Pre-programmed regulatory networks and poised chromatin potentiate lineage-specific differentiation of CD4+ T cells

Antony Christopher Raubitschek

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Steven F. Ziegler, Chair

John Stamatoyannopoulos

Daniel J. Campbell

Program Authorized to Offer Degree:

Department of Immunology

University of Washington

Abstract

Pre-programmed regulatory networks and poised chromatin potentiate lineage-specific differentiation of CD4+ T cells

Antony Christopher Raubitschek

Chair of the Supervisory Committee:
Affiliate Professor Steven F. Ziegler
Department of Immunology

How the interactions of transcription factors (TFs) with specific regions of chromatin control the basis for cellular function and identity remains an open question. Here, analysis of regions of accessible chromatin, which are identified globally across T cell subtypes, demonstrates that the chromatin of naïve CD4 T cells is poised towards the acquisition of effector function. Additionally, the development of genome-wide maps of *in vivo* transcription factor footprints allows the assembly of extensive regulatory networks comprising connections among 476 sequence-specific transcription factors, demonstrating the dynamics of these connections across mature human CD4 T cell subtypes. Well-characterized subset-specific TFs show prominent occupancy in subset-specific networks. Moreover, conserved regulatory architectures act through these

subset-specific regulatory factors, to potentiate, rather than direct, mature CD4 T cell differentiation. These results provide a model that explains seemingly conflicting experimental evidence of T cell subset stability and plasticity.

Table of Contents

List of Figures

Acknowledgements

The author wishes to express sincere appreciation to Dr. Steven F. Ziegler for his excellent mentorship throughout graduate school and charitable contribution to providing homes for wayward graduate students. In addition, thanks go to Dr. John Stamatoyannopoulos for providing extensive support and guidance and to Shane Neph, Richard Sandstrom, and Andrew Stergachis for technical and intellectual contributions to Chapter 3. The author must also acknowledge the rest of the Ziegler lab and the Stamatoyannopoulos lab for creating a productive and enriching learning environment.

Dedication

This dissertation is dedicated to my father Andrew Raubitschek, my first scientific mentor and most trusted advisor, and to my mother Joanne Watchie and wife Dannica Lin whose unwavering support and love have made this work possible.

Chapter 1: Introduction

Project Overview

The vertebrate immune system is composed of two arms: an innate system and an adaptive system [1]. The innate immune system identifies molecular patterns associated with pathogens via germline-encoded receptors called pattern recognition receptors [2]. The adaptive immune system identifies pathogens via the unique antigen receptors of B and T cells, and these receptors are generated via random gene rearrangement [3, 4].

The adaptive immune system must restrict the antigen specificity of the B cell rectors (BCRs) and T cell receptors (TCRs) to prevent self-reactivity while maintaining the ability to react to non-self antigen. The primary mechanism by which T cells control reactivity is via central tolerance, a process by which developing T cells undergo negative selective in the thymus [5, 6].

As an immature T cell undergoes selection, the TCR engages self-peptide containing major histocompatibility complex (MHC). The expression of a functional TCR protein capable of interacting with is necessary for the continued development of the thymocyte in a process called positive selections, and during this stage T cells are selected to be either CD8 or CD4 positive depending on the ability of the TCR to recognize peptide presented on the surface of MHC class I or class II proteins, respectively [5, 7, 8]. T cells expressing TCRs that react too

strongly are then deleted during negative selection to prevent the development of autoreactivity [9].

After undergoing positive and negative selection, now mature CD4 and CD8 T cells exit the thymus. Upon encountering their cognate antigen in the environment these cells integrate multiple signals and then generate appropriate immune responses.

If these cells receive TCRs in the absence of additional inflammatory signals, the cells may be peripherally deleted or become anergic.[10, 11]. In the presence of proper inflammatory signals and coreceptors CD8 T cells become cytotoxic T lymphocytes that demonstrate cytotoxic function, whereby they directly trigger the destruction of cells presenting foreign peptides consistent with infection with intracellular pathogens.[12] CD4 T cells possess more diversity in response, and depending on the specific signals received, can promote diverse immune responses that appropriately combat different classes of pathogens. These signals are provided primarily in the form on cytokine signaling from innate cells, which secreted differentially depending on the characteristic of the pathogen present. These cytokine signals than activate different CD4 T cell programs, each program being optimized to combat a specific class of pathogens.

After pathogen clearance a subset of CD4 T cell called memory T cells persist in both in secondary lymphoid tissue and in peripheral tissue. These T cells form the basis for cellular memory, and mediate rapid immune response on subsequent pathogen encounters. Remarkably, the memory T cells appear to reengage appropriate effector programs rapidly on pathogen reencounter,

suggesting that these cells have differentiated to a specific lineage of CD4 T cell. The study of how CD4 T cell rapidly reengage an advantageous immune response provides insight into an important mechanism of adaptive immune control, and additionally may reveal insights into how cellular lineage differentiation is generally regulated

Control of adaptive immunity by CD4 T cell subsets

The ability of the immune system to remember pathogenic challenges, preventing further illness from re-exposure to identical or similar pathogens has been appreciated for over two hundred years [13]. This ability is the primary function of the adaptive immune system, which is critically controlled by CD4 T cells. The requirement for these cells is clearly demonstrated in patients infected with HIV, who develop AIDS when CD4 T cells become depleted.

CD4 T cells are not a homogenous population of cells, but rather develop along defined programs to one of a multiple of lineage fates upon encountering cognate antigen. Originally two subsets of CD4 T cells were described from the analysis of mouse CD4 T cell clones, and were called Th1 and Th2 cells respectively [14, 15]. Th1 cells characteristically produce the cytokine IFNγ, as well as high amounts of IL-2 and coordinate immune responses to intracellular pathogens, while Th2 cells produce IL-4, IL-5, and IL-13 and coordinate immune responses to multicellular extracellular pathogens such as helminthes [16].

The stimulation of naïve CD4 T cells *in vitro* (with polyclonal stimulation such as anti-CD3 and anti-CD28) along with appropriate cytokines (IL-12 for Th1 cells, IL-4 for Th2 cells) leads to the differentiation into Th1 [17] and Th2 cells [18,

19]. Despite initial controversy over whether simulating cytokines lead individual naïve CD4 T cells to adopt the Th1 or Th2 fates or whether these cytokines function to support selective outgrowth of differentiated cells, it is clear from experiments in which single naïve CD4 T cells were primed *in vitro* that individual CD4 T cells can indeed be made to directly differentiate into either of these T helper fates [20-23]. Thus, the differentiation of CD4 T cells provided a readily accessible model to study lineage differentiation in primary human cells.

More recently the ability of CD4 T cells to develop along additional differentiation programs has been appreciated. Th17 cells, characterized by the production of IL-17A, IL-17F and IL-22 [24, 25], were shown to be derived from naïve CD4 T cells [26, 27]. These cells coordinate immune responses to extracellular bacteria, particularly at mucosal surfaces.

Another population of CD4 T cells, called Tregs, was shown to be critical for the preservation of immune tolerance. Importantly, two populations of Tregs have been described, one arising from the thymus (nTreg), and a second, iTreg, arising from naïve CD4 T cells in the periphery [28, 29] as in the differentiation of Th1, Th2, or Th17 cells. These iTregs possess similar suppressive capacities to nTregs [30]. With increasing numbers of potential CD4 phenotypes a naïve T cell may differentiate to, a more complex view of lineage differentiation is necessary.

Control of lineage differentiation by “master” regulatory factors

Lineage differentiation has long been hypothesized to be supported by the differential expression of key genes [31]. In CD4 T cells each described lineage has been associated with a “master” regulatory factor, which together are indispensable for lineage commitment and directly mediate many of the subset-specific functions of CD4 T cells. Commitment of naïve CD4+ T cells to the Th1, Th2, Th17, and Treg lineages is dependent on lineage-specific expression of T-bet, GATA3, RORγt, and FOXP3, respectively [32].

T-bet, a protein encoded by the *TBX21* gene, is required for the generation of Th1 cells, as Tbx21-/- cells show severe defects in Th1 cell differentiation both *in vitro* and *in vivo* [33, 34]. Overexpression of T-bet in differentiating T cells or even in fully differentiated Th2 cells results in the acquisition of IFN-g production and suppression of IL-4 production by these cells [35]. Importantly, T-bet can directly inhibit GATA3 expression [36].

While GATA3 is expressed by naïve CD4 T cells, and is an important factor in the differentiation of immature thymocytes into CD4 T cells [37], GATA3 is additionally required for the generation of Th2 cells [38]. GATA3 is further unregulated or downregulated during Th2 or Th1, respectively [39, 40]. The depletion of GATA3 in peripheral T cells results in the complete loss of Th2 differentiation *in vitro* and *in vivo* [41, 42], while over expression in Th1 cells induces production of IL-4 [40]. Furthermore ectopic expression of GATA3 induces further endogenous expression [43].

The indispensable roles of T-bet and GATA3 in the generation of Th1 and Th2 cells respectively, combined with the ability of each factor to regulate GATA3 expression in order to direct cells away from alternative lineages lead to the view of these factors as the master regulators of Th1 and Th2 differentiation, respectively. Additional support from this view comes from experiments in model cytokine loci which show that these factors are directly responsible for chromatin remodeling [35, 43-45]. However, the expression of these factors is not entirely restricted to their corresponding subsets, emphasizing the need for sophisticated modeling of dose dependent effects to explain their roles in CD4 T cell differentiation.

Peripheral expression of RORγt, in contrast appears more restricted to Th17 cells [46], although it is also expressed in double positive thymocytes [47]. This factor and the related family member RORα are required for the generation of Th17 cells [46, 48]. Whether RORγt expression can enforce commitment to the Th17 lineage, preventing acquisition of Th1 or Th2 cytokine production is not well-established.

FoxP3 is required both for the development of Tregs [49, 50]. Continuous expression of FoxP3 is required to maintain the suppressive activity of nTregs [51], and ectopic expression of FoxP3 converts conventional T cells to a Treg-like phenotype [52].

Lineage stability and plasticity

Classical models of lineage differentiation involve the irreversible commitment of multipotent precursors into differentiated lineages [31]. However, the model of committed lineage differentiation of CD4 T cells has recently been challenged by report of instability within these lineages. For example, Th17 cells readily acquire expression of IFNγ, a hallmark of Th1 cells, and can subsequently lose expression of their own signature cytokine, IL-17 [53], suggesting conversion of Th17 cells to Th1 cells. Treg cells with lowered FOXP3 expression and acquisition of effector functions have also been observed [54, 55]. In addition, unexpectedly diverse phenotypes in human CD4 T cells have been observed in *vivo*, including cells coexpressing IFNγ and IL-17 [56], cells coexpressing IFNγ and the Th2 characteristic cytokine IL4 [57], and cells coexpressing RORγt and Tbet [58], GATA3 and Tbet [59], FOXP3 and T-bet [60], and FOXP3 and RORγt [61]. The existence of T cell subsets co-expressing multiple lineage regulators highlights the importance of understanding the interactions of these factors with the larger network of TFs in the context of the genome-wide regulation.

Much attention has been paid to the ability of these TFs to directly interact via protein:protein interactions. For example, phosphorylation of T-bet at Y525 by Itk induces the interaction of T-bet with GATA3 and inhibition of GATA3 function [62], and FoxP3 interacts with RORγt and inhibits its function [63]. However, given that IL-4 and IFNγ coproducing cells can be observed in vivo in cell populations coexpressing T-bet and GATA3, these factors can clearly be

functional simultaneously, emphasizing the need to study the interactions of these factors within the chromatin context.

Despite the experimental data suggesting that mature T cells may alter phenotype, in the setting of allergy (Th2 cells), autoimmunity (Th1 and Th17 cells), or cancer (Treg cells) inappropriate responses have proven to be persistent and unmalleable. Consequently, understanding the regulation of stability and plasticity within these lineages is extremely relevant to the treatment of a wide range of illness.

Signaling transducers and activator of transcription (STAT) proteins mediate CD4 T cell differentiation

The most important determinant of CD4 T cell fate is the cytokine milieu present during TCR-mediated activation of naïve CD4 T cells. The cytokine milieu signals predominantly through the JAK-STAT pathway, and activated STAT proteins ultimately transduce these signals into alterations in transcription via interactions with DNA regulatory elements.

Activation of STAT1 by IFNγ induces T-bet differentiation in Th1 differentiation.[64, 65]; however, in vivo Th1 cells develop in the absence of STAT1 although they express lower levels of T-bet [66]. Activation of STAT3 by IL-6, IL-21, and IL-23 activates Th17 differentiation [67-70]; IL-17-producing cells fail to develop in the absence of STAT3 [71-73]. Activation of STAT4 by IL-12 mediates Th1 responses *in vitro* and *in vivo* [74, 75], and STAT4 expression is upregulated in Th1 cells vs. Th2 cells [76]. Low level STAT5 activation by

common γ chain cytokines is required for cell proliferation and survival [77], while strong STAT5 signaling is required for Th2 activation [78, 79] and also critical for Treg development [80, 81]. On the other hand, STAT5 suppresses Th17 cell differentiation [82]. Activation of STAT6 by IL-4 mediates Th2 differentiation [83, 84]; however, STAT6 independent Th2 differentiation can occur *in vivo* [85]. GATA3 on the other hand is indispensable for these responses. STAT3, STAT4, STAT5, and STAT6 bind differentially to diverse locations in genome during T cell differentiation mediated by their activation [86, 87].

Thus STAT proteins therefore are the critical mediators of cytokine mediated signals that establish master regulatory expression, although in some cases redundant mechanisms may exist that establish CD4 T cell lineages in the absence of specific STAT proteins.

Other Factors Involved in CD4 T cell Differentiation

A number of other factors show differential expression and/or activity between CD4 T cell subsets. These include Eomes, the Runx proteins Runx1 and Runx3, the interferon regulatory factors, IRF1 and IRF4, Growth factor independent 1 (GFI-1), c-Maf, JunB, Ikaros, and Helios.

Eomes functions similarly to T-bet, and stimulates IFNγ production in Th1 cells [88]. Eomes is involved in mediating IFNγ production downstream of IL-21 [89], so it may function in Th17 cells as well.

The Runx3 protein is most highly expressed in Th1 cells [90, 91], and functions to enhance IFNγ production [92], and repress IL-4 expression [90].

Runx1 has been shown to support the function of both Tregs and Th17 cells by enhancing expression of FoxP3 and RORγt respectively [93-95].

IRF4 is involved in Th2 differentiation and augments production of IL4, possibly by upregulating GATA3 [96]. In addition IRF4 is necessary for Th17 differentiation [97]. And its expression is necessary in Tregs for the proper control of Th2 driven immune responses [98]. IRF1 is involved in both Th1 and Th17 differentiation, as it mediates expression of IL-12Rα, a component of the receptors for IL-12 and IL-23 [99].

GFI-1 promotes Th2 differentiation [100], and also suppresses IFNγ production, IL-17 production, and Treg development [101, 102]. JunB and cMaf also promote Th2 differentiation, as they enhance production of IL-4 [103, 104].

Ikaros too is important in Th2 differentiation; in the absence of Ikaros, IFNg and T-bet expression are upregulated [105]. Helios is highly expressed in nTregs but not iTregs [106].

This listing of factors known to regulate CD4 T cell differentiation is incomplete, and doubtless many additional factors have as of yet unappreciated roles in the regulation of this process. The multitude of factors involved in regulation underscores the importance of describing the interaction of factors in a systems based approach in order to fully appreciate the complex interactions in specifying CD4 T cell differentiations.

Chromatin accessibility as a regulator of gene expression

Each cell in the human body shares an essentially identical genetic map, which encodes ~20,000 protein coding genes. The cell-type specific differential expression of these genes is therefore necessary to provide the divergent cellular phenotypes present in a complex multicellular organism. The basis for this differential regulation is two-fold: 1) the presence of DNA-binding proteins called transcription factors (TFs) within the nucleus and 2) the availability of the cognate regions of DNA to which TFs bind. The understanding of how chromatin accessibility and TF binding are regulated on a genomic scale has been facilitated greatly by the development of next generation sequencing technology. This has allowed for the examination of epigenetically regulated marks of chromatin accessibility and repression such as DNA methylation and histone methylation in CD4 T cells [107, 108]. Approximately 2.9 million of these regulatory elements are present in the human genome [109]. These regulatory elements selectively recruit TFs and their co-regulators to specific genomic loci, providing the foundation for cell-selective gene regulation.

Of particular interest is the dynamic regulation of the availability of regulatory regions during lineage differentiation. A classic experiment by Groudine and Weintraub demonstrated the propagation of beta-globin DNaseI hypersensitive sites (DHSs) in the absence of factors required for its induction, establishing the ability of chromatin to assert dominant roles in developmental decisions [110].

Chromatin accessibility has also been studied in the context of modifications to histones or directly to DNA which increase or decrease the accessibility of the chromatin. Direct methylation occurs at cytosine residues usually within CpG dinucleotides and generally opposes transcription [111]. Transcriptional silencing by DNA methylation is epigenetically heritable as the methylation status of DNA domains is faithfully propagated during development [112]. Thus alterations to DNA methylation may represent a fundamental mechanism by which chromatin accessibility is regulated during lineage differentiation.

In addition chromatin carries numerous histone modifications which have been correlated with active or repressed chromatin [113]. Of particular interest are H3K4me3, which marks areas of transcriptional activity and H3K27Me3, which marks areas of closed or repressed chromatin. The bivalent chromatin mark H3K4me3/H3K27me3 is found at several thousand developmentally poised regulatory elements across the genome [114] and appears to mark genes important for developmental fate decisions [115]. It is thought that genes with these bivalent marks are silent but remain poised for activation.

Together, these studies point to the role of regulatory region accessibility is guiding lineage differentiation. However, the degree to which histone modifications can be reliably inherited is unclear, as modifications appear readily alterable, and the scores different modifications, which may contribute in combinatorial fashion to the overall accessibility of an individual regulatory region, provide further complication.

Alternatively, chromatin accessibility can be directly measured via interaction with the non-specific exonuclease DNaseI. Sensitivity to DNaseI cleavage reliably detects regions of regulatory DNA characterized by the cooperative binding of sequence-specific transcriptional regulatory factors in place of a canonical nucleosome [116-119]. DNaseI hypersensitivity is the essential characteristic of all defined classes of active *cis*-regulatory elements including enhancers [120], promoters [118, 121], silencers [122], insulators [123], and locus control regions [124, 125]. Because DNaseI hypersensitivity overlies *cis*-regulatory elements directly and is maximal over the core region of regulatory factor occupancy, it enables precise delineation of the genomic *cis*-regulatory compartment. DNaseI treatment followed by next generation sequencing allows the systematic detection of regulatory elements [126].

Role of chromatin accessibility in promoting CD4 lineage stability vs. plasticity

Initial studies of CD4 T cell chromatin markers revealed marks consistent with irreversible commitment to lineages at key cytokine loci such as *IFNG* [127, 128] and *IL4* [129]. Similarly, *IL17* is found to possess activating marks in Th17 cells[130]. These studies noted active chromatin marks at in these signature cytokine genes in relevant lineages and repressive chromatin marks in opposing lineages. However, in support of a more plastic view of lineage differentiation in Th17 cells stimulated under Th1-promoting conditions, the *IL17* locus shows loss of activating marks and gain of repressive marks, while the *IFNG* locus shows

loss of repressive marks and gain of activating marks [53]. Underlying this and the other noted instances of lineage stability may be that even in firmly “lineage commited” CD4 T cells, chromatin marks at the key lineage regulating transcription factors demonstrate a broader spectrum of epigenetic states [107], suggesting that the expression of these factors may be more plastic.

Factors which control chromatin accessibility in CD4 T cells

Given the apparent abilities of “master regulatory factors” to directly regulate lineage commitment, it seems reasonable that these factors establish distinct cell fates by establishing the assessable regulatory landscape, as has been shown for in commitment to B and myeloid cell lineages [131]. Consistent with this view, T-bet can directly recruit chromatin remodeling complexes [132], and indeed T-bet, GATA3, RORC, and FoxP3 can act locally to regulate chromatin accessibility at key effector genes. However, other factors, such as the glucocorticoid receptor, have been recently described to act largely through utilization of a pre-existing regulatory DNA landscape [126].

Recent evidence shows that FoxP3 similarly acts predominantly through a pre-accessible chromatin landscape, where accessibility at specific FoxP3 binding sites is maintained by FoxO1 in naïve CD4 T cells [133]. In Th1 and Th2 cells, the enhancer landscape was found to be more dependent of STAT4 and STAT6, respectively, than T-bet and GATA3, although both factors clearly mediated remodeling at a large number of enhancers when overexpressed [134].

Currently lacking is a global description of how the regulatory landscape evolves in the differentiation of naïve CD4 T cells into the various effector lineages.

In this thesis we described the direct measurement of the accessibility of genomic regions to interaction with DNA binding proteins by mapping sensitivity to DNaseI cleavage genome wide in naïve CD4 T cells, and Th1, Th2, Th17, and Treg cells which have been isolated from healthy donors. This genome-wide mapping of potentially all regulatory regions active in CD4 T cells subsets, genome-wide, complements the available public data and provides insight into how these sites are regulated during lineage differentiation.

TFs act through collectively via a complex network of interactions

The most critical regulatory regions may be those involved in establishing lineage-specific TF expression. These regions regulate the interactions between transcription factors in which a given TF regulates other TFs or itself. Such mutual crossregulation among TFs defines regulatory sub-networks that underlie major features of cellular identity and complex functions such as pluripotency [135, 136], development [137] and differentiation [138]. On a broader level, these cross-regulatory interactions among the entire complement of TFs expressed in a given cell type form a core transcriptional regulatory network that endow the cell with the ability to integrate cellular signals and coordinate responses [139].In order to understand what transcription factors are associated with DNA at specific locations, chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq) has become the method of choice as it provides the ability to assay binding of a given factor genome-wide. Genome-wide analysis by ChIP-seq has been used to

derive subnetworks of small numbers of TFs, for example those involved in pluripotency {Kim, 2008 #615}, or larger-scale networks of dozens TFs {Gerstein, 2012 #891}. In CD4 T cells, the binding of the specific factors T-bet, GATA3, and STAT proteins has been assayed in CD4 T cells subsets by ChIP-seq [140, 141], and in Th17 cells, a small core network of factors has recently been described by this method [142]. However, such approaches are limited by three major factors: (1) the availability of suitable affinity reagents to perform precipitations; (2) the experimental complexity of interrogating the activities of multiple TFs within the same cellular environment; and (3) the sizable number individual experiments required to analyze each of the TFs within each of the cellular states that need to be studied.

A second, widely used method for constructing regulatory networks that attempts to address some of these issues, is studying the effects perturbations have on the expression of the individual network components. For example, the sea urchin endomesoderm regulatory network was constructed by individually perturbing the expression of dozens of TFs and analyzing the effect of these perturbations on the expression of downstream TF genes [137]. This method has more recently been applied at a larger scale by collecting several hundred independent gene expression perturbations for a single cell type and performing *de novo* network construction to generate the network of that cell type [143] including larger Th17-specific networks in the murine system [142, 144]. However, the networks generated from these studies are often confounded by

indirect edges, and they require hundreds of independent perturbations to build a single network.

We have described a method for identifying bound transcription factors genome-wide by DNaseI footprinting, which allows detection of binding of potentially all DNA bound factors in a single experiment [145]. Genomic footprinting enables systematic analysis of TF occupancy within the regulatory DNA of transcription factor genes, providing a direct and empirical approach for mapping cross-regulatory interactions (edges) between TF genes (nodes). Systematically applying this approach to all TFs with defined recognition sequences enables the construction of comprehensive TF-to-TF regulatory networks, and directed analysis of their organization, dynamics, and architectural features [146]. This approach has several notable advantages. Firstly, it precisely recapitulates well-validated TF-toTF regulatory connections [146]. Secondly, it provides directly visualizable and interpretable information. Thirdly, it is agnostic to whether any given TF-to-TF regulatory interaction is positive (activating) or negative (repressive), as these may vary conditionally even for a given TF. Finally, it accounts for the fact that a proportion of TF recognition sequences are partially redundant, and that simultaneous expression of cognate TFs for such elements results in alternating occupancy by all potentially occupying TFs [147].

Given the ready accessibility of CD4 T cells in the peripheral blood and the important contribution of CD4 T cell lineage control to maintaining immune homeostasis, CD4 T cells represent an important and available model to study

the dynamics of chromatin accessibility and regulatory networks in lineage differentiation.

Using genomic DNaseI footprinting, we define hundreds of thousands of transcription factor occupancy sites for over 400 TFs within CD4 T cell subsets, and compute comprehensive transcription factor regulatory networks for each cell type. The results expose a conserved CD4 T cell network architecture in which the chief role of canonical ‘master’ regulatory factors is to propagate or amplify mature CD4 T cell differentiation programs. This thesis provides an expansive resource for understanding transcriptional regulatory biology of lineage specification by CD4 T cells, and elucidating the general principles governing the action of ‘master’ lineage regulators within transcriptional regulatory networks.

Chapter 2: Materials and Methods

CD4 T cell subset isolation by FACS

PBMCs were isolated from healthy donors. CD4 T cells were enriched by MACS sorting using CD4 T cell isolation kit II (Miltenyi Biotec). CD4 T cells were isolated as in **Figure 3.1** on a FACS Aria II as follows: Naïve: CD3+CD4+ CD45RO-CD25-CCR4-CCR6-CXCR3-; Th1: CD3+CD4+CD45RO+CD25-CCR4-CCR6-CXCR3+; Th2: CD3+CD4+CD45RO+CD25-CCR4+CCR6-CXCR3-; Th17: CD3+CD4+CD45RO+CD25-CCR6+; Treg: CD4+CD4+CD25+CD127-. Naïve CD4 T cells were isolated directly for treatment with DNaseI. Other cell types were expanded *in vitro*.

Expansion of CD4 T cell subsets

Th1, Th2, Th17, and Treg subsets were stimulated with anti-CD3/28 Dynabeads (Invitrogen), and cultured in AIM-V media (Invitrogen) supplemented with 2% human AB serum and 50 U/mL IL-2 at 37^0C and 5% CO_2. Additionally, 10 ng/mL IL-12, 10 ng/mL IL-4, 10 ng/mL IL-1β, or IL-2 to 300 U/mL were added to Th1, Th2, Th17, and Treg cultures, respectively. Fresh media and cytokines were added on day 3 and cells were subsequently split every 2-3 days. On day eight, $1x10^5$ cells from each culture were collected and stimulated for 5 hours with PMA and ionomycin in the presence of brefeldin A. Th1, Th2, and Th17 cells were stained intracellularly for IFNγ, IL-17, IL-4, and IL-13. Populations were considered sufficiently pure if greater than 50% of cells were making lineage-

defining cytokines (IFNγ for Th1, IL-4 or IL-13 for Th2, IL-17 for Th17), and if fewer than 2% of cells were making cytokines specific to other lineages (Th1 cells producing IL-4 or IL-17, Th2 cells producing IFNγ or IL-17, Th17 cells producing IL-4). Tregs were considered sufficiently pure if greater than 80% were FOXP3+ by intracellular staining. Th1, Th2, and Treg cells were collected for DNaseI treatment. Th17 cells underwent a second round of purification by IL17 cytokine capture.

IL-17 Cytokine Capture

Th17 cells were stimulated for 90 min with .5 μM ionomycin and 5 ng/mL PMA. Cells were then stained sequentially with biotin-anti-CD45, streptavidin, and biotin-anti-IL17. Cells were incubated an additional 45 min on a rotator at 37^0C. Detection was performed with an anti-IL17-PE antibody which recognizes a motif independent of the capture antibody [148]. PE+ cells were collected on FACS Aria II (**Figure 3.2**). Cells were rested in AIM-V media (Invitrogen) supplemented with 2% human AB serum and 50 U/mL IL-2 for 48 hours at 37^0C, 5% CO_2, stimulated with anti-CD3/28 microbeads, and cultured in media additionally supplemented with 10 ng/mL IL-1β. After seven days if culture following this stimulation, 1x10^5 cells from each culture were collected and stimulated for 5 hours with PMA and ionomycin in the presence of brefeldin A, and cells were stained intracellularly for IFNγ, IL-17, IL-4, and IL-13. Populations were considered sufficiently pure if greater than 50% of cells were making IL-17,

and if fewer than 2% of cells were producing IL-4. Cells were collected eight days following restimulation for DNaseI treatment.

Polarization of naïve CD4+ T cells to generate Th1 and Th2 cells

Naive CD4 T cells were isolated as described. Cells were stimulated with anti-CD3/28 microbeads and cultured in AIM V media supplemented with 2% human AB serum at 37^0C, 5% CO_2. Additionally, 10 ng/mL IL-12 and 10 μg/mL anti-IL4 or 10 ng/mL IL-4, 10 μg/mL anti-IL-12, and 10 μg/mL anti-IFNG were added to Th1 and Th2 cultures, respectively. Fresh media and cytokines were added on day 2 with addition of 50 U/mL IL-2, and cells were subsequently split every 2-3 days. On day 8, polarizations were confirmed by intracellular cytokine staining as described, and cells were collected for DNaseI treatment.

DNaseI Mapping

DNaseI digestion and high-throughput sequencing were performed on intact human nuclei from various cell types, following published methods [149, 150]. Briefly, nuclei from roughly 10 million cells were extracted using NP-40 in an isotonic buffer. The NP-40 detergent was removed and the nuclei were incubated for 3 minutes at 37°C with limiting concentrations of the DNA endonuclease, deoxyribonuclease I (DNaseI) (Sigma) supplemented with Ca2+ and Mg2+. The digestion was stopped with EDTA and the samples were treated with proteinase K. The small ‘double-hit’ fragments (<500 bp) were recovered by sucrose ultra-centrifugation, end-repaired and ligated with adapters compatible

with the Illumina sequencing platform. The sequencing tags were aligned to the human reference genome and per-nucleotide cleavage counts were generated by summing the 5' ends of the aligned sequencing tags at each position in the genome.

DNaseI hypersensitive site (DHS) identification

We analyzed uniquely mapped -15 million uniquely mapped DNaseI cleavage events and verified reproducible patterns of DNaseI hypersensitivity at canonical *IFNG*, Th2, *IL17*, and *FOXP3* loci for biological replicates of each cell type. We chose replicates with highest data quality and increased sequencing depth to ~50 million uniquely mapped DNaseI cleavage events from representative samples of each cell type, which enabled delineation of identified and focused analyses within high DNaseI cleavage regions as identified by the *hotspot* algorithm and thresholded to a false discovery rate of 1%[126], within each cell type.

Identification of sites with differential regulated genomic regions

Sites of differential chromatin were defined pairwise between data sets (by cell type) as those DHSs found in one set and not in the other, as determined by the *bedops* program[151] with parameters -n -1. To identify highly differentially regulated DHSs, an expectation by rank approach was utilized[152, 153], and classified these into eight patterns showing predominant expression in 1 or 2 subsets by k-means clustering.

Identification of overrepresented motifs

Transcription factor (TF) binding motif sites were identified using *FIMO*[154] version 4.6.1, with p < 1e-5 and defaults for other parameters. The density of a motif in a cell type was calculated as the number of motif instances found within hotspot regions, normalized by the number of nucleotide bases within the hotspot regions. Subset-specific motifs were identified by dividing the motif density in a given DHS subset by the density in all subsets.

Association of regions with genes and function annotations

The association of DHSs with genes and function annotations was calculated using the Genomic Regions Enrichment of Annotations Tool (GREAT) [155]. Genomic regions were defined by a basal region plus extending regions, where a gene is assigned a basal regulatory domain of 5 kb upstream and 1kb downstream of the TSS (regardless of other nearby genes). The gene regulatory domain was extended in both directions to the nearest gene's basal domain but no more than 100 Mb in one direction. The association between the set of genes associated with differentially regulated regions and the set of genes with given functional annotations was calculated.

DNaseI footprinting

For each cell type, we computed the DNaseI cleavage per nucleotide by assigning to each base of the human genome an integer score equal to the number of uniquely mappable sequence tags with 5' ends mapping to that position. To identify DNaseI footprints comprehensively across the genome, we scanned for predicted TRANSFAC motif binding sites using FIMO [156], version 4.6.1, with a maximum *p*-value threshold of 1e-5 and defaults for other parameters. For each cell type, we filtered putative motif binding sites to those that overlapped footprints by at least 3 nucleotides (nt) as previously described [145]. Motif occupancy was calculated as the percentage of TF motifs within DHSs which overlapped footprints and normalized by the total number of detected footprints to account for differences due to sequencing depth and sample quality.

Visualization of aggregate DNaseI signal at binding motifs

TRANSFAC motifs [157] were used as inputs to *FIMO* to find all motif instances (P < 1e-5) within the DNaseI hotspots of each cell type. The left and right coordinates of each motif instance were padded by 35 nt. Using the *bedmap* tool from the BEDOPS suite, the per-nucleotide DNaseI cleavage values from deeply sequenced DNaseI-seq libraries were recovered for each motif occurrence. Aggregate plots were made by averaging DNaseI cleavages over all strand-oriented motif occurrences. Sequence logos were generated by assessing

the information content of the oriented genomic sequences from all motif occurrences [158].

Regulatory network construction

For each cell type, we computed DNaseI footprints as previously described [145]. We mapped motif-binding protein information found in TRANSFAC to 538 coding genes, using GeneCards [159] and UniProt Knowledgebase [160]. Some genes were indistinguishable when viewed from a potential motif-binding event perspective, as their respective gene products were annotated as binders to the same set of motif templates by TRANSFAC. In such cases, we chose a single gene as a representative and removed others which reduced the number of genes from 538 to 476. Additionally, we included a T-bet binding motif derived from ChIP-chip experiments. For each gene body, we padded the transcriptional start sites upstream by 5,000 nt and the transcriptional stop site by 1,000 nt downstream, and we identified all footprints within this gene regulatory region. Each network contained 476 vertices, one per gene. As described in Neph, Stergachis [161], a directed edge was drawn from a gene-vertex to another when a motif instance, potentially bound by the first gene's protein product, was found within the regulatory region of the second gene and overlapped a footprint by at least 3 nt.

Hive plots

We generated a hive plot [162] using the R library HiveR, version 0.2.1, to visualize directed interactions for factors known to differentially regulate Th17 and Treg cells among five CD4 cell types (Naïve, Th1, Th2, Th17, Treg). The hive plot was divided into five sections, one for each cell type. Reading the figure in clockwise fashion, a directed edge drawn from one axis to the next indicates the first gene regulating the second. Genes were oriented identically along each axis. A common interaction was defined as an interaction existing in two or more cell types. A second qualitative hive plot was created between the same five cell types and over all 476 genes.

Unique edge connectedness

We calculated the edges unique to each regulatory network. The sub-network of unique edges belonging to Th17 cells was plotted in Cytoscape [163]. Nodes were colored based on proximity to RORC.

Contribution of factors to network

We counted the total number of edges for every TF gene-node (sum of in- and out-edges) in a cell type and calculated the proportion relative to all edges (normalized network degree (NND)). We identified the factors for each CD4 T cell subset that had the highest relative NND.

Network clustering

We computed the pairwise Euclidean distances between cell types using NND vectors (NND values over all TFs) and grouped the cell types using complete linkage clustering.

Network Regulation of Effector Genes

For each transcriptional start site we interrogated regions of +/-10 kb for DNaseI footprints of factors within the network as described. We counted the total number of incoming edges for each gene and calculated the proportion relative to all interactions in each cell type. In order to identify network targets that showed high degrees of cell type specific regulatory interactions we calculated enrichment (modified Z-score) as [x - median(x)]/mad(x)] where x is the normalized input frequency for each cell type and defined significant enrichment as $z > 3.5$.

Analysis of factor position in network architecture

We utilized the mfinder software [164], version 1.20, to pull out all feed forward loop instances in regulatory networks. Prior to using the software, all self-edges, those from a TF gene-vertex to itself, were removed per the requirements of the software. The software parameters were set to *-ospmem 38 -maxmem 1000000 -s 3 -r 250 -z -2000.*

CNOT3 knockdown

Jurkat cells were cultured in RPMI 1640 medium supplemented with 10% FBS, 100 U/ml penicillin and 25 µg/ml streptomycin and maintained in a humidified incubator at 37 °C in the presence of 5% CO_2. Using the nucleofector II (Lonza) with X-001 program, $2x10^6$ cells were transfected in 100 µl of nucleofector® solution V (VAC-1003, Lonza) with 50 pmol of siRNA against human CNOT3 (HS_CNOT3_5 FlexiTubesiRNA, QIAGEN). siGENOME non-targeting siRNA pool (D-991206-13-20, Thermo Scientific) was used as a negative control. Total RNA was extracted using TRIzol (Ambion) 48 hours after transfection and reverse-transcribed (iScript™ cDNA synthesis kit, BIO-RAD). Strand-specific RNA libraries were prepared following a standard TruSeq stranded mRNA sample preparation guide with the indexed adapters (RS-122-2101, Illumina), and sequenced on an Illumina Genome Analyzer (Illumina) by the High-Throughput Genomics Center (University of Washington) according to a standard protocol. Genes showing differences in isoform abundance were identified using Cutdiff2[165] with a false discovery rate of 5%.

Chapter 3: Poised regulatory networks prime CD4+ T cell differentiation programs

Introduction

A hallmark of the adaptive immune system is the differentiation of naïve CD4 T cells into distinct effector lineages that orchestrate stereotypical immune responses to specific classes of pathogens. CD4+ T cell differentiation provides a well-studied paradigm for cell fate commitment controlled by lineage-selective expression of 'master' regulatory transcription factors, which in turn act across a broad landscape of downstream genes. Commitment of naïve CD4+ T cells to the Th1, Th2, Th17, and Treg lineages is dependent on lineage-specific expression of T-bet, GATA3, RORγt, and FOXP3, respectively [32]. T-bet, a protein encoded by the *TBX21* gene, is required for the generation of Th1 cells, which are crucial for the coordination of the host defense against viruses and intercellular pathogens [33]. GATA3 is required for the generation of Th2 cells, which coordinate anti-helminthic defenses [38], and RORγt and RORα are required for the generation of Th17 cells, which coordinate the removal of extracellular bacteria [46, 48]. FoxP3 is required for Tregs, which regulate inflammation in response to infection and prevent the development of

autoimmunity [49]. Each of these subsets is critical for the establishment of immunological memory and thereby facilitates the efficient removal of pathogens upon re-exposure. More rarely, these cell types can be inappropriately induced in the setting of allergy (Th2 cells), autoimmunity (Th1 and Th17 cells), or cancer (Treg cells), and these inappropriate responses have proven to be persistent and unmalleable. Additionally, commitment to a single lineage does not appear to be invariant [53-55], and the existence of T cell subsets co-expressing multiple lineage regulators [59-61] highlights the importance of understanding the interactions of these factors with the larger network of TFs in the context of the genome-wide regulation.

A fundamental question surrounding T cell differentiation from the naïve cell state is whether this process primarily represents the acquisition of new regulatory potential, the selective restriction of fates encoded within naïve cells, or some balance between these two. Additionally, how lineage-selective expression of ‘master’ regulators drives different differentiation programs is poorly understood. Recent studies have revealed a defining role for the pre-existing chromatin landscape of accessible enhancers and other regulatory DNA regions in modulating the binding patterns of induced transcription factors [126]. Defining this landscape within naïve CD4+ T cells and relating it to that of differentiated lineage populations should therefore provide insights into key commitment mechanisms.

Induced transcription factors (TFs) remodel gene expression patterns both through interaction with regulatory DNA of target genes that influence specific

cellular pathways, such as cytokine genes, and through binding within the regulatory DNA of other transcription factors. The latter creates networks of mutual cross regulation that have the potential to amplify, propagate, or reshape transcriptional regulatory signals [139, 166]. Currently, however, the transcription factor regulatory networks governing CD4+ T cell regulation are largely unexplored.

DNaseI hypersensitive sites (DHSs) are generic markers of all recognized classes of *cis*-regulatory DNA (enhancers, promoters, insulators, etc.). DHS mapping has been extensively employed to define with high precision the major regulatory elements controlling diverse immune genes including the T-cell receptor-alpha [167, 168], immunoglobulin [169-171], Th2 cytokine [172], and the interferon gamma [128] loci, among many others. High-quality genome-wide maps of DHSs can be generated using DNaseI-seq [126], providing comprehensive views of the regulatory landscapes of both specialized cell types and broader lineages [109].

DHSs result from the cooperative binding of sequence-specific TFs in place of a canonical nucleosome [173]. The binding of individual sequence-specific transcription factors within regulatory DNA can be visualized on a cellular scale using genomic DNaseI footprinting [149], which can be applied to specific human cell types to define the occupancy patterns of hundreds of transcription factors within a single experiment [145]. Expansive analyses of the regulatory DNA landscapes of hundreds of different cellular types and states has affirmed that core features of cell identity are reflected in the differential accessibility of

regulatory DNA, with different cell types typically distinguished from one another by patterned activation of thousands of predominantly distal regulatory elements [174].

How such landscapes evolve during the transition from naïve CD4 T cells to an effector lineage is not well defined, although it is clear that pre-existing patterns of regulatory DNA accessibility can exert a dominant influence on the genomic occupancy patterns of TFs that subsequently appear within the nucleus[126], including FoxP3[133]. Recent work has demonstrated that the enhancer landscape in Th1 and Th2 cells is dependent not just on T-bet and GATA3 but also on STAT factors[134]. Detailed delineation of accessibility patterns should therefore provide insight into the degree to which different lineage-regulating TFs act by exploiting available regulatory landscape vs. facilitating de novo acquisition of accessibility.

Of particular interest is the state of regulatory DNA controlling TF genes. TFs regulate both downstream effector genes and one another, forming, in the latter case, a cross-regulatory network that endows the cell with systems-level properties governing major features of cellular identity such as development [175] and differentiation [138, 146]. Systematic and large scale analyses of TF regulatory networks using approaches such as chromatin immunoprecipitation are hampered by the requirement for a separate experiment for every TF examined. Although widely employed, network inference from changes in RNA transcription following TF perturbation (e.g., knockdown) is largely indirect, and represents a major cellular disruption. While such an approach has been

employed to study gene networks governing murine Th17 differentiation [142, 144], even cursory analyses of the state and architecture of the TF network during alternative lineage differentiation from naïve cells has not been examined, nor has TF networks derived from human CD4 T cell data.

The recent advent of genomic DNaseI footprinting[147], which can detect the occupancy of hundreds of TFs simultaneously, has greatly facilitated the construction and comparative analysis of comprehensive TF regulatory networks in human cells[146]. Genomic footprinting enables the generation of detailed maps of TF-to-TF and TF-to-effector gene regulatory interaction networks spanning hundreds of TFs within the context of a single experiment, providing a unique and powerful platform for analyzing large-scale transcriptional regulatory mechanisms.[147]

Here, we apply DNaseI hypersensitivity mapping and genomic footprinting to define and analyze the lineage-selective propagation, acquisition, or restriction of regulatory DNA during the transition from human naïve CD4 T cells to effector progeny (Th1, Th2, Th17, and Treg) cells. We derive comprehensive transcription factor regulatory networks for CD4 T cells, exposing a conserved CD4 T cell network architecture in which the chief role of canonical lineage-selective ‘master’ regulatory TFs is to propagate or amplify upstream signals from combinations of more widely active regulators. Our results collectively provide an expansive resource for understanding the transcriptional regulatory programs of CD4 T cells and for elucidating the general principles governing the

action of ‘master’ lineage regulatory factors within transcriptional regulatory networks.

Results

Global identification of CD4+ T-cell regulatory DNA

We used DNaseI hypersensitivity mapping [126] to define comprehensive regulatory DNA landscapes in naïve CD4 T cells and Th1, Th2, Th17, and Treg subsets isolated from peripheral blood of normal healthy donors (**Figure 3.1-3.3**). To identify DHSs, we analyzed ~50 million uniquely mapped DNaseI cleavage events from biological replicates of each cell type, which enabled delineation of 97,007 to 127,610 DHSs per cell type (**Figure 3.4A)** at a false discovery rate (FDR) of 1%**.** In total we defined 229,117 distinct ~150bp DHSs, each of which was detected within one or more CD4 T cell type (**Figure 3.4B**). 41% of these (95,216 DHSs) were entirely specific to one of the five T cell subsets, with subset-specific DHSs ranging from 10.8K to 28.8K per cell type. Although 59% of DHSs were shared between two or more cell types, only 20.7% (47,585) were common to all subsets. The vast majority of these site were previously uncharacterized (**Figure 3.5**). The regulatory DNA specified in these *in vivo* derived CD4 T cells was similar but not identical to the regulatory DNA of CD4 T cell polarized *in vitro* as utilized in previous analyses[134, 176], with Th1 and Th2

cells showing approximately 80% overlap between *in vitro* and *in vivo* derived cells (**Figure 3.6**).

The well-characterized regulatory elements in the *IFNG-IL26-IL22* locus illustrate patterns of cell-selective DHS activation (**Figure 3.4C**). In this locus, all known regulatory sites (three gene promoters, seven intergenic enhancers, one intronic CTCF-binding enhancer, and two CTCF-binding insulators [128, 177]) are localized by DHSs. Many of these regions show clear subset-specific changes in chromatin accessibility, underscoring the influence of chromatin accessibility patterns in specifying dynamic transcriptional regulation. Subset-specific regulatory DNA is also prominent within lineage regulating transcription factor loci (**Figure 3.7A**) and more generally in the regulatory DNA regions of lineage-specific effector genes (**Figure 3.7B**), where relative differences in chromatin accessibility at the lineage-specific transcription factors and effector genes mirrors the known differential transcription of these genes.

Collectively, these DHS maps provide a comprehensive description of the cell-selective accessible chromatin landscapes in CD4 T cells which further specify cell-specific TF binding and gene expression in these cell types.

DNaseI hypersensitivity provides both a qualitative marker of regulatory DNA regions, and a quantitative marker of activation with >100-fold dynamic range [109, 150]. We next analyzed the most highly differentially regulated DHSs (i.e., those showing the greatest quantitative changes in DNaseI signal) to identify regulatory DNA that was strongly regulated in one or two cell subsets. We identified 43,703 regions with the strongest subset-specific patterns of

chromatin accessibility and classified these into seven patterns using a clustering approach (**Figure 3.8A**). The significant majority of elements localized in distal regions, either within downstream introns or within intergenic regions, consistent with prior reports that most of the cell-selective regulatory landscape lies external to promoter regions [109]. Of the sites identified, 3922 overlapped TSSs (**Figure 3.9A**); given the high sensitivity and statistical confidence ($p < 10^{-5}$) with which these promoters were identified, these correspond to a comprehensive set of differentially regulated transcripts between the CD4 subsets analyzed. Strikingly, the majority of the remaining ~50,000 sites were clustered within 5 kb of these 3933 TSSs (**Figure 3.9B**), indicating that expression of these transcripts is finely controlled by the interaction of multiple, local regulatory regions. Gene ontology analysis showed that collectively these elements are extremely strongly linked to genes involved in CD4 T cell functions such as T-cell activation and cytokine signaling (**Figure 3.8B** and **Table 1**).

Next we analyzed the regulatory information content of these highly differentially regulated elements by determining relative over-representation of known transcription factor recognition sequences within each of the seven clusters. Within the single cell type-specific clusters for Th1, Th2, Th17 and Treg cells, this analysis highlighted, respectively, T-bet, GATA, ROR and STAT5 recognition sequences (**Figure 3.8C**). This suggests that canonical lineage-specific TFs as well as STATs are involved in specifying effector cell-specific chromatin accessibility, and it highlights the ability to derive these relationships *de novo* from sequence analysis of the accessibility data.

DHSs selective for naive cells exhibited overrepresentation of recognition sequences for downstream effectors of Wnt signaling, which plays a role in maintaining the naïve T cell phenotype [178]. DHSs bi-selective for naive+Th1, Th2+Th17, or naive+Treg cells are populated with recognition sequences for general regulatory factors including CTCF, CCAAT/enhancer binding factors, and Homeobox factors, suggesting that TFs acting through these elements may be involved in establishing the broader regulatory landscape in CD4 T cells [177, 179, 180].

Acquisition of T-cell subset-specific regulatory DNA patterns

A fundamental question concerning the regulation of CD4 T lineage specification concerns whether differentiation from the naïve cell state represents primarily the acquisition of new regulatory potential, or the selective restriction of fates encoded within naïve cells, or some defined balance between these two processes. To identify regulatory DNA which informs the acquisition of lineage-specific function, we examined sets of DHSs that were completely lost or gained in the process of lineage differentiation. This analysis revealed expansion of the DHS landscape within each lineage, with acquisition of 51.0K to 66.4K DHSs (average 59.0K). This acquisition was balanced in each lineage by loss of an average of 33.3K DHSs (**Figure 3.10A**). The regulatory DNA acquired during differentiation was largely specific to the particular subtype into which naive cells differentiated (**Figure 3.10B**, left). However, the restricted elements lost from naïve cells were surprisingly consistent across effector subtypes (**Figure 3.10B**,

right), indicating the existence of a core regulatory DNA repertoire involved in maintenance of the naïve state. Taken together, these findings suggest that differentiation from naive to effector states involves both programmed loss of these core elements and simultaneous acquisition of new regulatory DNA landscapes that are highly specific to particular effector functions.

We next examined the regulatory information content of DHSs that were selectively lost vs. gained during lineage differentiation from naive CD4 T cells to effector cells (**Figure 3.10A**). In addition to lineage-specific DHSs, we also analyzed subset combinations (Th1+Th2, etc.; **Figure 3.10B**). For example, DHSs selectively gained in Th17 cells are significantly enriched in RORA sites (which can be bound by RORα or RORγ [95, 181]) and STAT sites. Similarly, GATA3 sites in combination with STAT6 sites are selectively enriched within Th2-acquired sites, both those that are specific to Th2 cells or those that are present in both Th2 and Th17 cells (**Figure 3.10C**). These results indicate that naïve CD4 T cells possess specific cohorts of regulatory DNA regions that are programmed, by virtue of their TF recognition sequence content, for rapid differentiation into specific lineages in responses to differentiation signals from various lineage-specific TFs. The results also indicate that STAT factors cooperate with multiple lineage-specific regulators to facilitate the acquisition of effector-selective regulatory DNA.

By contrast, examination of DHSs lost during the transition from naïve CD4 T cells to effector subsets revealed lineage-selective loss of major regulators of alternative lineage fates. For example, whereas ROR recognition

sites are selectively enriched in DHSs gained by Th17 cells, they are also enriched in DHSs that are selectively lost during differentiation to Th1 and Th2 cells. Similarly, DHSs enriched in T-bet and GATA3 recognition sites are selectively lost during differentiation of naïve CD4 T cells to lineages not dominantly regulated by these factors. DHSs lost in Th17 cells are enriched in AP2 and CNOT3 recognition sites, which are conversely enriched in DHSs gained during differentiation to Th1 and Th2 cells. Across all other lineages, DHSs lost from naive cells were substantially enriched in recognition sites for TCF/LEF, downstream effectors of Wnt signaling which plays a role in maintaining the naïve T cell phenotype [178], while DHSs gained were enriched in recognition sites for activating factors AP-1 and MAF.

Collectively, the above results indicate that differentiation from naive CD4 T cells is characterized by execution of three simultaneous regulatory programs: First, the maintenance of common regulatory DNA; second, the acquisition of novel lineage-selective regulatory DNA; and third, selective repression of regulatory regions associated with alternative lineage programs. Each of these regulatory programs is anchored in sequence features encoded within regulatory DNA that provide the basis for control by lineage-specific TFs.

Lineage commitment reflected in reciprocal TF enrichment in acquired vs. extinguished DHSs

The comparison of DHSs gained vs. lost during the transition from naïve CD4 T cells to effector subsets revealed reciprocal enrichment patterns for TF

recognition sequences. For example, DHSs selectively gained in Th1, Th2, or Th17 lineages show enrichment patterns for RORA, GATA3, STAT6, and STAT1 sites (**Figure. 10C** left) that are mirrored in DHSs selectively lost during transition from naïve cells to the alternative lineage fates (compare **Figure 3.10C** right). Thus, RORA binding site enrichment in novel DHSs *gained* during Th17 differentiation is mirrored by RORA enrichment in DHSs *lost* during transition to Th1 or Th2 fates. Strikingly, strong reciprocity (≥0.5) was evident for over 65% of 476 TFs examined and was most strongly evident for major lineage regulators. For example, **Figure 3.10C** illustrates reciprocity for GATA3 and RORA recognition sites in differentiation-gained vs. -lost DHS subsets, as evidenced by the left:right mirror symmetry of the normalized motif densities (**Figure 3.11**, left and middle panels). By contrast, general TFs such as p53 do not show this pattern (**Figure 3.11**, rightmost panels). Thus, the lineage-specific motif content of DHSs specifically gained in the transition to a given lineage (for example Th17 cells) and the lineage-specific motif content of DHSs specifically lost in the transition to alternative lineages (e.g., Th1 and Th2 cells) is remarkably similar. The fact that differentiation involves both the selective gain of lineage-specific regulatory DNA as well as the selective loss of regulatory DNA compatible with other lineage fates suggests a simple ‘door-closing’ mechanism for enforcing lineage commitment.

Transcription factor footprints in CD4 T-cell regulatory DNA

To analyze the *in vivo* occupancy of transcription factor recognition sequences within CD4 T-cell regulatory DNA, we performed genomic footprinting [145, 149]. Because DNaseI cleavages from DNase-seq are substantially concentrated within the ~2-3% of the genome defined by DHSs, significantly increasing the number of mapped cleavages enables large-scale recognition of DNaseI footprints. To visualize DNaseI footprints genome-wide in naive cells and effector subtypes, we deepened DNaseI cleavage coverage roughly 6-fold to an average of 300 million uniquely mapping sequence reads (DNaseI cleavage events) per sample. Alignment of DNaseI footprints with recognition sequences for transcription factors accurately and quantitatively recapitulates TF occupancy measured by ChIP-seq [145], and thus provides a powerful approach for analyzing the occupancy patterns of up to hundreds of TFs simultaneously. Using an automated approach, we delineated an average of 650,000 DNaseI footprints (6-40bp) in each cell type (range 483K to 788K) at a FDR of 1% (**Table 2)**.

Examples of DNasbeI footprints in the *IL12RB2-IL23R* locus demonstrate subset specificity of detected TF footprints (**Figure 3.12A-B**) Comparison with immune gene loci that have been heavily annotated by ChIP illustrates that genomic footprinting comprehensively detects the occupancy of lineage-regulating TFs in immune cells (**Figure 3.13A-B**).

Foortprint data also provide insights into the aggregate regulatory inputs of key immunomodulatory genes. For example, analysis of TF footprints within the

proximal intergenic and intronic DHSs of the *IL23R-IL12RB2* locus provides a general view of how the regulation of these genes develops as a function of naive vs. differentiated effector state (**Figure 3.12C-E**). In naïve CD4 T cells, in which both genes are quiescent, singleton T-bet and RORα/γt footprints are detected, compatible with a poised state (**Figure 3.12C**). In Th1 cells, upregulation of transcription of *IL12RB2* is accompanied by an expansion of the local regulatory DNA compartment and the appearance of footprints for activating factors such as IRF, RUNX, AP1, and SP1 as well as STAT factors and BCL6 (in Th17) (**Figure 3.12D**). In comparison, IL23R acquires an expanded DHS landscape and footprints for a similar cohort of activating factors (**Figure 3.12D**). Significant upregulation of IL23R expression In Th17 cells is accompanied by a specific increase in the number of ROR footprints, indicating that the increased availability of ROR within these cells is a vital trigger beyond the presence of general activating factors (**Figure 3.12E**).

The 'depth' of DNaseI footprints provides a quantitative measure of transcription factor occupancy, which can be leveraged to analyze the occupancy of specific factor recognition sequences, both within individual DHSs and in aggregate across the genomic regulatory DNA landscape [145]. To profile the aggregate occupancy of lineage regulating TFs in the context of each CD4 T cell subtype, we identified high stringency ($P<10^{-5}$) recognition sites for T-bet, GATA3, RORC, and FoxP3 within DHSs of naive cells and those of each effector cell type. We then computed the average occupancy of each TF recognition sequence by measuring the average depth of DNaseI footprints over these sites (**Figure**

3.14A), and visualizing base-level cleavage patterns which trace features of the protein-DNA interface (**Figure 3.14B**). RORA motifs showed highly Th17-specific occupancy (**Figure 3.14B** and **Figure 3.15A,** top), while GATA3 motifs were specifically occupied in Th2 cells, but showed some occupancy in naive cells and other effector subtypes, reflecting the essential role of low-level GATA3 expression in CD4 T cell development [182]. By contrast FoxP3 (forkhead) motifs were similarly occupied in all CD4 T cell subsets, likely indicating the binding of these motifs by factors with similar sequence specificity. Indeed, in murine Treg differentiation, FoxO1 is displaced by FOXP3 at prospective FOXP3 binding sites [183]. Similarly forkhead motifs could be occupied in Th1 cells by FoxO1 or other forkhead binding proteins such as FoxJ1, which may function in a cell intrinsic regulatory role in Th1 cells [184]. T-bet shows highest occupancy in Th1 cells and Treg cells, where the occupancy likely reflects the critical role of this factor in a subpopulation of Tregs [60]. We also detected differential occupancy of STAT motifs, consistent with differing cooperating roles with lineage regulators (**3.14B**). The results thus accord with the differential availability of canonical subset-specific factor motifs across CD4 T cell subsets, and provide additional insights into occupancy patterns of recognition sequences by related factors.

STAT family members differ slightly in their DNA binding specificity[185], suggesting subtle variation in DNA binding domain morphology. Averaging per nucleotide DNaseI cleavage across the recognition sequence landscape of a TF effectively traces the structural topology of protein:DNA interaction, paralleling co-crystal structures[147] (**Fig. 3.14A**). We therefore analyzed in detail the

aggregate cleavage patterns over 3- and 4-space STAT recognition sequences, which are differentially (and competitively) occupied by STAT family members including STAT3 (Th17), STAT4 (Th1), STAT6 (Th2), and STAT5 (increased activity in Tregs)[86, 87].

Although absolute aggregate per nucleotide cleavage over the recognition sequence of a given TF may vary widely between cell types (reflecting differential occupancy related to factor abundance), the relative cleavage pattern (reflecting the morphology of the protein:DNA interface) is generally stable[147]. For example, in spite of variable expression of IRF1 and GATA3 between CD4 T cell subsets resulting in differential occupancy, the binding interface morphology reflected in relative per nucleotide DNaseI cleavage patterns was unchanged (**Fig. 3.15C**). By contrast, we detected clear cell-selective differences in the STAT protein:DNA interface (**Fig. 3.15D**), compatible with preferential occupancy by different factors in different cell types (STAT 4/3/5 in Th1/Th17/Treg, respectively). Interestingly, these differences were concentrated in the central gap region between the half sites of the STAT dimer, suggesting that this intervening region is selectively distorted as a consequence of interactions between the two halves of the STAT dimer.

T-cell subset-specific transcriptional regulatory networks

We next sought to explore comprehensively the regulatory landscape of transcription factor genes themselves, including mutual cross-regulation, as a function of CD4 T cell subtype. Genomic footprinting enables systematic

analysis of TF occupancy within the regulatory DNA of transcription factor genes, providing a direct and empirical approach for mapping cross-regulatory interactions (edges) between TF genes (nodes). Systematically applying this approach to all TFs with defined recognition sequences enables the construction of comprehensive TF-to-TF regulatory networks, and directed analysis of their organization, dynamics, and architectural features[146]. For example, the regulatory DNA of the GATA3 locus in Th2 cells is populated by CNOT3, SP1, c-MAF, and JUNB footprints, the latter two of which mediate Th2 differentiation [186]. In turn, GATA3 cross regulates ETS1, MYC, and ZBTB7B (ThPOK) (**Figure 3.16**). These interactions define a node (GATA3) and incoming and outgoing edges of a transcriptional network. Iteration of this process over the universe of all TF genes with known recognition sequences (n=476) permits the construction of a comprehensive transcription factor regulatory network [161]. Despite the functional redundancy of some DNA binding motifs [187], the significant majority of interactions can be ascribed to a unique TF. Because DNaseI hypersensitivity at proximal regulatory sequences is tightly linked with gene expression (R=0.93 vs. RNA-seq [188], transcriptional regulatory network construction is naturally focused on the expressed TF complement of each cell type.

This approach has several notable advantages. Firstly, it precisely recapitulates well-validated TF-TF regulatory connections[146]. Secondly, it provides directly visualizable and interpretable information. Thirdly, it is agnostic to whether any given TF-to-TF regulatory interaction is positive (activating) or

negative (repressive), as these may vary conditionally even for a given TF. Finally, it accounts for the fact that a proportion of TF recognition sequences are partially redundant, and that simultaneous expression of cognate TFs for such elements results in alternating occupancy by all potentially occupying TFs[147].

We utilized this approach to generate transcription factor regulatory networks for CD4 naïve cells and for the four effector cell types. We defined an average of 14,893 TF-to-TF interactions (edges) between the 476 TF nodes per cell type (range 10,585-15,544), of which the majority were conserved among all subsets, with an average of 1,404 specific to a single cell subset (**Figure 3.18**). At higher resolution, examination of cell-selective interactions between eight factors known to regulate T cell function clearly highlighted the role of cell-restricted factors within their canonical cognate cell types (**Figure 3.17**).

We next focused on the sub-networks formed by the relatively small proportion of edges unique to each network. For example, the Th17-specific sub-network contains 6.2% of the total number of TF-to-TF interactions detected in Th17 cells (**Figure 3.18**). These in turn form a highly-connected sub-network (**Figure 3.19**), indicating that the Th17 regulatory program is governed not by one or more simple TF cascades, but rather by the coordinated activity of many factors. Among all TFs, RORC is the most highly connected (**Figure 3.20A**) and is centrally placed in the sub-network (**Figure 3.19**). 53% of the nodes in the network lie within two degrees of RORC and the edges to and from these nodes comprise 87% of the subset-specific network (**Figure 3.20B**). It is notable that the structure of the Th17-specific sub-network departs greatly from the complete

structure of the regulatory network in Th17 cells, as highlighted by the position of CTCF. CTCF is one of the most well-connected and central factors of the complete regulatory network but is peripheral within the sub-network, connected by only a single edge. Together, these results highlight RORC as the distinguishing regulatory of Th17 cells vs. other subsets, while providing an expanded view of the complex regulatory milieu through which it facilitates execution of the Th17 regulatory program.

Canonical lineage regulators are distinguished by high cell-selective connectivity

Network analysis provides a powerful platform for unbiased identification of critical cell-selective regulators, which appear as TFs that exhibit cell-specific increases in network connectivity [161]. To infer the factors critical to T cell subset differentiation, we identified TFs that exhibited subset-specific increases in network degree, i.e. the sum total of both incoming edges and outgoing edges. For all factors we calculated the normalized network degree (total number of incoming and outgoing connections divided by network size), which reflects the aggregate connectivity of a TF determined by both the relative occupancy of a TFs regulatory DNA by other factors, and the scope of regulation of other factors by that TF. The factors with the highest relative changes in connectivity between the networks of CD4 T cell subsets are shown in **Figure 3.21**. In the naïve CD4 T cell network, SATB1, NR4A2, LEF1, and TCF7 were found to have the highest relative connectivity when compared to networks in other T cell subtypes. SATB1

is a transcriptional repressor and chromatin modifying factor that has been shown to be critical for CD4 T cell development and persistence [189], and NR4A2 is reported to induce the development of iTreg cells from naïve CD4 T cells [190]. LEF1 and TCF7 (which encodes the TCF-1 protein) are both downstream effectors in Wnt signaling which are known to be highly expressed in naïve T cells and down regulated upon activation [191].

Across other T cell subsets we identified well characterized CD4 subset-regulating factors within their respective subtypes: GATA3 and the known regulator of Th2 cytokine production NFIL3 [192] in Th2 cells; RORC and PPARG, which has been shown to be important in restraining Th17 dependent autoimmunity [193], in Th17 cells; and FOXP3, STAT5, IKZF2 (which encodes the Helios protein), IRF4, and T-bet in Treg cells. While no factors showed unique increases in occupancy in the Th1 cell network, CEBPB and TBX21 (which encodes the protein T-bet) showed high relative occupancy in Th1-specific networks. The presence of TBX21 in Th1, Th17, and Treg cells, may indicate a shared role for this factor in regulatory networks (**Figure 3.21**). CEPB, which has been shown to target diverse cytokine gene loci such as IL-4 [194], IL-5 [195], IL-6 [196], and TNF-a [197] also showed high occupancy in Th2 cells. Connectivity analysis therefore reveals the prominent activity of known regulators of CD4 T cell subset function and implicates important novel roles for several additional factors.

We verified that observed network structures were delineated specifically by lineage-differentiation by comparing the global network structure of *in vitro*

polarized Th1 and Th2 cells to our *ex vivo* subsets. By comparing the total degree of every node in the network, which strikingly reproduces lineage relationships [161], we verify that global network structure is extremely similar in T cells subsets obtained through these different experimental methodologies (**Figure 3.22**).

TF network outputs to key immunomodulatory genes

The central TF network of each cell type collectively connects to diverse non-TF target genes mediating subset-specific cellular functions. Using an unbiased approach, we detected 2,695 non-TF genes (range 269-715 genes per cell type; **Methods**) that showed highly subset-specific incoming regulatory connections (i.e., occupancy of their proximal regulatory DNA by multiple TFs simultaneously in a cell-selective fashion) (**Figure 3.23A**). This analysis in turn disclosed many genes with well-known subset-specific effector function (**Figure 3.23A**, right). Importantly, for a given gene, network connectivity may vary considerably across cell types, even in cases where gene expression remains relatively uniform. For example, *CD4* is expressed at roughly equivalent levels in both naïve and effector cells, and yet shows high network connectivity selectively in naïve cells that coincides with the selective appearance of an upstream DHS (**Figure 3.23B**). Using this approach we identified numerous genes of previously uncharacterized immunomodulatory function, which included several poly-zinc-finger transcription factors (**Figure 3.24**), suggesting a role for these TFs in CD4 T cell subset function.

Poised network structures around canonical effector cytokine genes

We next examined the proximal TF networks that regulate key cytokine genes including *IL17A* and *IFNG* by combining the TF network with footprint occupancy analysis within these gene loci (**Figure 3.25A-D**). The interactions between the *IL17A* gene and TFs reported to regulate expression of this cytokine are shown in **Figure 3.25A,B**. Interactions between these TFs and *IL17A,* as well as with *RORC*, are detected exclusively in Th17 cells. However, connections between the regulating TFs including *RUNX1, ETS1, SMAD4, and BATF,* are already established in naïve CD4 T cells, and thus comprise a 'poised' sub-network encompassing interactions necessary for proper regulation of IL-17 upon appearance of RORC.

An analogous situation obtains at the *IFNG* locus, where T-bet, RUNX3 and CTCF footprints were found to be present at this locus even in naïve CD4 T cells, poising these cells for rapid expression of this gene (**Figure 3.25C,D**). In Th1 and Th17 cells, additional activating factors, such as STAT3 and STAT4, are recruited to this locus. Like the *IL17A* regulating network, the interactions between TFs in the proximal regulatory sub-network are pre-specified in naïve cells, establishing the regulatory framework for later expression of *IFNG*.

Driver vs. passenger roles of canonical lineage regulators in CD4 T-cell networks

Given the observed centrality of canonical lineage-defining TFs to cell type-specific TF regulatory networks, we next asked whether TBX21, GATA3, RORC, and FOXP3 function within the TF network primarily as drivers of cell-specific regulatory states, or whether they are themselves governed to varying degrees by upstream regulatory signals.

Feed-forward loops (FFLs) are perhaps the most important architectural motif in biological networks[198], and are widely exploited by diverse processes in which the forward propagation of signals is vital to biological outcome. FFLs comprise a basic three-node structure in which information is propagated from a top (driver) node through the middle (1st passenger) to the bottom (2nd passenger) node, with parallel reinforcing input from the driver node directly to the 2nd passenger node. Because of their prevalence, functional importance, and hierarchical structure, FFLs present a simple yet powerful framework for analyzing major regulatory cascades within complex network architectures.

We identified an average of 124,615 FFLs within each cell type TF network, and then computed the frequency with which each master regulatory TF was found in the driver vs. passenger positions (**Figure 3.26**), revealing a marked preference for canonical lineage-defining factors to reside in passenger positions of FFLs. In 94.4-99.0% of FFLs, TBX21, GATA3, FoxP3, and RORC were found in one of the two FFL passenger positions. This specificity contrasted sharply with non-lineage-selective regulators (e.g., ThPOK/*ZBTB7B* and *CNOT3*; **Figure 3.26**).

The striking concentration of canonical lineage regulators within passenger positions of FFLs indicates that in the context of the TF network, these factors function primarily to integrate and propagate upstream signals in a subset-specific fashion. Consistent with this model, we observed that the local sub-networks surrounding key effector cytokine genes such as *IL17A* or *IFNG* are largely pre-specified in naïve CD4 T cells (**Figure 3.25**), indicating that lineage-specific expression of major lineage regulators does not result in extensive feedback remodeling of the TF network.

Upstream drivers of canonical CD4 T lineage regulators

We next asked what TFs occupy the driver positions in FFLs that contain canonical lineage regulators in a passenger position. Computing these on a cell type specific basis revealed a surprising number of widely expressed TFs (e.g., BACH1, CNOT3, FOS, IRF1, KLF4, MAF, MYC, PATZ1, PPARG, RXRA, ZBTB7B; **Figure 3.27**). Moreover, these TFs were frequently shared, in different combinations, between multiple FFLs, while a smaller subset of factors was utilized uniquely to drive expression of a specific factor in a specific cell type (e.g., STAT5A driving FOXP3-containing loops; **Figure 3.27**). In order to determine if these factors, which are widely expressed could neverthe less display subset-specific alterations in expression we consulted the Gene Expression Atlas,[199] and found substantial evidence for differential expression of these factors in different subsets of CD4 T cells (see **Table 3**).

While c-MAF, PPARG, and STAT5 are known to play CD4 T subset-specific roles,[80, 186, 193] and FOS, MYC, KLF4, PPARG, and RXRA may play roles in activation and differentiation[193, 200-202], our analysis also exposed factors which have not previously been appreciated to play important roles in T cell lineage specification, including PATZ1 and CNOT3. Importantly, utilization of specific factor combinations as drivers was highly subset-specific (**Figure 3.28**). For example, T-bet containing FFLs show different driver TF preferences in each cellular context: in both Th1 and Treg cells, CNOT3 commonly drives T-bet-containing FFLs; while ZBTB7B functions as a Treg-specific driver (**Figure 3.28A**).

*Although CNOT3 previously lacked any recognized role in T cell differentiation and function, it has recently come to light as the most frequently mutated gene in adult T cell acute lymphoblastic leukemia (ALL)[203]. In order to verify the predicted functional network connectivity of CNOT3 in human T lymphoid cells, we performed siRNA knockdown of CNOT3 in Jurkat T cells, followed by RNA-seq. siRNA transfection resulted in ~50% knockdown of CNOT3 expression (**Figure 3.29A**). Comparison of global RNA expression in these cells vs. those infected with control siRNA revealed differential expression of transcripts associated with 32 TF genes (**Figure 3.29B**). Strikingly, 23 of these TFs were direct (i.e., first order) targets of CNOT3 in the Th1 network ($P < 4.6 \times 10^{-7}$). Of the remaining 9 factors, 8 were found within two connections downstream of CNOT3 (**Figure 3.29B**). These results confirm the function of CNOT3 within T cell regulatory networks.*

Discussion

The classical view of CD4 T cell specification involving selective expression of a lineage master regulatory factor that results in irreversible lineage specification has been challenged by multiple reports showing coexpression of such factors[59, 60] and plasticity between effector subsets[53-55, 107, 204, 205]. A comprehensive account of how lineage specification and plasticity are balanced will require, in turn, delineation of the interactions between canonical lineage-regulators and other expressed TFs, description of the interactions among these more broadly defined lineage regulators, and description of the interactions between the entire compliment of TFs and downstream effector genes.

Using DNaseI mapping and genomic footprinting, we have defined comprehensive regulatory DNA landscapes for human naïve CD4, Th1, Th2, Th17, and Treg cells at unprecedented scope and resolution, exposing an order of magnitude more regulatory DNA elements (of which the overwhelming majority are distal, non-promoter elements) than prior descriptions of enhancer landscapes in CD4 T cells (**Figure 3.5**).[114, 134, 176] Analysis of transitions from naïve cells to different effector lineages revealed systematic patterns of regulatory DNA extinction and acquisition, consistent with the action of specific transcription factor cohorts. In turn, genomic footprint maps enabled the construction of by far the most comprehensive transcription factor regulatory

networks for CD4 T cells. Importaintly, the direct comparison of this data to the majority of publish genome-wide CD4 T cell TF-binding or network architecture studies is impractical because these studies were performed in the mouse; specific regulatory connections may be divergent between mouse and man[206] . Collectively, our data, obtained from primary human *in vivo*-differentiated cells, provide a new and rich resource for elucidating both mechanisms of normal differentiation and those underlying variation linked to immune diseases and traits.[207]

Our results bring to light a number of findings with general significance for understanding mechanisms of CD4 T cell differentiation. A significant observation was that the accessible regulatory DNA of naïve CD4 T cells is enriched in recognition sites for the lineage master regulatory factors, and undergoes programmed remodeling during subset specification. During differentiation toward a specific effector lineage, lineage-relevant regulatory DNA inherited from naïve cells is maintained; new, predominantly lineage-specific elements arise; and accessibility at regulatory elements associated with other lineage trajectories is specifically extinguished, providing a mechanism for reinforcing lineage commitment at the level of cis-regulatory DNA.

Our analysis of TF networks in CD4 T cells provides novel general insights into the organization of high level transcriptional control pathways. Unlike one-by-one analyses such as TF ChIP-seq, genomic footprinting enables simultaneous profiling of the occupancy of hundreds of sequence-specific TFs,[147] enabling the construction of accurate, large-scale transcription factor

regulatory networks in primary human cells.[146] Strikingly, while master regulatory TFs T-bet, GATA3, FoxP3, and RORγt acquire high network connectivity in their cognate lineage, these factors are not positioned within TF networks as primary drivers of major downstream regulatory processes. Rather, they are themselves dynamically regulated by combinations of more widely expressed TFs, many of which were not previously appreciated to play significant roles in CD4 T cell regulation. One in particular – CNOT3 – was prevalent in network driver positions across all lineages, and knockdown of CNOT3 resulted in both up- and down-regulation of proximal network neighbors. CNOT3 was recently recognized as a tumor suppressor of T cell acute lymphoblastic leukemia[203], raising the intriguing possibility that CD4 T cell malignancies may be potentiated at the TF network level.

Taken together, our analyses of CD4 T cell regulatory landscapes and networks suggest a general framework under which lineage stability and plasticity are balanced by the interplay between networks of cross-regulatory TF interactions and the interactions of these networks with the broader accessible chromatin landscape (**Figure 3.30**). A network of commonly expressed transcription factors that nonetheless have are differentially expressed across CD4 T cells function upstream of lineage-specific mater regulatory expression. Critically, these master regulatory factors lack prominent interactions back on the specifying upstream network, which may allow for the persistence of network architectures necessary for future expression of factors associated with other lineages. Naïve CD4 T cells are enriched in accessible regulatory DNA that

contains recognition sites for these master regulatory factors, effectively ‘bookmarking’ them for rapid response to differentiation signals. During differentiation of a specific effector lineage, maintenance of lineage-appropriate regulatory DNA is coupled both to loss of elements associated with other lineage programs and to lineage-specific acquisition of specialized regulatory DNA. The passively supports lineage maintenance by reducing responsiveness to master regulatory factors associated with other lineages, many of which can be expressed in multiple effector types[59, 60, 204].

Chapter 4: Concluding Remarks

The comprehensive description of regulatory elements, in addition to providing the preceding insights into how multiple TFs interact in order to coordinate CD4 T cell lineage differentiation, will facilitate many future studies designed to better understand how CD4 T cell differentiation is controlled and how to intervene therapeutically in CD4 T cell differentiation. First, the cataloguing of active regulatory regions is the first step in designing studies to specifically understand the function of individual elements in regulating lineage differentiation. Indeed, the deletion of a single regulatory element may have a profound impact on the differentiation of CD4 T cells [208, 209].

Second DNaseI footprinting has greatly augmented the data generated, by directly detecting protein:DNA interactions. This allows the identity of DNA bound factors to be inferred. The availability of both more and higher quality recognition promises to considerably enhance human transcriptional network analysis. Such refined data may increase the number of TFs which can be included in this analysis and may enable differentiation of multiple factors that currently appear to bind the same recognition sequence, such as STAT proteins.

In the case of STAT proteins, many ChIP-seq studies have already been performed [86, 87], and the comparison of DNaseI footprinting and ChIP-seq will allow sites of direct and indirect binding to be differentiated. Currently, because these ChIP-seq studies were performed in mice, the direct comparison of data between these, and our profiling in human cells is limited by incomplete descriptions of analogous regions of the human and mouse genomes.

A third future utility for the genome-wide maps of regulatory elements is in the generation of extremely sophisticated maps of regions from human to mouse. By analyzing the total regulatory content of analogous cell types from the two species, regulatory regions can be directly compared. Furthermore, the detection of the active regulatory content of individual regulatory elements by DNaseI footprinting will allow regions to be mapped not only based on relative position near genes, but by their actual functional content. The development of such sophisticated maps could allow the correlation of murine models of disease in which the effect of single factors can be studies in isolation with the vast amount of human genetic data obtained from effectors such as genome-wide association studies (GWAS).

Currently the effect of GWAS identified SNPs that lie outside of the coding region of genes cannot be ascribed a function. These regions may directly affect transcriptional regulation by disruption TF binding sites, or alternatively they may merely function as markers of other genetic disruptions that they lie in linkage disequilibrium with the identified SNPs. The genome-wide maps of regulatory elements to single base pair resolution allow SNPs that disrupt TF binding sites to be directly identified. The theoretical value of this method can be clearly demonstrated by comparing the overlap of DHSs regions in the identified T cell subsets with GWAS SNPs correlated with an increased risk of Crohn's disease. T cells show greatly increased numbers of Crohn's disease associated SNPs relative to other cell types, and specifically Th1 cells and to an even greater extent Th17 cells have accessible chromatin that include SNPs associated with

disease (**Figure 4.1**). Thus, DNaseI sensitivity identifies regions in which mutations may result is phenotypic changes in gene expression.

References

1. Medzhitov, R. and C.A. Janeway, Jr., *Innate immune recognition and control of adaptive immune responses.* Semin Immunol, 1998. **10**(5): p. 351-3.
2. Janeway, C.A., Jr., *The immune system evolved to discriminate infectious nonself from noninfectious self.* Immunol Today, 1992. **13**(1): p. 11-6.
3. Kleinfield, R., et al., *Recombination between an expressed immunoglobulin heavy-chain gene and a germline variable gene segment in a Ly 1+ B-cell lymphoma.* Nature, 1986. **322**(6082): p. 843-6.
4. Marrack, P. and J. Kappler, *The T cell receptor.* Science, 1987. **238**(4830): p. 1073-9.
5. Jameson, S.C., K.A. Hogquist, and M.J. Bevan, *Positive selection of thymocytes.* Annu Rev Immunol, 1995. **13**: p. 93-126.
6. Siggs, O.M., L.E. Makaroff, and A. Liston, *The why and how of thymocyte negative selection.* Curr Opin Immunol, 2006. **18**(2): p. 175-83.
7. Fink, P.J. and M.J. Bevan, *Positive selection of thymocytes.* Adv Immunol, 1995. **59**: p. 99-133.
8. Huesmann, M., et al., *Kinetics and efficacy of positive selection in the thymus of normal and T cell receptor transgenic mice.* Cell, 1991. **66**(3): p. 533-40.
9. Kappler, J.W., N. Roehm, and P. Marrack, *T cell tolerance by clonal elimination in the thymus.* Cell, 1987. **49**(2): p. 273-80.
10. Carlow, D.A., et al., *Peripheral tolerance through clonal deletion of mature CD4-CD8+ T cells.* Int Immunol, 1992. **4**(5): p. 599-610.
11. Schwartz, R.H., *T cell anergy.* Annu Rev Immunol, 2003. **21**: p. 305-34.
12. Bennett, S.R., et al., *Help for cytotoxic-T-cell responses is mediated by CD40 signalling.* Nature, 1998. **393**(6684): p. 478-80.
13. Jenner, E., *An inquiry into the causes and effects of the Variolae vaccinae, 1798.* 1st ed. 1966, London,: Dawsons. 4 , iv, 77 p.
14. Mosmann, T.R., et al., *Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins.* J Immunol, 1986. **136**(7): p. 2348-57.
15. Killar, L., et al., *Cloned, Ia-restricted T cells that do not produce interleukin 4(IL 4)/B cell stimulatory factor 1(BSF-1) fail to help antigen-specific B cells.* J Immunol, 1987. **138**(6): p. 1674-9.
16. Abbas, A.K., K.M. Murphy, and A. Sher, *Functional diversity of helper T lymphocytes.* Nature, 1996. **383**(6603): p. 787-93.
17. Hsieh, C.S., et al., *Development of TH1 CD4+ T cells through IL-12 produced by Listeria-induced macrophages.* Science, 1993. **260**(5107): p. 547-9.
18. Le Gros, G., et al., *Generation of interleukin 4 (IL-4)-producing cells in vivo and in vitro: IL-2 and IL-4 are required for in vitro generation of IL-4-producing cells.* J Exp Med, 1990. **172**(3): p. 921-9.
19. Swain, S.L., et al., *IL-4 directs the development of Th2-like helper effectors.* J Immunol, 1990. **145**(11): p. 3796-806.

20. Rocken, M., J.H. Saurat, and C. Hauser, *A common precursor for CD4+ T cells producing IL-2 or IL-4.* J Immunol, 1992. **148**(4): p. 1031-6.
21. Sad, S. and T.R. Mosmann, *Single IL-2-secreting precursor CD4 T cell can develop into either Th1 or Th2 cytokine secretion phenotype.* J Immunol, 1994. **153**(8): p. 3514-22.
22. Grogan, J.L., et al., *Early transcription and silencing of cytokine genes underlie polarization of T helper cell subsets.* Immunity, 2001. **14**(3): p. 205-15.
23. Noben-Trauth, N., J. Hu-Li, and W.E. Paul, *IL-4 secreted from individual naive CD4+ T cells acts in an autocrine manner to induce Th2 differentiation.* Eur J Immunol, 2002. **32**(5): p. 1428-33.
24. Aggarwal, S., et al., *Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17.* J Biol Chem, 2003. **278**(3): p. 1910-4.
25. Park, H., et al., *A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17.* Nat Immunol, 2005. **6**(11): p. 1133-41.
26. Harrington, L.E., et al., *Interleukin 17-producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages.* Nat Immunol, 2005. **6**(11): p. 1123-32.
27. Veldhoen, M., et al., *TGFbeta in the context of an inflammatory cytokine milieu supports de novo differentiation of IL-17-producing T cells.* Immunity, 2006. **24**(2): p. 179-89.
28. Chen, W., et al., *Conversion of peripheral CD4+CD25- naive T cells to CD4+CD25+ regulatory T cells by TGF-beta induction of transcription factor Foxp3.* J Exp Med, 2003. **198**(12): p. 1875-86.
29. Fu, S., et al., *TGF-beta induces Foxp3 + T-regulatory cells from CD4 + CD25 - precursors.* Am J Transplant, 2004. **4**(10): p. 1614-27.
30. DiPaolo, R.J., et al., *Autoantigen-specific TGFbeta-induced Foxp3+ regulatory T cells prevent autoimmunity by inhibiting dendritic cells from activating autoreactive T cells.* J Immunol, 2007. **179**(7): p. 4685-93.
31. Waddington, H.K., *Fetal salvage in abruptio placentae.* Am J Obstet Gynecol, 1957. **73**(4): p. 816-21.
32. Zhu, J., H. Yamane, and W.E. Paul, *Differentiation of effector CD4 T cell populations (*).* Annu Rev Immunol, 2010. **28**: p. 445-89.
33. Szabo, S.J., et al., *A novel transcription factor, T-bet, directs Th1 lineage commitment.* Cell, 2000. **100**(6): p. 655-69.
34. Szabo, S.J., et al., *Distinct effects of T-bet in TH1 lineage commitment and IFN-gamma production in CD4 and CD8 T cells.* Science, 2002. **295**(5553): p. 338-42.
35. Mullen, A.C., et al., *Role of T-bet in commitment of TH1 cells before IL-12-dependent selection.* Science, 2001. **292**(5523): p. 1907-10.
36. Usui, T., et al., *T-bet regulates Th1 responses through essential effects on GATA-3 function rather than on IFNG gene acetylation and transcription.* J Exp Med, 2006. **203**(3): p. 755-66.
37. Ho, I.C., T.S. Tai, and S.Y. Pai, *GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation.* Nat Rev Immunol, 2009. **9**(2): p. 125-35.

38. Zheng, W. and R.A. Flavell, *The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells.* Cell, 1997. **89**(4): p. 587-96.
39. Zhang, D.H., et al., *Transcription factor GATA-3 is differentially expressed in murine Th1 and Th2 cells and controls Th2-specific expression of the interleukin-5 gene.* J Biol Chem, 1997. **272**(34): p. 21597-603.
40. Ouyang, W., et al., *Inhibition of Th1 development mediated by GATA-3 through an IL-4-independent mechanism.* Immunity, 1998. **9**(5): p. 745-55.
41. Zhu, J., et al., *Conditional deletion of Gata3 shows its essential function in T(H)1-T(H)2 responses.* Nat Immunol, 2004. **5**(11): p. 1157-65.
42. Pai, S.Y., M.L. Truitt, and I.C. Ho, *GATA-3 deficiency abrogates the development and maintenance of T helper type 2 cells.* Proc Natl Acad Sci U S A, 2004. **101**(7): p. 1993-8.
43. Ouyang, W., et al., *Stat6-independent GATA-3 autoactivation directs IL-4-independent Th2 development and commitment.* Immunity, 2000. **12**(1): p. 27-37.
44. Seki, N., et al., *IL-4-induced GATA-3 expression is a time-restricted instruction switch for Th2 cell differentiation.* J Immunol, 2004. **172**(10): p. 6158-66.
45. Lee, H.J., et al., *GATA-3 induces T helper cell type 2 (Th2) cytokine expression and chromatin remodeling in committed Th1 cells.* J Exp Med, 2000. **192**(1): p. 105-15.
46. Ivanov, II, et al., *The orphan nuclear receptor RORgammat directs the differentiation program of proinflammatory IL-17+ T helper cells.* Cell, 2006. **126**(6): p. 1121-33.
47. Sun, Z., et al., *Requirement for RORgamma in thymocyte survival and lymphoid organ development.* Science, 2000. **288**(5475): p. 2369-73.
48. Yang, X.O., et al., *T helper 17 lineage differentiation is programmed by orphan nuclear receptors ROR alpha and ROR gamma.* Immunity, 2008. **28**(1): p. 29-39.
49. Sakaguchi, S., *Naturally arising Foxp3-expressing CD25+CD4+ regulatory T cells in immunological tolerance to self and non-self.* Nat Immunol, 2005. **6**(4): p. 345-52.
50. Hori, S., T. Nomura, and S. Sakaguchi, *Control of regulatory T cell development by the transcription factor Foxp3.* Science, 2003. **299**(5609): p. 1057-61.
51. Williams, L.M. and A.Y. Rudensky, *Maintenance of the Foxp3-dependent developmental program in mature regulatory T cells requires continued expression of Foxp3.* Nat Immunol, 2007. **8**(3): p. 277-84.
52. Fontenot, J.D., M.A. Gavin, and A.Y. Rudensky, *Foxp3 programs the development and function of CD4+CD25+ regulatory T cells.* Nat Immunol, 2003. **4**(4): p. 330-6.
53. Mukasa, R., et al., *Epigenetic instability of cytokine and transcription factor gene loci underlies plasticity of the T helper 17 cell lineage.* Immunity, 2010. **32**(5): p. 616-27.
54. Miyara, M., et al., *Functional delineation and differentiation dynamics of human CD4+ T cells expressing the FoxP3 transcription factor.* Immunity, 2009. **30**(6): p. 899-911.

55. Zhou, X., et al., *Instability of the transcription factor Foxp3 leads to the generation of pathogenic memory T cells in vivo.* Nat Immunol, 2009. **10**(9): p. 1000-7.
56. Acosta-Rodriguez, E.V., et al., *Surface phenotype and antigenic specificity of human interleukin 17-producing T helper memory cells.* Nat Immunol, 2007. **8**(6): p. 639-46.
57. Yin, Z., et al., *The elevated ratio of interferon gamma-/interleukin-4-positive T cells found in synovial fluid and synovial membrane of rheumatoid arthritis patients can be changed by interleukin-4 but not by interleukin-10 or transforming growth factor beta.* Rheumatology (Oxford), 1999. **38**(11): p. 1058-67.
58. Nistala, K., et al., *Th17 plasticity in human autoimmune arthritis is driven by the inflammatory environment.* Proc Natl Acad Sci U S A, 2010. **107**(33): p. 14751-6.
59. Hegazy, A.N., et al., *Interferons direct Th2 cell reprogramming to generate a stable GATA-3(+)T-bet(+) cell subset with combined Th2 and Th1 cell functions.* Immunity, 2010. **32**(1): p. 116-28.
60. Koch, M.A., et al., *The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation.* Nat Immunol, 2009. **10**(6): p. 595-602.
61. Xu, L., et al., *Cutting edge: regulatory T cells induce CD4+CD25-Foxp3- T cells or are self-induced to become Th17 cells in the absence of exogenous TGF-beta.* J Immunol, 2007. **178**(11): p. 6725-9.
62. Hwang, E.S., et al., *T helper cell fate specified by kinase-mediated interaction of T-bet with GATA-3.* Science, 2005. **307**(5708): p. 430-3.
63. Zhou, L., et al., *TGF-beta-induced Foxp3 inhibits T(H)17 cell differentiation by antagonizing RORgammat function.* Nature, 2008. **453**(7192): p. 236-40.
64. Lighvani, A.A., et al., *T-bet is rapidly induced by interferon-gamma in lymphoid and myeloid cells.* Proc Natl Acad Sci U S A, 2001. **98**(26): p. 15137-42.
65. Afkarian, M., et al., *T-bet is a STAT1-induced regulator of IL-12R expression in naive CD4+ T cells.* Nat Immunol, 2002. **3**(6): p. 549-57.
66. Lieberman, L.A., et al., *STAT1 plays a critical role in the regulation of antimicrobial effector mechanisms, but not in the development of Th1-type responses during toxoplasmosis.* J Immunol, 2004. **172**(1): p. 457-63.
67. Zhou, L., et al., *IL-6 programs T(H)-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways.* Nat Immunol, 2007. **8**(9): p. 967-74.
68. Nurieva, R., et al., *Essential autocrine regulation by IL-21 in the generation of inflammatory T cells.* Nature, 2007. **448**(7152): p. 480-3.
69. Korn, T., et al., *IL-21 initiates an alternative pathway to induce proinflammatory T(H)17 cells.* Nature, 2007. **448**(7152): p. 484-7.
70. Bettelli, E., et al., *Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells.* Nature, 2006. **441**(7090): p. 235-8.
71. Harris, T.J., et al., *Cutting edge: An in vivo requirement for STAT3 signaling in TH17 development and TH17-dependent autoimmunity.* J Immunol, 2007. **179**(7): p. 4313-7.

72. Mathur, A.N., et al., *Stat3 and Stat4 direct development of IL-17-secreting Th cells.* J Immunol, 2007. **178**(8): p. 4901-7.
73. Yang, X.O., et al., *STAT3 regulates cytokine-mediated generation of inflammatory helper T cells.* J Biol Chem, 2007. **282**(13): p. 9358-63.
74. Thierfelder, W.E., et al., *Requirement for Stat4 in interleukin-12-mediated responses of natural killer and T cells.* Nature, 1996. **382**(6587): p. 171-4.
75. Cai, G., et al., *Identification of STAT4-dependent and independent mechanisms of resistance to Toxoplasma gondii.* J Immunol, 2000. **165**(5): p. 2619-27.
76. Usui, T., et al., *GATA-3 suppresses Th1 development by downregulation of Stat4 and not through effects on IL-12Rbeta2 chain or T-bet.* Immunity, 2003. **18**(3): p. 415-28.
77. Moriggl, R., et al., *Stat5 is required for IL-2-induced cell cycle progression of peripheral T cells.* Immunity, 1999. **10**(2): p. 249-59.
78. Cote-Sierra, J., et al., *Interleukin 2 plays a central role in Th2 differentiation.* Proc Natl Acad Sci U S A, 2004. **101**(11): p. 3880-5.
79. Zhu, J., et al., *Stat5 activation plays a critical role in Th2 differentiation.* Immunity, 2003. **19**(5): p. 739-48.
80. Burchill, M.A., et al., *IL-2 receptor beta-dependent STAT5 activation is required for the development of Foxp3+ regulatory T cells.* J Immunol, 2007. **178**(1): p. 280-90.
81. Burchill, M.A., et al., *Linked T cell receptor and cytokine signaling govern the development of the regulatory T cell repertoire.* Immunity, 2008. **28**(1): p. 112-21.
82. Laurence, A., et al., *Interleukin-2 signaling via STAT5 constrains T helper 17 cell generation.* Immunity, 2007. **26**(3): p. 371-81.
83. Kaplan, M.H., et al., *Stat6 is required for mediating responses to IL-4 and for development of Th2 cells.* Immunity, 1996. **4**(3): p. 313-9.
84. Takeda, K., et al., *Essential role of Stat6 in IL-4 signalling.* Nature, 1996. **380**(6575): p. 627-30.
85. Jankovic, D., et al., *Single cell analysis reveals that IL-4 receptor/Stat6 signaling is not required for the in vivo or in vitro development of CD4+ lymphocytes with a Th2 cytokine profile.* J Immunol, 2000. **164**(6): p. 3047-55.
86. Wei, L., et al., *Discrete roles of STAT4 and STAT6 transcription factors in tuning epigenetic modifications and transcription during T helper cell differentiation.* Immunity, 2010. **32**(6): p. 840-51.
87. Yang, X.P., et al., *Opposing regulation of the locus encoding IL-17 through direct, reciprocal actions of STAT3 and STAT5.* Nat Immunol, 2011. **12**(3): p. 247-54.
88. Intlekofer, A.M., et al., *Anomalous type 17 response to viral infection by CD8+ T cells lacking T-bet and eomesodermin.* Science, 2008. **321**(5887): p. 408-11.
89. Suto, A., et al., *IL-21 inhibits IFN-gamma production in developing Th1 cells through the repression of Eomesodermin expression.* J Immunol, 2006. **177**(6): p. 3721-7.
90. Djuretic, I.M., et al., *Transcription factors T-bet and Runx3 cooperate to activate Ifng and silence Il4 in T helper type 1 cells.* Nat Immunol, 2007. **8**(2): p. 145-53.
91. Naoe, Y., et al., *Repression of interleukin-4 in T helper type 1 cells by Runx/Cbf beta binding to the Il4 silencer.* J Exp Med, 2007. **204**(8): p. 1749-55.

92. Cruz-Guilloty, F., et al., *Runx3 and T-box proteins cooperate to establish the transcriptional program of effector CTLs.* J Exp Med, 2009. **206**(1): p. 51-9.
93. Kitoh, A., et al., *Indispensable role of the Runx1-Cbfbeta transcription complex for in vivo-suppressive function of FoxP3+ regulatory T cells.* Immunity, 2009. **31**(4): p. 609-20.
94. Rudra, D., et al., *Runx-CBFbeta complexes control expression of the transcription factor Foxp3 in regulatory T cells.* Nat Immunol, 2009. **10**(11): p. 1170-7.
95. Zhang, F., G. Meng, and W. Strober, *Interactions among the transcription factors Runx1, RORgammat and Foxp3 regulate the differentiation of interleukin 17-producing T cells.* Nat Immunol, 2008. **9**(11): p. 1297-306.
96. Lohoff, M., et al., *Dysregulated T helper cell differentiation in the absence of interferon regulatory factor 4.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11808-12.
97. Brustle, A., et al., *The development of inflammatory T(H)-17 cells requires interferon-regulatory factor 4.* Nat Immunol, 2007. **8**(9): p. 958-66.
98. Zheng, Y., et al., *Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T(H)2 responses.* Nature, 2009. **458**(7236): p. 351-6.
99. Kano, S., et al., *The contribution of transcription factor IRF1 to the interferon-gamma-interleukin 12 signaling axis and TH1 versus TH-17 differentiation of CD4+ T cells.* Nat Immunol, 2008. **9**(1): p. 34-41.
100. Zhu, J., et al., *Growth factor independent-1 induced by IL-4 regulates Th2 cell proliferation.* Immunity, 2002. **16**(5): p. 733-44.
101. Zhu, J., et al., *Gfi-1 plays an important role in IL-2-mediated Th2 cell expansion.* Proc Natl Acad Sci U S A, 2006. **103**(48): p. 18214-9.
102. Zhu, J., et al., *Down-regulation of Gfi-1 expression by TGF-beta is important for differentiation of Th17 and CD103+ inducible regulatory T cells.* J Exp Med, 2009. **206**(2): p. 329-41.
103. Kim, J.I., et al., *The transcription factor c-Maf controls the production of interleukin-4 but not other Th2 cytokines.* Immunity, 1999. **10**(6): p. 745-51.
104. Li, B., et al., *Regulation of IL-4 expression by the transcription factor JunB during T helper cell differentiation.* EMBO J, 1999. **18**(2): p. 420-32.
105. Quirion, M.R., et al., *Cutting edge: Ikaros is a regulator of Th2 cell differentiation.* J Immunol, 2009. **182**(2): p. 741-5.
106. Thornton, A.M., et al., *Expression of Helios, an Ikaros transcription factor family member, differentiates thymic-derived from peripherally induced Foxp3+ T regulatory cells.* J Immunol, 2010. **184**(7): p. 3433-41.
107. Wei, G., et al., *Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells.* Immunity, 2009. **30**(1): p. 155-67.
108. Schmidl, C., et al., *Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity.* Genome Res, 2009. **19**(7): p. 1165-74.
109. Thurman, R.E., et al., *The accessible chromatin landscape of the human genome.* Nature, 2012. **in press**.
110. Groudine, M. and H. Weintraub, *Propagation of globin DNAase I-hypersensitive sites in absence of factors required for induction: a possible mechanism for determination.* Cell, 1982. **30**(1): p. 131-9.

111. Bird, A., *DNA methylation patterns and epigenetic memory.* Genes Dev, 2002. **16**(1): p. 6-21.
112. Pfeifer, G.P., et al., *Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability.* Proc Natl Acad Sci U S A, 1990. **87**(21): p. 8252-6.
113. Suganuma, T. and J.L. Workman, *Signals and combinatorial functions of histone modifications.* Annu Rev Biochem, 2011. **80**: p. 473-99.
114. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.
115. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.
116. Weintraub, H. and M. Groudine, *Chromosomal subunits in active genes have an altered conformation.* Science, 1976. **193**(4256): p. 848-56.
117. Felsenfeld, G., et al., *Chromatin structure and gene expression.* Proc Natl Acad Sci U S A, 1996. **93**(18): p. 9384-8.
118. Keene, M.A., et al., *DNase I hypersensitive sites in Drosophila chromatin occur at the 5' ends of regions of transcription.* Proc Natl Acad Sci U S A, 1981. **78**(1): p. 143-6.
119. Wu, C., Y.C. Wong, and S.C. Elgin, *The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity.* Cell, 1979. **16**(4): p. 807-14.
120. Bulger, M. and M. Groudine, *Enhancers: the abundance and function of regulatory sequences beyond promoters.* Dev Biol, 2010. **339**(2): p. 250-7.
121. McGhee, J.D., et al., *A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion.* Cell, 1981. **27**(1 Pt 2): p. 45-55.
122. Petrykowska, H.M., C.M. Vockley, and L. Elnitski, *Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus.* Genome Res, 2008. **18**(8): p. 1238-46.
123. Gaszner, M. and G. Felsenfeld, *Insulators: exploiting transcriptional and epigenetic mechanisms.* Nat Rev Genet, 2006. **7**(9): p. 703-13.
124. Li, Q., S. Harju, and K.R. Peterson, *Locus control regions: coming of age at a decade plus.* Trends Genet, 1999. **15**(10): p. 403-8.
125. Grosveld, F., et al., *Position-independent, high-level expression of the human beta-globin gene in transgenic mice.* Cell, 1987. **51**(6): p. 975-85.
126. John, S., et al., *Chromatin accessibility pre-determines glucocorticoid receptor binding patterns.* Nat Genet, 2011. **43**(3): p. 264-8.
127. Hatton, R.D., et al., *A distal conserved sequence element controls Ifng gene expression by T cells and NK cells.* Immunity, 2006. **25**(5): p. 717-29.
128. Schoenborn, J.R., et al., *Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-gamma.* Nat Immunol, 2007. **8**(7): p. 732-42.
129. Ansel, K.M., et al., *Regulation of Th2 differentiation and Il4 locus accessibility.* Annu Rev Immunol, 2006. **24**: p. 607-56.

130. Akimzhanov, A.M., X.O. Yang, and C. Dong, *Chromatin remodeling of interleukin-17 (IL-17)-IL-17F cytokine gene locus during inflammatory helper T cell differentiation.* J Biol Chem, 2007. **282**(9): p. 5969-72.
131. Mercer, E.M., et al., *Multilineage priming of enhancer repertoires precedes commitment to the B and myeloid cell lineages in hematopoietic progenitors.* Immunity, 2011. **35**(3): p. 413-25.
132. Miller, S.A., S.E. Mohn, and A.S. Weinmann, *Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression.* Mol Cell, 2010. **40**(4): p. 594-605.
133. Samstein, R.M., et al., *Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification.* Cell, 2012. **151**(1): p. 153-66.
134. Vahedi, G., et al., *STATs Shape the Active Enhancer Landscape of T Cell Populations.* Cell, 2012. **151**(5): p. 981-93.
135. Boyer, L.A., et al., *Core transcriptional regulatory circuitry in human embryonic stem cells.* Cell, 2005. **122**(6): p. 947-56.
136. Kim, J., et al., *An extended transcriptional network for pluripotency of embryonic stem cells.* Cell, 2008. **132**(6): p. 1049-61.
137. Davidson, E.H., et al., *A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo.* Dev Biol, 2002. **246**(1): p. 162-90.
138. Yun, K. and B. Wold, *Skeletal muscle determination and differentiation: story of a core regulatory network and its context.* Curr Opin Cell Biol, 1996. **8**(6): p. 877-89.
139. Alon, U., *An introduction to systems biology : design principles of biological circuits*. Chapman & Hall/CRC mathematical and computational biology series. 2007, Boca Raton, FL: Chapman & Hall/CRC. xvi, 301 p., 4 p. of plates.
140. Jenner, R.G., et al., *The transcription factors T-bet and GATA-3 control alternative pathways of T-cell differentiation through a shared set of target genes.* Proc Natl Acad Sci U S A, 2009. **106**(42): p. 17876-81.
141. O'Shea, J.J., et al., *Genomic views of STAT function in CD4+ T helper cell differentiation.* Nat Rev Immunol, 2011. **11**(4): p. 239-50.
142. Ciofani, M., et al., *A validated regulatory network for Th17 cell specification.* Cell, 2012. **151**(2): p. 289-303.
143. Basso, K., et al., *Reverse engineering of regulatory networks in human B cells.* Nat Genet, 2005. **37**(4): p. 382-90.
144. Yosef, N., et al., *Dynamic regulatory network controlling TH17 cell differentiation.* Nature, 2013. **496**(7446): p. 461-8.
145. Neph, S., et al., *An expansive human regulatory lexicon encoded in transcription factor footprints.* Nature, 2012. **In Press**.
146. Neph, S., et al., *Circuitry and dynamics of human transcription factor regulatory networks.* Cell, 2012. **150**(6): p. 1274-86.
147. Neph, S., et al., *An expansive human regulatory lexicon encoded in transcription factor footprints.* Nature, 2012. **489**(7414): p. 83-90.
148. Streeck, H., et al., *Rapid ex vivo isolation and long-term culture of human Th17 cells.* J Immunol Methods, 2008. **333**(1-2): p. 115-25.
149. Hesselberth, J.R., et al., *Global mapping of protein-DNA interactions in vivo by digital genomic footprinting.* Nat Methods, 2009. **6**(4): p. 283-9.

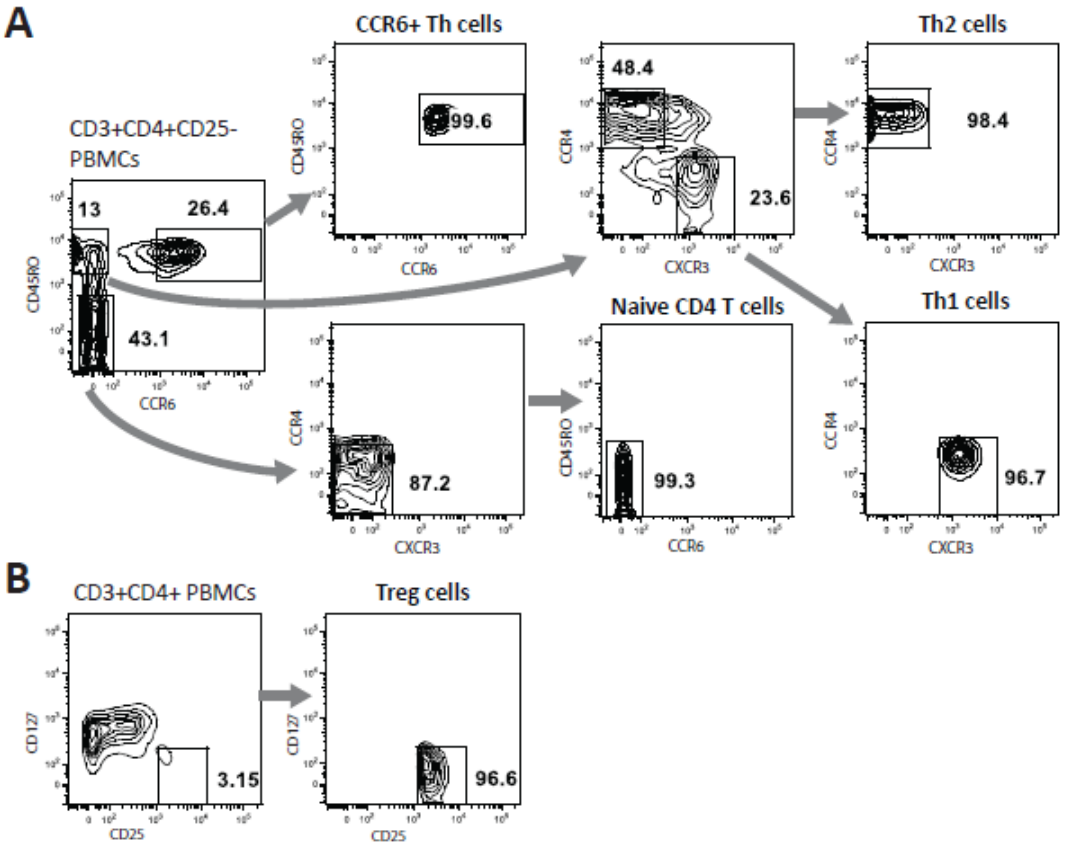
150. Sabo, P.J., et al., *Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays.* Nat Methods, 2006. **3**(7): p. 511-8.
151. Neph, S., et al., *BEDOPS: High performance genomic feature operations.* Bioinformatics, 2012.
152. Thomas, S., et al., *Dynamic reprogramming of chromatin accessibility during Drosophila embryo development.* Genome Biol, 2011. **12**(5): p. R43.
153. Paige, S.L., et al., *A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development.* Cell, 2012. **151**(1): p. 221-32.
154. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W202-8.
155. McLean, C.Y., et al., *GREAT improves functional interpretation of cis-regulatory regions.* Nat Biotechnol, 2010. **28**(5): p. 495-501.
156. Bryne, J.C., et al., *JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update.* Nucleic Acids Res, 2008. **36**(Database issue): p. D102-6.
157. Wingender, E., et al., *TRANSFAC: a database on transcription factors and their DNA binding sites.* Nucleic Acids Res, 1996. **24**(1): p. 238-41.
158. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res, 2004. **14**(6): p. 1188-90.
159. Rebhan, M., et al., *GeneCards: integrating information about genes, proteins and diseases.* Trends Genet, 1997. **13**(4): p. 163.
160. Magrane, M. and U. Consortium, *UniProt Knowledgebase: a hub of integrated protein data.* Database (Oxford), 2011. **2011**: p. bar009.
161. Neph, S., et al., *Circuitry and dynamics of human transcription factor regulatory networks.* Cell, 2012. **In Press**.
162. Krzywinski, M., et al., *Hive plots--rational approach to visualizing networks.* Brief Bioinform, 2011.
163. Smoot, M.E., et al., *Cytoscape 2.8: new features for data integration and network visualization.* Bioinformatics, 2011. **27**(3): p. 431-2.
164. Milo, R., et al., *Superfamilies of evolved and designed networks.* Science, 2004. **303**(5663): p. 1538-42.
165. Roberts, A., et al., *Identification of novel transcripts in annotated genomes using RNA-Seq.* Bioinformatics, 2011. **27**(17): p. 2325-9.
166. Lee, T.I., et al., *Transcriptional regulatory networks in Saccharomyces cerevisiae.* Science, 2002. **298**(5594): p. 799-804.
167. Chattopadhyay, S., et al., *Biochemical and functional analyses of chromatin changes at the TCR-beta gene locus during CD4-CD8- to CD4+CD8+ thymocyte differentiation.* J Immunol, 1998. **160**(3): p. 1256-67.
168. Ortiz, B.D., D. Cado, and A. Winoto, *A new element within the T-cell receptor alpha locus required for tissue-specific locus control region activity.* Mol Cell Biol, 1999. **19**(3): p. 1901-9.
169. Forrester, W.C., et al., *Dependence of enhancer-mediated transcription of the immunoglobulin mu gene on nuclear matrix attachment regions.* Science, 1994. **265**(5176): p. 1221-5.
170. Mills, F.C., et al., *DNase I hypersensitive sites in the chromatin of human mu immunoglobulin heavy-chain genes.* Nature, 1983. **306**(5945): p. 809-12.

171. Banerji, J., L. Olson, and W. Schaffner, *A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes.* Cell, 1983. **33**(3): p. 729-40.
172. Fields, P.E., et al., *Th2-specific chromatin remodeling and enhancer activity in the Th2 cytokine locus control region.* Immunity, 2004. **21**(6): p. 865-76.
173. Gross, D.S. and W.T. Garrard, *Nuclease hypersensitive sites in chromatin.* Annu Rev Biochem, 1988. **57**: p. 159-97.
174. Thurman, R.E., et al., *The accessible chromatin landscape of the human genome.* Nature, 2012. **489**(7414): p. 75-82.
175. Davidson, E.H., et al., *A genomic regulatory network for development.* Science, 2002. **295**(5560): p. 1669-78.
176. Hawkins, R.D., et al., *Global Chromatin State Analysis Reveals Lineage-Specific Enhancers during the Initiation of Human T helper 1 and T helper 2 Cell Polarization.* Immunity, 2013.
177. Sekimata, M., et al., *CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus.* Immunity, 2009. **31**(4): p. 551-64.
178. Muralidharan, S., et al., *Activation of Wnt signaling arrests effector differentiation in human peripheral and cord blood-derived T lymphocytes.* J Immunol, 2011. **187**(10): p. 5221-32.
179. Inamori, K., et al., *Identification of homeobox genes expressed in human T-lymphocytes.* Biochem Biophys Res Commun, 1993. **196**(1): p. 203-8.
180. Lekstrom-Himes, J. and K.G. Xanthopoulos, *Biological role of the CCAAT/enhancer-binding protein family of transcription factors.* J Biol Chem, 1998. **273**(44): p. 28545-8.
181. Jetten, A.M., S. Kurebayashi, and E. Ueda, *The ROR nuclear orphan receptor subfamily: critical regulators of multiple biological processes.* Prog Nucleic Acid Res Mol Biol, 2001. **69**: p. 205-47.
182. Wang, L., et al., *Distinct functions for the transcription factors GATA-3 and ThPOK during intrathymic differentiation of CD4(+) T cells.* Nat Immunol, 2008. **9**(10): p. 1122-30.
183. Samstein, R.M., et al., *Foxp3 exploits a preexistent enhancer landscape for regulatory T cell lineage specification.* Cell, 2012. **In Press**.
184. Lin, L., et al., *Modulation of Th1 activation and inflammation by the NF-kappaB repressor Foxj1.* Science, 2004. **303**(5660): p. 1017-20.
185. Ehret, G.B., et al., *DNA binding specificity of different STAT proteins. Comparison of in vitro specificity with natural target sites.* J Biol Chem, 2001. **276**(9): p. 6675-88.
186. Voice, J., et al., *c-Maf and JunB mediation of Th2 differentiation induced by the type 2 G protein-coupled receptor (VPAC2) for vasoactive intestinal peptide.* J Immunol, 2004. **172**(12): p. 7289-96.
187. Berger, M.F., et al., *Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences.* Cell, 2008. **133**(7): p. 1266-76.
188. The, et al., *Initial Analysis of the Encyclopedia of DNA Elements in the Human Genome.* Nature, 2012. **In Press**.

189. Alvarez, J.D., et al., *The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development.* Genes Dev, 2000. **14**(5): p. 521-35.
190. Sekiya, T., et al., *The nuclear orphan receptor Nr4a2 induces Foxp3 and regulates differentiation of CD4+ T cells.* Nat Commun, 2011. **2**: p. 269.
191. Willinger, T., et al., *Human naive CD8 T cells down-regulate expression of the WNT pathway transcription factors lymphoid enhancer binding factor 1 and transcription factor 7 (T cell factor-1) following antigen encounter in vitro and in vivo.* J Immunol, 2006. **176**(3): p. 1439-46.
192. Kashiwada, M., et al., *NFIL3/E4BP4 controls type 2 T helper cell cytokine expression.* EMBO J, 2011. **30**(10): p. 2071-82.
193. Klotz, L., et al., *The nuclear receptor PPAR gamma selectively inhibits Th17 differentiation in a T cell-intrinsic fashion and suppresses CNS autoimmunity.* J Exp Med, 2009. **206**(10): p. 2079-89.
194. Davydov, I.V., P.H. Krammer, and M. Li-Weber, *Nuclear factor-IL6 activates the human IL-4 promoter in T cells.* J Immunol, 1995. **155**(11): p. 5273-9.
195. van Dijk, T.B., et al., *A composite C/EBP binding site is essential for the activity of the promoter of the IL-3/IL-5/granulocyte-macrophage colony-stimulating factor receptor beta c gene.* J Immunol, 1999. **163**(5): p. 2674-80.
196. Natsuka, S., et al., *Macrophage differentiation-specific expression of NF-IL6, a transcription factor for interleukin-6.* Blood, 1992. **79**(2): p. 460-6.
197. Greenwel, P., et al., *Tumor necrosis factor alpha inhibits type I collagen synthesis through repressive CCAAT/enhancer-binding proteins.* Mol Cell Biol, 2000. **20**(3): p. 912-8.
198. Alon, N., et al., *Multi-node graphs: a framework for multiplexed biological assays.* J Comput Biol, 2006. **13**(10): p. 1659-72.
199. Kapushesky, M., et al., *Gene Expression Atlas update--a value-added database of microarray and sequencing-based functional genomics experiments.* Nucleic Acids Res, 2012. **40**(Database issue): p. D1077-81.
200. Granelli-Piperno, A., L. Andrus, and R.M. Steinman, *Lymphokine and nonlymphokine mRNA levels in stimulated human T cells. Kinetics, mitogen requirements, and effects of cyclosporin A.* J Exp Med, 1986. **163**(4): p. 922-37.
201. An, J., et al., *Kruppel-like factor 4 (KLF4) directly regulates proliferation in thymocyte development and IL-17 expression during Th17 differentiation.* FASEB J, 2011. **25**(10): p. 3634-45.
202. Stephensen, C.B., A.D. Borowsky, and K.C. Lloyd, *Disruption of Rxra gene in thymocytes and T lymphocytes modestly alters lymphocyte frequencies, proliferation, survival and T helper type 1/type 2 balance.* Immunology, 2007. **121**(4): p. 484-98.
203. De Keersmaecker, K., et al., *Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia.* Nat Genet, 2013. **45**(2): p. 186-90.
204. Lee, Y.K., et al., *Late developmental plasticity in the T helper 17 lineage.* Immunity, 2009. **30**(1): p. 92-107.
205. Lee, Y.K., et al., *Developmental plasticity of Th17 and Treg cells.* Curr Opin Immunol, 2009. **21**(3): p. 274-80.

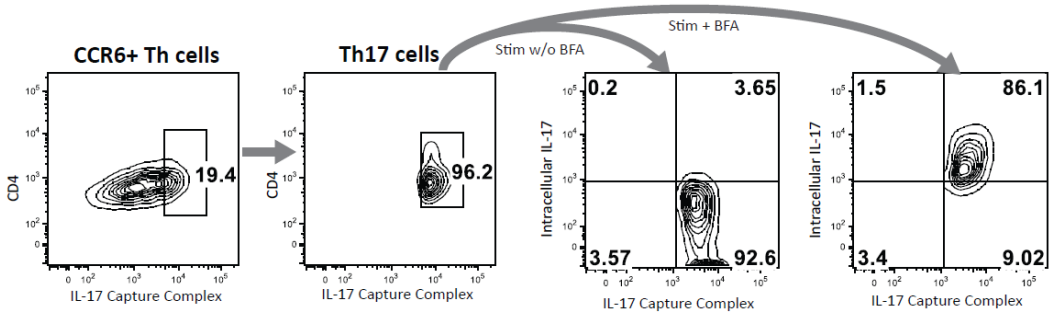
206. Vierstra, J., et al., *Organization and comparative evolutionary dynamics of the mouse regulatory DNA landscape.* Science, 2013. **In Review**.
207. Maurano, M.T., et al., *Systematic localization of common disease-associated variation in regulatory DNA.* Science, 2012. **337**(6099): p. 1190-5.
208. Zheng, Y., et al., *Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate.* Nature, 2010. **463**(7282): p. 808-12.
209. Wang, X., et al., *Transcription of Il17 and Il17f is controlled by conserved noncoding sequence 2.* Immunity, 2012. **36**(1): p. 23-31.

Figure 3.1



Isolation and expansion of CD4+ T cell subsets
(A) *Chemokine receptor sorting for isolation of Naïve, Th1, Th2, and Th17 CD4+ T cells*. Outline of sorting and gating strategy. Grey arrows indicate flow cytometric sorting, after which purity of chemokine receptor expression is verified.
(B) *IL-2R based sorting for isolation of Treg cells*. Outline of sorting and gating strategy. After sorting, cells are verified to be CD25+CD127-.

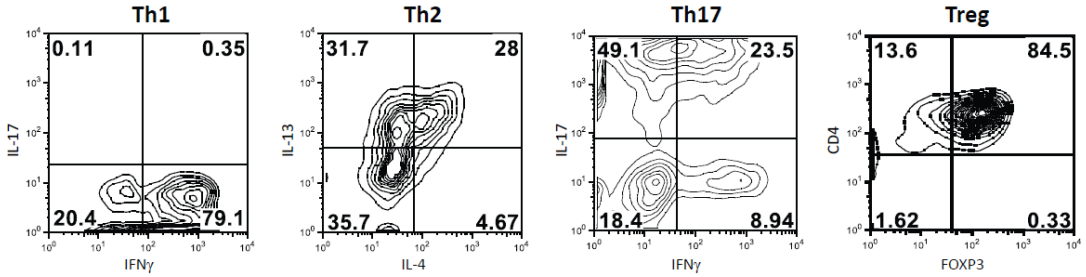
Figure 3.2



IL-17 Capture

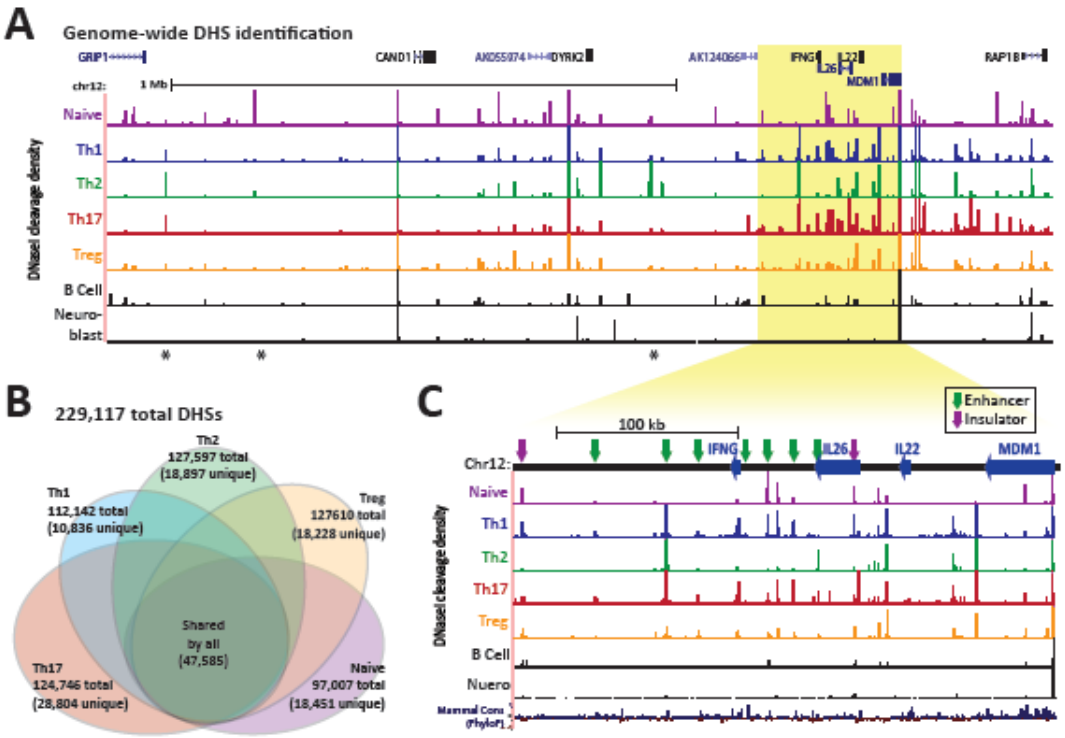
Th17 cells are sequentially stimulated for 90 min, treated with capture complex, incubated an additional 30 min, and stained with IL-17 detection antibody. IL-17+ cells are isolated by flow cytometric sorting. To demonstrate effectiveness of the technique of IL17 capture, cells undergo 4 hours of additional stimulation in the presence or absence of brefeldin A, and are stained for intracellular IL-17. IL-17 captured cells display specific enrichment in IL-17 producing cells. This is not due to secondary detection of surface IL-17, as significant intracellular IL-17 is not detected without treatment with brefeldin A.

Figure 3.3



Production of subset specific factors by T cell subsets
Th1, Th2, and Th17 cells are stimulated with PMA and ionomycin in the presence of brefeldin A. For Th1 cells, IFNγ and IL-17 production is shown. For Th2 cells, IL-4 and IL-13 production is shown. For Th17 cells, IFNγ and IL-17 production are shown. For Treg cells, FoxP3 production is shown along with surface expression of CD4.

Figure 3.4



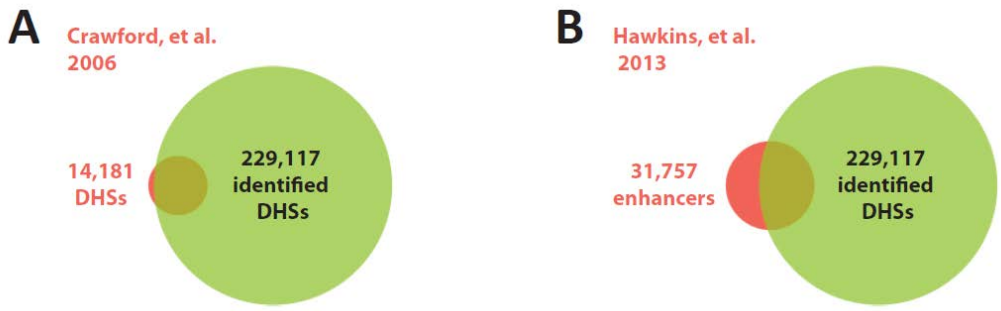
T cell subset specific chromatin accessibility

(A) *Genome-wide profiling of DHSs identifies cell selective regions of chromatin accessibility.* DNaseI cleavage density (mapped cleavages in a 150 bp sliding window) in a two Mb region of chromosome 12 is shown in naïve CD4 T cells, Th1, Th2, Th17, and Treg. Isolated CD20+ B cells and SKNHS neuroblastoma cells are shown for comparison. DNaseI cleavage density is normalized by total number of cleavage events mapped to DHSs genome wide. The area around the *IFNG* gene is highlighted, and example sites with differential sensitivity to DNaseI cleavage are marked with an asterisk.

(B) *CD4 T cell subsets contain shared and unique DHSs.* The total number of identified DHSs in naïve, Th1, Th2, Th17, and Treg CD4+ T cells, and the number that are common or unique to each cell type. Overlaps between subsets are represented in a pseudo-Venn diagram. Sizes of regions are proportional to number of DHSs in subset. Selected subsets of overlapping sites are excluded from diagram for simplicity.

(C) *Regulatory regions in the IFNG-IL26-IL22 locus are demarcated by regions of accessibility.* Top, schematic depicting location of the human *IFNG*, *IL26*, and *IL22* genes. Orthologous locations of previously characterized mouse enhancers are shown as green arrows, while insulators are shown in purple. Bottom, DNaseI cleavage density from indicated cell types and PhyloP conservation score across 23 mammalian species.

Figure 3.5

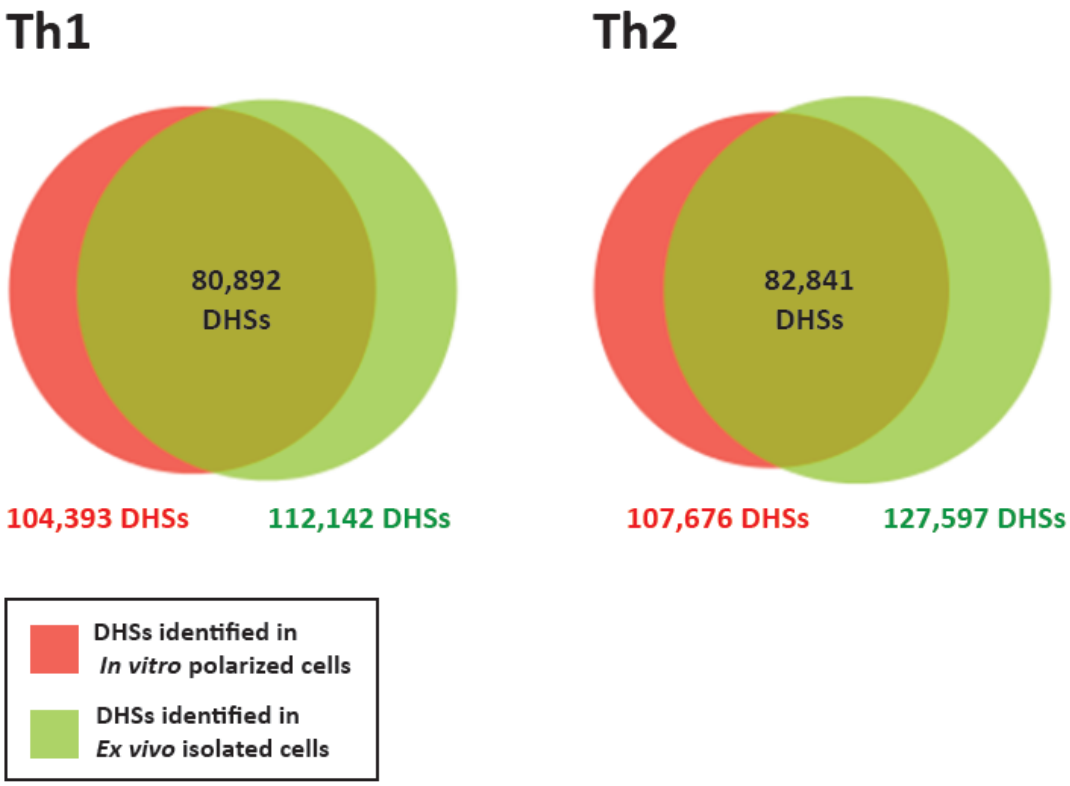


Identification of novel regulatory elements

(a) *A small fraction of t*he *229,117 identified sites to sites were previously identified by Crawford, et al. 2006.* 14,181 out of 14,190 sites identified in Crawford, et al were lifted over into NCBI build 37 (hg19). The sites that overlapped between studies was identified utilizing the bedops[1] tool with the –e -1 flags.

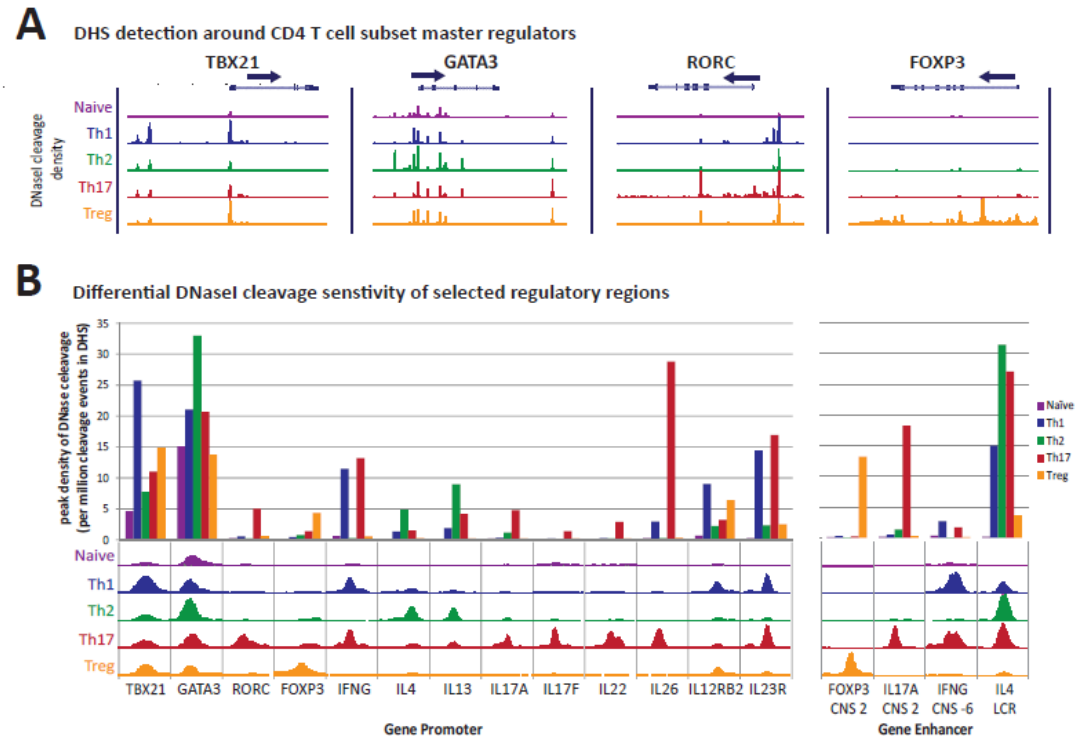
(b) *A small fraction of t*he *229,117 identified sites to sites were previously identified by Hawkins, et al. 2013.* The sites that overlapped between studies was identified utilizing the bedops[1] tool with the –e -1 flags.

Figure 3.6



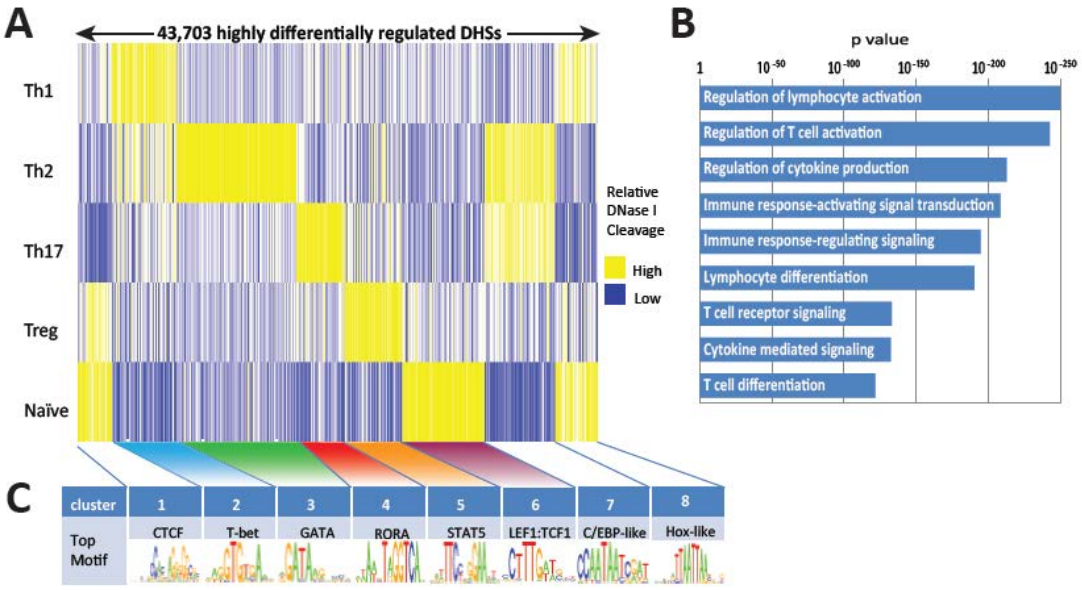
Comparison of DHSs identified in *in vitro* polarized versus *ex vivo* isolated Th1 and Th2 cells. DHSs were identified in naïve CD4 T cells polarized *in vitro* for 8 days and compared to DHSs identified in *ex vivo* isolated CD4 T cells. Overlap of DHSs is depicted by overlap of circles in Venn Diagrams.

Figure 3.7



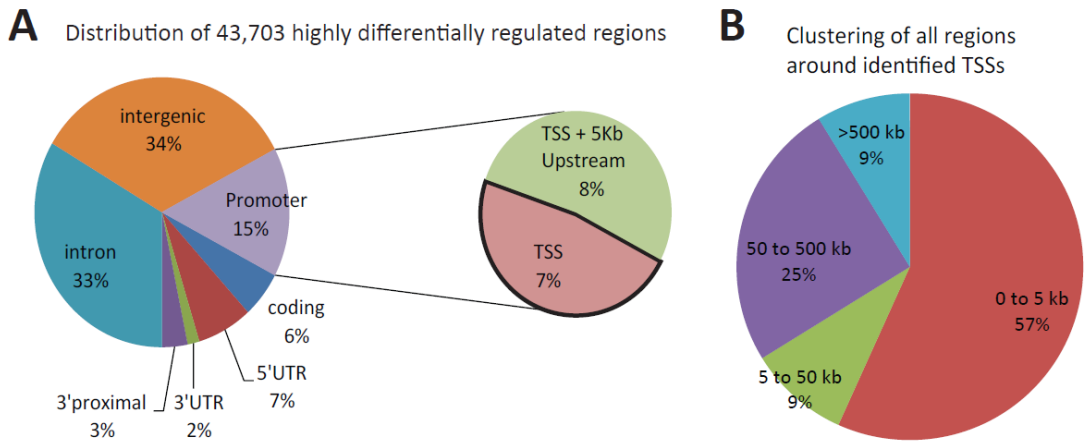
Cell selective patterns of chromatin accessibility
(A) *DHSs in the proximity of master regulatory genes display subset-specific regions of chromatin accessibility.* DHSs are detected as peaks of DNaseI cleavage density at promoters and *cis*-regulatory regions around CD4 T cells subset master regulatory factors.
(B) *Cell-selective patterns of chromatin accessibility mirror known differential transcription*. Top, peak density of DNaseI cleavage at indicated regions. Bottom, density of DNaseI cleavages in and around gene promoters or enhancers. Relative scale is constant.

Figure 3.8



Highly differentially regulated regulatory elements
(A) *Clustering of highly differentially regulated DHSs according to T cell subtype.* DHSs with large variation in relative degree of cleavage show highly significant ($p<10^{-5}$) differences in DNaseI cleavage density between naïve, Th1, Th2, Th17, and Treg CD4+ T cells. A heatmap was created by horizontally tiling the peak DNaseI cleavage densities at each DHS, organized by k means clustering.
(B) *Subset-specific DHSs are strongly associated with CD4 T cell gene expression.* DHSs were mapped to associated genes using GREAT [155]. Binomial raw p-values of region based associations of DHSs with genes annotated with T cell-specific GO biological processes.
(C) *Differentiation and cell-specific DHS clusters are highly enriched for subset-specific regulatory motifs.* The density of TF motifs was calculated over DHSs in each cluster. The ratio of the density of motifs in each cluster versus the density of motifs across all clusters was calculated, and the top scoring motif is shown; in cases where redundant motifs were detected among top two scoring motifs, the motif with highest information content is shown. T cell subset-specific clusters are highlighted, and show enrichment of motifs for known lineage regulating factors.

Figure 3.9

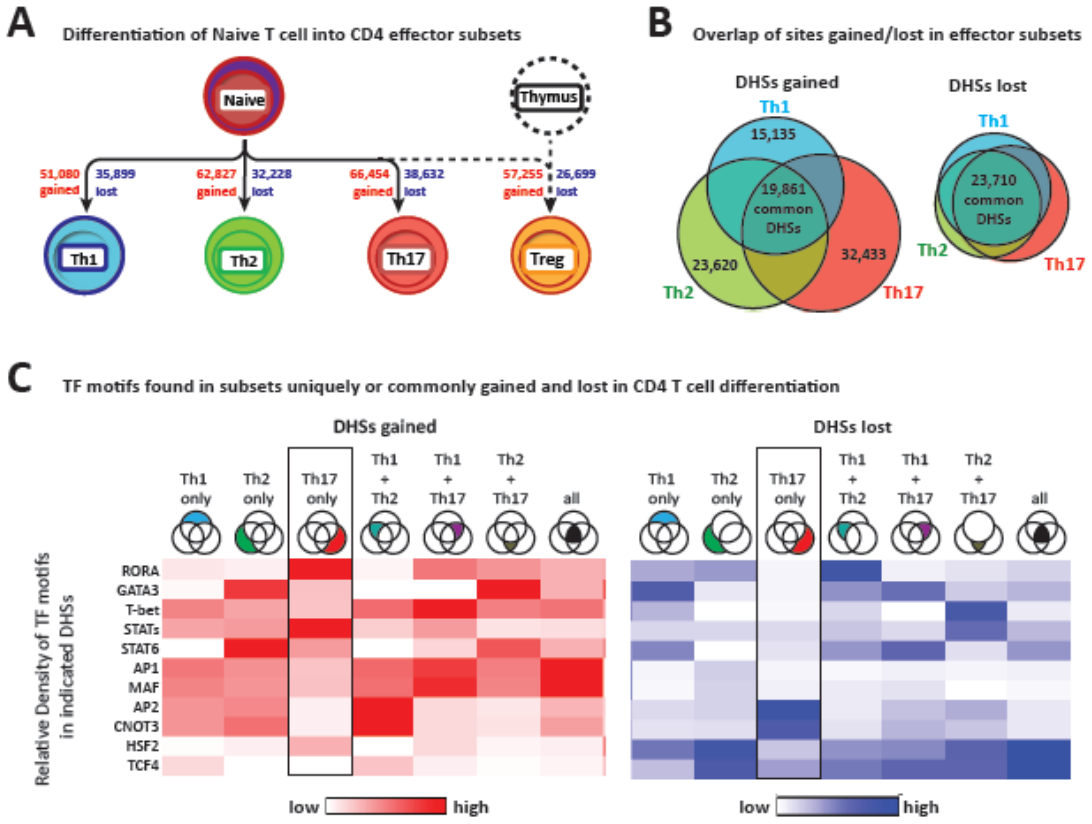


Distribution of regions with highly subset-specific patterns of DNaseI sensitivity

(A) *Distribution of sites relative to known genomic annotations.* Regions identified as having highly subset-specific DNaseI sensitivity were classified according to overlap with Genecode V7 annotations.

(B) *Subset-specific DHSs are primarily clustered around TSSs* . The majority of the 43,703 identified regions are clustered within 5 kb of the 3,218 transcriptional start sites that showed highly differential accessibility.

Figure 3.10



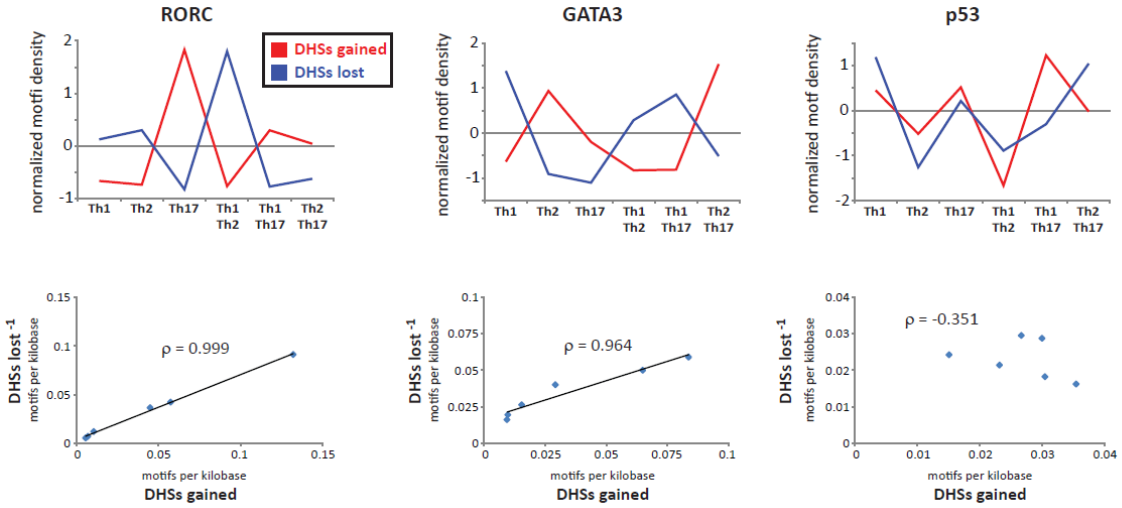
Differential regulatory content of lineage specific DHSs

(A) *DHSs are lost and gained during differentiation of naïve CD4 T cells to T cell effectors*. A schematic of differentiation of naïve T cells to effector lineages. Numbers of DHSs gained and lost in transition from naïve CD4 T cell to various effector subsets is indicated.

(B) *DHSs gained or lost during differentiation are highly restricted to specific T cell subsets.* Venn diagrams of DHSs gained and lost in the transition from naïve CD4 T cells to effector subtypes. Overlap of circles is proportional to number of overlapping of DHSs.

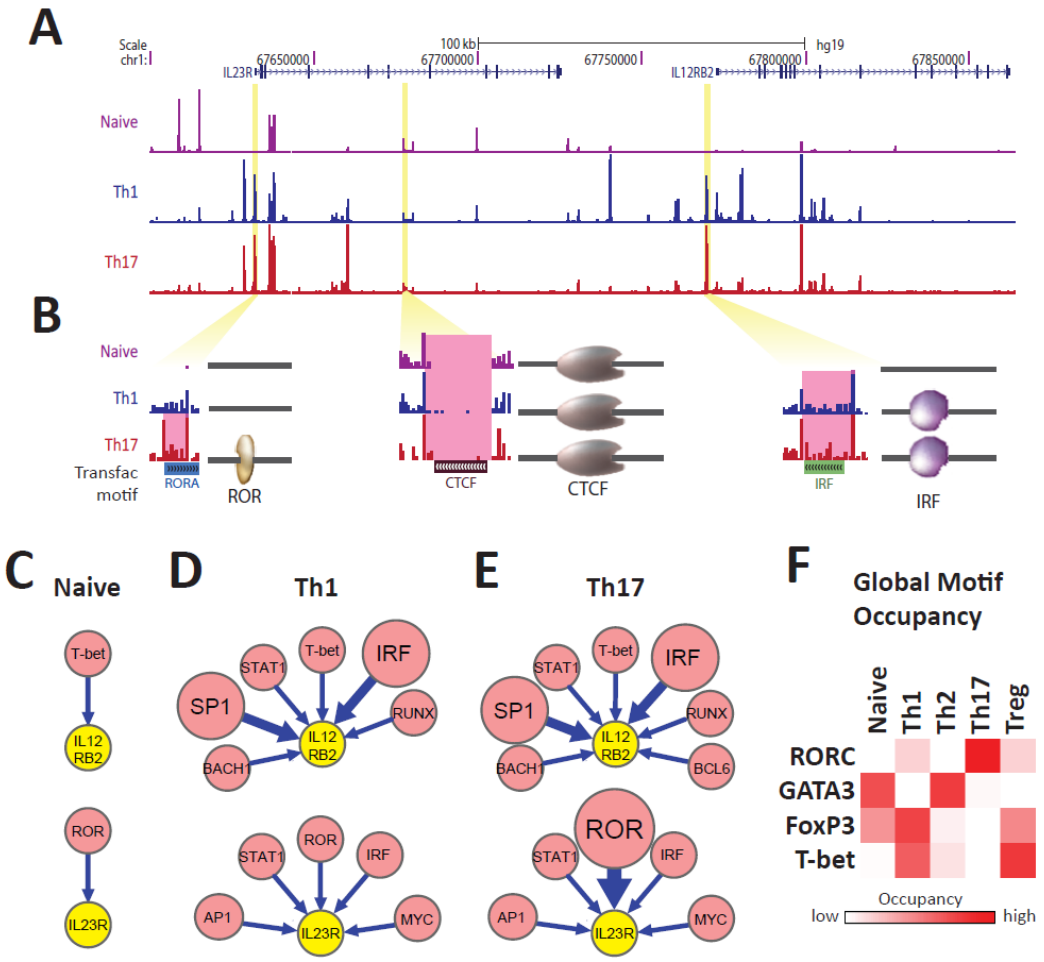
(C) *Distribution of multiple motifs across all T cell subsets enumerates subset-specific confinement of regulatory motifs.* Heatmap of the normalized motif density across all subsets for each selected transfac motifs and STAT6-specfic as well as a generic STAT motif described in Wei, Vahedi [86]. A higher density is represented graphically as increased color intensity. Total DHSs were divided based on their specific presence in indicated subsets. DHSs gained and lost uniquely in Th17 cells are highlighted for reference.

Figure 3.11



RORA and GATA3 motifs enrichment in subsets of gained DHSs are closely mirrored in DHSs lost in reciprocal subsets.
Above, the normalized density of RORA, GATA3, and p53 motifs in indicated subsets of DHSs. Below, the density of motifs in each subset of gained DHSs is plotted against the density of motifs in the reciprocal subset of lost DHSs (pairs of reciprocal subset are Th1:Th2+Th17, Th2:Th1+Th17, and Th17:Th1+Th2) and the Pearson correlation (ρ) is shown.

Figure 3.12



Genomic profiling of TF occupancy

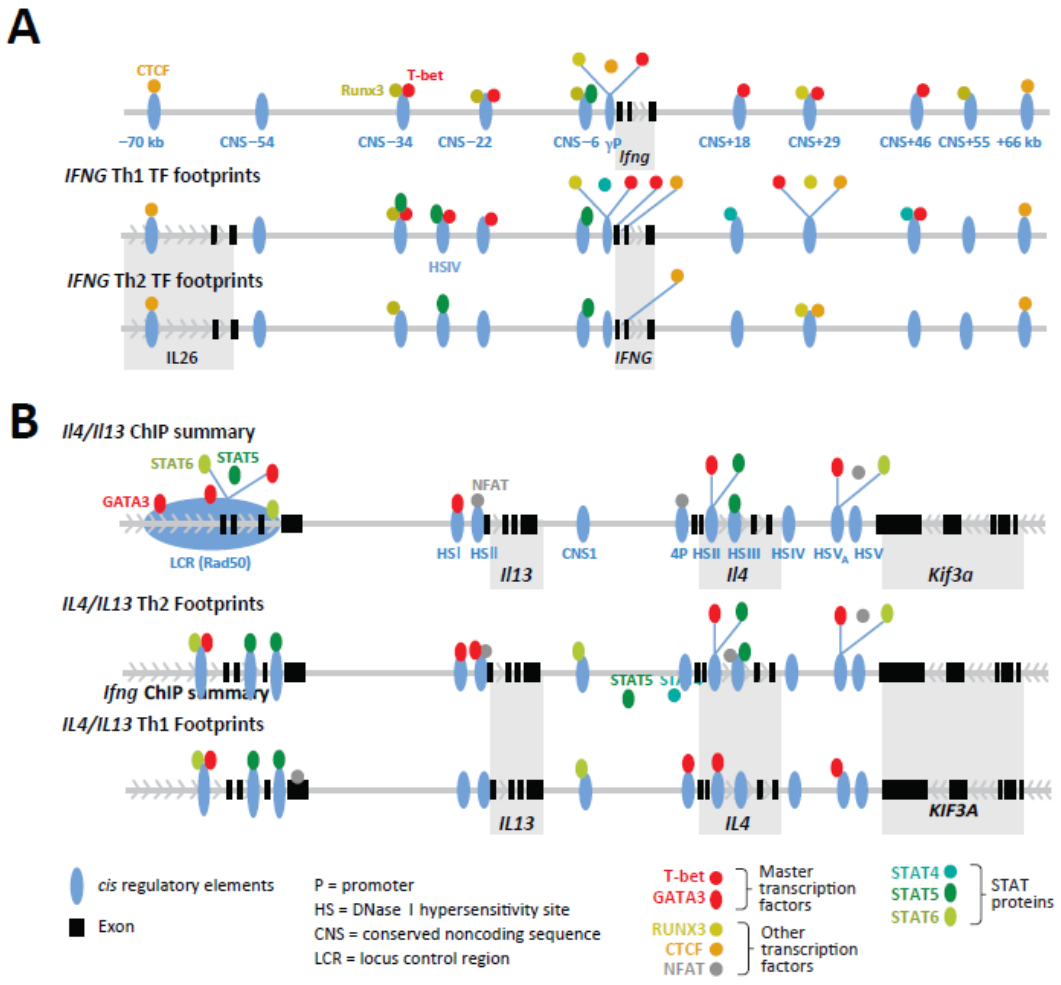
(A) *DNaseI identification of regulatory regions in IL23R-IL12RB2 locus*. Chromatin accessibility from human naïve, Th1, and Th17 CD4 T cells. DHSs correspond with peaks in the DNaseI cleavage density profile.

(B) *Example regulator regions harboring clearly defined DNaseI footprints*. Above, per nucleotide DNaseI cleavage in selected regulatory regions. Below, motifs detected within the region. A footprint overlapping a motifs indicates binding of the factor. Shown to the right of each region is a schematic view of the DNA protein interaction which results in the identified footprints.

(C-E) *Summary of detected interactions in regulatory regions for IL12RB2 and IL23R*. TF footprints are represented as nodes with directional connection indicating binding to the gene regulatory region. The size of the node is proportional to the number of TF footprints detected.

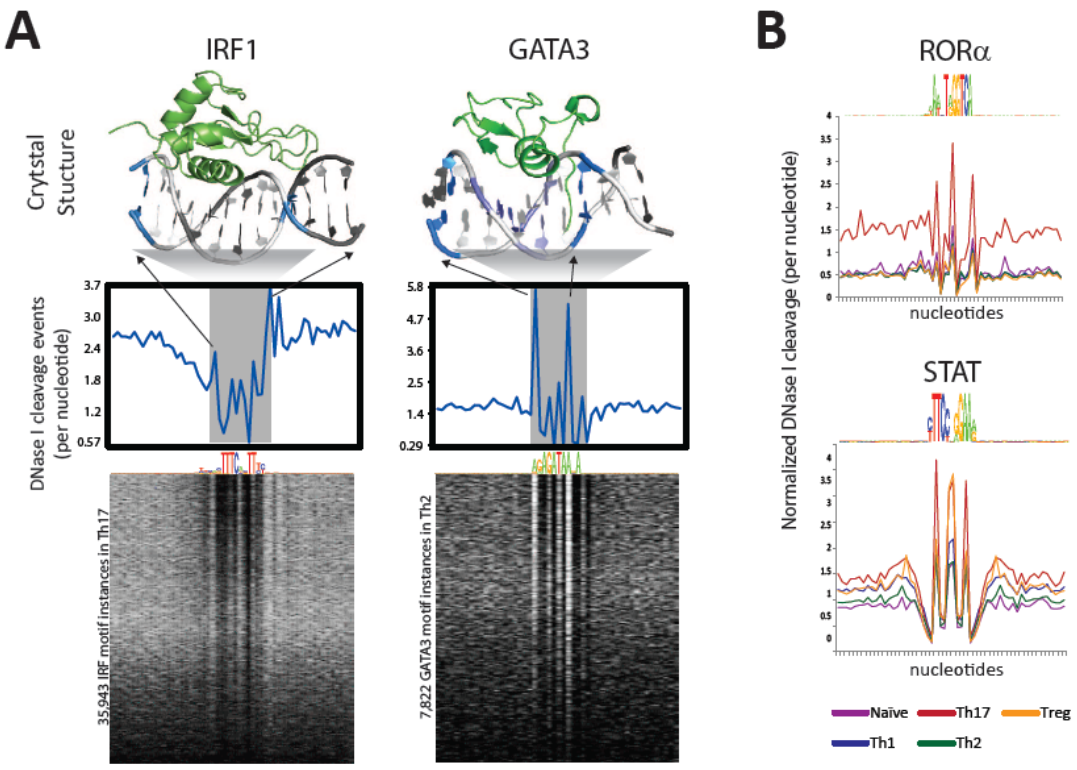
(F) *Lineage-specific differences in regulatory factor occupancy.* Occupancy of recognition sequences for major regulators is calculated as the percentage of motifs in DHSs which overlap footprints in each CD4 T cell subsets. Heat map of occupancy for GATA3, RORA, T-bet, and FoxP3 motifs, normalized by row and column

Figure 3.13



TF footprints identify direct DNA:protein interactions
(A and B) *ChIP binding profiles for an assortment of TFs at the IFNG and IL14/13 loci are mirrored by DNaseI TF footprints***.** The top panels are adapted from Zhu et al 2010 [32], and highlight locations at which factors are shown to associate by ChIP. Middle and bottom panels highlight TF footprints observed by deep-sequencing of DNaseI hypersensitive sites. Identified footprints are overlapped with FIMO predicted motif locations (p<1e-4).

Figure 3.14



DNaseI cleavage patterns result from direct protein:DNA interactions
(A) *DNaseI cleavage patterns parallel the topology of protein-DNA interfaces.* The co-crystal structures of IRF1 and GATA3 bound to DNA ligands is shown juxtaposed above the average nucleotide-level DNaseI cleavage patterns (blue) at motif instances of IRF1 and GATA3 in DHSs. Nucleotides that are sensitive to cleavage by DNaseI are colored as blue on the co-crystal structure while protected nucleotides are colored in grey. Interactions between DNA and the indicated protein are represented on the DNaseI cleavage pattern for interactions with alpha helices (green dotted circles) and interactions with IRF1 protein loops or the GATA3 C-terminal basic tail (green asterisks). The motif logo generated from IRF1 or GATA3 DNaseI footprints and a heatmap showing the per-nucleotide DNaseI cleavage for each motif instance in DHSs is displayed below the DNaseI cleavage pattern.
(B) *Nucleotide-level DNaseI cleavage patterns show cell-selective patterning*. The average nucleotide-level DNaseI cleavage pattern at motif instances of RORα and a generic 3-spaced STAT motif [86]. Cleavage events are normalized to number of events per million total cleavage events in DHSs.

Figure 3.15

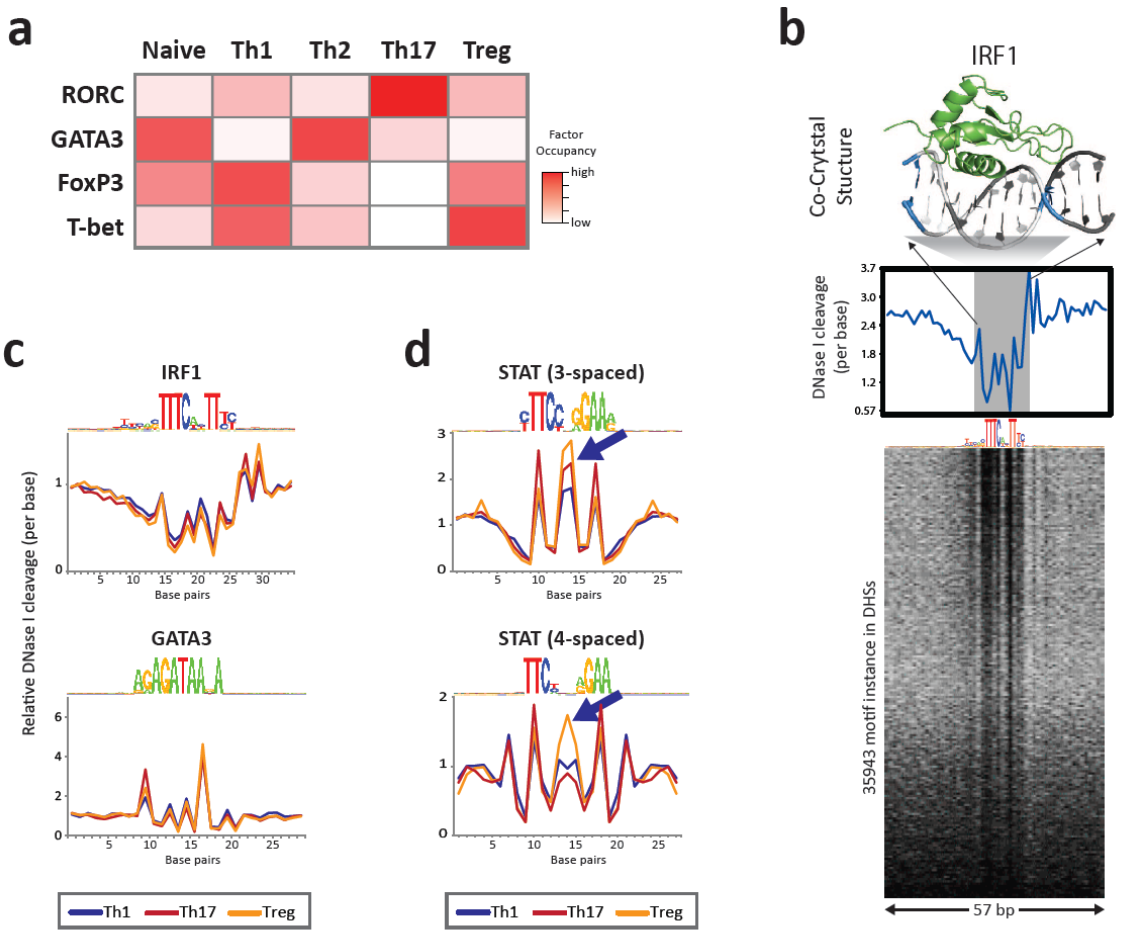


Figure 3. Profiling of TF occupancy identifies subset-specific occupancy at lineage-specific factor motifs

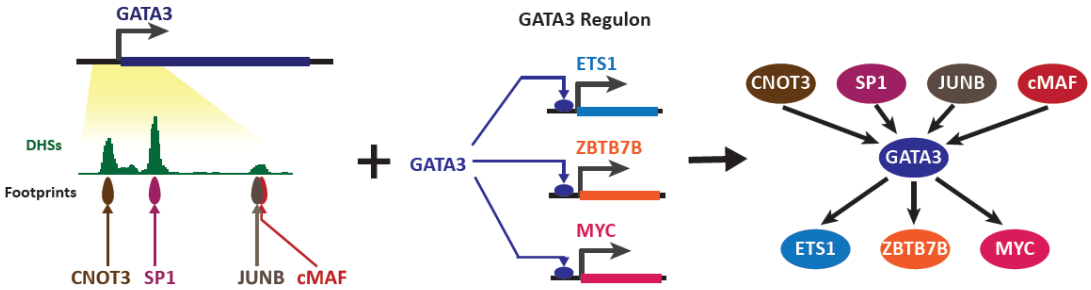
(a) *Lineage-specific differences in regulatory factor occupancy.* Occupancy of recognition sequences for major regulators is calculated as the percentage of motifs in DHSs which overlap footprints in each CD4 T cell subsets. Heat map of occupancy for GATA3, RORA, T-bet, and FoxP3 motifs, normalized by row and column.

(b) *DNaseI cleavage patterns parallel the topology of protein-DNA interfaces.*

(c) *Despite differential occupancy of motifs, cleavage patterns at bound motifs are consistent between cell types.* The nucleotide-level DNaseI cleavage pattern at motif instances of IRF1 and GATA3 motifs is normalized for local degree of DNaseI sensitivity.

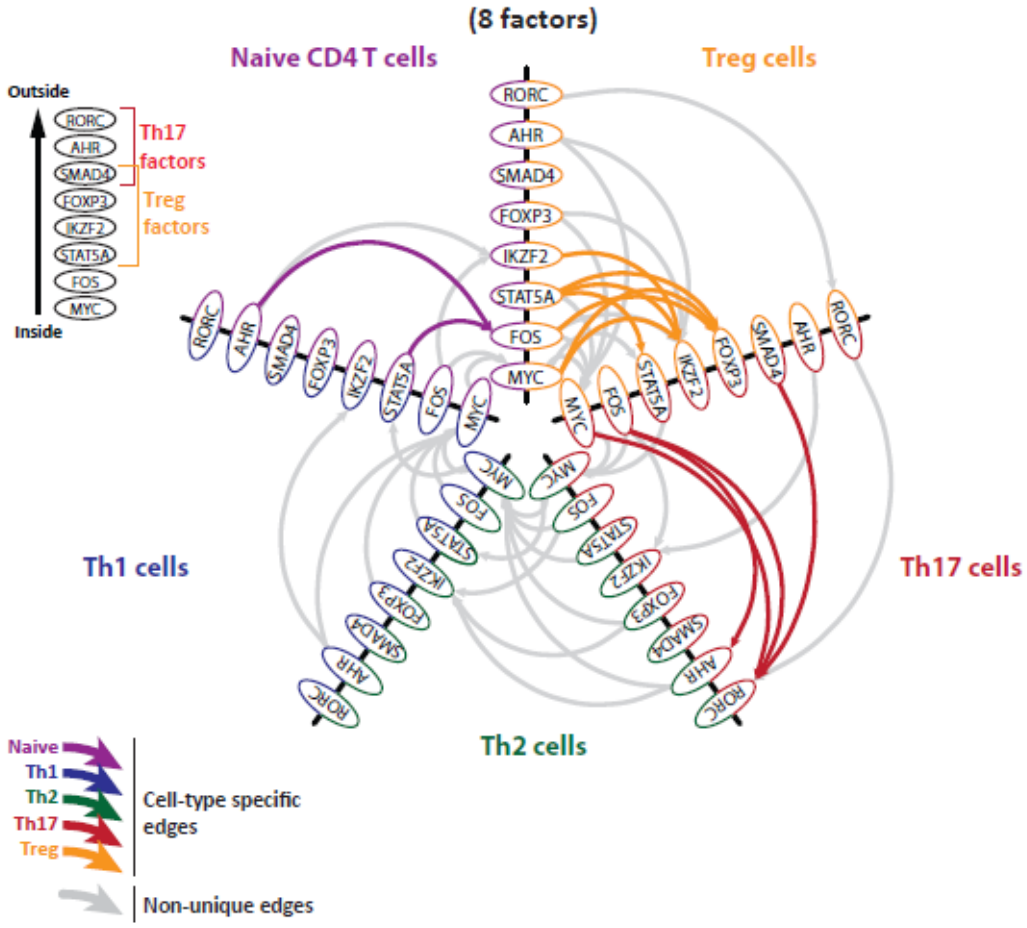
(d) *Nucleotide-level DNaseI cleavage patterns at STAT motifs show cell-selective patterning*. The nucleotide-level DNaseI cleavage pattern at motif instances of 3-spaced and 4-spaced STAT motifs[9] is normalized for local degree of DNaseI sensitivity. Cell-selective alterations in the STAT:DNA interface is focused in the center of the motifs between the binding interfaces for the STAT dimer (areas indicated with blue arrows).

Figure 3.16



DNaseI footprinting allows simultaneous detection of upstream and downstream interactions with GATA3
Left, CNOT3, SP1, JUNB, and cMAF footprints are identified in Th2 cell DHSs within the regulatory region of the GATA3 gene. Middle, GATA3 footprints are identified near the genes for ETS1, ZBTB7B, and MYC (among others). Right, this information is synthesized into a regulatory network around GATA3. Iteration over all TFs allows construction of a global regulatory network.

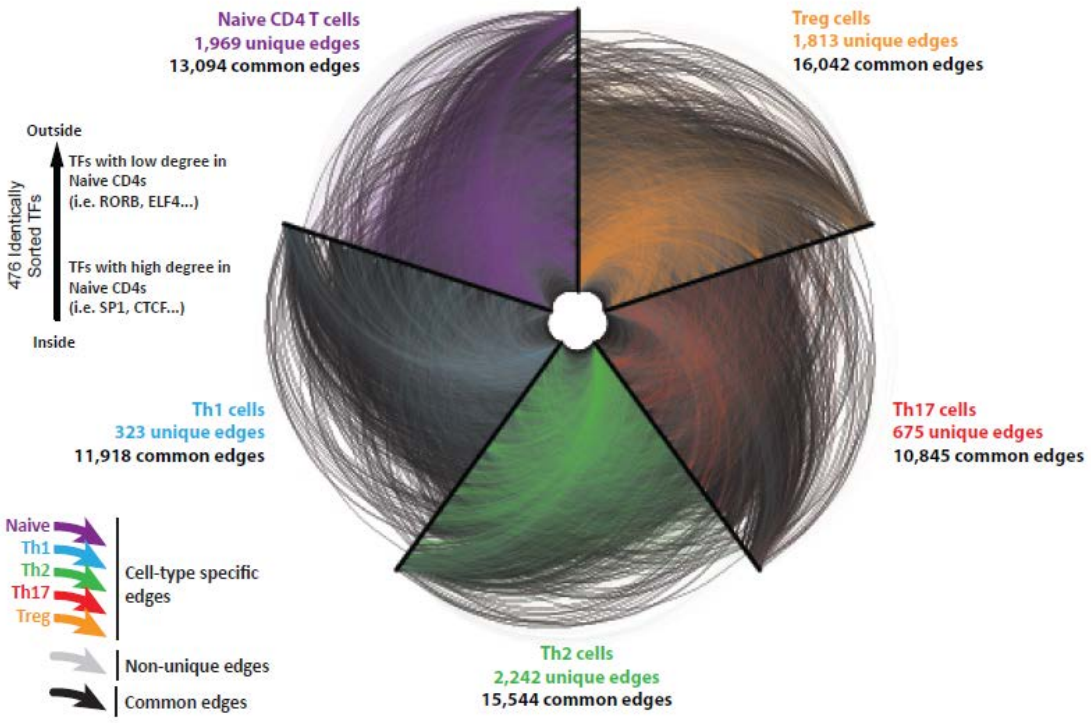
Figure 3.17



Cross-regulatory interactions in T cell networks reveal subset-specific connections

Eight factors known to regulate CD4 T cell function are arranged in the same order along each axis. Regulatory interactions (i.e., from regulat*or* to regulate*d*) are shown by arrows in clock-wise orientation. Cell type-specific edges are colored as indicated, whereas regulatory interactions present in two or more cell type networks are shown in grey.

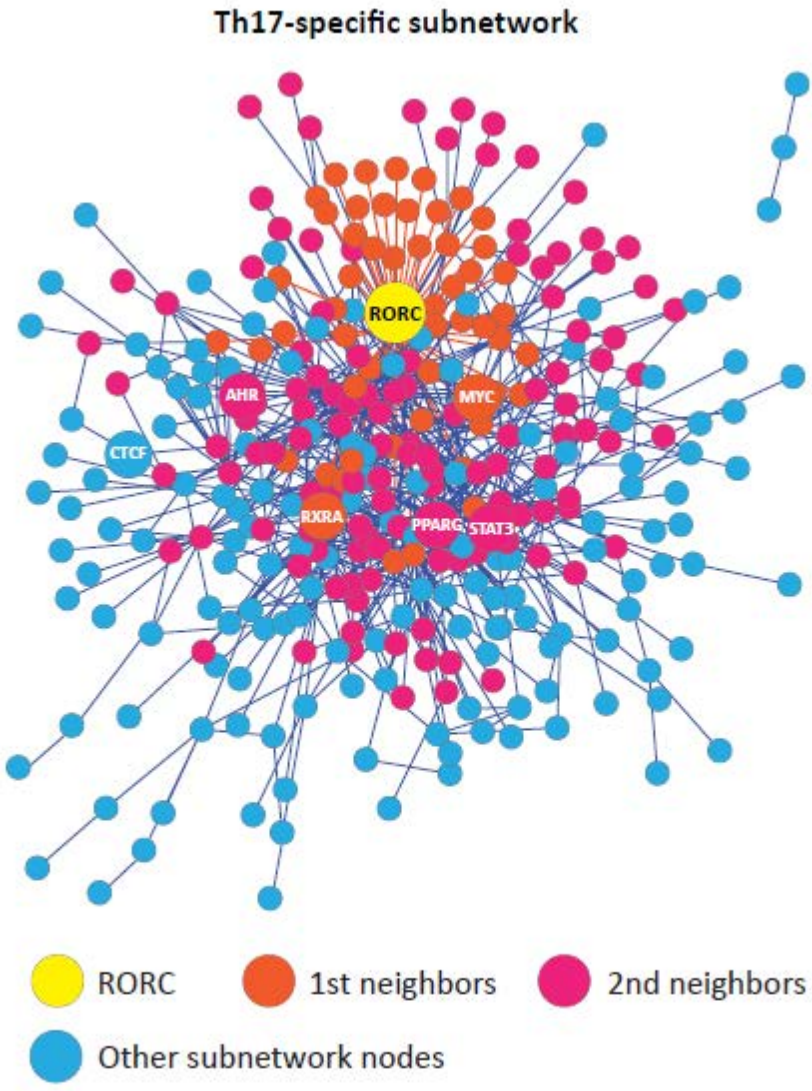
Figure 3.18



Cross-regulatory interactions between all 476 TFs in regulatory networks of CD4 T cell subsets

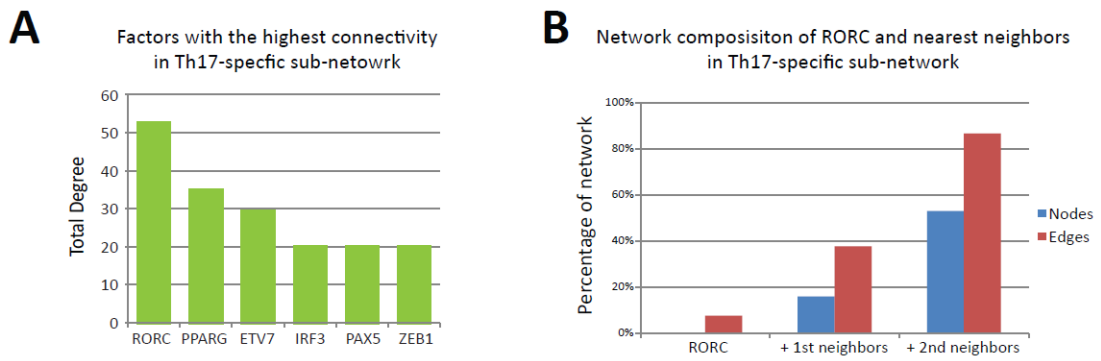
The 476 TFs are arranged in the same order along each axis, regulatory interactions are directed clockwise. Edges unique to a given cell type network are colored as indicated in the legend whereas regulatory interactions present in two or more networks are colored grey. Interactions present in all five cell type networks are colored black.

Figure 3.19



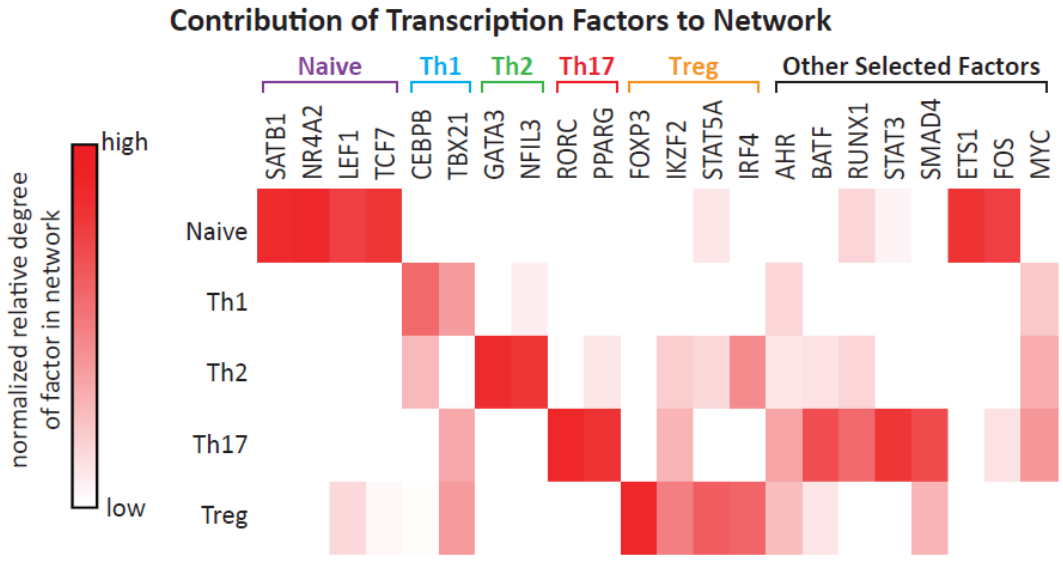
RORC is a prominently placed node in a well-connected Th17 sub-networks Nodes in network are TFs and edges are detected interactions as described. RORC is highlighted in yellow. Nodes directly and secondarily connected to RORC are highlighted as indicated. RXRA, MYC, AHR, STAT3, PPARG, and CTCF are labeled for reference. The sub-network of Th17-specific interactions composes 6.2% of the total Th17 network.

Figure 3.20



Th17-specific sub-network prominently features RORC
(A) *Factors with highest degree of connectivity in Th17-specific sub-network.* Total number of connections to and from the most prominent factors in the network formed by edges present uniquely in the Th17 network.
(B) *The majority of the Th17-specifc sub-network is directly or secondarily connected to RORC.* Percentage of network which is specified by RORC and its connecting edges, as well as the percentage of network specified by including the first neighbors (nodes directly connected to RORC) and second neighbors (nodes directly connected to first neighbors).

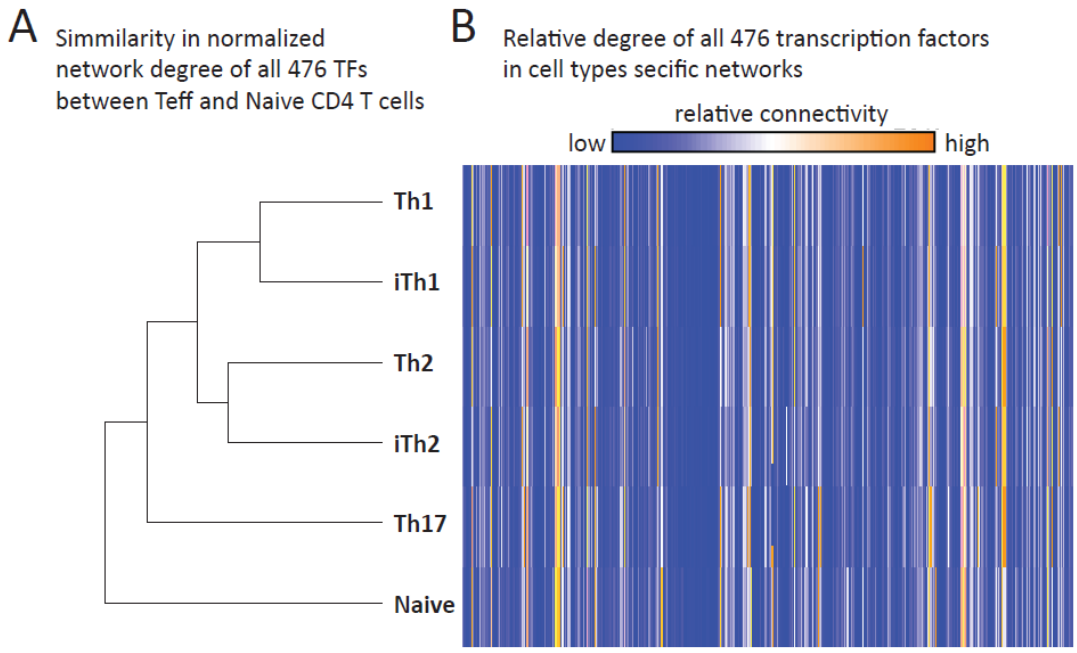
Figure 3.21



Transcription factors show lineage specific changes in network connectivity

Heatmap showing connectivity of factors within regulatory network. The relative connectivity (normalized network degree) for each TF was determined and the factors showing the greatest subset-specific increases in both incoming and outgoing edges for different CD4 T cell subsets are shown. Identified factors include well characterized lineage-defining factors. Selected other factors are also shown.

Figure 3.22



Comparison of overall network architecture shows high similarity between *ex vivo* isolated vs. *in vivo* derived T helper cells
(A) *Similarity in normalized network degree of all TFs between naïve CD4+ T cells and T effector cells*. Hierarchical clustering based on relative total degree of nodes in network. Th1 and Th2 cells isolated *ex vivo* are compared to iTh1 and iTh2 cells which were derived via polarization of naïve CD4 T cells following activation.
(B) *Heatmap showing normalized network degree of all TFs in network in naïve and T effector cells*. Normalized network degree of TFs in indicated cell types is tiled horizontally. TFs are arranged alphabetically along x-axis.

Figure 3.23

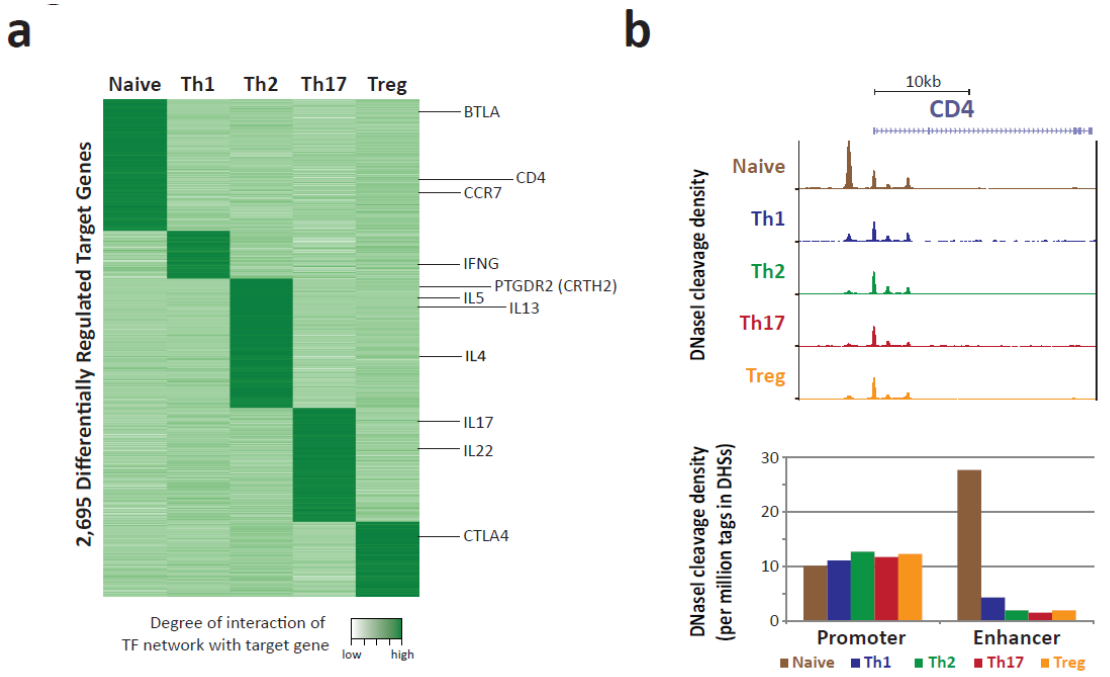
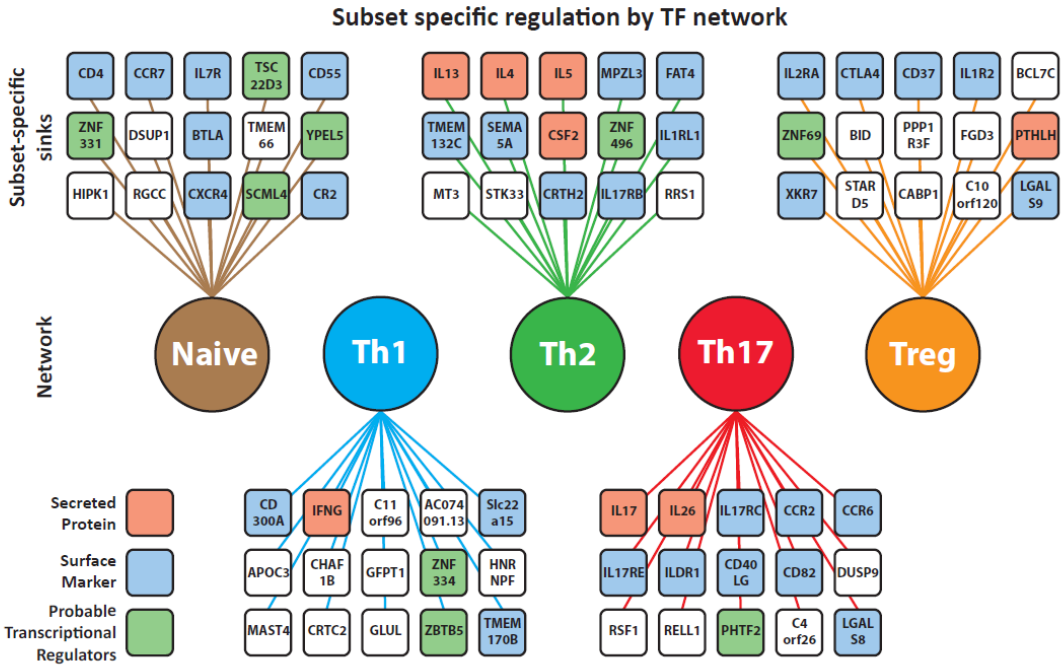


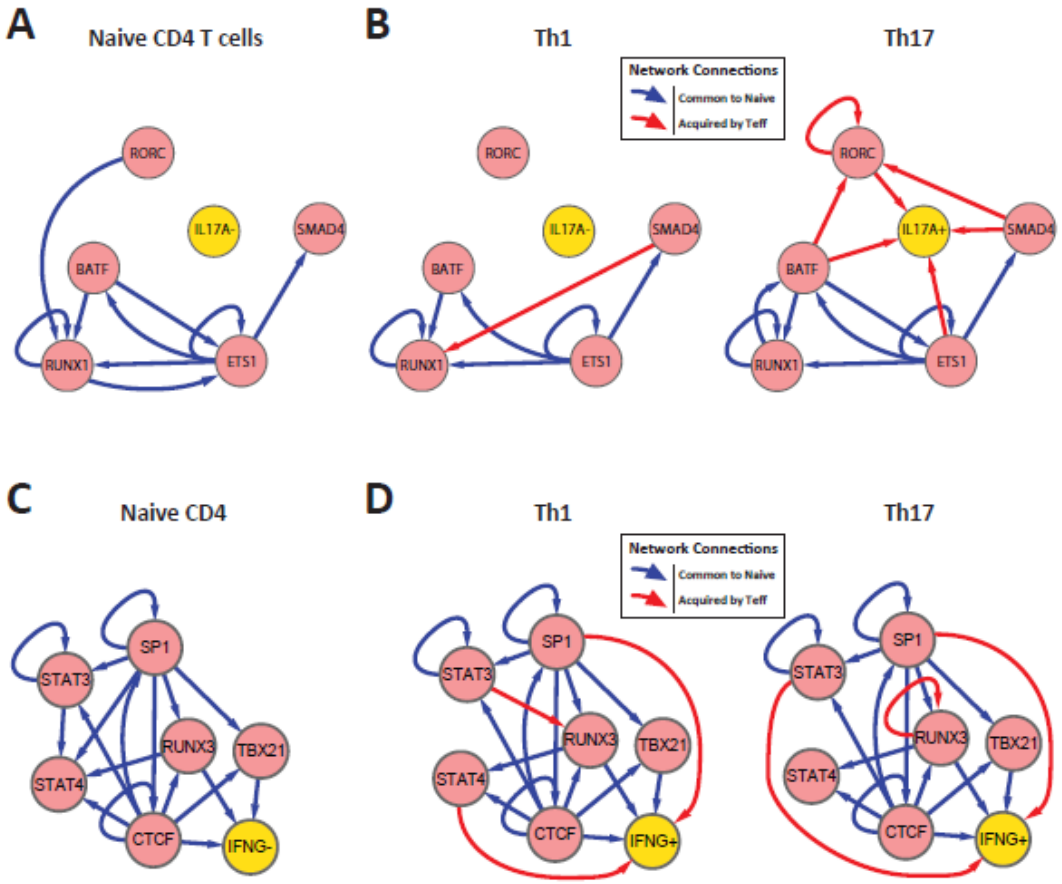
Figure 5. Network regulation of subset-specific target genes
(a) *Target genes show lineage-specific interactions with regulatory network*. Heatmap showing the relative connectivity (normalized "in" degree of network) of factors showing lineage-specific regulation by the network.
(b) *CD4 shows Naïve CD4 T cell specific regulatory input into upstream enhancer element.* Above, DNaseI cleavage density profile around *CD4* gene is shown in indicated cell types. Below, DNaseI accessibility of the *CD4* promoter and upstream enhancer is shown.

Figure 3.24



Lineage-specific target genes include known effector genes as well as uncharacterized effector genes and transcription factors. Lineage-specific target genes are tiled (squares), with connections (lines) to networks (circles) as indicated. Target genes are classified according to function as secreted protein, cell surface receptor, transcription factor, or other or unknown function.

Figure 3.25



Sub-networks are poised for rapid responses in a T cell effector-specific manner

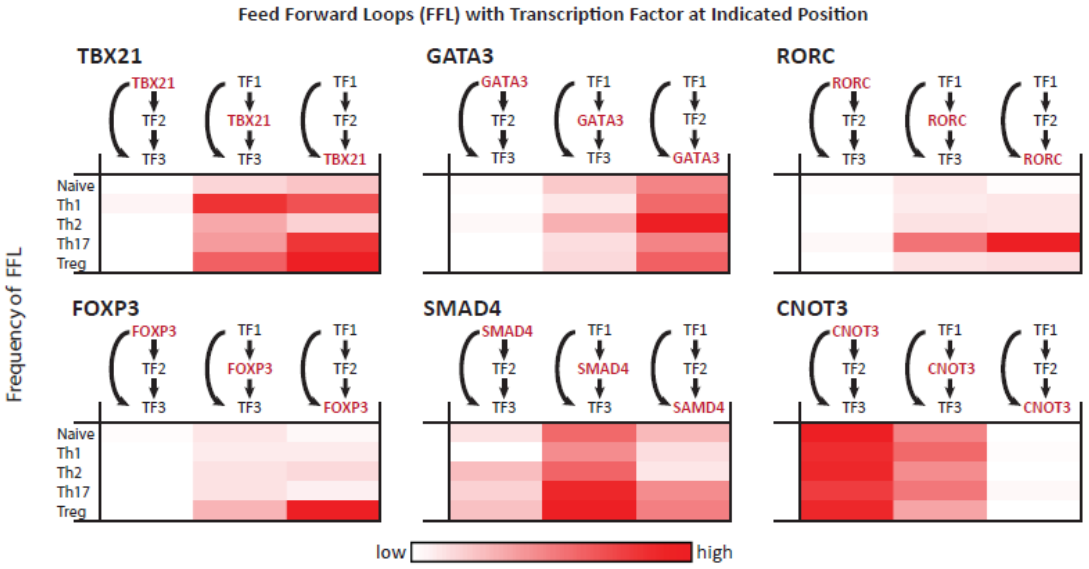
(A) *Footprints of factors in regulatory network within the IL17A locus in Naïve CD4 T cells*. Locus boundaries were defined according to published reports [130]. Interactions among selected factors and with the *IL17A* locus are shown as arrows.

(B) *Footprints of factors in regulatory network within the IL17A locus in Th1 and Th17 cells.* Interactions present in naïve CD4+ T cells are shown in blue; interactions acquired during lineage differentiation to Th1 or Th17 cells are shown in red. IL17 expression in Th17 cells is indicated with a "+" symbol.

(C) *Footprints of factors in regulatory network within the IFNG locus in Naïve CD4 T cells*. Locus boundaries were defined according to published reports [177].

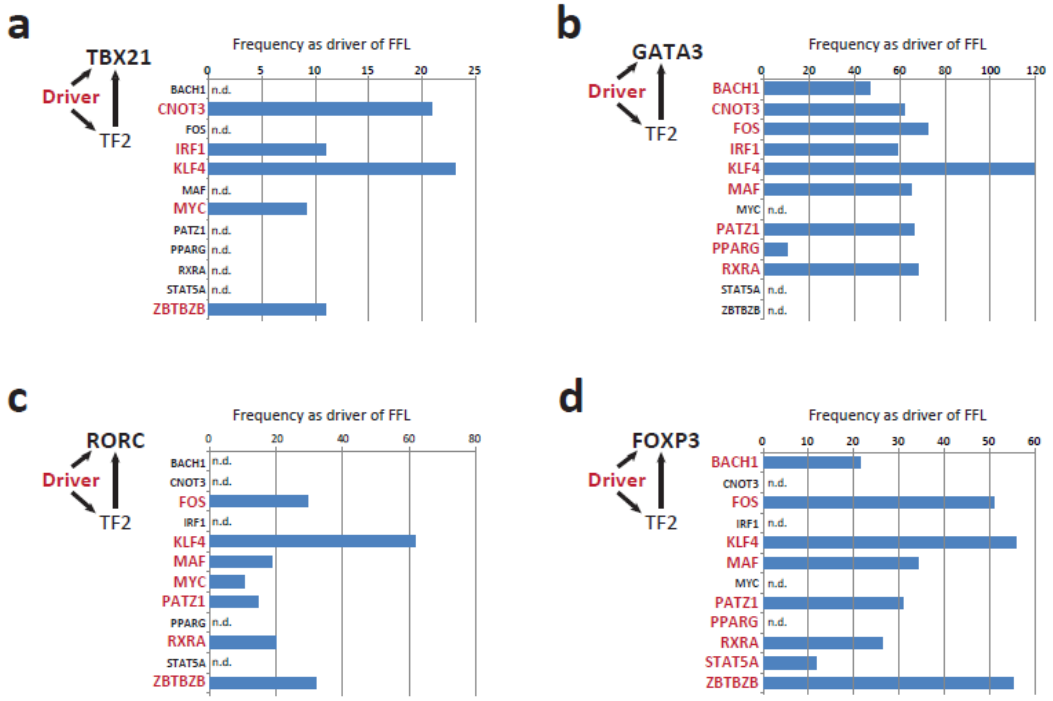
(D) *Footprints of factors in regulatory network within the IFNG locus in Th1 and Th17 cells.* IFNG expression in Th1 and Th17 cells is indicated with a "+" symbol.

Figure 3.26



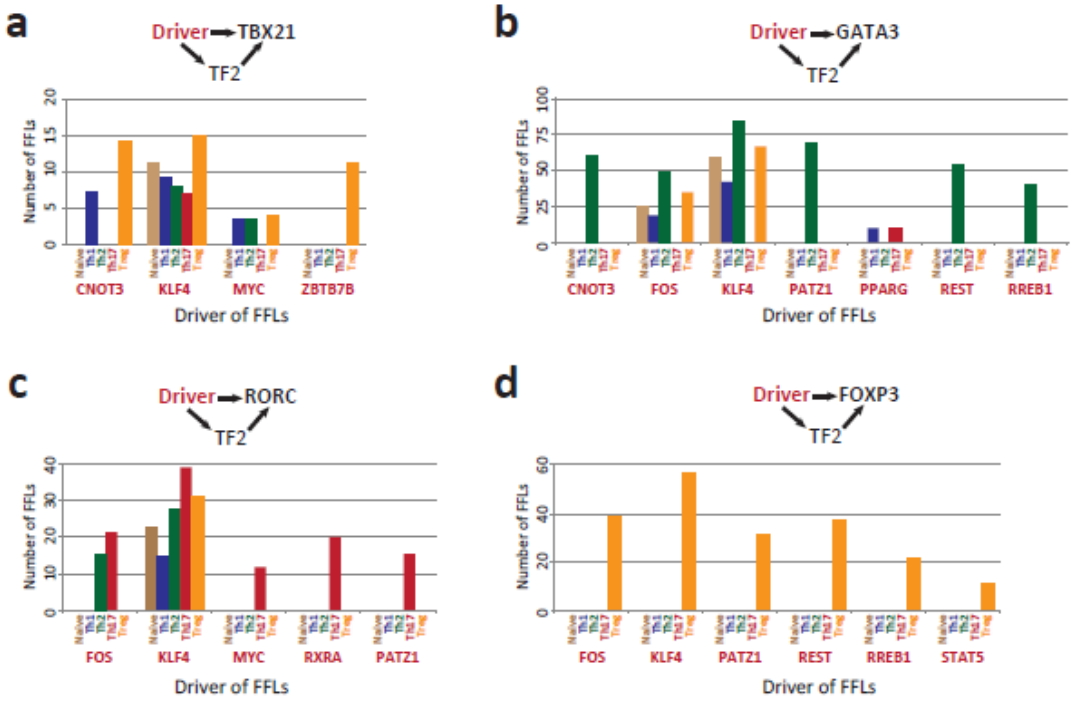
Lineage restricted factors adopt passenger positions in feed forward loops
Heatmap showing frequency with which each factor is present as the driver (top position) or passenger (middle or bottom position) within feed forward loops in each CD4+ T cell subset.

Figure 3.27



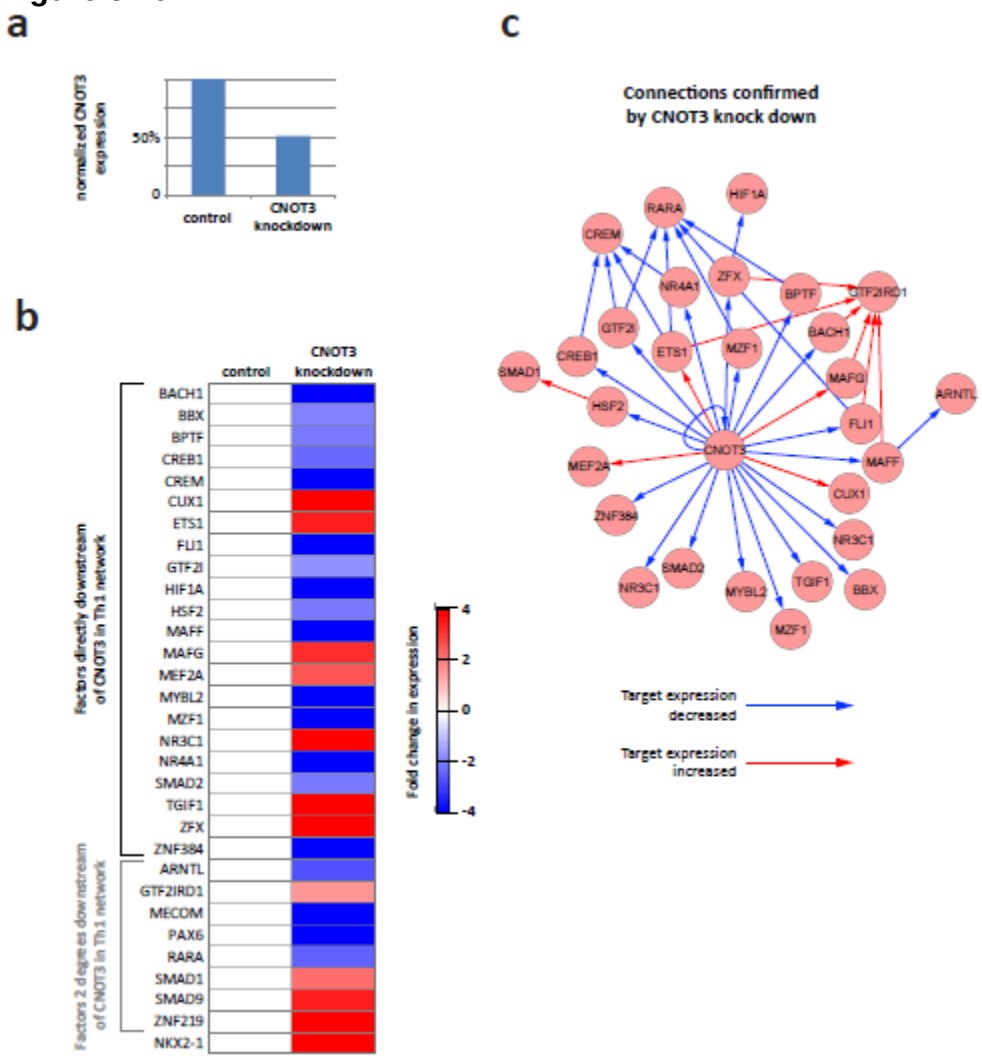
A common set of factors drive "master" regulatory factors.
For each CD4 T cell lineage-defining factor, T-bet (**a**), GATA3 (**b**), RORC (**c**), and FoxP3 (**d**) common drivers for loops containing the lineage-defining factors are shown along with number of such loops.

Figure 3.28



"Master" regulatory factor regulation is lineage-specific
For each CD4 T cell lineage-defining factor, *TBX21* (**a**), *GATA3* (**b**), *RORC* (**c**), and *FOXP3* (**d**), common drivers for loops containing the lineage-defining factors are shown along with number of such loops in each cell type.

Figure 3.29



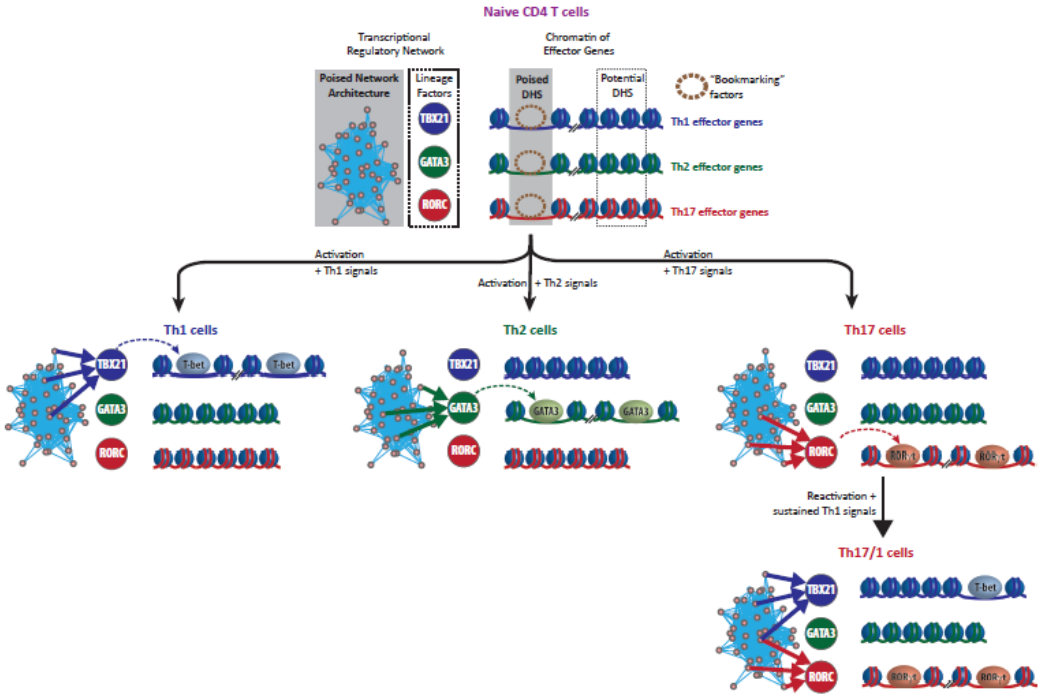
Alteration of transcription factor expression downstream of CNOT3 knockdown

(a) *CNOT3 siRNA knocks down transcription of CNOT3.* The total RNA-seq tags which mapped to the CNOT3 gene were summed in CNOT3-specific siRNA and control siRNA transfected cells. Expression shown is normalized to control siRNA infected level.

(b) *CNOT3 knockdown results in changes in expression in genes downstream of CNOT3 in the Th1 network.* Transcript isoforms with significant ($q < 0.05$) changes in expression are listed along with heat map showing degree and direction of change. When changes in multiple transcript isoforms from the same gene are detected, we show the most significant result. Genes are grouped based on location in the network with respect to CNOT3.

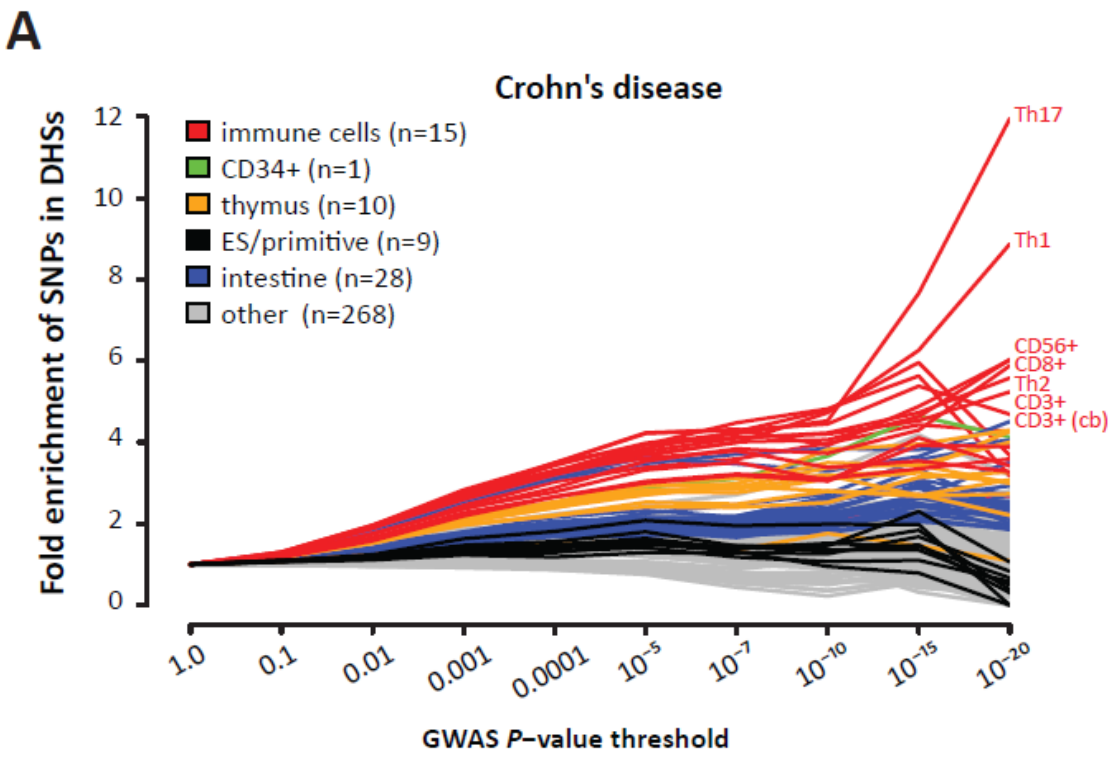
(c) *CNOT3 knockdown confirms network connections.* CNOT3 and downstream factors that show significant change in expression after CNOT3 knockdown are shown with connectivity as described in Th1 network. Decrease in expression following CNOT3 knockdown are indicated by blue arrows; increased expression is indicated by red arrows.

Figure 3.30



Proposed mechanism of helper T cell differentiation
Naïve CD4 T cell possess a poised regulatory network architecture and chromatin poised to respond to expression of sunset-specific transcription factors. Differentiation to Th1, Th2, and Th17 is accompanied by subset-specific expression of T-bet, GATA3, and RORC, respectively, through upstream regulation by the network and by binding of subset-specific factors to pre-available sites and acquired sites. Th17 cells can acquire Th1-like effector function via subsequent expression of T-bet.

Figure 4-1



Th17 and Th1 cells are specifically enriched in SNPs in DHSs
The fold-enrichment in the number SNPs in all DHSs present in the indicated subsets is calculated for SNPs having p-values which fall into each of the bins indicated on the x-axis.

Table 1

GO Biological Process	Binom FDR Q value
regulation of lymphocyte activation	**9.82E-248**
regulation of T cell activation	**9.59E-241**
lymphocyte activation	**1.39E-230**
positive regulation of lymphocyte activation	**6.97E-218**
positive regulation of cell activation	**1.27E-216**
positive regulation of leukocyte activation	**9.29E-216**
regulation of cytokine production	**3.54E-211**
positive regulation of T cell activation	**2.41E-210**
immune response-activating signal transduction	**6.20E-207**
immune response-regulating signaling pathway	**4.79E-193**
leukocyte differentiation	**8.55E-189**
activation of immune response	**1.57E-186**
lymphocyte differentiation	**3.65E-170**
immune response-activating cell surface receptor signaling pathway	**1.30E-162**
antigen receptor-mediated signaling pathway	**1.14E-161**
immune response-regulating cell surface receptor signaling pathway	**1.75E-153**
T cell activation	**1.92E-147**
regulation of lymphocyte differentiation	**2.15E-140**
T cell receptor signaling pathway	**7.07E-132**

Subset-specific DHSs are strongly associated with CD4 T cell gene expression. The association of highly differentially regulated DHSs and genes with indicated GO biological process annotations was calculated as FDR Q-values of region based associations by analysis with GREAT.

Table 2

Cell type	Unique reads	Footprints
Naïve CD4	287,232,507	709,387
Th1	261,414,709	482,219
Th2	404,747,076	885,537
Th17	201,846,523	542,535
Treg	226,446,617	788,078
Average	**276,337,486**	**681,551**

DNaseI footprints detected in CD4 T cell subsets.
The total number of uniquely mapped reads sequenced and total number of footprints detected for naïve CD4, Th1, Th2, Th17, and Treg cells are reported along with the average read depth and average number of footprints detected per cell type.

Table 3

Factor	T cell subsets with differential expression
BACH1	Th2 > Th1[1]
CNOT3	Treg > Teff[2]; Th1 > Th2[1]
FOS	Th2 > Th1[1]; Teff > Treg[3]
IRF1	Treg > Teff[2]
KLF4	Teff > Treg[2]; Th2 > Th1[1]
MAF	Treg > Naïve[4]; Treg > Teff[3]; Th2 > Th1[1]
MYC	Treg > Naïve[4]; Treg > Teff[2, 5]; Th1 > Th2[1]
PATZ1	Teff > Treg[2]; Th2 > Th1[1]
PPARG	Treg > Teff[2]; Th2 > Th1[1]
RXRA	Th2 > Th1[1]; Treg > Naïve[4]
STAT5A	Th1 > Th2[1]; Treg > Teff[2]

Differential expression of factors predicted to regulate T-bet, GATA3, RORgt, and/or FoxP3. Differential expression of factors in CD4 T cell subsets was identified from public data available in the Gene Expression Atlas[6]

References:

1. Rodriguez, A. *et al.* Requirement of bic/microRNA-155 for normal immune function. *Science* **316**, 608-611 (2007).
2. Alberts, R. *et al.* Expression QTL mapping in regulatory and helper T cells from the BXD family of strains reveals novel cell-specific genes, gene-gene interactions and candidate genes for auto-immune disease. *BMC genomics* **12**, 610 (2011).
3. Herman, A.E., Freeman, G.J., Mathis, D. & Benoist, C. CD4+CD25+ T regulatory cells dependent on ICOS promote regulation of effector cells in the prediabetic lesion. *The Journal of experimental medicine* **199**, 1479-1489 (2004).
4. Samstein, R.M. *et al.* Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell* **151**, 153-166 (2012).
5. Layland, L.E. *et al.* Pronounced phenotype in activated regulatory T cells during a chronic helminth infection. *J Immunol* **184**, 713-724 (2010).
6. Kapushesky, M. *et al.* Gene Expression Atlas update--a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic acids research* **40**, D1077-1081 (2012).