THE UNIVERSITY OF CHICAGO

COMPUTATIONAL GENOMIC ANALYSIS OF TRANSCRIPTIONAL REGULATION IN INNATE LYMPHOID CELL DEVELOPMENT

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
AND
THE FACULTY OF THE DIVISON OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOPHYSICAL SCIENCES

BY
HERMAN GUDJONSON

CHICAGO, ILLINOIS
DECEMBER 2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

αLP	$\alpha_4\beta_7$-expressing Lymphoid Progenitor
AhR	Aryl hydrocarbon Receptor
Arg1	Arginase 1
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
CHILP	Common Helper Innate Lymphoid Precursor
CLP	Common Lymphoid Progenitor
cNK	conventional Natural Killer Cell
CRM	Cis-Regulatory Module
EILP	Early Innate Lymphoid Precursor
FDR	False Discovery Rate
ftILCP	fetal transitional Innate Lymphoid Cell Precursor
HDLSS	High Dimensional Low Sample Size
HMG	High-mobility group
HSC	Hematopoietic Stem Cell
Id	Inhibitor of DNA binding
ILC	Innate Lymphoid Cell
ILC2P	Innate Lymphoid Cell 2 Precursor
ILCP	Innate Lymphoid Cell Precursor
iNK	immature NK
LTi	Lymphoid Tissue inducer Cell

LTiP	Lymphoid Tissue inducer Precursor Cell
MHC	Major Histocompatibility Complex
NCR	Natural Cytotoxicity Receptor
NGS	Next Generation Sequencing
NKT	Natural Killer T Cell
PLZF	Promyelocytic Leukaemia Zinc Finger
pre-NKP	pre-Natural Killer Precursor
RA	Retinoic Acid
RMA	Robust Multi-array Average
RNA-seq	RNA-sequencing
RPKM	Reads Per Kilobase of exon model per Million reads
rNKP	refined Natural Killer Precursor
SCID	Severe Combined Immunodeficiency
Th	T helper Cell
TMM	Trimmed Method of Means
TPM	Transcripts Per Million
TSLP	Thymic Stromal Lymphopoietin
t-SNE	t-distributed stochastic neighbor embedding

ACKNOWLEDGMENTS

I would first like to thank my faculty advisors Dr. Aaron Dinner and Dr. Albert Bendelac. Right from my first rotation project, Aaron had a remarkable willingness to point me towards a wealth of formative collaborations when I had little experience and so valuably shaped my introduction to the biological sciences. I am sincerely thankful for Aaron's patience, understanding and guidance as an advisor, and hold him in my mind as the example for an advisor with his student's interests truly at heart. I am also sincerely grateful for Aaron's unwavering support throughout my studies. I am very thankful for Albert's willingness to accept me in his lab and for all his patient guidance through the rich field of immunology. Albert is an extraordinary scientist and teacher, and I have only the utmost respect for his scientific intuition and knowledge.

I would like to thank members of the Bendelac lab Dr. Mike Constantinides and especially Dr. Isabel Ishizuka for working closely with me throughout my analysis of ILC development. Isabel devoted an incredible amount of time in both providing training during my foray into experiment and in discussion that constantly enriched my scientific understanding. Without Mike's seminal contributions and the work of our collaborators Dr. Silvestre Chea and Dr. Rachel Golub, my thesis work would not have been possible. I would also of course like to thank the many collaborators I have worked with associated with many other labs at the University of Chicago.

I would like to thank the Graduate Program in the Biophysical Sciences for giving me the opportunity to study at the University of Chicago and helping me navigate a dual-mentored thesis. I am especially grateful to Dr. Michele Wittels for her constant assistance and support. A very special thanks goes to the members of my Biophysics class for becoming my close friends in our early years. I cannot imagine working through these years without Eugene Leypunskiy and Kevin Song.

Many many thanks to my parents and my sister Hanna, who I could always count on for support and smiles.

Finally, I would like to thank the other members of my committee, Dr. Matthew Stephens and Dr. Marcus Clark, for there always insightful discussions and helpful input throughout my studies.

ABSTRACT

Innate lymphoid cells (ILCs) are a recently identified subset of the innate immune system found to have transformative roles in integrating innate and adaptive immune responses. In a wide distribution of tissues, ILCs are important sources of cytokines that promote inflammation, host defense against infection, tissue repair, regulation of microbiota, and physiological homeostasis. As a newly appreciated lineage, many aspects of ILC development and differentiation are not well understood. The Bendelac lab recently identified a common ILC precursor (ILCP) in the fetal liver and bone marrow on the basis of expression of PLZF, the NKT master regulator. The ILCP gives rise to all ILC lineages, including ILC1, ILC2, and ILC3, but not LTi or NK. The developmental stages and associated regulatory factors surrounding the emergence and differentiation of the ILCP have yet to be defined. We present three experimental designs that probe transcriptional regulation in ILC precursors. First we use genome-wide expression profiling to describe the role PLZF plays in distinguishing cNK and ILC1 lineages. Next we performed computational clustering of single ILC progenitor expression profiles to establish a hierarchy of ILC development and found: 1) There is a specific developmental progression of transcription factor induction upstream of the ILCP. 2) ILC differentiation occurs after acquisition of PLZF in the ILCP. 3) LTi specification immediately precedes ILC differentiation and is a distinct lineage decision. 4) ILC differentiation occurs through multilineage transcriptional priming. Finally, we compared chromatin accessibility transcriptional profiling in the ILPC to identify the predominant transcriptional regulators affiliated with ILC specification. The precise elaboration of ILC developmental stages and identification of novel ILC developmental regulators will improve our understanding of the functional requirements of ILCs and their roles in immunity.

CHAPTER 1

INTRODUCTION

The immune system is responsible for our bodys protection from disease-causing pathogens. The diverse set of effector cells and molecules that comprise the immune system form an interacting network that performs surveillance, generates specific response, and retains immunological memory of an infection. These immune functions are broadly accomplished through the two primary arms of the immune system, the adaptive and innate immune systems. Adaptive immunity is realized through cells that develop pathogen-specific responses and memory through non-deterministic rearrangement of antigen-binding receptors. In contrast, innate immune effectors non-specifically respond to pathogens by inducing immediate inflammation and priming the adaptive immune system for action.

Innate lymphoid cells (ILCs) are a group of innate immune effectors that have only been functionally appreciated in the past decade. ILCs are primarily found in mucosal barrier tissues such as the skin, lung, and intestine and mediate rapid cytokine response to infection and tissue injury. Distinct ILC subtypes are poised to generate cytokine response profiles that mirror those of the canonical T helper cell subtypes that control intracellular pathogens (type 1), large parasites (type 2), and extracellular microbes (type 3) [1]. Importantly, ILCs lack the characteristic antigen-specific receptors of adaptive T cells and respond via generic germline-encoded receptors [2]. Accordingly, while ILCs generate effector responses aligned with particular T cell programs, ILCs have separate functional roles in promoting protection from disease, preventing autoimmunity and asthma, and maintaining tissue homeostasis among other physiological functions [2].

1.1 ILC Development

Since ILC subtypes and functional roles have been only recently characterized, there has been considerable interest in understanding the developmental requirements of these new

lineages. All distinct cells of the immune system derive from pluripotent hematopoietic stem cells in the bone marrow, belonging to either the lymphoid or myeloid lineages. Interestingly, although ILCs belong to the innate immune system, ILCs differentiate from the common lymphoid progenitor (CLP) [3], which is also responsible for giving rise to the adaptive B and T cell lineages. In addition to shared cytokine response profiles, ILC effectors also share the respective transcriptional determinants of T helper cell effectors in addition to several common developmental regulators. First we will discuss the characterization of cell types comprising the three groups of ILCs, type 1, type 2, and type 3 ILCs [4], and then elaborate on the current understanding of ILC development and differentiation.

1.1.1 Types of ILCs

Type 1 ILCs include conventional natural killer cells (cNK) and ILC1 and are characterized by expression of the transcriptional master regulator T-bet (encoded by the Tbx21 gene) and effector cytokine IFNγ, similar to Th1 effector cells. Although cNK and ILC1 are phenotypically similar, they possess a few key distinguishing characteristics. cNK are circulating cells that express the transcription factor Eomes, while ILC1 are tissue-resident cells that do not express Eomes. ILC1s exclusively express CD49a (integrin α1 subunit of $\alpha_1\beta_1$) and Trail while cNK exclusively express DX5 (also known as CD49b or integrin α2 subunit of $\alpha_2\beta_1$) [5]. The strong similarities in expression profiles and relative instability of these distinguishing markers can make these lineages difficult to distinguish in practice.

Type 2 ILCs include ILC2 only, which is characterized by expression of the transcriptional master regulator Gata3 and effector cytokines Il-5, Il-13, and the growth factor amphiregulin, similar to Th2 effector cells. ILC2 respond to IL-25 and IL-33 signaling and have been shown to be protective in helminth infection independent of the adaptive immune system [6]. In addition to mediating response in lung and skin inflammation, ILC2 have also been shown to regulate fat homeostasis [7, 8]. Although Gata3 only remains predominantly expressed in ILC2s among differentiated ILCs, conditional deletion of Gata3 in hematopoietic cells

affects all ILC lineages except cNK [9] and among differentiated ILCs, Gata3 is required for the maintenance of ILC1s and ILC2s, but not ILC3s [10]. In T cell development, precise regulation of Gata3 expression is necessary for T cell lineage specification and complete activation of T cell-specific genes prior to effector differentiation as well as for maintenance of the Th2 program.

A number of transcription factors are specifically necessary for ILC2 development in addition to Gata3, including Bcl11b, Gfi1, and RORα. The deletion of any of these factors results in dramatically reduced numbers of peripheral ILC2s as well as bone marrow ILC2 precursors [11, 12, 13, 14, 15]. Gata3 and Gfi1 cooperatively upregulate and maintain each others expression while Bcl11b directly controls Gfi1 expression. Interestingly, these factors have distinctly different degrees of importance in T cell development. RORαonly has a minor role in Th17 development while Gfi1 particularly promotes Th2 over Th17 and Treg differentiation and similarly maintains Gata3 expression. In contrast, Bcl11b deficiency results in an early block in T cell development with aberrant derepression of innate lineage factors [16]. In fact, Bcl11b-deficient T cell precursors acquire cNK lineage phenotypes [16, 17].

Type 3 ILCs are comprised of the heterogeneous populations lymphoid tissue inducer cells (LTi), NCR$^+$ (natural cytotoxicity receptor) ILC3s, and NCR$^-$ ILC3s that are all characterized by expression of the transcriptional master regulator RORγt and effector cytokines IL-17 and IL-22, similar to Th17 effector cells. In developing fetal tissue, LTi prime stromal cells to attract lymphocytes and form secondary lymphoid tissue structures, such as lymph nodes, Peyers Patches and Isolated Lymphoid Follicles [18]. Among type 3 ILCs, LTi exclusively express Ccr6 and partially express CD4 [19]. NCR$^+$ ILC3s differentiate from NCR$^-$ ILC3s following the upregulation of T-bet and NK marker Nkp46, a process which does not occur in LTi [20, 21, 22, 23]. The differentiation or NCR$^-$ to NCR$^+$ ILC3s requires Notch signaling as well as Gata3 mediated repression of RORγt [24, 25]. Furthermore, downregulation of RORγt in NCR$^+$ ILC3s can lead to complete loss of the ICL3 phenotype, resulting in

conversion from NCR^+ ILC3 to ILC1. Accordingly, RORγt is required for the development of LTi and ILC3 but not cNK, ILC1, or ILC2, and also appears to modulate the function of type 3 ILCs. Similarly, RORγt is required for Th17 differentiation, although the nuclear receptor appears to influence earlier thymocyte maturation as well. In line with the observed plasticity between ILC1 and ILC3 lineages, the runt-related transcription factor Runx3 is necessary for the development of type 1 and type 3 ILC lineages [26]. Runt-related transcription factors such as Runx1, Runx3 and Th-POK (encoded by Zbtb7b gene) also play important roles in T cell development [27], as Runx1 is required for T cell maturation while Runx3 and Th-POK are antagonistic regulators in CD4-CD8 differentiation [28].

1.1.2 ILC Precursors

To better understand the differentiation pathways of these varied ILC subtypes, recent efforts have focused on identifying intermediate ILC precursor populations. These studies have not only uncovered important early development requirements of ILC lineages, but have also clarified lineage distinctions among phenotypically similar cell types. Here we survey a selection of ILC precursors and transcriptional regulators that influence these developmental stages.

ILCP

In recent years, the Bendelac group determined that ILC1, ILC2, and ILC3 are developmentally distinct lineages from cNK and LTi through investigation of the natural killer T (NKT) cell master regulator PLZF in ILC development [29]. Constantinides et al. fate mapped PLZF expression by expressing a GFP-Cre fusion gene under the control of the *Zbtb16* (which encodes PLZF) promoter in conjunction with a Rosa-26-flox-stop-flox-YFP strain, resulting in persistent YFP expression in any cells that previously upregulated PLZF. These experiments showed that peripheral ILCs but not cNK or LTi were marked by a history of transient PLZF expression, suggesting the existence of a PLZF-expressing precursor

to ILCs excluding cNK and LTi. The Zbtb16 reporter mouse strain identified a Lin$^-$ $\alpha_4\beta_7{}^+$ IL-7Rα^+ PLZF$^+$ precursor in the bone marrow and fetal liver termed the innate lymphoid cell precursor (ILCP). Transfer studies of ILCP confirmed that the cell could efficiently give rise to all ILCs but not cNK, LTI, B and T cells. Moreover, in in vitro culture single ILCP could produce all three ILC types, showing that the ILCP contains a true multi-potential precursor for all ILCs. Since ILCP produce multi-potential colonies at a low frequency (12.2% of all colonies) and PLZF is relatively rapidly downregulated in culture as effector cell markers arise, the ILCP is likely developmentally close to lineage trifurcation. Although PLZF is expressed in the common precursor to ILC lineages, deletion of PLZF primarily affects ILC2 and ILC1 but not ILC3. In NKT cells, PLZF regulates a number of factors responsible for ILC2 specification [30]. In T cells, PLZF represses the expression of Bach2 leading to an overall promotion of T effector programs over T regulatory programs [31]. In ILCs, PLZF may likewise be promoting ILC effector programs through repression of Bach2.

CHILP

The E protein inhibitor Id2 was recognized early as a requirement for the development of all ILC lineages including cNK and LTi [32, 33, 34]. E proteins are DNA binding proteins essential for B and T cell development [35]. In particular, the E protein E2A induces early expression of B cell lineage-defining transcription factors and is required for necessary notch signaling in T cell development [36]. Inhibitor of DNA binding (Id) proteins heterodimerize with E proteins and prevent them from binding to DNA. The overexpression of Id2 or related family members is sufficient to inhibit B and T cell differentiation from the CLP [37, 38], suggesting that in ILC development Id2 is responsible for repressing competing lineages. Although the ILCP expresses Id2, the additional dependence of cNK and LTi on Id2 suggests the existence of an upstream precursor with Id2 expression.

In pursuance of this reasoning, Klose et al. used an Id2 reporter mouse strain to identify a Lin$^-$ $\alpha_4\beta_7{}^+$ IL-7Rα^+ CD25$^-$ Id2$^{\text{high}}$ precursor that could give rise to ILC1, ILC2, ILC3

and LTi, termed the common helper innate lymphoid precursor (CHILP) [39]. The lineage potential of the CHILP was determined through transfer studies and single-cell cultures. Further dissection of the CHILP population suggests that is actually a mixture of distinct precursors. Given that ILCP express Id2, it is not surprising that CHILP present bimodal expression of PLZF and likely contains ILCP. While the CHILP was shown to not express RORγt in the bone marrow, bone marrow precursors generally show little RORγt expression and very inefficiently give rise to LTi [40]. In the fetal liver, where LTi are much more efficiently produced, the CHILP could be expected to express RORγt and potentially contain LTi precursors. Although the CHILP was shown to give rise to all ILC lineages in single-cell cultures, crucially LTi and cNK could not be differentiated from ILC3 and ILC1 lineages and so it remains to be determined whether the CHILP has ILC1, ILC2, ILC3 and LTi multi-potential or is a mixture of separately lineage-committed precursors.

EILP

A bone marrow precursor to all ILC lineages including cNK and LTi was recently discovered using a Tcf7 reporter mouse strain [41]. Identified on the basis of Tcf7 expression, the early innate lymphoid precursor (EILP) is further defined by Lin$^-$ $\alpha_4\beta_7{}^+$ Thy1$^-$ Il-7Rα^- and in transfer studies gives rise to all ILCs including cNK and LTi but not B or T cells, showing that it is committed to ILC lineages. The expression profile of EILP corroborates a role as an early precursor. The EILP does not express PLZF and expresses lower levels of Id2 than the CHILP. The EILP however does express high levels of the transcription factors Tox and Nfil3, which have both been shown to be required for development of all ILC lineages. In single-cell cultures, the EILP could give rise to ILC1, ILC2, and ILC3, although it is not clear if cNK and LTi can be reliably distinguished in these models. The increased proportion of multi-potential colonies generated from single EILP compared to ILCP and CHILP also points to an early developmental role for the EILP. The lack of Il-7Rα expression on the EILP is a noteworthy peculiarity given the expression of the cytokine receptor chain in all

other ILC progenitors including the CLP as well as in peripheral ILC lineages including the LTi. It is unclear if downregulation of Il-7Rα is an alternative ILC developmental path or a transient state that perhaps allows for differentiation of cNK, which do not expression Il-7R and do not require Il-7 signaling for development [42].

The high-mobility group (HMG) box transcription factor Tcf7 has an integral role in both early ILC development and in early T cell development. In ILC development, Tcf7 is required for the efficient differentiation of all ILC lineages but interestingly specifically represses the ILC3 effector program in differentiated ILCs [43, 44, 45]. In T cell development, Tcf7 cooperates with Lef1 and β-catenin and is necessary for efficient early T cell development as well as in later $CD4^+$ differentiation through regulation of Th-POK [46, 47, 48, 49]. While Tcf7 expression is directly regulated by Notch signaling in T cell development, Tcf7 appears capable of being induced independently of Notch signaling in ILC development [24]. Tox is also a HMG box transcription factor required for T cell development progression past double positive thymocyte precursors [50]. In ILC development, Tox deficiency leads to severe reductions in all ILC lineages including cNK and LTi as well as reductions in CHILP among other ILC precursor populations [51, 52]. The basic leucine zipper transcription factor Nfil3 on the other hand does not appear to have a direct role in T cell development. While Nfil3 deficiency also leads to substantially reduced cell numbers of all ILC lineages including LTi and particularly cNK [53, 54, 55, 56, 57], Nfil3 does not appear to be required for the maintenance and survival of mature ILCs in at least some lineages. Nfil3 appears to be capable of directly regulating cNK defining transcription factor Eomes and Id2 through promoter binding as well as indirectly regulating Id2 through transcriptional regulation of Tox [58, 59].

Two additional ILC precursors have been identified on the basis of expression of other proteins rather than particular transcriptional regulators. The $\alpha_4\beta_7$-expressing lymphoid progenitor (αLP) is an early precursor for all ILC lineages including cNK and LTi found in the bone marrow [59]. The αLP is defined through the profile Lin^- $cKit^{lo}$ $Sca1^{lo}$ $Flt3^-$

Il-7Rα^+ $\alpha_4\beta_7{}^+$ and does not express PLZF. Consistent with its interpretation as an early developmental stage, in single-cell cultures the aLP frequently generated multi-lineage ILC colonies like the EILP, although in these experiments cNK and LTi could not be reliably differentiated from ILC1 and ILC3 respectively. A fetal intestinal precursor to all ILCs was identified on the basis of expression of Arginase 1 (Arg1), a urea cycle enzyme that metabolizes the amino acid L-arginine [60]. Arg1 is expressed in all ILC types in the fetal intestine as well as in peripheral ILC2 and bone marrow ILC2P and defines an Arg1$^+$ intestinal precursor that is Lin$^-$ $\alpha_4\beta_7{}^+$ IL-7Rα^+ Arg1$^+$ NK1.1$^-$ ST2$^-$ and negatively fate-mapped for RORγt. These precursors coexpress T-bet, Gata3, and RORγt, which implies a degree of multi-lineage priming, and in single-cell culture generate dual but not triple ILC lineage colonies. This suggests that the Arg1$^+$ intestinal precursor is likely a recent emigrant from the fetal liver as they appear to the developmentally staged after the ILCP and closer to differentiation.

In addition to transcriptional regulators, the notch, aryl hydrocarbon receptor (AhR) and retinoic acid (RA) signaling pathways all have implicated roles in ILC development. While notch signaling is strictly required in thymic T cell development [61], it appears to be influential but somewhat dispensable in ILC development. Notch signaling has been suggested to contribute to upregulation of $\alpha_4\beta_7$ on CLP [62], RORγt in bone marrow ILC precursors [63], and promote differentiation of ILC2 at the αLP stage [64, 41, 14]. Although it has been suggested that Tcf7 upregulation depends on notch signaling, conditional deletion of the notch transducer Rbpj did not affect expression of Tcf7 in ILC precursors [24]. The AhR responds to dietary ligands and is expressed by intestinal ILC3 [65]. While it has been shown that AhR deficiency impacts ILC3 and LTi populations and AhR dietary ligands can regulate ILC3 population size [21, 66, 67], it is unclear if AhR signaling could act in bone marrow ILC precursors. RA signaling is also influential in peripheral ILC homeostasis, where retinoic acid, or vitamin A, augments the expansion and function of ILC3s over ILC2s [68]. In fetal liver, maternal retinoids are required for the early generation of type 3 ILCs through

regulation of RORγt, implying an impact in ILC development [69].

While it is apparent that many ILC developmental factors and precursors have been discovered, a missing component in understanding is the coherent picture of their relationships. This includes both defining the precise sequence of ILC developmental stages and identifying the ILC context specific molecular interactions and mechanism that drive the developmental progression.

This condensed survey is not a comprehensive review of all influential factors in ILC development. For further discussion, see the dissertation of Ishizuka (ref [70])

1.2 Methods of Genomic Analysis

Methods for interrogating transcriptional regulation are rapidly transforming. Where studies were typically modeled around targeted interrogation of single regulators and a few downstream targets, there is increasing reliance on the use of genomic methods for both identifying novel target regulators as well as interpreting global regulator influence. We highlight some considerations for analysis of modern genomic methods of investigation, focusing on transcriptional and chromatin accessibility profiling.

1.2.1 Population-level transcriptional profiling

The ability to assay complete transcriptomes at a time has had a transformative affect on biological experimental design and discovery. Large-scale transcriptional profiling methods were initially spearheaded by microarray technologies, in which gene transcripts are detected through hybridization to probe sequences on microchips, but have been largely supplanted by next generation sequencing (NGS) technologies. This shift in adoption has been driven by dramatic cost reductions in recent years for NGS experiments as well as their increased flexibility in experimental design, which has been adapted to survey a wide variety of sequence-associated quantities in addition to transcriptional profiles. For our purposes,

we will briefly overview a few important considerations for the analysis of population-level microarray and RNA-sequencing (RNA-seq) transcriptional profiling.

One of the initial analysis steps is performing quantification and normalization of raw transcriptional profiling data. For microarray experiments, quantification is typically straightforward as the measured signal is derived from fluorescence channel measurements. For quantification-focused RNA-seq experiments, raw sequencing reads need to first be mapped to their originating genomic locations and then counted according to their association with distinct transcripts. There are many algorithm choices for each of these steps depending on analysis objectives [71], for example gene-level or transcript-level quantification, as well as recent methods for quantification that bypasses the computationally costly read mapping step [72]. Following quantification, normalization of transcript levels between measured samples is crucial for reliable comparison. For microarray experiments, slight differences in imaging conditions or hybridization efficiencies can lead to large-scale technical biases in measured signals. For RNA-seq experiments, variation in read library sizes, transcript lengths, and potential sequencing biases need to be accounted for before comparing expression levels. To account for read library sizes and transcript lengths, raw RNA-seq read counts are typically converted to RPKM (reads per kilobase of exon model per million reads) or TPM (transcripts per million) expression measures, although substantial technical biases can still exist if samples have heterogeneous transcript distributions. For both microarray and RNA-seq experiments, substantial technical bias can be expected if samples are isolated in different facilities, measured on different machines, or even measured in different runs on the same machine. For this reason, between sample normalization procedures are often necessary before comparison. These normalization procedures are typically based on the notion that distribution of gene expression levels should have similar characteristics between samples and a few examples include quantile and robust multi-array average (RMA) normalization for microarrays [73] and UpperQuartile and trimmed method of means (TMM) normalization for RNA-seq.

Typically transcriptional profiling experiments are interested in differential expression, or the statistical identification of genes with changing expression levels between conditions. An important consideration for these experiments is that they are typically high dimensional low sample size (known as HDLSS), since whole transcriptome measurements can include tens of thousands of genes and individual conditions are often only replicated several times due to cost. Accordingly, these analyses require making a large number of comparisons with limited ability to estimate within-sample gene expression variance. While it might seem surprising that there is observable variance in samples that are each averages of millions of cells, the variance between biological replicates reflects whole-population shifts of gene expression in response to slight differences in experimental conditions or in the isolation of specific cell types, variability which affect different genes to different degrees. The difficulty in addressing multiple testing for differential expression is that across all genes, at a given level of significance the number of false positives could easily outnumber the number of true positives. Instead, we are often interested in controlling the false discovery rate (FDR), which instead estimates significance levels that correspond to different expected proportions of false positives [74]. A key concept for approaching the issue of imprecise gene-specific variance estimates with HDLSS data is variance shrinkage, in which variance estimates across all genes are used to regularize individual estimates [75]. Empirical Bayes strategies for variance shrinkage are commonly implemented in differential expression analysis pipelines. These analysis strategies have been packaged for easy application, as by limma for microarray experiments [76], and Deseq2 [77] and edgeR [78] for RNA-seq experiments.

1.2.2 Single-cell transcriptional profiling

Recent developments in optimizing sequencing assays to work at the level of individual cells and in producing new instruments that isolate and interrogate individual cells has opened the field of genomics to single-cell study. The predominant technologies that measure single-cell gene or protein expression levels include single-cell multiplex qPCR, single-cell RNA-

seq, and mass cytometry. In single-cell multiplex qPCR experiments, individual cells are isolated in microfluidic devices and tens of quantitative real-time qPCR gene expression measurements per cell are performed on hundreds to thousands of cells. In single-cell RNA-seq experiments, complete cDNA libraries for individual cells are isolated, amplified and then sequenced for hundreds to thousands of cells. In mass cytometry experiments, proteins are bound by tens of antibodies conjugated to heavy metal identifiers and then measured by mass spectrometry for hundreds of thousands of cells. The analysis of single-cell genomic data present distinct challenges from population-level transcriptional profiling but also enable new kinds of biological interpretation.

Compared to population-level measurements, single-cell expression measurements have a much higher degree of variability and require tailored quantification and normalization procedures before comparison. Single-cell multiplex qPCR measurements generate relative expression signals to housekeeping gene levels as an internal reference. Single-cell RNA-seq experiments are exploring the use new normalization methods based on molecular barcoding or spike-in standards to distinguish technical and biological variability. A new normalization challenge relevant to all single-cell measurements is the high rate of drop-out events, where expression measurements for individual genes can be entirely missed in measurement, and has only very recently been addressed [79]. With a set of single-cell transcriptional libraries, the analysis objectives are very different from population-level transcriptional profiling. Whereas in population-level transcriptional profiling experimentalists are typically interested in some form of differential gene expression between conditions, with single-cell profiles experimentalists are typically more interested in understanding the relative relationships between individual cell profiles [80]. When interrogating a heterogeneous population of cells, this could mean clustering individual cell profiles to identify distinct subpopulations. In a developmental context, we are often interested in understanding cellular trajectories through differentiation or developmental paths. Recently a number of algorithms have been developed that aim to organize single-cell transcriptional profiles according to the notion of

developmental pseudotime, a quantitative measure of biological progression through potentially branching trajectories, including Monocle 2 [81], Wishbone [82], Diffusion Pseudotime (DPT) [83], and SLICER [84]. These methods follow a roughly similar strategy for analysis: first, dimensionality reduction is performed on high-dimensional expression profiles; next, a graph or network is constructed that relates adjacent data points in the dimensionality reduced space; finally, the constructed graph is used to identify distinct branches and assign a developmental pseudotime to each cell. A critical aspect for any of these methods is the calculation of an appropriate distance measure between individual cell profiles. While single-cell RNA-seq experiments have the benefit of measuring all available transcripts without bias, it is necessary to perform feature selection so that tightly correlated gene modules do not dominate distance calculations, for example the large set of cell-cycle genes can confound a developmental trajectory if not accounted for [85]. For single-cell multiplex qPCR and single-cell mass cytometry, pre-selection of the panel of genes to be measured typically insures there will not be dominant clusters of highly correlated confounding genes but requires experimentalists to be careful the panel does not miss any crucial factors for distinguishing potential subpopulations. The output of single-cell trajectory analysis is extremely informative for studying transcriptional regulation, since a complete continuous developmental ordering of transcriptional profiles can indicate the timing of induction for transcriptional regulators with fine-grained resolution.

1.2.3 Chromatin accessibility profiling

In eukaryotes, the organization of regulatory DNA into regions of open or closed chromatin both modulate and reflect the binding of transcription factors and subsequent regulation of transcriptional activation. Methods for genome-wide survey of chromatin accessibility such as MNase-seq, FAIRE-seq, DNa-seq and ATAC-seq separate the genome by chemically or enzymatically isolating either accessible or protected regions and then quantify the degree of accessibility through NGS sequencing [86]. Here we specifically highlight ATAC-seq [87],

Assay for Transposase-Accessible Chromatin using sequencing, based on fragmentation of DNA and insertion of sequencing adapters by the transposase Tn5, since it is the most current method for assaying open chromatin and has dramatically reduced the number of cells required for interrogation from millions to just thousands. The comparison of changing enhancer regions with changing expression levels leads to more a detailed understanding of regulatory mechanisms.

A chromatin accessibility profiling experiment like ATAC-seq generates a library of reads that derive from genomic regions with increased chromatin accessibility. Like in RNA-seq experiments, these raw sequencing reads need to be mapped to their originating genomic locations and then quantified by feature for comparison. Accessible regions are not predefined like transcript exons in RNA-seq, so these features must first be algorithmically identified. There are a number of peak calling methods that effectively aim to identify short contiguous regions with accumulations in reads well above background levels [88]. The number of features identified depends on the significance threshold used and needs to be calibrated according to concordance between biological replicates. Currently, methods for identifying changes in chromatin accessibility are adapted from differential expression methods designed for RNA-seq, where read counts associated with peaks replace read counts associated with gene transcripts [89]. While these methods reliably identify changes in read-count distributions, they ignore lineshape information contained in accessibility profiles, which has been demonstrated to be predictive of transcription factor binding [90]. We are often interested in the presence of transcription factor binding motifs in the identified regulatory regions. Methods for identifying binding motifs either focus on de novo motif identification, where sequences are searched for overrepresented nucleotide combinations, or on the enrichment of known motif consensus sequences, where motif-defining position weight matrices are scanned for high probability matches in target sequences. An important consideration is that the presence of motif consensus sequences is not absolutely predictive of true transcription factor binding, which can depend on a variety of factors, such as interaction

with co-factors, binding of competing transcription factors, and the activation state of the transcription factor. Currently there remains a gap in understanding of how to connect regulatory enhancers defined by chromatin accessibility profiling and transcription.

1.3 AIMS AND SIGNIFICANCE

In a wide distribution of tissues, ILCs are important sources of cytokines that promote inflammation, host defense against infection, tissue repair, regulation of microbiota, and physiological homeostasis. This rich functional diversity is achieved through the actions of three distinct ILC effector programs that strikingly parallel the three types of T helper cell subsets in the adaptive immune system. However, in contrast with adaptive T helper cells, ILCs acquire their different effector programs during development, rather than in response to specific immune challenges in the periphery. In this way, not only is the ILC differentiation pathway a compelling developmental system for study because it involves the resolution of a lineage trifurcation, but it also is an instance of a biological system that takes an alternative path to variations of common phenotypic fates. In comparison with the developmental requirements of the NKT and T cell lineages, different developmental aspects are shared in ILC development. As illustrated in the introductory discussions, many common factors between ILC and T cell development have distinct impact and sufficiency requirements in the two lineages. Additionally, there is a clear set of unique developmental factors specific for ILC development. The further identification and characterization of ILC developmental regulators that govern the acquisition and resolution of distinct ILC effector programs will lead to a better understanding of how common phenotypic programs are diversely utilized by distinct cell lineages. Moreover, the association of new regulatory factors with specific association with ILC developmental stages will enable the generation of new experimental models to study the roles of ILCs without compromising other immune function.

We use modern genomic approaches, including both population-level and single-cell transcriptional profiling as well as chromatin accessibility profiling, to resolve the developmental

progression that occurs during ILC development.

The aims of my studies are:

- Dissect the contribution of PLZF expression to the divergence between ILC1 and NK lineages by microarray expression profiling
- Define transcriptional progression in early ILC development through single-cell profiling of PLZF-GFPCre$^+$ ILCP and $\alpha_4\beta_7{}^+$ ILC precursors.
- Computationally identify the branch point of LTi and ILCs.
- Characterize the stage of early ILC lineage trifurcation where cells acquire effector programs.
- Identify and characterize the predominant transcriptional regulators at the ILCP stage through comparison of transcriptional and chromatin accessibility profiling.

CHAPTER 2

PLZF EXPRESSION MAPS THE EARLY STAGES OF ILC1 LINEAGE DEVELOPMENT

Michael G. Constantinides[a], Herman Gudjonson[b], Benjamin D. McDonald[a], Isabel E. Ishizuka[a], Philip A. Verhoef[a], Aaron R. Dinner[a,b,c], and Albert Bendelac[a,1]

[a] Committee on Immunology,

[b] Institute for Biophysical Dynamics, and

[c] Department of Chemistry, University of Chicago, Chicago, IL 60637

[1] To whom correspondence should be addressed. Email abendela@bsd.uchicago.edu

Author contributions: M.G.C. and A.B. designed research; M.G.C., H.G., B.D.M., I.E.I., and P.A.V. performed research; M.G.C., H.G., A.R.D., and A.B. analyzed data; M.G.C. and A.B. wrote the paper; and H.G. performed bioinformatics analysis.

Published in PNAS March 11, 2015; doi:10.1073/pnas.1423244112

2.1 SIGNIFICANCE

Diverse populations of group 1 innate lymphocytes, which exert critical early cytolytic functions against virally infected cells, have recently been discovered, raising issues of lineage relationships. We used expression of the transcription factor promyelocytic leukaemia zinc finger (PLZF) to identify the developmental intermediates of innate lymphoid cells type 1 (ILC1s), a subset of innate lymphoid cells that are particularly abundant in the liver, and demonstrated that this lineage arises from a distinct precursor, but that its development partially overlaps with established classical NK stages. Using microarray analysis, we defined a set of PLZF-dependent genes that may contribute to lineage divergence between ILC1s and classical NK cells.

2.2 ABSTRACT

Among the variety of tissue-resident NK-like populations recently distinguished from recirculating classical NK (cNK) cells, liver innate lymphoid cells (ILC) type 1 (ILC1s) have been shown to represent a distinct lineage that originates from a novel promyelocytic leukaemia zinc finger (PLZF)-expressing ILC precursor (ILCP) strictly committed to the ILC1, ILC2, and ILC3 lineages. Here, using PLZF-reporter mice and cell transfer assays, we studied the developmental progression of ILC1s and demonstrated substantial overlap with stages previously ascribed to the cNK lineage, including prepro-NK, pre-NK precursor (pre-NKP), refined NKP (rNKP), and immature NK (iNK). Although they originated from different precursors, the ILC1 and cNK lineages followed a parallel progression at early stages and diverged later at the iNK stage, with a striking predominance of ILC1s over cNKs early in ontogeny. Although a limited set of ILC1 genes depended on PLZF for expression, characteristically including *Il7r*, most of these genes were also differentially expressed between ILC1s and cNKs, indicating that PLZF together with other, yet to be defined, factors contribute to the divergence between these lineages.

2.3 INTRODUCTION

NK cells represent the longest and best studied population of innate lymphocytes, whose roles in host defense against infections, particularly viral infection by cytomegalovirus, and in tumor or transplant rejection have been extensively characterized [91, 92, 93]. The core NK program is based on a panoply of NK-lineage receptors specific for MHC and MHC-like ligands that are induced by stress and infection to trigger perforin- and granzyme-mediated cytolysis of target cells and secretion of type 1 cytokines. This program is controlled in part by the transcription factor T-bet, the related factor eomesodermin, and the cytokine IL-15.

NK cells were long considered to be a single lineage phenotypically characterized by the surface CD3ε^- NK1.1$^+$ DX5$^+$ profile of its mature product. However, several recent studies have revealed the existence of distinct subpopulations of NK-like cells that, unlike the recirculating classical NK (cNK) cells found in spleen, blood, and lymph nodes, exhibited the unusual property of long-term residence in particular tissues, including liver, intestinal epithelium, uterus, skin, and salivary glands [5, 94, 95, 96, 20, 97]. For example, in parabiotic pairs of CD45$^-$ congenic mice, a fraction of CD3ε^- NK1.1$^+$ liver cells were found to be tissue-resident rather than recirculating cells (6). The resident cells were further distinguished from cNK cells by the expression of the TNF family member TNFSF10 (TRAIL), the presence of high levels of integrin α1 chain (CD49a), the absence of integrin α2 (CD49b, stained by the DX5 antibody), and the expression of CD160, a receptor for epithelium-expressed HVEM [5, 95, 20, 39]. Whereas, like cNKs, the liver-resident cells expressed T-bet and were endowed with cytolytic and type 1 cytokine secretion properties, they characteristically lacked eomesodermin and expressed some level of surface IL7Rα chain [5, 39, 98]. Although their DX5$^-$ phenotype initially suggested they were immature NK cells [99, 100, 101], further studies based on fate mapping and cell transfers into Rag$^{-/-}$ Il2rg$^{-/-}$ hosts established that they represented a different lineage, originating from a small subset of lineage (Lin)$^-$ IL7Rα^+ $\alpha_4\beta_7^{\text{high}}$ Id2$^{\text{high}}$ bone marrow or fetal liver cells that transiently but characteristically expressed high amounts of the transcription factor promyelocytic leukaemia zinc finger (PLZF)

[39, 29]. This PLZF-expressing precursor was termed the ILC precursor (ILCP) because it also generated ILC2s and ILC3s, but, importantly, generated only few cNK-like cells and did not give rise to LTis, B, or T cells [29]. Consistent with these findings, most innate lymphoid cells type 1 (ILC1s) were fate-mapped by PLZF, whereas, in contrast, cNKs were generally not fate-mapped, with the notable exception of a small fraction of cNKs, which perhaps originated from ILCPs that retained some cNK potential or, alternatively, were derived from rare cNK precursors that might express low level of PLZF. Thus, these findings altogether supported the notion that ILC1s and cNKs primarily originated from different precursors. Furthermore, although there remain a few discrepancies in the description of these cells by different groups, in particular regarding the degree of requirement for the transcription factor Nfil3 [5, 59], it seems clear that the liver-resident NK cells described by Yokoyama and coworkers are largely identical to liver ILC1s [5].

The early stages of NK cell development have long been elusive. A CD3ε^- CD122 (IL2Rβ)$^+$ NK1.1$^-$ NK precursor (NKP) subset was first identified based on its ability to generate NK1.1$^+$ cells upon single cell culture in vitro, albeit with a relatively low clonal frequency of $\sim$1 in 12 [102, 103]. More recent studies, however, have refined the phenotype of these precursors based on the coexpression of CD122 with CD127 (IL7Rα), CD244, and CD27, identifying a refined NKP (rNKP) as well as an earlier, so-called pre-NKP stage, with a similar phenotype, except for low or absent CD122 [104]. At the single cell level, both of these populations could generate CD3ε^- NK1.1$^+$ clones at a high frequency of $\sim$1 in 2 in culture with OP9 stromal cells and a mixture of the cytokines SCF, Flt3L, IL-7, and IL-15. In addition, after in vivo transfer into Rag$^{-/-}$ Il2rg$^{-/-}$ hosts, both populations generated exclusively CD3ε^- NK1.1$^+$ cells in the spleen and liver. Another study used an Id2-GFP reporter strain to identify a population of Lin$^-$ Id2$^{\text{high}}$ Sca1$^+$ CD117 (cKit)$^{\text{int}/-}$ CD135 (Flt3)$^-$ IL7Rα^+ cells, termed prepro-NKs [105], that generated CD3ε^- NK1.1$^+$ cells at a frequency of $\sim$1 in 2 upon single-cell culture in vitro with OP9 cells and a mixture of IL-7 and IL-15.

In this study, we used PLZF-IRES-GFPCre reporter mice and cell transfers to examine whether some of these early stages of development might be intermixed with, or reassigned to, ILC1s. We demonstrated considerable heterogeneity, with precursors to ILC1s but also ILC2s and ILC3s present in various proportions at different stages. Furthermore, by transcriptional analysis of PLZF-deficient ILC1s compared with WT ILC1s and cNKs, we identified a PLZF-dependent contribution to the genetic program that differentiates ILC1s from cNKs, which included the characteristic expression of the IL7Rα chain. The results help redefine and clarify the early stages of ILC and cNK development and the contribution of PLZF to the divergence of these lineages.

2.4 RESULTS

2.4.1 Early NK1.1$^-$ Stages of Development

We used PLZF-IRES-GFPCre mice to report and fate-map PLZF expression in lymphoid compartments. We previously established that GFP faithfully reported PLZF expression in the hematopoietic compartment, with high levels of expression in ILCP and in NKT cells, but not in other lymphoid subsets [29]. However, in fate-mapping experiments based on the ROSA26-fl-STOP-fl-YFP allele, we reported that, whereas a majority of mature ILCs and NKT cells were YFP$^+$, as expected, $\sim$30% of other hematopoietic cells were also YFP$^+$, including T, B, myeloid cells, and early hematopoietic cell precursors such as early HSC, LMPP, and CLP cells. This indiscriminate background mapping occurred before hematopoiesis, likely in early embryonic stem cells, because it was readily abrogated in radiation chimeras reconstituted with sorted YFP-negative Lin$^-$ Sca1$^+$ cKit$^+$ (LSK) precursors extracted from PLZF-IRES-GFPCre$^+$ ROSA26-fl-STOP-fl-YFP mice [29]. In the experiments reported below, we used these chimeras, termed YFP-negative chimeras, to faithfully fate-map PLZF expression in the hematopoietic system.

To precisely track the expression of PLZF side by side with fate-mapping at different

stages of lymphoid development, we compared GFP expression in straight PLZF-IRES-GFPCre^{+} mice with YFP expression in YFP-negative chimeras (Fig. 2.1). We zoomed in on the pre-NKP and rNKP stages, identified by their Lin^{-} IL7Rα^{+} CD27^{+} CD244^{+} CD117 (cKit)$^{\text{int/low}}$ Flt3^{-} profile, with the pre-NKP expressing weak or absent CD122 (IL2Rβ), and the rNKP expressing high CD122 [104] as gated in Fig. 2.1A. Both pre-NKPs and rNKPs were previously shown to selectively reconstitute CD3ε^{-} NKp46^{+} NK1.1^{+} Ly49D^{+} Ly49H^{+} cNK cells in the spleen and liver of Rag$^{-/-}$ Il2rg$^{-/-}$ mice after in vivo transfer, but a detailed analysis of markers that can identify ILC1s, such as CD49a, DX5, TRAIL, CD127 (IL7Rα), or eomesodermin, was not reported [104]. Whereas CLPs failed to express either GFP or YFP, as expected, we found that nearly half of pre-NKPs expressed high amounts of GFP (Fig. 2.1A), a property previously shown to be restricted to the ILCP [29]. Furthermore, at the subsequent rNKP stage, the expression of GFP was down-modulated, both in frequency and in intensity, whereas YFP fate-mapping increased from 15 to 36% on average (Fig. 2.1A). These findings suggested that cells with the pre-NKP phenotype in fact included a large fraction of PLZF-expressing ILCPs, which subsequently down-regulated PLZF and began to exhibit fate-mapping as they progressed into what was previously described as the rNKP stage.

We therefore extended the comparison between the GFP$^{\text{high}}$ fraction of pre-NKPs and the previously defined ILCPs. Gating on GFP$^{\text{high}}$ cells in the entire bone marrow compartment, we found that these cells not only uniformly expressed the Lin^{-} $\alpha_4\beta_7{}^{\text{high}}$ CD127 (IL7Rα)$^{\text{high}}$ CD117 (c-Kit)$^{+}$ profile, as previously described for ILCPs, but also were negative for CD135 (Flt3) and positive for CD27 and CD244 (Fig. 2.1B), the key markers of pre-NKPs [104]. In contrast, the GFP-negative fraction of pre-NKPs lacked $\alpha_4\beta_7$ expression (Fig. 2.1C). Because the transition between pre-NKPs and rNKPs is marked by the up-regulation of CD122, we reexamined CD122 expression with a bright antibody-biotin and streptavidin-phycoerythrin combination to detect low quantities of this protein on the cell surface. Indeed, we found that ILCPs expressed variable amounts of surface CD122 that ranged from negative

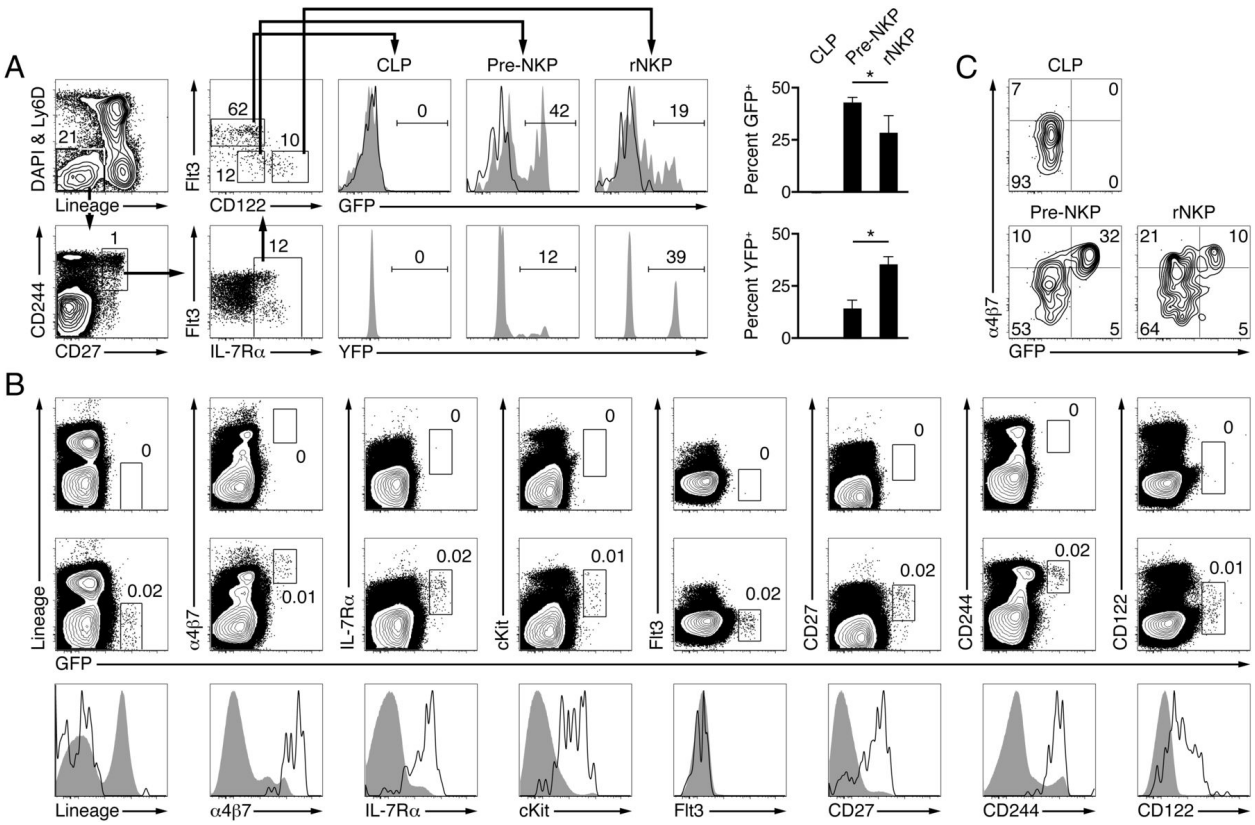


Figure 2.1: Pre-NKPs and rNKPs are heterogeneous populations.
(A) Gating strategy for CLPs (CD122$^-$ Flt3$^+$), pre-NKPs (CD122$^-$ Flt3$^-$), and rNKPs (CD122$^+$ Flt3$^-$) among Lin$^-$ Ly6D$^-$ CD27$^+$ CD244$^+$ IL-7Rα^+ bone marrow cells (left four dot plots). Histograms show expression of GFP (reporting PLZF; Top) by the indicated cell types from PLZF$^{\text{GFPcre+/+}}$ (filled gray) and wild-type (open) mice and YFP (reporting fate-mapping by PLZF; Bottom) by the same cell types in YFP-negative chimeras (radiation chimeras reconstituted with YFP$^-$ Lin$^-$ Sca-1$^+$ cKit$^+$ (LSK) bone marrow cells from PLZF$^{\text{GFPcre+/–}}$ ROSA26-YFP mice). Bar graphs show summary of results (mean ± SEM). (B) FACS analysis of Ly6D$^-$ bone marrow cells from adult WT (Top) and PLZF$^{\text{GFPcre+/+}}$ (Middle) mice. Bottom depicts the same results in histogram form with GFP$^{\text{high}}$ (open) and GFP$^-$ (filled gray) Ly6D$^-$ bone marrow cells from adult PLZF$^{\text{GFPcre+/+}}$ animals. (C) Expression of GFP and $\alpha_4\beta_7$ by the indicated cell types from PLZF$^{\text{GFPcre+/+}}$ mice. Representative of 38 mice analyzed in three or more independent experiments. $^*P < 0.05$.

to intermediate and, thus, might include cells already committing to the ILC1 sublineage (Fig. 2.1B, *Far Right*).

Thus, we conclude that the previously described pre-NKP population is composed of a mixture of ILCPs (precursors to ILC1s, ILC2s, and ILC3s) and cNK precursors. These two fractions can be distinguished by their differential pattern of expression of PLZF and $\alpha_4\beta_7$. The distinct progenies of these two fractions can be inferred from transfer experiments showing that the unfractionated pre-NKPs generated classical NK cells in the spleen [104], whereas the PLZF$^+$ $\alpha_4\beta_7{}^+$ (ILCP) fraction did not [29]. In retrospect, the failure to observe significant NK1.1$^-$ progeny (i.e., ILC2 and ILC3) after transfer of pre-NKPs into Rag$^{-/-}$ Il2rg$^{-/-}$ hosts in the study by Fathman et al. [104], is consistent with their focus on spleen and liver, whereas ILC2s and ILC3s predominate in the intestinal lamina propria.

We also examined the Lin$^-$ Sca1$^+$ CD117 (cKit)$^{\text{int}/-}$ CD135 (Flt3)$^-$ IL7Rα^+ cell, which was proposed to be an early NK precursor, termed prepro-NK [105]. Strikingly, we found uniform but low expression of GFP contrasting with a high frequency of YFP fate-mapping, suggesting that a large fraction of these cells had, in fact, originated from ILCPs and down-regulated PLZF (Fig. 2.S1). Because the majority of these cells also expressed CD25, ICOS, and T1/ST2, the combination of which is highly characteristic of ILC2s, we concluded that they mostly represented late developmental intermediates of ILC2s termed immature ILC2s (iILC2 or LSIG), rather than cNK precursors, consistent with their reported Id2high profile [52]. Indeed, cell transfers into Rag$^{-/-}$ Il2rg$^{-/-}$ hosts have demonstrated that this precursor population exclusively generates mature ILC2 in vivo [52].

2.4.2 Immature NKs

We next examined the so-called immature NK (iNK) cells, which are defined by expression of NKp46 and NK1.1, but not DX5. Although these cells were previously considered to be developmental intermediates of the cNK lineage before acquisition of DX5 [92], recent reports have emphasized that a fraction of iNKs expressed a phenotype similar to that

of mature ILC1s, including expression of IL7Rα and CD49a, but neither CD49b (DX5) nor eomesodermin [5, 39, 106]. In one study, transfer of eomesodermin-negative iNKs into $Rag^{-/-}$ $Il2rg^{-/-}$ hosts reconstituted liver ILC1s but not cNKs [39]. We found that nearly 80% of iNKs, defined by their CD3ε^- CD122$^+$ NKp46$^+$ NK1.1$^+$ DX5$^-$ profile, were YFP$^+$, but these cells did not express GFP, suggesting that a large proportion had originated from ILC1 rather than from cNK precursors (Fig. 2.2A). We further showed that the PLZF fate-mapped iNKs (isolated from YFP-negative chimeras) gave rise exclusively to CD49a$^+$ DX5$^-$ (ILC1s) but not to CD49a$^-$ DX5$^+$ (cNKs) liver cells after transfer into $Rag^{-/-}$ $Il2rg^{-/-}$ recipients (Fig. 2.2B). In contrast, CD45-congenic CLPs (coinjected with the YFP$^+$ iNKs) generated both cNKs and ILC1s in the liver (Fig. 2.2B). We also purified bone marrow CD3ε^- NKp46$^+$ NK1.1$^+$ DX5$^-$ iNKs, irrespective of their PLZF fate-mapping, and transferred them into $Rag^{-/-}$ $Il2rg^{-/-}$ hosts, to test whether they contained cNK precursors. Although these cells generated a majority of ILC1s, consistent with the predominance of ILC1 precursors, we also observed a substantial fraction of cNKs (Fig. 2.2B). Thus, CD3ε^- NKp46$^+$ NK1.1$^+$ DX5$^-$ bone marrow cells, previously termed iNKs, are a mixture of both ILC1 and cNK lineage cells, with a strong predominance of ILC1s. Upon stimulation with ionomycin and PMA, the bone marrow CD3ε^- NKp46$^+$ NK1.1$^+$ DX5$^-$ cells produced significantly less IFN-γ than their liver counterpart, further supporting the notion that they represented immature ILC1s (iILC1s) (Fig. 2.2C).

Altogether, our analysis of the developmental stages of ILC1s and cNKs suggests a new model whereby these lineages originate from distinct precursors, distinguished by the expression of PLZF and $\alpha_4\beta_7$, before undergoing a parallel sequence of progressive steps encompassing previously described populations and diverging between iNKs and iILC1s as proposed in Fig. 2.S2.

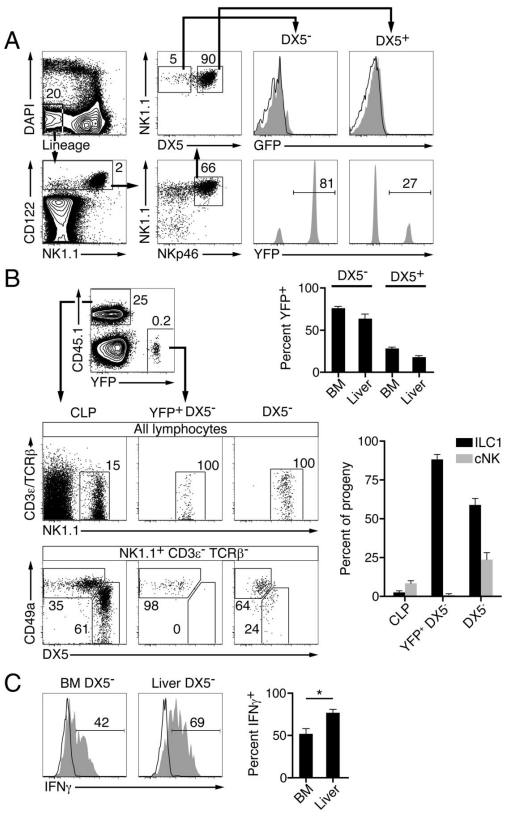


Figure 2.2: A majority of bone marrow NK1.1$^+$ DX5$^-$ immature NKs are ILC1 rather than cNK precursors.

(A) Gating strategy for DX5$^-$ (NK1.1$^+$ DX5$^-$) and DX5$^+$ (NK1.1$^+$ DX5$^+$) populations among Lin$^-$ CD122$^+$ NKp46$^+$ bone marrow cells (left four dot plots). Histograms show expression of GFP (reporting PLZF; Top) by the DX5$^-$ and DX5$^+$ populations from PLZF$^{GFPcre+/+}$ (filled gray) and WT (open) mice and YFP expression (reporting PLZF fate-mapping; Bottom). Bar graphs show summary of results (mean ± SEM) for bone marrow DX5$^-$ and DX5$^+$ cells (as gated) and for liver DX5$^-$ (CD3ε^- TCRβ^- NK1.1$^+$ DX5$^-$) and DX5$^+$ (CD3ε^- TCRβ^- NK1.1$^+$ DX5$^+$) populations. Data representative of 35 mice analyzed in three independent experiments. (B) CD45.2 Rag$^{-/-}$ Il2rg$^{-/-}$ mice were injected with a mixture of 750 CD45.1 CLPs + 1,500 YFP$^+$ Lin$^-$ CD122$^+$ NKp46$^+$ NK1.1$^+$ DX5$^-$ bone marrow cells from YFP$^-$ chimeras or with 1,500 CD45.1 Lin$^-$ CD122$^+$ NKp46$^+$ NK1.1$^+$ DX5$^-$ bone marrow cells alone, as indicated. The progeny of these populations within the liver was analyzed 34 wk later by FACS. Bar graphs show summary of results (mean ± SEM). Representative of 35 chimeras analyzed in two independent experiments. (C) The Lin$^-$ CD122$^+$ NKp46$^+$ NK1.1$^+$ DX5$^-$ bone marrow (BM DX5$^-$) and CD3ε^- TCRβ^- NK1.1$^+$ DX5$^-$ liver (liver DX5$^-$) populations were sorted from WT mice and stimulated with PMA (20 ng/mL) and ionomycin (5 μg/mL) for 4 h in the presence of 1x GolgiStop, and IFN-γ production was analyzed by intracellular FACS staining (filled gray). Control histograms (open) are from unstimulated cells. Results from three independent experiments summarized in the bar graph (mean ± SEM). $^*P < 0.05$.

2.4.3 Predominance of ILC1s Early in Ontogeny

Tissue-resident ILC1s are reminiscent of the early waves of tissue-resident $\gamma\delta$T cells that populate tissues such as skin, uterus, intestine, liver, and lung in the newborn (23). These tissue-resident $\gamma\delta$T cells express semiinvariant TCRs and are thought to represent a first line of defense in the perinatal period, before the generation of their blood recirculating counterparts, which have a more diverse repertoire. Indeed, we found that a great majority of CD3ε^- NK1.1$^+$ cells in both embryonic day (E)16 fetal liver and in 1-wk-old liver were YFP$^+$ fate-mapped, expressed CD49a and TRAIL, and lacked eomesodermin, whereas CD49a$^-$ TRAIL$^-$ eomesodermin$^+$ cNK cells appeared later in ontogeny between weeks 1 and 3, and were characterized by lower YFP fate-mapping (Fig. 2.3). Note that to study fetal and newborn hematopoiesis, these experiments used the straight PLZF-IRES-GFPCre crossed to ROSA-fl-STOP-fl-YFP mice, rather than the YFP-negative adult bone marrow chimeras. This experimental necessity explains in part the higher frequency of YFP labeling in cNKs, due to background labeling before the hematopoietic stem cell stage. Surprisingly, liver ILC1s tended to coexpress DX5 in fetal life, but not after birth, emphasizing that DX5 expression may not always be a reliable marker of cNKs. Altogether, these results therefore demonstrate that ILC1s strongly predominate over cNK cells in perinatal life, whereas cNK cells achieve greater frequency in adults.

Impact of PLZF on ILC1 Development and Maturation. We previously showed that PLZF-deficient ILC1s were decreased fourfold compared with their wild-type counterparts in the liver of competitive bone marrow chimeras [29]. By microarray analysis, we found that ILC1s isolated from the liver of PLZF-deficient mice exhibited a limited set of changes compared with those of wild-type littermates, because only 229 genes had altered expression by more than 1.4-fold (Fig. 2.4A). In contrast, and similar to a recent study [95], >2,000 genes exhibited differential expression between liver cNK cells and ILC1s, indicating that, as expected, the differences between cNKs and ILC1s do not solely depend on PLZF. Strikingly, however, 59% of the PLZF-dependent genes in ILC1s were also differentially expressed

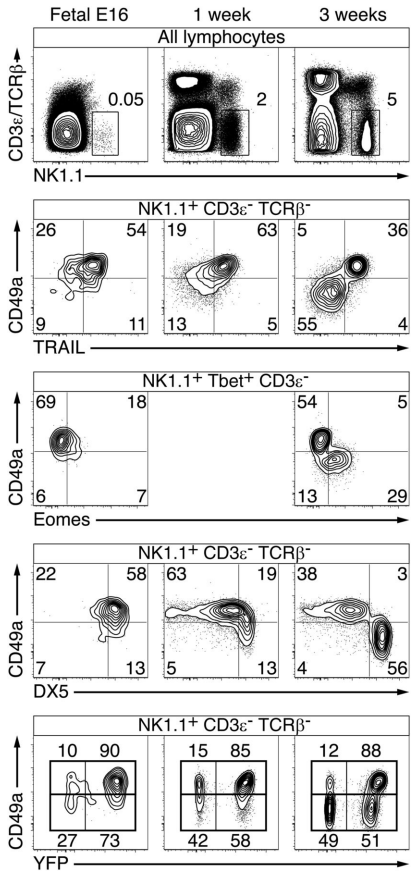


Figure 2.3: Predominance of ILC1s over cNKs in the fetal and newborn liver. FACS analysis of hepatic lymphocytes isolated from WT or PLZF$^{GFPcre+/-}$ ROSA26-YFP mice at E16 or at 1 and 3 wk after birth, as indicated. At Bottom, dot plots are divided into CD49a^{+} and CD49a^{-} fractions and the numbers indicate the percentages of YFP^{+} and YFP^{-} cells within each of these gates. All dot plots, except on the third row, are obtained from the same samples. Representative of 45 mice for each time point, analyzed in two independent experiments.

between ILC1s and cNKs, a highly significant overlap ($P = 2.7 \times 10^{93}$ by hypergeometric test). Because ILC1s and cNKs differ by their history of PLZF expression, these differentially expressed genes may be involved in the divergence between these two lineages. For example, PLZF-deficient ILC1s showed decreased expression of CD127 (IL7Rα), both at the mRNA level by microarray analysis and at the protein level by flow cytometry (Fig. 2.4 B and C). IL7Rα is expressed in the hematopoietic precursors to both ILC1s and cNKs, but its expression is characteristically maintained only in mature ILC1s [106]. Although the role of IL7Rα in ILC1 biology is not fully understood, because mature ILC1s seem mostly dependent on IL15 for their maintenance [5], it may be important for optimal survival during a developmental window, before expression of the IL15 receptor. Thus, the conspicuous defect in IL7Rα expression in PLZF-deficient ILC1s might contribute in part to their decreased frequency. Other ILC1 genes that depended on PLZF and were not expressed by cNKs included *Mmp9*, *Klrb1b*, *Klrk1*, and *Tnfrsf25*, as well as *Cd3g* and *Cd3e*, which were previously reported to be expressed by fetal but not adult NK cells [107]. This expression of Cd3 genes may reflect the predominance of ILC1s over cNKs in the fetus. Most of the genes that were differentially expressed between wild-type ILC1s and PLZF-deficient ILC1s as well as between wild-type ILC1s and cNKs cells showed concordant changes in the two comparisons (Fig. 2.4B, upper right and lower left quadrants), although there were examples of opposite variations (Fig. 2.4B, upper left and lower right quadrants), for example for *Il18r1*. Importantly, several of the key genes that were differentially expressed between cNKs and ILC1s, including *Eomes*, Itga1 encoding *CD49a*, *Tnfsf10* encoding TRAIL, *Sell* encoding CD62L, *Edg8* encoding S1P5, and *Cxcr6*, did not seem to be influenced by PLZF.

2.5 DISCUSSION

This study uses PLZF reporting and fate-mapping to identify developmental intermediates in the ILC1 lineage pathway and to define their relationship with previously proposed stages of cNK development. Whereas the history of expression of PLZF and $\alpha_4\beta_7$ clearly distin-

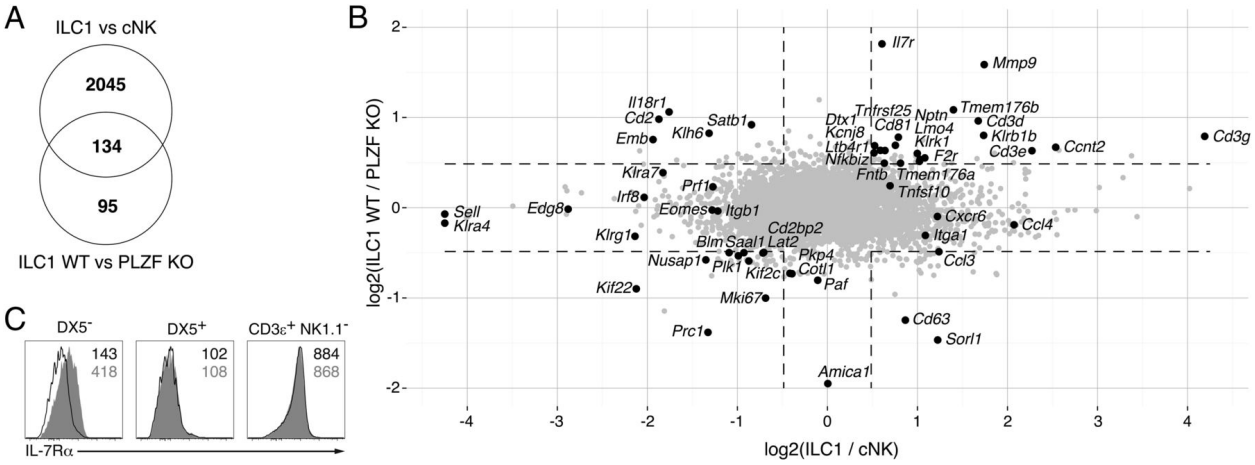


Figure 2.4: Most PLZF-dependent genes in ILC1s are also differentially expressed between ILC1s and cNKs.

Shown is a comparison of changes in gene expression due to PLZF deficiency in ILC1s (ILC1 WT/PLZF KO) and changes in gene expression between ILC1s versus cNKs (ILC1/cNK). (A) Venn diagram. The P value for 134/229 PLZF-dependent genes to be part of the 2,179 genes differentially expressed between ILC1s and cNKs is 2.7×10^{93} by hypergeometric test. (B) Scatter plot. Dotted lines delineate the 1.4-fold change threshold. Select immune genes are highlighted. (C) FACS analysis of IL7Rα expression by liver ILC1s (DX5$^-$) and cNKs (DX5$^+$) from PLZF$^{-/-}$ (open) and WT littermates (filled gray), with T cells as positive controls (CD3ε^+ NK1.1$^-$). Mean fluorescence intensity (MFI) is indicated in black for PLZF$^{-/-}$ and gray for WT. Representative of 4 WT and 4 PLZF$^{-/-}$ littermates examined in two separate experiments.

guished ILC1 from cNK precursors, the subsequent developmental intermediates appeared to be intermixed within the rNKP and iNK stages, which were previously assigned solely to cNKs. The earlier, pre-NKP stage appeared to be an even more complex mixture of PLZF-expressing ILCPs, which can give rise to ILC1s, ILC2s, and ILC3s, and of PLZF-negative cNK precursors. In contrast, the so-called prepro-NK cells, previously shown to generate NK1.1^{+} cells in single-cell culture with OP9 and IL-15, include a majority of immature ILC2s, suggesting that these cells might exhibit a degree of plasticity in forced cytokine environments, as reported recently for more mature ILC lineages [22, 108].

The earliest committed precursor to the cNK lineage has not been physically identified yet. Although the PLZF^{-} $\alpha_4\beta_7^{-}$ fraction of pre-NKPs is likely associated with the cNK lineage, recent studies using transfer experiments into Rag$^{-/-}$ Il2rg$^{-/-}$ hosts suggested that an earlier cNK precursor might be found at a post-CLP stage defined by a Lin^{-} IL7Rα^{+} $\alpha_4\beta_7^{\text{high}}$ Id2$^{\text{low}}$ PLZF$^{\text{low/neg}}$ profile [39, 59]. These cells might then rapidly down-regulate $\alpha_4\beta_7$ to become pre-NKPs in the cNK pathway. Thus, whereas ILC1s and cNKs both require NFIL3 and Tox at an early phase of their development, the divergence between these lineages is associated with the differential regulation of $\alpha_4\beta_7$ and PLZF, and possibly other factors such as Id2 and Gata3, which are also expressed at very high levels in ILCPs [29]. The divergence between these lineages and the parallel stages of their development are depicted in Fig. 2.S2.

Our study demonstrates that PLZF accompanies but is not absolutely required for this divergence. Indeed, PLZF has a detectable impact, as demonstrated by the observation that most of the PLZF-dependent genes in ILC1s, including the characteristic Il7r, are also differentially expressed between ILC1s and cNKs. However, this impact is relatively modest compared, for example, with NKT thymocyte development, where PLZF is absolutely essential for the acquisition of the T-helper polarized effector programs, which closely mirrors the trifurcation of ILCPs into ILC1, ILC2, and ILC3 lineages. This surprising difference may reflect the presence of redundant gene(s) in ILCPs or, alternatively, may suggest that

PLZF-deficient ILC1s are expanded from a fraction of precursors that overcome defective survival signals, for example through the IL7 receptor, but are otherwise relatively normal. Future studies of the redefined early stages of cNK and ILC development should help unravel the transcriptional events controlling these innate lymphoid lineages and elucidate the molecular basis of their divergence.

2.6 METHODS

Mice. C57BL/6J (stock no. 000664), B6.SJL-*Ptprc*^{a} *Pepc*^{b}/BoyJ (CD45.1; stock no. 002014), and B6.129 1-Gt(ROSA)26Sor$^{tm1(EYFP)Cos}$/J (stock no. 006148) mice were obtained from The Jackson Laboratory, whereas B6.Rag2^{tm1Fwa} II2rg^{tm1Wjl} (Rag$^{-/-}$ Il2rg$^{-/-}$, stock no. 4111) mice were obtained from Taconic. The PLZF^{GFPcre} strain has been described (15). PLZF$^{-/-}$ mice were a gift from P. P. Pandolfi, Beth Israel Deaconess Medical Center, Boston, and were backcrossed to C57BL/6J for at least nine generations. Unless noted otherwise, animals were 410 wk of age when analyzed and were compared with littermate controls obtained from crosses between heterozygous breeders. For fetal experiments, the morning a vaginal plug was observed was considered E0. Mice were housed in a specific pathogen-free environment at the University of Chicago, and experiments were performed in accordance with the guidelines of the Institutional Animal Care and Use Committee.

Preparation of Cell Suspensions. Spleen and liver were mechanically dissociated through 70-μm filters, and bone marrow was isolated by gently crushing femurs and tibias before filtration. Following dissociation, each liver was centrifuged at 400 g for 5 min, resuspended in 5 mL of 40% (vol/vol) Percoll (Sigma-Aldrich), and then centrifuged at 800 g for 10 min. The supernatant was aspirated and the cell pellet was resuspended in HBSS (Gibco) containing 0.25% BSA (Sigma-Aldrich) and 0.65 mg/L sodium azide (Sigma-Aldrich). Flow Cytometry. Cell suspensions were incubated with purified anti-CD16/32 (clone 93) for 10 min on ice to block Fc receptors. Fluorochrome- or biotin-labeled monoclonal antibodies (clones denoted in parentheses) against $\alpha_4\beta_7$ (DATK32), B220 (RA3-6B2), CD3ε (145-2C11), CD4 (RM4-5

or GK1.5), CD8α (53-6.7), CD11b (M1/70), CD11c (N418), CD19 (6D5), CD25 (PC61), CD27 (LG.7F9), CD45.1 (A20), CD45.2 (104), CD49a (HMα1), CD122 (5H4 or TM-β1), CD160 (7H1), CD244 (2B4), cKit (2B8), DX5 (DX5), Eomes (Dan11mag) Flt3 (A2F10), Gr-1 (RB6-8C5), ICOS (C398.4A), IFNγ (XMG1.2), IL-7Rα (A7R34), Ly-6D (49-H4), NK1.1 (PK136), NKp46 (29A1.4), Sca-1 (D7), T1/ST2 (D1H9), TCRβ (H57-597), Ter-119 (TER-119), Thy1.2 (53-2.1) and TRAIL (N2B2) were purchased from BD Biosciences, BioLegend, eBioscience, or R& D Systems. To exclude dead cells, 4',6-diamidino-2-phenylindole (DAPI; Molecular Probes) was added to all live samples. For intracellular staining, cells were fixed with 4% paraformaldehyde and permeabilized by using the Foxp3 Transcription Factor Staining Buffer Set (eBioscience). Cells were run on an LSRII (BD Biosciences) or sorted by using a FACS Aria II (BD Biosciences) and analyzed by using FlowJo software (Tree Star). Collected events were gated on DAPI lymphocytes and doublets were excluded. Unless specified otherwise, CLPs were gated Lin$^-$ Ly6D$^-$ CD244$^+$ CD27$^+$ IL-7Rα^+ Flt3$^+$ CD122$^-$ (19), pre-NKP were gated Lin$^-$ Ly6D$^-$ CD244$^+$ CD27$^+$ IL-7Rα^+ Flt3$^-$ CD122$^{\text{low/neg}}$ (19), rNKP were gated Lin$^-$ Ly6D$^-$ CD244$^+$ CD27$^+$ IL-7Rα^+ Flt3$^-$ CD122$^{\text{high}}$ (19), prepro-NK were gated Lin$^-$ Sca-1$^+$ cKit$^{\text{low}}$ IL-7Rα^+ Flt3$^-$ (20), bone marrow DX5$^-$ cells were gated Lin$^-$ CD122$^+$ NK1.1$^+$ NKp46$^+$ DX5$^-$ (14), bone marrow DX5$^+$ cells were gated Lin$^-$ CD122$^+$ NK1.1$^+$ NKp46$^+$ DX5$^+$, liver DX5$^+$ cells were gated CD3ε^- TCRβ^- NK1.1$^+$ DX5$^+$ CD49a$^-$, and liver DX5$^-$ cells were gated CD3ε^- TCRβ^- NK1.1$^+$ DX5$^-$ CD49a$^+$. Lineage mixtures for pre-NKPs, rNKPs, and CLPs included antibodies against CD3ε, CD11b, CD19, and NK1.1; for prepro-NK: CD3ε, CD8α, CD11b, CD19, Gr-1, and Ter119; for bone marrow NK1.1$^+$ subsets: CD3ε, CD4, CD8α, CD19, and Ter119.

Bone Marrow Chimeras. To generate YFP$^-$ chimeras, 510×10^3 LSK cells were sorted from the bone marrow of PLZF$^{\text{GFPcre+/–}}$ ROSA-YFP mice and injected retroorbitally into lethally irradiated (1,000 rads) CD45.1 recipients. Chimeras were analyzed 57 wk after reconstitution, gating on CD45.2$^+$ cells to exclude residual host cells.

Isolation and Adoptive Transfer of Bone Marrow CLPs and NK1.1$^+$ Subsets. For the iso-

lation of CLPs from CD45.1 bone marrow, lineage$^+$ cells were first depleted by using an autoMACS (Miltenyi Biotec) after staining with biotin-conjugated antibodies against B220, CD3ε, CD4, CD8α, CD11b, CD11c, CD19, Gr-1, NK1.1, TCRβ, and Ter119, followed by incubation with SAv microbeads (Miltenyi Biotec). CLPs were then sorted as DAPI$^-$ Lin$^-$ IL-7Rα^+ cKit$^{\text{int}}$ Sca-1$^{\text{int}}$ Flt3$^{\text{high}}$ to greater than 95% purity by using a FACS Aria II (BD Biosciences). For the isolation of YFP$^+$ NK1.1$^+$ DX5$^-$ cells, bone marrow from YFP$^-$ chimeras (described above) was stained with APC-conjugated anti-NK1.1 antibody, bound to anti-APC microbeads (Miltenyi Biotec), subjected to double-column enrichment, and sorted as DAPI$^-$ Lin$^-$ CD122$^+$ NK1.1$^+$ NKp46$^+$ DX5$^-$ YFP$^+$ to greater than 95% purity. For the isolation of NK1.1$^+$ DX5$^-$ cells, CD45.1 bone marrow was stained with APC-conjugated anti-NK1.1 antibody, bound to anti-APC microbeads, subjected to double-column enrichment, and sorted as DAPI$^-$ Lin$^-$ CD122$^+$ NK1.1$^+$ NKp46$^+$ DX5$^-$ to greater than 95% purity. For the NK1.1$^+$ subsets, the lineage mixture antibodies used were as follows: CD3ε, CD4, CD8α, CD19, and Ter119. Following isolation, 1,500 YFP$^+$ NK1.1$^+$ DX5$^-$ cells mixed with 750 CD45.1-congenic CLPs, or 1,500 CD45.1 NK1.1$^+$ DX5$^-$ cells alone were injected retroorbitally into sublethally irradiated (400 rads) 6- to 10-wk-old CD45.2 Rag$^{-/-}$ Il2rg$^{-/-}$ mice. Recipient mice were analyzed 34 wk after transfer.

Isolation and in Vitro Stimulation of NK1.1$^+$ DX5$^-$ Cells. Following enrichment using FITC-NK1.1 and anti-FITC microbeads (Miltenyi Biotec), NK1.1$^+$ DX5$^-$ cells were sorted as DAPI$^-$ Lin$^-$ NK1.1$^+$ NKp46$^+$ DX5$^-$ from the bone marrow or DAPI$^-$ CD3ε^- TCRβ^- NK1.1$^+$ NKp46$^+$ DX5$^-$ from the liver of WT mice (>95% purity). For the bone marrow cells, the lineage mixture antibodies used were the following: CD3ε, CD4, CD8α, CD19, TCRβ, and Ter119. Up to 10,000 sorted cells were resuspended in 500 muL of RPMI medium 1640 (Cellgro) containing 10% FCS (Biowest) and 1 GolgiStop (BD). Samples were stimulated with phorbol 12-myristate 13-acetate (PMA; 20 ng/mL) and ionomycin (5 μg/mL) for 4 h at 37 C. Following the stimulation, samples were fixed and permeabilized by using the Cytofix/Cytoperm kit (BD).

Microarray Analysis. Liver lymphocytes from pools of $PLZF^{+/+}$ or $PLZF^{-/-}$ mice were sorted into ILC1s ($CD3\varepsilon^{-}$ $NK1.1^{+}$ $CD49a^{+}$ $DX5^{-}$) and cNKs ($CD3\varepsilon^{-}$ $NK1.1^{+}$ $CD49a^{-}$ $DX5^{+}$) and stored in TRIzol (Life Technologies). RNA was isolated by using the RNeasy mini kit (Qiagen), processed, and annealed to the mouse WG-6 array (Illumina). Illumina MouseWG-6 v2.0 Expression BeadChip data were quality-checked by using the lumi package [109] and quality control metrics in the R/Bioconductor statistical software package (v3.0.0). Data were background corrected and normalized by variance-stabilizing transformation [110] followed by quantile normalization. Data were log2 transformed and fit to linear models by using the limma package [76] for fold change comparison between groups. A fold-change cutoff of 1.4 was used to distinguish differentially expressed genes. Microarray data has been deposited with the National Center for Biotechnology Information Gene Expression Omnibus repository under the accession number GSE65898.

Statistical Analysis. Two-tailed Students t test was performed by using Prism (GraphPad Software) to determine whether data differed from the expected value. $^{*}P < 0.05$.

Acknowledgments

This work was supported by NIH Grants R01 AI108643, AI038339, and HL118092 (to A.B.) and University of Chicago Digestive Diseases Research Core Center Grant P30 DK42086.

2.7 SUPPORTING INFORMATION

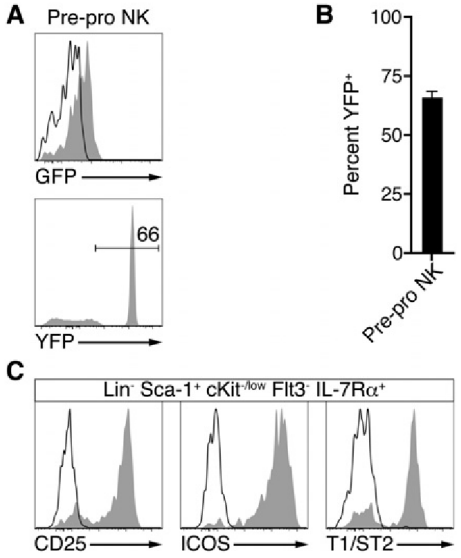


Figure 2.S1: Pre-pro-NKs contain a majority of immature ILC2s.
(A, Top) Expression of GFP (reporting PLZF) by the pre-pro-NK population (Lin^- Sca-1^+ $cKit^{-/low}$ $Flt3^-$ IL-$7R\alpha^+$) from $PLZF^{GFPcre+/+}$ (filled gray) and WT (open) mice. (Bottom) YFP expression (reporting fate-mapping by PLZF) in YFP-negative radiation chimeras. (B) Summary of results (mean ± SEM) shown in A. (C) FACS analysis of pre-pro-NK cells, stained for indicated markers (filled gray) or with isotype controls (open). Data representative of at least three mice analyzed in two or more independent experiments.

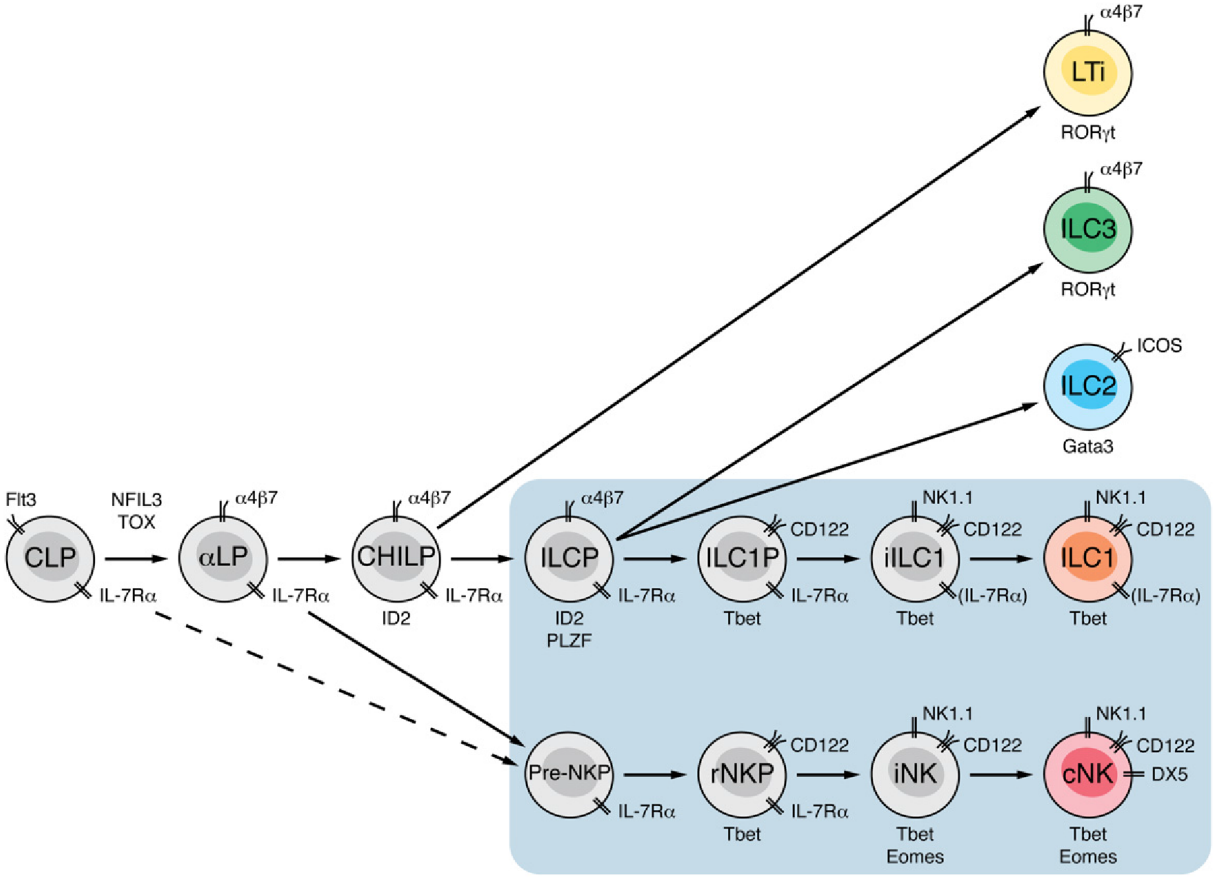


Figure 2.S2: Divergence and parallels in the development of ILC1s and cNKs. The diagram highlights, within the shaded rectangle, the parallel progression sequence of ILC1 and cNK progenitors, with ILCP, ILC1P, and iILC1 previously included in, and now distinguished from, pre-NKP, rNKP, and iNK, respectively. The abbreviated names of developmental intermediates include CLP (common lymphoid precursor), αLP ($\alpha_4\beta_7{}^+$ lymphoid precursor), CHILP (common helper innate lymphocyte precursor), ILCP (innate lymphoid cell precursor), ILC1P (ILC1 precursor), iILC1 (immature ILC1), pre-NKP (pre-NK precursor), rNKP (refined NKP), iNK (immature NK), and mNK (mature NK).

CHAPTER 3

SINGLE–CELL ANALYSIS DEFINES THE DIVERGENCE BETWEEN THE INNATE LYMPHOID CELL LINEAGE AND LYMPHOID TISSUE–INDUCER CELL LINEAGE

Isabel E Ishizuka[1,2,6], Sylvestre Chea[3,6], Herman Gudjonson[1,4-6], Michael G Constantinides[1,2], Aaron R Dinner[4,5], Albert Bendelac[1,2,7] and Rachel Golub[3,7]

[1] Committee on Immunology, University of Chicago, Chicago, Illinois, USA. [2] Department of Pathology, University of Chicago, Chicago, Illinois, USA. [3] Institut Pasteur, Immunology Department, Lymphopoiesis Unit, Inserm U668, University Paris Diderot, Paris, France. [4] Institute of Biophysical Dynamics, University of Chicago, Chicago, Illinois, USA. [5] Department of Chemistry, University of Chicago, Chicago, Illinois, USA. [6] These authors contributed equally to this work. [7] These authors jointly directed this work.
Correspondence should be addressed to A.B. (abendela@bsd.uchicago.edu).

Author contributions: I.E.I., H.G., M.G.C. and A.B. designed the experiments; I.E.I. performed single-cell sorting and culture experiments; S.C. designed and performed the lymphoid Biomark assay; H.G. performed computational analysis of the Biomark experiments; M.G.C. designed and performed experiments; A.R.D. supervised computational analysis; A.B. supervised experiments; R.G. supervised Biomark experiments; and I.E.I., H.G. and A.B. wrote the manuscript with contributions from all authors.

Published in Nature Immunology January 18, 2016; doi:10.1038/ni.3344

3.1 ABSTRACT

The precise lineage relationship between innate lymphoid cells (ILCs) and lymphoid tissue-inducer (LTi) cells is poorly understood. Using single-cell multiplex transcriptional analysis of 100 lymphoid genes and single-cell cultures of fetal liver precursor cells, we identified the common proximal precursor to these lineages and found that its bifurcation was marked by differential induction of the transcription factors PLZF and TCF1. Acquisition of individual effector programs specific to the ILC subsets ILC1, ILC2 and ILC3 was initiated later, at the common ILC precursor stage, by transient expression of mixed ILC1, ILC2 and ILC3 transcriptional patterns, whereas, in contrast, the development of LTi cells did not go through multilineage priming. Our findings provide insight into the divergent mechanisms of the differentiation of the ILC lineage and LTi cell lineage and establish a high-resolution 'blueprint' of their development.

3.2 INTRODUCTION

Innate lymphocytes lack B cell or T cell antigen receptors and exert effector functions at mucosal barriers [106, 111]. These populations segregate into three general groups on the basis of their expression of the transcription factors T-bet, GATA-3 and RORγt. However, there is considerable heterogeneity among T-bet-expressing group 1 lymphocytes, which comprise conventional (or classical) natural killer (NK) cells (cNK cells), group 1 innate lymphoid cells (ILC1 cells) and tissue-resident NK cells, and among RORγt-expressing group 3 lymphocytes, which comprise CCR6$^+$ lymphoid tissue-inducer (LTi) cells and CCR6$^-$ ILC3 cells. In addition, some plasticity has been reported among CCR6$^-$ ILC3 cells, which can upregulate T-bet expression and acquire group 1 properties [22], and among some populations of ILC2 cells, which can acquire group 3 properties [108].

Lineage-tracing and cell-transfer studies have suggested that ILC1, ILC2 and ILC3 cells, but not LTi cells or cNK cells, are derived from a common dedicated precursor, the ILC

precursor (ILCP), characterized by expression of the transcription factor PLZF [29]. Similar to the LTi precursor (LTiP), the ILCP originates from a lymphoid precursor that expresses the integrin $\alpha_4\beta_7$ and is itself derived from the common lymphoid precursor (CLP). The $Id2^{hi}$ fraction of $\alpha_4\beta_7{}^+$ lymphoid precursors, called 'common helper innate lymphoid precursors' (CHILPs), is a heterogeneous population that includes the PLZF-expressing ILCPs as well as precursors of LTi cells [39], but whether the CHILP population includes a common precursor of both ILCs and LTi cells or separate precursors of each lineage has not yet been determined. A study has suggested that cNK cells might originate from an earlier $Id2^{lo}$ $CXCR6^+$ fraction of $\alpha_4\beta_7$-expressing lymphoid precursors (αLPs) [59]. Thus, the developmental relationships between these lineages remain incompletely established.

Several transcription factors, including Nfil3, Tox, Id2, GATA-3, TCF-1 (encoded by *Tcf7*) and PLZF (encoded by *Zbtb16*) [29, 59, 57, 55, 51, 52, 112, 9, 44, 43, 32], are required for the development of all or several of these innate lineages, which suggests they have an effect at a common precursor stage. However, partial defects, rather than complete defects, have often been reported in mice lacking these transcription factors, which suggests substantial redundancy and complexity of this early transcriptional network. Other transcription factors have been found to selectively affect individual ILC lineages, such as the effect of RORα (encoded by *Rora*), Bcl-11b and Gfi1 in ILC2 cells [14, 15, 113], which suggests more distal effects in the ILC-differentiation pathway. Precise understanding of the general hierarchy of expression of these factors is missing, however, which limits the design and interpretation of mechanistic studies aiming at delineating their interplay.

Here we used cultures of single cells purified from the fetal livers of a mouse reporter strain expressing green fluorescent protein (GFP) and Cre recombinase from the gene encoding PLZF (*Zbtb16*-GFP-Cre) to precisely define the differentiation stages between CLP and ILC and to identify the stage of divergence between ILCs and LTi cells. We also performed multiplex quantitative single-cell transcriptional analysis of 100 lymphoid genes to characterize the complexity of molecular events associated with this differentiation path-

way. We derived a high-resolution map and inferred a precise ordering of the induction of transcription factor expression and identified previously unknown stages of development before and after PLZF expression and the stage of bifurcation between the ILC lineage and LTi cell lineage. Notably, transcriptional priming for the different cytokine effector programs occurred at the ILCP stage itself through multilineage priming. In contrast, the LTiP proceeded to directly acquire its type 3 program without undergoing mixed transcriptional priming. Together our findings further define the dichotomy between ILCs and LTi cells and provide new insight into the stages and mechanisms of their development and the interplay of transcription factors that direct their differentiation.

3.3 RESULTS

3.3.1 Bifurcation of αLPs into ILCPs and LTiPs

Lin$^-$ IL-7Rα^+ fetal liver cells (where lineage (Lin) is defined by a 'cocktail' of antibodies to CD3ε, TCRβ, CD19, CD11c, GR-1, Ter119 and NK1.1) included CLPs, identified by a Flt3$^+$ $\alpha_4\beta_7{}^-$ profile, and a cell population expressing $\alpha_4\beta_7$, which included the precursors to innate lymphoid lineage, thought to arise from the CLP (Fig. 3.1a). This $\alpha_4\beta_7$-expressing population was originally called 'αLP', but the identification of ILCPs and LTiPs among αLPs prompted us to limit the designation of 'αLP' to the $\alpha_4\beta_7$-expressing cells that were neither ILCP nor LTiP but include their precursors [111, 114]. We further subcategorized this αLP population as an Flt3$^+$ subset and Flt3$^-$ subset to identify early precursors and late precursors, respectively. We detected these populations by flow cytometry among fetal liver cells obtained from the *Zbtb16*-GFP-Cre reporter strain at embryonic day 15 (E15) (Fig. 3.1a). Among the Lin$^-$ IL-7Rα^+ $\alpha_4\beta_7{}^+$ population, the ILCPs were identified by GFP expression, and the LTiPs were identified by their GFP$^-$ CXCR5$^+$ profile [29, 62]. The LTiPs were clearly distinguishable from the other subsets in this staining, although a small fraction of these cells seemed to have low expression of GFP (Fig. 3.1a). As expected, only

the LTiPs coexpressed the chemokine CCR6 and the coreceptor CD4 (Fig. 3.1b).

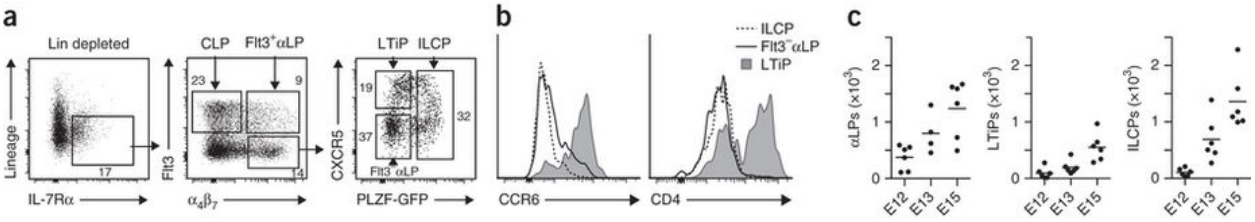


Figure 3.1: Identification of distinct subpopulations of $\alpha_4\beta_7$-expressing lymphoid precursors in fetal liver.
(a) Flow cytometry of fetal liver cells from E15 Z *Zbtb16*-GFP-Cre reporter mice after depletion of Lin^+ cells (expressing CD3ε, CD11c, CD19, NK1.1, TCRβ, Ter119 and GR-1) by magnetic beadbased cell separation and staining for IL-7Rα (left), Flt3 and $\alpha_4\beta_7$ (middle) and CXCR5 (right), for the identification of distinct subpopulations representing CLPs, αLPs, LTiPs and ILCPs (middle and right) after gating on Lin^- IL-7Rα^+ cells (left). Numbers adjacent to outlined areas indicate percent cells in each subpopulation. (b) Flow cytometry of ILCPs, LTiPs and $Flt3^-$ αLPs stained for CCR6 or CD4. (c) Quantification of αLPs, LTiPs and ILCPs in fetal liver at E12, E13 and E15. Each symbol represents an individual fetus; small horizontal lines indicate the mean. Data are representative of twelve independent experiments (a) or three experiments (b) or are pooled from two independent experiments (c).

Both LTiPs and ILCPs could already be detected at a low frequency among Lin^- IL-7Rα^+ $\alpha_4\beta_7{}^+$ fetal liver cells at E12, the earliest time point of our analysis, and their absolute numbers increased five- to tenfold by E15 (Fig. 3.1c). To establish the lineage relationships among αLPs, ILCPs and LTiPs, we performed single-cell cultures of αLP subsets ($Flt3^+$ and $Flt3^-$) on OP9 stromal cells in the presence of interleukin 7 (IL-7) and stem-cell factor, as described5. We categorized the growing cells as ILC1 colonies, ILC2 colonies or ILC3 colonies by high expression of the activating NK cell receptor NK1.1, the costimulatory receptor ICOS or $\alpha_4\beta_7$, respectively, as reported5. We distinguished LTi cells from ILC3 colonies by expression of CD4, which was expressed by nearly half the LTi cells but not by ILC3 cells (Fig. 3.1b), although this method underestimated the real frequency of LTi cell colonies by approximately twofold. We were unable to distinguish cNK cells from ILC1 cells by expression of eomesodermin (Eomes), because expression of this transcription factor was induced in ILC1 cells in our culture conditions (data not shown). However, because fetal

progenitor cells do not produce cNK cells [115], we counted all NK1.1$^+$ colonies as ILC1 cells.

ILCPs gave rise nearly exclusively to single and mixed colonies of ILC1 cells, ILC2 cells and ILC3 cells (Fig. 3.2), consistent with this cell type's being a common precursor to these lineages, as reported [29]. In contrast, Flt3$^+$ and Flt3$^-$ αLPs also generated a sizeable proportion of LTi cells, which indicated that this group of cells included precursors to the LTi cell lineage. Notably, many wells containing LTi cells also included ILC1 cells or ILC2 cells (the presence of ILC3 cells could not be ascertained in wells containing LTi cells), which indicated that a single Flt3$^+$ or Flt3$^-$ αLP precursor could generate both LTi cells and ILC lineages. Thus, while the reported CHILP population [39] included a heterogeneous mixture of ILCPs and other precursors, our observations identified the Flt3$^-$ αLP as the common proximal precursor, at the single-cell level, of both ILCPs and LTiPs.

Notably, when we performed the cultures with OP9 stromal cells lacking the DL1 ligand of Notch receptors, we found that many fewer ILC2 colonies were derived from Flt3$^+$ αLPs or Flt3$^-$ αLPs than from ILCPs (Fig. 3.2a, bottom). However, this defect was largely corrected in OP9-DL1 cultures (Fig. 3.2b, bottom). This finding was consistent with the proposal that a Notch signal is essential for ILC2 differentiation [44, 14] and further established that the signal must be delivered early at the αLP stage, rather than late at the ILCP stage. Together these experiments demonstrated that the fetal liver αLP pathway was dedicated to the formation of the ILC and LTi cell lineages. They also identified the late Flt3$^-$ αLP as the common proximal precursor to ILCPs and LTiPs.

3.3.2 Single-cell multiplex quantitative RT-PCR of ILC precursors

We used multiplex quantitative RT-PCR for transcriptional analysis of single cells along the fetal pathway linking αLPs to ILCPs and LTiPs, with a panel of 100 probes for genes encoding lymphoid development factors, including transcription factors, cytokine receptors and chemokine receptors. Data from 339 single cells, including 157 αLPs, 168 ILCPs and

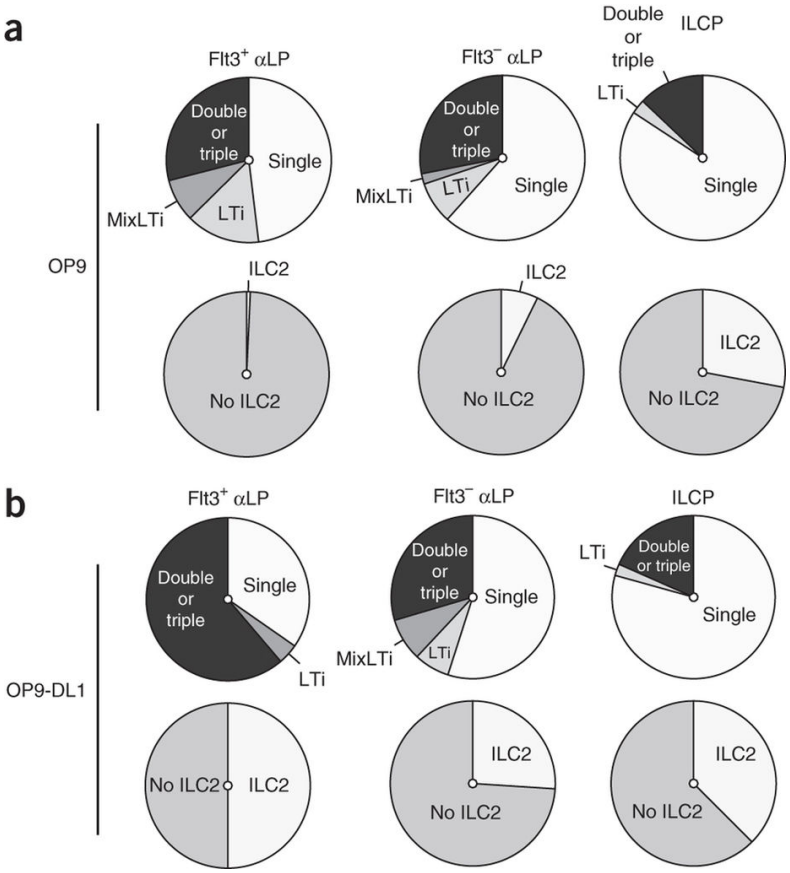


Figure 3.2: Colonies derived from single-cell cultures of $\alpha_4\beta_7$-expressing lymphoid precursors in fetal liver.
Frequency of single ILC1, ILC2 or ILC3 cells (Single) or two or more ILC subsets (Double or triple), LTi cells, and mixed colonies of LTi ($CD4^+$) cells plus either ILC1 cells or ILC2 cells (the presence of ILC3 cells could not be ascertained in these wells) (MixLTi) (top), and frequency of wells containing ILC2 colonies (whether single or mixed with other ILC lineages) (bottom), among single-cell cultures of various precursors sorted from *Zbtb16*-GFP-Cre fetal livers at E14E16 and cultured with OP9 cells (a) or OP9-DL1 cells (b). Average cloning efficiency, 31-64%; <3% of colonies could not be characterized by staining and are not presented here. In OP9-DL1 cultures, colonies of pro-T cells were also observed in 47% of $Flt3^+$ αLP wells, 19.7% of $Flt3^-$ αLP wells and 14.5% of ILCP wells; these were identified by a $CD4^+$ or $CD8^+$, $ICOS^-$ $\alpha_4\beta_7^-$ $NK1.1^-$ phenotype (data not shown). Data are pooled from at least three independent experiments (OP9) or one to three independent experiments (OP9-DL1) with 51-447 total colonies for each progenitor population.

14 LTiPs, were compiled from two independent single-cell sorting experiments with pooled E15 fetal livers. After filtering results by the expression of 'housekeeping' (control) genes, we considered the transcriptional profiles from 299 single cells for downstream analysis.

Unsupervised hierarchical clustering analysis of these transcriptional profiles defined groups of cells that seemed to be in distinct developmental stages (Fig. 3.3). Consistent with the proposal that acquisition of PLZF expression signified a distinctive developmental transition for ILCs, we found strong, visually evident separation of the transcriptional profiles of the αLP populations (Fig. 3.3, left) and ILCP populations (Fig. 3.3, right). We used contiguous branches of the hierarchical clustering dendrogram to define three clusters of predominantly αLPs (AI-AIII), one cluster composed of mixed αLPs and ILCPs (B), and four clusters of predominantly ILCPs (CI-CIV), on the basis our understanding of ILC development. All clusters were distinct from one another and were composed of a substantial number (>20) of similar cells (Supplementary Fig. 3.S1). One additional cluster of about 20 cells, all with conspicuously high expression of *Il1rl1* (which encodes the IL-33 receptor chain IL-33Rα), was removed from the study because it was unrelated to the other clusters and instead seemed to represent contaminating mast cell precursors with low expression of $\alpha_4\beta_7$and PLZF (Supplementary Fig. 3.S2). Thus, this analysis identified further heterogeneity among precursor cells and generated a 'blueprint' of their temporal sequence during ILC development.

3.3.3 Early developmental transitions

To facilitate analysis of the clusters identified above, we generated a condensed heat map of gene expression in all 299 single cells, limited to a set of 20 key informative genes. Consistent with αLPs' being early precursors to ILCPs and LTiPs, there was sparse expression of genes encoding transcription factors and cytokines specific for these lineages in the A clusters. For example, *Zbtb16*, *Tcf7*, *Gata3*, *Bcl11b*, *Cxcr5*, *Rorc*, *Rora* and *Tbx21* were not expressed by clusters AI-AIII (Fig. 3.4). In contrast, clusters AI-AIII expressed genes encoding transcrip-

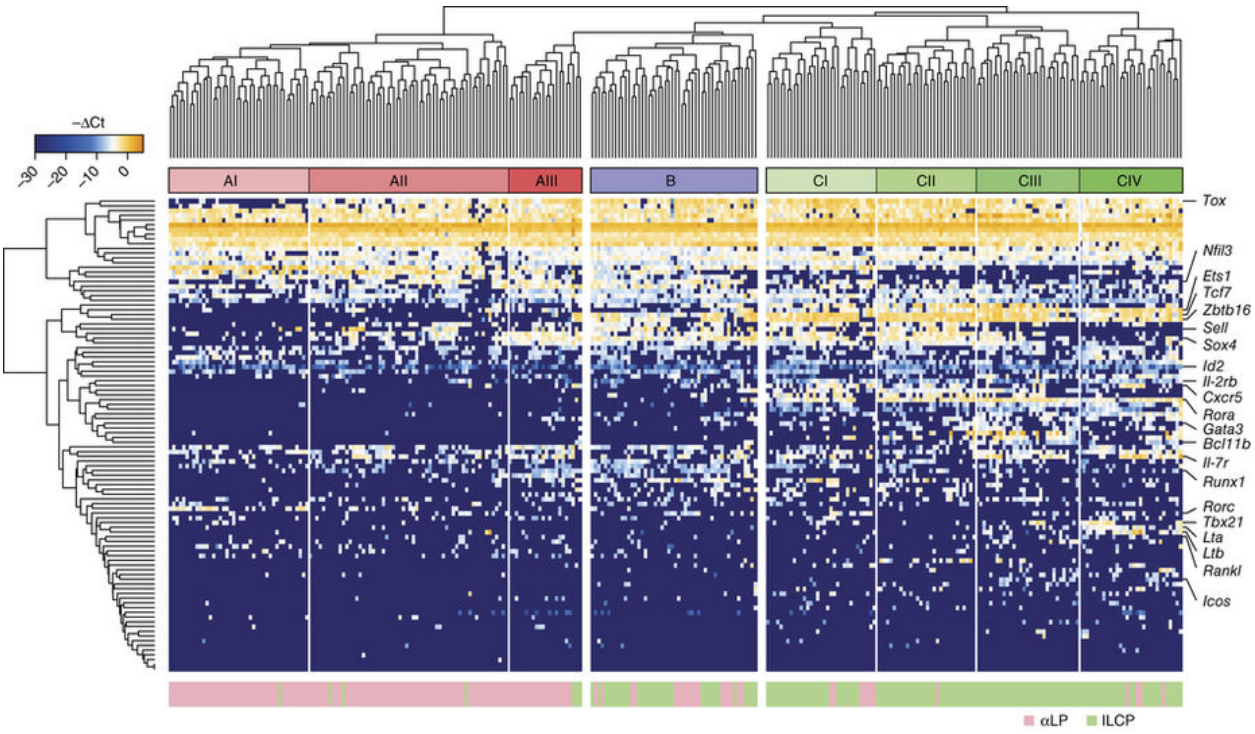


Figure 3.3: Hierarchical clustering distinguishes αLP and ILCP transcriptional profiles. Hierarchical clustering dendrogram (top and left margin) derived from single-cell multiplex quantitative PCR analysis of 299 single αLPs and ILCPs, distinguishing clusters AI-AIII (αLPs; left), cluster B (mixed αLPs and ILCPs; middle), and clusters CI-CIV (ILCPs; right); each column represents a single cell (bottom, sorted cell type; key, bottom right), and each row represents one gene (of 100 total; right margin, select genes encoding products with known or anticipated roles in ILC development). ΔCt (top left key), difference in threshold cycle relative to the average value for the 'housekeeping' (control) gene. Data are from two independent experiments with 157 αLPs and 168 ILCPs from pooled livers.

tion factors linked to early ILC and LTi development, including *Id2*, *Tox*, *Nfil3*, *Sox4*, *Ets1* and *Runx1* (Fig. 4a). That conclusion was confirmed by plotting of the average mRNA expression per cell (Fig. 3.4b) and the frequency of cells expressing genes encoding these transcription factors in each cluster (Supplementary Fig. 3.S3). Notably, a distinct temporal pattern of expression could be inferred from these plots. Thus, cells in cluster AI had low expression of *Id2*, but most lacked expression of *Tox* and *Nfil3* (Fig. 3.4). Nearly all cells in clusters AII and AIII, however, expressed *Tox* and *Nfil3* (Fig. 3.4). Cells in clusters AI and AII lacked expression of *Ets1*, *Sox4* and *Runx1*, whereas a majority of cells in cluster AIII expressed these genes (Fig. 3.4). Finally, *Tcf7* and *Zbtb16* were not expressed by clusters

AI-AIII (αLPs) but were widely expressed in cluster B and the C clusters (ILCPs) (Fig. 3.4). We measured low Id2 expression across clusters AI-AIII and found a tendency toward more frequent and higher expression in cluster B and clusters CI-CIV (Fig. 3.4), in line with the suggestion that increased *Id2* expression is correlated with commitment to innate lineages [39]. Thus, our analysis suggested that the temporal patterns of expression of these transcription factors were precisely regulated. Furthermore, unlike *Tox*, whose expression remained high, *Nfil3* was ultimately downregulated in ILCP clusters (Fig. 3.4b), consistent with its temporally limited requirement, as suggested by experiments of late gene ablation [57].

3.3.4 Bifurcation of ILC and LTi cell branches

Cluster B comprised the most mixed representation of αLPs and ILCPs and seemed to be a developmental transition state linking early developmental events and specification to the ILC lineage versus the LTi cell lineage. Notably, cells in cluster B expressed *Tcf7* and *Zbtb16* but largely lacked expression of genes encoding ILC lineagespecific factors (Fig. 3.4). In fact, the expression of *Tcf7* and *Zbtb16* was lower in cluster B than that in ILCP clusters (Fig. 3.4), which confirmed the conclusion that cluster B corresponded to a transitional stage. A similar pattern was found for *Id2* (Fig. 3.4), suggestive of developmental continuity between αLP clusters and ILCP clusters, with cluster B probably including the first cells to acquire *Zbtb16* expression before specification to the ILC lineage. Although the induction of *Tcf7* and *Zbtb16* seemed to be nearly synchronous, a fraction of cluster B cells expressed *Tcf7* without *Zbtb16* (Fig. 3.4 and Supplementary Fig. 3.S3), which suggested that this fraction might have included cells destined to the LTi cell lineage. In fact, detailed analysis of cells in cluster B that expressed *Tcf7* but not *Zbtb16* showed that most of these cells had a tendency to express genes encoding factors associated with the LTi cell lineage, including *Rorc*, while conspicuously lacking expression of *Rora*, which is expressed in more mature LTiPs (Fig. 3.5). In contrast, cells in cluster B that expressed both *Tcf7* and *Zbtb16* rarely

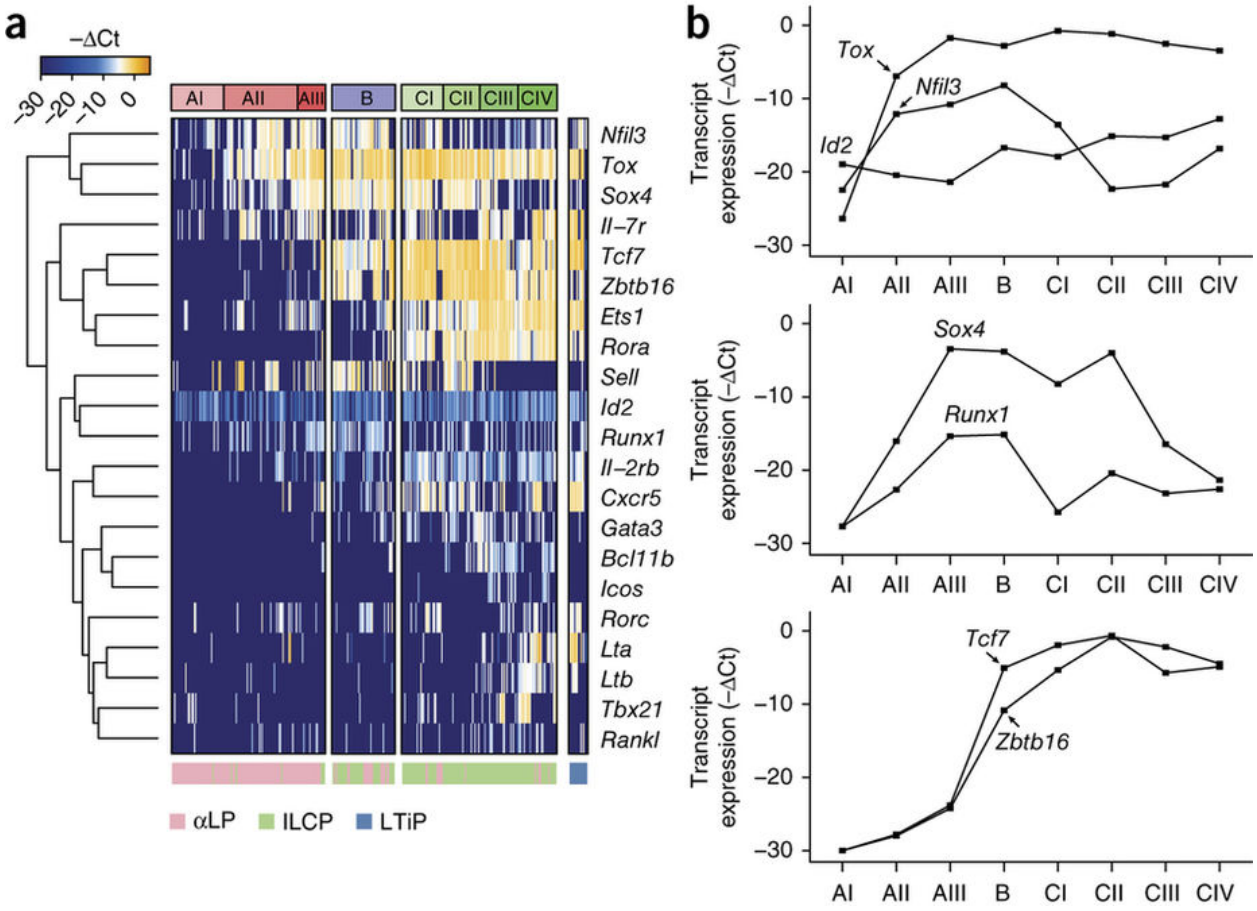


Figure 3.4: Clusters define the developmental progression of key transcription factors. (a) Single-cell multiplex quantitative PCR (as in Fig. 3) of transcripts from genes encoding products with known or anticipated roles in ILC development, ordered by hierarchical clustering of αLPs and ILCPs as in (Fig. 3.3), and of LTiPs (far right); below, sorted cell type (key, bottom left). (b) Transcript expression (average ΔCt values) for genes encoding products that define early developmental transitions in each cluster.

expressed *Rorc* at that stage and instead displayed some expression of genes encoding factors associated with ILC lineages, such as *Gata3*, while they lacked *Rora* expression (Fig. 3.5); this suggested that they were less mature than most ILCPs. Thus, we concluded that cluster B represented the stage of bifurcation of the αLP into the LTi cell and ILC lineages.

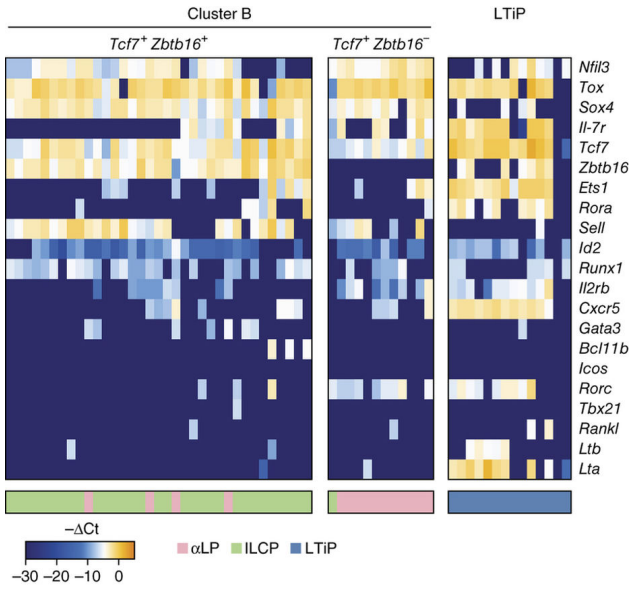


Figure 3.5: Transitional cluster B includes two distinct subsets, on the basis of expression of *Tcf7* and *Zbtb16*.
Transcript expression by cells from cluster B that express both *Tcf7* and *Zbtb16* (*Tcf7*$^+$ *Zbtb16*$^+$) or express *Tcf7* but not *Zbtb16* (*Tcf7*$^+$ *Zbtb16*$^-$) (Fig. 3.3), and of LTiPs (right).

3.3.5 ILC lineage differentiation originates in the ILCP

Cells in the ILCP clusters (CI-CIV) showed signs of ILC maturation and expressed genes encoding ILC lineage-defining transcription factors and cytokines (Fig. 3.4). These cells had almost universally high expression of *Zbtb16*, *Tcf7* and *Id2*, and additional transcripts further distinguished them from the developmentally intermediate cluster B. Almost all cells in ILCP clusters expressed *Ets1* and *Rora* and tended to have low *Nfil3* expression. The ILCP clusters delineated by hierarchical clustering probably reflected sequential stages of ILC maturation, rather than association with particular ILC lineages. Cells in clusters CI and CII had high expression of *Sell* and *Sox4* that diminished in clusters CIII and CIV. Conversely, cells in clusters CIII and CIV had higher expression of Ets1 and Il7r than that of cells in clusters CI and CII. Characteristically, cells in clusters CIII and CIV more frequently expressed genes encoding ILC lineage-defining transcription factors, including *Tbx21*, *Bcl11b* and *Rorc*, than did cells in clusters CI and CII. Our finding of the expression of genes encoding ILC lineage-

defining transcription factors in multiple ILCP clusters indicated that we were probably capturing developmental stages of ILC differentiation. Furthermore, cells that expressed genes encoding ILC lineage-defining transcription factors in later clusters (clusters CIII and CIV) tended to express genes encoding other lineage-associated cytokines and transcription factors and thus seemed to be more differentiated.

3.3.6 ILCPs undergo multilineage transcriptional priming

To better assess the range of cells among the ILCP clusters that seemed to be differentiating toward ILC lineages, we compiled a subset of cells that expressed genes encoding ILC lineage-associated factors (Fig. 3.6). Specifically, we identified cells with ILC1, ILC2 or ILC3 markers by their expression of *Tbx21* (ILC1) or *Bcl11b* (ILC2) or at least two genes among *Cxcr5*, *Rorc*, *Lta* and *Ltb* (ILC3). In this population, we observed many cells coexpressing genes encoding markers of different ILC lineages, such as *Tbx21* and *Bcl11b*, or *Tbx21* and *Cxcr5*, or *Tbx21* and *Rorc*, and even cells expressing multiple markers of the three different lineages. In total, among 120 total cells in ILCP clusters, we found 60 cells that expressed genes encoding lineage-differentiation markers, of which nearly one third (19) coexpressed genes encoding markers of different lineages.

The multilineage priming noted above was substantiated by the expression of genes encoding additional lineage-specific factors (Fig. 3.6). In particular, the simultaneous expression of *Bcl11b* and *Icos* was associated with expression of canonical ILC2 genes, including *Il1rl1* and *Il13*. We also observed increased expression of *Gata3* in *Bcl11b*-expressing cells, although *Gata3* was expressed throughout ILCPs without strict lineage association (as shown before by flow cytometry [29]). Across cells in ILCP clusters, we observed a relatively strong correlation between the expression of *Rxrg* and *Igf2* and that of *Bcl11b*. The expression of *Tbx21* was associated with higher expression of *Il2rb* and, in some cases, with expression of *Eomes* and *Ncr1* (which encodes the activating receptor NKp46). We also found that *Irf8* expression seemed to be associated with *Tbx21* expression in differentiating ILCPs. Finally,

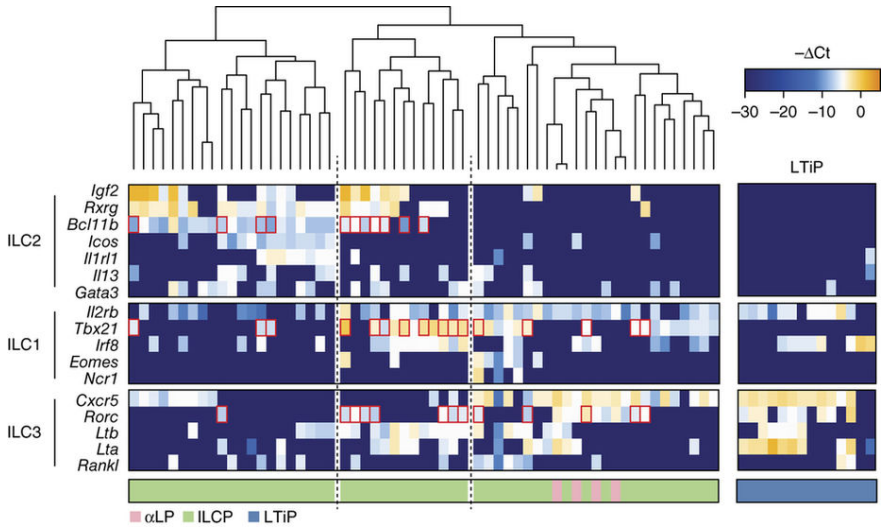


Figure 3.6: Multilineage transcriptional priming in ILCPs.
Expression of ILC lineage-specific transcripts (left margin) in differentiating ILCPs and in LTiPs (far right); red outlines indicate single cells expressing multiple genes encoding lineage-specific transcription factors (two or more genes among *Bcl11b*, *Tbx21* or *Rorc*). Below, sorted cell type (key, bottom left).

we found that expression of *Rorc* was associated with high expression of *Cxcr5*, as well as with expression of *Lta*, *Ltb* and *Rankl*. Generally, expression of *Il2rb* and *Cxcr5* was observed throughout the ILCP clusters, with higher expression and greater frequency of expression associated with the expression of genes encoding other ILC1 and ILC3 factors. There was substantial overlap between the expression of genes encoding ILC1 and ILC3 factors, in particular for the expression of *Tbx21* and *Cxcr5*, *Lta*, *Ltb* and *Rankl*.

The results noted above were in contrast to the expression profile of LTiPs in the same fetal livers (Figs. 3.5 and 3.6). These cells showed no evidence of multilineage priming. In particular, they conspicuously lacked expression of *Gata3*, *Bcl11b* or *Tbx21*, which further supported the proposal that they follow a distinct lineage-differentiation pathway.

Collectively, the expression of genes encoding multiple lineage factors in ILCP clusters was suggestive of multilineage potential. This further confirmed the proposal that cells navigating ILC lineage 'decisions' were in the ILCP population. Moreover, it suggested that during the initial lineage 'decisions', the relationships among distinct ILC lineage factors

are more complicated than might have been anticipated from studies of helper T cells, for example.

3.3.7 Transcriptional programs correlate with lineage 'decisions'

To functionally evaluate the relevance of the transcriptional programs noted above in ILCPs, we assessed their lineage potential in single-cell cultures. We sorted fetal ILCP subsets on the basis of their expression of the cytokine receptor chain CD122 (IL-2Rβ), ICOS and the chemokine receptor CXCR5 (Fig. 3.7a) and studied their progeny in single-cell cultures. We found that these subsets showed some bias toward the corresponding ILC1 or ILC2 program (Fig. 3.7b, top). Thus, the earliest precursors with a bias toward the ILC1 lineage or ILC2 lineage could be discerned at the ILCP stage by their CD122$^+$ CXCR5$^-$ profile or ICOS^{hi} profile, respectively (Fig. 3.7a). However, the biases were incomplete, indicative of the residual multipotency of these populations. For example, while ICOS^{hi} ILCPs generated mostly ILC2 colonies, they also gave rise to some single ILC1 and ILC3 colonies, as well as to dual and triple colonies (Fig. 3.7b). In contrast, CD122$^+$ CXCR5$^-$ ILCPs generated mainly ILC1 colonies but also generated some single ILC3 colonies as well as dual colonies (Fig. 3.7b). Even the ILCPs that produced dual colonies maintained a corresponding bias as, for example, all dual colonies originating from ICOS^{hi} precursors included an ILC2 colony, and those originating from a CD122$^+$ CXCR5$^-$ precursor included an ILC1 colony (Fig. 3.7b, bottom). Together with the multilineage priming observed at the transcriptional level, these observations indicated that lineage polarization was proceeding by stepwise restriction of alternative programs from multipotential precursors, which ultimately led to the canonical polarized lineages. Using intracellular staining of transcription factors, we further confirmed that a substantial fraction of fetal ILCPs coexpressed T-bet, RORγt and GATA-3 or ICOS, indicative of multilineage priming, whereas mature ILCs were strictly limited to expression of their lineage-specific transcription factors (Fig. 3.7c). This scenario is different from current models of the polarization of helper T cells, wherein acquisition of cytokine effector

programs does not typically involve intermediates with mixed lineage patterns.

3.4 DISCUSSION

Using the *Zbtb16*-GFP-Cre reporter system, we have identified the stage at which the αLP bifurcates into ILCPs or LTiPs during early fetal lymphopoiesis. We also used single-cell multiplex quantitative PCR analysis to produce a high-resolution map of the development of innate lymphocyte lineages. We emphasize that the proposed developmental progression discussed below is inferred on the basis of the continuity of gene expression between clusters but remains to be experimentally demonstrated.

We found that *Zbtb16* and *Tcf7* were simultaneously upregulated in cluster B at the bifurcation of the ILC and LTi cell lineages, with *Tcf7* expression marking both lineages and *Zbtb16* expression identifying the ILC lineage. A fraction of the precursor cells that expressed *Tcf7* but did not express *Zbtb16* expressed *Rorc* and seemed closer to LTiPs than to ILCPs. Notably, cells in cluster B lacked expression of *Rora*, which is characteristically induced later in development, in further support of the conclusion that cells in cluster B navigate the bifurcation between the LTi cell lineage and ILC lineage.

The A clusters before the induction of *Zbtb16* and *Tcf7* must therefore represent earlier precursor cells, on the basis of the expression of *Id2*, *Tox*, *Nfil3* and *Runx1* (refs. [59, 57, 55, 51, 32, 116, 117, 33]), which has been linked to both lineages. *Id2* seemed to be the earliest expressed transcription factorencoding gene with substantial expression in cluster AI, a stage at which *Nfil3* and *Tox* transcripts were barely detected. Expression of *Nfil3* and *Tox* rapidly ascended during the transition to cluster AII and reached maximal levels in cluster AIII. These findings stand in apparent contrast to a published report showing that Nfil3 can bind to and induce Id2 and that ablation of Nfil3 can be complemented by Id2 (ref. [57]) but are consistent with the presence of *Id2* transcripts reported in arrested precursor cells lacking expression of *Nfil3* or *Tox* [57, 51]. Nfil3 might therefore exert a positive feedback loop rather than being a primary trigger for *Id2* expression. The expression of Nfil3 and

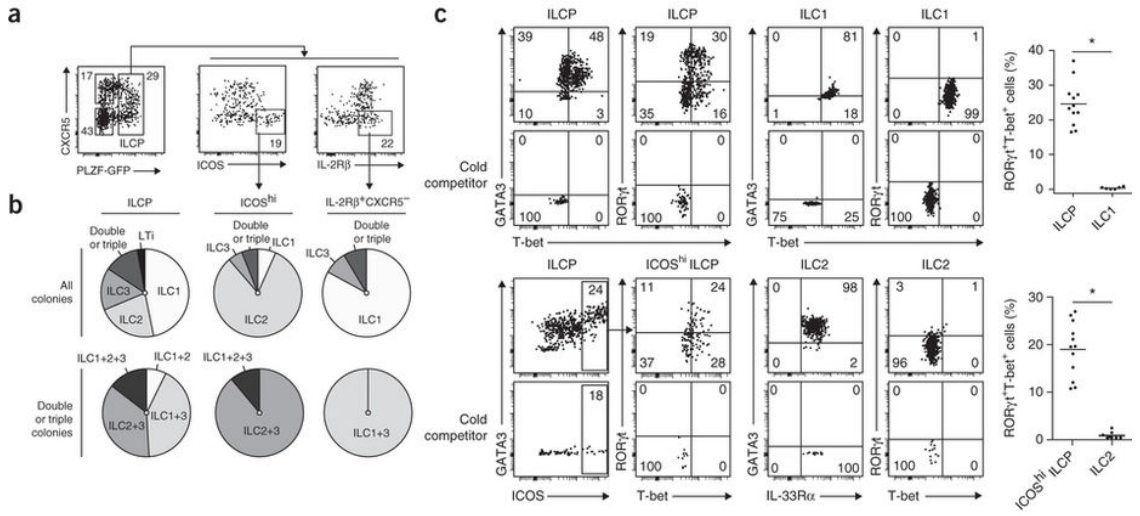


Figure 3.7: ILCP subsets with biased progeny in single-cell cultures.
(a) Flow cytometry of fetal liver ILCPs, identified by GFP (PLZF) expression (as in Fig. 3.1) and sorted on the basis of expression of CXCR5, ICOS and IL-2Rβ. Numbers adjacent to outlined areas indicate percent cells in each subpopulation. (b) Distribution of single colonies and double or triple colonies (top) and the composition of double or triple colonies (bottom) from each progenitor cell. ILCP, n = 425 colonies; IL-2Rβ^+ CXCR5$^-$ ILCP, n = 82 colonies; ICOS$^{\text{hi}}$ ILCP, n = 150 colonies. (c) Intracellular staining of various transcription factors in fetal ILCPs at E14 (n = 12; gated as Lin$^-$ IL-7Rα^+ $\alpha_4\beta_7{}^+$ PLZF$^+$) and in ILC1 cells from mature adult liver (n = 6; gated as CD3ε^- NK1.1$^+$ CD49a$^+$) or ILC2 cells from adult bone marrow (n = 7; gated as Lin$^-$ IL-7Rα^+ Thy1$^+$ Sca1$^+$), in the presence of unlabeled antibody (cold competitor) added before the conjugated antibody used for staining (bottom row of each group) or in the absence of such unlabeled antibody (top row of each group). Numbers adjacent to outlined areas or in quadrants indicate percent cells in each. Each symbol (far right) represents an individual mouse; small horizontal lines indicate the mean. *$P < 0.0001$ (two-tailed Student's t-test). Data representative of at least three experiments (a), are pooled from one to six independent experiments (b) or are representative of two independent experiments (c).

that of Tox also seemed to be induced nearly simultaneously. Thus, a report that Nfil3 can induce *Tox* and that ablation of *Nfil3* can be complemented by Tox7 might also reflect a positive feedback mechanism.

Additional transcription factorencoding genes that were induced early after *Id2*, *Tox* and *Nfil3* but before *Zbtb16* and *Tcf7* included *Sox4*, *Runx1* and *Ets1*. Although the function of the transcription factors encoded is currently unknown, Sox4 has been associated with the development of fetal $\gamma\delta$T cells in conjunction with TCF-1 (ref. [45]), whereas Runx1 has been shown to be important for the development of NK cells and LTi cells [116], and Ets1 has been associated with NK cell development [118]. Thus, these factors are probably integral components of the early transcriptional network of innate lymphocytes.

Clusters CI-CIV defined several stages of ILCPs, with cluster CI associated with the induction of *Rora*, which was maintained in all ILC lineages, although it seems to be required only for ILC2 cells [14, 13]. Cluster CII was associated with *Gata3* expression, which was maintained at a high level during the remaining ILCP stages and ultimately is downregulated in mature ILC1 and ILC3 cells [29]. Therefore, transient but high *Gata3* expression is an intrinsic developmental event that might distinguish ILC3 cells from LTi cells. Clusters CIII and CIV were marked by emergence of the expression of genes encoding lineage-specific factors. Notably, there was a distinct and extensive pattern of coexpression of genes encoding factors from the three different lineages in nearly a third of these cells, which indicated that a phase of multilineage transcriptional priming preceded the polarization of mature lineages. In contrast, the LTiPs did not go through this singular process and directly expressed Rorc and other genes encoding attributes of the LTi cell lineage without coexpression of genes encoding alternative ILC lineage markers. The extent of multilineage priming revealed by our study goes beyond the coexpression of low levels of T-bet and RORγt by GATA3$^+$ ILC precursors detected in the fetal intestine by flow cytometry [119]. Indeed, we found that multilineage priming included an extended list of genes encoding canonical markers of all three lineages. ILCPs sorted on the basis of the expression of genes encoding some of these lineage

differentiation markers showed a relative bias in differentiation toward the corresponding lineages after single-cell culture, although these subsets also maintained some multipotency, as shown by the generation of mixed colonies. These findings suggested that the terminal differentiation of ILCPs into polarized ILC lineages is more complex than previously envisaged. Instead of directly acquiring one of three programs, precursors of ILCs first activated multiple effector programs simultaneously and then progressively turned off priming of alternative lineages. In the absence of exogenous polarizing cytokines, such a developmental strategy can take advantage of the well-established antagonistic effects between these programs [120]. In ILCPs, external or internal inputs that might influence lineage 'decisions' include retinoids, which have been shown to favor type 3 innate lymphocytes [69, 68], or Notch signals, which are needed for ILC2 cells [44, 121]. Multilineage transcriptional priming has been proposed as a more general strategy for developmental 'decisions', such as the differentiation of various hematopoietic lineages from a common progenitor [122, 123, 124, 125]. It has not been widely reported for programs used in the differentiation of $CD4^+$ helper T cells into the TH1, TH2 or TH17 subset of helper T cells, which instead rely on polarizing cytokines released during infection or allergy [120], although mixed transcriptional patterns have been observed in cells cultured under mixed cytokine conditions [126].

Our studies further emphasize that although they share many functional properties, ILC3 cells and LTi cells have distinct precursors and different developmental histories. Only precursors of ILC3 cells transited through a stage of high expression of *Zbtb16* and *Gata3* with multilineage priming at the single-cell level. Our results should encourage studies aimed at further elucidation of the different functions of LTi cells and ILC3 cells, as illustrated, for example, by a report that ILC3 cells and LTi cells have different topographical locations in the lamina propria [127].

In summary, our study has allowed the identification of several previously unknown developmental transitions and transcription factors sequentially associated with the lineage progression of innate lymphocytes. In particular, our results characterized in cellular and

molecular detail the previously poorly defined bifurcation of lymphoid precursors into LTiPs and ILCPs. They have also demonstrated differential mechanisms of the acquisition of polarized effector programs by these lineages.

3.5 METHODS

Mice. Zbtb16-GFP-Cre reporter mice were generated in our laboratory as described and contain an internal ribosome entry site after the final exon of *Zbtb16*, followed by a fusion gene encoding enhanced GFPCre recombinase[29]. For fetal ontogeny experiments, the morning a vaginal plug was identified was counted as embryonic day 0 (E0). Mice were housed in a specific pathogenfree environment at the University of Chicago and experiments were performed in accordance with the guidelines of the Institutional Animal Care and Use Committee.

Preparation of cell suspensions. Fetal livers were mechanically dissociated through a 70-μm cell strainer and resuspended in HBSS (Gibco) containing 0.25% BSA (Sigma-Aldrich) and 5 mM sodium azide (Sigma-Aldrich).

Flow cytometry. Cell suspensions were pre-incubated with BD Fc Block for 10 min on ice. Fetal liver cells were pre-enriched for $\alpha_4\beta_7$-expressing cells unless otherwise indicated. For pre-enrichment of $\alpha_4\beta_7{}^+$ cells, fetal liver or bone marrow cells were stained with allophycocyanin (APC)-conjugated antibody and were subsequently bound to anti-APC microbeads (Miltenyi Biotec). The cells were then enriched using the autoMACS (Miltenyi Biotec) positive-selection double-sensitive program. Lineage (CD3ε, CD11c, CD19, NK1.1, TCRβ, Ter119 and GR-1) depletion of fetal liver or bone marrow cells was accomplished by incubation of cells with biotinylated antibodies to lineage markers antibodies (identified below) and then binding to streptavidin-conjugated microbeads (Miltenyi Biotec) and separation using the autoMACS depletion-sensitive program. Fluorochrome- or biotin-conjugated monoclonal antibodies to the following were used (clone number in parentheses): to mouse $\alpha_4\beta_7$(DATK32), CCR6 (29-2L17), CD11c (N418), CD19 (6D5), CD27 (LG.3A10), CD90.2 (Thy-1.2; 53-2.1), CD122 (5H4), CD127 (IL-7Rα; A7R34), CXCR5 (L138D7), FcϵRIα (MAR-1), GR-1 (RB6-

8C5), ICOS (C398.4A), IL-33R (ST2; DIH9), NKp46 (29A1.4), Sca-1 (D7), Ter119 (TER-119), T-bet (4B10), mouse IgG1 κ-chain (MG1-45), mouse IgG2a κ-chain (MG2a-53) and rat IgG2b κ-chain (RTK45-30) (all from BioLegend); CD3ε(145-2C11), CD4 (GK1.5), CD8α (53-6.7), CD45.1 (A20), CD45.2 (104), CD49a (Ha31/8), NK1.1 (PIK136), TCRβ (H57-597) and RORγt(Q31-378) (all from BD Biosciences); and GATA3 (TWAJ; eBioscience). The D-9 antibody to PLZF was conjugated to Alexa Fluor 647 using the labeling kit from Molecular Probes Life Technologies. For intracellular staining, cells were fixed and permeabilized using the Foxp3 Transcription Factor Staining Buffer Set (eBioscience). Cells were then blocked with unlabeled isotype-matched control antibody (identified above). As negative control, a 20-fold excess of unlabeled antibody to the transcription factor was added before the conjugated antibody ('cold' competition). Data were acquired on a LSRII (BD Biosciences) or cells were sorted using a FACSAria II (BD Biosciences) and data were analyzed using FlowJo software (TreeStar).

Single-cell cultures. Stocks of OP9 and OP9-DL1 stromal cells were a gift from J.C. Ziga-Pflcker. All experiments were performed in Opti-MEM with GlutaMAX (Gibco) containing 10% FCS (Gibco), 1% penicillin-streptomycin (Gibco), and 60 mM 2-mercaptoethanol (Sigma-Aldrich) and were maintained in a 37 C incubator (Thermo Scientific) with 5% CO2. Stromal cells were plated at 70% confluency. Before addition of lymphocytes, stromal cells were irradiated (1,500 rads) and culture media supplemented with murine IL-7 (25 ng/ml; BioLegend) and stem-cell factor (25 ng/ml; BioLegend). Fetal liver lymphocyte samples underwent enrichment for $\alpha_4\beta_7{}^+$ cells by MACS and were sorted as single cells onto 96-well plates containing stromal cells and cytokines as described above. Cultures were analyzed after 6 or 10 d of culture and only colonies with more than ten CD45.2$^+$ cells were considered.

Biomark. Cells were sorted in 96-well quantitative PCR plates in 10 l of a CellsDirect One-Step quantitative RT-PCR Kit (Life Technologies), containing mixtures of diluted primers (0.05x final concentration). Preamplified cDNA was obtained after reverse transcription (15 min at 40 C, 15 min at 50 C and 15 min at 60 C), and preamplification (22 cycles: 15 s at 95

C and 4 min at 60 C), and diluted 1: 5 in TE pH8 Buffer (Ambion). Sample mix was as follows: diluted cDNA (2.9 μl), Sample Loading Reagent (0.29 μl; Fluidigm), TaqmanUniversal PCR Master Mix (3.3 μl; Applied Biosystem) or Solaris quantitative PCR Low ROX Master Mix (3.3 μl; GE Dharmacon). The assay mix was as follows: Assay Loading Reagent (2.5 μl; Fluidigm), Taqman (2.5 μl; Applied Biosystem) or Solaris (2.5 μl; GE Dharmacon). A 48.48 or 96.96 dynamic array integrated fluidic circuit (IFC; Fluidigm) was primed with control line fluid, and the chip was loaded with assays (either Taqman or Solaris) and samples using an HX IFC controller (Fluidigm). The experiments were run on a Biomark (Fluidigm) for amplification and detection (2 min at 50 C, 10 min for Taqman reagents or 15 min for Solaris reagents at 95 C, 40 cycles: 15 s at 95 C and 60 s at 60 C).

Analysis of single-cell multiplex quantitative PCR data. Independent single-cell quantitative PCR experiments were performed for αLPs, ILCPs and LTiPs from two different pools of *Zbtb16*-GFP-Cre E15 fetal livers. Cells not expressing detectable levels of all three housekeeping genes (*Actb*, *Gapdh* and *Hprt*) were removed from further downstream analysis. To appropriately compare measurements of distinct cells, expression levels of each gene for a given cell were normalized with respect to the average housekeeping gene expression level for that cell. Specifically, the cycle threshold value for each gene (Ct_g) was converted to the measure $-\Delta Ct_g =< Ct_{HKG} > Ct_g$, the difference between the average housekeeping gene cycle threshold (Ct_{HKG}) and the gene cycle threshold. For quantification purposes, genes that were not detected were given a $-\Delta Ct$ value of –30, close to the minimum value detected. Hierarchical clustering was performed using the Euclidian distance metric with complete-linkage agglomeration.

Visual inspection of the intercellular distances used for hierarchical clustering corroborates the strong distinction between αLP and ILCP clusters and the similarities between adjacent clusters in the prescribed developmental order (Supplementary Fig. 3.S1). We performed permutation analysis to empirically evaluate the significance of our clustering assignments. Specifically, we calculated the sum of square Euclidean distances from the group mean for

each cluster and used the total sum of squares within groups (SSW) value to compare our clustering assignment to 10,000 randomly permuted clustering assignments. We found the SSW of our clustering assignment to be substantially lower than those of all the random permutations, implying $P < 10^4$. We further directly compared all pairs of clusters using the same approach and similarly found that all clusters were significantly distinguishable from one another with $P < 10^4$. Finally, we performed hierarchical clustering with each gene iteratively removed from our complete data set to evaluate the stability of our clustering method. We found that hierarchical intercellular relationships were widely preserved under these perturbations. For instance, in ∼85% of all cell pairs, the number of branches connecting these pairs of cells to their common branch point on the clustering dendrogram were changed by less than two on average. Hierarchical clustering and analysis of quantitative PCR data was performed with custom scripts using the base packages in R (v3.1.2) and heat map displays were generated using the NMF package (v0.20.6).

Statistical analysis. Two-tailed Student's t-test was performed using Prism (GraphPad Software).

Accession codes. GEO: microarray data, GSE76407.

Acknowledgements

We thank J.C. Ziga-Pflcker (University of Toronto) for stocks of OP9 and OP9-DL1 stromal cells. Supported by the US National Institutes of Health (R01 HL118092, AI038339 and AI108643) and by the Digestive Diseases Research Center of Excellence (P30DK42086).

3.6 SUPPORTING INFORMATION

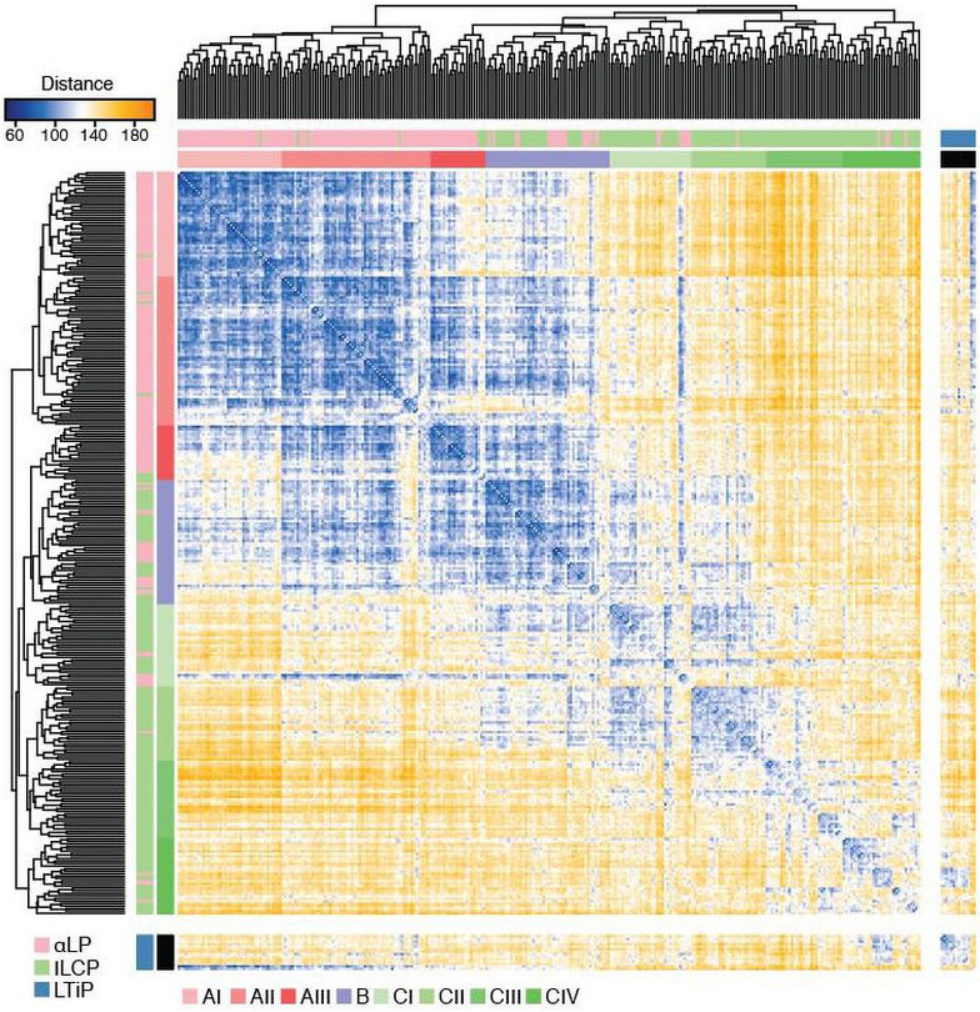


Figure 3.S1: Intercellular transcriptional distances confirm hierarchical clustering analysis. Euclidean distance between all pairs of measured transcriptional profiles. Dendrograms shown are those obtained from the hierarchical clustering of these intercellular distances. Outer bar, sorted cell type; αLP (red), ILCP (green), LTiP (blue). Inner bar, cluster assignment of each cell; AI–III (red), B (purple), CI–IV (green).

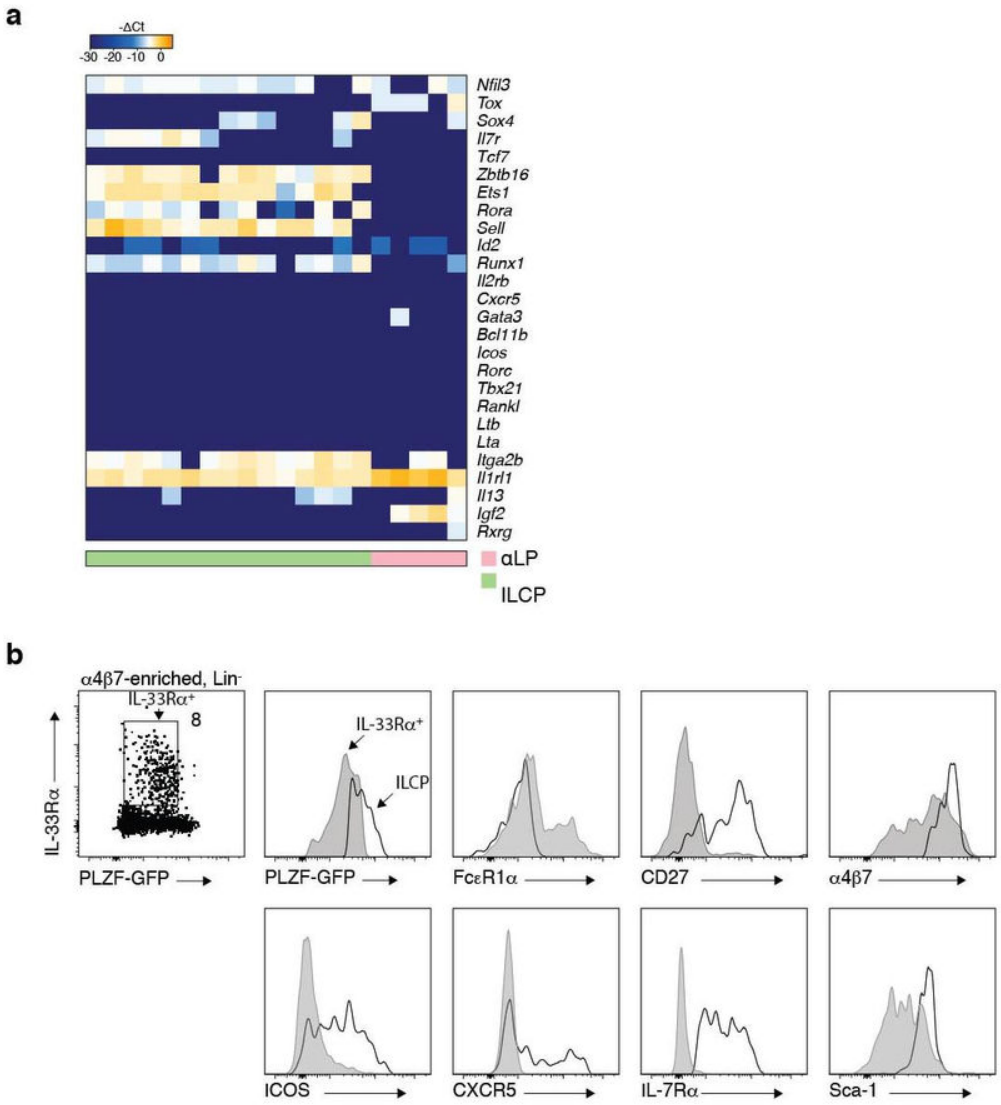


Figure 3.S2: $\alpha_4\beta_7^+$ IL-33Rα^{hi} cells represent contaminating mast cell precursors. a, Biomark analysis of the *Il33r*$^+$ cluster excluded from the study suggested that they were not innate lymphoid cells, as evidenced by the nearly uniform lack of *Tcf7* and *Tox*, as well as other key markers. b, FACS analysis suggested the mast cell nature of these cells. $\alpha_4\beta_7$-MACS-enriched fetal liver cells obtained from the *Zbtb16*-GFPCre reporter strain were gated as Lin$^-$ cells (this is the same population that was used to sort single αLP and ILCP) and stained for IL-33Rα. The IL-33Rα^+ cells expressed intermediate amounts of PLZF (GFP) and $\alpha_4\beta_7$, possibly explaining why they were found as contaminants in the Biomark analysis of purified αLP and ILCP. 16% of these cells expressed the mast cell marker FcεRIα, indicating that they were mast cell lineage precursors.

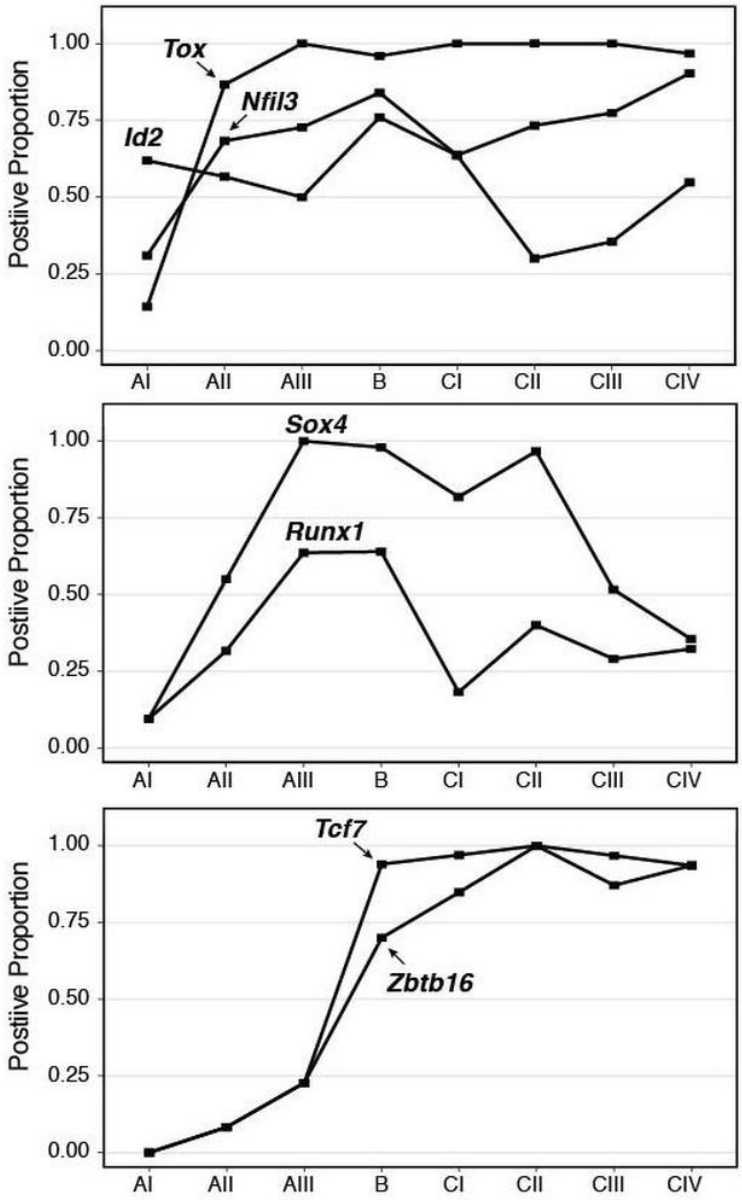


Figure 3.S3: Expression frequencies of key transcription factors in αLP and ILCP transcription-state clusters define their developmental progression.
In each cluster, the proportion of cells with detectable levels of transcript is calculated for genes defining early developmental transitions. Similar to the trends observed in average transcript levels by cluster, *Id2* is first expressed at a low level in AI, followed by *Tox* and *Nfil3* in AII, *Sox4* and *Runx1* in AIII, and finally by *Tcf7* and *Zbtb16* in B.

APPENDIX

3.A Continuous representation of single-cell expression profiles through dimensionality reduction

Following our hierarchical clustering analysis of single ILC precursor expression profiles, we aimed to develop a more precise characterization of the developmental relationships between individual expression profiles. Initially, we focused on improving our treatment of filtering and normalization routines for single-cell multiplex qPCR data to achieve more accurate expression comparisons. Next, applied a non-linear dimensionality reduction technique to our revised expression profiles to illustrate the continuous developmental trajectory structure in our data and enhance our interpretation of ILC differentiation.

3.A.1 Revisited filtering and normalization for single-cell multiplex qPCR

Standard methods for interpreting expression measurements from quantitative real-time PCR (qPCR) data rely on comparison to a set of common housekeeping gene standards. qPCR measurements report the number of PCR amplification cycles required for a transcript to reach a pre-specified threshold level. The cycle number for a given transcript is then compared to the cycle number associated with a housekeeping gene to report a relative measure of gene expression. Specifically, these cycle number thresholds are typically subtracted to produce a logarithmic scale expression measure called ΔCt, or the change in cycle thresholds. Single-cell multiplex qPCR perform tens of qPCR measurements in parallel from transcript material portioned from single cells.

Due to the small amount of starting material used in single-cell multiplex qPCR experiments, it is not surprising that measurement preparation could fail entirely for individual cells. In these experiments, typically a few housekeeping genes are measured and cells are filtered if they contain a failed measurement for any housekeeping gene. In our experiment, we measured three housekeeping genes, *Actb*, *Gapdh*, and *Hprt*, and found that this filtering

metric is not always the most representative of outlier expression profiles. For some cells, expression measurements appear to be within normal ranges but a single housekeeping gene happens to be missing. Instead of relying on accurate measurement of only three housekeeping genes, we instead filter outlier cells based on extreme shifts in expression and low gene coverage. Specifically, when we plot the proportion of total genes expressed by the median z-scaled deviation in expression for measured genes for each cell, we clearly find outliers with low gene coverage and large shifts in measured expression that correspond to failed measurement (Figure 3.A.1A). We also observe some cells that have large shifts in measured expression despite having relatively normal gene coverage. We assume these cells were simply measured with reduced efficiency and that normalization procedures can regularize their expression values. We expect this filtering procedure is more robust than using a strict housekeeping gene requirement and retains some cells where a housekeeping gene measurement failed sporadically.

The most basic normalization strategy drawn from low throughput qPCR measurements is to report expression levels relative to housekeeping gene expression level. This strategy relies on accurate measurement of relatively few housekeeping genes as well as the assumption of their stable expression and fails to incorporate additional expression information from the large number of other genes measured. For high-throughput multiplex qPCR experiments, recent normalization methods have been developed based on those used for analysis of bulk expression profiling experiments [128]. However, these methods, such as quantile normalization, tend to produce expression values that are no longer as interpretable in terms of cycle threshold units and may be sensitive large shifts in gene coverage.

We developed a normalization strategy for single-cell multiplex qPCR based on an explicit regression model that accounts for different measurement efficiencies for each cell. Any scaling of the amount of transcript material capable of being measured results in a constant shift in all cycle threshold values measured for a given cell. For all genes with measured expression values in each cell, we fit the following regression model:

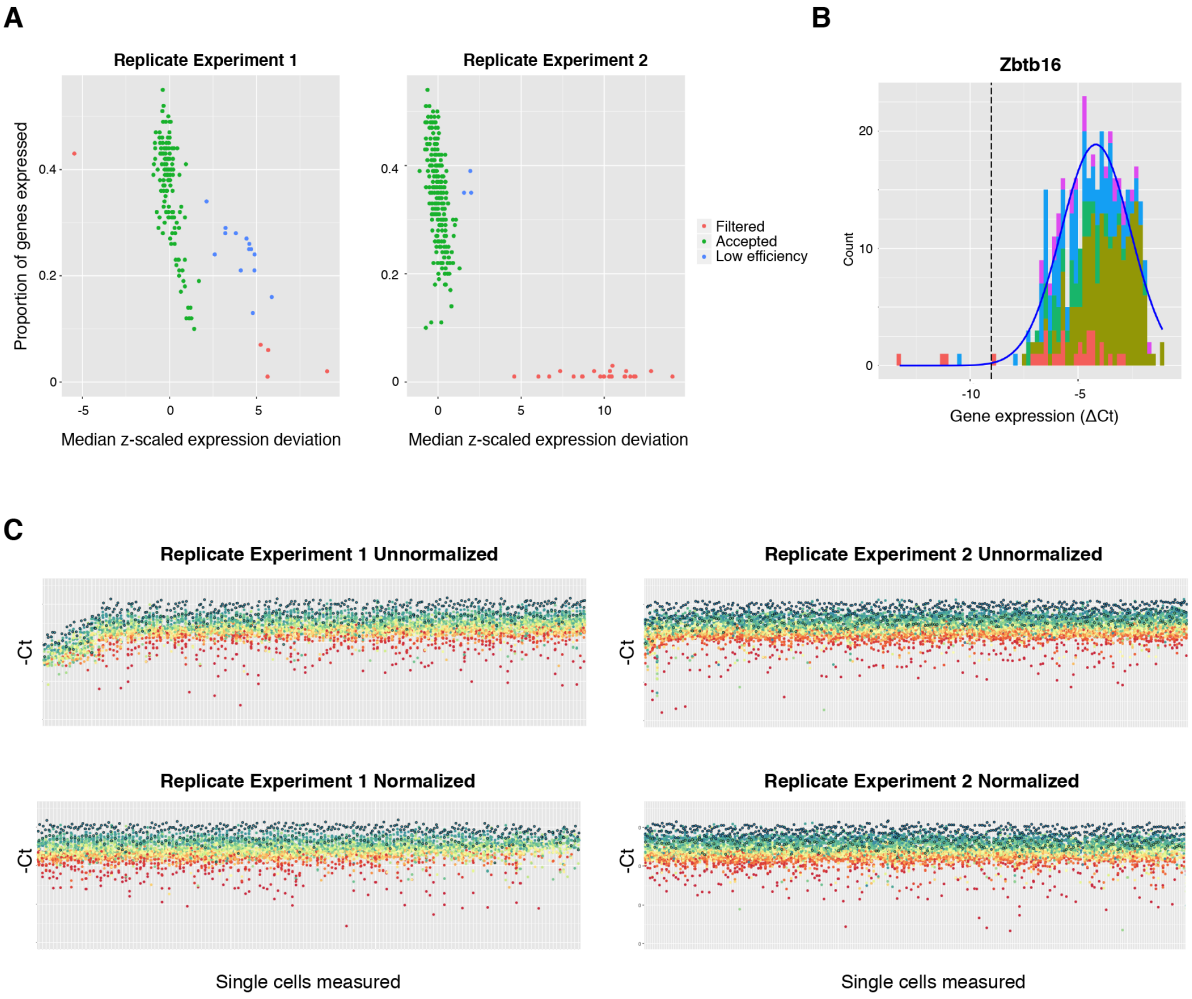


Figure 3.A.1: Data-driven single-cell multiplex qPCR filtering and normalization. A) Outlier cells were filtered if they had low gene coverage and large shifts in gene expression measurements. Low efficiency cells with normal gene coverage but large shifts in gene expression measurements were retained for further analysis. B) Distribution of measured ΔCt values for *Zbtb16* with Gaussian fit shown in blue and low expression threshold ($\mu - 3\sigma$) shown by the dotted line. Colors distinguish separate experiments. C) Regression model based normalization corrects for large shifts in gene expression. Scatterplot of measured expression values (–Ct) shown for each cell before and after normalization. Colors correspond to different genes.

$$Ct_{ij} = X_i + C_j + \varepsilon_i \tag{3.1}$$

Where Ct_{ij} is the cycle threshold measurement for gene i of cell j, X_i is the average cycle threshold value for gene i, and C_j is the constant shift in measurement efficiency for cell j. To obtain normalized expression values, we simply remove the contributions to expression measurements from the cell-specific and experiment specific constants and provide the remaining residual expression variance per gene. The distributions of measured expression values before and after normalization are illustrated in Figure 3.A.1C.

This procedure is analogous to standard housekeeping gene normalization, but utilizes the complete set of expression information to account for differing measurement efficiencies per cell. Accordingly, we expect this procedure to be more robust to possible inaccuracies in housekeeping gene measurement. We found that this regression-based method reduced measured gene expression variance for many more genes than achieved through housekeeping gene normalization (data not shown). For consistency with standard single-cell qPCR reporting, we calculate final expression values relative to the collective average housekeeping levels for consistency with typically reported expression measurements. This regression-based procedure could produce bias if there is a dominant and consistent shift in most measured expression values between different cell stages. We do not expect this would be a problem if we have a sufficiently diverse representation of genes with different expression dynamics in our assayed panel. This method could be restricted to only use a gene subset with desired distributional properties for further refinement.

Across all measured cells, the expression values for most genes are normally distributed in Ct with clear outlier values among low expression range, and can be directly seen in the distribution of *Zbtb16* measured values for example (Figure 3.A.1B). It is apparent these outlier values are unphysical, as they purport thousand fold reduction in expression compared to average gene expression levels. To regularize the expression values we use for downstream analysis, for each gene, we iteratively calculate a robust Gaussian fit to the

expression value distribution and remove outlier values beyond three standard deviations. We consider removed outlier expression values as equivalent to non-expressed measurements. In the calculation of distance measures between expression profiles, there is the question of how to treat non-expressed measurements relative to measured expression values. This choice effectively determines how we balance a difference in measured expression values to the cost of entirely turning the expression of a gene on or off. For distance metric calculation, we chose to impose a relatively strong penalty for switching gene expression on or off by placing non-expressed measurements 10 Ct below the 3σ low expression threshold defined by our Gaussian distribution fits for each gene. This choice is justified since across our data, we found that most expression variance is due to on-off gene expression, even when calculated with non-expressed genes most conservatively placed right at the low expression thresholds.

3.A.2 Dimensionality reduced representation of ILC precursor expression profiles

We identified a few salient features of our measured population of single cell precursors through hierarchical clustering. We observed dynamic changes in transcription for many genes across ordered developmental stages. In particular, we found a set of transcription factors that are sequentially induced early in ILC specification, as well as a number of transcription factors with transient expression dynamics. Additionally, we found a substantial portion of ILCP showed signs of early lineage differentiation, indicating that within our data the singular developmental path in early ILC specification is branching in later stages. These characteristic properties of a developing precursor population are indicative that representation of this data by a group of discrete clusters may not be the most appropriate. Although we might expect there to be developmental transitions distinguished by induction or down-regulation of transcription factors and cell markers, we would expect these transitions should appear to be continuous with sufficient single-cell sampling. In particular, in studying developmental trajectories we dont expect cell stages to be distinctly well-separated clusters,

which is typically assumed in the design of clustering algorithms. Furthermore, clustering metrics often measure the degree of separation between clusters, which makes evaluation of clustering quality between adjacent stages difficult. For these reasons, other methods for interpretation besides a clustering-based approach could be more appropriate for a continuous developmental process.

Essentially, we would like to reconstruct cellular trajectories that are representative of the neighbor relationships between single-cell expression profiles. Dimensionality reduction techniques are often used to visualize and identify relationships between subgroups in a collection of high-dimensional data; the most commonly applied technique is principal components analysis. Recently, a number of single-cell analysis methods utilize forms of dimensionality reduction to identify the geometries of developmental trajectories and have been further extended to associate individual expression profiles with developmental pseudotime, quantitative measure of biological progression that effectively order transiting cells by developmental stage [80]. For our purposes, we apply a non-linear dimensionality reduction technique, t-distributed stochastic neighbor embedding (t-SNE), to illustrate the interpretive benefits a dimensionality reduction-based strategy has in developmental analysis. The aim of t-SNE is to identify a non-linear mapping of data points in a high-dimensional space to points in a low-dimensional space (2D or 3D) in a way that best preserves high-dimensional neighbor relationships in the low-dimensional representation [129]. Visualization of single-cell population structure has been successfully performed using a technique based on t-SNE [130]. In our case, we expect preservation of local neighbor relationships in a low-dimensional representation will correspond to adjacent placement of developmentally similar precursor cells, and any preservation of longer-range expression profile similarities should reflect the large-scale geometry of developmental trajectories. The balance between preservation of local and long-range cell relationships is controlled by a parameter called the perplexity in t-SNE, and is essentially the extent of neighbor relationships intended to be satisfied in the mapping.

We calculated the complete set of Euclidean distances between single-cell expression profiles using the normalized and adjusted qPCR expression values as described above and performed t-SNE projection to two dimensions. We found that a perplexity value of 30 was capable of resolving the lineage differentiation and branching structure discussed below. For clarity, mast cell contaminants discussed in the previous chapter are removed from all dimensionality reduction figures presented in this section.

RESULTS

By coloring cells by their assigned hierarchical clusters defined in the previous chapter, it is clear that this continuous representation recapitulates the same developmental progression we previously inferred (Figure 3.A.2A). Specifically, we see a contiguous developmental trajectory that starts with αLP cluster cells, then transitions through cluster B cells, and is completed by ILCP cluster cells. This continuous representation emphasizes that these cluster transitions appear quite continuous, a feature that could not be easily determined through clustering. It is also apparent that our previously described developmental stages appear to be necessary transition stages that span the primary developmental path. For example, any minimal trajectory from αLP to ILCP must transit through cluster B, corroborating our previous interpretation of this intermediate stage. Overall, there are no apparent gaps or large jumps between developmental stages, which would indicate potentially missing transition states. This is somewhat surprising given that we know that EILP precursors were not surveyed in our experiment due to their lack of IL-7r expression and might point to the existence of alternative developmental paths.

This continuous developmental representation enables the visualization of gene expression induction without the need to perform precise stage-specific clustering. In a clustering-based approach, improper cluster resolution could be detrimental to the interpretation of developmental progression. If clusters are too large and inclusive, we might average over and effectively miss a transitional stage; on the other hand, if clusters are too small and overfit-

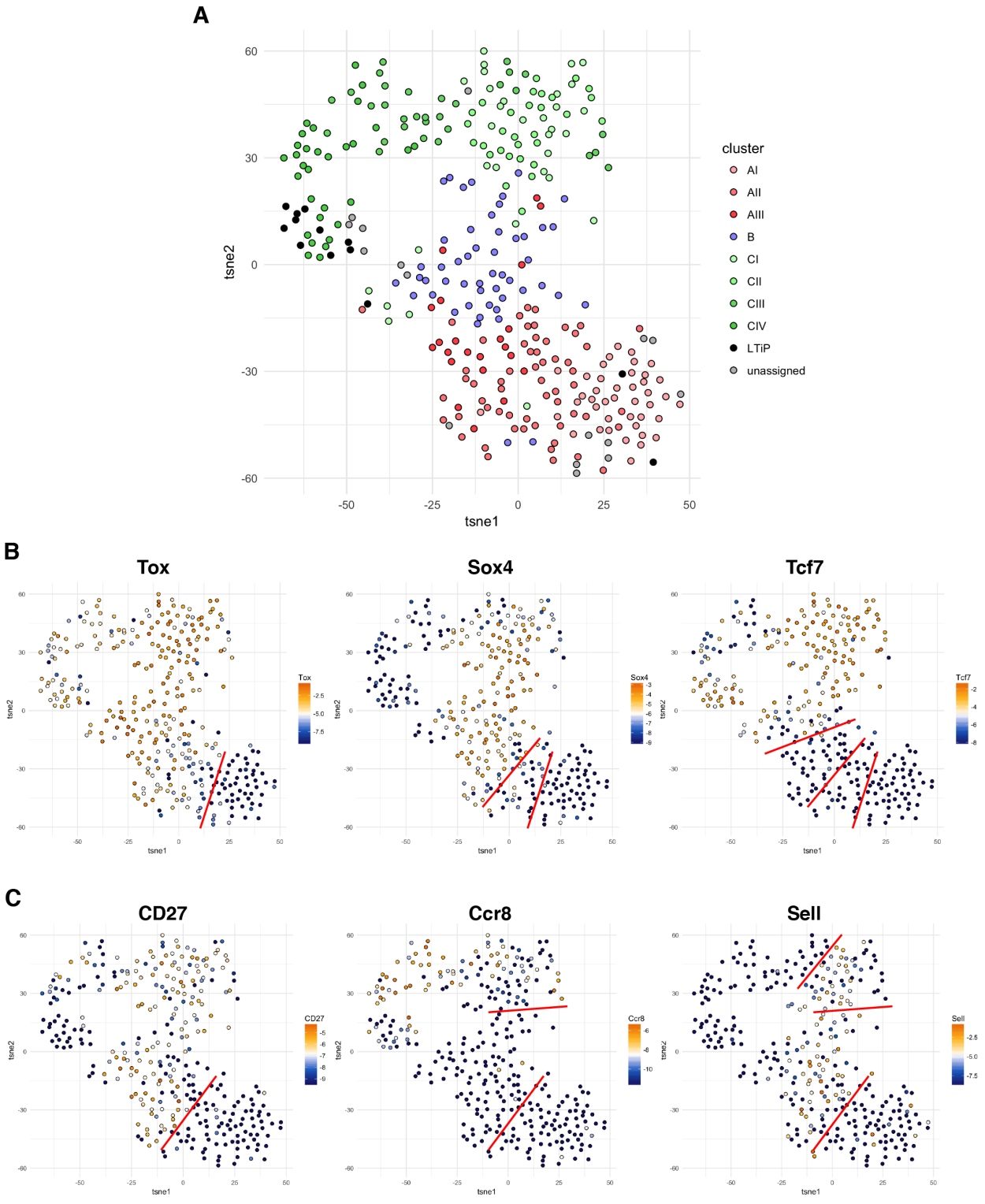


Figure 3.A.2: ILC developmental progression along dimensionality reduced (t-SNE) representation.
A) Continuous developmental representation of ILC precursors with hierarchical clusters colored. B) Sequence of transcription factor induction along primary developmental branch. C) Sequence of surface marker expression along primary developmental branch. Stages of induction highlighted in red.

ting, we might erroneously interpret heterogeneity as distinct developmental stages. These concerns are analogous to proper choice of the perplexity parameter in dimensionality reduction by t-SNE, although it is much easier to determine an appropriate range of perplexity values through inspection of the resulting visualizations than it is to determine the most appropriate number of clusters. Accordingly, we expect inference of the relative timing of gene expression induction based on our continuous developmental representation to be less error-prone and more precise. On our continuous representation, by examining the expression *Tcf7*, *Sox4* and *Tox*, we can see the same early sequence of transcription factor induction in αLP as we had found before but with clear lines of demarcation across the primary developmental branch (Figure 3.A.2B). While the measured cells are destroyed in single-cell expression profiling, these experiments can reveal potential surface markers that delineate developmental stages and are compatible with further experimental manipulation like cell sorting. Among our measured genes, we found that expression of the surface markers *CD27*, *Ccr8*, and *Sell*, tend to have distinct expression patterns along the ILC developmental path (Figure 3.A.2C). The use of these markers could enable better mapping of developmental potential to precursor expression profiles as well as detailed investigation of particular stages in differentiation.

A clearly missing facet in our hierarchical clusters was the resolution of ILC lineage branches. Although we observed early differentiation towards all three types of ILC lineages through direct examination of lineage-defining factors, no hierarchically defined clusters directly corresponded to these distinct lineages. In contrast, these distinct lineage branches are clearly resolved in our continuous developmental representation (Figure 3.A.3A-D). Previously, cells belonging to these differentiating lineages were split between clusters that broadly correspond to earlier and later stages of ILC differentiation after induction of PLZF. Essentially, the gene expression commonalities that correlate with differentiation extent were more dominant in clustering than lineage-specific similarities in gene expression among these cells. Through a two-dimensional continuous representation, we can simultaneously represent de-

velopmental progression from earlier to later states of differentiation in concert with divergence between lineage branches. Specifically, we see that expression of *Icos* distinguishes the differentiating ILC2 lineage branch, expression of *Tbx21* distinguishes the differentiating ILC1 lineage branch, and expression of *Rorc* and *Lta* expression distinguishes the differentiating ILC3 branch while simultaneously marking separately sorted LTiP cells. Among differentiating ILCP primed towards the ILC2 lineage, we can observe a variable extent of commitment, where some cells express effector genes such as *Il13* and *Il33r* in addition to *Icos*, *Bcl11b* and *Gata3* (data not shown).

Interestingly, our continuous developmental representation also provides a very direct visualization of branching between the LTi and ILC lineages. In our hierarchical clustering analysis, cells diverging towards the LTi lineage had to be separately extracted from the B cluster population through filtering on a PLZF^{-} Tcf7^{+} signature. In our continuous representation, expression of *Rorc* clearly highlights a protrusion of cells from cluster B diverging from the primary ILC developmental path. Furthermore, this protrusion extends directly towards the differentiated ILC3 and LTiP lineage branch, illustrating shared characteristics between these diverging cells and the type 3 effector program. In this case, there is a gap between the LTi primed cells diverging from cluster B and the differentiated ILC3 and LTiP, confirming that we did not sample precursor cells that span LTiP differentiation. The clear identification of LTi divergence from cluster B through dimensionality reduction further substantiates our inference that LTi branching closely follows the late αLP developmental stages.

Finally, the placement of multilineage transcriptional primed precursor cells in our continuous developmental representation can provide an indication as to whether these states are required for ILC differentiation. If multilineage ILC precursors form a compact developmental stage that ILCP must transit through before lineage commitment, we might expect that multilineage transcriptional priming is a required developmental state for ILC differentiation. Instead, we observe sporadic multilineage expression in both early and late stage

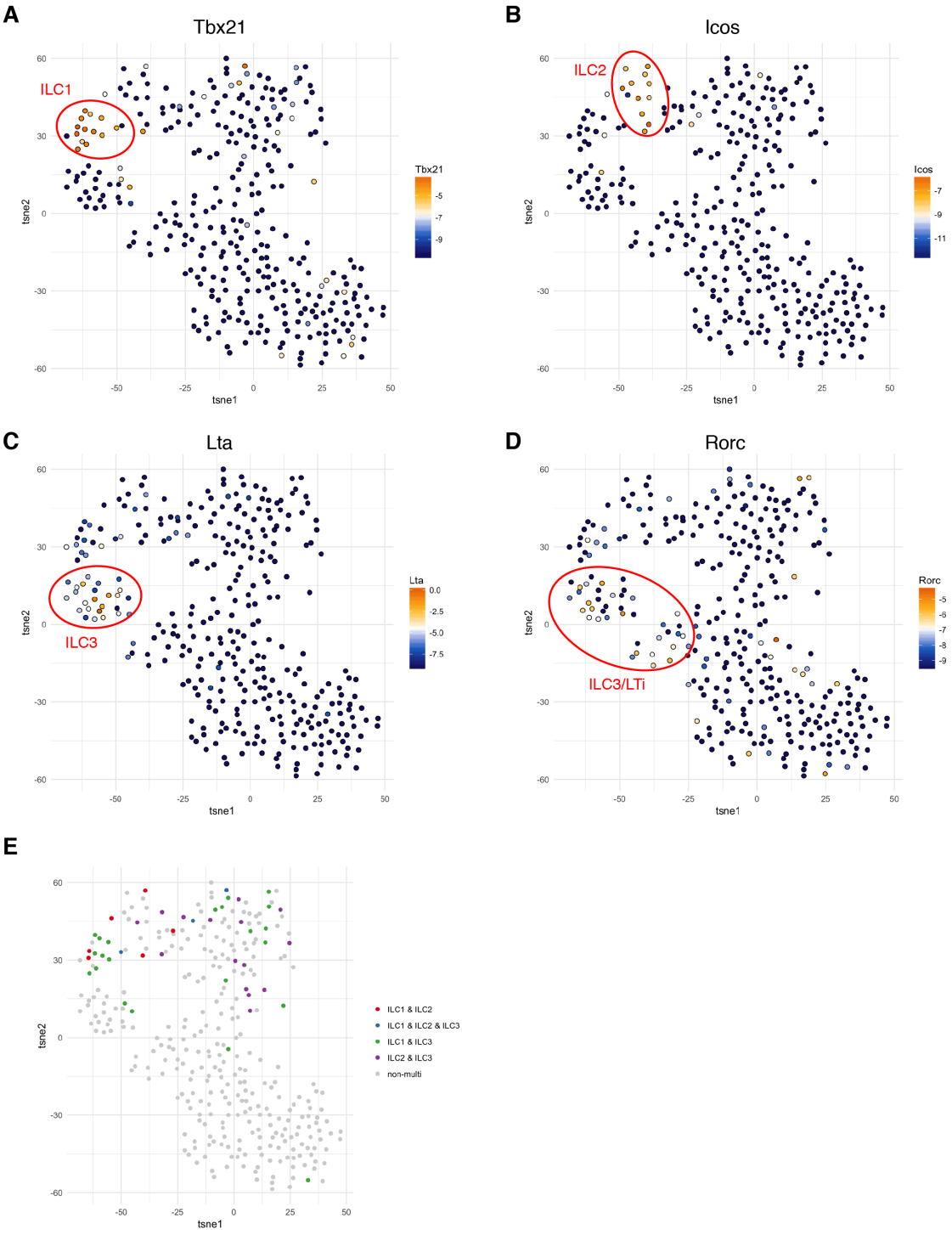


Figure 3.A.3: ILC lineage branching on dimensionality reduced (t-SNE) representation. A-D) Expression of lineage-defining factors identify branching trajectories on continuous developmental representation. Specific lineage branches highlighted in red. E) Multilineage transcriptionally primed ILC precursors are distributed sporadically on ILC developmental path. Colors indicate simulation expression of a specific combination of lineage-defining factors (*Tbx21* for ILC1, *Bcl11b* for ILC2, and *Rorc* for ILC3).

ILCP as well as even among seemingly lineage-committed cells largely without a consistent pattern (Figure 3.A.3E). The only consolidated set of multilineage precursors are those with bilineage priming towards the ILC1 and ILC3 lineages, which interestingly almost completely overlap with high expression of *Tbx21*. Accordingly, it appears that ILC1 lineage primed ILCP also tend to express *Rorc*. Overall, this suggests a sporadic or stochastic mechanism for multilineage transcriptional priming rather than deterministic staging. While we might be undersampling multilineage primed states due to gene expression dropout associated with single-cell measurements, the multilineage states we do measure are sporadically placed relative to one another.

In summary, we extend our hierarchical clustering analysis of single ILC precursor expression profiles by refining our treatment of single-cell multiplex qPCR data filtering and normalization, and then applying a dimensionality reduction procedure to generate a continuous representation of the evidenced ILC developmental trajectories. This dimensionality reduction strategy enhances our interpretation of developmental stage progression, lineage divergence and differentiation as well as multilineage transcriptional priming in ILC precursor populations.

CHAPTER 4

GENOME–WIDE CHROMATIN ACCESSIBILITY CHANGES REVEAL INFLUENTIAL REGULATORY FACTORS IN ILC PROGENITOR

4.1 ABSTRACT

Innate lymphoid cells (ILCs) develop from common precursor downstream of lymphoid progenitors. Restriction of developmental potential to innate lymphoid lineages from the common lymphoid precursor is accompanied by a rich set of changes in transcriptional regulation. By a combination of genome-wide transcriptional and chromatin accessibility profiling, we identify regulatory factors affiliated with early epigenetic features that define the ILC progenitor. Our findings highlight the predominant roles TCF, ROR, GATA, and RUNX family members have in promoting ILC differentiation while elaborating their global appearance patterns in enhancers.

4.2 INTRODUCTION

Innate lymphoid cells (ILCs) monitor peripheral mucosal tissues and are responsible for rapid cytokine secretion in response to inflammatory signals. As innate effectors, ILCs mediate immunity in the absence of somatically rearranged antigen receptors characteristic of adaptive B and T cells. Interestingly, ILCs have a shared developmental origin with the B and T cell lineages and are derived from the common lymphoid progenitor (CLP) [3, 131]. Following divergence from the adaptive cell fates, CLP are capable of differentiating into three groups of differentiated ILC forms analogous to the three classes of $CD4^+$ T helper (Th) cell subsets in both cytokine production and transcriptional determination [1]. Type 1 ILCs, including conventional natural killer cells (cNKs) and ILC1, produce IFNγ and express the lineage-determining transcription factor Tbx21 (also called T-bet) similar to Th1 cells.

ILC2s produce IL5 and IL13 and express the lineage-determining transcription factor Gata3 similar to Th2 cells. Finally, type 3 ILCs, including lymphoid tissue inducer cells (LTi) and a variety of ILC3 cells, produce IL17 and IL22 and express the lineage-determining transcription factor RORγt.

In the effort to comprehensively map the developmental branching structure for this diverse set of innate lineages downstream of the CLP, there has been a recent surge in reported ILC progenitor populations with varying degrees of developmental restriction [132]. In this study, we focus on the innate lymphoid cell precursor (ILCP), a committed ILC progenitor defined by expression of the transcription factor PLZF (encoded by *Zbtb16*) [29]. Transfer studies demonstrated that the ILCP efficiently gives rise to all peripheral ILC lineages, ILC1, ILC2 and ILC3, but not LTi, NK, B and T cells. Furthermore, in an in vitro differentiation system, single ILCP progenitors simultaneously gave rise to all three ILC subtypes, confirming a multi-lineage developmental potential. Thus, the ILCP is the farthest downstream ILC progenitor prior to ILC effector differentiation. Accordingly, we are studying the developmental transition of CLP to the ILCP to identify transcriptional and epigenetic changes that specifically define commitment to the common ILC lineage.

Currently, the studies that define ILC developmental precursors have exclusively characterized gene transcription and protein expression in different ILC developmental stages. Distinct epigenetic landscapes, including DNA and histone modifications as well as chromatin accessibility, found in different cell types effectively contextualize the action of expressed regulatory factors, since non-coding regulatory elements with bound transcription factors can control gene expression from long genomic distances. In this way, analysis of regulatory enhancers can reveal both factors induced downstream of transcriptional regulators as well as regulatory network interactions. Chromatin accessibility profiling of differentiated ILC effector subsets has been used to show that differentiated ILC contain open chromatin at effector cytokines prior to their induction [133]. Additionally, chromatin accessibility profiling of intestinal ILC effectors characterized the degree of versatility in ILC response

to diverse microbial environments [134]. While these studies profile differentiated forms of ILCs, systematic study of chromatin regulation during early ILC development has not yet been performed to our knowledge. It is clear that many regulatory factors play important roles during ILC development and that regulatory factor expression is also very dynamic in early ILC specification [132]. Given that epigenetic information can propagate over longer time-scales than gene expression during development [135, 136], by examining chromatin regulation in ILC progenitors we can deepen our understanding of ILC regulatory factor interactions and genome-wide influences.

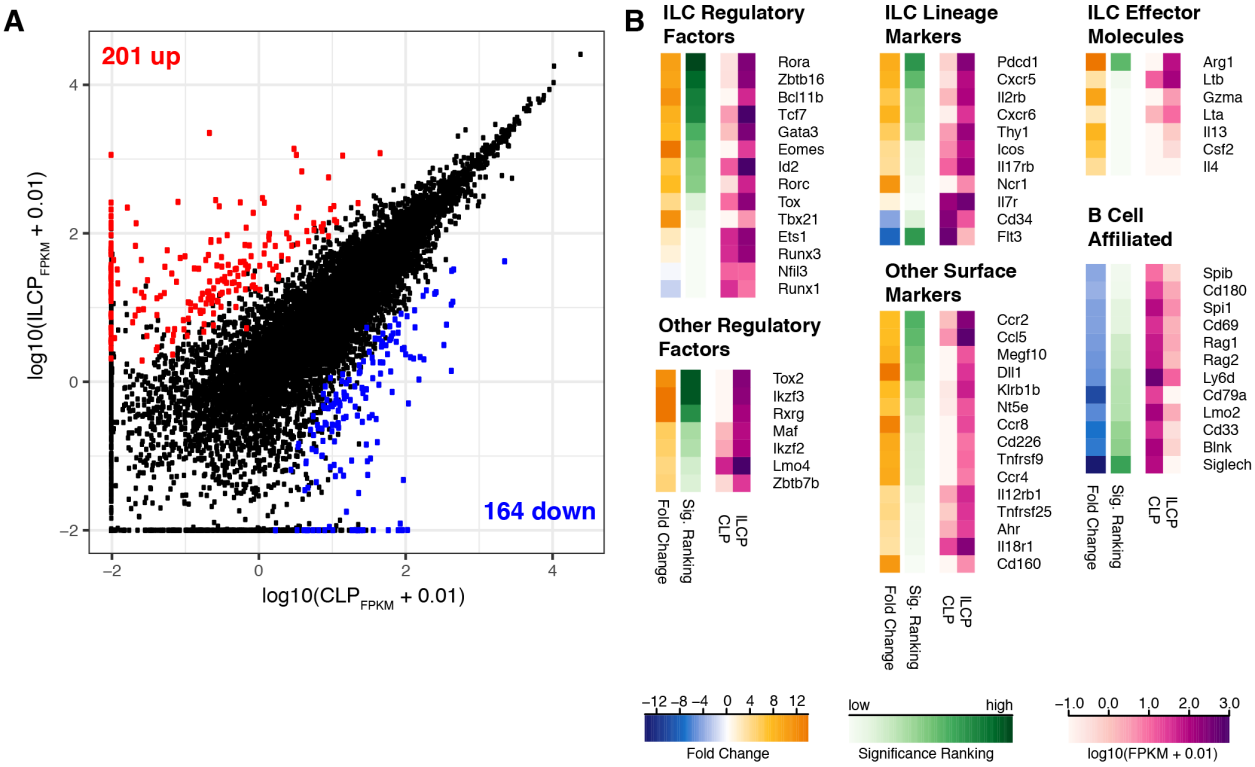
Here we present the simultaneous characterization of genome-wide transcription and chromatin accessibility changes that occur during ILC lineage restriction. We identify the comprehensive set of transcriptional changes that define the transition of CLP to ILCP, including several transcription factors with uncharacterized roles in ILC development. Our corresponding genome-wide regulome analysis revealed that many distal regulatory elements are substantially changed during this transition. We performed statistical modeling of transcription factor motif presence in regulatory elements to uncover the transcription factors most strongly associated with particular chromatin dynamics. In addition to regulatory motifs that enrich for opening or closing chromatin, we also found a large number of regulatory motifs that enrich specifically for unchanged chromatin. Our systematic survey of regulatory factors predictive of chromatin dynamics revealed that TCF, ROR, GATA, and RUNX family members are widely impactful regulators during ILC specification.

4.3 RESULTS

4.3.1 Transcriptional changes from CLP to ILCP

We performed genome-wide RNA sequencing (RNA-Seq) analysis to establish a comprehensive transcriptional characterization of the ILCP. We isolated ILCP (defined as Lin^- Il7r^+ $\alpha_4\beta_7{}^+$ PLZF^+ Cxcr5^-) and CLP (defined as Lin^- Il7r^+ cKit^{int} $\text{Sca-1}^{\text{int}}$ $\text{Flt3}^{\text{high}}$) populations

in adult bone marrow extracted from PLZF-GFP reporter mice as previously described [29]. Due to the rarity of the ILCP (only thousands of cells per mouse), we could only perform sequencing of a single biological replicate of ILCP. To distinguish substantial changes in transcription, we performed differential expression analysis with a fixed biological coefficient of variation and chose a significance cutoff that yielded 201 upregulated and 164 downregulated genes (see methods for details, Figure 4.1A).



C

ID	Name	Size	Enrichment Score	FDR
KEGG:hsa04060	Cytokine-Cytokine Receptor Interaction	148	3.62	<0.001
KEGG:hsa04514	Cell Adhesion Molecules (CAMs)	76	3.07	<0.001
KEGG:hsa04640	Hematopoietic Cell Lineage	63	2.8	<0.001
GO:0006954	Inflammatory Response	290	4.03	<0.001

Figure 4.1: Transcriptional changes from CLP to ILCP.
A) Log scale FPKM values in CLP and ILCP. Red, upregulated; blue, downregulated. B) Select transcriptional characterization of CLP and ILCP. C) Top enriched KEGG and GO pathways between CLP and ILCP.

Despite only affording comparison between single replicates, the derived gene rankings demonstrate the quality of transcriptional profiling (Figure 4.1B). Many key transcription

factors that regulate ILC development appear among the most substantially changing genes, including *Zbtb16* (encoding PLZF), *Tcf7*, *Id2*, *Tox*, *Bcl11b* and *Gata3*. Each of these regulatory factors has been individually shown to be necessary for proper ILC development and each is highly expressed in ILCP. Other key regulators like *Nfil3* and *Runx1* were expressed in both CLP and ILCP but appeared unchanged in the transition from CLP to ILCP. In single-cell transcriptional studies [64], these factors have transient transcriptional dynamics immediately prior to ILCP state, which were likely averaged out when sequenced in bulk. We also observe substantial upregulation of ILC lineage-specific transcription factors such as *Tbx21*, *Rorc*, and *Rora*. From single-cell profiling, we expect one third of ILCP to be undergoing initial steps of differentiation and so the appearance of these lineage restricted transcription factors likely derive from these differentiating subpopulations.

We found an additional set of transcription factors without well-characterized roles in ILC development, including *Ikzf2*, *Ikzf3*, *Rxrg*, *Maf*, *Lmo4*, *Tox2* and *Zbtb7b*. Two members of the Ikaros family of zinc-finger transcription factors, Ikzf2 and Ikzf3, are substantially upregulated in ILCP. Members of the Ikaros family of transcriptional regulators are widely expressed throughout hematopoiesis and influence a variety of cell-fate decisions [137]. Ikzf3, also known as Aiolos, can influence the development and survival of the B cell fate and is one of the most upregulated genes in ILCP. Rxrg appears strongly upregulated in ILCP, though from single-cell investigation we know Rxrg expression is affiliated with the ILC2 lineage [64]. Since Maf also plays important roles in the differentiation of T helper cell subsets [138], expression of Maf in ILCP could represent an additional shared developmental determinant between the lineages. Many of these transcription factors have appeared consistently in ILC profiling studies [139, 115, 64, 51], but whether these factors have precise roles remain to be determined.

These changes in expression of key regulators coincide with expected changes in surface-expressed lineage markers, including upregulation of *Pdcd1*, *Il7r* and *Thy1* as well as down-regulation of *Cd34* and *Flt3*. Additionally, we observe upregulation of many ILC lineage

markers, such as *Ncr1*, *Il2rb*, *Icos*, *Il17rb*, *Cxcr5* and *Cxcr6*. The inhibitory receptor Pdcd1 has been recently revealed to coincide with PLZF expression and therefore acts as an effective alternative marker to Zbtb16-GFP for ILC progenitor populations [12]. In our comparison, *Pdcd1* is one of the most strongly upregulated genes in ILCP. Among other immunoregulatory surface-expressed molecules we observed upregulation of interleukin receptors *Il18r1* and *Il12rb1*, *Ccl5*, chemokine receptors *Ccr2*, *Ccr4* and *Ccr8*, tumor necrosis factors *Tnfrsf9* and *Tnfrsf25*, as well as *Ahr*, *Megf10*, *Dll1*, *Klrb1b*, *Nt5e*, *Cd160* and *Cd226*. As the ILCP represents the developmental stage just prior to significant differentiation, we only observe slight upregulation of a few ILC effector cytokines, including *Arg1*, *Gzma*, *Il4*, *Il13*, *Csf2*, *Lta*, and *Ltb*. At this stage, the expression of ILC specific transcription factors are much more substantial that those of effector molecules as expected.

The most substantially downregulated genes in ILCP included many B cell development affiliated factors. The ETS family transcription factors Spi1 (or PU.1) and Spib are required for functional lymphoid lineage priming and B cell differentiation and were much more highly expressed in CLP than in ILCP. The recombination-activating genes Rag1 and Rag2, which are responsible for adaptive antigen receptor rearrangement were also notably downregulated in ILCP. One of the most strongly downregulated genes in ILCP was Blnk, a linker protein in B cell receptor signaling pathways required for complete B cell development [140]. We also observed downregulation of B cell associated surface markers, including *Cd180*, *Cd69*, *Cd79a*, *Ly6d* and Siglec family members *Cd33* and *Siglech*. The downregulation of B cell affiliated factors in ILCP was contrasted by the upregulation of a few of genes associated with T cell differentiation, including T cell receptor constant region genes, *Themis*, and T cell signaling genes *Txk* and *Itk*. As these T cell genes are associated with adaptive signaling complexes, it is unlikely they play functional roles in ILCs, and perhaps more likely indicate a low level of sequencing contamination.

Gene set enrichment analysis confirmed that the predominant biological processes being altered are immune-related, as we found cytokine-cytokine receptor interaction, cell adhesion

molecules, and hematopoietic cell lineage comprise the three most enriched KEGG pathways while inflammatory response was the most enriched GO pathway (Figure 4.1C). These transcriptional profiles confirm the relevance of known ILC developmental factors as well as suggest new regulatory factors and markers by global profiling.

4.3.2 Enhanced distal chromatin accessibility during ILC specification

We further examined the global regulatory landscape during ILC specification by an assay for transposase-accessible chromatin using sequencing (ATAC-seq) [87]. The modest requirements for starting material allowed us to perform this assay with the rare ILC progenitor population isolated ex vivo. To allow direct comparison to transcriptional profiling, we analyzed the transition from CLP (3 replicates) to ILCP (2 replicates). We found 50,000 total candidate regulatory elements shared among all replicates of either CLP or ILCP. Analysis of differential read-counts revealed 12,367 significantly changing regulatory element peaks between stages, with 6404 upregulated and 5963 downregulated (Figure 4.3A). The greater proportion of ILCP-specific peaks and their larger absolute fold changes illustrates an overall bias towards chromatin opening in ILCP by both peak number and strength.

We next examined the genomic locations of identified regulatory elements and found an overall enrichment (20-fold) of called peaks in promoter regions ($\pm$1kb of transcription start sites (TSSs)). More specifically, unchanged peaks retained in CLP and ILCP at similar strengths are particularly enriched near gene TSS (Figure 4.3B). Peaks with less than 2 fold change occur more than 30 times expected by chance (Figure 4.3C-D). This location bias suggests that ILC-specific chromatin changes are almost exclusively distal.

For a few key ILC lineage-associated genes, we observe largely concordant regulation of local chromatin and transcription (Figure 4.2). For Zbtb16, many new regulatory elements appear adjacent to the TSS and in a region of $\sim$200kb extending into intronic regions of the gene. A number of regulatory regions adjacent to Id2 have greatly increased openness in ILCP as well. As Zbtb16 and Id2 are strongly upregulated in ILCP, we would expect these

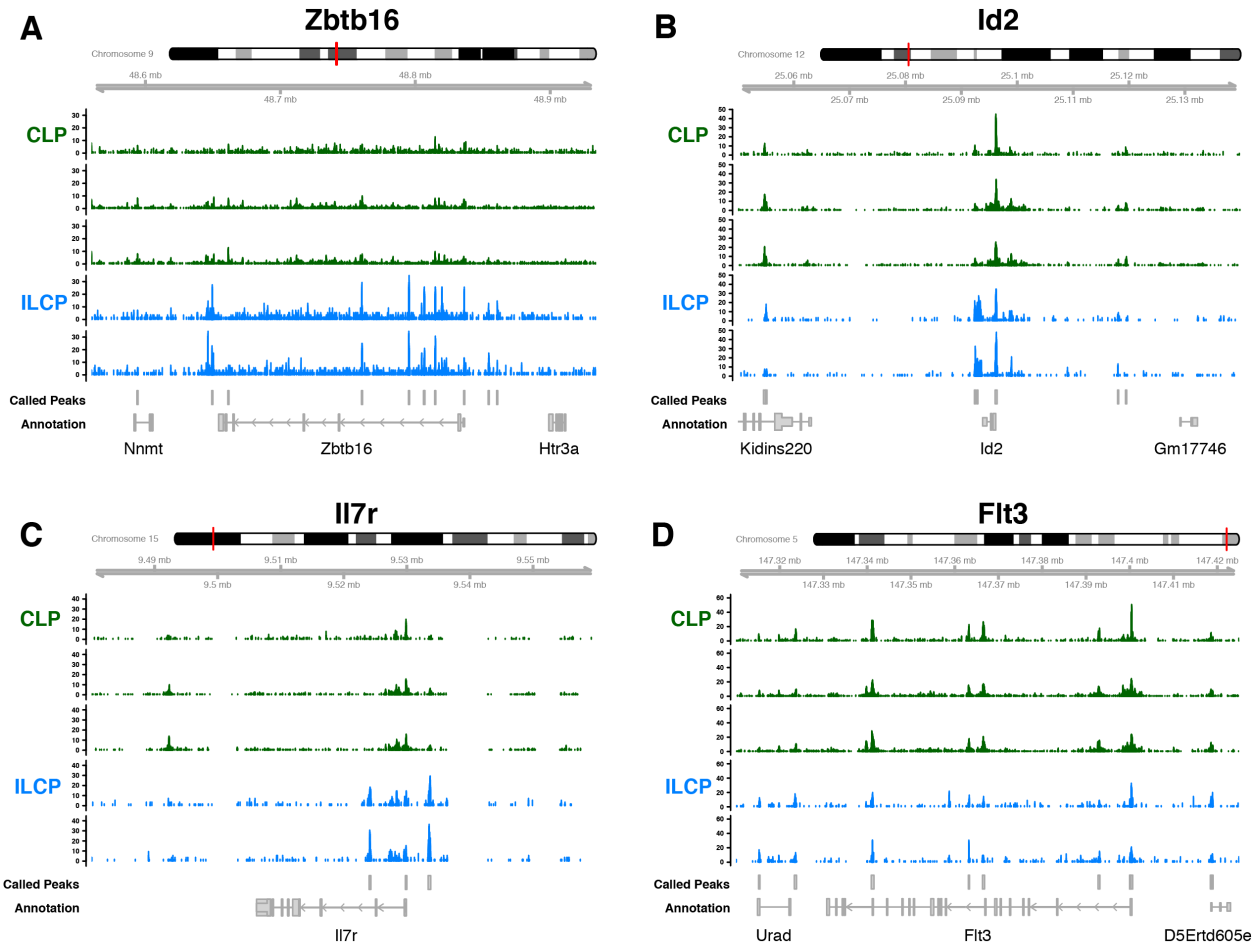


Figure 4.2: Chromatin accessibility changes from CLP to ILCP.
ATAC-seq coverage for Zbtb16, Id2, Il7r, and Flt3 in CLP and ILCP samples. Peak features called in either all CLP or all ILCP samples annotated in grey.

ILCP-specific regulatory elements to be responsible for their upregulation in ILCP. Interestingly, while Il7r is already transcribed in CLP, ILCP-specific regulatory elements that appear near the gene TSS accompany more solidified Il7r expression at the ILCP stage. In the other direction, most of the strong regulatory peaks in CLP adjacent to the downregulated marker gene Flt3 have a greatly reduced footprint in ILCP. The general correspondence between chromatin regulation and transcription is not as direct as that illustrated in the aforementioned examples. The presence of discordant regulatory elements even in the aforementioned examples, including reduced peaks near Id2 and Il7r as well as new peaks near Flt3, illustrates the complication.

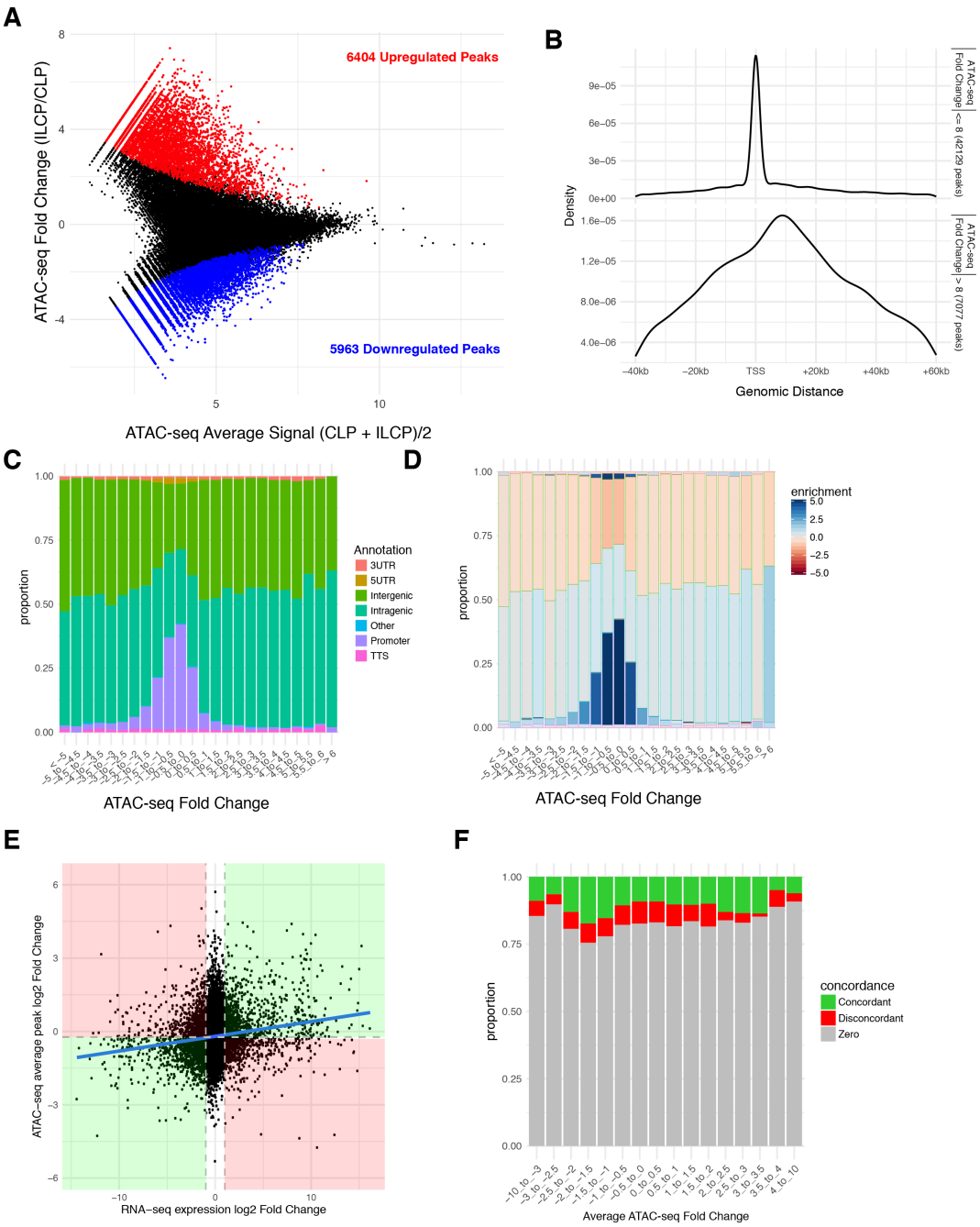


Figure 4.3: Enhanced distal chromatin accessibility during ILC specification. A) Increasing and decreasing regulatory elements as called by edgeR. B) Density distribution of unchanged compared to strongly changing peaks around aligned gene TSS. C) Frequency of peaks found in annotated genomic locations and D) relative enrichment of peak frequency in each genomic location. E) Correspondence between transcriptional fold change and average peak fold change in 50kb window around gene TSS. Concordant region highlighted in green and discordant region highlighted in red. Linear fit line shown in blue with Pearson correlation coefficient of 0.15. F) Proportions of genes with concordant or discordant transcriptional and chromatin dynamics grouped by average chromatin accessibility fold change.

We observe many more substantial changes in chromatin accessibility than transcriptional changes across the genome. The average fold change of regulatory elements within 50kb of a gene TSS is weakly correlated with the corresponding transcriptional change (Pearson correlation coefficient of 0.15) (Figure 4.3E). By using this local regulatory element fold change statistic to categorize genes with concordant chromatin and transcriptional changes, only ∼20% have substantial transcriptional changes. Among these, most tend to be concordant with a greater bias associated with larger chromatin categories (Figure 4.3F).

This simplified analysis shows that among genes with large changes in chromatin accessibility, a small subset have concordant transcriptional changes while most remain transcriptionally unchanged. These trends point to fact that more specific attributes of peaks besides fold change and genomic location are necessary to understand how the observed chromatin changes influence gene expression.

4.3.3 Motifs associated with distinct chromatin profiles

To identify predominant regulatory factors acting in ILCP transition, we surveyed all identified regulatory elements for transcription factor motif consensus sequences. We scanned all annotated transcription factor consensus sequences from five transcription factor repositories (JASPAR, Homer, Chen, Hocomoco, Uniprobe) with FIMO to obtain locations of highly significant motif matches [141]. We used a hierarchical clustering strategy on motif overlap proportions to empirically group largely redundant motifs.

While we initially expected there to be motifs correlating with opening and closing of chromatin, we were surprised to find a set of motifs distinctly associated with unchanging chromatin accessibility between CLP and ILCP. In particular, there were a number of motifs whose appearance strongly enriches for peaks with limited absolute fold change (between 1.3 to 0.7) (Figure 4.4A).

To screen for the most influential motifs on peak dynamics, we constructed a multinomial logistic regression model that accounts for three distinct groups of peaks (increasing,

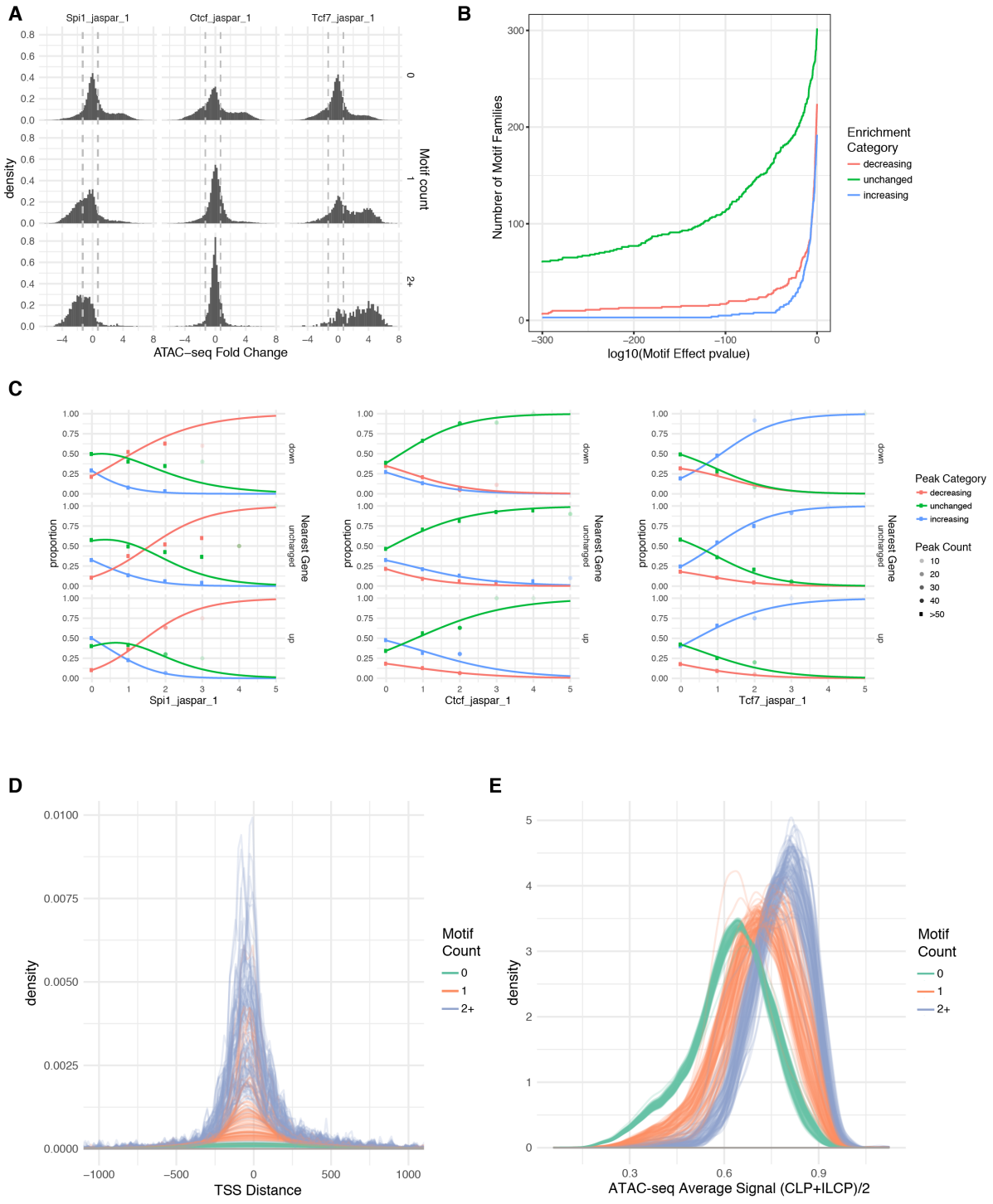


Figure 4.4: Transcription factor motifs associated with distinct chromatin profiles. A) Enrichment of delineated decreasing, unchanged and increasing regulatory elements by Spi1, Ctcf, and Tcf7, respectively. B) Number of significantly influential motif families enriching for peak group categories. C) Multinomial modeling of peak group categories. D) Motifs that enrich for unchanged regulatory elements are preferentially found in promoter regions. E) Motifs that enrich for unchanged regulatory elements are preferentially found in stable peaks with high read-counts.

unchanging, decreasing) partitioned by the previously mentioned fold change delimiters (Figure 4.4C). We use this model to evaluate whether the number of times a motif is found in a regulatory element significantly influences the probability that a regulatory element belongs to one of the increasing, unchanging, or decreasing sets of peaks. We found it was necessary to include the RNA expression status (upregulated, unchanged, or downregulated) of the nearest gene as confounder and interaction term, due to the previously described baseline correspondence between transcriptional changes and local chromatin openness. For each motif, we fit this multinomial model and calculated the significance of the motif coefficients from type I ANOVA with interaction. We Benjamini-Hochberg corrected the full set of significance values to associate motifs with corresponding false discovery rates (FDRs).

Generally, many motifs were associated with significant multinomial regression models, likely due to the large number of predictive peaks in dataset. There were slightly more motifs that tend to enrich for increasing regulatory elements than those that enrich for decreasing regulatory elements, but by far most motifs actually enrich for the empirically defined set of unchanged regulatory elements (Figure 4.4B). The group of motifs that enrich for unchanged peaks is comprised of transcription factors with seemingly mixed functional origins. Notably, this group includes CTCF, an architectural protein responsible for defining chromatin looping structures and organization. We also observed a number of E2F family transcription factors as well as Myc, which have roles in cell cycle progression. Other transcription factor families found notably bind GC-rich motif sequences, including SP, EGR, KLF, and AP-2. Collectively, we observed that the presence of these motifs particularly enrich for regulatory elements in promoter regions (Figure 4.4D), and also for regulatory elements with high average read-count in CLP and ILCP (Figure 4.4E), which corroborates that these motifs specifically enrich for persistent regulatory features.

Among all motif groups that significantly enrich for increasing regulatory elements, the TCF, ROR, GATA and RUNX motif families stand out with exceptionally large influences on peak dynamics (Figure 4.5A). Each of these families has multiple individual motif mem-

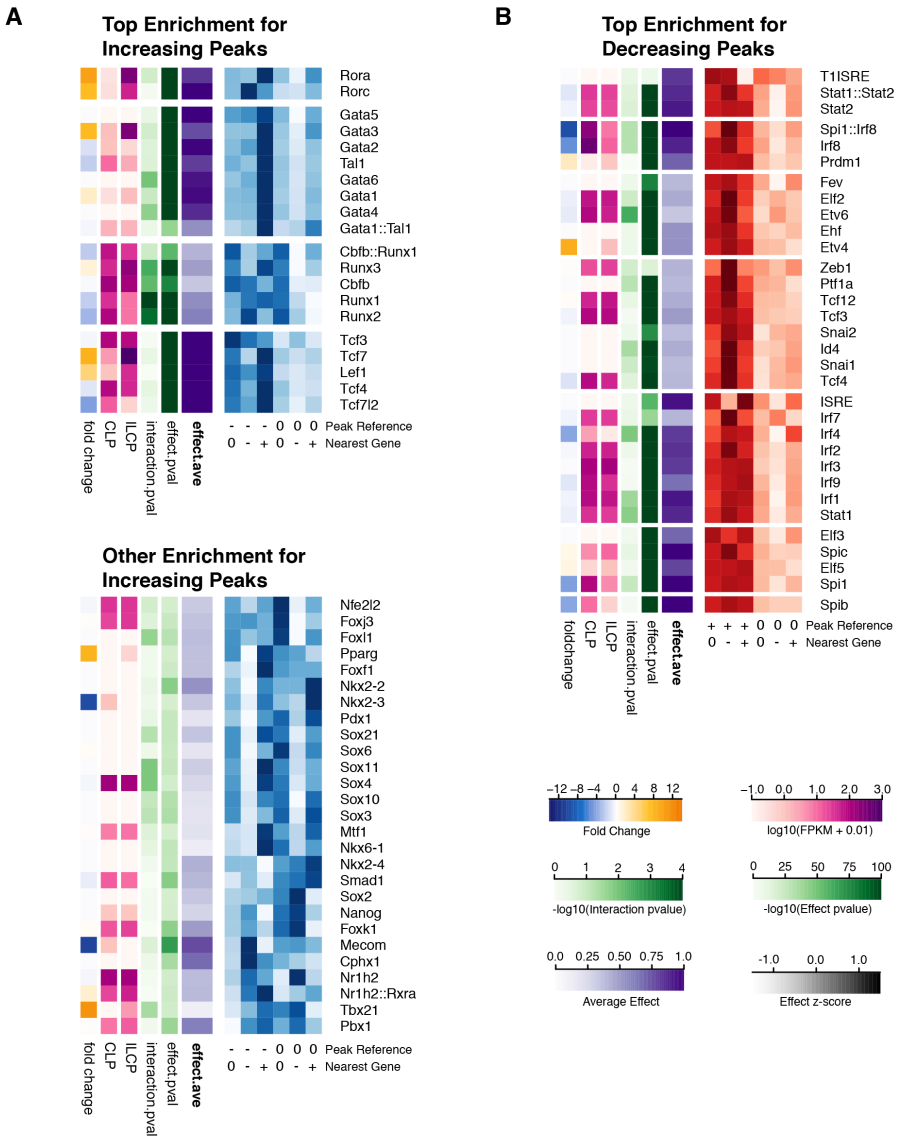


Figure 4.5: Transcription factor motifs enrich for changing chromatin.
A) Top motifs that enrich for increasing regulatory elements. B) Top motifs that enrich for decreasing regulatory elements.

bers with significance levels far beyond all others surveyed. Motifs affiliated with the TCF family are among the most strongly enriching with an average multinomial model coefficient of 1.5, implying that each TCF motif enhances the odds of belonging to the set of increasing peaks by more than a factor of 4. We anticipate this strong influence on peak dynamics is driven by Tcf7, given that Tcf7 transcription is the most significantly upregulated of any transcription factor. Interestingly, both ROR family members identified, nuclear receptors RORα and RORγt, have integral roles in ILC development and are substantially upregulated at the ILCP stage. As a master regulator of both ILC3 and LTi lineages, RORγt has spotted transcription prior to differentiation in ILCP. While defective RORα specifically impacts ILC2 differentiation, RORα is widely expressed in most ILCP after induction of PLZF and in peripheral ILCs and LTi. Among GATA family members, we anticipate enrichment of increasing peaks to be driven by Gata3, master regulator of the ILC2 lineage, since Gata3 is also widely expressed in ILCP prior to differentiation and impacts multiple ILC lineages when perturbed. These ROR and GATA family results in particular highlight how lineage specific transcription factors could play early regulatory roles prior to ILC differentiation. Among RUNX family members, both Runx1 and Runx3 have transcriptional dynamics that could fit with early regulatory roles. In single-cell study, Runx1 was observed to have a transient transcriptional peak prior to PLZF induction. Recently, it has been found that Runx3 is expressed at the ILCP stage and is essential for the development of ILC1s and ILC3s. While TCF, ROR and GATA family members have relatively equal enrichment rates despite the transcriptional dynamics of the nearest gene, RUNX family members interestingly have significantly altered effect coefficients depending on these categories of transcriptional change. Runx1 in particular tends to have enhanced enrichment for increasing peaks near downregulated genes. The predominant four motif families whose presence enriches for increasing peaks recapitulate known influencers in ILC development, but point to their specific association with new regulatory elements found in ILCP.

We detected a number of motifs in addition to these relatively known factors that specif-

ically enrich for increasing regulatory elements (Figure 4.5A). Interestingly, these include Tbx21, master regulator of the ILC1 lineage, and SOX family motifs, of which Sox4 is transcriptionally induced just prior to the ILCP. We also found Nanog and Sox2, key pluripotency regulators in embryonic stem cells, as well as a number of motifs with reported roles in regulating hematopoietic stem cell differentiation, such as FOX family genes, Nfe2l2 (or Nrf2), Mecom (or Evi1), and Pbx1. As a mediator of Notch pathway signaling, Nrf2 could be of particular interest given the necessary role of Notch signaling in early ILC specification. Finally, we found Nr1h2 and Pparg, which are known to interact with retinoid receptors. Further investigation is required to determine if some of these factors have roles in ILC development or if their genomic locations are similarly regulated by other ILC specific factors.

Most of the motifs that significantly enrich for decreasing peaks are affiliated with B cell development (Figure 4.5B). Two motifs that most strongly enriched for decreasing peaks include Spi1 (or PU.1) and Spib, complementary ETS family transcription factors associated with the B cell lineage. We also found IRF family members that strongly enriched for decreasing peaks, which are known to participate in B cell fate decisions in concert with Spi1. Both Spi1 and Irf8 in particular were substantially transcriptionally downregulated, suggesting decreasing regulatory elements could be directly explained by decreased occupancy of these factors. Interestingly, we also found a motif family that includes Tcf3 (or E2A) that enriches for decreasing peaks. While Tcf3 was not substantially transcriptionally downregulated, Tcf3 is expressed in both CLP and ILCP and is known for mutual antagonism with (the strongly upregulated) ILC lineage-specific factor Id2. While Ebf1 motifs trended towards enrichment for against increasing peaks, the effect sizes were not substantial enough to be strongly significant (not shown). This is somewhat surprising given the relatively strong transcriptional downregulation of Ebf1 and its established role in early B cell development, and highlights stronger regulatory influence of earlier lymphoid factors at the CLP stage.

4.4 DISCUSSION

In this study, we directly compare transcriptomic and epigenomic information to better describe the global regulatory changes that occur during ILC specification. Our genome-wide transcriptional characterization of the transition between CLP and ILCP highlights predominant transcriptional changes that define the ILCP developmental stage, including ILC specific transcriptional regulators and surface markers. By surveying the ILCP, we present a comprehensive transcriptional profile of the stage committed to the ILC lineages, but just prior to extensive lineage differentiation.

Given the intricate set of successive transcriptional changes that occurs throughout early ILC specification, this survey of changes occurring at ILCP stage reveals useful markers that can tease out finer stages of development. For instance, one of the most highly upregulated surface markers in ILCP that we identified, Pdcd1, has been shown to be an effective proxy for PLZF expression in ILC progenitor cells, allowing for the isolation of the ILCP without specific reporter mice [139, 12]. We anticipate that some of the uncharacterized transcription factors and surface markers that we found associated with the ILCP could help distinguish ILC progenitor stages or be associated with branching of specific ILC lineages.

The low cell number requirements for ATAC-seq enabled the survey of chromatin accessibility for the rare ILC precursor population. While the ILCP has a more restricted developmental potential than the CLP, we generally observed more opening regulatory enhancers with larger fold-changes than closing enhancers at the ILCP stage. Consistent with other studies, we found that the most significantly changing regulatory enhancers tend to be distal rather than proximal to gene transcription start sites. More broadly across hematopoiesis, distal enhancers are a more reliable predictor of cell identity than proximal enhancers or even transcriptional profiles [142]. In the comparison of chromatin accessibility between differentiated ILC effectors, common regulatory enhancers tend to be found in promoter regions while variable regulatory enhancers that distinguish different ILC lineages are found in distal genomic locations [133]. We also observed that changes in chromatin accessibility surround-

ing gene TSS tend to be more prevalent than corresponding changes in gene transcription. Further study is required to determine if these widespread chromatin accessibility changes can be assigned functional roles in differentiated ILCs or other lineages.

By examining the distribution of transcription factor motifs among regulatory elements, we were able to learn more specific information about transcriptional regulation occurring during ILC specification. We demonstrate through statistical modeling that the presence of motif sequences in regulatory regions is highly predictive of changes in chromatin accessibility in those regions. Moreover, the models demonstrate the direct relationship between motif count and accessibility. This is compelling since we can identify these strong associations between motifs and chromatin regulation without needing to restrict our evaluation to true binding events, as could be measured by CHIP-seq for instance. This affords us the ability to scan and evaluate the influence of hundreds of motifs systematically, which would be prohibitively expensive to perform experimentally. Granted, we are precluded from evaluating motifs without well-established motif consensus sequences. Still, from this large set of motif comparisons, we identified the most influential set of motifs that should be prioritized for further investigation.

In specifying a logistic regression model for examining the association between motif presence and chromatin accessibility, we uncovered a set of motifs that enrich for unchanging regulatory elements that we did not anticipate. By considering unchanging ATAC-seq peaks as a distinct group from increasing or decreasing peaks, we could isolate a large number of motifs with roles that are likely separate from lineage specification. While it is difficult to provide a collective interpretation for this set of motifs, it notably includes the architectural protein CTCF, cell cycle regulators such as E2A, and generic transcriptional promoters of development and proliferation. We also noticed that a number of these motifs bind GC-rich sequences, which leads us to speculate that they might be associated with CpG islands found in a large portion (40%) of mammalian promoter regions. Given that the CLP and ILCP populations that we analyzed are both progenitor populations, it is not clear if the large

number of developmental factors found in this group is specific to progenitor cells or could be as extensively observed in differentiated cells as well. The different types of motifs found in this group illustrate the diverse functional categories of regulatory enhancers that are simultaneously surveyed in chromatin profiling studies.

Our statistical modeling also highlighted the predominant enrichment for opening chromatin by the TCF, ROR, GATA and RUNX motif families. Of all TCF family members, Tcf7 is highly expressed at the ILCP stage and required for development of all known ILC subsets [41, 43, 44]. Similarly, Gata3 of the GATA family is highly expressed at the ILCP stage and required for development of all ILC subsets and LTi [9]. Multiple members of the RUNX family could be acting at the ILCP stage. Runx3 expression is particularly affiliated with differentiated ILC1 and ILC3 lineages but is also present in ILCP. From our single-cell studies, Runx1 appears to have peak expression just prior to the upregulation of PLZF and could be acting earlier in ILC development. We also note that we observed upregulation of Zbtb7b (or Th-POK), an important T cell differentiation factor that also belongs to the RUNX family. Two members of the ROR family, RORα and RORγt, have necessary roles in promoting ILC2 and ILC3 development, respectively. While all of these highlighted motif families contain factors known to influence ILC development, our findings point to their specific association with remodeled chromatin in the ILCP stage. Given the increased transcription of regulatory factors belonging to each of these four motif families, the most direct interpretation for their enrichment of opening chromatin in ILCP is increased occupancy in these new regulatory regions.

Besides just indicating which transcription factors are associated with remodeled chromatin, our statistical modeling provides an initial indication of how these factors might be acting in regulatory elements. Generally, the relatively consistent influence of motif presence on peak dynamics regardless of the transcriptional dynamics of the nearest gene suggests that these regulatory factors can have either activating or inhibitory roles depending on context. For instance, Tcf7 motifs widely appear preferentially in opening chromatin near both

upregulated and downregulated genes. In fact, Tcf7 motifs tend to appear more frequently near downregulated genes (data not shown). Interestingly, another study found that perturbation of Tcf7 in early ILC progenitors results in upregulation of B cell affiliated genes, illustrating both the inhibitory role of Tcf7 and regulation of competing lineages [139].

Collectively, our study characterizes the genome-wide set of transcriptional and epigenetic features coincident with ILC specification. While uncovering both known and novel transcriptional regulators, our statistical modeling provides further insight into the potentially important roles these factors have in ILC development.

4.5 METHODS

RNA-seq data acquisition Pooled samples of CLP and ILCP with unpaired 50bp reads were sequenced with Illumina Hi-seq. Read alignment was performed with bowtie2 and transcript quantification with eXpress. Without biological replicates for measured cell types, we could not perform traditional differential expression analysis and calculate the significance of expression changes as we cannot estimate expression variance. To obtain ranking of transcriptional changes, we performed edgeR differential expression analysis with a fixed biological coefficient of variation suggested in edgeR manual. We then determined the significance ranking with these estimated significance values. We anticipate this procedure yields a more realistic prioritization than fold change ordering since it incorporates read-count information. Since the overall significance scale returned by this procedure arbitrarily depends on the chosen biological coefficient of variation, we chose a significance cutoff that yielded 201 upregulated and 164 downregulated genes. Gene set enrichment analysis was performed using the desktop java application based on the fixed ordering of transcriptional change just described using an unweighted statistic. Gene set size was limited to 300 to filter very large gene set annotations.

ATAC-seq data acquisition Three CLP and two ILCp replicate samples with paired 50bp reads were sequenced by Illumina Hi-seq. Read alignment was performed with bowtie2.

After alignment, read locations were shifted according to the tn5 transposase strand bias (+4bp for reads on the + strand, and -5bp for reads on the - strand). Peak features were called for each sample separately using MACS2 at FDR 0.01 with 5 allowed duplicates at any position. The complete peak set was subsequently filtered to only include peaks called in at least all CLP or all ILCP samples. Differential read-count analysis was performed using the edgeR option in the DiffBind R package. Fold change thresholds for demarcating increasing, decreasing, or unchanged groups of ATAC-seq peaks were determined empirically through observation of the enriched set of unchanged peaks by particular motifs. Sp4 in particular was widely found in peaks and the fold change boundaries were based on the unchanged peak subset strongly enriched by Sp4.

Motif database scanning Motif consensus sequences were aggregated from the JASPAR, Homer, Chen, Hocomoco, and Uniprobe transcription factor repositories. If possible, motifs were assigned the nearest mouse homolog for interpretation and to assign expression values. Motif sequences were scanned across all regulatory element regions as called above with FIMO from the MEME suite [141] with a p-value threshold of 1e-4. To account for redundant motif entries between and within databases, we defined an empirical grouping of motif families based on overlap. For each pair of motifs, we calculated the proportion of occurrences where the two motifs were found in overlapping locations. To define motif families, we then performed hierarchical clustering on this proportion overlap data, using a complete linkage criterion that corresponds to at least 20% overlap required among all family members.

Multinomial logistic model A multinomial logistic regression model was used to test for peak group (increasing, decreasing, unchanged) enrichment by motif presence. The following sys-

tem of equations was fit for each tested motif:

$$\log\left(\frac{P_{increasing}}{P_{decreasing}}\right) = C + C_{RNA} + C_M \cdot MOTIF + C_{M,RNA} \cdot MOTIF \tag{4.1}$$

$$\log\left(\frac{P_{increasing}}{P_{unchanged}}\right) = D + D_{RNA} + D_M \cdot MOTIF + D_{M,RNA} \cdot MOTIF \tag{4.2}$$

where P_{group} is the probability that the regulatory element belongs to the labeled peak group, RNA is the transcriptional status of nearest gene (upregulated, downregulated, unchanged), and $MOTIF$ is the number of motif sequences found in the regulatory element. For motifs in which some multinomial categories did not have a sufficient number of peaks, a set of peaks were pre-populated to avoid instabilities that derive from fitting 0 proportions in multinomial regression. Specifically, the complete set of peaks with all possible combinations of peak group, RNA, and 0 and 1 $MOTIF$ numbers were included when pre-populating. We evaluated model significance with type I ANOVA, which successively tests for significant attribution of the total sum of squares (SS) to the components $SS(RNA)$, $SS(MOTIF|RNA)$, and $SS(MOTIF * RNA|RNA, MOTIF)$. We refer to the last two corresponding p-values as the motif effect p-value and the interaction p-value, respectively. To identify the strongest enrichment group for each motif, we compare coefficients from each of three possible reference group directions and choose the one with the largest absolute coefficients. We calculate the weighted average of motif effect coefficients, with weights defined as inverse estimated variance, for the overall motif effect measure. Logistic regression models were fit using the nnet package in R.

CHAPTER 5

DISCUSSION AND FUTURE DIRECTIONS

Collectively, our studies span diverse scales of investigation into ILC transcriptional regulation. Our microarray expression profiling study of the ILC1 and NK lineages focused on defining the regulatory responsibilities of a single transcriptional regulator, PLZF. Our single-cell expression profiling study of ILCP and $\alpha_4\beta_7{}^+$ ILC precursors expanded our purview and defined the developmental progression of many transcriptional regulators while refining our placement of ILC developmental branch points. Finally, our RNA-seq and chromatin accessibility profiling study of ILCP characterized the genome-wide influence ILC transcriptional regulators.

5.1 Role of PLZF in ILC1 development

The ILC1 and NK are phenotypically similar effector cells distinguished by expression of PLZF in ILC1 development. We compared the results from two contrasts, ILC1 to NK and ILC1 to PLZF-deficient ILC1, to determine the degree to which expression differences between ILC1 and NK cells can be explained by prior expression of PLZF in ILC1 development. In this way, we used a specific experimental design to establish that a majority of PLZF-dependent genes in ILC1 including Il7rα is part of the differentially regulated program between ILC1 and NK.

Since this study represents a comparison between two contrasts, we point to the extension to multiple comparison analyses. For instance, in immune development many differentiating cell types can be influenced by single transcription factor or other perturbations. If we are interested in understanding the effect of perturbation in many contexts, we would be interested in the comparison of many related differential expression measurements. A method named CorMotif has been developed to jointly assess differential expression in many contexts simultaneously. CorMotif identifies and leverages patterns of shared differential expression

between studies [143], a strategy that has been demonstrated to increase statistical power in identifying eQTL in multiple tissues [144]. We anticipate the use of these multiple comparison analyses will be increasingly useful as technical advancement in sequencing has reduced the cell requirements for transcriptional profiling and so greatly expanded the range of accessible cell types for comparison.

5.2 Single-cell transcriptional profiling of ILCP

Through single-cell transcriptional profiling of ILC precursor populations we established a map of ILC developmental progression. This reconstruction was based on observable cell state heterogeneity among ILCP and αLP precursor populations. By clustering single-cell transcriptional profiles of 100 immune factors, we could infer the sequence of developmental stages from continuity in expression of key regulatory factors and by profile similarity of adjacent clusters. We confirmed early induction of Id2, Nfil3 and Tox occurs among αLP precursors and that the expression of PLZF and Tcf7 mark a transitional stage where LTiP branches from the ILCP. We also show that ILC lineage trifurcation occurs in the ILCP, and that ILC differentiation can occur through multi-lineage transcriptional priming.

5.2.1 Id2, Nfil3 and Tox mark the earliest stages of ILC development

Early clusters comprised almost entirely of αLP (cluster A) were marked by their expression of the early ILC developmental factors *Id2*, *Nfil3*, and *Tox*. These clusters also did not express *Zbtb16*, *Tcf7*, or ILC effector markers corroborating their developmental placement prior to the ILCP. Ordering of these clusters further implied Id2 is the earliest expressed transcription factor, and is then followed by simultaneous induction of Nfil3 and Tox. These findings contrast suggestions that Nfil3 and Tox are responsible for the induction of Id2. This notion has been evidenced by the fact that Tox or Nfil3-deficiency leads to reduction in Id2 expression and that Id2 expression can rescue ILC-related defects in these backgrounds

[117, 145, 57, 146]. Further, Nfil3 has been shown to bind the Id2 promoter in CHILP and peripheral LTi. We note that these findings are not contradictory to Id2 expression that precedes Nfil3 and Tox. To account for all observations, early low-level expression of Id2 is likely later enforced and upregulated through positive feedback with Nfil3 and Tox. In our single-cell experiments we in fact observe increased transcriptional levels of Id2 following acquisition of Nfil3 and Tox. Further, if Nfil3 and Tox were to precede Id2 induction, we would have expected to observe a distinct group of cells that expressed Nfil3 and Tox but lacked Id2 expression. These observations might also be explained by possible discrepancies in Id2 transcript and protein expression levels, or by a difference in when Id2 is initially induced and when Id2 expression is required for developmental progression. While there is evidence that Nfil3 can regulate and potentially induce Tox, it is still unclear what factors or signaling are responsible for induction of Nfil3 and initial induction of Id2.

5.2.2 Bifurcation of ILC and LTi cell branches

We identified a transitional cluster (cluster B) composed of mixed aLP and ILCP that has particular relevance to branching between the LTi and ILCP lineages. While it was understood based on the developmental potentials of known precursors that LTi branching must occur prior to ILC differentiation, it was not clear where the common precursor to the ILC and LTi lineages would be found and what factors drive their bifurcation. The transitional cluster was marked by the earliest induction of *Zbtb16* and *Tcf7*, and was further characterized by expression of early ILC developmental factors including *Id2*, *Nfil3* and *Tox*, while lacking expression of ILC differentiation markers. The partitioning of the cluster by *Zbtb16* expression provided the indication that cells in this stage are poised to differentiate into LTiP, since *Zbtb16* negative cluster B cells almost uniformly expressed *Rorc* and were uniformly *Gata3* negative. These *Zbtb16* negative cluster B cells also did not express *Rorα* and *Cxcr5*, which are both expressed in more mature LTiP. The identity of this transitional cluster as the branch point of ILC and LTi is further corroborated by our single-cell culture

experiments, which demonstrated that the late Flt3$^-$ αLP stage is the last precursor that gives rise to both ILC and LTi.

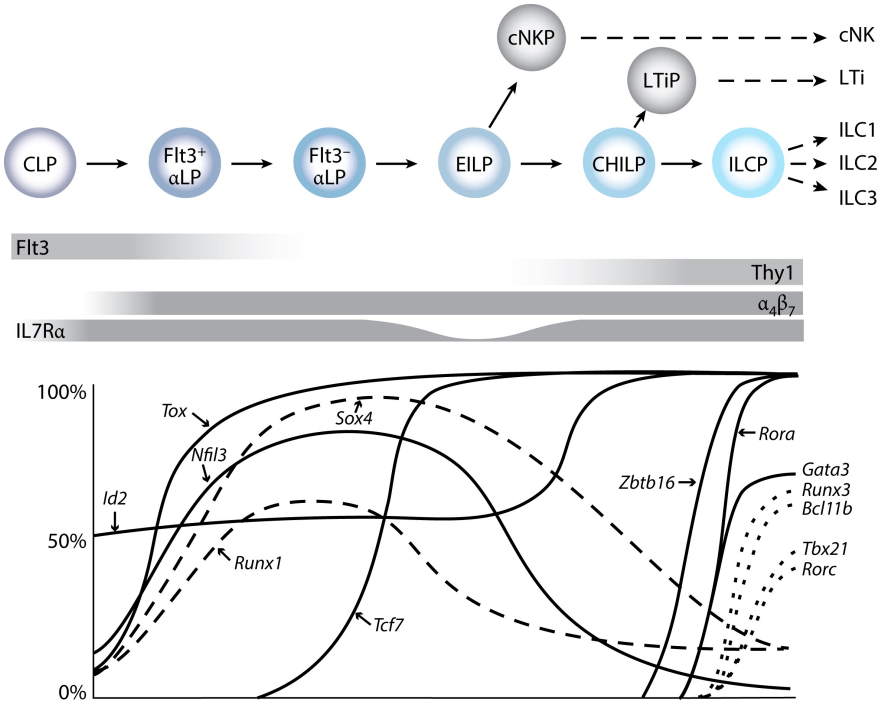
The expression profile of this transition cluster leads to the question of where EILP and CHILP precursors fit into this developmental progression. Since both *Zbtb16* expressing and *Zbtb16* negative cluster B cells express *Tcf7* and cluster B represents the earliest induction of *Tcf7*, the EILP would appear to be developmentally between the αLP clusters and cluster B. The difficulty in placing the EILP derives from its lack of IL7rα expression, which is present on all known ILC precursors including the CLP. This suggests that an early ILC precursor with ILC, cNK and LTi potential transiently downregulates IL7rα to give rise to the EILP. Transient downregulation of IL7rα has precedence in both early T cell precursors as well as DP thymocytes in T cell development [147, 148, 149]. Given that the original description of the CHILP is likely contaminated by ILCP and LTiP in the fetal liver, the question of where a true CHILP precursor to the ILC and LTi lineages would fit remains as well. We would expect this precursor to express high levels of Id2 and Tcf7 while not expressing PLZF and RORγt. If these cells exist among Il7rα^+ αLP, it is surprising we did not observe them since we observe LTi primed late αLP, although this could have been due to limited sampling in that developmental stage. Our clustering analysis and further investigations into dimensionality reduced representation of the single-cell transcriptional profiles suggest that late aLP cluster cells relatively contiguous with transitional cluster B cells. If the EILP and CHILP progress linearly between these stages, the fact that we do not observe a discontinuity between these clusters implies that we are not measuring enough alternate developmental factors that would corroborate the transition. An alternative possibility is that the downregulation of Il7rα and generation of the EILP and CHILP represent an alternative developmental pathway from the late aLP. The validity of these hypotheses need to be evaluated either through more extensive single-cell profiling that would incorporate an Il7rα^- EILP or through combined examination of Tcf7 and Id2 reporter strains.

We can now construct a consolidated linear developmental scheme (Figure 5.1) that

merges transcriptional transitions determined through single-cell transcriptional profiling as well as those inferred from characteristics of other ILC precursors. In this model, we place the EILP immediately after the Flt3$^-$ αLP following downregulation of Il7rα and upregulation of Tcf7. We then presume a bona fide CHILP precursor with high Id2 expression before upregulation of PLZF that gives rise to both LTiP and ILCP progenitors. From our single-cell transcriptional profiling, this specific placement of these precursor stages appears required for consistency with a linear developmental scheme. It will be important to corroborate this model.

5.2.3 Lineage trifurcation and multilineage transcriptional priming

Single-cell transcriptional profiling also established that initial ILC differentiation occurs in the ILCP. In fact, around half of ILCP express lineage-associated markers and many express markers associated with distinct lineages simultaneously. These multilineage transcriptionally primed precursors coexpressed *Tbx21*, *Gata3*, and *Rorc*, in addition to other differentiation markers including *Il2rβ* (ILC1), *Bcl11b* and *Icos* (ILC2), and *Lta*/*Ltb* and *Cxcr5* in mixed patterns. Similar coexpression of T-bet, Gata3, and RORγt was observed in the Arg1$^+$ fetal intestinal ILC precursor [119]. Multilineage transcriptional priming during ILC development strongly contrasts direct polarization of T helper cell subsets by exogenous cytokine signaling during immune response [120]. While these results demonstrate that ILC differentiation occurs through multilineage transcriptional priming, it does not appear to be a required developmental state. The fraction of multilineage primed precursors are dispersed throughout the dimensionality reduce ILC developmental trajectory. If all ILC precursors pass through a multilineage primed state, we would expect these cells to be tightly grouped and form a distinct state. This suggests that multilineage transcriptional priming in ILCP is stochastic process, either induced transcriptionally or through exogenous signals. Interestingly, the mixed patterns of multilineage priming imply that all combinations of lineages need to be resolved into individual fates.



Ishizuka IE, et al. 2016.
Annu. Rev. Immunol. 34:299–316

Figure 5.1: Revised model of ILC developmental stages.
(Top) Model depicting the stages of development from common lymphoid progenitor (CLP) to early (Flt3$^+$) and late (Flt3$^-$) $\alpha_4\beta_7$–expressing lymphoid precursor (αLP), early innate lymphoid progenitor (EILP), common helper innate lymphoid progenitor (CHILP), and innate lymphoid cell precursor (ILCP). The conventional natural killer cell precursor (cNKP) and the lymphoid tissue inducer precursor (LTiP) are shown branching off of the EILP and CHILP, respectively. (Bottom) Sequential induction of transcription factors throughout ILCP ontogeny. The graph shows the percentage of cells positive for the indicated transcription factors as detected by single-cell transcriptomic analysis or transcription factor reporter strains [64, 41].

5.2.4 Transcription factor expression profiles

Single-cell transcriptional profiling revealed precise stage specific patterns of expression for known and novel transcriptional regulators. Gata3 is promiscuously expressed at the ILCP stage and can be found at intermediate to high levels in roughly 70% of ILCP by staining [29]. By multiplex qPCR we observed *Gata3* transcript expression in about half of ILCP, but in diverse transcriptionally primed states, further corroborating that Gata3 is not strictly associated with the ILC2 fate in development. Similarly, we observed *Rorα* expression in almost all ILCs following induction of PLZF as well as in LTiP. While Gata3 deficiency results in defects in all ILCs as well as LTi, Rora deficiency only affects the generation of ILC2s [13]. We also observed close association of *Bcl11b* and *Rxrg* expression with the ILC2 primed ILCPs, indicating that the two transcription factors are early markers for ILC2 commitment. Conditional deletion of Bcl11b has confirmed that loss of Bcl11b blocks ILC2 development at the ILCP stage [12]. Any specific role of Rxrg in ILC development is unknown. *Irf8* appeared to be preferentially associated with *Tbx21* primed ILCPs, although any role it has is also unknown. The transcription factors *Sox4* and *Runx1* were both transiently expressed and induced just prior to expression of PLZF. The peak expression of these factors is suggestive of the possibility that they may influence the bifurcation of LTi and ILCP. Known targets of Runx1 including Id2, Ets1 and Sox4 are all upregulated at the transitional cluster B stage. In T cell development, Runx1 has been shown to enhance RORγt expression and negatively regulate Gata3, implying a capability to influence LTi [150]. Similarly, Sox4 is also capable of upregulating RORγt and repressing Gata3 [151].

In aggregate, our current understanding of ILC lineage-specific regulators is sufficient to provide an initial picture of how lineage differentiation is likely enforced (Figure 5.2). The early requirement of Notch signaling for ILC2 development draws analogy to the Notch-Tcf7-Gata3-Bcl11b cascade established in thymocyte development [152]. Although Notch signaling does not appear to be required for Tcf7 upregulation in ILC precursors, it may still play a role in enforcing expression of Tcf7 expression as well as subsequent expression

of Gata3 and Bcl11b. Downstream of Tcf7, it is apparent that transcription factors Gata3, Bcl11b and Gfi1 form a positive reinforcement loop for the ILC2 lineage. The ILC2 associated regulators Gata3 and Gfi1 have well-established roles in suppressing genes associated with Th1 and Th17 differentiation. Among transcription factors, Gfi1 directly inhibits Sox4 [153, 154] while Gata3 antagonizes Runx3 at the protein level and may indirectly inhibit expression of T-bet [155]. In the inverse direction, Runx3 antagonizes Gata3 at the protein level and T-bet can suppress Gata3 expression. It is apparent that these lineage-enforcing mechanisms are only representative of cross-inhibition between ILC2 and ILC1/3 programs, and further investigation is required to establish whether ILC1 and ILC3 lineages also enforce their bifurcation.

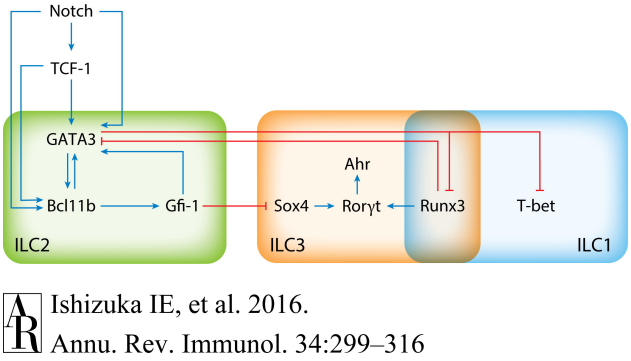


Figure 5.2: ILC lineage transcription factor network.
Transcription factor networks regulating cytokine effector programs in innate lymphoid cell (ILC) lineages. Blue arrows depict positive interactions, whereas red lines depict inhibitory interactions at the gene or protein level.

5.2.5 Requirements for ILC differentiation and development

It is informative to view the observed characteristics of ILC differentiation through the lens of ILC developmental requirements. As the roles of ILCs are further defined, it will become important to understand what specific properties of the collective innate immune system need to be maintained for effective immunity. Besides functionality in individual ILC lineages, it has been demonstrated that modulation of relative cell numbers of ILC2 and ILC3

lineages is important for type-specific pathogen clearance in the periphery. Accordingly, we might expect that it is important to achieve full representation of required peripheral ILC types in acceptable proportions. If particular ratios of ILC cell types require regulation, the natural extension is to consider whether this imposes constraints on ILC development. It is apparent that dietary ligands such as AhR ligands and retinoic acid can modulate the sizes of peripheral pools of ILCs in the intestine, although it is unclear if these peripheral expansions require contributions from hematopoietic-compartment ILC progenitors. Parabiotic mice studies have demonstrated that there is minimal replacement of tissue-resident ILCs by circulating hematopoietic cells over the timescale of weeks, although inflammatory conditions promotes some recruitment of presumably newly generated bone-marrow derived ILC2 cells [156]. This suggests that the peripheral balance of ILC types can be primarily maintained through expansion and contraction of differentiated ILC populations, which would imply that developmental output from hematopoietic ILC progenitors does not need to be strictly regulated. Instead, if hematopoietic ILC progenitors are mostly required for re-population of specific lineages during inflammatory stress, modulation of ILC developmental output by external signals might be a stronger developmental requirement than proportional generation of ILC types. A more general demonstration of compensation among components of the innate immune system can be seen in a study of patients with severe combined immunodeficiency. Although stem cell transplants reconstituted B, T, cNK and ILC1 cells but not ILC2 and ILC3 cells, these patients did not contract common herpes virus at higher rates [157].

Given the slow turnover of peripheral tissue-resident ILCs, perhaps the most crucial phase of ILC development is the seeding of peripheral tissues from fetal ILC progenitors. The importance of the initial embryonic waves of ILCs and related lineages is best exemplified by the strict requirement for the LTi lineage in embryonic development. Inadequate generation of LTi precursors associated with maternal retinoid deficiency ultimately results in insufficient production of secondary lymphoid structures and permanently impaired immune

fitness [69]. Interestingly, this retinoic acid dependent generation of fetal LTi precursors is accomplished through expression of RORγt, which is almost entirely absent in adult ILC precursors. The existence of distinct developmental hierarchies in fetal and adult hematopoiesis has precedence in upstream differentiation of the HSC, where potential for erythroid and megakaryocytic lineage branching occurs throughout the early cellular hierarchy in the fetus but is restricted to the stem cell compartment in the adult [158]. In the fetus then, it is important to ensure that regulation of the developmental choice between the LTi and ILC lineages results in sufficient numbers of LTi. Further investigation is required to determine whether this developmental decision is actively monitored through feedback via retinoic acid or other molecules, or if adequate LTi cell numbers are achieved passively in development. Among other ILC lineages however, there are not strong reasons to believe that their differentiation is strictly regulated aside from the requirement that all ILC types are generated in the fetus for seeding peripheral cells.

This interpretation of ILC developmental requirements can shape our understanding of ILC developmental stages. Generally speaking, in transit from a multipotent stem cell to a differentiated effector cell, we should expect that intermediate developmental stages are only strictly required if sequential induction of lineage-specific factors is necessary or if input signaling is needed to make an appropriate developmental branching decision. In ILC development, transit through the early sequence of Id2, Nfil3, and Tox induction prior to ILC differentiation might be necessary for proper divergence from the adaptive lymphoid lineages and upregulation of Tcf7 and PLZF. Given that development of LTi is responsive to retinoic acid signaling and required in embryonic development but dispensable in adult, we would anticipate there is a necessary developmental stage that processes the decision between LTi and ILC lineages. However, if the downstream differentiation of ILC1, ILC2 and ILC3 lineages does not require strict regulation as suggested above, then the differentiation process could be resolved through a variety of developmental paths rather than through a rigid developmental sequence.

In fact, a few characteristics of multilineage transcriptional priming observed in ILCP support the notion that ILC differentiation may be resolved non-deterministically. First, we observed all combinations of ILC lineage master regulators (*Tbx21*, *Gata3*, and *Rorc*) among multilineage transcriptionally primed cells in ILCP. This includes tri-lineage primed cells as well as all possible combinations of bi-lineage primed cells, which implies that all possible ILC lineage bifurcations are resolved during differentiation. Second, dimensionality reduction based visualization of the ILC developmental path revealed that all types of multilineage primed precursors are interspersed throughout ILC differentiation rather than tightly grouped, as might be expected if a multilineage primed state was a required developmental stage. Moreover, multilineage priming appeared to be possible even in cells ostensibly committed to a particular lineage, as a few ILC2 primed cells that express *Gata3*, *Bcl11b*, *Icos*, *Il13*, and *Il1rl1* could still be found with *Tbx21* or *Rorc* expression. This observation is more consistent with sporadic generation of multiple lineage regulators rather than a developmental model where multilineage regulator competition must be first resolved before further differentiation.

Stochastic fluctuations of lineage regulators has precedence in both early HSC differentiation and in embryonic stem cells [159, 160], and multiple developmental paths to particular lineages has also been observed in early HSC differentiation [159, 158, 122]. We will note that the suggestion that the resolution of three ILC lineages occurs through stochastic competition between lineage regulators is uniquely representative of a lineage trifurcation, where restriction of developmental potential is typically thought of as successive bifurcations. Among ILC lineages, the spectrum of functional plasticity between the ILC1 and ILC3 lineages as well as the observed cross-inhibition between ILC2 and ILC1/3 factors might be indicative that resolution of the bifurcation between the ILC2 and ILC1/3 lineages is most important.

The widespread expression of lineage-defining transcription factors Gata3 and RORγt in early ILCP further speaks to the unconstrained nature of ILC differentiation. Both of these transcriptional regulators are expressed immediately following acquisition of PLZF and are

observed simultaneously with the expression of genes associated with all lineages. Despite posited roles of these factors in inhibiting competing lineages, it seems likely that these cross-inhibitory mechanisms might be effective only in later stages of differentiation given the degree of multilineage transcriptional priming observed. Modulation of the strength of these inhibitory interactions could be realized through a few mechanisms. Cross-inhibition could depend on the activity of other mediating proteins, for example indirect inhibition of RORγt by Gata3 could be primarily mediated through inhibition of Runx3. Modulation of the epigenetic enhancer landscape could also adjust the degree to which transcription factors influence their downstream targets. As an example, the regulatory enhancer landscape surrounding ILC2 effector genes Il4, Il5 and Il13 shifts through progression from ILCP through ILC2P to peripheral ILC2s, presumably corresponding to regulatory enforcement of the lineage. Further investigation is needed to see if similar regulatory adjustments occur around developmental transcription factors as well.

The developmental requirements for ILC differentiation also appear less stringent in comparison to other studied developmental systems. Among the hematopoietic compartment, the development of the adaptive B and T cell lineages requires specific coordination with recombined receptor signaling. Since the specificity of recombined antigen receptors cannot be anticipated, developing B and T cells must repeatedly assess receptor gene rearrangement status and signaling competence at distinct developmental checkpoints [161]. Furthermore, the measured signaling specificity at these required developmental checkpoints determines whether development can proceed and which cell fates are accessible. Both adaptive B and T cell lineages sequentially rearrange the two component chains of their antigen receptors in a strictly specified developmental progression to test for successful gene rearrangement, guard against self-reactivity, and, for T cells, ensure sufficient specificity for antigen-presenting molecules [162, 163]. As another illustration, $\alpha\beta$T cells differentiate into helper ($CD4^+$) and cytotoxic ($CD8^+$) T cell subsets whose functions depend on recognition of two different classes of antigen presenting molecules MHCII and MHCI, respectively. Accordingly,

the CD4-CD8 lineage decision must match receptor specificity to MHCII or MHCI, and is accomplished through a specific progression from a $CD4^+$ $CD8^+$ double-positive cell state that essentially tests whether downregulation of coreceptor CD8 influences receptor signaling. Other antigen receptor specificities results in the production of other lineages, for example the expression of particular invariant chain antigen receptors with specificity to MHC class I-like Cd1d results in the differentiation of invariant NK-like T cells [164] while self-reactivity in the thymus can lead to differentiation of $CD4^-$ $CD8^-$ double-negative TCR$\alpha\beta$ IELs and regulatory T cells [165, 166]. In other well-studied non-hematopoietic developmental systems, such as zygote development and intestinal crypt formation in mammals as well as retinal neuron specification and embryonic stripe formation in Drosophila, cellular differentiation is strongly constrained by spatial relationships that must to be maintained among differentiated cells, granted these systems are typically studied for their precise developmental regulation. In these systems, imprecise lineage specification of adjacent cells can lead to functional defects, which is typically not of concern among mobile hematopoietic cells.

5.2.6 Considerations in single-cell transcriptional profiling analysis

Overall, our analysis of single-cell transcriptional profiling of ILC precursors is demonstrative of the many advantages single-cell studies possess compared to population-level transcriptional studies. From a heterogeneous ILC precursor population, we isolated a previously unaccounted for mast-like contaminating population. With the remaining single-cell profiles, we were able to construct a developmental progression of transcriptional states and identify the branch point of LTi and ILCp as well as the stage of ILC lineage trifurcation. Our developmental progression uncovered transient expression dynamics of important transcription factors including Nfil3, Runx1 and Sox4, all of which are missed in population-level transcriptional profiling of CLP and ILCP. Single-cell transcriptional profiles at the stage of lineage trifurcation revealed that individual ILC precursors undergo multilineage transcriptional priming, which could not be inferred from population-level profiling. Further

representation of our developmental trajectory using a dimensionality reduction based approach revealed the limitations of a clustering based strategy. In our clustering analysis, individual ILC lineage branches were not resolved but rather clustered together according to their degree of maturity. In our dimensionality reduced representation, developmental maturation as well as individual ILC differentiation branches could be resolved simultaneously. The single contiguous picture of developmental stages that a single-cell study produces is very informative for the interpretation of precursor relationships, since it imposes restrictions on where precursor snapshots can potentially reside.

Our observations in producing developmental pathway representations of single-cell transcriptional profiling data point towards several ways these analyses might be improved or extended. A relatively new technical challenge in the analysis of single-cell data is the prevalence of expression dropout events, or transcript levels below the limit of detection. In our single-cell multiplex qPCR data we can clearly observe measurement artifacts in the low expression regime, which indicate that complete dropout events are also likely. Appropriate accounting for dropout noise will likely need to be specific to the single-cell technology used as different measurement methodologies have different transcript sampling characteristics. In the case of technologies like single-cell multiplex qPCR and mass cytometry with a targeted set of genes assayed, we imagine noise models will need to be assay and probe specific. Expression dropout events can be accounted for with a zero-inflated expression model, where the distribution of expression values is assumed to be a explicit mixture of zero values for dropout events and of a second continuous distribution for measurable expression. For single-cell qPCR data, there has been some work that suggests treating expression dropout explicitly with a zero-inflated expression model increases statistical power in differential expression analysis [167, 79]. For single-cell RNA-seq data, a few studies have developed mathematical models to explicitly deconvolve the relative contributions of technical and biological variability in expression measurements [168, 169]. A recent intriguing suggestion for single-cell RNA-seq is to simultaneously perform clustering, normalization and dropout

imputation, essentially contextualizing noise model to specific cell subpopulations [170]. Importantly, explicit noise modeling in these datasets will be necessary to establish the degree to which measured heterogeneity reflects true biological variability. By extension, this detailed expression modeling might also need to be considered in interpretation, as they could indicate that particular developmental transitions are unresolvable if dominated by technical noise.

In the study of continuous developmental transitions, there have been recent shifts in how to represent and interpret these processes from single-cell expression data. Dimensionality reduction techniques for single-cell expression data typically produce a visualization of developmental paths in two dimensions by default. We note that in our ILC differentiation data, there is evidence of a potentially non-deterministic lineage trifurcation. The three differentiation possibilities with expression profile representation among all possible bilineage combinations would require at least three dimensions for representation of all developmental paths, although we expect the two dimensional projection highlights the predominant lineage relationships. Developmental paths are also typically represented linearly, where sampled cells are assigned a single pseudotime metric that infers their placement earlier or later in differentiation. The sporadic representation of multilineage ILC states would suggest that higher-dimensional structures might be more representative of developmental transitions rather than linear paths, particularly at differentiation decisions. In other words, particularly around loosely regulated differentiation decisions, it may be more accurate to consider the set of potential developmental trajectories that result in lineage specification rather than the average linear trajectory. In fact, within the remaining variability surrounding linear developmental paths, it would not be surprising to find possibly parallel developmental trajectories. More explicit physical modeling of the molecular components on developmental paths could best inform what trajectories are possible and has been demonstrated to some degree in dynamical systems based models of lineage bifurcations and transitions with single-cell data [171, 172]. These dynamical systems models can be further informed by

incorporating restrictions on the true timescale of expression dynamics for different genes as well as by incorporating information that establishes the developmental potential observed in different regions of a trajectory.

In current developmental trajectory analyses, lineage branches are assumed to diverge rather than reconnect at downstream points. However, the discussion above suggests that developmental trajectory analyses may need to accommodate divergent paths to common differentiated states, or developmental cycles. In our ILC data, we observe distinct lineage branching for the phenotypically similar LTi and ILC3 lineages. Furthermore, the anticipated expression profiles of EILP and CHILP entertain the possibility of alternative developmental paths prior to generation of the ILCP. Alternative developmental paths also have precedence in early HSC differentiation, as discussed above. It should not be difficult to extend branching analyses to handle lineage convergence, as branching detection methods could be simply applied along the reverse developmental trajectory of a differentiated lineage that converged.

In the future, single-cell trajectory analysis should focus on enhancing the analysis of genes that govern cell state transitions. The first step towards this goal is the identification of differentially expressed genes along continuous developmental trajectories. This differential expression problem is fundamentally different from the typical design in bulk microarray or RNA-seq experiments, where discrete pairs of conditions are typically compared. Differential expression along developmental trajectories requires statistics that reveal the association of a genes expression with a continuous developmental pseudotime-like metric. We also anticipate there will increasingly be experimental designs in which developmental trajectories are compared following perturbation of a gene or change in condition. The comparison of continuous trajectories is very different from bulk differential expression since we are interested in understanding how lineage paths are shifted or truncated, as observed following perturbation of Bc11b in ILC precursors [12]. For these comparisons, we would imagine it might be necessary to pair equivalent regions on developmental paths to contextualize any differential expression that may be present. To determine how genes govern cell state

transitions, we anticipate there will be substantial computational focus on regulatory network reconstruction based on the ample cell state transition data provided by single-cell trajectory reconstructions. The accessibility of bulk expression profiling data initially led to substantial interest in regulatory network reconstruction, although these efforts were ultimately stymied by insufficient samples for inference across thousands of genes as well as the substantial variability lost in bulk averaging. Accordingly, the large datasets being generated for high-dimensional expression states in single cells makes single-cell expression profiling studies an attractive prospect to revisit the network reconstruction problem. A few studies already demonstrate initial efforts into network reconstruction strategies, including a Boolean switch-based network model [173] and Bayesian network expression model [174]. We anticipate these strategies will be further refined on larger single-cell expression datasets in the future.

5.3 Chromatin accessibility profiling of ILCP

We directly compared transcriptomic and epigenomic information to better describe the global regulatory changes that occur during ILC specification. We present a genome-wide characterization of ILCP transcription that reveal upregulation of several transcription factors with uncharacterized roles in ILC development, including Ikzf2, Ikzf3, Rxrg, Maf, Lmo4, Tox2 and Zbtb7b. Consistent with other studies, we found that the most significantly changing regulatory enhancers tend to be distal rather than proximal. Statistical modeling demonstrated that the presence of motif sequences in regulatory regions is highly predictive of changes in chromatin accessibility in those regions. Our modeling revealed a set of motifs that particularly enrich for unchanged ATAC-seq peaks, highlighting the diverse set of chromatin accessibility features assayed in these experiments. The predominant motif families that enrich for opening chromatin were the TCF, ROR, GATA, and RUNX motif families which all contain well-substantiated developmental factors with roles in ILC development. Our single-cell transcriptional profiling study particularly highlight the widespread expression of Rora and Gata3 in ILCP as well as the transient induction of Runx1 prior to PLZF

expression, substantiating roles for these motifs in opening regulatory enhancers in ILCP.

Further comparison of ILCP regulatory enhancers to those found in other cell types could greatly enhance our understanding of the regulatory requirements for different ILC lineages. As we have compared expression and necessity of transcriptional regulators between the ILC and T cell lineages, the comparison of regulatory enhancers at the ILCP stage to those of T cell precursors could distinguish the specific regulatory mechanisms that are shared and unique to the lineages. Similarly, since the ILCP is restricted to ILC lineages not including NK and LTi, the comparison of ILCP regulatory enhancers to early NK precursor and LTiP regulatory enhancers should yield valuable insight into the particular regulatory mechanisms responsible for lineage branching. Finally, the comparison of ILCP regulatory enhancers to those of other ILC precursors could distinguish whether a linear precursor progression exists. For instance, we observe opening of chromatin near IL-7Rα in ILCP. If EILP derives from aLP and progresses to ILCP while transiently downregulating IL-7Rα, we would expect the chromatin accessibility near Il-7Rα in EILP to remain open. Identifying lineage-specific regulatory enhancers can be extremely useful for the generation of novel experimental models. Instead of having to rely on perturbation of required developmental factors with lineage-specific expression, the perturbation of a lineage-specific regulatory enhancer can also acutely target a lineage-specific regulatory mechanism and perturb a specific developmental branch.

While our presented motif analysis only evaluated motifs independently, regulatory enhancers often contain tens of motifs at a time. Future work examining motif interactions could provide additional mechanistic insights as there could be core regulatory units re-used in multiple enhancers. Naturally, motif interactions could be more predictive of chromatin accessibility dynamics and nearby gene transcription than independent motifs. A variety of computational methods have been designed to search for commonly occurring cis-regulatory modules (CRMs) [175]. Broadly, these methods tend to utilize three strategies: sequence conservation across related species, motif-based transcription factor binding site clustering, and motif-blind sequence analysis. A primary difficulty facing recognition of CRMs lies in the

variation that could be present in transcription factor binding site arrangement, composition and relative spacing. Moreover, as has been demonstrated by the outperformance of motif-based approaches by motif-blind approaches, inference of true transcription factor binding based on consensus motif sequences can be unreliable. Still, the genome-wide catalog of transcription factor motif presence in ILC precursor regulatory regions we produced is a high dimensional dataset that might have a reduced representation as functional transcription factor units. Since many of the relevant ILC transcription factors have well-characterized consensus motif sequences, we expect that use of them in cis-regulatory module prediction procedures should identify informative transcription factor associations.

As introduced in the discussion of additional chromatin accessibility comparisons, the integrated analysis of chromatin accessibility in other cell stages will likely be necessary to distinguish lineage-specific regulatory enhancers. In our analysis, we simply present the association of motif presence with peak change in a single comparison. Future analyses would have to account for the much more complicated regulatory peak dynamics that we would expect across hierarchical cell stages. Instead of assigning peaks to three categories for a single comparison, we anticipate it will be necessary to perform more general clustering of peak dynamics across many samples to discern the predominant patterns of chromatin accessibility. Similarly, we anticipate it will also be necessary to perform clustering of the corresponding genome-wide expression profiling for measured cell stages for interpretation. Further investigation will be necessary to determine it will be more effective to match motif presence to chromatin accessibility and gene expression clusters, or if it might be more informative to successively compare changes in chromatin accessibility and gene expression between sequential developmental stages.

In our chromatin accessibility analysis, we demonstrate that motif presence is highly predictive of regulatory enhancer peak dynamics and identify a subset of transcription factors that are likely responsible for ILC regulatory changes. Ultimately, we are interested in producing a model that can integrate motif presence and peak dynamics information to

predict gene expression. Such a model would allow us to infer relevant regulatory interactions between developmental transcription factors and downstream targets. Two outstanding challenges in determination of such a predictive model are affiliating enhancers with specific regulatory functions and integrating the activities of many regulatory enhancers surrounding target genes. Given that diverse collections of transcription factor binding sites comprise cis-regulatory modules, they are capable of performing intricate integration of regulatory inputs that results in complex regulatory functions. A simplistic categorization of enhancer regulatory functions includes transcriptional activation, transcriptional inhibition, and potentially silencing of nearby enhancer activity. A surprising observation from our chromatin accessibility analysis was the generally indiscriminant presence of influential transcription factor motifs with regard to nearby gene transcription. For example, Tcf7 motifs were essentially as predictive of increasing regulatory peaks near upregulated genes as near downregulated genes. This either suggests that these transcription factors are capable of playing opposite regulatory roles in different contexts or that other enhancers must be suppressing the activity of other potentially contradictory regulatory elements. The direct comparison of enhancers where the same motif is found in activating or inhibiting contexts would most likely be informative in understanding if the activity is contextualized by other transcription factors. Due to the established relevance of long-range (>100kb) enhancer activity, assigning regulatory enhancers to target genes is a difficult problem. It has been demonstrated that enriched patterns of gene expression adjacent to TF binding sites can identify the regulatory action and targets of particular transcription factors [176]. In our analysis, we use the common simplifying association of the nearest gene to regulatory enhancers, although large-scale study has demonstrated that a large portion of enhancers target genes that are not their immediate neighbors [177]. We anticipate the best way to circumvent this assignment challenge will be to use an external source of data describing potential DNA interactions. In particular, the integration of data from chromatin conformation capture (3C) assays can identify topologically associated domains and specific DNA looping structures that would

inform the inference of regulatory enhancer interaction with DNA promoters.

Currently, the precise integration of regulatory enhancer and transcription factor binding site information into a predictive model of gene regulation remains a difficult challenge, although a recent study has demonstrated that changes in gene expression can be predicted through regression of nearby enhancer characteristics for genes with many regulatory enhancers [178]. The widespread re-use of regulatory motifs across the genome is an indication that enhancer functions can be computationally inferred from genome-wide experiments. More generally, we might be most interested in developing a dynamical model that undergoes transcriptional transitions based on our inferred understanding of enhancer activity. The specification of a dynamical model would ultimately test whether our understanding of regulatory interactions is sufficient to drive the expected course of development and provide a platform to computationally evaluate the roles of specific transcriptional regulators and regulatory enhancers.

5.4 CONCLUSION

In summary, our studies demonstrate distinct data-driven approaches to better understand the determinants of transcriptional regulation in ILC development. Through analysis of bulk expression data, single-cell transcriptional profiles, and genome-wide chromatin accessibility, we further defined the regulatory roles of known ILC developmental factors and identified novel regulatory candidates whose future study will elaborate the drivers of ILC specification. These studies have not only advanced our understanding of ILC development but have also broadly illustrated how continued development of computational methods for the analysis of single-cell and genome-wide data will enhance our investigative capabilities in many biological contexts.

REFERENCES

[1] David Artis and Hergen Spits. The biology of innate lymphoid cells. *Nature*, 517(7534):293–301, 2015.

[2] Gérard Eberl, Marco Colonna, James P Di Santo, and Andrew NJ McKenzie. Innate lymphoid cells: A new paradigm in immunology. *Science*, 348(6237):aaa6566, 2015.

[3] Qi Yang, Steven A Saenz, Daniel A Zlotoff, David Artis, and Avinash Bhandoola. Cutting edge: Natural helper cells derive from lymphoid progenitors. *The Journal of Immunology*, 187(11):5505–5509, 2011.

[4] Hergen Spits, David Artis, Marco Colonna, Andreas Diefenbach, James P Di Santo, Gerard Eberl, Shigeo Koyasu, Richard M Locksley, Andrew NJ McKenzie, Reina E Mebius, et al. Innate lymphoid cells—a proposal for uniform nomenclature. *Nature Reviews Immunology*, 13(2):145–149, 2013.

[5] Dorothy K Sojka, Beatrice Plougastel-Douglas, Liping Yang, Melissa A Pak-Wittel, Maxim N Artyomov, Yulia Ivanova, Chao Zhong, Julie M Chase, Paul B Rothman, Jenny Yu, et al. Tissue-resident natural killer (nk) cells are cell lineages distinct from thymic and conventional splenic nk cells. *Elife*, 3:e01659, 2014.

[6] K Oeser, C Schwartz, and D Voehringer. Conditional il-4/il-13-deficient mice reveal a critical role of innate immune cells for protective immunity against gastrointestinal helminths. *Mucosal immunology*, 8(3):672–682, 2015.

[7] Jonathan R Brestoff, Brian S Kim, Steven A Saenz, Rachel R Stine, Laurel A Monticelli, Gregory F Sonnenberg, Joseph J Thome, Donna L Farber, Kabirullah Lutfy, Patrick Seale, et al. Group 2 innate lymphoid cells promote beiging of white adipose tissue and limit obesity. *Nature*, 519(7542):242–246, 2015.

[8] Min-Woo Lee, Justin I Odegaard, Lata Mukundan, Yifu Qiu, Ari B Molofsky, Jesse C Nussbaum, Karen Yun, Richard M Locksley, and Ajay Chawla. Activated type 2 innate lymphoid cells regulate beige fat biogenesis. *Cell*, 160(1):74–87, 2015.

[9] Ryoji Yagi, Chao Zhong, Daniel L Northrup, Fang Yu, Nicolas Bouladoux, Sean Spencer, Gangqing Hu, Luke Barron, Suveena Sharma, Toshinori Nakayama, et al. The transcription factor gata3 is critical for the development of all il-7rα-expressing innate lymphoid cells. *Immunity*, 40(3):378–388, 2014.

[10] Maria Elena De Obaldia and Avinash Bhandoola. Transcriptional regulation of innate and adaptive lymphocyte lineages. *Annual review of immunology*, 33:607–642, 2015.

[11] Yong Yu, Cui Wang, Simon Clare, Juexuan Wang, Song-Choon Lee, Cordelia Brandt, Shannon Burke, Liming Lu, Daqian He, Nancy A Jenkins, et al. The transcription factor bcl11b is specifically expressed in group 2 innate lymphoid cells and is essential for their development. *Journal of Experimental Medicine*, pages jem–20142318, 2015.

[12] Yong Yu, Jason CH Tsang, Cui Wang, Simon Clare, Juexuan Wang, Xi Chen, Cordelia Brandt, Leanne Kane, Lia S Campos, Liming Lu, et al. Single-cell rna-seq identifies a pd-1hi ilc progenitor and defines its development pathway. *Nature*, 539(7627):102–106, 2016.

[13] Timotheus YF Halim, Aric MacLaren, Mark T Romanish, Matthew J Gold, Kelly M McNagny, and Fumio Takei. Retinoic-acid-receptor-related orphan nuclear receptor alpha is required for natural helper cell development and allergic inflammation. *Immunity*, 37(3):463–474, 2012.

[14] See Heng Wong, Jennifer A Walker, Helen E Jolin, Lesley F Drynan, Emily Hams, Ana Camelo, Jillian L Barlow, Daniel R Neill, Veera Panova, Ute Koch, et al. Transcription factor ror [alpha] is critical for nuocyte development. *Nature immunology*, 13(3):229–236, 2012.

[15] Chauncey J Spooner, Justin Lesch, Donghong Yan, Aly A Khan, Alex Abbas, Vladimir Ramirez-Carrozzi, Meijuan Zhou, Robert Soriano, Jeffrey Eastham-Anderson, Lauri Diehl, et al. Specification of type 2 innate lymphocytes by the transcriptional determinant gfi1. *Nature immunology*, 14(12):1229–1236, 2013.

[16] Long Li, Mark Leid, and Ellen V Rothenberg. An early t cell lineage commitment checkpoint dependent on the transcription factor bcl11b. *Science*, 329(5987):89–93, 2010.

[17] Peng Li, Shannon Burke, Juexuan Wang, Xiongfeng Chen, Mariaestela Ortiz, Song-Choon Lee, Dong Lu, Lia Campos, David Goulding, Bee Ling Ng, et al. Reprogramming of t cells to natural killer–like cells upon bcl11b deletion. *Science*, 329(5987):85–89, 2010.

[18] Masayuki Tsuji, Keiichiro Suzuki, Hiroshi Kitamura, Mikako Maruya, Kazuo Kinoshita, Ivaylo I Ivanov, Kikuji Itoh, Dan R Littman, and Sidonia Fagarasan. Requirement for lymphoid tissue-inducer cells in isolated follicle formation and t cell-independent immunoglobulin a generation in the gut. *Immunity*, 29(2):261–271, 2008.

[19] A Lügering, M Ross, M Sieker, J Heidemann, IR Williams, W Domschke, and T Kucharzik. Ccr6 identifies lymphoid tissue inducer cells within cryptopatches. *Clinical & Experimental Immunology*, 160(3):440–449, 2010.

[20] Anja Fuchs, William Vermi, Jacob S Lee, Silvia Lonardi, Susan Gilfillan, Rodney D Newberry, Marina Cella, and Marco Colonna. Intraepithelial type 1 innate lymphoid cells are a unique subset of il-12-and il-15-responsive ifn-γ-producing cells. *Immunity*, 38(4):769–781, 2013.

[21] Marina Cella, Anja Fuchs, William Vermi, Fabio Facchetti, Karel Otero, Jochen KM Lennerz, Jason M Doherty, Jason C Mills, and Marco Colonna. A human natural killer cell subset provides an innate source of il-22 for mucosal immunity. *Nature*, 457(7230):722–725, 2009.

[22] Christoph SN Klose, Elina A Kiss, Vera Schwierzeck, Karolina Ebert, Thomas Hoyler, Yannick d'Hargues, Nathalie Göppert, Andrew L Croxford, Ari Waisman, Yakup Tanriver, et al. A t-bet gradient controls the fate and function of ccr6-ror [ggr] t+ innate lymphoid cells. *Nature*, 494(7436):261–265, 2013.

[23] Lucille C Rankin, Joanna R Groom, Michaël Chopin, Marco J Herold, Jennifer A Walker, Lisa A Mielke, Andrew NJ McKenzie, Sebastian Carotta, Stephen L Nutt, and Gabrielle T Belz. The transcription factor t-bet is essential for the development of nkp46+ innate lymphocytes via the notch pathway. *Nature immunology*, 14(4):389–395, 2013.

[24] Sylvestre Chea, Thibaut Perchet, Maxime Petit, Thomas Verrier, Delphine Guy-Grand, Elena-Gaia Banchi, Christian AJ Vosshenrich, James P Di Santo, Ana Cumano, and Rachel Golub. Notch signaling in group 3 innate lymphoid cells modulates their plasticity. *Science signaling*, 9(426):ra45, 2016.

[25] Chao Zhong, Kairong Cui, Christoph Wilhelm, Gangqing Hu, Kairui Mao, Yasmine Belkaid, Keji Zhao, and Jinfang Zhu. Group 3 innate lymphoid cells continuously require the transcription factor gata-3 after commitment. *Nature immunology*, 17(2):169–178, 2016.

[26] Takashi Ebihara, Christina Song, Stacy H Ryu, Beatrice Plougastel-Douglas, Liping Yang, Ditsa Levanon, Yoram Groner, Michael D Bern, Thaddeus S Stappenbeck, Marco Colonna, et al. Runx3 specifies lineage commitment of innate lymphoid cells. *Nature immunology*, 16(11):1124–1133, 2015.

[27] Dominic Chih-Cheng Voon, Yit Teng Hor, and Yoshiaki Ito. The runx complex: reaching beyond haematopoiesis into immunity. *Immunology*, 146(4):523–536, 2015.

[28] Amélie Collins, Dan R Littman, and Ichiro Taniuchi. Runx proteins in transcription factor networks that regulate t-cell lineage choice. *Nature Reviews Immunology*, 9(2):106–115, 2009.

[29] Michael G Constantinides, Benjamin D McDonald, Philip A Verhoef, and Albert Bendelac. A committed precursor to innate lymphoid cells. *Nature*, 508(7496):397–401, 2014.

[30] Ai-Ping Mao, Michael G Constantinides, Rebecca Mathew, Zhixiang Zuo, Xiaoting Chen, Matthew T Weirauch, and Albert Bendelac. Multiple layers of transcriptional regulation by plzf in nkt-cell development. *Proceedings of the National Academy of Sciences*, 113(27):7602–7607, 2016.

[31] Rahul Roychoudhuri, Kiyoshi Hirahara, Kambiz Mousavi, David Clever, Michael Bonelli, Christopher Klebanoff, Vittorio Sartorelli, Yuka Kanno, Luca Gattinoni, Ena Wang, et al. Bach2 represses effector programmes to stabilize treg-mediated immune homeostasis-a new target in tumor immunotherapy? *Journal for immunotherapy of cancer*, 1(1):O14, 2013.

[32] Kazuyo Moro, Taketo Yamada, Masanobu Tanabe, Tsutomu Takeuchi, Tomokatsu Ikawa, Hiroshi Kawamoto, Jun-ichi Furusawa, Masashi Ohtani, Hideki Fujii, and Shigeo Koyasu. Innate production of th2 cytokines by adipose tissue-associated c-kit+ sca-1+ lymphoid cells. *Nature*, 463(7280):540–544, 2010.

[33] Yoshifumi Yokota, Ahmed Mansouri, Seiichi Mori, Seiichi Sugawara, Satoko Adachi, Shin-Ichi Nishikawa, and Peter Gruss. Development of peripheral lymphoid organs and natural killer cells depends on the helix–loop–helix inhibitor id2. *Nature*, 397(6721):702–706, 1999.

[34] JK Kim, M Takeuchi, and Y Yokota. Impairment of intestinal intraepithelial lymphocytes in id2 deficient mice. *Gut*, 53(4):480–486, 2004.

[35] Barbara L Kee. E and id proteins branch out. *Nature Reviews Immunology*, 9(3):175–184, 2009.

[36] Tomokatsu Ikawa, Hiroshi Kawamoto, Ananda W Goldrath, and Cornelis Murre. E proteins and notch signaling cooperate to promote t cell lineage specification and commitment. *Journal of Experimental Medicine*, 203(5):1329–1342, 2006.

[37] Ana Pereira de Sousa, Claire Berthault, Alessandra Granato, Sheila Dias, Cyrille Ramond, Barbara L Kee, Ana Cumano, and Paulo Vieira. Inhibitors of dna binding proteins restrict t cell potential by repressing notch1 expression in flt3-negative common lymphoid progenitors. *The Journal of Immunology*, 189(8):3822–3830, 2012.

[38] Ming Ji, Huajie Li, Hyung Chan Suh, Kimberly D Klarmann, Yoshifumi Yokota, and Jonathan R Keller. Id2 intrinsically regulates lymphoid and erythroid development via interaction with different target proteins. *Blood*, 112(4):1068–1077, 2008.

[39] Christoph SN Klose, Melanie Flach, Luisa Möhle, Leif Rogell, Thomas Hoyler, Karolina Ebert, Carola Fabiunke, Dietmar Pfeifer, Veronika Sexl, Diogo Fonseca-Pereira, et al. Differentiation of type 1 ilcs from a common progenitor to all helper-like innate lymphoid cell lineages. *Cell*, 157(2):340–356, 2014.

[40] Sandrine Schmutz, Nabil Bosco, Stephane Chappaz, Onur Boyman, Hans Acha-Orbea, Rhodri Ceredig, Antonius G Rolink, and Daniela Finke. Cutting edge: Il-7 regulates the peripheral pool of adult rorγ+ lymphoid tissue inducer cells. *The Journal of Immunology*, 183(4):2217–2221, 2009.

[41] Qi Yang, Fengyin Li, Christelle Harly, Shaojun Xing, Longyun Ye, Xuefeng Xia, Haikun Wang, Xinxin Wang, Shuyang Yu, Xinyuan Zhou, et al. Tcf-1 upregulation identifies early innate lymphoid progenitors in the bone marrow. *Nature immunology*, 16(10):1044–1050, 2015.

[42] Naoko Satoh-Takayama, Sarah Lesjean-Pottier, Paulo Vieira, Shinichiro Sawa, Gerard Eberl, Christian AJ Vosshenrich, and James P Di Santo. Il-7 and il-15 independently program the differentiation of intestinal cd3- nkp46+ cell subsets from id2-dependent precursors. *Journal of Experimental Medicine*, pages jem–20092029, 2010.

[43] Lisa A Mielke, Joanna R Groom, Lucille C Rankin, Cyril Seillet, Frederick Masson, Tracy Putoczki, and Gabrielle T Belz. Tcf-1 controls ilc2 and nkp46+ rorγt+ innate lymphocyte differentiation and protection in intestinal inflammation. *The Journal of Immunology*, 191(8):4383–4391, 2013.

[44] Qi Yang, Laurel A Monticelli, Steven A Saenz, Anthony Wei-Shine Chi, Gregory F Sonnenberg, Jiangbo Tang, Maria Elena De Obaldia, Will Bailis, Jerrod L Bryson, Kristin Toscano, et al. T cell factor 1 is required for group 2 innate lymphoid cell generation. *Immunity*, 38(4):694–704, 2013.

[45] Nidhi Malhotra, Kavitha Narayan, Ok Hyun Cho, Katelyn E Sylvia, Catherine Yin, Heather Melichar, Mehdi Rashighi, Veronique Lefebvre, John E Harris, Leslie J Berg, et al. A network of high-mobility group box transcription factors programs innate interleukin-17 production. *Immunity*, 38(4):681–693, 2013.

[46] Sjef Verbeek, David Izon, Frans Hofhuis, Els Robanus-Maandag, Hein Te Riele, Marc van de Watering, Mariette Oosterwegel, Anne Wilson, H Robson MacDonald, and Hans Clevers. An hmg-box-containing t-cell factor required for thymocyte differentiation. *Nature*, 374(6517):70–74, 1995.

[47] Brittany Nicole Weber, Anthony Wei-Shine Chi, Alejandro Chavez, Yumi Yashiro-Ohtani, Qi Yang, Olga Shestova, and Avinash Bhandoola. A critical role for tcf-1 in t-lineage specification and differentiation. *Nature*, 476(7358):63–68, 2011.

[48] Kristine Germar, Marei Dose, Tassos Konstantinou, Jiangwen Zhang, Hongfang Wang, Camille Lobry, Kelly L Arnett, Stephen C Blacklow, Iannis Aifantis, Jon C Aster, et al. T-cell factor 1 is a gatekeeper for t-cell specification in response to notch signaling. *Proceedings of the National Academy of Sciences*, 108(50):20060–20065, 2011.

[49] Farrah C Steinke, Shuyang Yu, Xinyuan Zhou, Bing He, Wenjing Yang, Bo Zhou, Hiroshi Kawamoto, Jun Zhu, Kai Tan, and Hai-Hui Xue. Tcf-1 and lef-1 act upstream of th-pok to promote the cd4+ t cell fate and interact with runx3 to silence cd4 in cd8+ t cells. *Nature immunology*, 15(7):646–656, 2014.

[50] Parinaz Aliahmad and Jonathan Kaye. Development of all cd4 t lineages requires nuclear factor tox. *Journal of Experimental Medicine*, 205(1):245–256, 2008.

[51] Corey R Seehus, Parinaz Aliahmad, Brian De La Torre, Iliyan D Iliev, Lindsay Spurka, Vincent A Funari, and Jonathan Kaye. The development of innate lymphoid cells requires tox-dependent generation of a common innate lymphoid cell progenitor. *Nature immunology*, 16(6):599–608, 2015.

[52] Thomas Hoyler, Christoph SN Klose, Abdallah Souabni, Adriana Turqueti-Neves, Dietmar Pfeifer, Emma L Rawlins, David Voehringer, Meinrad Busslinger, and Andreas Diefenbach. The transcription factor gata-3 controls cell fate and maintenance of type 2 innate lymphoid cells. *Immunity*, 37(4):634–648, 2012.

[53] Shintaro Kamizono, Gordon S Duncan, Markus G Seidel, Akira Morimoto, Koichi Hamada, Gerard Grosveld, Koichi Akashi, Evan F Lind, Jillian P Haight, Pamela S Ohashi, et al. Nfil3/e4bp4 is required for the development and maturation of nk cells in vivo. *Journal of Experimental Medicine*, 206(13):2977–2986, 2009.

[54] Duncan M Gascoyne, Elaine Long, Henrique Veiga-Fernandes, Jasper De Boer, Owen Williams, Benedict Seddon, Mark Coles, Dimitris Kioussis, and Hugh JM Brady. The basic leucine zipper transcription factor e4bp4 is essential for natural killer cell development. *Nature immunology*, 10(10):1118–1124, 2009.

[55] Cyril Seillet, Lucille C Rankin, Joanna R Groom, Lisa A Mielke, Julie Tellier, Michael Chopin, Nicholas D Huntington, Gabrielle T Belz, and Sebastian Carotta. Nfil3 is required for the development of all innate lymphoid cell subsets. *Journal of Experimental Medicine*, 211(9):1733–1740, 2014.

[56] Theresa L Geiger, Michael C Abt, Georg Gasteiger, Matthew A Firth, Margaret H O'Connor, Clair D Geary, Timothy E O'Sullivan, Marcel R van den Brink, Eric G Pamer, Alan M Hanash, et al. Nfil3 is crucial for development of innate lymphoid cells and host protection against intestinal pathogens. *Journal of Experimental Medicine*, pages jem–20140212, 2014.

[57] Wei Xu, Rita G Domingues, Diogo Fonseca-Pereira, Manuela Ferreira, Helder Ribeiro, Silvia Lopez-Lastra, Yasutaka Motomura, Lara Moreira-Santos, Franck Bihl, Veronique Braud, et al. Nfil3 orchestrates the emergence of common helper innate lymphoid cell precursors. *Cell reports*, 10(12):2043–2054, 2015.

[58] Victoria Male, Ilaria Nisoli, Duncan M Gascoyne, and Hugh JM Brady. E4bp4: an unexpected player in the immune response. *Trends in immunology*, 33(2):98–102, 2012.

[59] Xiaofei Yu, Yuhao Wang, Mi Deng, Yun Li, Kelly A Ruhn, Cheng Cheng Zhang, and Lora V Hooper. The basic leucine zipper transcription factor nfil3 directs the development of a common innate lymphoid cell precursor. *Elife*, 3:e04406, 2014.

[60] Laurel A Monticelli, Michael D Buck, Anne-Laure Flamar, Steven A Saenz, Elia D Tait Wojno, Naomi A Yudanin, Lisa C Osborne, Matthew R Hepworth, Sara V Tran, Hans-Reimer Rodewald, et al. Arginase 1 is an innate lymphoid-cell-intrinsic metabolic checkpoint controlling type 2 inflammation. *Nature immunology*, 17(6):656–665, 2016.

[61] Arivazhagan Sambandam, Ivan Maillard, Valerie P Zediak, Lanwei Xu, Rachel M Gerstein, Jon C Aster, Warren S Pear, and Avinash Bhandoola. Notch signaling controls the generation and differentiation of early t lineage progenitors. *Nature immunology*, 6(7):663–670, 2005.

[62] Marie Cherrier, Shinichiro Sawa, and Gérard Eberl. Notch, id2, and rorγt sequentially orchestrate the fetal development of lymphoid tissue inducer cells. *Journal of Experimental Medicine*, 209(4):729–740, 2012.

[63] Cécilie Possot, Sandrine Schmutz, Sylvestre Chea, Laurent Boucontet, Anne Louise, Ana Cumano, and Rachel Golub. Notch signaling is necessary for adult, but not fetal, development of ror [gamma] t+ innate lymphoid cells. *Nature immunology*, 12(10):949–958, 2011.

[64] Isabel E Ishizuka, Sylvestre Chea, Herman Gudjonson, Michael G Constantinides, Aaron R Dinner, Albert Bendelac, and Rachel Golub. Single-cell analysis defines the divergence between the innate lymphoid cell lineage and lymphoid tissue-inducer cell lineage. *Nature immunology*, 17(3):269–276, 2016.

[65] Fanny L Casado, Kameshwar P Singh, and Thomas A Gasiewicz. The aryl hydrocarbon receptor: regulation of hematopoiesis and involvement in the progression of blood diseases. *Blood Cells, Molecules, and Diseases*, 44(4):199–206, 2010.

[66] Elina A Kiss, Cedric Vonarbourg, Stefanie Kopfmann, Elias Hobeika, Daniela Finke, Charlotte Esser, and Andreas Diefenbach. Natural aryl hydrocarbon receptor ligands control organogenesis of intestinal lymphoid follicles. *Science*, 334(6062):1561–1565, 2011.

[67] Jacob S Lee, Marina Cella, and Marco Colonna. Ahr and the transcriptional regulation of type-17/22 ilc. *Frontiers in immunology*, 3, 2012.

[68] SP Spencer, C Wilhelm, Q Yang, JA Hall, N Bouladoux, A Boyd, TB Nutman, JF Urban, J Wang, TR Ramalingam, et al. Adaptation of innate lymphoid cells to a micronutrient deficiency promotes type 2 barrier immunity. *Science*, 343(6169):432–437, 2014.

[69] Serge A van de Pavert, Manuela Ferreira, Rita G Domingues, Hélder Ribeiro, Rosalie Molenaar, Lara Moreira-Santos, Francisca F Almeida, Sales Ibiza, Inês Barbosa, Gera Goverse, et al. Maternal retinoids control type 3 innate lymphoid cells and set the offspring immunity. *Nature*, 508(7494):123–127, 2014.

[70] Isabel E Ishizuka. *Stages and Transcriptional Regulation of Innate Lymphoid Cell Development*. PhD thesis, UChicago, March 2017.

[71] Ana Conesa, Pedro Madrigal, Sonia Tarazona, David Gomez-Cabrero, Alejandra Cervera, Andrew McPherson, Michał Wojciech Szcześniak, Daniel J Gaffney, Laura L Elo, Xuegong Zhang, et al. A survey of best practices for rna-seq data analysis. *Genome biology*, 17(1):13, 2016.

[72] Rob Patro, Stephen M Mount, and Carl Kingsford. Sailfish enables alignment-free isoform quantification from rna-seq reads using lightweight algorithms. *Nature biotechnology*, 32(5):462–464, 2014.

[73] David B Allison, Xiangqin Cui, Grier P Page, and Mahyar Sabripour. Microarray data analysis: from disarray to consolidation and consensus. *Nature reviews genetics*, 7(1):55–65, 2006.

[74] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.

[75] Gordon K Smyth. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology*, 3(1):1–25, 2004.

[76] Gordon K Smyth, Matthew Ritchie, Natalie Thorne, and James Wettenhall. Limma: linear models for microarray data. in bioinformatics and computational biology solutions using r and bioconductor. statistics for biology and health. 2005.

[77] Michael Love, Simon Anders, and Wolfgang Huber. Differential analysis of count data–the deseq2 package. *Genome Biol*, 15:550, 2014.

[78] Mark D Robinson, Davis J McCarthy, and Gordon K Smyth. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, 2010.

[79] Peter V Kharchenko, Lev Silberstein, and David T Scadden. Bayesian approach to single-cell differential expression analysis. *Nature methods*, 11(7):740–742, 2014.

[80] Cole Trapnell. Defining cell types and states with single-cell genomics. *Genome research*, 25(10):1491–1498, 2015.

[81] Xiaojie Qiu, Qi Mao, Ying Tang, Li Wang, Raghav Chawla, Hannah Pliner, and Cole Trapnell. Reversed graph embedding resolves complex single-cell developmental trajectories. *bioRxiv*, page 110668, 2017.

[82] Manu Setty, Michelle D Tadmor, Shlomit Reich-Zeliger, Omer Angel, Tomer Meir Salame, Pooja Kathail, Kristy Choi, Sean Bendall, Nir Friedman, and Dana Pe'er. Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nature biotechnology*, 34(6):637–645, 2016.

[83] Laleh Haghverdi, Maren Buettner, F Alexander Wolf, Florian Buettner, and Fabian J Theis. Diffusion pseudotime robustly reconstructs lineage branching. *Nature methods*, 13(10):845–848, 2016.

[84] Joshua D Welch, Alexander J Hartemink, and Jan F Prins. Slicer: inferring branched, nonlinear cellular trajectories from single cell rna-seq data. *Genome biology*, 17(1):106, 2016.

[85] Florian Buettner, Kedar N Natarajan, F Paolo Casale, Valentina Proserpio, Antonio Scialdone, Fabian J Theis, Sarah A Teichmann, John C Marioni, and Oliver Stegle. Computational analysis of cell-to-cell heterogeneity in single-cell rna-sequencing data reveals hidden subpopulations of cells. *Nature biotechnology*, 33(2):155–160, 2015.

[86] Maria Tsompana and Michael J Buck. Chromatin accessibility: a window into the genome. *Epigenetics & chromatin*, 7(1):33, 2014.

[87] Jason D Buenrostro, Paul G Giresi, Lisa C Zaba, Howard Y Chang, and William J Greenleaf. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, dna-binding proteins and nucleosome position. *Nature methods*, 10(12):1213–1218, 2013.

[88] Elizabeth G Wilbanks and Marc T Facciotti. Evaluation of algorithm performance in chip-seq peak detection. *PloS one*, 5(7):e11471, 2010.

[89] Rory Stark and Gordon Brown. Diffbind: differential binding analysis of chip-seq peak data. *R package version*, 100, 2011.

[90] Roger Pique-Regi, Jacob F Degner, Athma A Pai, Daniel J Gaffney, Yoav Gilad, and Jonathan K Pritchard. Accurate inference of transcription factor binding from dna sequence and chromatin accessibility data. *Genome research*, 21(3):447–455, 2011.

[91] Joseph C Sun and Lewis L Lanier. The natural selection of herpesviruses and virus-specific nk cell receptors. *Viruses*, 1(3):362–382, 2009.

[92] Wayne M Yokoyama, Sungjin Kim, and Anthony R French. The dynamic life of natural killer cells. *Annu. Rev. Immunol.*, 22:405–429, 2004.

[93] Eric Vivier, David H Raulet, Alessandro Moretta, Michael A Caligiuri, Laurence Zitvogel, Lewis L Lanier, Wayne M Yokoyama, and Sophie Ugolini. Innate or adaptive immunity? the example of natural killer cells. *Science*, 331(6013):44–49, 2011.

[94] Wayne M Yokoyama, Dorothy K Sojka, Hui Peng, and Zhigang Tian. Tissue-resident natural killer cells. In *Cold Spring Harbor symposia on quantitative biology*, volume 78, pages 149–156. Cold Spring Harbor Laboratory Press, 2013.

[95] Hui Peng, Xiaojun Jiang, Yonglin Chen, Dorothy K Sojka, Haiming Wei, Xiang Gao, Rui Sun, Wayne M Yokoyama, and Zhigang Tian. Liver-resident nk cells confer adaptive immunity in skin-contact inflammation. *The Journal of clinical investigation*, 123(4):1444, 2013.

[96] Victor S Cortez, Anja Fuchs, Marina Cella, Susan Gilfillan, and Marco Colonna. Cutting edge: Salivary gland nk cells develop independently of nfil3 in steady-state. *The Journal of Immunology*, 192(10):4487–4491, 2014.

[97] Dorothy K Sojka, Zhigang Tian, and Wayne M Yokoyama. Tissue-resident natural killer cells and their potential diversity. In *Seminars in immunology*, volume 26, pages 127–131. Elsevier, 2014.

[98] Cécile Daussy, Fabrice Faure, Katia Mayol, Sébastien Viel, Georg Gasteiger, Emily Charrier, Jacques Bienvenu, Thomas Henry, Emilie Debien, Uzma A Hasan, et al. T-bet and eomes instruct the development of two distinct natural killer cell lineages in the liver and in the bone marrow. *Journal of Experimental Medicine*, pages jem–20131560, 2014.

[99] Kazuyoshi Takeda, Erika Cretney, Yoshihiro Hayakawa, Tsuyoshi Ota, Hisaya Akiba, Kouetsu Ogasawara, Hideo Yagita, Katsuyuki Kinoshita, Ko Okumura, and Mark J Smyth. Trail identifies immature natural killer cells in newborn mice and adult mouse liver. *Blood*, 105(5):2082–2089, 2005.

[100] Scott M Gordon, Julie Chaix, Levi J Rupp, Junmin Wu, Sharline Madera, Joseph C Sun, Tullia Lindsten, and Steven L Reiner. The transcription factors t-bet and eomes control key checkpoints of natural killer cell maturation. *Immunity*, 36(1):55–67, 2012.

[101] Christian AJ Vosshenrich and James P Di Santo. Developmental programming of natural killer and innate lymphoid cells. *Current opinion in immunology*, 25(2):130–138, 2013.

[102] Sungjin Kim, Koho Iizuka, Hyun-Seok P Kang, Ayotunde Dokun, Anthony R French, Suellen Greco, and Wayne M Yokoyama. In vivo developmental stages in murine natural killer cell maturation. *Nature immunology*, 3(6):523–528, 2002.

[103] Eleftheria E Rosmaraki, Iyadh Douagi, Claude Roth, Francesco Colucci, Ana Cumano, and James P Di Santo. Identification of committed nk cell progenitors in adult murine bone marrow. *European journal of immunology*, 31(6):1900–1909, 2001.

[104] John W Fathman, Deepta Bhattacharya, Matthew A Inlay, Jun Seita, Holger Karsunky, and Irving L Weissman. Identification of the earliest natural killer cell–committed progenitor in murine bone marrow. *Blood*, 118(20):5439–5447, 2011.

[105] Sebastian Carotta, Swee Heng Milon Pang, Stephen L Nutt, and Gabrielle T Belz. Identification of the earliest nk-cell precursor in the mouse bm. *Blood*, 117(20):5449–5452, 2011.

[106] Andreas Diefenbach, Marco Colonna, and Shigeo Koyasu. Development, differentiation, and diversity of innate lymphoid cells. *Immunity*, 41(3):354–365, 2014.

[107] LEWIS L Lanier, CHIWEN Chang, HERGEN Spits, and JOSEPH H Phillips. Expression of cytoplasmic cd3 epsilon proteins in activated human adult natural killer (nk) cells and cd3 gamma, delta, epsilon complexes in fetal nk cells. implications for the relationship of nk and t lymphocytes. *The Journal of Immunology*, 149(6):1876–1880, 1992.

[108] Yuefeng Huang, Liying Guo, Jin Qiu, Xi Chen, Jane Hu-Li, Ulrich Siebenlist, Peter R Williamson, Joseph F Urban Jr, and William E Paul. Il-25-responsive, lineage-negative klrg1hi cells are multipotential/inflammatory/'type 2 innate lymphoid cells. *Nature immunology*, 16(2):161–169, 2015.

[109] Pan Du, Warren A Kibbe, and Simon M Lin. lumi: a pipeline for processing illumina microarray. *Bioinformatics*, 24(13):1547–1548, 2008.

[110] Simon M Lin, Pan Du, Wolfgang Huber, and Warren A Kibbe. Model-based variance-stabilizing transformation for illumina microarray data. *Nucleic acids research*, 36(2):e11–e11, 2008.

[111] Nicolas Serafini, Christian AJ Vosshenrich, and James P Di Santo. Transcriptional regulation of innate lymphoid cell fate. *Nature Reviews Immunology*, 15(7):415–428, 2015.

[112] Nicolas Serafini, Roel GJ Klein Wolterink, Naoko Satoh-Takayama, Wei Xu, Christian AJ Vosshenrich, Rudi W Hendriks, and James P Di Santo. Gata3 drives development of rorγt+ group 3 innate lymphoid cells. *Journal of Experimental Medicine*, pages jem–20131038, 2014.

[113] Jennifer A Walker, Christopher J Oliphant, Alexandros Englezakis, Yong Yu, Simon Clare, Hans-Reimer Rodewald, Gabrielle Belz, Pentao Liu, Padraic G Fallon, and Andrew NJ McKenzie. Bcl11b is essential for group 2 innate lymphoid cell development. *Journal of Experimental Medicine*, 212(6):875–882, 2015.

[114] Kazuyo Moro and Shigeo Koyasu. Innate lymphoid cells, possible interaction with microbiota. In *Seminars in immunopathology*, volume 37, pages 27–37. Springer, 2015.

[115] Michael G Constantinides, Herman Gudjonson, Benjamin D McDonald, Isabel E Ishizuka, Philip A Verhoef, Aaron R Dinner, and Albert Bendelac. Plzf expression maps the early stages of ilc1 lineage development. *Proceedings of the National Academy of Sciences*, 112(16):5123–5128, 2015.

[116] Masashi Tachibana, Mari Tenno, Chieko Tezuka, Machiko Sugiyama, Hisahiro Yoshida, and Ichiro Taniuchi. Runx1/cbfβ2 complexes are required for lymphoid tissue inducer cell differentiation at two developmental stages. *The Journal of Immunology*, 186(3):1450–1457, 2011.

[117] Parinaz Aliahmad, Brian De La Torre, and Jonathan Kaye. Shared dependence on the dna-binding factor tox for the development of lymphoid tissue-inducer cell and nk cell lineages. *Nature immunology*, 11(10):945–952, 2010.

[118] Kevin Ramirez, Katherine J Chandler, Christina Spaulding, Sasan Zandi, Mikael Sigvardsson, Barbara J Graves, and Barbara L Kee. Gene deregulation and chronic activation in natural killer cells deficient in the transcription factor ets1. *Immunity*, 36(6):921–932, 2012.

[119] Jennifer K Bando, Hong-Erh Liang, and Richard M Locksley. Identification and distribution of developing innate lymphoid cells in the fetal mouse intestine. *Nature immunology*, 16(2):153–160, 2015.

[120] Jinfang Zhu and William E Paul. Peripheral cd4+ t-cell differentiation regulated by networks of cytokines and transcription factors. *Immunological reviews*, 238(1):247–262, 2010.

[121] Daniel R Neill, See Heng Wong, Agustin Bellosi, Robin J Flynn, Maria Daly, Theresa KA Langford, Christine Bucks, Colleen M Kane, Padraic G Fallon, Richard Pannell, et al. Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. *Nature*, 464(7293):1367–1370, 2010.

[122] Peter Laslo, Chauncey J Spooner, Aryeh Warmflash, David W Lancki, Hyun-Jun Lee, Roger Sciammas, Benjamin N Gantner, Aaron R Dinner, and Harinder Singh. Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell*, 126(4):755–766, 2006.

[123] Ming Hu, Diane Krause, Mel Greaves, Saul Sharkis, Michael Dexter, Clare Heyworth, and Tariq Enver. Multilineage gene expression precedes commitment in the hemopoietic system. *Genes & development*, 11(6):774–785, 1997.

[124] Toshihiro Miyamoto, Hiromi Iwasaki, Boris Reizis, Min Ye, Thomas Graf, Irving L Weissman, and Koichi Akashi. Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. *Developmental cell*, 3(1):137–147, 2002.

[125] Samuel Yao-Ming Ng, Toshimi Yoshida, Jiangwen Zhang, and Katia Georgopoulos. Genome-wide lineage-specific transcriptional networks underscore ikaros-dependent lymphoid priming in hematopoietic stem cells. *Immunity*, 30(4):493–507, 2009.

[126] Yaron E Antebi, Shlomit Reich-Zeliger, Yuval Hart, Avi Mayo, Inbal Eizenberg, Jacob Rimer, Prabhakar Putheti, Dana Pe'er, and Nir Friedman. Mapping differentiation under mixed culture conditions reveals a tunable continuum of t cell fates. *PLoS biology*, 11(7):e1001616, 2013.

[127] Naoko Satoh-Takayama, Nicolas Serafini, Thomas Verrier, Abdessalem Rekiki, Jean-Christophe Renauld, Gad Frankel, and James P Di Santo. The chemokine receptor cxcr6 controls the functional topography of interleukin-22 producing intestinal innate lymphoid cells. *Immunity*, 41(5):776–788, 2014.

[128] Jessica C Mar, Yasumasa Kimura, Kate Schroder, Katharine M Irvine, Yoshihide Hayashizaki, Harukazu Suzuki, David Hume, and John Quackenbush. Data-driven normalization strategies for high-throughput quantitative rt-pcr. *BMC bioinformatics*, 10(1):110, 2009.

[129] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[130] El-ad David Amir, Kara L Davis, Michelle D Tadmor, Erin F Simonds, Jacob H Levine, Sean C Bendall, Daniel K Shenfeld, Smita Krishnaswamy, Garry P Nolan, and Dana Pe'er. visne enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature biotechnology*, 31(6):545–552, 2013.

[131] Reina E Mebius, Toshihiro Miyamoto, Julie Christensen, Jos Domen, Tom Cupedo, Irving L Weissman, and Koichi Akashi. The fetal liver counterpart of adult common lymphoid progenitors gives rise to all lymphoid lineages, cd45+ cd4+ cd3- cells, as well as macrophages. *The Journal of Immunology*, 166(11):6593–6601, 2001.

[132] Isabel E Ishizuka, Michael G Constantinides, Herman Gudjonson, and Albert Bendelac. The innate lymphoid cell precursor. *Annual review of immunology*, 34:299–316, 2016.

[133] Han-Yu Shih, Giuseppe Sciumè, Yohei Mikami, Liying Guo, Hong-Wei Sun, Stephen R Brooks, Joseph F Urban, Fred P Davis, Yuka Kanno, and John J O'Shea. Developmental acquisition of regulomes underlies innate lymphoid cell functionality. *Cell*, 165(5):1120–1133, 2016.

[134] Meital Gury-BenAri, Christoph A Thaiss, Nicolas Serafini, Deborah R Winter, Amir Giladi, David Lara-Astiaso, Maayan Levy, Tomer Meir Salame, Assaf Weiner, Eyal David, et al. The spectrum and regulatory landscape of intestinal innate lymphoid cells are shaped by the microbiome. *Cell*, 166(5):1231–1246, 2016.

[135] Chamutal Bornstein, Deborah Winter, Zohar Barnett-Itzhaki, Eyal David, Sabah Kadri, Manuel Garber, and Ido Amit. A negative feedback loop of transcription factors specifies alternative dendritic cell chromatin states. *Molecular cell*, 56(6):749–762, 2014.

[136] David Lara-Astiaso, Assaf Weiner, Erika Lorenzo-Vivas, Irina Zaretsky, Diego Adhemar Jaitin, Eyal David, Hadas Keren-Shaul, Alexander Mildner, Deborah Winter, Steffen Jung, et al. Chromatin state dynamics during blood formation. *Science*, 345(6199):943–949, 2014.

[137] Liza B John and Alister C Ward. The ikaros gene family: transcriptional regulators of hematopoiesis and immunity. *Molecular immunology*, 48(9):1272–1278, 2011.

[138] Fernando Macian. Nfat proteins: key regulators of t-cell development and function. *Nature Reviews Immunology*, 5(6):472–484, 2005.

[139] Cyril Seillet, Lisa A Mielke, Daniela B Amann-Zalcenstein, Shian Su, Jerry Gao, Francisca F Almeida, Wei Shi, Matthew E Ritchie, Shalin H Naik, Nicholas D Huntington, et al. Deciphering the innate lymphoid cell transcriptional program. *Cell reports*, 17(2):436–447, 2016.

[140] Yoshiyuki Minegishi, Jurg Rohrer, Elaine Coustan-Smith, Howard M Lederman, Rajita Pappu, Dario Campana, Andrew C Chan, and Mary Ellen Conley. An essential role for blnk in human b cell development. *Science*, 286(5446):1954–1957, 1999.

[141] Charles E Grant, Timothy L Bailey, and William Stafford Noble. Fimo: scanning for occurrences of a given motif. *Bioinformatics*, 27(7):1017–1018, 2011.

[142] M Ryan Corces, Jason D Buenrostro, Beijing Wu, Peyton G Greenside, Steven M Chan, Julie L Koenig, Michael P Snyder, Jonathan K Pritchard, Anshul Kundaje, William J Greenleaf, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature genetics*, 48(10):1193–1203, 2016.

[143] Yingying Wei, Toyoaki Tenzen, and Hongkai Ji. Joint analysis of differential gene expression in multiple studies using correlation motifs. *Biostatistics*, 16(1):31–46, 2014.

[144] Timothée Flutre, Xiaoquan Wen, Jonathan Pritchard, and Matthew Stephens. A statistical framework for joint eqtl analysis in multiple tissues. *PLoS genetics*, 9(5):e1003486, 2013.

[145] Parinaz Aliahmad, Asha Kadavallore, Brian de la Torre, Dietmar Kappes, and Jonathan Kaye. Tox is required for development of the cd4 t cell lineage gene program. *The Journal of Immunology*, 187(11):5931–5940, 2011.

[146] Victoria Male, Ilaria Nisoli, Tomasz Kostrzewski, David SJ Allan, James R Carlyle, Graham M Lord, Andreas Wack, and Hugh JM Brady. The transcription factor e4bp4/nfil3 controls commitment to the nk lineage and directly regulates eomes and id2 expression. *Journal of Experimental Medicine*, 211(4):635–642, 2014.

[147] Susan M Schlenner, Vikas Madan, Katrin Busch, Annette Tietz, Carolin Läufle, Celine Costa, Carmen Blum, Hans Jörg Fehling, and Hans-Reimer Rodewald. Fate mapping reveals separate origins of t cells and myeloid lineages in the thymus. *Immunity*, 32(3):426–436, 2010.

[148] David Allman, Arivazhagan Sambandam, Sungjune Kim, Juli P Miller, Antonio Pagan, David Well, Anita Meraz, and Avinash Bhandoola. Thymopoiesis independent of common lymphoid progenitors. *Nature immunology*, 4(2):168–174, 2003.

[149] Renata Mazzucchelli and Scott K Durum. Interleukin-7 receptor expression: intelligent design. *Nature Reviews Immunology*, 7(2):144–154, 2007.

[150] Vanja Lazarevic, Xi Chen, Jae-Hyuck Shim, Eun-Sook Hwang, Eunjung Jang, Alexandra N Bolm, Mohamed Oukka, Vijay K Kuchroo, and Laurie H Glimcher. T-bet represses th17 differentiation by preventing runx1-mediated activation of the gene encoding ror [gamma] t. *Nature immunology*, 12(1):96–104, 2011.

[151] Makoto Kuwahara, Masakatsu Yamashita, Kenta Shinoda, Soichi Tofukuji, Atsushi Onodera, Ryo Shinnakasu, Shinichiro Motohashi, Hiroyuki Hosokawa, Damon Tumes, Chiaki Iwamura, et al. The transcription factor sox4 is a downstream target of signaling by the cytokine tgf-[beta] and suppresses th2 differentiation. *Nature immunology*, 13(8):778–786, 2012.

[152] Ellen V Rothenberg, Ameya Champhekar, Sagar Damle, Marissa Morales Del Real, Hao Yuan Kueh, Long Li, and Mary A Yui. Transcriptional establishment of cell-type identity: dynamics and causal mechanisms of t-cell lineage commitment. In *Cold Spring Harbor symposia on quantitative biology*, volume 78, pages 31–41. Cold Spring Harbor Laboratory Press, 2013.

[153] Jinfang Zhu, Dragana Jankovic, Alex Grinberg, Liying Guo, and William E Paul. Gfi-1 plays an important role in il-2-mediated th2 cell expansion. *Proceedings of the National Academy of Sciences*, 103(48):18214–18219, 2006.

[154] Jinfang Zhu, Hidehiro Yamane, and William E Paul. Differentiation of effector cd4 t cell populations. *Annual review of immunology*, 28:445–489, 2009.

[155] Ryoji Yagi, Ilkka S Junttila, Gang Wei, Joseph F Urban, Keji Zhao, William E Paul, and Jinfang Zhu. The transcription factor gata3 actively represses runx3 protein-regulated production of interferon-γ. *Immunity*, 32(4):507–517, 2010.

[156] Georg Gasteiger, Xiying Fan, Stanislav Dikiy, Sue Y Lee, and Alexander Y Rudensky. Tissue residency of innate lymphoid cells in lymphoid and nonlymphoid organs. *Science*, 350(6263):981–985, 2015.

[157] Frédéric Vély, Vincent Barlogis, Blandine Vallentin, Bénédicte Neven, Christelle Piperoglou, Mikael Ebbo, Thibaut Perchet, Maxime Petit, Nadia Yessaad, Fabien Touzot, et al. Evidence of innate lymphoid cell redundancy in humans. *Nature immunology*, 17(11):1291–1299, 2016.

[158] Faiyaz Notta, Sasan Zandi, Naoya Takayama, Stephanie Dobson, Olga I Gan, Gavin Wilson, Kerstin B Kaufmann, Jessica McLeod, Elisa Laurenti, Cyrille F Dunant, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*, 351(6269):aab2116, 2016.

[159] Cristina Pina, Cristina Fugazza, Alex J Tipping, John Brown, Shamit Soneji, Jose Teles, Carsten Peterson, and Tariq Enver. Inferring rules of lineage commitment in haematopoiesis. *Nature cell biology*, 14(3):287–294, 2012.

[160] Elsa Abranches, Ana MV Guedes, Martin Moravec, Hedia Maamar, Petr Svoboda, Arjun Raj, and Domingos Henrique. Stochastic nanog fluctuations allow mouse embryonic stem cells to explore pluripotency. *Development*, 141(14):2770–2779, 2014.

[161] Ellen V Rothenberg. Transcriptional control of early t and b cell developmental choices. *Annual review of immunology*, 32:283–321, 2014.

[162] Fritz Melchers. Checkpoints that control b cell development. *The Journal of clinical investigation*, 125(6):2203, 2015.

[163] Andrea C Carpenter and Rémy Bosselut. Decision checkpoints in the thymus. *Nature immunology*, 11(8):666–673, 2010.

[164] Takeshi Egawa, Gerard Eberl, Ichiro Taniuchi, Kamel Benlagha, Frederic Geissmann, Lothar Hennighausen, Albert Bendelac, and Dan R Littman. Genetic evidence supporting selection of the vα14i nkt cell lineage from double-positive thymocyte precursors. *Immunity*, 22(6):705–716, 2005.

[165] Benjamin D McDonald, Jeffrey J Bunker, Isabel E Ishizuka, Bana Jabri, and Albert Bendelac. Elevated t cell receptor signaling identifies a thymic precursor to the tcr$\alpha\beta$+ cd4- cd8β- intraepithelial lymphocyte lineage. *Immunity*, 41(2):219–229, 2014.

[166] Gretta L Stritesky, Stephen C Jameson, and Kristin A Hogquist. Selection of self-reactive t cells in the thymus. *Annual review of immunology*, 30:95–114, 2012.

[167] Andrew McDavid, Greg Finak, Pratip K Chattopadyay, Maria Dominguez, Laurie Lamoreaux, Steven S Ma, Mario Roederer, and Raphael Gottardo. Data exploration, quality control and testing in single-cell qpcr-based gene expression experiments. *Bioinformatics*, 29(4):461–467, 2012.

[168] Philip Brennecke, Simon Anders, Jong Kyoung Kim, Aleksandra A Kołodziejczyk, Xiuwei Zhang, Valentina Proserpio, Bianka Baying, Vladimir Benes, Sarah A Teichmann, John C Marioni, et al. Accounting for technical noise in single-cell rna-seq experiments. *Nature methods*, 10(11):1093–1095, 2013.

[169] Dominic Grün, Lennart Kester, and Alexander Van Oudenaarden. Validation of noise models for single-cell transcriptomics. *Nature methods*, 11(6):637–640, 2014.

[170] Elham Azizi, Sandhya Prabhakaran, Ambrose Carr, and Dana Pe'er. Bayesian inference for single-cell clustering and imputing. *Genomics and Computational Biology*, 3(1):46, 2017.

[171] Eugenio Marco, Robert L Karp, Guoji Guo, Paul Robson, Adam H Hart, Lorenzo Trippa, and Guo-Cheng Yuan. Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proceedings of the National Academy of Sciences*, 111(52):E5643–E5650, 2014.

[172] Rob Morris, Ignacio Sancho-Martinez, Tatyana O Sharpee, and Juan Carlos Izpisua Belmonte. Mathematical approaches to modeling development and reprogramming. *Proceedings of the National Academy of Sciences*, 111(14):5076–5082, 2014.

[173] Victoria Moignard, Steven Woodhouse, Laleh Haghverdi, Andrew J Lilly, Yosuke Tanaka, Adam C Wilkinson, Florian Buettner, Iain C Macaulay, Wajid Jawaid, Evangelia Diamanti, et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nature biotechnology*, 33(3):269–276, 2015.

[174] Yosef Buganim, Dina A Faddah, Albert W Cheng, Elena Itskovich, Styliani Markoulaki, Kibibi Ganz, Sandy L Klemm, Alexander van Oudenaarden, and Rudolf Jaenisch. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*, 150(6):1209–1222, 2012.

[175] Kushal Suryamohan and Marc S Halfon. Identifying transcriptional cis-regulatory modules in animal genomes. *Wiley Interdisciplinary Reviews: Developmental Biology*, 4(2):59–84, 2015.

[176] Mark Maienschein-Cline, Jie Zhou, Kevin P White, Roger Sciammas, and Aaron R Dinner. Discovering transcription factor regulatory targets using gene expression and binding data. *Bioinformatics*, 28(2):206–213, 2011.

[177] Evgeny Z Kvon, Tomas Kazmar, Gerald Stampfel, J Omar Yáñez-Cuna, Michaela Pagani, Katharina Schernhuber, Barry J Dickson, and Alexander Stark. Genome-scale functional characterization of drosophila developmental enhancers in vivo. *Nature*, 512(7512):91–95, 2014.

[178] Alvaro J González, Manu Setty, and Christina S Leslie. Early enhancer establishment and regulatory locus complexity shape transcriptional programs in hematopoietic differentiation. *Nature genetics*, 47(11):1249–1259, 2015.