

TRANSCRIPTIONAL LANDSCAPES IN LYMPHOCYTE DEVELOPMENT AND  
DIFFERENTIATION: TCF-1 ENFORCES EPIGENETIC IDENTITY IN DEVELOPING T CELLS  
AND T-BET RESOLVES FUNCTIONALLY DISTINCT MEMORY B CELLS

John L. Johnson

A DISSERTATION

in

Immunology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation

  
\_\_\_\_\_

Michael P. Cancro, Ph.D.

Professor of Pathology and Laboratory Medicine

Graduate Group Chairperson

  
\_\_\_\_\_

David Allman, Ph.D.

Professor of Pathology and Laboratory Medicine

Dissertation Committee

David Allman, Ph.D., Professor of Pathology and Laboratory Medicine

Christopher A. Hunter, Ph.D., Mindy Halikman Heyer Distinguished Professor of Pathobiology

Laurence C., Eisenlohr, V.M.D., Ph.D., Professor of Pathology and Laboratory Medicine

Will Bailis, Ph.D., Assistant Professor of Microbiology

## *DEDICATION*

TO

*My mother and father*

*This dissertation is the fruit which grew from humble beginnings. You labored to nourish me with an ample supply of books and the instruments of experimentation. The seed of curiosity planted in my earliest childhood was nurtured by your hands in tender care and affection. More importantly, I was cultivated in the light of wisdom and right action.*

## ACKNOWLEDGMENT

My profoundest gratitude goes out to my fellow co-first authors. The publications presented in the pages of this dissertation are not the result of my individual effort, but the product of the cooperation of competent and intelligent people aiming to create something great. I was just in the right place at the right time. To George Georgakilas for your patience and willingness to teach. To Jelena Petrovic for your words of wisdom and your attention to detail. To Rebecca Rosenthal for your perseverance and your friendship. To Jamie Knox for your knowledge and expertise. To all the members of the Vahedi lab, the Wherry lab, the Hunter lab, and the Cancro lab including, but not limited to, Stanley Cai, Makoto and Junko Kurachi, Gretchen Harms Pritchard, and Mariya Kostiv.

To the Penn Immunology Graduate Group, especially Mary Taylor, for providing the perfect environment to encourage my scientific growth. I would also like to thank my IGG cohort for being my friends, for believing in me, and for defending me. I will cherish our memories together and I hope our bond will be strengthened as time goes on. To Naomi Goldman, for your endless support and love.

To Drs. David Allman, Chris Hunter, John Wherry, and Golnaz Vahedi for providing mentorship and training. To Dr. Taku Kambayashi for believing in me enough to give me the opportunity.

Finally, to my mentor, Dr. Michael Paul Cancro, thank you for providing the capstone to my PhD that I was seeking.

# ABSTRACT

## TRANSCRIPTIONAL LANDSCAPES IN LYMPHOCYTE DEVELOPMENT AND DIFFERENTIATION: TCF-1 ENFORCES EPIGENETIC IDENTITY IN DEVELOPING T CELLS AND T-BET RESOLVES FUNCTIONALLY DISTINCT MEMORY B CELLS

John L. Johnson

Michael P. Cancro

An important role of cellular differentiation is to establish distinct and durable cell subsets that serve different functions over the course of an immune response. Here, I investigate the problem of cellular differentiation by considering 1) how epigenetic repression is overcome to establish unique preimmune lymphocyte identity and 2) the durability of intraclonal and interclonal diversification resulting from an immune response. The epigenetic states of hematopoietic cells contain cell-type specific accessible chromatin structures which are developmentally constructed from repressive, compacted chromatin. However, these structures feature binding sites for lineage-specific transcription factors, suggesting these factors play a role in their generation. I used measurements of chromatin accessibility in sequential stages of T cell development from bone-marrow derived progenitors alongside alternative lymphocyte lineages to identify the central role that TCF-1 plays in creating T-cell specific chromatin during differentiation. Genetic deficiency of TCF-1 reduced the accessible T cell chromatin state and the T cell gene program, whereas the ectopic expression of TCF-1 in fibroblasts caused T cell chromatin to become accessible and T cell genes to be expressed. These findings demonstrate that TCF-1 can overcome repressive chromatin to establish a naïve T cell identity distinct from other lymphocyte lineages. Despite our improved understanding of preimmune lymphocyte differentiation, much less is known about the course of lymphocyte differentiation beyond the naïve stage. During immune

responses, some activated B lymphocytes express the transcription factor T-bet, but the clonal relationship to their T-bet<sup>-</sup> counterparts and the durability of the T-bet<sup>+</sup> phenotype is unclear. I found that T-bet<sup>+</sup> B cells are generated early after influenza infection and develop into a persistent memory pool. Immune repertoire profiling of influenza hemagglutinin-specific T-bet<sup>+</sup> and T-bet<sup>-</sup> memory B cells demonstrates that most clones are unique to their respective subset, but lineage tree analysis of the remaining shared clones shows that T-bet<sup>+</sup> clones can stably bifurcate from T-bet<sup>-</sup> cells. Further, genetic fate-mapping indicates that T-bet expression in B cells is stable. Together, these and other findings suggest that T-bet<sup>+</sup> B cells are a distinct and durable memory subset and uniquely contribute to the anti-viral humoral response.

# TABLE OF CONTENTS

<b>ACKNOWLEDGMENT</b>	III
<b>ABSTRACT</b>	IV
<b>LIST OF TABLES</b>	X
<b>LIST OF FIGURES</b>	XI
<b>LIST OF ABBREVIATIONS</b>	XII
<b>CHAPTER 1: OVERVIEW OF CELLULAR DIFFERENTIATION</b>	1
1.1 Cellular differentiation in the immune system and embryology	1
1.2 Cellular differentiation, regional specification, and morphogenesis	3
1.3 Cellular differentiation is a developmental hierarchy	4
1.4 Differentiation is organized at the genetic level by transcription factors	6
1.5 Additional embryological concepts: the fate map and clonal analysis	8
1.6 Cellular differentiation is central to adaptive immune system function	10
<b>CHAPTER 2: TCF-1 ESTABLISHES T CELL EPIGENETIC IDENTITY</b>	12
<b>2.1 PREFACE</b>	12
<b>2.2 INTRODUCTION</b>	13
2.2.1 The diversity of blood cells develops from hematopoietic progenitors	13
2.2.2 T cell development occurs in the thymus	15
2.2.3 Transcription factors acting in T cell development	17
2.2.4 HMG Box transcription factors: TCF-1 and LEF-1	20
2.2.5 Scope	23
<b>2.3 RESULTS</b>	24

2.3.1 Three waves of chromatin remodeling during T cell development	24
2.3.2 TCF-1 is the top enriched transcription factor in mature T cell clusters	25
2.3.4 TCF-1 binding exerts a strong harmonizing impact on the chromatin of single T cells	34
2.3.4 TCF-1 can create <i>de novo</i> chromatin accessibility in fibroblasts	39
2.3.5 TCF-1 binding at chromatin domains with H3K27me3 and H3K9me3 repressive marks	44
2.3.7 Genes up-regulated by TCF-1 reside in previously repressed chromatin domains in fibroblast	49
<b>2.4 DISCUSSION</b>	<b>53</b>
<b>2.5 MATERIALS AND METHODS</b>	<b>57</b>
<b>CHAPTER 3: DIFFERENTIATION OF A DISTINCT AND STABLE T-BET<sup>+</sup> MEMORY B CELL SUBSET</b>	<b>77</b>
<b>3.1 PREFACE</b>	<b>77</b>
<b>3.2: INTRODUCTION</b>	<b>78</b>
3.2.1 The innate and adaptive immune responses	78
3.2.2 B lymphocytes of the adaptive immune response: subsets and development	80
3.2.3 Thymus dependent and independent B cell responses	83
3.2.4 Plasma cells and memory B cells in humoral immunity	86
3.2.5 A T-bet expressing B cell subset accumulates during aging	87
3.2.6 Scope	88
<b>3.3 RESULTS</b>	<b>89</b>
3.3.1 T-bet expression distinguishes unique influenza-specific memory B cell populations	89
3.3.2 Human T-bet <sup>hi</sup> B cells are an anatomically compartmentalized component of influenza-specific memory	96
3.3.3 T-bet <sup>hi</sup> HA-specific memory B cells are resident in the spleen	101

3.3.4 Established T-bet <sup>+</sup> and T-bet <sup>-</sup> memory B cells undergo minimal interconversion	105
3.3.5 HA stalk-specific antibody is derived primarily from the T-bet-expressing B cell compartment	110
<b>3.4 DISCUSSION</b>	<b>115</b>
<b>3.5 MATERIALS AND METHODS</b>	<b>119</b>
<b>CHAPTER 4: PERSPECTIVE</b>	<b>128</b>
4.1 Investigating cellular differentiation in the immune system	128
4.2 Mechanisms of shaping epigenetic cellular identity in T cell development	129
4.3 TCF-1 establishes T cell chromatin during T cell development	133
4.4 Perspective on epigenetic engineering to control differentiation in immune responses	135
4.4 Cellular differentiation in the memory B cell response	136
4.5 T-bet <sup>+</sup> memory B cells arise independently from T-bet <sup>-</sup> memory B cells	140
4.6 Perspective on the future research of T-bet <sup>+</sup> B cells	142
<b>APPENDIX: SUPPLEMENTARY FIGURES</b>	<b>144</b>
Figure S1. TCF-1 binding occurs at three waves of chromatin remodeling during T cell development, Related to Figure 2	145
Figure S2. TCF-1-deficient DP T cells cannot establish the open chromatin landscape and transcriptional output of normal DP T cells, Related to Figure 3	146
Figure S3. TCF-1 binding coordinates open chromatin between single DP T cells, Related to Figure 4	147
Figure S4. TCF-1 can bind to nucleosomes and create chromatin accessibility in fibroblasts, Related to Figure 5	149
Figure S5. TCF-1 can bind to repressed chromatin and promote accessibility, Related to Figure 6	151
Figure S6. T cell-specific genes innately repressed in fibroblasts are upregulated by TCF-1, Related to Figure 7	153
Figure S7. Characterization of HA-specific B cells after influenza infection, Related to Figure 8.	155
Figure S8. Characterization of human B cell subsets, Related to Figure 9.	157

<b>Figure S9. Enumeration of HA-specific B cells in parabiotic partners, Related to Figure 10.</b>	<b>158</b>
<b>Figure S10. Immune repertoire analysis by IgH sequencing, Related to Figure 11.</b>	<b>160</b>
<b>BIBLIOGRAPHY</b>	<b>161</b>

## LIST OF TABLES

<b>Table I: Number of hemagglutinin-specific B cells after influenza infection. Related to Figure 8.</b>	<b>95</b>
<b>Table II: Serum antibody titers to various influenza antigens after infection. Related to Figure 12</b>	<b>114</b>

## LIST OF FIGURES

<b>Figure 1: Developmental hierarchy in hematopoiesis</b>	<b>5</b>
<b>Figure 2: TCF-1 binding occurs at three waves of chromatin remodeling during T cell development</b>	<b>28</b>
<b>Figure 3: TCF-1-deficient T cells cannot establish the open chromatin landscape and transcriptional output of normal T cells</b>	<b>32</b>
<b>Figure 4: TCF-1 binding has the highest coordinate impact on open chromatin of single T cells</b>	<b>37</b>
<b>Figure 5: TCF-1 can bind to nucleosomes and create chromatin accessibility in fibroblasts</b>	<b>42</b>
<b>Figure 6: TCF-1 can bind to repressed chromatin and make it open</b>	<b>46</b>
<b>Figure 7: T cell-specific genes innately repressed in fibroblasts are upregulated by TCF-1</b>	<b>51</b>
<b>Figure 8: T-bet expression identifies memory B cell populations with unique tissue distribution</b>	<b>93</b>
<b>Figure 9: Human T-bet<sup>hi</sup> B cells do not recirculate via the lymphatics and maintain influenza-specific memory in the spleen</b>	<b>99</b>
<b>Figure 10: T-bet expression resolves spleen resident versus recirculating MBC pools</b>	<b>103</b>
<b>Figure 11: T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs are selected from a shared pre-immune lineage but do not interconvert</b>	<b>108</b>
<b>Figure 12: T-bet<sup>+</sup> B cells are required for optimal influenza antibody responses and HA stalk-specific antibody in mice</b>	<b>113</b>
<b>Figure 13: Current working model: Forging T cell epigenetic identity using TCF-1</b>	<b>134</b>
<b>Figure 14: Current working model: The differentiation of T-bet<sup>+</sup> memory B cell subsets</b>	<b>141</b>

## LIST OF ABBREVIATIONS

ABC	Age-associated B Cells
AD	Agglutination dose
ADCC	Antibody-Dependent Cell mediated Cytotoxicity
AID	Activation-Induced Deaminase
APC	Antigen Presenting Cell
APRIL	A Proliferation Inducing Ligand
ASW	Average Silhouette Width
ATAC	Assay for Transposase-Accessible Chromatin
BAFF	B cell Activating Factor
BCR	B Cell Receptor
bHLH	beta Helix-Loop-Helix
BM	Bone Marrow
BPL	beta-Propiolactone
CCR7	C-chemokine receptor-7
CCR9	C-chemokine receptor-9
CDR	Complementarity Determining Region
CLP	Common Lymphoid Progenitor
CMP	Common Myeloid Progenitor
CRM	<i>Cis</i> -Regulatory Module
CSL	CBF1/Su(H)/LAG1
CSR	Class-Switch Recombination
DC	Dendritic Cell
DL	Delta-like
DP	Double Positive
DN	Double Negative
EBF	Early B cell Factor
ETP	Early Thymic Progenitor
FACS	Fluorescence Activated Cell Sorting
Fc	Fragment Crystallizable
FcR	Fragment Crystallizable Receptor
FDC	Follicular Dendritic Cell
GCB	Germinal Center B
GFP	Green Fluorescent Protein
GMP	Granulocyte-Monocyte Progenitor
Gn	Granulocyte
HA	Hemagglutinin

HAI	Hemagglutinin Inhibition
HAU	Hemagglutination units
HDAC	Histone Deacetylase
HIV	Human Immunodeficiency Virus
HMG	High Mobility Group
HPAP	Human Pancreas Analysis Program
HSC	Hematopoietic Stem Cell
ICN	intracellular domain of Notch1
IDR	irreproducible discovery rate
IFN $\gamma$	Interferon Gamma
LDTF	Lineage-Determining Transcription Factor
Ig	Immunoglobulin
IL	Interleukin
i.n.	Intranasal
KO	Knockout
Lin	Lineage
LLPC	Long Lived Plasma Cell
LMPP	Lymphoid-Primed Multipotent Progenitor
LN	Lymph Node
LNP	Lipid Nano Particle
LT-HSC	Long-Term Hematopoietic Stem Cell
MACS	Model-based analysis of ChIP-seq
MAML	mastermind-like
MBC	memory B cell
medLN	Mediastinal lymph node
MEP	Megakaryocyte-erythrocyte progenitor
mesLN	Mesenteric lymph node
MK	Megakaryocyte
MHC	Major Histocompatibility Complex
Mo	Monocyte
MPP	Multipotent Progenitor
M $\phi$	Macrophage
MZ	Marginal Zone
NK	Natural Killer
OAS	Original Antigenic Sin
PAMP	Pathogen Associated Molecular Pattern
PB	Plasmablast
PCA	Principal Component Analysis

pLN	Peripheral lymph node
PNA	Peanut Agglutinin
PSGL	P-selectin glycoprotein 1
PR8	Influenza strain A/PR/8/34
PRC	Polycomb Repressive Complex
PRR	Pattern recognition receptor
PTF	Pioneer Transcription Factor
RAG	Recombination activating gene
RBC	Red Blood Cell
RBPJ	Recombination Signal Combining Protein-J
RPKM	Reads Per Kilobase of Transcript per Million
RV	Retrovirus
scATAC	Single Cell ATAC
SEM	Standard Error of the Mean
SHM	Somatic Hypermutation
SLPC	Short-Lived Plasma Cell
ST-HSC	Short-Term Hematopoietic Stem Cell
T-BET	T-box Expressed in T cells
TCR	T cell receptor
TCID50	50% Tissue Culture Infective Dose
TD	Thymus-dependent
TH	T Helper
TFH	T follicular helper
TI	Thymus-independent
TLR	Toll Like Receptors
TPM	Tags per million
Treg	Regulatory T cell
VH	Heavy chain Variable gene
VST	variance stabilizing transformation
WSS	Within Sum of Squares

## CHAPTER 1: Overview of cellular differentiation

### 1.1 Cellular differentiation in the immune system and embryology

The theme of the 8<sup>th</sup> Midwinter Conference of Immunologists, held in California in 1969, was the Regulation of Cellular Differentiation in the Immune System. Seminal discoveries made in the 1960s shed light on the importance of cooperation between two distinct populations of lymphocytes, B and T cells, in the function of the adaptive immune response. One can only speculate that the organizers felt it worthwhile to provoke new patterns of thought on the matter by exploring connections to more established disciplines. In fact, Ray Owen chaired the first session on “Principles of Cellular Differentiation” with only two speakers: Clifford Grobstein (from UCSD) and Robert Auerbach (from University of Wisconsin, Madison). The session was reported as providing a “very fruitful interaction between two speakers disciplined in embryology and an audience who were specialists in immunology (1).” These fruitful interactions included discussion of the replication of the differentiated state, the different phases of cellular development, and mechanisms of communication. Parallels were drawn between embryonic development and the cellular interactions of thymus and bone marrow derived lymphocytes as well as the stem cell concept in antibody formation; that is, restrictive differentiation, environmental responsiveness, and cell multipotentiality. In his closing remark, Grobstein noted, “both disciplines seek answers to similar questions of differentiation, for example, the kind of regulation that occurs in cells ... the nature of the cue...and the steps in the instructive pathway (1).”

The intertwining of embryology and immunology has not faltered; it has instead grown stronger with the continued discovery of important and distinct lymphocyte

subsets and the increasing appreciation for the epigenetic control of immune cell differentiation, even serving at times as a model for metazoan development. Therefore, the purpose of this overview is to accomplish a number of objectives: 1) to frame current problems in lymphocyte differentiation in the time-tested framework of embryology 2) to highlight the importance of studying lymphocyte differentiation, and 3) to describe the systems I will use to investigate two aspects of lymphocyte differentiation.

The central theme of this dissertation is the cellular differentiation of the lymphocytes of the adaptive immune system, both in their initial pre-immune development and in their further antigen-experienced specialization. Comparisons will be made in this overview to the developmental process occurring in embryogenesis, as this is a tradition of thought extending back to Aristotle and greatly enhanced by brilliant minds that followed. For both embryogenesis and the adaptive immune system, the process of cellular differentiation is fundamental for the establishment of proper function and form. In each process, a group of apparently homogenous cells becomes permanently transformed in character and acquires a specialized function. In gastrulation, the inner cell mass of the blastocyst forms the nervous system, the notochord, the integument, and the gut. In hematopoiesis, stem cells form myeloid cells, erythrocytes, granulocytes, and the T and B lymphocytes of the adaptive immune system. In both cases, all of the parts contain specific types of cells, in the correct proportions and position relative to one another, and carrying out the proper specialized function. Therefore, the form of the question has remained the same in both processes of development: to what extent and by what process do cells differentiate from their progenitors and from each other?

## 1.2 Cellular differentiation, regional specification, and morphogenesis

Cell differentiation is the expression of gene products by a population of cells that are different than those made by their progenitors and different from those made by other populations of cells. Differentiation pervades nearly every aspect of multicellular organisms occurring not only in the embryo, but also in the continual process of tissue repair and homeostasis. Cell differentiation is a complex process requiring genetic and epigenetic mechanisms of gene control, and embryonic development has laid the conceptual groundwork, defined terms, and provided a framework for exploring these mechanisms in all forms of differentiation.

Other developmental processes occur in the embryo but have more limited parallels to lymphocyte differentiation. They are important, and I will touch on them briefly. In early development, the core problem is regional specification. This is the process whereby cells in certain regions of the embryo are turned onto certain pathways of development. The most obvious parallel in the development of the adaptive immune system is the regional specification that occurs when lymphocyte precursors enter the thymus and are specified to become T lymphocytes. Whether a type of regional specification occurs in the peripheral lymphatics during the lifetime of mature lymphocytes is an interesting question, but not one that will be addressed in this dissertation. Regional specification is not to be confused with cell differentiation, as these are both important but distinct problems.

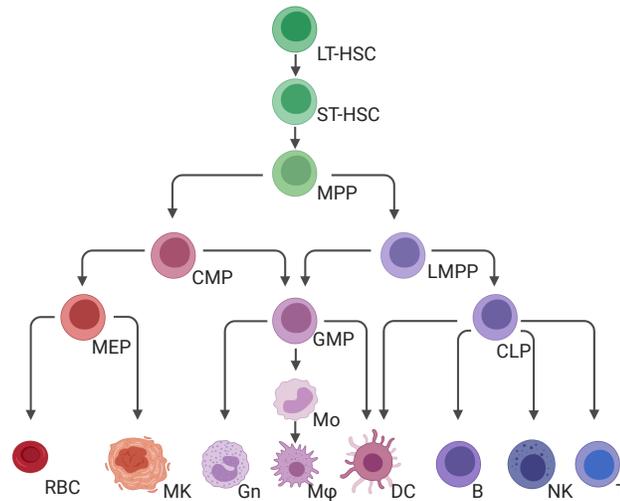
The developmental process of least concern to us, because it has the least obvious parallel, is morphogenesis or the creation of form. This term is used to describe the cell and tissue movements that shape the organism. However, leukocytes are unique

in that cells operate more or less as individual, mobile units. As such, comparisons made of lymphocyte movements to the behavior of tissues would have to be made at an abstract level. However, lymphocytes clearly demonstrate coordinated behaviors such as the orderly trafficking of immature T lymphocytes through thymic structures during development or the aggregation of mature B cells in follicles in peripheral lymphatics. Although the mechanical and physical components of morphogenesis are not as directly applicable to lymphocytes, the chemically directed components of this process are more closely shared. Despite the importance of morphogenesis in embryonic development, morphogenesis is largely the consequence of cellular differentiation.

### 1.3 Cellular differentiation is a developmental hierarchy

In embryogenesis, the parts to be developed in the basic body plan are not specified all at once but are formed as a hierarchy of developmental choices. Perhaps not surprisingly, a developmental hierarchy also exists in hematopoiesis. **Figure 1** illustrates the subdivisions of the developmental choices made during hematopoiesis. Setting aside discussions about details of the diagram, the familiar cell-types the developed immune system are found at the bottom of the tree. These cell-types are preceded by a series of prior commitment choices. At each of these a choice is made between increasingly restricted alternative states arising from a subdivision of an earlier less committed progenitor. Every decision is made among a small number of alternatives and a new state of commitment is adopted with further restriction of potency. Upon reaching terminal differentiation, potential has been exhausted, and the cell persists in its final state until its death. The difficulty of deducing the arrangement of these fate decisions lies in the largely covert nature of the cell state that can only be uncovered by

experimentally probing for remaining lineage potential. Despite this difficulty, the organization of cell fate choices into a hierarchy has provided a useful framework for understanding the gene control mechanisms of differentiation, and these molecular mechanisms will be discussed below.



**Figure 1. The hierarchy of development in hematopoiesis.**

The formation of blood cells occurs by a process of progressive determination and fate restriction. The property of self-renewal is lost at the onset of differentiation, but hematopoietic progenitors initially maintain their pluripotency for all subsequent lineages. As differentiation proceeds alternative fate choices are gradually lost as cells commit to a lineage. The hematopoietic system is therefore built up as a result of a hierarchy of decisions and several fate choices will be made before the cells differentiate into the mature cell types shown at the bottom. The mechanisms underlying this process are discussed in Chapter 2. For some cell-types, such as B and T cells, further differentiation will occur after encountering antigen in an immune response and this will be discussed in Chapter 3.

#### 1.4 Differentiation is organized at the genetic level by transcription factors

The process of restrictive differentiation that occurs during both embryogenesis and hematopoiesis is the consequence of the regulatory system controlling gene expression. Differentiation is the establishment of a new regulatory state that can be thought of as the cumulative activity of particular sets of DNA-binding transcription factors coordinating their activity at non-coding DNA and determining gene expression. Regulatory states are interpreted by the DNA sequence elements composed of clusters of transcription factor binding sites called *cis*-regulatory modules (CRM) or enhancers. It is the binding of a transcription factor to a CRM that allows the transcription factor to modulate the expression of nearby target genes, and the integration of all transcription factors at the CRM produces a unique regulatory output. However, the availability of a CRM for binding by the transcription factor, and therefore the contribution of the CRM to the regulatory state, is controlled by the proteins that package DNA, collectively called chromatin. The regulatory state encoding the prior cell identity can therefore be made irrecoverable by two mechanisms: 1) transcription factors acting in the prior regulatory state either cease to be expressed or 2) the previously available CRMs are made inaccessible by the closing and compaction of chromatin. These mechanisms provide directionality to differentiation, allowing signaling inputs to activate transcription factors on the newly accessible chromatin and not at previously active chromatin. Thus, restrictive differentiation is a consequence of the interplay between transcription factors, DNA sequence elements known as CRMs, and the chromatin.

The hierarchical and ordered nature of development is also the consequence of the integration of transcription factor activity at chromatin-controlled CRMs. The

interpretation of the current regulatory state at the available CRMs results in a regulatory output that may induce new transcription factors or other regulatory genes. The addition of transcription factors to a regulatory state specifies new modes of gene regulation that can have multiple outcomes: 1) the process of differentiation continues by inducing the expression of additional lineage-related regulatory genes or transcription factors, 2) the cell is made competent for an alternative lineage by becoming receptive to new signal inputs, or 3) the developmental process ends without inducing new transcription factors and by expressing structural genes associated with terminal differentiation. Thus, the contribution of transcription factors to differentiation is inextricable from the developmental sequence because their activity is dependent on the developmental history of the cell to establish a responsive regulatory state and to prime the chromatin for action at appropriate CRMs.

However, cellular differentiation is rarely the result of the expression of a single gene. Thomas Hunt Morgan used the term “gene battery” in 1934 to refer to the functionally related effector genes that are coordinately expressed in a given cell type upon differentiation. The CRMs regulating these gene batteries share lineage-specific transcription factor binding sites, but the chromatin is often closed at these CRMs. How is the cell-type specific accessibility at these gene batteries established during development? Recent research has demonstrated that lineage-specific transcription factors known as Pioneer Transcription Factors (PTFs) or Lineage-Determining Transcription Factors (LDTFs) establish accessible chromatin at the CRMs of their corresponding lineage through positive interactions with the normally repressive chromatin (2, 3). LDTFs add new modes of regulation to the already existing regulatory state by acting on the “tabula rasa” of the chromatin to create *de novo* accessibility and

establish competent and active CRMs. Because LDTFs are not inhibited by closed chromatin they can reprogram cells or cause trans-determination when ectopically expressed, and their identification has opened the door for new therapeutic options involving cellular engineering through the manipulation of cell identity. Identifying the LDTFs acting in T cell development will be the focus of chapter 2.

Questions of mechanism have remained at the forefront of the field since the advent of molecular biology in the 1980s, and for good reason. The models of genetic control proposed by Jacob and Monod in 1961 (4) and Britten and Davidson in 1969 (5) cannot be adequately tested at the cellular level. Advanced molecular techniques such as microarrays and next generation sequencing have only intensified the investigation into the molecular workings of differentiation. However, the questions being addressed by these tools were the questions initially proposed by earlier embryologists. The basic observation of development is the cell is the fundamental unit by which biological systems are organized at the organismal level. As such, breaking down any cellular system, such as the immune system, requires an understanding of the differentiative events at the cellular level. Without this framework, making sense of the molecular mechanisms driving differentiation would not be possible. For that reason, I will discuss two additional embryological concepts, albeit familiar to an immunologist, that will be explored in chapter 3.

### 1.5 Additional embryological concepts: the fate map and clonal analysis

The first of these concepts is the fate map. Fate mapping is fundamental to embryology and just as important to immunology. Regional specification is the core problem of embryonic development and the fate map is a diagram indicating what each

specified region of the early embryo will turn into. To create a fate map, the trajectory of the cells must be tracked throughout development. For some organisms, the fate of each individual cell can be tracked with great precision whereas other organisms tolerate some degree of cell mixing between adjacent regions and thereby reducing the resolution of the fate map. As long as the random mixing of cells is minimal, the fate map tells you what organs and limbs each cell will become and allows faithful cell lineages to be constructed. To the immunologist, fate mapping can establish critical lineage relationships that are normally obscure for a number of reasons: 1) the longitudinal study of immune cell differentiation, especially those located in tissues, is often not possible, 2) a specific immune response does not develop in isolation as an embryo does but is surrounded by cells differentiating at various stages from other immune responses, and 3) the terminal fate is normally indistinguishable even while differentiating cells take different developmental trajectories to the terminal fate. The fate map can also help determine whether a cell state is stable or merely represents a temporary phenotype, and this is especially useful for studying lymphocyte differentiation. For both embryology and immunology, the cell lineage established by fate mapping places the genetic and epigenetic mechanisms of differentiation into the context of prior and future molecular events. Thus, the interpretation of experiments on developmental decisions depends on the fate map. However, fate mapping alone does not provide information on commitment as it only reveals what will become of cells if left in place.

The second concept, clonal analysis, is related to the fate map, but differs because it allows us to say something about commitment and determination. Determination means that a cell is intrinsically committed to develop into a specific cell type or a structure. To illustrate, a single cell may form cell type A or cell type B, or

neither, but the undifferentiated cell could not have been determined if the clone develops into both cell types. An important consideration to make concerns the inverse of this principle: if the cell differentiates into one cell type, but not the other, then this clonal restriction does not mean determination has occurred. However, not all developmentally significant forms of clonal restriction need occur through an internal, genetic mechanism of determination. Clonal restriction can also occur by the existence of an extracellular barrier limiting mixing of unrestricted cells and thereby maintaining stable and distinct cell identities. Moreover, fate mapping is often done in the embryo by labelling groups of cells, but the analysis of a clone provides information about the individual behavior, the resulting lineages, and the contribution of a cell to a particular structure. This information is particularly important for studying lymphocyte immune responses as they are typically polyclonal, and clonal analysis can distinguish which clones contribute to certain effector or memory cell populations.

#### 1.6 Cellular differentiation is central to adaptive immune system function

Why frame the lymphocytes of the adaptive immune system immunology in terms of embryonic development? Given the importance of the immune system in protecting us from pathogens, it's no surprise the mammalian immune system relies on cellular specialization through stable cellular differentiation as the basis of its functional organization. Therefore, the study of differentiation itself and the mechanisms governing differentiation, are of utmost concern to an immunologist. As encapsulated in the clonal selection theory (6), the differentiation of lymphocytes occurs in two phases. The first phase is the establishment of a preimmune pool of T and B cells bearing clonally distributed receptors for antigen. At homeostasis, prior to antigen or pathogen exposure,

the cells of the immune system develop from the progressive determination of post-embryonic stem cells according to control principles similar to those in operation during organismal development. The development of the embryo is a self-directed, hierarchical process, and similarly the immune system differentiates according to an internally regulated and genetically encoded developmental hierarchy. As such, both developmental programs are conducted without much regard to the outside world. The precise gene regulation necessary for this process of differentiation is supported by the regulated packaging and post-translational modification of chromatin and histones directed by lineage-specific transcription factors. In chapter 2, I explore the lineage-specific transcription factors guiding chromatin accessibility and chromatin modifications in the context of the developmental sequence that transforms uncommitted hematopoietic progenitors to mature, preimmune T cells.

In the second phase of lymphocyte differentiation, sufficient antigen receptor occupancy causes the activation of the lymphocyte, initiates mitosis, and primes the cell to receive differentiative signals. The response is not random; activated lymphocytes differentiate according to the circumstances of the infectious challenge and the type of pathogen encountered. Although stereotyped responses exist, new subsets of cells are still discovered as techniques are developed that afford more precise resolution. Accordingly, chapter 3 addresses the extent to which lymphocytes, particularly memory B cells, develop into multiple subsets in response to infection, and interrogates the differentiative relationship between these subsets.

The concepts discussed in this overview have been adapted from Jonathan Slack's *From Egg to Embryo* (7) and Eric Davidson's *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (8).

## CHAPTER 2: TCF-1 ESTABLISHES T CELL EPIGENETIC IDENTITY

### 2.1 PREFACE

The manuscript presented in this chapter was originally published in *Immunity* (9). It has been reformatted here in accordance with the University of Pennsylvania dissertation formatting guidelines and the introduction has been expanded.

Authors:

John L. Johnson<sup>\*1,2,3</sup>, George Georgakilas<sup>\*1,2,3</sup>, Jelena Petrovic<sup>\*4</sup>, Makoto Kurachi<sup>2,5</sup>, Christelle Harly<sup>6</sup>, Warren Pear<sup>2,4</sup>, Avinash Bhandoola<sup>6</sup>, E. John Wherry<sup>2,5</sup>, Golnaz Vahedi<sup>1,2,3</sup>

Affiliations:

<sup>1</sup>Department of Genetics, <sup>2</sup>Institute for Immunology, <sup>3</sup>Epigenetics Institute, <sup>4</sup>Department of Pathology and Laboratory Medicine, <sup>5</sup>Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; <sup>6</sup>National Cancer Institute, National Institutes of Health, Bethesda, Maryland, 20184

*\*Co-first authors*

Respective Contributions:

All authors contributed extensively to the work presented in this paper. J.L.J. designed and conducted the experiments and helped analyze data. G.G. performed computational analysis, wrote the code and analyzed the data. J.P. performed ChIP-seq. M.K. designed and validated the RV constructs. C.H. provided mice. W.P., A.B., and E.J.W provided technical support, reagents, and conceptual advice. G.V. conceived the project, administered the experiments and analyses, and wrote the manuscript.

## 2.2 INTRODUCTION

### 2.2.1 The diversity of blood cells develops from hematopoietic progenitors

Eukaryotic organisms express genes in incredibly diverse patterns that are necessary for biological complexity (10). This transcriptional diversity is largely controlled by the interactions between transcription factors and their cognate DNA binding sites within accessible chromatin regions. However, eukaryotic genomes are compacted to fit over a meter of DNA within the limited volume of the nucleus and this compaction is inherently repressive to processes that require access to the DNA sequence (11). Despite the inherently repressive state of chromatin, a number of lineage-instructive transcription factors alone or in cooperation with their partners can access a subset of their binding sites even if it is partially occluded by nucleosomes, recruiting chromatin-remodeling enzymes and exposing the underlying DNA. The distinctive collection of such accessible sequences controls the transcriptional output of a cell type and determines its functional characteristics (12).

Hematopoiesis is an excellent system for studying lineage-instructive transcription factors and their roles in establishing chromatin accessibility as the differentiation of the diverse and well-defined cell-types of the blood is continuous throughout life. Hematopoiesis originates primarily in the fetal liver but is relocated to the bone marrow at birth, serving as the source of hematopoiesis through adult life (13). The process of hematopoiesis begins with the hematopoietic stem cell (HSC) (14) that is subdivided by lifespan into two major cell subsets. Long-lived HSCs (LT-HSCs) are self-renewing and can generate all the major lineages of the blood through asymmetric division, whereas short-term HSCs (ST-HSCs) generate all blood lineages but self-

renew only for approximately eight weeks (15). A mixture of LT-HSCs and ST-HSCs can be distinguished by the absence of surface markers expressed on mature cell types (Lineage, or Lin), and the presence of Sca-1 and the cytokine receptor Kit (LSK, Lin<sup>-</sup> Sca1<sup>+</sup>Kit<sup>+</sup>). Further, the SLAM marker CD150 more definitively identifies LT-HSCs within the LSK compartment (16).

As in other developmental systems, hematopoiesis follows a developmental hierarchy in which lineage potential and self-renewal is lost as cells differentiate. Thus, LT-HSCs give rise to ST-HSCs with diminished capacity for self-renewal and ST-HSCs in turn give rise to multipotent progenitors (MPPs) which have no capacity for self-renewal but are multipotent (17). At this point, a fate choice is available, and the cell either continues differentiation towards an erythrocyte lineage or the myeloid and lymphoid lineage depending on cytokine cues. Cell that lose megakaryocyte potential but retain lymphoid and myeloid potential are thus named lymphoid-primed MPPs (LMPPs) (18). The LMPP subset expresses Flt3 and gives rise to the common lymphoid progenitor (CLP) that was originally thought to be lymphoid restricted, but actually retains significant myeloid cell potential (19). The cell surface marker Ly6D further differentiates CLPs into those with restricted potential for the B lineage (LyD<sup>+</sup> CLPs) versus those with T, NK, B, and DC potential (20). However, many developmental decisions at this early stage are still plastic and not yet fully determined. For instance, a subset of common myeloid progenitors (CMP) expressing Flt3, but lacking CD150 expression possesses T cell potential that can be revealed using single cell assays *in vitro* under T-inductive conditions (21). Despite possessing this T cell potential, this subset does not have the ability to home to the thymus (21), the site of T cell specification and development (22, 23). This clonal analysis is a good illustration that an

apparent clonal restriction *in vivo* does not necessarily coincide with developmental restriction as potential is determined by both intrinsic genetic mechanisms and external instructive mechanisms. Therefore, progenitors must possess T lineage potential as well as the ability to reach the organ of T cell specification to develop into T cells.

### 2.2.2 T cell development occurs in the thymus

T cell development is unique among the blood lineages because it is completed in a specialized organ, the thymus, and homing to the thymus from blood-mobilized bone marrow progenitors is necessary for T lineage specification (24). T cell development begins when these rare progenitors settle the thymus, but very few are estimated to reach the thymus each day (25). However, migration to the thymus appears to be a regulated process and not a stochastic one as not all bone marrow progenitors are equally capable of thymic settling. For instance, HSCs do not settle the thymus yet hold profound T cell potential. Conversely, LMPPs and CLPs both settle the thymus and possess T cell potential (26). However, it is unclear which, if either, are the true progenitors. Unlike HSCs, both subsets express the chemokine receptors C-chemokine receptor-7 (CCR7) and C-chemokine receptor-9 (CCR9) which are essential for thymic settling (27, 28). P-selectin glycoprotein 1 (PSGL) is also necessary for thymic homing and promotes homing when the thymic niche empties (29, 30).

The inception of T-lineage cells occurs when bone marrow-derived multipotent precursors seed the thymus and give rise to early thymic progenitors (ETP) (31, 32). ETPs reside within the CD4<sup>-</sup>CD8<sup>-</sup>CD44<sup>+</sup>CD25<sup>-</sup>Kit<sup>hi</sup> double negative 1 (DN1) subset and are defined by the absence of lineage-associated markers and CD25 while also expressing Kit and CD44 (Lin<sup>lo</sup>Kit<sup>hi</sup>CD25<sup>-</sup>). ETPs are rare, making up 0.01% of

thymocytes, but can generate all thymic lymphoid populations. However, ETPs are not determined or developmentally restricted to the T lineage. ETPs must pass through additional stages of differentiation before becoming mature T cells and they retain differentiative potential for alternative fates, including NK and myeloid cells, which is revealed when ETPs are removed from T-inductive signals (33-36).

The potential to develop into alternative fates is gradually lost as T cell differentiation proceeds, and the process of development is associated with an orderly trafficking through anatomic structures of the thymus. ETPs migrate from the corticomedullary junction in the perimedullary cortex to the inner cortex while differentiating into DN2 cells (CD44<sup>+</sup>Kit<sup>+</sup>CD25<sup>+</sup>) (37). However, DN2 cells are heterogeneous as was first demonstrated by examining the expression of the lymphocyte-restricted kinase Lck using a GFP reporter (38). Lck<sup>-</sup> DN2 cells retain DC and NK lineage potential that is revealed after removing Notch signals whereas Lck<sup>+</sup> DN2 cells do not, even if removed from Notch signals. Thus, a key regulatory event seems to occur in the DN2 compartment. DN2 heterogeneity is further refined by using the level of Kit expression with DN2a cells being Kit<sup>hi</sup> and retaining alternative lineage potential whereas DN2b cells are Kit<sup>int</sup> and are firmly committed to T cell development (39). Further, the regulatory event occurring between DN2a and DN2b that enforces lineage commitment is the initiation of Bcl11b expression through the combinatorial activity of transcription factors at the *Bcl11b* locus (40).

DN2b cells then become DN3 cells, move to the subcapsular zone, and undergo rearrangement of the T cell receptor (TCR) at the  $\beta$ ,  $\gamma$ , and  $\delta$  loci as these loci become accessible and the recombination activating gene 1 (RAG1) and RAG2 are expressed (41, 42). A small number of developing T cells will rearrange the  $\gamma$  and  $\delta$  loci and

become  $\gamma\delta$  T cells, but most DN3 cells (>95%) will rearrange the  $\beta$  locus and continue development to become conventional  $\alpha\beta$  T cells (43, 44). Cells rearranging the  $\beta$  locus are held at the  $\beta$  selection checkpoint until a productive, in-frame, recombination of the  $\beta$  locus leads to expression of the  $\beta$  chain of the TCR that pairs with a surrogate  $\alpha$  chain (pre-T- $\alpha$ ) for trafficking to the cell surface (45). The signaling accompanying this process causes DN3 cells to downregulate RAG expression and enforce allelic exclusion of the other  $\beta$  chain allele (46). DN3 cells passing  $\beta$  selection differentiate into DN4 cells and undergo a burst of proliferation that replicates the successful  $\beta$ -chain rearrangement into multiple daughter cells before the TCR co-receptors CD4 and CD8 are upregulated and the DN4 cell becomes a CD4<sup>+</sup>CD8<sup>+</sup> double positive (DP) cell and migrating back to the cortex (47). The TCRs are tested for MHC binding in a process of positive selection wherein DP cells that bind MHC II become CD4 single positive cells and MHC II restricted whereas DP cells that bind MHC I become CD8 single positive cells and MHC I restricted (48-50). Cells that fail to bind MHC will fail to receive survival signals from the TCR and therefore 'die by neglect.' However, binding that is too strong to MHC will cause the cell to undergo negative selection and undergo apoptosis or possibly differentiate into regulatory T cells if MHC II restricted (51). Finally, T cell development ends in the medulla and fully competent mature T cells emigrate from the thymus by entering the blood stream (52).

### 2.2.3 Transcription factors acting in T cell development

Notch1, an evolutionarily conserved signaling pathway involved in multiple developmental processes, initiates the T cell fate decision when Notch1 on the surface of bone marrow progenitors settling the thymus interacts with Delta-like Notch ligands on

the epithelial cells in the thymus (53). Mammals have four Notch homologues, Notch 1-4, and two families of ligands, Delta-like and Jagged, but the engagement of Notch1 with Delta-like ligands in the thymus is critical for T cell development. DL-1 and DL-4 both induce Notch1 signals and induce T cell development, but DL-4 is the most abundant ligand in the thymus (54). Notch receptors that engage ligand undergo cleavage events mediated by a metalloprotease and  $\gamma$ -secretase to release the intracellular domain of Notch1 (ICN) for translocation to the nucleus. ICN binds the transcription factor CSL (CBF1/Su(H)LAG1), also known as RBPJ (recombination signal combining protein-J), and recruits mastermind-like (MAML) protein to act as a scaffold for the binding of co-activators such as p300 and the activation of the T cell gene program (55, 56). Notch signals and transcriptional activation by ICN is crucial for generating ETPs and ectopic expression of ICN is sufficient to induce T cell development in bone marrow progenitors outside the thymus but is also oncogenic (57, 58). Notch1 also represses the development of alternative fates. For instance, Notch1 inhibits B cell development in the BM when ICN is ectopically expressed in bone marrow progenitors (59). Removing ETP or DN2a cells from Notch1 signals allows non-T cell fates to develop indicating that alternative fates are suppressed by Notch1 signals in early T cell progenitors (60).

The Ets family transcription factor PU.1 is a winged helix-turn-helix transcription factor critical for establishing a gene regulatory program in bone marrow progenitors conducive for T cell specification (61). The role of PU.1 in B and myeloid cell development is well-established, and both lineages highly express PU.1 (62). However, PU.1 is also expressed in early T cell progenitors in the DN stages but must decline for T cell commitment to proceed (63). The dose of PU.1 in T cell progenitors strongly influences fate commitment, and enforced PU.1 expression blocks commitment and

diverts DN3 thymocytes to the myeloid lineage, but only when Notch1 signaling is abrogated (63). Thus, while necessary for bone marrow progenitors to initiate T cell specification, PU.1 must be precisely regulated to remove undue influence on pre-T cells to develop into alternative fates (64). As such, PU.1 is downregulated at the DN2a and DN2b transition, and TCF-1 and RUNX1 are important for this repression (65, 66). However, the mechanisms of the timing of this downregulation are unclear as TCF-1 and RUNX1 are both expressed highly in ETPs.

Other ubiquitous transcription factors, including E proteins, play an important role in early T cell development (67). Members of the basic helix-loop-helix family (bHLH), E proteins regulate transcription in many hematopoietic lineages and are suppressed by ID proteins that dimerize with E proteins but lack a DNA-binding domain. Two splice variants of the E2A gene, E12 and E47, are encoded by the *Tcf2a* locus and are critical for early B and T cell development (68). E2A regulates Notch1 transcription in early T cell progenitors and synergizes with Notch to regulate key T cell genes (69). Another bHLH member, HEB, is highly expressed in the thymus, but HEB deficiency manifests at a later block in T cell development compared to E47 deficiency (70). ID3, an Id family member, is induced by pre-TCR signals and inhibits E protein activity and Notch1 transcription (71). As such, E2A binding drops dramatically in  $\beta$ -selected thymocytes.

Zinc finger transcription factors, such as GATA3, also play important roles in T development (72). GATA3-deficient hematopoietic progenitors generate normal numbers of LMPPs and CLPs but show a deficiency in early T cell development (73). Conversely, GATA3 does not enhance T cell development when overexpressed in bone marrow progenitors unlike Notch and other key T cell transcription factors. Similar to PU.1, GATA3 levels must be precisely controlled as overexpression of GATA3 diverts cells to

the mast cell lineage when Notch signals are removed (74). Another zinc finger transcription factor, Bcl11b, is likewise important for T-lineage commitment. Bcl11b is expressed in DN2a cells at the point of T cell commitment and deletion of Bcl11b in DN T cells causes an incomplete developmental block at the DN2a stage (75, 76). Importantly, stem cell genes and alternative lineage fates such as myeloid and NK cells increase when Bcl11b is deleted in DN2a progenitors (77). Deletion of Bcl11b later, in DN3 and DP stages, causes diversion of T cells to the NK fate while also losing the T cell gene program (77). Together, this indicates that the role Bcl11b plays in T cell commitment is primarily in repressing alternative lineages and maintaining T cell identity in a manner similar to PAX5 in the B lineage (78).

#### 2.2.4 HMG Box transcription factors: TCF-1 and LEF-1

The High Mobility Group (HMG) Box transcription factors are a superfamily dating back 1 billion years and can be divided into two major groups based on sequence-dependent and sequence-independent DNA recognition, including the TCF/SOX and HMG/UBF families, respectively (79). Members of the TCF/SOX family have one HMG Box binding domain that binds the (A/T)(A/T)CAAAG motif whereas HMG/UBF members contain multiple HMG Box domains and bind DNA non-specifically (80). HMG Box domains bind unwound non-B-type DNA and alter DNA architecture by inducing bends in the DNA backbone upon binding (81). This distortion of DNA is the result of DNA contacts made between the HMG Box and the minor groove of the DNA helix. HMG Box containing transcription factors have been proposed to play an architectural role, bending the DNA backbone to allow binding of other transcriptional regulators to the DNA and the formation of nucleoprotein complexes and occurs, for

example, at the TCR $\alpha$  enhancer (82).

The two lymphoid-specific HMG Box transcription factors are T cell factor 1 (TCF-1) and lymphoid-enhancing factor-1 (LEF-1). TCF-1 is encoded by the gene *Tcf7* and was identified in the early 1990s by Hans Clever who cloned TCF-1 from a human T cell line as a transcription factor bound to the CD3 $\epsilon$  enhancer (83). LEF-1 was also identified in 1990 soon after at the human immunodeficiency virus (HIV) promoter and was originally named TCF-1 $\alpha$  while the mouse homologue was denoted as LEF-1 (84). TCF-1 and LEF-1 are highly conserved in the HMG Box domain and share 98% similarity and are believed to have arisen through gene duplication (85). High level of TCF-1 expression is restricted to the T-lineage and LEF-1 is B and T-specific (86). Genetic deletion of TCF-1 results in dramatically reduced thymic cellularity and multiple blocks in T cell development (87). T cell development in LEF-1 deficient mice is largely normal, but exhibit defects in B cell development (88). Moreover, LEF-1 may compensate for TCF-1 as the genetic deficiency of both transcription factors causes an absolute block in T cell development (89).

Canonically, TCF-1 and LEF-1 are the transcriptional effectors of the evolutionarily conserved Wnt signaling pathway (90). The Wnt pathway regulates numerous developmental systems including embryonic patterning, cell-fate decisions, and tissue homeostasis (91). Wnt proteins are released and bind to Frizzled/low density lipoprotein receptor related protein complex on the cell surface of target cells (92). In the target cell, the transcriptional regulator  $\beta$ -catenin is maintained at low levels in the cytoplasm through constant degradation. Degradation is lifted when Wnt signals are received by the complex containing glycogen synthase kinase-3 $\beta$  (GSK-3), Axin, Adenomatous Polyposis Coli. The accumulation of  $\beta$ -catenin allows for nuclear

translocation and binding to TCF-1 and LEF-1 to activate transcription. In the absence of nuclear  $\beta$ -catenin, TCF-1 and LEF-1 are believed to be transcriptionally repressive by associating with the Groucho/TLE family of repressors (93). Thus,  $\beta$ -catenin displaces Groucho/TLE and converts TCF-1 and LEF-1 to transcriptional activators (94). However, loss-of-function studies for  $\beta$ -catenin does not diminish T cell development despite the clear role of TCF-1 in activating target T cell genes (95, 96).

Finally, there are multiple isoforms for TCF-1 including multiple splice variants of TCF-1 ranging from 25-55 kD. Further, TCF-1 has two promoters, 1 kb apart, that drive transcription of two major isoforms (97). The upstream promoter transcribes the 'full-length' TCF-1 isoform, called p45, and the downstream promoter transcribes the shorter, p33, isoform of TCF-1. The 5' coding region of the full-length isoform encodes the  $\beta$ -catenin interaction domain that the p33 isoform lacks. Thus, the p33 isoform is speculated to be the repressive isoform of TCF-1. However, the mechanisms governing the abundance of each isoforms is unclear.

The distinct phases of T cell development in the thymus are controlled by the upregulation of transcription factors including TCF-1, GATA3, and Bcl11b as well as the repression of alternative-lineage factors such as PU.1 and Bcl11a. Notch signaling is indispensable in driving specification but is transient and only active up through the  $\beta$ -selection checkpoint. The earliest T cell-specific transcription factor is TCF-1, encoded by *Tcf7*, which is steeply upregulated in T cell progenitors by Notch1 signaling and sustained until maturation. TCF-1 can positively regulate GATA3 in addition to Bcl11b, which is necessary for T lineage commitment (40, 75). Transcription factors required in other hematopoietic differentiation programs such as E2A and its relatives, Ikaros, Gfi1, Myb, and RUNX1 are also essential in T cell development (98, 99). Despite the broad

knowledge on the functions of these transcription factors at distinct developmental stages, it remains unclear which ones shape the chromatin accessibility of mature T cells in the thymus.

Numerous studies in macrophages and B cells illustrate the emergence of accessible chromatin commanded by lineage-determining transcription factors (12, 100-106). The pervasive patterns of PU.1 binding to thousands of genomic regions are closely related to the permissive chromatin state in macrophages (101). EBF1 can induce lineage-specific chromatin accessibility in B cell progenitors (103, 104). In addition to instructing development, transcription factors can also play key roles in cell reprogramming. For example, C/EBP $\alpha$  can induce transdifferentiation of B cells into macrophages at high efficiency by activating regulatory elements of macrophages (105).

Despite numerous studies of CD4<sup>+</sup> T helper cell differentiation (107-111) and CD8<sup>+</sup> T effector responses (112-117), and many reports on the dynamics of histone modifications during T cell development (118-120), we have a limited understanding of transcription factors shaping the chromatin accessibility of mature T cells in the thymus.

### 2.2.5 Scope

By mapping chromatin accessibility at eight stages of thymic T cell development in mice, we found the significant enrichment of TCF-1 motif and binding events at genomic regions that become accessible at the earliest stage of development and persist until T cell maturation. While TCF-1-deficient mice show a severe reduction in thymocyte numbers (87) and ectopic TCF-1 in bone marrow progenitors can drive the expression of T-lineage genes (121), the mechanism through which TCF-1 controls T

cell identity remains unclear. Some T-like cells continue to develop in the absence of TCF-1 although they are functionally limited in terms of differentiation, and memory T cells lacking TCF-1 are also defective (122). In line with these studies, we found that T-like cells in TCF-1-deficient mice cannot establish the open chromatin landscape and transcriptional profile of normal T cells. Moreover, TCF-1, but not RUNX1 or GATA3, could dictate a coordinate chromatin opening in single cells that follow a T cell trajectory amongst a vast landscape of possible states, suggesting a unique property for this lineage-determining transcription factor. Gain of function experiments in fibroblasts further revealed the ability of TCF-1 to bind to previously occupied nucleosomes, generating *de novo* chromatin accessibility even at condensed chromatin regions and inducing the expression of T cell-restricted genes ordinarily silenced in fibroblasts. Remarkably, a subset of these binding events further erased the pre-existing repressive marks in fibroblasts, highlighting the ability of this transcription factor to substantially target closed chromatin. Collectively, our results identify the role of TCF-1 in the making of chromatin accessibility at T cell genes and reveal an unprecedented means through which this protein controls the epigenetic identity of T cells.

## 2.3 RESULTS

### 2.3.1 Three waves of chromatin remodeling during T cell development

To elucidate the developmental stages in which the open chromatin landscapes of mature CD4<sup>+</sup> and CD8<sup>+</sup> T cells are established in the thymus, we assessed chromatin accessibility at eight stages of development including ETP (also referred to as DN1), DN2a, DN2b, DN3, DN4, DP, CD4<sup>+</sup>, and CD8<sup>+</sup> T cells using ATAC-seq (STAR Method).

To identify T cell-specific regulatory elements, we compared these maps with those of progenitor cells including hematopoietic stem cells (HSC), multipotent progenitors (MPP), and common lymphoid progenitors (CLP) in addition to B and NK cells. Initial steps of the analysis led to the characterization of 35,869 open chromatin regions with differential accessibility levels across cell states. Our unsupervised clustering of these regulatory elements revealed patterns of gain and loss of chromatin accessibility as cells progress from early to terminal stages of T cell fate determination (**Figures 2A and S1A-B**). We aggregated patterns of gain and loss in chromatin accessibility into broader meta-clusters capturing selective opening in early, intermediate, and late phases of development. Our data show that the sustained accessibility of mature T cells is established in three distinct waves: “early” at ETP (1,705 regulatory elements, cluster 9), “intermediate” after commitment at DN2b (1,399 regulatory elements, cluster 19), and “late” at the single-positive stage (1,917 regulatory elements, cluster 10) (**Figure 2 A-B**). In addition, a set of genomic regions that became open early was shared between T and NK cells (1,445 regulatory elements, cluster 7). Notably, our data revealed a pattern of gain and loss of accessibility as 75% (9,071) of regulatory elements that became accessible at the early ETP stage were dismantled before T cell maturation (“Open Early in T” meta-cluster, **Figure 2A**). These results demonstrate an unexpected dynamic in the remodeling of the regulatory landscape with distinct expansions and restrictions of regulatory elements during T cell development.

### 2.3.2 TCF-1 is the top enriched transcription factor in mature T cell clusters

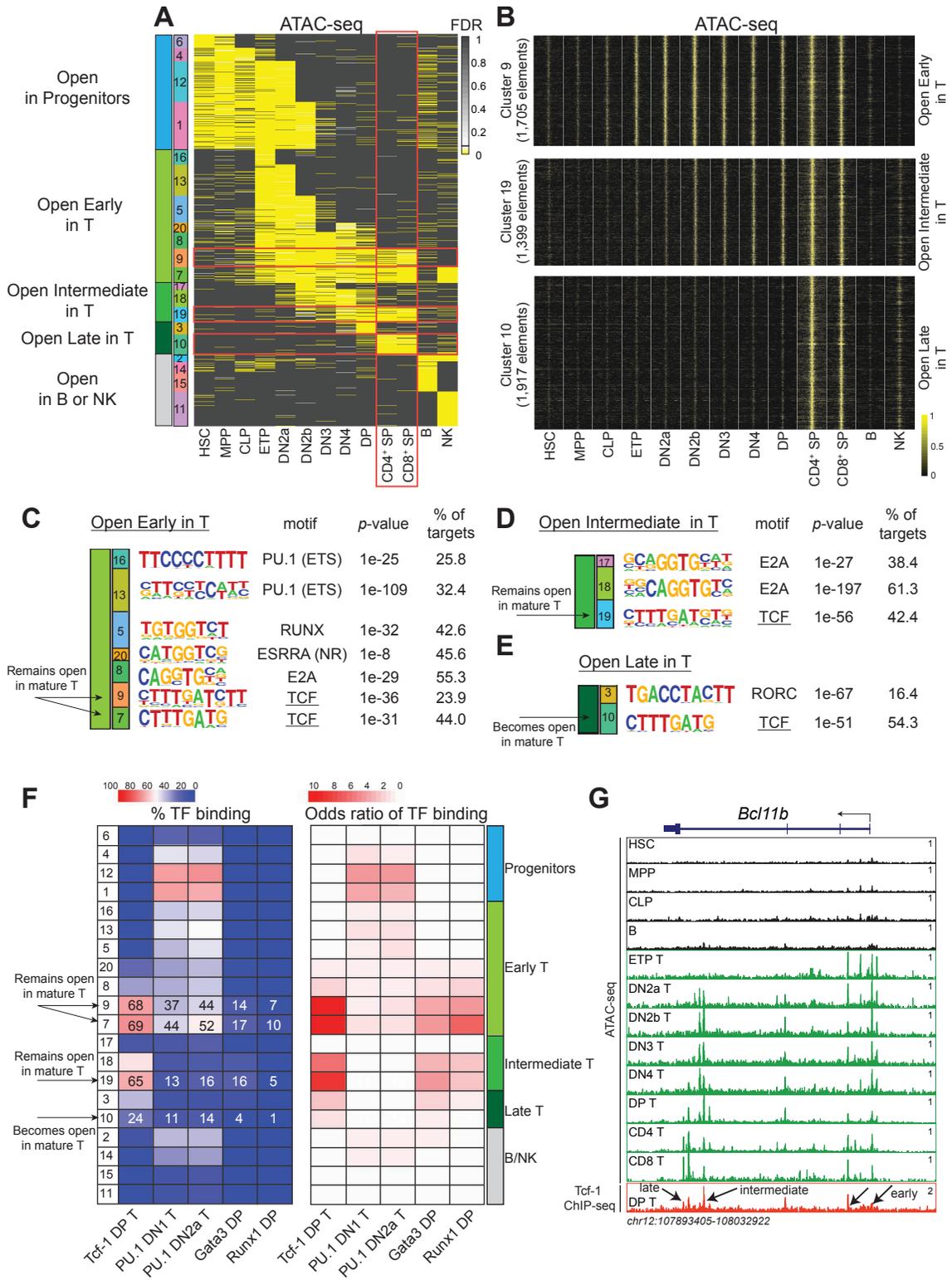
We reasoned that the transcription factors that can bind to nucleosomal DNA in progenitors and create the chromatin accessibility landscape of terminally differentiated

cells should be enriched within regulatory elements that selectively become open in that lineage. To find transcription factors with such characteristics, we inferred their occupancy in cell and stage-specific regulatory elements by performing motif analysis (106). B-cell specific open chromatin regions were enriched with motifs of EBF1, a transcription factor which has been previously reported to create the accessibility of regulatory elements in B cells (**Figure S1C**) (103, 104). Furthermore, Tbox, ETS, and GATA motifs were highly enriched among regulatory elements of NK and progenitor-specific cells (**Figure S1C**). In T cells, recognition sites for TCF, a high-mobility group (HMG) family of proteins, were the top enriched motif in the early, intermediate, and late waves of chromatin opening that persisted until T cell maturation (clusters 9, 19, 7, and 10) (**Figure 2C-E**). Notably, E2A, ETS and Runx recognition sites were among the second and third motifs in these clusters (**Figure S1C**). Similar analysis on chromatin accessibility maps of human T cells revealed the enrichment of TCF motifs within T cell-specific open chromatin of human naïve T cells, suggesting the conserved role of this transcription factor in humans and mice (**Figure S1D**).

Among TCF family transcription factors, TCF-1 is induced early at the inception of T lineage cells. To further substantiate direct binding of TCF-1 in comparison to other T cell related transcription factors including GATA3, RUNX1, and PU.1, we calculated the number of genomic regions within each cluster bound by these transcription factors using ChIP-seq (STAR method). As predicted by the enrichment of its motif, TCF-1 bound to around 70% of the genomic regions within the early and intermediate T cell specific clusters in addition to 24% of the late T cell cluster. This contrasts with RUNX1 and GATA3 binding events at less than 17% of the genomic regions within T cell-specific clusters (**Figure 2F**). Moreover, the highest odds ratio was associated with TCF-1

binding events in early and intermediate T cell clusters in particular clusters 9, 7, and 19 (**Figure 2F**). Of note, the early regulatory elements deactivated before maturation were enriched with PU.1 binding, reminiscent of earlier findings that most active chromatin features at PU.1 binding events are 'dismantled' as PU.1 is downregulated in early DN stages (120). Together, the pervasive binding of TCF-1 corroborates the strong enrichment of TCF motifs at accessible regulatory elements of mature T cells.

We further sought to explore the relationship between the activation of regulatory elements and their associated genes. The ontology of genes proximal to T cell-specific clusters was mostly related to T cell receptor signaling and naïve T cell development with no ontology distinguishing different waves of chromatin opening (**Figure S1E**). The gene expression levels proximal to dynamic regulatory elements did not present significant differences during development, suggesting a larger transformation for the regulatory landscape than the transcriptional output (**Figure S1F**). While the T cell commitment factor *Bcl11b* has low expression levels in ETP, multiple T cell-specific regulatory elements proximal to this gene became accessible at the earliest stage and co-localized with TCF-1 binding (**Figure 2G**). The three waves of chromatin remodeling during development is attested to the *Bcl11b* locus as the rightmost elements including the *Bcl11b* promoter became accessible in the ETP and were retained until T cell maturation. The middle of the locus was mostly accessible in intermediate stages, and the leftmost elements of the locus gained accessibility late in the developmental process. Collectively, these results demonstrate the dynamic of expansions and restrictions of regulatory elements during T cell development and foreshadow the importance of TCF-1 in patterning the regulatory landscape from early thymic progenitors to mature T cells.



**Figure 2. TCF-1 binding occurs at three waves of chromatin remodeling during T cell development.**

**(A)** A dynamic remodeling of the chromatin landscape with distinct expansions and restrictions of regulatory elements at early, intermediate and late stages of T cell development. Accessibility heatmap of 35,869 enhancers are measured by bulk ATAC-seq in HSC, MPP, CLP, ETP, DN2a-b, DN3, DN4, DP, SP, B and NK cells. Rows represent regulatory loci and columns the significance of each element's accessibility level in every sample. Enhancers are organized in groups with k-means (k=20) clustering using FDR as a proxy for signal enrichment. Lower values represent higher chromatin accessibility. Number of clusters was chosen based on Average Silhouette Width statistic. Clusters were further assembled into meta-clusters depending on their accessibility pattern such as open in progenitor, early, intermediate and late in T as well as B or NK cells. Clusters that are open in mature T cells and specific to T cell development are highlighted in red.

**(B)** ATAC-seq profiles of normalized tag counts around enhancers (+/- 2kb window and 10bp bin size) in clusters 9, 19 and 10 across all 13 cell types.

**(C-E)** TCF is the top enriched motif at T cell-specific regions that become accessible at early, intermediate and late waves of gain in chromatin accessibility during T cell development. *De novo* motif discovery using HOMER in each group using remaining elements in other clusters as background unveiled putative cell-type specific transcription regulators.

**(F)** Widespread binding of TCF-1 at the open chromatin of T cells. Percentage of enhancers in each cluster that are bound by TCF-1, PU.1, GATA3 and RUNX1 ChIP-seq peaks (left) and their corresponding odds ratio (right). Contingency tables were calculated using ChIP-seq data summarized in STAR Method.

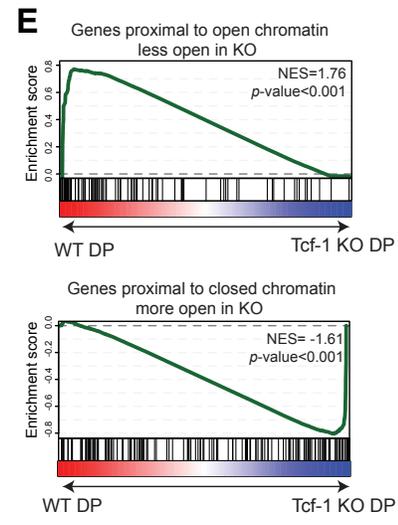
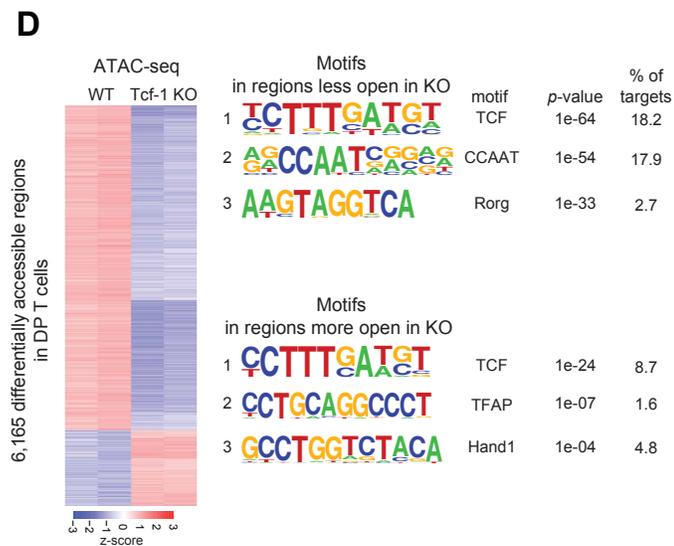
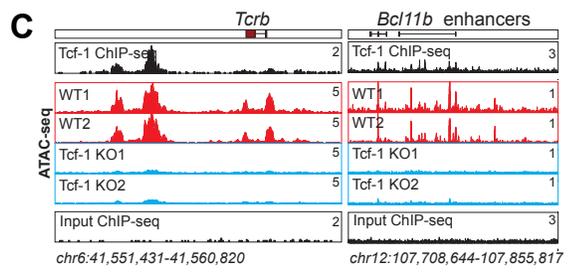
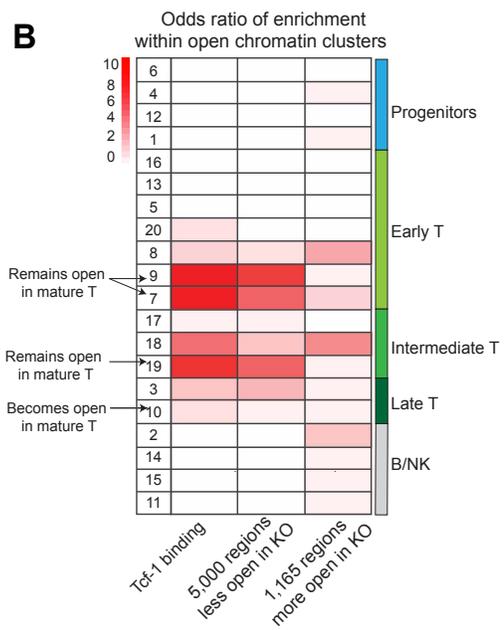
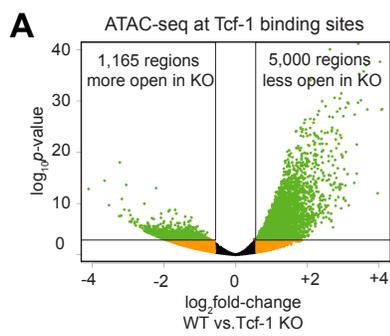
**(G)** Example of the *Bcl11b* locus and several TCF-1 bound enhancers that exhibit diverse accessibility levels during T cell development. The three waves of chromatin remodeling during development is attested to at the *Bcl11b* gene as the rightmost elements including the *Bcl11b* promoter became accessible in ETP and were retained until T cell maturation, the middle elements were mostly accessible in intermediate stages, and the leftmost elements gained accessibility late in the developmental process.

### 2.3.3 TCF-1-deficient T cells cannot establish the open chromatin landscape of normal T cells

Germline deletion of TCF-1 leads to a severe reduction in thymocyte numbers (87). Although some T lineage-like cells continue to develop in the thymus of TCF-1-deficient mice, they are functionally limited in terms of differentiation and persistence of memory T cells during infection (87, 122). It remains unclear whether the chromatin accessibility landscape and transcriptional outputs of these T-like cells is different from those of normal T cells. Therefore, we next measured chromatin accessibility at TCF-1 binding events in wildtype and TCF-1-knockout DP T cells. Our data revealed the loss of chromatin accessibility at 5,000 regulatory elements and the gain at 1,165 genomic loci in TCF-1-deficient T cells (**Figures 3A and S2A**). We sought to elucidate the relationship between regulatory elements that required TCF-1 for their accessibility and the three waves of chromatin opening during T cell development (clusters in **Figure 2A**). Regulatory elements that lost chromatin accessibility in the absence of TCF-1 were strongly enriched within early or intermediate waves of chromatin opening during T cell development, suggesting that this transcription factor is required for patterning the chromatin at early stages (clusters 7, 9 and 19) (**Figure 3B**). Examples of affected regions included the well-annotated *Tcrb* enhancer (123) and the distal *Bcl11b* enhancers (124) (**Figure 3C**). Performing *de novo* motif analysis revealed TCF as the top enriched motif in the lost sites supporting the notion that TCF-1 is directly responsible for chromatin accessibility (**Figure 3D**). TCF-1-bound regions with gains in accessibility in the knockout cells were also enriched with the TCF motif but were associated with elements accessible in B and NK cells or T cell regulatory elements deactivated in mature T cells, supporting the previously reported repressive role of TCF-1 at some genomic locations (**Figures 3B, 3D**) (125). Together, these data demonstrate

that TCF-1 is required for patterning the chromatin of T cells at early stages of development in the thymus.

To elucidate how changes in chromatin accessibility relate to the dynamics of gene expression, we evaluated the transcriptome of wildtype and TCF-1-deficient T cells using RNA-seq (**Figure S2B-C**). We then interrogated changes in the expression of genes proximal to TCF-1-dependent open chromatin regions using gene-set-enrichment analysis. Genes proximal to regions that became less accessible in the absence of TCF-1, such as *Tcrb* and *Bcl11b*, displayed reduced expression in cells lacking this transcription factor (**Figure 3E**). Conversely, genes such as *Adam19* that became more accessible also showed an increase in transcription in TCF-1 deficient T cells (**Figure S2D**). Together, these results indicate that while some T-like cells continue to develop in the absence of TCF-1 in the thymus, they cannot establish the open chromatin landscape and transcriptional profiles of normal T cells.



**Figure 3. TCF-1-deficient T cells cannot establish the open chromatin landscape and transcriptional output of normal T cells.**

(A) We generated three replicates of ATAC-seq in wildtype and TCF-1 germline deleted DP T cells and evaluated the chromatin accessibility levels at TCF-1 binding sites based on ChIP-seq. We applied variance stabilizing transformation and library size normalization on the raw ATAC-seq counts and used DESeq2 to delineate differentially accessible regions at TCF-1 binding sites(126) (fold-change > 1.5 and  $p$ -value < 1e-3). While 5,000 genomic regions were less accessible, 1,165 regions were more accessible in TCF-1 deficient T cells.

(B) Loss of TCF-1 selectively diminishes the accessibility of genomic regions that become open at early or intermediate stages of development and sustain accessibility in mature T cells. The overlapping genomic regions identified by DESeq in (A) and open chromatin clusters in **Figure 2** were found and used to measure the odds ratio.

(C) Representative examples included the well-established enhancers of *Tcrb* and *Bcl11b*.

(D) *De novo* motif analysis using HOMER unveiled TCF-1 as the most significantly enriched motif in regions that both gained and lost accessibility.

(E) TCF-1 dependent changes in gene expression correlate with changes in chromatin accessibility. We measured gene expression using RNA-seq in replicates for wildtype and TCF-1 deficient T cells. We used DESeq2 to identify differentially expressed genes (fold-change > 1.5 and  $p$ -value < 1e-3). Our analysis unveiled 1,167 down- and 1,293 up-regulated genes in TCF-1 deficient compared to wildtype T cells. Genes were ranked based on  $\log_2$  fold change estimated by DESeq2 and used as the pre-ranked gene list in GSEA analysis. The GSEA gene sets were genes within 10kb of top 200 regions with highest fold-change in chromatin accessibility between wildtype and KO cells.

#### 2.3.4 TCF-1 binding exerts a strong harmonizing impact on the chromatin of single T cells

If a transcription factor is required for patterning the regulatory landscape of a lineage, it may need to exert a harmonizing impact on the chromatin of individual cells making the same fate decision. To interrogate which T cell transcription factor may have such features, we first exploited maps of chromatin accessibility at the population level and reasoned that at a given regulatory element, the strength of bulk ATAC-seq signal reflects the fraction of cells in the population with open chromatin. We compared the normalized intensity of chromatin accessibility in bulk ATAC-seq at genomic regions uniquely bound by T lineage transcription factors TCF-1, GATA3, or RUNX1 (**Figure 4A**). Our analysis revealed that TCF-1 binding events rendered the highest average level of chromatin opening in comparison to RUNX1 and GATA3, advancing the notion that TCF-1 may unify chromatin accessibility across single T cells (**Figure 4A**).

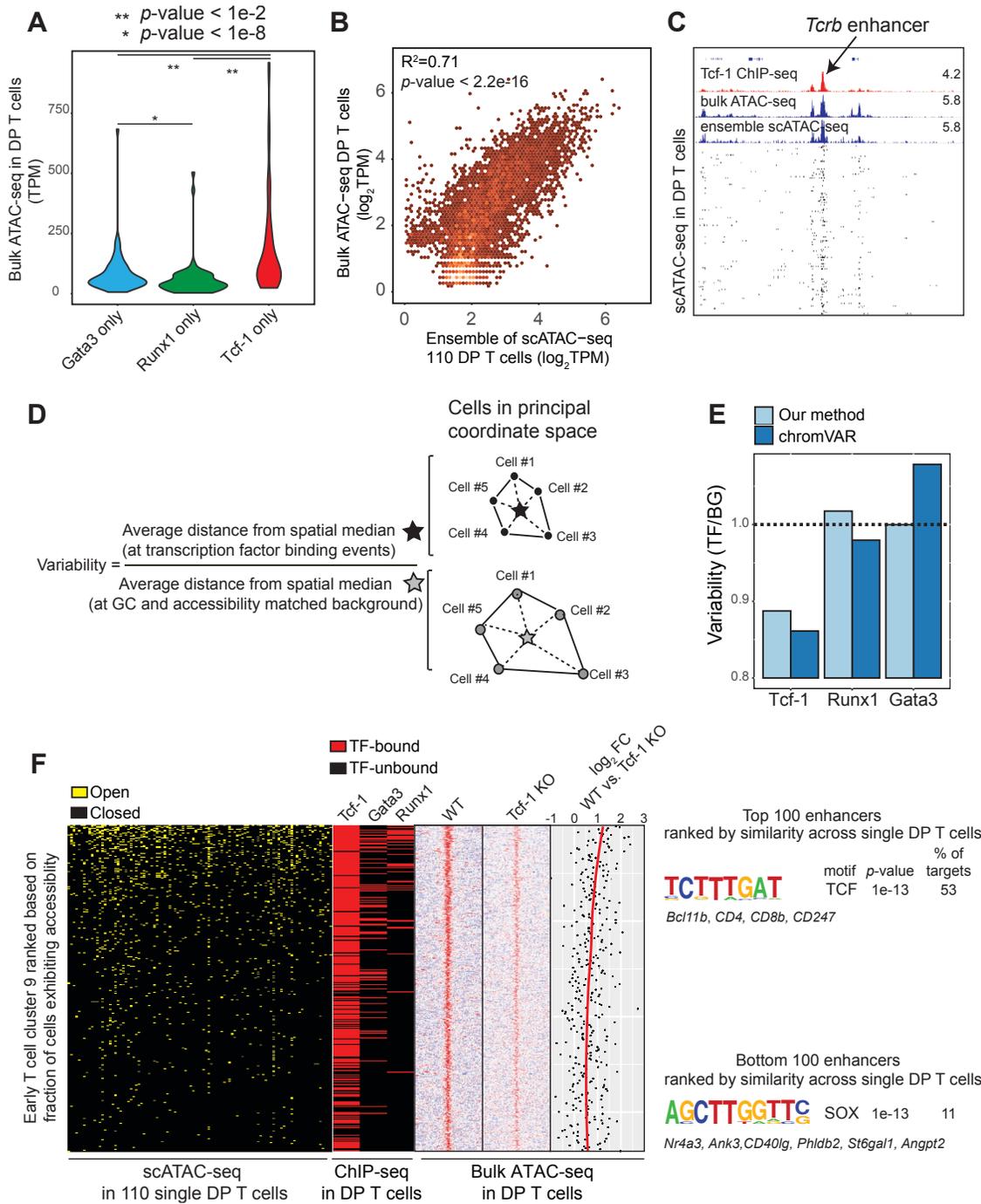
While chromatin accessibility maps of bulk T cells measure the average patterns of open regulatory elements at the population level, it remains unclear if Tn5 insertions linearly reflected the fraction of individual cells with open chromatin. To address this concern, we tested our hypothesis using single-cell (sc)ATAC-seq (127). In this approach, individual cells stained for viability were captured and assayed using a programmable microfluidics platform (Fluidigm) (**Figure S3A-B**). Collapsing reads from single T cells to aggregate scATAC-seq data closely reproduced measures of accessibility profiled by ATAC-seq generated from 50,000 T cells (**Figure 4B**). A representative genomic region such as the *Tcrb* enhancer confirmed the strong correlation between bulk and single-cell measurements (**Figure 4C**). Furthermore, data from single T cells recapitulated several characteristics of bulk ATAC-seq data, including

fragment-size periodicity corresponding to integer multiples of nucleosomes (**Figure S3C**). Together, we performed three independent single-cell captures and 110 T cells at the DP stage passed various quality control thresholds, suggesting high-confidence single-cell chromatin accessibility maps in T cells (**Figure S3D**).

Single-cell chromatin accessibility data are sparse, binary, and high dimensional, leading to unique computational challenges. To overcome these difficulties, we developed a method using a geometric distance metric and quantified cell-to-cell chromatin accessibility variation (**Figure 4D**, STAR Methods). To interrogate which T cell transcription factor can create harmonizing effects, we exploited our method on binarized scATAC-seq count data in every cell and calculated the average distance between pairs of T cells at genomic regions uniquely bound by TCF-1, RUNX1 or GATA3. We reasoned that binarizing scATAC-seq count data at transcription factor binding events reflects the openness or closeness (1 or 0) of a locus in a single cell. Due to biases in the number of observed fragment counts between cells based on the GC content or mean accessibility of a given peak set, we normalized the distance between individual cells at each set of transcription factor binding events to that of a background set comprising an equal number of peaks with matching GC content and mean accessibility. Our single-cell analysis revealed that TCF-1-bound regions were associated with the least variability among individual T cells in comparison with GATA3 and RUNX1 (**Figure 4E**). We further applied another analytical technique called “chromVAR” which was recently developed to address the same question (128). Unlike our method in which the difference in accessibility of a genomic region between every cell-pair contributes to the variability score, chromVAR relies on the aggregate of accessibility signal across a genomic set. Despite differences in the inference of

variability at transcription factor binding sites, chromVAR also identified TCF-1 as the least variable transcription factor in exerting chromatin accessibility across single T cells (**Figure 4E**). Together, two analytical strategies developed by us and others corroborate the enrichment of TCF-1 binding at regulatory elements that their accessibility is conserved across single T cells.

As an alternative strategy, we ranked T cell specific genomic regions in the early T cell cluster (cluster 9) based on the fraction of cells harboring open chromatin and evaluated whether they were bound by T cell transcription factors TCF-1, GATA3, and RUNX1 (**Figure 4F**). The top regulatory elements open across majority of single cells were bound consistently by TCF-1 in contrast with GATA3 and RUNX1 (**Figure 4F**). We reasoned if TCF-1 indeed plays a role in creating accessibility at genomic regions with the highest similarity across individual cells, its deletion should have a stronger effect on the accessibility of these regions at the bulk level. Indeed, the most similar genomic regions across individual T cells, i.e. being open at the highest fraction of cells, were more affected by loss of TCF-1 compared to the least similar genomic regions (**Figure 4F**). In line with consistent TCF-1 binding and a stronger effect size in chromatin accessibility in the absence of TCF-1, the TCF motif was selectively enriched within the top 100 most similar genomic regions. Furthermore, the genes proximal to these genomic regions with the highest similarity across individual T cells were associated with T cell biology and included T cell relevant genes such as *Bcl11b* (**Figure 4F**). Together, studying maps of chromatin accessibility at bulk and single cell levels with distinct analytical strategies suggests that TCF-1 could dictate a harmonizing impact on the chromatin of individual T cells.



**Figure 4. TCF-1 binding has the highest coordinate impact on open chromatin of single T cells.**

(A) TCF-1 binding events harbor the strongest chromatin accessibility measured by bulk ATAC-seq in DP T cells. Genome-scale binding of TCF-1, RUNX1, and GATA3 in DP T cells was measured by ChIP-seq. An equal number of genomic regions with unique binding of each transcription factor were subsampled from ChIP-seq data sets. The normalized tag count for ATAC-seq in DP T cells was calculated for each group of transcription factor binding.

(B) The aggregate maps of scATAC-seq data closely reproduced measures of accessibility profiled by ATAC-seq generated from 50,000 DP T cells. Open sites identified from bulk ATAC-seq in 50,000 DP T cells were merged with peaks characterized by aggregating the samples from 110 single DP T cells passing QC measures. Normalized enrichment was subsequently calculated in downsampled bulk ATAC-seq and aggregated scATAC-seq enabling the assessment of the correlation level between the two assays.

(C) Aggregated single cell ATAC-seq profile recapitulates chromatin accessibility on the bulk level at *Tcrb* enhancer.

(D) A novel method to infer transcription factor-associated chromatin accessibility variation across single cells.

(E) Chromatin accessibility across individual T cells is the least variable at TCF-1 binding events using our method (D) or chromVAR (128).

(F) Fraction of cells with binarized open chromatin was measured across all pairs of elements to rank regulatory elements. TCF-1, GATA3 and RUNX1 ChIP-seq enrichments were assessed in the same order as well as changes in chromatin accessibility based on bulk ATAC-seq signal in wildtype and TCF-1 KO T cells. *De novo* motif analysis using HOMER at the 100 enhancers exhibiting the highest similarity at the single cell level revealed the enrichment of TCF while TCF was not enriched at 100 least similar enhancers. T cell related genes were associated with the top enhancers.

#### 2.3.4 TCF-1 can create *de novo* chromatin accessibility in fibroblasts

It has been shown that when TCF-1 is forcibly expressed in bone marrow progenitors, it can drive the expression of T-lineage genes (121). Yet, it is not clear whether this alteration in the gene expression program of multipotent progenitors relates to the ability of TCF-1 to bind to silent chromatin and/or drive the epigenetic commitment to the T cell lineage. To examine if TCF-1 can create *de novo* open chromatin, we assessed this transcription factor in a gain-of-function model in nonhematopoietic somatic cells. We reasoned that fibroblasts could serve as an ideal model since the chromatin state in fibroblasts is distinct from cells of the hematopoietic system and T cell-specific genes are repressed in these somatic cells, allowing us to better evaluate the role of TCF-1 in targeting condensed chromatin.

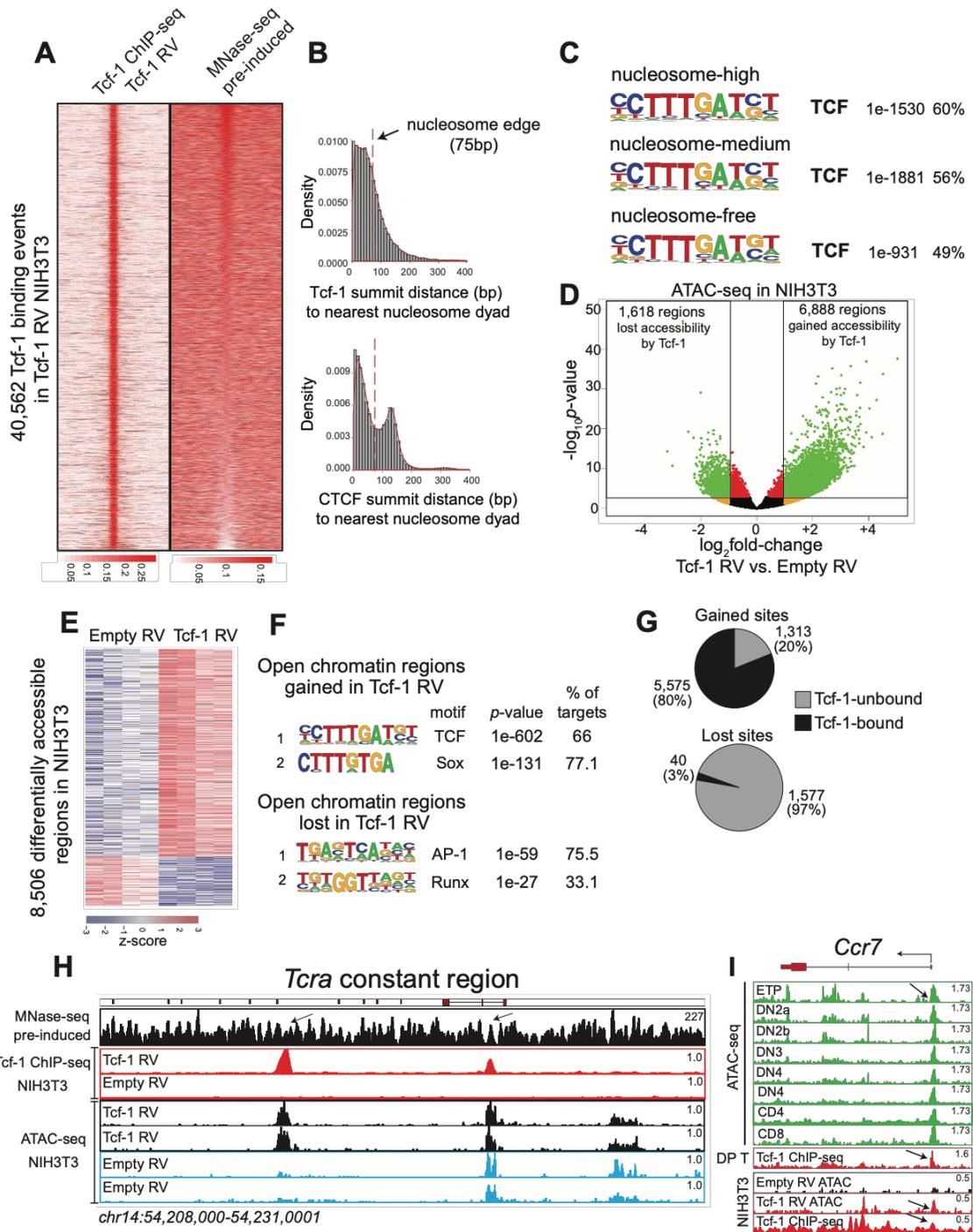
To evaluate the genome-scale binding of TCF-1, we ectopically expressed this transcription factor in a fibroblast cell line using a retroviral transduction system and performed TCF-1 ChIP-seq (**Figure S4A**). To define genome-scale TCF-1 binding events, we used the irreproducible discovery rate (IDR) method with a threshold of 2% (129) (**Figure S4B**). We further mapped the position of nucleosomes using micrococcal nuclease (MNase)-seq in pre-induced cells. The ectopic expression of TCF-1 led to more than 40,000 TCF-1 binding events across the genome of fibroblasts where 73% of these events colocalized with previously nucleosome-occupied DNA (**Figures 5A and S4C**). The extent to which TCF-1 bound to nucleosome-occupied regions in fibroblasts was comparable to reprogramming transcription factors such as Oct4 (85%), Sox2 (80%), and Klf4 (65%) (130). As an independent measure, we found that 67% of TCF-1 summits, the center of TCF-1 peak, were within 75bp of a nucleosome dyad in contrast with CTCF binding which is favored towards nucleosome-free regions, suggesting that

TCF-1 binding is selectively enriched at previously occupied nucleosomes (**Figure 5B**). Furthermore, TCF was the strongest motif within TCF-1-bound sites with different levels of nucleosome occupancy ( $p$ -value $<1e-930$ ) (**Figure 5C**). TCF recognition sites bound by TCF-1 in fibroblasts were significantly closer to the nucleosome dyads compared to random TCF sites not bound by this transcription factor, reminiscent of PU.1 binding events being shielded by nucleosomes in cells that do not express this protein (131) (**Figure S4D**). Together, the ectopic expression of TCF-1 in fibroblasts revealed the widespread binding of TCF-1 at genomic regions previously occupied by nucleosomes harboring TCF consensus binding sites.

To measure the impact of widespread TCF-1 binding on silent genomic loci, we mapped the accessibility of chromatin by ATAC-seq post transduction with Empty or TCF-1 vectors. Using differential enrichment analysis, we found that 6,888 genomic regions previously occupied by nucleosomes gained accessibility while 1,618 sites became less accessible after TCF-1 expression in fibroblasts (**Figures 5D-E, S4E**). We further performed *de novo* motif analysis and observed that more than 80% of the gained sites harbored a TCF motif while the lost sites were enriched with AP-1 and Runx family motifs (**Figure 5F**). In concordance with motif presence, 80% of the gained sites were also bound by TCF-1 based on the TCF-1 ChIP-seq while only 3% of lost sites colocalized with TCF-1 binding (**Figure 5G**), suggesting an indirect role of TCF-1 on sites losing chromatin accessibility. To infer nucleosome position and occupancy within TCF-1 binding events, we further applied the NucleoATAC algorithm (132) to our chromatin accessibility data and found 7,395 genomic regions with significant loss of nucleosomes after TCF-1 expression (**Figure S4F**). A striking example of *de novo* regulatory elements induced by TCF-1 includes the T cell receptor alpha locus where the

binding of TCF-1 at previously occupied nucleosomes led to gains in chromatin accessibility at multiple genomic regions (**Figure 5H**). Together, our data suggest that TCF-1 can bind to thousands of previously nucleosome-occupied DNA and this binding can lead to *de novo* chromatin accessibility.

We next sought to examine whether *de novo* chromatin accessibility in fibroblasts has any relevance to T cell biology. Our data revealed that TCF-1 binding events in T cells and fibroblasts are highly correlated (**Figure S4G**) and more than 800 *de novo* regulatory elements in fibroblasts (~11%) overlapped with open chromatin in T cells while only 40 regions (~0.5%) corresponded to the open chromatin in B cells (**Figure S4H**). Furthermore, the *de novo* regulatory elements in fibroblasts were selectively enriched for regions belonging to the early wave of chromatin opening during T cell development (cluster 9) (**Figure S4I**). For example, the promoter of *Ccr7*, which is among the regulatory elements that gain accessibility at the early cluster 9, is bound by TCF-1 and becomes accessible in TCF-1-expressing fibroblasts (**Figure 5I**). Together, TCF-1 can invoke a subset of T cell regulatory elements to become open in distant somatic cells like fibroblasts.



**Figure 5. TCF-1 can bind to nucleosomes and create chromatin accessibility in fibroblasts.**

(A) TCF-1 ChIP-seq TCF-1 (p33) expressing NIH3T3 using retrovirus (RV) as well as in Empty vector controls 48 hours post transduction resulting in the identification of 40,562 reproducible peaks. The region surrounding TCF-1 summits was segmented in three non-overlapping 200bp windows centered around each summit. Normalized MNase-seq enrichment was calculated for each window and summits were ordered from high to low enrichment. (B) The majority of TCF-1 binding events occur within the boundaries of DNA wrapped around nucleosome. The distance between TCF-1 as well as CTCF (serving as control) and the closest nucleosome summits were calculated as an alternative strategy of assessing the ability of TCF-1 to directly bind nucleosomes. The vertical dashed red line is set to 75bp which is typically half the size of histone octamer bound DNA denoting the edge of nucleosomes. (C) TCF-1 motif is equally preserved in genomic loci presenting nucleosome high, medium and low enrichment. K-means clustering (k=3) was applied on TCF-1 summits using the normalized MNase-seq enrichment in the three non-overlapping 200bp windows centered around each summit. We chose the open chromatin regions presenting no overlap with TCF-1 summits as background in our *de novo* motif analysis using HOMER and found TCF as the only motif significantly enriched in each cluster.

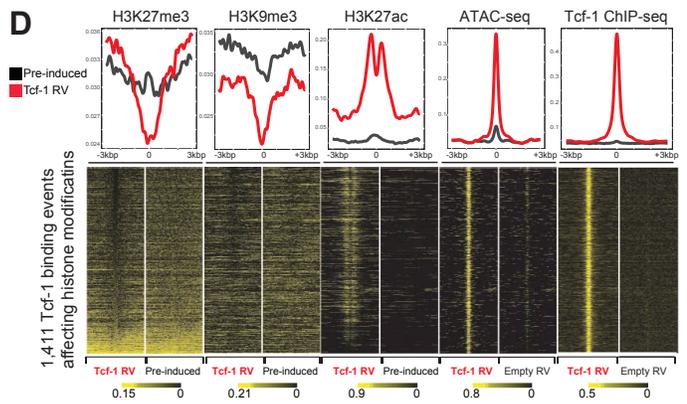
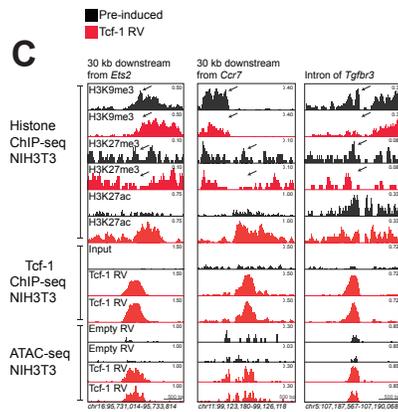
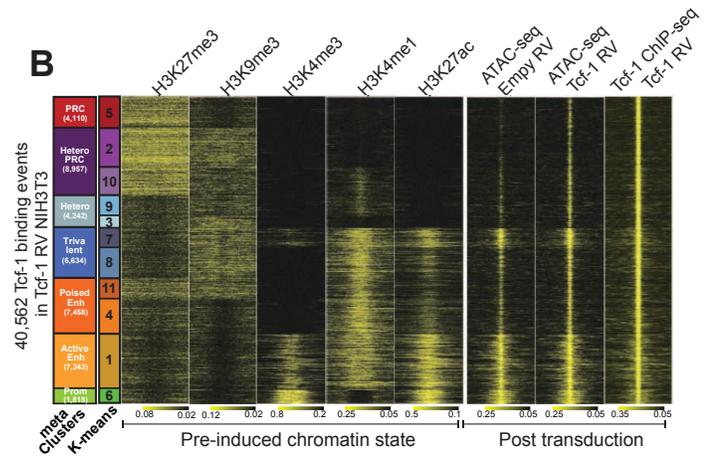
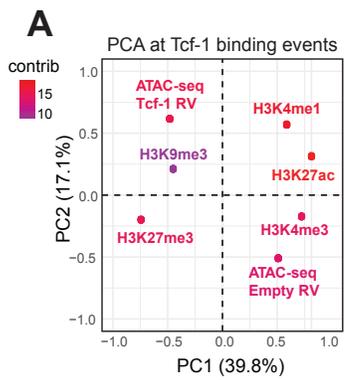
(D-I) TCF-1 can create *de novo* open chromatin in fibroblasts. (E) We performed ATAC-seq in triplicates in Empty and TCF-1 RV NIH3T3 cells. To identify differentially accessible regions, TCF-1 ChIP-seq and ATAC-seq peaks were merged to facilitate differential enrichment at both TCF-1 bound and unbound regions of the genome. We used DESeq2 and based on fold-change > 1.5 and *p*-value < 1e-3, 6,888 regions gained while 1,618 lost accessibility in TCF-1 RV cells. (F) *De novo* motif discovery unveiled TCF-1 as the most significantly enriched motif in gained sites and AP-1 as well as Runx motifs in lost sites. (G) TCF-1 bound to 5,575 (80%) gained sites in contrast to only 40 (3%) lost sites. (H) TCF-1 directly binds nucleosomes at *Tcra* related enhancers and creates *de novo* open chromatin. (I) Example of the *Ccr7* gene promoter that becomes accessible at the earliest stage of T cell developmental and is naturally bound by TCF-1 in T and Tcf1-RV NIH3T3 cells with a corresponding increase in accessibility.

### 2.3.5 TCF-1 binding at chromatin domains with H3K27me3 and H3K9me3 repressive marks

The widespread binding of TCF-1 in fibroblasts led to thousands of *de novo* open chromatin regions. Yet, it is not clear whether these TCF-1-dependent regulatory elements were previously repressed or instead poised for activation with permissive histone modifications in fibroblasts. To address this question, we examined the pre-existing patterns of histone modifications in fibroblasts using maps of 5 histone modifications including: H3K4me3, primarily associated with promoters; H3K4me1 and H3K27ac characteristic of poised and active promoters and enhancers; and the repressive marks H3K9me3 and H3K27me3. Correlation and principal component analysis (PCA) at TCF-1 bound sites indicated a preferential colocalization of gained sites with previously repressed domains containing H3K27me3 or H3K9me3 modifications (**Figures 6A and S5A-B**). To create a more quantitative picture of the chromatin state prior to TCF-1 binding, we developed an unsupervised learning workflow and partitioned TCF-1 binding events into 11 clusters corresponding to 7 distinct chromatin states (**Figures 6B, S5C-D**, STAR Methods). Although less than half of TCF-1 binding events associated with active and poised enhancers or promoters (~40%), 16,800 (~42%) occurred within repressed and heterochromatin genomic regions. Strikingly, the gains in chromatin accessibility by TCF-1 were strongly enriched at these repressed domains (**Figures 6B and S5E**).

The widespread binding of TCF-1 at genomic regions with pre-existing repressive marks was unexpected. To further assess whether TCF-1 is also capable of erasing the repressive histone modifications, we next mapped H3K27me3 and H3K9me3 repressive marks in addition to the active enhancer mark H3K27ac in TCF-1 expressing cells. We

found that more than 1,400 TCF-1 binding events overlapping *de novo* open chromatin were associated with gain in H3K27ac and loss of H3K27me3 and H3K9me3 repressive marks at the center of TCF-1 binding (**Figure 6C-D**). Together, the integration of nucleosome mapping, chromatin accessibility, transcription factor binding, and histone modifications in fibroblasts suggest a fundamental role of TCF-1 in creating *de novo* chromatin accessibility because of its binding to previously repressed chromatin domains.



**Figure 6. TCF-1 can bind to repressed chromatin and make it open.**

**(A-B)** TCF-1 binds to repressed chromatin and promote accessibility. **(A)** The enrichment of ATAC-seq in NIH3T3 TCF-1 RV versus Empty RV cells and vice versa was also calculated around each summit for assessing different levels of chromatin accessibility. Principal component analysis showed that gain in accessibility occurs at TCF-1 binding sites located in repressed chromatin and loss in accessibility happens in previously active regulatory regions. **(B)** K-means clustering ( $k=11$ , designated as the optimal number of clusters by Average Silhouette Width coefficient) of TCF-1 summits on the adjusted significance levels of the enrichment in each histone mark identified chromatin states ranging from PRC (H3K27me3) (4,110, 10.2%), hetero/PRC (H3K27me3 and H3K9me3) (8,957, 22%), hetero (H3K9me3) (4,242, 10.4%), trivalent (H3K27ac, H3K4me1 and H3K9me3) (6,634, 16.4%), poised enhancers (H3K4me1) (7,458, 18.3%), active enhancers (H3K4me1 and H3K27ac) (7,343, 18.2%) and promoters (H3K4me3) (1,818, 4.5%). Normalized enrichment profiles of histone modification using ChIP-seq as well as ATAC-seq were also calculated for 10bp non-overlapping bins spanning the +/- 3kb region centered around TCF-1 summits.

**(C-D)** More than 1,400 TCF-1 binding events colocalize with a gain in both chromatin accessibility and the active mark H3K27ac with a corresponding loss of H3K27me3/H3K9me3 repressive marks. To assess differences in the enrichment of H3K9me3, H3K27me3 and H3K27ac ChIP-seq signal around TCF-1 binding events between pre-induced and TCF-1 RV NIH3T3 cells, we used the diffR function from normR package using an FDR threshold of  $5e-2$ . The average profiles for histone modifications, chromatin accessibility and TCF-1 binding using ChIP-seq were generated and demonstrated above each heatmap.

### 2.3.6 T cell-restricted genes are actively transcribed after TCF-1 expression

To evaluate whether the ectopic expression of TCF-1 and its widespread binding at over forty thousand genomic regions corresponds to any change in gene expression, we measured the transcriptional changes in fibroblasts (**Figure S6A**). After TCF-1 transduction, we found that 1,478 genes were upregulated while 1,296 genes were downregulated (**Figure S6B**). To further assess the identity of these up- and down-regulated genes, we generated two gene sets containing top “T cell genes” and “fibroblast genes” by performing differential expression analysis in DP T cells and pre-induced fibroblasts. Using gene-set-enrichment analysis, we found that the fibroblast gene-set was enriched within the down-regulated genes, suggesting the repression of the fibroblast gene expression program by TCF-1 (**Figure 7A**). Conversely, the T cell gene set was enriched within genes upregulated by TCF-1 (**Figure 7B**). The leading edge in this enrichment analysis included genes essential for T cell commitment and development including *Bcl11b*, *Rorc*, and *Cd247* (**Figure 7C**). Together, TCF-1 can initiate the reprogramming of fibroblasts towards T cells.

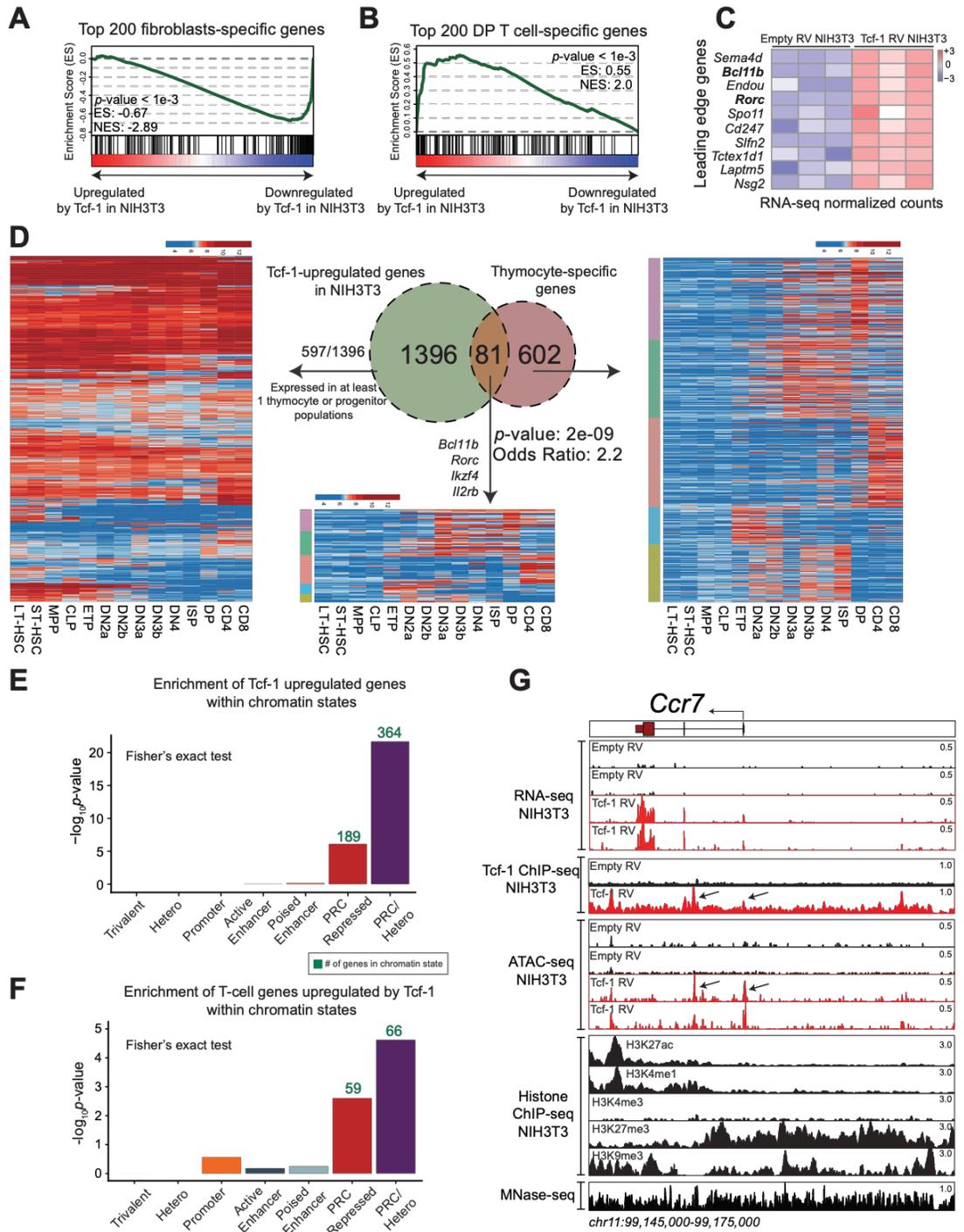
To examine whether TCF-1 upregulated genes in fibroblasts have any relevance to transcriptional profiles during T cell development, we delineated ‘thymocyte-specific genes’ as a group of genes that were selectively expressed in at least one stage of T cell development but not in bone-marrow progenitors using the ImmGen expression data (133) (**Figure 7D**). We found that TCF-1 was capable of upregulating 81 thymocyte-specific genes with ontologies associated with tissue development, cell proliferation and immune system processes (**Figures 7D and S6C**). Examples include *Bcl11b*, *Ikzf4*, *Il2rb*, *Klf4*, and *Rorc*. Additional 597 genes upregulated by TCF-1 were expressed at

multiple cellular states (**Figures 7D** left panel). It is well established that TCF-1 has recurring roles in T cell development, peripheral T cells and cells with stem properties (134, 135). We further evaluated the expression of the 1,478 genes up-regulated by TCF-1 in fibroblasts for their expression in hematopoietic progenitors together with naïve CD4<sup>+</sup> and naïve, effector and memory CD8<sup>+</sup> T cells using RNA-seq data (**Figures S6D-E**). After performing unsupervised clustering, we found that 753 genes are ordinarily expressed in one of these hematopoietic stages. Among these, 475 genes (63%) including *Ccr7*, *Il15ra*, and *Icosl* are selectively expressed in the T cell program (**Figures S6D-E**). In addition, 42 genes that were upregulated by TCF-1 in fibroblasts were selectively downregulated in TCF-1-deficient DP T cells, suggesting that TCF-1 is required and sufficient for the expression of these genes (**Figure S6F**). Together, our data suggest that *de novo* open chromatin regions are invoked by TCF-1 to induce the T cell-specific gene expression program in fibroblasts.

### 2.3.7 Genes up-regulated by TCF-1 reside in previously repressed chromatin domains in fibroblast

Our data in TCF-1 expressing fibroblasts led to two unexpected observations: (a) TCF-1 can generate chromatin accessibility at previously repressed domains and (b) TCF-1 can induce the expression of thousands of genes. To relate the chromatin state at the TCF-1 binding events to changes in transcriptional outputs in fibroblasts, we calculated the enrichment of up- and down-regulated genes among genes whose 5kb extended regions fell within TCF-1 binding events in different chromatin states. We found that the genes up-regulated by TCF-1 were significantly enriched for TCF-1 binding events at chromatin domains with repressive chromatin marks (**Figure 7E**). Conversely, genes downregulated by TCF-1 were mostly associated with promoters and

the trivalent state with high H3K4me1 and H3K27ac surrounded by H3K9me3 (**Figure S6G**). Of note, a statistically significant proportion of genes were proximal to TCF-1 binding events that led to gain in H3K27ac and loss of H3K27me3/H3K9me3 modifications in contrast to those that did not alter the chromatin state (**Figure S6H**). Remarkably, genes of the T cell program were strongly enriched within genomic regions previously within repressed chromatin domains or harboring high nucleosome occupancy (**Figures 7F and S6I**). Examples of T cell genes ordinarily blanketed by repressive H3K27me3 and H3K9me3 in fibroblasts and actively transcribed after TCF-1 expression included the receptor required for cell trafficking within and out of the thymus, *Ccr7*, and an essential transcription factor for T cell development *Rorc* (**Figures 7G and S6J**). Thus, TCF-1 can induce the expression of T cell genes in an unrelated non-hematopoietic cell type by accessing repressive chromatin domains and converting these regions to open, transcriptionally active loci.



**Figure 7. T cell-specific genes innately repressed in fibroblasts are upregulated by TCF-1.**

**(A-C)** Ectopic expression of TCF-1 upregulates T cell genes and downregulates fibroblast genes. Differential expression analysis between WT DP T cells and Empty RV NIH3T3 cells delineated T cell and fibroblast gene sets (STAR methods). RNA-seq between Empty and TCF-1 RV NIH3T3 cells (see **Figure S6**) and GSEA analysis on the fibroblast gene set (**A**) and the T cell gene set (**B**). Leading edge analysis (**C**) in top T cell genes included essential genes in T cell development and commitment such as *Bcl11b* and *Rorc*.

**(D)** TCF-1 upregulates thymocyte-specific genes that are expressed at different stages during normal T cell development. Thymocyte-specific genes were defined (STAR methods) and the overlap tested between TCF-1 RV upregulated genes in NIH3T3 (see **Figure S6B**) and thymocyte-specific genes. These genes were clustered using ImmGen microarray expression profiles (middle and right). Gene expression profiles of genes not overlapping thymocyte-specific genes but expressed in progenitors (597 genes) were also plotted (left).

**(E-F)** TCF-1 can upregulate genes initially buried within repressed chromatin. TCF-1 summits assigned to chromatin states (see **Figure 6B**) were linked to proximal genes (STAR methods). **(E)** Enrichment of upregulated genes by TCF-1 within each chromatin state. **(F)** Enrichment of T cell genes upregulated by TCF-1 (**B**) in each chromatin state was compared to fibroblast genes (**A**).

**(G)** Example of T cell genes ordinarily blanketed by H3K27me3 and H3K9me3 in fibroblasts and induced after TCF-1 expression included *Ccr7*.

## 2.4 DISCUSSION

It has been known for more than 2 decades that TCF-1 is a key transcription factor in T cell development (87). As a major mediator of Notch signaling in the specification of bone-marrow progenitors to a T cell fate, TCF-1 is required for the expression of transcription factors essential for T cell commitment and specification such as GATA3 and Bcl11b (121, 136). Yet, it has been unclear whether the mechanism by which TCF-1 controls T cell fate is the specific transcriptional regulation of a small number of genes or whether this protein has a more fundamental role shaping the global epigenetic identity of T cells. Here, by reading between the 'open' lines of the genome during thymocyte development, we found that TCF-1 is the most enriched transcription factor at thousands of regulatory elements that become accessible at the earliest stage and persist until T cell maturation. While it remains unclear whether all genomic regions with H3K9me3 modifications comprise heterochromatin (137), we found that TCF-1 binding across the genome of fibroblasts leads to gains in chromatin accessibility at genomic regions enriched with H3K27me3 and H3K9me3 repressive marks and lacking H3K27ac and H3K4me1/3 activating marks. This unique ability of TCF-1 targeting repressed chromatin might be attributed to the ability of HMG proteins to introduce a strong bend into DNA (82). The *de novo* change in accessible elements caused by TCF-1 coincided with transcription of hundreds of T cell genes. A subset of *de novo* open chromatin regions was also associated with gain of the active enhancer mark H3K27ac and loss of the repressive marks H3K27me3 and H3K9me3, corroborating the ability of TCF-1 in targeting silent chromatin. Transcription of *Ccr7*, a gene that plays an essential role in intrathymic migration and proper T cell development (138), was induced when TCF-1 was ectopically expressed in fibroblasts. Notably, the promoter of *Ccr7* that

became accessible in TCF-1 expressing fibroblasts is among the first wave of chromatin remodeling in T cell development. Similarly, the up-regulation of *Bcl11b*, a T-cell restricted transcription factor essential for T cell commitment, highlights the role of TCF-1 as an early inducer and stabilizer of T cell identity by promoting epigenetic changes that drive key transcriptional regulators of the T cell program. Although TCF-1 has been known to induce selected genes involved in T cell biology, these results reveal a previously unappreciated broad mechanism by which TCF-1 controls T cell fate through genome-wide programming of the epigenetic identity of T cells. Our integrative strategy exploiting development and reprogramming in T cell progenitors may also provide a foundation to delineate cell fate determining transcription factors shaping the accessibility landscape of other cell types.

It has been recently shown that TCF-1 is essential for repressing CD4<sup>+</sup> related genes in CD8<sup>+</sup> T cells through intrinsic HDAC activity (125). Of all TCF-1 binding events that had differential accessibility in the absence of TCF-1 in T cells, we found that a majority (80%) exerts an activating role (i.e., losing accessibility in knockout cells) with a smaller number gaining accessibility, supporting this previously reported repressive role of TCF-1. Strikingly, both gained and lost sites in our data were enriched with TCF-1 binding and TCF motif, suggesting the direct role of this transcription factor at recognizing its binding sites across the genome. While further analysis is required to examine the sequence features and epigenetic modifications classifying the activating versus repressive TCF-1 binding events, our work reveals the pervasive role of TCF-1 at establishing *de novo* open chromatin during development and reprogramming.

Conrad Waddington proposed a metaphor for cellular differentiation coining the term “epigenetic landscape” and envisioning a cell rolling down a hill like a ball with

successive bifurcations that resulted in irreversible cell fate` decisions (139). Exploiting the single cell technology, we interrogated whether a lineage-determining transcription factor can exert harmonizing and coordinate impact on the chromatin of single cells following the T cell trajectory. To infer cell-to-cell variability on open chromatin associated with transcription factors, we developed an analytical method and found that TCF-1 target sites but not those of RUNX1 or GATA3 confer the lowest cell-to-cell variability across individual T cells. Stated in a different way, open chromatin events that are highly conserved across single cells (revealed by single cell ATAC-seq) are likely to be causal to the identity of that cell type since, in this case, T cells appear not to function effectively without TCF-1 driven epigenetic events (122). Despite the limitation that our knowledge of transcription factor binding is still gathered from bulk assays such as ChIP-seq, our data demonstrate a distinct pattern at genomic regions with TCF recognition sites and TCF-1 binding, suggesting the role of this transcription factor at coordinating the chromatin accessibility of individual cells.

Our data clearly demonstrate that the TCF motif and TCF-1 binding events are strongly enriched at T-cell specific regulatory elements that become accessible early and persist until T cell maturation. Furthermore, loss of TCF-1 selectively affected the accessibility of the early regulatory elements. These findings together with the early up-regulation of TCF-1 in T cell development and the ability of this protein to reprogram the gene expression profile of fibroblasts may describe TCF-1 as a “pioneer” transcription factor (140). Nonetheless, we propose that the epigenetic complexities and the requirement for combinatoriality among transcription factors suggest that even ‘master regulators’ or ‘pioneer factors’ may require additional events to fully enact the program of cell lineage that they initiate (141-143). Here, we found that TCF-1 was endowed with an

unprecedented ability to target chromatin regions with repressive marks and in this manner, is more potent than the previously characterized pioneer factors in other developmental settings which are often impeded by heterochromatin (130, 144). Nevertheless, not the entire collection of ~1 million TCF recognition sites are bound by TCF-1 in fibroblasts and only a fraction of the T cell-specific regulatory elements became accessible in this context. It is worth noting that no other transcription factor including the previously studied pioneer factors has been reported to bind to the entire set of possible binding sites present in the genome (130, 145). We postulate that higher order chromatin conformation and epigenetic modifications such as DNA methylation may impede TCF-1 binding to the entire set of its cognate sites (146, 147). Moreover, the three waves of chromatin remodeling during T cell development enriched with TCF-1 binding suggest multiple modes of action for this transcription factor. The regulatory elements in the intermediate wave that remain closed at an earlier stage may indicate a requirement for the cooperation between TCF-1 and its partners. Similarly, although more than thousand TCF-1 binding events in fibroblasts abolished the pre-existing repressive marks, the remaining TCF-1 binding events did not modify the chromatin state, indicating the requirement of cooperating partners at these regulatory sequences. The regulatory syntax that TCF-1 follows to read the genetic code may be ascertained by machine learning techniques delineating rules of transcription factor engagement from DNA sequence and shape, histone modification, chromatin conformation, and transcription factor binding data during development and reprogramming. Collectively, our integrative data highlight a widespread means by which TCF-1 initiates the T lineage program through genome-wide epigenetic programming and induction of T cell identity genes.

## 2.5 MATERIALS AND METHODS

**Mice:** Mice used were C57BL/6J (CD45.2<sup>+</sup>) and B6.SJL-*Ptprc*<sup>a</sup> *Pepc*<sup>b</sup>/BoyJ (CD45.1<sup>+</sup>) purchased from the US National Cancer Institute animal facility. All mice analyzed were 6-12 weeks and were used without randomization or 'blinding' of researchers to mouse or sample identity. *Tcf7*<sup>-/-</sup> (*TCF-1*<sup>-/-</sup>  $\Delta$ VII) mice were kindly provided by A. Bhandoola (87). All animal work was in accordance with the Institutional Animal Care and Use Committee for the University of Pennsylvania in accordance with guidelines set forth by the NIH.

**Cell culture:** NIH3T3 cells were purchased from ATCC for this study and used at a low passage number (<12) and were maintained in high glucose DMEM 1x medium with L-glutamine and sodium pyruvate (Corning) with 100 U mL<sup>-1</sup> penicillin and 100  $\mu$ g mL<sup>-1</sup> streptomycin (Gibco) and 10% bovine serum (Gibco). 293T (ATCC) cells were maintained in high glucose DMEM 1x medium with L-glutamine and sodium pyruvate (Corning), and 100 U mL<sup>-1</sup> penicillin and 100  $\mu$ g mL<sup>-1</sup> streptomycin (Gibco) with 10% fetal calf serum (Gemini). All cells were grown at 37°C and 5% CO<sub>2</sub>.

**Retroviral Transductions:** Gateway compatible MSCV-IRES-VEX (MSCV-ccdB-VEX) and empty vector controls (MSCV-VEX) retroviral vectors were obtained from A. Bhandoola (121). Mouse *Tcf7* cDNA (NM\_009331) of the short isoform of TCF-1 (p33) was obtained from Origene and cloned into MSCV-ccdB-VEX (MSCV-TCF7-VEX) according to Gateway Clonase II instructions (Invitrogen). Sequences were verified using MacVector v15.5.0. Cells were transduced by addition of virions to culture media supplemented with polybrene at 8  $\mu$ g mL<sup>-1</sup> and 10 mM HEPES. As transduction efficiency in NIH3T3 was >99%, all assays on transduced NIH3T3 cells were performed without cell sorting.

**Retroviral Packaging:** 293T cells were plated in 4 mL DMEM media in 10 cm dishes prior to transfection. Immediately prior to transfection, chloroquine was added to a final concentration of 25  $\mu$ M. The retroviral construct and the pCL-Eco plasmid were transiently co-transfected using Lipofectamine 3000 (Invitrogen). The cells were returned to the incubator for 6 hours. Subsequently, the medium was changed to fresh media. Virions were collected 24 and 48 hr after transfection, snap-frozen, and stored at  $-80^{\circ}\text{C}$  for future use.

**Assay for Transposase-Accessible Chromatin (ATAC):** ATAC-seq was performed as previously described with minor modifications (148). Fifty thousand cells were pelleted at 550 x g and washed with 1 mL 1x PBS, followed by treatment with 50  $\mu$ L lysis buffer (10 mM Tris-HCl [pH 7.4], 10 mM NaCl, 3 mM  $\text{MgCl}_2$ , 0.1% IGEPAL CA-630). After pelleting nuclei, the pellets were resuspended in 50  $\mu$ L transposition reaction with 2.5  $\mu$ L Tn5 transposase (FC-121-1030; Illumina) to tag and fragment accessible chromatin. The reaction was incubated in a  $37^{\circ}\text{C}$  water bath for 45 minutes. Tagmented DNA was purified using a MinElute Reaction Cleanup Kit (Qiagen) and amplified with 12 cycles of PCR. Libraries were purified using a QIAQuick PCR Purification Kit (Qiagen). Libraries were paired-end sequenced (38bp+37bp) on a NextSeq 550 (Illumina). For accessibility in NIH3T3 cells, two biological replicates were performed at both 48 and 96 hr time points after transduction. Three technical replicates were performed between WT and TCF-1 KO DP T cells.

**Single Cell ATAC:** Single cell ATAC-seq was performed as previously described (127) using the C<sub>1</sub> Single-Cell Auto Prep System with the C<sub>1</sub> Open App™ program (Fluidigm). Briefly, cells were FACS sorted to high viability and purity. Cells were then stained with mammalian LIVE/DEAD Viability/Cytotoxicity Kit (Invitrogen) for 10 minutes on ice at a

final concentration of 5  $\mu$ M Ethidium homodimer-1 and 5  $\mu$ M Calcein AM in 1x PBS. After staining, cells were diluted in RPMI-1640 to a concentration of 400,000 cells mL<sup>-1</sup>. C<sub>1</sub> Cell Suspension Reagent (Fluidigm) was added to a final concentration of 20%. Brightfield and fluorescent images of each capture site was taken with a Leica DMI8. The Lysis/Tagmentation step in the C<sub>1</sub> protocol was lengthened to a duration of 60 minutes using the Open App™ software (Fluidigm). After single cell ATAC-seq chemistry was performed on the Fluidigm C<sub>1</sub>, tagmented DNA was harvested and amplified for 14 PCR cycles (Fluidigm). Libraries were paired-end sequenced (38bp+37bp) on a NextSeq 550. Three captures of DP T cells were performed over the course of this study.

**Chromatin Immunoprecipitation (ChIP) assay:** ChIP-seq was performed as previously described (149). Briefly, chromatin samples prepared from fixed cells were immunoprecipitated with antibodies recognizing mouse TCF-1 (C46C7; CST), H3K9me3 (AM39161; Active Motif), and H3K27me3 (07-449; EMD Millipore). Antibody-chromatin complexes were captured with protein G–conjugated beads, washed, and eluted. After reversal of cross-linking, RNase and proteinase K treatment were performed and DNA was purified and quantified for library preparation. Input sample was prepared by the same approach without immunoprecipitation. Libraries were then prepared using the Ultra™ DNA Library Prep Kit (NEB). Two replicates were performed for each condition. Indexed libraries were validated for quality and size distribution using a TapeStation 2200 (Agilent). Single end sequencing (75 bp) was performed on a NextSeq 550.

**RNA-seq:** Cells were washed once with 1x PBS before resuspending pellet in 350  $\mu$ L Buffer RLT Plus (Qiagen) with 10% 2-Mercaptoethanol (Sigma), vortexed briefly, snap-frozen on dry ice, and stored at -80°C. Subsequently, total RNA was isolated using the RNeasy Plus Micro Kit (Qiagen). RNA integrity numbers were determined using a

TapeStation 2200 (Agilent), and all samples used for RNA-seq library preparation had RIN numbers greater than 9.5. Libraries were prepared using the SMARTer® High-Input Strand-Specific Total RNA-seq for Illumina kit (Clontech). Libraries were single-end sequenced (75 bp) on a NextSeq 550. Three biological replicates were performed for TCF-1 RV and Empty RV transduced NIH3T3 cells. Two technical replicates were performed in WT and TCF-1 KO DP T cells.

**Cell staining and flow cytometry:** Single-cell suspensions were prepared from thymi of mice by dissociation of tissue through 70  $\mu$ M mesh filters (Falcon) in RPMI 1640 (Corning) +1% FBS (Gemini), and surfaces were stained following standard protocols. The fluorochrome-conjugated, anti-mouse antibodies were as follows: PE CD4 (RM4-4), APC CD8a (53-6.7), PE c-Kit (2B8), APC CD25 (PC61), and Streptavidin BV605. For intracellular detection of TCF-1 in RV-transduced NIH3T3, cells were harvested after trypsin dissociation (Gibco), fixed with 1% PFA for 10 minutes on ice to preserve VEX signal, fixed and permeabilized with the FoxP3/Transcription Factor Staining Buffer Set (eBioscience), and incubated with PE-conjugated anti-TCF-1 (S33-966). All antibodies used for flow cytometry were purchased from BioLegend or BD Biosciences. Data were collected on an LSRII running DIVA software (BD Biosciences) and were analyzed with FlowJo software v10.2 (TreeStar).

**Cell sorting:** Antibodies used in the lineage cocktail (Lin) include biotinylated antibodies against B220 (RA3-6B2), CD19 (1D3), CD11b/Mac1 (M1/70), Gr1 (8C5), CD11c (HL3), NK1.1 (PK136), TER119 (TER-119), CD3 $\epsilon$  (2C11), CD8 $\alpha$  (53-6.7), CD8 $\beta$  (53-5.8), TCR $\beta$  (H57),  $\gamma\delta$ TCR (GL-3). After surface staining with the lineage cocktail, cells were incubated with Streptavidin Microbeads (Miltenyi Biotec). DN cells were then negatively isolated from total thymocytes using magnetic separation columns (Miltenyi Biotec).

Negatively selected cells were then stained with c-Kit and CD25 followed by Streptavidin BV605 to reveal escaping Lin<sup>+</sup> cells. The DN3 population was defined and cell-sorted as Lin<sup>-</sup> Kit<sup>-</sup> CD25<sup>+</sup>. Total thymocytes were stained with CD4<sup>+</sup> CD8<sup>+</sup> to define and sort the DP population. Dead cells were excluded through 7-amino-actinomycin D (7-AAD) uptake. Doublets were excluded through forward scatter–height by forward scatter–width and side scatter–height by side scatter–width parameters. Purity was verified after sorting, and all cell populations were sorted to a purity of >98%. Sorting was performed on FACS Aria II (BD Biosciences) and were analyzed with FlowJo v10.2 (TreeStar).

**High-throughput sequencing data pre-processing:** Quality assessment of raw reads was achieved with FastQC (150) and contaminants were removed using Trimalore (151) with parameters '*-q 15 --length 20 --stringency 5*'. For RNA-Seq samples, '*--clip\_R1 3'*' was added to the Trimalore parameters facilitating the removal of the 3nt bias introduced to the 5' end of reads. Human (GRCh37, November 17 2015) and mouse (GRCm38, May 23 2014) reference genomes were downloaded from UCSC repository (152) and mouse gene models were derived from Gencode vM11 (153).

Bulk ATAC-seq samples were mapped to the reference genomes using Bowtie2 v2.2.9 (154) with *-X2000*. STAR v2.5 (155) was used for aligning single-cell ATAC, RNA, ChIP and MNase-seq reads with parameters specifically optimized based on the properties of each protocol. RNA-seq samples were analyzed with parameters '*--outFilterMultimapNmax 1 --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0 --alignEndsType Local*'. On the other hand, ChIP-seq raw reads were aligned with parameters '*--alignSJDBoverhangMin 999 --alignSJoverhangMin 999 --alignIntronMax 1 --outFilterMultimapNmax 1 --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0 --alignEndsType Local*'

to disable the usage of known and prevent calling novel splice junctions. The same parameters were also applied for mapping scATAC-seq and MNase-seq data combined with '*--alignMatesGapMax 2000*' which limits the distance between aligned read mates to 2,000bp.

Reads aligned to the mitochondrial genome as well as reads mapping to multiple genomic loci were discarded from downstream analyses. Additionally, Picard (156) minimized the PCR amplification bias in ATAC-, ChIP- and MNase-seq samples. In cases of paired-end MNase-seq samples, fragments smaller than 75bp were also filtered out.

ATAC-seq samples derived from single DP T cells were filtered using previously described quality standards (127). In brief, libraries containing less than 10,000 fragments or libraries with less than 15% of their fragments falling in open chromatin (as defined in the single cell accessibility section) were also removed from subsequent analyses (**Figure S3C-D**).

**Differential gene expression analysis:** HTSeq v0.6.1 (157) facilitated counting RNA-seq reads on Gencode vM11 (153) gene models with parameters '*-s yes -t exon -m intersection-nonempty*'. DESeq2 (126) was subsequently applied on gene counts to identify genes differentially expressed between DP WT and DP TCF-1 KO (**Figure S2C**), NIH3T3 Empty RV and NIH3T3 TCF-1 RV (**Figure S6B**) as well as DP TCF-1 WT and NIH3T3 Empty RV cells after removing entries that exhibited zero counts in all replicates. The quality of replicates was assessed by calculating pairwise spearman correlation coefficient (**Figures S2B and S6A**) as well as plotting the variability explained by the first two principal components (data not shown).

Additionally, gene expression levels were calculated in a variety of cell types ranging from hematopoietic stem cells to effector and memory T cells and normalized using the variance stabilizing transformation (VST) (126). K-means (k=12) clustering was then applied on the VST expression values of genes upregulated by TCF-1 in NIH3T3 cells to identify cell state specific patterns (i.e., clusters) of TCF-1 regulated gene expression. For the same set of genes, we also calculated RPKM normalized expression values that were used to filter out lowly expressed genes (RPKM < 0.5 in all samples) and visualizing the clusters (**Figure S6D-E**). Cluster 1 was removed from the analysis due to low expression levels in all hematopoietic lineages. Genes downregulated in the TCF-1 KO DP T cells were overlapped with the genes upregulated by TCF-1 RV in NIH3T3 and the significance of the overlap was tested by Fisher's exact test. (**Figure S6F**).

**Defining thymocyte-specific gene program:** Normalized microarray expression data for bone marrow stem cell and thymocyte populations was downloaded from the Immunological Genome Project Consortium (133). Microarray probe IDs (affy mogene 1.0st v1) were converted to Ensembl gene IDs using the Ensembl mouse gene mart (GRCm38.p5) in biomaRt (158, 159). Genes were considered expressed in a population if expression values were above 120 indicating >95% probability of true expression (160). To define thymocyte-specific genes (**Figure 7D**), genes were filtered based on expression values lower than 120 in all considered progenitor populations (LT-HSC, ST-HSC, MPP, CLP) and with expression values higher than 120 in at least 1 thymocyte population (ETP, DN2a, DN2b, DN3a, DN3b, DN4, ISP, DP, CD4<sup>+</sup>, CD8<sup>+</sup>). Genes were further filtered based on having at least a 2-fold increase in expression between any two populations. The overlap of thymocyte-specific genes and genes upregulated by TCF-1 RV in NIH3T3 was determined using the GeneOverlap package (161). Genes

upregulated by TCF-1 RV in NIH3T3, described in previous sections, but not overlapping with thymocyte-specific genes were filtered based on expression >120 in at least one progenitor population and plotted (**Figure 7D**). Thymocyte genes were grouped into patterns of expression by combining thymocyte-specific genes with both overlapping and non-overlapping with genes upregulated by TCF-1 RV in NIH3T3 and performing k-means clustering using 5 centers. Gene ontology analysis (**Figures S1E and S6C**) was performed using the Gene Ontology gene set collection in MSigDB database v6.1 (162, 163).

**Peak calling:** Following ENCODE guidelines, for the characterization of reproducible TCF-1 peaks in NIH3T3 TCF-1 RV cells, macs2 v2.1.1 (164) was initially applied separately on each of the two ChIP-seq replicates as well as after merging both replicates with parameters '*--nomodel --extsize 300 --keep-dup all --call-summits -q 0.9*' using the TCF-1 ChIP-seq on NIH3T3 Empty RV cells as control. The identified peaks were filtered with Irreproducible Discovery Rate (IDR) v2.0.2 (165) using an IDR threshold of  $2e-2$  resulting in a high-quality set of 40,562 reproducible peaks.

TCF-1, GATA3, RUNX1 and PU.1 binding sites in mouse thymocytes were identified by applying macs2 with parameters '*-p 1e-3 -q 0.05*' using the corresponding Input samples as control resulting in 56,817 TCF-1 peaks, 54,475 GATA3, 67,915 RUNX1, 98,036 PU.1 in DN1 and 92,660 in DN2a.

A proximity-based strategy was adopted for linking genes to regulatory elements and transcription factor binding sites. Gene models were downloaded from Gencode M11 and both ends of each gene were extended by 5kbp. Open chromatin sites identified by

ATAC-seq as well as ChIP-seq derived transcription factor binding sites were assigned to genes if they were found to overlap with their extended models.

**Differentially accessible chromatin between DP WT and TCF-1 KO as well as**

**between NIH3T3 TCF-1 and Empty RV cells: Macs2 with *'-p 1e-7 --nolambda --***

*nomodel'* was applied on each DP WT and DP TCF-1 KO ATAC-seq replicate separately to identify accessible chromatin. Peaks were subsequently merged using BEDTools (166) and ATAC-seq read counts were calculated in the merged peaks for every replicate. The resulting count table was used to identify 6,165 (1,165 presenting more and 5,000 less enrichment in DP TCF-1 KO) loci differentially enriched in ATAC-seq signal between DP WT and DP TCF-1 KO with DESeq2 after applying a 0.001 and 0.55 cutoff on *p*-value and logFC respectively (**Figure 3A**).

The same approach and cutoff were applied in NIH3T3 cells (**Figure S4D**) for identifying 8,506 genomic regions presenting differential ATAC-seq signal enrichment between Empty and TCF-1 RV (6,888 presenting more and 1,618 less enrichment in TCF-1 RV).

**Characterization of cell-state specific accessible chromatin:** An IDR threshold of  $5e-2$  was used, following the pipeline described in previous section, to identify accessible chromatin for every murine ATAC-seq sample (HSC, MPP, CLP, B, NK and all stages of T cell development from DN1 to naïve CD4+ and naïve CD8+ cells). Peaks were merged and filtered based on their overlap with annotated promoters (Gencode M11 TSSs extended by +4kb/-2kb) resulting in a collection of 55,481 distal regulatory elements. The FDR value of each peak in every cell type was used as a proxy for the level of accessibility.

Each peak was assigned a 13-dimensional vector containing the ATAC-seq enrichment proxy in every cell type. Average Silhouette Width (ASW) statistic was used for deciding on the number of clusters prior to applying k-means. The initial set of regulatory regions was reduced after removing the members of clusters 7, 8, 10, 11, 13, 17 and 23 (**Figure S1A**). The remaining 35,869 loci were re-clustered after re-calculating ASW (data not shown) to produce the final set of groups (**Figure 2A**). Normalized (TPM) ATAC-seq profile for every regulatory element was calculated by segmenting a +/- 2,000bp window around its center in 10bp bins and calculating the normalized overlapping ATAC-seq tag counts (**Figures 2B and S1B**).

*De novo* motif analysis using Homer with '*-size given -len 6,8,10*' was applied on each cluster separately using the excluded set of clusters as background (**Figure 2C-E**, **Figure S1C**). Additionally, odds ratio and percentage of binding of TCF-1, GATA3, RUNX1 and PU.1 (DN1 and DN2a) was calculated for each cluster based on publicly available CHIP-seq data (**Figure 2F**).

An alternative approach was used for identifying T cell specific accessible chromatin in human cells (**Figure S1D**). The lack of replicates for certain cell types restricted the use of IDR. Therefore, macs2 with parameters '*-p 1e-7 --nolambda --nomodel*' was used for every cell type (HSC, MPP, CLP, B, NK, Naïve CD4<sup>+</sup> and Naïve CD8<sup>+</sup> cells) on each replicate separately. Peaks were merged with BEDTools and normalized ATAC-seq enrichment for every cell type was calculated after merging the replicate samples within each cell type. Gencode M11 gene models were used to separate the set of ATAC-seq peaks into distal and promoter related loci after extending the annotated gene transcription start sites by -4kb/+2kb.

Each peak was assigned a 7-dimensional vector containing the normalized ATAC-seq enrichment in every cell type. Within Sum of Squares (WSS) statistic was used (data not shown) for deciding on the number of clusters prior to applying k-means (k=10 for the distal sets and k=5 for the promoter sets). De novo motif analysis using Homer with '*size given -len 6,8,10,12*' was applied on each cluster separately with remaining peaks in other clusters as background (**Figure S1D**).

**Querying chromatin accessibility at the single-cell level:** To assess whether TCF-1 binding events harbor the strongest chromatin accessibility as measured by ATAC-seq in DP T cells, we measured genome-wide binding of TCF-1, RUNX1 and GATA3 by ChIP-seq as previously described. An equal number of genomic regions with unique binding of each transcription factor were subsampled and the normalized tag count enrichment from ATAC-seq in DP T cells facilitated the comparison of the 3 regulatory proteins (**Figure 4A**).

Based on this analysis, TCF-1 bound open chromatin was found to exhibit the highest levels of accessibility compared to RUNX1 and GATA3. This observation inspired us to further investigate with a single cell analysis. ATAC-seq data from 110 single DP T cells passing previously defined (127) quality standards (**Figure S3D**) were utilized to test the hypothesis that TCF-1 exerts a deterministic effect on the chromatin, forcing T cell fate commitment. Following pre-processing and alignment, DP single cell ATAC-seq reads were merged and using macs2 with parameters '*-p 1e-3*', 22,774 accessible sites were identified.

To assess the correlation between aggregated single cell and bulk ATAC-seq enriched sites identified from both experimental procedures were merged. Normalized enrichment

was subsequently calculated in bulk (downsampled to 11.6 million reads using samtools) and aggregated scATAC-seq with 11.6 million reads enabling the correlation level quantification between the two assays (**Figure 4B**).

Our objective was to assess whether TCF-1-bound open chromatin had lower accessibility variance than background noise and chromatin bound by RUNX1 or GATA3. To this end, we generated 4 disjoint sets comprising of ATAC-seq peaks uniquely bound by TCF-1, RUNX1, GATA3 as well as peaks not bound by any of these three transcription factors. For each subset, binarized accessibility matrices were calculated based on the overlap between the identified peaks and ATAC-seq reads from each cell, thus 1 translates to accessible and 0 to inaccessible regions.

TCF-1 binding events overlapped with more ATAC-seq peaks than RUNX1 or GATA3, therefore we subsampled 30 peaks from each TF-bound peak set. We repeated the subsampling process 500 times to increase accuracy. We then calculated the accessibility variance between cells at each subsample as follows. For each subsample, the binary accessibility vector of each cell formed a 30-dimensional vector. To measure cell to cell differences in accessibility levels, we calculated the pairwise Manhattan distance between accessibility vectors, forming a distance matrix.

We subsequently centered the Manhattan distance matrix by subtracting column and row means and adding the overall mean. Then we spectrally decomposed the centered matrix to define principal coordinates and mapped all accessibility vectors to full principal coordinate space. We identified the location that minimized the average distance to all vectors, termed the spatial median (**Figure 4D**). Then, we calculated each vector's

distance from the spatial median. Finally, we calculated the average distance from accessibility vectors to the spatial median using the R package *vegan* (167).

**Correction for Technical Biases:** Variation associated with technical factors such as GC content and mean accessibility differences can often introduce obstacles in interpreting NGS data. To overcome such limitations, for every original peak, we selected 30 “technical control” ones. The set of peaks not bound by any TF were divided into 2-percentiles based on GC content. Every original peak was subsequently placed into a 2-percentile, and 30 technical control peaks within a 2-percentile of GC content were randomly subsampled with replacement. All technical control peaks were also within +/- 0.01 of the overall mean accessibility of the original 30 peaks.

**Correction for Background Noise:** To measure accessibility variation beyond background noise, we calculated accessibility variation (with technical controls) for 500 randomly selected subsamples of peaks bound by no TF. This can be viewed as a negative control.

A variability equal to 1 implied that a TF was associated with no more variation than background noise. A variability below 1 implied that a TF was associated with less variation than background noise, and a variability above 1 implied greater variation than background noise.

In addition to the methodology described above, we also applied *chromVAR* (168) for assessing the deterministic effect of TCF-1 on shaping the chromatin landscape during T cell development (**Figure 4E**).

**ChIP-seq oriented approach for assessing the deterministic effect of TCF-1 during T cell development:** An alternative, unbiased strategy was also adopted which, unlike

the previous approach, was not formed on the basis of TCF-1 binding (**Figure 4F**). The T cell specific sites in cluster 9 (**Figure 2A**) were ranked based on the sum of binary counts across individual T cells. Using default parameters of *bedtools intersect*, the overlap of regions with ChIP-seq signal from transcription factors known to be important in T cell development such as TCF-1, GATA3 and RUNX1 was assessed. *De novo* motif analysis was performed on the top and bottom 100 enhancers that exhibit the highest and lowest homogeneity respectively at the single cell level using Homer with parameters '-size given -len 6,8,10,12'. Background control in this motif analysis was any other open chromatin sites in DP T cells. The top/bottom 100 enhancers were also linked to genes based on proximity (<10kbp) in order to enable GO term enrichment analysis using the GSEA software.

**Identifying the nucleosome occupancy level on TCF-1 binding sites:** MNase-seq in mouse embryonic fibroblasts was used as a proxy for observing the nucleosome enrichment surrounding TCF-1 binding sites. To this end, the region around TCF-1 peak summits was divided into 3 windows of 200bp each; -300/-100, -100/+100 and +100/+300 following a upstream-central-downstream rationale. The nucleosome enrichment in every window was approximated by calculating the number of overlapping MNase-seq reads after extending their 3' end to 147bp and normalizing based on the number of uniquely mapped reads in each sample. TCF-1 summits were subsequently ranked from high to low enrichment by summing the values of left, central and right windows. For visualization purposes, the normalized MNase-seq enrichment was also calculated for 10bp non-overlapping bins spanning the +/- 3kb region centered around TCF-1 summits (**Figure 5A**).

In the case of mouse embryonic fibroblasts, visualizing the prior nucleosome enrichment status on the genomic loci bound by TCF-1 after *Tcf7* retroviral transduction clearly suggests that TCF-1 binding occurs on: a) nucleosome dense, b) nucleosome free and c) regions of intermediate nucleosome occupancy (**Figure 5A**). Instead of choosing an arbitrary threshold on the ratio of central versus left and right window nucleosome enrichment, k-means ( $k = 3$ ) clustering was applied resulting in the formation of 3 TCF-1 summit groups and validating the previously described observation (**Figure S4C**). A total of 29,661 (73.2%) TCF-1 binding events occur on sites with dense (10,593, 26.2%) or intermediate (19,068, 47%) nucleosome enrichment and 10,901 (26.8%) on nucleosome free regions.

Dpos module from Danpos2 (169) was applied on the MNase-seq data with default settings to identify nucleosome positioning as well as calculate the nucleosome enrichment profile on a genome-wide scale. Regions called as nucleosomes exhibiting increased fuzziness (Dpos score less than 80) were removed from subsequent analyses. The distance of 40,562 TCF-1 summits in mouse embryonic fibroblasts to the closest nucleosome summit was calculated as an alternative strategy of assessing the ability of TCF-1 to directly bind on nucleosomes (**Figure 5B**). The typical length of DNA fragments wrapped around nucleosomes is 147bp. This allowed us to classify 27,145 TCF-1 summits (66.9%) located less than 75bp (vertical dashed red line) away from nucleosome summits as bound to nucleosomes and 13,417 (33.1%) summits as unbound. As a control, we applied the same bound/unbound to nucleosomes classification scheme on CTCF summits derived from analyzing public ChIP-seq data, resulting in 20,370 (56.6%) bound and 15,616 (43.4%) unbound summits.

To assess the difference of nucleosome occupancy level around TCF-1 ChIP-seq peak summits between Empty RV and TCF-1 RV NIH3T3 cells, TCF-1 summits (IDR less than or equal to 0.02) were intersected with ATAC-seq enriched regions in both conditions. Summits overlapping ATAC-seq peaks in either set (n=15,763) were extended by +/- 500 bases and nucleosome occupancy in Empty and TCF-1 RV NIH3T3 cells was measured using NucleoATAC algorithm (132). NucleoATAC infers nucleosome enrichment by integrating large and small ATAC-seq fragment positioning in accessible chromatin. Therefore, to quantitate nucleosome enrichment around TCF-1 summits with NucleoATAC algorithm, ATAC-seq signal from both Empty and TCF-1 RV NIH3T3 samples is required. Out of 15,763 queried summits 7,395 were found to exhibit at least 1.5 fold-change difference in nucleosome occupancy signal between Empty and TCF-1 RV NIH3T3 cells (**Figure S4F**).

**T cell gene enrichment in nucleosome enriched based clusters of TCF-1 summits:**

Based on the previously described analysis regarding the pre-induced nucleosome enrichment levels around TCF-1 ChIP-seq derived binding events in TCF-1 RV NIH3T3 cells, we identified 3 clusters of TCF-1 summits (**Figure S4C**). TCF-1 summits with high (n=10,593), intermediate (n=19,068) and low (n=10,901) nucleosome enrichment. In parallel, the previously described differential gene expression analysis between Empty RV NIH3T3 and DP T cells, identified 3,349 genes as DP T and 4,040 as NIH3T3 cell-specific. To calculate the enrichment of the 2 gene sets in the 3 nucleosome enrichment clusters, TCF-1 peak summits were associated with genes, as described in previous section, resulting in 27,794 interactions between 24,330 TCF-1 summits and 10,212 genes. To remove redundancy in the association between genes and nucleosome clusters we filtered out genes associated to zero or more than one clusters. The

remaining were used to calculate the enrichment of DP T cell-specific genes in high, intermediate and low MNase clusters with Fisher's exact test (**Figure S6I**).

**Motif distances from nucleosome summit:** MEME-FIMO (165) and TCF-1 position probability matrix (MA0769.1) from JASPAR (170) facilitated the discovery of 1,102,896 putative TCF-1 binding sites (motifs) in the mouse genome using a  $p$ -value threshold of  $1e-4$ . 17,816 motifs were found to overlap with TCF-1 ChIP-seq peaks specific to TCF-1 RV NIH3T3 and 7,782 with peaks specific to DP T cells. To avoid biases associated to imbalanced number of motif occurrences in peaks, a one-to-one association between motifs and summits was created by selecting the closest to summit motif per peak with a maximum distance threshold of 100bp. This resulted in the finalized sets of motifs bound in TCF-1 RV NIH3T3 ( $n=10,665$ ) and DP T cells ( $n=6,115$ ). The remaining unbound putative TCF-1 sites were grouped into motif hotspots using a distance threshold of 500bp. For every hotspot the motif with the highest FIMO score was selected as its representative (random selection for ties) resulting in the formation of the final 'Random' set of unbound motifs ( $n=862,733$ ) that were used as control.

Nucleosome positions were called using Danpos2 (169) as previously described. The distance between motifs from NIH3T3, DP T and Random sets to their closest nucleosome dyad was calculated using BEDTools (166). The visual comparison of the distribution of distances between each cell type specific set and the Random set was achieved by randomly selecting 1,000 samples from each set with replacement, plotting the density of distances and repeating this process 1,000 times (**Figure S4D**). To assess whether there is a statistically significant difference in the median motif distance from the nucleosome dyad between each cell-specific (target) set and the Random set, we carried out two separate bootstrapping procedures, one for each target set. Distances

from the target and Random set were combined into a pooled vector. Both target and Random sets were transformed by subtracting each set's mean from every member of the relevant set and adding the mean of the pooled vector. This way, both sets are first centered around their mean and then shifted by the pooled mean resulting in the proper transformation for testing the null-hypothesis (no difference between median motif distances from the nucleosome center of the two sets) without making any assumptions about their variance. Subsequently, we randomly selected 1,000 samples (with replacement) from each transformed set and compared the difference between median distances. After repeating this process 100,000 times we divided the number of times we observed a difference between the median distances larger than (or equal to) the raw difference (no subsampling) to calculate the  $p$ -value (**Figure S4D**).

#### **Characterization of the chromatin state in NIH3T3 cells prior to *Tcf7* retroviral**

**transduction:** In addition to having established the pre-*Tcf7* overexpression nucleosome occupancy environment in NIH3T3 cells, querying the chromatin state landscape is a critical step towards unveiling the properties of TCF-1 binding in a genome-wide, quantitative way. To this end, the enrichment of H3K4me3, H3K4me1, H3K9me3, H3K27me3 and H3K27ac versus Input in the +/- 1kb (and +/- 250bp) region surrounding TCF-1 summits has been calculated by modeling read counts with a binomial mixture model of two components with normR (171). The first component models the background and the second one the signal, independently for each histone mark, resulting in a five-dimensional vector of  $p$ -values adjusted for multiple comparisons for every summit. H3K27me3, H3K9me3, and H3K27ac enrichment in TCF-1 RV NIH3T3 cells was calculated as well (**Figure 6C-D**). Furthermore, the

enrichment of ATAC-seq in NIH3T3 TCF-1 RV versus Empty RV cells and vice versa has also been calculated around summits.

These enrichment results facilitated the assessment of correlations between the chromatin status and chromatin accessibility before and after *Tcf7* overexpression (**Figure 6, Figure S5**). Additionally, k-means clustering has been applied on TCF-1 summits based on the enrichment level of the 5 chromatin marks in pre-induced cells resulting in the formation of 11 clusters (**Figure 6B, Figure S5D, Table S6**), following silhouette coefficient analysis (**Figure S5C**). For visualization purposes, the normalized histone mark ChIP-seq as well as ATAC-seq enrichment was also calculated for 10bp non-overlapping bins spanning the +/- 3kb region centered around TCF-1 summits separately for each cluster.

Deregulated genes in NIH3T3 TCF-1 RV cells were linked to TCF-1 binding sites based on the proximity strategy described in previous sections. Consequently, genes were also connected to chromatin states. This enabled the calculation of the significance of up- and down-regulated genes enrichment in each chromatin state using Fisher's exact test (**Figures 7E and S6G**). To assess differences in the enrichment of H3K9me3, H3K27me3 and H3K27ac ChIP-seq signal around TCF-1 binding events between pre-induced and TCF-1 RV NIH3T3 cells, we used *diffR* function from *normR* package using an FDR threshold of 5e-2.

**Gene set enrichment analysis:** Pre-ranked lists of genes were used by ranking genes using estimated log2 fold-change in DESeq2. GSEA v2.2.4 with default parameters was used to perform gene set enrichment analysis.

**Quantification and Statistical Analysis:** All statistical analyses were performed from packages from R's basic installation.

**Data and Software Availability:** The accession number for the ChIP-seq, RNA-seq and ATAC-seq reported in this study is NCBI GEO: GSE99149

## CHAPTER 3: Differentiation of a distinct and stable T-bet<sup>+</sup> memory B cell subset

### 3.1 PREFACE

The manuscript presented in this chapter was originally published in *Immunity* (172). It has been reformatted here in accordance with the University of Pennsylvania dissertation formatting guidelines and the introduction has been expanded.

#### Authors:

John L. Johnson<sup>1†</sup>, Rebecca L. Rosenthal<sup>1†</sup>, James J. Knox<sup>1†</sup>, Arpita Myles<sup>1†</sup>, Martin S. Naradikian<sup>2</sup>, Joanna Madej<sup>1</sup>, Mariya Kostiv<sup>1</sup>, Aaron M. Rosenfeld<sup>1</sup>, Wenzhao Meng<sup>1</sup>, Shannon R. Christensen<sup>3</sup>, Scott E. Hensley<sup>3</sup>, Jonathan Yewdell<sup>4</sup>, David H. Canaday<sup>5</sup>, Jinfang Zhu<sup>6</sup>, Adrian B. McDermott<sup>7</sup>, Yoav Dori<sup>8</sup>, Max Itkin<sup>9</sup>, E. John Wherry<sup>10</sup>, Norbert Pardi<sup>11</sup>, Drew Weissman<sup>11</sup>, Ali Najji<sup>12</sup>, Eline T. Luning Prak<sup>1</sup>, Michael R. Betts<sup>3</sup>, and Michael P. Cancro<sup>1</sup>

#### Affiliations:

<sup>1</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; <sup>2</sup>La Jolla Institute For Allergy and Immunology, La Jolla, CA 92037, USA; <sup>3</sup>Department of Microbiology, University of Pennsylvania, Philadelphia, PA 19104, USA; <sup>4</sup>Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA; <sup>5</sup>Division of Infectious Disease, Case Western Reserve University School of Medicine, and Cleveland VA Hospital, Cleveland, OH 45106, USA; <sup>6</sup>Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA; <sup>7</sup>Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA; <sup>8</sup>Center for Lymphatic Imaging and Intervention, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; <sup>9</sup>Division of Interventional Radiology, Department of Radiology, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA; <sup>10</sup>Department of Systems Pharmacology and Translational Therapeutics,

University of Pennsylvania, 19104, USA; <sup>11</sup>Department of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; <sup>12</sup>Department of Surgery, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA.

† *Co-first authors*

Respective Contributions:

JLJ, RLR, AM, MPC designed the murine portion of the study; JJK, MRB designed the human portion; JLJ, RLR, JJK, AM, MPC wrote the manuscript; DHC, YD, MI, AN, MRB provided human tissues; JLJ, JJK, RLR, MN, JM, MK, SRC performed experiments; JLJ, RLR, JJK analyzed the data; WM, AR and ELP generated and analyzed the immune repertoire profiling data, contributed the immune repertoire models, and helped edit the manuscript; AM, NP, DW, SH, ABM, JZ helped secure reagents critical to the study; MPC, MRB, AN, and ELP secured the funding.

## 3.2: INTRODUCTION

### 3.2.1 The innate and adaptive immune responses

Protective immunity depends on selecting appropriate effector function and perpetuating those choices in long-lived memory cells. While non-specialized effector function may afford some degree of protection, the adaptive immune response is most protective when effector function is pathogen-tailored and subsequent memory cells develop sustained specialization. These fate choices are set early in the immune response due to the interplay between the innate and adaptive immune systems (173). The signals exchanged between these two systems are often classified as activating and regulatory interactions but can also include interactions that specify effector cell identity and function. Cells of the innate immune system are the first line of defense and can

respond rapidly to the presence of a pathogen. Innate cells such as macrophages and DCs express a variety of germ-line encoded PRRs that bind evolutionarily conserved PAMPs (174). The engagement of PAMPs with PRRs transmits an activation signal that triggers an immediate innate cell response. Once activated, the innate cell response provides inductive signals to cells of the adaptive immune response to direct effector choices (175, 176).

Unlike the innate immune response, the adaptive immune response is directed by antigen-specific lymphocytes. T cells and B cells bear antigen receptors (TCRs and BCRs, respectively) that are assembled by gene rearrangement during their development in the thymus and bone marrow. According to the clonal selection theory (6), these unique antigen receptors are clonally distributed throughout T and B cell populations. The interactions of antigen receptor with antigens are reversible, but receptor occupation needs to exceed a threshold for lymphocyte activation to occur. This requirement provides selectivity, as only the antigens with the right stereochemical and electrostatic configuration will bind the receptor with a high enough affinity to cross-link receptors or engage co-receptors (177). Antigen binding to the receptor triggers a cascade of intracellular signaling events that culminates in the activation of transcription factors to induce mitosis. Activated T and B cell clones retain their antigen specificity as they expand by 10- to 100-fold and gain effector function. A fraction of these cells persists indefinitely as memory cells, sustaining ongoing effector functions and participating in responses to subsequent pathogen challenges. Deleting cells with antigen receptors that respond to self-antigens or preventing their acquisition of effector function sustains immunological tolerance. These four pillars of the adaptive immune response—inducibility, specificity, memory, and non-responsiveness to self—

are shared between the cell-mediated and humoral aspects of the immune response, mediated by T and B cells, respectively.

The T cells of the cell-mediated immune response coordinate their effector functions via interactions of their TCR with peptide:MHC complexes on antigen presenting cells. T cells are subdivided based on the mutually exclusive expression of either CD8 or CD4, co-receptors involved in the binding of TCR to MHC I or MHC II, respectively. The distribution of MHC I and MHC II expression varies amongst cells. Essentially all nucleated cells express MHC I and peptides derived from cytosolic protein synthesis are presented on MHC I. The immune surveillance of intracellular activity is therefore carried out by CD8<sup>+</sup> cytotoxic T cells, or CTL, that lyse cells presenting non-self peptides on MHC I. Conversely, extracellular activity is monitored by CD4<sup>+</sup> T cells that bind peptides derived from the endocytic pathway presented on MHC II. Thus, MHC II expression and presentation are limited to antigen presenting cells that regularly internalize components of the extracellular environment, such as B cells, macrophages, and DCs. CD4<sup>+</sup> T cells, or helper T cells, produce membrane-associated proteins and cytokines to orchestrate the immune response and are capable of differentiating into effector subsets such as T<sub>H</sub>1, T<sub>H</sub>2, T<sub>H</sub>17, T<sub>FH</sub>, or Tregs (178-180). The differentiation programs associated with these effector subsets are carried out by master transcriptional regulators in response to certain cues shaped by the nature of the pathogen and its PAMPs, the PRRs that are engaged, and the interactions between innate and adaptive immune cells. TCR signal strength, costimulation, and cytokine milieu guide CD4<sup>+</sup> helper T cells to adopt effector fates (181-184).

### 3.2.2 B lymphocytes of the adaptive immune response: subsets and development

The humoral response is mediated by B cells that, following activation and differentiation to plasma cells, produce highly specific antibodies. Antibodies serve as the antigen receptor of B cells, but their exacting specificity for the molecular conformation of the antigen means that a highly diverse set of antibodies need to be represented in the primary pool of B cells (185). To achieve this diversity, antibodies are composed of two chains, the heavy and light chain, and gene rearrangement of V, D, and J gene segments for the heavy chain and V, J segments for the light chain greatly diversifies the coding sequence (186). When translated, the recombined VDJ and VJ regions of the heavy and light chains fold and come together to create an antigen binding region called the combining site. The junctional sequence spanning the recombined segments, known as the complementarity defining region 3, or CDR3, is the most variable sequence of the antibody and plays a major role in determining specificity (187). Further, an antibody is bivalent being comprised of two pairs of heavy and light chains joined together by disulfide bonds (188). For antigen binding, the heavy and light chains are made up of a variable region and a constant region. The joining of the heavy and light chain variable regions contacts the antigen and confers specificity whereas the constant region of the heavy chain, but not the light chain, dictates effector function.

There are five Ig heavy chain constant region isotypes,  $\mu$ ,  $\delta$ ,  $\alpha$ ,  $\gamma$ ,  $\epsilon$  that combine with  $\lambda$  or  $\kappa$  side chains to make an IgM, IgD, IgA, IgG, or IgE antibody, respectively. The IgG isotypes can be further subdivided into IgG1, IgG2a/c (IgG2a in BALB/c and IgG2c for C57Bl/6 (189)), IgG2b, or IgG3. The selection of the Ig heavy chain isotype has profound implications for the response quality. Antibodies can protect through multiple mechanisms such as neutralization, opsonization, complement deposition, degranulation, and ADCC, and it is the heavy chain constant region that dictates this

effector action. The heavy chain constant region links the adaptive immune system with the innate immune system as innate and other cell types express isotype-specific Fc receptors (FcR). FcRs can be activating or inhibitory and can modulate cell behavior based on the FcR expression profile (190). Moreover, the selection of isotype is based on the cytokine milieu and the expression of master transcriptional regulators.

Immunoglobulin class switching, also known as class switch recombination (CSR) or isotype switching, is the process by which antibody heavy chain constant regions are changed in an ongoing immune response, involves gene rearrangement and is thus irreversible. Therefore, tight regulation of isotype switching is critical for effective humoral immunity and to avoid effector mechanisms that cause autoimmune or inflammatory disease. Generating a protective antibody response of the proper effector isotype requires strict regulation of B cell development, commitment, and homeostasis.

There are two major types of mature B cells: B1 and B2 cells. B1 B cells develop from fetal liver precursors during embryogenesis and contribute to the bulk of IgM titers to “natural” antigens as an extension of the innate immune system (191-193). B2 B cells are derived from bone marrow stem cells and are constantly produced throughout life except during old age when lymphopoiesis declines (194, 195). Multiple transcription factors coordinate the transcriptional network of B cell development. The transcription factor PU.1 (196) is expressed in B cell precursors to establish a regulatory state for E2A (197) and EBF (198) to carry out the transcriptional program of early B cell specification with PAX5 (199, 200) maintaining B cell identity. As B cells develop in the bone marrow, they migrate from the endosteum to the central sinus of the bone marrow and become immature B cells when they acquire IgM. At this time, immature B cells exit the bone marrow and become transitional B cells (201, 202). In the periphery, transitional B cells

undergo the final stage of development into mature B cells and can differentiate into either marginal zone B cells or follicular B cells which occupy different niches in the peripheral lymphatics (203-205). The cues required to adopt the MZ instead of the FO B cell fate are the strength of tonic BCR signals, transcriptional programs from Notch2 (206), and interactions with Delta-Like-1 (207).

Millions of B cells are produced every day, but a small proportion of them enter into the mature pool. A tenth of developing B cells become immature B cells and only a third of those immature B cells will make it through the transitional B cell stage and become mature B cells (202). This loss is due to stringent selection based on BCR specificity at the immature and transitional stages (208-210). Negative selection and positive selection reduce the frequency of polyreactive and self-reactive specificities (209-211) while also selecting for BCRs with near threshold signal strength (212-215). Therefore, autoreactive specificities decrease with differentiation (216, 217), and even among peripheral B cells, transitional B cells are a source of autoreactive BCR specificities (218, 219).

### 3.2.3 Thymus dependent and independent B cell responses

The immunoglobulin constructed by recombination during development serves as the extracellular B cell receptor (BCR) driving antigen-specific B cell responses. The chemical structures recognized by the BCR include proteins, glycoproteins, polysaccharides, viral particles, and bacterial cells. Cross-linking of BCRs through binding to epitopes on these multivalent chemical structures leads to the internalization of the BCR as well as the bound antigen and initiates intracellular signaling events that culminate in the activation of nuclear transcription factors and gene expression for B cell

activation (reviewed in (220)). B cells also express PRRs internally and externally and can become activated by the binding of PRRs to PAMPs to produce an innate-like immune response or to tune the adaptive response (221). Thus, the internalization of antigen by BCR cross-linking serves two purposes: 1) to deliver antigen to the endocytic compartment for protein degradation and the loading of peptides onto MHC II and 2) to expose PAMPs to intracellular PRRs. The quality of the ensuing B cell response depends on the mode of interaction of the antigen with the BCR and the chemical composition of the antigen. The integration of these parameters follows the two-signal model whereby the cross-linking of the BCR provides Signal 1 and additional activation or survival signals delivered by other receptors provide Signal 2 (222). There are two major types of B cell responses that are largely determined by the nature of Signal 2. In thymus-dependent (TD) responses, Signal 2 is provided by antigen-specific CD4<sup>+</sup> T cells that bind peptide:MHC presented by B cells in the form of surface receptor engagement or cytokine secretion. In thymus-independent (TI) responses, Signal 2 can come from B cell engagement of PRRs with PAMPs (TI-1) or through extensive cross-linking of the BCR itself (TI-2) to foster a short-lived response composed of low affinity antibodies to microbial antigens (223, 224).

There are substantial differences between TI and TD responses in terms of antibody affinity, isotype production, effector differentiation, and memory cell generation, and different subsets of B cells contribute differentially to each type of response. In the B-2 lineage, follicular B cells make up the bulk of TD responses. Conversely, marginal zone B cells of the B-2 lineage and B-1 B cells primarily give rise to TI responses (225) likely because of the characteristics of their BCR signaling, their differentiative potential, and their enhanced sensitivity to detect PAMPs. Activated B cells in TI responses rapidly

differentiate into extrafollicular plasmablasts for the short-lived secretion of low-affinity antibody, primarily of the IgM isotype (226, 227). Conversely, B cell activity in TD responses is protracted and eventually produces IgG (228). Intriguingly, the antibody affinity of a TD response is not static but gradually increases over time in a process known as affinity maturation (229). As with TI responses, B cells in a TD response differentiate into extrafollicular PCs, relatively long-lived antibody secreting cells, and memory cell B cells (MBCs) (230-233). Some activated B cells in a TD response will migrate to the border between the follicle and the T cell zone to seek out cognate help from CD4 T cells primed by dendritic cells (234, 235). Once engaged, B cell proliferation continues and forms a germinal center, a transient structure composed of proliferating B cells, CD4 T cells, and other myeloid cells such as follicular dendritic cells.

The germinal center is a structure unique to TD responses and fosters the affinity maturation of germinal center B (GCB) cells (236, 237) and their differentiation to MBCs and LLPCs (238). The exchange of signals between B cells and T cells is limited to the CD4 T cells with specificity for the peptide antigens derived from the antigen internalized by the B cell (239, 240). Thus, cognate interactions are provided by the CD4 T cell and includes the engagement of CD40/CD40L and the production of key cytokines. These signals foster the upregulation of BCL6 in B cells to drive a germinal center transcriptional program (241-244). Another critical gene to be upregulated in GCB cells is activation-induced deaminase (AID), which creates point mutations in the variable (V) regions of the Ig (245, 246). The process of accumulating these point mutations is referred to as somatic hypermutation (SHM) and results in clonal variants with altered B cell receptors (247). An altered receptor can change the affinity and specificity of the receptor and thereby influence the fate of GCB cells. In the GC, clones compete for

survival signals and those with high-affinity receptors are preserved to continue SHM or to differentiate whereas the clones with low-affinity receptors die (248). SHM and affinity maturation is thought to be more efficient in GCs, but instances of SHM occurring in extrafollicular sites have been reported (249, 250). How selective competition is created between clones of differing affinities is still a matter of intense investigation, but clearly clones compete for antigen and the opportunity to present that antigen (251). In the cyclic reentry model, somatic hypermutation is thought to occur in the proliferating B cells of the dark zone of the GC before upregulating the altered BCR and migrating to the light zone (251). In the light zone, B cells find FDCs bearing antigen and compete with other clones for the antigen. Successful acquisition of antigen allows the B cell to present antigen to nearby CD4 T cells to receive survival signals and instructive cues. The CD4 T cell subset guiding the GC reaction is transcriptionally distinct from other T helper cells (252) and are known as T follicular helper ( $T_{FH}$ ) cells, which migrate to the border of the B cell follicle via CXCR5 (253) and select high affinity GC clones through BLyS secretion (239).

### 3.2.4 Plasma cells and memory B cells in humoral immunity

Humoral immunity depends on the continual production of circulating antibody as antibodies only have half-lives measured in days or weeks. Therefore, the differentiation of activated B cells into LLPCs or the sustained differentiation of SLPCs due to persistent antigen is critical for long term protective immunity (254-256). The differentiation of activated B cells into plasma cells depends on the integration of various instructive cues resulting in the upregulation of the transcription factor BLIMP-1 (257, 258) and extinguishing the B lineage transcription factor PAX5 (259, 260). These

transcription factor dynamics facilitate a gene expression program that direct B cells to terminally differentiate into plasma cells and acquire the specialized functions associated with antibody production. The heterogeneity in the lifespan of PCs has been elucidated in recent years as well. T cell independent PCs are longer lived than initially thought and the turnover kinetics in BM PC pools is more complex than previously appreciated (261, 262). Moreover, PCs are heterogeneous and have additional functions beyond antibody production including cytokine and antimicrobial secretion (263, 264).

Similarly, accumulating evidence shows that memory cells are not monolithic populations, but instead consist of functionally distinct subsets that play different roles in protective immunity. Thus, several subsets of memory T cells have been defined, reflecting differences in phenotype, function, and migration patterns (265, 266). Memory B cell (MBC) subsets have also been described based on differential expression of CD73, CD80 and PD-L2 (267); MBCs expressing both CD80 and PD-L2 form plasma cells upon re-challenge, whereas the double-negative cells join germinal centers (268).

Different memory fates can be determined by cytokine milieu, metabolic cues and transcriptional programs. For example, reciprocal patterns of T-bet and Eomesodermin expression underlie differentiation of T cells to effector versus memory subsets (269, 270). While the demarcation of T cell memory subsets by transcription factor expression is well established, analogous relationships have not been extensively explored in MBCs.

### 3.2.5 A T-bet expressing B cell subset accumulates during aging

The discovery of a T-bet<sup>+</sup> B cell subset in both mice and humans has piqued interest in the origin and role of these cells in primary and secondary humoral immune

responses. T-bet<sup>+</sup> B cells were first described in the context of murine aging and were thus termed “Age-associated B Cells,” or ABCs (271, 272). Subsequent analyses revealed roles for cognate T cell help and antigen presentation in their development. This, as well as a high frequency of somatically mutated immunoglobulin (Ig) genes in these cells, suggested that T-bet<sup>+</sup> ABCs are MBCs formed during T-dependent B cell responses (273). Whether T-bet<sup>+</sup> versus T-bet<sup>-</sup> MBCs differ in their origins, kinetics of generation, trafficking patterns, and functional roles remains unclear. We previously showed that T-bet<sup>+</sup> B cells appear and persist following influenza immunization or infection in mice (273, 274), providing a means to track T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs in a defined antigen system. Moreover, most humans have been exposed to influenza through immunization and infection and thus have standing influenza hemagglutinin (HA)-specific MBCs, enabling direct comparative analyses between human and murine MBC subsets.

### 3.2.6 Scope

Our results reveal multiple MBC subsets distinguished by T-bet expression, whose phenotypic and functional attributes are largely shared between mice and humans. We show that T-bet expression divides influenza-specific MBCs into T-bet<sup>-</sup>, T-bet<sup>lo</sup>, and T-bet<sup>hi</sup> populations with differing anatomic localization, residency patterns, and antigenic specificity. Upon infection, all three subsets are initially observed in draining lymph nodes, spleen, and infected tissues, whereas T-bet<sup>hi</sup> MBCs are selectively maintained in the spleen, remain resident, and are excluded from the lymphatics. In addition, B cell receptor sequencing shows that HA-specific T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs are largely clonally distinct, with infrequent sharing of clones. Divergence within clonal

lineages, in conjunction with genetic fate-mapping, demonstrates that T-bet expression in T-bet<sup>+</sup> MBCs is stable. Finally, we show in mice that T-bet expression in the B lineage is required for the development of HA-specific IgG2c and nearly all HA stalk-specific antibody. Together, these results establish T-bet expression as a distinguishing feature of MBC subsets that have profoundly different homing and functional properties and mediate distinct aspects of humoral immune memory.

### 3.3 RESULTS

#### 3.3.1 T-bet expression distinguishes unique influenza-specific memory B cell populations

Prior studies suggested T-bet-expressing B cells are an antigen-experienced population, but the functional differences between T-bet<sup>+</sup> and T-bet<sup>-</sup> MBC subsets remain unclear. Thus, we set out to define T-bet<sup>-</sup> and T-bet<sup>+</sup> B cell generation and persistence using influenza infection in a T-bet-ZsGreen reporter system (275). We infected T-bet-ZsGreen reporter mice with 30 TCID<sub>50</sub> of influenza A/Puerto Rico/8/1934 (PR8) and observed weight loss and recovery over a period of 4 weeks post infection, with the nadir at 9 days post infection (dpi; **Figure S7A**). We harvested mediastinal, mesenteric, and pooled peripheral (superficial cervical, axillary, brachial, and inguinal) lymph nodes (LN), spleen, lungs, and blood from infected mice at multiple time points, and identified influenza hemagglutinin (HA)-specific B cells using biotinylated PR8 HA probes modified to prevent sialic acid binding (276). The HA probes were separately conjugated to two streptavidin-fluorophore conjugates to exclude fluorophore-specific B cells during flow cytometric analysis (**Figure S7B**).

We identified low numbers of HA-specific B cells in lymphoid organs of naïve mice, in agreement with previous estimates of the primary HA-responsive repertoire (277); these were uniformly T-bet<sup>-</sup> (**Figure S7B**). To exclude this primary pool, we focused subsequent analyses on IgD<sup>-</sup> B cells (**Figure S7C**). IgD<sup>-</sup> HA-specific B cells were detected in spleen, mediastinal LN, and lungs of all mice at both acute infection and memory time points (**Figure 8A**). Examination of T-bet-ZsGreen and CD11c expression in HA-specific B cells indicated that T-bet<sup>+</sup> B cells can be phenotypically subdivided into T-bet<sup>hi</sup> and T-bet<sup>lo</sup> subsets with different tissues being variably comprised of T-bet<sup>-</sup>, T-bet<sup>lo</sup>, and T-bet<sup>hi</sup> subsets across infection (**Figure 8A**). The T-bet<sup>hi</sup> subset contained both CD11c<sup>+</sup> and CD11c<sup>-</sup> cells with a phenotype and level of T-bet expression matching Age-associated B Cells (**Figure S7C**). Furthermore, we confirmed that T-bet<sup>lo</sup> B cells expressed increased T-bet mRNA transcripts versus T-bet<sup>-</sup> B cells (**Figure S7D**).

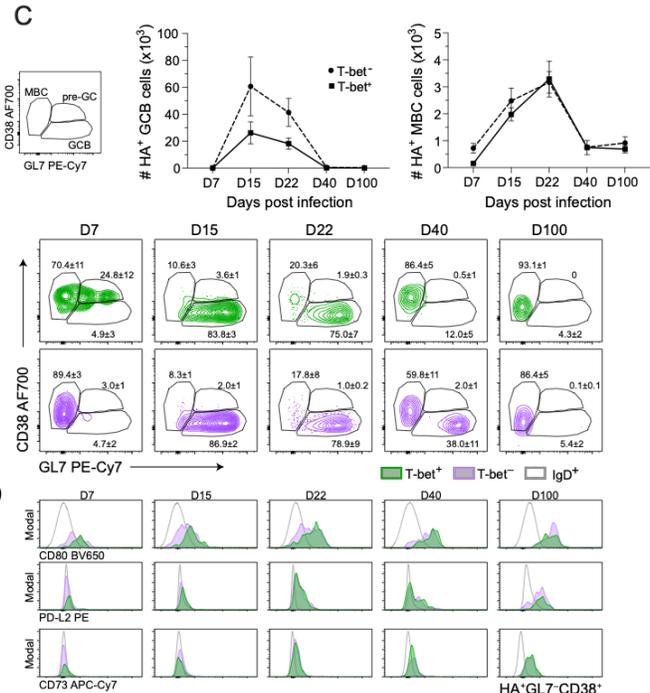
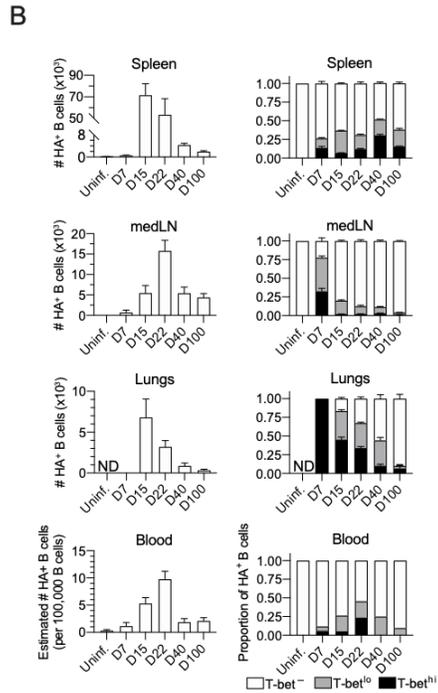
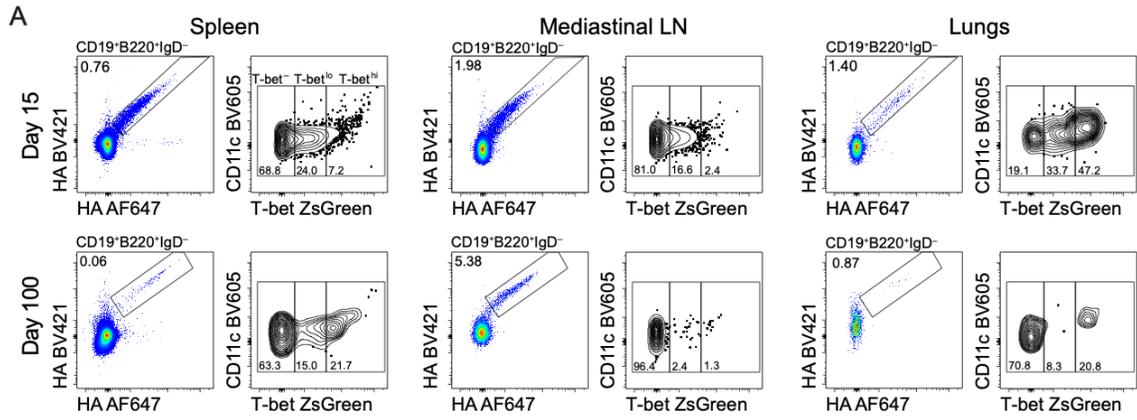
In agreement with prior studies (278, 279), HA-specific B cells were readily identified in spleen and mediastinal LN by 7 days post infection (dpi), peaked in number and frequency at 15 dpi in spleen and 22 dpi in mediastinal LNs, and then declined to steady state numbers in both organs by 40 dpi (**Figure 8B**). HA-specific B cells were occasionally detected in the lungs of some mice as early as day 7, but cell numbers peaked in lungs of all mice by 15 dpi and displayed a gradual decline continuing at least through 100 dpi (**Figure 8B**). Small numbers of HA<sup>+</sup> B cells were also detected in mesenteric and peripheral LNs, but these were dwarfed by spleen, mediastinal LN, and lung responses (**Figure S7E**).

Our longitudinal analysis of the HA-specific B cell pool identified differential induction and maintenance properties for the T-bet<sup>-</sup>, T-bet<sup>lo</sup>, and T-bet<sup>hi</sup> B cell subsets across tissues. The lung HA-specific response was entirely comprised of T-bet<sup>hi</sup> cells at

7 dpi; however, HA-specific cells were not detected in lungs of all mice at this time (**Figure 8B**). The lung HA-specific population remained T-bet-dominated throughout acute infection but was primarily T-bet<sup>-</sup> by 100 dpi (**Figures 8A and 8B**). The majority of mediastinal LN HA<sup>+</sup> B cells also expressed T-bet at 7 dpi, but T-bet<sup>hi</sup> cells rapidly declined by 15 dpi, suggesting rapid tissue exit or differentiation. T-bet<sup>lo</sup> B cells similarly declined by 15 dpi, albeit more slowly, and were nearly undetectable by 100 dpi in this tissue. In accord with possible tissue egress, blood T-bet<sup>hi</sup> cells peaked in frequency by 22 dpi and rapidly declined to undetectable levels by 40 dpi. In contrast, T-bet<sup>lo</sup> and T-bet<sup>hi</sup> subsets were consistently maintained in the spleen from 7 to 100 dpi comprising 27% to 52% (with an average of 37%) of the splenic HA-specific B cell response (**Figures 8A and 8B**). These findings identify early but transient T-bet<sup>+</sup> B cell responses in lungs and mediastinal LN and suggest T-bet<sup>+</sup> HA-specific B cell memory is primarily sustained in the spleen.

We also assessed germinal center B (GCB) cells and MBC marker expression in splenic HA-specific cells at each time-point. T-bet<sup>lo</sup> and T-bet<sup>hi</sup> cells were pooled for these analyses since they displayed similar memory marker expression throughout infection (data not shown). We delineated GCB cells as CD38<sup>-</sup>GL7<sup>+</sup> and presumptive MBCs as CD38<sup>+</sup>GL7<sup>-</sup> (280) (**Figure 8C**). Notably, GL7 expression closely correlated with other established GCB markers including CD95 and Peanut Agglutinin (PNA) (**Figure S7F**; data not shown). At 7 dpi, GL7 was present primarily on T-bet<sup>+</sup> cells without concomitant CD38 downregulation (**Figure 8C**), suggesting a pre-GC phenotype (281-283). Nearly all HA-specific cells exhibited a GCB phenotype by 15 dpi and maintained this through 22 dpi, at which time an MBC population begins to emerge (**Figure 8C**). At 40 dpi, the majority of T-bet<sup>+</sup> cells have a memory phenotype, whereas

nearly half of T-bet<sup>+</sup> cells still maintain GC markers (**Figure 8C**). Regardless of T-bet expression, nearly all splenic HA-specific B cells acquired an MBC phenotype by 100 dpi. In contrast, GCB cells persisted in the mediastinal LN, and to a lesser extent in the lung, out to 100 dpi; however, these were T-bet<sup>-</sup>. (**Figure S7G**). Further analyses of splenic MBC-phenotype cells at each time point identified upregulation of the previously characterized MBC markers CD80, PD-L2, and CD73 (268) beginning by 15-22 dpi and increasing further by 100 dpi, suggesting formation of stable T-bet<sup>-</sup> and T-bet<sup>+</sup> memory pools by the latter time point (**Figure 8D**). We also observed early expression of CD80 in CD38<sup>+</sup>GL7<sup>-</sup> cells as early as 7 dpi (**Figure 8D**); these may represent other non-GC cells such as extrafollicular plasmablasts, since they were not omitted by our gating strategy. These findings indicate that GC and MBC differentiation is similar between T-bet<sup>-</sup> and T-bet<sup>+</sup> subsets during the influenza response, except the T-bet<sup>+</sup> subset loses GC characteristics and transitions to a memory phenotype earlier than the T-bet<sup>-</sup> population. Moreover, T-bet<sup>hi</sup> HA-specific MBCs appear to be spleen-compartmentalized upon resolution of infection.



**Figure 8. T-bet expression identifies memory B cell populations with unique tissue distribution.**

T-bet-ZsGreen reporters were intranasally infected with 30 TCID<sub>50</sub> influenza A/Puerto Rico/8/1934 (PR8). **(A)** Fluorescently-conjugated PR8 hemagglutinin (HA) detects HA-specific (HA<sup>+</sup>) B cells, and T-bet-ZsGreen expression in HA<sup>+</sup> B cells resolves T-bet<sup>-</sup>, T-bet<sup>lo</sup>, and T-bet<sup>hi</sup> subsets across tissues at acute (day 15) and memory (day 100) timepoints. **(B)** Number of HA<sup>+</sup> B cells in spleen, pooled mediastinal lymph nodes (medLN), lungs, and blood at different time points after infection (left column), and proportions of HA<sup>+</sup> B cells that are T-bet<sup>-</sup>, T-bet<sup>lo</sup>, and T-bet<sup>hi</sup> in each tissue (right column). The number of HA<sup>+</sup> B cells in blood was estimated by calculating their frequency per 100,000 B cells, and proportions of T-bet-defined subsets in blood were calculated after concatenation due to low cell number. **(C)** Gating scheme identifies splenic HA<sup>+</sup> GCB cell (GL7<sup>+</sup>CD38<sup>-</sup>), MBC (GL7<sup>-</sup>CD38<sup>+</sup>), and pre-GC cell (CD38<sup>+</sup>GL7<sup>+</sup>) subsets; concatenated flow plots (bottom) depict CD38 and GL7 expression of T-bet<sup>+</sup> (pooled T-bet<sup>lo</sup> and T-bet<sup>hi</sup>; green) and T-bet<sup>-</sup> (purple) HA<sup>+</sup> B cells at each time point (bottom). Line plots (top) depict number of HA<sup>+</sup> GCB cells and MBCs separated by T-bet expression phenotype over time. **(D)** Expression of memory markers (CD80, PD-L2, CD73) in T-bet<sup>+</sup> (green) and T-bet<sup>-</sup> (purple) splenic HA<sup>+</sup> MBCs (GL7<sup>-</sup>CD38<sup>+</sup>) and naive follicular B cells (IgD<sup>+</sup>; grey). Data in **(B)** and **(C)** are compiled from 2 independent experiments with at least 3 mice per experiment. Data in **(A)** and **(D)** are representative of 2 independent experiments with at least 3 mice per experiment. Data in **(B)** and **(C)** are plotted as mean ± SEM. HA<sup>+</sup> B cells were identified as live, singlet, DUMP<sup>-</sup>, B220<sup>+</sup>, CD19<sup>+</sup>, IgD<sup>-</sup> cells, HA-BV421<sup>+</sup>, HA-AF647<sup>+</sup> cells. DUMP gate includes CD4, CD8, Gr-1, and F4/80.

**Table I: Number of hemagglutinin-specific B cells after influenza infection subset by level of T-bet expression. Related to Figure 8.**

Tissue	Days p.i.	T-bet <sup>-</sup>		T-bet <sup>lo</sup>		T-bet <sup>hi</sup>	
		Average	SEM	Average	SEM	Average	SEM
Spleen	0	<b>199</b>	60	<b>0</b>	0	<b>0</b>	0
	7	<b>565</b>	240	<b>87</b>	29	<b>91</b>	47
	15	<b>63,118</b>	20,498	<b>24,923</b>	5,878	<b>5,874</b>	770
	22	<b>36,188</b>	10,807	<b>10,795</b>	3,649	<b>6,468</b>	2,317
	40	<b>3,077</b>	2,056	<b>565</b>	176	<b>701</b>	224
	100	<b>1,937</b>	532	<b>862</b>	431	<b>603</b>	238
medLN	0	<b>14</b>	14	<b>0</b>	0	<b>0</b>	0
	7	<b>146</b>	119	<b>199</b>	167	<b>78</b>	58
	15	<b>4,283</b>	1,339	<b>1,057</b>	271	<b>162</b>	43
	22	<b>13,484</b>	1,735	<b>1,868</b>	956	<b>397</b>	143
	40	<b>5,520</b>	2,050	<b>557</b>	173	<b>104</b>	24
	100	<b>4,740</b>	2,027	<b>105</b>	38	<b>62</b>	15
Lungs	0	<b>0</b>	0	<b>0</b>	0	<b>0</b>	0
	7	<b>0</b>	0	<b>0</b>	0	<b>9</b>	6
	15	<b>1,529</b>	885	<b>2,844</b>	1,241	<b>3,798</b>	1,701
	22	<b>1,067</b>	293	<b>990</b>	186	<b>1,137</b>	347
	40	<b>531</b>	215	<b>382</b>	93	<b>87</b>	54
	100	<b>217</b>	112	<b>21</b>	21	<b>42</b>	42
mesLN	0	<b>6</b>	6	<b>0</b>	0	<b>0</b>	0
	7	<b>11</b>	5	<b>2</b>	2	<b>0</b>	0
	15	<b>55</b>	27	<b>20</b>	17	<b>5</b>	5
	22	<b>69</b>	16	<b>23</b>	7	<b>39</b>	19
	40	<b>70</b>	23	<b>26</b>	4	<b>26</b>	4
	100	<b>57</b>	17	<b>34</b>	11	<b>7</b>	5
pLN	0	<b>8</b>	5	<b>0</b>	0	<b>0</b>	0
	7	<b>15</b>	8	<b>2</b>	2	<b>10</b>	7
	15	<b>48</b>	23	<b>16</b>	13	<b>26</b>	11
	22	<b>50</b>	13	<b>41</b>	6	<b>30</b>	13
	40	<b>30</b>	13	<b>23</b>	8	<b>24</b>	11
	100	<b>38</b>	10	<b>26</b>	9	<b>0</b>	0

### 3.3.2 Human T-bet<sup>hi</sup> B cells are an anatomically compartmentalized component of influenza-specific memory

*The following section on the distribution of human B cell subsets is the work of James J. Knox who acquired this data as a PhD student in Michael Bett's lab and as a post-doc in the lab of Michael P. Cancro.*

Having identified discrete influenza-specific MBC subsets with differential tissue localization properties in mice, we questioned whether analogous human T-bet-expressing MBCs show a similar anatomical distribution. In humans, T-bet-expressing B cells have been identified alongside T-bet<sup>-</sup> MBCs in peripheral blood during active viral infections and vaccinations, malaria infection, and autoimmune disease (284-288). Since T-bet<sup>hi</sup> B cells (CD21<sup>-</sup>T-bet<sup>high</sup>) display a unique trafficking receptor profile (CD11c<sup>+</sup>CXCR3<sup>+/-</sup>CXCR5<sup>low</sup>CD62L<sup>low</sup>) compared to classical MBCs (287, 288), we hypothesized, as in mice, human T-bet<sup>hi</sup> B cells might have a distinct tissue distribution pattern. To test this, we obtained peripheral blood, tonsil, iliac and mesenteric lymph nodes, spleen, and bone marrow as donated or discarded surgical tissue and examined B cell phenotypes in these tissues (gating in **Figure S8A**). In agreement with our previous study (288), we observed T-bet<sup>hi</sup> B cells in the peripheral blood of all subjects (**Figures 9A and 9B**). We also readily identified T-bet<sup>hi</sup> B cells within spleen and bone marrow compartments (**Figures 9A and 9B**) and confirmed their antigen-experienced phenotype in spleen (**Figure S8B**). Conversely, T-bet<sup>hi</sup> B cells were largely absent from tonsils and both iliac and mesenteric lymph nodes (**Figures 9A and 9B**).

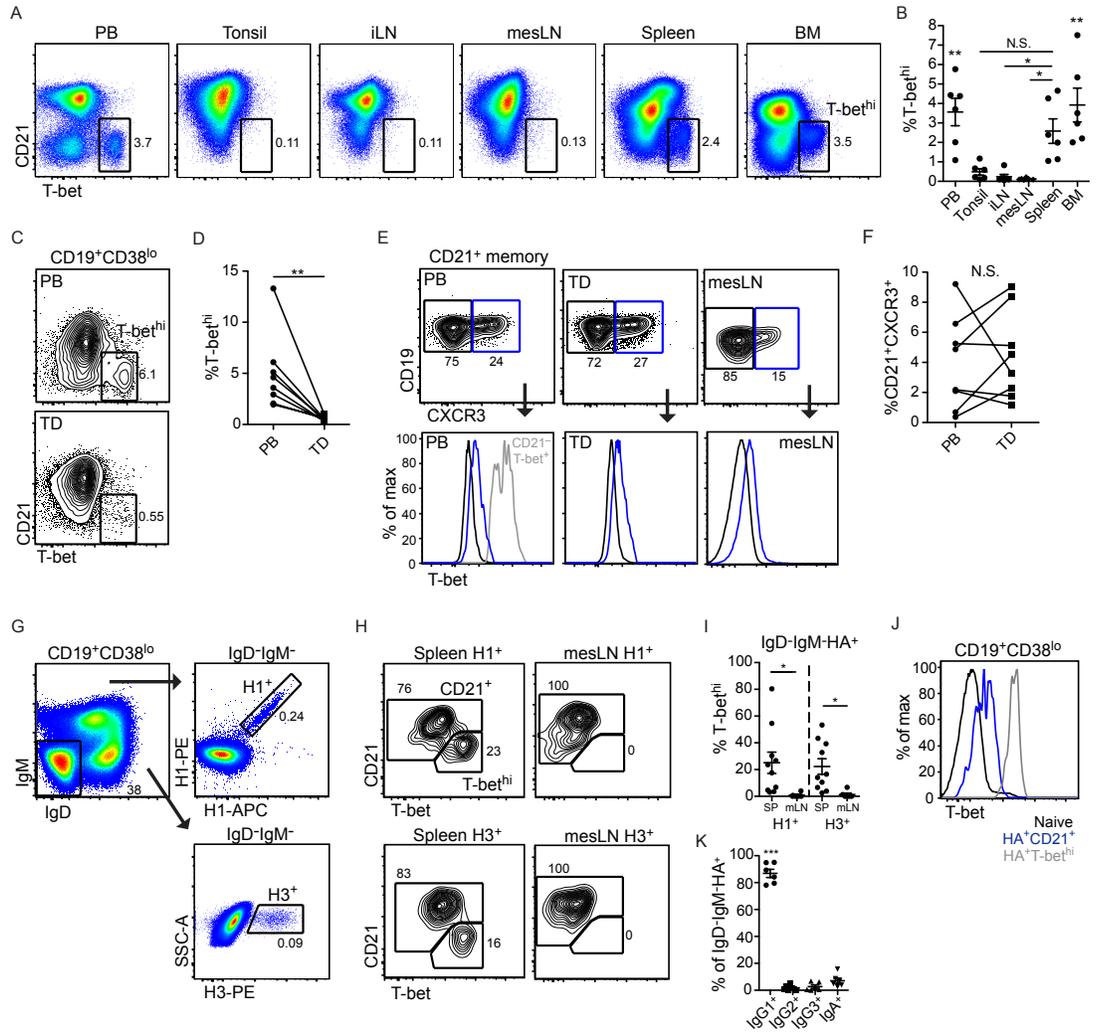
These findings suggested restricted trafficking properties of the T-bet<sup>hi</sup> B cell subset. To determine whether human peripheral blood T-bet<sup>hi</sup> B cells recirculate through tissues, we obtained matched peripheral blood and thoracic duct fluid (chyle) samples

from individuals undergoing thoracic duct cannulation. The thoracic duct is the body's largest lymphatic vessel that collects draining lymph from both lymphoid and non-lymphoid tissues for return to the blood; thus, its contents represent cells undergoing lymphatic recirculation. Despite consistent identification of T-bet<sup>hi</sup> B cells in the peripheral blood of these individuals, these cells were essentially absent in matched thoracic duct fluid (**Figures 9C and 9D**). In contrast, naïve B cells and various CD21<sup>+</sup> MBC subsets (IgM<sup>+</sup>CD27<sup>+</sup>, IgM<sup>-</sup>CD27<sup>+</sup>, and IgM<sup>-</sup>CD27<sup>-</sup>) were detected at similar frequencies in both peripheral blood and thoracic duct fluid of all subjects, suggesting this anatomical compartmentalization is a feature specific to T-bet<sup>hi</sup> B cells, analogous to what we had observed in mice (**Figures S8C-E**).

We next asked whether T-bet expression per se is associated with tissue restriction. We previously identified two distinct populations of T-bet-expressing B cells in blood of healthy individuals: T-bet<sup>hi</sup> B cells (CD21<sup>-</sup>T-bet<sup>high</sup>) and T-bet<sup>lo</sup> B cells (CD21<sup>+</sup>T-bet<sup>low</sup>; (288)), which likely correspond to the observed T-bet<sup>hi</sup> and T-bet<sup>lo</sup> MBC pools in mice (**Figure 8**). Here, we extended these findings to show that CXCR3 expression enriches for T-bet<sup>lo</sup> cells within the greater CD21<sup>+</sup> memory population (**Figure 9E**). Using the CD21<sup>+</sup>CXCR3<sup>+</sup> phenotype, we find that T-bet<sup>lo</sup> B cells are present in human blood, thoracic duct fluid, and lymph nodes (**Figure 9E**), and at similar frequencies between matched blood and thoracic duct fluid samples (**Figure 9F**). Thus, human T-bet<sup>lo</sup> B cells recirculate through all lymphoid tissues, while T-bet<sup>hi</sup> B cells are restricted to the spleen, blood, and bone marrow in healthy individuals. Further analyses will be necessary to determine the relationship between T-bet<sup>lo</sup> and T-bet<sup>hi</sup> B cells; however, we propose that these CD21<sup>+</sup>T-bet<sup>lo</sup> cells represent the human equivalent of the T-bet<sup>lo</sup> B cells observed in T-bet-ZsGreen reporter mice (**Figure 8**). Taken together, these findings identify

human T-bet<sup>hi</sup> B cells as a unique, tissue-restricted subset that does not recirculate via the lymphatic system.

Since murine T-bet<sup>hi</sup> HA-specific B cells preferentially populate the spleen at memory time points (**Figure 8**), we next asked whether the human spleen harbors an HA-specific T-bet<sup>hi</sup> MBC population. Using fluorophore-conjugated HA probes from two distantly-related influenza strains, A/California/07/2009 (H1) and A/Wisconsin/67/2005 (H3) (276, 289), we assessed T-bet expression by HA-specific class-switched (IgD<sup>-</sup>IgM<sup>-</sup>) B cells in the spleen (**Figure 9G**) and mesenteric lymph nodes (**Figure S8F**). Despite our efforts, we were unable to obtain human mediastinal lymph node samples without significant blood contamination for analysis of lung-draining lymphoid tissue. HA-specific T-bet<sup>hi</sup> B cells recognizing H1 or H3 strains were identified in the spleens of all donors but were rarely detected in mesenteric lymph nodes, whereas T-bet<sup>lo</sup>CD21<sup>+</sup> and T-bet<sup>-</sup>CD21<sup>+</sup> HA-specific memory B cells were present in all assessed tissues (**Figures 9H-J**; data not shown). The relative representation of T-bet<sup>hi</sup> B cells within the splenic HA-specific population varied considerably (~3-80% of H1<sup>+</sup> B cells and ~3-53% of H3<sup>+</sup> B cells; **Figure 9I**) and positively correlated with age (**Figures S8G and S8H**). Lastly, we assessed the isotype distribution of the human splenic HA-specific MBC compartment and found that human IgG1, the analog of murine IgG2a/c, dominated the class-switched memory response to influenza (**Figure 9K**; isotype gating in **Figure S8I**). IgG3<sup>+</sup> and IgA<sup>+</sup> HA-specific cells could be detected at low levels in some donors; however, IgG2<sup>+</sup> HA-specific B cells were rarely identified (**Figure 9K, Figure S8I**). Together, these findings suggest T-bet expressing B cells are a critical component of human influenza HA-specific B cell memory and, as in mice, identify the human T-bet<sup>hi</sup> HA-specific MBC pool as spleen-localized and absent from lymphatics.



**Figure 9. Human T-bet<sup>hi</sup> B cells do not recirculate via the lymphatics and maintain influenza-specific memory in the spleen.**

(A) Identification of human CD21<sup>-</sup>T-bet<sup>hi</sup> B cells within total CD19<sup>+</sup> B cells from peripheral blood (PB), tonsil, iliac lymph node (iLN), mesenteric lymph node (mesLN), spleen, and bone marrow (BM) of representative donors. Different tissue types in (A) or (B) are not matched. (B) Frequency of T-bet<sup>hi</sup> B cells in various tissues (n=6 per tissue group). Statistics represent comparisons between PB, spleen, or BM with tonsil, iLN, and mLN; frequencies within PB, spleen, and BM are not statistically different from one another. (C) Identification of T-bet<sup>hi</sup> B cells in matched peripheral blood (PB) and thoracic duct fluid (TD) samples from a representative donor. (D) Frequency of T-bet<sup>hi</sup> B cells in matched PB and TD samples (n=8). (E) Identification of CD21<sup>+</sup>CXCR3<sup>+</sup>T-bet<sup>lo</sup> (blue) and CD21<sup>+</sup>CXCR3<sup>-</sup>T-bet<sup>-</sup> (black) subsets of memory (IgD<sup>-</sup>/IgD<sup>+</sup>CD27<sup>+</sup>) B cells in matched PB and TD from a representative donor, and mesLN from another donor; T-bet expression by these populations is shown as a histogram. Blood T-bet<sup>hi</sup> B cells are included for comparison in grey. (F) Frequency of the CD21<sup>+</sup>CXCR3<sup>+</sup> population within PB and TD CD19<sup>+</sup> B cell pools from an 8-donor cohort. (G) Identification of HA-specific, IgD<sup>-</sup>IgM<sup>-</sup> B cells within CD19<sup>+</sup>CD38<sup>low</sup> splenic B cells using two fluorescently-labelled A/California/07/2009 HA probes (H1 strain) or a single fluorescently-labelled A/Wisconsin/67/2005 HA probe (H3 strain). (H) CD21 and T-bet expression in IgD<sup>-</sup>IgM<sup>-</sup> HA<sup>+</sup> B cells in spleens and mesLNs from representative donors using H1 or H3 probes. (I) Frequency of T-bet<sup>hi</sup> phenotype within IgD<sup>-</sup>IgM<sup>-</sup>H1<sup>+</sup> or H3<sup>+</sup> B cells in spleens from two 10-donor cohorts and mLN from a 6-donor cohort. (J) T-bet MFI of splenic naïve (IgD<sup>+</sup>CD27<sup>-</sup>) B cells and switched (IgD<sup>-</sup>IgM<sup>-</sup>) H1-HA-specific CD21<sup>+</sup> and CD21<sup>-</sup>T-bet<sup>hi</sup> B cells from a representative donor. (K) Frequency of isotype expression within human splenic IgD<sup>-</sup>IgM<sup>-</sup>HA<sup>+</sup> B cells (n=6). Statistical comparisons performed using one-way ANOVA with Tukey post-test (B), paired t-test (D and F), unpaired t-test (I), and repeated measures ANOVA with Tukey post-test (K). Lines depict mean ± SEM. N.S. = not significant, \*p<0.05; \*\*p<0.01; \*\*\*p<0.001. *Credit goes to James J. Knox for generating this figure.*

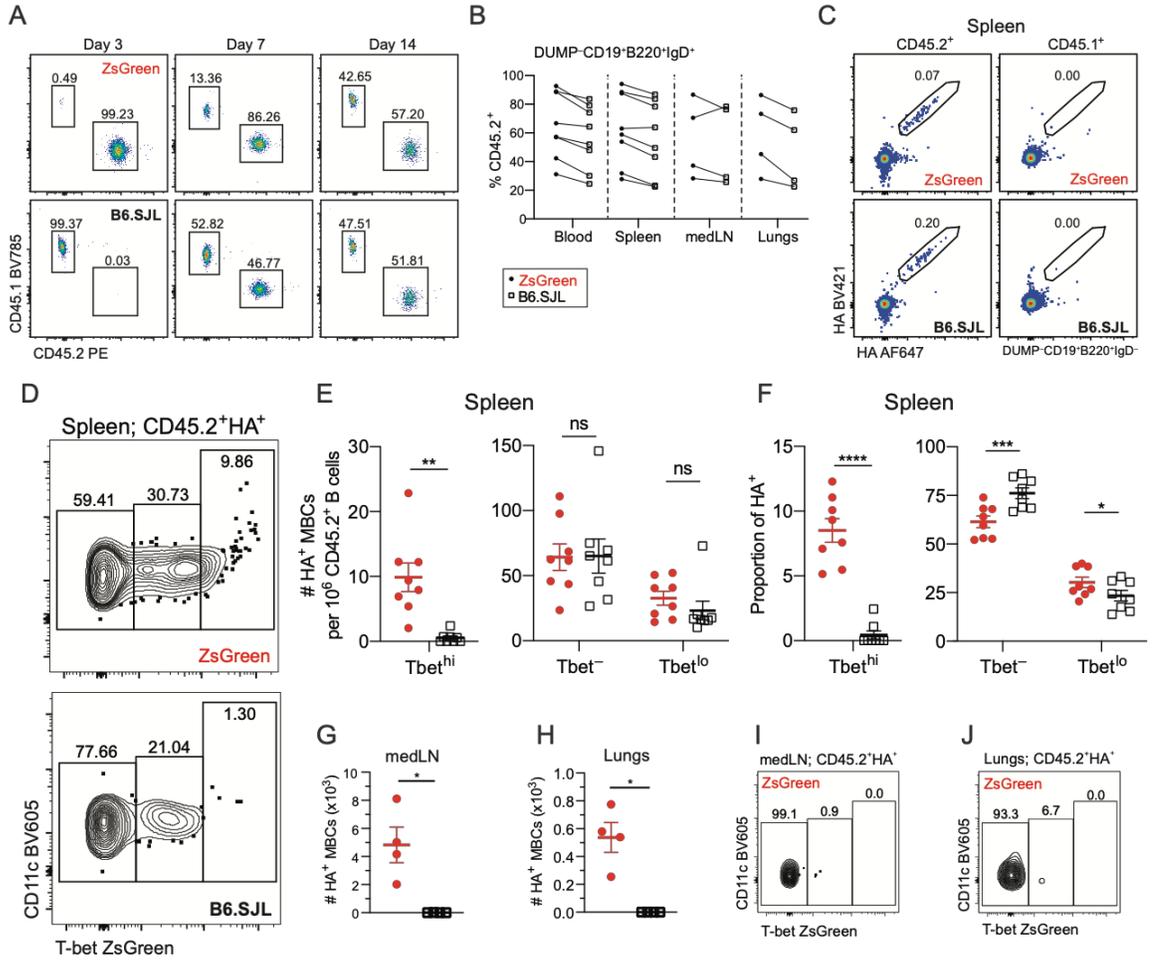
### 3.3.3 T-bet<sup>hi</sup> HA-specific memory B cells are resident in the spleen

The apparent splenic localization of T-bet<sup>hi</sup> HA-specific MBCs in both mice and humans led us to rigorously assess tissue residency using a parabiosis-based approach. Thus, T-bet-ZsGreen reporter mice infected  $\geq 40$  days prior with PR8 were surgically joined to naïve congenic B6.SJL partners. We reasoned that all HA-specific MBC will originate in the T-bet-ZsGreen partner, so their presence in the B6.SJL partner would indicate that they are a recirculating population. Conjoined mice were monitored by serial bleeds to assess the portion of circulating CD45.2<sup>+</sup> (T-bet-ZsGreen reporter origin) versus CD45.1<sup>+</sup> (B6.SJL origin) B cells in each partner. Both partners demonstrated mixing of CD45.1<sup>+</sup> and CD45.2<sup>+</sup> B cells in the blood as early as day 7, achieving stable proportions between the partners by day 14 (**Figure 10A**). Accordingly, parabiosed pairs were euthanized  $\geq 17$  days post-surgery. We observed similar frequencies of CD45.2<sup>+</sup>IgD<sup>+</sup> B cells – a pool anticipated to circulate freely – in spleen, lungs, and mediastinal lymph nodes of each partner (**Figure 10B**), suggesting equilibration of recirculating B cells by day 17.

HA-specific MBCs were observed in the spleens of both partners, and virtually all of these were CD45.2<sup>+</sup> (**Figure 10C**), consistent with their origin in the previously infected T-bet-ZsGreen partner. T-bet<sup>+</sup> and T-bet<sup>lo</sup> HA-specific B cells were identified in the spleens of both partners, suggesting these subsets recirculate (**Figures 10D-F, Figure S9A**). In contrast, T-bet<sup>hi</sup> HA-specific MBCs were absent from the naïve B6.SJL partner spleens but remained in spleens of previously infected T-bet-ZsGreen mice (**Figures 10E, 10F, S9A**), even when data were concatenated from 7 parabiotic pairs (**Figure 10D**). To confirm that ZsGreen-expressing cells were not being rejected in the B6.SJL mice, we measured frequencies of donor T-bet-ZsGreen<sup>+</sup>CXCR3<sup>+</sup>CD8<sup>+</sup>

lymphocytes, which highly express the ZsGreen protein, and found similar frequencies of these cells in both partners (**Figure S9B**). Moreover, T-bet<sup>lo</sup> HA-specific B cells were also present in spleens of both mice (**Figures 10D and 10E**). Thus, broad rejection of ZsGreen<sup>+</sup> cells is not occurring, consistent with previous studies (290). Taken together, these findings identify splenic T-bet<sup>hi</sup> HA<sup>+</sup> MBCs as a tissue-resident memory pool.

We also investigated whether HA-specific B cells showed evidence of residency in mediastinal LNs and lung, the other primary locations of influenza memory cells, as others recently demonstrated (291). We identified significant HA-specific B cell numbers in mediastinal LNs of the previously infected partner that were absent in the naïve partner (**Figure 10G**). We suspect this reflects the extended maintenance of HA-specific GCs in mediastinal LNs, as the local HA-specific B cell population retains a GC phenotype at least through 100 dpi (**Figure S7G**). This phenomenon appears to be mediastinal LN-specific, as HA-specific GC B cells were not identified at this late time point in any other lymphoid tissues examined. We also identified HA-specific MBCs in the lungs of the previously infected partner (**Figure 10H**), but these cells were absent in the naïve partner. Notably, nearly all mediastinal LN- and lung-localized cells were T-bet<sup>hi</sup> (**Figures 10I-J**). Thus, HA-specific memory is anatomically compartmentalized, encompassing tissue-resident T-bet<sup>hi</sup> B cells in the spleen and circulating T-bet<sup>lo</sup> and T-bet<sup>hi</sup> B cell populations.



**Figure 10. T-bet expression resolves spleen resident versus recirculating MBC pools.**

(A). T-bet-ZsGreen reporters (CD45.2<sup>+</sup>; ≥ 40 dpi) and naïve B6.SJL (CD45.1<sup>+</sup>) were surgically conjoined and showed evidence of blood sharing by day 7, with equilibrium reached by day 14. Parabionts were euthanized at ≥ 17 days post-surgery for analysis. (B) Frequencies of naïve follicular (IgD<sup>+</sup>) B cells expressing CD45.2 in lymphoid and non-lymphoid tissues from each parabiosis pair. (C) Identification of HA<sup>+</sup>IgD<sup>-</sup> B cells expressing either CD45.1 or CD45.2 in parabiosis partners. (D) Identification of T-bet-ZsGreen reporter-derived (CD45.2<sup>+</sup>) T-bet<sup>-</sup>, T-bet<sup>lo</sup>, and T-bet<sup>hi</sup> HA<sup>+</sup> MBCs in spleens of T-bet-ZsGreen and B6.SJL partners; data concatenated from 7 pairs. (E) Numbers of T-bet<sup>-</sup>, T-bet<sup>lo</sup>, and T-bet<sup>hi</sup> HA<sup>+</sup> splenic MBCs in T-bet-ZsGreen (red) and B6.SJL (black) partners. (F) Percentage of splenic HA<sup>+</sup> MBCs that are T-bet<sup>-</sup>, T-bet<sup>lo</sup>, or T-bet<sup>hi</sup> in each partner. (G and H) Number of HA<sup>+</sup> MBCs in medLN (G) and lungs (H) of parabiosis partners. (I and J) T-bet-ZsGreen expression in HA<sup>+</sup> MBCs from medLN (I) and lungs (J) of T-bet-ZsGreen partner. HA<sup>+</sup> MBCs were not detected in the medLN or lung of the B6.SJL partner. Data displayed are from 8 pairs across three independent experiments for spleen and 4 pairs across two-independent experiments for medLN and lungs. HA<sup>+</sup> B cells were identified as live, singlet, DUMP<sup>-</sup>, B220<sup>+</sup>, CD19<sup>+</sup>, CD45.2<sup>+</sup>, IgD<sup>-</sup>, HA-BV421<sup>+</sup>, HA-AF647<sup>+</sup> cells. Data in (E), (F), (G), and (H) show individual points with the mean ± SEM indicated. Statistical comparisons performed using paired two-tailed t-test. ns = not significant, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.001

### 3.3.4 Established T-bet<sup>+</sup> and T-bet<sup>-</sup> memory B cells undergo minimal interconversion

The different residency and recirculation properties of T-bet<sup>-</sup> and T-bet<sup>+</sup> MBCs raised the question of how T-bet<sup>-</sup> and T-bet<sup>+</sup> MBCs arise. We considered four possibilities (**Figure S10A**): 1) T-bet<sup>-</sup> and T-bet<sup>+</sup> MBCs arise independently and are stable, separate subsets; 2) T-bet<sup>+</sup> cells give rise to T-bet<sup>-</sup> cells (or vice versa); 3) T-bet<sup>+</sup> and T-bet<sup>-</sup> cells undergo shared selection followed by stable commitment to either a T-bet<sup>+</sup> or T-bet<sup>-</sup> long-lived MBC population; 4) T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs interconvert by modifying T-bet expression as needed to change localization or functional properties. Immune repertoire profiling of antibody heavy chain variable region gene (VH) rearrangements can be used to distinguish between these four models, as each model makes distinct predictions regarding differences in VH usage (model 1), somatic hypermutation (model 2) and clonal overlap (models 3 and 4; **Figure S10A**).

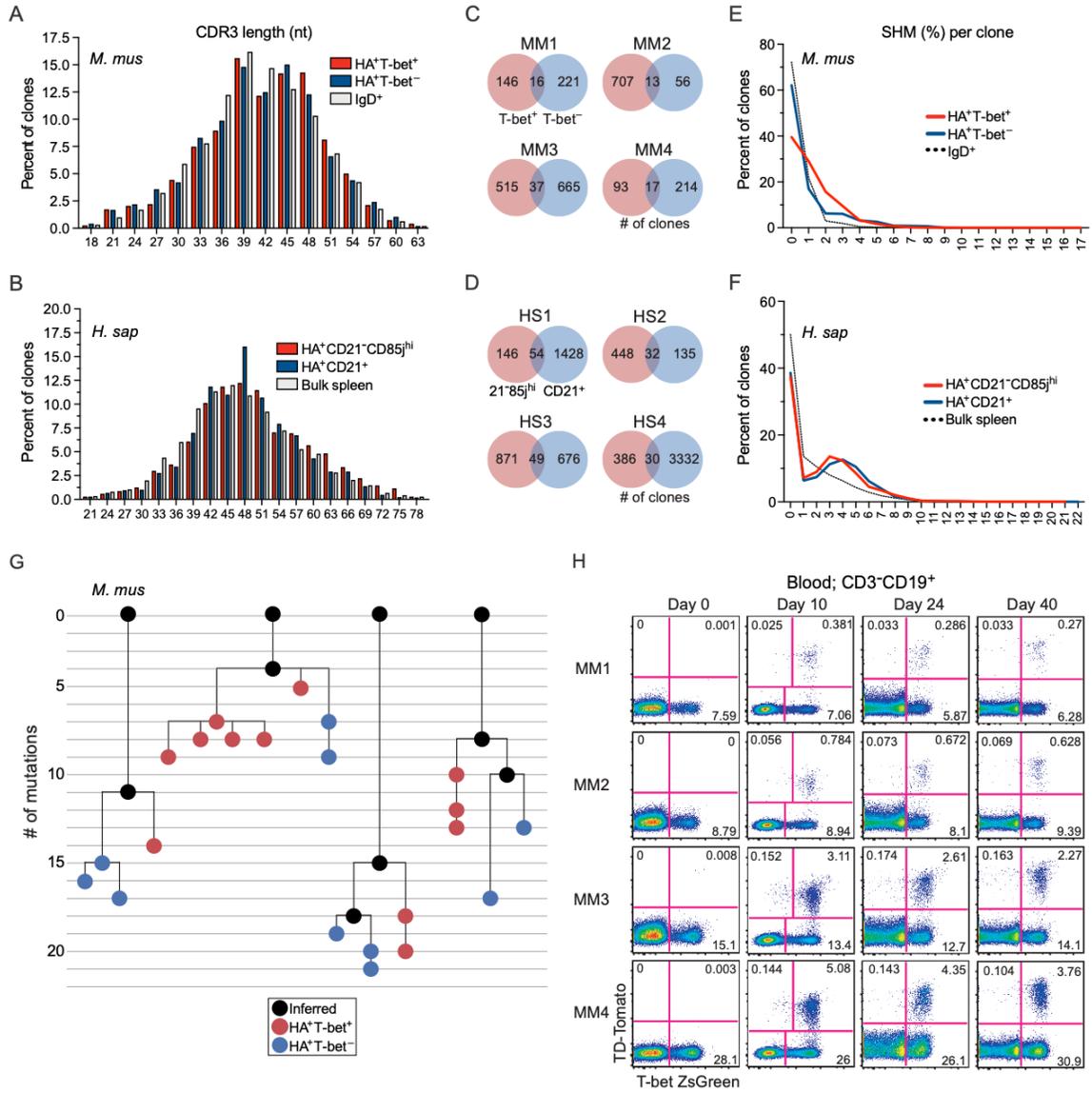
We therefore sequenced VH rearrangements of HA-specific MBCs separated by T-bet expression (**Figure S10B**) in T-bet-ZsGreen reporter mice immunized intradermally with 30 µg of lipid nano particles (LNP) loaded with HA mRNA (292). The LNP platform generates both T-bet<sup>-</sup> and T-bet<sup>+</sup> MBCs (**Figure S10B**), and we confirmed an HA-antibody response by hemagglutination inhibition (**Figure S10C**). At 90 days post immunization, all splenic IgD<sup>-</sup> HA-specific B cells were sorted based upon T-bet-ZsGreen expression into T-bet<sup>-</sup> and T-bet<sup>+</sup> memory subsets, and antibody VH rearrangements of both subsets and IgD<sup>+</sup> naïve follicular control B cells were sequenced. Similarly, we used the CD21<sup>-</sup>CD85j<sup>hi</sup> surface phenotype, which specifically identifies human T-bet<sup>hi</sup> B cells (288), to sort human splenic HA-specific IgD<sup>-</sup>IgM<sup>-</sup> MBCs into CD21<sup>-</sup>CD85j<sup>hi</sup> and CD21<sup>+</sup> subsets (**Figure S10D**) and sequenced the VH rearrangements of these populations along with control bulk splenocytes.

VH analysis revealed comparable VH usage (**Figures S10E and S10F**) and CDR3 length distributions (**Figures 11A and 11B**) in T-bet<sup>+</sup> and T-bet<sup>-</sup> populations, and there was some clonal overlap between the two populations (**Figures 11C and 11D**), ruling out a strict separate lineage model (model 1). T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs harbored similar levels of somatic hypermutation, suggesting that one population was not a precursor to the other (model 2), and both populations showed significantly more mutations than naïve B cells or unsorted splenocytes (**Figures 11E and 11F**).

Next, we scrutinized the lineage trees of clones that contained T-bet<sup>+</sup> and T-bet<sup>-</sup> members. In mice, overlapping clones between T-bet<sup>+</sup> and T-bet<sup>-</sup> populations were not as frequent as they were within replicate sequencing libraries from the same subset (**Figure S10G**), suggesting that many T-bet<sup>+</sup> and T-bet<sup>-</sup> clones arise independently, rather than being fully intermingled. The same clonal analysis in humans did not reach statistical significance, likely due to our restricted sample size – we sampled a small portion of the spleen and therefore missed many clonal members (**Figure S10H**). In further support of this separation, analyses of mouse clonal lineages containing both T-bet<sup>+</sup> and T-bet<sup>-</sup> cells revealed that nearly all exhibited segregation of T-bet<sup>+</sup> and T-bet<sup>-</sup> sequences onto separate branches (**Figure 11G**). Taken together, these findings in mice following immunization and in established HA<sup>+</sup> MBCs in human spleen favor model 3, in which T-bet<sup>+</sup> and T-bet<sup>-</sup> MBC precursors undergo shared selection, subsequently commit to a T-bet<sup>+</sup> or T-bet<sup>-</sup> MBC population, and thereafter remain stable with respect to T-bet expression status, with minimal, if any, interconversion between established T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs.

To further verify the stability of T-bet<sup>+</sup> B cells, we used a combined reporter/fate mapper mouse (293). These T-bet-sufficient mice contain the ZsGreen construct fused

to cre<sup>ERT2</sup> under control of the T-bet promoter, such that treatment with tamoxifen during active T-bet transcription causes irreversible tdTomato expression. Using these mice, one can delineate cells that expressed T-bet during the tamoxifen treatment period and have subsequently lost expression (tdTomato<sup>+</sup>T-bet<sup>-</sup>ZsGreen<sup>-</sup>) from those that retained it (tdTomato<sup>+</sup>ZsGreen<sup>+</sup>). We treated ≥ 20-week-old T-bet-ZsGreen/tdTomato mice with tamoxifen on days 0, 2, and 4 and performed serial bleeds to assess stability of tdTomato<sup>+</sup>ZsGreen<sup>+</sup> B cells. All mice demonstrated tdTomato labeling at day 10, with T-bet<sup>+</sup> cells outnumbering T-bet<sup>-</sup> cells 10:1 within the tdTomato<sup>+</sup> B cell population (**Figure 11H**). The ratio of T-bet<sup>+</sup> to T-bet<sup>-</sup> cells was maintained steadily in all mice (**Figure 11H**), suggesting most B cells expressing T-bet at day 0 maintained expression for 40 days, interconverting rarely if at all during this period. In combination with our clonal overlap and lineage tree analyses, these data show that established T-bet<sup>+</sup> MBCs represent a separate, stable population.



**Figure 11. T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs are selected from a shared pre-immune lineage but do not interconvert.**

HA-specific splenic MBCs from T-bet-ZsGreen reporters (day 100 post immunization) were sorted into T-bet<sup>-</sup> and T-bet<sup>+</sup> subsets, with naïve follicular (IgD<sup>+</sup>) B cell controls, for immunoglobulin heavy chain genomic sequencing. Human HA-specific splenic MBCs were similarly sorted into CD21<sup>+</sup> and CD21<sup>-</sup>CD85<sup>hi</sup> subsets; CD21<sup>-</sup>CD85<sup>hi</sup> phenotype identifies human T-bet<sup>hi</sup> B cells (288) subsets. **(A)** CDR3 lengths (in nucleotides) of in-frame sequences from murine T-bet<sup>-</sup> and T-bet<sup>+</sup> HA<sup>+</sup> MBCs and naïve follicular (IgD<sup>+</sup>) B cell controls after all replicates were pooled. **(B)** CDR3 lengths of in-frame sequences from CD21<sup>+</sup> and CD21<sup>-</sup>CD85<sup>hi</sup> HA<sup>+</sup> MBC subsets were quantified (in nucleotides). Bulk splenocytes (largely naïve follicular B cells) served as a control. **(C)** The number of clones that overlap between T-bet<sup>-</sup> (blue) and T-bet<sup>+</sup> (red) HA<sup>+</sup> MBCs in mouse (*M. mus*, MM). **(D)** The number of clones that overlap between CD21<sup>+</sup> (blue) and CD21<sup>-</sup>CD85<sup>hi</sup> (red) HA<sup>+</sup> MBCs in humans (*H. sap*; HS). **(E)** Percentages of clones binned by the level of somatic mutation (expressed as the percent difference in nucleotide sequence to the nearest germline VH gene) in mouse T-bet<sup>-</sup> and T-bet<sup>+</sup> HA<sup>+</sup> MBCs and naïve follicular B cells. **(F)** Percent of the heavy chain V-gene that is mutated from germline in CD21<sup>+</sup> and CD21<sup>-</sup>CD85<sup>hi</sup> HA<sup>+</sup> MBCs and bulk splenocytes in humans. **(G)** Representative lineage trees of shared clones between T-bet<sup>-</sup> and T-bet<sup>+</sup> HA<sup>+</sup> murine MBCs, with inferred nodes (black), T-bet<sup>-</sup> nodes (blue), and T-bet<sup>+</sup> nodes (red). Trees were generated in ImmuneDB and visualized with ETE3 (see Methods). Lineages had to contain at least 10 copies of T-bet<sup>+</sup> and T-bet<sup>-</sup> and have at least 4 trunk mutations (shared SHMs) to be included. Numbers indicate the number of mutations compared to the preceding vertical node. The inferred node at the top of the tree indicates the nearest germline sequence. **(H)** T-bet-ZsGreen fate-mapping mice (293) were treated with tamoxifen to mark T-bet expressing cells with permanent, Rosa21-driven, tdTomato expression and the status of T-bet expression of marked B cells in the blood was tracked over 40 days. For panels **(A)**, **(C)**, **(E)**, and **(G)**, two independent experiments were carried out with at least 4 mice per group. Each gave similar results, and the results for the more recent experiment are shown. For panels **(B)**, **(D)**, and **(F)**, the splenocytes from 4 adult subjects were sorted and sequenced. For genetic fate mapping **(H)**, two independent experiments were carried out with at least 4 mice per group; one experiment is shown here.

### 3.3.5 HA stalk-specific antibody is derived primarily from the T-bet-expressing B cell compartment

The distinct localization and phenotypic stability of T-bet<sup>+</sup> HA-specific MBCs led us to assess the contribution of the T-bet-expressing B cell compartment to the influenza humoral response. In mice, T-bet promotes antibody class-switching to IgG2a/c (294-301), the dominant isotype in influenza and other anti-viral responses (298, 302). In accordance with this, we observed a greater increase in PR8-specific and HA-specific IgG2c (**Figures 12A and 12B**) compared to IgG1, evident by day 12 and 15, respectively. This isotype bias confirmed previous studies (279) and suggested a key role for T-bet in regulating influenza antibody production (303). Therefore, we tested the contribution of T-bet in the B lineage to HA-specific humoral responses by infecting B cell-specific conditional T-bet knockouts (T-bet<sup>flox/flox</sup> CD19<sup>Cre/WT</sup>; hereafter referred to as cKOs), CD19-Cre controls (CD19<sup>Cre/WT</sup> T-bet<sup>WT/WT</sup>), and wild type mice with PR8 and examined antibody levels and characteristics. All three groups showed similar weight loss kinetics (**Figure 12C**), total HA-specific B cell numbers (**Figure 12D**), and HA-specific GCB cell numbers (**Figure 12E**), although CD19-cre controls recovered weight more quickly (**Figure 12C**). As such, CD19 heterozygosity does not appear to significantly impair the influenza response, and initiation of the humoral response does not require T-bet expression in B cells.

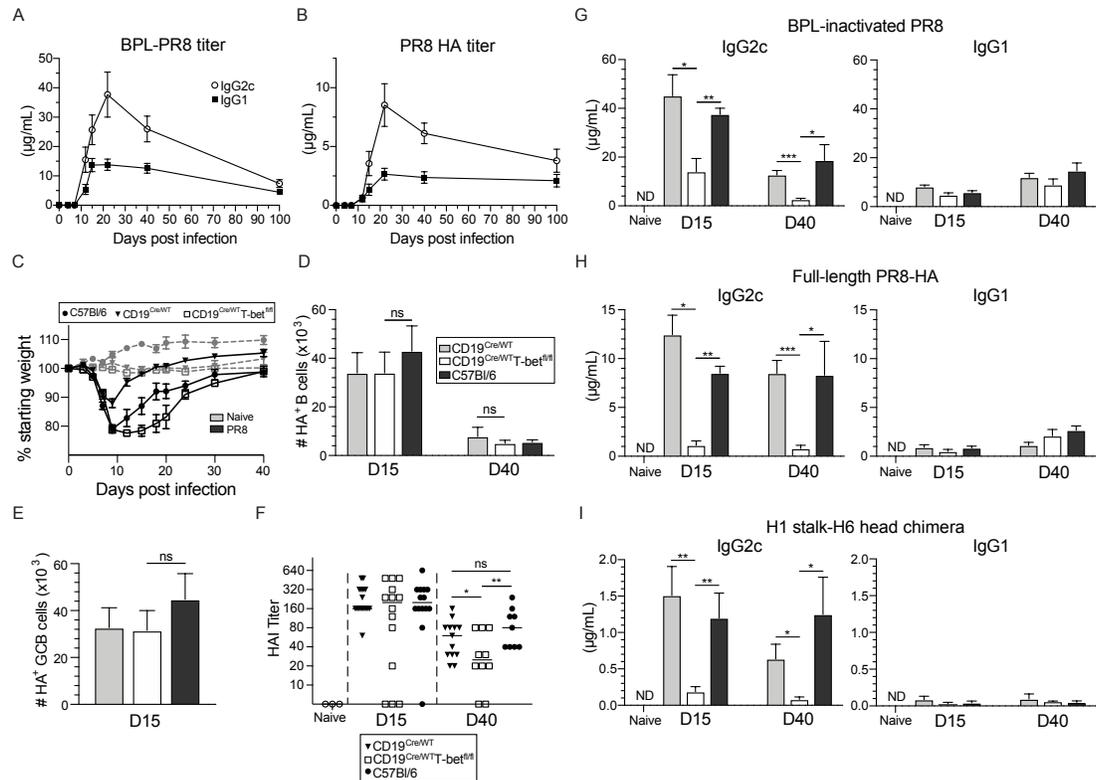
To assess the functionality of antibodies generated in the absence of B lineage T-bet expression, we performed hemagglutination inhibition (HAI) assays from serum. At 15 dpi, the majority of mice displayed HAI titers greater than 40, the level associated in human studies with protection (304, 305), although one wild type and several cKO mice had titers ranging from 20 to undetectable (**Figure 12F**). HAI titers declined in all groups

by 40 dpi (**Figure 12F**), likely due to the loss of acute infection-generated IgM titers (302). However, cKOs displayed significantly reduced HAI titers versus the wild type and CD19-Cre groups at 40 dpi, with 70% of mice showing titers below 40 (**Figure 12F**). These findings suggest T-bet expression in B cells may be necessary for the development of sustained protective influenza-specific titers.

We hypothesized that decreased HAI titers in cKOs may reflect a loss of specific components of the antibody response. We next assessed antibody titers and found a significant reduction in total PR8-specific IgG2c in cKOs, as expected (**Figure 12G**). Low IgG2c titers remained in cKOs at 15 dpi but were nearly absent by 40 dpi, suggesting T-bet-independent mechanisms can initiate a degree of IgG2c switching during acute infection (**Figure 12G**). We focused subsequent analyses on the HA protein, the antigenic target relevant for protective humoral immunity to influenza and identified significantly reduced IgG2c titers to full-length HA in cKOs at both 15 and 40 dpi compared to wild type and CD19-Cre control groups (**Figure 12H**). PR8- and HA-specific titers of IgG1, a T-bet-independent isotype, were unaffected in cKOs and did not increase to compensate for IgG2c loss (**Figures 12G and 12H**). These findings confirm that the majority of HA-specific IgG2c antibody is derived from T-bet-expressing B cells.

Lastly, we questioned whether T-bet expressing B cells are important for influenza antibody responses to certain specificities. Recent studies highlight a critical role for HA-specific IgG2a/c antibodies for *in vivo* influenza protection, which are primarily skewed toward stalk recognition (306, 307). Thus, we assessed stalk reactivity of IgG1 and IgG2c using a chimeric construct comprised of the PR8-related H1 stalk and unrelated H6 head (308-310). This chimera is bound primarily by stalk-specific antibodies since most PR8-generated HA head-binding antibodies are strain-specific

and do not recognize H6 head. We found that the stalk response is dominated by IgG2c in wild type mice at both 15 and 40 dpi, while IgG1 stalk titers were negligible (**Figure 12I**). Moreover, cKOs largely lost IgG2c stalk-reactive titers (**Figure 12I**), indicating that the bulk of the influenza stalk-specific antibody response arises from T-bet-expressing B cells.



**Figure 12. T-bet<sup>+</sup> B cells are required for optimal influenza antibody responses and HA stalk-specific antibody in mice.**

(A and B) Total betapropiolactone (BPL)-inactivated PR8-specific IgG1 and IgG2c (A) and PR8 hemagglutinin (HA)-specific IgG1 and IgG2c (B) in sera from infected T-bet-ZsGreen mice over time. (C) Weight loss and recovery from influenza infection in wild type C57Bl/6, CD19<sup>cre/WT</sup>T-bet<sup>fl/fl</sup>, and CD19<sup>cre/WT</sup> mice compared to PBS-treated controls. (D) Number of HA-specific splenic B cells at day 15 and 40 dpi. (E) Number of HA-specific splenic GCB cells at 15 dpi. (F) Hemagglutination inhibition (HAI) titers at 15 and 40 dpi. (G-I) Antibody titers to BPL-inactivated PR8 (G), full-length PR8-HA (H), or chimeric construct comprised of H1 stalk and H6 head (I). Wild type C57Bl/6 were used for naïve controls in (F-I). Data are represented as mean ± SEM from 3 independent experiments with at least 3-5 mice in each group. Statistical comparisons performed using two-sided t-test (G-I) and Wilcoxon rank-sum test (F). \*p<0.05, \*\*p<0.01, \*\*\*p<0.001. Cells in (D, E) were identified as DUMP<sup>-</sup>, CD19<sup>+</sup>, B220<sup>+</sup>, CD138<sup>-</sup>, IgD<sup>-</sup>, HA-PE<sup>+</sup>, with the additional definition of GC cells in (E) as PNA<sup>+</sup>CD95<sup>+</sup>.

Table II: Serum antibody titers to various influenza antigens after infection. Related to Figure 12.

Antigen	Isotype ( $\mu\text{g/mL}$ )	Day 15 p.i.			Day 40 p.i.		
		CD19 <sup>cre/WT</sup>	<b>CD19<sup>cre/WT</sup>Tbet<sup>fl/fl</sup></b>	C57Bl/6	CD19 <sup>cre/WT</sup>	<b>CD19<sup>cre/WT</sup>Tbet<sup>fl/fl</sup></b>	C57Bl/6
BPL-PR8	IgG2c	45.1 $\pm$ 8.7	<b>14.0 <math>\pm</math> 5.4</b>	37.5 $\pm$ 2.6	12.7 $\pm$ 1.8	<b>2.6 <math>\pm</math> 0.5</b>	18.6 $\pm$ 6.5
	IgG1	8.2 $\pm$ 0.7	<b>4.8 <math>\pm</math> 0.8</b>	5.8 $\pm$ 0.7	11.7 $\pm$ 3.2	<b>8.9 <math>\pm</math> 2.4</b>	14.7 $\pm$ 3.2
HA	IgG2c	13.6 $\pm$ 2.4	<b>2.8 <math>\pm</math> 0.9</b>	8.5 $\pm$ 0.7	8.5 $\pm$ 1.3	<b>0.8 <math>\pm</math> 0.4</b>	8.3 $\pm$ 3.5
	IgG1	0.9 $\pm$ 0.3	<b>0.5 <math>\pm</math> 0.2</b>	0.9 $\pm$ 0.2	1.1 $\pm$ 0.3	<b>2.1 <math>\pm</math> 0.7</b>	2.7 $\pm$ 0.5
HA-stalk	IgG2c	1.5 $\pm$ 0.4	<b>0.2 <math>\pm</math> 0.1</b>	1.2 $\pm$ 0.3	0.6 $\pm$ 0.2	<b>0.1 <math>\pm</math> 0.0</b>	1.2 $\pm$ 0.5
	IgG1	0.1 $\pm$ 0.0	<b>0.0 <math>\pm</math> 0.0</b>	0.0 $\pm$ 0.0	0.1 $\pm$ 0.1	<b>0.1 <math>\pm</math> 0.0</b>	0.0 $\pm$ 0.0

\* note: values  $\pm$  SEM

### 3.4 DISCUSSION

Our study reveals multiple MBC subsets delineated by T-bet expression, whose distinct phenotypic and functional attributes are shared by mice and humans. T-bet expression status divides MBCs by anatomic localization and residency, as well as effector function and epitope specificity. Thus, T-bet<sup>-</sup> and T-bet<sup>lo</sup> MBCs originate in all secondary lymphoid tissues and freely recirculate, whereas T-bet<sup>hi</sup> MBCs reside in the spleen and are excluded from the lymphatics. Further, clonal and *in vivo* lineage tracing analyses show that while T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs likely arise from common pre-immune pools, they diverge after antigen encounter and thereafter remain as separate and stable pools. Finally, we show that the development of mouse IgG2c HA- and HA stalk-specific antibodies, as well as durable neutralizing titers, require T-bet expression in the B lineage. Taken together, these findings show that T-bet expression is a conserved feature of an MBC subset with differential circulatory properties, tissue-residency, and epitope specificity.

Pathogen-driven responses generate both isotype-switched and unswitched T-bet expressing B cells (301, 311, 312), but detailed analyses of the generation, fate and anatomic characteristics of T-bet<sup>+</sup> B cells have not been conducted. Our results formally demonstrate antigen-mediated and antigen-specific generation of T-bet<sup>+</sup> GC B cells during viral infection, followed by the establishment of somatically mutated, antigen-specific T-bet<sup>+</sup> and T-bet<sup>-</sup> MBC pools whose numbers are maintained indefinitely. Consistent with memory character, both T-bet<sup>+</sup> and T-bet<sup>-</sup> HA-specific B cells express the MBC markers CD73, CD80 and PD-L2 with kinetics similar to those in hapten-carrier responses (280). Despite these surface phenotypic similarities, our clonal analyses and

genetic fate-mapping experiments suggest it is unlikely that a progenitor-successor relationship exists, or that frequent interconversion occurs, between T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs. Thus, while T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs both result from antigen-driven naïve B cell activation, they most often arise independently and remain distinct, rather than representing different stages in a common differentiation pathway. In addition, the role for these HA-specific MBC subsets in recall responses remains an open question. Rechallenge studies in multiple mouse models have found that both CD80<sup>+</sup>PD-L2<sup>+</sup> MBCs and HA-specific MBCs preferentially differentiate into early antibody-secreting cells (ASCs) as opposed to re-entering germinal centers following antigen encounter (268). These studies suggest that T-bet<sup>+</sup> MBCs are primed for ASC differentiation, but what influence T-bet has on this fate decision compared to T-bet independent factors such as receptor affinity remains to be determined.

Our tissue distribution analyses indicate that memory B cells are anatomically compartmentalized: T-bet<sup>-</sup> and T-bet<sup>lo</sup> MBCs are found in all secondary lymphoid tissues, whereas T-bet<sup>hi</sup> MBCs are primarily in the spleen, blood, and bone marrow. Parabiosis experiments further confirmed that established influenza-specific T-bet<sup>hi</sup> MBCs neither exit the spleen to populate the lymphatic system, nor home to the spleen from blood or other anatomical locations. However, T-bet<sup>hi</sup> B cells were identified transiently in mediastinal LN and lungs early after infection, suggesting T-bet<sup>hi</sup> B cell generation can occur outside the spleen. We have previously reported the critical role of innate sensors, such as nucleic acid sensing Toll-like receptors (TLRs), and common gamma chain cytokines in regulating T-bet<sup>+</sup> B cell fate (274). Thus, the generative signals for T-bet expressing B cells are not spleen-specific per se, and the differential

anatomic distribution of established T-bet<sup>hi</sup> MBCs is not an immediate consequence of early antigen encounter specifically within the spleen.

Chemokine receptors and integrins regulate the anatomic distribution of immune cells and may contribute to T-bet<sup>hi</sup> B cells' characteristic localization properties. Studies examining human peripheral blood samples found that T-bet<sup>+</sup> B cells express the integrin CD11c, the chemokine receptor CXCR3, and low levels of CXCR4, CXCR5, and CCR7, chemokine receptors associated with homing to lymphoid organs (285, 287, 313). Thus, the specific combination of these and other surface receptors may impede lymphatic entry and help recruit T-bet<sup>hi</sup> B cells to the spleen. Via mechanisms that are unclear, T-bet<sup>hi</sup> B cells also appear to enter the blood following activation or recent tissue egress. Consistent with this idea, we observed early loss of HA-specific T-bet<sup>hi</sup> B cells in the mediastinal LN and lungs in infected mice, coupled with a temporary wave of HA-specific T-bet<sup>hi</sup> B cells in blood, and we previously described an increase in peripheral blood HA-specific T-bet<sup>hi</sup> B cells following influenza vaccination in humans (314). While they are normally absent from the lymphatics, recent evidence suggests consistent viremia and/or chronic immune activation may be able to override T-bet<sup>hi</sup> MBC compartmentalization: lymphoid tissue infections with pathogens such as HIV and *Toxoplasma gondii* are associated with a local enrichment of T-bet-expressing B cells in lymph nodes (315, 316).

The splenic residency associated with T-bet<sup>hi</sup> MBCs leads to some intriguing possibilities regarding the role of T-bet<sup>+</sup> MBCs in immune surveillance. While tissue residency may be critical to protect from local reinfection, this role seems unlikely for spleen-resident HA-specific T-bet<sup>hi</sup> MBCs, inasmuch as influenza is a respiratory infection and the virus is not known to replicate in the spleen (317). Instead, T-bet<sup>hi</sup>

splenic resident MBCs may be uniquely positioned to support broad immune surveillance and rapidly produce antibody for systemic dissemination upon reinfection. In support of this notion, T-bet<sup>+</sup> B cells express elevated quantities of BLIMP-1 (287) and, when isolated from the blood of SLE patients, quickly differentiate into plasma cells upon TLR7 stimulation without obligate division (318). It is tempting to speculate that circulating T-bet<sup>lo</sup> B cells are short-lived cells derived from a stationary, self-renewing T-bet<sup>hi</sup> population; indeed, the possibility of self-renewal and multipotency of T-bet<sup>+</sup> MBCs has been reported by others (311), and our findings confirm that T-bet<sup>+</sup> MBCs are a persistent population. Alternatively, T-bet<sup>lo</sup> B cells might be a stable and persistent population with separate maintenance requirements from the T-bet<sup>hi</sup> subset. These possibilities are not mutually exclusive and resolving their relative merits and contributions will require examination of the turnover rates and clonal composition of these MBC subsets.

Given the striking parallels between human and mouse T-bet<sup>+</sup> MBCs, we propose that T-bet expression is a conserved divisor for memory B cell subsets, and that the relative contributions of T-bet<sup>+</sup> vs T-bet<sup>-</sup> memory B cells to various aspects of humoral immunity merits detailed investigation. Importantly, in addition to their differences in anatomic localization, T-bet<sup>+</sup> and T-bet<sup>-</sup> B cells differ in the quality and specificity of antibodies they generate. Our studies with conditional knockout mice show that T-bet drives influenza-specific IgG2c production to HA and the HA stalk. It is tempting to speculate that these subsets could differentially contribute to immunodominance, cross-reactivity, or original antigenic sin, and may thus play distinct roles in immune responses to heterologous challenges. The loss of the IgG2c component of anti-influenza responses in cKOs suggests a link between T-bet<sup>+</sup> MBCs

and influenza-specific antibody production; however, the direct contribution of these established memory cells to antibody titer maintenance is unclear, as are the implications of tissue-restriction. T-bet<sup>+</sup> B cells are known to arise post influenza vaccination in humans (287, 314, 319); therefore, based on recent interest in developing HA-stalk-reactive vaccines for broad protection against influenza, we posit that focusing vaccine design efforts on driving T-bet expression in HA-specific B cells and maintaining this population long-term might lead to the development of more effective prophylactic agents and vaccination regimens for influenza.

### 3.5 MATERIALS AND METHODS

**Mice:** C57BL/6 and B6.SJL (10-12 weeks old, females, purchased from The Jackson Laboratory) T-bet<sup>flox/flox</sup> CD19<sup>Cre/WT</sup> (from the laboratory of E. John Wherry, University of Pennsylvania) and T-bet-ZsGreen (as previously reported (275)) were maintained and used in accordance with the University of Pennsylvania Institutional Animal Care and Use Committee guidelines.

**Infections:** Mice were infected by intranasal infection with 30 tissue culture infectious dose<sub>50</sub> (TCID<sub>50</sub>) of influenza strain A/Puerto Rico/8/1934 (PR8; American Type Culture Collection).

**Human samples:** All study participants provided written informed consent. Tissue samples were collected with IRB approval at the University of Pennsylvania (809316; 815056; 822686) and Case Western Reserve University (10-09-12). Human peripheral blood mononuclear cell samples were obtained from the University of Pennsylvania Human Immunology core. Human bone marrow mononuclear cell samples were obtained from the University of Pennsylvania Stem Cell and Xenograft core. Paired

blood and thoracic duct fluid samples were obtained from individuals with idiopathic or trauma-based chylopericardium or chylothorax requiring intervention at the Hospital of the University of Pennsylvania. Lymphoid tissue samples (mesenteric lymph node, iliac lymph node, tonsil, and spleen) were obtained at the Hospital of the University of Pennsylvania and Case Western Reserve University: mesenteric and iliac lymph nodes were obtained during abdominal surgery and kidney transplant surgery, respectively. Non-enlarged tonsils were obtained from sleep apnea patients. Spleens were removed and obtained due to trauma or surgical intervention. Additional spleen samples were obtained from the Human Pancreas Analysis Program (HPAP) at the University of Pennsylvania. Mononuclear cells were mechanically separated from solid tissues and enriched using a ficoll gradient.

**Parabiotic surgery:** Age-matched T-bet-ZsGreen reporters and B6.SJL adult female mice were conjoined as described previously (320). Briefly, a skin incision was made from the olecranon to the knee of each of the mice to be joined. The elbows and knees of the two paired mice were then tied together with surgical suture, followed by connecting of the skin with surgical sutures and staples. For pain control, mice were given buprenorphine (0.1 mg/kg every 6 hours for 36 hours) and meloxicam (5 mg/kg every 12 hours for 72 hours) and provided with sulfamexathole (400mg/L) and trimethoprim (800mg/L) antibiotics in their drinking water to prevent infection. Mice were monitored for signs of pain, infection, or damage to sutures. Blood was periodically drawn from the tail to check for anastomoses, which appeared complete by d14, therefore, mice were euthanized at day 17. The spleen was harvested from both partners for all pairs, and the lungs and mediastinal lymph nodes were also collected from some pairs.

**Flow cytometry:** Flow cytometry reagents were purchased from BioLegend (BL), BD Biosciences (BD), eBioscience (eBio), Southern Biotech (SB), or Invitrogen (Inv). The following antibodies were used for mouse studies: T-bet (4B10; BL), CD11c (N418; BL), IgM (R6-60.2; BD), CD38 (90; eBio), CD73 (TY/11.8; BL), CD80 (16-10A1; BD or BL), PD-L2 (TY25; BL), CD138 (281-2; BL), IgD (11–26c.2a; BL), B220 (RA3-6B2; BL or eBio), CD19 (1D3; BD or eBio), CD19 (6D5; BL) peanut agglutinin–FITC (Sigma), CD45.1 (A20; BL), CD45.2 (104; BL), CD183/CXCR3 (CXCR3-173; BL) and CD3 (17A2; BL). DUMP gate comprised CD8 (53-6.7; eBio), CD4 (H129.19; BL), F4/80 (BM8; eBio), Ly-6G/GR1 (RB6-8C5; eBio). The following antibodies were used for human studies: CD38 (HIT2; BL), CD85j (GHI/75; BD; HP-F1; eBio), T-bet (4B10; eBio and BL), IgM (MHM-88; BL), IgD (IA6-2; BD), CD10 (CB-CALLA; eBio), CD27 (O323; BL), CXCR3 (G025H7; BL), IgG (G18-145, BD), CD21 (Bu32, BL; B-ly4, BD), CD19 (HIB19, BL), CD3 (UCHT1, BL), CD14 (MØP9, BD), CD16 (3G8, BD), CD11c (3.9, eBio), Bcl-6 (K112-91, BD), Ki67 (56, BD), IgG1 (HP6069, Inv), IgG2 (HP6002, SB), IgG3 (HP6050, SB), and IgA (polyclonal, Inv). For detection of murine influenza-binding B cells, recombinant HA PR8 (276) was obtained from the laboratory of Dr. Barney Graham, National Institute of Allergy and Infectious Disease, biotinylated, and conjugated to streptavidin-fluorophores as previously described (276), or was directly conjugated using the R-phycoerythrin conjugation kit from Abcam (catalog ab102918) as per manufacturer's instruction. Human HA-specific B cell staining was performed using A/California/07/2009 and A/Wisconsin/67/2005 HA probes prepared as previously described (276, 289). Mouse samples were prepared for flow cytometry as follows: Mouse Fc fragment (Jackson ImmunoResearch; 015-000-008) was added to all staining cocktails at a final concentration of 1:200. Mouse spleens were homogenized, on ice, in staining buffer

(PBS + 0.5%BSA + 2mM EDTA) and passed through nylon mesh (50 $\mu$ M) to obtain single cell suspension. Red blood cells were lysed using ACK lysing buffer (Lonza, cat 10-548E) as per manufacturer's instructions. Cells were washed with PBS and stained as described previously(271, 274). Live/dead discrimination was done using Zombie Aqua fixable viability kit (BL). Prior to T-bet staining, cells were fixed and permeabilized using eBioscience™ Foxp3 / Transcription Factor Staining Buffer Set, at 4°C for 45min-1hr. Human samples were prepared for flow cytometry as previously described (288). Data were acquired on BD LSR II flow cytometer and FACS analyses were performed using FlowJo v9 and v10 (Becton Dickinson Co., Ashland, OR).

**Serum antibody titers:** Serum was harvested by spinning whole blood at 13000g for 10 minutes and stored at -20°C until use. Antibody titers were assessed using ELISA as previously described(271, 274) with the following modifications: 96-well medium-binding plates were coated with either 20HAU/well of BPL-inactivated PR8, 2  $\mu$ g/mL of PR8 HA, or 2  $\mu$ g/mL of H6/H1 chimeric constructs (expressed in baculovirus system as previously described (321)). HA-specific monoclonal antibodies (from Dr. Jonathan Yewdell, National Institute of Allergy and Infectious Diseases) were used as standards to determine concentration of IgG1 and IgG2a/c. Standards were used at a starting concentration of 100 ng/mL for IgG2a and 10 ng/mL for IgG1 and diluted 2-fold across.

**HAU (hemagglutination unit) and HAI assays:** Viral HAU titers were determined before every HAI assay. All dilutions were prepared in PBS. 50  $\mu$ L diluted virus, 50  $\mu$ L heat-inactivated sera and 12.5  $\mu$ L of 2% turkey erythrocytes were used per well for all assays, which were performed in round-bottom plates.

Starting with a 1:100 dilution of live virus, 2-fold dilutions were mixed with 2% turkey erythrocytes (Lampire biologicals) and incubated for 1 hour at room temperature. Agglutination dose (AD) was determined at the end of the incubation period, and confirmed by repeating the process with a 2-fold dilution series of virus, ranging from 4AD to 0.25 AD. This dose was subsequently used for the HAI assay.

Sera were heat-treated at 55°C for 30 minutes, diluted 2-fold in PBS (starting dilution 1:20), mixed with 4AD and 2% turkey erythrocytes, and incubated as for HAU assay. HAI titers are expressed as inverse of the highest dilution that inhibited agglutination.

**mRNA production:** The sequence of the Puerto Rico/8/1934 influenza virus hemagglutinin (pTEV-PR8 HA-A101) was codon-optimized, synthesized and cloned to the mRNA production plasmid. The mRNA was produced using T7 RNA polymerase (Megascript, Ambion) on linearized plasmids. The mRNA was transcribed to contain 101 nucleotide-long poly(A) tails. One-methylpseudouridine (m<sup>1</sup>Ψ)-5'-triphosphate (TriLink) instead of UTP was used to generate modified nucleoside-containing mRNA. Capping of the *in vitro* transcribed mRNAs was performed co-transcriptionally using the trinucleotide cap1 analog, CleanCap (TriLink). mRNA was purified by cellulose purification, as described (322). All mRNAs were analyzed by denaturing or native agarose gel electrophoresis and were stored frozen at -20°C.

**LNP formulation of the mRNA:** Cellulose-purified m<sup>1</sup>Ψ-containing RNAs were encapsulated in LNPs using a self-assembly process as previously described wherein an ethanolic lipid mixture of ionizable cationic lipid, phosphatidylcholine, cholesterol and polyethylene glycol-lipid was rapidly mixed with an aqueous solution containing mRNA at

acidic pH (323). The RNA-loaded particles were characterized and subsequently stored at -80°C at a concentration of 1 µg µl<sup>-1</sup>. The mean hydrodynamic diameter of these mRNA-LNP was ~80 nm with a polydispersity index of 0.02-0.06 and an encapsulation efficiency of ~95%.

**Mouse B cell receptor sequencing:** Genomic DNA was extracted from sorted cells using the Qiagen Genra DNA purification kit (Qiagen, Germantown, MD, Cat. No.158689). Primers used were adapted from Wang et al. (324) at the beginning of the FW1 region of VH and were modified to include adaptor sequences for the Illumina NexteraXT kit (sequences are provided below). Samples were amplified in duplicate (2 biological replicates per sample).

Primers:

VHmix (MH1) 5'- GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGSARGTNMAGCTGSAGSAGTC-3'

JH1,JH4 mix 5'- TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTANTGAGGAGACGGTGAC-3'

JH2 5' -TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAGGAGACTGTGAGAGTGG-3'.

The mouse IgH library was generated with one VH primer and a cocktail of JH1,2,4 primers. The VH and JH primer mixes were used at 0.6 µM in a reaction volume of 25 µL using a Multiplex PCR kit (Qiagen, Valencia, CA, Cat. No. 158388). Amplification conditions for the PCR were: primary denaturation at 95°C for 10 minutes, cycling at 95°C 45s, 60°C 45s, and 72°C for 90s for 35 cycles, and a final extension step at 72°C for 10 minutes.

Amplicons were purified using the Agencourt AMPure XP beads system (Beckman Coulter, Inc., Indianapolis, IN), and second-round PCRs were performed as described (325) to add Illumina NexteraXT adaptors to the IgH library. Final sequencing libraries were quantified by Qubit Fluorometric Quantitation (Thermo Fisher Scientific, Grand

Island, NY) and loaded onto an Illumina MiSeq instrument in the Human Immunology Core facility at the University of Pennsylvania and sequenced using 2x300 bp paired end kits (Illumina MiSeq Reagent Kit v3, 600 cycle, Illumina Inc., San Diego, Cat. No. MS-102-3003).

**Human B cell receptor sequencing:** Genomic DNA was extracted from sorted cells using the Qiagen Genra DNA purification kit. Sequences were generated from genomic DNA using primers that were situated at FR1 and JH (BIOMED2) for IgH V region sequencing. Samples were amplified in duplicate (2 biological replicates per sample). Second-round amplification to generate sequencing libraries used Illumina Nextera XT kit as previously described (325, 326). Sequencing were performed on an Illumina MiSeq instrument in the Human Immunology Core facility at the University of Pennsylvania using a 2x300 bp paired end kit.

**Sequencing data analysis:** Raw sequence data (FASTQ files) were processed through pRESTO version 0.5.10 (327). First, paired reads (R1 & R2) were aligned. Then sequences with an average Phred quality score of less than 30 (an error rate of 1 in 1000 bases) were removed. Of the remaining sequences, the 5' and 3' ends were trimmed until a window of 20 nucleotides had an average quality score of at least 30. Short reads of less than 100 bases were discarded after the trimming. Finally, nucleotides with a quality score of less than 30 were masked with an "N," and any sequence with more than 10 such N's were discarded.

**Code 1** shows a script performing these filtering steps.

```
PairSeq.py -1 *R1* *R2*
AssemblePairs.py align -1 *R1_pair-pass* -2 *R2_pair-pass* --coord illumina --rc tail
FilterSeq.py quality -s *assemble-pass* -q 30
```

```
FilterSeq.py trimqual -s *quality-pass* -q 30 --win 20
```

```
FilterSeq.py length -s *trimqual-pass* -n 100
```

```
FilterSeq.py maskqual -s *length-pass* -q 30
```

```
FilterSeq.py missing -s *maskqual-pass* -n 10
```

ImmuneDB (328) was used for gene identification and clonal inference of heavy chain sequencing data in both humans (using v0.26.0) and mice (using v0.28.0). Sequences were trimmed to IMGT position 20 in mice and 80 in humans to remove 5' primer sequences. Clones were assembled by grouping sequences with the same V-gene, J-gene, and 85% CDR3 amino-acid similarity as described in (326).

In the murine dataset, all mice contained a common CDR3 amino-acid string, CARGNRYWYFDVW (or a truncated variant of CARGNRYWYFDV or CARGNRYWYFD), possibly due to contamination, and were excluded from further analysis. Further excluded were two clones that had over 20% mutation in the V-region, due to incorrect V-gene assignment.

For all further analysis of both human and murine data, clones in each subject/subset combination were only included if they contained more than half the mean frequency of copies in that subject/subset.

**Quantification and statistical analyses:** All p values were determined using one of the following as mentioned in figure legends: unpaired non-parametric t-test or one-way ANOVA with Tukey post hoc test, paired t-test or repeated measures ANOVA with Tukey post hoc test, or Spearman correlation, using GraphPad Prism version 7 or version 8 (GraphPad Software, La Jolla, CA 92037 USA). \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001. Data are represented as mean  $\pm$  SEM. The number of mice and human subjects used in each

experiment, as well as the exact number of times an experiment was repeated, is mentioned in the figure legends

## CHAPTER 4: PERSPECTIVE

### 4.1 Investigating cellular differentiation in the immune system

Cellular differentiation is the principle problem in metazoan biology, and in my graduate work I have tried to address the question of how and to what extent lymphocytes differentiate in the establishment and functioning of the adaptive immune system. In order to function properly, the immune system requires the establishment of diverse cell-types. These cell-types result from the continual differentiation of hematopoietic stem cells throughout life, but the mechanisms that forge unique cell identity are still unclear. Despite the diversity, this range of cell-types alone is not sufficient to ensure functional immune responses to the wide variety of pathogenic insults encountered. As such, further differentiation and specialization of effector and memory immune cells is required to carry out the type of immune response best suited to the specifics of the pathogen challenge. However, extending the hematopoietic tree beyond the naïve lymphocyte stage to the mature subsets of an ongoing immune response cannot be done without firmly understanding their differentiative relationships and the stability of their identity. The problem of differentiation in the immune system is then two-fold: 1) to uncover the mechanisms leading to unique and stable cell identity and 2) to identify new subsets and define the developmental relationships between the diverse cell-types.

Questions about the control of cell identity have been long on the minds of embryologists and seeking answers to these questions has led to the understanding that chromatin acts as a biochemical filter for developmental cell-types by determining which

genes are transcribed into RNA (329). The application of genome-wide techniques such as ChIP-seq and ATAC-seq to differentiated cell-types in the hematopoietic system reveal that chromatin states vary profoundly between hematopoietic lineages and are more closely tied to cell identity than gene expression (119, 330). However, the chromatin states of mature cell-types are not pre-established in progenitors, begging the question of how they are established. Therefore, I first set out to address the question of how the chromatin state of a cell identity is shaped during development. Significant work had been done on erythrocyte (331, 332), macrophage (106), and B cell development (104, 333), but comparatively little work had been done on T cell development. As such, I focused my attention on T lymphopoiesis. As the proteins that modify histones are non-specific, they must be guided by sequence-specific transcription factors to cell-specific loci (334, 335). However, sequence-specific transcription factors are occluded by nucleosomes and cell-specific loci are not accessible in progenitors (119, 336-338). We hypothesized that a special class of transcription factors, uninhibited by nucleosomal DNA, acts during T cell development to create chromatin accessibility for the establishment of a T cell chromatin state and the activation of a T cell program. Therefore, we set out to examine the DNA sequences of developmentally regulated T cell chromatin to identify this mystery transcription factor.

#### 4.2 Mechanisms of shaping epigenetic cellular identity in T cell development

By investigating genome-wide measurements of accessibility at multiple T cell developmental stages we found that T cell-specific accessible chromatin defining T cell identity is more enriched for TCF motifs compared to the motifs of other developmentally important transcription factors. Further, T cell-specific chromatin was highly bound by TCF-1 and deletion of TCF-1 severely abrogated the accessibility of T cell chromatin in

the few surviving DP T cells. The chromatin accessibility mediated by TCF-1 regulated components of the T cell gene program and changes to accessibility were reflected by a compromised transcriptome including downregulation of important T cell genes such as *Bcl11b* and others. At the single cell level, TCF-1 enforced a coordinate accessibility among individual cells whereas sites bound by other transcription factors were more heterogeneous in the population. Finally, ectopic expression of TCF-1 in fibroblasts converted inaccessible T cell-specific heterochromatin into active and accessible chromatin and induced expression of T cell-specific genes.

Previously, TCF-1 was shown to induce the gene program of T cell specification in bone marrow progenitors even in the absence of Notch signaling (121). The question then arises of whether TCF-1 initiates a T cell gene program that has been primed in the chromatin of bone marrow progenitors prior to settling in the thymus. Our results indicate the opposite. Inasmuch as we could measure it, chromatin accessibility in bone marrow progenitors (as measured in CLPs in this study) was devoid of most T cell accessible chromatin. Thus, it is likely that the T cell specification events mediated by TCF-1 is the result of sweeping changes to chromatin that make the T cell gene program possible. Consistent with this notion, we observed a profound increase in accessibility between the CLP and ETP stage where TCF-1 is first induced to high levels of expression from Notch1 signaling (136).

Together, our results indicate that TCF-1 acts as a lineage-determining transcription factor to create accessible T cell chromatin and shape the epigenetic identity of T cells. More broadly, our results suggest that the adoption of cell identity during hematopoiesis is more than just the adoption of a specific gene program but is established through extensive chromatin changes outnumbering even gene expression changes. Similar conclusions have been drawn in other developmental contexts in

hematopoiesis and embryogenesis (330, 338, 339). Moreover, our results indicate that this process of development is carried out by lineage-specific transcription factors that make lineage-specific chromatin accessible to organize the more ubiquitous transcription factors to function at lineage-specific gene loci.

Transcription factors that mediate the selection and accessibility of chromatin have been termed 'pioneer transcription factors' (PTFs) (2) or 'lineage-determining transcription factors' (LDTFs) (3). Both labels describe transcription factors that act before other transcription factors to establish lineage-specific accessible chromatin. The difference between them is that PTFs have been shown to have an intrinsic biochemical affinity to nucleosomal DNA (340), usually mediated through a special protein domain (341), whereas LDTFs have not. The activity of TCF-1 during T cell development is consistent with these descriptions, especially of the LDTF class. Although we have not yet shown the binding of purified TCF-1 to nucleosomal DNA arrays *in vitro*, our results in the fibroblast model demonstrated that almost half of TCF-1 binding events occurred at a known nucleosome position, near the dyad. Thus, it is likely that TCF-1 will be classified as a PTF in the future. In fact, a recent publication by Howard Xue's group demonstrated a vestigial, but functional histone deacetylase (HDAC) domain in TCF-1 (125), providing evidence of a biochemical link to the histone proteins comprising the nucleosome. Whether this HDAC domain is required for TCF-1 to create *de novo* chromatin accessibility remains to be determined.

Surprisingly, we found that  $\beta$ -catenin and Wnt signaling was not required for TCF-1 to create accessible chromatin in the fibroblast model because we used the p33 isoform that does not bind  $\beta$ -catenin (97). The Wnt/TCF pathway is an evolutionarily ancient signaling pathway used in multiple developmental contexts (91). The role of  $\beta$ -catenin in T cell development is a complex one with conflicting reports. One study found

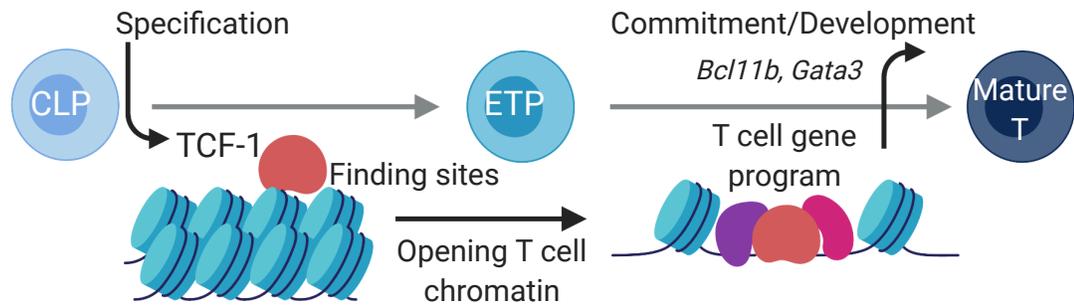
that  $\beta$ -catenin was essential for T cell development (342) whereas others show that T cell development is largely intact in the absence of  $\beta$ -catenin (95, 96). Moreover, the p33 isoform has been assumed to be transcriptionally repressive because  $\beta$ -catenin is not able to displace the recruitment of the Groucho/TLE family of transcriptional repressors (93). Again, our results conflicted with this classification. We found that expression of the p33 isoform in fibroblasts was able to create *de novo* chromatin accessibility and activate T cell genes suggesting that  $\beta$ -catenin is not required for either function. Another study by Howard Xue's group suggests the role of  $\beta$ -catenin is limited to the regulation of cell survival signals during positive selection but is dispensable for T cell differentiation (343). Thus, it is possible that the effect of TCF-1 on shaping T cell chromatin and activating T cell gene expression is largely independent from  $\beta$ -catenin except to regulate a small suite of survival genes at a specific stage of development.

Because we utilized an unbiased, genome-wide approach, our analysis turned up other unexpected findings. We found the chromatin that becomes accessible during T cell development is dynamically regulated and becomes accessible in waves, occurring at the ETP stage, the DN3-DN4 stage, and the mature single positive stage. The timing corresponds to important developmental events occurring at these stages including T cell specification,  $\beta$ -selection, and MHC restriction, respectively (45, 50, 121). Importantly, TCF-1 binding events and motifs were enriched in all three waves of accessible chromatin. Although this observation seems incongruous with the role of TCF-1 in creating accessible chromatin, PTFs and LDTFs are known to require additional regulatory events to specify their activity as even PTFs bind only a fraction of their total binding sites (344). Such events could include the cooperation of other transcription factors functioning in a combinatorial manner with TCF-1 (3). PTFs are repelled by highly condensed heterochromatin (144), but we found that TCF-1 could

overcome the repressive heterochromatin and establish accessibility. However, other epigenetic factors including DNA methylation and chromatin folding could be responsible in limiting the ability of TCF-1 to create accessible chromatin. In support of this idea, a paper published in the same issue of *Immunity* by Keji Zhao's group showed that the 3-dimensional organization of the chromatin is dynamically regulated at key stages of T cell development suggesting that the coordinated accessibility of these waves could be the result of positive changes to the higher-order chromatin architecture (345).

#### 4.3 TCF-1 establishes T cell chromatin during T cell development

In summary, my studies on the transcription factor TCF-1 in T cell development has strengthened our understanding of how cell identity is managed at the chromatin level during development. The role of TCF-1 in activating the gene program of T cell specification has been known for several years, but our work has expanded the role of TCF-1 beyond a simple model of gene activation. Our expanded model argues that TCF-1 opens unprimed or repressed chromatin in progenitor cells to establish an accessible and active chromatin state for the execution of a T cell gene program, and a graphical depiction of our model is shown in **Figure 13**. Although our work sheds light on the factors involved in creating a cell-specific chromatin state, it also raises additional questions.



**Figure 13: Current working model: Forging T cell epigenetic identity using TCF-1.**

In the thymus, Notch signals induce TCF-1 expression in thymic settling progenitors such as CLP or LMPP. TCF-1 carries out T lineage specification by finding T cell-specific CRMs buried in closed chromatin by binding its motif on nucleosome wrapped DNA. Through an uncharacterized mechanism, TCF-1 establishes accessible chromatin at T cell-specific CRMs, allowing the binding of additional transcription factors that are normally repelled by the chromatin. The addition of new CRMs transforms the gene program to induce the expression of *Bcl11b* for T cell commitment and *Gata3* for further T cell development. TCF-1 guides individual cells along the T cell trajectory by reducing heterogeneity of accessibility at T cell specific CRMs and the germline removal of TCF-1 severely abrogates chromatin accessibility, reduces the T cell gene program, and negatively impacts mature T cell function. TCF-1 is sufficient to create *de novo* accessibility even in unrelated cell types such as fibroblasts suggesting that modulation of TCF-1 levels could be used to tune the T cell chromatin state.

#### 4.4 Perspective on epigenetic engineering to control differentiation in immune responses

The T cell chromatin state is dependent on TCF-1, and the requirement of TCF-1 expression for robust memory CD8<sup>+</sup> T cell responses suggests cell function beyond the naïve stage is a function of the chromatin state established during development (122). Recent work demonstrates that naïve, memory, effector, and exhausted CD8<sup>+</sup> T cells have unique chromatin identities, but the epigenetic profile of naïve and memory CD8<sup>+</sup> T cells share many features including enrichment for TCF motifs (113-115). Conversely, the accessible chromatin of effector and exhausted CD8<sup>+</sup> T cells, which do not express TCF-1, lack TCF motifs suggesting that restoring TCF-1 expression could revert the chromatin state of these cells to a naïve or memory state and restore immune function and durability. The epigenetic reprogramming of immune cells to modulate aspects of their function and survival has major therapeutic potential.

However, the transcription factors and regulatory events responsible for creating different chromatin states between effector and memory subsets are not fully elucidated. Unique chromatin states are established in a stepwise fashion through the hierarchical activity of transcription factors to induce a distinct and stable gene program. Thus, identifying the transcription factors that define cell identity requires an understanding of the developmental sequence and the relationship between subsets. Generally, there is much less acceptance over the developmental relationships of effector and memory cell subsets beyond the naïve stage for both T and B cells. Part of the reason for this is the many different types of effector and memory cells that can develop from the different types of pathogens that can be encountered. Therefore, defining the developmental trajectory of these subsets is a major open-ended question. In the second half of my

graduate work I set to work on defining the differentiative relationships between subsets of memory B cells resolved by the transcription factor T-bet.

#### 4.4 Cellular differentiation in the memory B cell response

We used a *Tbx21* transcriptional reporter to track T-bet expression in influenza hemagglutinin (HA)-specific B cells at various time-points after PR8 infection. We determined that T-bet<sup>-</sup> and T-bet<sup>+</sup> HA-specific B cells arose early, acquired a germinal center phenotype, and persisted long-term where both subsets acquired the memory B cell markers CD73, PD-L2, and CD80. We also noted that the T-bet<sup>+</sup> population was composed of T-bet<sup>hi</sup> and T-bet<sup>lo</sup> subsets with T-bet<sup>hi</sup> memory B cells (MBCs) having a different anatomic distribution and recirculation properties in mice and humans compared to T-bet<sup>lo</sup> and T-bet<sup>-</sup> MBCs. Thus, we found that T-bet<sup>hi</sup> MBCs were restricted to the spleen while T-bet<sup>lo</sup> and T-bet<sup>-</sup> MBCs circulated throughout all secondary lymphoid organs. Through parabiosis, we determined that T-bet<sup>hi</sup> MBCs were spleen-resident whereas T-bet<sup>lo</sup> and T-bet<sup>-</sup> MBCs circulated freely. To assess the differentiative relationship between T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs we performed a clonal analysis by IgH sequencing and found that most clones were not shared between the subsets. In the instances where we found clonal sharing, lineage tree analysis revealed that clonal daughters eventually adopted either T-bet<sup>+</sup> or a T-bet<sup>-</sup> fate. Further fate-mapping using an inducible *Tbx21* fate-mapping mouse confirmed the stability of the T-bet<sup>+</sup> fate. Finally, B cell expression of T-bet was important for the humoral response to influenza and B lineage deletion of T-bet caused a reduction in influenza-specific antibody titers, especially to the HA stalk.

Previously, Naradikian et al. (274) determined that the signals necessary to engender a T-bet<sup>+</sup> B cell phenotype include BCR signaling, B cell intrinsic nucleic-acid

detection through TLR engagement, and T<sub>H</sub>1 inflammatory cytokine signals. Moreover, a T<sub>H</sub>2-mediated response does not produce T-bet<sup>+</sup> B cells. Influenza is an RNA virus that produces a T<sub>H</sub>1-mediated response (346), and our results confirm that influenza infection meets the conditions necessary to generate T-bet<sup>+</sup> B cells. In the presence of IFN $\gamma$  and TLR7/9 agonists, T-bet expression occurs within 48 hours of B cell activation *in vitro* suggesting that T-bet expression is an effector phenotype linked to the events of B cell activation. Analogously, B cells *in vivo* adopt an activated T-bet<sup>+</sup> effector phenotype likely driven by BCR cross-linking with TLR7/9 signals and induced by IFN $\gamma$  or IL-21 produced by T<sub>FH</sub> during a T<sub>H</sub>1-type response (274). The events of signaling, and not an intrinsic activation program, drive the differentiation of T-bet<sup>+</sup> B cells, and the adoption of a T-bet<sup>+</sup> effector phenotype does not appear to be confined to a certain pre-immune or antigen-experienced subset of B cells. Thus, the differentiation of T-bet<sup>+</sup> effector B cells result from the integration of innate and adaptive activation signals with inflammatory cytokines in unrestricted B cell populations.

By tracking antigen-specific B cells, our results extend previous findings by demonstrating that T-bet<sup>+</sup> B cells that arise during infection persist as a stable memory pool. Therefore, the description of T-bet<sup>+</sup> B cells is not confined to recently activated effector B cells but also encompasses *bona fide* memory cells in both mice and humans. Although definitive MBC markers in mice are lacking, expression of CD73, CD80, and PD-L2 have been demonstrated on MBCs generated with an alum-adjuvanted hapten-carrier response (267, 268). In mice, we found that nearly all HA-specific T-bet<sup>+</sup> and T-bet<sup>-</sup> B cells express these MBC markers by day 100 post infection and are maintained at a steady state, consistent with memory cell identity. Our results indicate MBCs are not a monolithic population but are composed of discrete subsets that reflect fate choices shaped by the events of early infection and differentiate according to pathogen-driven

cues. However, our results also indicate that T-bet<sup>+</sup> and T-bet<sup>-</sup> memory pools are distinct and do not interchange, and T-bet<sup>+</sup> MBCs do not arise from T-bet<sup>-</sup> MBCs or vice versa. Thus, the effector functions selected during the humoral response are perpetuated by the differentiation of distinct pools of long-lived memory cells.

B cells expressing T-bet have been described as “Age-associated B cells” (ABCs) and are driven by similar innate and adaptive signals in both normal and autoimmune humoral responses (347). Moreover, ABCs require cognate help from CD40/CD40L interactions to form, suggesting they are the product of T dependent B cell responses. Confirming this notion, we observed that T-bet<sup>+</sup> B cells had a GC phenotype, formed long-lived memory cells, and were hypermutated—processes that require T cell help. Using the immune response to influenza, our results suggest that ABCs, regardless of their setting of generation, are a durable MBC population. Moreover, we found that T-bet<sup>hi</sup> MBCs were spleen-resident in mice and increased in proportion with age in human spleens suggesting that studies on T-bet<sup>hi</sup> B cells isolated from the blood alone may not be an accurate representation of the ABC pool in either settings of normal immune responses or autoimmunity. Indeed, some have noted that T-bet<sup>hi</sup> ABCs isolated from the blood reflect an activated, effector phenotype rather than a memory phenotype and, when isolated from the blood, do not demonstrate a proportional increase with age in humans (348). Thus, studies on this critical population underlying normal and autoimmune humoral responses need to be conducted in the tissues such as the spleen rather than the circulation where possible.

The function of the T-bet<sup>+</sup> MBC subset is unknown as of yet. The role of memory B cells in humoral immunity may appear redundant alongside established plasma cells and elevated antibody titers. However, memory B cells act as a second line of defense through rapid reactivation and differentiation into plasma cells (349-351). Thus, the

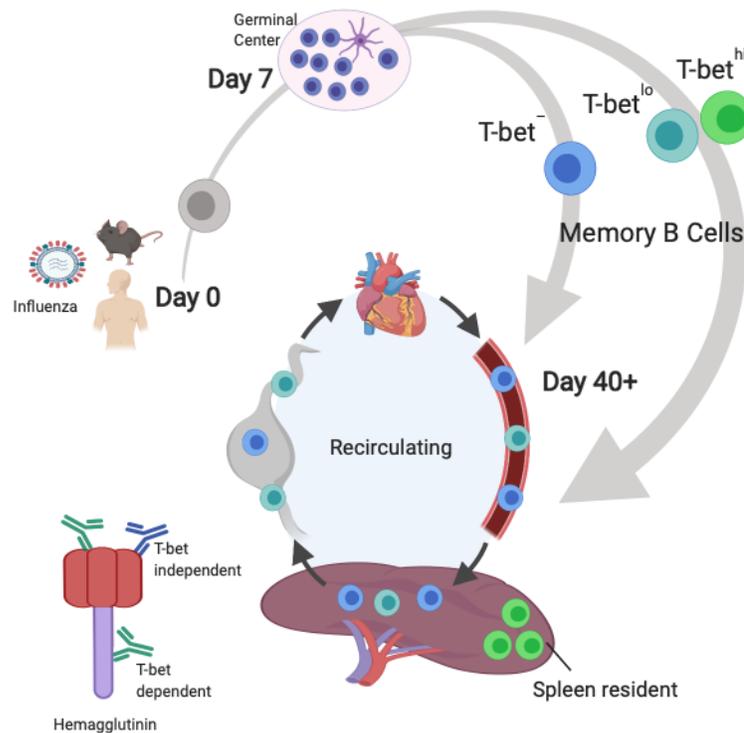
secondary humoral response is typically faster, of greater magnitude, and consists of isotype-switched antibodies of higher affinity. The characteristics of the secondary response are thought to be the consequence of the activity of MBCs, although some aspects have been challenged recently (352). Nonetheless, MBCs make up a large proportion of the peripheral B cell pool in adults. Moreover, the role of MBCs in recall responses may contribute to the phenomenon of Original Antigenic Sin (OAS) or the re-emergence of autoimmune antibodies after B cell depletion (353-356). Thus, the role of T-bet<sup>+</sup> MBCs in these aspects of the recall response needs to be addressed.

We show that MBC subsets are functionally divided by their anatomic distribution and recirculation properties with T-bet<sup>-</sup> and T-bet<sup>lo</sup> MBCs recirculating freely and T-bet<sup>hi</sup> MBCs residing in the spleen. The functional significance of T-bet<sup>hi</sup> tissue residency is unknown. Tissue-resident memory T cells, which survey intracellular environments, are thought to mediate rapid effector responses to local antigen reencounter (357). Presumably, soluble antigen recirculates more rapidly than the migration of cells and antibodies freely diffuse through the blood. Therefore, the significance of resident memory B cells is unclear, but is an emerging and active area of research (291). However, the residency of T-bet<sup>hi</sup> MBCs in the spleen may be related to their function. The spleen is a unique secondary lymphoid organ that surveys both the circulatory and lymphatic systems. Thus, we speculate the splenic-residency of T-bet<sup>hi</sup> MBCs may function to support rapid recall and differentiation to antibody secreting cells, especially when antigens are mobilized in circulating lymph and/or blood in a manner similar to the function of marginal zone B cells. Indeed, a study by Shlomchik's group has demonstrated that T-bet<sup>+</sup> MBCs reside in the marginal sinus and appear to displace marginal zone B cells (358). Degradation of the marginal sinus structure with age has been reported previously, and we speculate the accumulation of ABCs may be at least

partially responsible for this decline (359, 360). Experiments addressing the tissue localization and the plasma cell potential of T-bet<sup>hi</sup> MBCs can help answer these questions.

#### 4.5 T-bet<sup>+</sup> memory B cells arise independently from T-bet<sup>-</sup> memory B cells

A clear function for B cell expression of T-bet in the humoral response is to facilitate isotype-switching to IgG2c (294-301), and our results were consistent with this function. Importantly, we did not see an increase in IgG1 titers, a T-bet independent isotype, in the absence of B lineage T-bet expression. Moreover, nearly all of the antibodies specific to the HA-stalk were IgG2c. In conjunction with the clonal and fate-mapping analysis, these results suggest that T-bet<sup>+</sup> B cells do not develop from the T-bet<sup>-</sup> pool. Instead, we propose that T-bet<sup>+</sup> and T-bet<sup>-</sup> B cells mostly arise independently, and our model is described in **Figure 14**. In a linear model of differentiation, cells adopt an alternative identity from the accumulation of, or increased exposure to, a differentiative factor. In the absence of the differentiative factor, cells do not differentiate but remain, nonetheless. Our results indicate that the development of T-bet<sup>+</sup> MBCs does not follow a linear differentiation model but instead follows an 'all-or-none' model. Thus, when T-bet is deleted in the B lineage, most HA-specific titers are absent and are not produced by a T-bet independent subset of MBCs.



**Figure 14: Current working model: The differentiation of T-bet<sup>+</sup> memory B cell subsets**

T-bet<sup>+</sup> B cells differentiate early as a distinct pool from T-bet<sup>-</sup> B cells after influenza infection. T-bet<sup>+</sup> B cells possess unique specificities including the stalk of influenza hemagglutinin, and the expression of T-bet facilitates isotype switching to IgG2a/c. Both T-bet<sup>+</sup> and T-bet<sup>-</sup> pools enter germinal centers, undergo somatic hypermutation, and persist indefinitely as memory B cells. However, they possess different recirculation properties and tissue distributions in mice and humans. T-bet<sup>lo</sup> and T-bet<sup>-</sup> memory B cells recirculate through the blood and lymphatics whereas T-bet<sup>hi</sup> memory B cells are spleen resident. Without B cell expression of T-bet, neutralizing titers to influenza fail to be maintained and the bulk of the influenza-specific antibody response is absent.

#### 4.6 Perspective on the future research of T-bet<sup>+</sup> B cells

An alternative explanation to the reduced HA-specific titers in the absence of B lineage T-bet expression is a potential role of T-bet<sup>+</sup> MBCs in generating or maintaining plasma cells (318). In the humoral response, pathogen-specific titers can remain elevated and provide protection either through the establishment of LLPCs or the continual differentiation of SLPCs from an effector or memory pool (256). No T-bet<sup>+</sup> plasma cell subset has been identified and it is well accepted that plasma cells do not express T-bet. Thus, T-bet dependent antibody responses depend on the differentiation of T-bet<sup>+</sup> B cells into plasma cells while losing T-bet expression in the process. We identified T-bet<sup>lo</sup> and T-bet<sup>hi</sup> subsets with different circulatory properties and hypothesize that T-bet<sup>hi</sup> B cells are a tissue-resident stem-like pool that populate the circulating T-bet<sup>lo</sup> pool on the way to plasma cell differentiation. A recent study by the Winslow group demonstrates that T-bet<sup>+</sup> memory B cells are multipotent, supporting this hypothesis (311). Indeed, preliminary results in the lab indicate established T-bet<sup>hi</sup> MBCs turnover and differentiate into T-bet<sup>lo</sup> MBCs and eventually plasma cells when adoptively transferred. A progenitor-successor relationship between T-bet<sup>hi</sup> MBCs, T-bet<sup>lo</sup> MBCs, and plasma cells suggests that supporting or reducing antibody production from T-bet<sup>+</sup> B cell populations requires targeting the spleen-resident T-bet<sup>hi</sup> MBC pool.

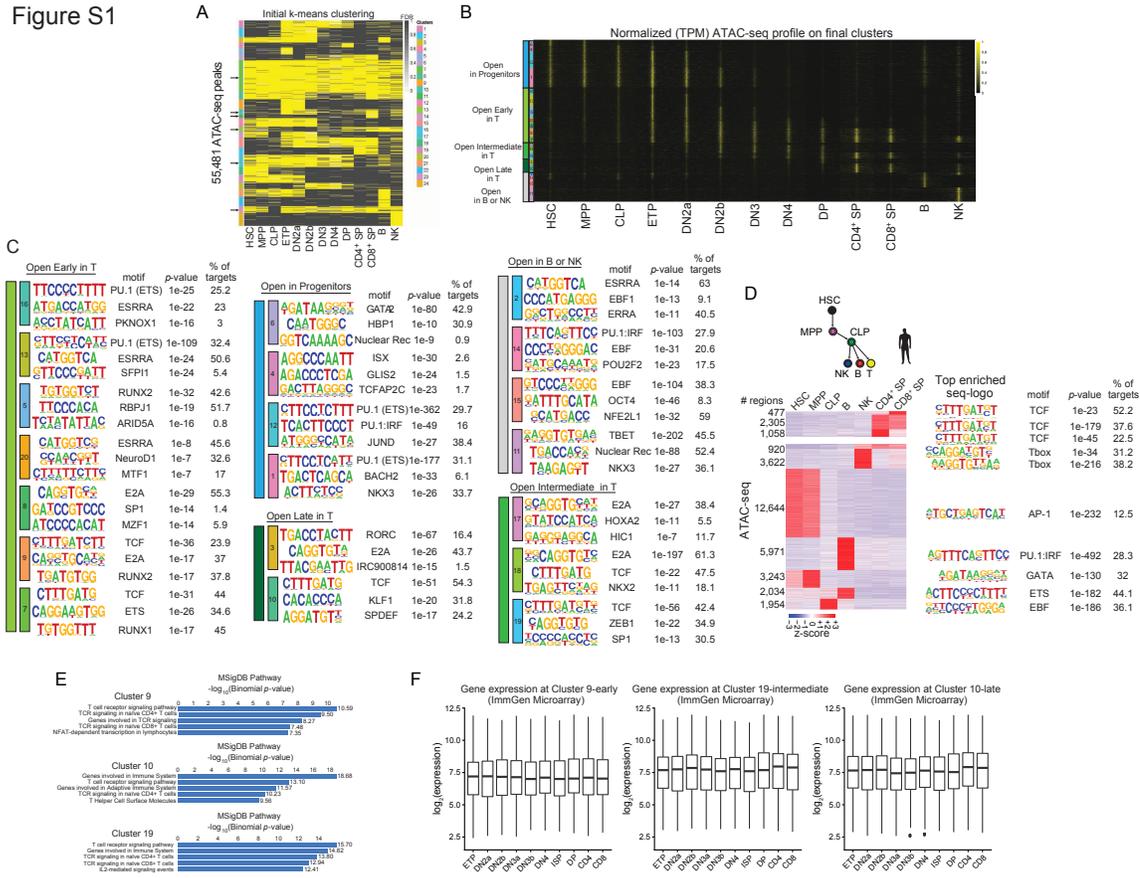
Finally, what is the nuclear role of T-bet? As a transcription factor, T-bet binds to the DNA in a sequence-specific manner to regulate the expression of genes (361). However, transcription factor activity is filtered by the chromatin state and T-bet does not appear to be sufficient or necessary to alter the chromatin state (111). A recent study by the Lund group (362) has demonstrated that the chromatin state of T-bet<sup>+</sup> B cells is considerably different from the chromatin state of T-bet<sup>-</sup> B cells in an influenza response

suggesting that unique modes of gene regulation operate in T-bet<sup>+</sup> cells. Identifying the transcription factors that act in the T-bet<sup>+</sup> chromatin landscape and the gene loci regulated by unique T-bet<sup>+</sup> B cell CRMs will provide insight into their regulatory program as well their function.

Naturally, future research will likely focus on the mechanisms that establish the unique T-bet<sup>+</sup> chromatin state. However, when setting out to understand the regulatory events leading to the divergent chromatin states of T-bet<sup>+</sup> and T-bet<sup>-</sup> B cells, it must be noted that our results demonstrate that T-bet<sup>+</sup> MBCs do not develop from T-bet<sup>-</sup> MBCs. Epigenetic cell identity is built progressively by the incremental activity of transcription factors on the chromatin. Therefore, an in-depth analysis of the establishment of T-bet<sup>+</sup> cells should not focus primarily on the chromatin differences between T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs, but on the events leading to the adoption of a T-bet<sup>+</sup> fate. The transcriptional events downstream of TLR signaling and/or inflammatory cytokine signaling are a good place to start because STATs can shape the chromatin landscape (111). As suggested by my investigation on immune cell identity at the chromatin level and the differentiative relationships between immune cell subsets as outlined in this thesis, the establishment of unique chromatin identity is best decoded when the fate choices and lineage relationships between subsets are properly understood.

# APPENDIX: Supplementary Figures

Figure S1



**Figure S1. TCF-1 binding occurs at three waves of chromatin remodeling during T cell development, Related to Figure 2**

(A) Accessible regulatory elements were identified in mouse across multiple cell types including progenitors (HSC, MPP and CLP), T cell development (ETP, DN2a, DN2b, DN3, DN4, DP, CD4<sup>+</sup> and CD8<sup>+</sup> SP) and non-T cell lineages (B and NK) using bulk ATAC-seq data. Peaks were called with macs2 and their reproducibility was assessed using IDR. Accessible regions were filtered based on annotated gene promoters (see STAR methods) into distal and proximal regulatory elements. Proximal elements were filtered out and the remaining 55,481 regions were clustered into 24 groups with k-means using the FDR value in each sample as a proxy of ATAC-seq enrichment. Black arrows indicate clusters that have been filtered out from the final set of regions shown in **Figure 2A**.

(B) ATAC-seq profiles (+/- 2kb window) around summits of all 20 k-means clusters identified in mouse from **Figure 2A**.

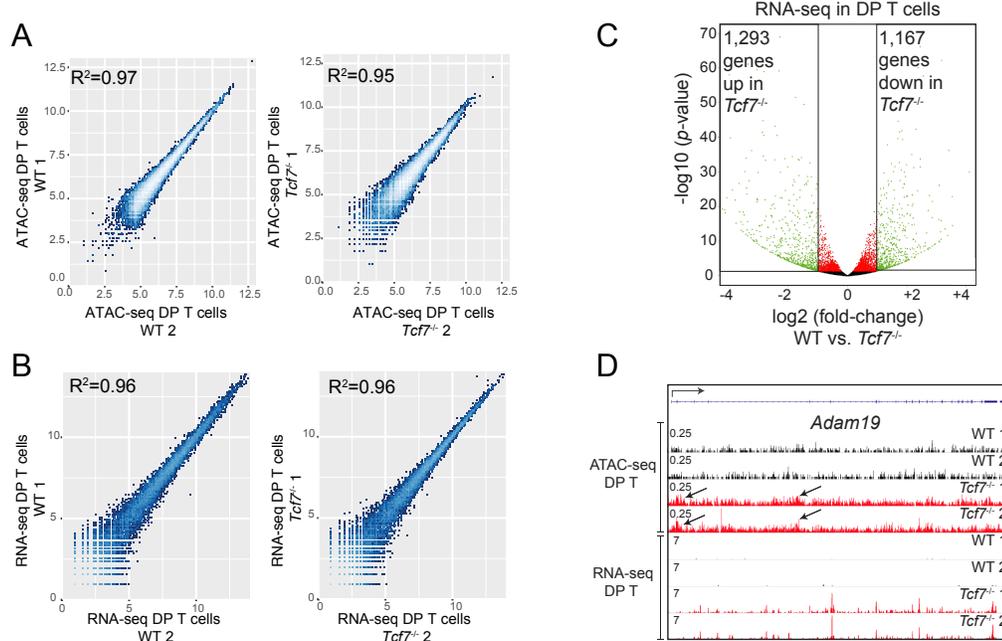
(C) Top three enriched motifs from *de novo* motif analysis using HOMER on all clusters in Figure 2A.

(D) Clustering of accessible regulatory elements based on public ATAC-seq data in human HSC, MPP, CLP, B, NK, CD4<sup>+</sup> and CD8<sup>+</sup> T cells. Peaks were called with macs2. Due to the lack of replicates, IDR assessment of peak reproducibility was not feasible. Normalized tag counts were used as a proxy of ATAC-seq enrichment to identify clusters (see STAR methods). HOMER facilitated the *de novo* motif discovery in different clusters.

(E) Pathway analysis for genes proximal to clusters 9, 19, and 10 (see **Figure 2A**).

(F) Distribution of the expression levels for genes proximal to clusters 9, 19, and 10 using microarray data from ImmGen.

Figure S2



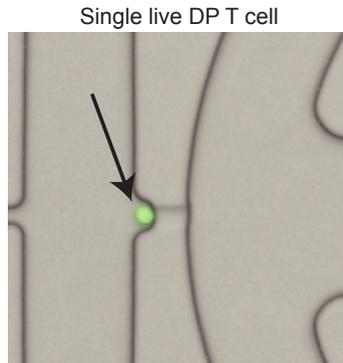
**Figure S2. TCF-1-deficient DP T cells cannot establish the open chromatin landscape and transcriptional output of normal DP T cells, Related to Figure 3**  
**(A-B)** Pairwise Pearson correlation plots for replicates in **(A)** ATAC-seq and **(B)** RNA-seq experiments in WT and *Tcf7*<sup>-/-</sup> DP T cells.

**(C)** Volcano plot demonstrating differentially regulated gene expression in WT and *Tcf7*<sup>-/-</sup> DP T cells. DESeq2 was used to identify 1,167 down- and 1,293 up-regulated (fold-change > 1.5 and *p*-value < 1e-3) in *Tcf7*<sup>-/-</sup> DP T cells.

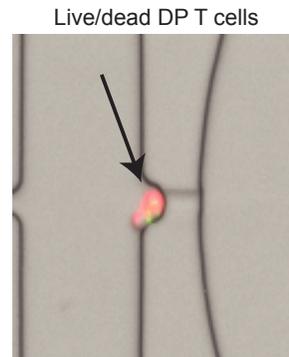
**(D)** Example of ATAC- and RNA-seq profiles from WT and *Tcf7*<sup>-/-</sup> DP T cells in *Adam19* locus.

Figure S3

A

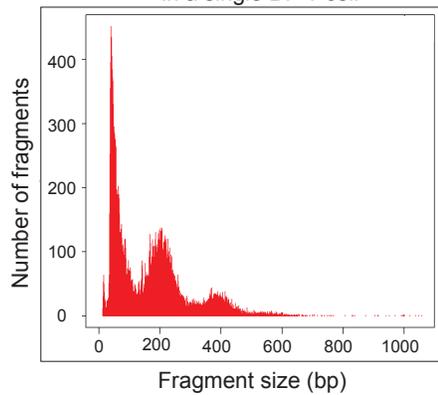


B



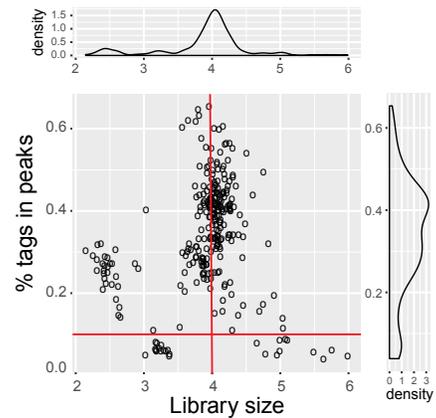
C

Fragment size distribution of ATAC-seq library in a single DP T cell



D

Single cell ATAC-seq in DP T cells in 3 captures



**Figure S3. TCF-1 binding coordinates open chromatin between single DP T cells, Related to Figure 4**

(A-B) Examples of live/dead stained cells captured with Fluidigm C1 platform.

(C) Fragment size periodicity from scATAC-seq data derived from single DP T cells.

(D) Single DP T cells plotted based on two quality assessment metrics derived from scATAC-seq samples. Single cell libraries containing less than 10,000 fragments or with less than 15% of their fragments falling in open chromatin did not reach the minimum quality criteria. Together, we performed three independent single-cell captures and 110 T cells at the DP stage passed various quality control thresholds.



**Figure S4. TCF-1 can bind to nucleosomes and create chromatin accessibility in fibroblasts, Related to Figure 5**

(A) Ectopic expression of TCF-1 in NIH3T3 using retroviral transduction and subsequent verification of expression by intracellular staining and flow cytometry.

(B) Log<sub>10</sub> TCF-1 peak score as defined by the IDR package in replicates of ChIP-seq in TCF-1 RV NIH3T3. Red signifies peaks that fail to pass the IDR cutoff while black reflects the reproducible subset of TCF-1 peaks in multiple thresholds.

(C) K-means clustering (k=3) of TCF-1 summits based on the normalized MNase-seq enrichment in three non-overlapping 200bp windows centered around summits (see **Figure 5**).

(D) Distribution of distance between TCF motifs unbound by TCF-1 (red) and uniquely bound by TCF-1 ChIP-seq (green) in TCF-1 RV NIH3T3 cells (upper panel) or DP T cells (lower panel). Nucleosome summits were identified using Danpos2 on public MNase-seq data in fibroblasts (see STAR methods). TCF motif occurrences were identified genome-wide using FIMO and the TCF-1 PWM derived from JASPAR database. Statistical significance of the difference between the TCF-1 bound and unbound motifs to the nucleosome summits was calculated using a bootstrap approach (see STAR methods).

(E) Pairwise Pearson correlation plots between ATAC-seq replicates in Empty and TCF-1 RV NIH3T3 cells.

(F) Nucleosome enrichment profile around TCF-1 ChIP-seq binding summits assessed using the NucleoATAC algorithm in Empty and TCF-1 RV NIH3T3 cells (see STAR methods).

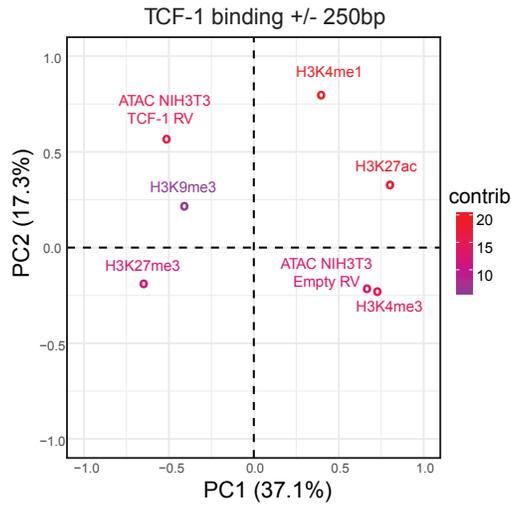
(G) Normalized TCF-1 ChIP-seq enrichment profile around merged TCF-1 binding events in TCF-1 RV NIH3T3 and DP T cells with a Pearson correlation coefficient of 0.3.

(H) Regions that gained accessibility in TCF-1 RV NIH3T3 were overlapped with ATAC-seq peaks specific to HSC, MPP, CLP, B, NK, naïve CD4+, naïve CD8+, effector CD8+ and memory CD8+ cells. “Union T cells” was generated by merging the open chromatin regions of all T cell datasets.

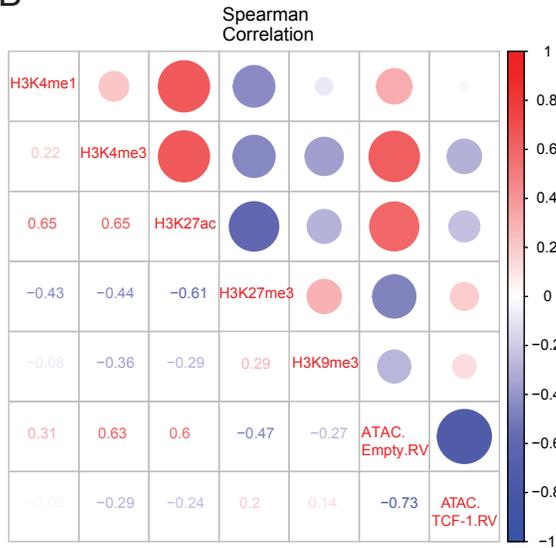
(I) Enrichment analysis of TCF-1 ChIP-seq binding events in: (1) DP T cells, (2) genomic loci that lose accessibility in *Tcf7*<sup>-/-</sup> DP T cells and (3) regions that gained accessibility in TCF-1 RV NIH3T3 cells at the clusters shown in **Figure 2A**.

Figure S5

A

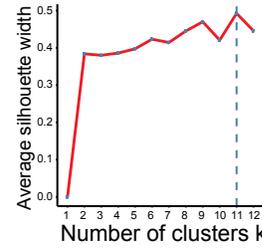


B

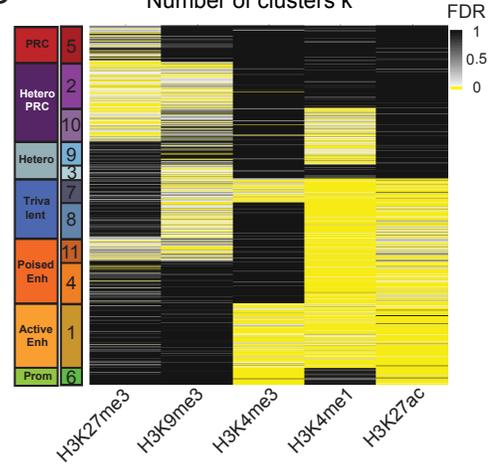


C

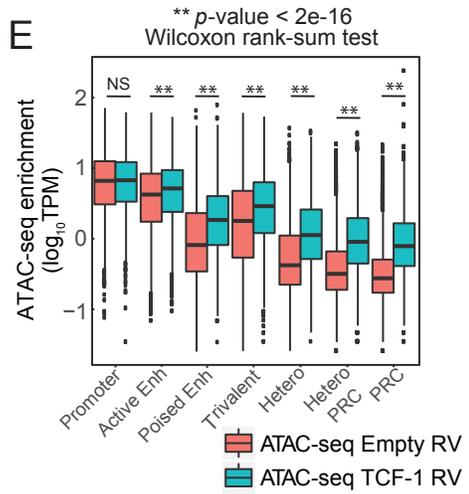
Optimal number of clusters



D



E



**Figure S5. TCF-1 can bind to repressed chromatin and promote accessibility,  
Related to Figure 6**

(A) Two replicates of H3K9me3 and H3K27me3 ChIP-seq from NIH3T3 cells were generated as well as an input control (see STAR methods) and combined with public H3K4me3, H3K4me1 and H3K27ac in NIH3T3 and their corresponding control input. The enrichment of histone mark signal was calculated in a window (+/- 250bp) around TCF-1 ChIP-seq summits from TCF-1 RV NIH3T3 cells with normR algorithm and used as input for the Principal component analysis (see **Figure 6A**). The enrichment of ATAC-seq in TCF-1 RV vs Empty RV NIH3T3 and vice versa was also calculated and used in the same analysis.

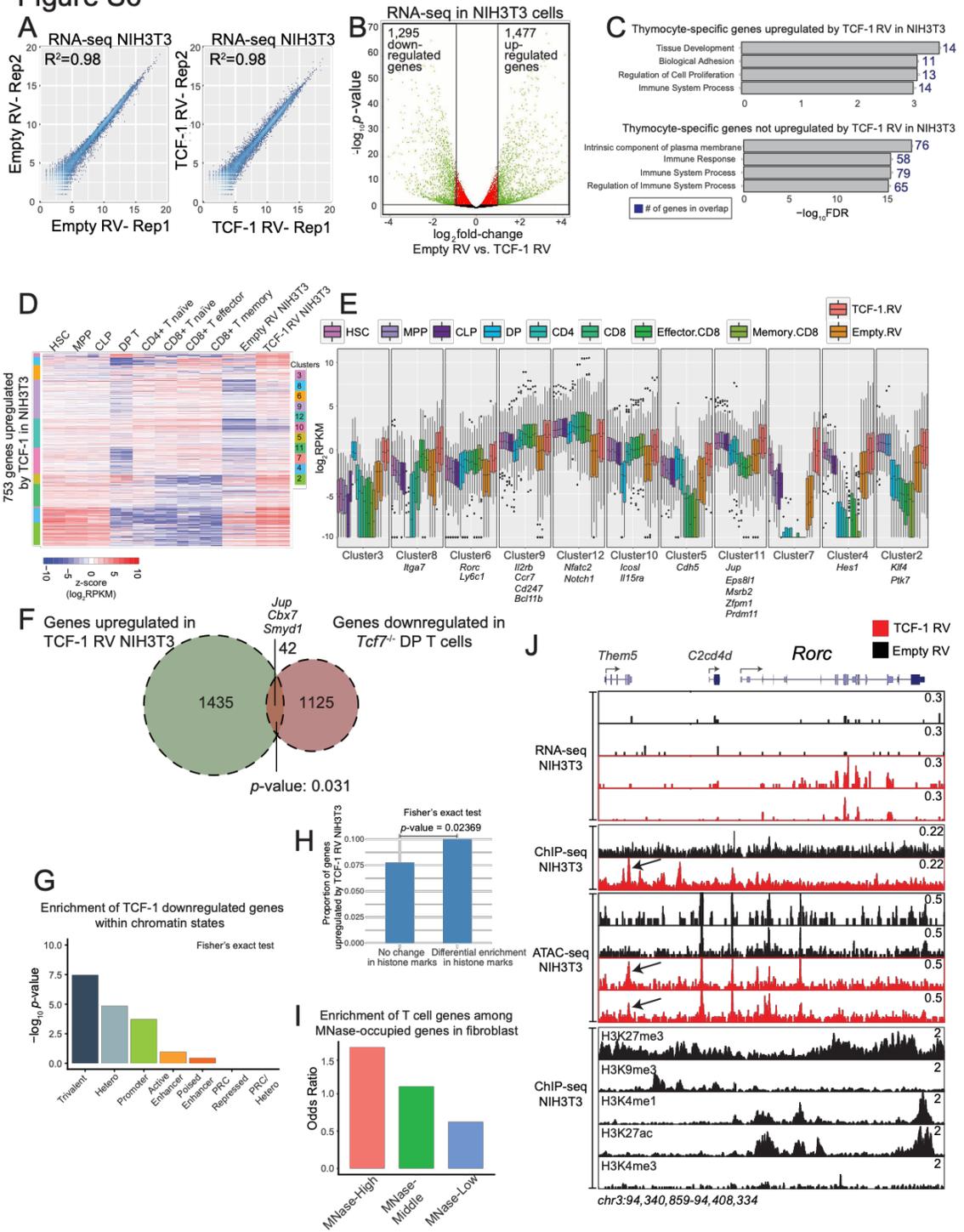
(B) Pairwise Spearman correlation coefficient between the enrichment (A) of different histone marks and chromatin accessibility surrounding TCF-1 ChIP-seq binding sites.

(C) Average silhouette width plot for the identification of the optimal number of clusters when applying k-means on the adjusted significance level of enrichment of H3K4me1, H3K4me3, H3K9me3, H3K27me3 and H3K27ac ChIP-seq signal in the 2kbp window centered on TCF-1 ChIP-seq summits.

(D) Heatmap of the adjusted significance level of histone mark enrichment around TCF-1 summits separated into the identified 11 clusters (C). Chromatin states were defined based on the enrichment level of each histone mark in each cluster.

(E) Wilcoxon rank-sum test to assess the statistical significance of the difference in ATAC-seq signal enrichment between Empty and TCF-1 RV NIH3T3 in the chromatin states from (D).

Figure S6



**Figure S6. T cell-specific genes innately repressed in fibroblasts are upregulated by TCF-1, Related to Figure 7**

(A) Pearson correlation plots between RNA-seq replicates in Empty RV NIH3T3 and TCF-1 RV NIH3T3 cells for assessing the reproducibility of the data.

(B) Volcano plot demonstrating differential gene expression using DESeq2 in three RNA-seq replicates of Empty RV and TCF-1 RV NIH3T3 cells with the upregulation of 1,477 genes and the downregulation of 1,295 genes (fold-change > 1.5 and  $p$ -value <  $1e-3$ ).

(C) Gene ontology analysis for thymocyte-specific genes (see **Figure 7D**) overlapping TCF-1 upregulated genes in NIH3T3 (top) and thymocyte-specific genes not upregulated by TCF-1 in TCF-1 RV NIH3T3 (bottom).

(D) RNA-seq counts on gene exons from HSC, MPP, CLP, DP, naïve CD4<sup>+</sup> SP, naïve CD8<sup>+</sup> SP, effector CD8<sup>+</sup> and memory CD8<sup>+</sup> T cells were calculated for genes upregulated by TCF-1 in TCF-1 RV NIH3T3 cells (B). Genes were clustered in 12 groups using k-means after reducing the heteroskedasticity of the data by applying variance-stabilizing transformation with DESeq2. We subsequently calculated normalized expression values (RPKM) for filtering lowly expressed genes (RPKM < 0.5 in all hematopoietic cell types) and visualizing the results. Cluster 1 was excluded from any further analysis due to low expression levels.

(E) Boxplots of the normalized gene expression levels (D) with representative examples for each cluster.

(F) TCF-1 is sufficient to upregulate TCF-1 dependent T cell genes in fibroblasts. Genes downregulated in *Tcf7*<sup>-/-</sup> DP T cells (see **Figure S2**) were overlapped with genes upregulated by TCF-1 RV in NIH3T3 and the statistical significance of the enrichment was calculated with Fisher's exact test.

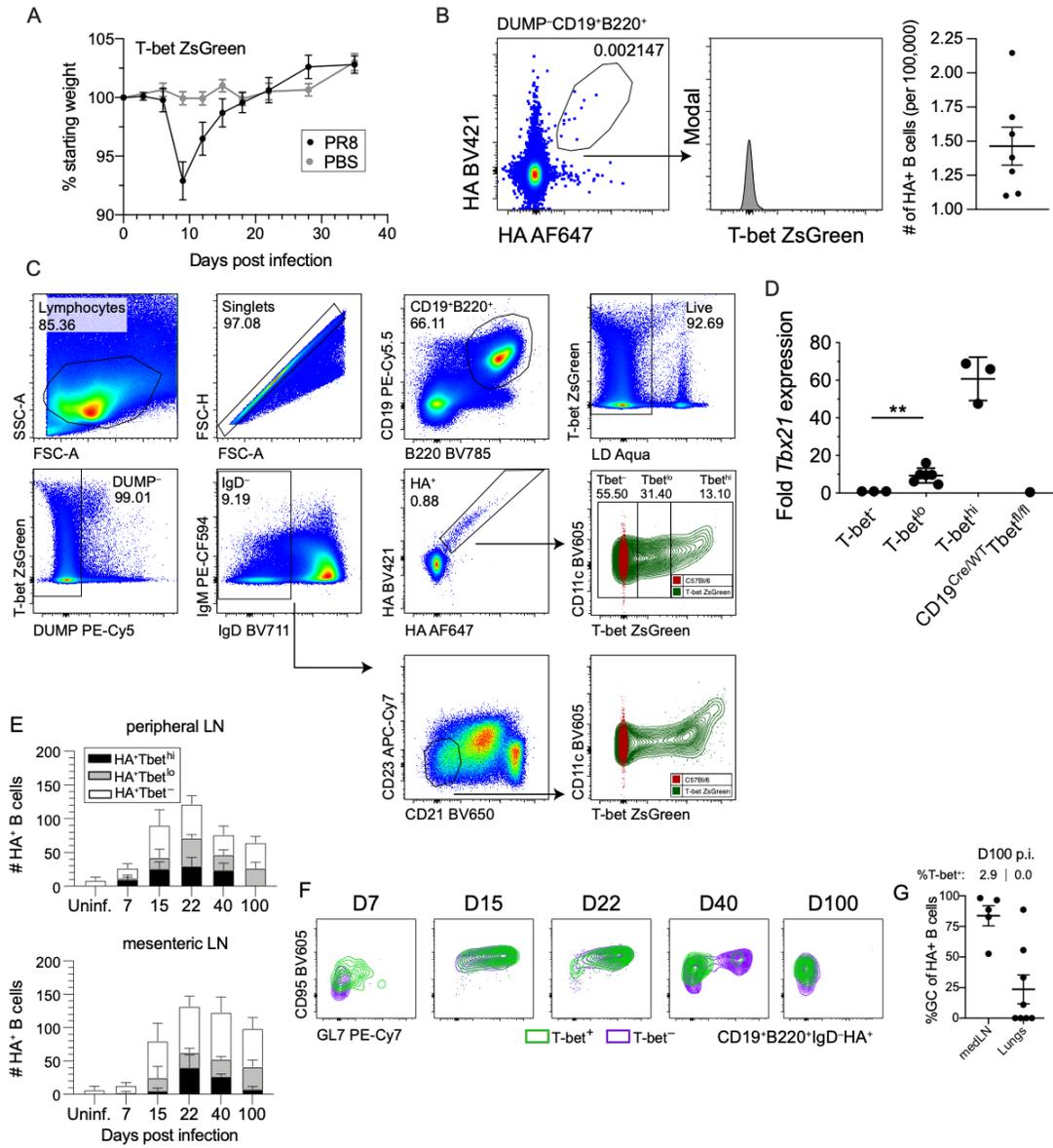
(G) TCF-1 ChIP-seq summits (see **Figure 5A**) were assigned to chromatin states (see **Figure 6B**) and linked to proximal genes (STAR methods). The enrichment of downregulated genes by TCF-1 within each chromatin state was assessed with Fisher's exact test.

(H) A statistically significant (tested with Fisher's exact test) higher proportion of genes were proximal to TCF-1 binding events that led to gain in H3K27ac and loss of H3K27me3/H3K9me3 modifications (**Figure 6D**) in contrast to those binding events that did not alter the chromatin state.

(I) The enrichment of T cell genes in different levels of nucleosome occupancy (see **Figure 7B**, **S4C**, and STAR methods) was calculated using Fisher's exact test.

(J) ATAC-, RNA- and ChIP-seq (histones and TCF-1) profiles in Empty and TCF-1 RV NIH3T3 cells in *Rorc* locus as an example of a key T cell gene that is innately not expressed in fibroblasts.

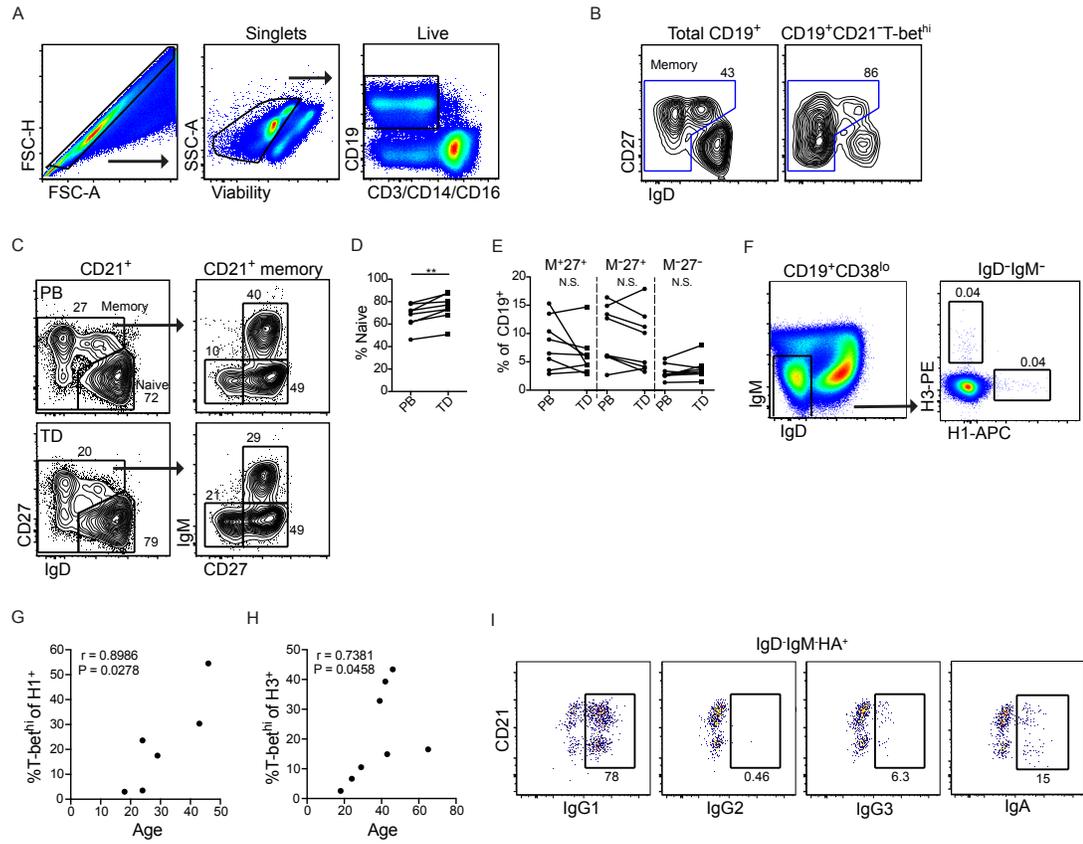
Figure S7



**Figure S7. Characterization of HA-specific B cells after influenza infection, Related to Figure 8.**

(A) Weight loss and recovery from PR8 influenza infection in T-bet-ZsGreen mice compared to PBS-treated controls. (B) Fluorescently-conjugated PR8 hemagglutinin (HA) detects the precursor frequency of HA-specific B cells in naïve T-bet-ZsGreen mice, which are uniformly T-bet<sup>+</sup>. The naïve precursor frequency per 100,000 B cells is plotted. (C) Gating scheme for the identification of T-bet-ZsGreen mouse HA<sup>+</sup> B cells and subsetting into T-bet<sup>-</sup>, T-bet<sup>lo</sup>, and T-bet<sup>hi</sup> populations via flow cytometry. C57Bl/6 mice are included in the T-bet-ZsGreen expression plot as a control. An identical gating scheme was used for all tissues in **Figure 8**. (D) *Tbx21* expression in sorted T-bet<sup>-</sup>, T-bet<sup>lo</sup>, and T-bet<sup>hi</sup> B cell subsets via qPCR. CD19<sup>+</sup> B cells were sorted into the corresponding subsets according to ZsGreen expression (C), and RNA was isolated and cDNA prepared for qRT-PCR analysis. (E) Number of HA<sup>+</sup> B cells in peripheral and mesenteric lymph nodes by T-bet expression phenotype at different time points after infection; mice are the same as in **Figure 8B**. (F) Expression of GL7 and CD95 on T-bet<sup>-</sup> and T-bet<sup>+</sup> splenic HA<sup>+</sup> B cells at the indicated time points post PR8 infection. (G) The percentage of HA<sup>+</sup> B cells that are GL7<sup>+</sup>CD38<sup>-</sup> in the mediastinal LN and lungs of T-bet-ZsGreen mice 100 days post PR8 infection, and the percentage that are T-bet<sup>+</sup> in each tissue (top). Data in (D) are plotted as mean ± SEM. HA<sup>+</sup> B cells were identified as live, singlet, DUMP<sup>-</sup>, B220<sup>+</sup>, CD19<sup>+</sup>, IgD<sup>-</sup>, HA-AF647<sup>+</sup>, HA-BV421<sup>+</sup> cells. DUMP gate includes CD4, CD8, Gr-1, and F4/80.

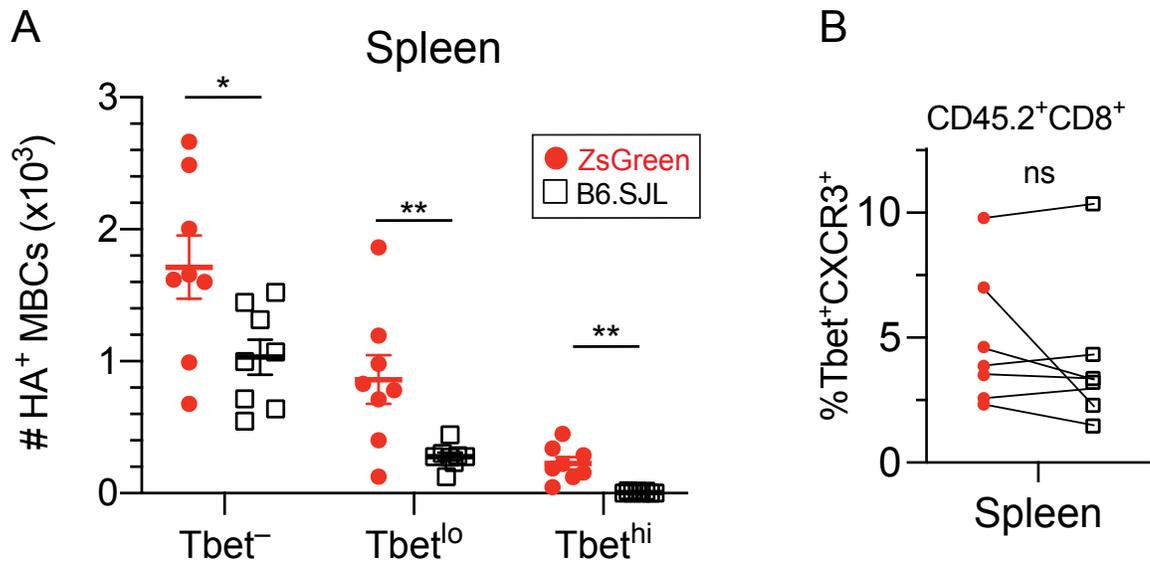
Figure S8



**Figure S8. Characterization of human B cell subsets, Related to Figure 9.**

(A) Gating scheme for the identification of human peripheral blood CD19<sup>+</sup> B cells via flow cytometry. An identical gating scheme and identical gates were used for all tissues in Figure 9. (B) Memory (IgD<sup>-</sup>/IgD<sup>+</sup>CD27<sup>+</sup>) phenotype of total CD19<sup>+</sup> and T-bet<sup>hi</sup> B cells from spleen of a representative donor. (C) Identification of naïve (CD21<sup>+</sup>IgD<sup>+</sup>CD27<sup>-</sup>) and CD21<sup>+</sup> memory (IgD<sup>-</sup>/IgD<sup>+</sup>CD27<sup>+</sup>) B cells in paired peripheral blood (PB) and thoracic duct fluid (TD) from a representative donor. Memory cells are further subsetted into IgM<sup>+</sup>CD27<sup>+</sup>, IgM<sup>-</sup>CD27<sup>+</sup>, and IgM<sup>-</sup>CD27<sup>-</sup>. (D and E) Frequency of naïve B cells (D) and CD21<sup>+</sup> memory subsets (E) gated in (C) within paired PB and TD sample cohort (n=8). (F) Identification of HA-specific, class-switched (IgD<sup>-</sup>IgM<sup>-</sup>) B cells within CD19<sup>+</sup>CD38<sup>low</sup> mesLN B cells using fluorescently-labelled A/California/07/2009 HA probe (H1 strain) or a A/Wisconsin/67/2005 HA probe (H3 strain). (G and H) Correlation between frequency of T-bet<sup>hi</sup> phenotype within class-switched H1-binding (G) or H3-binding (H) B cells and subject age. Subjects are the same as in Figure 9I; those without age information were omitted from this analysis. (I) IgG1, IgG2, IgG3, and IgA expression profile by class-switched H1-HA<sup>+</sup> splenic B cells from a representative human donor. Statistical comparisons performed using paired t-test (D and E) and Spearman correlation (G and H). N.S. = not significant, \*\*p<0.01. *Credit goes to James J. Knox for generating this figure.*

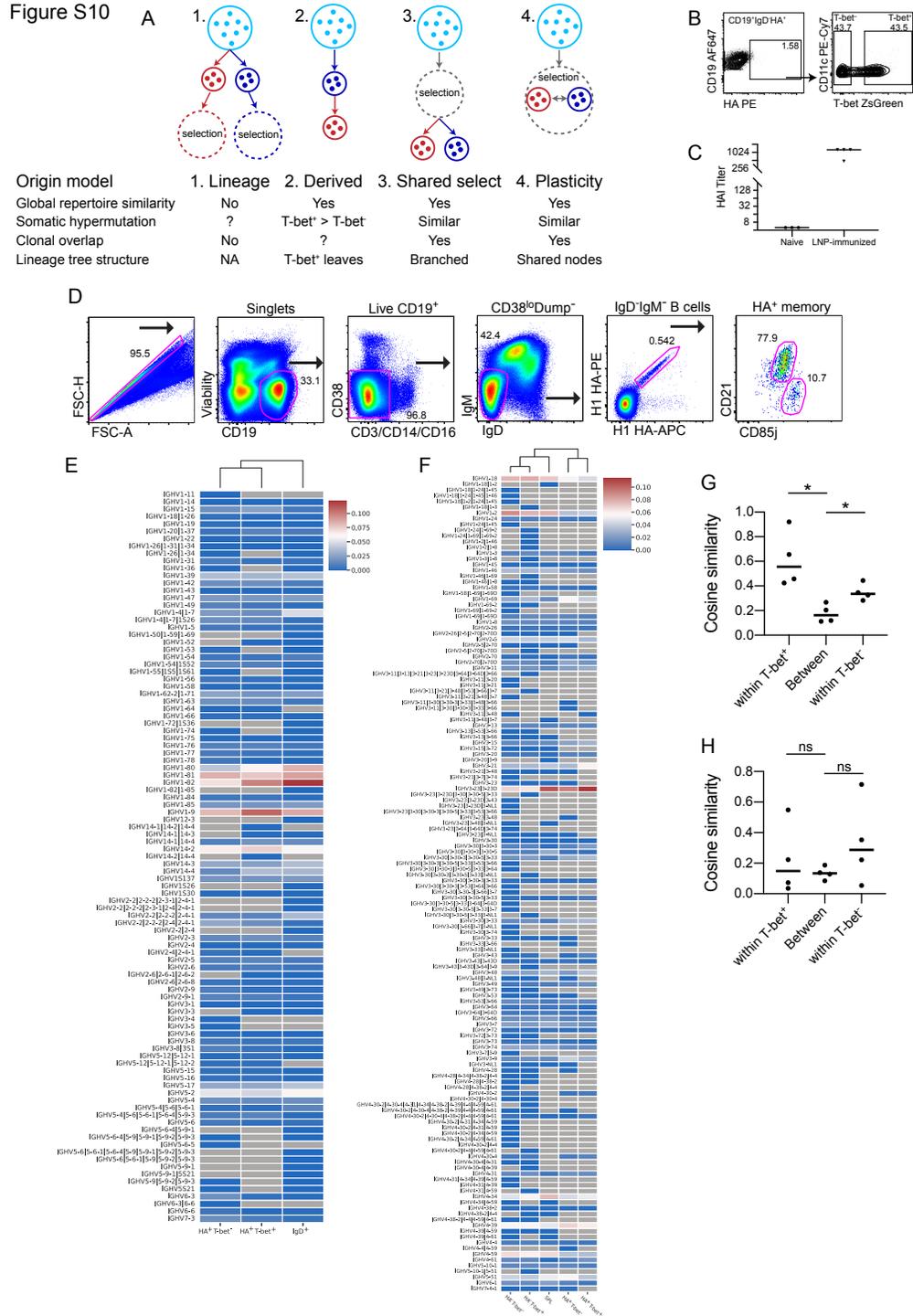
Figure S9



**Figure S9. Enumeration of HA-specific B cells in parabiotic partners, Related to Figure 10.**

(A) Absolute numbers of Tbet<sup>-</sup>, Tbet<sup>lo</sup> and Tbet<sup>hi</sup> HA<sup>+</sup> splenic MBCs in ZsGreen (red) and B6.SJL (black) parabiosis partners from Figure 10. (B) The percent of splenic donor-derived CD8<sup>+</sup> T cells that are Tbet-ZsGreen<sup>+</sup>CXCR3<sup>+</sup> in each partner  $\geq 17$  days of parabiosis from 8 parabiotic pairs. Statistical comparisons performed using paired two-tailed t-test. ns = not significant, \* $p < 0.05$ , \*\* $p < 0.01$

Figure S10



**Figure S10. Immune repertoire analysis by IgH sequencing, Related to Figure 11.**

(A) Models regarding the possible origin of T-bet<sup>+</sup> and T-bet<sup>-</sup> MBCs and their distinguishing characteristics. (B) Scheme for sorting T-bet<sup>+</sup> and T-bet<sup>-</sup> HA-specific MBCs from negatively depleted, CD19<sup>+</sup>IgD<sup>-</sup>HA<sup>+</sup> splenocytes 90 days post HA mRNA LNP immunization. (C) HAI titers at 21 days post immunization. (D) Gating scheme for the sorting of human splenic IgD<sup>-</sup>IgM<sup>-</sup>HA<sup>+</sup> memory B cell subsets, defined by CD21 and CD85j expression. (E and F) VH gene usage in mice (E) and humans (F) computed based upon clonal usage (each clone is only counted once). Sequencing libraries from the same subject (human or mouse) and subset were pooled. Clones having less than half the mean copy number frequency within that subject/subset sequencing library were excluded from the analysis. Clones from the same species and subset were then pooled. Gray cells indicate no data and the VH usage was normalized by subset in both panels. Clone counts for mice (E) are HA<sup>+</sup>T-bet<sup>+</sup> 1,605; HA<sup>+</sup>T-bet<sup>-</sup> 1,993; and IgD<sup>+</sup> 21,889. Clone counts for humans (F) are HA<sup>+</sup>CD21<sup>+</sup> 5,739; HA<sup>+</sup>CD21<sup>-</sup>CD85j<sup>hi</sup> 2,018; HA<sup>-</sup>CD21<sup>+</sup> 31,396; HA<sup>-</sup>CD21<sup>-</sup>CD85j<sup>hi</sup> 21,442; and SPL 10,125. (G and H) Cosine similarity between sequencing replicates of each subset and between subsets in mice (G) and humans (H). Statistical comparisons in (G) and (H) performed using paired t-test, \*p<0.05.

## BIBLIOGRAPHY

1. Garvey JS. Cellular Differentiation in the Immune System. *Science*. 1969;165:205-211.
2. Iwafuchi-Doi M, and Zaret KS. Cell fate control by pioneer transcription factors. *Development*. 2016;143:1833-1837.
3. Heinz S, Romanoski CE, Benner C, et al. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol*. 2015;16:144-154.
4. Jacob F, and Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology*. 1961;3:318-356.
5. Britten RJ, and Davidson EH. Gene regulation for higher cells: a theory. *Science*. 1969;165:349-357.
6. Burnet FM. A Modification of Jerne's Theory of Antibody Production using the Concept of Clonal Selection. *The Australian Journal of Science*. 1957;20:67-68.
7. Slack JMW. *From Egg to Embryo: Regional Specification in Early Development*, 2 Edition. 1991. (Cambridge: Cambridge University Press).(1991).
8. Davidson EH. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. 2006. (Academic).(2006).
9. Johnson JL, Georgakilas G, Petrovic J, et al. Lineage-Determining Transcription Factor TCF-1 Initiates the Epigenetic Identity of T Cells. *Immunity*. 2018;48:243-257.e210.
10. Struhl K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell*. 1999;98:1-4.
11. Horn PJ, and Peterson CL. Molecular biology. Chromatin higher order folding--wrapping up transcription. *Science*. 2002;297:1824-1827.

12. Monticelli S, and Natoli G. Transcriptional determination and functional specificity of myeloid cells: making sense of diversity. *Nat Rev Immunol*. 2017.
13. Morrison SJ, Uchida N, and Weissman IL. The biology of hematopoietic stem cells. *Annual review of cell and developmental biology*. 1995;11:35-71.
14. Spangrude G, Heimfeld S, and Weissman I. Purification and characterization of mouse hematopoietic stem cells. *Science*. 1988;241:58-62.
15. Morrison SJ, and Weissman IL. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity*. 1994;1:661-673.
16. Kiel MJ, Yilmaz OH, Iwashita T, et al. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell*. 2005;121:1109-1121.
17. Adolfsson J, Borge OJ, Bryder D, et al. Upregulation of Flt3 expression within the bone marrow Lin(-)Sca1(+)c-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity. *Immunity*. 2001;15:659-669.
18. Adolfsson J, Månsson R, Buza-Vidas N, et al. Identification of Flt3+ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential: A Revised Road Map for Adult Blood Lineage Commitment. *Cell*. 2005;121:295-306.
19. Kondo M, Weissman IL, and Akashi K. Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow. *Cell*. 1997;91:661-672.
20. Mansson R, Zandi S, Welinder E, et al. Single-cell analysis of the common lymphoid progenitor compartment reveals functional and molecular heterogeneity. *Blood*. 2010;115:2601-2609.

21. Chi AW, Chavez A, Xu L, et al. Identification of Flt3(+)CD150(-) myeloid progenitors in adult mouse bone marrow that harbor T lymphoid developmental potential. *Blood*. 2011;118:2723-2732.
22. Miller JF, and Osoba D. Current concepts of the immunological function of the thymus. *Physiological reviews*. 1967;47:437-520.
23. Koch U, and Radtke F. Mechanisms of T Cell Development and Transformation. *Annual review of cell and developmental biology*. 2011;27:539-562.
24. Donskoy E, and Goldschneider I. Thymocytopoiesis is maintained by blood-borne precursors throughout postnatal life. A study in parabiotic mice. *J Immunol*. 1992;148:1604-1612.
25. Spangrude GJ, and Scollay R. Differentiation of hematopoietic stem cells in irradiated mouse thymic lobes. Kinetics and phenotype of progeny. *J Immunol*. 1990;145:3661-3668.
26. Schwarz BA, Sambandam A, Maillard I, et al. Selective thymus settling regulated by cytokine and chemokine receptors. *J Immunol*. 2007;178:2008-2017.
27. Zlotoff DA, Sambandam A, Logan TD, et al. CCR7 and CCR9 together recruit hematopoietic progenitors to the adult thymus. *Blood*. 2010;115:1897-1905.
28. Krueger A, Willenzon S, Lyszkiewicz M, et al. CC chemokine receptor 7 and 9 double-deficient hematopoietic progenitors are severely impaired in seeding the adult thymus. *Blood*. 2010;115:1906-1912.
29. Gossens K, Naus S, Corbel SY, et al. Thymic progenitor homing and lymphocyte homeostasis are linked via S1P-controlled expression of thymic P-selectin/CCL25. *J Exp Med*. 2009;206:761-778.

30. Rossi FM, Corbel SY, Merzaban JS, et al. Recruitment of adult thymic progenitors is regulated by P-selectin and its ligand PSGL-1. *Nat Immunol.* 2005;6:626-634.
31. Allman D, Sambandam A, Kim S, et al. Thymopoiesis independent of common lymphoid progenitors. *Nat Immunol.* 2003;4:168-174.
32. Porritt HE, Rumpf LL, Tabrizifard S, et al. Heterogeneity among DN1 prothymocytes reveals multiple progenitors with different capacities to generate T cell and non-T cell lineages. *Immunity.* 2004;20:735-745.
33. Benz C, and Bleul CC. A multipotent precursor in the thymus maps to the branching point of the T versus B lineage decision. *J Exp Med.* 2005;202:21-31.
34. Heinzl K, Benz C, Martins VC, et al. Bone marrow-derived hemopoietic precursors commit to the T cell lineage only after arrival in the thymic microenvironment. *J Immunol.* 2007;178:858-868.
35. Wada H, Masuda K, Satoh R, et al. Adult T-cell progenitors retain myeloid potential. *Nature.* 2008;452:768-772.
36. Bell JJ, and Bhandoola A. The earliest thymic progenitors for T cells possess myeloid lineage potential. *Nature.* 2008;452:764-767.
37. Porritt HE, Gordon K, and Petrie HT. Kinetics of steady-state differentiation and mapping of intrathymic-signaling environments by stem cell transplantation in nonirradiated mice. *J Exp Med.* 2003;198:957-962.
38. Masuda K, Kakugawa K, Nakayama T, et al. T cell lineage determination precedes the initiation of TCR beta gene rearrangement. *J Immunol.* 2007;179:3699-3706.
39. Yui MA, Feng N, and Rothenberg EV. Fine-scale staging of T cell lineage commitment in adult mouse thymus. *J Immunol.* 2010;185:284-293.

40. Kueh HY, Yui MA, Ng KK, et al. Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. *Nat Immunol.* 2016;17:956-965.
41. Bassing CH, Swat W, and Alt FW. The mechanism and regulation of chromosomal V(D)J recombination. *Cell.* 2002;109 Suppl:S45-55.
42. Sleckman BP, Bassing CH, Bardon CG, et al. Accessibility control of variable region gene assembly during T-cell development. *Immunol Rev.* 1998;165:121-130.
43. Capone M, Hockett RD, Jr., and Zlotnik A. Kinetics of T cell receptor beta, gamma, and delta rearrangements during adult thymic development: T cell receptor rearrangements are present in CD44(+)CD25(+) Pro-T thymocytes. *Proc Natl Acad Sci U S A.* 1998;95:12522-12527.
44. Taghon T, Yui MA, Pant R, et al. Developmental and molecular characterization of emerging beta- and gammadelta-selected pre-T cells in the adult mouse thymus. *Immunity.* 2006;24:53-64.
45. Aifantis I, Mandal M, Sawai K, et al. Regulation of T-cell progenitor survival and cell-cycle entry by the pre-T-cell receptor. *Immunol Rev.* 2006;209:159-169.
46. Michie AM, and Zuniga-Pflucker JC. Regulation of thymocyte differentiation: pre-TCR signals and beta-selection. *Semin Immunol.* 2002;14:311-323.
47. Godfrey DI, Kennedy J, Suda T, et al. A developmental pathway involving four phenotypically and functionally distinct subsets of CD3-CD4-CD8- triple-negative adult mouse thymocytes defined by CD44 and CD25 expression. *J Immunol.* 1993;150:4244-4252.

48. Singer A, Adoro S, and Park JH. Lineage fate and intense debate: myths, models and mechanisms of CD4- versus CD8-lineage choice. *Nat Rev Immunol.* 2008;8:788-801.
49. Kisielow P, Teh HS, Bluthmann H, et al. Positive selection of antigen-specific T cells in thymus by restricting MHC molecules. *Nature.* 1988;335:730-733.
50. Jameson SC, Hogquist KA, and Bevan MJ. Positive selection of thymocytes. *Annu Rev Immunol.* 1995;13:93-126.
51. Klein L, Kyewski B, Allen PM, et al. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol.* 2014;14:377-391.
52. Matloubian M, Lo CG, Cinamon G, et al. Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature.* 2004;427:355-360.
53. Yashiro-Ohtani Y, Ohtani T, and Pear WS. Notch regulation of early thymocyte development. *Semin Immunol.* 2010;22:261-269.
54. Besseyrias V, Fiorini E, Strobl LJ, et al. Hierarchy of Notch-Delta interactions promoting T cell lineage commitment and maturation. *J Exp Med.* 2007;204:331-343.
55. Nam Y, Sliz P, Song L, et al. Structural basis for cooperativity in recruitment of MAML coactivators to Notch transcription complexes. *Cell.* 2006;124:973-983.
56. Wallberg AE, Pedersen K, Lendahl U, et al. p300 and PCAF act cooperatively to mediate transcriptional activation from chromatin templates by notch intracellular domains in vitro. *Mol Cell Biol.* 2002;22:7812-7819.

57. Sambandam A, Maillard I, Zediak VP, et al. Notch signaling controls the generation and differentiation of early T lineage progenitors. *Nat Immunol.* 2005;6:663-670.
58. Pear WS, Aster JC, Scott ML, et al. Exclusive development of T cell neoplasms in mice transplanted with bone marrow expressing activated Notch alleles. *J Exp Med.* 1996;183:2283-2291.
59. Pui JC, Allman D, Xu L, et al. Notch1 expression in early lymphopoiesis influences B versus T lineage determination. *Immunity.* 1999;11:299-308.
60. Radtke F, Wilson A, Stark G, et al. Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity.* 1999;10:547-558.
61. Spain LM, Guerriero A, Kunjibettu S, et al. T cell development in PU.1-deficient mice. *J Immunol.* 1999;163:2681-2687.
62. Klemsz MJ, McKercher SR, Celada A, et al. The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. *Cell.* 1990;61:113-124.
63. Dionne CJ, Tse KY, Weiss AH, et al. Subversion of T lineage commitment by PU.1 in a clonal cell line system. *Developmental biology.* 2005;280:448-466.
64. Anderson MK, Hernandez-Hoyos G, Diamond RA, et al. Precise developmental regulation of Ets family transcription factors during specification and commitment to the T cell lineage. *Development.* 1999;126:3131-3148.
65. Rosenbauer F, Owens BM, Yu L, et al. Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU.1. *Nat Genet.* 2006;38:27-37.
66. Zarnegar MA, Chen J, and Rothenberg EV. Cell-type-specific activation and repression of PU.1 by a complex of discrete, functionally specialized cis-regulatory elements. *Mol Cell Biol.* 2010;30:4922-4939.

67. Quong MW, Romanow WJ, and Murre C. E protein function in lymphocyte development. *Annu Rev Immunol.* 2002;20:301-322.
68. Murre C, McCaw PS, and Baltimore D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell.* 1989;56:777-783.
69. Ikawa T, Kawamoto H, Goldrath AW, et al. E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment. *J Exp Med.* 2006;203:1329-1342.
70. Barndt R, Dai MF, and Zhuang Y. A novel role for HEB downstream or parallel to the pre-TCR signaling pathway during alpha beta thymopoiesis. *J Immunol.* 1999;163:3331-3343.
71. Yashiro-Ohtani Y, He Y, Ohtani T, et al. Pre-TCR signaling inactivates Notch1 transcription by antagonizing E2A. *Genes Dev.* 2009;23:1665-1676.
72. Ting CN, Olson MC, Barton KP, et al. Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature.* 1996;384:474-478.
73. Hosoya T, Kuroha T, Moriguchi T, et al. GATA-3 is required for early T lineage progenitor development. *J Exp Med.* 2009;206:2987-3000.
74. Taghon T, Yui MA, and Rothenberg EV. Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. *Nat Immunol.* 2007;8:845-855.
75. Li L, Leid M, and Rothenberg EV. An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science.* 2010;329:89-93.
76. Ikawa T, Hirose S, Masuda K, et al. An essential developmental checkpoint for production of the T cell lineage. *Science.* 2010;329:93-96.

77. Li P, Burke S, Wang J, et al. Reprogramming of T cells to natural killer-like cells upon Bcl11b deletion. *Science*. 2010;329:85-89.
78. Cobaleda C, Jochum W, and Busslinger M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature*. 2007;449:473-477.
79. Laudet V, Stehelin D, and Clevers H. Ancestry and diversity of the HMG box superfamily. *Nucleic Acids Res*. 1993;21:2493-2501.
80. van de Wetering M, and Clevers H. Sequence-specific interaction of the HMG box proteins TCF-1 and SRY occurs within the minor groove of a Watson-Crick double helix. *Embo j*. 1992;11:3039-3044.
81. Giese K, Cox J, and Grosschedl R. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell*. 1992;69:185-195.
82. Love JJ, Li X, Case DA, et al. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*. 1995;376:791-795.
83. van de Wetering M, Oosterwegel M, Dooijes D, et al. Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. *Embo j*. 1991;10:123-132.
84. Waterman ML, and Jones KA. Purification of TCF-1 alpha, a T-cell-specific transcription factor that activates the T-cell receptor C alpha gene enhancer in a context-dependent manner. *The New biologist*. 1990;2:621-636.
85. Gastrop J, Hoevenagel R, Young JR, et al. A common ancestor of the mammalian transcription factors TCF-1 and TCF-1 alpha/LEF-1 expressed in chicken T cells. *Eur J Immunol*. 1992;22:1327-1330.

86. Oosterwegel M, van de Wetering M, Timmerman J, et al. Differential expression of the HMG box factors TCF-1 and LEF-1 during murine embryogenesis. *Development*. 1993;118:439-448.
87. Verbeek S, Izon D, Hofhuis F, et al. An HMG-box-containing T-cell factor required for thymocyte differentiation. *Nature*. 1995;374:70-74.
88. van Genderen C, Okamura RM, Farinas I, et al. Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes Dev*. 1994;8:2691-2703.
89. Okamura RM, Sigvardsson M, Galceran J, et al. Redundant regulation of T cell differentiation and TCRalpha gene expression by the transcription factors LEF-1 and TCF-1. *Immunity*. 1998;8:11-20.
90. Cadigan KM, and Waterman ML. TCF/LEFs and Wnt signaling in the nucleus. *Cold Spring Harb Perspect Biol*. 2012;4.
91. Cadigan KM, and Nusse R. Wnt signaling: a common theme in animal development. *Genes Dev*. 1997;11:3286-3305.
92. Bhanot P, Brink M, Samos CH, et al. A new member of the frizzled family from *Drosophila* functions as a Wingless receptor. *Nature*. 1996;382:225-230.
93. Cavallo RA, Cox RT, Moline MM, et al. *Drosophila* Tcf and Groucho interact to repress Wingless signalling activity. *Nature*. 1998;395:604-608.
94. Behrens J, von Kries JP, Kuhl M, et al. Functional interaction of beta-catenin with the transcription factor LEF-1. *Nature*. 1996;382:638-642.
95. Cobas M, Wilson A, Ernst B, et al. Beta-catenin is dispensable for hematopoiesis and lymphopoiesis. *J Exp Med*. 2004;199:221-229.

96. Koch U, Wilson A, Cobas M, et al. Simultaneous loss of beta- and gamma-catenin does not perturb hematopoiesis or lymphopoiesis. *Blood*. 2008;111:160-164.
97. Van de Wetering M, Castrop J, Korinek V, et al. Extensive alternative splicing and dual promoter usage generate Tcf-1 protein isoforms with differential transcription control properties. *Mol Cell Biol*. 1996;16:745-752.
98. Rothenberg EV, Moore JE, and Yui MA. Launching the T-cell-lineage developmental programme. *Nat Rev Immunol*. 2008;8:9-21.
99. Miyazaki M, Miyazaki K, Chen K, et al. The E-Id Protein Axis Specifies Adaptive Lymphoid Cell Identity and Suppresses Thymic Innate Lymphoid Cell Development. *Immunity*. 2017;46:818-834 e814.
100. van Oevelen C, Collombet S, Vicent G, et al. C/EBPalpha Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. *Stem Cell Reports*. 2015;5:232-247.
101. Ghisletti S, Barozzi I, Mietton F, et al. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity*. 2010;32:317-328.
102. Ostuni R, Piccolo V, Barozzi I, et al. Latent enhancers activated by stimulation in differentiated cells. *Cell*. 2013;152:157-171.
103. Gyory I, Boller S, Nechanitzky R, et al. Transcription factor Ebf1 regulates differentiation stage-specific signaling, proliferation, and survival of B cells. *Genes Dev*. 2012;26:668-682.
104. Boller S, Ramamoorthy S, Akbas D, et al. Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. *Immunity*. 2016;44:527-541.

105. Di Stefano B, Sardina JL, van Oevelen C, et al. C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature*. 2014;506:235-239.
106. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576-589.
107. Ciofani M, Madar A, Galan C, et al. A validated regulatory network for th17 cell specification. *Cell*. 2012;151:289-303.
108. Samstein RM, Arvey A, Josefowicz SZ, et al. Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell*. 2012;151:153-166.
109. Shih HY, Sciume G, Mikami Y, et al. Developmental Acquisition of Regulomes Underlies Innate Lymphoid Cell Functionality. *Cell*. 2016;165:1120-1133.
110. Vahedi G, Kanno Y, Furumoto Y, et al. Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature*. 2015;520:558-562.
111. Vahedi G, Takahashi H, Nakayamada S, et al. STATs shape the active enhancer landscape of T cell populations. *Cell*. 2012;151:981-993.
112. Kurachi M, Barnitz RA, Yosef N, et al. The transcription factor BATF operates as an essential differentiation checkpoint in early effector CD8+ T cells. *Nat Immunol*. 2014;15:373-383.
113. Pauken KE, Sammons MA, Odorizzi PM, et al. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science*. 2016;354:1160-1165.
114. Yu B, Zhang K, Milner JJ, et al. Epigenetic landscapes reveal transcription factors that regulate CD8+ T cell differentiation. *Nat Immunol*. 2017;18:573-582.

115. Scott-Browne JP, Lopez-Moyado IF, Trifari S, et al. Dynamic Changes in Chromatin Accessibility Occur in CD8+ T Cells Responding to Viral Infection. *Immunity*. 2016;45:1327-1340.
116. Scharer CD, Bally AP, Gandham B, et al. Cutting Edge: Chromatin Accessibility Programs CD8 T Cell Memory. *J Immunol*. 2017;198:2238-2243.
117. Philip M, Fairchild L, Sun L, et al. Chromatin states define tumour-specific T cell dysfunction and reprogramming. *Nature*. 2017;545:452-456.
118. Dose M, Emmanuel AO, Chaumeil J, et al. beta-Catenin induces T-cell transformation by promoting genomic instability. *Proc Natl Acad Sci U S A*. 2014;111:391-396.
119. Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, et al. Chromatin state dynamics during blood formation. *Science*. 2014;345:943-949.
120. Zhang JA, Mortazavi A, Williams BA, et al. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell*. 2012;149:467-482.
121. Weber BN, Chi AW, Chavez A, et al. A critical role for TCF-1 in T-lineage specification and differentiation. *Nature*. 2011;476:63-68.
122. Zhou X, Yu S, Zhao DM, et al. Differentiation and persistence of memory CD8(+) T cells depend on T cell factor 1. *Immunity*. 2010;33:229-240.
123. Osipovich O, and Oltz EM. Regulation of antigen receptor gene assembly by genetic-epigenetic crosstalk. *Semin Immunol*. 2010;22:313-322.
124. Li L, Zhang JA, Dose M, et al. A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. *Blood*. 2013;122:902-911.
125. Xing S, Li F, Zeng Z, et al. Tcf1 and Lef1 transcription factors establish CD8(+) T cell identity through intrinsic HDAC activity. *Nat Immunol*. 2016;17:695-703.

126. Love MI, Huber W, and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
127. Buenrostro JD, Wu B, Littenberger UM, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature.* 2015;523:486-490.
128. Schep AN, Wu B, Buenrostro JD, et al. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods.* 2017.
129. Landt SG, Marinov GK, Kundaje A, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 2012;22:1813-1831.
130. Soufi A, Garcia MF, Jaroszewicz A, et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell.* 2015;161:555-568.
131. Barozzi I, Simonatto M, Bonifacio S, et al. Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol Cell.* 2014;54:844-857.
132. Schep AN, Buenrostro JD, Denny SK, et al. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res.* 2015;25:1757-1770.
133. Heng TS, Painter MW, and Immunological Genome Project C. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol.* 2008;9:1091-1094.
134. Im SJ, Hashimoto M, Gerner MY, et al. Defining CD8<sup>+</sup> T cells that provide the proliferative burst after PD-1 therapy. *Nature.* 2016;537:417-421.
135. Steinke FC, and Xue HH. From inception to output, Tcf1 and Lef1 safeguard development of T cells and innate immune cells. *Immunol Res.* 2014;59:45-55.

136. Germar K, Dose M, Konstantinou T, et al. T-cell factor 1 is a gatekeeper for T-cell specification in response to Notch signaling. *Proc Natl Acad Sci U S A*. 2011;108:20060-20065.
137. Zhu J, Adli M, Zou JY, et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell*. 2013;152:642-654.
138. Misslitz A, Pabst O, Hintzen G, et al. Thymic T cell development and progenitor localization depend on CCR7. *J Exp Med*. 2004;200:481-491.
139. Trapnell C. Defining cell types and states with single-cell genomics. *Genome Res*. 2015;25:1491-1498.
140. Zaret KS, and Carroll JS. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*. 2011;25:2227-2241.
141. Oestreich KJ, and Weinmann AS. Master regulators or lineage-specifying? Changing views on CD4+ T cell transcription factors. *Nat Rev Immunol*. 2012;12:799-804.
142. Rothenberg EV. The chromatin landscape and transcription factors in T cell programming. *Trends Immunol*. 2014;35:195-204.
143. Vahedi G, Kanno Y, Sartorelli V, et al. Transcription factors and CD4 T cells seeking identity: masters, minions, setters and spikers. *Immunology*. 2013;139:294-298.
144. Soufi A, Donahue G, and Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*. 2012;151:994-1004.
145. Chronis C, Fiziev P, Papp B, et al. Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell*. 2017;168:442-459 e420.

146. Wohrle S, Wallmen B, and Hecht A. Differential control of Wnt target genes involves epigenetic mechanisms and selective promoter occupancy by T-cell factors. *Mol Cell Biol.* 2007;27:8164-8177.
147. Guo C, and Morris SA. Engineering cell identity: establishing new gene regulatory and chromatin landscapes. *Curr Opin Genet Dev.* 2017;46:50-57.
148. Buenrostro JD, Giresi PG, Zaba LC, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013;10:1213-1218.
149. Yashiro-Ohtani Y, Wang H, Zang C, et al. Long-range enhancer activity determines Myc sensitivity to Notch inhibitors in T cell leukemia. *Proc Natl Acad Sci U S A.* 2014;111:E4946-4953.
150. Andrews S (2017). FastQC.
151. Krueger F. Trim Galore! 2017.
152. Rosenbloom KR, Armstrong J, Barber GP, et al. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* 2015;43:D670-681.
153. Mudge JM, and Harrow J. Creating reference gene annotation for the mouse C57BL6/J genome assembly. *Mamm Genome.* 2015;26:366-378.
154. Langmead B, and Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357-359.
155. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15-21.
156. BroadInstitute (2017). Picard Tools.
157. Anders S, Pyl PT, and Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31:166-169.

158. Aken BL, Ayling S, Barrell D, et al. The Ensembl gene annotation system. Database (Oxford). 2016;2016.
159. Durinck S, Spellman PT, Birney E, et al. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4:1184-1191.
160. Ericson J, Davis S, Lesh J, et al. (2015). ImmGen microarray gene expression data: Data Generation and Quality Control pipeline.
161. Shen L, and Sinai M (2013). GeneOverlap: Test and visualize gene overlaps. R package version 1.10.0 Edition.
162. Liberzon A, Subramanian A, Pinchback R, et al. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27:1739-1740.
163. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545-15550.
164. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137.
165. Grant CE, Bailey TL, and Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011;27:1017-1018.
166. Quinlan AR, and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841-842.
167. Oksanen J, Blanchet FG, Friendly M, et al. (2017). vegan: Community Ecology Package. 2.4-3 Edition.
168. Schep A (2016). chromVAR: Chromatin Variation Across Regions. 0.5.0 Edition.
169. Chen J, Liu H, Liu J, et al. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat Genet. 2013;45:34-42.

170. Khan A, Fornes O, Stigliani A, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* 2017.
171. Helmuth J, and Chung HR (2016). normr: Normalization and difference calling in CHIP-seq data.
172. Johnson JL, Rosenthal RL, Knox JJ, et al. The Transcription Factor T-bet Resolves Memory B Cell Subsets with Distinct Tissue Distributions and Antibody Specificities in Mice and Humans. *Immunity.* 2020.
173. Iwasaki A, and Medzhitov R. Control of adaptive immunity by the innate immune system. *Nat Immunol.* 2015;16:343-353.
174. Medzhitov R, and Janeway CA, Jr. Decoding the patterns of self and nonself by the innate immune system. *Science.* 2002;296:298-300.
175. Schenten D, and Medzhitov R. The control of adaptive immune responses by the innate immune system. *Adv Immunol.* 2011;109:87-124.
176. Matzinger P. The danger model: a renewed sense of self. *Science.* 2002;296:301-305.
177. Klinman NR. The mechanism of antigenic stimulation of primary and secondary clonal precursor cells. *J Exp Med.* 1972;136:241-260.
178. Murphy KM, and Reiner SL. The lineage decisions of helper T cells. *Nat Rev Immunol.* 2002;2:933-944.
179. Szabo SJ, Sullivan BM, Peng SL, et al. Molecular mechanisms regulating Th1 immune responses. *Annu Rev Immunol.* 2003;21:713-758.
180. Ho IC, and Glimcher LH. Transcription: tantalizing times for T cells. *Cell.* 2002;109 Suppl:S109-120.

181. Smith-Garvin JE, Koretzky GA, and Jordan MS. T cell activation. Annual review of immunology. 2009;27:591-619.
182. Constant S, Pfeiffer C, Woodard A, et al. Extent of T cell receptor ligation can determine the functional differentiation of naive CD4+ T cells. J Exp Med. 1995;182:1591-1596.
183. van Panhuys N, Klauschen F, and Germain RN. T-cell-receptor-dependent signal intensity dominantly controls CD4(+) T cell polarization In Vivo. Immunity. 2014;41:63-74.
184. Sharpe AH. Mechanisms of costimulation. Immunol Rev. 2009;229:5-11.
185. Kreth HW, and Williamson AR. The extent of diversity of anti-hapten antibodies in inbred mice: anti-NIP (4-hydroxy-5-iodo-3-nitro-phenacetyl) antibodies in CBA/H mice. Eur J Immunol. 1973;3:141-146.
186. Tonegawa S. Somatic generation of antibody diversity. Nature. 1983;302:575-581.
187. Wu TT, and Kabat EA. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J Exp Med. 1970;132:211-250.
188. Cohen S, and Milstein C. Structure of antibody molecules. Nature. 1967;214:449-452 passim.
189. Zhang Z, Goldschmidt T, and Salter H. Possible allelic structure of IgG2a and IgG2c in mice. Mol Immunol. 2012;50:169-171.
190. Nimmerjahn F, and Ravetch JV. Fcγ receptors as regulators of immune responses. Nat Rev Immunol. 2008;8:34-47.

191. Ghosn EE, Sadate-Ngatchou P, Yang Y, et al. Distinct progenitors for B-1 and B-2 cells are present in adult mouse spleen. *Proc Natl Acad Sci U S A*. 2011;108:2879-2884.
192. Berland R, and Wortis HH. Origins and functions of B-1 cells with notes on the role of CD5. *Annu Rev Immunol*. 2002;20:253-300.
193. Montecino-Rodriguez E, and Dorshkind K. B-1 B cell development in the fetus and adult. *Immunity*. 2012;36:13-21.
194. Krop I, de Fougerolles AR, Hardy RR, et al. Self-renewal of B-1 lymphocytes is dependent on CD19. *Eur J Immunol*. 1996;26:238-242.
195. Stephan RP, Sanders VM, and Witte PL. Stage-specific alterations in murine B lymphopoiesis with age. *Int Immunol*. 1996;8:509-518.
196. Iwasaki H, Somoza C, Shigematsu H, et al. Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. *Blood*. 2005;106:1590-1600.
197. Bain G, Maandag EC, Izon DJ, et al. E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell*. 1994;79:885-892.
198. Kee BL, and Murre C. Induction of early B cell factor (EBF) and multiple B lineage genes by the basic helix-loop-helix transcription factor E12. *J Exp Med*. 1998;188:699-713.
199. Urbanek P, Wang ZQ, Fetka I, et al. Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell*. 1994;79:901-912.
200. Mikkola I, Heavey B, Horcher M, et al. Reversion of B cell commitment upon loss of Pax5 expression. *Science*. 2002;297:110-113.

201. Allman DM, Ferguson SE, and Cancro MP. Peripheral B cell maturation. I. Immature peripheral B cells in adults are heat-stable antigenhi and exhibit unique signaling characteristics. *J Immunol.* 1992;149:2533-2540.
202. Allman DM, Ferguson SE, Lentz VM, et al. Peripheral B cell maturation. II. Heat-stable antigen(hi) splenic B cells are an immature developmental intermediate in the production of long-lived marrow-derived B cells. *J Immunol.* 1993;151:4431-4444.
203. Carsetti R, Rosado MM, and Wardmann H. Peripheral development of B cells in mouse and man. *Immunol Rev.* 2004;197:179-191.
204. Allman D, Lindsley RC, DeMuth W, et al. Resolution of three nonproliferative immature splenic B cell subsets reveals multiple selection points during peripheral B cell maturation. *J Immunol.* 2001;167:6834-6840.
205. Cancro MP. Peripheral B-cell maturation: the intersection of selection and homeostasis. *Immunol Rev.* 2004;197:89-101.
206. Saito T, Chiba S, Ichikawa M, et al. Notch2 is preferentially expressed in mature B cells and indispensable for marginal zone B lineage development. *Immunity.* 2003;18:675-685.
207. Hozumi K, Negishi N, Suzuki D, et al. Delta-like 1 is necessary for the generation of marginal zone B cells but not T cells in vivo. *Nat Immunol.* 2004;5:638-644.
208. Goodnow CC, Crosbie J, Adelstein S, et al. Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice. *Nature.* 1988;334:676-682.
209. Nemazee DA, and Burki K. Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes. *Nature.* 1989;337:562-566.

210. Hartley SB, Crosbie J, Brink R, et al. Elimination from peripheral lymphoid tissues of self-reactive B lymphocytes recognizing membrane-bound antigens. *Nature*. 1991;353:765-769.
211. Nossal GJ, and Pike BL. Evidence for the clonal abortion theory of B-lymphocyte tolerance. *J Exp Med*. 1975;141:904-917.
212. Gu H, Tarlinton D, Muller W, et al. Most peripheral B cells in mice are ligand selected. *J Exp Med*. 1991;173:1357-1371.
213. Lam KP, Kuhn R, and Rajewsky K. In vivo ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death. *Cell*. 1997;90:1073-1083.
214. Levine MH, Haberman AM, Sant'Angelo DB, et al. A B-cell receptor-specific selection step governs immature to mature B cell differentiation. *Proc Natl Acad Sci U S A*. 2000;97:2743-2748.
215. Cancro MP, and Kearney JF. B cell positive selection: road map to the primary repertoire? *J Immunol*. 2004;173:15-19.
216. Yurasov S, Wardemann H, Hammersen J, et al. Defective B cell tolerance checkpoints in systemic lupus erythematosus. *J Exp Med*. 2005;201:703-711.
217. Hondowicz BD, Alexander ST, Quinn WJ, 3rd, et al. The role of BLYS/BLYS receptors in anti-chromatin B cell regulation. *Int Immunol*. 2007;19:465-475.
218. Carsetti R, Kohler G, and Lamers MC. Transitional B cells are the target of negative selection in the B cell compartment. *J Exp Med*. 1995;181:2129-2140.
219. Wardemann H, Yurasov S, Schaefer A, et al. Predominant autoantibody production by early human B cell precursors. *Science*. 2003;301:1374-1377.
220. Dal Porto JM, Gauld SB, Merrell KT, et al. B cell antigen receptor signaling 101. *Mol Immunol*. 2004;41:599-613.

221. Pasare C, and Medzhitov R. Control of B-cell responses by Toll-like receptors. *Nature*. 2005;438:364-368.
222. Bretscher P, and Cohn M. A theory of self-nonself discrimination. *Science*. 1970;169:1042-1049.
223. Mosier DE, and Subbarao B. Thymus-independent antigens: complexity of B-lymphocyte activation revealed. *Immunol Today*. 1982;3:217-222.
224. Mond JJ, Lees A, and Snapper CM. T cell-independent antigens type 2. *Annu Rev Immunol*. 1995;13:655-692.
225. Martin F, Oliver AM, and Kearney JF. Marginal zone and B1 B cells unite in the early response against T-independent blood-borne particulate antigens. *Immunity*. 2001;14:617-629.
226. Garcia de Vinuesa C, O'Leary P, Sze DM, et al. T-independent type 2 antigens induce B cell proliferation in multiple splenic sites, but exponential growth is confined to extrafollicular foci. *Eur J Immunol*. 1999;29:1314-1323.
227. de Vinuesa CG, Cook MC, Ball J, et al. Germinal centers without T cells. *J Exp Med*. 2000;191:485-494.
228. McHeyzer-Williams LJ, and McHeyzer-Williams MG. Antigen-specific memory B cell development. *Annu Rev Immunol*. 2005;23:487-513.
229. Eisen HN, and Siskind GW. VARIATIONS IN AFFINITIES OF ANTIBODIES DURING THE IMMUNE RESPONSE. *Biochemistry*. 1964;3:996-1008.
230. Kocks C, and Rajewsky K. Stable expression and somatic hypermutation of antibody V regions in B-cell developmental pathways. *Annu Rev Immunol*. 1989;7:537-559.
231. Inamine A, Takahashi Y, Baba N, et al. Two waves of memory B-cell generation in the primary immune response. *Int Immunol*. 2005;17:581-589.

232. Blink EJ, Light A, Kallies A, et al. Early appearance of germinal center-derived memory B cells and plasma cells in blood after primary immunization. *J Exp Med*. 2005;201:545-554.
233. Hosokawa T. Studies on B-cell memory. II. T-cell independent antigen can induce B-cell memory. *Immunology*. 1979;38:291-299.
234. Nieuwenhuis P, and Opstelten D. Functional anatomy of germinal centers. *The American journal of anatomy*. 1984;170:421-435.
235. Jacob J, Przylepa J, Miller C, et al. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. III. The kinetics of V region mutation and selection in germinal center B cells. *J Exp Med*. 1993;178:1293-1307.
236. Jacob J, Kelsoe G, Rajewsky K, et al. Intraclonal generation of antibody mutants in germinal centres. *Nature*. 1991;354:389-392.
237. Berek C, Berger A, and Apel M. Maturation of the immune response in germinal centers. *Cell*. 1991;67:1121-1129.
238. Takahashi Y, Dutta PR, Cerasoli DM, et al. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. V. Affinity maturation develops in two stages of clonal selection. *J Exp Med*. 1998;187:885-895.
239. Goenka R, Barnett LG, Silver JS, et al. Cutting edge: dendritic cell-restricted antigen presentation initiates the follicular helper T cell program but cannot complete ultimate effector differentiation. *Journal of immunology*. 2011;187:1091-1095.
240. Victora GD, and Nussenzweig MC. Germinal centers. *Annu Rev Immunol*. 2012;30:429-457.
241. Allman D, Jain A, Dent A, et al. BCL-6 expression during B-cell activation. *Blood*. 1996;87:5257-5268.

242. Dent AL, Shaffer AL, Yu X, et al. Control of inflammation, cytokine expression, and germinal center formation by BCL-6. *Science*. 1997;276:589-592.
243. Shaffer AL, Yu X, He Y, et al. BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. *Immunity*. 2000;13:199-212.
244. Basso K, and Dalla-Favera R. BCL6: master regulator of the germinal center reaction and key oncogene in B cell lymphomagenesis. *Adv Immunol*. 2010;105:193-210.
245. Muramatsu M, Kinoshita K, Fagarasan S, et al. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*. 2000;102:553-563.
246. Pavri R, Gazumyan A, Jankovic M, et al. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell*. 2010;143:122-133.
247. Pavri R, and Nussenzweig MC. AID targeting in antibody diversity. *Adv Immunol*. 2011;110:1-26.
248. Zotos D, and Tarlinton DM. Determining germinal centre B cell fate. *Trends Immunol*. 2012;33:281-288.
249. Cattoretti G, Buttner M, Shaknovich R, et al. Nuclear and cytoplasmic AID in extrafollicular and germinal center B cells. *Blood*. 2006;107:3967-3975.
250. Di Niro R, Lee SJ, Vander Heiden JA, et al. Salmonella Infection Drives Promiscuous B Cell Activation Followed by Extrafollicular Affinity Maturation. *Immunity*. 2015;43:120-131.
251. Gitlin AD, Shulman Z, and Nussenzweig MC. Clonal selection in the germinal centre by regulated proliferation and hypermutation. *Nature*. 2014;509:637-640.

252. Chtanova T, Tangye SG, Newton R, et al. T follicular helper cells express a distinctive transcriptional profile, reflecting their role as non-Th1/Th2 effector cells that provide help for B cells. *J Immunol.* 2004;173:68-78.
253. Ansel KM, Ngo VN, Hyman PL, et al. A chemokine-driven positive feedback loop organizes lymphoid follicles. *Nature.* 2000;406:309-314.
254. Manz RA, Thiel A, and Radbruch A. Lifetime of plasma cells in the bone marrow. *Nature.* 1997;388:133-134.
255. Amanna IJ, Carlson NE, and Slifka MK. Duration of humoral immunity to common viral and vaccine antigens. *N Engl J Med.* 2007;357:1903-1915.
256. Slifka MK, Antia R, Whitmire JK, et al. Humoral immunity due to long-lived plasma cells. *Immunity.* 1998;8:363-372.
257. Shapiro-Shelef M, Lin KI, McHeyzer-Williams LJ, et al. Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity.* 2003;19:607-620.
258. Kallies A, Hasbold J, Fairfax K, et al. Initiation of plasma-cell differentiation is independent of the transcription factor Blimp-1. *Immunity.* 2007;26:555-566.
259. Lin KI, Angelin-Duclos C, Kuo TC, et al. Blimp-1-dependent repression of Pax-5 is required for differentiation of B cells to immunoglobulin M-secreting plasma cells. *Mol Cell Biol.* 2002;22:4771-4780.
260. Shaffer AL, Lin KI, Kuo TC, et al. Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity.* 2002;17:51-62.
261. Bortnick A, Chernova I, Quinn WJ, 3rd, et al. Long-lived bone marrow plasma cells are induced early in response to T cell-independent or T cell-dependent antigens. *J Immunol.* 2012;188:5389-5396.

262. Chernova I, Jones DD, Wilmore JR, et al. Lasting antibody responses are mediated by a combination of newly formed and established bone marrow plasma cells drawn from clonally distinct precursors. *J Immunol.* 2014;193:4971-4979.
263. Fritz JH, Rojas OL, Simard N, et al. Acquisition of a multifunctional IgA<sup>+</sup> plasma cell phenotype in the gut. *Nature.* 2012;481:199-203.
264. Radbruch A, Muehlinghaus G, Luger EO, et al. Competence and competition: the challenge of becoming a long-lived plasma cell. *Nat Rev Immunol.* 2006;6:741-750.
265. Mueller SN, Gebhardt T, Carbone FR, et al. Memory T cell subsets, migration patterns, and tissue residence. *Annu Rev Immunol.* 2013;31:137-161.
266. Sallusto F, Lenig D, Forster R, et al. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature.* 1999;401:708-712.
267. Tomayko MM, Steinel NC, Anderson SM, et al. Cutting edge: Hierarchy of maturity of murine memory B cell subsets. *J Immunol.* 2010;185:7146-7150.
268. Zuccarino-Catania GV, Sadanand S, Weisel FJ, et al. CD80 and PD-L2 define functionally distinct memory B cell subsets that are independent of antibody isotype. *Nat Immunol.* 2014;15:631-637.
269. Best JA, Blair DA, Knell J, et al. Transcriptional insights into the CD8(+) T cell response to infection and memory T cell formation. *Nat Immunol.* 2013;14:404-412.
270. Hu G, and Chen J. A genome-wide regulatory network identifies key transcription factors for memory CD8(+) T-cell development. *Nat Commun.* 2013;4:2830.
271. Hao Y, O'Neill PJ, Naradikian MS, et al. A B-cell subset uniquely responsive to innate stimuli accumulates in aged mice. *Blood.* 2011.

272. Rubtsov AV, Rubtsova K, Fischer A, et al. Toll-like receptor 7 (TLR7)-driven accumulation of a novel CD11c(+) B-cell population is important for the development of autoimmunity. *Blood*. 2011;118:1305-1315.
273. Russell Knode LM, Naradikian MS, Myles A, et al. Age-Associated B Cells Express a Diverse Repertoire of VH and Vkappa Genes with Somatic Hypermutation. *J Immunol*. 2017;198:1921-1927.
274. Naradikian MS, Myles A, Beiting DP, et al. Cutting Edge: IL-4, IL-21, and IFN-gamma Interact To Govern T-bet and CD11c Expression in TLR-Activated B Cells. *J Immunol*. 2016;197:1023-1028.
275. Zhu J, Jankovic D, Oler AJ, et al. The transcription factor T-bet is induced by multiple pathways and prevents an endogenous Th2 cell program during Th1 cell responses. *Immunity*. 2012;37:660-673.
276. Whittle JR, Wheatley AK, Wu L, et al. Flow cytometry reveals that H5N1 vaccination elicits cross-reactive stem-directed antibodies from multiple Ig heavy-chain lineages. *Journal of virology*. 2014;88:4047-4057.
277. Cancro MP, Gerhard W, and Klinman NR. The diversity of the influenza-specific primary B-cell repertoire in BALB/c mice. *J Exp Med*. 1978;147:776-787.
278. Frank GM, Angeletti D, Ince WL, et al. A Simple Flow-Cytometric Method Measuring B Cell Surface Immunoglobulin Avidity Enables Characterization of Affinity Maturation to Influenza A Virus. *mBio*. 2015;6:e01156.
279. Boyden AW, Legge KL, and Waldschmidt TJ. Pulmonary infection with influenza A virus induces site-specific germinal center and T follicular helper cell responses. *PloS one*. 2012;7:e40733.

280. Weisel FJ, Zuccarino-Catania GV, Chikina M, et al. A Temporal Switch in the Germinal Center Determines Differential Output of Memory B and Plasma Cells. *Immunity*. 2016;44:116-130.
281. Taylor JJ, Pape KA, and Jenkins MK. A germinal center-independent pathway generates unswitched memory B cells early in the primary response. *J Exp Med*. 2012;209:597-606.
282. Shinall SM, Gonzalez-Fernandez M, Noelle RJ, et al. Identification of murine germinal center B cell subsets defined by the expression of surface isotypes and differentiation antigens. *J Immunol*. 2000;164:5729-5738.
283. Sheikh AA, Cooper L, Feng M, et al. Context-Dependent Role for T-bet in T Follicular Helper Differentiation and Germinal Center Function following Viral Infection. *Cell Rep*. 2019;28:1758-1772 e1754.
284. Obeng-Adjei N, Portugal S, Holla P, et al. Malaria-induced interferon-gamma drives the expansion of Tbethi atypical memory B cells. *PLoS Pathog*. 2017;13:e1006576.
285. Wang S, Wang J, Kumar V, et al. IL-21 drives expansion and plasma cell differentiation of autoreactive CD11c(hi)T-bet(+) B cells in SLE. *Nat Commun*. 2018;9:1758.
286. Chang LY, Li Y, and Kaplan DE. Hepatitis C viraemia reversibly maintains subset of antigen-specific T-bet+ tissue-like memory B cells. *Journal of viral hepatitis*. 2017;24:389-396.
287. Lau D, Lan LY-L, Andrews SF, et al. Low CD21 expression defines a population of recent germinal center graduates primed for plasma cell differentiation. *Sci Immunol*. 2017;2:eaai8153.

288. Knox JJ, Buggert M, Kardava L, et al. T-bet<sup>+</sup> B cells are induced by human viral infections and dominate the HIV gp140 response. *JCI insight*. 2017;2.
289. Joyce MG, Wheatley AK, Thomas PV, et al. Vaccine-Induced Antibodies that Neutralize Group 1 and Group 2 Influenza A Viruses. *Cell*. 2016;166:609-623.
290. Bell P, Vandenberghe LH, Wu D, et al. A comparative analysis of novel fluorescent proteins as reporters for gene transfer studies. *J Histochem Cytochem*. 2007;55:931-939.
291. Allie SR, Bradley JE, Mudunuru U, et al. The establishment of resident memory B cells in the lung requires local antigen encounter. *Nat Immunol*. 2019;20:97-108.
292. Pardi N, Hogan MJ, Naradikian MS, et al. Nucleoside-modified mRNA vaccines induce potent T follicular helper and germinal center B cell responses. *J Exp Med*. 2018;215:1571-1588.
293. Yu F, Sharma S, Edwards J, et al. Dynamic expression of transcription factors T-bet and GATA-3 by regulatory T cells maintains immunotolerance. *Nat Immunol*. 2015;16:197-206.
294. Gerth AJ, Lin L, and Peng SL. T-bet regulates T-independent IgG2a class switching. *Int Immunol*. 2003;15:937-944.
295. Liu N, Ohnishi N, Ni L, et al. CpG directly induces T-bet expression and inhibits IgG1 and IgE switching in B cells. *Nat Immunol*. 2003;4:687-693.
296. Peng SL, Szabo SJ, and Glimcher LH. T-bet regulates IgG class switching and pathogenic autoantibody production. *Proc Natl Acad Sci U S A*. 2002;99:5545-5550.
297. Wang NS, McHeyzer-Williams LJ, Okitsu SL, et al. Divergent transcriptional programming of class-specific B cell memory by T-bet and ROR $\alpha$ . *Nat Immunol*. 2012;13:604-611.

298. Coutelier JP, van der Logt JT, Heessen FW, et al. IgG2a restriction of murine antibodies elicited by viral infections. *J Exp Med*. 1987;165:64-69.
299. Hocart MJ, Mackenzie JS, and Stewart GA. The IgG subclass responses to influenza virus haemagglutinin in the mouse: effect of route of inoculation. *J Gen Virol*. 1989;70 ( Pt 4):809-818.
300. Markine-Goriaynoff D, and Coutelier JP. Increased efficacy of the immunoglobulin G2a subclass in antibody-mediated protection against lactate dehydrogenase-elevating virus-induced polioencephalomyelitis revealed with switch mutants. *Journal of virology*. 2002;76:432-435.
301. Rubtsova K, Rubtsov AV, van Dyk LF, et al. T-box transcription factor T-bet, a key player in a unique type of B-cell activation essential for effective viral clearance. *Proc Natl Acad Sci U S A*. 2013;110:E3216-3224.
302. Fazekas G, Rosenwirth B, Dukor P, et al. IgG isotype distribution of local and systemic immune responses induced by influenza virus infection. *Eur J Immunol*. 1994;24:3063-3067.
303. Piovesan D, Tempany J, Di Pietro A, et al. c-Myb Regulates the T-Bet-Dependent Differentiation Program in B Cells to Coordinate Antibody Responses. *Cell reports*. 2017;19:461-470.
304. de Jong JC, Palache AM, Beyer WE, et al. Haemagglutination-inhibiting antibody to influenza virus. *Dev Biol (Basel)*. 2003;115:63-73.
305. Hobson D, Curry RL, Beare AS, et al. The role of serum haemagglutination-inhibiting antibody in protection against challenge infection with influenza A2 and B viruses. *J Hyg (Lond)*. 1972;70:767-777.

306. DiLillo DJ, Tan GS, Palese P, et al. Broadly neutralizing hemagglutinin stalk-specific antibodies require FcγR interactions for protection against influenza virus in vivo. *Nat Med.* 2014;20:143-151.
307. DiLillo DJ, Palese P, Wilson PC, et al. Broadly neutralizing anti-influenza antibodies require Fc receptor engagement for in vivo protection. *The Journal of clinical investigation.* 2016;126:605-610.
308. Antibodies cross-reactive to influenza A (H3N2) variant virus and impact of 2010-11 seasonal influenza vaccine on cross-reactive antibodies - United States. *MMWR. Morbidity and mortality weekly report.* 2012;61:237-241.
309. Pica N, Hai R, Krammer F, et al. Hemagglutinin stalk antibodies elicited by the 2009 pandemic influenza virus as a mechanism for the extinction of seasonal H1N1 viruses. *Proc Natl Acad Sci U S A.* 2012;109:2573-2578.
310. Krammer F, Margine I, Tan GS, et al. A carboxy-terminal trimerization domain stabilizes conformational epitopes on the stalk domain of soluble recombinant hemagglutinin substrates. *PloS one.* 2012;7:e43603.
311. Kenderes KJ, Levack RC, Papillion AM, et al. T-Bet(+) IgM Memory Cells Generate Multi-lineage Effector B Cells. *Cell reports.* 2018;24:824-837.e823.
312. Barnett BE, Staupé RP, Odorizzi PM, et al. Cutting Edge: B Cell-Intrinsic T-bet Expression Is Required To Control Chronic Viral Infection. *J Immunol.* 2016;197:1017-1022.
313. Isnardi I, Ng YS, Menard L, et al. Complement receptor 2/CD21- human naive B cells contain mostly autoreactive unresponsive clones. *Blood.* 2010;115:5026-5036.

314. Andrews SF, Chambers MJ, Schramm CA, et al. Activation Dynamics and Immunoglobulin Evolution of Pre-existing and Newly Generated Human Memory B cell Responses to Influenza Hemagglutinin. *Immunity*. 2019;51:398-410 e395.
315. Johrens K, Anagnostopoulos I, Durkop H, et al. Different T-bet expression patterns characterize particular reactive lymphoid tissue lesions. *Histopathology*. 2006;48:343-352.
316. Austin JW, Buckner CM, Kardava L, et al. Overexpression of T-bet in HIV infection is associated with accumulation of B cells outside germinal centers and poor affinity maturation. *Science translational medicine*. 2019;11.
317. Turner DL, Bickham KL, Farber DL, et al. Splenic priming of virus-specific CD8 T cells following influenza virus infection. *Journal of virology*. 2013;87:4496-4506.
318. Jenks SA, Cashman KS, Zumaquero E, et al. Distinct Effector B Cells Induced by Unregulated Toll-like Receptor 7 Contribute to Pathogenic Responses in Systemic Lupus Erythematosus. *Immunity*. 2018;49:725-739.e726.
319. Koutsakos M, Wheatley AK, Loh L, et al. Circulating T<sub>FH</sub> cells, serological memory, and tissue compartmentalization shape human influenza-specific B cell immunity. *Science Translational Medicine*. 2018;10.
320. Kamran P, Sereti KI, Zhao P, et al. Parabiosis in mice: a detailed protocol. *Journal of visualized experiments : JoVE*. 2013.
321. Margine I, Palese P, and Krammer F. Expression of functional recombinant hemagglutinin and neuraminidase proteins from the novel H7N9 influenza virus using the baculovirus expression system. *Journal of visualized experiments : JoVE*. 2013:e51112.

322. Baierdorfer M, Boros G, Muramatsu H, et al. A Facile Method for the Removal of dsRNA Contaminant from In Vitro-Transcribed mRNA. *Mol Ther Nucleic Acids*. 2019;15:26-35.
323. Maier MA, Jayaraman M, Matsuda S, et al. Biodegradable lipids enabling rapidly eliminated lipid nanoparticles for systemic delivery of RNAi therapeutics. *Mol Ther*. 2013;21:1570-1578.
324. Wang Z, Raifu M, Howard M, et al. Universal PCR amplification of mouse immunoglobulin gene variable regions: the design of degenerate primers and an assessment of the effect of DNA polymerase 3' to 5' exonuclease activity. *Journal of Immunological Methods*. 2000;233:167-177.
325. Meng W, Zhang B, Schwartz GW, et al. An atlas of B-cell clonal distribution in the human body. *Nat Biotechnol*. 2017;35:879-884.
326. Rosenfeld AM, Meng W, Chen DY, et al. Computational Evaluation of B-Cell Clone Sizes in Bulk Populations. *Front Immunol*. 2018;9:1472.
327. Vander Heiden JA, Yaari G, Uduman M, et al. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics*. 2014;30:1930-1932.
328. Rosenfeld AM, Meng W, Luning Prak ET, et al. ImmuneDB, a Novel Tool for the Analysis, Storage, and Dissemination of Immune Repertoire Sequencing Data. *Front Immunol*. 2018;9:2107.
329. Perino M, and Veenstra GJ. Chromatin Control of Developmental Dynamics and Plasticity. *Dev Cell*. 2016;38:610-620.
330. Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*. 2016;48:1193-1203.

331. Groudine M, and Weintraub H. Activation of globin genes during chicken development. *Cell*. 1981;24:393-401.
332. Fujiwara T, O'Geen H, Keles S, et al. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell*. 2009;36:667-681.
333. Decker T, Pasca di Magliano M, McManus S, et al. Stepwise activation of enhancer and promoter regions of the B cell commitment gene Pax5 in early lymphopoiesis. *Immunity*. 2009;30:508-520.
334. Ptashne M. Epigenetics: core misconception. *Proc Natl Acad Sci U S A*. 2013;110:7101-7103.
335. Henikoff S, and Shilatifard A. Histone modification: cause or cog? *Trends Genet*. 2011;27:389-396.
336. Han M, and Grunstein M. Nucleosome loss activates yeast downstream promoters in vivo. *Cell*. 1988;55:1137-1145.
337. Wang X, Bai L, Bryant GO, et al. Nucleosomes and the accessibility problem. *Trends Genet*. 2011;27:487-492.
338. Thurman RE, Rynes E, Humbert R, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012;489:75-82.
339. Burton A, and Torres-Padilla ME. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat Rev Mol Cell Biol*. 2014;15:723-734.
340. Cirillo LA, and Zaret KS. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell*. 1999;4:961-969.

341. Cirillo LA, Lin FR, Cuesta I, et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell*. 2002;9:279-289.
342. Xu Y, Banerjee D, Huelsken J, et al. Deletion of beta-catenin impairs T cell development. *Nat Immunol*. 2003;4:1177-1182.
343. Xu Z, Xing S, Shan Q, et al. Cutting Edge: beta-Catenin-Interacting Tcf1 Isoforms Are Essential for Thymocyte Survival but Dispensable for Thymic Maturation Transitions. *J Immunol*. 2017;198:3404-3409.
344. Lupien M, Eeckhoute J, Meyer CA, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*. 2008;132:958-970.
345. Hu G, Cui K, Fang D, et al. Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. *Immunity*. 2018;48:227-242.e228.
346. Strutt TM, McKinstry KK, Marshall NB, et al. Multipronged CD4(+) T-cell effector and memory responses cooperate to provide potent immunity against respiratory virus. *Immunol Rev*. 2013;255:149-164.
347. Naradikian MS, Hao Y, and Cancro MP. Age-associated B cells: key mediators of both protective and autoreactive humoral responses. *Immunol Rev*. 2016;269:118-129.
348. Sanz I, Wei C, Jenks SA, et al. Challenges and Opportunities for Consistent Classification of Human B Cell and Plasma Cell Populations. *Front Immunol*. 2019;10:2458.
349. Arpin C, Banchereau J, and Liu YJ. Memory B cells are biased towards terminal differentiation: a strategy that may prevent repertoire freezing. *J Exp Med*. 1997;186:931-940.

350. Kurosaki T, Kometani K, and Ise W. Memory B cells. *Nat Rev Immunol.* 2015;15:149-159.
351. Kometani K, Nakagawa R, Shinnakasu R, et al. Repression of the transcription factor Bach2 contributes to predisposition of IgG1 memory B cells toward plasma cell differentiation. *Immunity.* 2013;39:136-147.
352. Mesin L, Schiepers A, Ersching J, et al. Restricted Clonality and Limited Germinal Center Reentry Characterize Memory B Cell Reactivation by Boosting. *Cell.* 2020;180:92-106 e111.
353. Anolik JH, Barnard J, Cappione A, et al. Rituximab improves peripheral B cell abnormalities in human systemic lupus erythematosus. *Arthritis Rheum.* 2004;50:3580-3590.
354. Fazekas de St G, and Webster RG. Disquisitions on Original Antigenic Sin. II. Proof in lower creatures. *J Exp Med.* 1966;124:347-361.
355. DiLillo DJ, Hamaguchi Y, Ueda Y, et al. Maintenance of long-lived plasma cells and serological memory despite mature and memory B cell depletion during CD20 immunotherapy in mice. *J Immunol.* 2008;180:361-371.
356. Purtha WE, Tedder TF, Johnson S, et al. Memory B cells, but not long-lived plasma cells, possess antigen specificities for viral escape mutants. *J Exp Med.* 2011;208:2599-2606.
357. Schenkel JM, and Masopust D. Tissue-resident memory T cells. *Immunity.* 2014;41:886-897.
358. Trivedi N, Weisel F, Smita S, et al. Liver Is a Generative Site for the B Cell Response to *Ehrlichia muris*. *Immunity.* 2019;51:1088-1101.e1085.

359. Birjandi SZ, Ippolito JA, Ramadorai AK, et al. Alterations in marginal zone macrophages and marginal zone B cells in old mice. *J Immunol.* 2011;186:3441-3451.
360. Turner VM, and Mabbott NA. Ageing adversely affects the migration and function of marginal zone B cells. *Immunology.* 2017;151:349-362.
361. Szabo SJ, Kim ST, Costa GL, et al. A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell.* 2000;100:655-669.
362. Stone SL, Peel JN, Scharer CD, et al. T-bet Transcription Factor Promotes Antibody-Secreting Cell Differentiation by Limiting the Inflammatory Effects of IFN-gamma on B Cells. *Immunity.* 2019;50:1172-1187 e1177.