioned "DN2a-like" genes *Sox13* was not fully activated due to the loss of Bcl11b (Figure 2B).

Sox13 is a high- mobility group (HMG) transcription factor that promotes γδT differentiation but inhibits abT cells (Melichar et al., 2007). This may help explain why Bcl11b-deficient pro-T cells are not as competent in making γδT cells as their wild-type counterparts either (Li et al., 2010a).

The "intermediate" genes were markedly downregulated or upregulated in Bcl11b-deficient DN2a cells compared to those in the wild-type DN2a cells, but failed to reach the DN2b level (Figure 2B). This result implies the regulatory apparatus for these genes could be initiated without the involvement of Bcl11b, however, at a critical point between DN2a and DN2b the input from Bcl11b becomes absolutely essential. This group of genes once again includes a few progenitor- and/or alternative lineage-associated genes, such as *Tal*, *Gfi1b*, *Sfpi1* and *Bcl11a* (in C3), and a cohort of T-cell identity genes, such as *Cd2*, *Cd3d*, *Cd3e*, *Cd247*, *Dntt*, *Zap79*, *Thy1* and *Egr1* (in C8), many of which are involved in the assembly of TCR complex.

One profound defect in Bcl11b-deleted T-cell progenitor is the increased NK lineage potential. Many NK-lineage signature factors are abnormally augmented in Bcl11b-deficient cells even before these cells actually differentiate into NK-like cells (Li et al., 2010a; Li et al., 2010b). Indeed, a number of NK genes, such as *Id2, Zbtb16, Nfil3* and *Il2rb*, were found in the last group of genes, the "Bcl11b-deficiency-specific upregulated" genes. As discussed earlier, *Nfil3* and *Il2rb* are either silent or expressed at the background level in both wild-type DN2a and DN2b cells, as well as in DN1 cells where Bcl11b is not expressed (see Chapter 2). Thus, the loss of Bcl11b by itself is not sufficient to increase access to NK program; additional players are likely involved.

Taken together, the global expression profiling showed that loss of Bcl11b uncoupled several components of T-lineage commitment process. With sustained Notch-Delta signaling Bcl11b-deficient pro-T cell were able to initiate T-cell specification program. However, it appeared that Bcl11b was required for the completion of specification (Rothenberg et al., 2010). Furthermore, Bcl11b was absolutely essential to silence stem cell/progenitor- and alternative lineage-associated expression programs, thus, without it pro-T cells were trapped in a DN2a-like immature state (Rothenberg et al., 2010).

Normal Expression of *Notch1 and Notch3* during Commitment Requires Bcl11b

Although Notch-Delta signaling is the main driving force for T-lineage specification, our data suggested that in the absence of Bcl11b, Notch-Delta signaling failed to advance T-lineage program beyond the commitment. Many known Notch targets, such as *Hes1, Tcf7* and *Cd25,* were apparently not affected by the loss of Bcl11b. Nevertheless, one Notch function indicator, *Ptcra*, failed to properly respond to the long-term sustained Notch-Delta signaling that is capable of inducing *Ptcra* in normal DN2b cells. Another Notch downstream target *Dtx1* (in C1) was also substantially downregulated in Bcl11b-deficient DN2a cells compared to that in wild-type DN2a cells (Figure 2B). One possibility is that Bcl11b is a potent downstream mediator for part of Notch signaling pathway during the commitment (Rothenberg et al., 2010).

Another plausible explanation is Notch signaling itself was impaired by the loss of Bcl11b. Although still being expressed, both *Notch1* and, to lesser extent, *Notch3* were downregulated in Bcl11b-deficient cell compared to those in the wild-type counterparts (Notch1 is one of 46 Bcl11b-deficiency-specific downregulated genes, with more than 2x

decrease in mRNA level compared to both wild-type DN2a and DN2b) (Figure 2B & 3A). Notch signaling has been suggested to be a primary activator of Bcl11b expression (Tydell et al., 2007; Li et al., 2010b). Conceivably, a positive feedback loop may arise between Notch1 and Bcl11b during the normal T-cell commitment.

Gata3 is Upregulated in the Absence of Bcl11b

In striking contrast, GATA-3, a T-cell-specific transcription factor, was one of the 202 genes that were augmented only in Bcl11b-deficient cells. To test whether the upregulated GATA-3 expression might contribute to some of the abnormalities in Bcl11b-deficient cells, we performed ChIP-seq analysis to survey global GATA-3 occupancy in Bcl11b-deficient DN2a cells. Intriguingly, of 202 "Bcl11b-deficiency-specific upregulated" genes, 50 genes (~25%) were bound by GATA-3 in Bcl11b-deficient DN2a cells. These genes include *Zbtb16*, *Il2rb*, *Id2*, *Zbtb7b*, *Cpa3*, *Ikzf2* and *Gata3* itself (Figures 2B & S1; Table S2). Increase in CD4SP gene *Zbtb7b* and mast cell gene *Cpa3* was not unexpected; these two have been shown as direct downstream targets of GATA-3, and either can be induced by exogenous expression of GATA-3 (Taghon et al., 2007) or requires GATA-3 for normal expression (Wang et al., 2008). In addition, the result suggests that, in the absence of Bcl11b GATA-3 may directly upregulated *Zbtb16*, *Il2rb* and *Id2*, factors that favor NK-lineage differentiation. The seven aforementioned GATA-3 targets are perfectly bound by GATA-3 in wild-type DN2b (see Chapter 2; Figure S1). All but *Il2rb* and *Zbtb7b* are expressed at moderate level in the wild-type counterparts (see Chapter 2; Figure S1). Thus, under normal condition these genes are accessible to GATA-3, as well as to the transcriptional apparatus. It is possible that in wild-type pro-T cell Bcl11b directly or indirectly controls the transcriptional activities of

these genes at the physiological levels that actually help modulate T-lineage commitment program. Taking away of Bcl11b not only unleashes *Gata3,* but also removes the inhibitory mechanism of NK-lineage program mediated by Bcl11b, which is in turn reinforced by GATA-3.

Compelling evidence has shown that transcription factor global binding is largely cell type/stage-dependent (Ghisletti et al., 2010; Heinz et al., 2010; Lin et al., 2010; Treiber et al., 2010; Wei et al., 2011). To exclude the possibility that Bcl11b is so essential for maintaining T-cell identity that loss of it might have created a catastrophic effect on the cellular state including epigenetic landscape such that GATA-3 is no longer capable of finding its proper targets, we compared the global GATA-3 occupancy in Bcl11b-deficient DN2a-like cells with that in the wild-type cells generated from previous study (Figure 3D). Our data showed that increased GATA-3 did not increase the total number of GATA-3 bound sites in Bcl11b-deficient DN2a (Table 2). Moreover, GATA-3 global occupancy in Bcl11b-deficient DN2a-like cells was better correlated to that in wild-type DN2b than that in either DN1 or DP cells ($r = 0.51$, 0.38, and -0.015, respectively; Figure 3C). Complementary to the gene profile analysis, this result strongly indicates although T-lineage commitment profoundly depends on Bcl11b, the deletion of Bcl11b does not severely alter pro-T cell's identity.

DISCUSSION

T-lineage differentiation is regulated by a gene network consisting of multiple T-specific and non-T-specific regulators. We have shown here while not being involved in the initiation of T-lineage program, T-specific factor Bcl11b is indispensible for T-cell precursors' ability to pass the commitment checkpoint en route to become fully

functional T cells. When Bcl11b was deleted before pro-T cells had any chance to encounter it, the development was blocked at a DN2a-like stage with the presence of Notch-Delta signaling, when Bcl11b is normally in the process of being rapidly upregulated. The developmental block occurs before the commitment, as reflected by the enhanced NK potential, sustained progenitor-associated program and incomplete T-cell specification process.

Notch-Delta signaling is the central driver for the early T-cell development. Our genome-wide gene expression analysis suggests that Notch pathway was compromised due to the loss of Bcl11b, highlighted by the substantial reduction in *Notch1* and *Notch3* mRNA expression, failure to maintain *Dtx1* expression, and lack of proper *Ptcra* upregulation. In contrast, many known T-lineage differentiation regulators, such as TCF-1, Myb and E2A, were not affected in the absence of Bcl11b. The undermined Notch signaling therefore was one likely reason that Bcl11b-deficient pro-T cells were unable to complete the T-lineage specification despite being on OP9-DL culture for the same length of time as the controls. On the other hand, NK and myeloid cells have been shown to arise more robustly from Bcl11b-deficient cells if being transferred to OP9 control culture or treated with γ-secretase inhibitor (Li et al., 2010a). Together with the fact that no B-cell potential was observed in Bcl11b-deficient cell, these evidences signify that although being weakened, Notch-Delta signaling or its downstream mediator(s) is still able to partially restrain the access of pro-T cell to disparate lineages.

That being said, one cannot exclude the possibility of Bcl11b directly involving T-cell specification process. Normally Bcl11b mRNA is not detectable in DN1 stage, and only at modest level in DN2a stage. Bcl11b is sharply upregulated from DN2a to DN2b,

which is accompanied by the T-lineage commitment. Deletion of Bcl11b for most part does not prevent the initiation of T-lineage specification program. Nevertheless, without Bcl11b, T-lineage specification program losses the momentum to continue. It is possible that there is a critical point between DN2a and DN2b where Bcl11b reaches certain level and becomes absolutely necessary, and lack of it will halt the program.

Another notable defect caused by Bcl11b deletion is the enhanced NK potential in pro-T cells. We have pinpointed here GATA-3 was upregulated in Bcl11b-deficient pro-T cells, and increased GATA-3 in response to Bcl11b deletion might impede T-cell development by promoting NK-like gene program. Like Bcl11b, GATA-3 is one of the most T cell-specific transcription factors. GATA-3 function in early T-cell development is context and dosage-dependent. Germline deletion or conditional knockout of GATA-3 in multipotent progenitors blocks T-cell development at DN2a stage (Rothenberg and Scripture-Adams, 2008). Paradoxically, instead of enhancing T-cell develpmental progression, GATA-3 in DN2a stage either blocks cell survival when Notch-Delta signaling is activated, or promotes mast cell differentiation under non-T-cell permissive condition (Taghon et al., 2007). Interestingly, GATA-3 has been shown to directly engage in multiple aspects of NK development. GATA-3 is required for generating thymus-derived NK cells, which uniquely express both GATA-3 and IL7r (Vosshenrich et al., 2006). GATA-3 is also important for peripheral NK cell maturation (Samson et al., 2003). All of these findings suggest that GATA-3 is a link closely connecting T and NK programs. With the exception of *Cpa3*, many mast cell-associated genes that in previous study response rapidly to the exogenous GATA-3, such as *Gata1*, *Gata2* and *Mitf* (Taghon et al., 2007), were not induced in this Bcl11b-deficient context. One reasonable

explanation is that GATA-3 level is not high enough to initiate mast cell program in Bcl11b-deficient cells, whereas, on the other hand, the moderate elevation of GATA-3 in the absence of Bcl11b may help enhance NK potential by maintaining or even upregulating NK-associated gene expression program. Thus, our data suggest the dosage dependency of GATA-3 is rather sophisticated than previously thought. It is possible that different levels of GATA-3 dictate the passes to different lineages, including T, NK and mast cell, in a context-dependent manner. Since GATA-3 can be expressed in much higher level in peripheral T helper 2 cells compared to thymocytes (Asnagli et al., 2002), there must exist an inhibitory mechanism(s) to tightly control GATA-3 expression level in pro-T cells. Deletion of Bcl11b could partially liberate *Gata3* from the transcriptional inhibition.

Again, it is unlikely that the upregulation of NK program is a merely a by-product of Bcl11b deletion. Exogenous expression of GATA-3 appears unable to increase *Id2* expression in the presence of Bcl11b (Taghon et al., 2007). In addition, while ~20% Bcl11b-deficient specifically upregulated genes were directly bound by GATA-3, they only counted for ~5% of the total GATA-3 target in Bcl11b-deleted DN2 cells. For instance, although GATA-3 bound to *Tcf7* and *Tcfe2a* in both wild-type and knockout cells as indicated by ChIP-seq (Table S2; see Chapter 2), no significant impact had been observed for the mRNA expression of either genes in Bcl11b-deficient cells. As a result, imbalanced E2A (encoded by *Tcfe2a*) and Id2 could help divert T-cell development toward NK pathway (Boos et al., 2007). Collectively, these observations suggest that besides keeping *Gata3* expression in check, Bcl11b may add an additional layer of safeguard to specifically block the access of pro-T cell to NK lineage.

In contrast to the developmental arrest, our analyses imply that the loss of Bcl11b does not dramatically change pro-T cell cellular status at least in the short-term *in vitro* cultured cells (10~14 days). This may due to the fact that most T-lineage regulators, such as TCF-1 and E2A, were largely unaffected by the loss of Bcl11b. In the absence of Bcl11b these regulators collaborated with the weakened but not completely demolished Notch signaling to maintain T-cell identity, yet failed to push T-lineage differentiation forward to pass the commitment.

Reference

Asnagli, H., Afkarian, M., and Murphy, K.M. (2002). Cutting edge: Identification of an alternative GATA-3 promoter directing tissue-specific gene expression in mouse and human. J Immunol *168*, 4268-4271.

Boos, M.D., Yokota, Y., Eberl, G., and Kee, B.L. (2007). Mature natural killer cell and lymphoid tissue-inducing cell development requires Id2-mediated suppression of E protein activity. J Exp Med *204*, 1119-1130.

Cismasiu, V.B., Adamo, K., Gecewicz, J., Duque, J., Lin, Q., and Avram, D. (2005). BCL11B functionally associates with the NuRD complex in T lymphocytes to repress targeted promoter. Oncogene *24*, 6753-6764.

Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.L., Ragoussis, J., and Natoli, G. (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity *32*, 317-328.

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.

Iacovelli, S., Iosue, I., Di Cesare, S., and Guttinger, M. (2009). Lymphoid EVA1 expression is required for DN1-DN3 thymocytes transition. PLoS ONE *4*, e7586.

Ikawa, T., Hirose, S., Masuda, K., Kakugawa, K., Satoh, R., Shibano-Satoh, A., Kominami, R., Katsura, Y., and Kawamoto, H. (2010). An essential developmental checkpoint for production of the T cell lineage. Science *329*, 93-96.

Li, L., Leid, M., and Rothenberg, E.V. (2010a). An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. Science *329*, 89-93.

Li, P., Burke, S., Wang, J., Chen, X., Ortiz, M., Lee, S.C., Lu, D., Campos, L., Goulding, D., Ng, B.L., Dougan, G., Huntly, B., Gottgens, B., Jenkins, N.A., Copeland, N.G., Colucci, F., and Liu, P. (2010b). Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. Science *329*, 85-89.

Lin, Y.C., Jhunjhunwala, S., Benner, C., Heinz, S., Welinder, E., Mansson, R., Sigvardsson, M., Hagman, J., Espinoza, C.A., Dutkowski, J., Ideker, T., Glass, C.K., and Murre, C. (2010). A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. Nat Immunol *11*, 635-643.

Melichar, H.J., Narayan, K., Der, S.D., Hiraoka, Y., Gardiol, N., Jeannet, G., Held, W., Chambers, C.A., and Kang, J. (2007). Regulation of gammadelta versus alphabeta T lymphocyte differentiation by the transcription factor SOX13. Science *315*, 230-233.

Rothenberg, E.V., and Scripture-Adams, D.D. (2008). Competition and collaboration: GATA-3, PU.1, and Notch signaling in early T-cell fate determination. Semin Immunol *20*, 236-246.

Rothenberg, E.V., Zhang, J., and Li, L. (2010). Multilayered specification of the T-cell lineage fate. Immunol Rev *238*, 150-168.

Sambandam, A., Maillard, I., Zediak, V.P., Xu, L., Gerstein, R.M., Aster, J.C., Pear, W.S., and Bhandoola, A. (2005). Notch signaling controls the generation and differentiation of early T lineage progenitors. Nat Immunol *6*, 663-670.

Samson, S.I., Richard, O., Tavian, M., Ranson, T., Vosshenrich, C.A., Colucci, F., Buer, J., Grosveld, F., Godin, I., and Di Santo, J.P. (2003). GATA-3 promotes maturation, IFN-gamma production, and liver-specific homing of NK cells. Immunity *19*, 701-711.

Taghon, T., Yui, M.A., and Rothenberg, E.V. (2007). Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. Nat Immunol *8*, 845-855.

Treiber, T., Mandel, E.M., Pott, S., Gyory, I., Firner, S., Liu, E.T., and Grosschedl, R. (2010). Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. Immunity *32*, 714-725.

Tydell, C.C., David-Fung, E.S., Moore, J.E., Rowen, L., Taghon, T., and Rothenberg, E.V. (2007). Molecular dissection of prethymic progenitor entry into the T lymphocyte developmental pathway. J Immunol *179*, 421-438.

Vosshenrich, C.A., Garcia-Ojeda, M.E., Samson-Villeger, S.I., Pasqualetto, V., Enault, L., Richard-Le Goff, O., Corcuff, E., Guy-Grand, D., Rocha, B., Cumano, A., Rogge, L., Ezine, S., and Di Santo, J.P. (2006). A thymic pathway of mouse natural killer cell development characterized by expression of GATA-3 and CD127. Nat Immunol *7*, 1217-1224.

Wakabayashi, Y., Watanabe, H., Inoue, J., Takeda, N., Sakata, J., Mishima, Y., Hitomi, J., Yamamoto, T., Utsuyama, M., Niwa, O., Aizawa, S., and Kominami, R. (2003). Bcl11b is required for differentiation and survival of alphabeta T lymphocytes. Nat Immunol *4*, 533-539.

Wang, L., Wildt, K.F., Zhu, J., Zhang, X., Feigenbaum, L., Tessarollo, L., Paul, W.E., Fowlkes, B.J., and Bosselut, R. (2008). Distinct functions for the transcription factors GATA-3 and ThPOK during intrathymic differentiation of CD4(+) T cells. Nat Immunol *9*, 1122-1130.

Weber, B.N., Chi, A.W., Chavez, A., Yashiro-Ohtani, Y., Yang, Q., Shestova, O., and Bhandoola, A. (2011). A critical role for TCF-1 in T-lineage specification and differentiation. Nature *476*, 63-68.

Wei, G., Abraham, B.J., Yagi, R., Jothi, R., Cui, K., Sharma, S., Narlikar, L., Northrup, D.L., Tang, Q., Paul, W.E., Zhu, J., and Zhao, K. (2011). Genome-wide Analyses of

Transcription Factor GATA3-Mediated Gene Regulation in Distinct T Cell Types. Immunity *35*, 299-311.

Yashiro-Ohtani, Y., Ohtani, T., and Pear, W.S. (2010). Notch regulation of early thymocyte development. Semin Immunol *22*, 261-269.

METHODS

Cell Culture

In vitro tracking development of Bcl11b-deficient cells was performed as described (Li et al., 2010a). Briefly, fetal liver (FL) cell samples (E13.5) from $Bcl11b^{fl/fl}$-ROSA26R-YFP or C57BL/6-ROSA26R-YFP mouse embryos were first depleted of cells positive for Ter119, CD19 and Gr1 by magnetic selection. c-$Kit^{+}CD27^{+}$ multi-lineage precursors were then sorted from lineage-depleted FL cells by FACS and retrovirally transduced with recombinase Cre. Treated precursors were co-cultured with OP9-DL1 stromal cells in the presence of 5 ng/mL each of Flt3L, IL-7 and SCF for 24 hours. Successfully transduced cells were sorted for YFP expression together with lymphoid progenitor-enriched phenotype (c-$Kit^{+}CD27^{+}$) and returned to the OP9-DL1 or OP9-DL4 co-culture. After 10~14 days of culture, the cells were harvested and sorted for DN2a (YFP^{+}c-$Kit^{hi}CD45^{+}CD25^{+}Lin^{-}$) from $Bcl11b^{fl/fl}$ and DN2b (YFP^{+}c-Kit^{int} $CD45^{+}CD25^{+}Lin^{-}$) from C57BL/6.

Chromatin Immunoprecipitation; mRNA Purification and cDNA Library Building; High Throughput Sequencing; and ChIP-seq and RNA-seq Data Analysis

All were performed as described (see Chapter 2).

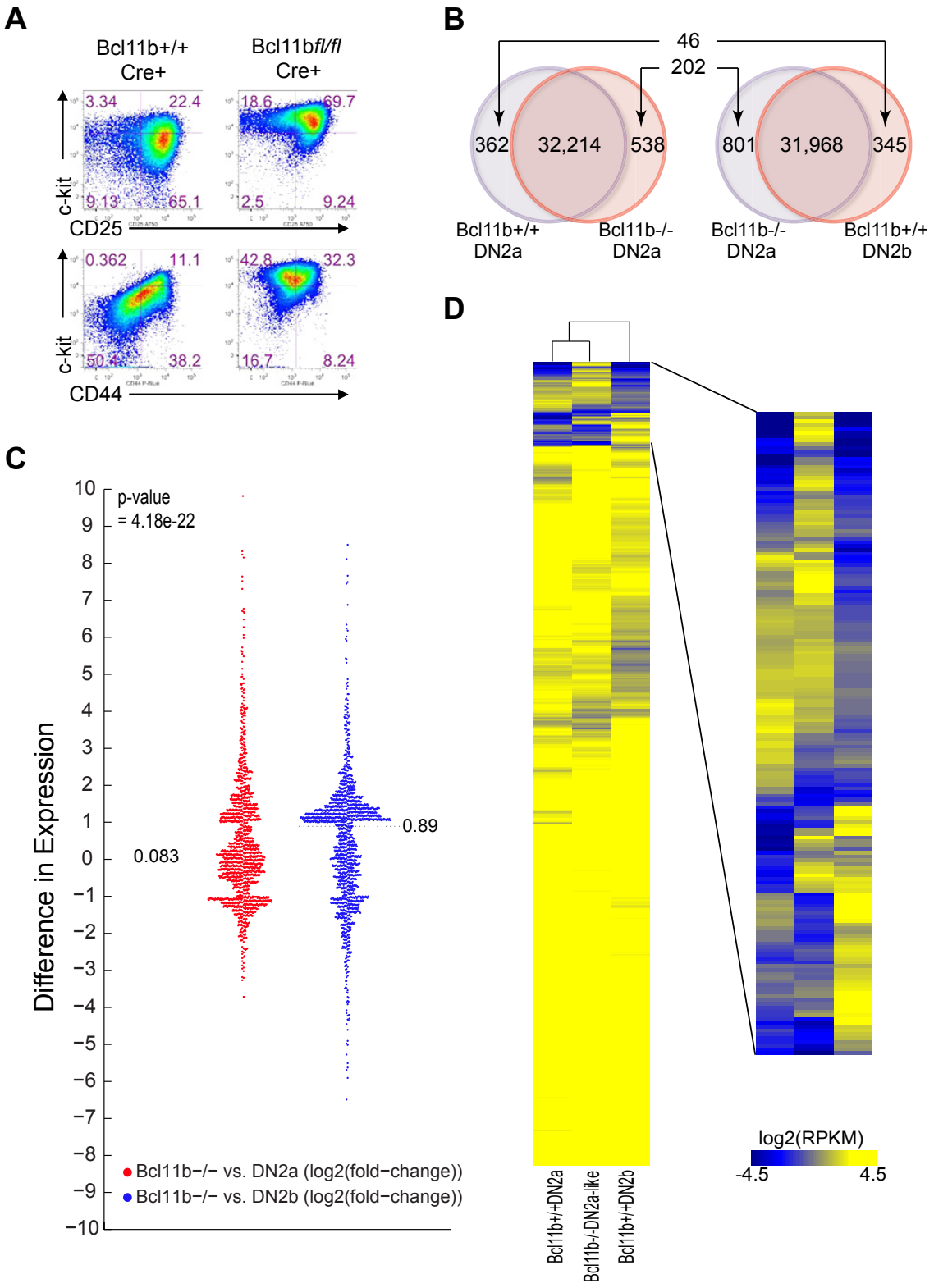

A
Bcl11b+/+
Cre+
Bcl11bfl/fl
Cre+
3.34
22.4
9.13
65.1
18.6
69.7
2.5
9.24
c-kit
CD25
0.362
11.1
50.4
38.2
42.8
32.3
16.7
8.24
c-kit
CD44
B
46
202
362
32,214
538
801
31,968
345
Bcl11b+/+
DN2a
Bcl11b-/-
DN2a
Bcl11b-/-
DN2a
Bcl11b+/+
DN2b
C
p-value
= 4.18e-22
Difference in Expression
0.083
0.89
Bcl11b−/− vs. DN2a (log2(fold−change))
Bcl11b−/− vs. DN2b (log2(fold−change))
D
log2(RPKM)
-4.5
4.5
Bcl11b+/+DN2a
Bcl11b-/-DN2a-like
Bcl11b+/+DN2b

Figure 1. Comparison of global expression profiles between Bcl11b-deficient DN2a-like cells and the wild-type counter parts.

A. Flow cytometric analysis of fetal liver precursors *in vitro* derived pro-T cells after 9 days of OP9-DL1 or DL4 co-culture. Left panels are wild-type control, and right panels are Bcl11b-deleted cells.

B. Pair-wise comparisons in gene expression between Bcl11b-deficient DN2-like cells (Bcl11b-/-DN2a) and either wild-type DN2a (Bcl11b+/+DN2a) or wild-type DN2b (Bcl11b+/+DN2b). Statistically significantly changed genes were defined by DEGseq ($p < 0.001$, ≥2x change in expression).

C. Difference in expression ($\log_2$(fold-change)) of individual differentially expressed genes between Bcl11b-deficient cells and either wild-type DN2a or DN2b cells were calculated and plotted. Each colored dot represents one gene. Dashed line indicates where the median (in parenthesis) of each data set is. Difference in expression between Bcl11b-deficient cells and wild-type DN2b cells were more spread out (K-S two-sided test; p = 4.18e-22).

D. Hierarchical clustering of expression patterns of 1,746 differentially expressed genes. Expression levels of differentially expressed genes are hierarchically clustered along both sample and gene dimensions and displayed as a heatmap. A subcluster representing the most changed genes is zoomed in to highlight the distinct patterns.

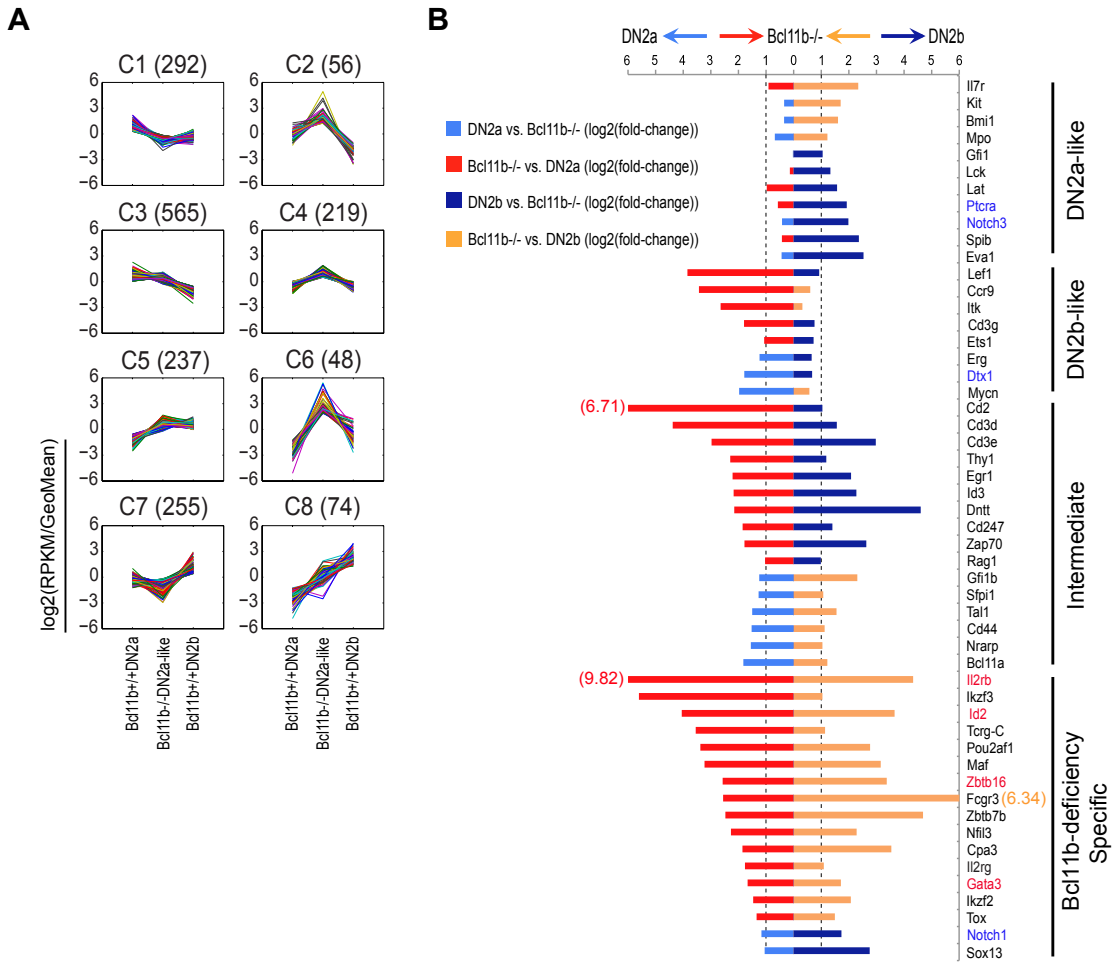


Figure 2. Distinct patterns of differentially expressed genes in Bcl11b-deficient cells.

A. K-means clustering for differentially expressed genes. 1,746 differentially expressed genes were subjected to K-means clustering analysis. The number of genes in each cluster is included in each parenthesis. Genes in each cluster are listed in order in Table S1.

B. Some known early T-cell development related genes were selected from each of four subgroups, and plotted accordingly. Colored bars represent the level (bar length) and direction (bar color) of change in mRNA expression for individual genes. Light blue indicates downregulation from DN2a to Bcl11b-deficient DN2a; red indicates

upregulation from DN2a to Bcl11b-deficient DN2a; light brown indicates downregulation from Bcl11b-deficient DN2a to DN2b; and navy blue indicates upregulation from Bcl11b-deficient DN2a to DN2b. Dashed lines are the boundaries for ≥2x change in expression.

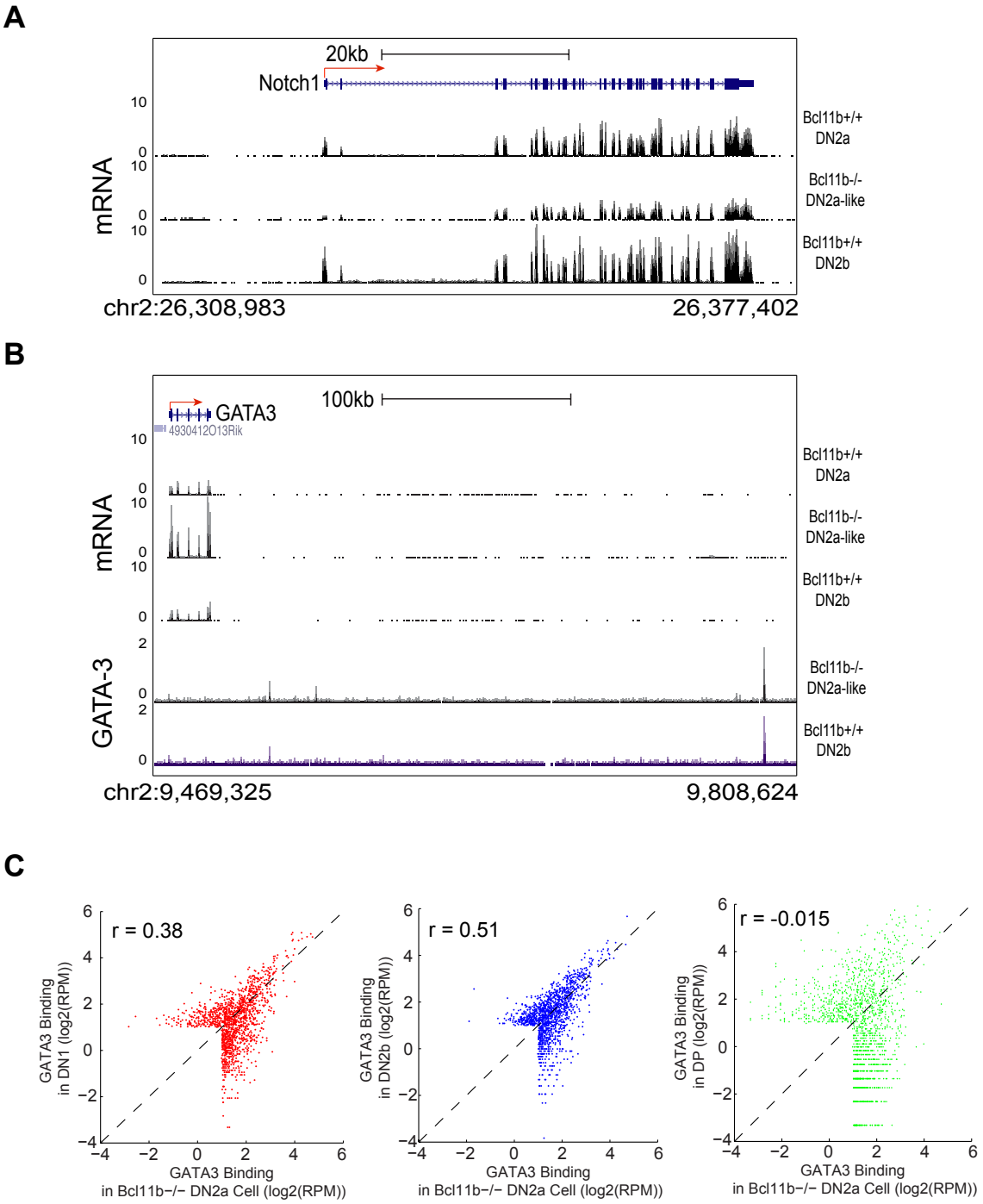


Figure 3. Bcl11b deletion leads to changes in *Notch1* and *Gata3* expression, but not pro-T cell identity.

A. UCSC genome browser track image depicting expression pattern at *Notch1* locus.

B. UCSC genome browser track image depicting expression pattern (top three) and GATA-3 occupancy (bottom 2) at *Gata3* locus.

Order from top to bottom: RNA-seq of Bcl11b+/+DN2a, Bcl11b-/-DN2a and Bcl11b+/+DN2b, and ChIP-seq of Bcl11b-/-DN2a and Bcl11b+/+DN2b. Red arrow: TSS and direction of transcription. Uniform scales are used for mRNA or ChIP in all cell types, as shown by y-axis labels (RPKM for RNA-seq, RPM for ChIP-seq). Chromosomal coordinates are indicated below each panel.

C. Scatter plots depicting the comparisons in GATA-3 binding between Bcl11b-deficient DN2a with wild-type DN1, DN2b or DP. Pearson correlation coefficients are shown for each comparison.

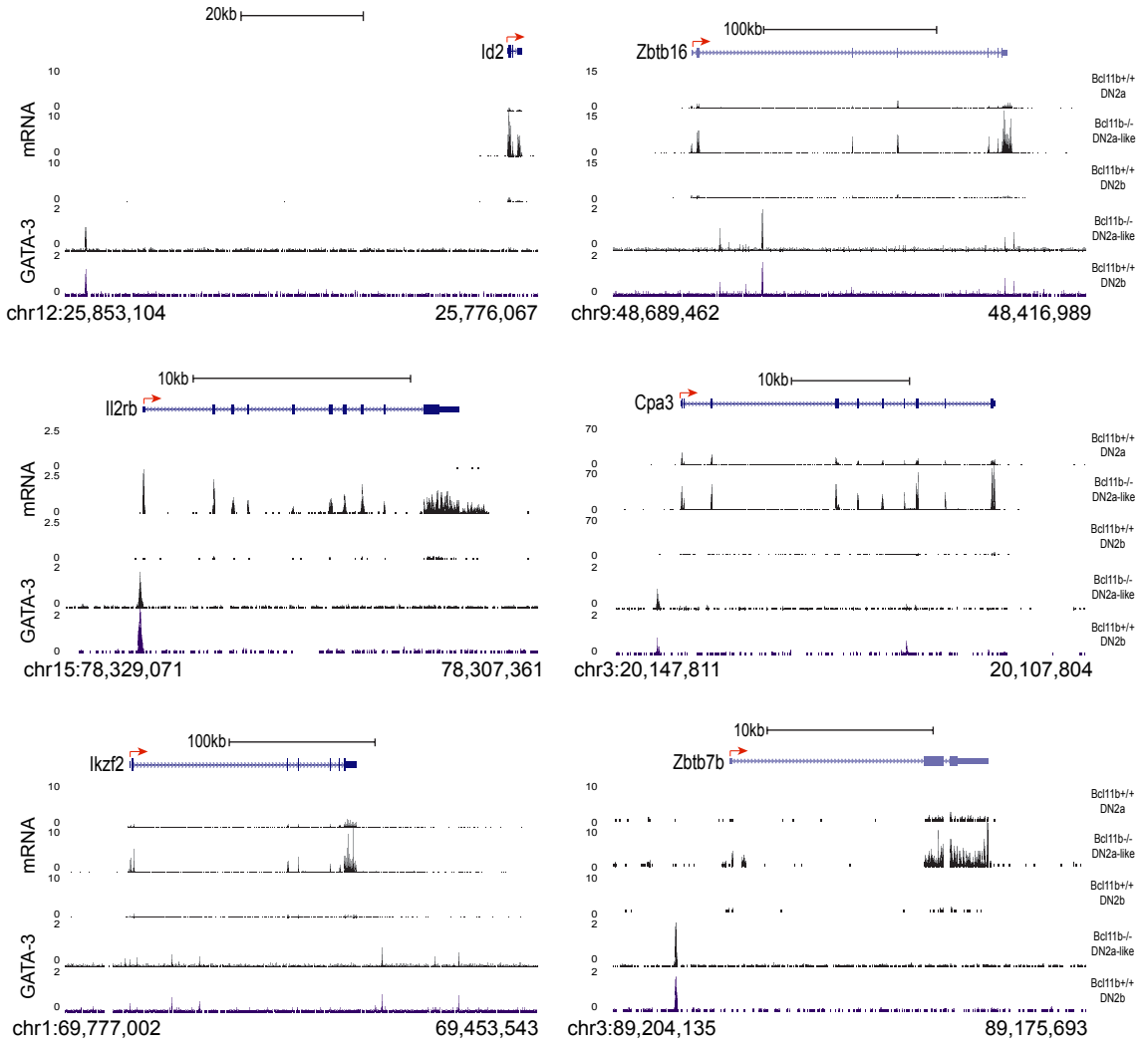


Figure S1. GATA3 targets that are specifically upregulated in Bcl11b-deficient DN2a cells.

A-F. UCSC genome browser track image depicting expression pattern (top three) and GATA-3 occupancy (bottom 2) at selected GATA-3 targets gene loci that are specifically upregulated in Bcl11b-deficient DN2a cells: *Id2*, *Zbtb16*, *Il2rb*, *Cpa3*, *Ikzf2* and *Zbtb7b.* Order from top to bottom: RNA-seq of Bcl11b+/+DN2a, Bcl11b-/-DN2a and Bcl11b+/+DN2b, and ChIP-seq of Bcl11b-/-DN2a and Bcl11b+/+DN2b. Red arrow: TSS

and direction of transcription. Uniform scales are used for ChIP in all cell types, and mRNA scales are uniform within each panel, as shown by y-axis labels (RPKM for RNA-seq, RPM for ChIP-seq). Chromosomal coordinates are indicated below each panel.

Supplementary Tables 1-2

Table S1. K-means clustering for differentially expressed genes in Bcl11b-deficient DN2a compared to wild-type cells (see supplementary files).

Table S2. Genome-wide GATA3 binding in Bcl11b-deficient DN2a (see supplementary files).

Chapter 4

Discussion

The purpose of my thesis is to study the molecular mechanism underlying T-cell lineage specification and commitment. Using genome-wide sequencing and proper statistical analysis tools, we mapped and analyzed the global distributions and longitudinal changing patterns of histone modifications, the correlated transcriptional activities, as well as the occupancies of two transcription factors, PU.1 and GATA-3, in five successive stages of early T-cell development. Our results delineate mechanistically separated and sequentially occurred downregulation as well as upregulation events during the critical T-lineage commitment process, which help explain why there is not a single "master" regulator that dictates the whole process. Our data also provide a comprehensive resource of discrete histone-modified genomic regions, some of which have been utilized in a stage-specific manner and may define at least partially the regulatory network during early T-cell development. Furthermore, combining this global approach with gene perturbation, we investigated the function of Bcl11b, a T-cell specific regulator, on T-lineage commitment. The gene expression profiling and GATA-3 global occupancy analysis in Bcl11b-deficent cells indicate that Bcl11b is required for the proper expression of *Notch1*, *Notch3* and *Gata3* during this process, and the cooperatively interplay between Bcl11b, Notch and GATA-3 is essential for the completion of T-cell lineage specification and the elimination of NK lineage potential.

On the other hand, from our genome-wide studies several interesting questions have emerged, and are worth for further exploration.

H3K27me3 vs. H3K9me3 and DNA Methylation

A major feature linked to T-lineage commitment is the gradual repression of progenitor-associated genes. In Chapter 2, we found a series of progressive epigenetic

changes in establish a stable repressive chromatin structure at some newly repressed gene loci, including promoter regions and/or distal elements. The likely epigenetic event that initiates repression process is the gradual deacetylation, which often slightly precedes or is parallel with the downregulation of mRNA expression. The reduction in H3K4me2 enrichment usually delays one or two developmental stages compared to that in H3Ac and mRNA, sometimes being accompanied by focal then slowly expanded H3K27me3 enrichment. But at some gene loci H3K27me3 is not on until the genes are completely or almost completely silent. These observations indicate that while relinquishing of active epigenetic modification is the default method for gene silencing, additional H3K27me3 may act as a stabilizer for the repression. Presumably, this added epigenetic repression reduces the likelihood of escaping gene silencing and lineage reprogramming.

On the other hand, however, while silent transcription factors are specifically favored by H3K27me3, only one third of all silent gene loci and even lower fraction of constitutively silent nonhematopoietic genes are enriched with H3K27me3 in any particular stage. Besides H3K27me3, H3K9me3 and DNA methylation are also implicated in the developmentally regulated gene repression (Hawkins et al., 2010; Ji et al., 2010). It is reasonable to believe these two modifications may provide stable epigenetic repression for some silent gene loci that are not enriched with H3K27me3.

H3K27me3 and H3K9me3 have been shown to selectively target different sets of genes in human embryonic stem cells (hESCs) as well as the terminal differentiated fetal lung fibroblasts (IMR90) with very small percentage of overlapping (Hawkins et al., 2010). And both modifications are associated with lineage and/or developmentally

related genes in hESCs, and with nondevelopmentally related genes in IMR90 (Hawkins et al., 2010).

DNA methylation represses gene expression by methylating promoter CpG islands, and developmentally programmed CpG methylation is associated with lineage restriction (Ji et al., 2010). Interestingly, in one study that compares DNA methylation pattern with different histone modification patterns, while most active promoters are hypomethylated, a subset of silent promoters that lack any histone modifications are hypermethylated (Hawkins et al., 2010). Since DNA methylation is believed to be less flexible than histone modifications, and can be maintained during cell division, it may provide long-term lineage stability when initial factors required for silencing are no longer present (Cedar and Bergman, 2011).

Possible Roles of Zbtb7b and Bcl11a in the Absence of Bcl11b

In Chapter 3, we found a group of about 50 genes that were both upregulated specifically in Bcl11b-deficient DN2a cell and bound by GATA-3. *Gata3* itself is among this group. Since three genes from this group, *Zbtb16*, *Id2* and *IL2rb*, are strongly connected to the enhanced NK-potential in response to Bcl11b deletion, this finding implies the potential NK-lineage promoting function of GATA-3 in the absence of Bcl11b.

Another gene that also belongs to this group is *Zbtb7b*. Zbtb7b (or Thpok) is a CD4 T cell lineage-specific regulator. The sequential activity of GATA-3 and Zbtb7b leads DP cells into CD4 SP (single positive) pathway (Wang et al., 2008). Thpok deficiency abolishes CD4 T cell development, and directs DP cells into CD8 lineage without affecting positive selection (Kappes, 2010). However, it is unclear whether

Zbtb7b plays any role in early T-cell development (in particular, αβ T cell) prior to positive selection.

Our genome-wide studies show that *Zbtb7b* is expressed at low level in DN1 cells, and immediately diminished to essentially undetectable in DN2a and onward until the positive selection. *Zbtb7b* is actively repressed by H3K27me3 from DN2a to pre-positive selection DP cells, though is also constitutively bound by GATA-3 throughout early T-cell development (see Chapter 2). GATA-3 has been shown to be a direct inducer of Zbtb7b during positive selection (Wang et al., 2008). The upregulation of Zbtb7b might be just a byproduct of increased GATA-3 with no negative effect, which is supported by the absence of *Cd4* expression in Bcl11b-deficient DN2a cells. Or, the presence of Zbtb7b actually contributes to the overall phenotype of Bcl11b-deficient cells that is unrelated with CD4 SP development.

Interestingly, recent studies show that the GATA-3 and Zbtb7b cascade is required in producing $CD4^+$ CD1d-restricted invariant NKT (iNKT) cell development (Engel et al., 2010; Wang et al., 2010). Furthermore, Zbtb7b is highly expressed in $NK1.1^+$ γδ T cell, implying a potential role in γδ T cell maturation and function (Kappes, 2010). Collectively, these findings suggest Zbtb7b is not confined to solely promoting CD4 SP lineage differentiation.

So, why do Bcl11b-deficient DN2a cells not express CD4? The most likely explanation is the presence of Runx3 in Bcl11b-deficient DN2a (see Chapter 2). Runx3 is a known repressor of CD4 expression (Collins et al., 2009). Since Runx3 expression is not affected with the deletion of Bcl11b (see Chapter 3), it is possible that the amount of Zbtb7b in Bcl11b-deficient DN2a cells is unable to overpower Runx3 and activate CD4

expression. Although more experiments are needed, it should not be ruled out that Zbtb7b plays a possible role in Bcl11b-deficient DN2a specific phenotypes, which, if is true, may highlight the importance of the programmed epigenetic repression at *Zbtb7b* locus during normal T-lineage commitment.

Another remaining question worth to be addressed in the future is why Bcl11b deletion leads to the upregulation of some genes that are not expressed or expressed at low or background levels in pro-T cell (DN1, or DN2a, or both), such as *Zbtb7b* and *Il2rb*. Since Bcl11b is not expressed in DN1 cells, and expressed at modest level in DN2a cells, there must be additional mechanism(s), i.e., a specific regulator equivalent to Bcl11b. One candidate is Bcl11a, a paralog of Bcl11b (Avram et al., 2000; Avram et al., 2002).

Bcl11a is an essential B-lineage regulator and plays a key role in promoting fetal-to-adult switch in hemoglobin expression (Liu et al., 2003; Sankaran et al., 2010). Bcl11a is expressed in DN1 cell, and gradually downregulated afterward until being completely turned off during DN2b to DN3 transition. And it is one of the "intermediate" genes whose expression levels in Bcl11b-deficient DN2a cells lie in somewhere between those in DN2a and DN3a (see Chapter 3).

Bcl11a and Bcl11b share a high percentage of amino acid sequence similarity, in particular the two centrally located zinc fingers, zinc finger 3 and 4 (94% identity) (Avram et al., 2002). It is reasonable to predict that the two transcription factors overlap in their binding targets. Intriguingly, in Bcl11a mutant fetal thymus, the number of γδT cells increases 3~4 fold while αβT cells are reduced compared to those in the wild-type control (Liu et al., 2003). The unique complement expression pattern of Bcl11a and

Bcl11b and the perturbation study results strongly suggest a functional connection between these two factors in early T-cell development.

A Potential Stage-Specific Cis-regulatory Element of *Gata3*

One of the major contributions from our genomic study is having identified ~20,000 histone modified potential cis-regulatory elements (distal) that might be part of the gene network regulating early T-cell development. A number of previously identified cis-regulatory elements are among this list (see Chapter 2).

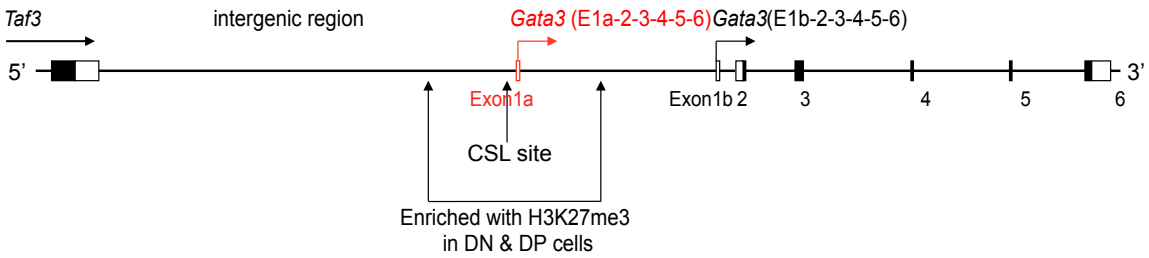


Figure1. Schematic representation of *Gata3* locus.

The two isoforms of *Gata3* utilize either exon1a (E1a, in red) or exon1b (E1b, in black). Exon1a and its flanking regions are enriched with H3K27me3 (indicated by two dark arrows) in DN and DP cells. One CSL DNA-binding site (CSL site) resides just upstream of exon1a, being buried by the H3K27me3 enrichment in DN and DP cells. Arrows on the top indicate the directions of transcription. Filled boxes are the translated regions

In Chapter 3, we found that the elevated GATA-3 expression may mediate at least partially the phenotypic defect caused by Bcl11b deletion. And as opposite to its essential role in early T-cell development at the normal level, high dosage of GATA-3 has

previously shown to divert pro-T into mast cell lineage (Taghon et al., 2007). In sharp contrast, overexpression of GATA-3 in CD4 naïve or Th1 cells is able to promote Th2 differentiation (Ouyang et al., 1998; Ouyang et al., 2000).

GATA3 utilizes an alternative promoter in a stage-specific manner. While both distal E1a (exon 1a) and proximal E1b (exon 1b) are expressed in Th2 cells, the isoform containing the proximal E1b is the predominate form in thymocytes (Asnagli et al., 2002) (Figure 1). GATA-3 (E1b isoform) is among the earliest factors induced by Notch-Delta signaling (Taghon et al., 2005), suggesting *Gata3* is a downstream target of Notch-Delta signaling. However, the distal E1a promoter can be upregulated by Notch1 in Th2 cells but not in thymocytes (Amsen et al., 2007; Fang et al., 2007). It is unclear why the E1a promoter does not response to Notch-Delta signaling in these early T cells. Nevertheless, these data suggest there is a cell-intrinsic inhibitory mechanism in play to restrain GATA-3 from inappropriate upregulation by Notch signaling. Our genome-wide studies demonstrate that as no or very few transcripts are generated from the distal E1a in all five populations, the region spanning from the upstream of E1a promoter, which includes a known CSL binding site, to the downstream intronic region between E1a and E1b is correspondingly heavily covered with H3K27me3 (see Chapter 2). Within this region, a smaller distinct island just upstream of the E1a TSS (transcription start site) was enriched with both H3K4me2 and H3K27me3, but low for H3Ac, suggesting that the E1a promoter is already "poised" for activation as early as in DN1 stage. Interestingly, the H3K27me3 enrichment is retained across this region in CD4 naïve T cell, and abolished in Th2 cells where E1a is upregulated (Wei et al., 2009). Our data indicate that during early T-cell development, the widespread distal H3K27me3 prevents the E1a promoter

from being activated, resulting in a tightly controlled expression of GATA-3. And more importantly, there might actually be a potential repressor element(s) masked with H3K27me3 enrichment within this region.

Reference

Amsen, D., Antov, A., Jankovic, D., Sher, A., Radtke, F., Souabni, A., Busslinger, M., McCright, B., Gridley, T., and Flavell, R.A. (2007). Direct regulation of Gata3 expression determines the T helper differentiation potential of Notch. Immunity *27*, 89-99.

Asnagli, H., Afkarian, M., and Murphy, K.M. (2002). Cutting edge: Identification of an alternative GATA-3 promoter directing tissue-specific gene expression in mouse and human. J Immunol *168*, 4268-4271.

Avram, D., Fields, A., Pretty On Top, K., Nevrivy, D.J., Ishmael, J.E., and Leid, M. (2000). Isolation of a novel family of C(2)H(2) zinc finger proteins implicated in transcriptional repression mediated by chicken ovalbumin upstream promoter transcription factor (COUP-TF) orphan nuclear receptors. J Biol Chem *275*, 10315-10322.

Avram, D., Fields, A., Senawong, T., Topark-Ngarm, A., and Leid, M. (2002). COUP-TF (chicken ovalbumin upstream promoter transcription factor)-interacting protein 1 (CTIP1) is a sequence-specific DNA binding protein. Biochem J *368*, 555-563.

Cedar, H., and Bergman, Y. (2011). Epigenetics of haematopoietic cell development. Nat Rev Immunol *11*, 478-488.

Collins, A., Littman, D.R., and Taniuchi, I. (2009). RUNX proteins in transcription factor networks that regulate T-cell lineage choice. Nat Rev Immunol *9*, 106-115.

Engel, I., Hammond, K., Sullivan, B.A., He, X., Taniuchi, I., Kappes, D., and Kronenberg, M. (2010). Co-receptor choice by V alpha14i NKT cells is driven by Th-POK expression rather than avoidance of CD8-mediated negative selection. J Exp Med *207*, 1015-1029.

Fang, T.C., Yashiro-Ohtani, Y., Del Bianco, C., Knoblock, D.M., Blacklow, S.C., and Pear, W.S. (2007). Notch directly regulates Gata3 expression during T helper 2 cell differentiation. Immunity *27*, 100-110.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., Antosiewicz-Bourget, J., Ye, Z., Espinoza, C., Agarwahl, S., Shen, L., Ruotti, V., Wang, W., Stewart, R., Thomson, J.A., Ecker, J.R., and Ren, B. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell *6*, 479-491.

Ji, H., Ehrlich, L.I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K., Rossi, D.J., Inlay, M.A., Serwold, T., Karsunky, H., Ho, L., Daley, G.Q., Weissman, I.L., and Feinberg, A.P. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature *467*, 338-342.

Kappes, D.J. (2010). Expanding roles for ThPOK in thymic development. Immunol Rev *238*, 182-194.

Liu, P., Keller, J.R., Ortiz, M., Tessarollo, L., Rachel, R.A., Nakamura, T., Jenkins, N.A., and Copeland, N.G. (2003). Bcl11a is essential for normal lymphoid development. Nat Immunol *4*, 525-532.

Ouyang, W., Lohning, M., Gao, Z., Assenmacher, M., Ranganath, S., Radbruch, A., and Murphy, K.M. (2000). Stat6-independent GATA-3 autoactivation directs IL-4-independent Th2 development and commitment. Immunity *12*, 27-37.

Ouyang, W., Ranganath, S.H., Weindel, K., Bhattacharya, D., Murphy, T.L., Sha, W.C., and Murphy, K.M. (1998). Inhibition of Th1 development mediated by GATA-3 through an IL-4-independent mechanism. Immunity *9*, 745-755.

Sankaran, V.G., Xu, J., and Orkin, S.H. (2010). Transcriptional silencing of fetal hemoglobin by BCL11A. Ann N Y Acad Sci *1202*, 64-68.

Taghon, T., Yui, M.A., and Rothenberg, E.V. (2007). Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. Nat Immunol *8*, 845-855.

Taghon, T.N., David, E.S., Zuniga-Pflucker, J.C., and Rothenberg, E.V. (2005). Delayed, asynchronous, and reversible T-lineage specification induced by Notch/Delta signaling. Genes Dev *19*, 965-978.

Wang, L., Carr, T., Xiong, Y., Wildt, K.F., Zhu, J., Feigenbaum, L., Bendelac, A., and Bosselut, R. (2010). The sequential activity of Gata3 and Thpok is required for the differentiation of CD1d-restricted CD4+ NKT cells. Eur J Immunol *40*, 2385-2390.

Wang, L., Wildt, K.F., Zhu, J., Zhang, X., Feigenbaum, L., Tessarollo, L., Paul, W.E., Fowlkes, B.J., and Bosselut, R. (2008). Distinct functions for the transcription factors GATA-3 and ThPOK during intrathymic differentiation of CD4(+) T cells. Nat Immunol *9*, 1122-1130.

Wei, G., Wei, L., Zhu, J., Zang, C., Hu-Li, J., Yao, Z., Cui, K., Kanno, Y., Roh, T.Y., Watford, W.T., Schones, D.E., Peng, W., Sun, H.W., Paul, W.E., O'Shea, J.J., and Zhao, K. (2009). Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity *30*, 155-167.