Title
The cis-Regulation of Aire Expression

Permalink
https://escholarship.org/uc/item/3wj6f647

Author
LaFlam, Taylor

Publication Date
2015

Peer reviewed|Thesis/dissertation

The *cis*-Regulation of *Aire* Expression

by

Taylor Nicholas LaFlam

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

iv

To my parents, who always encouraged my curiosity

Acknowledgments

During my time at UCSF, I have benefited greatly from numerous sources of support, both scientific and personal. It is no exaggeration to say that my work would have been impossible without them.

To begin, I would like to thank Mark Anderson, my thesis advisor, for his mentorship during these past several years. Mark always succeeded at striking that delicate balance between supporting my efforts and allowing me enough space to learn how to operate independently. I feel that I have grown tremendously as a scientist during graduate school, and I owe much of this to Mark's enthusiastic, discerning, and collaborative approach to science.

I would also like to thank the other members of my thesis committee, Mark Ansel and Brian Black, for their consistently thoughtful and helpful comments. I found my committee meetings to be fun and rewarding, and this research was improved by their advice.

I have very much enjoyed being part of the Anderson lab, and much of this is due to my wonderful fellow lab members. They were always a pleasure to be around. Thanks to Kellsey Johannes, Jessica Cortez, Kayla Fasano, and Wint Lwin for all their help, including diligent maintenance of the mouse colony through the years—I certainly created no shortage of work for them. Thanks to Alice Chan, Corey Miller, Jennifer Lu, Imran Khan, Maria Mouchess, Mickie Cheng, Ruth Taniguchi, Taka Hanafusa, Todd Metzger, and especially Mike Waterfield for technical assistance and advice on my project. I'd like to single out Whitney Purtha, longtime resident of the desk behind me, for our many delightful chats, sometimes scientific but mostly not.

I want to acknowledge the administrative and logistical support that I have received, which, when going smoothly, you scarcely notice. And I scarcely noticed it. So, thank you to

Contributions to Presented Work

Chapter 2 is adapted from a manuscript in preparation: "Identification of a *cis*-regulatory element essential for immune tolerance," by Taylor N. LaFlam[1], Grégory Seumois[2], Corey N. Miller[1], Kayla J. Fasano[1], Wint Lwin[1], Irina Proekt[3], Anjana Rao[2], Pandurangan Vijayanand[2], and Mark S. Anderson[1]. In the ChIP-seq experiments (Figure 2.1 and S2.3), all steps after chromatin sonication were performed at Anjana Rao's lab at the La Jolla Institute for Allergy and Immunology. Greg Seumois and Vijay Pandurangan developed the protocol, and Greg Seumois took the lead in performing the chromatin immunoprecipitation. Jeremy Day oversaw the high-throughput sequencing. Amin Gholami and Jason Greenbaum performed bioinformatic analysis. Kayla Fasano and Wint Lwin performed the autoantibody radioligand binding assays (Figure 2.5B). Corey Miller and Irina Proekt performed fundoscopy (Figure 2.5C) The UCSF Mouse Molecular Pathology Core embedded, sectioned, and stained tissues with hematoxylin and eosin (Figure 2.5D).

Chong Park and Kendall Bell of the UCSF ES Cell Targeting Core tested candidate CRISPR guide RNAs for efficacy and prepared RNA for injection. Helen Lu of the UCSF LARC Transgenic Mouse Core performed all microinjections and mouse associated with the generation of the *Eaig*, *A4dig*, *Adig*Δ*CNS1*, *ACNS1*[-/-], *AdigFlox61-94*, and *AdigFlox111-145* mice.

Todd Metzger and Imran Khan treated *Aire* lineage-tracing mice, sorted mTECs, and prepared RNA for sequencing. This RNA-seq (Figure 3.1B) was performed by the UCSF SABRE Functional Genomics Facility, led by David Erle and managed by Andrea Barczak.

[1] Diabetes Center, University of California, San Francisco, San Francisco, CA, USA

[2] La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA

[3] Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA, USA

Rebecca Barbeau prepared and sequenced the samples. Josh Pollack performed bioinformatic analysis.

The *cis*-Regulation of *Aire* Expression

Taylor N. LaFlam

Abstract

Immune self-tolerance, the absence of an immune response against the body, is an active process involving several complementary mechanisms. The autoimmune regulator gene (*Aire*) plays a critical role in this process by inducing the expression of a diverse set of tissue-specific antigens (TSAs) in medullary thymic epithelial cells (mTECs). These TSAs are presented and help eliminate or control self-reactive thymocytes. The expression of *Aire* is highly restricted, occurring in adults only in mTECs and extrathymic *Aire*-expressing cells (eTACs), which are found in secondary lymphoid tissues. How this specific pattern of expression is achieved has remained unclear. Although RANK and non-canonical NF-κB signaling are known to be important for the development and maintenance of *Aire*-expressing mTECs, the *cis*-regulation of *Aire* is poorly understood. We aimed to discover distal *cis*-regulatory elements that control *Aire*. Here, we identify a conserved DNA element that is essential for *Aire* expression. We find enrichment of enhancer-associated histone marks near this element, which we call ACNS1 and show to be NF-κB-responsive. We demonstrate that a 4.3-kb region, which includes ACNS1 and the *Aire* promoter, is sufficient to recapitulate the cell-type specificity of *Aire* expression. We show that ACNS1 is essential for *Aire* expression in mTECs and eTACs in vivo and is required to prevent spontaneous autoimmunity. We also observe that *Aire* and *Dnmt3l* are co-expressed and that ACNS1 is necessary for thymic *Dnmt3l* expression. Finally, we present preliminary evidence for additional *Aire cis*-regulatory elements. Together, these findings further our understanding of the control of immune tolerance and may lead to improvements in the diagnosis and treatment of autoimmune disease.

Table of Contents

List of Tables

List of Figures

Chapter 4: Preliminary evidence for additional *Aire cis*-regulatory elements

Chapter 5: Discussion and Concluding Remarks

Chapter 1: Immune Tolerance and *Aire*

ΓΝΩΘΙ ΣΕΑΥΤΟΝ

"KNOW THYSELF"

-inscribed at the ancient Temple of Apollo at Delphi

Overview:

We begin with a brief discussion of immunity, focusing on how the immune system distinguishes self from non-self and on general mechanisms of immune tolerance. We next summarize previous research on the autoimmune regulator (*Aire*) gene, including its discovery, role in immune tolerance, pattern of expression, and mechanisms of action. We close by introducing the central topic of this dissertation, the transcriptional regulation of *Aire*.

Immunity:

Life is an ongoing struggle against disorder. Organisms face the never-ending challenges of obtaining enough energy, avoiding or enduring environmental hazards, and finding a mate. To this, add the difficulty of hostile neighbors.

Since soon after they arose, organisms have faced the threat of other living things seeking to survive at their expense. When these attackers are large and try to kill quickly, we call them predators; when they are small and seek to take life bit by bit, we call them parasites or pathogens. Only those living things that evolved ways to recognize and combat these invaders survived. Eventually, in some multicellular organisms, these systems also came to be used to

1

fight the dangers that arise within, eliminating dead and cancerous cells. Today, members of all kingdoms of life are equipped with immunity in one form or another. These range from the various tools bacteria use against bacteriophages—such as restriction enzymes to chop up invading viral genomes (1)—to the elaborate network of cells and extracellular molecules found in vertebrates (2). Given that recent research has revealed great complexity in the immune systems of various phyla of invertebrates (3,4), the idea that mammals have the ne plus ultra of immune systems is somewhat overstated. Nevertheless, the vertebrate immune system possesses a particular DNA-rearranging mechanism that is both critical to immune function and foundational to the profound biological need for the gene at the heart of this dissertation.

A critical aspect of immunity in all cases is the ability to distinguish the self from the non-self, to know what to protect and what to reject. In vertebrates, such means of recognition can be divided into two major groups, aligning with the separation of the immune system as a whole into innate and adaptive immunity—a convenient if sometimes criticized framework (5). Recognizing non-self in the innate immune system depends on use of invariant, germline-encoded pattern recognition receptors (PRRs), which include the toll-like receptors, C-type lectin receptors, NOD-like receptors, and RIG-I-like receptors (6). These receptors recognize widespread, structurally conserved molecules, such as lipopolysaccharide, flagellin, and single-stranded RNA. Collectively, these PRR ligands are traditionally called pathogen-associated molecular patterns, though they would be more accurately described as simply microbe-associated molecular patterns because they are not restricted to pathogenic organisms. Some of these receptors also help direct the immune system to distressed, dying, or abnormal cells by detecting what are termed damage-associated molecular patterns (7). The ubiquity and conservation of microbe-associated molecular patterns allows a rather limited repertoire of PRRs

to be effective. An intricate network of factors regulates innate immunity, resulting in, for example, appropriately differential responses to a pathogenic bacterium and a commensal bacterium, each of which expresses components recognized by PRRs (8).

The innate immune system, equipped with these PRRs, acts as the first line of protection. It begins with the epithelial barriers of the skin and mucosa. It includes a variety of hematopoietic cells, including neutrophils, monocytes, macrophages, eosinophils, basophils, dendritic cells (DCs), and natural killer (NK) cells. It also includes secreted proteins, such as complement and α-defensins, which act to both detect and damage pathogens. The innate immune system is able to successfully control many infections. Indeed, the importance of this aspect of the innate immune system has been underlined by the discovery that certain immunodeficiency syndromes are caused by mutations in genes encoding innate immunity effector molecules. For example, defects in NADPH oxidase lead to chronic granulomatous disease, which is characterized by severe, recurrent fungal and bacterial infections (9). Some pathogens, however, are able to evade or overwhelm the innate immune system. As such, another important role for the innate immune system is to initiate and guide the response of the adaptive immune system. (The adaptive immune system can in turn stimulate and shape the innate immune response.)

The adaptive immune system of jawed vertebrates (gnathostomes) includes two major types of cells: B lymphocytes (or simply B cells) and T lymphocytes (T cells). Both cell lineages are characterized by expression of antigen receptors, which are membrane-bound proteins composed of immunoglobulin domains and whose expression depends on a process of somatic DNA rearrangement called V(D)J recombination. These antigen receptors are called, intuitively if uncreatively, the B cell receptor (BCR) and the T cell receptor (TCR). The lymphocytes of

jawless vertebrates (agnathans), such as the lamprey, also have antigen receptors, though they evolved independently and are made up of leucine-rich-repeat domains assembled by a different mechanism of DNA recombination (10). The presence of T- and B-like lymphocytes in both jawed and jawless vertebrates indicates that these cell lineages preceded the evolution of antigen receptors.

Each of these lineages has distinct roles. B cells are the agents of humoral immunity--that is, immunity mediated by the actions of antibodies present in the fluids (i.e., humors) of the body; antibodies are secreted forms of the BCR. T cells, in contrast, effect cell-mediated immunity. This includes the actions of cytolytic (killer) T cells in eliminating damaged and virally infected cells, as well as the actions of helper T cells, which facilitate the activities of several types of immune cells, including B cells, other T cells, and macrophages. The defining characteristics of the adaptive immune system are the antigen specificity of the response, discussed more below, and the generation of immunologic memory, in which the adaptive immune system reacts more quickly and robustly upon subsequent encounters with a given antigen than occurred at the time of the initial infection (2).

Antigen receptors and immune tolerance:

The PRRs of the innate immune system can bind only a limited, consistent set of ligands. In contrast, each lymphocyte has a unique antigen receptor, creating an enormous anticipatory diversity of receptors that is capable of recognizing components of almost any pathogen. The TCR, with certain limited exceptions, recognizes as its antigen a peptide loaded onto the heterodimeric membrane-bound proteins called major histocompatibility complex I (MHC-I) and MHC-II. In contrast, BCRs bind many classes of molecules, including carbohydrates, proteins, lipids, and small molecules. Both the TCR and the BCR contain a variable region, which binds

4

the antigen, and a constant region, which anchors the receptor to the membrane and interacts with co-receptor proteins. The antigen-binding component of both antigen receptors is a heterodimer. In most T cells, the TCR consists of one TCRα chain and one TCRβ chain; there is also a population of T cells expressing TCRγ with TCRδ. The BCR is a dimer of heterodimers, each of which has an IgH chain and either an Igκ or Igλ chain.

A functional antigen receptor gene is created by a process of DNA rearrangement, called V(D)J recombination, that occurs in developing lymphocytes (11,12). The variable region of each antigen receptor gene is encoded by multiple gene segments: two gene segments, called V and J, in *Tcra, Tcrc, Igk,* and *Igl*; three gene segments, called V, D, and J, in *Tcrb, Tcrd*, and *Igh*. In each of these loci, there are multiple slightly different copies of each of these gene segments. In V(D)J recombination, double-strand breaks are made in the DNA, and the DNA is rearranged such that one of each segment comes together to form an exon encoding an entire variable region. This process provides three sources for antigen receptor diversity. The first is from all the different possible combinations of V and J, or V and D and J, segments present in each locus. For example, the human *TCRB* locus has 52 V segments, 2 D segments, and 13 J segments. The second is from the dimerization of the receptor, wherein any given TCRα might be partnered with any possible TCRβ. Finally, the recombination involves error-prone end joining that usually results in the deletion or insertion of a small number of nucleotides at the junction, further increasing receptor diversity. The end result is that V(D)J recombination can yield any of approximately 5 x 10^{13} possible BCRs and 10^{18} possible αβ TCRs in humans and similar numbers in mice (2).

Having such a broad array of receptors leaves the immune system ready to specifically recognize and respond to almost any threat and is essential for survival. Mutations that prevent

V(D)J recombination or other critical aspects of lymphocyte biology produce severe immunodeficiencies that are invariably fatal without bone marrow transplantation (13). However, the very diversity of antigen receptors presents a significant risk because there will inevitably be receptors specific for self-antigens. The presence of self-reactive lymphocytes can in turn result in autoimmunity (14,15). A variety of overlapping safeguards are present to control this risk without rendering the adaptive immune system so constrained as to prevent it from effectively fighting infection. One such checkpoint occurs at the binding of the TCR to its antigen. T cell activation begins with interaction between a T cell and an antigen-presenting cell (APC), such as a DC, that is displaying both the cognate antigen (i.e., the antigen recognized by the TCR) on its surface and additional costimulatory molecules, such as CD28. If only the cognate antigen is present, the T cell will not become activated (16). In order for there to be adequate expression of costimulatory molecules, the APC must itself be activated. The stimulus to adopt an activated state involves the PRRs discussed above. Although this requirement for costimulation serves as an important regulator of T cell responses, additional mechanisms are also necessary.

The mechanisms by which immune tolerance is established and maintained can be divided into central tolerance, which occurs at the site of lymphocyte development, and peripheral tolerance, which occurs elsewhere in the body (16,17). With regards to T cell tolerance, which will be the primary concern of this dissertation, central tolerance refers to processes occurring in the thymus. The two checkpoints that developing T cells must clear after undergoing recombination to form a TCR are positive selection and negative selection (18). Positive selection refers to the need for the T cell to bind with at least low affinity to some peptide (any peptide) presented by cortical thymic epithelial cells (cTECs) in the thymic cortex. Thymocytes whose TCRs fail to bind adequately lack critical survival signals and apoptose; they

are said to die by neglect. Cells that pass positive selection continue to mature and proceed to the medulla. There, they remain for a few days, interacting with numerous medullary thymic epithelial cells (mTECs) and DCs. If a given cell's TCR recognizes a self-antigen strongly, it will die, which is termed negative selection. If it recognizes a self-antigen moderately, but not strongly enough to be driven to die, it can develop into a regulatory T cell. T cells can also adopt a regulatory T cell identity after leaving the thymus (19). Regulatory T cells act to control immune responses and prevent autoimmunity through multiple mechanisms (20).

Although negative selection had been hypothesized for decades, the first definitive evidence for clonal deletion of developing self-reactive T cells came less than 30 years ago. In 1987, Kappler and colleagues observed that mice that express the MHC-II molecule I-E, in comparison to mice that do not express I-E, have significantly fewer mature thymocytes and peripheral T cells bearing a TCRβ V region associated with reactivity to I-E (21). The following year, another group reported deletion of double positive (CD4$^+$ CD8$^+$) thymocytes in male mice transgenic for a TCR that recognizes H-Y, which is expressed only in males (22). Once it was recognized that strong binding to an antigen in the thymus leads to deletion, it was hypothesized that T cells recognizing ubiquitous or thymic antigens would be negatively selected in the thymus whereas cells with reactivity to peripheral antigens would be controlled outside of the thymus. Unexpectedly, however, it was noted that numerous genes that would not be predicted to be transcribed in the thymus given their narrow pattern of expression, such as insulin or desmoglein 1b, were being expressed within the thymus (23,24). This expression of a wide variety of these tissue-specific antigens (TSAs) has been termed promiscuous gene expression. Further investigation revealed that this gene expression was concentrated in mTECs (25). While

expression of TSAs by mTECs was being discovered, the study of a rare genetic disorder led to the identification of an important agent in this promiscuous gene expression.

The autoimmune regulator (*Aire*) gene:

Beginning in the middle of the 20th century, sporadic case reports emerged describing rare children with both hypoparathyroidism and Addison's disease (adrenal insufficiency) (26,27). In time, this disease was recognized to be an autosomal recessive organ-specific autoimmune disorder (28), which was called autoimmune polyendocrine syndrome type 1 (APS1, also called autoimmune polyglandular syndrome type 1 or autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy [APECED]). Generally, diagnosis requires the presence of at least two of hypoparathyroidism, adrenocortical insufficiency, and mucocutaneous candidiasis (29). The first sign of disease, most often candidiasis, typically occurs in early childhood. As they age, patients frequently develop additional manifestations, such as ovarian failure, diabetes mellitus, vitiligo, and pernicious anemia (Table 1.1). Autoantibodies are prevalent in APS1 patients, including anti-cytokine antibodies. A majority of patients have anti-IL-17A, anti-IL-17F, and anti-IL22 antibodies, which are suspected to underlie their increased susceptibility to candidiasis. Anti-type I interferon antibodies, particularly against IFN-ω and IFN-α2, are present so consistently as to be considered diagnostic (30). Due to founder effects, there are elevated rates of APS1 in several populations, particularly Finns, Sardinians, and Iranian Jews (31).

By 1994, linkage analysis of 14 Finnish families had narrowed the location of the mutation responsible for APS1 to an approximately 500-kb region at chromosome 21q22.3 (32). A positional cloning approach culminated three years later with two groups publishing simultaneous papers, both of which described a novel gene, which they called autoimmune regulator (*AIRE*). Patients with APS1 but not related controls had homozygous or compound

Table 1.1. Clinical manifestations of APS1

Modified from (29).

Manifestation	Prevalence at age 10 yrs (%)	Prevalence at age 40 yrs (%)
Candidiasis	83	100
Hypoparathyroidism	65	87
Adrenal insufficiency	40	81
Ovarian failure		69
Alopecia	16	39
Vitiligo	9	31
Testicular failure		28
Pernicious anemia	3	28
Severe obstipation	10	26
Diabetes mellitus	3	23
Keratoconjunctivitis	18	22
Chronic diarrhea	13	22
Hypothyroidism	1	21
Hepatitis	12	18
Rash with fever	12	15

In addition, 77% of patients show enamel hypoplasia (33).

APS1 patients are also predisposed to developing oral squamous cell carcinoma (34).

heterozygous mutations in the gene (34,35). Since then, numerous mutations have been identified in *AIRE* in APS1 patients, located throughout the gene (36,37). Soon after *AIRE* was cloned, the murine orthologue, *Aire*, was identified on chromosome 10 and found to have substantial conservation with the human gene (38–40). Both human and murine *AIRE* consist of 14 exons, encoding a 57-kDa protein of 545 amino acids in humans and a 59-kDa protein of 552 amino acids in mice. The human and murine proteins share 73% sequence identity and 76% sequence similarity (38).

In both papers identifying *AIRE*, the authors noted the presence of two PHD-type zinc finger domains and hypothesized that Aire acts as a transcriptional regulator (34,35). Further analysis revealed that Aire contains a domain known to be DNA-binding in other proteins (41) and that Aire localizes to non-PML dots in the nucleus (42,43), strengthening this hypothesis. The critical step in deciphering the function of Aire, however, was creating and characterizing an Aire-deficient mouse. Targeted deletion of exon 2 of *Aire* led to organ-specific autoimmunity reminiscent of APS1, with autoantibodies against and infiltrates in multiple organs, most consistently the retina and salivary gland (44). Another targeted knockout, in which a mutation created a stop codon in exon 6, also resulted in organ-specific autoimmunity, though the affected tissues differed somewhat (45). In fact, the heterogeneity of APS1 is echoed in mice by the finding that there are significant differences in the severity and pattern of disease in different genetic backgrounds. C57BL/6 mice are relatively resistant to disease, developing non-fatal lacrimal, retinal, and salivary inflammation, whereas NOD mice are affected in more tissues and die by several months of age (46).

Critically, it was found that the autoimmune phenotype of *Aire*$^{-/-}$ mice required that Aire be deficient in the thymic stroma, not hematopoietic cells. Microarray-based global expression

analysis of mTECs revealed that Aire-deficient cells had decreased expression of many TSAs (44). This defect in promiscuous gene expression in the absence of Aire creates holes in the repertoire of antigens presented in the thymus, allowing the escape of T cells that would normally be negatively selected (47,48). In many cases, organ-specific autoimmunity can be pinpointed to loss of specific antigens in the thymus, including IRBP for retinitis (49), Obp1a for dacryoadenitis (50), mucin 6 for gastritis (51), and vomeromodulin for interstitial lung disease (52). Wholesale loss of antigen expression is not necessary for disease development, however. In heterozygosity, a dominant-negative mutation of *Aire*, also seen in some APS1 patients, leads to substantially decreased but not absent TSA expression; these mice still develop autoimmune disease, albeit with some differences from the *Aire*$^{-/-}$ mice (53). Indeed, even more mild reductions in *Aire* activity lead to increased escape of self-reactive T cells in mice (54). Furthermore, reduced thymic expression of the acetylcholine receptor and insulin in humans has been linked to increased risks of developing myasthenia gravis and diabetes, respectively (55,56). *Aire*-expressing mTECs are also involved in inducing self-reactive T cells to become regulatory T cells (57,58). Both negative selection of and regulatory T cell generation from T cells specific for TSAs occurs through both direct interactions with mTECs and indirectly via DCs that cross-present antigens from mTECs (59,60).

Tissue distribution of *Aire* expression*:*

In the first decade after its discovery, a number of groups used various techniques, including Northern blots, reverse transcription polymerase chain reaction (RT-PCR), in situ hybridization, and immunohistochemistry and immunofluorescence with anti-Aire antibodies, to assess expression of *AIRE* in both human and murine tissues (34,35,38,43,61–67) (Table S1.1). Collectively, these finding were disparate and at times outright contradictory. The primary point

of consensus, however, was that *Aire* is expressed in the thymus, particularly (and probably exclusively) in a subset of mTECs. Several markers can be used to identify mTECs. These cells are CD45[-], EpCAM[+], and MHC-II[+]. cTECs are Ly51[+] whereas mTECs are Ly51[-]. mTECs can in turn be divided into MHC-II[lo] and MHC-II[hi] populations, making up approximately 30% and 70% of mTECs, respectively, in young adult mice. The majority of these MHC-II[hi] mTECs are Aire[+]. Studies of mTEC ontogeny and experiments involving BrdU pulse-chase and fetal thymic organ cultures (FTOCs) revealed that mTECs begin as MHC-II[lo] and then mature to become MHC-II[hi] Aire[-] followed by MHC-II[hi] Aire[+] (68–70). Upregulation of CD80 and CD86 mirrors the increase in MHC-II. Aire[+] mTECs were originally thought to be a terminally differentiated cell, but a few groups found evidence that there was a post-Aire mTEC (71,72). An inducible *Aire-Cre* lineage-tracing mouse definitively revealed this post-Aire population, indicating that at least some Aire[+] mTECs downregulate first Aire and then MHC-II while retaining some TSA expression (73).

In 2008, in the wake of a pair of contradictory papers (66,67), it was unclear whether *Aire* was expressed in lymph node stromal cells. The issue was clarified using a transgenic reporter mouse in which an IGRP-GFP fusion protein was expressed in place of *Aire* in an *Aire*-containing bacterial artificial chromosome (BAC). The GFP expression generally recapitulated *Aire* expression in the thymus—nearly all Aire[+] cells were GFP[+] while most GFP[+] cells were Aire[+], a discrepancy we attribute to GFP having a longer half-life. This reporter mouse proved to be useful in detecting Aire in the periphery. A tissue survey involving immunofluorescent staining for GFP found almost all tissues to be devoid of GFP. Rare GFP[+] cells, however, were present in the spleen and lymph nodes (74). These GFP[+] cells, which were confirmed to be Aire[+] by immunofluorescence and named extra-thymic *Aire*-expressing cells (eTACs), are also present

in human secondary lymphoid tissues (75). Like mTECs, eTACs express a large number of TSAs in an Aire-dependent fashion and help control self-reactive T cells. *Aire*-expressing cells therefore play a role in both central and peripheral tolerance (Figure 1.1). Interestingly, Aire induces distinct sets of TSAs in these two cell types. eTACs were originally classified as stromal cells, like mTECs, because they express low levels of CD45 and are radioresistant. Upon further investigation, however, it became clear that these cells are actually of hematopoietic origin and are a previously unidentified cell type within the conventional DC lineage (75).

Aire is expressed in both testes and early in development. Using a monoclonal anti-Aire antibody, Aire was observed in some spermatogonia and spermatocytes (but not spermatids or Sertoli cells). There, Aire did not control TSA expression. In $Aire^{-/-}$ mice, however, an early wave of germ cell apoptosis in the testis proceeds abnormally (76). The discovery that Aire is expressed in early embryogenesis occurred after the Matsumoto group was surprised to discover global GFP expression in *Aire-Cre* x *loxP-STOP-loxP-eGFP* mice. Using RT-PCR and an *Aire* GFP reporter, they found that *Aire* is expressed throughout early development: at the 2-cell-stage, in the blastocyst, in about half of embryonic stem (ES) cells, and in the epiblast at embryonic day 6.5 (E6.5). They also observed *AIRE* transcript in human ES cells (77). The role of *Aire* in early embryogenesis is wholly unknown, though the absence of obvious developmental defects in $Aire^{-/-}$ mice suggests that its functions at that stage are at least largely redundant.

In summary, *Aire* has an unusual and highly restricted pattern of expression: mTECs, eTACs, some developing sperm, and early in embryogenesis. Nearly all study has focused on its role in the first two cell types, where it promotes TSA expression and thereby immune tolerance.

Figure 1.1. The role of *Aire* in T-cell tolerance.

(A) Expression of *Aire* in mTECs leads to induction of a diverse set of TSAs, which are displayed on MHC-I and -II, where they are bound by the TCRs of self-reactive, cognate T cells.

(B) Maintenance of T-cell immune tolerance involves mechanisms in both the thymus and the periphery. In both settings, *Aire*-expressing cells present a diverse array of TSAs. In the thymus, mTECs drive self-reactive thymocytes towards cell death or a regulatory T cell fate. In secondary lymphoid tissues, eTACs induce anergy or cell death in self-reactive T cells. Regulatory T cells are also an essential component of peripheral tolerance. Tconv means conventional T cell; Treg means regulatory T cell.

Figure 1.1

A
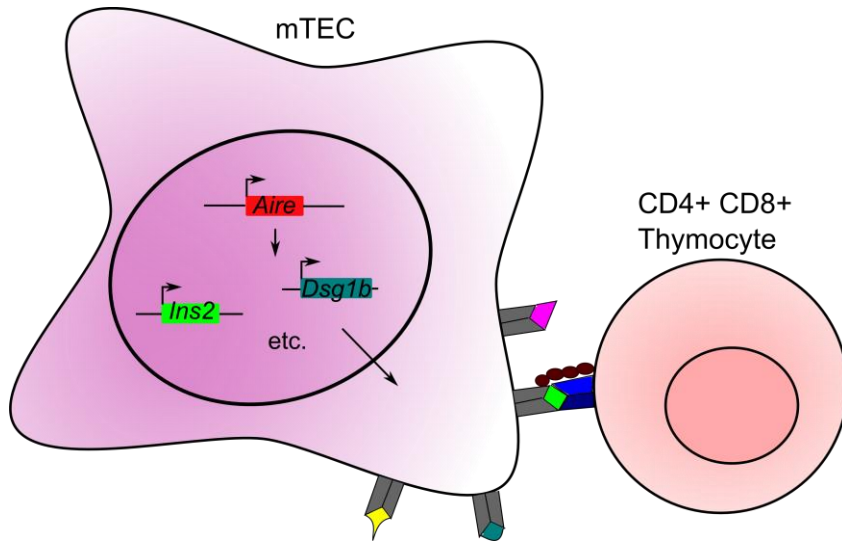
B
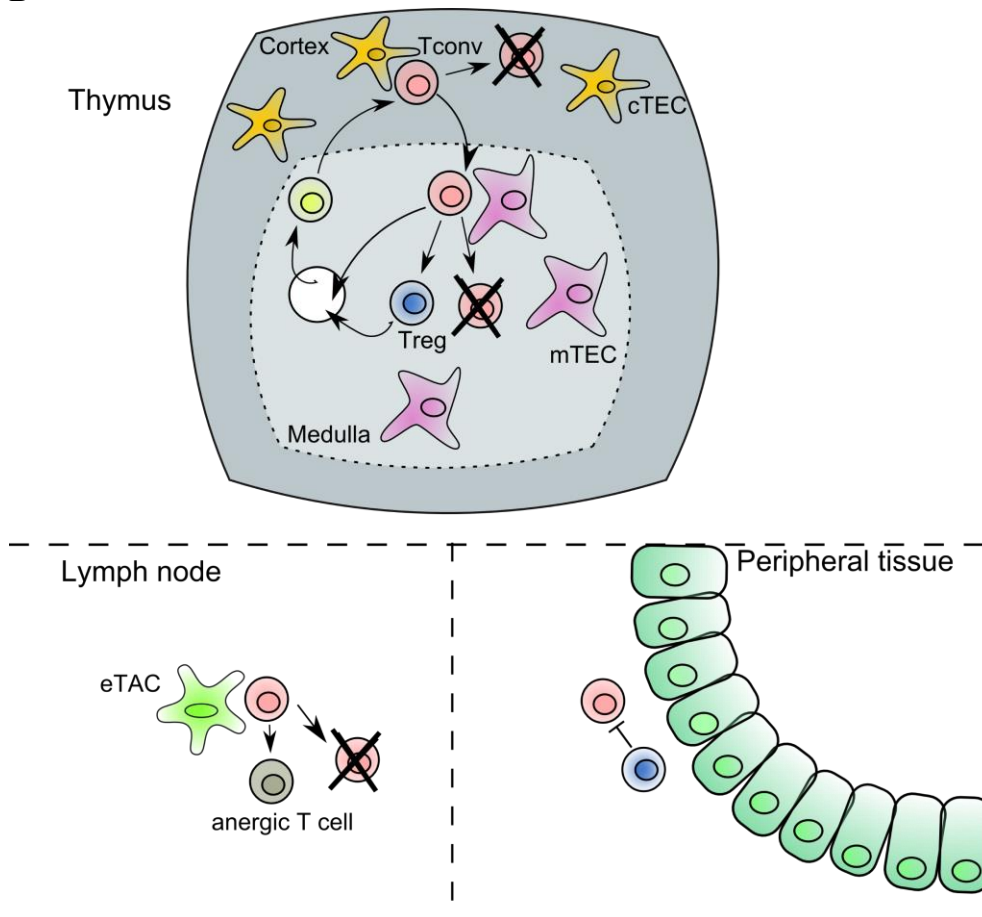
Molecular mechanisms of *Aire*:

The mechanism by which *Aire* drives TSA expression remains an active area of research. The Aire protein contains four major domains and several smaller motifs (78). The first 100 or so amino acids of the protein make up an HSR domain, first identified in the Sp100 protein (79). (All amino acid spans given here are for human Aire; domains are of very similar size and spacing in murine Aire.) Subsequent re-analysis of this domain revealed it to be a type of CARD domain based on predicted structure (80). Aire has been found to multimerize using this domain (81). Immediately following the HSR domain is a bipartite nuclear localization sequence (NLS) spanning amino acids 110-133 and a second NLS spanning amino acids 155-167 (82). Amino acids 189-280 make up a SAND (Sp100, Aire, NucP41/75, Deaf1) domain. The SAND domain is often involved in direct DNA binding; however, a four-residue motif critical to DNA binding in other SAND domain-containing proteins is not present in Aire. As such, it was hypothesized that in Aire this domain might be involved in protein-protein interactions. Recently, this hypothesis was affirmed by the finding that the SAND domain of Aire interacts with Atf7ip, which, in complex with MBD1, targets Aire to regions with methylated CpGs (83). There are two PHD domains, at amino acids 299-340 (PHD1) and amino acids 434-475 (PHD2), separated by a proline-rich region. PHD1 allows Aire to bind to unmethylated lysine 4 of histone 3 (H3K4me0), which is thought to help target Aire to repressed genes and is crucial for its function (84,85). The PHD2 domain does not mediate chromatin binding, but its deletion instead leads to decreased interaction with certain other protein partners of Aire and diminished TSA induction (86). Aire also contains four LXXLL motifs, scattered throughout the protein.

Some early studies hypothesized that Aire is a transcription factor. In vitro experiments found evidence for sequence-specific DNA binding by Aire (87) and transcriptional

transactivation by the two PHD domains (81). Chromatin immunoprecipitation (ChIP) followed by PCR suggested that Aire binds at the promoters of *Aire*-dependent TSAs (88). However, another study found that Aire interacts with DNA in a non-sequence-specific fashion (85) and a yeast one-hybrid screen performed several years ago by Kellsey Johannes of the Anderson lab failed to find any sequences to be preferentially bound by Aire (data not shown). Moreover, the tremendous diversity of the genes whose thymic expression depends on Aire, and the variegated expression of these genes, is inconsistent with Aire acting as a traditional transcription factor, though it certainly affects transcription of numerous genes.

Analyses focused on identifying the protein partners of Aire have found evidence that Aire enhances target gene expression via both transcriptional elongation and pre-mRNA processing pathways. Anti-Aire co-immunoprecipitation followed by mass spectrometry using cultured cells expressing ectopic Aire resulted in the identification of dozens of Aire binding partners, which can be grouped into four categories: nuclear transport, chromatin binding, transcription, and pre-mRNA processing (89). A number of these factors are known to be involved in transcriptional elongation, including the DNA-PK-containing complex, and DNA-PK-deficient mTECs have defective TSA expression. Furthermore, Aire promotes TOP2a-initiated double-strand breaks, which can help support more efficient transcription by relaxing DNA (89). Also, in vitro assays have found that Aire can induce transcriptional elongation but not initiation, via binding and recruitment of the elongation factor P-TEFb (90). More recently, microarray analysis (with multiple probes per gene) showed that RNA polymerase II (PolII) stalled 50-100 bp from the transcriptional start site of Aire-activated genes in Aire-deficient mTECs, whereas when Aire was present PolII was able to proceed (91). Aire also interacts with a number of different splicing factors and, for at least some TSAs, there appears to be a

substantial decrease in spliced but not unspliced transcripts in the absence of Aire (89,92). In addition, Aire and some microRNAs also appear to cooperate to induce promiscuous gene expression (93).

Two groups have presented evidence that the activity of Aire is modulated by post-translational modifications. Phosphorylation of Aire by DNA-PK appears to be necessary for normal induction of gene expression by Aire (94). In addition, Aire can be acetylated by p300 at several sites in the NLS and SAND domain; Aire with mutations that mimic this acetylation has reduced transactivation activity in vitro (95). Conversely, a recent paper finds that the deacetylase Sirt1 is highly expressed in Aire$^+$ mTECs, and expression of Aire-dependent TSAs is markedly reduced in mice lacking Sirt1 in the thymic stroma (96).

Intriguingly, ectopic *Aire* expression is sufficient to induce TSA expression in cells such as 293T cells, which lack other factors characteristic of mTECs and eTACs. However, the precise set of genes regulated by Aire is greatly affected by the other factors present in the cell, such that mTECs, eTACs, 293T, and pancreatic beta cells (the latter two expressing ectopic *Aire*) show only a modestly overlapping set of Aire-induced genes (74,83,97). In mTECs, there is a large pool of Aire-independent TSAs in addition to the numerous Aire-dependent TSAs (98). Although as a population, mTECs express thousands of TSA genes, in any given *Aire*-expressing cell only a fraction of these genes are transcribed, in an apparently stochastic manner (99,100). This expression, however, is not completely random. The TSAs expressed by a given cell exhibit clustering, wherein certain sets of TSAs are typically expressed together (101).

The regulation of *Aire* expression:

Aire is expressed in a very limited set of tissues, with expression in mTECs and eTACs being most prominent. The expression of a gene can be regulated at multiple levels: the rate of

transcription, RNA processing, mRNA stability, and the rate of translation. Post-translational means of controlling protein activity include proteasomal degradation; proteolytic cleavage; activating or inhibitory post-translational modifications such as phosphorylation, acetylation, or ubiquitination; binding of activating or inhibitory protein partners; and localization to a particular cell compartment. There are no reports of dynamic changes in *Aire* activity in vivo independent of changes in transcript levels, suggesting that *Aire* expression is regulated primarily at the transcriptional level. Historically, it was thought that the most prevalent point of regulation was at transcriptional initiation (102). In recent years, however, it has become clear that control of transcriptional elongation is very important, particularly for many developmentally regulated genes (103).

Transcription is regulated by the interactions of various *trans*-acting regulatory factors— such as proteins and long noncoding RNAs—and *cis*-regulatory (DNA) elements (102). Classes of protein *trans*-acting factors include general transcription factors, activators, and coactivators. The *cis*-regulatory elements (CREs) can be divided into promoters, each of which contains a core promoter and proximal promoter elements, and distal CREs. General transcription factors are the essential elements of the transcriptional machinery and include PolII and associated proteins such as TFIIA, TFIIB, and TFIID. These assemble at the core promoter, which defines the transcriptional start site, to form the pre-initiation complex (PIC). Only a very low level of transcription can be supported by the PIC alone. However, transcription is greatly increased by the addition of activators. An activator protein generally consists of both a DNA-binding domain, which allows it to bind in a sequence-specific fashion to one of the proximal promoter elements, and a separate activation domain that allows it to stimulate transcription. Activators can stimulate transcription through a variety of mechanisms, including increasing PIC formation, promoting

19

transcriptional initiation, promoting transcriptional elongation, or modifying chromatin. Activators include members of the Hox, Ets, and POU protein families, among others. Coactivators, the third class of *trans* elements, do not bind DNA directly. Rather, they rely on protein-protein interactions to gather at the promoter, where they modulate activator activity.

Distal CREs are generally classified as enhancers, silencers, insulators, and locus control regions (LCRs). Much like proximal promoter elements, enhancers bind activators (and in turn coactivators) that act to increase transcription. Classically, enhancers are defined as being able to heighten transcription independent of the position or orientation of the enhancer. Some enhancers are located very far, even hundreds of kb, from their target gene. They are thought to act through a DNA looping mechanism which allows direct contact between the complex of proteins bound to the enhancer and the proteins at the promoter. Silencers are basically the opposite of enhancers, binding repressor proteins that act to decrease transcription. Insulators act to prevent the transcription status of genes in one region from spreading to those in the adjacent region. CTCF is classically associated with insulators. Finally, LCRs are groups of distal CREs that are able to produce the tissue-specific physiological levels of expression of a gene cluster. By definition, they act in a position-independent, copy-number-dependent manner. Well-characterized LCRs include the CD2, Th2, and β-globin LCRs.

Cell-type-specific expression of genes relies on combinatorial, synergistic activation by *trans* elements. Any single transcription factor involved in driving a particular gene might be expressed in many tissues, but the full set of activators and coactivators necessary to support transcription of a given gene are only present in the specific tissue in which the gene is expressed.

The expression pattern of *Aire* is highly unusual, restricted to a small number of cell types that are nevertheless members of vastly different lineages, including undifferentiated

embryonic cells, hematopoietic cells, differentiated non-hematopoietic somatic cells, and germ cells. Ectopic expression of *Aire* results in substantial changes in gene expression, and there is some evidence that Aire itself can stimulate apoptosis (104,105). This, together with the harmful effects of Aire deficiency, suggests that achieving appropriate expression of *Aire* is critical. To this point, however, relatively little is known about how *Aire* transcription is regulated.

The most expansive body of work touching on this regulation has been a collection of studies connecting TNF receptor superfamily signaling and non-canonical NF-κB to the development of *Aire*-expressing mTECs. The development of mTECs requires cross-talk with RANKL (*Tnfrsf11*)-expressing hematopoietic cells, namely lymphoid tissue inducer cells and γδ T cell progenitors during fetal development (70,106) and positively selected thymocytes after birth (107,108). In RANK (*Tnfrsf11a)*-deficient mice, there are fewer mTECs and a more-or-less complete loss of Aire[+] mTECs (70,109). Conversely, the addition of exogenous RANKL to FTOCs leads to increased numbers of mTECs (MHC-II[lo], MHC-II[hi], and Aire[+]). In addition, mice deficient in OPG (*Tnfrsf11b*), a secreted RANKL-binding protein, have enlarged thymic medullae and more mTECs, including more Aire[+] mTECs (107). Furthermore, administration of anti-RANK antibody in mice results in depletion of mTECs, particularly Aire[+] mTECs (110). Fellow TNFRSF member CD40 is also expressed by mTECs and appears to cooperate with RANK. Although loss or exogenous stimulation of CD40 has only a modest impact on mTEC frequency—somewhat more impact on immature than mature mTECs—deletion or stimulation of both RANK and CD40 has a greater effect than manipulating RANK alone (107,111,112). A third TNFRSF member, LTβR, is also expressed by mTECs and plays a role in mTEC development (71,112–114).

NF-κB is a family of transcription factors that consists of *Rela* (p65), *Relb, Rel* (c-Rel), *Nfkb1* (p105/p50), and *Nfkb2* (p100/p52). They play an important role in inflammation and immunity, among other processes (115). They form various dimers that can bind κB sites and, usually, drive transcription. Normally, NF-κB members are kept out of the nucleus by various inhibitory proteins. Activation of NF-κB can be broadly divided into two pathways, the canonical (or classical) and the non-canonical (or alternative). In the stereotypical canonical pathway, degradation of IκB proteins upon phosphorylation by IKK leads to nuclear translocation of p50/p65 dimers. In contrast, in the stereotypical non-canonical pathway, induction of proteolytic processing of p100 to p52 leads to nuclear translocation of p52/RelB dimers (116). A number of studies have identified a role for non-canonical NF-κB in mTEC biology. Mice with global deletion of *Relb* have disrupted thymic architecture and lack Aire[+] mTECs (109,117). In FTOCs, RANK stimulation induces *Relb* and leads to *Relb*-dependent increases in mTECs (111). Similarly, thymi from *Nfkb2*-deficient mice have decreased *Aire* and TSA expression and develop autoimmunity affecting various organs that can be transferred by thymic transplant (118). The NF-κB signaling pathway members NIK and TRAF6 have also been implicated in RANK-dependent *Relb* induction and *Aire*-expressing mTEC development (111,112,119,120).

Investigation of *cis*-regulatory elements controlling *Aire* has been limited to a few studies performed by the Peterson group. They used in vitro assays with various *AIRE* promoter luciferase constructs to identify a basal promoter that begins somewhere between position -350 and -248 relative to the translation start site; a TATA box is at -162 and the transcriptional start site is at -128. Mutant analysis and elecrophoretic mobility shift assays (EMSAs) suggest that there are AP-1, NF-Y, Sp1, and two Ets proximal promoter elements (121,122). Like many

promoters, the proximal human *AIRE* promoter is a CpG island (the murine *Aire* promoter falls a little short of the definition of a CpG island), methylation of which is generally associated with an absence of expression. This region is scantly methylated in mTECs, consistent with it being permissive for *Aire* expression. However, the region is also unmethylated in several other tissues in which *Aire* is not expressed, indicating that although perhaps demethylation of this region is required for *Aire* expression, it is but one requirement (123).

Although these studies provide important insights, a number of questions regarding the transcriptional regulation of *Aire* remain. It is not known the degree to which the loss of Aire[+] cells in mice with defects in non-canonical NF-κB reflects a direct role for NF-κB in driving *Aire* transcription versus an indirect role in which NF-κB supports maturation of mTECs to the point that other factors become expressed that do directly drive *Aire* expression. It is unclear what additional factors regulate *Aire* in *trans*. It is unknown whether there is a common set of transcription factors and CREs that induce *Aire* in all cell types in which it is expressed or whether some or all of these elements are unique to a particular cell type.

We approached our study of the transcriptional regulation of *Aire* by focusing on identifying distal CREs that govern *Aire* expression. In so doing, we sought to increase our understanding of immune regulation, to facilitate future studies of *Aire* regulation in *trans* by finding candidate transcription factor binding sites, and to aid in the diagnosis of APS1 by finding sites in which noncoding mutations might disrupt *Aire* expression and lead to disease. In the following chapters, we describe a novel CRE that is necessary for *Aire* expression, identify an unexpected additional role for this CRE, and present evidence for additional *Aire* regulatory regions.

Conclusion:

The immune system is essential to identify and eliminate pathogens and neoplastic cells. An essential component of this system are the antigen-receptor cells of the adaptive immune system. The very mechanism that provides each T and B cell with a unique receptor, however, also leads to the generation of a substantial number of potentially self-reactive cells, necessitating an active process to preserve immune self-tolerance. Critical to this immune tolerance is the *Aire* gene. In both humans and mice, *Aire* is expressed in a subset of mTECs and eTACs, where it induces the expression of a diverse array of TSAs, supporting elimination and control of self-reactive T cells. The regulation of *Aire* expression is poorly understood, though it is known that RANK and non-canonical NF-κB signaling are important for the development and maintenance of *Aire*-expressing mTECs. We focused on identifying distal *Aire* CREs, none of which were previously known. In subsequent chapters, we present new findings on the *cis*-regulation of *Aire.*

Table S1.1. Selected published findings on *AIRE* expression

Reference	Species	Methods	Findings
(34)	Human	NB	Strong bands in thymus, pancreas, and adrenal cortex; weak bands in spleen, LN, PBMC, fetal liver, thyroid and testis
(35)	Human	NB	Strong bands in thymus and LNs; faint bands in appendix and fetal liver
(43)	Human	IHC	Strong staining in Hassall's corpuscles in thymus; moderate staining in mTECs; also present in splenic red pulp and PBMCs
(61)	Human	ISH; IHC	Expressed in medulla, LN medulla and paracortex, spleen, and fetal liver. Also observed that 90% of thymic Aire$^+$ cells were cytokeratin$^+$ and MHC-II^{hi} and that 20% were CD80$^+$/CD86$^+$
(62)	Human	RT-PCR	Highest expression in mTECs, followed by whole thymus and LNs, still lower in PBMCs, spleen, and purified CD14$^-$ cells
(38)	Mouse	NB; RT-PCR ISH	NB: Thymus RT-PCR: Consistently in E11 fetuses, spleens, ES cells, and often in lungs ISH: In mouse embryo, detected from E14.5 in small number of cells in thymic medulla
(63)	Mouse	NB; RT-PCR	NB: not detected RT-PCR: Highest in thymus, next highest in ovary, lower still in lung, testis, kidney, adrenal, thyroid, and heart
(64)	Mouse	RT-PCR ISH, IHC	Detected by all three in thymus, spleen, LNs, and brain; numerous other tissues detected by one or two of the methods
(65)	Mouse	RT-PCR ISH, IHC	RT-PCR: high in thymus, moderate in spleen and testis, and low in several other tissues; ISH: only in thymic medulla IHC: mTECs, other thymic cells, LNs, spleen, lymphocytes, and various other tissues In fetus, thymic expression from E14.5; fetal liver expression at E14.5 that declines with time
(66)	Mouse	RT-PCR	Highest expression in thymus, followed by LNs, followed by spleen; in LNs, much higher in CD45$^-$ fraction than CD45$^+$
(67)	Mouse	FC, IF RT-PCR	FC: almost all thymic CD45$^-$ MHC-II^{hi} Ly51^{lo} cells IF: scattered cells in medulla and cortico-medullary junction of thymus; not present in LNs or spleen RT-PCR: By far highest in CD45$^-$ cells in thymus; weakly in thymic and splenic DCs

NB: Northern blot; IHC: immunohistochemistry; ISH: in situ hybridization; RT-PCR: reverse transcription polymerase chain reaction; FC: flow cytometry; LN: lymph node

Chapter 2: Identification of an *Aire*-regulating *cis*-regulatory element required for immune tolerance

"The true method of knowledge is experiment…"

-William Blake, in "All Religions Are One"

Overview:

Aire plays an essential role in establishing immune tolerance by inducing the expression of a diverse set of TSAs in mTECs. In post-embryonic mice, *Aire* expression is highly restricted, but how this pattern of expression is achieved is unclear. It is known that non-canonical NF-κB stimulated by RANK is essential for the development and maintenance of *Aire*-expressing mTECs, but the *cis*-regulation of *Aire* is poorly understood. We aimed to identify distal *cis*-regulatory elements (CREs) governing *Aire* expression. We located a highly conserved candidate *Aire* CRE, ACNS1, that is flanked in mTECs by enhancer-associated histone marks. We show that this element is NF-κB-responsive. We also identify a 4.3-kb region that is sufficient to reproduce the cell-type specificity of *Aire*. We show ACNS1 is required for *Aire* expression in both mTECs and eTACs. Finally, we observe that loss of ACNS1 leads to spontaneous autoimmunity.

Introduction:

The establishment and maintenance of immune tolerance requires the successful detection and control of self-reactive T cells (17). In the thymus, mTECs play a crucial role in

this process, presenting a varied repertoire of antigens, causing the negative selection of high-affinity self-reactive thymocytes and driving those with more moderate affinities to a regulatory T cell fate (25,47,48,57,58). mTECs express and display thousands of TSAs, some of which are Aire-independent and some of which are Aire-dependent, thereby largely reproducing within the thymus the antigenic profile of the peripheral self (25,44). *Aire* also induces a significantly different set of TSAs in eTACs, a rare CD45^{int}, MHC-II^{hi}, EpCAM^{+}, CD86^{-}, CD11c^{lo} hematopoietic cell that develops from the conventional DC lineage and is present in secondary lymphoid tissues; these cells likewise can tolerize self-reactive T cells (74,75). People with mutations in *AIRE* develop the autosomal recessive organ-specific autoimmune disorder, APS1, highlighting the physiologic importance of this gene (34,35). Aire-deficient mice also develop spontaneous organ-specific autoimmunity (44,45).

As discussed in Chapter 1, *Aire* itself has an unusual and restricted pattern of expression. It is most prominently expressed in mTECs and eTACs, as noted above; it is also present in a subset of developing spermatocytes (76) and early in embryogenesis (77). Ectopic *Aire* introduced into cells as disparate as pancreatic beta cells and 293T cells induces expression of large numbers of genes (83,89,97), illustrating the importance of not only ensuring appropriate initiation of *Aire* expression but also preventing inappropriate activation. Previous research has demonstrated the importance of RANK and CD40 signaling to the development and maintenance of *Aire*-expressing mTECs (70,107,111) and RANKL blockade is sufficient to deplete these cells in vivo (110). In addition, various elements of the non-canonical NF-κB pathway, which can be activated by RANK or CD40, have been found to be essential for the development of *Aire*-expressing mTECs (109,111,117–120). The *AIRE* promoter has also been characterized in vitro

(121,122). Still, much about the regulation of *Aire* expression remains to be determined, including the identity of any distal CREs.

Gene expression can be controlled through modulation of multiple processes, including transcription, transcript processing, translation, and protein degradation. Although there does appear to be a role for post-translation modifications in shaping the level of Aire activity (94–96), it seems that the predominant mechanism of regulating *Aire* is transcriptional: we have not observed *Aire* transcription without corresponding protein expression. In addition, various reporter constructs, both transgenic and knock-in, appear to faithfully recapitulate *Aire* expression (74,75,77,124). That recapitulation suggests transcription is the dominant point of *Aire* regulation because such transgenes would be expected to be transcribed similarly to *Aire* but would not be expected to undergo any subsequent regulatory steps in common with *Aire*.

Transcriptional regulation itself is a complicated dance, involving both *trans*-acting regulatory elements—DNA-associated molecules such as transcription factors—and CREs—DNA regions on the same chromosome as the gene. These CREs are divided into those that are proximal to the transcription start site and make up the promoter and those that are distal, from 1 kb to as much as 1 Mb away (125). Classes of distal CREs include enhancers, which increase transcription; silencers, which decrease transcription; and insulators, which block the spread of a particular set of genomic conditions, such as active transcription and heterochromatin. Collectively, these distal CREs, found scattered throughout introns and intergenic regions—more densely in gene-rich regions—appear to make up at least several percent of the genome, exceeding the amount of sequence that encodes proteins (126).

Although definitively determining that a region is a distal CRE requires experimental evidence of function, several methods have been found to be reasonably successful at predicting

CREs. One approach is to look for conserved non-coding sequences (CNSs), which takes advantage of the fact that CREs, like protein-coding genes, are subject to purifying selection, which tends to stabilize the sequence, relative to the drift observed in non-functional sequence. Purifying selection occurs because mutations more often have a negative effect than a positive effect on the organism's fitness (127). Consequently, conserved areas are enriched for functional elements (128), though even between relatively phylogenetically close species, such as humans and mice, a large proportion (perhaps half) of functional elements are not conserved (129,130). Thus, analysis of sequence conservation is a useful but imperfect tool.

Another widely used method for identifying candidate CREs is ChIP-seq against histone modifications or other DNA-binding proteins associated with enhancers (127,131). Monomethylated lysine 4 of histone 3 (H3K4me1) is frequently present at enhancers. H3K27ac is an even more useful marker, as it tends to be present at enhancers only in the tissues in which the enhancer is active (132). The acetyltransferase p300 is also frequently found at enhancers (133). It has also recently been observed that active enhancers are frequently the site of transcription (134). Finally, the recently developed assay for transposase-accessible chromatin using sequencing (ATAC-seq) permits DNase I hypersensitivity-like mapping of open chromatin but with finer mapping and using many fewer cells (135).

We chose to use sequence conservation analysis and ChIP-seq against H3K27ac to identify candidate *Aire* CREs. We identified a promising candidate a few kb from *Aire* and found it was NF-κB-responsive. We proceeded to functional assays and determined that this element is required for *Aire* expression and immune self-tolerance.

Results and Discussion:

Identification of candidate Aire cis-regulatory elements

Distal CREs, by definition, are located some distance from the promoters of genes that they control. Enhancers have been identified up to 1 Mb from their target gene (125). Although CREs are usually a fair bit closer than that, the possibility that they may be so far away makes for a rather daunting search in the absence of any additional guiding information. Fortuitously, an *Aire* reporter mouse we generated several years ago allowed us to narrow our search to a smaller region. The *Adig (Aire*-driven IGRP-GFP) reporter mouse was created by injection of a modified BAC into a fertilized oocyte. The BAC, RP23-461E7, is from a BAC library created from genomic DNA pooled from three female, C57BL/6J mice (136). This BAC contains an approximately 180 kb insert that is roughly centered on *Aire*. Additional genes contained in this BAC are *Icosl, Dnmt3l, Pfkl, 1810043G02Rik,* and part of *Trpm2*. In the *Adig* construct, genomic sequence beginning with the coding sequence of *Aire* in exon 1 and extending partway into exon 3 was replaced with a reporter cassette consisting of a β-globin intron, an IGRP (*G6pc2*)-GFP fusion gene, and an SV40 poly A sequence ((74) and Fig. 2.1A).

Analysis of GFP expression in this transgenic mouse revealed faithful recapitulation of *Aire* expression in mTECs and eTACs ((74) and Fig. 2.1B). More specifically, approximately 95% of Aire$^+$ cells are GFP$^+$ whereas approximately 75% of GFP$^+$ cells are Aire$^+$ (74). We hypothesize that this discrepancy is due to a greater half-life of IGRP-GFP compared to Aire, such that GFP lingers in post-Aire cells**.** This same 461E7 BAC has also served as the backbone of additional *Aire*-recapitulating transgenes, including an *Aire-Cre* and an *Aire-DTR* (73). This recapitulation of *Aire* expression strongly suggests that all essential CREs governing *Aire* are within the 180-kb span contained within the BAC (as well as implying that there is not an

essential CRE in the first two introns of *Aire*). Going forward, we focused our search on this genomic region.

We began by using two complementary approaches to identify conserved non-coding sequences in this region. The first was the online sequence alignment tool mVista (137). In this tool, we aligned regions homologous to the 461E7 BAC, drawn from seven disparate placental mammals (Table S2.1), using the LAGAN global alignment algorithm (138) followed by the RankVista algorithm (139) to identify conserved regions based on this alignment. The analysis was restricted to placental mammals because of incomplete synteny between *Aire*-flanking regions when comparing placental mammals with other vertebrates. *Aire* resides within a large block of synteny among placental mammals. For example, in comparing mice and humans, the region of synteny extends approximately 2.5 Mb in one direction from *Aire* and 500 kb in the other direction. In contrast, *Aire* was flanked by different genes in at least one direction when comparing placental mammals to marsupials, monotremes, and other more distant vertebrates. We anticipated that the lack of conserved genes surrounding *Aire* when comparing other vertebrates to placental mammals would hinder genome alignment.

We used the Ensembl genome browser to examine the flanking genes of *Aire* in a variety of vertebrate species whose genomes have been sequenced, in order to deduce the phylogeny of the genomic environment of *Aire* (Figure S2.1A). In determining the *Aire*-flanking region phylogeny, we noted that it appears that in the ancestral jawed vertebrate, *Aire* was adjacent to the ancestor of the four members of the *Sp100* family. Similarities between the structure of Aire and Sp100 have long been noted (41); Aire and Sp100 proteins share certain protein domains, including the SAND domain and PHD-like domain (Figure S2.1B). We hypothesize that *Aire* arose as a gene duplication of the ancestral *Sp100*. In addition, we note that *Aire* is found in all

Figure 2.1. Identification of candidate *Aire cis*-regulatory elements.

(A) Schematic of the *Adig* transgene, in which an IGRP-GFP cassette replaces the coding portion of *Aire* exon 1, all of exon 2, and part of exon 3 in the RP23-461E7 BAC. (B) Immunofluorescent staining for GFP (green) and Aire (red) in frozen thymic and splenic sections from *Adig* mice. Scale bars are 50 μm. (C) Alignment of conservation and ChIP-seq data for the region spanned by the 461E7 BAC (mm10, chr10:77961078-78140698). The upper three rows show unique reads from H3K27ac ChIP-seq of GFP$^+$ and GFP$^-$ mTECs from *Adig* mice and the D10 T cell line. Below are selected H3K27ac tracks (signal tracks, dense setting) of ENCODE/LICR, publicly available on the UCSC genome browser; from top to bottom: brain, bone marrow, brown adipose tissue, heart, ES-Bruce4, kidney, limb, liver, placenta, small intestine, spleen, testis, and thymus. (These tracks were aligned to the mm9 mouse genome but are here aligned with H3K27ac reads generated and aligned to mm10, as this 180-kb region differs between mm9 and mm10 by a single nucleotide.) Below that, the PhastCons placental mammal track and the RepeatMasker track of the UCSC Genome Browser; Refseq genes; conserved noncoding sequences (p<0.01) identified using mVista; and non-exon PhastCons conserved elements. The expanded region below shows the overlap of a particular mVista-identified CNS and three PhastCons elements along with H3K27ac ChIP-seq. Depicted H3K27ac tracks of mTECs are representative of three analyzed samples each.

Figure 2.1
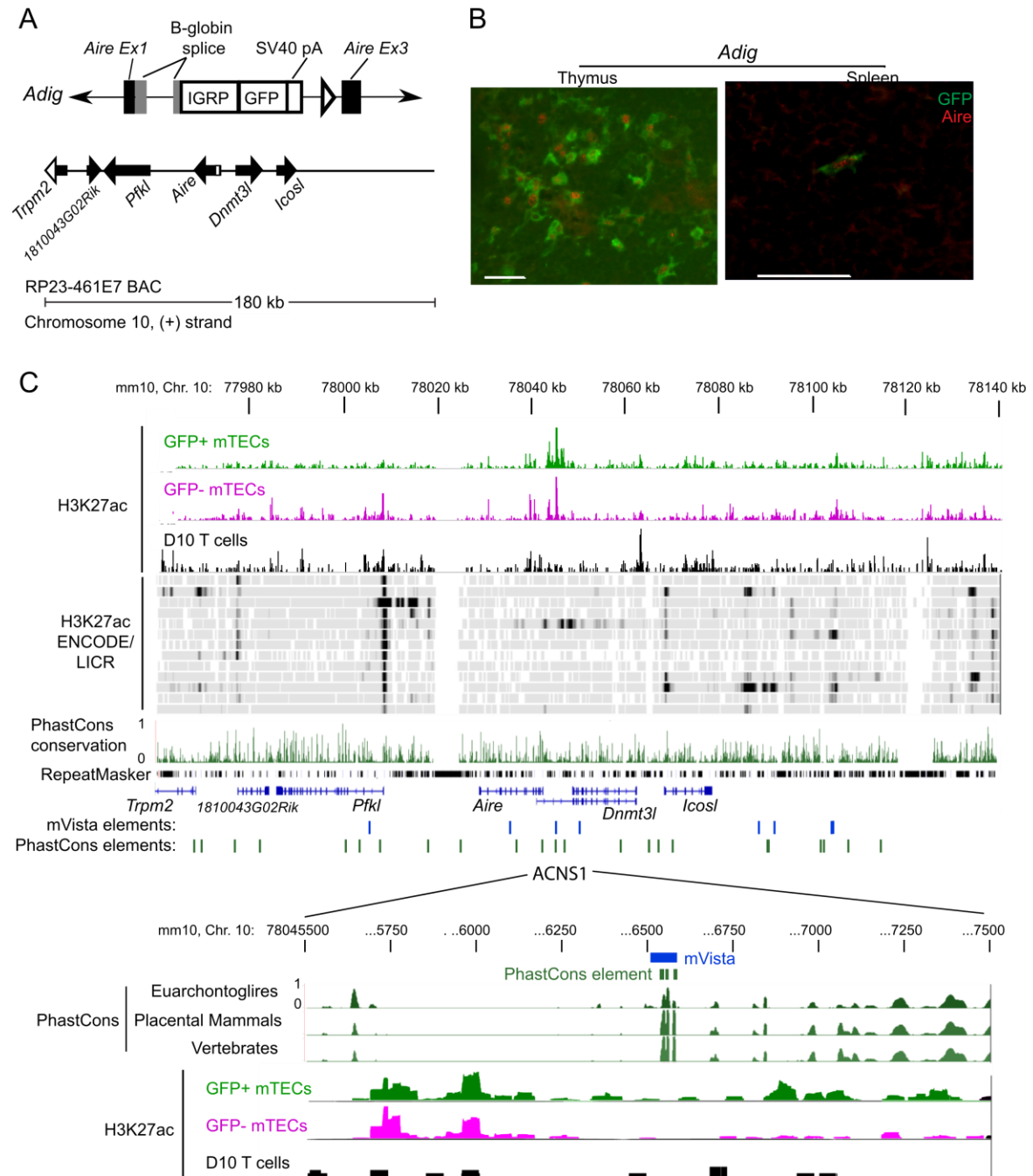
jawed but not jawless vertebrates, indicating that the advent of *Aire* coincides, at least generally, with the evolution of V(D)J-recombination-generated antigen-receptor lymphocytes.

The mVista analysis of the seven placental mammal sequences identified eight conserved non-coding sequences in the 180-kb region ($p < 0.01$) (Table S2.2). We complemented this analysis by comparing these CNSs to the non-exon conserved elements already identified by the PhastCons algorithm (140), which is publicly available through the UCSC Genome Browser (Figure 2.1C). Our attention was caught by a location at which an mVISTA-identified element and a PhastCons element overlap. We called this site *Aire* CNS1 (ACNS1). (There are three levels of PhastCons analysis: euarchontoglires, placental mammals, and vertebrates. These particular overlapping PhastCons elements are present all three tracks.)

In order to investigate further the possibility that ACNS1 is an active enhancer in mTECs, as well as to identify candidate enhancers that may not be conserved, we decided to take a non-biased genome-wide approach by performing ChIP-seq against H3K27ac. There have been several recent efforts to comprehensively identify CREs in a variety of human and mouse tissues (141–144). These publicly available data sets are a valuable resource. However, none of them includes an *Aire*-expressing tissue, limiting their utility in identifying *Aire* CREs. (A number of them include thymus, but since the vast majority of cells in the thymus are thymocytes, the epigenetics of rarer thymic epithelial cells are masked.) We sorted GFP$^+$ and GFP$^-$ mTECs from the *Adig* (*Aire-GFP* reporter) mouse (Figure S2.2A and B). Most GFP$^+$ mTECs also express *Aire*; most of the rest have probably only recently downregulated *Aire* (73). In contrast, the GFP$^-$ mTECs are a more diverse population, consisting of both immature mTECs that have not yet expressed *Aire* and mTECs that previously expressed *Aire*. We anticipated that this diversity might make it difficult to interpret differences in ChIP-seq data between the two mTEC subsets

that we sorted, so we also used the Th2-skewed D10 T cell line as a non-*Aire*-expressing control. In light of the relatively small numbers of mTECs that we were able to isolate (on the order of 100,000 cells), we used a ChIP-seq approach that has been shown to work with as few as 10,000 T cells (145).

The graph of unique reads over the 461E7 BAC region revealed substantial enrichment of H3K27ac near but not immediately at ACNS1 (Figure 2.1C). Genome-wide analysis has shown that the H3K27ac peaks are typically offset from the CRE itself (i.e., where the transcription factors bind) by a few hundred base pairs (143). No other locations within the *Aire*-recapitulating region showed consistent enrichment for this modification. In contrast to mTECs, D10 cells did not have H3K27ac enrichment near ACNS1. We also examined a number of publicly available H3K27ac analyses created as part of the ENCODE consortium (Figure 2.1C). These assays, performed on a number of non-Aire-expressing tissues, similarly do not show enrichment of H3K27ac near ACNS1, consistent with the possibility that ACNS1 is an enhancer active in mTECs but few other tissues. The one ENCODE sample that does show H3K27ac enrichment near ACNS1 is the ES cell line ES-Bruce4. *Aire* is known to be expressed early in embryonic development and in transcribed in some, though not all, ES cell lines (77). Interestingly, the accompanying RNA-seq track for these particular ES cells does not show transcription of *Aire,* but does show transcription of *Dnmt3l* (data not shown). In Chapter 3, we consider the similarities in the expression and regulation of these two genes.

We also initially performed ChIP-seq against H3K4me2, both as a means of optimizing the ChIP-seq protocol on mTECs using an antibody that the Rao lab had worked with extensively and because H3K4me2 is also enriched at enhancers, though with less specificity for the tissues in which the enhancer is actually active compared to H3K27ac (146,147). Although

the results were not as striking as the H3K27ac data, there is also enrichment of H3K4me2 near ACNS1 (Figure S2.3). Based on the presence of these modifications near ACNS1 in mTECs but not in other tissues, we hypothesized that ACNS1 is an *Aire* enhancer. One caveat to this interpretation is that this mark was present in both GFP$^+$ and GFP$^-$ mTECs. One possible explanation for this discrepancy is that ACNS1 regulates a gene aside from *Aire.* Another possibility, however, is that this simply reflects retention of the H3K27ac mark in the aftermath of *Aire* expression in the post-Aire mTECs that constitute a sizable portion of the GFP$^-$ subset.

ACNS1 is an NF-κB-responsive element

We examined the sequence of ACNS1 to look for putative transcription factor binding sites. We observed two well-conserved candidate NF-κB binding sites (Figure 2.2A). This was particularly striking in light of the already established role for non-canonical NF-κB in mTEC development and maintenance (107,109,117,118). In order to determine whether NF-κB can actually bind these candidate κB sites, we performed electrophoretic mobility shift assays (EMSAs). We observed a large shift in biotinylated ACNS1 probe after incubation with nuclear lysates from p52-FLAG-transfected 293T cells (Figure 2.2B). We used p52 because this protein can bind DNA as homodimers and was previously implicated in mTEC biology (118). This shift could be almost completely abrogated by a 200-fold excess of untagged ACNS1 probe but not by an excess of probe in which the κB sites were mutated. In addition, the inclusion of anti-FLAG antibody in the binding reaction led to a supershift. Together, these results indicate specific binding by p52 to the κB sites of ACNS1.

We then assayed the ability of ACNS1 to act as an enhancer in an NF-κB-dependent manner. We inserted ACNS1 into a thymidine kinase minimal promoter (TK)-beta-galactosidase reporter plasmid. Co-transfection of ACNS1-TK-beta-gal, p52-FLAG, and RelB-FLAG resulted

in high amounts of beta-galactosidase expression, whereas there was little transcription when ACNS1 was not present or when p52 and RelB were not transfected (Figure 2.2C). Like p52, RelB is important for *Aire*-expressing mTECs (117). We also mutated each NF-κB binding site alone or in combination and found that each site has an equal and approximately additive effect on reporter expression (Figure 2.2D).

A small region upstream of Aire can recapitulate Aire cell-type specificity

One traditional method of testing a DNA sequence for enhancer function is to clone the sequence into a reporter vector, use it to make a transgenic mouse, and screen the mouse for reporter expression (128). We did not attempt to assay the small ACNS1 region in isolation in this manner. However, we did create two small transgenes that had the *Aire* promoter and additional upstream sequence, including ACNS1. In our initial attempt, the included sequence ran from within the 5' UTR of *Aire* to approximately 100 bp distal to ACNS1 (Figure S2.4A). Three *Eaig* (extended *Aire* promoter IGRP-GFP) founder mice were born and analyzed. Immunofluorescent staining of frozen thymic and splenic sections showed no GFP expression in any of these founders (Figure S2.4B and C). Since we only assayed this limited number of founder lines, we cannot say with confidence that this transgene is completely unable to recapitulate *Aire* expression; we can say, however, that it is not consistently able to do so.

We followed up on this by using a slightly larger transgene that extended 1.2 kb beyond ACNS1. We called this transgene *A4dig* (*Aire* 4.3-kb-driven IGRP-GFP) (Figure 2.3A). Eight founder lines were generated. Flow cytometric analysis of mTECs showed no GFP expression in four of these lines; in the other four, however, approximately 2% of mTECs expressed GFP (Figure 2.3B and Figure S2.5A). Remarkably, expression of GFP was limited to the MHC-II^{hi}, Aire^{+} subset of mTECs (Figure 2.3B and C). None of these mice expressed GFP in eTACs

Figure 2.2: ACNS1 is an NF-κB-responsive element.

(A) Comparison of consensus ACNS1 sequence to κB sequence motif. (B) Electrophoretic mobility shift assay using nuclear lysates from 293T cells transfected with p52-FLAG, incubated with biotinylated murine ACNS1 probe. Untagged WT and mutant probes (altered first, second, or both κB sites, respectively) at 200-fold excess to biotinylated probe, and anti-FLAG antibody included in lanes as specified. (C) Relative luminescence after chemiluminescent detection of beta-galactosidase in cellular lysates of 293T cells, 48 hours after transfection with TK-beta-gal or ACNS1-TK-beta-gal plasmid ± RelB-FLAG and p52-FLAG. (D) Relative luminescence after chemiluminescent detection of beta-galactosidase in cellular lysates of 293T cells, 48 hours after transfection with RelB-FLAG and p52-FLAG and WT or mutant ACNS1-TK-beta-gal. In (C) and (D), normalized for transfection efficiency using pRL-CMV. All data are representative of at least three independent experiments. Data in (C) and (D) analyzed by Student's *t*-test. ns denotes not significant, * for $p < 0.05$, ** for $p < 0.01$, and **** for $p < 0.0001$.

Figure 2.2
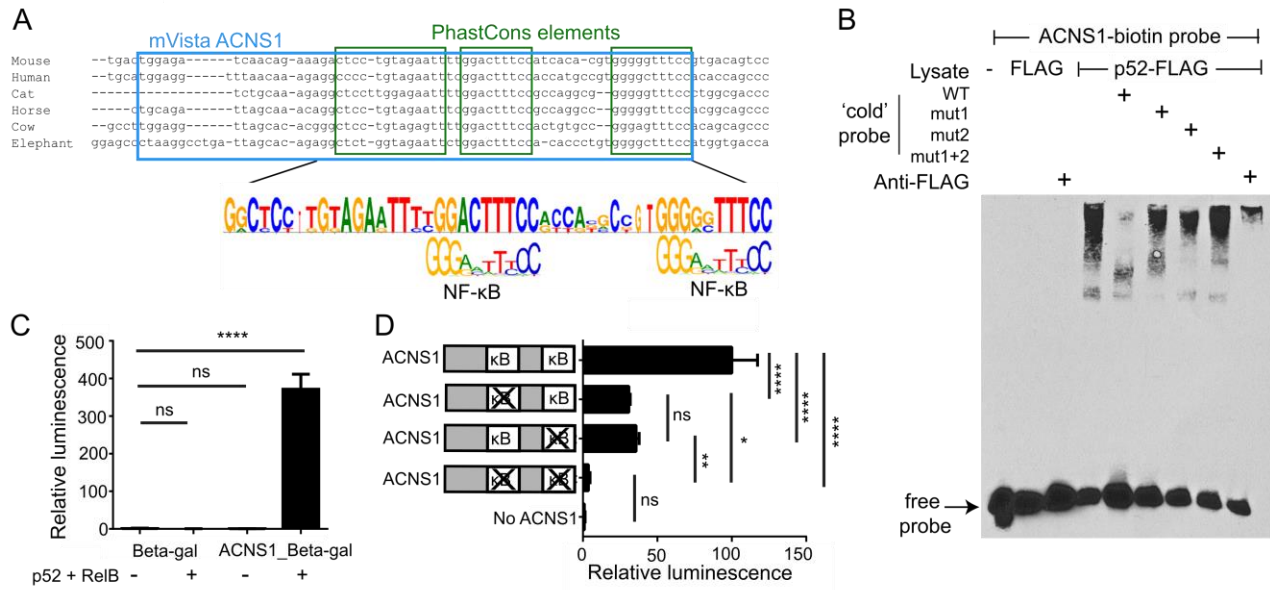
(Figure 2.3D and Figure S2.5B). Together, these results argue that this small transgene contains all critical elements to regulate the thymic cell-type specificity of *Aire* expression, albeit in only a minority of cells. As is typical with non-targeted transgenesis, the founders showed a wide range of transgene copy number (from one to 22). There was a trend toward correlation between the frequency of GFP-expressing mTECs and the transgene copy number, but the relationship was not statistically significant (Figure S2.5C). The median fluorescence intensity of GFP in GFP$^+$ cells of the *A4dig* mice was lower than that of *Adig* mice but not dramatically so; it is not as though the *A4dig* GFP$^+$ cells are all just barely past the threshold to be considered GFP$^+$ (Figure S2.5A and D). As such, the infrequency of GFP$^+$ cells does not seem to simply be a case of an intrinsically low-expressing transgene causing many cells to have a small increase in GFP signal such that a few are now just within the GFP gate. In fact, the specific, variegated expression seen with *A4dig* has been observed in other transgenes and in endogenous genes upon deletion of certain regulatory regions (148). We hypothesize that the inconsistency of transgene expression in the *A4dig* mouse compared to the *Adig* mouse is due to *A4dig* lacking an important regulatory element located elsewhere in the *Aire*-recapitulating BAC, which normally helps stabilize expression. Despite the variegated expression, however, we were nevertheless struck by how this relatively small transgene was able to replicate a highly specific pattern of expression, restricted even to the point of not being present in Aire$^-$ mTECs.

Two smaller transgenic *Aire* reporters have been published: one in which *Cre* was flanked by 23 kb from upstream of *Aire* plus 11 kb from downstream of *Aire* (149). This *Aire-Cre* mouse did not show evidence of activation early in embryogenesis, in contrast to the *Aire-Cre* created using the RP23-461E7 BAC (77). Another, smaller, transgene used just 6.3 kb of upstream sequence (almost the entire sequence between *Aire* and *Dnmt3l*) to drive SV40 large T

antigen, with the goal of creating an immortalized *Aire*-expressing TEC cell line (150). These transgenic mice developed extremely hyperplastic thymi, leading to death in young adulthood when the thymi became so large that they interfered with breathing. The authors of this study do not report that the mice had enlarged lymph nodes. This suggests that the 6.3-kb region is sufficient to drive expression in mTECs but not be enough to drive expression in eTACs. However, there are a number of alternatives to this interpretation. It may be that the large T antigen drives more proliferation in mTECs than eTACs at a given level of expression. Alternatively, in line with *Aire* expression being lower in eTACs, the construct may be expressed at a lower level in eTACs than mTECs such that, especially given that eTACs are rarer than mTECs, the effect of T antigen in the periphery takes longer to become grossly apparent.

Together, the various patterns of expression of these different transgenes suggest that *Aire* is regulated by multiple distal CREs. The absence of embryonic Cre activity in the 23-11 construct suggests that a regulatory element needed for embryonic but not adult expression is somewhere in the more distal parts of the BAC: in the region extending from -23 kb to -97 kb or the region extending from +11 kb (after stop codon) to +70 kb. The lack of eTAC expression with the *A4dig* construct, along with the results of the 6.3-kb construct, similarly suggests that a CRE is located outside of this small region that is essential for expression in eTACs. This eTAC CRE could be the same or different from the putative embryonic CRE. Finally, the variegated expression in mTECs of *A4dig* mice suggests there may be another element outside of the 4.3-kb region that affects thymic *Aire* expression.

In conclusion, a small transgene including 4.3 kb of *Aire* upstream sequence is able to create a highly specific pattern of expression, restricted even to the point of not being present in non-Aire-expressing mTECs.

Figure 2.3. A small transgene is able to reproduce the cell-type specificity of *Aire* expression.

(A) Schematic of the *A4dig* transgene, in which a 4.3-kb region extending upstream from the translation start site of *Aire* precedes an *IGRP-GFP* cassette. (B) Flow cytometry plots of mTECs, gated as DAPI⁻, CD45⁻, EpCAM⁺, MHC-II⁺, Ly51⁻, showing frequency (%) of GFP⁺ cells in *Adig,* representative *A4dig*, and WT mice. (C) Immunofluorescent staining for GFP (green) and Aire (red) in frozen thymic sections from *Adig* and *A4dig* mice. Scale bars are 50 μm. (D) Flow cytometry plots of eTACs, gated as DAPI⁻, CD45^{int}, MHC-II^{hi}, EpCAM⁺, CD86⁻, showing frequency (%) of GFP⁺ cells in *Adig,* representative *A4dig*, and WT mice. Data in (B) and (D) are representative of eight distinct *A4dig* founder lines analyzed across three experiments. Data in (C) are representative of two independent experiments totaling four mice of each genotype.

Figure 2.3
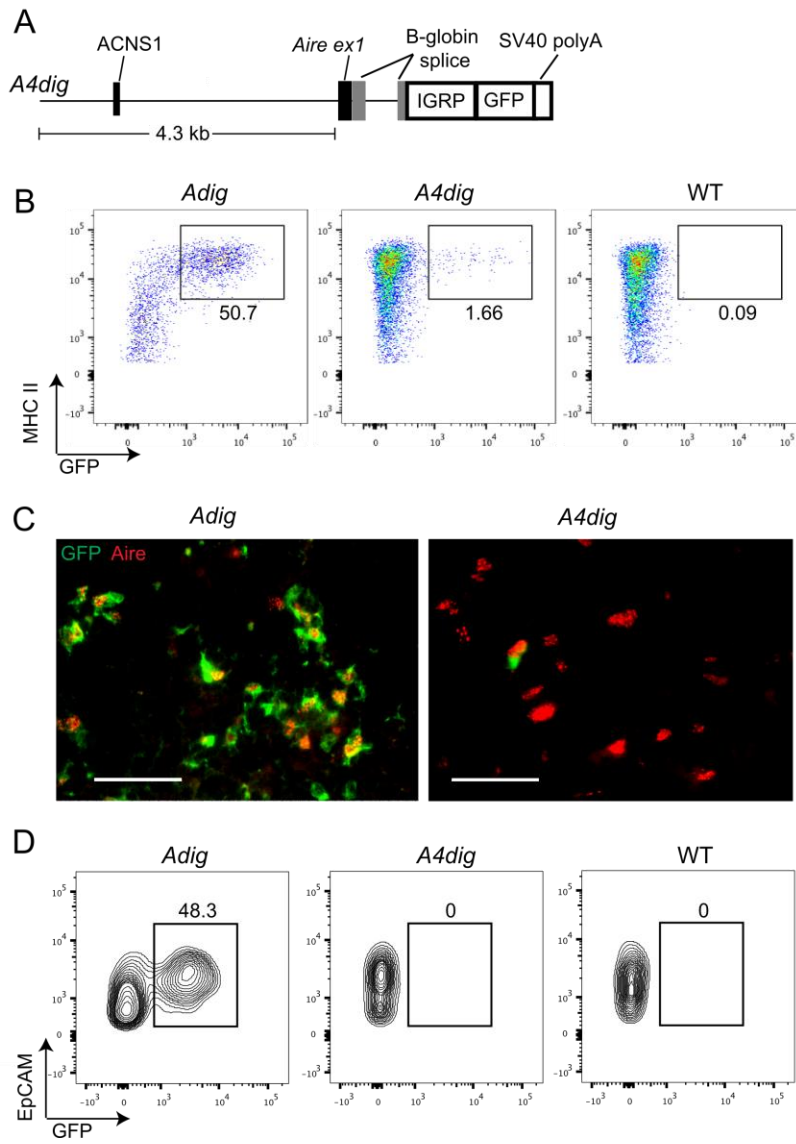
ACNS1 is necessary for Aire expression

After exploring the sufficiency of the 4.3-kb region, which includes ACNS1, we investigated the converse by determining whether ACNS1 is necessary for *Aire* expression. We began by deleting a 45-bp section of ACNS1, containing all three PhastCons elements and both κB sites, from the *Aire* reporter BAC and generating a new transgenic mouse line, *AdigΔACNS1* (Figure S2.6A). We observed no expression of GFP in either the mTECs or eTACs of *AdigΔACNS1* mice (Figure S2.6B and C). We also crossed them to NOD *8.3* TCR transgenic mice. This TCR is specific for a peptide in IGRP (the "i" in *Adig,* encoded by *G6pc2*). Whereas there was nearly complete negative selection of *8.3*-expressing thymocytes in (B6xNOD)F1 *Adig$^+$ 8.3$^+$* mice, there was no decrease in *8.3* cells in (FVBxNOD)F1 *AdigΔACNS1$^+$ 8.3$^+$* mice (Fig. S2.6D), indicating that the transgene is not being expressed. The critical caveat in interpreting these results is that only a single founder mouse was generated, so it is possible that the lack of reporter expression is due not to an intrinsic defect in the transgene but is instead because the transgene is located in a transcriptionally repressive location.

Nevertheless, we were encouraged enough by these results to pursue germline deletion of endogenous ACNS1. To do so, we took advantage of the recently developed genome editing technology CRISPR-Cas9 (151,152). This technology, derived from a means of viral defense present in many prokaryotes, allows targeted induction of a double-strand break using the Cas9 nuclease and a guide RNA complementary to the site to be targeted. Error-prone non-homologous end joining of these double-strand breaks frequently results in indel formation. In addition, it is possible to introduce specific point mutations or new sequences, such as a *loxP* site or reporter construct, by including an oligonucleotide to introduce this change through

homology-directed repair (153), a process that can be facilitated by use of a modified Cas9 that acts as a nickase (154).

Introduction of an indel using a single guide RNA is efficient for knocking out protein-coding genes. Since enhancers are relatively more resistant to small alterations, however, we decided to delete all of ACNS1, as well as some flanking sequence, using two guide RNAs simultaneously, one targeting upstream and one targeting downstream. We designed eight candidate guide sequences using the online tool created by the Zhang lab (Figure S2.7A). This tool (www.crispr.mit.edu) allowed us to select sequences predicted to have a lower likelihood of generating off-target mutations because there are fewer sequences elsewhere in the genome with only minor differences in insensitive positions of the guide RNA. We cloned the guide RNA sequences into the pX330 CRISPR-Cas9 plasmid (155,156). The UCSF ES Cell Targeting Core tested each of these guide RNAs for cutting efficiency in ES cells (Figure S2.7B). The two most efficient guide RNAs were injected with Cas9 mRNA into the cytoplasm of fertilized C57BL/6 oocytes. The resulting pups were screened by PCR and Sanger sequenced (Figure S2.7C). We identified seven mice with the desired deletion, in each case in heterozygosity, which was a success rate of approximately one in seven (Figure S2.7D). We performed our subsequent analyses on descendents of an arbitrarily chosen founder, with the exception of some analyses of compound heterozygotes from crosses of this line and another founder line, as discussed below.

We observed that whereas approximately 45% of mTECs are Aire$^+$ in a WT mouse, ACNS1$^{-/-}$ mice lacked Aire$^+$ mTECs (Figure 2.4B). Immunofluorescent staining of frozen thymic sections confirmed these findings, showing none of the characteristic speckled nuclear staining of Aire in ACNS1$^{-/-}$ mice (Figure 2.4C). Quantitative RT-PCR analysis of RNA from FACS-isolated mTECs showed that *Aire* transcript was also greatly reduced (Figure 2.4D). We noticed

an average of 39% (Figure 2.4A); when we also included WT and ACNS1$^{+/-}$ mice that were part of the experiments investigating two forms of compound heterozygous mouse, described below, this decrease in Aire$^+$ cells was statistically significant (Figure S2.8A). We also observed reduced Aire signal in those mTECs that were Aire$^+$ in ACNS1$^{+/-}$ mice (Figure S2.8B) and a trend toward decreased *Aire* transcript, as determined by quantitative RT-PCR (Figure S2.8C).

We also observed that there are fewer MHC-II^{lo} mTECs in ACNS1$^{-/-}$ mice (Figure S2.8D). We also see fewer of these cells in *Aire*$^{-/-}$ mice (data not shown). We have previously observed that a substantial component of the MHC-II^{lo} mTEC subset consists of post-Aire cells (73), and preliminary data suggest that *Aire*$^{-/-}$ mice have fewer post-Aire cells (data not shown). One possible explanation for the decreased number of MHC-II^{lo} mTECs in ACNS1$^{-/-}$ mice, then, is that there are fewer post-Aire cells because of a partial developmental block in the mTECs of these mice due to their failure to express *Aire*.

We also found by immunofluorescent staining that there is a large reduction in Aire$^+$ cells in the lymph nodes of ACNS1$^{-/-}$ mice (Figure 2.4E and F). We are following up on some equivocal quantitative RT-PCR comparing *Aire* expression in sorted eTACs from WT and ACNS1$^{-/-}$ mice. We are also seeking to optimize flow cytometric detection of Aire in eTACs, in order to confirm this finding by a more rigorously quantitative method.

In the course of generating ACNS1$^{-/-}$ mice, we identified a founder with a smaller, partially overlapping deletion that spares ACNS1 itself (Figure 2.4G). In homozygosity, this ACNS1adjacent deletion allele results in no change in the frequency of Aire$^+$ mTECs (Figure 2.4H). This suggests that the essential *Aire*-regulating element is even smaller than 269-bp region absent in the ACNS1$^{-/-}$ mouse, which is consistent with the lack of expression from the *Aire* reporter BAC from which we deleted just 45 bp of ACNS1.

Analyses of CRISPR-Cas9-mediated genome editing have demonstrated the risk of undesired off-target mutations (157), raising the risk that the observed phenotype is not due to the deletion of ACNS1 but a cryptic mutation elsewhere. However, we have found several pieces of evidence that together strongly support the interpretation that this loss of *Aire* expression is specifically due to deletion of ACNS1. First, analyses were performed on mice that were at least two generations removed from the founder. Given that mice used in breeding in each generation were selected for ACNS1 genotype, the likelihood of the analyzed ACNS1$^{-/-}$ mice also being homozygous for a different causative mutation is quite low unless this mutation is strongly linked to ACNS1. We also found that ACNS1$^{-/-}$ compound heterozygote mice, containing one allele derived from each of two different founders, also lacked Aire^{+} mTECs (Figure S2.8E). In addition, we crossed ACNS1$^{+/-}$ mice to *Aire*$^{+/-}$ mice. The resulting compound heterozygote mice, with missing ACNS1 on one chromosome and exon 2 of *Aire* on the other, also showed a loss of Aire expression (Figure 2.4I). Together, these results indicate that the phenotype seen in ACNS1$^{-/-}$ mice is due to the deletion of ACNS1, not an unobserved off-target mutation. The *Aire,* ACNS1 compound heterozygote results further suggest that ACNS1 has a direct role in controlling *Aire* transcription, rather than acting indirectly by regulating a different gene that in turn is essential for *Aire*.

The finding that ACNS1 regulates *Aire* gives new significance to the finding that ACNS1 is an NF-κB-responsive element (Figure 2.2). Previous research connecting non-canonical NF-κB to the development of *Aire*-expressing mTECs was unable to distinguish the degree to which this reflected a direct role for NF-κB in promoting *Aire* expression versus an indirect role in which NF-κB is required for mTEC maturation to the state at which other factors become expressed that directly drive *Aire*. The observation that ACNS1 responds to NF-κB provides

Figure 2.4. ACNS1 is required for *Aire* expression.

(A) Schematic of the CRISPR-Cas9-mediated germline deletion of ACNS1. Arrows show the sites targeted by the two simultaneously injected guide RNAs. (B) Flow cytometry plots of mTECs, gated as DAPI$^-$, CD45$^-$, EpCAM$^+$, MHC-II$^+$, Ly51$^-$, showing frequency (%) of Aire$^+$ cells. Bar graph (right) shows mean ± SD frequencies of Aire$^+$ cells among mTECs from WT, ACNS1$^{+/-}$, and ACNS1$^{-/-}$ mice. (C) Immunofluorescent staining for cytokeratin-5 (red) and Aire (green) in frozen thymic sections from WT and ACNS1$^{-/-}$ mice. Scale bars are 50 μm. (D) Quantitative PCR analysis of *Aire* and various TSA genes, comparing RNA from WT and ACNS1$^{-/-}$ mTECs, showing mean ± SD (technical replicates), normalized to *Actb*. (E) Immunofluorescent staining for Aire (green) with DAPI counterstain in frozen lymph node sections from WT and ACNS1$^{-/-}$ mice. Scale bars are 25 μm. Arrowhead indicates nuclear Aire staining. (F) Graph summarizes the results in (E), showing the number of Aire$^+$ cells observed in 80 lymph node sections each from WT and ACNS1$^{-/-}$ mice. (G) Schematic showing location of ACNS1adjacent deletion relative to ACNS1 deletion. (H) Flow cytometry plots of mTECs, gated as in (B), showing frequency (%) of Aire$^+$ cells. Bar graph (right) shows mean ± SD frequencies of Aire$^+$ cells among mTECs from WT and ACNS1adjacent$^{-/-}$ mice. (I) Flow cytometry plots of mTECs, gated as in (B), showing frequency (%) of Aire$^+$ cells. Bar graph (right) shows mean ± SD frequencies of Aire$^+$ cells among mTECs from WT and *Aire*$^{+/-}$ ACNS1$^{+/-}$ (compound heterozygote) mice. Data in (B), (G), and (I) summarize three independent experiments containing one or more mice within each group and totaling at least five mice per group. Data in (C), (E), and (F) are representative of two independent experiments of at least three mice per group. Data in (B), (H), and (I) analyzed by Student's *t*-test; data in (F) analyzed by Garwood method of Poisson distribution confidence interval calculation. ns denotes not significant and **** for p < 0.0001.

Figure 2.4
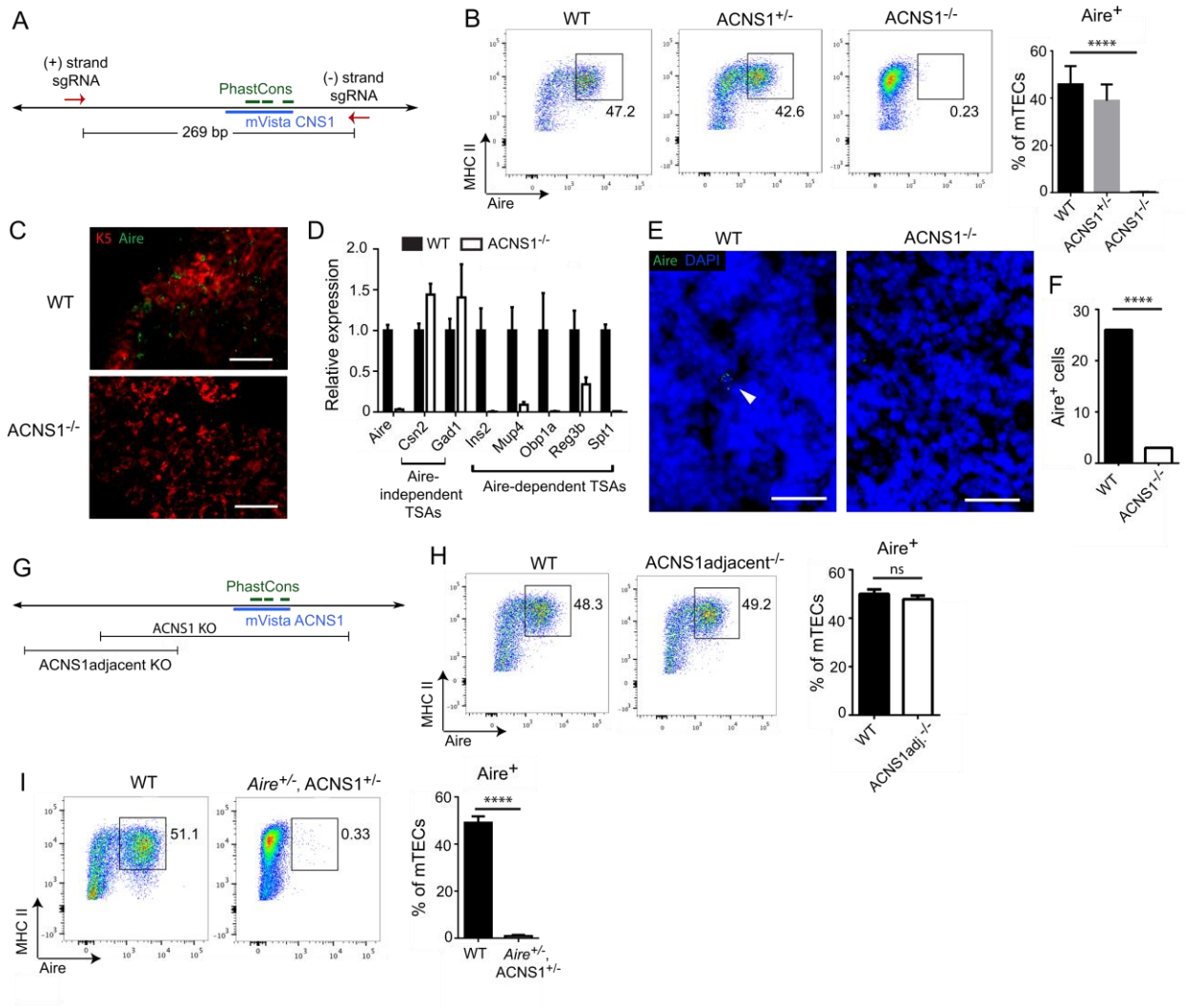
evidence for a direct role for NF-κB in turning on *Aire*. Thus far, however, it has proven to be challenging to directly observe NF-κB binding at ACNS1 in *Aire*-expressing cells. There is no widely available TEC cell line that expresses endogenous *Aire*. In addition, due to technical challenges, we have not successfully performed anti-NF-κB ChIP on mTECs. As such, the mechanism by which CNS1 regulates *Aire* remains an important open question.

Loss of CNS1 leads to spontaneous autoimmunity

Aire acts to prevent autoimmunity by inducing TSA expression, which in turn permits successful deletion and control of self-reactive thymocytes. We followed up on the finding that ACNS1$^{-/-}$ is essential for *Aire* expression by analyzing the expression of several TSA genes. We found substantially reduced expression of Aire-dependent TSAs but not Aire-independent TSAs in mTECs from ACNS1$^{-/-}$ mice (Figure 2.4D). In *Aire*$^{-/-}$ mice, it has been shown that loss of particular TSAs can be sufficient for specific organ-specific autoimmunity. For example, the loss of IRBP (*Rbp3*) expression leads to retinal disease (49). Given the substantial reduction in Aire-dependent TSA expression in ACNS1$^{-/-}$ mice, we hypothesized that these mice would develop autoimmune disease.

We examined ACNS1$^{-/-}$ mice for manifestations of autoimmunity often observed in Aire-deficient mice on the C57BL/6 background: retinopathy, salivary gland disease (sialadenitis), and lacrimal gland disease (dacryoadenitis) (44,46). We began by looking for evidence of the escape of self-reactive T cells by immunizing mice with the P2 peptide of IRBP. T cells specific for this peptide undergo Aire-dependent negative selection (158). Ten days after immunization, we quantitated P2-I-A^{b} tetramer-specific T cells in draining lymph nodes and spleen by flow cytometry. We observed a significant increase in the number of these cells in ACNS1$^{-/-}$ mice

compared to WT mice, to a level comparable to that seen in *Aire*$^{-/-}$ mice (Figure 2.5A). These results are consistent with a defect in negative selection in ACNS1$^{-/-}$ mice.

Consistent with this defect in negative selection, ACNS1$^{-/-}$ mice developed spontaneous autoimmune disease. Approximately 20 percent of assayed mice had anti-IRBP autoantibodies at 10 weeks of age (Fig. 2.5B). Fundoscopic examination of 10- to 15-week-old ACNS1$^{-/-}$ mice showed a similar incidence of inflammation-associated retinal disease (Figure 2.5C). Finally, histologic examination of approximately 15-week-old ACNS1$^{-/-}$ mice frequently revealed significant immune infiltrates in their salivary and lacrimal glands and retinal destruction (Figure 2.5D). In summary, we find that the deletion of ACNS1 results in a pattern of spontaneous autoimmunity akin to that previously observed in the traditionally targeted Aire-deficient mouse.

Conclusion:

In summary, we have identified a *cis*-regulatory element that governs *Aire* expression, and, consequently, is essential for immune self-tolerance. This element, ACNS1, is conserved across placental mammals. Consistent with its role in regulating *Aire*, it is flanked by chromatin enriched for the enhancer-associated histone modification H3K27ac. Deletion of endogenous ACNS1 led to a loss of *Aire* expression in both the thymus and the periphery. ACNS1$^{-/-}$ mTECs showed decreased expression of *Aire*-dependent TSAs. Mice lacking ACNS1 showed escape of self-reactive T cells from the thymus and spontaneously developed several manifestations of autoimmune disease. In addition, ACNS1 is NF-κB-responsive, providing the first evidence for a direct role for non-canonical NF-κB in *Aire*-expressing mTECs.

NF-κB is active in a variety of cell types, nearly all of which do not express *Aire*. Consequently, additional factors must also be necessary for the induction of *Aire* expression. The identity of these additional factors and whether they bind at ACNS1 or elsewhere within the

Figure 2.5. Mice lacking ACNS1 develop spontaneous autoimmunity.

(A) WT, ACNS1[-/-], and *Aire*[-/-] mice, the last serving as a positive control, were immunized with the IRBP P2 peptide. Draining lymph nodes and spleen were harvested 9-10 d later for P2-I-A^b tetramer-based quantitation of P2-specific T cells. Each dot on the graph represents an individual mouse, the short horizontal lines show the means, and the dotted line represents the limit of detection. (B) Radioligand binding assay of anti-IRBP antibodies in sera collected from 10-wk-old WT and ACNS1[-/-] mice. Autoantibody index of 1 is defined by a sample with 367 µg/ml of polyclonal rabbit anti-IRBP antibody. (C) Representative fundoscopic images of retinas of 10- to 15-wk-old WT and ACNS1[-/-] mice. Graph summarizes the incidence of retinopathy in each genotype. (D) Representative hematoxylin- and eosin-stained retinal, salivary, and lacrimal sections from 14- to 16-wk-old WT and ACNS1[-/-] mice; arrows indicate mononuclear infiltrates. Scale bars are 100 µm. Graphs (right) show retinal degeneration scores and infiltrate scores of salivary and lacrimal glands; each dot is a single mouse and short horizontal lines show the means. (E) Correlation of four manifestations of autoimmunity in a cohort of WT and ACNS1[-/-] mice. Data in (A) are representative of two independent experiments. Data in (B) are pooled from two experiments; WT, n=7; ACNS1[-/-], n=8. Data in (C) are pooled from several experiments; WT, n=22, ACNS1[-/-], n=8. Data in (D) and (E) are from one cohort of 6 WT and 6 ACNS1[-/-] mice. Data in (A) and (D) analyzed by Mann-Whitney test; data in (C) analyzed by chi-square test. ns denotes not significant, * for $p < 0.05$ and ** for $p < 0.01$.

Figure 2.5
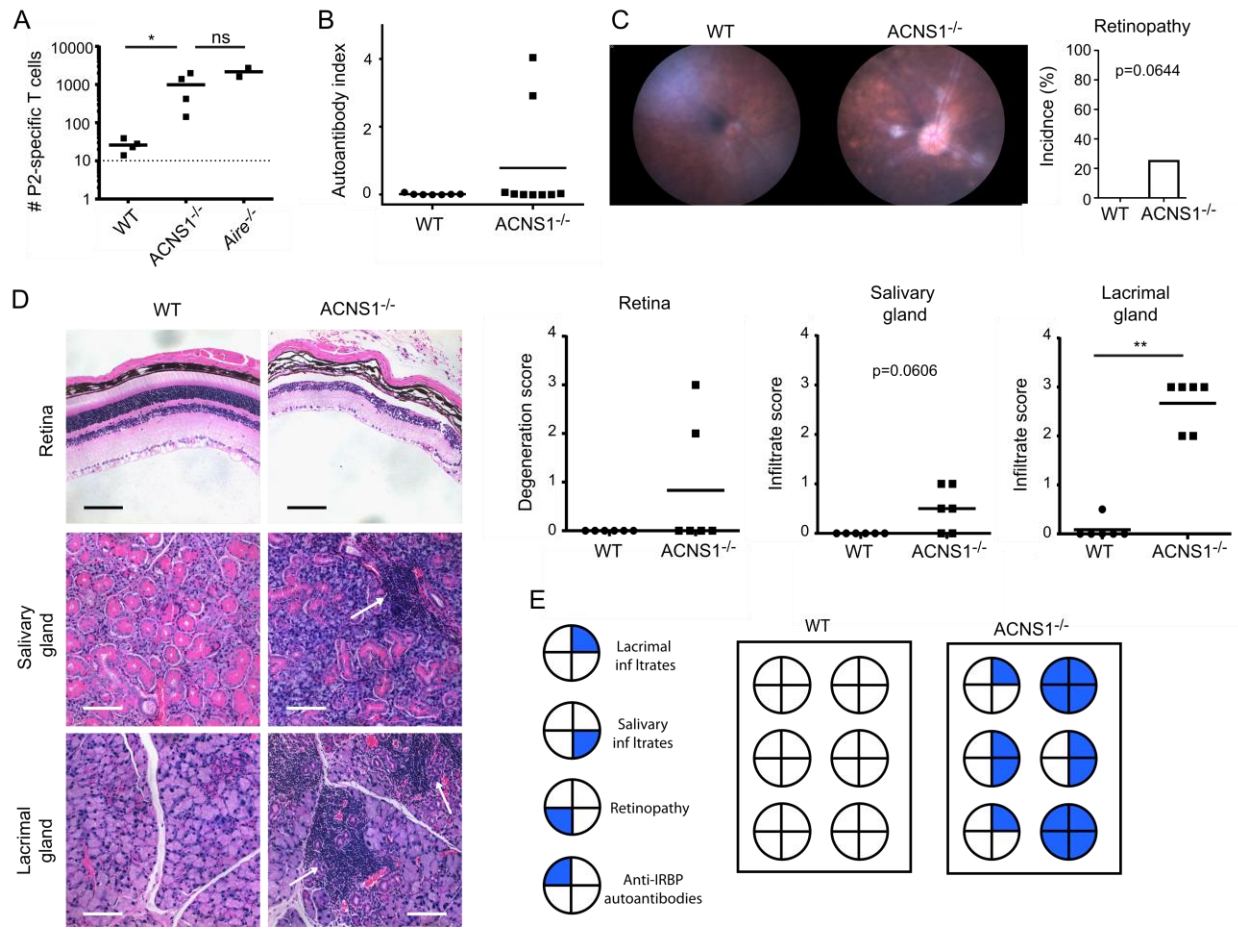
4.3-kb *A4dig* region remain important and unresolved questions. Moreover, additional research is needed to determine whether there are other distal *Aire cis*-regulatory elements. The variegated GFP expression in mTECs of *A4dig* mice suggests that additional elements do exist. In addition, the lack of GFP expression in eTACs of *A4dig* mice may reflect differences in the mechanism of *cis*-regulation of *Aire* in mTECs and eTACs, a topic that merits further investigation.

Although on a population level *Aire* induces the expression of thousands of genes, in an individual cell its effect is limited and stochastic, such that any given TSA is only expressed in a minority of mTECs (101). It follows that the regulatory region of the *A4dig* transgene might be able to support sufficient expression of a given antigen to permit effective negative selection and thereby serve as the basis for targeted induction of antigen-specific tolerance in vivo. Finally, these results imply that some individuals may develop autoimmune disease due to deletion of ACNS1. Indeed, certain patients with clinical features of APS1 have been identified who lack mutations in the coding sequence of *AIRE*, and further investigation into the role of ACNS1 in human health is warranted.

Materials and Methods:

Mice. C57BL/6.*Adig* mice were generated as described (74). NOD.*8.3* mice (159) were a gift from Jeff Bluestone. C57BL/6.*Aire*$^{-/-}$ mice were generated as described (44).

FVB.*Eaig* mice were generated by modification of the JDLNL targeting construct used to generate *Adig* mice (74). First, PvuI and MluI sites were inserted immediately upstream of the 5' *Aire* homology region of JDLNL using standard cloning methods. Then, additional homology sequence corresponding to mm9:chr10:77508148-77509503 was cloned using PCR and inserted into the MluI site. The resulting *Eaig* plasmid was isolated using Plasmid Maxi Kit (Qiagen), digested with PvuI to separate the transgene (Fig. 2.4A) from the plasmid backbone, purified by gel electrophoresis, Qiaquick Gel

Extraction Kit (Qiagen), followed by Qiaquick PCR Purification Kit (Qiagen), and injected into fertilized FVB/N oocytes by Helen Lu of the UCSF LARC Transgenic Mouse Core.

FVB.*A4dig* mice were generated by modification of the *Eaig* construct. A short homology region corresponding to mm9:chr10:77510085-77510611 (- strand) was cloned using PCR and inserted at the MluI site using InFusion (Clontech), adding a SalI site at the 3' end. The resulting plasmid was linearized with SalI and underwent recombineering-based gap repair in which the plasmid was electroporated into heat-induced SW105 bacteria (Biological Resources Branch, NCI at Frederick) already transformed with the 461E7 BAC, followed by kanamycin selection, and plasmid isolation and sequencing. The resulting *A4dig* plasmid was isolated using Plasmid Maxi Kit (Qiagen), digested with PvuI to separate the transgene (Fig. 2.5A) from the plasmid backbone, purified by gel electrophoresis, Quiquick Gel Extraction Kit (Qiagen), followed by Qiaquick PCR Purification Kit (Qiagen), and injected into fertilized FVB/N oocytes by Helen Lu of the UCSF Laboratory Animal Resource Center (LARC).

FVB.*Adig∆ACNS1* mice were also made by BAC recombineering. A 45-bp region (corresponding to mm9:chr10:77509287-77509323) encompassing most of ACNS1 was deleted from the *Eaig* plasmid by a PCR-mediated deletion method (160) and linearized plasmid was transformed into heat-induced EL350 bacteria (a gift from Jeroen Roose) already transformed with the 461E7 BAC, selected with kanamycin, and arabinose treated to induce Cre, causing *NeoR* excision. The BAC was isolated using Nucleobond BAC 100 kit (Macherey-Nagel) and injected into fertilized FVB/N oocytes by the UCSF LARC Transgenic Mouse Core.

C57BL/6.*ACNS1*$^{-/-}$ mice were created by CRISPR-Cas9-mediated deletion in fertilized oocytes. More specifically, the online tool designed by lab of Feng Zhang at MIT (www.crispr.mit.edu) was used to identify candidate guide RNA sequences with a low predicted incidence of off-target mutations. These candidates were cloned by the method described in (156) into pX330-U6-Chimeric_BB-CBh-hSpCas9, which was a gift from Feng Zhang (Addgene plasmid #42230). These candidate guide RNAs were then tested by Chong Park and colleagues at the UCSF ES Cell Targeting Core by transfecting into ES cells and cutting activity assessed using Cel1-based surveyor assay. Guide RNAs and Cas9 mRNA were in

vitro transcribed (Promega), isolated, and injected into cytoplasm of fertilized B6 oocytes at either 50 ng/µl Cas9 and 20 ng/µl each sgRNA or 25 ng/µl Cas9 and 10 ng/µl sgRNA—both resulted in pups with the desired deletion.

All animals were housed and bred in specific pathogen-free conditions at UCSF. All experiments were approved by the Institutional Animal Care and Use Committee of UCSF.

ChIP-seq. ChIP was performed as described in (145). Briefly, sorted cells were fixed with 1% formaldehyde, washed, snap frozen, and stored at -80 °C. Frozen samples were lysed and chromatin was sheared by sonication, using a Bioruptor (Diagenode), to yield 100-500 bp fragments. Chromatin was diluted, incubated overnight with protein A-coated magnetic beads (Invitrogen) pre-coated with anti-H3K27ac (Abcam**,** Ab4729, lot GR167929-1) or H3K4me2 (Abcam, clone Y47). Beads underwent a series of washes. Chromatin was then eluted, treated with RNase and proteinase K, and purified by affinity column (Zymo Research). Whole genome amplification of DNA was performed using WGA-SEQX (Sigma). Sequencing library was prepared and the samples were sequenced using Illumina Hi-Seq. Sequences were filtered, aligned, and unique reads selected by the La Jolla Institute for Allergy and Immunology bioinformatics core, using the Cufflinks (161), followed by peak identification using MACS (162).

Quantitative PCR. RNA was isolated from sorted mTECs using RNeasy Plus Micro Kit (Qiagen) and reverse transcribed using oligo-dT primers and the SuperScript III Kit (Life Technologies). TaqMan gene expression assays (Applied Biosystems) were used for all targets. Quantitative PCR was performed using 7500 Fast Real-Time PCR System (Applied Biosystems). Transcript abundance was normalized to *Actb* and analyzed using the ΔΔCt method.

Electrophoretic mobility shift assay. 293T cells were cultured in high-glucose DMEM, 10% heat-inactivated FCS, and penicillin/streptomycin, at 37 °C, 10% CO_2. 293T cells were transfected with p52-FLAG using Trans-IT 293 (Mirus). After 24 hours, cells were collected, and nuclear extracts prepared

using NE-PER Kit (Thermo Scientific). Nuclear extract protein concentration was determined using BCA Protein Assay Kit (Thermo Scientific). p52-FLAG pcDNA3 was a gift from Stephen Smale (Addgene plasmid #20019). EMSAs were performed using LightShift Kit (Thermo Scientific). Binding reactions contained 5 µg of nuclear extract, 10 nM biotinylated probe, 2 µM unlabeled probe, and 1 µg anti-FLAG (M2 clone, Sigma). Binding reactions were electrophoresed on 6% DNA Retardation Gels (Invitrogen) and transferred to Biodyne B nylon membrane (Thermo Scientific). DNA was cross-linked using UV light, and detected using the Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific). Probes were ordered from IDT. Probes sequences were as follows:

WT-CNS1-F: 5'- (biotin)-AATTTTGGACTTTCCATCACACGTGGGGGTTTCCGTGAC,

WT-CNS1-R: 5'- GTCACGGAAACCCCCACGTGTGATGGAAAGTCCAAAATT,

mut1-CNS1-F: 5'- AATTT**CTC**ACTTTCCATCACACGTGGGGGTTTCCGTGAC,

mut1-CNS1-R: 5'- GTCACGGAAACCCCCACGTGTGATGGAAAGT**GAG**AAATT,

mut2-CNS1-F: 5'- AATTTTGGACTTTCCATCACACGT**CTC**GGTTTCCGTGAC,

mut2-CNS1-R: 5'- GTCACGGAAACC**GAG**ACGTGTGATGGAAAGTCCAAAATT,

mut1+2-CNS1-F: 5'- AATTT**CTC**ACTTTCCATCACACGT**CTC**GGTTTCCGTGAC, and

mut1+2-CNS1-R: 5'- GTCACGGAAACC**GAG**ACGTGTGATGGAAAGT**GAG**AAATT.

In vitro reporter assays. TK-beta-gal was a gift from Brian L. Black. ACNS1 sequence (mm9, chr10:77509283-77509327) or ACNS1 with mutated kB sites (same mutations as in EMSA probes) was cloned into Kpn-linearized plasmid using standard cloning methods. p52-FLAG pcDNA3 was a gift from Stephen Smale (Addgene plasmid #20019). RelB-FLAG pcDNA was a gift from Stephen Smale (Addgene plasmid 20017). pRL-CMV was purchased from Promega. 293T cells were cultured in high-glucose DMEM, 10% heat-inactivated FCS, and penicillin/streptomycin, at 37 °C, 10% CO_2. 293T cells in 24-well plates were transfected with 12.5 µg pRL-CMV and 250 µg of (ACNS1)-TK-beta-gal, p52, and RelB plasmids using Trans-IT 293 (Mirus). After 48 hours, cells were briefly washed and lysed using Renilla Luciferase Assay System (Promega). Samples were then aliquoted and analyzed for Renilla

57

luciferase using Renilla Luciferase Assay System and for beta-galactosidase activity using Luminescent Beta-galactosidase Detection Kit II (Clontech). Luminescence was measured using Victor^2 1420 Multilabel Counter (PerkinElmer). In each experiment, all conditions were performed in triplicate.

Flow cytometry. TECs and eTACs were isolated as previously described (74). Briefly, thymi or peripheral lymph nodes and spleens were isolated, minced, and digested with DNase I and Liberase TM (Roche) before centrifugation on gradient of Percoll PLUS (GE Healthcare). The enriched stromal cells were stained with the antibodies against the surface markers indicated in the figure legends (BioLegend). For intracellular staining, cells were fixed using Foxp3 Staining Buffer Kit (eBioscience) and stained with anti-Aire (clone 5H12, eBioscience). Data were collected using an LSRII flow cytometry (BD) and analyzed using FACSDiva (BD) and FlowJo (TreeStar). Cell sorting was performed using a FACS Aria III (BD).

IRBP P2 peptide immunization and tetramer analysis. Mice were immunized with 100 μg P2 peptide (IRBP, aa 271-290), emulsified in complete Freund's adjuvant, subcutaneously at the chest. Tetramer analysis was performed 9-10 d after immunization on pooled axillary and cervical lymph nodes and spleen using P2-I-A^b tetramer, as previously described (158). The tetramer was generated by the NIH Tetramer Core Facility. The limit of detection was defined as 3 SD above the mean number of P2 tetramer-staining CD8+ T cells observed.

Immunofluorescence. Thymi and lymph nodes were harvested and embedded into OCT media (Tissue-Tek) and frozen. 8-μm sections were fixed in 1:1 acetone:methanol, blocked with 10% normal goat serum, and stained for cytokeratin-5 (Abcam), GFP (Invitrogen), or Aire (eBioscience, clone 5H12), followed by goat secondary conjugated with A488 or A594 (Invitrogen). Immunofluorescent slides were imaged using Axio Imager M2 widefield fluorescence microscope (Zeiss) with CoolSnap HQ2 camera (Photometrics). Images were analyzed using ImageJ (NIH).

Histology. Eyes, lacrimal glands, and salivary glands were harvested and fixed overnight with 10% formalin, transferred to 30% ethanol for 30 minutes, and stored long-term in 70% ethanol. Tissues were embedded in paraffin, sectioned, and stained with hematoxylin and eosin by the UCSF Mouse Pathology Core. Tissues were scored by an observed blinded to sample identity using the following criteria: Retinas: 0, no lesion; 1: loss of <50% of photoreceptor layer; 2: loss of >50% of photoreceptor layer; 3: loss of >50% of photoreceptor layer and loss of a minority of the outer nuclear layer; 4: loss of >50% of photoreceptor layer and >50% of outer nuclear layer. Lacrimal and salivary glands: 0, no infiltrate; 0.5 trace infiltrate; 1: minor infiltrate; 2: moderate infiltrate; 3: severe infiltrate; 4: complete tissue destruction.

Radioligand binding assay. Murine IRBP cDNA was in vitro transcribed and translated with [^{35}S] labeling using TNT Quick-Coupled Transcription/Translation Kit (Promega) and assays performed as described (52). Each well was loaded with 20,000 cpm of labeled protein. Polyclonal rabbit anti-IRBP (Proteintech, 14352-1-AP) was used as the positive standard. Autoantibody index was calculated as [cpm sample – cpm negative control] / [cpm positive standard – cpm negative standard]. Samples were positive if 3 SD above the mean for the WT mice.

Statistical analysis. Statistical analysis was performed using Prism 6.0 (GraphPad Software). Tetramer analysis and histology scoring were analyzed by Mann-Whitney rank-sum testing, mTEC flow cytometry and in vitro reporter assays were analyzed by Student's *t* test, retinopathy incidence was analyzed by chi-square test, and eTAC quantitation was analyzed by Garwood method of Poisson distribution confidence interval calculation. ns denotes not significant, * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$, and **** for $p < 0.0001$.

Supplemental Data:

Table S2.1. Sequences aligned and compared in mVista conservation analysis.

Species	Sequence
Bos taurus (cow)	bosTau6, chr1:145622665-145829200
Canis lupus familiaris (dog)	canFam2, chr31:40378450-40553800
Equus caballus (horse)	equCab2, chr26:39415250-39592500
Homo sapiens (human)	hg19, chr21:45573763-45787110
Loxodonta africana (African elephant)	loxAfr3, scaffold_67:11282750-11459800
Mus musculus (house mouse)	mm9, chr10:77423776-77603325
Myotis lucifugus (little brown bat)	myoLuc2, GL430058:304000-484000

Table S2.2. RankVista-identified conserved noncoding sequences in RP23-461E7 BAC

p-value ≤0.01.

Start coordinate (mm9, chr10)	End coordinate (mm9, chr10)	Length (bp)	p-value	Annotation
77469790	77469883	94	0.004	ACNS4, *Pfkl* intron 1
77499541	77499689	149	0.006	ACNS2, *Aire* intron 8
77509267	77509327	61	3.4E-6	ACNS1, intergenic, *Aire* & *Dnmt3l*
77514377	77514442	66	0.00093	ACNS3, *Dnmt3l* intron 3
77552190	77552330	141	7.9E-5	ACNS5, intergenic, *Icosl* and *D10Jhu81e*
77555547	77555661	115	4.8E-6	ACNS6, intergenic, *Icosl* and *D10Jhu81e*
77567645	77567755	111	0.01	ACNS7, intergenic, *Icosl* and *D10Jhu81e*
77567982	77568071	90	0.0028	ACNS8, intergenic, *Icosl* and *D10Jhu81e*

Figure S2.1. Phylogeny of *AIRE*-flanking regions and protein sequence similarities suggest *Aire* and *Sp100* family resulted from gene duplication.

(A) Each color represents a particular syntenic region. Legend shows list of representative genes from each region, using the name of the human orthologue. The graph depicts relative phylogenetic relationships, but the length of each arm does not correspond to the degree of genomic difference or time since last common ancestor. (B) Comparison of protein domains in *Aire* and *Sp140* orthologues from various vertebrates.

Figure S2.1

A

■ (purple) KIF1A, ILKAP, FAM132A, ESPNL, SCLY	■ (yellow) EFHD1, GIGYF2, KCNJ13, NGEF
■ (orange) SP100/110/140/140L orthologue	■ (green) PFKL, C21orf2, TRPM2, LRRC3
■ (red) AIRE	■ (blue) PWP2, C21orf33, ICOSLG, DNMT3L

B

■ (salmon) HSR (aka Sp100)	■ (purple) PHD or PHD-like
■ (cyan) SAND	■ (tan) Bromodomain

63
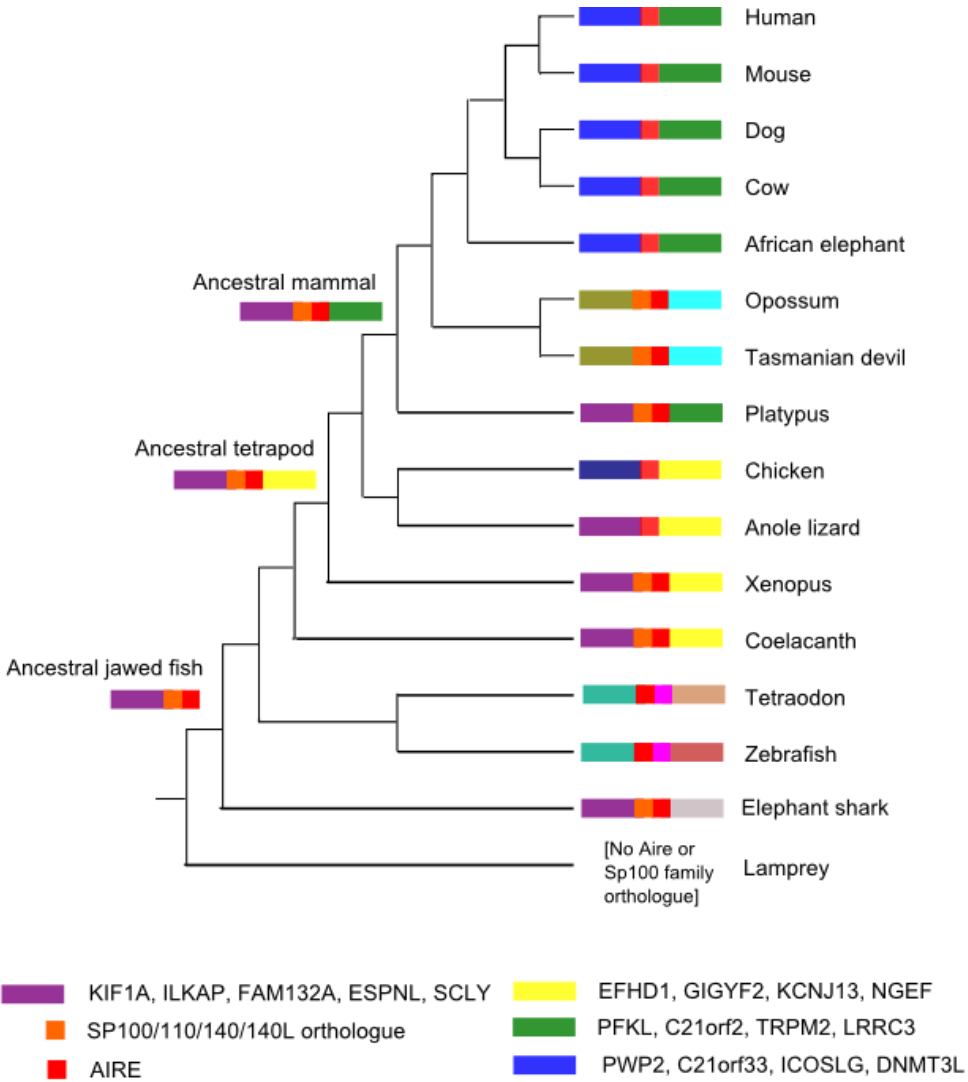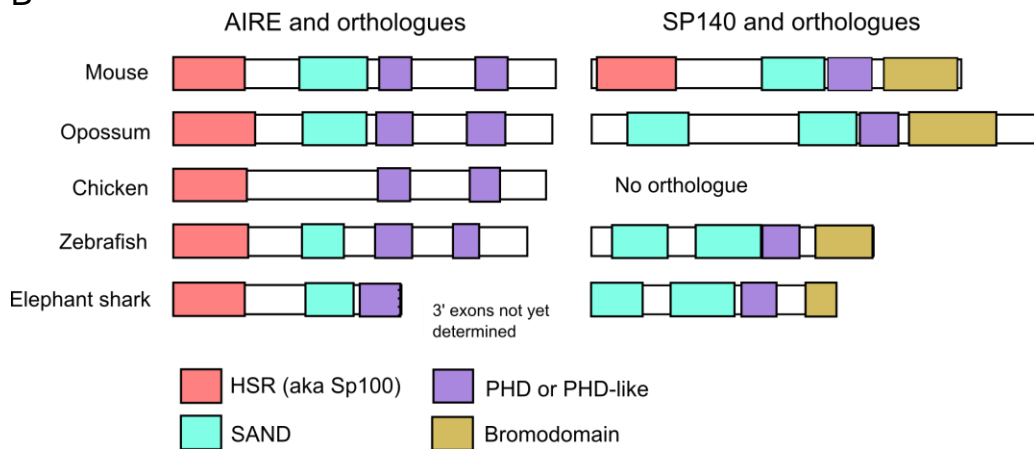
Figure S2.2. FACS of mTECs for ChIP-seq.

(A) Fluorescence-assisted cell sorting gating strategy for obtaining GFP$^-$ and GFP$^+$ mTECs for ChIP-seq. A typical experiment involved thymi pooled from 6-12 *Adig* mice yielding 150,000 to 250,000 GFP$^+$ and 100,000 to 200,000 GFP$^-$ cells. (B) A small number of sorted cells from a representative sort were analyzed for purity. The way in which the very small portion of GFP$^-$ sorted cells were loaded is suspicious for carry-over from the pre-sorted sample being the source of much of the high proportion of CD45$^+$ EpCAM$^-$ in the GFP$^-$ purity check.

Figure S2.2

A

B

GFP+ mTECs

Tube: GFP+			
Population	#Events	%Parent	%Total
All Events	1,240	####	100.0
size	1,141	92.0	92.0
DAPI	1,085	95.1	87.5
CD11c neg	1,085	100.0	87.5
Epcam 45	1,014	93.5	81.8
TECs	1,012	99.8	81.6
mTECs	1,004	99.2	81.0
mTEC hi	923	91.9	74.4
mTEC lo	73	7.3	5.9
GFP+	942	93.8	76.0
GFP alt	987	98.3	79.6
NOT(GFF	17	1.7	1.4

GFP- mTECs

Tube: GFP-			
Population	#Events	%Parent	%Total
All Events	235	####	100.0
size	104	44.3	44.3
DAPI	74	71.2	31.5
CD11c neg	74	100.0	31.5
Epcam 45	16	21.6	6.8
TECs	15	93.8	6.4
mTECs	15	100.0	6.4
mTEC hi	6	40.0	2.6
mTEC lo	9	60.0	3.8
GFP+	0	0.0	0.0
GFP alt	2	13.3	0.9
NOT(GFF	13	86.7	5.5

65

Figure S2.3. Enrichment of H3K4me2 and H3K27ac near ACNS1 in *Aire*-expressing cells.

Alignment of ChIP-seq results for the region spanned by the 461E7 BAC. Upper rows show unique reads from H3K27ac ChIP-seq of GFP$^+$ and GFP$^-$ mTECs from *Adig* mice (performed in triplicate) and D10 cells and lower rows show unique reads from H3K4me2 ChIP-seq of GFP$^+$ and GFP$^-$ mTECs from *Adig* mice (performed in triplicate).

Figure S2.3
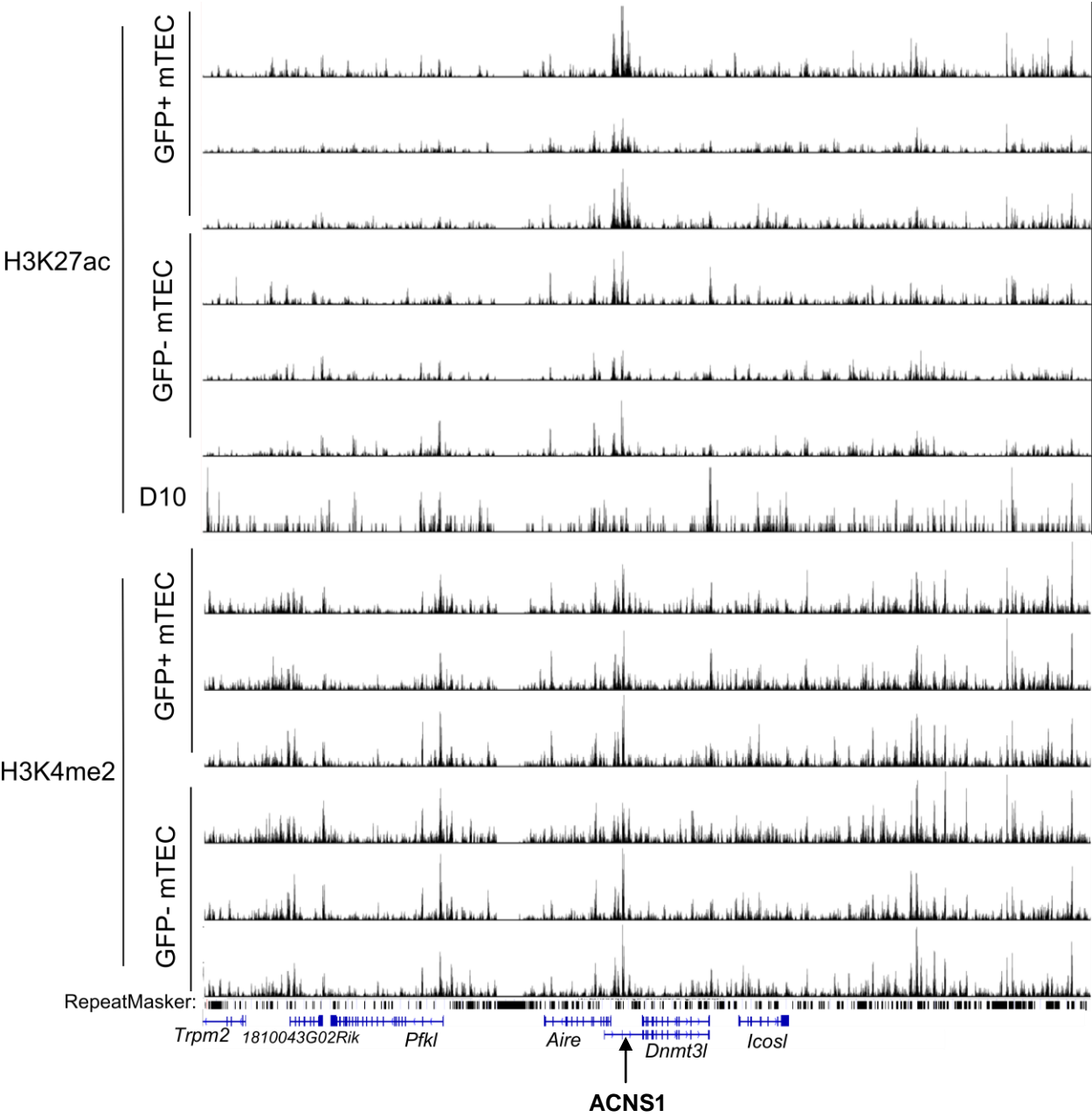
Figure S2.4. A transgene including 3 kb of sequence upstream of *Aire* was not expressed.

(A) Schematic of the *Eaig* (extended *Aire*-promoter-driven IGRP-GFP) transgene, modified from the *Adig* reporter construct. (B) Immunofluorescent staining for GFP (green) and Aire (red), in frozen thymic sections from *Adig*, *Eaig*, and WT mice. (C) Immunofluorescent staining for GFP (green) and Aire (red) in frozen splenic sections from *Adig* and *Eaig* mice. Arrows indicate nuclear Aire staining.

Figure S2.4

A

EAIG CNS1 *Aire Ex1, part of 5' UTR* B-globin splice SV40 pA | IGRP | GFP |

├──── 3.2 kb ────┤

B *Adig* *EAIG* founder 1 *EAIG* founder 2 *EAIG* founder 3 WT

Thymus

GFP Aire

C *Adig* *EAIG* founder 1 *EAIG* founder 2 *EAIG* founder 3

GFP Aire
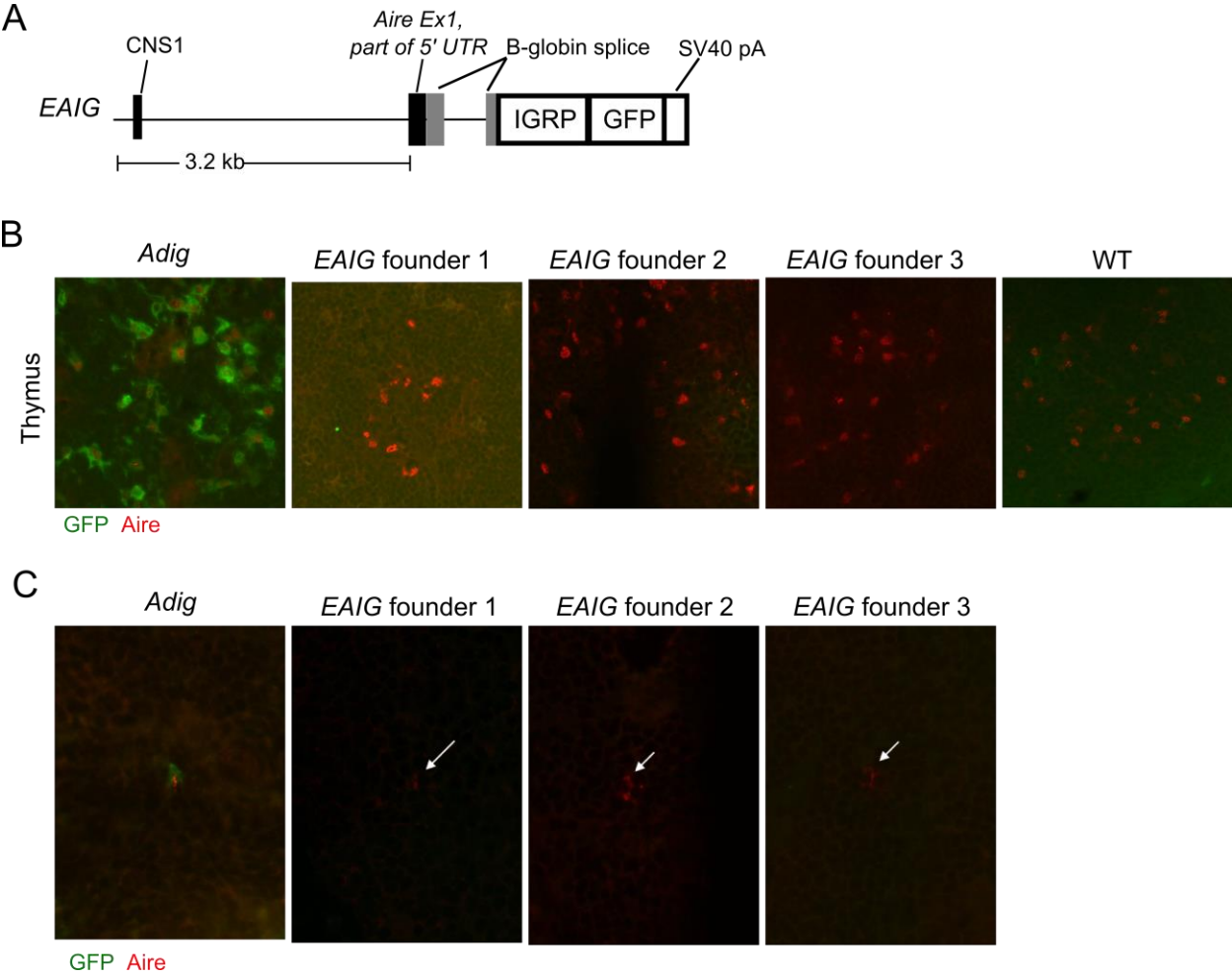
Figure S2.5. The small reporter transgene, *A4dig*, recapitulates *Aire* expression in mTECs in half of founder lines.

(A) Flow cytometry of mTECs, gated as CD45$^-$, EpCAM$^+$, MHC-II$^+$, Ly51$^-$, showing the frequency (%) of GFP$^+$ cells (left) and the frequency (%) of Aire$^+$ cells among these GFP$^+$ mTECs (right), in *Adig*, all *A4dig* lines, and WT mice. (B) Flow cytometry of eTACs, gated as CD45^{int}, MHC-II^{hi}, EpCAM$^+$, CD86$^-$, showing the frequency (%) of GFP$^+$ cells in the same mice as (A). (C) Median fluorescence intensity of GFP among GFP$^+$ cells in *Adig* and *A4dig* mice. (D) Plot correlating transgene copy number, determined by quantitative PCR, to the frequency of GFP$^+$ cells among mTECs. Data were gathered over three experiments. .

Figure S2.5
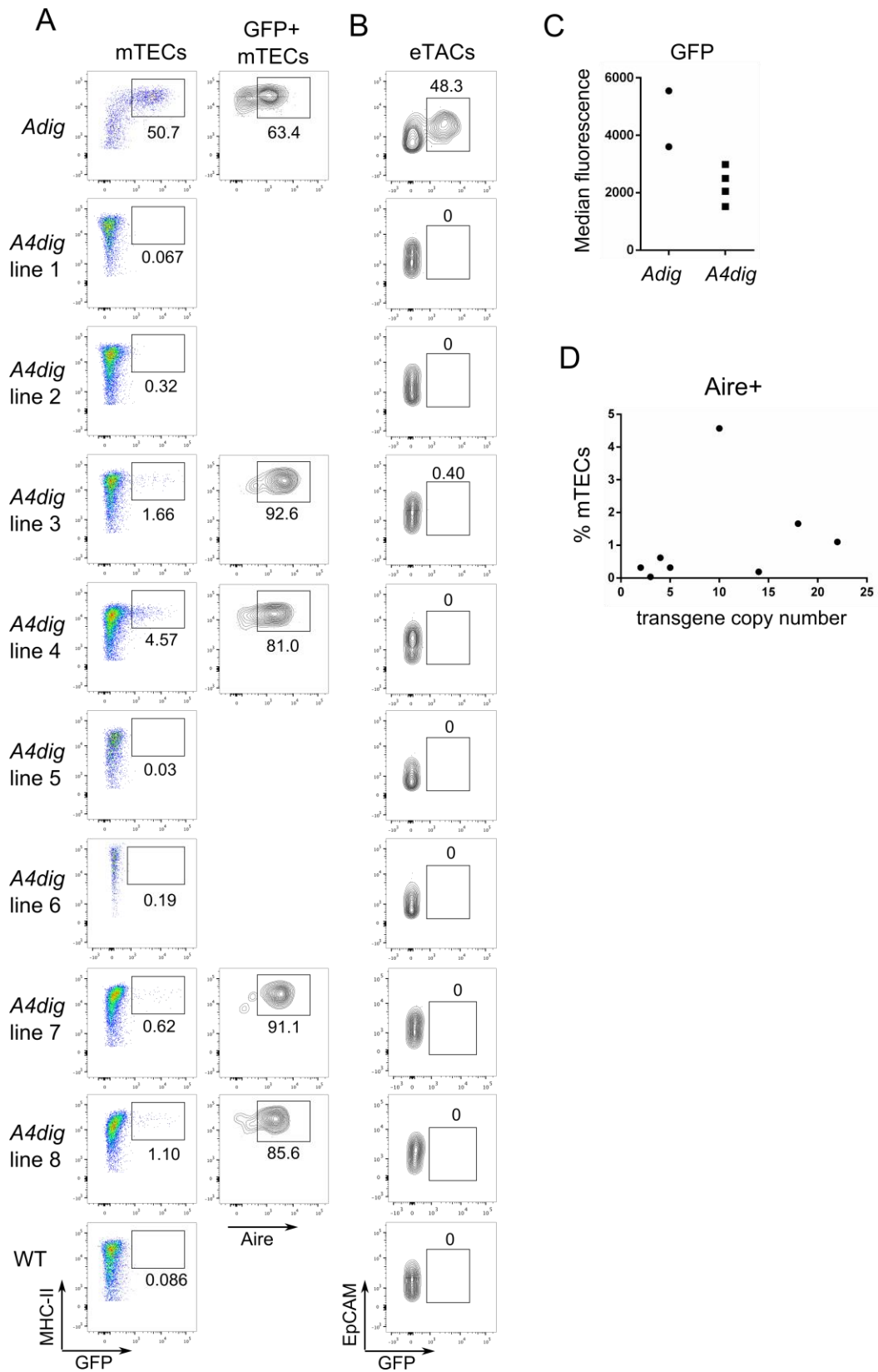
Figure S2.6. *Aire* reporter BAC lacking *ACNS1* is not expressed.

(A) Schematic of *AdigΔACNS1* transgene, showing how plasmid was used to modify 461E7 BAC by recombineering. (B) Flow cytometry of mTECs, gated as CD45$^-$, EpCAM$^+$, MHC-II$^+$, Ly51$^-$, showing MHC-II x Aire subset gating and the frequency (%) of GFP$^+$ cells among Aire$^+$ mTECs of WT, *Adig*, and *AdigΔACNS1* mice. Graph (below) summarizes data from three independent experiments; each dot is one mouse; short horizontal lines are means. (C) Flow cytometry of eTACs, gated as CD45^{int}, MHC-II^{hi}, EpCAM$^+$, CD86$^-$, showing (right) the frequency (%) of GFP$^+$ cells among eTACs of WT, *Adig*, and *AdigΔACNS1* mice. Graph (below) summarizes data from three independent experiments; each dot is one mouse; short horizontal lines are means. (D) Flow cytometric tetramer analysis of CD4 SP thymocytes, gated as DAPI-, CD4+, CD8-, showing histograms of 8.3 TCR-specific tetramer-binding cells and. irrelevant tetramer-binding cells in WT, *Adig,* and *AdigΔACNS1* mice.
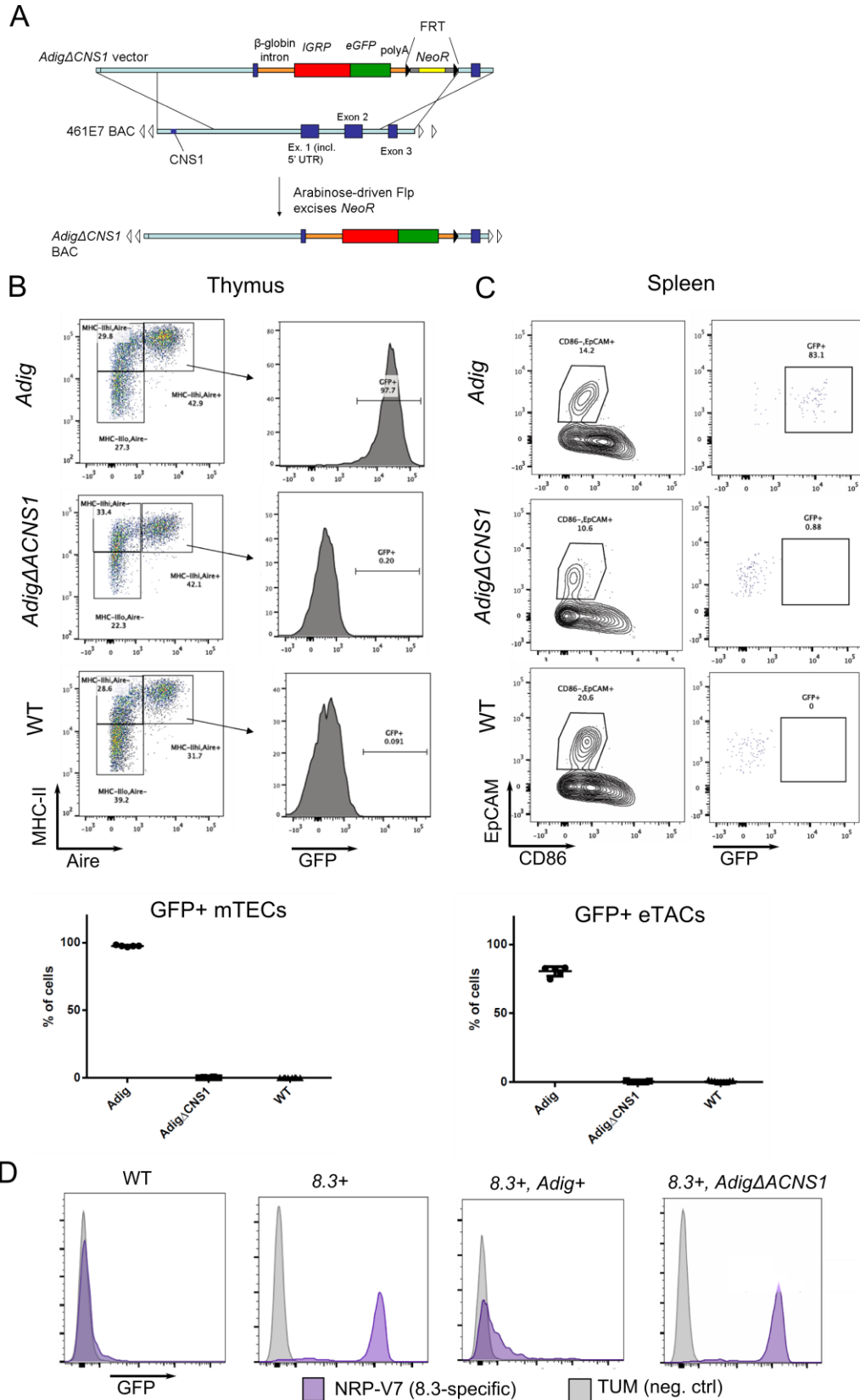
Figure S2.6

Figure S2.7. CRISPR-Cas9-mediated deletion of ACNS1.

(A) Schematic of ACNS1 locus, with PhastCons elements and mVista-identified CNS labeled. Arrows indicate locations of candidate guide RNAs. Stars mark guide RNAs ultimately used to create ACNS1$^{-/-}$ mouse. (B) Surveyor assay of ES cells transfected with Cas9 construct that included one of the guide RNAs specified in (A). + and - indicate positive and negative control lanes, respectively. Performed by UCSF ES Cell Targeting Core. (C) Representative gel electrophoresis of products of ACNS1 genotyping PCR. (D) Table summarizing the ACNS1-targeting CRISPR-Cas9 injection conditions and results.

Figure S2.7

A

sgRNA1 *

mm9, chr10:77509050-77509450

PhastCons

mVista ACNS1

B

Surveyor assay

uncleaved amplicon →

Cel1 cleavage
products of
heteroduplexes
(indel-containing
and WT)

C

ACNS1: -/- +/- +/+ +/-

1000 bp
700
500

D

CRISPR deletion		
injection date	5/15/14	5/22/14
cas9 mRNAs	50ng/ul	25ng/ul
sgRNAs	20ng/ul each	10ng/ul
# of injected	140	180
# of planted	96	155
pregnancy	2/3 females pregrant	5/5 females
# of pups born	7	37
# of edited pups	2	6
# of pups with expected D	1	6*

*one pup also had another editing event on the other chromosome

Figure S2.8. Additional effects of ACNS1 deletion on mTECs.

(A) Graph summarizes flow cytometry results of frequency (%) of Aire$^+$ cells among mTECs, gated as CD45$^-$, EpCAM$^+$, MHC-II$^+$, Ly51$^-$, of WT and ACNS1$^{+/-}$ mice. (B) Graph summarizes median fluorescence intensity of Aire, normalized within each experiment to that of WT mice, in Aire$^+$ mTECs, gated as CD45$^-$, EpCAM$^+$, MHC-II$^+$, Ly51$^-$, of WT and ACNS1$^{+/-}$ mice. (C) Quantitative PCR analysis of *Aire*, comparing RNA from WT and ACNS1$^{+/-}$ mTECs, showing mean ± SD from three independent experiments, normalized to *Actb*. (D) Flow cytometry of mTECs, gated as CD45$^-$, EpCAM$^+$, MHC-II$^+$, Ly51$^-$, showing frequency (%) of MHC-II^{lo} cells. Bar graph (right) shows mean ± SD frequencies of MHC-II^{lo} cells among mTECs from WT and ACNS1$^{-/-}$ mice. (E) Flow cytometry of mTECs, gated as CD45$^-$, EpCAM$^+$, MHC-II$^+$, Ly51$^-$, showing frequency (%) of Aire$^+$ cells. Bar graph (right) shows mean ± SD frequencies of Aire$^+$ cells among mTECs from WT and compound heterozygous ACNS1$^{-/-}$ mice. Data in (A) and (B) pooled from seven independent experiments, for a total of 16 WT and 15 ACNS1$^{+/-}$ mice. Data in (D) and (E) are from at least two or three independent experiments containing one or more individual mice within each group and totaling five or six individual mice per group. Data analyzed with Student's *t*-test. * $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$

Figure S2.8
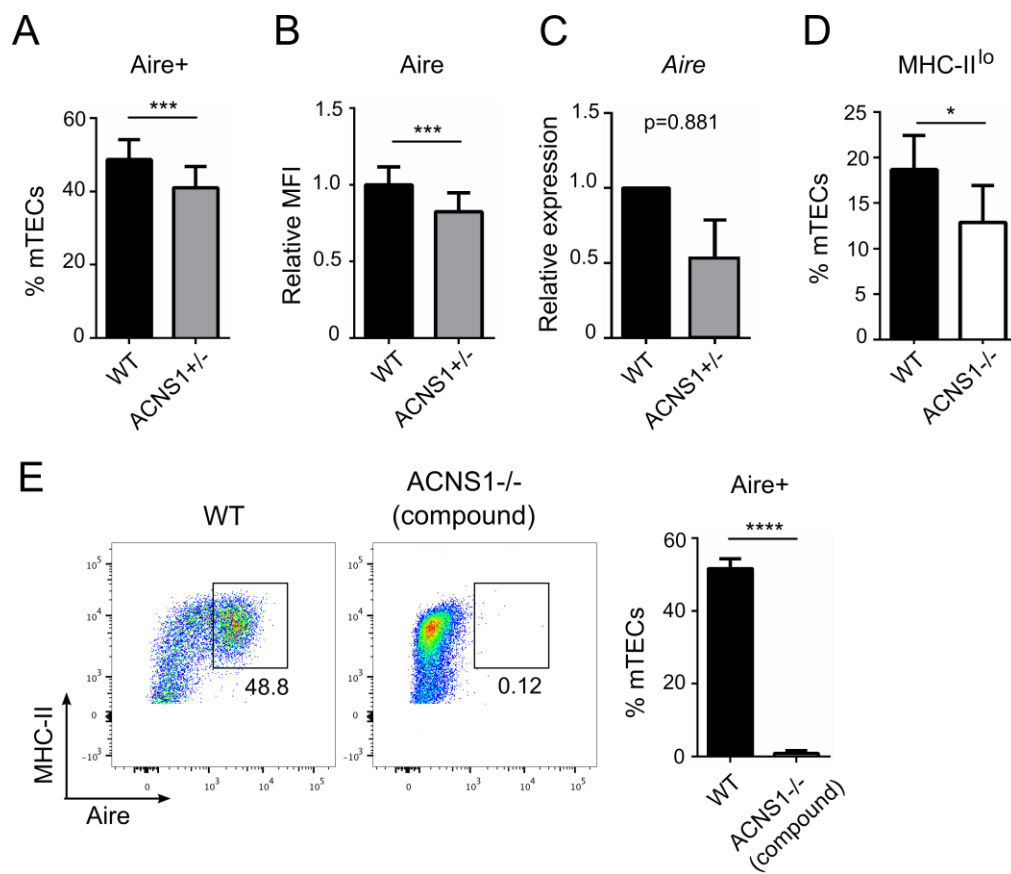
Chapter 3: ACNS1 is essential for thymic *Dnmt3l* expression

Overview:

The DNA (cytosine-5-)-methyltransferase 3-like (*Dnmt3l*) gene is a paralogue of the de novo DNA methyltransferases *Dnmt3a* and *Dnmt3b*. Dnmt3L is enzymatically inactive but facilitates the activities of Dnmt3a and Dnmt3b and is essential for maternal imprinting and methylation of retrotransposons. *Dnmt3l* and *Aire*, which are adjacent in the genomes of placental mammals, have similar, highly restricted patterns of expression, including co-expression in mature mTECs. However, regulation of *Dnmt3l* expression, particularly in the thymus, is poorly understood. We observe that *Dnmt3l* expression does not require *Aire* and, interestingly, *Dnmt3l* transcript is increased in *Aire$^{-/-}$* mice. We examined *Dnmt3l* expression in ACNS1$^{-/-}$ mice, which lack a DNA element required for *Aire* expression, and found that these mice also had a complete loss of *Dnmt3l* expression in the thymus. We present evidence that ACNS directly co-regulates both *Aire* and *Dnmt3l*, suggesting a possible role for *Dnmt3l* in TSA expression.

Introduction:

Aire is located within a gene-rich region, with five genes within 100 kb. *Pfkl, 18100423G02Rik ,* and *Trpm2* are downstream of *Aire,* and *Dnmt3l* and *Icosl* are upstream. This presents a clear example of finely tuned transcriptional regulation in mammals, as these genes have distinct transcriptional profiles despite their proximity.

At the time of the discovery of *AIRE,* researchers had already mapped *PFKL*, the gene immediately downstream of *AIRE* (34,35). A few years later, the gene immediately upstream of *AIRE* was identified in humans (163) and mice (164). The gene was called *DNMT3L* because it showed significant conservation with *DNMT3A* and *DNMT3B*. In placental mammals, *AIRE* and *DNMT3L* are oriented head-to-head such that they transcribe away from each other. In humans, the translation start sites of these two genes are located 24 kb apart. In contrast, in mice the translation start sites are separated by just 6 kb; indeed, one isoform of *Dnmt3l* includes an untranslated exon located in the third intron of *Aire* (165)*.* Much of the larger gap between *AIRE* and *DNMT3L* in humans can be explained by the presence of a number of repetitive sequences that are absent from that location in the mouse genome.

Although it is a paralogue of *Dnmt3a* and *Dnmt3b*, which perform de novo methylation of cytosines, *Dnmt3l* lacks certain motifs required for eynzymatic activity (163). Instead, Dnmt3l participates in de novo methylation by directly interacting with Dnmt3a and Dnmt3b and enhancing their activity (166,167). The C-terminal half of Dnmt3L is sufficient for this facilitation of methyltransferase activity. The N-terminal half of Dnmt3L largely consists of a PHD-like domain and binds to H3K4me0, which is, interestingly, the same histone bound by the PHD1 domain of Aire (85,168). Dnmt3a- and Dnmt3b-mediated methylation is essential for development. *Dnmt3a$^{-/-}$, Dnmt3b$^{-/-}$* mice die by embryonic day 10.5 (E10.5). These two genes show some redundancy, as *Dnmt3a$^{-/-}$* mice appear normal at birth but fail to thrive and die by several weeks of age, and *Dnmt3b$^{-/-}$* die mice in utero a few days before birth (169). In contrast to the severe phenotypes of *Dnmt3a$^{-/-}$* and *Dnmt3b$^{-/-}$* mice, *Dnmt3l$^{-/-}$* mice grow normally after birth and are viable, with their only obvious defect being infertility of both males and females (170). In female *Dnmt3l$^{-/-}$* mice, this infertility emerges from a failure to establish appropriate

maternal imprinting in oocytes, leading to death of resulting embryos at E9.5 or E10.5, even if the embryo is itself *Dnmt3l*[+/-] or is transferred to a WT mother for gestation (170,171). In male germ cells, the principal problem is insufficient methylation of retrotransposons, which allows their reactivation and results in meiotic catastrophe in spermatocytes (172,173).

The initial studies of *Dnmt3l* observed it to be prominently expressed in the testis, ovary, and thymus (163,164). Since then, careful studies of *Dnmt3l* expression in germ cells have carefully plotted its expression over time . It is expressed in the developing testis by E13.5, peaks at E15.5, and is largely absent by post-natal day 6 (P6); *Dnmt3l* expression therefore coincides with the primary period of de novo methylation of the male germ cell (174). In the ovary, *Dnmt3l* expression begins in the first few days after birth, peaks at P16, and is present at a moderate level into adulthood; like the expression observed in testes, *Dnmt3l* expression in oocytes coincides with a period of peak germ cell de novo methylation, in this case during the oocyte growth phase (174). Prospermatogonia and oocytes express different isoforms of *Dnmt3l* (165). The originally identified transcript, consisting of 12 exons and beginning transcription roughly 6 kb away from *Aire*, is seen in prospermatogonia and ES cells. In contrast, oocytes express a longer isoform of *Dnmt3l* that includes two novel exons and omits the first exon of the prospermatogonia isoform. The first of these novel exons is located in the third intron of *Aire* and the second is roughly midway between *Aire* and the translation start site of *Dnmt3l*. Both the oocyte and prospermatogonia isoforms appear to encode identical protein sequences. Finally, three short, untranslated isoforms of *Dnmt3l* are observed in spermatocytes. The methylation status of the *Dnmt3l* promoter plays an important role in determining which isoform is transcribed. When the primary *Dnmt3l* promoter is unmethylated, as it is in ES cells and prospermatogonia, transcription occurs from this point. In growing oocytes, however, this

promoter is methylated, leading to transcriptional initiation at the distal exon located within an intron of *Aire* (175). The regulation of *Dnmt3l* in *trans* more poorly understood, although it is clear that the promoter methylation observed in oocytes is at least partially dependent on *Dnmt3l* itself (175).

The most prominent defect in *Dnmt3l*$^{-/-}$ mice is their infertility, so it is unsurprising that research has focused on elucidating the role and regulation of this gene in germ cells. Such research has ignored thymic expression *Dnmt3l,* however, despite recognition since the discovery of the gene that *Dnmt3l* is expressed there (163,164). We were drawn to this topic after observing the strong resemblance between the expression patterns of *Aire* and *Dnmt3l* and began to investigate thymic *Dnmt3l* expression. After determining that expression of both genes is concentrated in mature mTECs, we proceeded to characterize the expression of *Dnmt3l* in the absence of ACNS1, a conserved CRE that is essential for *Aire* expression, and found that ACNS1 is likewise required for *Dnmt3l* expression.

Results and Discussion:

Aire and Dnmt3l are co-expressed

The highly specific expression patterns of *Dnmt3l* and *Aire* are quite similar. *Dnmt3l* is principally expressed in prospermatogonia, growing oocytes, ES cells, and the thymus—the thymic cell type(s) in which *Dnmt3l* is expressed is less well identified; one publication asserted that quantitative RT-PCR of thymic epithelial cells revealed high levels of *Dnmt3l*, but no data were shown (176). *Aire* is expressed most prominently in mature mTECs and in eTACs, a hematopoietic cell type found in secondary lymphoid tissues; *Aire* is also expressed in the early embryo, ES cells, and some developing spermatocytes (74–77). Publicly available expression data sets contained more unbiased evidence that these two genes are generally co-expressed. We

began by using a microarray-based mouse expression atlas of 91 tissues and cell types, which is available at www.biogps.org,to (177,178). According to the website's expression correlation algorithm, *Aire* and *Dnmt3l* have a correlation of 0.9208, making *Dnmt3l* the fourth-best correlated gene to *Aire* and *Aire* the 28th-best correlated to *Dnmt3l.* In order to better understand whether *Aire* and *Dnmt3l* were expressed in the same cells in the thymus, we examined data gathered by the Immunological Genome Project, which has performed RNA-seq on dozens of lymphoid, myeloid, and stromal cell types (www.immgen.org and (179,180)). We observed that *Dnmt3l* expression largely correlated with *Aire* expression, with both *Aire* and *Dnmt3l* being most highly expressed in MHC-II^{hi} mTECs (Figure 3.1A). In contrast, expression of *Icosl* and *Pfkl*, the other two genes nearest *Aire* on the genome, was not as well correlated with that of *Aire*. These data, which suggest that thymic *Dnmt3l* expression is restricted to mature mTECs, builds on the previously mentioned report that thymic *Dnmt3l* expression is limited to thymic epithelial cells.

We validated these findings by examining RNA-seq data obtained using our recently published *Aire* lineage-tracing mouse (73). This mouse has an *Aire-Cre* transgene, in which the first few exons of *Aire* in the RP23-461E7 BAC are replaced with *Cre*, fused to two ERT2 domains to prevent nuclear localization in the absence of tamoxifen. The mouse also a *Rosa26-loxP-STOP-loxP-TdTomato* transgene. Upon administration of tamoxifen, Cre localizes to the nucleus, excises the STOP cassette, and permanently activates RFP expression in the *Aire*-expressing cells and their progeny. We performed high-throughput sequencing of RNA (RNA-seq) on four subsets of lineage-traced mTECs that were harvested one week after tamoxifen administration: MHC-II^{lo} RFP$^-$, which we call pre-Aire mTECs; MHC-II^{hi}, RFP$^-$, which we call early Aire mTECs; MHC-II^{hi}, RFP$^+$, which we call late Aire mTECs; and MHC-II^{lo}, RFP$^+$,

which we call post-Aire mTECs. The RNA-seq data revealed that both *Aire* and *Dnmt3l* are dramatically upregulated in both MHC-II^{hi} mTEC subsets compared to pre-Aire cells; *Aire* and *Dnmt3l* both show low levels of expression in post-Aire cells (Figure 3.1B). Neither *Icosl* nor *Pfkl*, which are also located near *Aire*, show the same large changes in expression.

One possible explanation for this shared pattern of thymic expression is that *Dnmt3l* is simply a TSA, a large number of which are expressed in mTECs, where they help to tolerize self-reactive T cells. Several microarray-based genome-wide expression profiles have been made comparing WT and *Aire*$^{-/-}$ mTECs. Although the first such assay did not include probes against *Dnmt3l* (44), more recent analyses do evaluate *Dnmt3l* expression. We examined the results of one arbitrarily selected study that included both WT and *Aire*$^{-/-}$ MHC-II^{hi} mTECs (114) and found that, far from being dependent on Aire for expression, *Dnmt3l* transcripts were more abundant in *Aire*$^{-/-}$ mice (Figure 3.1C). This study also revealed increased *Aire* transcript in *Aire*$^{-/-}$ mice compared to WT; this particular *Aire*$^{-/-}$ mouse was created by targeted deletion of exon 2, which leads to a frame shift and a premature stop codon after amino acid 48 but does not prevent transcription (44). One possible explanation for this moderately increased amount of *Aire* transcript in *Aire*$^{-/-}$ mice is the existence of a feedback mechanism in which an absence of Aire protein or activity spurs increased transcription. Two other genes near *Aire*, *Pfkl* and *Icosl* do not show the same dramatic increase in transcript as *Dnmt3l*. Another study that compared gene expression of MHC-II^{hi} mTECs from *Aire*$^{-/-}$ mice to those from WT mice similarly found a large increase in *Dnmt3L* expression in *Aire*$^{-/-}$ mTECs (Figure S3.1A and (91)). Expression profiling of a different *Aire*-deficient mouse, in which a 13 bp was deleted from exon 8, did not show the same upregulation of *Dnmt3l* but also did not show a decrease in its expression (Figure S3.1B and (181)). We validated these studies by performing quantitative RT-PCR against RNA

83

Figure 3.1. *Aire* and *Dnmt3l* are co-expressed.

(A) Heat map showing expression of *Aire* and the *Aire*-flanking genes *Dnmt3l, Icosl*, and *Pfkl* across the range of cell types profiled in the Immunological Genome Project; each column is a different cell type; columns are ordered by level of *Aire* expression; color indicates level of expression, from red (high) to blue (low). (B) Expression of *Aire* and *Aire*-flanking genes in various mTEC subsets, determined by RNA-seq of sorted mTECs from tamoxifen-treated *Aire* lineage-tracing (*Aire-Cre*, *Rosa26-loxP-STOP-loxP-TdTomato*) mice. Subsets were defined as: MHC-II^{lo}, RFP$^-$ = "pre-Aire", MHC-II^{hi}, RFP$^-$ = "early Aire", MHC-II^{hi}, RFP$^+$ = "late Aire", and MHC-II^{lo}, RFP$^+$ = "post-Aire". Graph shows mean ± SD of number of reads aligned to each gene, normalized to control for variation in the total number of reads per sample, and additionally normalized such that the mean of the pre-*Aire* samples equals 1. (C) Expression of *Aire* and *Aire*-flanking genes in sorted MHC-II^{hi} mTECs from WT or *Aire*$^{-/-}$ mice, assessed by microarray, bars showing mean ± SD. This data set was previously published (114). The raw expression values are normalized such that the mean of the WT samples equals 1. (D) Quantitative RT-PCR analysis of *Aire* and *Dnmt3l,* comparing RNA from WT and *Aire*$^{-/-}$ mTECs, , showing mean ± SD (technical replicates), normalized to *Actb*. In both (C) and (D), the *Aire*$^{-/-}$ mice lack exon 2, as described in (44).

Figure 3.1
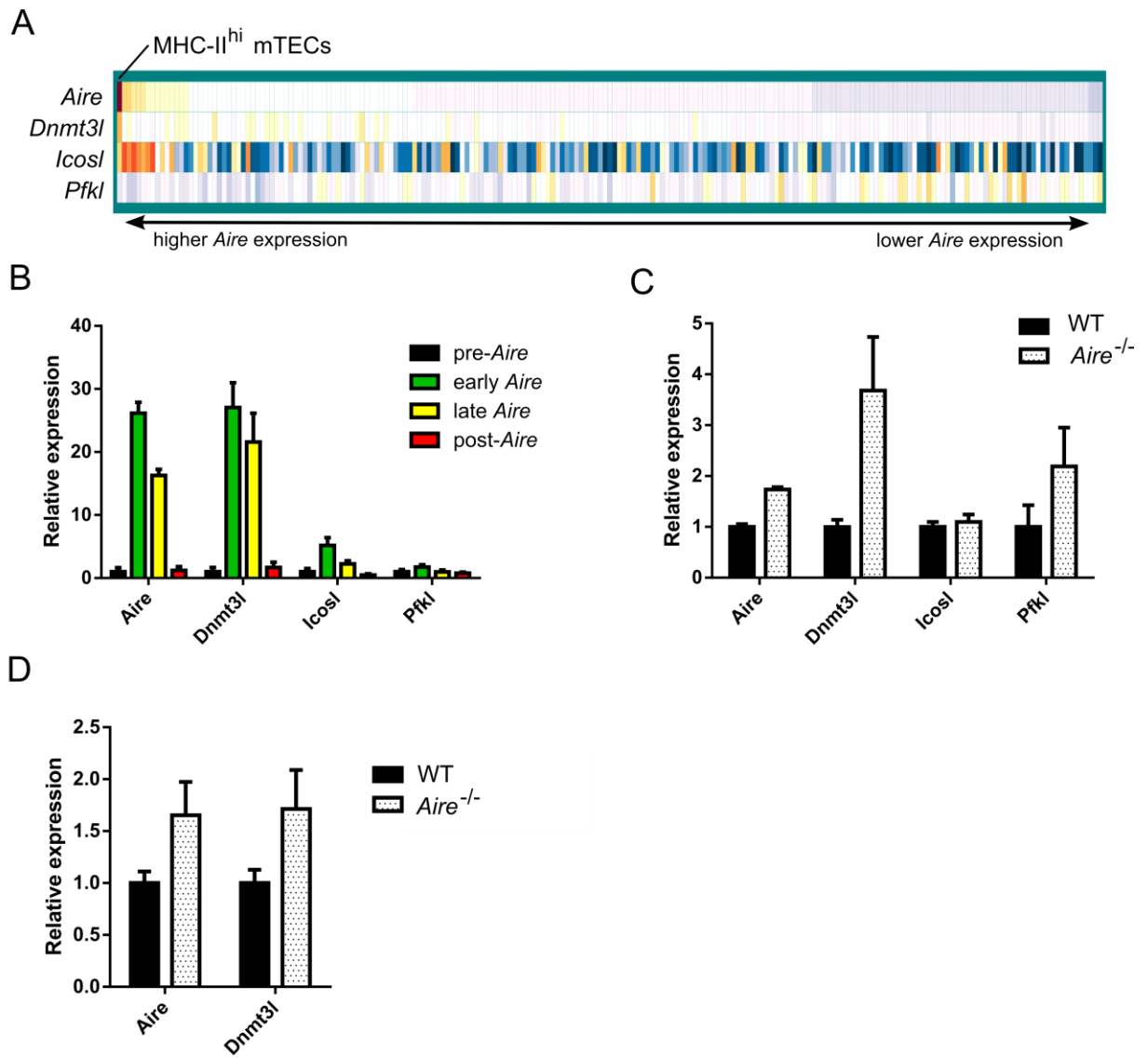
from sorted WT and *Aire*^{-/-} mTECs and found that both *Aire* and *Dnmt3l* transcript levels were approximately 70% higher in *Aire*^{-/-} mTECs, which is generally in line with previous studies (Figure 3.1C). Together, these results show that *Dnmt3l* is not an *Aire*-dependent TSA.

It is more difficult to rule out that *Dnmt3l* is an *Aire*-independent TSA, and in fact brings up the question of how a TSA is defined. If the pattern of expression alone determines what a TSA is, then *Aire* itself would be considered to be one. Another reasonable definition, however, might specify that the definition of a TSA requires a functional component—namely, that the role of the TSA in mTECs is to be expressed and presented on MHC such that thymocytes specific for the antigen are eliminated. In order to demonstrate that *Dnmt3l* is not an *Aire*-dependent TSA according to this definition, it would be necessary to identify an additional function for *Dnmt3l* in mTECs.

ACNS1 is required for the thymic expression of Dnmt3l

The observation that *Dnmt3l* expression is increased in *Aire*^{-/-} mice, to at least the same degree as the increase in *Aire* transcript, led us to wonder whether these two genes might not just be co-expressed but also co-regulated. We decided to assay the expression of *Dnmt3l* in ACNS1^{-/-} mice, which lack a conserved DNA element that is essential for *Aire* expression and immune tolerance, as discussed in the previous chapter. In contrast to the traditionally targeted *Aire*^{-/-} mouse (which have deletion of exon 2), mTECs from ACNS1^{-/-} mice lack not only Aire protein but also *Aire* transcript (Figure 2.4D). ACNS1^{-/-} mice also have a nearly complete loss of *Dnmt3l* expression in mTECs (Figure 3.2A). *Pfkl* and *Icosl*, the other two genes closest to *Aire*, were still expressed in ACNS1^{-/-}, suggesting that the loss of *Dnmt3l* expression was not simply due to a complete collapse of gene expression in the general region near *Aire*.

There are several general mechanisms by which deletion of ACNS1 could lead to loss of both *Aire* and *Dnmt3l* expression. First, ACNS1 could directly enhance *Aire* expression and *Aire* in turn enhance *Dnmt3l* expression. Assuming that the *Aire* transcript itself does not function as a regulator in *trans*, this possibility was eliminated by the observation that *Dnmt3l* is expressed in *Aire*[-/-] mice, which have nearly full-length *Aire* transcript but severely truncated Aire protein. A second possible explanation is that ACNS1 consists of a single CRE that co-regulates both *Aire* and *Dnmt3l*. Third, the 269-bp region absent in the ACNS1[-/-] mouse contains two distinct, separable CREs, one regulating *Aire* and another regulating *Dnmt3l*. Further research is needed to distinguish between these two possibilities. One point in favor of the model of a single element co-regulating both genes is that the highly conserved region within ACNS1 is quite small, only approximately 60 bp. Another possibility is that ACNS1 helps preserve a topological domain that contains both *Aire* and *Dnmt3l*, which collapses in its absence. The final possibility is that ACNS1 could control *Dnmt3l* and *Dnmt3l* in turn be required for *Aire* expression.

The most direct way to assay this last possibility would be to determine whether *Dnmt3l*[-/-] mice are able to express *Aire*. We have recently acquired *Dnmt3l*[+/-] mice and are breeding them to homozygosity in order to carry out this study. In the interim, however, we also have noted two less direct piece of evidence that argues against *Aire* expression depending on *Dnmt3l*. We previously observed that compound heterozygous mice that lack ACNS1 on one chromosome and exon 2 of *Aire* on the other chromosome (ACNS1[+/-], *Aire*[+/-] mice) lack Aire+ mTECs to the same degree that homozygous ACNS1[-/-] mice do (Figure 2.4I). Using quantitative RT-PCR, we found that mTECs from ACNS1[+/-], *Aire*[+/-] mice express *Dnmt3l* (Figure 3.2B). In other words, these mice are Aire-deficient despite having a nearly normal amount of *Dnmt3l* transcript. This result is not what we would expect to see in ACNS1[+/-], *Aire*[+/-] mice if the failure of ACNS1[-/-]

Figure 3.2. ACNS1 co-regulates *Aire* and *Dnmt3 in mTECs*.

(A) Quantitative RT-PCR analysis of *Aire* and *Aire*-flanking genes, including *Dnmt3l,* comparing RNA from mTECs from WT and ACNS1$^{-/-}$ mice, showing mean ± SD (technical replicates), normalized to *Actb*. (B) Quantitative RT-PCR analysis of *Aire* and *Dnmt3l*, comparing RNA from mTECs from WT and *Aire*$^{+/-}$*,* ACNS1$^{+/-}$ (compound heterozygote) mice, showing mean ± SD (technical replicates), normalized to *Actb*. (C) Expression of *Aire* and *Dnmt3l* in ES cells treated with an irrelevant shRNA (against GFP) or one of three shRNAs targeting *Dnmt3l*, showing mean ± SD, assessed by oligonucleotide beadchip. This data set was previously published (182).

Figure 3.2
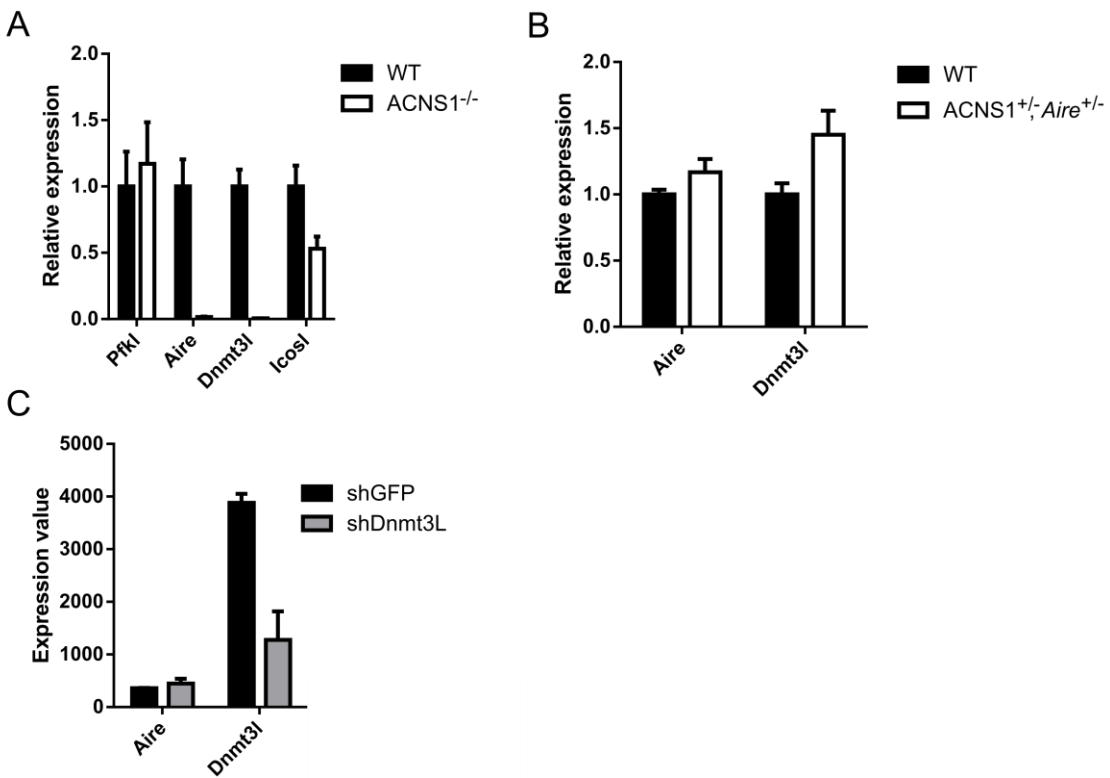
mice to express *Aire* were due to *Aire* expression being dependent on *Dnmt3l,* instead of ACNS1 directly regulating *Aire,* because ACNS1$^{+/-}$, *Aire*$^{+/-}$ mice still have one normal *Aire* allele at which *Dnmt3l*-dependent transcription would be able to occur. As a complement to these data, we considered publicly available expression data from shRNA-mediated knockdown of *Dnmt3l* in ES cells (182). This knockdown, which reduced *Dnmt3l* transcript levels by approximately 70%, did not have any effect on *Aire* transcript levels (Figure 3C). This suggests that *Aire* in mTECs does not require *Dnmt3l*, with the two major caveats being that it could be that *Aire* requires *Dnmt3l* in the thymus but not in ES cells and that it might be that the abundance of *Dnmt3l* transcripts was not decreased enough to reveal its role in inducing *Aire.* Still, in conjunction with the results with the ACNS1$^{+/-}$, *Aire*$^{+/-}$ mice, these findings provide strong evidence that the loss of *Aire* in ACNS1$^{-/-}$ mice is not an indirect effect mediated by loss of *Dnmt3l* expression. Instead, ACNS1 directly regulates both *Aire* and *Dnmt3l.*

ACNS1$^{-/-}$ mice do not recapitulate the Dnmt3l $^{-/-}$ mouse

Upon realizing that ACNS1$^{-/-}$ mice have defects in thymic *Dnmt3l* expression, we began examining these mice to see whether they recapitulate the defects of *Dnmt3l*$^{-/-}$ mice: namely, male hypogonadism and infertility in both sexes. The testes of two 15-wk-old ACNS1$^{-/-}$ were grossly indistinguishable from age-matched WT controls (data not shown). In addition, we set up four breedings involving one ACNS1$^{-/-}$ mouse and one WT or ACNS1$^{+/-}$ mouse, using two male ACNS1$^{-/-}$ mice and two female ACNS1$^{-/-}$ mice. Mice were six or seven weeks old at the time they were placed into breedings. All four breedings yielded two litters of viable mice within two months. These results demonstrate that the phenotype previously observed in *Dnmt3l*-deficient mice is not present in ACNS1$^{-/-}$ mice despite their lack of *Dnmt3l* in the thymus.

We hypothesize that this discrepancy between ACNS1$^{-/-}$ and *Dnmt3l*$^{-/-}$ mice reflects distinct regulatory mechanisms for thymic and germ cell expression of *Dnmt3l*, such that ACNS1 is dispensable for *Dnmt3l* expression in the testis and ovary. We are investigating this point in collaboration with the laboratory of Diana Laird, beginning with immunohistochemical characterization of *Dnmt3l* and *Aire* expression in germ cells from WT compared to ACNS1$^{-/-}$ neonatal mice.

Conclusion:

Dnmt3l is an enzymatically inactive de novo methyltransferase paralogue that facilitates the activity of *Dnmt3a* and *Dnmt3b*. Mice deficient in *Dnmt3l* are viable and appear grossly normal. However, both males and females are infertile, due to defects in methylation that prevent normal establishment of maternal imprinting and restraint of retrotransposons. *Dnmt3l* lies in head-to-head orientation and in close proximity with *Aire,* and has previously been observed to have several commonalities with *Aire*, including binding to H3K4me0 by a PHD domain and highly restricted patterns of expression that include ES cells, thymus, and gonads.

By examining publicly available data and our RNA-seq results from *Aire* lineage-tracing mice, we conclude that thymic *Dnmt3l* expression is restricted to mature mTECs, just as *Aire* expression is. We determine that *Dnmt3l* expression is not *Aire*-dependent and that *Dnmt3l* is, in fact, upregulated in the *Aire*$^{-/-}$ mouse in which there is deletion of exon 2. Intrigued by this co-expression of *Aire* and *Dnmt3l*, we next assayed *Dnmt3l* expression in ACNS1$^{-/-}$ mice, which lack an essential *Aire*-regulating CRE, and found a wholesale loss of thymic *Dnmt3l* expression. *Dnmt3l* expression is not lost in compound heterozygous ACNS1$^{+/-}$, *Aire*$^{+/-}$ mice, although these mice lack Aire$^+$ cells. Conversely, shRNA-mediated knockdown of *Dnmt3l* in ES cells does not result in less *Aire* transcript. Together, these two findings suggest that the lack of both *Aire* and

Dnmt3l expression in the absence of ACNS1 reveals co-regulation of both genes by this DNA element rather indirect regulation of *Aire* by ACNS1 via *Dnmt3l*. Finally, we note that despite the loss of *Dnmt3l* expression in the thymus, ACNS1$^{-/-}$ mice are not infertile. This raises the possibility that thymic and gonadal *Dnmt3l* are regulated by distinct CREs, an important question that we are currently seeking to answer. The finding that ACNS1 regulates both *Aire* and *Dnmt3l* in mTECs suggests a shared function for these two genes, both of which are already known to have large, genome-wide epigenetic or transcriptional effects. We will discuss this intriguing possibility in Chapter 5.

In sum, ACNS1, a highly conserved DNA element required for *Aire* expression, is also essential for thymic expression of *Dnmt3l*, hinting at the exciting possibility that *Dnmt3l* has a role in the induction of TSA expression in mTECs.

Materials and Methods:

Mice: C57BL/6.*Aire-Cre* mice were generated by standard cloning methods in conjunction with BAC recombineering and was previously described in (73). Briefly, an ERT2-Cre-ERT2 cassette ligated into the pAireFNF targeting construct, which was in turn created from JDFNF (74), which contains 5' and 3' *Aire* homology arms. The resulting plasmid was linearized and transformed into heat-induced SW105 bacteria already transformed with the 461E7 BAC, selected with kanamycin, and arabinose treated to induce Flp, causing *NeoR* excision. Highly purified *Aire-Cre* BAC was isolated and and injected into fertilized FVB/N oocytes by the UCSF Mouse Transgenic core. Mice analyzed were backcrossed onto C57BL/6 at least six generations. *Aire-Cre* mice were crossed B6;129.*Ai14* mice (Jackson), which have a *Rosa26-LoxP-STOP-LoxP-TdTomato* cassette and were previously described (183). All animals were housed and bred in specific pathogen-free conditions at UCSF. All experiments were approved by the Institutional Animal Care and Use Committee of UCSF.

RNA-seq of *Aire* lineage-traced subsets: C57BL/6.*Aire-Cre* mice received 2 mg of tamoxifen (Sigma-Aldrich), dissolved in corn oil (Sigma-Aldrich), by oral gavage. mTECs were collected 7 d later, isolated as previously described (74). Briefly, thymi were isolated, minced, and digested with DNase I and Liberase TM (Roche) before centrifugation on gradient of Percoll PLUS (GE Healthcare). The enriched stromal cells were stained with the antibodies against the surface markers indicated in the figure legends (BioLegend).). Cell sorting was performed using a FACS Aria III (BD). RNA was isolated from the sorted cells using the Quick-RNA MiniPrep Kit (Zymo Research, R1054S), omitting the on-column DNase I treatment. RNA was then provided to UCSF SABRE Functional Genomics Core. where. Rebecca Barbeau prepared the library using the Ovation RNA-Seq System (Nugen), which was followed by 50 bp SE RNA-seq on an Illumina platform, to an average read depth of 28.8 million reads per sample. Bioinformatic analysis was performed by Josh Pollack.

Quantitative PCR. RNA was isolated from sorted mTECs using RNeasy Plus Micro Kit (Qiagen) and reverse transcribed using oligo-dT primers and the SuperScript III Kit (Life Technologies). TaqMan gene expression assays (Applied Biosystems) were used for all targets. Quantitative PCR was performed using 7500 Fast Real-Time PCR System (Applied Biosystems). Transcript abundance was normalized to *Actb* and analyzed using the ΔΔCt method.

Publicly available microarray data: Heat map of Immunological Genome Project data was created using My GeneSet tool. All possible cell types were included. Microarray data were obtained using the Profile Graph tool of Geo2R platform, part of the Gene Expression Omnibus website (www.ncbi.nlm.nih.gov/geo). Gene sets: Figure 3.1C, GSE8563; Figure 3.2C, GSE44644; Figure S3.1A, GSE33878; Figure S3.1B, GSE28393.

Supplemental Data:

Figure S3.1: *Dnmt3L* expression is not *Aire*-dependent.

(A) Expression of *Aire* and *Aire*-flanking genes in sorted MHC-II^{hi} mTECs from WT or *Aire*$^{-/-}$ mice, assessed by microarray, bars showing mean ± SD. This data set was previously published (91). The raw expression values are normalized such that the mean of the WT samples equals 1.

(B) Expression of *Aire* and *Aire*-flanking genes in sorted mTECs from WT or *Aire*$^{-/-}$ mice, assessed by microarray, bars showing mean ± SD. This data set was previously published (181). The raw expression values are normalized such that the mean of the WT samples equals 1. The *Aire*$^{-/-}$ mice in (A) lack exon 2 whereas the *Aire*$^{-/-}$ mice in (B) have a 13-bp deletion from exon 8.

Figure S3.1

A

B
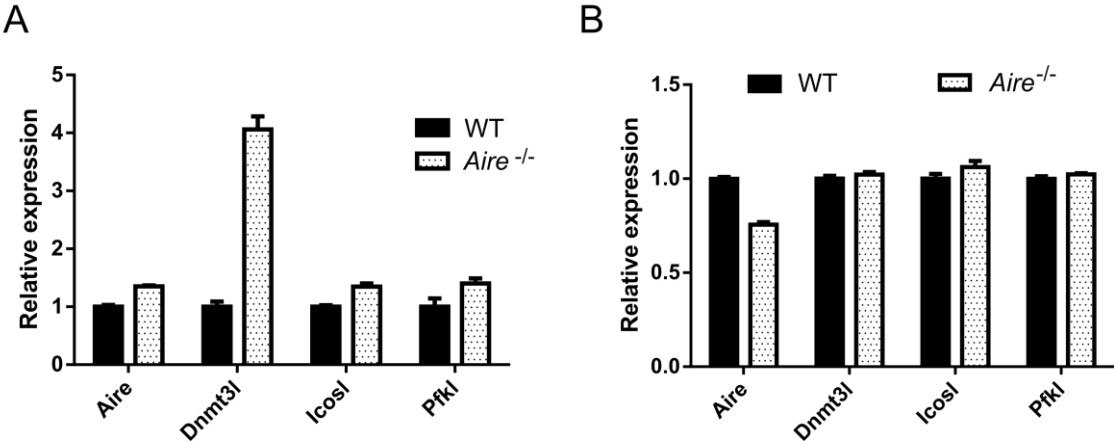
Chapter 4: Preliminary evidence for additional *Aire cis*-regulatory elements

Overview:

Genes are frequently regulated by multiple distal CREs. Although ACNS1 is clearly required for *Aire* expression, the thymic-restricted, variegated expression of the short, ACNS-containing transgene of the *A4dig* mouse suggests that additional *Aire*-regulating distal CREs exist. We made transgenic mice in which large regions of an *Aire* reporter BAC can be deleted by Cre-mediated recombination to determine whether these regions contain CREs affecting *Aire* expression. Our preliminary results suggest there are at least two additional CREs involved in *Aire* expression, one upstream and one downstream of *Aire*.

Introduction:

Just as a single CRE frequently has multiple transcription factor binding sites, a gene is frequently regulated by multiple distal CREs. At times, these CREs have distinct roles, acting to enhance expression of the gene in particular settings. For example, three distal CREs have been identified that control *Foxp3*, the defining gene of regulatory T cells (184). One of these elements is required for both thymic and peripheral regulatory T cells, one is dispensable in the thymus but required for peripheral induction, and one is required to maintain *Foxp3* expression in dividing regulatory T cells. It is plausible that a similar, modular system of regulation underlies regulation of *Aire* expression. Indeed, in light of the diversity of the cell types that express *Aire*— undifferentiated cells, developing germ cells (spermatocytes), differentiated endoderm-derived epithelium (mTECs), and differentiated hematopoietic cells.(eTACs)—it

seems that convergence of several CREs upon a single gene might be more parsimonious than expression of a common set of essential transcription factors which then bind the same CRE(s) in each of these cell types.

As mentioned in previous chapters, there is some circumstantial evidence that *Aire* is controlled by more than one CRE. A transgene containing a total of 37 kb of *Aire*-flanking sequence produces consistent thymic expression but does not seem to be expressed in early embryogenesis, in contrast to transgenes created using the entire 461E7 BAC (77,149). This implies that at least one additional CRE exists, either located further from *Aire* than these 37 kb extend or located within an intron of *Aire*. Similarly, a 6.3-kb upstream fragment was able to support robust expression in mTECs but not, at first glance, expression in eTACs (150). Finally, our own work provides support for the hypothesis that there are additional distal CREs. In particular, we find that a 4.3-kb upstream fragment is able to support specific, intermittently achieved expression in mTECs but not in eTACs (Figures 2.3 and S2.5). We hypothesize that at least one additional element is necessary for induction of *Aire* expression in eTACs and the same or different CRE increases the probability of stable, continuous *Aire* expression.

Our examination of the *Aire*-recapitulating 461E7 region led us right to ACNS1 (Figure 2.1). There was no single other candidate that caught our attention so strongly. No other region showed the same striking enrichment for H3K27ac in mTECs. Many additional CNSs were identified by PhastCons or mVista but there was little to make one stand apart from the others. Moreover, even if some of these CNSs are functional elements, we have no evidence beyond proximity that they regulate *Aire*—and given the number of other genes present near *Aire*, there are several other very plausible targets. As such, we decided that instead of deleting small

97

candidate regions, we would assay the effect of two large deletions, with the plan to follow up with finer deletions if loss of these large regions had an effect.

While investigating the role of ACNS1 in *Aire* regulation, we first evaluated its function in vivo by creating a new transgenic mouse using the *Aire*-recapitulating *Adig* BAC from which we had deleted most of ACNS1. The *AdigΔACNS1* mice did not express the reporter transgene (Figure S2.6), leading us to proceed to germline deletion of endogenous ACNS1. Although this decision was vindicated by the finding that ACNS1$^{-/-}$ mice do not express *Aire* (Figure 2.4), we cannot be certain that the lack of *AdigΔACNS1* expression is due to the missing ACNS1 rather than the site in which the transgene randomly integrated because we were only able to analyze the descendants of a single founder mouse. Indeed, we could not confidently compare the *AdigΔACNS1* mice to the previously generated *Adig* mouse (74) because of one of the central challenges when using transgenic constructs: large variability in expression caused by differences in integration site and copy number. This problem is particularly acute when using small (plasmid-sized) transgenes but is also present when using artificial chromosome-based transgenes (185). In response to this challenge, several strategies have been developed to allow use of large transgenes while avoiding the position-dependent effects observed with random integration of transgenes. These include the use of transchromosomes, in which microcell-mediated chromosome transfer allows the introduction of large, independently segregating chromosome fragments into a target cells; homologous recombination into the ES cell genome (i.e., the conventional targeted knock-in or knockout approach), which is possible even with large pieces of DNA such as those in BACs; Cre-mediated recombination into a site at which *loxP* sites were previously inserted in a targeted fashion; and, similarly, a system for φC31-

mediated site-directed transgenesis using previously introduced *attP* sites in the *Rosa26* or *H11* locus (185,186).

We ultimately decided to use the Cre/*loxP* system to inducibly generate deletions in randomly integrated transgenes, allowing us to measure the effect of removing this region while controlling for copy number and integration site. We generated *Aire-GFP* reporter mice in which approximately 30 kb of sequence upstream or downstream of *Aire* could be removed by Cre-mediated recombination. Our preliminary results suggest that both regions contribute substantially to *Aire* expression.

Results and Discussion:

Generation of AdigFlox mice

In order to control for position- and copy number-dependent effects on expression, we decided to make mice bearing a modifiable reporter BAC built from the *Aire*-recapitulating BAC, RP23-461E7 (73,74). To accomplish this, we made BACs that differed in two important ways from the previously validated *Adig* (*Aire-GFP*) reporter mice (74): they had a single *FRT* site, instead of a single *loxP* site, just downstream from the SV40 polyA or the reporter gene cassette, and they had two *loxP* sites flanking a slightly more than 30-kb region of the BAC (Figure 4.1A). In one case, these *loxP* sites were located approximately 4 kb and 37 kb upstream of *Aire*; this transgene was called *AdigFlox61-94* (because the *loxP* sites are located 61 and 94 kb from the end of the BAC insert upstream of *Aire*); ACNS1 is not within the floxed region. In the other case, the *loxP* sites were located approximately 13 kb and 47 kb downstream from the *Aire* start site; this transgene was called *AdigFlox111-145*. Injection of these BACs into fertilized oocytes yielded several transgenic founder mice.

Figure 4.1. Generation of *AdigFlox* mice.

(A) Schematic of the *AdigFlox61-94* and *AdigFlox111-145* BACs, with the *Adig* BAC included for reference. (B) Schematic depicting the breedings used to generate single-copy *AdigFlox* and *AdigΔ* mice. (C) Quantitative PCR amplification curves; CD86 and *Aire* exon 5 were amplified in the same well. Both *AdigFlox* and *AdigFlox, Flp* mice are from the *AdigFlox61-94_D* line. Data are representative of 3 technical replicates each. (D) Flow cytometry histograms, showing GFP expression of Aire$^+$ mTECs, gated as CD45$^-$, EpCAM$^+$, Ly51$^-$, MHC-II^{hi}, Aire$^+$, from lines of WT, *AdigFlox61-94*, and *AdigFlox111-145* mice.

Figure 4.1

(a) 2 endogenous copies => ΔCt 4.0
(b) ΔCt 1.3 => 2 * 2 ^ (4.0 - 1.3) = 13 copies (2 endogenous, 11 transgenic)
(c) ΔCt 3.6 => 2 * 2 ^ (4.0 - 3.5) = 3 copies (2 endogenous, 1 transgenic)
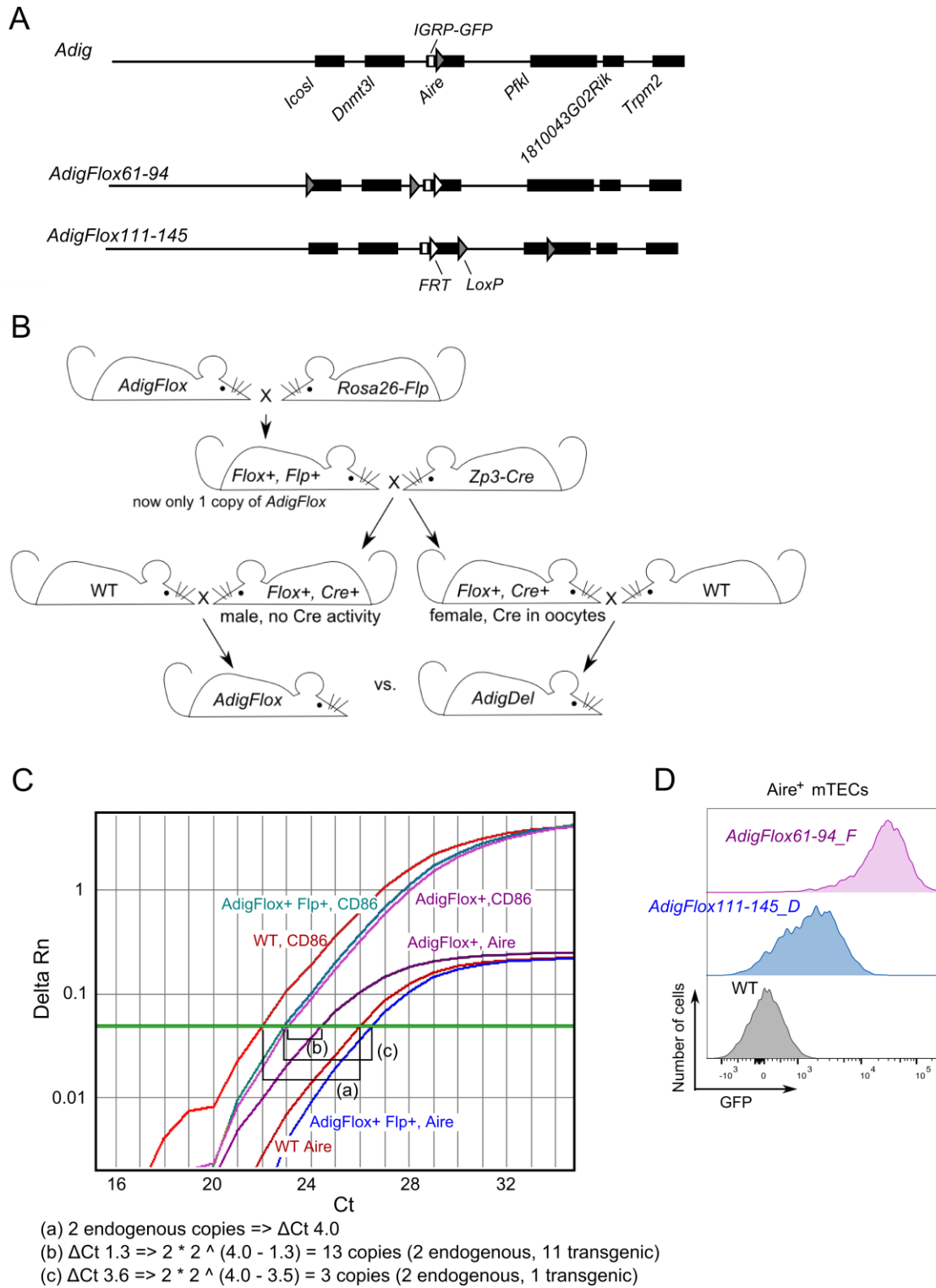
We then proceeded to a series of four generations of breeding to yield mice that differed only by the presence or absence of the floxed region (Figure 3.1B). Founder mice were first crossed to WT B6 mice, and *AdigFlox*[+] offspring were then crossed to B6.*Rosa26-Flp* mice (187). *Rosa26-Flp* mice ubiquitously expressed the Flp recombinase, which caused recombination among the *FRT* sites and reduced the transgene to a single copy (Figure 3.1C). It was important to have only a single copy so that the mice that underwent Cre-mediated recombination would have the same number of copies of the transgene as mice that did not— recombination between the *loxP* sites of different copies of the transgene would result in the mouse having fewer copies of the transgene. The resulting single-copy *AdigFlox* mice were then bred to B6.*Zp3-Cre* mice (188). These mice express Cre in developing oocytes, such that the floxed region was expected to be excised in offspring of female *Zp3-Cre, AdigFlox* mice. In contrast the offspring of male *Zp3-Cre, AdigFlox* mice were expected to still have the floxed portion of the transgene. Multiple founders were generated for each of these two transgenes, but the number of lines under study was reduced (in response to inefficient transmission of the transgene, poor expression of the transgene prior to recombination, and logistical limitations) to three *AdigFlox61-94* lines (lines A, B, and F) and two *AdigFlox111-145* lines (lines B and D). *AdigFlox* lines were checked for GFP expression in Aire[+] cells prior to Flp- and Cre-mediated recombination (Figure 3.1D).

Preliminary evidence that 61-94 and 111-145 regions contribute to Aire expression

Thus far, we have analyzed a limited number of *AdigFlox* and *AdigΔ* mice. A single experiment has been performed on each of the five lines mentioned above. One potential source of error is that each mouse was classified as *AdigFlox* or *AdigΔ* based solely on whether its father or mother was *Zp3-Cre*[+]. After the first couple of experiments, we began performing PCR

assays with goal of confirming that the floxed region had been deleted or remained present, depending on the predicted status. Although in some cases the results aligned neatly with what we expected, this was not always the case. It was not simply a matter of incomplete deletion in offspring of female *Zp3-Cre*[+] mice, although this was sometimes observed. Results that were challenging to unambiguously interpret were observed, including some that seemed to indicate that the 61-94 region was missing from *AdigFlox61-94_F* mice even before they were crossed to *Zp3-Cre*-expressing mice. Despite this fact, we observed differences in GFP expression between *AdigFlox61-94_F* mice with *Zp3-Cre*[+] mothers compared to those with *Zp3-Cre*[+] fathers. At this time, we believe these experiments need a bit of a reset, with new breedings and more rigorous genotyping. In addition, we will focus on a more limited number of lines: *AdigFlox61-94_A*, *AdigFlox111- 145_B*, and *AdigFlox111-145_D*. These are lines that are breeding well and appear by PCR-based interrogation to have intact, floxed 61-94 and 111-145 regions, respectively.

We are very interested in continuing to pursue this line of investigation because of the preliminary results we have produced. These results suggest that the 61-94 region and particularly the 111-145 region have significant enhancer activity on the reporter gene and, by extension, *Aire.* The pooled results from both experiments with *AdigΔ11-145* mTECs showed a substantial decrease in GFP expression relative to those from *AdigFlox111-145* mice, down almost to the same level seen with WT mice, as determined using both quantitation of the frequency of GFP[+] cells among Aire[+] mTECs and median GFP fluorescence intensity (Figure 4.2A and B). The frequency of GFP[+] cells was more a proxy for average levels of GFP expression than a reflection of how the population was divided into two subpopulations; the lower bound of the GFP gate is set based on WT mice, and the amount of GFP signal is not sufficient to prevent overlap between the WT and *AdigFlox* populations, limiting the proportion

Figure 4.2. Large regions near *Aire* contribute to *Aire* expression.

(A) Flow cytometry of Aire$^+$ mTECs, gated as CD45$^-$, EpCAM$^+$, MHC-II$^+$, Ly51$^-$, showing the frequency (%) of GFP$^+$ cells. Bar graph (right) summarizes these results, showing mean ± SD for WT, *AdigFlox111-145,* and *AdigΔ111-145* mice. (B) Flow cytometry of Aire$^+$ mTECs, gated as CD45$^-$, EpCAM$^+$, MHC-II$^+$, Ly51$^-$, showing overlapping histograms of GFP expression. Bar graph (right) summarizes these data by showing mean ± SD of median fluorescence intensity of GFP for WT, *AdigFlox111-145*, and *AdigΔ111-145* mice. (C) The same as (A) except with WT, *AdigFlox61-94*, and *AdigΔ61-94* mice. (D) The same as (B) except with WT, *AdigFlox61-94*, and *AdigΔ61-94* mice. (E) Flow cytometry of eTACs, gated as CD45^{int}, MHC-II^{hi}, EpCAM$^+$, CD86$^-$, showing the frequency (%) of GFP$^+$ cells. Bar graph (right) summarizes these results, showing mean ± SD for WT, *AdigFlox111-145,* and *AdigΔ111-145* mice. (B) Flow cytometry of eTACs, gated as CD45$^-$, EpCAM$^+$, MHC-II$^+$, Ly51$^-$, showing overlapping histograms of GFP expression. Bar graph (right) summarizes these data by showing mean ± SD of median fluorescence intensity of GFP for WT, *AdigFlox111-145*, and *AdigΔ111-145* mice. Data represent pooling of two or three experiments, each involving a different line of transgenic mice, because levels of GFP expression in different lines of a given transgene (*AdigFlox61-94* or *AdigFlox111-145*) were quite similar.

Figure 4.2
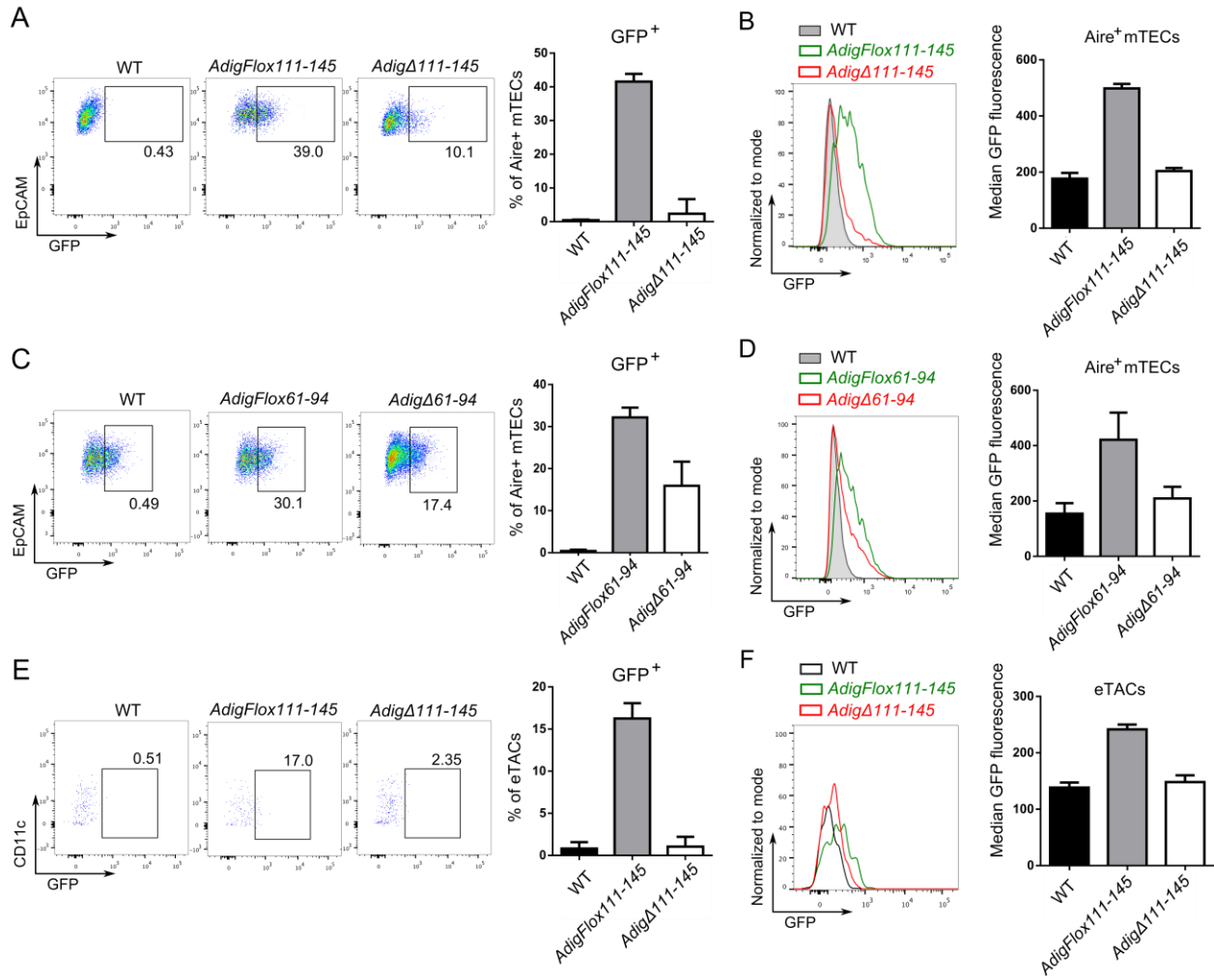
of cells in the GFP gate. *Adig△61-94* mice also show decreased GFP expression, though not to the same degree as seen with loss of the 111-145 region (Figure 4.2C and D). We saw little GFP expression in eTACs of the small number of *AdigFlox61-94* mice that we analyzed, preventing informative comparison to *Adig△61-94*. However, analysis of eTACs from *Adig△111-145* mice, seemed to show a sharp decrease in GFP expression relative to *AdigFlox111-145* mice (Figure 4.2E and F). Together, these results suggest that both the 61-94 and 111-145 regions have CREs that affect *Aire*.

Conclusion:

Previous observations led us to suspect that additional *Aire*-regulating CREs exist beyond ACNS1. In particular, the lack of GFP expression in eTACs and most, but not all, Aire$^+$ mTECs in mice with a small transgene with 4.3 kb of sequence from upstream of *Aire* (Figure 2.3) strongly suggested to us that there are additional DNA elements that act to enhance *Aire* expression. It may be that, just as ACNS1 is essential for thymic and peripheral *Aire* expression, any additional CREs are also play a role in all *Aire*-expressing tissues. Alternatively, there may be CREs that are specific for particular tissues.

We generated mice that allow for analysis of the role of large regions near *Aire* using a BAC reporter transgene while controlling for copy number and position-dependent effects. These mice have *loxP*-flanked 33-kb regions located slightly upstream and slightly downstream of *Aire* (*AdigFlox61-94* and *AdigFlox111-145*, respectively). Preliminary results comparing GFP expression in mice in which there has been Cre-mediated excision of these regions (yielding *Adig△61-94* and *Adig△111-145*) are very intriguing. Deletion of either region appears to lead to decreased reporter expression, with loss of 111-145 having a stronger effect. There are also preliminary data suggesting that the 111-145 region is important for reporter expression in

eTACs. Certain technical difficulties and the small number of analyzed mice prevent us from drawing any firm conclusion from these results, but we do believe that these findings indicate that further study of these mice is warranted.

Materials and Methods:

Mice. B6;FVB.AdigFlox61-94 mice were created by a combination of standard cloning methods and BAC recombineering. We previously removed the *loxP* site from the backbone of the RP23-461E7 BAC. We began by making the *AdigF* BAC using the previously generated JDFNF plasmid (74), which consisted of a 5' Aire homology arm, a beta-globin intron, an IGRP-GFP fusion protein, an SV40 polyA, an FRT site, NeoR, an FRT site, and 3' Aire homology arm. We transformed linearized JDFNF into SW105 bacteria that had been heat-induced and already had the *loxP*-free 461E7 BAC. We selected for recombination with kanamycin, screened by PCR and restriction digest, and then treated with arabinose to induce Flp expression, causing excision of *NeoR*. This yielded the *AdigF* BAC, which was then transformed into EL350 cells, which have arabinose-inducible Cre. The *AdigF* was then modified further by recombineering using constructs that each consisted of a 5' homology arm, a *loxP-NeoR-loxP* cassette, and a 3' homology arm. The homology arms targeted the LNL cassettes to 61 kb, 94 kb, 111 kb, and 145 kb from the end of the 461E7 insert (when considered in the orientation such that Aire is on the (+) strand). After LNL was successfully inserted into the appropriate location in *AdigF*, it was recombined down to just a single *loxP* element. When the second LNL cassette had been and the bacteria were treated with arabinose to induce recombination, careful screening was essential to ensure that recombination had not occurred between either part of this LNL cassette and the more distant previously inserted *loxP*. Once completed, the *AdigFlox61-94* and *AdigFlox111-145*

BACs were isolated using Nucleobond BAC 100 kit (Macherey-Nagel) and injected into fertilized FVB/N oocytes by Helen Lu of the UCSF LARC Transgenic Mouse Core. *Rosa26-Flp* mice (B6.129S4-Gt(ROSA)26Sor^{tm2(FLP*)Sor}/J) were obtained from Jackson Laboratory and previously described (187). B6.*Zp3-Cre* mice (C57BL/6-Tg(Zp3-cre)93Knw/J) were also obtained from Jackson Laboratory and previously described (188). All animals were housed and bred in specific pathogen-free conditions at UCSF. All experiments were approved by the Institutional Animal Care and Use Committee of UCSF.

Flow cytometry. mTECs and eTACs were isolated as previously described (74). Briefly, thymi or peripheral lymph nodes and spleens were isolated, minced, and digested with DNase I and Liberase TM (Roche) before centrifugation on gradient of Percoll PLUS (GE Healthcare). The enriched stromal cells were stained with the antibodies against the surface markers indicated in the figure legends (BioLegend). For intracellular staining, cells were fixed using Foxp3 Staining Buffer Kit (eBioscience) and stained with anti-Aire (clone 5H12, eBioscience). Data were collected using an LSRII flow cytometry (BD) and analyzed using FACSDiva (BD) and FlowJo (TreeStar).

Transgene Copy Number Assays: The number of copies of the *AdigFlox* transgenes was determined using duplex quantitative PCR assays. Each well contained both primers and a FAM-labeled probe specific for *Cd86* (the same used in many Anderson lab genotyping reactions) and primers and a Cy5-labeled probe specific for exon 5 of *Aire*:

qAireEx5_F: 5'-CCATGGCAGCTTCTGTCCAGA

qAireEx5_R: 5'-GCAGCCTGCTTCCATGCATTTAC

qAireEx5_Pr: 5'-Cy5-CCTCGGGTTCCTGGAACATCCCCAGAGG-BHQ2

Aire copy number was determined by normalization to *Cd86* followed by comparison to WT mice, which have 2 copies of *Aire*, by the $\Delta\Delta C_t$ method. Therefore the number of copies of the transgene present was calculated to be two less than the total number of copies of *Aire* as determined using this method.

Chapter 5: Discussion and Concluding Remarks

Overview:

We summarize the results of our investigation into the *cis*-regulation of *Aire,* as presented in the previous chapters. In particular, we focus on our identification of ACNS1, a highly conserved DNA element that is essential for thymic and peripheral *Aire* expression, as well as our unexpected finding that it is also necessary for thymic *Dnmt3l* expression. We discuss promising directions for further research and the implications of our findings on clinical medicine, with a goal of placing them in the larger context of our understanding of the biology of immune tolerance.

Aire's role in the balancing act of immunity:

Infectious disease has been a grave threat throughout most of human history and, indeed, in the history of life on earth. Even today, infection is the cause of approximately 20% of all deaths worldwide and is the leading cause of death of young children (189). Shaped by this strong evolutionary pressure, our immune system is prepared to react vigorously to foreign threats. T cells are an important component of this response. V(D)J recombination, the DNA rearrangement process that produces functional TCR genes (and BCR genes, in B cells), yields an extraordinarily diverse repertoire of receptors that are capable of recognizing a similarly vast array of foreign proteins (11). Inevitably, however, this same immense pool of receptors includes many that are specific for self-antigens (14,15). Most individuals are free of autoimmunity, indicating that these numerous self-reactive T cells are somehow prevented from doing harm. In

a substantial minority of the population, however—approaching 10% of people in some developed countries—this self-tolerance is broken and autoimmune disease develops (190).

The most common autoimmune diseases, such as Graves' disease, type 1 diabetes, systemic lupus erythematosus, and multiple sclerosis, are multifactorial disorders whose precise causes remain obscure. Select monogenic disorder, however, reveal the role of specific components of the elegant machinery that establishes and maintains immune self-tolerance, as well as illustrating the damage that may result from dysregulation of a single critical protein (191). The devastating disease immunodysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) is caused by mutations in *FOXP3,* a transcription factor that is essential for the development of regulatory T cells, and a mutation of the murine orthologue causes the autoimmunity observed in the scurfy mouse (192–194). This autoimmunity emerges from a defect in peripheral tolerance, which refers to control of self-reactive T cells after they have exited the thymus. In contrast, mutations in *AIRE* produce defects in central tolerance, the process by which self-reactive thymocytes are detected and controlled, either through elimination or differentiation into regulatory T cells (17). APS1, the disorder resulting from the absence of *AIRE,* is typified by hypoparathyroidism, adrenocortical insufficiency, and mucocutaneous candidiasis, but may also involve various other organ-specific autoimmune manifestations (Table 1.1 and (29)). In mice and patients with genetic defects in *Aire*, many self-reactive T cells escape from the thymus because mTECs do not express the full complement of TSAs that they normally display (44,47,48,98). eTACs, rare hematopoietic *Aire*-expressing cells located in secondary lymphoid tissues, are also capable of tolerizing self-reactive T cells through expression of a non-redundant set of TSAs (74,75). It is clear that *Aire* expression in mTECs is indispensable because

transplantation of *Aire*[-/-] thymi produces autoimmunity in mice (44). The precise biological role of eTACs remains an important area of investigation.

Identification of an *Aire*-regulating DNA element, ACNS1*:*

In addition to being present in mTECs and eTACs, *Aire* is also transcribed in many ES cell lines, in early embryogenesis through the epiblast stage, and in some developing spermatocytes (76,77). There is also some weak evidence for expression in the ovary (44,63). This highly specific pattern of expression is rather unusual given the anatomical and ontological diversity of these cell types. Previous research effectively characterized the *Aire* promoter in vitro and revealed a role for TNFRSF members RANK and, to a lesser degree, CD40 along with non-canonical NF-κB family members p52 and RelB in *Aire*-expressing mTECs (70,107,109,111,112,117–119,121,122). And yet, the distal *cis*-regulation of *Aire* was wholly unclear. We aimed to address this gap by identifying distal CREs that govern *Aire* expression.

We found a non-coding region, ACNS1, that is well conserved among placental mammals and located approximately 3 kb upstream of *Aire* in mice and humans. ChIP-seq of sorted mTECs revealed enrichment of the enhancer-associated histone mark H3K27ac (132) near this region. Such enrichment was present in neither the D10 T cell line that we also analyzed nor in any of a number adult tissues for which H3K27ac ChIP-seq data were publicly available. This suggested that ACNS1 regulated a gene with a highly specific pattern of expression that included mTECs. Interestingly, the two genes nearest ACNS1, *Aire* and *Dnmt3l,* both matched this description. Still, given our longstanding research focus, our initial assays were focused on investigating whether ACNS1 played a role in the transcriptional regulation of *Aire.*

After some encouraging results with transgenic mice, we successfully deleted endogenous ACNS1 by injection of two guide RNAs and Cas9 mRNA into fertilized C57BL/6

oocytes. Upon crossing to homozygosity, we observed a complete (or very nearly so) loss of *Aire* expression in both mTECs and eTACs. Concomitant with this loss of *Aire* expression, we observed greatly decreased expression of *Aire*-dependent TSAs and the development of spontaneous autoimmunity closely resembling that observed in *Aire*$^{-/-}$ mice. We have not yet determined whether ACNS1 is also required for the expression of *Aire* in early embryogenesis or testes. To address that, we recently began collaborating with the Hebrok lab on performing CRISPR-Cas-mediated deletion of ACNS1 from *Aire*-expressing ES cells, including human ES cell lines. If deletion leads to loss of *Aire* expression, it will both reveal a role for ACNS1 in another *Aire*-expressing cell type and demonstrate that ACNS1 is a functional element in humans.

We were impressed by the ease and efficiency of performing CRISPR-Cas-mediated knockout. This was our first use of the technology, and we were nevertheless able to have C57BL/6 mice heterozygous for the desired deletion within six months of deciding to use this approach, a considerable reduction in the time and effort that would be required if using homologous recombination-based approaches to create knockouts. Indeed, even the cloning required for the CRISPR-Cas approach was significantly easier. Since then, additional attempts to use CRISPR-Cas by members of the Anderson lab have confirmed the increased challenge of knocking in a new piece of DNA using the technology, but we remain convinced that it represents a substantial improvement over previous techniques.

We observed two conserved κB sites within ACNS1—corresponding to two of the three overlapping PhastCons elements—and showed by EMSA that they can bind p52, a non-canonical NF-κB member. We further used an in vitro reporter assays to demonstrate that ACNS1 has NF-κB-responsive enhancer activity. These findings provide the first evidence for a direct connection between non-canonical NF-κB and *Aire* transcription. At this point, the

evidence is still circumstantial: in the absence of non-canonical NF-κB, *Aire*-expressing mTECs do not develop; ACNS1 can mediate NF-κB transactivation; ACNS1 is required for *Aire* expression. Two pieces of evidence that would strengthen this connection would be if ChIP-seq of NF-κB in *Aire*-expressing cells showed binding at ACNS1 and if selective mutation of the κB sites in endogenous ACNS1 led to reduced *Aire* expression.

The specificity of *Aire* expression obviously cannot be explained by NF-κB alone, given how frequently it is active in non-*Aire*-expressing cells. Although some specificity may be encoded in additional, putative CREs, we hypothesize that there may be additional transcription factors that act through ACNS1. One intriguing candidate is AP-2β (*Tcfap2b*). We note that while two of the PhastCons elements aligning with ACNS1 match up with the κB sites, the role of the third, which encompasses a conserved 13-bp sequence, is unclear (Figure 2.2). It seems very plausible that this small sequence is another transcription factor binding site. Preliminary evaluation with the LASAGNA transcription factor binding prediction tool (http://biogrid-head.engr.uconn.edu/lasagna_search and (195)) identified it as a candidate AP-2β site, albeit one less closely matched to the consensus sequence than is the case with the κB sites (data not shown). AP-2β is expressed in the neural crest, peripheral nervous system, facial and limb mesenchyme, various developing epithelia, and trophoectoderm (196), as well as, most importantly, mature mTECs; this last point is supported by both our RNA-seq of *Aire* lineage-tracing mice and the Immunological Genome Project's expression profile of MHC-II^hi mTECs. It would be worthwhile to investigate whether this candidate AP-2β site is functional by using EMSAs and in vitro reporter assays, as performed with the κB sites.

***Aire* and *Dnmt3l*:**

We noticed similarities in the expression patterns of *Aire* and *Dnmt3l*, as well as H3K27ac enrichment near ACNS1 in the Bruce4 ES cell line (Figure 2.1), which expresses *Dnmt3l* but not *Aire* (per ENCODE-associated track on UCSC genome browser; data not shown). This led us to investigate whether *Dnmt3l* expression was altered in ACNS1$^{-/-}$ mice and subsequently discover that thymic *Dnmt3l* expression is lost in these mice. We are currently investigating whether ACNS1 also regulates *Dnmt3l* in germ cells, though the observation that ACNS1$^{-/-}$ are fertile strongly suggests that *Dnmt3l* expression is not wholly abrogated in the germ cells of these mice.

The finding that *Aire* and *Dnmt3l* are co-regulated naturally provokes the question of whether *Dnmt3l* also plays a role in TSA expression. This is not necessarily the case. Perhaps the two genes are functionally independent and their co-regulation is an evolutionary accident in which *Aire* is incidentally expressed in germ cells and early in embryogenesis, where *Dnmt3l* plays an important role, and, conversely, *Dnmt3l* is incidentally expressed in mTECs. Still, the fact that *Dnmt3l* is involved in epigenetic regulation allows us to readily hypothesize about how these two factors might interact. Of interest is recent evidence for a role for Dnmt3L beyond increasing the activity of Dnmt3a and Dnmt3b. In ES cells, Dnmt3L competes with Dnmt3a and Dnmt3b at bivalent promoters, which tend to be associated with genes with important developmental functions (182). Knockdown of *Dnmt3l* results in increased methylation at these promoters and impaired differentiation of these ES cells into primordial germ cells.

We know that both Aire and Dnmt3L bind H3K4me0, which is usually found at repressed genes (85,168,197). It is also known that Aire interacts with Atf7ip and MBD1, which form part of a repressive complex and which are recruited to methylated CpGs, which are

enriched at the promoters of repressed genes (83). One hypothesis is that Aire and Dnmt3L are targeted to the same repressed TSA genes, those marked by H3K4me0 and methylated CpGs. There, Dnmt3L may act to appropriately modulate methylation of these promoters such that a limited number of TSAs come to be expressed in any given cell. Such hypotheses are at this point highly speculative, but further investigation into the thymic function of *Dnmt3l* is certainly merited.

All together, these findings lead us to a new model for the transcriptional regulation of *Aire* (Figure 5.1). In this model, RANK-induced activation of non-canonical NF-κB converges with other transcription factors, which are not expressed in other cells with activated NF-κB, to bind ACNS1 and at least one other CRE. These CREs in turn enhance expression of *Aire*, which then induces expression of a diverse set of TSAs and thereby supports tolerance induction. Simultaneously, *Dnmt3l* expression is induced by interaction between ACNS1 and the *Dnmt3l* promoter, and perhaps Dnmt3L also contributes to TSA expression.

Clinical implications:

Most patients with clinical APS1 have been found to have homozygous or compound heterozygous loss-of-function mutations in the coding portion of *AIRE* (36). In a few cases, disease has been found in heterozygous individuals due to dominant-negative gain-of-function mutations of *AIRE* (53,198). There remain some other individuals with clinical disease characteristic of APS1 in whom no *AIRE* mutations have not been found (Michail Lionakis and Bergithe Oftedal, personal communication). Genetic defects involving ACNS1, including deletion of part or all of the element or insertions or point mutations that interfere with the κB sites, may be the source of disease in some of these patients. We suggest that sequencing,

perhaps copy number variant testing of ACNS1 is warranted in patients with clinical APS1 in whom causative *AIRE* mutations are not apparent.

We also propose that our research may serve as the foundation for developing a method for targeted, antigen-specific tolerance induction. We observed inconsistent but appropriately restricted expression of GFP in the *A4dig* mouse. We note that mTECs effectively enforce negative selection even though each TSA is only expressed in approximately 1-3% of mTECs (99,199). This is the same proportion of mTECs that express GFP in the *A4dig* mouse, leading us to hypothesize that an antigen driven by the *A4dig* regulatory region would be expressed sufficiently to successfully induce tolerance to this antigen. We speculate that the *A4dig* transgene could consequently serve as the basis for a gene therapy agent or other genetic manipulation that would produce mTEC-restricted expression of a new antigen, such as the mismatched HLA in individuals facing an allogeneic transplant, and thereby decrease the risk of developing autoimmunity. Identification of a small transgene capable of producing specific expression in eTACs might also serve as the basis for a tolerance-inducing therapeutic.

Future research:

A number of research directions are possible as follow up to the findings presented here, ranging from inquiry into biochemical mechanism to further exploration into the evolution of the *Aire* locus. We hold that the following experiments address questions of substantial scientific interest and play to the strengths of the Anderson lab.

1. ***Research on the broader roles of ACNS1:*** The preliminary finding that ACNS1 controls both *Aire* and *Dnmt3L* expression in the thymus is an exciting one. So far, we have only used quantitative RT-PCR to document decreased expression of *Dnmt3l.* Anti-Dnmt3L antibodies are commercially available, and an important first step will be to develop

Figure 5.1. Model of *cis*-regulation of thymic *Aire* expression.

This cartoon portrays our model for the transcriptional regulation of *Aire* in mTECs. RANK and CD40 signaling promote nuclear localization of non-canonical NF-κB members p52 and RelB, which bind at ACNS1 together with additional transcription factor(s). ACNS1 then acts as a critical enhancer of *Aire* and *Dnmt3l* activity. Aire induces expression of a diverse set of TSAs, which are processed and displayed on MHC, allowing tolerization of self-reactive thymocytes. Additional CRE(s) likely also contribute to *Aire* expression. Dashed lines are speculative relationships.

Figure 5.1
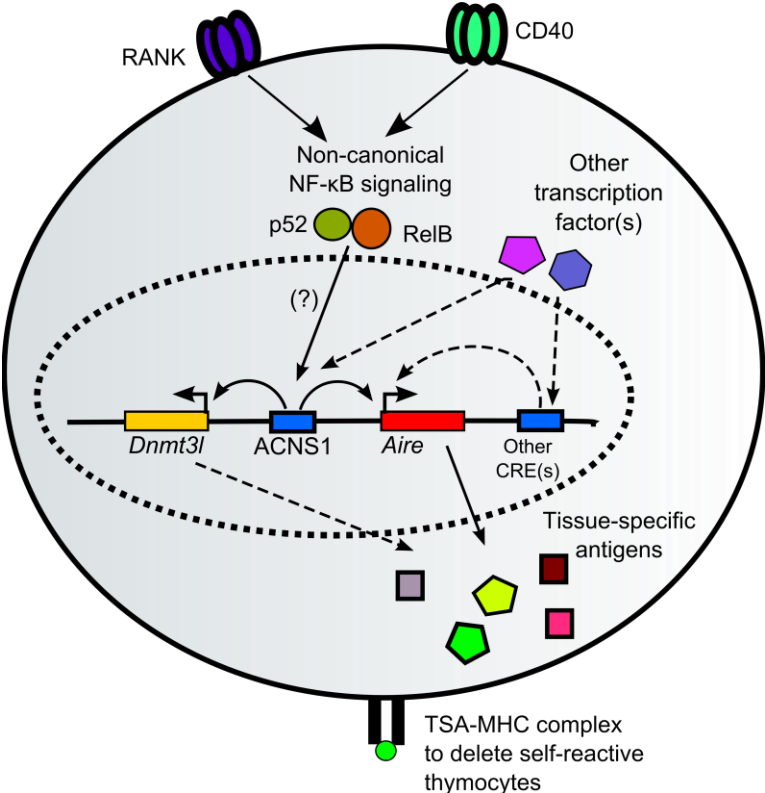
immunofluorescent tissue staining and flow cytometry protocols for assaying Dnmt3L expression in mTECs and elsewhere. From there, three preeminent questions concern whether ACNS1 regulates embryonic and testicular *Aire* expression, whether ACNS1 regulates germ cell *Dnmt3l* expression, and what is the function of thymic *Dnmt3l*. We have recently acquired the *Dnmt3l$^{-/-}$* mouse from (Jackson Labs, previously described (170)). One productive line of inquiry will be to determine whether these *Dnmt3l$^{-/-}$* mice show any defects in central tolerance. Although an autoimmune phenotype has never been reported in them, there may nevertheless be a mild but detectable phenotype that will be apparent upon careful study. Indeed, on the C57BL/6 background even *Aire$^{-/-}$* mice have disease almost so benign that they could escape detection if immunologic phenotyping were not performed (though reduced fertility would probably give them away even without this). As such, assays of these *Dnmt3l$^{-/-}$* mice should include the frequencies of mTEC subsets, levels of *Aire*-dependent and *Aire*-independent TSA expression, and checking for autoimmunity by the techniques used in profiling the ACNS1$^{-/-}$ mouse. In addition, RNA-seq of sorted mTECs from WT, *Aire$^{-/-}$*, ACNS1$^{-/-}$, and *Dnmt3l$^{-/-}$* mice will help determine whether there is a synergistic effect from the loss of both *Aire* and *Dnmt3l* expression and will provide a rich data set from which to generate additional hypotheses.

2. ***Identification of additional Aire CREs:*** One aspect of this aim will be to complete the analysis of the *AdigFlox61-94* and *AdigFlox111-145* mice. As discussed in the previous chapter, preliminary results indicate a role for both regions in the regulation of *Aire* expression, and repeated, more careful analysis of these mice is warranted. Another approach to looking for *Aire*-regulating CREs would be to use CRISPR-Cas-mediated screening in ES cells. That is, *Aire*-expressing ES cells (which also express an *Aire-GFP* reporter) could be targeted with a population of guide RNAs that can produce any of a number of moderately sized deletions,

perhaps 1-3 kb, followed by sorting and deep sequencing of cells showing significantly decreased or increased Aire. This would generate a list of high probability candidate regions that could then be probed more finely to exactly identify the CRE. One limitation to this CRISPR approach is that it can only identify CREs that are important for the expression of *Aire* in ES cells, and CREs active only in mTECs or eTACs may exist. However, a notable advantage of this approach is that one can directly investigate regulation of human *AIRE* by simply using human ES cells.

3. ***The A4dig transgene as an inducer of tolerance:*** As discussed above, we hypothesize that there is adequate expression from the *A4dig* transgene to support negative selection. Since the *A4dig* IGRP-GFP fusion protein contains the cognate antigen to the NOD *8.3* TCR, we propose first crossing *A4dig* mice to these mice and using tetramer analysis to measure whether negative selection is occurring. If robust negative selection is observed, the next step might be to create a lentiviral construct with thymic epithelial cell tropism and to determine whether transduction of the thymus leads to successful negative selection of 8.3^+ thymocytes.

Conclusion:

The immune system must find an appropriate balance between the extremes of autoimmunity and immunodeficiency. *AIRE* plays an important role in this process by significantly decreasing the frequency of self-reactive T cells that could cause disease in the periphery. It does so through an impressive and incompletely understood induction of numerous TSAs, thereby facilitating thymic recapitulation of the antigens of the peripheral self. Concordant with this powerful and unusual function, *Aire* expression is highly restricted. Here, we have presented several new findings on how this pattern of expression is achieved. We identified a conserved DNA element, ACNS1, that is essential for thymic and peripheral *Aire* expression and

the prevention of spontaneous autoimmunity. We further showed that ACNS1 regulates thymic expression of *Dnmt3l,* whose function in the thymus remains an important and open question. We also identified a small transgene that produces specific but variegated expression in *Aire*-expressing mTECs. This led us to hypothesize that additional more distal *Aire*-regulating CREs exist, and we generated new mouse lines that will permit evaluation of large regions flanking *Aire* for regulatory function.

In sum, these findings increase our understanding of the regulation of immune tolerance, suggest new directions for further research, and may eventually result in improvements in the diagnosis and treatment of autoimmune disease. We look forward to seeing where future research leads.

Appendix: Selected Protocols

Stromal Preps:

Purpose: To enrich for mTECs or eTACs for sorting or flow cytometric analysis

1. Sac mice. Remove thymi and/or lymph nodes and/or spleens (lymph nodes are more work to gather but tend to make for easier, less clumpy preps) and put into 6 cm dishes containing 5 mL of digestion medium (DMEM, 2% FCS, 100 µg/ml DNase I (Roche), 0.5 units/ml Liberase TM (Roche)). If pooling for sorting purposes, can put two or maybe three thymi or spleens etc. into one dish. Dice finely with razor blades into pieces small enough to be pulled up into 9" Pasteur pipette (Fisherbrand). Transfer to 15 ml conicals.

2. Incubate at 37 °C water bath, mixing by pipetting up and down (with Pasteur pipette) every 7 to 10 minutes. Incubate spleen or lymph nodes for 45 to 50 minutes and thymi for 25 to 30 minutes – you want there to no longer be visible pieces of tissue.

   If it is critical to maximize your yield (e.g. sorting for ChIP-seq), make the following modifications to steps 1 and 2: harvest into DMEM (not digestion buffer); after transfer to 15 ml conical, centrifuge **briefly** (several seconds) to get tissue to bottom; aspirate most of the DMEM; add 3 ml digestion medium; one-third of the way through incubation, pull most of digestion medium off and add to equal volume of AutoMACS buffer (see step 3) on ice and add 3 mL more of digestion medium; repeat two-thirds of the way through incubation period.

3. After incubation, centrifuge in large centrifuge at 1200 rpm, 4 °C, 5 min. Unless otherwise stated, all spins from here on should be like this. Decant supernatant and resuspend with 5 ml of cold 'AutoMACS' (1 x PBS [CMF = calcium, magnesium free], 0.5% BSA, 2 mM EDTA). Incubate on ice for 10 minutes. Spin down and decant. .

4. Resuspend with 4 ml of 1.115 Percoll (83.47% Percoll Plus (GE Healthcare), 9.27% 10x PBS [CMF], 7.26% 1x PBS [CMF]). Filter through 70 µm cell strainers into FCS-coated 15 ml conicals (add FCS, tip tube around to coat, and pour out FCS). Incline tubes at ~45° and slowly (<100 µl/sec) layered 2 ml 1.065 Percoll (47.16% Percoll Plus, 5.24% 10x PBS [CMF], 47.6% 1x PBS [CMF]) followed by 2 ml 1x PBS [CMF] on top. Centrifuge 2700 rpm, 4 °C, 30 min., acceleration of 5/9, brake of 0/9.

5. Transfer cells at 1.065/PBS interface (aka light fraction, enriched for stromal cells and DCs, vs. dense fraction, enriched for thymocytes) to tube containing 8 ml FACS buffer (or less if you're pooling several samples at this point for sorting, etc.). Filter through 70 µm cell strainer into another 15 ml conical. Spin down. Resuspend with 3.5 ml FACS buffer (if individual mouse) or 10 ml FACS buffer (if pooled sample) and transfer to filter-top FACS tube or through 70 µm cell strainer to 15 ml conical, respectively. Add 100 µl to 900 µl FACS buffer and count using Vi-Cell. Meanwhile, spin down sample. Usually, get ~5-10 million cells per thymus at this point.

6. Resuspend cells with 20 µg/ml 2.4G2 (Fc block, from UCSF core) in FACS buffer, 10 µl per million cells (but a minimum of 30 µl). Incubate on ice for 10 minutes. Add equal volume of FACS buffer containing 1:100 dilutions of antibodies for surface staining (all BioLegend). For thymi: anti-CD45 PerCP (30-F11), anti-Ly51 PE (6C3), anti-CD11c PE-Cy7 (N418), anti-EpCAM APC-Cy7 (G8.8), and anti-I-A^b A647 (if keeping cells alive) or Pac Blue (if fixing) (AF6-120.1); if FVB mice, can use anti-I-A/E PE (M5/115.14.2) and omit Ly51. For spleens/LNs: same staining except anti-CD86 PE (GL-1), or can use anti-CD11c in PE and anti-CD86 in PE-Cy7. Incubate on ice for 30 min. Wash twice with FACS buffer (if not fixing) or once with FACS buffer (if fixing). If

not fixing, resuspend in ~350 µl (one thymus) or ~500 µl (spleen / set of lymph nodes) or a couple ml (large pooled sample for sorting) FACS buffer. Add DAPI (no more than 1 µL) shortly and pass into new filter-top tube shortly before running on flow cytometer.

7. If fixing, resuspend in 500 µl of 1x (i.e. diluted 4x from stock) fixative in FoxP3 staining kit (eBioscience or equivalent Tonbo kit). Incubate on ice for 30 minutes. Historically, we've been unable to successfully stain for Aire in eTACs. Based on Jennifer Lu's very recent work, it appears that the key may be to fix with 1% PFA in overnight. This method of fixation also works for mTECs. Wash twice with 1x perm. buffer.

8. Resuspend with 1:100 anti-Aire e660 (5H12, eBioscience), same volume as surface staining antibody mix (10 µl per millions cells, minimum of 30 µl). Incubate on ice for 30 minutes. Wash twice with perm. buffer. Can hold at this point overnight at 4 °C, covered. Wash once with FACS buffer. Resuspend with ~350 µl (thymus) or ~500 µl (spleen/LN) and run on flow cytometer. Usually gather 100,000 to 300,000 thymic events and 1 to 1.5 million splenic / LN events.

BAC Recombineering and Preparation for Injection:

Purpose: To modify BAC or to add sequence from BAC to plasmid by gap repair, taking advantage of heat-inducible homologous recombination machinery of special bacteria in order to create a large transgene.

1. Inoculate from glycerol stock to get culture of recombineering bacteria (e.g. SW105 or EL350) in LB with chloramphenicol (12.5 µg/ml) at **30 °C**. Inoculate 100 µl of dense culture into two 10 ml aliquots of LB in 14 ml round-bottom tubes. Incubate at 30 °C, 200 rpm, for a few hours (until culture is slightly but definitely cloudy). Swirl one sample in 42 °C water bath for 20 minutes (keep other at 30 °C) and then cool both by swirling in watery ice.

2. Spin down in large, table-top centrifuge at 4,000 rpm, 4 °C, 5 min. Decant. Resuspend with 1 ml of ice-cold milliQ water by swirling then add 5 more ml. Repeat spin / decant / resuspend (note that pellet will be less solid). Repeat spin a third time and decant gently. Resuspend by swirling with the few hundred µl of liquid that remain in the tube.

3. Aliquot 200 µl to 1.5 ml tube on ice containing 20 ng of linearized plasmid. Two steps of linearization, with gel-based purification after first and simply PCR clean-up kit after second, will reduce background. Then transfer 75 µl of bacteria + DNA to electroporation cuvette. Electroporate (1.75 kV, 25 µF, 200 Ω) and as quickly as possible added 1 ml SOC medium and transferred to 14 ml round-bottom tube. Want time constant of 4.8 to 5.0. Incubate at 30 °C, 200 rpm for 1 hour. Spread 200 µL on LB chloramphenicol (12.5 µg/ml) kanamycin (25 µg/ml) plates.

4. Incubate at 30 °C. Colonies should be apparent in 30 to 48 hours. Should get many more colonies on plate using heat-induced bacteria than non-heat-induced. Pick several colonies and inoculate in 3 ml LB chloramphenicol kanamycin. Incubate at 30 °C 200 rpm for 24 -36 hours. Isolate DNA using Qiaprep Mini Prep Kit (Qiagen). Depending on what you're inserting, you can usually perform initial screening of these mini preps by PCR. To ensure you have the intact BAC, you will need to grow up a larger amount and perform restriction digest with a rarely cutting enzyme (XhoI and SpeI often work well with 461E7-based BACs); need to digest at least 1-2 µg to see full set of bands.

5. To excise *NeoR* using arabinose-inducible Cre (EL350) or Flp (SW105), inoculate 3 ml of LB chloramphenicol and when moderately cloudy add 30 µl of 10% arabinose. After 30 minutes at 30 °C, 200 rpm, make several dilutions of sample (e.g., 1,000-fold and 20,000-fold and plate 10 µl + 90 µl LB on LB chloramphenicol plates). Incubate at 30 °C. Check for excision of *NeoR* by verifying kanamycin sensitivity, PCR screening, and restriction digest as needed.

6. To prepare for microinjection, grow up 3 ml mini prep, then dilute 1:2,000 into 125 to 150 ml of LB chloramphenicol. Incubate 30 °C, 200 rpm, overnight. Isolate DNA from cultures using Nucleobond BAC-100 kit (Macherey-Nagel), per its instruction. That is:

    a. Pour culture into 500 ml centrifuge bottles. Centrifuge in large centrifuge (on floor) 5,000 x *g*, 4 °C, 15 min. Decant supernatant. Resuspend with 24 ml buffer S1.

b. Add 24 ml buffer S2; mix by gentle inversion. After about 3 minutes, add 24 ml cooled buffer S3, mix by gentle inversion, and place on ice.

c. Equilibrate column with 6 ml buffer N2. Place Nucleobond filter in funnel, wet the tip with N2, and filter lysate through into column. Do not wait for every bit of lysate to get through (maybe 5 ml left in the funnel). If isolating multiple BACs, clean funnel with bleach and dI water in between samples.

d. Wash column with 2 x 18 ml buffer N3. Elute into angled 50 ml conical by applying 15 ml buffer NS warmed to 50 °C. Transfer eluate to round-bottom centrifuge tube. Add 10.5 ml isopropanol and mix by inversion. Balance tubes. Centrifuge 15,000 x g, 4 °C, 30 min. Decant. Add 5 ml 70% ethanol. Centrifuge 15,000 x g, 4 °C, 15 min. Decant. Let air dry ~ 15 minutes. May carefully use Kimwipe to get some excess ethanol from inside tube.

e. Add 100 µl 10 mM Tris 0.1 mM EDTA to pellet incubate at 4 °C overnight. Quantify and check for purity using Nanovue. For microinjection, Helen prefers to get at least 50 µl of 100 ng/µl.

Anti-histone ChIP-seq of mTECs:

Performed in collaboration with Greg Seumois and others in Anjana Rao's lab at LIAI. Previously developed by Greg Seumois and Vijay Pandurangan, see (145).

FACS:

In brief, isolate thymi from 10 four- to five-week-old mice, mince and digest with Liberase TM and DNase I, wash, centrifuge on Percoll three-layer gradient, wash light fraction, Fc block and stain (anti-Ly51-PE, anti-CD11c-PE-Cy7, anti-EpCAM-APC-Cy7, anti-I-A^b-A647, anti-CD45-PerCP), wash, and sort on BD
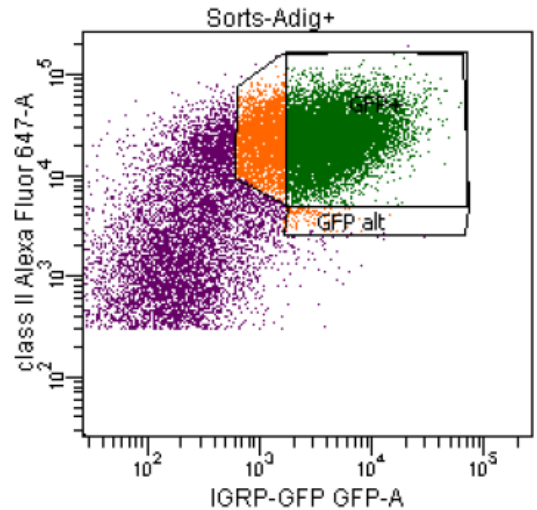
FACSAriaIIIu. Gate on FSC x SSC, DAPI-, EpCAM+, CD45-, MHC-II+, Ly51^{lo}, and using GFP+ and MHC-II (GFP+ in green, GFP- in purple on flow plot; orange show intermediate prevalence of Aire-expressing cells and are discarded). Sort (precision setting) into chilled 1.5 mL low-retention microcentrifuge tubes containing 500 µL of 1:1 DMEM/FCS.

Fixation:

1. After sorting, mix by gentle inversion. Centrifuge at 528 x g in swinging bucket rotor at 4 °C for ten minutes. Aspirate all but ~ 50 µl of supernatant with micropipettor. Centrifuge at 8,000 x g for ~ 10 s and aspirate remaining supernatant with micropipettor.

2. Gently resuspend with DMEM 10% FCS using 100 µL per 10^5 cells. Add 10x fixation buffer (11% formaldehyde, 100 mM NaCl, 1 mM EDTA, 500 µM EGTA, 50 mM HEPES) to 1x, briefly vortex, and incubate at room temperature for 10 minutes, occasionally inverting the tubes. Add 20x glycine buffer (2.5 M glycine) to 1x, briefly vortex and place on ice for 5 minutes.

3. Centrifuge 1000 x g, 4 °C, 4 minutes. Aspirate supernatant with micropipettor. Add 1.5 mL cold (4 °C) PBS (calcium, magnesium free), flick on tube rack several times and invert tube several times. Incubate on ice 2 minutes.

4. Centrifuge 1250 x g, 4 °C, 4 minutes. Aspirate all but about 50 µl supernatant with micropipettor. Gently resuspend with 950 µl cold PBS.

5. Centrifuge 1250 x g, 4 °C, 4 minutes. Aspirate supernatant with micropipettor. Flash freeze by placing tip of tube into liquid nitrogen. Store at -80 °C.

Sonication:

1. Immediately beforehand, make lysis buffer: 1% SDS, 50 mM Tris-HCl pH 8.0, 140 mM NaCl, 10 mM EDTA, 5 mM EGTA, 1x Sigma protease inhibitor, 1 mM PMSF, 20 mM NaBu (protease inhibitors added from frozen stock aliquot thawed immediately before use).

2. Remove cell samples from -80 °C, place on dry ice. Add 65 µl lysis buffer and place at room temperature. After 90 seconds, gently resuspend and place on ice. After approximately 5 minutes, transfer to Axygen low-retention 0.6 mL tubes. Place back on ice.

3. Set up Bioruptor: Fill chamber with ice. After several minutes, remove some ice and add water such that there is a layer of ice on the surface. Set to 'high', 30 sec cycle on, 30 sec cycle off. Place samples at room temperature for ~ 4 minutes to help SDS redissolve as load onto adaptor.

4. Load into sonicator and began sonicating. After fourth on cycle, remove samples, remove some water and replace with ice, very briefly vortex and briefly centrifuge each sample. Samples out of sonicator for 90 to 120 seconds.

5. Repeat step 4 four times, so sample exposed to a total of 20 sonication cycles. After final sonication cycle, can store samples at -80 °C or ship on dry ice. To verify successful

sonication, take 2.5 µl from sample, add 12.5 µl TE buffer, and incubate at 95 °C for 15 minutes. Allow a slow decrease of temperature and run 10 µL on 1.2% agarose gel. Stain the gel with 1:20,000 SYBR Gold (Life Technologies) in TAE for 10 minutes, destain for 15 minutes, then image with UV. For histone ChIP-seq, want fragments of 100-500 bp at this point. Can also quantify chromatin concentration using Picogreen at this point (see later below for protocol).

Chromatin immunoprecipitation:

1. Determine the amount of beads required for the experiment. Will be based on 1 µl antibody : 10 µl beads. Strictly speaking, need to optimize for each antibody. With anti-H3K27ac (Ab4729, lot GR167929-1), we had success with 500 ng = 0.5 µL per sample (each sample was from ~100,000 cells). Add beads on tube, place on magnetic rack, wait 1 minute, and discard supernatant. Add 1 ml of ice-cold RIPA (10 mM Tris-HCl pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% SDS, 0.1 % sodium deoxycholate, 0.1% Triton), vortex briefly, spin briefly, and incubate for 5 minutes at 4 °C. Repeat twice and resuspend beads in the initial volume with RIPA.

2. Split beads into different "antibody" tubes. Dilute 10x with blocking solution (1x PBS, 0.5% BSA) and appropriate amount of antibody. Add beads to blocking buffer then beads to that. Do not vortex. Need to validate the amount of antibody needed for each antibody used. Incubate at 4 °C, rotating, for 4 to 6 hours.

3. Wash beads with 3 x 1 ml ice-cold RIPA, rocking several minutes at 4 °C, and changing tubes each time. Resuspend beads in the initial volume with RIPA and add beads to diluted chromatin. (We used 5 µL per sample for anti-H3K27ac.).

4. Take fragmented chromatin in lysis buffer, dilute to 1 mL with RIPA + 0.25x protease inhibitors (Sigma), 0.25 mM PMSF, and 5 mM sodium butyrate, vortex. Make sure that there is no precipitated SDS. Ideally, your sample would be 500 ng of chromatin. Spin max speed, 4 °C, 15 min. Transfer the supernatant (leaving 10-20 μl in the bottom) to a new pre-chilled low-retention tube and centrifuge again max speed, 4 °C, 15 min. Transfer the supernatant to a new pre-chilled low-retention tube. Take the appropriate amount of diluted chromatin (when small number of cells, as with mTECs, will probably use 900 or 950 μl) and add appropriate amount of bead/antibody complexes (step 3). Store the remaining chromatin at -20 °C for use as input sample. Incubate overnight (~ 10 hrs) at 4 °C, rotating.

5. Washes (perform in cold room or on ice): Pre-chill 1.5 ml low-retention tubes. Transfer ChIP samples to new tubes. Let tubes sit in magnetic rack to collect beads, remove supernatant, add 950 μl RIPA. Invert tubes gently to resuspend beads. Rotate at 4 °C, 5 minutes. Replace tubes in magnetic device to collect beads and remove supernatant. Repeat washes with high salt buffer (50 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate) then with LiCl buffer (50 mM Tris-HCl pH 8, 250 mM LiCl, 1 mM EDTA, 1% NP-40, 0.5% sodium deoxycholate), and then with low salt buffer (10 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM EDTA). Collect beads, resuspend with 1 mL of TE, incubate rotating at 4 °C for several minutes. Wash a new set of low-retention tubes with 200 μl room temperature elution buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS) by putting the buffer in the tubes, vortexing 10 sec, putting in thermomixer for 5 min. at 65 °C, quickly spinning, and discarding buffer.

Transfer the beads + TE suspension into these tubes, place in magnetic rack to capture beads, and remove supernatant.

6. Elution: Resuspend beads in 100 µl of elution buffer by vortexing. Place tubes in thermomixer (or water bath) at 65 °C for 15 min. at 800-1000 rpm. (If in water bath, vortex every 2 minutes.) Place tubes in magnetic rack and collect supernatant into fresh 1.5 ml low-retention tubes. Wash beads with additional 100 µl of 55-65 °C elution buffer, vortex, spin down, collect beads on magnetic rack, and add supernatant to previous 100 µl. For inputs, take 50 µL of input chromatin and add 150 µl elution buffer and 5 µl of 10% SDS.

7. Add 2 µl RNase A (20 mg/ml) to every tube. Mix and incubate in thermomixer or water bath for 30 minutes, 37 °C, 500 rpm. Add 7 µl 300 mM $CaCl_2$ in 10 mM Tris pH 8, followed by 2 µl 20 mg/ml proteinase K. Mix and incubate in water bath or thermomixer for 4 hours at 55 °C (800 rpm). Reverse crosslinks by incubating in a water bath or thermomixer at 65 °C for 4 hours to overnight (800 rpm).

8. DNA purification: Clean up DNA using a Zymo Research Kit. When have numerous samples, use ZR-96 ChIP DNA Clean & Concentrator. Can also use individual column based Zymo ChIP DNA Clean & Concentrator kit with the following modifications: Add 1000 µl DNA binding buffer (so 1200 µl total); load half of sample and spin 15 sec 10,000 rpm, reload flow through and spin 30 sec 10,000 rpm, load remaining sample and spin 15 sec 10,000 rpm, and reload flow through and spin 60 sec 10,000 rpm. Add 300 µl wash buffer to column, spin 30 sec 10,000 rpm, then 300 µl and spin 2 min 10,000 rpm. Add 12 µL elution buffer and spin 15 sec 12,000 rpm then 12 µL 2 min. 12,000 rpm. For long-term storage, screw-top tubes work best.

Quality Control:

1. Picogreen quantification: Use black, flat-bottom plates. Add 4 µl sample to 17 µl TE and aliquot 10 µL to each of two wells. Dilute 40 ng/µl λ DNA 1:4 six times with TE to make standard curve, ending with no DNA, for 8 total dilutions. Resuspend 4 µl Picogreen dye (Life Technologies) in 2 mL TE. Add 190 µL of this to 10 µl DNA, being sure to add to the center of the well not the side. Using a multichannel is okay. Read plate and graph data bi-exponentially and find best-fit curve.

2. Often, before library preparation, it is wise to perform QC qPCRs against some positive and negative control regions. The appropriate regions will depend on the cell type and target of the ChIP.

Whole Genome Amplification:

1. Use the WGA SeqPlex Kit (Sigma, SEQXE). To begin, dilute 1.2 ng of (post-ChIP) DNA in 12 µL of TE. Add 2 µl of Library Preparation Buffer (LP100). Vortex for 10 seconds, spin down, and place in thermoblock for 2 min. 95 °C then hold at 4 °C or on ice. Once cooled down, add 1 µL of Library Preparation Enzyme (E0531), cap, and mix by vortexing. On thermoblock: 16 °C 20 min, 24 °C 20 min, 37 °C 20 min, 75 °C 5 min, 4 °C hold.

2. Determine optimum number of amplification cycles: for each well, 1.5 µl 5x Amplification Buffer (A5112), 0.15 µl DNA Amplification Enzyme (AE), 4.1 µl nuclease-free water, 0.1 µl of 1:1000 dilution of SYBR Green (Life Technologies), 0.15 µL ROX passive dye (Invitrogen), and 1.5 µL template. Run on qPCR machine at 94 °C for 20 min, (94 °C for 15 sec + 70 °C 5 min) x 35; good to follow with a melt curve to check not just getting a bunch of primer dimers. To analyze, determine the Ct value where curve peaks, this Ct + 2 is the

optimal number of amplification cycles (for ease of setting up program when have multiple samples, can also do +1 or +3 as needed).

3. Set up amplifications: in each well, 15 µL 5x Amplification Buffer, 1.5 µL DNA Amplification Enzyme, 37.5 µL nuclease-free water, and 13.5 µL template. Run 94 °C 2 minutes, (94 °C 15 sec + 70 °C 5 min) x X cycles (see step 2), 4 °C hold.

4. DNA Clean-up: If have many samples, use ZR-96 DNA Clean-up Kit (Zymo Research); alternatively, just use individual column DNA Clean and Concentrator Kit (Zymo). Generally follow directions of kit, except be sure to pass sample through column twice (reload flow through after first spin) and when elute, add 25 µl pre-warmed TE to column, let rest for 2 min, spin, and then add another 25 µL pre-warmed TE, let sit, and spin. Quantify DNA with Nanodrop. You should get more than 3 µg total DNA.

5. Primer Removal: Dilute 1 µg DNA to 30.5 µl with water. Add 3.7 µl 10x Primer Removal Buffer and 0.75 µl Primer Removal Solution (SR-400) to each sample. Vortex, quickly spin, Transfer 2.5 µl to QC plate. Add 2.63 µl water, 0.5 µL 10x Primer Removal Buffer (SR401), and 1.88 µL Primer Removal Enzyme to remaining 32.5 µL. Incubate at 37 °C for 60 min, then 65 °C for 20 min, then hold at 4 °C. Transfer 2.5 µl to QC plate. To remaining sample, add 45 µl Ampure XP beads, mix by pipetting, incubate 3 min. room temperature, capture beads with magnet, remove supernatant, wash 3 times with 200 µl 80% ethanol (do not remove plate from magnet), remove as much ethanol as possible, let pellet dry for 10 min., elute with 12 µl of water (vortex, quick spin, back on magnet) and the a second time with warmed 65 °C TE (optional). Measure DNA using Nanodrop and Picogreen.

6. You should get more than 300 ng of DNA. Also, run 5-10 ng on Bioanalyzer to see change in size after primer removal. Should see shift of ~140 bp. For QC PCR, add 197.5 µl TE to the

2.5 µl amplified or primer-removed samples, then dilute another 80-fold. Perform qPCR against primers (as though amplifying again)– want Ct value difference between amplified and primer removed to be > 4.

7. For sequencing, want 150 ng of DNA in 60 µl. We had success using TruSeq Nano Library Prep Kit (Illumina) followed by sequencing on Illumina High-Seq. This was performed by the high-throughput sequencing core at LIAI.

References:

1. Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. Nat Rev Microbiol. 2010 May;8(5):317–27.

2. Janeway C, Murphy K, Travers P, Walport M. Janeway's Immunobiology. 7th ed. New York, NY: Garland Science, Taylor & Francisc Group, LLC; 2008.

3. Ghosh J, Lun CM, Majeske AJ, Sacchi S, Schrankel CS, Smith LC. Invertebrate immune diversity. Dev Comp Immunol. 2011 Sep;35(9):959–74.

4. Loker ES, Adema CM, Zhang S-M, Kepler TB. Invertebrate immune systems--not homogeneous, not simple, not well understood. Immunol Rev. 2004 Apr;198:10–24.

5. Lanier LL, Sun JC. Do the terms innate and adaptive immunity create conceptual barriers? Nat Rev Immunol. 2009 May;9(5):302–3.

6. Kumar H, Kawai T, Akira S. Pathogen recognition by the innate immune system. Int Rev Immunol. 2011 Feb;30(1):16–34.

7. Woo S-R, Corrales L, Gajewski TF. Innate immune recognition of cancer. Annu Rev Immunol. 2015 Mar 21;33:445–74.

8. Sansonetti PJ. The innate signaling of dangers and the dangers of innate signaling. Nat Immunol. 2006 Dec;7(12):1237–42.

9. Heyworth PG, Cross AR, Curnutte JT. Chronic granulomatous disease. Curr Opin Immunol. 2003 Oct;15(5):578–84.

10. Boehm T. Evolution of vertebrate immunity. Curr Biol CB. 2012 Sep 11;22(17):R722–32.

11. Bassing CH, Swat W, Alt FW. The mechanism and regulation of chromosomal V(D)J recombination. Cell. 2002 Apr;109 Suppl:S45–55.

12. Schatz DG, Swanson PC. V(D)J Recombination: Mechanisms of Initiation. Annu Rev Genet. 2011;45(1):167–202.

13. Buckley RH. Molecular defects in human severe combined immunodeficiency and approaches to immune reconstitution. Annu Rev Immunol. 2004;22:625–55.

14. Basten A, Silveira PA. B-cell tolerance: mechanisms and implications. Curr Opin Immunol. 2010 Oct;22(5):566–74.

15. Carpenter AC, Bosselut R. Decision checkpoints in the thymus. Nat Immunol. 2010 Aug;11(8):666–73.

16. Bour-Jordan H, Bluestone JA. Regulating the regulators: costimulatory signals control the homeostasis and function of regulatory T cells. Immunol Rev. 2009 May;229(1):41–66.

17.   Kyewski B, Klein L. A CENTRAL ROLE FOR CENTRAL TOLERANCE. Annu Rev Immunol. 2006 Apr;24(1):571–606.

18.   Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). Nat Rev Immunol. 2014 Jun;14(6):377–91.

19.   Shevach EM, Thornton AM. tTregs, pTregs, and iTregs: similarities and differences. Immunol Rev. 2014 May;259(1):88–102.

20.   Josefowicz SZ, Lu L-F, Rudensky AY. Regulatory T Cells: Mechanisms of Differentiation and Function. Annu Rev Immunol. 2012;30(1):531–64.

21.   Kappler JW, Roehm N, Marrack P. T cell tolerance by clonal elimination in the thymus. Cell. 1987 Apr 24;49(2):273–80.

22.   Kisielow P, Blüthmann H, Staerz UD, Steinmetz M, von Boehmer H. Tolerance in T-cell-receptor transgenic mice involves deletion of nonmature CD4+8+ thymocytes. Nature. 1988 Jun 23;333(6175):742–6.

23.   Jolicoeur C, Hanahan D, Smith KM. T-cell tolerance toward a transgenic beta-cell antigen and transcription of endogenous pancreatic genes in thymus. Proc Natl Acad Sci U S A. 1994 Jul 5;91(14):6707–11.

24.   Sospedra M, Ferrer-Francesch X, Domínguez O, Juan M, Foz-Sala M, Pujol-Borrell R. Transcription of a Broad Range of Self-Antigens in Human Thymus Suggests a Role for Central Mechanisms in Tolerance Toward Peripheral Antigens. J Immunol. 1998 Dec 1;161(11):5918–29.

25.   Derbinski J, Schulte A, Kyewski B, Klein L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. Nat Immunol. 2001 Oct 15;2(11):1032–9.

26.   Talbot NB, Butler AM, Maclachlan EA. THE EFFECT OF TESTOSTERONE AND ALLIED COMPOUNDS ON THE MINERAL, NITROGEN, AND CARBOHYDRATE METABOLISM OF A GIRL WITH ADDISON'S DISEASE. J Clin Invest. 1943 Jul;22(4):583–93.

27.   Leonard MF. Chronic idiopathic hypoparathyroidism with superimposed Addison's disease in a child. J Clin Endocrinol Metab. 1946 Jul;6:493–506.

28.   Ahonen P, Koskimies S, Lokki M, Tiilikainen A, Perheentupa J. The Expression of Autoimmune Polyglandular Disease Type-I Appears Associated with Several Hla-a Antigens but Not with Hla-Dr. J Clin Endocrinol Metab. 1988 Jun;66(6):1152–7.

29.   Perheentupa J. Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy. J Clin Endocrinol Metab. 2006 Aug;91(8):2843–50.

30. Meloni A, Furcas M, Cetani F, Marcocci C, Falorni A, Perniola R, et al. Autoantibodies against type I interferons as an additional diagnostic criterion for autoimmune polyendocrine syndrome type I. J Clin Endocrinol Metab. 2008 Nov;93(11):4389–97.

31. Mathis D, Benoist C. Aire. Annu Rev Immunol. 2009 Apr;27:287–312.

32. Aaltonen J, Bjorses P, Sandkuijl L, Perheentupa J, Peltonen L. An Autosomal Locus Causing Autoimmune-Disease - Autoimmune Polyglandular Disease Type-I Assigned to Chromosome-21. Nat Genet. 1994 Sep;8(1):83–7.

33. Ahonen P, Myllarniemi S, Sipila I, Perheentupa J. Clinical Variation of Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy (apeced) in a Series of 68 Patients. N Engl J Med. 1990 Jun 28;322(26):1829–36.

34. Aaltonen J, Bjorses P, Perheentupa J, Horelli-Kuitunen N, Palotie A, Peltonen L, et al. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. Nat Genet. 1997 Dec;17(4):399–403.

35. Nagamine K, Peterson P, Scott HS, Kudoh J, Minoshima S, Heino M, et al. Positional cloning of the APECED gene. Nat Genet. 1997 Dec;17(4):393–8.

36. Heino M, Peterson P, Kudoh J, Shimizu N, Antonarakis SE, Scott HS, et al. APECED mutations in the autoimmune regulator (AIRE) gene. Hum Mutat. 2001 Sep;18(3):205–11.

37. Faiyaz-Ul-Haque M, Bin-Abbas B, Al-Abdullatif A, Abdullah Abalkhail H, Toulimat M, Al-Gazlan S, et al. Novel and recurrent mutations in the AIRE gene of autoimmune polyendocrinopathy syndrome type 1 (APS1) patients. Clin Genet. 2009 Nov;76(5):431–40.

38. Blechschmidt K, Schweiger M, Wertz K, Poulson R, Christensen HM, Rosenthal A, et al. The mouse AIRE gene: Comparative genomic sequencing, gene organization, and expression. Genome Res. 1999 Feb;9(2):158–66.

39. Mittaz L, Rossier C, Heino M, Peterson P, Krohn KJE, Gos A, et al. Isolation and characterization of the mouse Aire gene. Biochem Biophys Res Commun. 1999 Feb 16;255(2):483–90.

40. Wang C-Y, Shi J-D, Davoodi-Semiromi A, She J-X. Cloning of Aire, the Mouse Homologue of the Autoimmune Regulator (AIRE) Gene Responsible for Autoimmune Polyglandular Syndrome Type 1 (APS1). Genomics. 1999 Feb 1;55(3):322–6.

41. Gibson TJ, Ramu C, Gemünd C, Aasland R. The APECED polyglandular autoimmune syndrome protein, AIRE-1, contains the SAND domain and is probably a transcription factor. Trends Biochem Sci. 1998 Jul 1;23(7):242–4.

42. Rinderle C, Christensen HM, Schweiger S, Lehrach H, Yaspo ML. AIRE encodes a nuclear protein co-localizing with cytoskeletal filaments: altered sub-cellular distribution of mutants lacking the PHD zinc fingers. Hum Mol Genet. 1999 Feb;8(2):277–90.

43. Bjorses P, Pelto-Huikko M, Kaukonen J, Aaltonen J, Peltonen L, Ulmanen I. Localization of the APECED protein in distinct nuclear structures. Hum Mol Genet. 1999 Feb;8(2):259–66.

44. Anderson MS, Venanzi ES, Klein L, Chen Z, Berzins SP, Turley SJ, et al. Projection of an Immunological Self Shadow Within the Thymus by the Aire Protein. Science. 2002 Nov 15;298(5597):1395–401.

45. Ramsey C, Winqvist O, Puhakka L, Halonen M, Moro A, Kämpe O, et al. Aire deficient mice develop multiple features of APECED phenotype and show altered immune response. Hum Mol Genet. 2002 Feb 15;11(4):397–409.

46. Jiang W, Anderson MS, Bronson R, Mathis D, Benoist C. Modifier loci condition autoimmunity provoked by Aire deficiency. J Exp Med. 2005 Sep 19;202(6):805–15.

47. Liston A, Lesage S, Wilson J, Peltonen L, Goodnow CC. Aire regulates negative selection of organ-specific T cells. Nat Immunol. 2003 Apr;4(4):350–4.

48. Anderson MS, Venanzi ES, Chen Z, Berzins SP, Benoist C, Mathis D. The Cellular Mechanism of Aire Control of T Cell Tolerance. Immunity. 2005 Aug;23(2):227–39.

49. DeVoss J, Hou Y, Johannes K, Lu W, Liou GI, Rinn J, et al. Spontaneous autoimmunity prevented by thymic expression of a single self-antigen. J Exp Med. 2006 Nov 27;203(12):2727–35.

50. DeVoss JJ, LeClair NP, Hou Y, Grewal NK, Johannes KP, Lu W, et al. An autoimmune response to odorant binding protein 1a is associated with dry eye in the Aire-deficient mouse. J Immunol Baltim Md 1950. 2010 Apr 15;184(8):4236–46.

51. Gavanescu I, Kessler B, Ploegh H, Benoist C, Mathis D. Loss of Aire-dependent thymic expression of a peripheral tissue antigen renders it a target of autoimmunity. Proc Natl Acad Sci U S A. 2007 Mar 13;104(11):4583–7.

52. Shum AK, DeVoss J, Tan CL, Hou Y, Johannes K, O'Gorman CS, et al. Identification of an autoantigen demonstrates a link between interstitial lung disease and a defect in central tolerance. Sci Transl Med. 2009 Dec 2;1(9):9ra20.

53. Su MA, Giang K, Zumer K, Jiang H, Oven I, Rinn JL, et al. Mechanisms of an autoimmunity syndrome in mice caused by a dominant mutation in Aire. J Clin Invest. 2008 May;118(5):1712–26.

54. Liston A, Gray DHD, Lesage S, Fletcher AL, Wilson J, Webster KE, et al. Gene dosage-limiting role of Aire in thymic expression, clonal deletion, and organ-specific autoimmunity. J Exp Med. 2004 Oct 18;200(8):1015–26.

55. Giraud M, Taubert R, Vandiedonck C, Ke X, Lévi-Strauss M, Pagani F, et al. An IRF8-binding promoter variant and AIRE control CHRNA1 promiscuous expression in thymus. Nature. 2007 Aug 23;448(7156):934–7.

56. Vafiadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, Goodyer CG, et al. Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. Nat Genet. 1997 Mar;15(3):289–92.

57. Aschenbrenner K, D'Cruz LM, Vollmann EH, Hinterberger M, Emmerich J, Swee LK, et al. Selection of Foxp3+ regulatory T cells specific for self antigen expressed and presented by Aire+ medullary thymic epithelial cells. Nat Immunol. 2007 Apr;8(4):351–8.

58. Malchow S, Leventhal DS, Nishi S, Fischer BI, Shen L, Paner GP, et al. Aire-Dependent Thymic Development of Tumor-Associated Regulatory T Cells. Science. 2013 Mar 8;339(6124):1219–24.

59. Klein L, Hinterberger M, von Rohrscheidt J, Aichinger M. Autonomous versus dendritic cell-dependent contributions of medullary thymic epithelial cells to central tolerance. Trends Immunol. 2011 May;32(5):188–93.

60. Perry JSA, Lio C-WJ, Kau AL, Nutsch K, Yang Z, Gordon JI, et al. Distinct contributions of Aire and antigen-presenting-cell subsets to the generation of self-tolerance in the thymus. Immunity. 2014 Sep 18;41(3):414–26.

61. Heino M, Peterson P, Kudoh J, Nagamine K, Lagerstedt A, Ovod V, et al. Autoimmune Regulator Is Expressed in the Cells Regulating Immune Tolerance in Thymus Medulla. Biochem Biophys Res Commun. 1999 Apr 21;257(3):821–5.

62. Klamp T, Sahin U, Kyewski B, Schwendemann J, Dhaene K, Türeci Ã. Expression profiling of autoimmune regulator AIRE mRNA in a comprehensive set of human normal and neoplastic tissues. Immunol Lett. 2006 Aug 15;106(2):172–9.

63. Ruan Q-G, Wang C-Y, Shi J-D, She J-X. Expression and Alternative Splicing of the Mouse Autoimmune Regulator Gene (Aire). J Autoimmun. 1999 Nov;13(3):307–13.

64. Halonen M, Pelto–Huikko M, Eskelin P, Peltonen L, Ulmanen I, Kolmer M. Subcellular Location and Expression Pattern of Autoimmune Regulator (Aire), the Mouse Orthologue for Human Gene Defective in Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy (APECED). J Histochem Cytochem. 2001 Feb 1;49(2):197–208.

65. Adamson KA, Pearce SHS, Lamb JR, Seckl JR, Howie SEM. A comparative study of mRNA and protein expression of the autoimmune regulator gene (Aire) in embryonic and adult murine tissues. J Pathol. 2004 Feb 1;202(2):180–7.

66. Lee J-W, Epardaud M, Sun J, Becker JE, Cheng AC, Yonekura A, et al. Peripheral antigen display by lymph node stroma promotes T cell tolerance to intestinal self. Nat Immunol. 2007 Feb;8(2):181–90.

67. Hubert F-X, Kinkel SA, Webster KE, Cannon P, Crewther PE, Proeitto AI, et al. A specific anti-Aire antibody reveals aire expression is restricted to medullary thymic epithelial cells and not expressed in periphery. J Immunol Baltim Md 1950. 2008 Mar 15;180(6):3824–32.

68. Gäbler J, Arnold J, Kyewski B. Promiscuous gene expression and the developmental dynamics of medullary thymic epithelial cells. Eur J Immunol. 2007 Dec 1;37(12):3363–72.

69. Gray D, Abramson J, Benoist C, Mathis D. Proliferative arrest and rapid turnover of thymic epithelial cells expressing Aire. J Exp Med. 2007 Oct 29;204(11):2521–8.

70. Rossi SW, Kim M-Y, Leibbrandt A, Parnell SM, Jenkinson WE, Glanville SH, et al. RANK signals from CD4+3− inducer cells regulate development of Aire-expressing epithelial cells in the thymic medulla. J Exp Med. 2007 Jun 11;204(6):1267–72.

71. White AJ, Nakamura K, Jenkinson WE, Saini M, Sinclair C, Seddon B, et al. Lymphotoxin Signals from Positively Selected Thymocytes Regulate the Terminal Differentiation of Medullary Thymic Epithelial Cells. J Immunol. 2010 Oct 15;185(8):4769–76.

72. Wang X, Laan M, Kisand K, Scott HS, Peterson P. Post-Aire maturation of thymic medullary epithelial cells involves selective expression of keratinocyte-specific autoantigens. Front Immunol Toler. 2012;3:19.

73. Metzger TC, Khan IS, Gardner JM, Mouchess ML, Johannes KP, Krawisz AK, et al. Lineage tracing and cell ablation identify a post-Aire-expressing thymic epithelial cell population. Cell Rep. 2013 Oct 17;5(1):166–79.

74. Gardner JM, DeVoss JJ, Friedman RS, Wong DJ, Tan YX, Zhou X, et al. Deletional Tolerance Mediated by Extrathymic Aire-Expressing Cells. Science. 2008;321(5890):843–7.

75. Gardner JM, Metzger TC, McMahon EJ, Au-Yeung BB, Krawisz AK, Lu W, et al. Extrathymic Aire-Expressing Cells Are a Distinct Bone Marrow-Derived Population that Induce Functional Inactivation of CD4(+) T Cells. Immunity. 2013 Aug 27;

76. Schaller CE, Wang CL, Beck-Engeser G, Goss L, Scott HS, Anderson MS, et al. Expression of Aire and the Early Wave of Apoptosis in Spermatogenesis. J Immunol. 2008 Feb 1;180(3):1338–43.

77. Nishikawa Y, Hirota F, Yano M, Kitajima H, Miyazaki J, Kawamoto H, et al. Biphasic Aire expression in early embryos and in medullary thymic epithelial cells before end-stage terminal differentiation. J Exp Med. 2010 May 10;207(5):963–71.

78. Perniola R, Musco G. The biophysical and biochemical properties of the autoimmune regulator (AIRE) protein. Biochim Biophys Acta BBA - Mol Basis Dis. 2014 Feb;1842(2):326–37.

79. Sternsdorf T, Jensen K, Reich B, Will H. The nuclear dot protein sp100, characterization of domains necessary for dimerization, subcellular localization, and modification by small ubiquitin-like modifiers. J Biol Chem. 1999 Apr 30;274(18):12555–66.

80. Ferguson BJ, Alexander C, Rossi SW, Liiv I, Rebane A, Worth CL, et al. AIRE's CARD revealed, a new structure for central tolerance provokes transcriptional plasticity. J Biol Chem. 2008 Jan 18;283(3):1723–31.

81. Pitkanen J, Doucas V, Sternsdorf T, Nakajima T, Aratani S, Jensen K, et al. The autoimmune regulator protein has transcriptional transactivating properties and interacts with the common coactivator CREB-binding protein. J Biol Chem. 2000 Jun 2;275(22):16802–9.

82. Saltis M, Criscitiello MF, Ohta Y, Keefe M, Trede NS, Goitsuka R, et al. Evolutionarily conserved and divergent regions of the Autoimmune Regulator (Aire) gene: a comparative analysis. Immunogenetics. 2008 Jan 23;60(2):105–14.

83. Waterfield M, Khan IS, Cortez JT, Fan U, Metzger T, Greer A, et al. The transcriptional regulator Aire coopts the repressive ATF7ip-MBD1 complex for the induction of immunotolerance. Nat Immunol. 2014 Mar;15(3):258–65.

84. Org T, Chignola F, Hetényi C, Gaetani M, Rebane A, Liiv I, et al. The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. EMBO Rep. 2008 Apr;9(4):370–6.

85. Koh AS, Kuo AJ, Park SY, Cheung P, Abramson J, Bua D, et al. Aire employs a histone-binding module to mediate immunological tolerance, linking chromatin regulation with organ-specific autoimmunity. Proc Natl Acad Sci U S A. 2008 Oct 14;105(41):15878–83.

86. Yang S, Bansal K, Lopes J, Benoist C, Mathis D. Aire's plant homeodomain(PHD)-2 is critical for induction of immunological tolerance. Proc Natl Acad Sci U S A. 2013 Jan 29;110(5):1833–8.

87. Kumar PG, Laloraya M, Wang CY, Ruan QG, Davoodi-Semiromi A, Kao KJ, et al. The autoimmune regulator (AIRE) is a DNA-binding protein. J Biol Chem. 2001 Nov 2;276(44):41357–64.

88. Ruan Q-G, Tung K, Eisenman D, Setiady Y, Eckenrode S, Yi B, et al. The autoimmune regulator directly controls the expression of genes critical for thymic epithelial function. J Immunol Baltim Md 1950. 2007 Jun 1;178(11):7173–80.

89. Abramson J, Giraud M, Benoist C, Mathis D. Aire's partners in the molecular control of immunological tolerance. Cell. 2010 Jan 8;140(1):123–35.

90. Oven I, Brdicková N, Kohoutek J, Vaupotic T, Narat M, Peterlin BM. AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. Mol Cell Biol. 2007 Dec;27(24):8815–23.

91. Giraud M, Yoshida H, Abramson J, Rahl PB, Young RA, Mathis D, et al. Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. Proc Natl Acad Sci [Internet]. 2011 Dec 27 [cited 2012 Jan 9]; Available from: http://www.pnas.org/content/early/2011/12/23/1119351109.abstract

92. Giraud M, Jmari N, Du L, Carallis F, Nieland TJF, Perez-Campo FM, et al. An RNAi screen for Aire cofactors reveals a role for Hnrnpl in polymerase release and Aire-activated ectopic transcription. Proc Natl Acad Sci U S A. 2014 Jan 28;111(4):1491–6.

93. Ucar O, Tykocinski L-O, Dooley J, Liston A, Kyewski B. An evolutionarily conserved mutual interdependence between Aire and microRNAs in promiscuous gene expression. Eur J Immunol. 2013 Jul;43(7):1769–78.

94. Liiv I, Rebane A, Org T, Saare M, Maslovskaja J, Kisand K, et al. DNA-PK contributes to the phosphorylation of AIRE: importance in transcriptional activity. Biochim Biophys Acta. 2008 Jan;1783(1):74–83.

95. Saare M, Rebane A, Rajashekar B, Vilo J, Peterson P. Autoimmune regulator is acetylated by transcription coactivator CBP/p300. Exp Cell Res. 2012 Aug 15;318(14):1767–78.

96. Chuprin A, Avin A, Goldfarb Y, Herzig Y, Levi B, Jacob A, et al. The deacetylase Sirt1 is an essential regulator of Aire-mediated induction of central immunological tolerance. Nat Immunol. 2015 May 25;

97. Guerau-de-Arellano M, Mathis D, Benoist C. Transcriptional impact of Aire varies with cell type. Proc Natl Acad Sci U S A. 2008 Sep 16;105(37):14011–6.

98. Derbinski J, Gäbler J, Brors B, Tierling S, Jonnakuty S, Hergenhahn M, et al. Promiscuous Gene Expression in Thymic Epithelial Cells Is Regulated at Multiple Levels. J Exp Med. 2005 Jul 4;202(1):33–45.

99. Derbinski J, Pinto S, Rösch S, Hexel K, Kyewski B. Promiscuous gene expression patterns in single medullary thymic epithelial cells argue for a stochastic mechanism. Proc Natl Acad Sci U S A. 2008 Jan 15;105(2):657–62.

100. Villaseñor J, Besse W, Benoist C, Mathis D. Ectopic expression of peripheral-tissue antigens in the thymic epithelium: probabilistic, monoallelic, misinitiated. Proc Natl Acad Sci U S A. 2008 Oct 14;105(41):15854–9.

101. Pinto S, Michel C, Schmidt-Glenewinkel H, Harder N, Rohr K, Wild S, et al. Overlapping gene coexpression patterns in human medullary thymic epithelial cells generate self-antigen diversity. Proc Natl Acad Sci U S A. 2013 Sep 10;110(37):E3497–505.

102. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet. 2006;7:29–59.

103. Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol. 2015 Mar;16(3):167–77.

104. Colomé N, Collado J, Bech-Serra JJ, Liiv I, Antón LC, Peterson P, et al. Increased apoptosis after autoimmune regulator expression in epithelial cells revealed by a combined quantitative proteomics approach. J Proteome Res. 2010 May 7;9(5):2600–9.

105. Liiv I, Haljasorg U, Kisand K, Maslovskaja J, Laan M, Peterson P. AIRE-induced apoptosis is associated with nuclear translocation of stress sensor protein GAPDH. Biochem Biophys Res Commun. 2012 Jun 22;423(1):32–7.

106. Roberts NA, White AJ, Jenkinson WE, Turchinovich G, Nakamura K, Withers DR, et al. Rank Signaling Links the Development of Invariant γδ T Cell Progenitors and Aire+ Medullary Epithelium. Immunity. 2012 Mar 23;36(3):427–37.

107. Hikosaka Y, Nitta T, Ohigashi I, Yano K, Ishimaru N, Hayashi Y, et al. The Cytokine RANKL Produced by Positively Selected Thymocytes Fosters Medullary Thymic Epithelial Cells that Express Autoimmune Regulator. Immunity. 2008 Sep 19;29(3):438–50.

108. Irla M, Hugues S, Gill J, Nitta T, Hikosaka Y, Williams IR, et al. Autoantigen-Specific Interactions with CD4+ Thymocytes Control Mature Medullary Thymic Epithelial Cell Cellularity. Immunity. 2008 Sep 19;29(3):451–63.

109. Burkly L, Hession C, Ogata L, Reilly C, Marconl LA, Olson D, et al. Expression of relB is required for the development of thymic medulla and dendritic cells. Nature. 1995 Feb 9;373(6514):531–6.

110. Khan IS, Mouchess ML, Zhu M-L, Conley B, Fasano KJ, Hou Y, et al. Enhancement of an anti-tumor immune response by transient blockade of central T cell tolerance. J Exp Med. 2014 Apr 21;

111. Akiyama T, Shimo Y, Yanai H, Qin J, Ohshima D, Maruyama Y, et al. The Tumor Necrosis Factor Family Receptors RANK and CD40 Cooperatively Establish the Thymic Medullary Microenvironment and Self-Tolerance. Immunity. 2008 Sep 19;29(3):423–37.

112. Mouri Y, Yano M, Shinzawa M, Shimo Y, Hirota F, Nishikawa Y, et al. Lymphotoxin Signal Promotes Thymic Organogenesis by Eliciting RANK Expression in the Embryonic Thymic Stroma. J Immunol. 2011 May 1;186(9):5047–57.

113. Chin RK, Lo JC, Kim O, Blink SE, Christiansen PA, Peterson P, et al. Lymphotoxin pathway directs thymic Aire expression. Nat Immunol. 2003 Nov;4(11):1121–7.

114. Venanzi ES, Gray DHD, Benoist C, Mathis D. Lymphotoxin Pathway and Aire Influences on Thymic Medullary Epithelial Cells Are Unconnected. J Immunol. 2007 Nov 1;179(9):5693–700.

115. Vallabhapurapu S, Karin M. Regulation and Function of NF-κB Transcription Factors in the Immune System. Annu Rev Immunol. 2009;27(1):693–733.

116. Sun S-C. Non-canonical NF-|[kappa]|B signaling pathway. Cell Res. 2010 Dec 21;21(1):71–85.

117. Heino M, Peterson P, Sillanpaa N, Guerin S, Wu L, Anderson G, et al. RNA and protein expression of the murine autoimmune regulator gene (Aire) in normal, RelB-deficient and in NOD mouse. Eur J Immunol. 2000 Jul;30(7):1884–93.

118. Zhu M, Chin RK, Christiansen PA, Lo JC, Liu X, Ware C, et al. NF-κB2 is required for the establishment of central tolerance through an Aire-dependent pathway. J Clin Invest. 2006 Nov 1;116(11):2964–71.

119. Akiyama T, Maeda S, Yamane S, Ogino K, Kasai M, Kajiura F, et al. Dependence of Self-Tolerance on TRAF6-Directed Development of Thymic Stroma. Science. 2005 Apr 8;308(5719):248–51.

120. Kajiura F, Sun S, Nomura T, Izumi K, Ueno T, Bando Y, et al. NF-κB-Inducing Kinase Establishes Self-Tolerance in a Thymic Stroma-Dependent Manner. J Immunol. 2004 Feb 15;172(4):2067–75.

121. Murumägi A, Vähämurto P, Peterson P. Characterization of Regulatory Elements and Methylation Pattern of the Autoimmune Regulator (AIRE) Promoter. J Biol Chem. 2003 May 30;278(22):19784–90.

122. Murumägi A, Silvennoinen O, Peterson P. Ets transcription factors regulate AIRE gene promoter. Biochem Biophys Res Commun. 2006 Sep 22;348(2):768–74.

123. Kont V, Murumägi A, Tykocinski L-O, Kinkel SA, Webster KE, Kisand K, et al. DNA methylation signatures of the AIRE promoter in thymic epithelial cells, thymomas and normal tissues. Mol Immunol. 2011 Dec;49(3):518–26.

124. Yano M, Kuroda N, Han H, Meguro-Horike M, Nishikawa Y, Kiyonari H, et al. Aire Controls the Differentiation Program of Thymic Epithelial Cells in the Medulla for the Establishment of Self-Tolerance. J Exp Med. 2008 Nov 24;205(12):2827–38.

125. Lettice LA, Heaney SJH, Purdie LA, Li L, de Beer P, Oostra BA, et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet. 2003 Jul 15;12(14):1725–35.

126. Bulger M, Groudine M. Enhancers: the abundance and function of regulatory sequences beyond promoters. Dev Biol. 2010 Mar 15;339(2):250–7.

127. Maston GA, Landt SG, Snyder M, Green MR. Characterization of Enhancer Function from Genome-Wide Analyses. Annu Rev Genomics Hum Genet. 2012;13(1):null.

128. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, et al. In vivo enhancer analysis of human conserved non-coding sequences. Nature. 2006 Nov 23;444(7118):499–502.

129. Noonan JP, McCallion AS. Genomics of Long-Range Regulatory Elements. Annu Rev Genomics Hum Genet. 2010;11(1):1–23.

130. Denas O, Sandstrom R, Cheng Y, Beal K, Herrero J, Hardison RC, et al. Genome-wide comparative analysis reveals human-mouse regulatory landscape and evolution. BMC Genomics. 2015 Feb 14;16(1):87.

131. Campos EI, Reinberg D. Histones: Annotating Chromatin. Annu Rev Genet. 2009;43(1):559–99.

132. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci. 2010 Dec 14;107(50):21931–6.

133. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature. 2009 Feb 12;457(7231):854–8.

134. Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010 May 13;465(7295):182–7.

135. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013 Dec;10(12):1213–8.

136. Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser AG, Catanese JJ, et al. Bacterial Artificial Chromosome Libraries for Mouse Sequencing and Functional Analysis. Genome Res. 2000 Jan 1;10(1):116–28.

137. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W273–9.

138. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, NISC Comparative Sequencing Program, et al. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 2003 Apr;13(4):721–31.

139. Prabhakar S, Poulin F, Shoukry M, Afzal V, Rubin EM, Couronne O, et al. Close sequence comparisons are sufficient to identify human cis-regulatory elements. Genome Res. 2006 Jul;16(7):855–63.

140. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005 Aug;15(8):1034–50.

141. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. A map of the cis-regulatory sequences in the mouse genome. Nature [Internet]. 2012 Jul 1 [cited 2012 Aug 7]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/22763441

142. Consortium TEP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57–74.

143. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014 Mar 27;507(7493):455–61.

144. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015 Feb 19;518(7539):317–30.

145. Seumois G, Chavez L, Gerasimova A, Lienhard M, Omran N, Kalinke L, et al. Epigenomic analysis of primary human T cells reveals enhancers associated with TH2 memory cell differentiation and asthma susceptibility. Nat Immunol. 2014 Aug;15(8):777–88.

146. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet. 2008 Jul;40(7):897–903.

147. Pekowska A, Benoukraf T, Ferrier P, Spicuglia S. A unique H3K4me2 profile marks tissue-specific gene regulation. Genome Res. 2010 Nov 1;20(11):1493–502.

148. Ronai D, Berru M, Shulman MJ. Variegated expression of the endogenous immunoglobulin heavy-chain gene in the absence of the intronic locus control region. Mol Cell Biol. 1999 Oct;19(10):7031–40.

149. Fan Y, Rudert WA, Grupillo M, He J, Sisino G, Trucco M. Thymus-specific deletion of insulin induces autoimmune diabetes. EMBO J. 2009 Aug 13;28(18):2812–24.

150. Yamaguchi Y, Takayanagi A, Chen J, Sakai K, Kudoh J, Shimizu N. Mouse thymic epithelial cell lines expressing "Aire" and peripheral tissue-specific antigens reproduce in vitro negative selection of T cells. Exp Cell Res. 2011 Aug 15;317(14):2019–30.

151. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012 Aug 17;337(6096):816–21.

152. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 2013 May 9;153(4):910–8.

153. Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. Cell. 2013 Sep 12;154(6):1370–9.

154. Ran FA, Hsu PD, Lin C-Y, Gootenberg JS, Konermann S, Trevino AE, et al. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell. 2013 Sep 12;154(6):1380–9.

155. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013 Feb 15;339(6121):819–23.

156. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013 Nov;8(11):2281–308.

157. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol. 2013 Sep;31(9):822–6.

158. Taniguchi RT, DeVoss JJ, Moon JJ, Sidney J, Sette A, Jenkins MK, et al. Detection of an Autoreactive T-Cell Population Within the Polyclonal Repertoire That Undergoes Distinct Autoimmune Regulator (Aire)-Mediated Selection. Proc Natl Acad Sci [Internet]. 2012 May 2 [cited 2012 May 7]; Available from: http://www.pnas.org/content/early/2012/05/01/1120607109

159. Verdaguer J, Schmidt D, Amrani A, Anderson B, Averill N, Santamaria P. Spontaneous autoimmune diabetes in monoclonal T cell nonobese diabetic mice. J Exp Med. 1997 Nov 17;186(10):1663–76.

160. Hansson MD, Rzeznicka K, Rosenbäck M, Hansson M, Sirijovski N. PCR-mediated deletion of plasmid DNA. Anal Biochem. 2008 Apr 15;375(2):373–5.

161. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562–78.

162. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc. 2012 Sep;7(9):1728–40.

163. Aapola U, Shibuya K, Scott HS, Ollila J, Vihinen M, Heino M, et al. Isolation and Initial Characterization of a Novel Zinc Finger Gene, DNMT3L, on 21q22.3, Related to the Cytosine-5- Methyltransferase 3 Gene Family. Genomics. 2000 May 1;65(3):293–8.

164. Aapola U, Lyle R, Krohn K, Antonarakis SE, Peterson P. Isolation and initial characterization of the mouse Dnmt3l gene. Cytogenet Cell Genet. 2001;92(1-2):122–6.

165. Shovlin TC, Bourc'his D, La Salle S, O'Doherty A, Trasler JM, Bestor TH, et al. Sex-specific promoters regulate Dnmt3L expression in mouse germ cells. Hum Reprod. 2007 Feb 1;22(2):457–67.

166. Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem. 2004 Jun 25;279(26):27816–23.

167. Chedin F, Lieber MR, Hsieh C-L. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc Natl Acad Sci U S A. 2002 Dec 24;99(26):16916–21.

168. Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007 Aug 9;448(7154):714–7.

169. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999 Oct 29;99(3):247–57.

170. Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L and the establishment of maternal genomic imprints. Science. 2001 Dec 21;294(5551):2536–9.

171. Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Dev Camb Engl. 2002 Apr;129(8):1983–93.

172. Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004 Sep 2;431(7004):96–9.

173. Webster KE, O'Bryan MK, Fletcher S, Crewther PE, Aapola U, Craig J, et al. Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. Proc Natl Acad Sci U S A. 2005 Mar 15;102(11):4068–73.

174. La Salle S, Mertineit C, Taketo T, Moens PB, Bestor TH, Trasler JM. Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells. Dev Biol. 2004 Apr 15;268(2):403–15.

175. O'Doherty AM, Rutledge CE, Sato S, Thakur A, Lees-Murdock DJ, Hata K, et al. DNA methylation plays an important role in promoter choice and protein production at the mouse Dnmt3L locus. Dev Biol. 2011 Aug 15;356(2):411–20.

176. Aapola U, Mäenpää K, Kaipia A, Peterson P. Epigenetic modifications affect Dnmt3L expression. Biochem J. 2004 Jun 15;380(Pt 3):705–13.

177. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, et al. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. 2009;10(11):R130.

178. Lattin JE, Schroder K, Su AI, Walker JR, Zhang J, Wiltshire T, et al. Expression analysis of G Protein-Coupled Receptors in mouse macrophages. Immunome Res. 2008;4:5.

179. Heng TSP, Painter MW. The Immunological Genome Project: networks of gene expression in immune cells. Nat Immunol. 2008 Oct;9(10):1091–4.

180. Shay T, Kang J. Immunological Genome Project and systems immunology. Trends Immunol. 2013 Dec;34(12):602–9.

181. Hubert F-X, Kinkel SA, Crewther PE, Cannon PZF, Webster KE, Link M, et al. Aire-deficient C57BL/6 mice mimicking the common human 13-base pair deletion mutation

present with only a mild autoimmune phenotype. J Immunol Baltim Md 1950. 2009 Mar 15;182(6):3902–18.

182. Neri F, Krepelova A, Incarnato D, Maldotti M, Parlato C, Galvagni F, et al. Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. Cell. 2013 Sep 26;155(1):121–34.

183. Madisen L, Zwingman TA, Sunkin SM, Oh SW, Zariwala HA, Gu H, et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat Neurosci. 2010 Jan;13(1):133–40.

184. Zheng Y, Josefowicz S, Chaudhry A, Peng XP, Forbush K, Rudensky AY. Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate. Nature. 2010 Jan 13;463(7282):808–12.

185. Heaney JD, Bronson SK. Artificial chromosome-based transgenes in the study of genome function. Mamm Genome Off J Int Mamm Genome Soc. 2006 Aug;17(8):791–807.

186. Tasic B, Hippenmeyer S, Wang C, Gamboa M, Zong H, Chen-Tsai Y, et al. Site-specific integrase-mediated transgenesis in mice via pronuclear injection. Proc Natl Acad Sci U S A. 2011 May 10;108(19):7902–7.

187. Raymond CS, Soriano P. High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. PloS One. 2007;2(1):e162.

188. De Vries WN, Binns LT, Fancher KS, Dean J, Moore R, Kemler R, et al. Expression of Cre recombinase in mouse oocytes: a means to study maternal effect genes. Genes N Y N 2000. 2000 Feb;26(2):110–2.

189. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012 Dec 15;380(9859):2095–128.

190. Cooper GS, Bynum MLK, Somers EC. Recent Insights in the Epidemiology of Autoimmune Diseases: Improved Prevalence Estimates and Understanding of Clustering of Diseases. J Autoimmun. 2009;33(3-4):197–207.

191. Cheng MH, Anderson MS. Monogenic Autoimmunity. Annu Rev Immunol. 2012;30(1):393–427.

192. Bennett CL, Christie J, Ramsdell F, Brunkow ME, Ferguson PJ, Whitesell L, et al. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. Nat Genet. 2001 Jan;27(1):20–1.

193. Wildin RS, Ramsdell F, Peake J, Faravelli F, Casanova JL, Buist N, et al. X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. Nat Genet. 2001 Jan;27(1):18–20.
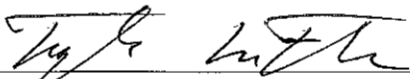
194. Brunkow ME, Jeffery EW, Hjerrild KA, Paeper B, Clark LB, Yasayko SA, et al. Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. Nat Genet. 2001 Jan;27(1):68–73.

195. Lee C, Huang C-H. LASAGNA-Search: an integrated web tool for transcription factor binding site search and visualization. BioTechniques. 2013 Mar;54(3):141–53.

196. Eckert D, Buhl S, Weber S, Jäger R, Schorle H. The AP-2 family of transcription factors. Genome Biol. 2005;6(13):246.

197. Žumer K, Low AK, Jiang H, Saksela K, Peterlin BM. Unmodified histone H3K4 and DNA-dependent protein kinase recruit autoimmune regulator to target genes. Mol Cell Biol. 2012 Apr;32(8):1354–62.

198. Ilmarinen T, Eskelin P, Halonen M, Ruppell T, Kilpikari R, Torres GD, et al. Functional analysis of SAND mutations in AIRE supports dominant inheritance of the G228W mutation. Hum Mutat. 2005 Oct;26(4):322–31.

199. Pinto S, Michel C, Schmidt-Glenewinkel H, Harder N, Rohr K, Wild S, et al. Overlapping gene coexpression patterns in human medullary thymic epithelial cells generate self-antigen diversity. Proc Natl Acad Sci U S A. 2013 Sep 10;110(37):E3497–505.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

_____   6/9/15
Author Signature   Date

153