

THE UNIVERSITY OF CHICAGO

DYNAMIC CONTROL OF TRANSCRIPTION FACTOR NETWORKS GOVERNS
SELF-ANTIGEN DIVERSITY FOR IMMUNE TOLERANCE

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

COMMITTEE ON MOLECULAR METABOLISM AND NUTRITION

BY
JOSHUA AARON MCKEEVER

CHICAGO, ILLINOIS

AUGUST 2026

ABSTRACT

The immune system has evolved to defend against pathogens while maintaining tolerance to self. Yet immune cells are not tailored to recognize specific pathogens. Instead, they possess a broad and continuous spectrum of antigen reactivity. This raises a fundamental question: how does the immune system achieve effective pathogen recognition while avoiding self-reactivity, and what molecular mechanisms tune this balance?

In medullary thymic epithelial cells (mTECs), we demonstrate that the diversity of self-antigens presented to developing thymocytes shapes the self-reactivity of the resulting T cell repertoire. We identify the Vitamin D receptor (VDR) as a key regulator of this process, acting to modulate the amplitude of NF- κ B signaling and, in turn, the heterogeneity of tissue-specific antigen expression programs that underlie central tolerance. Mechanistically, VDR restricts access to higher-affinity NF- κ B binding sites, thereby promoting transcriptional diversity. Loss of VDR disrupts this balance and compromises immunological self-tolerance.

Finally, we propose a transcriptional relay between NF- κ B and AP-1 that establishes tissue-specific gene expression. We propose a model in which NF- κ B activity—particularly RELA—primes the chromatin landscape, enabling subsequent redistribution of AP-1 binding. At composite NF- κ B:AP-1 sites, RELA guides AP-1 to specific genomic loci, after which AP-1 potentially replaces RELA and recruits additional transcriptional machinery to drive tissue-specific gene expression. Consistent with this model, deletion of RELA or the m6A reader YTHDF2—an mRNA decay factor that limits expression of the AP-1 component FOS—impairs AP-1 redistribution and leads to autoimmunity.

Together, these studies reveal how dynamic regulation of transcription factor activity governs both the breadth and variability of gene expression within individual cells, establishing principles that extend across immune tolerance, development, and disease.

DEDICATION

I am deeply grateful to my family for their unwavering support throughout this journey, and I dedicate this work to them.

This dissertation is especially dedicated to the young men in my family who may sometimes doubt their own potential or lack a role model who reminds them of what they are capable of achieving. I hope this accomplishment stands as an example that success is not reserved for a select few—it is built through persistence, resilience, and the courage to keep moving forward. My hope is that when you face uncertainty, you remember that your goals are within reach.

In the words of Grammy Award-winning artist and producer Pharrell Williams: “You can do it too.”

TABLE OF CONTENTS

LIST OF FIGURES	vii
ACKNOWLEDGMENTS	viii
1 INTRODUCTION	1
1.1 Central T Cell Tolerance and the Role of the Thymus	2
1.1.1 Discovery of immune tolerance	2
1.1.2 The role of T cells in immunity	3
1.1.3 How T cell acquire the ability to react to antigens	4
1.1.4 How T cells distinguish between self and non-self	5
1.1.5 Importance of a probabilistic filtration process	6
1.2 Thymic Epithelial Cells as a Source of Self-Antigen Diversity	8
1.3 Molecular Control of TSG Expression: Beyond AIRE	10
1.3.1 The role of Aire in ectopic expression	10
1.3.2 NF- κ B and tissue-specific gene expression	11
1.4 Cellular and Transcriptional Heterogeneity in mTECs	13
1.4.1 Development heterogeneity in mTECs	13
1.4.2 Transcriptional heterogeneity in mTECs	15
1.4.3 Modes of regulation of transcriptional heterogeneity in other cellular contexts	15
1.5 Expanding on signal-dependent transcription factors	18
1.5.1 Nuclear Receptors and integration of external signals	18
1.5.2 Vitamin D Receptor in metabolism and immunity	19
1.5.3 VDR and Autoimmunity	23
1.5.4 Relationship between VDR and NF- κ B	24
1.6 Significance of transcriptional heterogeneity in mTECs	25
1.6.1 Importance of self-antigen diversity in negative selection	25
1.6.2 Venturing into the unknowns of mTEC transcriptional heterogeneity ..	26
1.6.3 Regulation of expression frequency and amplitude	27
1.6.4 Assessing transcriptional heterogeneity	28
1.7 Aims & Significance	29
2 REGULATION OF TRANSCRIPTIONAL HETEROGENEITY IN MEDULLARY THYMIC EPITHELIAL CELLS	31
2.1 Abstract	32
2.2 Introduction	33
2.3 Results	34
2.3.1 TSGs strongly contribute to high transcriptional variance exhibited by Mature mTECs	34
2.3.2 Vitamin D Receptor emerges as a regulator of mTEC transcriptional heterogeneity	35
2.3.3 VDR loss limits range and amplitude of tissue-specific gene expression	37

2.3.4	VDR increases bandwidth of NF- κ B signaling to promote transcriptional heterogeneity	39
2.3.5	VDR restricts access to chromatin regions with high RELA motif density to regulate mTEC transcriptional heterogeneity	40
2.3.6	VDR loss in mTECs causes multi-organ autoimmunity	42
2.4	Discussion	43
2.5	Figures	46
2.6	Supplemental Figures	52
3	ACTIVATION OF PLEIOTROPIC TRANSCRIPTION FACTORS FOR ECTOPIC EXPRESSION	59
3.1	Abstract	59
3.2	Introduction	60
3.3	Results	61
3.3.1	AP-1 and NF- κ B factors exhibit dynamic changes across mTEC development	61
3.3.2	FOS redistributes across mTEC development to tissue-specific gene loci	63
3.3.3	RELA facilitates redistribution of FOS binding to mTEChi	63
3.3.4	YTHDF2 clears Fos transcriptions to facilitate redistribution of FOS ...	65
3.3.5	YTHDF2 deletion causes multi-autoimmunity due to loss of AP-1 associated tissue-specific gene expression	68
3.4	Discussion	69
3.5	Figures	71
4	Materials and Methods	77
5	DISCUSSION & FUTURE DIRECTIONS	90
5.1	Summary	90
5.2	Discussion	93
5.2.1	Regulation of TF networks for a particular fate	93
5.2.2	Reframing central tolerance as a problem of controlled heterogeneity ..	94
5.2.3	Defining an expanded role of NF- κ B in mTECs	95
5.2.4	Transcription factor redistribution as a driver of TSG expression	96
5.3	Future Directions	97
5.3.1	Consequences of modulation of individual TSG expression frequency or amplitude	97
5.3.2	Determining the role of AP-1 in TSG expression	98
5.3.3	Reassessing the weight of thymic VDR function in immune reactivity ..	100
5.4	Conclusion	101
	REFERENCES	103

LIST OF FIGURES

1.1.1	Overlap in self and pathogenic reactivity in individual T cell receptor	7
1.2.1	mTEC-mediated negative selection of autorreactive T cells	9
2.5.1	Tissue-specific genes are robustly expressed by individual cells and contribute to high transcriptional variance in Mature mTECs	46
2.5.2	Nuclear receptor Vitamin D receptor emerges as a potential regulator of transcriptional heterogeneity	47
2.5.3	VDR regulates transcriptional heterogeneity of tissue-specific gene expression	48
2.5.4	VDR controls amplitude and bandwidth of NF- κ B activity	49
2.5.5	VDR restricts access to chromatin regions with high RELA motif density . . .	50
2.5.6	VDR loss leads to multi-organ autoimmunity	51
S1	mTEC scRNA shows defined subsets with distinct changes in transcriptional variance	52
S2	mTECs exhibit broad changes in transcription factor activity during maturation	53
S3	VDR is highly active in mature mTECs and strongly correlates with TSG expression	54
S4	VDR loss collapses transcriptional heterogeneity prioritizing expression of particular TSGs	55
S5	VDR represses genes that are more highly and frequently expressed	56
S6	Defining scATAC clusters and high NF- κ B activity through various stages of mTEC development	57
S7	Effects of thymic Vdr deletion on peripheral T cell populations and mouse weights	58
3.1	Dynamic and inverse activity of NF- κ B and AP-1 factors across mTEC development	71
3.2	FOS redistributes to regions enriched for NF- κ B motifs	72
3.3	RELA-mediated chromatin accessibility changes facilitate FOS redistribution .	73
3.4	Model of AP-1 and NF- κ B activity at TSG loci	74
3.5	YTHDF2 loss compromises <i>Fos</i> clearance and redistribution	75
3.6	YTHDF2 loss compromises TSG expression and lead to multi-organ autoimmunity	76
5.1	<i>Vdr</i> limits the amplitude and increases the range of NF- κ B activity in mature mTECs	91
5.2	AP-1 redistribution mediated by YTHDF2 clearance of Fos transcripts and chromatin landscape changes by RelA	93

ACKNOWLEDGMENTS

Numerous people have contributed to this work by offering mentorship, resources, critical feedback, and friendship. All meaningful work is built on the foundation of what comes before and the support that surrounds it. I feel incredibly fortunate to be part of such a generous and supportive community, whose influence has shaped my professional, academic, and personal growth. I am deeply grateful.

My Advisors and Mentors

I would like to thank my primary mentor, Andrew Koh, PhD, for his unwavering support and encouragement throughout my graduate training. I am especially grateful that he took a chance on me as someone who initially had little background in this field and invested deeply in my development as a scientist. Throughout the challenges of this dissertation, Andrew consistently expressed confidence in my ability to grow and to make meaningful contributions, which had a lasting impact on my work and perspective. I also thank my dissertation committee—Yan Chun Li, PhD, Matthew Brady, PhD, and Amy Weinmann, PhD—for their thoughtful guidance, insight, and continued support throughout this project and my graduate career.

This dissertation was also shaped by the education and support I received from former mentors and professors. I am especially grateful to Hansel Fletcher, PhD for giving me my first experience in a biological science laboratory. He provided the space to learn from my mistakes and to develop an authentic love for research. I am thankful to Elaine Vanterpool, PhD for formally introducing me to the wonders of the immune system and planting seeds of interest in Immunology. I want to express my gratitude to Ricky Bluthenthal, PhD for solidifying my interest in a research career by encouraging me to push myself further—his guidance at a pivotal moment in my academic journey profoundly shaped my trajectory. Finally, I am deeply appreciative of Michael Grillo, PhD, whose mentorship provided me

with the research experience and confidence to pursue a career in science. He took a chance on me, and I was determined to honor that trust.

Institutional and Financial Support

I am grateful for the financial support and resources that made this work possible. This research was supported in part by the NIH NIGMS R35 Diversity Supplement, the NIH T32 Multi-Disciplinary Training Grant in Cancer Research, and the NIH T32 Initiative for Maximizing Student Development (IMSD).

My Colleagues

I am extremely grateful to my colleagues in the Koh lab, who have been unbelievably generous with their time and expertise. In particular, lab alumni Jason Caldwell and Noah Gamble, as well as Busayo Bolonduro and Caroline Kaiser, have contributed greatly to my dissertation work. Their assistance with experiments, critical insights, and genuine encouragement made the challenges of this project far more manageable. I also want to thank the rest of the current and former Koh lab member for the enrichment they have provided over the years.

I am also thankful to the students in the Molecular Metabolism program, who have been a source of friendship, constructive criticism, and inspiration. It has been an honor to be surrounded by such curious and motivated minds, and I have no doubt that they will go on to make meaningful contributions to the world. I would additionally like to thank Matthew Brady, chair of the program, for curating such a wonderful group and for the selfless support he provides to each student in the program.

My Friends and Family

I am deeply thankful to my immediate family—my mom, dad, and brother Jonathan—who have been my unwavering support system throughout my life. Their sacrifices have paved the way for all of my accomplishments.

I am also grateful to my extended family for providing a source of respite and comfort. They have shared countless meals, laughter, and a safe space where I can always be myself. I am especially indebted to my grandparents, Thelma McKeever and Raphael and Vanice Ellis, whose financial and personal sacrifices allowed their children—my parents—to thrive, ultimately creating a foundation that made my own path possible.

I am filled with profound gratitude. If it takes a village to raise a child, it takes a fleet of devoted and supportive people to nurture the development of a scientist.

CHAPTER 1:

INTRODUCTION

A central question in immunology concerns how the immune system achieves protection against a vast array of pathogens while maintaining tolerance to self. Unlike engineered systems designed to recognize predefined targets, T cells are generated with a broad and continuous spectrum of antigen reactivity. This fundamental property raises a key challenge. The immune system must be tuned so that responsiveness is sufficient to mount effective defenses against infection, yet constrained enough to avoid damaging self.

The establishment of self-tolerance is generally thought to occur during T cell development in the thymus, where thymocytes are selected based on their interactions with self-antigens. This process is mediated in part by medullary thymic epithelial cells (mTECs), which express a diverse repertoire of tissue-specific genes (TSGs) and present a representation of peripheral self within a single microenvironment. Through this mechanism, strongly self-reactive T cells are eliminated or diverted, ensuring that the emerging T cell pool is broadly functional yet self-tolerant.

However, self-antigen presentation in mTECs is not uniform. Individual mTECs express only a fraction of the total TSG repertoire, giving rise to substantial heterogeneity in antigen presentation across the population.¹⁻³ This raises a fundamental question: how does a distributed and variable system of self-antigen expression reliably shape a continuous spectrum of T cell reactivity?

Emerging evidence suggests that this heterogeneity is not simply stochastic, but instead reflects tightly regulated and dynamic control of gene expression⁴, which is driven by transcription factors. Yet, how the activity of these factors is quantitatively tuned to control not only the presence but also the diversity and variability of gene expression across cells remains poorly understood. More broadly, the mechanisms by which transcription factor dynamics are translated into heterogeneous yet functionally coherent patterns of self-antigen

expression remain to be defined.

In this work, we investigate how transcription factor activity is regulated to shape the breadth and heterogeneity of tissue-specific gene expression in mTECs. We identify the Vitamin D receptor as a key modulator of nuclear factor kappa B (NF- κ B) signaling amplitude, controlling access to regulatory elements and promoting transcriptional diversity. In addition, we uncover a transcriptional relay between NF- κ B and AP-1 that links chromatin priming to gene activation, providing a mechanistic framework for how dynamic transcription factor interactions establish tissue-specific expression programs. Together, these findings define how quantitative control of transcription factor activity governs antigen diversity in the thymus and, in turn, the calibration of T cell self-reactivity.

1.1 Central T Cell Tolerance and the Role of the Thymus

1.1.1 *Discovery of immune tolerance*

The modern understanding of how the immune system distinguishes self from non-self while retaining the capacity to respond to pathogens emerged gradually from a series of foundational observations in transplantation biology. Early studies demonstrated that foreign tissues could survive in immunoincompetent hosts but were rejected in immunocompetent ones, establishing that immune recognition is both specific and actively regulated.⁵ These findings suggested that the immune system is capable of discriminating between self and foreign, but did not yet explain how this distinction is established.

A major advance came with the work of R. D. Owen, who described naturally occurring chimerism in fraternal cattle twins.⁶ He observed that twin animals shared circulating red blood cells derived from each other, and that these foreign cells persisted throughout life without being rejected. This provided the first evidence that the immune system could tolerate genetically distinct cells if exposure occurred during development, suggesting that the definition of “self” is not fixed but acquired.

Building on this observation, Peter Medawar and colleagues demonstrated experimen-

tally that immune tolerance could be induced.⁷ By introducing donor-derived cells into fetal or neonatal mice, they showed that these animals would later accept tissue grafts from the same donor while rejecting grafts from unrelated individuals. These experiments established several key principles: tolerance is acquired rather than hardwired, it is highly dependent on developmental timing, and it is antigen-specific. Medawar’s work provided the first direct evidence that the immune system is educated during development, laying the foundation for the concept of central tolerance.

Theoretical insight into the underlying mechanism came from Frank Macfarlane Burnet, who proposed the clonal selection theory. Burnet suggested that each lymphocyte expresses a unique antigen receptor and that self-reactive clones are eliminated during development—a process termed clonal deletion.⁸ This framework provided a mechanistic explanation for Medawar’s findings, linking developmental exposure to antigen with the removal of autoreactive cells from the repertoire.

Together, these studies transformed the concept of immune tolerance from a descriptive phenomenon into a developmental and mechanistic process, establishing the foundation for our current understanding of how self-tolerance is generated and maintained.

1.1.2 The role of T cells in immunity

To understand why immune tolerance is essential, it is first necessary to consider the basic organization of the immune system. T cells play a central role in coordinating adaptive immune responses. Early detection of infection is mediated by innate immune cells, including macrophages and dendritic cells, which continuously survey tissues and barrier sites. These cells can engulf pathogens, process them into antigenic peptides, and present these peptides on major histocompatibility complex (MHC) molecules for recognition by T cells.⁹

Among antigen-presenting cells, dendritic cells play a particularly critical role in initiating and amplifying immune responses. After capturing antigens in peripheral tissues, dendritic cells migrate to lymph nodes along chemokine gradients, where they present pro-

cessed antigens to T cells. This interaction serves as a key checkpoint in determining whether an immune response will be activated.

T cells are broadly divided into $CD4^+$ helper T cells and $CD8^+$ cytotoxic T cells, each with distinct effector functions. When a TCR recognizes its cognate antigen with sufficient affinity, the T cell becomes activated, leading to clonal expansion and differentiation into specialized subsets. $CD4^+$ T cells orchestrate immune responses by supporting other immune cell types, while $CD8^+$ T cells directly eliminate infected or aberrant cells.^{10,11}

This system depends on an extraordinary degree of molecular specificity: each T cell expresses a unique TCR generated by somatic recombination, enabling recognition of specific peptide–MHC complexes^{12–14} It is this specificity that enables precise and effective immune responses. However, it also raises a fundamental problem—how does the immune system ensure that T cells respond to foreign pathogens while avoiding recognition of self?

1.1.3 How T cell acquire the ability to react to antigens

Understanding how T cells distinguish between self and non-self requires examining their development. Hematopoietic stem cells reside in the bone marrow, where they give rise to common lymphoid progenitors that enter the circulation and seed the thymus. Thymic entry is guided by chemokine cues, particularly the CCL25–CCR9 axis¹⁵, which directs progenitor homing to the thymic epithelium. Within the thymus, additional chemokine signals such as CCL19 and CCL21 help organize thymocyte migration and positioning as cells initiate the T cell developmental program.¹⁶

T cell development begins in the thymus with $CD4^- CD8^-$ double-negative (DN) progenitors that progress through DN1–DN4 stages. DN1 cells retain multipotent potential, while DN2 cells commit to the T cell lineage under sustained Notch signaling, which suppresses alternative fates and activates the T cell gene program. As cells transition from DN2 to DN3, they initiate V(D)J recombination at the TCR loci, generating the diversity required for antigen recognition.¹⁷

V(D)J recombination is the central mechanism that generates TCR diversity. Through the activity of the RAG1/2 recombinase complex, variable (V), diversity (D), and joining (J) gene segments are somatically rearranged at the TCR β locus, along with the addition and deletion of nucleotides at junctions.¹⁸ This process produces an enormous repertoire of unique TCR sequences, enabling the immune system to recognize a vast array of potential antigens. V(D)J recombination can theoretically generate 10^1 – 10^1 distinct TCR sequences.¹⁹

At the DN3 stage, thymocytes undergo β -selection, a critical checkpoint that tests the success of TCR β -chain recombination. Cells that produce a functional TCR β chain assemble the pre-TCR complex (TCR β paired with pre-T α), which delivers survival and proliferative signals and enforces allelic exclusion to ensure expression of a single β -chain.¹⁷ Cells that fail to generate a productive rearrangement are eliminated.

Cells that pass β -selection progress through DN4, proliferate, and subsequently rearrange the TCR α locus, further expanding receptor diversity.¹⁷ Together, these staged recombination events generate a highly diverse TCR repertoire, forming the molecular basis for antigen specificity in the adaptive immune system.

1.1.4 How T cells distinguish between self and non-self

Because rearrangement of TCRs is a stochastic mechanism generating enormous diversity of TCR clones, inevitably some will have reactivity towards self. the immune system has evolved to address this issue through selection processes.

Thymic selection is classically divided into two main stages. During positive selection, double-positive (CD4⁺CD8⁺) thymocytes are tested for their ability to recognize self-MHC molecules, ensuring that surviving T cells are functionally competent. Thymocytes that fail to bind self-MHC with sufficient affinity do not receive survival signals and undergo apoptosis, a process often referred to as “death by neglect.”²⁰

As a result, the vast majority of developing thymocytes—over 90–95%—are elimi-

nated at this stage.²¹ This selection step is critical because T cells recognize antigens only when they are presented on MHC molecules; without this capability, they would be unable to participate in adaptive immune responses.

In negative selection, thymocytes with strong reactivity toward self-antigens are actively counterselected, providing a critical safeguard against autoimmunity.²² Unlike in the periphery, where strong TCR signaling typically reflects productive immune recognition, in the thymus high-affinity interactions with self peptide–MHC complexes trigger apoptosis. This process eliminates potentially autoreactive T cells before they can enter the circulation.

However, negative selection is not exclusively eliminative. A subset of self-reactive CD4⁺ thymocytes receiving strong but context-dependent signals is instead diverted into the regulatory T cell (Treg) lineage. These cells upregulate FOXP3 and differentiate into a suppressive population that exits the thymus to enforce peripheral tolerance and restrain immune activation.²³

Together with positive selection—which ensures thymocytes are MHC-restricted and functionally competent—these processes shape the developing repertoire into a population that can recognize antigen in the context of self-MHC while remaining tolerant to self-antigens. In combination, they establish a T cell repertoire that is both self-restricted and self-tolerant.

1.1.5 Importance of a probabilistic filtration process

No single receptor–ligand interaction can perfectly encode the distinction between “self” and “non-self”; instead, thymocytes sample a continuum of affinities for self peptide–MHC complexes. As a result, negative selection operates as an inherently probabilistic filtration process that allows the immune system to balance two competing requirements: broad antigen recognition and stringent self-tolerance. Rather than enforcing binary outcomes, thymic selection is governed by graded signaling thresholds, where the strength, duration, and context of TCR engagement collectively influence cell fate in a probabilistic manner.²⁴ In this

framework, selection functions as a statistical filter, enriching for T cells with intermediate affinity that are capable of recognizing self-MHC while excluding cells at the extremes of either non-recognition or excessive self-reactivity.

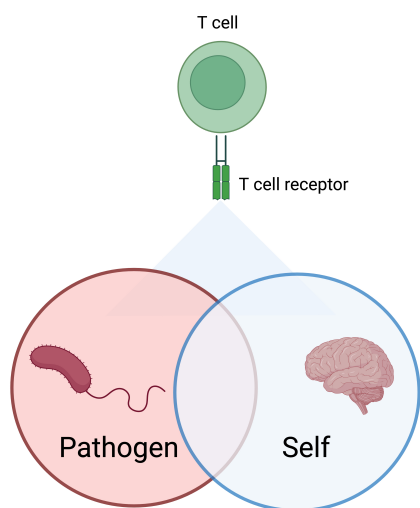


Figure 1.1.1 | Overlap in self and pathogenic reactivity in individual T cell receptor

An important and often underappreciated function of T cells is their role in cancer surveillance.²⁵ Cytotoxic T cells are capable of recognizing and eliminating malignant cells, yet this raises a conceptual paradox: if T cells are selected for self-tolerance, how can they also respond to tumors derived from self tissue? To resolve this, it is important to consider the nature of antigenicity in cancer. Tumor cells frequently generate mutated proteins and neo-antigens that are not present during thymic selection, allowing them to be perceived as foreign and thereby eliciting immune responses that lead to their elimination.²⁶ However, tumor recognition is not limited to neo-antigens. In some cases, cancer cells are targeted through elevated presentation or release of otherwise self-derived antigens. T cells with even modest affinity for these antigens can achieve effective activation through increased antigen density and avidity effects, where cumulative weak interactions across many TCR–pMHC engagements produce strong overall signaling.²⁷

Thus, thymic selection must not only enforce self-tolerance but also preserve a repertoire capable of responding to pathogens, altered self, and aberrant antigen presentation. By

operating as a probabilistic rather than binary checkpoint, it functions as a multi-objective filtering system that balances self-tolerance, immune responsiveness, and tumor surveillance within a single framework. This threshold-based, graded selection process maximizes coverage of potential foreign antigens while minimizing the risk of autoimmunity, while also enabling additional tuning through regulatory T cell differentiation. In this way, immune tolerance emerges not from fixed rules, but from distributed probabilistic constraints acting across a large and diverse population of developing T cells.

1.2 Thymic Epithelial Cells as a Source of Self-Antigen Diversity

A key conceptual point is that thymic selection must test T cells for MHC competence and self-tolerance in a context that is physically segregated from peripheral tissues and pathogen-derived antigens. Developing thymocytes are largely confined to the thymus, an environment composed predominantly of hematopoietic and stromal cells rather than peripheral tissue types. As a result, effective selection requires a specialized system for representing self within the thymic microenvironment. This is achieved through the evolution of thymic antigen-presenting networks that ectopically express and display a broad array of self-derived peptides.

As thymocytes progress from the DN stages into the $CD4^+CD8^+$ double-positive (DP) stage, they enter the cortex and encounter cortical thymic epithelial cells (cTECs), which are the primary mediators of positive selection. At this stage, DP thymocytes are tested for their ability to recognize self-MHC molecules using their newly generated $\alpha\beta$ TCR. cTECs present self peptide-MHC complexes that thymocytes scan through transient TCR interactions, and the outcome is determined by the strength and duration of signaling.

This same signaling event drives both survival and lineage commitment from the DP state to a single-positive (SP) fate. Thymocytes that recognize MHC class I preferentially downregulate CD4 and commit to the $CD8^+$ cytotoxic lineage, while those that recognize MHC class II downregulate CD8 and adopt a $CD4^+$ helper fate.²⁸ This process is orchestrated

by cTECs, which provide the essential self-peptide–MHC landscape required for both MHC restriction and early lineage specification.

Medullary thymic epithelial cells serve as a critical source of self-antigens for negative selection by functionally mirroring the peripheral tissues that developing T cells cannot directly access. Although thymocytes are confined to the thymus, mTECs overcome this limitation through the ectopic expression of a remarkably broad array of tissue-specific genes, encompassing a substantial fraction of the protein-coding genome.²⁹

This promiscuous gene expression enables mTECs to present peptides derived from proteins that are otherwise restricted to diverse peripheral organs, effectively creating a molecular representation of “self” within the thymic medulla. As a result, developing thymocytes can be screened against antigens from across the body, allowing for the deletion of T cells with high affinity for self-antigens during negative selection. In this way, mTECs act as a surrogate for peripheral tissues, ensuring that central tolerance is established against a wide and physiologically relevant repertoire of self-antigens before T cells enter circulation.

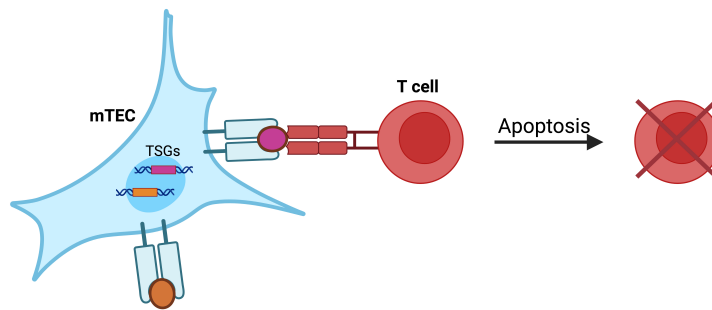


Figure 1.2.1 | mTEC-mediated negative selection of autoreactive T cells

1.3 Molecular Control of TSG Expression: Beyond AIRE

1.3.1 *The role of Aire in ectopic expression*

Regulation of TSG expression in mTECs has remained a central and unresolved question in immunology. A major breakthrough came with the discovery that Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy (APECED; APS-1) is caused by mutations in the human gene *AIRE*.^{30,31} Subsequent work in mice established that Aire is a key regulator of TSG expression, as loss of *Aire* results in reduced thymic expression of peripheral antigens and the development of multi-organ autoimmunity.³² These findings positioned Aire as a central mediator of self-antigen presentation and sparked extensive investigation into its molecular mechanism of action.

Structurally, AIRE contains a Caspase Activation and Recruitment Domain (CARD/HSR), a SAND domain, and two Plant Homeodomain (PHD) fingers.³³ The CARD domain mediates homomultimerization, which is required for Aire’s functional activity.³⁴ The SAND domain contributes to DNA and protein interactions, while the PHD fingers act as epigenetic readers. Notably, PHD1 preferentially binds unmethylated histone H3 lysine 4 (H3K4me0)³⁵, a mark associated with transcriptionally inactive chromatin, and both PHD domains are important for stable chromatin association. These observations led to a model in which AIRE targets epigenetically silent loci and facilitates their transcriptional activation. Importantly, AIRE lacks strong sequence-specific DNA binding, suggesting that its genomic targeting is dictated primarily by chromatin state rather than DNA sequence.

This chromatin-centric model raised the possibility that AIRE might function as a pioneer factor to open inaccessible chromatin. However, chromatin accessibility profiling demonstrated that *Aire*-deficient mTECs largely retain accessible chromatin landscapes, arguing against a primary role in chromatin opening.³⁶ Instead, AIRE appears to act on pre-accessible regulatory regions. Consistent with this, AIRE-bound regions are enriched for transcription factor motifs, suggesting that AIRE cooperates with other transcription factors

to localize to specific genomic loci. Thus, AIRE function is likely constrained not only by chromatin context but also by the availability of collaborating transcription factors.

Mechanistically, accumulating evidence indicates that Aire operates at a late stage of the transcription cycle, particularly at the level of transcriptional elongation. AIRE preferentially associates with genes that are already occupied by RNA polymerase II, and has been shown to promote the release of paused polymerase into productive elongation.³⁷ A key component of this process is AIRE's interaction with the positive transcription elongation factor b (P-TEFb) through its C-terminal domain. P-TEFb phosphorylates RNA polymerase II and negative elongation factors, thereby enabling transcriptional elongation, as well as coordinating downstream processes such as chromatin modification and RNA processing. The functional importance of this interaction is underscored by APS-1-associated mutations that truncate the C-terminus of AIRE, disrupting P-TEFb binding and resulting in severe autoimmunity.³⁸

In addition to P-TEFb, Aire engages a broader network of transcriptional regulators. AIRE can recruit bromodomain-containing proteins such as BRD4, which further enhances P-TEFb activity³⁹, and interacts with topoisomerases to alleviate torsional stress during transcription.⁴⁰ AIRE has also been implicated in modulating higher-order chromatin features, including binding to super-enhancer-like regions and influencing chromatin architecture through displacement of factors such as CTCF.⁴¹ Together, these findings support a model in which AIRE functions as a chromatin-associated scaffold that coordinates multiple components of the transcriptional machinery to enable the expression of otherwise silent genes.

1.3.2 NF- κ B and tissue-specific gene expression

Considering factors beyond AIRE that may regulate TSG expression, binding data indicate that AIRE preferentially localizes to genomic regions enriched for NF- κ B motifs.³⁶ This observation suggests two important points: first, that NF- κ B activity may precede and

facilitate AIRE recruitment; and second, that NF- κ B signaling itself may play a direct role in regulating TSG expression. Supporting the former, the promoter of *Aire* contains NF- κ B binding sites, indicating that NF- κ B activity may act upstream of *Aire* expression.

NF- κ B represents a well-characterized family of transcription factors central to immune signaling and cellular responses to inflammatory stimuli. NF- κ B signaling is broadly divided into two pathways that converge on distinct transcription factor subunits. The canonical pathway, typically associated with acute inflammatory responses, primarily activates RELA, which commonly heterodimerizes with NF- κ B1. In contrast, the non-canonical pathway converges on RELB, which often forms dimers with NF- κ B2.^{42,43}

Within mTECs, NF- κ B signaling has been most extensively studied in the context of cellular development. RELB is essential for the formation of the thymic medulla, and its loss results in impaired mTEC maturation⁴⁴, reduced *Aire* expression⁴⁵, and the development of autoimmunity.⁴⁶ Similarly, disruption of key components of the non-canonical NF- κ B pathway—including Lymphotoxin Beta Receptor(LT β R), TNF Receptor-Associated Factor 6 (TRAF6), Receptor Activator of Nuclear Factor κ B (RANK), and CD40^{47–50}—leads to a marked loss of mature mTECs and a substantial reduction in the size of the mTEC compartment. Together, these findings establish the non-canonical NF- κ B pathway as a critical regulator of mTEC development.

Activation of NF- κ B signaling in mTECs is thought to be driven largely by interactions with developing thymocytes. In particular, RANK ligand (RANKL) expressed by thymocytes engages RANK on immature mTECs, triggering non-canonical NF- κ B signaling and promoting their differentiation into mature mTECs.⁵⁰ Consistent with this model, loss of RELB results in a near-complete absence of the mature mTEC compartment. In contrast, loss of the canonical subunit RELA does not ablate mTECs, although it does reduce their abundance. Notably, RELA deficiency has been associated with decreased expression of tissue-specific genes⁵¹, suggesting that while the non-canonical pathway is indispensable for mTEC development, the canonical pathway may play a more direct role in regulating

TSG expression.

Together, these observations support a model in which distinct NF- κ B pathways contribute differentially to mTEC biology, with non-canonical signaling governing cellular development and canonical signaling contributing to the regulation of TSG expression, potentially in coordination with Aire.

1.4 Cellular and Transcriptional Heterogeneity in mTECs

To date, much of our understanding of mTEC heterogeneity has been framed in terms of discrete developmental states rather than transcriptional diversity per se. In this view, heterogeneity is largely interpreted as arising from progression along developmental trajectories that define the composition of the mTEC compartment.

However, this perspective largely overlooks how transcriptional heterogeneity itself is generated and regulated. While mTECs are known to traverse well-defined developmental stages, comparatively little attention has been given to the extent to which individual cells within a given state vary in their tissue-specific gene expression programs.

We therefore distinguish here between cellular heterogeneity, referring to variation in developmental or phenotypic state across the mTEC population, and transcriptional heterogeneity, referring to variability in gene expression output among cells that may otherwise occupy similar states. This distinction provides a useful framework for considering how regulatory mechanisms might act at multiple levels to shape mTEC function.

In the following sections, we review current models of mTEC developmental heterogeneity and then consider how these frameworks may be extended to account for transcriptional variability within the system.

1.4.1 *Development heterogeneity in mTECs*

mTECs arise from bipotent thymic epithelial progenitors capable of giving rise to both cortical and medullary lineages⁵², with lineage commitment toward the medullary fate driven by

signals such as RANK, CD40, and lymphotoxin receptor engagement. Within the medullary lineage, early immature mTEC_{lo} cells represent a functionally distinct population characterized by relatively low expression of MHC class II and costimulatory molecules, alongside high expression of chemokines such as *Ccl21a*.⁵³ These cells play a critical organizational role in the thymic microenvironment by establishing chemokine gradients that recruit CCR7⁺ positively selected thymocytes from the cortex into the medulla. This thymocyte–epithelial crosstalk is not merely positional but instructive: engagement of RANK on mTEC_{lo} cells by RANKL-expressing thymocytes promotes their differentiation and expansion, initiating the transition toward a mature phenotype.

As mTECs progress to the mTEC_{hi} stage, they undergo a major amplification event which facilitates their drastic phenotypic changes. This transit-amplification is characterized by marked expression of *Top2a* and other cell-cycle indicators like *Mki67*.⁴ As they progress further, they upregulate MHC class II, CD80,⁵⁴ and other antigen presentation machinery, reflecting their increasing specialization in mediating central tolerance. Within this compartment, a key subset is defined by expression of the transcriptional regulator *Aire*. AIRE⁺ mTECs are the principal drivers of promiscuous gene expression, enabling the ectopic transcription of a broad array of tissue-specific antigens.³² This state is transient, however, and lineage tracing studies have demonstrated that AIRE⁺ cells give rise to a series of post-Aire populations.⁵⁵

Post-Aire mTECs exhibit downregulation of *Aire* alongside the acquisition of distinct transcriptional and phenotypic features, including expression of markers associated with terminal differentiation. These cells can further diversify into specialized subsets, such as keratinocyte-like and tuft-like mTECs, which display gene expression programs reminiscent of peripheral epithelial lineages. These "mimetic" cell states are driven by increased activity of lineage defining transcription factors for skin, lung, liver, and intestinal cells like, GRHL, FOXA, FOXJ1, HNF4, SOX8, and SPIB.⁵⁶ The emergence of these terminally differentiated populations suggests that mTEC development is not a linear endpoint but a branching

process that generates additional cellular diversity within the medulla.

Importantly, genetic perturbations can reshape this developmental landscape. Loss of Aire, for example, not only impairs tissue-specific antigen expression but also alters the composition and progression of mTEC states, leading to reduced cellular and transcriptional heterogeneity.⁵⁷ Together, these studies establish a well-defined developmental framework for mTEC differentiation, yet they do not fully explain how extensive diversity in tissue-specific gene expression is generated within and across these states.

1.4.2 Transcriptional heterogeneity in mTECs

The regulatory logic underlying TSG expression has been the subject of considerable debate. Early models proposed that TSG expression in mTECs is largely stochastic, arising from random activation events across the genome.¹ However, the reproducible frequency with which individual TSGs are expressed argues against a purely random process. Instead, accumulating evidence suggests the presence of structured regulation, including the organization of TSGs into co-expression modules.^{3,4} Surprisingly, analyses of regulatory elements have shown that cells expressing a given TSG do not exhibit strong differences in chromatin accessibility within defined "peaks" compared to non-expressing cells.⁵⁸ Rather than being driven by chromatin accessibility alone, TSG activation appears to be influenced by more nuanced local chromatin features, with increased accessibility or instability in the surrounding chromatin correlating with a higher likelihood of gene expression.

1.4.3 Modes of regulation of transcriptional heterogeneity in other cellular contexts

Regulation of gene expression is a layered process involving signaling pathways that converge on transcription factors⁵⁹, endogenous chromatin context⁶⁰, and combinatorial interactions of various transcription factors in networks⁶¹, each providing an opportunity for encoding of regulatory information.

Gene transcription is regulated by transcription factors that bind specific DNA se-

quences and either recruit or interact with the transcriptional machinery to initiate gene expression.⁶² These factors act at regulatory elements such as enhancers, which can increase the probability of transcription by engaging in long-range chromatin interactions that bring distal regulatory regions into proximity with target gene promoters.^{63,64} Differential enhancer usage contributes to both the frequency and magnitude of gene expression, highlighting the importance of regulatory architecture in shaping transcriptional output. Consistent with this, a large fraction of disease-associated variants map to enhancer regions, underscoring their functional relevance.⁶⁵ Importantly, enhancer-mediated regulation is not solely determined by the presence or absence of enhancer–promoter contact; rather, the strength and frequency of these interactions provide an additional layer of control over gene expression.⁶⁶ Differential usage of enhancers is a key determinant of gene expression. Perturbations such as deletion, mutation, or rearrangement of specific enhancers can have profound effects on transcriptional output.⁶⁷ Moreover, individual genes are often regulated by multiple enhancers^{68,69}, introducing an additional combinatorial layer of control over gene expression.

Enhancer availability is dependent on chromatin accessibility, as regulatory elements must reside within open chromatin to permit transcription factor binding and activation.^{70,71} Consistent with this, enhancers comprise a substantial fraction of accessible chromatin regions⁷², indicating that accessibility landscapes reflect the underlying regulatory potential of the genome. Importantly, chromatin accessibility is itself dynamically regulated through mechanisms such as nucleosome positioning⁷³, histone modifications⁷⁴, and the activity of ATP-dependent chromatin remodeling complexes⁷⁵, which together modulate the exposure of regulatory DNA to transcription factors. These processes can vary across cells and developmental states, leading to heterogeneity in the accessibility of enhancer elements. As a result, even in the presence of the same genomic repertoire of enhancers, differences in chromatin accessibility can constrain which regulatory elements are available for activation, thereby influencing both the likelihood and magnitude of gene expression.⁷⁶ In this way, chromatin context acts as a critical determinant of transcriptional heterogeneity by shaping

the set of accessible enhancers that can participate in gene regulation within individual cells.

Another source of transcriptional heterogeneity arises from differences in transcription factor activity across cells. This is not independent of the role of chromatin accessibility and enhancers in gene regulation, as transcription factors both bind to enhancers and actively shape the chromatin landscape by displacing nucleosomes to maintain accessibility at regulatory regions.⁷⁷ In this way, transcription factors can reinforce or stabilize accessible chromatin states, linking TF activity directly to enhancer availability.

Transcription factors also operate within interconnected regulatory networks, where factors can act cooperatively to promote gene expression or antagonistically to repress alternative transcriptional programs.^{78–80} Consequently, variation in the presence, activity, and stoichiometry of transcription factors across cells can lead to differences in which genes are expressed and at what levels. This provides an additional layer of regulatory control that contributes to transcriptional heterogeneity.

A clear example of this is observed in mouse embryonic stem cells, where the transcription factors Klf4 and Zfp281 exert opposing effects on both transcriptional output and chromatin state, driving distinct regulatory programs.⁸¹ Such antagonistic interactions highlight how shifts in transcription factor balance can reconfigure gene regulatory networks and contribute to cell-to-cell variability in gene expression.

Various signaling inputs can converge on a core program that fans into multiple potential outcomes.⁸² This 'bow-tie' architecture is a common source of transcriptional heterogeneity and subsequent functional heterogeneity within and between cell populations. A well studied model of this phenomenon is the NF- κ B pathway. NF- κ B is relevant in various cellular and disease contexts including inflammation, immune responses, development and cancer, yet the gene expression outcomes under variable conditions within and between each process and disease can differ significantly. In unstimulated cells, NF- κ B is stuck in the cytosol in an inactive state due to its association with inhibitors of NF- κ B (I κ Bs). After stimulation, phosphorylation-induced degradation of the I κ Bs by IKK β releases NF- κ B for nuclear

translocation, where it regulates transcription of target genes. NF- κ B activity is further regulated by nuclear phosphorylation and multiple feedback loops in the network, resulting in complex dynamical activity profiles.^{83,84} In primary macrophages, NF- κ B dynamics are shown to be ligand- and dose-specific where different immune stimulatory ligands yield different signaling patterns and differential gene expression.⁸⁵ In Multiple myeloma, sustained NF- κ B activation upregulates proliferative genes, such as *IL6* and cyclin D1 (*CCND1*).⁸⁶

This principle of dynamic, context-dependent signaling extends beyond NF- κ B to other transcriptional regulators, including MSN2⁸⁷, CRZ1⁸⁸, NFAT⁸⁹, P53⁹⁰, all of which exhibit stimulus-specific or frequency-encoded transcriptional responses that shape downstream gene expression programs.

1.5 Expanding on signal-dependent transcription factors

1.5.1 Nuclear Receptors and integration of external signals

Considering transcription factors that directly integrate external signals into gene regulatory programs, Nuclear Receptors (NRs) provide a particularly instructive model. This unique class of ligand-activated transcription factors responds to small, membrane-permeable molecules such as steroid hormones, retinoids, and lipid-derived metabolites, allowing cells to couple environmental or systemic cues to transcriptional outputs.⁹¹

NRs are grouped together based on a conserved modular structure. Most family members share a central DNA-binding domain (DBD) composed of two zinc finger motifs that confer sequence-specific DNA recognition, a flexible hinge region, and a C-terminal ligand-binding domain (LBD) that mediates ligand-dependent conformational changes.^{92,93} The LBD also serves as a platform for co-regulator recruitment, enabling dynamic switching between transcriptional activation and repression depending on ligand availability.⁹⁴ This modular architecture underlies a common mechanistic logic: ligand binding induces structural rearrangements that alter DNA binding stability and cofactor interactions, thereby modulating transcriptional activity.

A direct consequence of this shared structure is the recognition of highly related DNA motifs. Nuclear receptors typically bind to hormone response elements (HREs) composed of direct repeats of a hexameric consensus sequence (often AGGTCA). Specificity among receptors is achieved in part through variation in the spacing between these repeats—commonly referred to as DR (direct repeat) elements, where the number denotes the number of nucleotides separating the repeats (e.g., DR1, DR4).⁹⁵ This seemingly simple variation provides a combinatorial mechanism for differential genomic targeting, even among receptors with highly similar DNA-binding domains.

Despite these structural and mechanistic similarities, NRs span a remarkably broad functional landscape. They regulate diverse biological processes including metabolism, immune function, and development, reflecting both their ligand diversity and context-dependent cofactor interactions.⁹⁶ For example, the glucocorticoid receptor (GR) modulates inflammatory gene expression in response to cortisol, playing a central role in immune suppression⁹⁷, while peroxisome proliferator-activated receptors (PPARs) regulate lipid metabolism and energy homeostasis in response to fatty acid derivatives.⁹⁸ Similarly, retinoic acid receptors (RARs) guide developmental gene expression programs by responding to vitamin A metabolites.⁹⁹

This combination of conserved structural features, shared DNA-binding logic, and divergent functional outputs makes nuclear receptors a compelling system for understanding how transcription factors can maintain mechanistic consistency while achieving context-specific regulatory effects.

1.5.2 Vitamin D Receptor in metabolism and immunity

One of the most well-characterized members of the nuclear receptor family is the Vitamin D Receptor (VDR), which exemplifies how environmental inputs can be translated into transcriptional responses. VDR is primarily activated by calcitriol (1,25-dihydroxyvitamin D), the biologically active form of vitamin D, and regulates gene programs involved in calcium

homeostasis¹⁰⁰, immune function¹⁰¹, and epithelial biology.¹⁰²

The production of calcitriol is a multi-step process that integrates environmental exposure with systemic metabolism. It begins in the skin, where 7-dehydrocholesterol undergoes a non-enzymatic photochemical conversion to vitamin D (cholecalciferol) upon exposure to ultraviolet B (UVB) radiation. This initial step directly links sunlight exposure to downstream transcriptional regulation. Vitamin D then enters circulation and is transported to the liver, where it is hydroxylated by 25-hydroxylase (primarily CYP2R1) to form 25-hydroxyvitamin D (calcidiol), the major circulating form. Subsequently, calcidiol is delivered to the kidneys, where it undergoes a second hydroxylation by 1α -hydroxylase (CYP27B1) to generate calcitriol, the high-affinity ligand for VDR.¹⁰³

Upon ligand binding, VDR undergoes conformational changes that promote heterodimerization with the retinoid X receptor (RXR) and binding to vitamin D response elements (VDREs), typically arranged as direct repeat motifs separated by three nucleotides (DR3 elements). This interaction facilitates the recruitment of coactivators or corepressors, enabling precise regulation of target gene expression in a tissue- and context-dependent manner.¹⁰⁴

Much of VDR's function is executed through its role as a ligand-activated transcription factor, enabling coordinated regulation of gene expression programs that maintain calcium homeostasis. This function is essential, as tightly controlled calcium levels are required for a wide range of physiological processes, including neuromuscular signaling, blood clotting, and intracellular signaling dynamics.

A major site of VDR-mediated regulation is the skeleton, which serves as the body's primary reservoir of calcium. Within bone, VDR activity in osteoblasts promotes the expression of receptor activator of NF- κ B ligand (RANKL), a key cytokine that drives the differentiation and activation of osteoclasts. Osteoclasts, in turn, resorb bone matrix, releasing stored calcium into the bloodstream.¹⁰⁵ Through this mechanism, VDR indirectly facilitates the mobilization of calcium from bone to maintain systemic calcium levels, par-

ticularly under conditions of deficiency.

However, the role of VDR in calcium regulation extends well beyond bone resorption. In the intestine, VDR plays a central role in promoting dietary calcium absorption. Upon activation by calcitriol, VDR induces the expression of calcium transport proteins such as TRPV6 (an apical calcium channel), calbindin-D9k (an intracellular calcium-binding protein), and PMCA1b (a basolateral calcium ATPase).¹⁰⁶ Together, these factors enable efficient transcellular transport of calcium from the intestinal lumen into circulation, representing one of the most important mechanisms for maintaining calcium balance under normal physiological conditions.

In the kidney, VDR contributes to calcium conservation by regulating reabsorption pathways in the distal nephron. VDR activation promotes expression of proteins such as TRPV5 and calbindin-D28k, which facilitate calcium reuptake from the filtrate back into the bloodstream, reducing urinary calcium loss.¹⁰⁷ This renal function complements intestinal absorption and bone remodeling to fine-tune circulating calcium levels. Importantly, VDR operates within a broader endocrine network that includes parathyroid hormone (PTH). When blood calcium levels drop, PTH secretion increases, stimulating renal production of calcitriol. Elevated calcitriol then activates VDR in target tissues to enhance intestinal absorption, promote bone resorption (via RANKL), and increase renal reabsorption. In turn, VDR signaling can suppress PTH gene expression in the parathyroid gland, forming a negative feedback loop that prevents excessive calcium mobilization.¹⁰⁸

Beyond its classical role in calcium homeostasis, the Vitamin D Receptor (VDR) has emerged as an important regulator of immune function, acting at the interface of environmental sensing and immune homeostasis. Notably, VDR is expressed across a broad range of immune cell types, including dendritic cells, macrophages, B cells, and T cells, positioning it to directly influence both innate and adaptive immune responses.¹⁰⁹

In innate immune cells such as dendritic cells and macrophages, VDR activation promotes a more tolerogenic and less inflammatory phenotype. Calcitriol signaling inhibits the

maturation of dendritic cells, reducing their expression of MHC class II and co-stimulatory molecules (e.g., CD80 and CD86), which are required for robust T cell activation.¹¹⁰ Concurrently, VDR signaling in macrophages can enhance antimicrobial responses—such as induction of antimicrobial peptides like cathelicidin—while limiting excessive inflammatory cytokine production.¹¹¹ This dual role reflects a broader theme in vitamin D biology: enhancing host defense while constraining immunopathology.

In the adaptive immune system, VDR exerts profound effects on T cell function. Both vitamin D and its active metabolite, 1,25(OH)D, suppress T cell proliferation and inhibit the production of pro-inflammatory cytokines such as IFN- γ and IL-2, which are typically associated with Th1 responses. At the same time, VDR signaling promotes the development and function of regulatory T cells (Tregs), in part by enhancing expression of IL-10, a key anti-inflammatory cytokine. VDR activation can also skew T helper cell differentiation away from Th1 and Th17 phenotypes toward more regulatory or Th2-like states, thereby reshaping the overall immune response toward tolerance rather than inflammation.¹¹²

B cells are similarly influenced by VDR activity. Vitamin D signaling has been shown to inhibit B cell proliferation, plasma cell differentiation, and immunoglobulin production, suggesting that VDR contributes to limiting excessive humoral responses and maintaining immune balance.¹¹³

These cellular effects translate into important physiological and pathological consequences. The immunoregulatory functions of vitamin D are increasingly recognized in the context of immune-mediated diseases. VDR signaling has been associated with suppression or modulation of conditions such as inflammatory bowel disease, type 1 diabetes, and multiple sclerosis, where dysregulated immune activation drives tissue damage. In these settings, vitamin D appears to function as a modulatory signal that dampens pathogenic immune responses while preserving essential host defense mechanisms.

1.5.3 VDR and Autoimmunity

The immunoregulatory functions of the Vitamin D Receptor (VDR) are particularly relevant in the context of autoimmune disease, where breakdowns in immune tolerance lead to tissue-specific pathology. Among these, type 1 diabetes (T1D) and multiple sclerosis (MS) represent two well-studied examples in which VDR signaling has been implicated in modulating disease risk and progression.

In T1D, autoimmune destruction of pancreatic β -cells is driven largely by autoreactive T cells with a strong Th1 and cytotoxic phenotype. VDR activation counteracts several of the key processes underlying this pathology. By suppressing T cell proliferation and reducing production of pro-inflammatory cytokines such as IFN- γ and IL-2, VDR signaling dampens the effector functions that mediate β -cell killing.¹¹⁴ At the same time, VDR promotes the differentiation and stability of regulatory T cells (Tregs), which are essential for maintaining peripheral tolerance to self-antigens, including those expressed in pancreatic islets. In antigen-presenting cells, such as dendritic cells, VDR signaling reduces co-stimulatory capacity and favors a tolerogenic phenotype, further limiting the activation of autoreactive T cells.¹¹⁵ Collectively, these effects suggest that adequate vitamin D signaling may help restrain the initiation and amplification of autoimmune responses targeting β -cells.

In MS, a similar immunological framework applies, though the target tissue is the central nervous system. Disease progression is driven by autoreactive T cells—particularly Th1 and Th17 subsets—that infiltrate the CNS and mediate demyelination.¹¹⁶ VDR signaling has been shown to suppress both Th1 and Th17 differentiation while promoting Treg responses, thereby shifting the balance away from inflammatory and toward regulatory immunity. Additionally, VDR activity in innate immune cells, including microglia and infiltrating macrophages, can reduce the production of inflammatory mediators that contribute to neuronal damage.^{mirarchi'microglia'2023chi}

Epidemiological observations further support a role for vitamin D signaling in these diseases. Both T1D and MS show geographic and seasonal patterns consistent with variation

in sunlight exposure, suggesting that reduced vitamin D synthesis may increase disease susceptibility. Moreover, genetic variation in components of the vitamin D pathway, including polymorphisms in the VDR gene, has been associated with altered risk for both conditions, although these associations are complex and context-dependent.¹¹⁷

1.5.4 Relationship between VDR and NF- κ B

Another well-characterized mechanism of VDR-mediated immune regulation is its ability to antagonize NF- κ B signaling pathway activity, a central driver of inflammatory gene expression. NF- κ B activity is tightly controlled at multiple levels, with one of the key regulatory steps being the cytoplasmic sequestration of NF- κ B dimers (most commonly p65/p50) by inhibitor of κ B (I κ B α). In resting cells, this interaction prevents NF- κ B from translocating to the nucleus and activating pro-inflammatory target genes.¹¹⁸

Activation of NF- κ B requires phosphorylation of I κ B α by the I κ B kinase (IKK) complex—particularly the catalytic subunit IKK β —which targets I κ B α for ubiquitination and proteasomal degradation. This degradation releases NF- κ B dimers, allowing their nuclear translocation and subsequent transcriptional activation of genes encoding cytokines, chemokines, and other mediators of inflammation.¹¹⁹

VDR interferes with this pathway at a critical upstream step. Upon activation by calcitriol, VDR can directly interact with IKK β , inhibiting its kinase activity.¹²⁰ This prevents phosphorylation of I κ B α , thereby stabilizing the inhibitor and maintaining NF- κ B dimers in the cytoplasm. As a result, nuclear translocation of NF- κ B is reduced, leading to diminished transcription of pro-inflammatory genes. This mechanism provides a rapid, non-genomic means by which VDR can suppress inflammatory signaling independently of direct DNA binding.

In addition to this cytoplasmic effect, VDR can also modulate NF- κ B activity at the transcriptional level. VDR has been reported to interfere with NF- κ B-driven gene expression through competition for shared coactivators or through direct interaction with NF- κ B sub-

units in the nucleus, further dampening transcriptional output. Moreover, VDR target genes can include inhibitors of inflammatory signaling, reinforcing this suppressive effect through secondary transcriptional programs.¹²¹

Functionally, this inhibition of NF- κ B signaling contributes to the broader anti-inflammatory and tolerogenic effects of vitamin D. Given that aberrant NF- κ B activation is a hallmark of many inflammatory and autoimmune diseases, VDR-mediated suppression of this pathway represents a key molecular link between vitamin D signaling and immune homeostasis.

1.6 Significance of transcriptional heterogeneity in mTECs

Building on these principles of signal-dependent transcriptional regulation, this work seeks to understand how transcription factors shape heterogeneity in self-antigen presentation. In particular, it focuses on how regulatory mechanisms control the diversity, frequency, and magnitude of TSG expression in mTECs.

1.6.1 Importance of self-antigen diversity in negative selection

To understand how autoreactive T cells are eliminated from the repertoire, it is important to consider several fundamental features of this process. First, T cell receptors (TCRs) are not pre-configured to recognize pathogens, but are instead generated through random gene rearrangement. As a result, TCR specificities are not binary but exist along a spectrum of affinities for a wide range of antigens. This means that multiple T cells can recognize the same antigen, while individual T cells may exhibit cross-reactivity to multiple distinct antigens.

This inherent variability presents a significant challenge for thymic selection. T cells with high affinity for self-antigens must be eliminated to prevent autoimmunity, whereas those with lower affinity must be preserved to maintain a functional immune repertoire. These selection outcomes depend critically on the extent to which developing T cells en-

counter self-antigens during their maturation.

Within the thymic medulla, developing T cells migrate and interact with antigen-presenting cells, sampling the available pool of self-antigens. Thus, the probability of a given TCR engaging its cognate antigen is directly influenced by the amplitude and frequency of antigen presentation. Effective tolerance therefore requires a balance: sufficient presentation of self-antigens to eliminate highly self-reactive T cells, but not so extensive as to deplete the repertoire of potentially useful, low-affinity clones.

This “Goldilocks” requirement implies that antigen presentation must be finely tuned. In this context, mTECs must exhibit substantial transcriptional heterogeneity, enabling the distributed expression of self-antigens necessary to achieve an optimal balance between tolerance and immune competence.

1.6.2 Venturing into the unknowns of mTEC transcriptional heterogeneity

The breadth of TSG expression is a key determinant of how effectively central tolerance is established. In principle, the quality of self-antigen presentation by mTECs could be achieved through several distinct strategies. One possibility is that individual mTECs express a broad range of TSGs at relatively low levels (high frequency, low amplitude), thereby increasing the likelihood that diverse autoreactive T cells encounter their cognate antigens. Alternatively, TSG expression could be more restricted (low frequency) but occur at higher levels (high amplitude), prioritizing the elimination of strongly autoreactive clones against a subset of high-risk self-antigens. A third possibility is a finely tuned system in which both the frequency and expression level of each TSG are precisely regulated across the mTEC population.

Each of these strategies presents distinct advantages and trade-offs. A high-frequency, low-amplitude model would promote broad antigen coverage with relatively uniform expression across cells, reducing the need for tight regulatory control. However, this approach may be insufficient for eliminating all autoreactive clones, as some T cells may require higher

levels of antigen presentation to trigger deletion.

In contrast, a low-frequency, high-amplitude model could enable more efficient deletion of T cells reactive to particularly critical self-antigens. For example, increased representation of genes such as *Ins2* could enhance protection against diseases like Type 1 Diabetes, which have significant physiological consequences. However, this biased strategy may come at the cost of reduced coverage of other self-antigens, potentially leaving gaps in tolerance and increasing susceptibility to other autoimmune conditions.

1.6.3 Regulation of expression frequency and amplitude

The most effective system is therefore likely to involve a balance between these extremes, in which both the frequency and magnitude of TSG expression are tuned across the mTEC population. Such a strategy would allow sufficient presentation of diverse self-antigens while ensuring that particularly sensitive or high-risk targets are expressed at levels adequate to drive robust negative selection.

A model in which both the frequency and amplitude of TSG expression are independently and tightly regulated represents the most mechanistically sophisticated framework—and is strongly supported by existing data. Early bulk RNA-seq studies revealed mosaic TSG expression across mature mTECs, implying heterogeneity in either the probability of expression, the level of expression, or both.²⁹ Subsequent single-cell analyses refined this view, demonstrating that individual TSGs are expressed at reproducible but gene-specific frequencies across cells¹²², while also exhibiting substantial variability in transcript abundance within expressing cells.⁴ Together, these findings argue that mTECs do not simply toggle TSGs on or off, but instead operate within a coordinated regulatory regime that modulates both how often genes are expressed and how strongly they are transcribed.

It remains unclear how the amplitude of TSG expression is regulated. As described above, AIRE has been characterized as a central driver of TSG expression, in part by promoting the release of poised, paused RNA polymerases at TSG loci. While this mechanism

explains how transcription is initiated, it may also influence transcriptional output. One possibility is that local AIRE abundance or activity modulates transcriptional amplitude, such that variation in AIRE dosage contributes directly to differences in expression levels across cells. Notably,³⁷, implying that its function depends on recruitment to chromatin through interactions with other transcription factors or cofactors. This raises the possibility that transcription factor networks, rather than Aire alone, play a key role in tuning the amplitude of TSG expression.

In contrast, the regulation of TSG expression frequency across the mTEC population remains even less well defined. Drawing on Waddington’s epigenetic landscape¹²³, one can conceptualize gene expression as constrained by chromatin-defined “barriers” that canalize cell states. In this framework, chromatin accessibility and transcription factor activity could restrict the repertoire of genes available for expression within individual cells, thereby shaping the frequency with which specific TSGs are expressed across the population. Such constraints have been observed in other cellular systems^{124–126}, where chromatin state biases gene expression programs, but this principle has not been fully explored in the context of controlled population heterogeneity in mTECs.

An additional open question is whether the frequency and amplitude of TSG expression are governed by shared regulatory mechanisms or are independently controlled at distinct molecular levels. Here, we investigate these possibilities, aiming to disentangle the regulatory architecture underlying TSG expression heterogeneity.

1.6.4 *Assessing transcriptional heterogeneity*

The most fundamental metric for assessing transcriptional heterogeneity within a cell population is variance.¹²⁷ Variance is mathematically defined as

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

where s^2 is the sample variance, n is the total number of observations, x_i is the value

of the i^{th} observation, and $\bar{x}$ is the sample mean. Variance quantifies the average squared deviation of observations from their mean and therefore provides a measure of dispersion within a dataset.

In sequencing-based gene expression data, however, variance is strongly dependent on mean expression level, with highly expressed genes generally exhibiting greater variance than lowly expressed genes.¹²⁸ As a result, direct comparisons of variance across genes can be misleading, as highly expressed genes may appear more heterogeneous simply due to their abundance rather than true biological variability.

To account for this mean–variance relationship, single-cell analysis frameworks commonly employ standardized measures of variance. In the original Seurat¹²⁹ highly variable gene (HVG) workflow (`FindVariableFeatures`, `vst` method), the mean expression and variance of each gene are first calculated across all cells. Seurat then fits a mean–variance trend to estimate the expected variance associated with a given level of expression. The observed variance of each gene is subsequently compared with this expectation, generating a standardized variance metric that quantifies the extent to which a gene exhibits greater variability than predicted based on its average expression level.

Conceptually, standardized variance can be represented as

$$\text{Standardized Variance} = \frac{\text{Observed Variance}}{\text{Expected Variance at the Same Mean Expression}}.$$

Genes with standardized variance values greater than 1 exhibit more variability than expected and are therefore prioritized as highly variable genes, whereas genes with values near 1 display variability consistent with the expected mean–variance relationship.

1.7 Aims & Significance

This work seeks to uncover molecular mechanisms of tissue-specific gene regulation for immune tolerance.

Aim 1: Determine how the diversity of self-antigen expression in mTECs is controlled

The establishment of central tolerance depends on the presentation of a diverse repertoire of self-antigens by mTECs. Here, we investigate how the Vitamin D receptor (VDR) regulates this process by modulating the amplitude of NF- κ B signaling. We quantify how VDR shapes the diversity of antigen expression programs and assess the consequences for thymocyte selection and T cell self-reactivity.

Aim 2: Develop a model of the NF- κ B–AP-1 transcriptional relay that establishes tissue-specific gene expression in mTECs.

Tissue-specific gene expression in mTECs requires coordinated activity of multiple transcription factors, yet the mechanisms underlying this coordination remain unclear. We propose a model in which NF- κ B, specifically RELA, primes the chromatin landscape and directs the redistribution of AP-1 to tissue-specific loci via composite NF- κ B:AP-1 binding sites. Furthermore, we assess how perturbations of this relay—through deletion of RELA or the m6A reader YTHDF2, which regulates expression of the AP-1 component *Fos*—disrupt tissue-specific antigen expression and impair immune tolerance.

Together, these aims will define how dynamic regulation of transcription factor activity controls the diversity and heterogeneity of self-antigen expression in mTECs, providing mechanistic insight into the establishment of central tolerance and the prevention of autoimmunity

CHAPTER 2:

The Vitamin D receptor promotes the diversification of self-antigens presented by thymic epithelia for immune tolerance

Joshua McKeever^{1,2} , Jason A. Caldwell¹ , Noah Gamble^{1,3} , Olubusayo Bolonduro^{1,4} , Alexander Hoffmann⁵ , Marc Timmers⁶ , Michael Downes⁷ , Ronald M. Evans⁷ , and Andrew S. Koh^{1,8,*}

¹Department of Pathology, University of Chicago, Chicago, IL, USA

²Committee on Molecular Metabolism and Nutrition, University of Chicago, Chicago, IL, USA

³Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL, USA

⁴Committee on Genetics, Genomics, Systems Biology, University of Chicago, Chicago, IL, USA

⁵Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA 90095, USA

⁶German Cancer Consortium (DKTK) Partner Site Freiburg, German Cancer Research Center (DKFZ), Heidelberg, Germany

⁷Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA

⁸Institute for Biophysical Dynamics, University of Chicago, Chicago, IL, USA

*Correspondence: akoh@uchicago.edu

2.1 Abstract

T cell repertoire selection in the thymus directly shapes the immune system’s capacity to mount a response against pathogens while maintaining tolerance to healthy tissues^{.23} Medullary thymic epithelial cells (mTECs) play a central role in this process by presenting a broad array of self-antigens, yet the mechanisms that generate and regulate this diversity remain incompletely understood.^{29,32} Here, we identify the Vitamin D receptor (VDR) as a key regulator of mTEC transcriptional heterogeneity and self-antigen diversity. Conditional deletion of *Vdr* in mTECs reduces single-cell transcriptional heterogeneity, leading to the loss of expression of nearly one thousand tissue-specific genes and overrepresentation of a distinct subset of others. This imbalance in ectopic expression disrupts central tolerance, causing the escape of autoreactive T cells and multi-organ autoimmunity. Single-cell transcriptomic and chromatin accessibility analyses reveal that VDR constrains the variance and amplitude of NF- κ B signaling by restricting access to chromatin regions enriched for NF- κ B motifs. Through this mechanism, VDR limits excessive activation of select tissue-specific genes and promotes balanced, heterogeneous self-antigen expression in mature mTECs. Together, these findings define VDR as a critical determinant of mTEC heterogeneity and thymic tolerance.

2.2 Introduction

The diversity of T cell receptors (TCRs) generated in the thymus underlies the adaptive immune system’s ability to recognize a vast array of pathogens.^{18,130} However, this same process inevitably generates TCRs with the capacity to recognize self-antigens, creating an inherent risk of autoimmunity.^{131,132} As a result, the immune system must balance the need for broad antigen recognition with the requirement to eliminate or control self-reactive clones.

Central tolerance mechanisms, particularly negative selection in the thymic medulla, are critical for maintaining this balance.²³ During this process, developing thymocytes are exposed to self-antigens and those bearing high-affinity receptors for self are deleted or diverted into regulatory lineages. Importantly, this selection occurs against a backdrop of graded TCR affinities rather than binary recognition, suggesting that the strength and diversity of self-antigen presentation play a key role in shaping the resulting T cell repertoire. Thus, the quality and structure of antigen presentation in the thymus are central determinants of immune tolerance.

mTECs play a central role in this process by ectopically expressing thousands of TSGs that are otherwise restricted to peripheral tissues.²⁹ Through this promiscuous gene expression program, mTECs provide a molecular representation of self within a single anatomical site. Notably, this expression is not uniform across cells. Instead, individual mTECs express only a subset of the total TSG repertoire, resulting in a highly heterogeneous but population-level comprehensive display of self-antigens. This organization suggests that antigen presentation is distributed across the mTEC population rather than consolidated within individual cells.

Despite this heterogeneity, TSG expression follows reproducible patterns at the population level, with individual genes expressed at characteristic frequencies across cells.^{1,3} This implies that transcriptional diversity in mTECs is not purely stochastic, but instead reflects regulated processes that control which genes are expressed, and in how many cells. Such regulation is likely to be critical for ensuring that the collective output of the mTEC pop-

ulation provides sufficient coverage of self-antigens to enforce tolerance, while maintaining the variability needed to sample a broad antigenic space.

However, how mTECs achieve this predictable yet heterogeneous gene expression program remains poorly understood. In particular, the mechanisms that control the distribution of gene expression across cells, and how these mechanisms interface with signaling pathways and chromatin dynamics during mTEC differentiation, are not well defined. More broadly, how transcriptional regulatory networks are tuned to generate diversity at the single-cell level while maintaining robustness at the population level remains an open question.

Here, we sought to investigate the mechanisms controlling transcriptional diversity in mTECs and how this diversity contributes to self-antigen representation and the establishment of central immune tolerance.

2.3 Results

2.3.1 TSGs strongly contribute to high transcriptional variance exhibited by Mature mTECs

To profile transcriptional heterogeneity in subpopulations of mTECs, we used the 10X Genomics Chromium Gene Expression platform in mTECs of a four-week-old C57BL/6 mouse that were sorted by fluorescence-activated cell sorting (FACS) (Supp Fig 1a), then identified distinct clusters based on stage-specific marker genes (Figure 2.5.1a and Supplemental Figure 1b-f). Within each group, we calculated the variance in gene expression for all genes captured in the assay as a measure of transcriptional heterogeneity in the populations (Supplemental Figure 1h). In comparing the top quartiles of gene expression variance between mature mTECs and immature mTECs, we found that transcriptional variance for all genes dramatically increases upon maturation (Figure 2.5.1b and Supplemental Figure 1i). Mature mTECs have been reported to express TSGs at low, predicable frequencies.^{1,3,133,134} This suggests ordered, heterogenous expression of TSGs at the mature mTEC stage. This increase in variance coincides with high levels TSG expression (Supplemental Figure 1g).

Thus, we asked whether TSG expression accounts for much of the variance observed in mature mTECs. Separately considering TSG expression variance and variance of other genes revealed that TSGs significantly contributes to the overall high variance observed in mature mTECs. Without TSGs, mature mTEC expression variance is similar to the variance exhibited in immature mTECs (Figure 2.5.1c). To reinforce that TSG expression is highly variable, we took the difference in gene expression variance for all genes between mature and immature mTECs and find that almost 7000 genes increase in variance across the developmental transition and approximately 60 percent of these genes are TSGs (Figure 2.5.1e). Interestingly, genes related to the epithelial program like those encoding for NF- κ B subunits and master regulator FOXN1¹³⁵ do not exhibit changes in expression variance (Figure 2.5.1e).

To determine whether the high variance observed among TSGs reflects transcriptional sparsity rather than uniformly low expression, we first examined how many mature mTECs are required to capture the full repertoire of expressed TSGs. Compared to other genes, TSGs required more cells—approximately 450—to reach saturation, indicating that their expression is distributed across distinct subsets of mature mTECs (Figure 2.5.1d). We then calculated the mean expression level of each TSG across all mature mTECs in which it was detected and normalized these values to the housekeeping gene *Hprt*. This analysis revealed that, when expressed, TSGs are transcribed at levels comparable to *Hprt*, but are detected in far fewer cells, reflecting the sparse and heterogeneous nature of TSG expression across the mTEC population (Figure 2.5.1f).

2.3.2 *Vitamin D Receptor emerges as a regulator of mTEC transcriptional heterogeneity*

Transcriptional heterogeneity within cells is often governed by complex networks of transcription factors (TFs).^{81,136} In our analyses, we observed broad shifts in TF activity during mTEC maturation (Supplementary Fig. 2b)³⁶—a developmental stage characterized by a marked increase in transcriptional heterogeneity. We hypothesized that TFs whose activity increases during this transition drive the diversification of gene expression programs in

mature mTECs.

To identify candidate regulators, we analyzed mTEC Multiome data, using marker gene expression to define clusters (Fig. 2.5.2a; Supplementary Figs. 1b–f and 2a). We then quantified TF activity changes from mTEC_{lo} to mTEC_{hi} using chromVAR deviation scores and assessed changes in deviation score variability (Supplementary Fig. 2d). Notably, members of the NF- κ B and nuclear receptor (NR) families exhibited some of the largest increases in both activity and variability across maturation (Fig. 2.5.2b; Supplementary Fig. 2c). These findings were corroborated by transcription factor footprinting of bulk inTAC-seq data¹³⁷, which revealed enhanced motif accessibility and inferred binding at NR sites in mature mTECs (Fig. 2.5.2c).

NRs are a well-characterized family of ligand-activated transcription factors with established roles in cell differentiation, metabolism, and immune regulation.^{94,138,139} Given these functions, NRs represent compelling candidates for modulating transcriptional variability in mTECs. Because many nuclear receptors act as ligand-sensitive transcriptional switches, their activity can introduce context-dependent variation in gene expression programs¹⁴⁰. Such tunable transcriptional control could underlie the increased heterogeneity observed during mTEC maturation. However, the high degree of sequence homology among NR family members often makes it difficult to pinpoint which specific receptors are transcriptionally active in a given context.¹⁴¹

To determine which nuclear receptor might be active during mTEC maturation, we examined the expression of all known NRs in mature mTECs, focusing on those with both high motif accessibility and appreciable transcript abundance. Among these, the Vitamin D Receptor (*Vdr*) emerged as the most highly expressed NR in mature mTECs in both our bulk RNA-seq dataset and the ImmGen database¹⁴² (Figure 2.5.2d; Supplementary Figure 3a), consistent with functional activation at this stage. Notably, *Vdr* expression in mature mTECs exceeded that of all 82 other cell types profiled by ImmGen (Supplementary Figure 3b). Furthermore, VDR deviation scores were highest among cells in the top quartile of TSG

expression within *Aire*⁺ mature mTECs (Figure 2.5.2e). Conversely, cells with high VDR deviation scores displayed elevated cumulative TSG expression relative to those with low VDR activity (Supplementary Figures 3d,e). Collectively, these findings support a model in which VDR contributes to the regulation of TSG expression variance in mature mTECs.

2.3.3 *VDR loss limits range and amplitude of tissue-specific gene expression*

To determine whether VDR regulates transcriptional heterogeneity, we performed scRNA-seq on *Vdr* WT and cKO mTECs using *Vdr*^{fl/fl}; *Foxn1*^{Cre -/-} and *Vdr*^{fl/fl}; *Foxn1*^{Cre +/-} mice. Default clustering of WT and KO cells showed mTECs at various developmental stages based on marker gene expression (Figure 3a and Supplemental Figure 4b-d). Interestingly, when split by condition, we find a cluster of cells almost entirely specific to the *Vdr* KO sample (Figure 2.5.2b and Supplemental Figure 4a). Surprisingly, this cluster of cells accounted for 50% of *Vdr* KO mTECs (Figure 2.5.3c). Given this convergence on a gene expression profile in VDR-deficient mTECs, we investigated whether tissue-specific gene expression was affected. The fact that the majority (80%) of cells in the *Vdr* KO specific cluster express *Aire*, suggests that these cells are able to reach the point of tissue-specific gene expression in development (Supplemental Figure 2d). In comparing TSG expression scores across all population clusters in the dataset, we found that the *Vdr* KO -specific cluster exhibited the highest scores (Figure 2.5.3d).

To define how VDR influences tissue-specific gene expression, we compared transcriptional profiles between *Vdr* KO and WT AIRE⁺ mTECs. Many genes were expressed at higher frequencies in VDR-deficient cells, a substantial fraction of which were TSGs (Figure 2.5.3e). Focusing on TSGs that exhibited an increase in expression frequency of at least 5% in KO cells, we identified 195 genes, including *Ins2* and *Gal*, that met this criterion (Figure 2.5.3e). To determine the cellular source of these changes, we examined the distribution of up-regulated genes across transcriptional clusters and found that their expression was concentrated primarily within the *Vdr* KO-specific cluster (Supplementary Figures 4e-g,j-l).

One possible explanation for the increased frequency of TSG expression was expansion of the AIRE⁺ mTEC compartment. However, flow cytometric analysis revealed no difference in the frequency of AIRE⁺ mTECs between *Vdr* WT and KO mice (Supplementary Figures 4h,i), indicating that the observed transcriptional changes are unlikely to result from altered population composition and instead reflect cell-intrinsic molecular differences.

Consistent with this interpretation, quantitative comparisons of KO and WT cells showed that TSGs exhibited larger increases in both expression frequency and expression variance than non-TSGs (Figures 2.5.3f,g). These findings suggest that loss of VDR promotes the emergence of a distinct transcriptional state characterized by altered patterns of promiscuous gene expression. Despite the increased frequency of expression of individual TSGs across the population, fewer distinct TSGs were detected within individual KO cells. Consequently, the ratio of total TSG transcript abundance to the number of expressed TSGs was increased in KO cells (Figures 2.5.3h,i), indicating that a smaller set of TSGs accounted for a greater fraction of total TSG expression. Together, these results suggest that loss of VDR shifts mTECs toward more concentrated expression of specific TSG programs at the expense of overall TSG diversity.

Analysis of expression magnitude revealed reduced expression of over one thousand TSGs in the *Vdr* KO, indicating widespread loss of self-antigen transcription. This reduction was not limited to lowly expressed genes, but affected a broad portion of the TSG repertoire, suggesting a substantial contraction in the diversity of self-antigens available for presentation. Alongside this loss, several hundred TSGs exhibited increased expression magnitude (Figure 3j). Notably, many genes overrepresented in the KO overlapped with the VDR-repressed gene set (Figure 3k). The magnitude of repression among down-regulated genes exceeded the induction observed for up-regulated genes in both datasets (Supplementary Figures 5b–e). Consistent with this imbalance, the VDR-repressed subset displayed stronger $-\log p$ values, indicating that *Vdr* loss disproportionately destabilizes genes with otherwise robust expression, a pattern observed in both single-cell and bulk RNA-seq data (Figure

2.5.3j; Supplementary Figure 5a). Collectively, these data indicate that VDR deficiency leads primarily to the loss and destabilization of a large fraction of the TSG repertoire, with more limited compensatory up-regulation of a smaller subset of highly active genes.

2.3.4 *VDR increases bandwidth of NF- κ B signaling to promote transcriptional heterogeneity*

To identify factors that could activate genes in regions repressed by VDR, we performed scATAC-seq on *Vdr* WT and cKO mTECs. Using accessibility at loci of known mTEC stage markers and known cell type-biased TFs, we annotated clusters (Supplemental Figure 6a-c). UMAP projections split by genotype revealed divergence in transit-amplifying and mature mTECs, suggesting strong differences in chromatin accessibility and transcription factor activity between WT and KO cells at these stages (Figure 2.5.4a).

Within this cluster, motif deviation analysis showed that NF- κ B motifs had the strongest increase in accessibility in *Vdr* KO cells (Figure 2.5.4b), indicating that VDR can inhibit NF- κ B activity. VDR-deficient cells exhibited higher NF- κ B activity at the transit-amplifying, mature *Aire*+, and late/post *Aire* stages (Supplemental figure 6g-k). This is consistent with prior work showing VDR inhibition of NF- κ B, particularly the RELA (p65):p50 dimer.^{120,121} Notably, NF- κ B deviations were most variable in transit-amplifying and mature mTECs (Figure 2b), suggesting that variability in their activity could be linked to transcriptional heterogeneity. Among NF- κ B family members, RELA and REL displayed the highest variability.

Prior studies have shown that subunit RELB is essential for mTEC maturation, as its loss ablates the mature mTEC compartment, whereas RELA deficiency permits maturation but reduces TSG expression.⁵¹ These findings highlight distinct roles of NF- κ B subunits. Consistent with this, RELA motifs were enriched in regions that gain accessibility during mTEC maturation and near TSG loci, compared with regions that remain static (Figure 2.5.4c). RELA also showed greater accessibility and footprinting depth at TSG-associated

regions relative to controls (Figure 2.5.4d), supporting its direct action at or near TSG loci. RELA green CUTandRun¹⁴³ in mTECs shows that RELA binding at regions near TSGs, as well as enhanced binding at regions near *Vdr* KO overrepresented genes (Figure 2.5.4e).

Because VDR loss increases both the frequency of TSG expression and NF- κ B activity, we next tested whether NF- κ B activity levels correspond to expression of different TSGs. Using Multiome data from *Aire*+ mTECs, we focused on the three NF- κ B subunits with transactivation domains.¹⁴⁴ Stratifying *Aire*+ cells by quartiles of REL, RELA, and RELB deviation scores, we observed that NF- κ B activity levels correlated with distinct patterns of TSG expression (Figure 2.5.4f).

Given the importance of activity levels for TSG expression, we asked whether NF- κ B activity is constrained in VDR-deficient cells. To do this, we assessed the range of RELA motif deviation scores in mature mTECs by taking the deviation score for each cell and normalizing it to the average deviation score for the population. We observed a significant decrease in deviation score variance *Vdr* cKO cells (Figure 2.5.4g). Indeed, this suggests that VDR-deficient cells exhibit a higher, yet more constrained range of RELA activity. To provide additional support across datasets, we analyzed the most variably expressed NF- κ B targets in WT cells and found that transcriptional variance of these genes was reduced in *Vdr* cKO cells, suggesting a narrower dynamic range of NF- κ B activity (Figure 2.5.4h). To directly connect this to TSG expression, we compared NF- κ B target expression in cells with high expression of *Vdr* KO-overrepresented TSGs versus other TSGs. Strikingly, cells expressing overrepresented *Vdr* KO TSGs exhibited uniformly high NF- κ B target gene expression, but within a restricted range (Figure 2.5.4i).

2.3.5 *VDR restricts access to chromatin regions with high RELA motif density to regulate mTEC transcriptional heterogeneity*

The range and amplitude of NF- κ B activity may be regulated by controlling where RELA can bind in mTECs, as transcription factor activity in signaling pathways is ultimately decoded

by chromatin accessibility.¹⁴⁵ We found that VDR regulates accessibility at thousands of RELA motifs, and these are generally more accessible than regions without motifs, and the changes in accessibility between WT and KO is greater for peaks containing RELA motifs. (Figure 2.5.5b, 5c).

This regulation could occur indirectly through chromatin remodeling or directly through VDR binding. To distinguish between these possibilities, we performed CUT&RUN for VDR in mature mTECs. VDR binding was enriched at differentially accessible regions compared with non-differentially accessible sites, and specifically at repressed peaks containing RELA motifs, relative to other repressed peaks and RELA-associated induced peaks (Figures 2.5.5d–f). Consistent with prior studies showing that VDR functions as a repressor in complex with nuclear co-repressor complex (NCOR)¹⁴⁶, these results indicate that VDR directly engages RELA motif-containing regions to restrict their accessibility.³

We next asked why these particular RELA motif-containing regions must be repressed to maintain mTEC heterogeneity. One possibility is that such regions represent “high-affinity” RELA binding sites, such that their accessibility could bias RELA occupancy and reduce transcriptional diversity. Supporting this idea, VDR-repressed regions with RELA motifs contained a higher density of motifs compared to VDR-independent RELA regions (Figure 2.5.5g). Moreover, regions proximal to the 195 TSGs overrepresented in *Vdr* KO cells had a higher RELA motif density compared to regions near other TSGs or housekeeping genes (Figure 2.5.5i). Consistently, scATAC fragments near *Vdr* KO-overrepresented TSGs were more enriched for RELA motifs than those near all TSGs (Figure 2.5.5h). We hypothesize that VDR restricts accessibility at RELA motif-dense regions to prevent overrepresentation of particular TSGs, thereby permitting RELA to distribute across additional genomic targets with lower motif density.

To test the prediction that RELA motif-dense regions are less accessible in wild-type cells, we compared transcription factor motif enrichment between longer ATAC fragments (>150 bp), which arise from nucleosome-dense and therefore less accessible regions¹⁴⁷, and

shorter fragments corresponding to more accessible chromatin (Supplemental Figure 7a). RELA motifs were the most strongly enriched in long fragments relative to short fragments, with VDR motifs also enriched, consistent with a model in which VDR represses accessibility at RELA motif-dense regions (Figure 2.5.5j).

We next asked whether this pattern was preferentially associated with genes dysregulated upon VDR loss. As a control, we examined motif enrichment specifically near TSGs overrepresented in *Vdr* KO cells compared to other TSGs. Consistent with our hypothesis, long fragments near *Vdr* KO-overrepresented genes were further enriched for both RELA and VDR motifs, suggesting that these regions are selectively constrained in the presence of VDR (Figure 2.5.5i). Finally, to determine whether loss of VDR leads to increased RELA occupancy at these sites, we performed greenCUT&RUN for RELA in wild-type and VDR-deficient contexts using mVenus-RelA-VDR mice. Consistent with increased accessibility, RELA binding was elevated at regions proximal to *Vdr* KO-overrepresented genes (Figure 2.5.5k).

Together, these data support a model in which VDR directly regulates accessibility at RELA motif-dense regions to balance TSG expression. By restricting RELA from preferentially binding at high-density sites, VDR ensures a broader distribution of NF- κ B activity, thereby maintaining transcriptional heterogeneity in the mTEC population.

2.3.6 *VDR loss in mTECs causes multi-organ autoimmunity*

We next sought to determine whether the compromised activation of certain TSGs caused by VDR-deficiency affected T cell repertoire selection in the thymi of *Vdr* KO mice. To address this, we assayed for evidence of the escape of self-reactive T cells from central tolerance induction in the peripheral tissues and secondary lymphoid organs of *Vdr*-KO mice and N11.5 NOD/*Vdr*^{fl/fl}; *Foxn1*-*Cre*^{+/-} mice. The NOD background was chosen due to its heightened susceptibility to autoimmunity¹⁴⁸ and because autoimmune phenotypes are exacerbated in NOD mice lacking Aire¹⁴⁹. In NOD/*Vdr*^{fl/fl}; *Foxn1*-*Cre* mice, we observed increases in

CD44^{hi}CD62L^{lo} activated T cells and reciprocal decreases in CD44^{lo}CD62L^{hi} naïve T cells in both CD4⁺ T helper and CD8⁺ cytotoxic compartments compared to WT littermates (Figure 2.5.6a,b). However, these changes were not consistent across replicates (Supplemental Figure 7b,c). Splenic Treg populations were also unaffected by *Vdr* loss (Supplemental Figure 7a). We next assessed whether these activated T cells could provoke autoreactive tissue damage. Substantial lymphocytic infiltration was observed in multiple peripheral organs—including the salivary gland, lung, kidney, lacrimal gland, pancreas, and liver—in both *Vdr* KO and NOD/*Vdr*^{fl/fl}; Foxn1-Cre mice, whereas infiltration was absent or less severe in WT mice (Figure 2.5.6e,f). Interestingly, NOD/*Vdr* cKO mice exhibited significant lower body weights compared to WT (Supplemental Figure 7d). Collectively, these results indicate that loss of VDR activity in mTECs impairs central tolerance induction in the thymus, leading to systemic hyperactivity of peripheral T cells and autoimmune manifestations across multiple tissues.

2.4 Discussion

Our study identifies the VDR as a central regulator of transcriptional heterogeneity in mTECs. Using single-cell transcriptomics, we show that transcriptional variance increases markedly during mTEC maturation, driven in large part by tissue-specific gene TSG expression. While TSGs are expressed at levels comparable to housekeeping genes when active, they appear in only a subset of mature mTECs, highlighting a sparse but robust pattern of expression. This heterogeneity likely underlies the ability of the thymic medulla to present a diverse repertoire of self-antigens to developing T cells, a key requirement for central tolerance.

Our analyses reveal that among nuclear receptors, VDR stands out as both highly expressed and functionally active in mature mTECs, with deviation scores correlating with high TSG expression. Loss of VDR skews the transcriptional landscape—a discrete *Vdr* KO-specific mTEC cluster emerges, characterized by altered TSG expression patterns. Notably,

although fewer distinct TSGs are expressed per cell in the absence of VDR, certain TSGs are expressed at higher frequency and magnitude, suggesting that VDR functions primarily as a transcriptional repressor that limits overrepresentation of particular genes. Mechanistically, VDR constrains chromatin accessibility at RELA motif-dense regions, redistributing NF- κ B activity across the genome and thereby broadening the dynamic range of transcriptional activation. In VDR-deficient mTECs, NF- κ B activity is concentrated at a subset of TSG loci, narrowing the amplitude and range of gene expression and ultimately reducing transcriptional heterogeneity.

These findings place VDR at the intersection of chromatin regulation and transcription factor signaling. By selectively repressing "high-affinity" RELA motif-containing regions, VDR prevents the monopolization of NF- κ B by a limited set of target genes, allowing the stochastic yet coordinated expression of TSGs necessary for comprehensive self-antigen presentation. Our CUT&RUN and scATAC-seq data confirm that VDR directly binds and represses chromatin at these regions, consistent with prior reports of VDR functioning in complex with nuclear co-repressor complexes¹⁴⁶.

Importantly, the functional consequences of VDR loss extend beyond the thymus. VDR-deficient mTECs fail to properly regulate TSG expression, leading to impaired central tolerance and the escape of self-reactive T cells. In both C57BL/6 and NOD backgrounds VDR deficiency resulted in elevated frequencies of activated CD4⁺ and CD8⁺T cells and widespread lymphocytic infiltration in multiple peripheral organs. These results highlight a critical role for VDR in maintaining immune homeostasis and preventing multi-organ autoimmunity.

Collectively, our data support a model in which VDR maintains transcriptional heterogeneity in mature mTECs by restricting access to high-density NF- κ B binding regions, thereby distributing transcriptional activity more broadly across the genome. This mechanism ensures balanced, diverse TSG expression, which is essential for central tolerance and immune self-recognition. Beyond mTECs, our findings suggest a general principle whereby

nuclear receptors can modulate transcriptional heterogeneity by constraining high-affinity transcription factor binding, providing a tunable layer of regulation for developmental and immune gene programs.

2.5 Figures

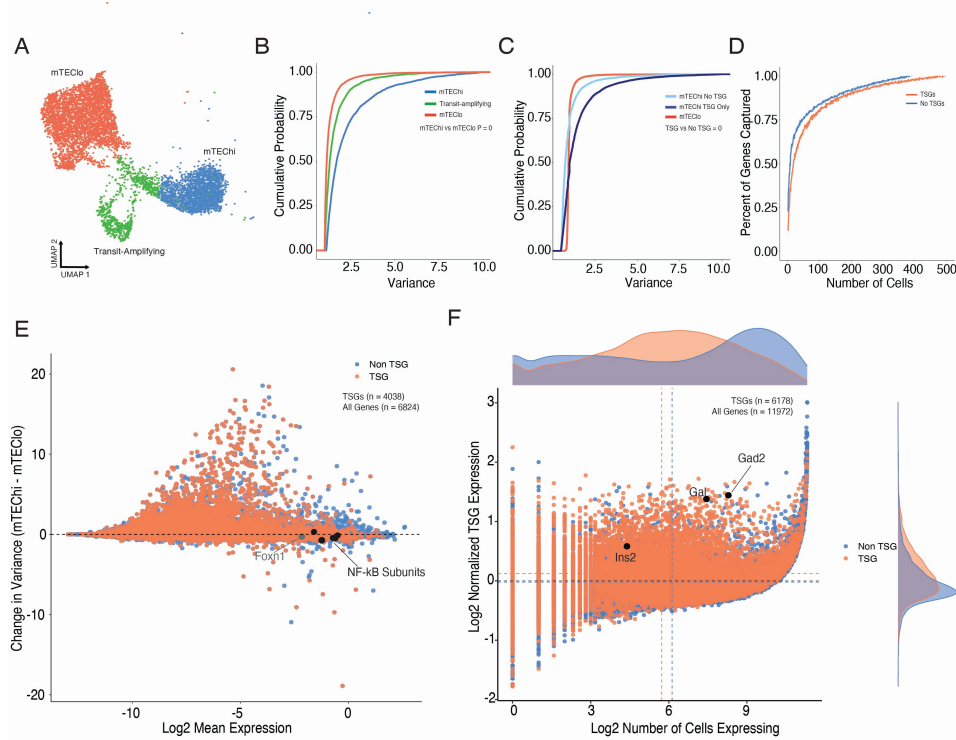


Figure 2.5.1 | Tissue-specific genes are robustly expressed by individual cells and contribute to high transcriptional variance in Mature mTECs : (a) Single cell RNA UMAP of three mTEC subpopulations. (b) CDF of standardized variance of all genes in each population. (c) CDF of standardized variance in mature mTECs excluding TSGs or with only TSGs compared to mTEClo. (d) Plot showing percentage of genes captured by number of cells in the mature mTEC population, using a TSG list and a non-TSG list. (e) Change in variance between mTEChi and mTEClo by Log2 mean expression of each gene. TSGs and non-TSGs highlighted. (f) Log2 mean expression of TSGs relative to *Hprt* expression by Log2 number of cells expressing each gene. TSGs and non-TSGs highlighted. Density of points indicated on the axes.

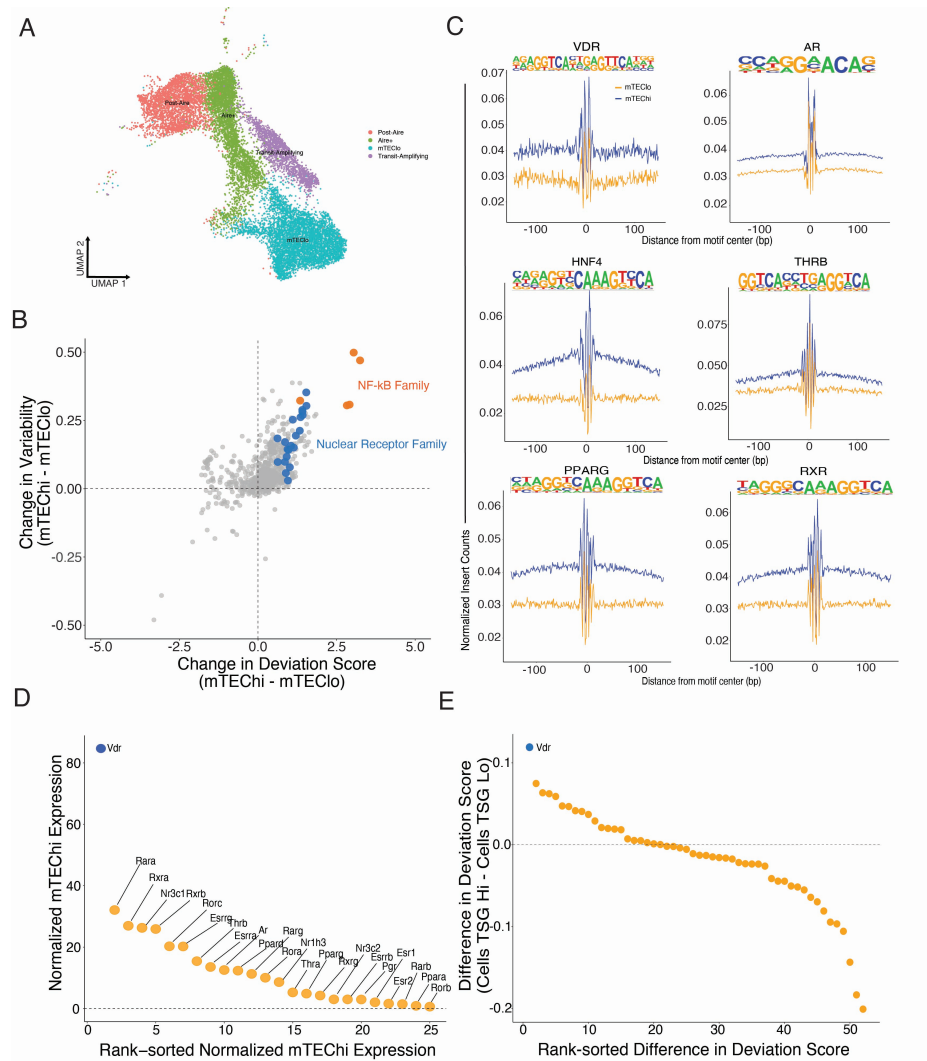


Figure 2.5.2 | Nuclear receptor Vitamin D receptor emerges as a potential regulator of transcriptional heterogeneity: (a) RNA clustered Multiome UMAP with annotated mTEC clusters. (b) Change in motif deviation score variability by change in motif deviation score. Nuclear receptor and NF- κ B families highlighted. (c) Transcription factor footprints of select nuclear receptors at mTEClo and mTEChi stages. (d) Bulk RNA-seq expression of nuclear receptors in mature mTECs. *Vdr* highlighted as the highest expressed nuclear receptor. (e) Change in deviation score of nuclear receptors in Aire+ cells exhibiting high TSG expression versus low TSG expression. VDR highlighted as the TF with the highest deviation score.

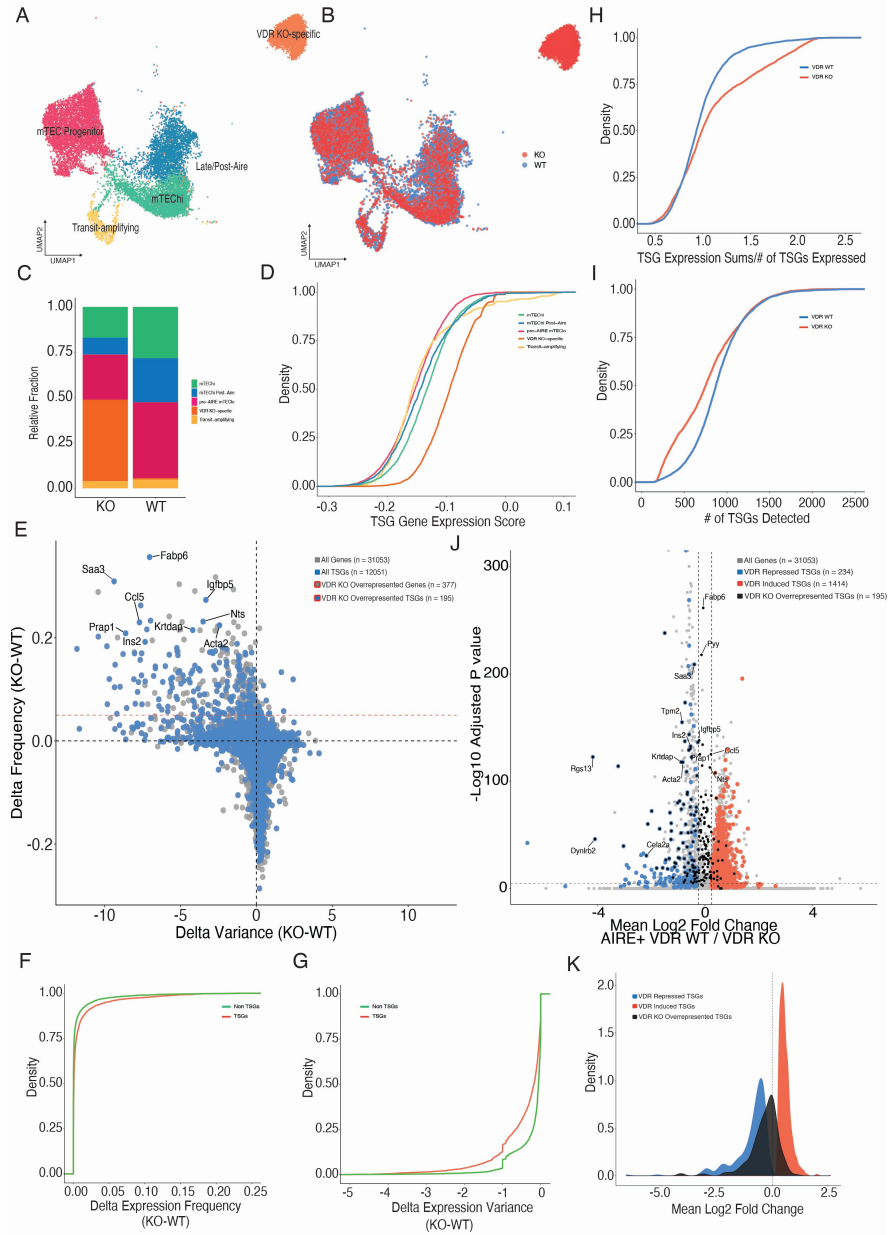


Figure 2.5.3 | VDR regulates transcriptional heterogeneity of tissue-specific gene expression: (a) Single-cell RNA UMAP of *Vdr* WT and KO cells merged. Clusters labeled based on marker gene expression. (b) RNA UMAP split by genotype. (c) Stacked barplot of fraction of each population in samples. Samples split by genotype. (d) CDF of TSG expression score calculated for each identified cluster. (e) Plot of delta expression frequency (KO - WT) and delta expression variance (KO - WT) of all genes in Aire+ cells. Highlighted are particular subsets of genes. (f) CDF of positive changes in frequency (KO - WT), split by TSG and non-TSGs. (g) CDF of change in variance (KO - WT) plotting only losses in variance, split by TSGs and non-TSGs. (h) CDF of TSG expression sums of each TSG normalized by the number of TSGs expressed in *Vdr* WT or KO Aire+ cells. (i) CDF of the number of TSGs detected in *Vdr* WT or KO Aire+ cells. (j) Volcano plot of Log2 fold change of gene expression in *Vdr* KO vs *Vdr* WT Aire+ cells. (k) Distribution of mean log2 fold change of VDR-repressed TSGs, VDR-induced TSGs, and *Vdr* KO overrepresented genes.

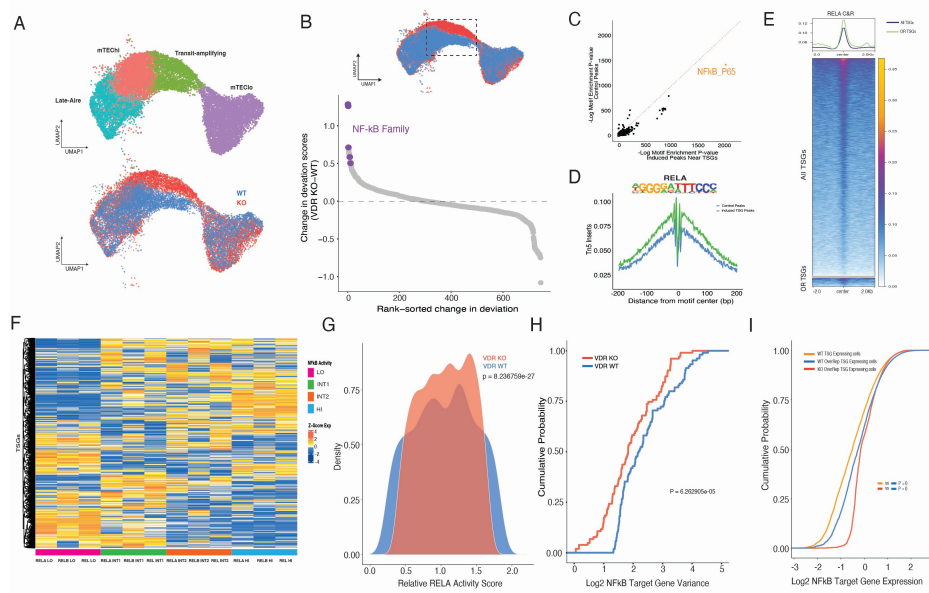


Figure 2.5.4 | VDR controls amplitude and bandwidth of NF- κ B activity: (a) Single-cell ATAC UMAPs of *Vdr* WT and KO mTECs with clusters labeled, and cells split by genotype. (b) Plot of change in motif deviation score between WT and KO cells in the indicated cluster (UMAP above plot). NF- κ B family factors labeled as the highest difference. (c) HOMER -Log motif enrichment of peaks induced from mTECto to mTEChi that neighbor TSGs versus control peaks. NF- κ B motif labeled as the most enriched. (d) Transcription factor footprint at RELA motifs in peaks near TSGs versus control peaks. (e) RELA greenCUT&RUN at peaks near all TSGs and peaks near *Vdr* KO overrepresented genes. (f) Heatmap of row z-score TSG expression in *Aire*+ cells exhibiting quartile binned motif deviation scores for NF- κ B. (g) Mean normalized NF- κ B deviation scores in *Vdr* WT or KO *Aire*+ cells. (h) CDF of Log2 NF- κ B target gene expression in *Aire*+ *Vdr* WT or KO cells. (i) Log2 NF- κ B target gene expression in *Aire*+ WT cells expressing TSGs, WT cells expressing Overrepresented TSGs, and KO cells expressing Overrepresented TSGs.

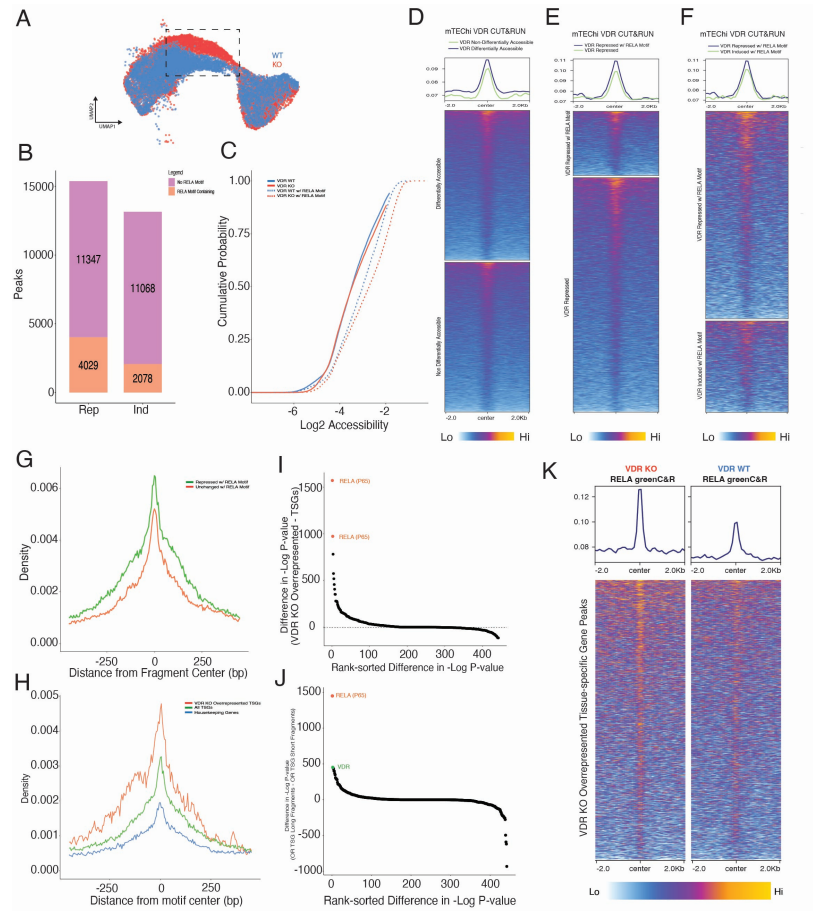


Figure 2.5.5 | VDR restricts access to chromatin regions with high RELA motif density: (a) Single cell UMAP of *Vdr* WT and KO cells split by genotype. (b) Stacked barplot of differentially accessibility peaks (WT vs KO), separated by peaks containing RELA motifs. (c) CDF of log2 accessibility of *Vdr* WT or KO peaks, and those with or without a RELA motif by genotype. (d) Dyad heat map of mTEChi VDR CUT&RUN at differentially accessible peaks vs non-differentially accessible peaks. (e) Dyad heat map of mTEChi VDR CUT&RUN at VDR-repressed peaks and VDR-repressed with a RELA motif. (f) Dyad heat map of mTEChi VDR CUT&RUN at VDR-repressed with a RELA motif and VDR-induced with a RELA motif. (g) Density plot of RELA motif density in scATAC fragments mapping to VDR-repressed regions with a RELA motif versus those unchanged with a RELA motif. (h) Density plot of RELA motif density in scATAC fragments mapping to *Vdr* KO overrepresented TSGs or all TSGs or housekeeping genes. (i) Ranked sorted change in -Log enrichment of motifs in *Vdr* KO overrepresented fragments vs all TSG fragments. (j) Ranked sorted change in -Log enrichment of motifs in long *Vdr* KO overrepresented fragments vs short *Vdr* KO overrepresented fragments. (k) Dyad heat map of *Vdr* WT and KO RELA CUT&RUN at *Vdr* KO overrepresented TSGs.

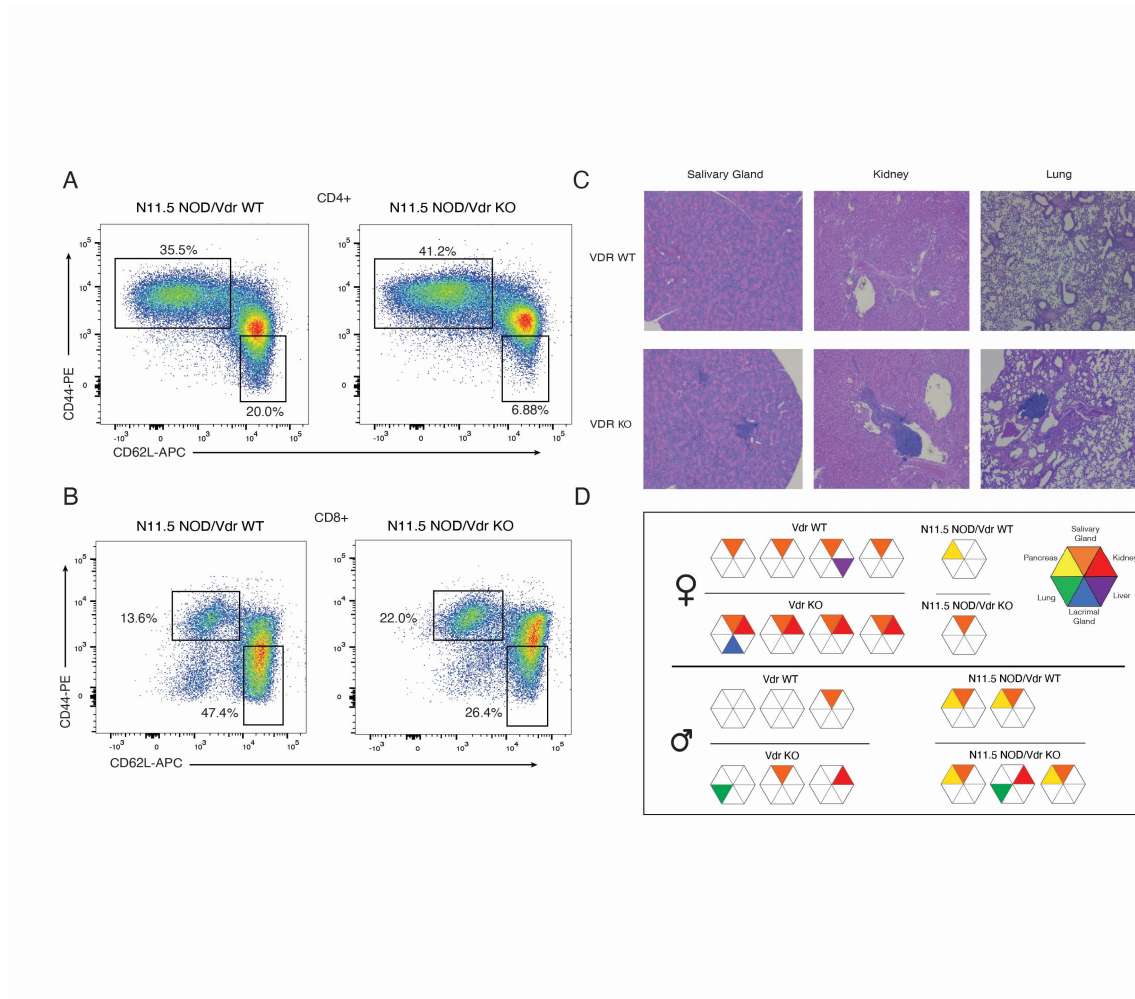


Figure 2.5.6 | *VDR* loss leads to multi-organ autoimmunity: (a) Representative flow cytometry plots of T effector memory (TEM) and naive (TN) compartments of splenic CD4+ T cells from 9-month-old *Vdr* WT and *Vdr* KO sex-matched littermates. (b) Representative flow cytometry plots of T effector memory (TEM) and naive (TN) compartments of splenic CD8+ T cells from 9-month-old *Vdr* WT and *Vdr* KO sex-matched littermates. (c) Representative H&E staining of indicated organs from *Vdr* WT and KO mice. (d) Histological analysis of indicated tissues for infiltrating lymphocytes via hematoxylin and eosin (H&E) staining from 6-12-month-old *Vdr* WT and KO mice. Each hexagon represents an individual mouse.

2.6 Supplemental Figures

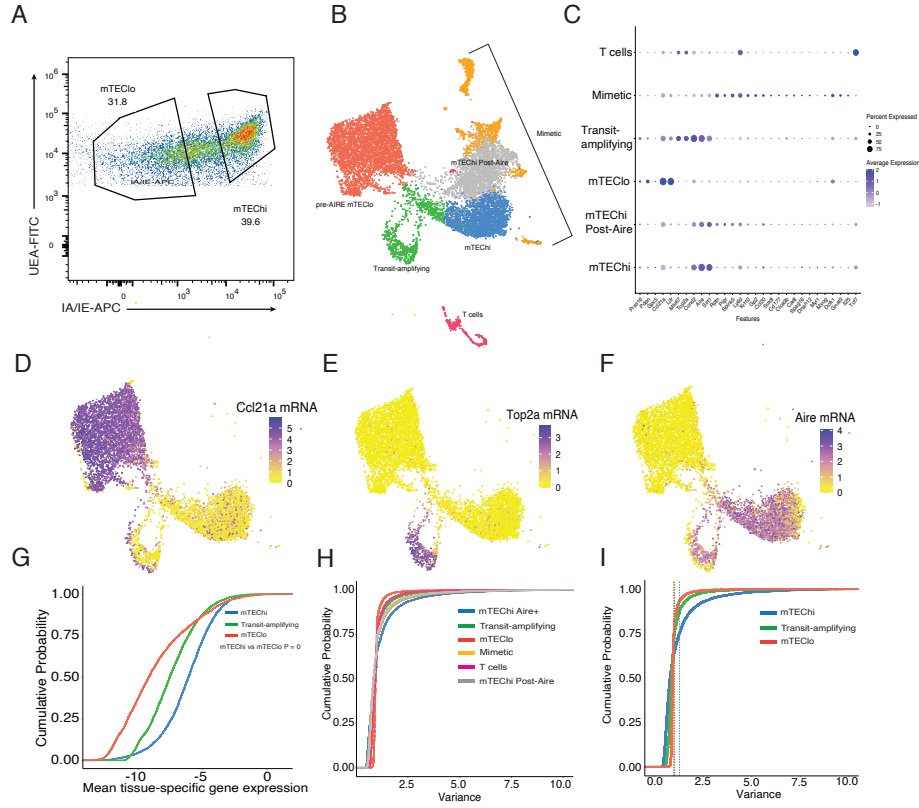


Figure S1 | mTEC scRNA shows defined subsets with distinct changes in transcriptional variance: (a) Flow cytometry of mTECs gated on mTEClo and mTEChi (b) Annotated scRNA-seq clusters with labeled mTEC subpopulations. (c) Dotplot of expression of markers genes in each cluster defined. (d-f) Expression of markers genes on scRNA UMAP. (d) Ccl21a (e) Top2a (f) Aire. (g) CDF of mean TSG expression in mTEClo, Transit-amplifying, and mTEChi. (h) CDF of standardized transcriptional variance for all mTEC clustered identified. (i) Full range of standardized variance for three mTEC populations: mTEClo, transit-amplifying, and mTEChi.

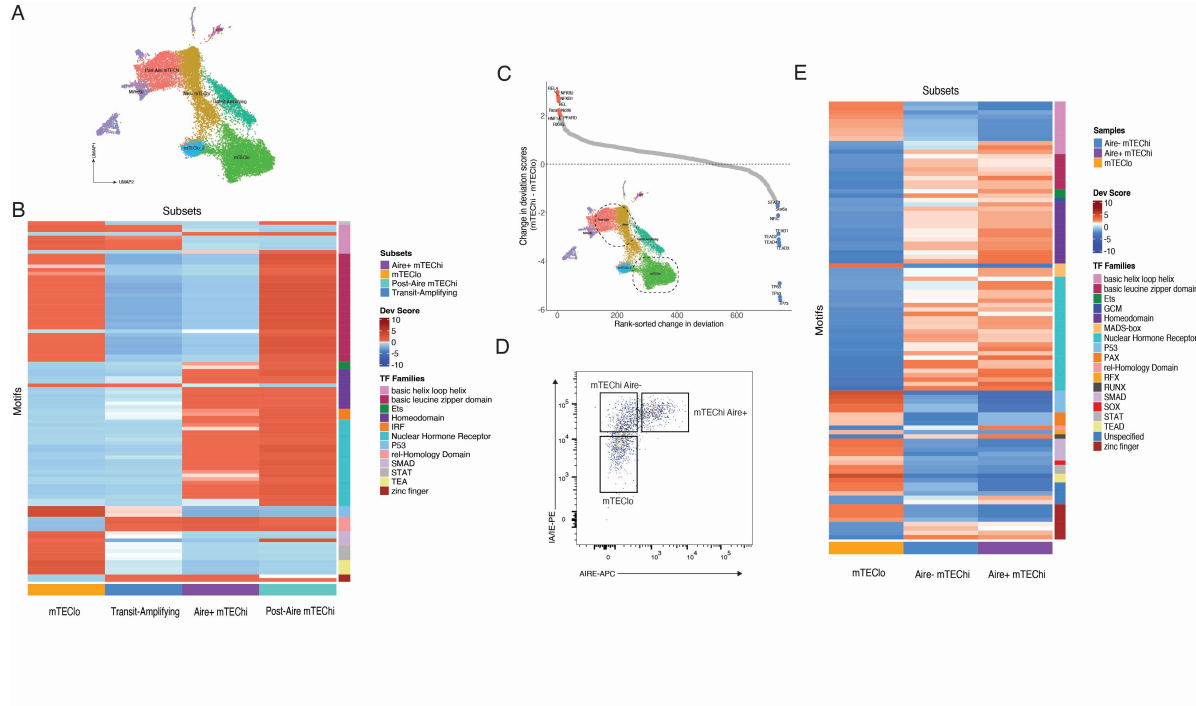


Figure S2 | mTECs exhibit broad changes in transcription factor activity during maturation: (a) mTEC multiome UMAP with clusters labeled based on marker gene expression. (b) Heatmap of top 100 most variable TF motif deviation scores across four subpopulations with TF families labeled. (c) Rank-sorted change in deviation score between mTEChi (*Aire*+ and Post-*Aire*) and mTEClo. Top motifs labeled. (d) Flow cytometry of mTECs gated on mTEClo, AIRE-mTEChi, and AIRE+ mTEChi. (e) Heatmap of top 100 most variable TF motif deviation scores across three subpopulations with TF families labeled.

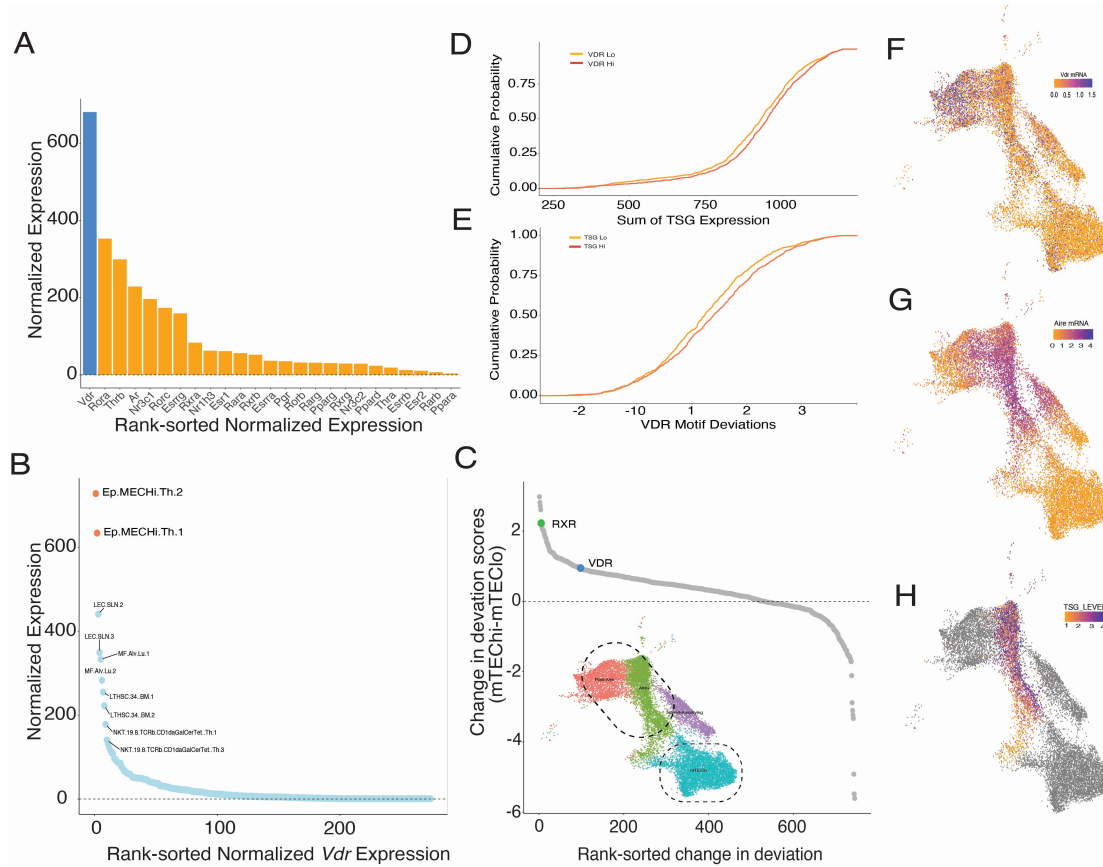


Figure S3 | VDR is highly active in mature mTECs and strongly correlates with TSG expression: (a) mTEChi nuclear receptor expression data from Immgen dataset. *Vdr* highlighted as the most highly expressed. (b) *Vdr* expression in 83 cell types from Immgen dataset. mTEChi highlighted as the cell type with the highest expression. (c) Rank-sorted change in deviation score between mTEChi (*Aire*+ and Post-*Aire*) and mTEClo. VDR and RXR labeled. (d) CDF of sum of TSG expression in cells exhibiting high and low VDR motif deviation scores. (e) CDF of VDR deviation scores in cells exhibiting high TSG expression or low TSG expression. (f) Multiome featureplot of *Vdr* expression (g) Multiome featureplot of *Aire* expression (h) Featureplot of TSG score.

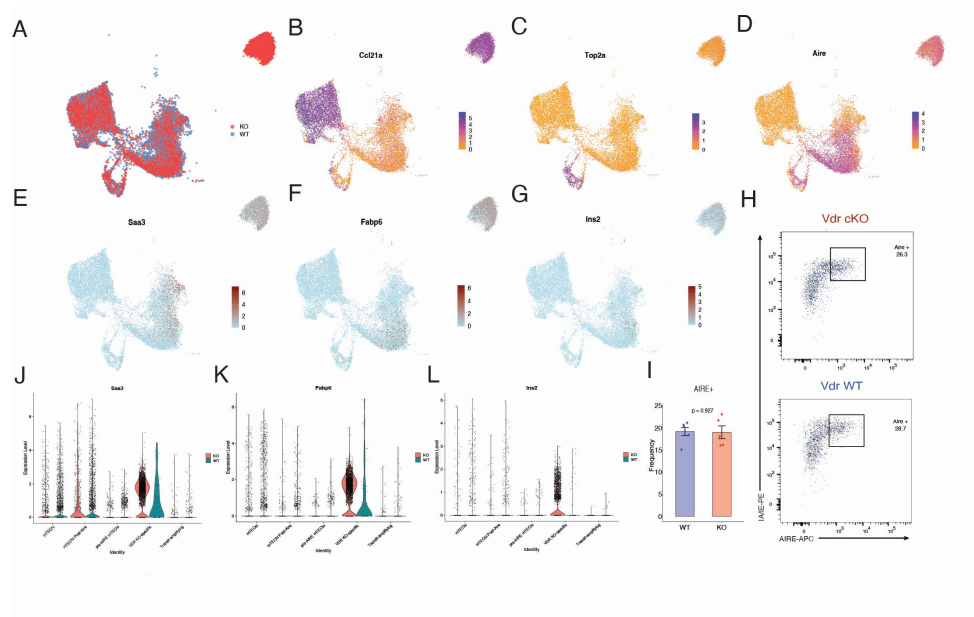


Figure S4 | VDR loss collapses transcriptional heterogeneity prioritizing expression of particular TSGs: (a) scRNA UMAP of *Vdr* WT and KO cells split by genotype. (b-d) scRNA featureplots of stage-specific marker genes. (b) *Ccl21a* expression (c) *Top2a* expression (d) *Aire* expression. (e-g) scRNA featureplots of selected *Vdr* KO Overrepresented TSGs. (e) *Saa3* expression (f) *Fabp6* expression (g) *Ins2* expression (h) Flow cytometry plot of AIRE+ cells in *Vdr* WT and cKO mTECs. (i) Frequency of AIRE+ cells in *Vdr* WT and cKO mTECs. (j-l) scRNA violin plots of overrepresented TSG expression (j) *Saa3* expression (k) *Fabp6* expression (l) *Ins2* expression.

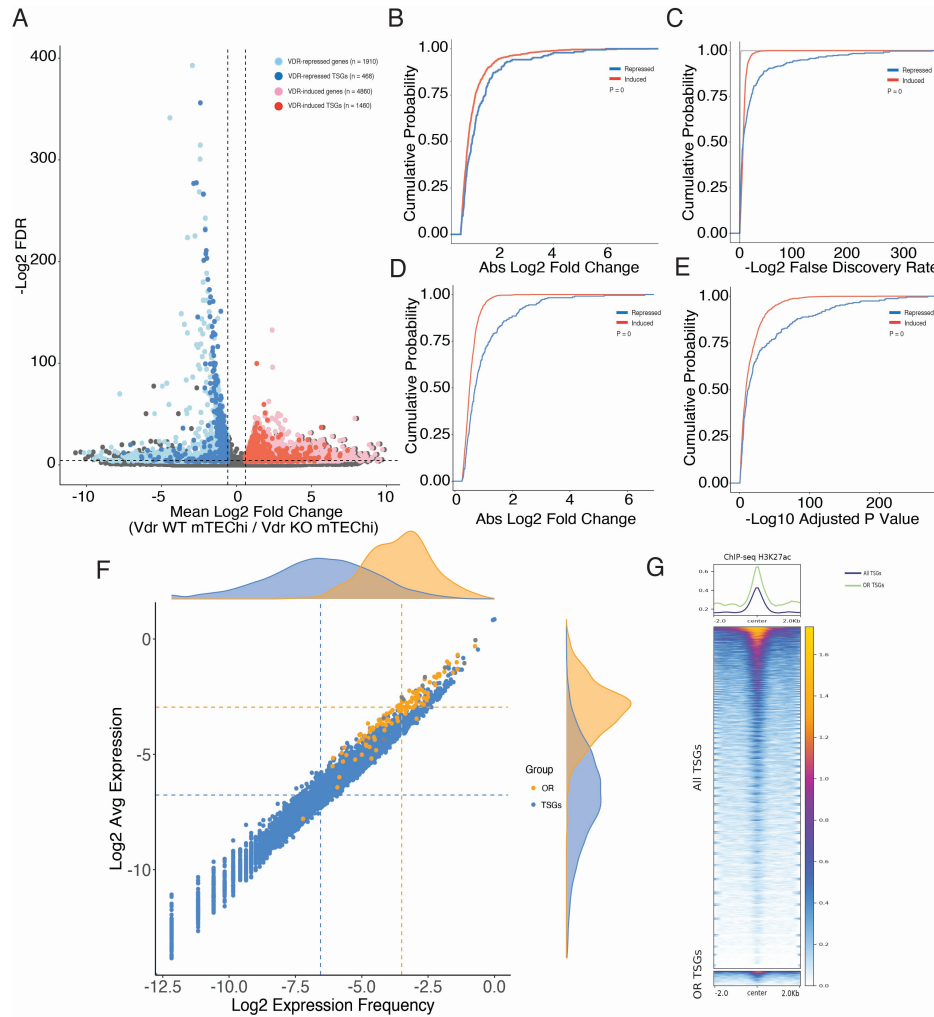


Figure S5 | VDR represses genes that are more highly and frequently expressed: (a) Volcano plot of Bulk RNA-seq *Vdr* WT vs KO with select subset of genes highlighted. (b) CDF of absolute value Log2 fold change TSG expression *Vdr* WT vs KO from bulk RNA-seq. Repressed and induced subsets labeled. (c) CDF of -Log2 false discovery rate of TSG expression *Vdr* WT vs KO from bulk RNA-seq. Repressed and induced subsets labeled. (d) CDF of absolute value Log2 fold change TSG expression *Vdr* WT vs KO from scRNA-seq. Repressed and induced subsets labeled. (e) CDF of -Log2 adjusted P-value of TSG expression *Vdr* WT vs KO from scRNA-seq. Repressed and induced subsets labeled. (f) Scatter plot of log2 mean TSG expression by log2 expression frequency. *Vdr* KO overrepresented genes along with other TSGs labeled. Density plots along axes. (g) Dyad heatmaps of H3K27ac at all TSG peaks or overrepresented TSG peaks.

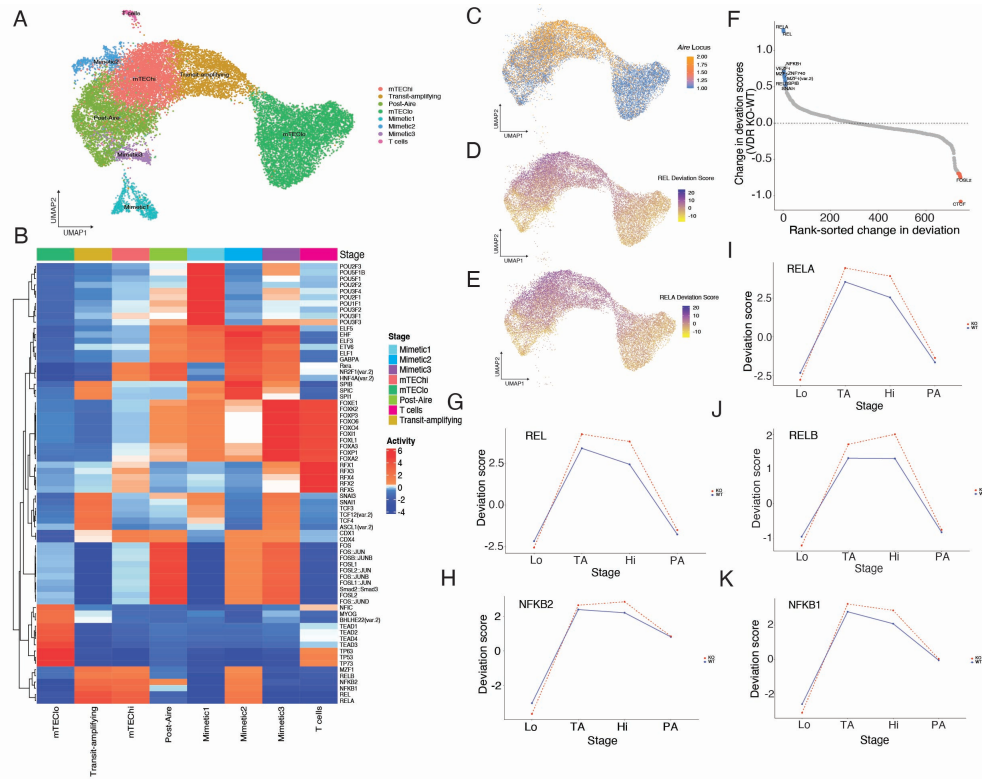


Figure S6 | Defining scATAC clusters and high NF- κ B activity through various stages of mTEC development: (a) Full scATAC-seq UMAP from *Vdr* WT and KO mTECs with all clusters labeled. (b) Heatmap of top variable transcription factor motifs across all clusters called. (c) Accessibility at *Aire* locus (d) Featureplot of REL and (e) RELA motif deviation scores. (f) Rank-sorted change in deviation score of VDR KO vs WT cells with the top and bottom 10 motifs labeled. (g-k) NF- κ B motif deviation score changes over development (g) REL score (h) NFKB2 score (i) RELA score (j) RELB score (k) NFKB1 score.

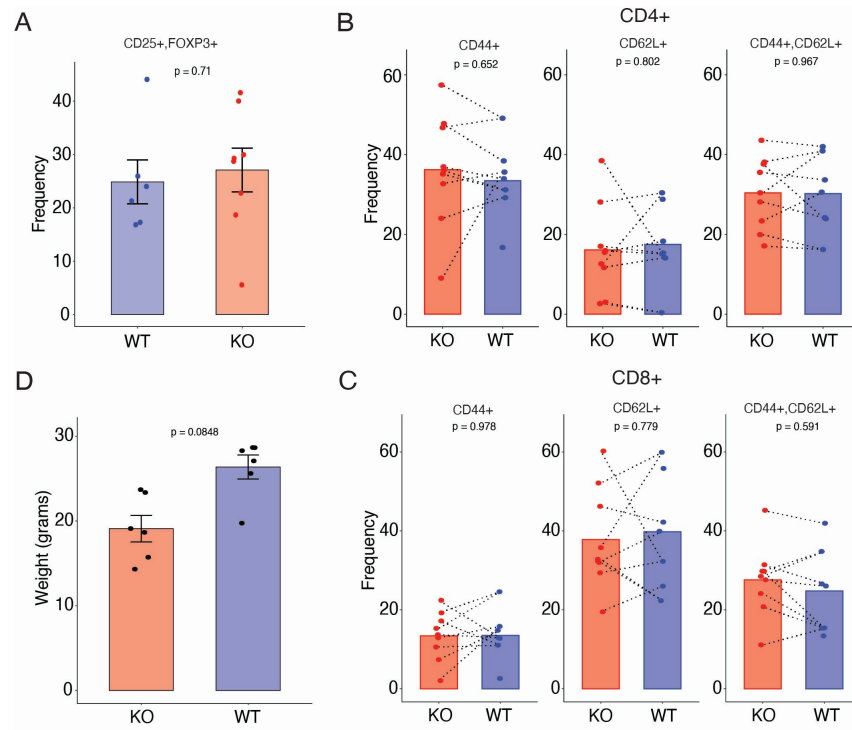


Figure S7 | Effects of thymic *Vdr* deletion on peripheral T cell populations and mouse weights: (a) Frequency of splenic CD4⁺ Tregs in *Vdr* WT and cKO in sex-matched littermates.. (b) Frequency of splenic CD4⁺ T effector memory (TEM), central memory (TCM), and naive (TN) compartments of *Vdr* WT and *Vdr* KO sex-matched littermates.(c) Frequency of splenic CD8⁺ T effector memory (TEM), central memory (TCM), and naive (TN) compartments of *Vdr* WT and *Vdr* KO sex-matched littermates. (d) Weights of 6-9 month *Vdr* WT and *Vdr* KO sex-matched littermates.

CHAPTER 3:

mTECs leverage pleiotropic transcription factors AP-1 and NF- κ B for ectopic gene expression

3.1 Abstract

Medullary thymic epithelial cells (mTECs) establish central tolerance through tissue-specific gene (TSG) expression, a process requiring precise coordination of transcription factor activity. Using multiomic profiling across mTEC development, we identify dynamic and stage-specific regulation of NF- κ B and AP-1 activity. NF- κ B activity peaks early, while AP-1 activity—marked by FOS—is transiently reduced before re-emerging in mature mTECs, coinciding with TSG expression. We show that FOS binding redistributes during development to TSG-associated loci, following prior occupancy by the NF- κ B subunit RELA, which primes chromatin accessibility. Loss of *Rela* disrupts this process, impairing FOS targeting and TSG expression. We further identify a post-transcriptional mechanism regulating this transition: the m6A reader YTHDF2 promotes degradation of *Fos* transcripts, enabling dynamic AP-1 regulation. YTHDF2 deficiency leads to sustained AP-1 activity, defective chromatin remodeling, and impaired FOS redistribution. Functionally, loss of YTHDF2 results in reduced expression of FOS-dependent TSGs and the development of multi-organ autoimmunity. Together, these findings define a coordinated mechanism linking chromatin priming and mRNA decay to transcriptional control of immune tolerance.

3.2 Introduction

Cell states and identities are governed by the transcription factors active at specific .^{79,150}. These factors do not act in isolation, but function within gene regulatory networks composed of interacting transcription factors that can act synergistically or antagonistically. The specific combinations of transcription factors, together with their expression levels, determine which genes are transcribed, ultimately shaping cellular phenotype. Understanding the mechanisms that regulate transcription factor networks across multiple levels is therefore essential for elucidating how cell identity is established, maintained, and dynamically modulated in response to internal and external cues.

Changes in transcription factor (TF) networks drive alterations in cellular identity, a phenomenon widely observed during development and cell state transitions.^{151,152} Interestingly, not all TFs in a network carry equal weight. Some act as master regulators, such as FOXP3 in regulatory T cells or the OCT4/SOX2/NANOG network in pluripotent stem cells, where they coordinate the activity of many downstream factors.^{151,153} Modulating these key TFs can effectively reshape the network, extinguishing one transcriptional program while enabling another, and thereby facilitating transitions between distinct cell states. Such network rewiring can occur through multiple mechanisms. One mechanism underlying these transitions is the transcriptional repression of TFs within an existing network, which allows new regulatory architectures to emerge. For example, BLIMP-1, a transcriptional repressor, promotes the terminal differentiation of B cells into plasma cells by silencing genes required for B cell receptor signaling, germinal center function, and proliferation, while permitting the expression of plasma cell-specific genes such as XBP-1.¹⁵⁴ Clearance of TFs can also occur through post-translational modifications: NF- κ B subunits such as RelA undergo phosphorylation-coupled ubiquitination and proteasomal degradation, providing a mechanism to terminate transcriptional responses and fine-tune signal duration.¹⁵⁵ At the RNA level, the m6A RNA-binding protein YTHDF2 targets transcripts encoding the TF Tal1 for degradation, thereby regulating hematopoietic stem cell differentiation.¹⁵⁶ In addition,

AP-1 factors can act as pioneer factors, establishing accessibility at new enhancers and passively redistributing TFs from previous binding sites to activate distinct gene programs.¹⁵⁷ How these diverse mechanisms—transcriptional repression, protein and RNA clearance, and chromatin remodeling—coordinate to extinguish existing networks and facilitate cell state transitions remains an open question.

Here, we explore how similar principles regulate TF networks over time to drive cell state transitions in mTECs. mTECs arise from immature mTEC_{lo} precursors and progress through defined developmental stages to become mature mTECs, which express thousands of TSGs normally restricted to peripheral tissues.²⁹ This precise temporal regulation of TSG expression is critical for exposing developing thymocytes to a comprehensive representation of self, a prerequisite for effective negative selection.³ Understanding how TF networks are dynamically regulated during mTEC maturation—and how they extinguish one program while promoting another—provides insight into the mechanisms underlying cell state transitions and the establishment of central tolerance.

3.3 Results

3.3.1 *AP-1 and NF- κ B factors exhibit dynamic changes across mTEC development*

Using multiome data, we assessed changes in transcription factor TF activity across four developmental stages of mTECs: mTEC_{lo}, transit-amplifying, Aire+ mTECs, and post-Aire mTECs. TF activity was inferred from chromatin accessibility using deviation scores. Focusing on the 100 most variable TFs across these stages, we found that motifs associated with the NF- κ B and AP-1 families exhibited the most dynamic changes in activity (Figure 3.1a). Notably, both TF families are enriched at accessible chromatin regions proximal to TSGs, suggesting a potential role in regulating promiscuous gene expression in mTECs (Figure 3.1d).

When directly comparing these two TF families, we observed a stage-specific divergence in activity. At the transit-amplifying stage, AP-1 activity was relatively decreased,

while NF- κ B activity increased. This trend shifted at the Aire+ stage, where the activities of both TF families overlapped—coinciding with the stage at which TSG expression is known to occur. This observation is intriguing, as NF- κ B and AP-1 are often described as cooperative regulators in inflammatory contexts, and their binding sites frequently co-occur in composite AP-1:NF- κ B motifs.^{158,159} However, our data suggest a more nuanced, potentially context-dependent relationship during mTEC development.

Interestingly, expression of AP-1 family members did not uniformly mirror their motif deviation scores, indicating that only a subset of AP-1 factors may be transcriptionally active in mTECs. To refine candidate regulators, we examined the correlation between TF expression and motif activity. This analysis identified several AP-1 family members whose expression positively correlated with their inferred activity (Figure 3.1e).

Given that AP-1 proteins typically function as heterodimers—most commonly between FOS and JUN family members^{160,161}—we next assessed co-expression relationships among these candidates. Because AP-1 family members share highly similar DNA-binding motifs, we focused on expression correlations to infer likely dimer pairs. Among the candidates, *Atf3*, *Fos*, and *Junb* exhibited the strongest co-expression (Figure 3.1f). Of these, the FOS:JUNB dimer represents the most likely functional pairing. Supporting this, *Junb* is among the most highly expressed AP-1 genes in mature mTECs, while *Fos* is also strongly expressed and represents a plausible dimerization partner (Figure 3.1g).

Importantly, *Fos* is an immediate early gene whose expression is rapidly induced in response to extracellular signals.¹⁶² As such, its activity provides a useful proxy for real-time AP-1 activation. Using FOS activity as a readout, we revisited the relationship between AP-1 and NF- κ B signaling. Prior work has highlighted a role for the NF- κ B subunit RELA in promoting TSG expression⁵¹, making it a key comparator in this context.

Consistent with the stage-dependent divergence observed earlier, we found an inverse relationship between RELA and FOS activity at the single-cell level. Specifically, cells with high FOS activity tended to exhibit low RELA activity, and vice versa, with the strongest

anti-correlation observed across the dataset (Figure 3.1h). This relationship extended beyond these two factors: cells with elevated FOS activity also showed coordinated decreases in the activity of multiple other TFs associated with NF- κ B signaling (Figure 3.1i).

3.3.2 *FOS redistributes across mTEC development to tissue-specific gene loci*

The enrichment of both NF- κ B and AP-1 motifs at TSGs, despite their inverse activity patterns, suggests that these factors act at shared loci but at distinct developmental stages. To determine when these factors bind, we performed FOS CUT&RUN in mTEC_{lo} and mTEC_{hi} cells. We observed a pronounced redistribution of FOS binding from mTEC_{lo}-associated sites to mTEC_{hi}-specific regions (Figure 3.2a).

This redistribution is accompanied by changes in the transcription factor landscape, as motif enrichment differs between FOS peaks in mTEC_{lo} and mTEC_{hi} cells (Figure 3.2b), suggesting stage-specific cofactor interactions. Consistent with this developmental shift, FOS binding near TSGs is substantially increased in mTEC_{hi} cells, with nearly twice as many TSG-proximal regions bound compared to mTEC_{lo} cells (Figure 3.2c,d).

We next examined motif enrichment in accessible chromatin near TSGs that are bound by FOS versus those that are not. NF- κ B motifs were significantly enriched in FOS-bound regions (Figure 3.2e), supporting the idea that FOS and NF- κ B co-occupy TSG loci. However, when comparing FOS peaks annotated to TSGs with other FOS-bound regions in mTEC_{hi} cells, NF- κ B motif enrichment was reduced, while homeodomain motifs, including DLX and LHX family factors, were more prominent (Figure 3.2f).

Together, these results suggest that FOS can bind both NF- κ B-associated and independent sites. One possible model is that NF- κ B binding precedes and facilitates subsequent recruitment of FOS to a subset of TSG loci.

3.3.3 *RELA facilitates redistribution of FOS binding to mTEC_{hi}*

sites

Given the enrichment of NF- κ B motifs at FOS-bound TSGs, we hypothesized that NF- κ B—specifically the RELA subunit—guides the redistribution of FOS to developmentally appropriate binding sites. To test this, we examined the temporal relationship between RELA and FOS binding.

RELA CUT&RUN revealed strong binding at chromatin regions accessible during the transit-amplifying (TA) stage, with reduced occupancy in mTEChi and post-Aire stages (Figure 3.3a), consistent with NF- κ B motif deviation trends (Figure 3.1b). In contrast, FOS binding decreases from mTEClo to the TA stage (Figure 3.3b), mirroring AP-1 activity dynamics.

To determine whether FOS binds regions previously occupied by RELA, we assessed FOS occupancy in mTEChi cells at sites bound by RELA during the TA stage. FOS binding was significantly enriched at these prior RELA-bound regions compared to sites bound by RELA in mTEChi cells (Figure 3.3c), suggesting a sequential binding model in which RELA precedes FOS at shared loci.

We next tested whether RELA regulates chromatin accessibility during the transition from mTEClo to mTEChi, thereby enabling FOS redistribution. Using inTAC-seq in *Rela* WT and KO cells, we found that RELA loss significantly alters chromatin accessibility at the pre-AIRE mTEChi stage, with a much smaller effect in AIRE⁺ cells (Figure 3.3d,e). These results indicate that RELA primarily functions to establish accessibility prior to full maturation.

We further asked whether loss of RELA disrupts the normal redistribution of FOS by maintaining accessibility at mTEClo-specific binding sites. Indeed, regions corresponding to mTEClo FOS peaks were significantly enriched among sites with increased accessibility in RELA knockout cells at the pre-AIRE stage—regions that would normally lose accessibility during differentiation (Figure 3.3d). Consistent with this, AP-1 motifs were enriched in regions with increased accessibility in RELA-deficient pre-AIRE and AIRE⁺ cells, suggesting persistence of an immature chromatin state and continued FOS binding at inappropriate loci.

Functionally, impaired redistribution of FOS is expected to affect TSG expression. Although chromatin accessibility increases in *Rela* knockout cells, this appears to reflect sustained accessibility at mTEC_{lo}-specific regions rather than appropriate targeting to mTEC_{hi} TSG loci. Supporting this, both NF- κ B and AP-1 motifs were enriched in TSG-associated regions that lose accessibility in the absence of RELA (Figure 3.3h).

Consistent with these findings, prior work from our lab has shown that *Rela* loss compromises the expression of hundreds of TSGs (Kaiser et al., unpublished). To determine whether these genes are associated with FOS activity, we stratified TSGs based on FOS activity using multiome data, defining activity groups by expression in at least 30 percent of cells. RELA-dependent TSGs were depleted from low FOS activity groups and enriched in intermediate and high FOS activity groups (Figure 3.3i), indicating a functional link between RELA-mediated chromatin priming and FOS-dependent gene expression.

Together, these data support a model in which RELA binds chromatin at the pre-AIRE stage to establish accessibility at future mTEC_{hi} binding sites. Following RELA dissociation, FOS is redistributed to these regions, where it contributes to activation of tissue-specific gene expression (Figure 3.4).

3.3.4 *YTHDF2 clears Fos transcriptions to facilitate redistribution of FOS*

Passive redistribution of AP-1 factors to newly accessible binding sites would likely be inefficient if driven solely by RELA-mediated chromatin remodeling. In other systems, such redistribution has been observed to occur gradually when dependent only on changes in chromatin accessibility.^{163,164} However, the sharp reduction in AP-1 motif activity at the transit-amplifying stage (Figure 3.1b) suggests a more active mechanism limiting FOS activity during this transition. This decrease could reflect reduced nuclear abundance of FOS protein, either through protein degradation or transcriptional downregulation. Given that Fos is an immediate early gene characterized by rapid and transient expression dynamics, we investigated whether post-transcriptional regulation contributes to the observed decline

in FOS activity.

mRNA clearance is a key mechanism for rapidly reshaping gene expression programs and can be mediated by RNA-binding proteins, including those that recognize N6-methyladenosine (m6A) modifications.¹⁶⁵ m6A is a widespread and dynamic RNA modification that adds an additional regulatory layer controlling mRNA stability, localization, and translation.¹⁶⁶ Proteins that bind m6A-modified transcripts—referred to as “readers”—can either stabilize transcripts or promote their decay. Among these, the YTH domain family proteins are well-characterized m6A readers, with YTHDF2 specifically implicated in mRNA degradation through recruitment of target transcripts to processing bodies (P-bodies).^{167,168} Notably, YTHDF2 has been shown to facilitate cell state transitions by clearing transcripts encoding transcription factors associated with prior regulatory programs, thereby enabling the establishment of new gene expression networks.^{156,169}

Based on these observations, we hypothesized that FOS activity is regulated through YTHDF2-mediated mRNA decay. To assess this, we first examined expression of YTH domain family members across mTEC development and found that *Ythdf2* is the most highly expressed family member in mature mTECs (Figure 3.5a). This expression pattern is consistent with a potential role in facilitating late-stage transcriptional reprogramming.

To directly test whether YTHDF2 contributes to clearance of AP-1 factors, we performed single-cell RNA-seq on wild-type and YTHDF2 knockout mTECs and identified populations corresponding to distinct developmental stages (Figure 3.5b,c). To identify direct RNA targets of YTHDF2, we complemented this analysis with ARTR-seq¹⁷⁰, which enables transcriptome-wide identification of YTHDF2-bound mRNAs. Integration of these datasets revealed that *Fos* mRNA is a direct target of YTHDF2, as it is enriched in ARTR-seq and exhibits increased mRNA abundance in YTHDF2-deficient cells (Figure 3.5c). These findings indicate that YTHDF2 directly regulates *Fos* transcript stability.

We next asked whether increased *Fos* mRNA in the absence of YTHDF2 translates into elevated AP-1 activity. Single-cell ATAC-seq revealed clear separation between wild-

type and knockout cells (Figure 3.5e,f), indicating widespread chromatin differences. At the TA stage, YTHDF2-deficient cells exhibited increased AP-1 motif activity, consistent with elevated FOS levels (Figure 3.5g). Interestingly, NF- κ B motif activity was also increased in the knockout, despite the absence of direct targeting of NF- κ B pathway components by YTHDF2. This suggests that elevated AP-1 activity or broader disruptions in chromatin state may indirectly influence NF- κ B signaling.

Strikingly, loss of YTHDF2 resulted in an accumulation of immature mTECs (Figure 3.5i), indicating a defect in developmental progression. This finding suggests that clearance of FOS is not only required for its proper redistribution to mTEChi-specific binding sites but is also critical for enabling the transition from mTEClo to more mature states. Notably, the transition from mTEClo to the TA stage corresponds to the point at which AP-1 activity sharply declines, highlighting this stage as a key regulatory checkpoint.

To determine whether YTHDF2-mediated mRNA decay facilitates chromatin remodeling and FOS redistribution, we examined accessibility dynamics between mTEClo and TA stages. In YTHDF2 knockout cells, thousands of regions corresponding to mTEClo-specific FOS binding sites remained aberrantly accessible at the TA stage (Figure 3.5j). These regions would normally lose accessibility during differentiation, suggesting a failure to properly exit the earlier chromatin state.

Taken together, these results support a model in which YTHDF2 targets Fos transcripts for degradation, thereby reducing AP-1 activity at a critical developmental transition. This clearance enables both the redistribution of FOS to mTEChi-specific regulatory regions and the broader chromatin and transcriptional reprogramming required for progression through mTEC development.

3.3.5 *YTHDF2 deletion causes multi-autoimmunity due to loss of AP-1 associated tissue-specific gene expression*

To determine whether loss of YTHDF2 contributes to the development of autoimmunity, we aged *Ythdf2* WT and KO sex-matched littermates for 6–9 months prior to sacrifice and organ collection. This aging period allowed sufficient time for potential defects in immune tolerance to manifest. Splenocytes were isolated and analyzed by flow cytometry to assess T cell activation states, including naïve, central memory, and activated populations. We observed a significant increase in the frequency of activated CD4⁺ T cells (CD44^{hi} CD62L^{lo}) in *Ythdf2* KO mice relative to WT controls (Figure 3.6a), consistent with systemic immune activation.

To determine whether this activation was associated with autoimmune pathology, we performed histological analysis of peripheral tissues. *Ythdf2* KO mice exhibited lymphocytic infiltration in multiple organs, including the lung, salivary gland, and kidney (Figure 3.6b), indicating a breakdown in immune tolerance and the development of multi-organ autoimmunity.

Given the central role of mTECs in establishing tolerance through TSG expression, we next asked whether loss of YTHDF2 disrupts TSG expression programs. Using scRNA-seq data, we subsetted for *Aire*⁺ mTECs and quantified the frequency of TSG expression in *Ythdf2* WT and KO cells. We observed both gains and losses of TSG expression in *Ythdf2* KO cells; however, there was an overall bias toward loss, with more TSGs showing reduced expression frequency than increased expression (Figure 3.6c). This suggests that YTHDF2 is broadly required to maintain normal TSG expression patterns.

We next sought to investigate the mechanism underlying this defect, focusing on the regulation of AP-1 activity, particularly FOS, which we hypothesized must be dynamically cleared and redistributed to enable TSG expression at appropriate loci. To test this, we leveraged multiome data to stratify mature mTECs into quartiles based on inferred FOS activity levels. Within each quartile, TSGs were assigned if they were expressed in at least

30 percent of cells, allowing us to link TSG expression patterns to FOS activity states.

This analysis revealed that TSGs associated with high FOS activity were preferentially reduced in *Ythdf2* KO cells (Figure 3.6d). In contrast, TSGs associated with lower FOS activity showed less pronounced changes. These results indicate that a subset of TSGs is particularly dependent on high AP-1 activity for expression and that this program is disrupted in the absence of YTHDF2.

Taken together, these findings support a model in which YTHDF2 is required to properly regulate AP-1 dynamics, enabling the redistribution of FOS to TSG loci in mature mTECs. In its absence, this redistribution is impaired, leading to reduced expression of FOS-dependent TSGs, defective central tolerance, and the emergence of multi-organ autoimmunity.

3.4 Discussion

Here we define a coordinated regulatory mechanism linking chromatin priming, transcription factor dynamics, and post-transcriptional control in the establishment of central immune tolerance. Our results support a model in which NF- κ B and AP-1 transcription factor families act in a temporally ordered sequence to regulate tissue-specific gene TSG expression during mTEC development.

A central finding is that NF- κ B, through the RELA subunit, binds early in development to establish accessible chromatin at loci that will later become active in mature mTECs. This primed chromatin landscape is subsequently utilized by AP-1, whose activity—marked by FOS—is dynamically redistributed during maturation. We show that FOS binding shifts from immature to mature regulatory regions and becomes enriched at TSG-associated loci in mature mTECs. Importantly, this redistribution follows prior RELA occupancy, supporting a sequential model in which NF- κ B establishes regulatory potential that is later executed by AP-1.

We further demonstrate that AP-1 activity is not only controlled at the level of

chromatin but also regulated post-transcriptionally. YTHDF2 emerges as a key regulator of this process by promoting degradation of *Fos* transcripts. This transient clearance of *Fos* mRNA appears necessary to properly reset AP-1 activity during the transit-amplifying stage. In the absence of YTHDF2, sustained *Fos* expression leads to elevated AP-1 activity, failure to appropriately remodel chromatin states, and impaired redistribution of FOS to mature regulatory elements.

Functionally, disruption of this axis results in defective TSG expression and breakdown of immune tolerance. YTHDF2-deficient mice exhibit increased T cell activation and multi-organ lymphocytic infiltration, consistent with systemic autoimmunity. These findings connect molecular-level dysregulation of transcription factor dynamics to organism-level failure of central tolerance.

Together, our data support a multilayered regulatory framework in which NF- κ B-mediated chromatin priming, YTHDF2-dependent RNA decay, and AP-1 redistribution are integrated to control the timing and specificity of TSG expression. This highlights the importance of temporal coordination across epigenetic, transcriptional, and post-transcriptional layers in immune tolerance.

Future work should determine how YTHDF2 activity is regulated during mTEC development and whether additional RNA-binding proteins contribute to shaping AP-1 dynamics. It will also be important to further dissect how NF- κ B and AP-1 cooperate at individual TSG loci and whether this regulatory logic extends to other systems involving promiscuous gene expression.

3.5 Figures

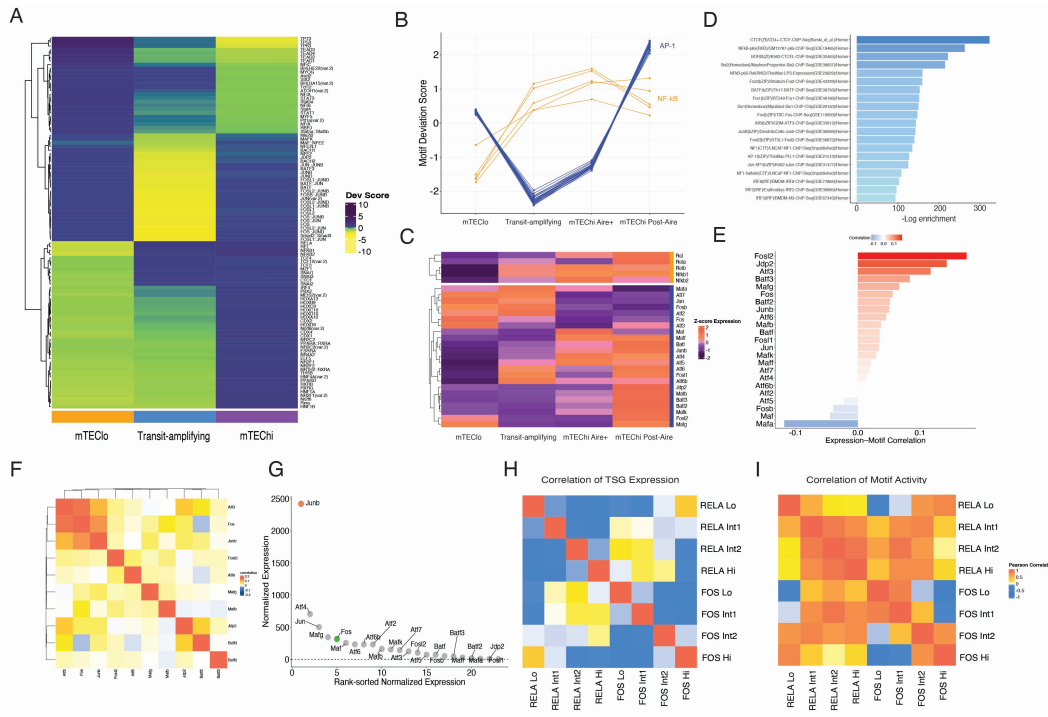


Figure 3.1 | Dynamic and inverse activity of NF- κ B and AP-1 factors across mTEC development: (a) Heatmaps depicting top 100 most variable TFs across three mTEC populations—multiome data. (b) Motif deviation scores of NF- κ B and AP-1 factor motifs across mTEC stages. (c) Z-score expression of NF- κ B and AP-1 subunits across mTEC development. (d) HOMER enrichment of transcription factor motifs near TSGs. (e) Ranked correlation of AP-1 factor expression and motif deviation score. (f) Correlation of expression of AP-1 factors. (g) Bulk RNA-seq expression of AP-1 factors. *Fos* and *Junb* highlighted. (h) Correlation of TSG expression in cells exhibiting four levels of RELA activity and FOS activity. Cells binned by deviation score quartiles. (i) Correlation of TF activity in cells exhibiting four levels of RELA activity and FOS activity. Cells binned by deviation score quartiles.

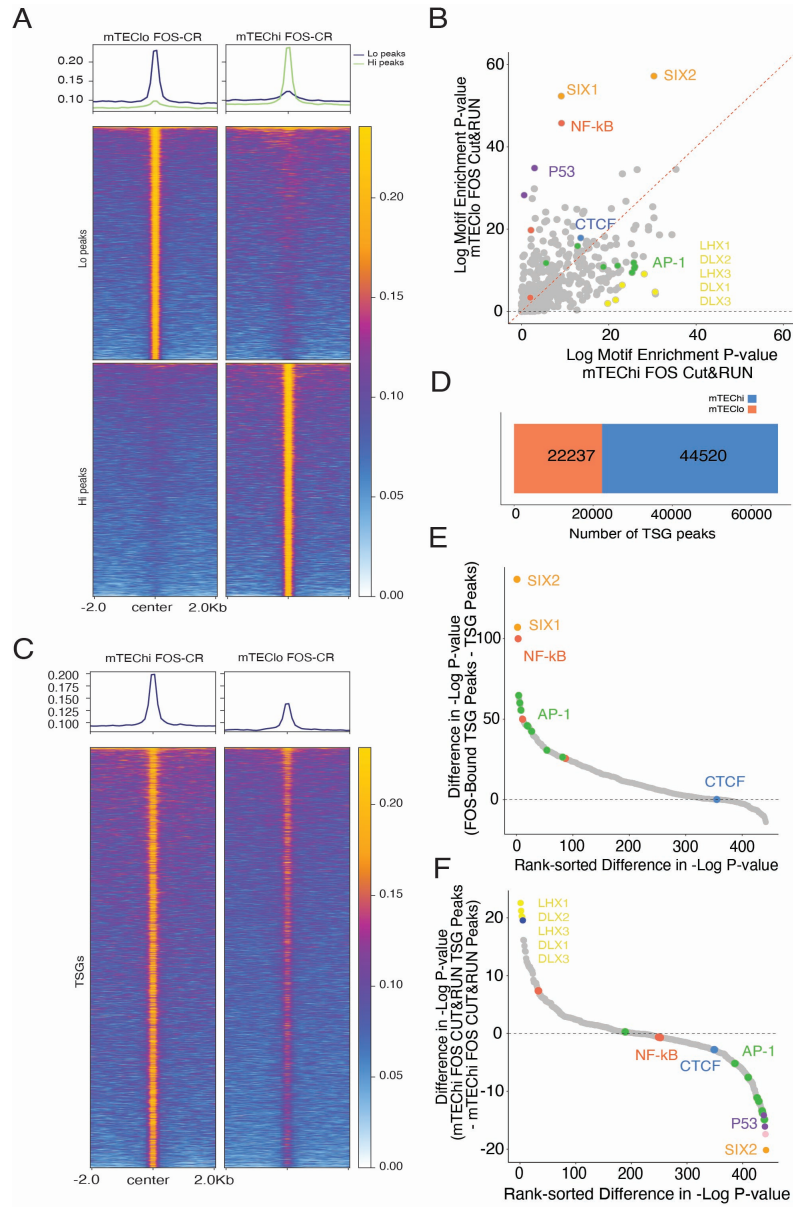


Figure 3.2 | *FOS* redistributes to regions enriched for *NF-κB* motifs: (a) Dyad heatmaps of mTECclo and mTEChi FOS CUT&RUN at peaks called for each sample. (b) Motifs enriched at mTECclo and mTEChi FOS peaks. Peaks were size-matched for comparison. (c) Dyad heatmaps of mTECclo and mTEChi FOS CUT&RUN at TSG peaks. All FOS peaks were merged and annotated for TSGs. (d) Stacked barplot of the number of TSG peaks contributed by each sample. (e) Rank-sorted change in HOMER motif enrichment of ATAC TSG peaks that intersect with mTEChi peaks versus other TSG peaks. (f) Rank-sorted change in HOMER motif enrichment of mTEChi FOS CUT&RUN TSG peaks versus other mTEChi FOS CUT&RUN peaks.

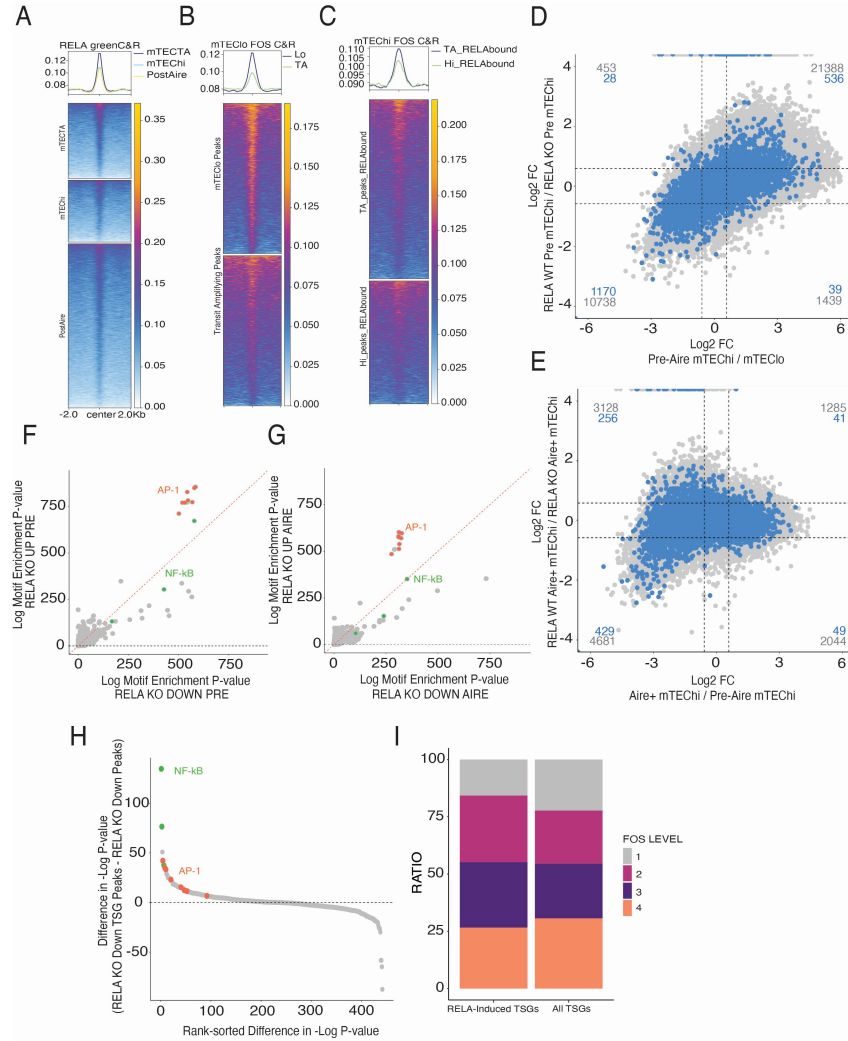


Figure 3.3 | *RELA*-mediated chromatin accessibility changes facilitate *FOS* redistribution: (a) Dyad heatmaps of *RELA* greenCUT&RUN at peaks specific to Transit-amplifying, AIRE+ mTEChi, and Post-AIRE stages. (b) Dyad heatmap of mTEClo *FOS* CUT&RUN at peaks specific to mTEClo and Transit-amplifying stages. (c) Dyad heatmap of mTEChi *FOS* CUT&RUN at peaks specific and Transit-amplifying and AIRE+ mTEChi stages that intersect with *RELA* greenCUT&RUN peaks. (d) Scatter plot of pre-AIRE *Rela* WT and KO inTAC-seq accessibility changes by changes from mTEClo to pre-Aire. Blue points highlighted are peaks that intersect with mTEClo *FOS* CUT&RUN. (e) Scatter plot of AIRE+ *Rela* WT and KO inTAC-seq accessibility changes by changes from pre-AIRE to AIRE+. Blue points highlighted are peaks that intersect with mTEClo *FOS* CUT&RUN. (f) HOMER motifs enriched at pre-AIRE *RELA*-repressed peaks vs pre-Aire *RELA*-induced peaks. Peaks were size-matched for comparison. (g) HOMER motifs enriched at Aire+ *RELA*-repressed peaks vs AIRE+ *RELA*-induced peaks. Peaks were size-matched for comparison. (h) Rank-sorted change in HOMER motif enrichment of AIRE+ *RELA*-induced TSG peaks versus other AIRE+ *RELA*-induced peaks. (i) Stacked barplot of TSGs assigned to cells exhibiting four levels of *FOS* activity, binned by quartiles. *RELA*-induced TSGs and all TSGs considered.

A

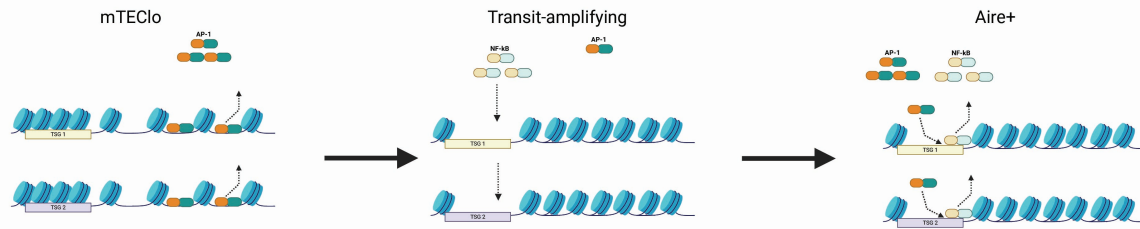


Figure 3.4 | Model of AP-1 and NF- κ B activity at TSG loci: (a) Model depicting AP-1 redistribution to TSG loci, facilitated by NF- κ B-mediated changes to chromatin landscape.

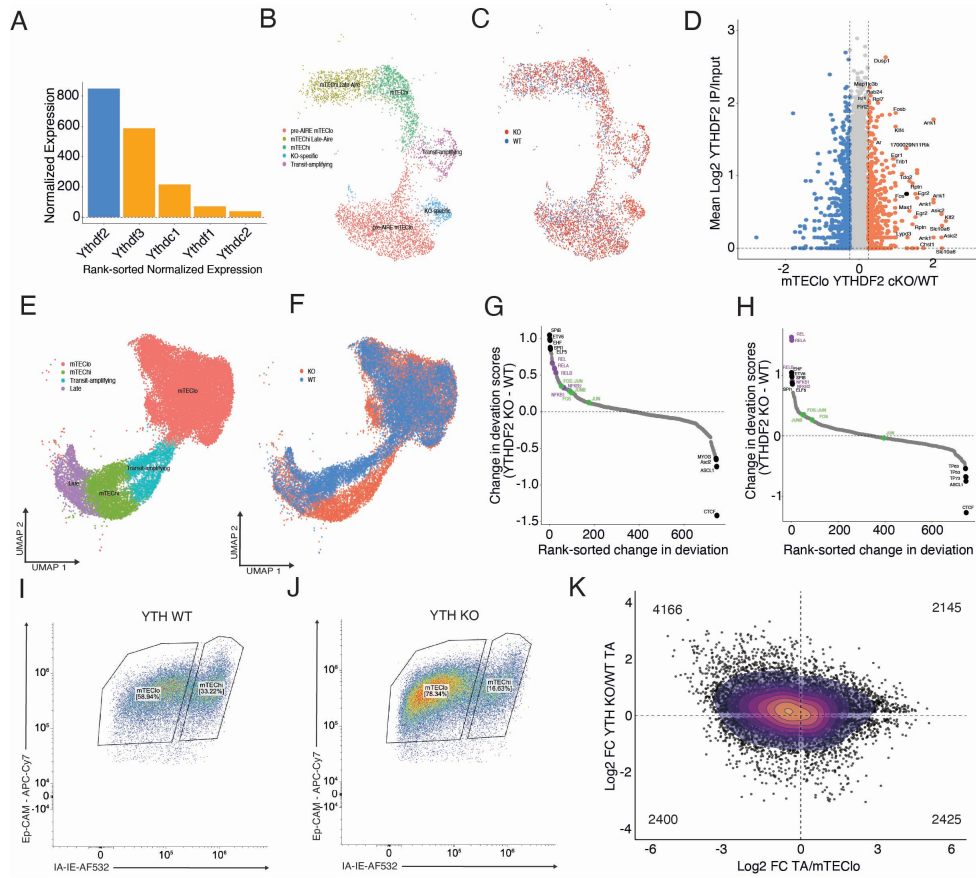


Figure 3.5 | *YTHDF2* loss compromises *Fos* clearance and redistribution: (a) Expression of all YTH factors in mature mTECs, data from Immgen. (b) scRNA UMAP of *Ythdf2* WT and KO mTECs labeled my cluster. (c) scRNA UMAP of *Ythdf2* WT and KO mTECs split by genotype. (d) ARTR-seq determined YTHDF2 binding of mRNA by log2 fold change of mRNA expression in *Ythdf2* WT vs KO mTEClo. (e) scATAC UMAP of *Ythdf2* WT vs KO mTECs with labels. (f) scATAC UMAP of *Ythdf2* WT vs KO mTECs split by genotype. (g) Rank-sorted changes in deviation score in Transit-amplifying mTECs (KO - WT). NF- κ B and AP-1 factors highlighted along with top and bottom 5. (h) Rank-sorted changes in deviation score in mature mTECs (KO - WT). NF- κ B and AP-1 factors highlighted along with top and bottom 5. (i) Flow cytometry of mTECs gated on mTEClo and mTEChi in *Ythdf2* WT and KO. (j) Scatter plot of accessibility changes in *Ythdf2* wt and KO transit-amplifying cells by changes from mTEClo to transit-amplifying cells. All peaks intersect with mTEClo FOS CUT&RUN.

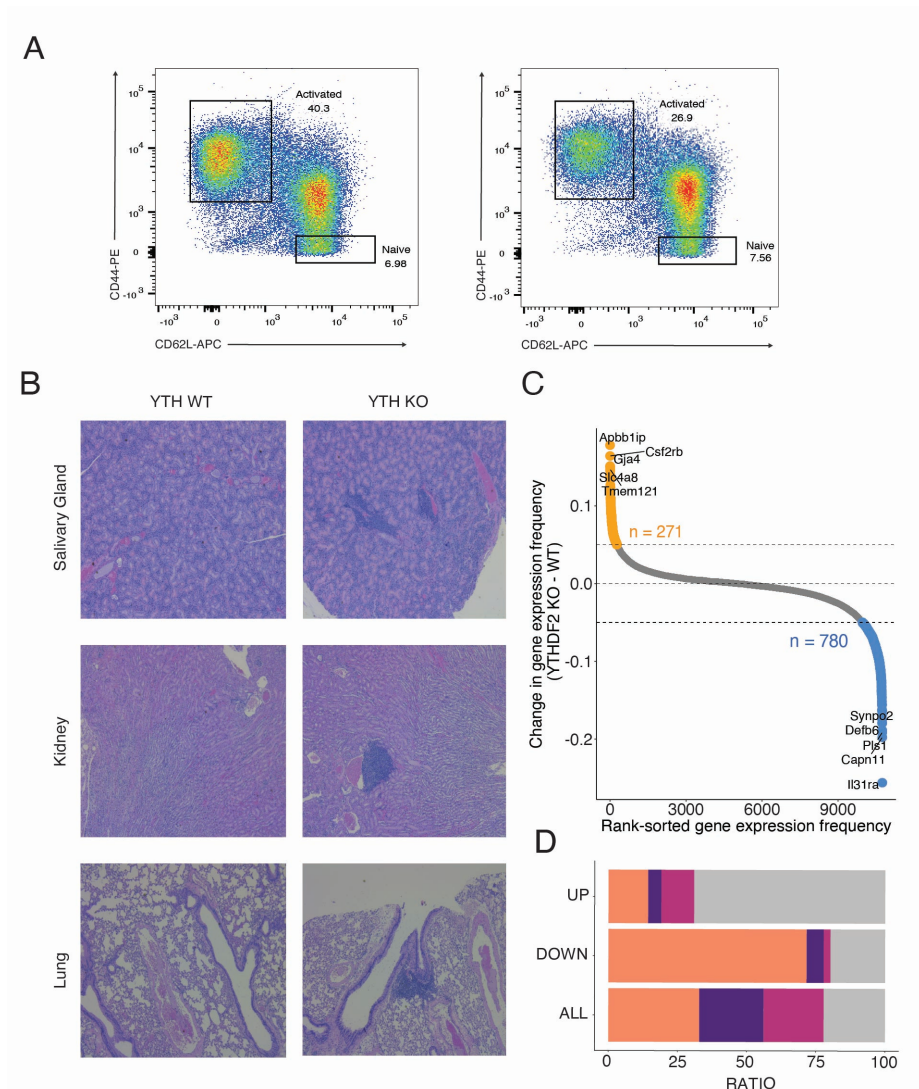


Figure 3.6 | *YTHDF2* loss compromises TSG expression and lead to multi-organ autoimmunity: (a) Representative flow cytometry plots of T effector memory (TEM) and naive (TN) compartments of splenic CD4⁺ T cells from 9-month-old sex-matched littermates. (b) Representative HE staining of indicated organs from *Ythdf2* WT and KO mice. (c) Ranked sorted change in TSG expression frequency (KO - WT) with genes with difference ≥ 0.05 labeled in orange, and ≤ -0.05 in blue. Top and bottom 5 genes highlighted. (d) Stacked barplot of association of differentially expressed genes with cells exhibiting four levels of FOS activity, binned by quartiles. Each gene was assigned to a quartile if it was expressed at a frequency of 30 percent or greater.

CHAPTER 4:

Materials and Methods

Mice

Mice used in this study were housed in pathogen-free facilities at the University of Chicago. All experiments and animal use procedures were conducted in compliance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Chicago. $Vdr^{flx/flx}$ mice were provided by Ron Evans (Salk Institute) and were bred with B6- $Foxn1^{Cre}$ homozygous mice purchased from Jackson Laboratories to generate $Vdr^{fl/fl}; Foxn1^{Cre -/-}$ and $Vdr^{fl/fl}; Foxn1^{Cre +/-}$ littermates. mVenus-RelA^{+/+} (provided by Alexander Hoffmann, UCLA) and $Vdr^{fl/fl}; Foxn1^{Cre +/-}$ mice were crossed to generate $Vdr^{fl/fl}; Foxn1^{Cre +/-}; mVenus-RelA^{+/+}$ mice. NOD/ShiLtJ and $Vdr^{fl/fl}; Foxn1^{Cre +/-}$ mice were crossed and backcrossed to NOD/ShiLtJ to generate N8-N11.5 progeny. NOD/ShiLtJ and C57BL/6J mice were purchased from Jackson Laboratories. $Ythdf2^{flx/flx}$ mice were provided by Chuan He (University of Chicago) and were bred with B6- $Foxn1^{Cre}$ homozygous mice purchased from Jackson Laboratories to generate $Ythdf2^{fl/fl}; Foxn1^{Cre -/-}$ and $Ythdf2^{fl/fl}; Foxn1^{Cre +/-}$ mice. Sex-matched littermates were used for all comparisons of genetic perturbations.

Isolation, sorting and analysis of murine medullary thymic epithelial cells

Thymi from 4-5 week-old mice were harvested and connective tissue was removed. Stromal tissue was perforated using scissors and incubated with rotation in DMEM-F12 (Gibco) at room temperature for 10 min to liberate thymocytes. Remaining stromal tissue was enzymatically digested (0.5 mg/mL Collagenase D (MilliporeSigma), 0.2 mg/mL DNaseI (MilliporeSigma), 0.5 mg/mL Papain (Worthington Biochemical). Cells were stained with anti-EpCAM antibodies conjugated to APC-Cy7 (BioLegend, clone 9C4) and EpCAM⁺ cells

were enriched via positive selection using magnetic anti-Cy7 beads (Miltenyi). Enriched fraction was stained with fluorochrome-conjugated antibodies (BioLegend) to CD45 (30-F11), Ly-51 (6C3), MHC-II I-A/I-E (M5/114.15.2) and EpCAM (G8.8), along with fluorescein-labeled UEA-I (Vector Labs) and DAPI (Invitrogen). Intracellular staining for AIRE (eBioscience, clone 5H12) was performed using the eBioscience FoxP3 transcription factor staining kit (Invitrogen) according to manufacturer’s instructions. Cell sorting was performed using FACS Symphony S6, FACS Aria Fusion, or FACS Aria II equipped with a 100 m nozzle (BD Biosciences). Flow cytometry data were acquired using BD LSR II or Fortessa and analyzed using FlowJo (v10).

Flow cytometry of splenocytes

Spleens from 4 week-old to 12 month-old mice were isolated in RPMI (Gibco) supplemented with 10% FCS. Cells were liberated by mincing with syringe plunger and filtered through 40 um strainer. Following red blood cell lysis (BD PharmLyse), cells were stained with fluorochrome-conjugated antibodies specific for mouse CD4 (GK1.5), CD8 (53-6.7), CD25 (PC61), CD44 (IM7), CD69 (H1.2F3), CD62L (MEL-14) and TCR (H57-597). Intracellular staining for FoxP3 (eBioscience, clone FJK-16s) was performed using the eBioscience FoxP3 transcription factor staining kit (Invitrogen) according to manufacturer’s instructions. Flow cytometry data were acquired using BD LSR II or Fortessa and analyzed using FlowJo (v10).

inTAC-seq

InTAC-seq sample preparation and sequencing

InTAC-seq was performed as described previously¹³⁷ with minor modifications. After extracellular staining, cells were incubated in 1.5% paraformaldehyde (Pierce 28908) for 2 min at room temperature. Cells were washed in 1X Perm buffer (eBioscience 00-5523-00) and stained with Cells were incubated with 1:500 diluted AIRE (eBioscience, clone 5H12) antibody in 1X Perm buffer for 30 min at room temperature, then washed with 2%

FCS/PBS. 60,000 cells were sorted into 2% FCS/PBS. Cells were pelleted then resuspended in the transposition mix (1× TD buffer, 5 L transposase (Illumina), 0.01% Tween-20), and reaction was carried out for 30 min at 37° C. DNA was reverse-crosslinked by adding 1X DEDA (Sigma-Aldrich D158003), 0.1% SDS, 0.5 uL Proteinase K (NEB P8107S) at 37° C. overnight. DNA was purified using Zymo DNA Clean & Concentrator kit. Libraries were constructed as described previously⁷⁰ and sequenced by paired-end, dual-index sequencing using the Element AVITI.

Bulk RNA-seq sample preparation

75,000 primary mTECs were FACS-sorted directly into RULT lysis buffer (Qiagen RNEasy UCP Micro Kit) and total RNA was extracted using manufacturer’s instructions. mRNA was enriched and RNA-seq libraries were constructed using the Illumina TruSeq Stranded mRNA Kit. Paired-end, dual-index sequencing was performed on the Illumina NovaSeq 6000 platform.

Bulk RNA-seq data processing

RNA-seq reads were mapped to the mm10 mouse genome assembly using *TopHat* (v2.1.1) with setting –microexon-search. Unmapped, unpaired and low quality reads (MAPQ 5) were removed using *samtools* (v1.9) view with settings -q 5 -f 2. Paired reads were counted for each gene using featureCounts from *Subread* (v2.0.1). ‘Transcripts-per-million’ (TPM) values were calculated for each gene to quantify the relative abundance of transcripts for clustering analysis. “Trimmed mean of M values” (TMM) was calculated for each gene for differential comparisons across samples using *edgeR* (v4.0.2) (calcNormFactors()). Common dispersions were estimated using estimateCommonDisp() and Benjamini-Hochberg false discovery rates (FDRs) were calculated for pairwise comparisons using exactTest(). Genes exhibiting FDR 0.05 were regarded as significant.

scATAC-seq library preparation

Single-cell ATAC-seq libraries were generated from freshly isolated thymic epithelial cells (TECs) obtained from wild-type (WT) and *Vdr*- or *Ythdf2*-knockout (KO) mice. Nuclei were isolated following the 10x Genomics Chromium Single Cell ATAC protocol (version X.X). Nuclei were washed, counted, and assessed for integrity using trypan blue exclusion and fluorescence microscopy. For each sample, approximately 50,000 nuclei were loaded into a single lane of a Chromium Next GEM Chip H, where individual nuclei were encapsulated into Gel Bead-in-Emulsions (GEMs). Within each GEM, nuclei underwent Tn5-mediated tagmentation, fragmenting accessible chromatin and appending adapters carrying unique barcodes. After tagmentation, GEMs were broken, and barcoded DNA fragments were purified, PCR-amplified, and indexed to generate sequencing-ready libraries. Library fragment size and quality were assessed using an Agilent Bioanalyzer High Sensitivity DNA kit, and concentrations were quantified with a Qubit fluorometer (Thermo Fisher Scientific). Libraries were sequenced on an Illumina NovaSeq 6000 (or NextSeq 2000) using paired-end 50 bp reads, targeting an average depth of 25,000–50,000 fragments per nucleus. Raw base call files were processed with the Cell Ranger ATAC pipeline (10x Genomics, version X.X) using the `cellranger-atac count` function and aligned to the mm10 reference genome. Feature-barcode matrices from WT and KO samples were merged prior to downstream processing in R using the Signac package (vX.X.X). Data were processed using default Signac workflows, including quality control, normalization, dimensional reduction, and peak calling.

scRNA-seq library preparation

Single-cell suspensions were prepared from freshly isolated thymic epithelial cells (TECs) obtained from wild-type and *Vdr*- or *Ythdf2*- knockout (KO) mice. Cells were enriched by enzymatic digestion and filtration to obtain a viable single-cell suspension (>90 percent viability, determined by trypan blue exclusion). Cell concentration was adjusted to approximately 1,000 cells/μL in PBS containing 0.04 percent BSA. Libraries were prepared using

the Chromium Next GEM Single Cell 3 Kit (10x Genomics, version X.X) according to the manufacturer’s protocol. Briefly, single cells were encapsulated into nanoliter-scale Gel Bead-in-Emulsions (GEMs) using the Chromium Controller, where cell lysis and barcoded reverse transcription of mRNA occurred within individual droplets. After cDNA synthesis, emulsions were broken, and barcoded cDNA was purified with silane magnetic beads. Amplified cDNA was then fragmented, end-repaired, A-tailed, and ligated with Illumina sequencing adapters. Libraries were indexed via PCR and quantified using a Bioanalyzer (Agilent) and Qubit fluorometer (Thermo Fisher Scientific) prior to sequencing. Libraries were sequenced on an Illumina NovaSeq 6000 (or NextSeq 2000) using paired-end 150 bp reads with a targeted depth of approximately 50,000–100,000 reads per cell. Raw base call files were processed with the Cell Ranger pipeline (10x Genomics, version X.X) using the cellranger count function and the mm10 (or hg38) reference genome to generate feature-barcode matrices for downstream analysis.

Single-cell ATAC-seq processing and analysis

Single-cell ATAC-seq libraries were processed using Cell Ranger ATAC (10x Genomics, version X.X) with default parameters. Reads were aligned to the mm10 reference genome provided by 10x Genomics using the BWA-MEM algorithm, and fragment files containing the genomic coordinates of uniquely mapped read pairs were generated for each sample. Only fragments with a valid barcode and mapping quality score ≥ 30 were retained. Barcode quality metrics (e.g., fraction of reads in peaks, transcription start site [TSS] enrichment, nucleosome signal) were computed. Downstream analyses were performed in R (v4.4.0) using Signac (v1.13.0)^{Stuart2021} and Seurat (v5.0.0)^{Hao2021}. Fragment files were imported using CreateChromatinAssay() and combined into a Seurat object using CreateSeuratObject(). Low-quality nuclei were removed based on standard quality-control thresholds: cells with fewer than 3,000 fragments, ≥ 5 percent mitochondrial reads, nucleosome signal ≥ 2 , or TSS enrichment ≥ 2 were excluded. Peaks overlapping genomic blacklist regions (ENCODE v2)

were also filtered out.. To normalize for differences in sequencing depth and accessibility bias, counts were transformed using TF-IDF (term frequency–inverse document frequency) normalization via `RunTFIDF()`, followed by dimensional reduction with singular value decomposition (SVD) on the top 50,000 most accessible peaks using `RunSVD()`. Low-rank representations were used to compute a neighborhood graph (`FindNeighbors()`) and to perform Louvain clustering (`FindClusters()`). Two-dimensional embeddings were generated with Uniform Manifold Approximation and Projection (UMAP) (`RunUMAP()`) based on the first 30 latent semantic indexing (LSI) components.

ChromVAR analysis of transcription factor activity

To infer transcription factor (TF) activity dynamics, we analyzed single-cell chromatin accessibility data using chromVAR^{Schep2017} implemented via the Signac package (v1.13.0)^{Stuart2021}. Peak-by-cell count matrices were generated from 10x Genomics single-cell ATAC-seq data and used to construct a chromatin accessibility Seurat object. Peaks overlapping mitochondrial DNA or genomic blacklist regions were excluded. Counts were normalized using term frequency–inverse document frequency (TF-IDF) normalization and reduced via singular value decomposition (SVD) prior to UMAP visualization. TF motif positions were obtained from the JASPAR 2022 database^{Castro-Mondragon2022} and matched to accessible regions using `AddMotifs()`. Deviation scores and z-scores for each motif were computed using `RunChromVAR()`, which estimates per-cell deviations from expected accessibility based on GC bias-corrected background models. TF deviation scores were used to quantify motif accessibility and variability across mTEC maturation states. Differential TF activity between mTEC_{lo} and mTEC_{hi} clusters was assessed using Wilcoxon tests, and variance estimates were used to identify TFs associated with transcriptional heterogeneity. Deviation variability was further compared across TF families (e.g., NF- κ B, Nuclear Receptor) to identify regulators showing both increased activity and heterogeneity during maturation.

scRNA-seq preprocessing, visualization, and clustering

Single-cell RNA-sequencing data were processed using the Seurat R package (v5.0.0)^{Hao2021}. Count matrices from wild-type (WT) and *Vdr*- or *Ythdf2*-knockout (KO) samples were imported using `Read10X()` and merged prior to processing using Seurat’s `merge()` function to generate a single unified object. All subsequent quality control, normalization, and analysis steps were performed jointly on the merged dataset to enable direct comparison between genotypes. Cells with fewer than 200 detected genes or with ≥ 10 percent mitochondrial gene content were excluded, and genes expressed in fewer than 3 cells were removed. Data were normalized using Seurat’s default log-normalization (`NormalizeData()`, `method = "LogNormalize"`, `scale.factor = 10,000`). The top 2,000 highly variable genes (HVGs) were identified using the “vst” method in `FindVariableFeatures()`. Expression values were scaled and centered with `ScaleData()`, regressing out total UMI counts and mitochondrial percentage to minimize technical variation. Dimensionality reduction was performed using Principal Component Analysis (PCA) on the HVGs (`RunPCA()`), and the number of significant principal components was determined by elbow plot inspection. Uniform Manifold Approximation and Projection (UMAP) was used for visualization (`RunUMAP()`), and clustering was performed using Seurat’s graph-based approach (`FindNeighbors()` and `FindClusters()`, Louvain algorithm, `resolution = 0.4–1.0`, adjusted per dataset). Cluster marker genes were identified using `FindAllMarkers()` (Wilcoxon rank-sum test, $\log_2$ fold-change threshold = 0.25, adjusted $p \leq 0.05$), and cluster identities were assigned based on canonical thymic epithelial cell markers and previously published single-cell references. Visualization was performed using Seurat’s built-in plotting functions (`DimPlot`, `FeaturePlot`, `VlnPlot`, and `DotPlot`).

Assessment of Gene Expression Heterogeneity

To quantify gene expression heterogeneity within each cell population, we applied Seurat’s *FindVariableFeatures* function (v5.0.1) separately to each cluster. For each cluster, cells were first subset into an independent Seurat object, and highly variable genes were identified

using the `vst` method following log-normalization. Variance calculations were therefore performed within clusters, ensuring that estimates of heterogeneity reflected intra-cluster variation rather than differences across the broader dataset. Default parameters were used unless otherwise specified.¹⁷¹

Multiome sample preparation & sequencing

For all multiome experiments, we used the ATAC+GEX single-cell kit and protocol (10X Genomics 1000236 with protocol CG000338 RevE) with minor modifications to sample preparation. Briefly, 40,000 mTECs were FACS-sorted into 1X PBS supplemented with 2% BSA and centrifuged at 300 rcf for 5 min. Cells were gently washed in 50 L Lysis Buffer (10mM Tris, 10mM NaCl, 3mM MgCl in nuclease-free water) and centrifuged at 300 rcf for 5 min. Cells were resuspended in 50 L permeabilization buffer (10mM Tris, 10mM NaCl, 3mM MgCl, 0.1% Tween20, 0.01% digitonin, and RNase inhibitor (Invitrogen) in nuclease-free water) and incubated for 5 min on ice. Nuclei were gently washed with wash buffer (10mM Tris, 10mM NaCl, 3mM MgCl, 0.1% Tween20, and RNase inhibitor in nuclease-free water) and centrifuged at 500 rcf for 5 min. Finally, nuclei were resuspended in 5 L chilled diluted nuclei buffer (10X Genomics) and added to transposition mix. Paired-end, dual-index sequencing was performed on the Illumina NovaSeq 6000 platform.

Multiome data quality-control

Following sequencing, bcl files were converted to fastq using `cellranger-arc` (v2.0.2) `mkfastq`. FASTQ files were aligned to the mm10 or hg38 genome assembly using `cellranger-arc` `count`. ATAC-seq fragment files were used as inputs to *ArchR* (v1.0.2) analysis pipeline in R (v4.3.2). Transcript count matrices were used as inputs to the *Seurat* gene expression analysis pipeline. For gene expression quality control, cells with `nFeature_RNA` 250 & 6,000, `nCount_RNA` 25,000 and `percent_mitochondrial` 25 were included for downstream analyses. Transcript counts were log-normalized. For scATAC-seq quality control, cells with `n_ATAC_Frags` 3,000

& TSS.Score 10 were included for downstream analyses. Doublet inference was conducted using *ArchR*'s `addDoubletScores()`, and presumed doublets were excluded. Cells that passed each filter were admitted for downstream analyses.

CUT&RUN sample preparation

Briefly, 350-500k cells were washed 3x in Wash Buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1x EDTA-free protease inhibitor cocktail (Roche)), then bound to Concanavalin-A beads (Bangs Laboratories) according to manufacturer's instructions. Cells were incubated with 1:100 dilution of antibody for 2h or overnight at 4°C in Permeabilization Buffer (1x Permeabilization Buffer (eBioscience), 0.5 mM spermidine, 1x EDTA-free protease inhibitor cocktail, 2 mM EDTA). Sample was then incubated with 700 ng/mL pA-MNase (S. Henikoff) in Permeabilization Buffer at 4°C for 1h. Digestion was performed in 0.5x Permeabilization Buffer supplemented with 2 mM CaCl₂ at 4°C for 1h. The reaction was stopped by the addition of 2x Stop Buffer (final concentration 100 mM NaCl, 10 mM EDTA, 2 mM EGTA, 20mg/mL glycogen, 25mg/mL RNase A (Thermo Fisher)) and the sample was incubated at 37°C for 20 min. Proteins in sample was then digested in 0.1% SDS and 250 mg/mL Proteinase K (New England Biolabs) for 2h at 56°C, shaking gently. CUT&RUN fragments were purified by phenol chloroform extraction. CUT&RUN libraries were generated using NEBNext UltraII DNA Library Prep Kit for Illumina coupled with NEBNext Multiplex Oligos for Illumina (New England Biolabs) with modifications optimized for small fragments as detailed in <https://dx.doi.org/10.17504/protocols.io.wvgfe3w>. Paired-end, dual-index sequencing was performed on the Illumina NextSeq500 platform.

greenCUT&RUN sample preparation

GreenCUT&RUN was performed as described¹⁴³ with some modifications. Briefly, 250,000 mammalian cells were washed 3x in Wash Buffer (20mM HEPESpH7.5, 150mM NaCl, 0.5mM spermidine,1x EDTA-free protease inhibitor cocktail (Roche)) and bound

to ConcanavalinA-coated magnetic beads. Cells were incubated with high-affinity GFP-nanobody-MNase fusion protein (targeting GFP-tagged factors) in Permeabilization Buffer (1x Permeabilization Buffer (eBioscience), 0.5 mM spermidine, 1x EDTA-free protease inhibitor cocktail, 2 mM EDTA) at 4°C for 20min. MNase cleavage was initiated 0.5x Permeabilization Buffer supplemented with 2 mM CaCl₂ at 4°C for 30min. The reaction was stopped by addition of 2x Stop Buffer (final concentration 100 mM NaCl, 10 mM EDTA, 2 mM EGTA, 20mg/mL glycogen, 25mg/mL RNase A (Thermo Fisher)) and the sample was incubated at 37°C for 20 min. Proteins in sample were then digested in 0.1% SDS and 250 mg/mL Proteinase K (New England Biolabs) for 2h or overnight at 56°C, shaking gently. Released fragments were extracted using phenol-chloroform/ethanol precipitation. Libraries were generated using NEBNext UltraII DNA Library Prep Kit for Illumina coupled with NEBNext Multiplex Oligos for Illumina (New England Biolabs) with modifications optimized for small fragments as detailed in <https://dx.doi.org/10.17504/protocols.io.wvgfe3w>. Paired-end, dual-index sequencing was performed on the Illumina NextSeq500 platform.

CUT&RUN data processing

CUT&RUN reads were mapped to mm10 mouse genome assembly using *Bowtie2* (v2.2.9) with settings *-local -very-sensitive-local -no-unal -no-mixed -no-discordant -phred33 -I 10 -X 700*. PCR duplicates were removed using *Picard* (v2.21.8) *MarkDuplicates REMOVE_DUPLICATES=true VALIDATION_STRINGENCY=LENIENT*. Reads with MAPQ scores below 30 were purged and excluded from downstream analysis using *samtools* (v1.9) *view -b -q 30 -f 2 -F 1804*. Peaks were called for each sample using *MACS2* (v2.2.7.1) with settings *-shift 0 -extsize 200 -nomodel -call-summits -keep-dup all -p 0.01*. For each sample, a 301 bp fixed-width peak set was generated by extending the MACS2 summits by 150 bp in both directions. Peaks were ranked by significance (MACS2 peak score) and overlapping peaks with lower peak scores were removed iteratively to create non-overlapping sample peak sets. Peaks mapping to chrY as well as any that spanned genomic

regions containing “N” nucleotides were removed. Robust peaks were defined by a ‘score-per-million’ (SPM) (each peak score divided by the sum of all peak scores in the sample, divided by 1 million), and we only retained peaks with an SPM value ≥ 5 . CUT&RUN fragment counts across regions of interest were normalized by the number of unique fragments in the sample library.

Reverse transcription-based RNA binding protein binding sites sequencing

ARTR-seq was performed according to the previously published procedure.¹⁷⁰ Briefly, mTECs were fixed with 1.5% paraformaldehyde (PFA) for 10 min at room temperature, quenched with 125 mM glycine, and permeabilized with 0.5% Triton X-100 on ice for 10 min. Samples were blocked with UltraPure bovine serum albumin (BSA) (1 mg/ml, Thermo Fisher Scientific), stained with YTHDF2 antibody (details) at room temperature for 1 hour, and then stained with fluorophore-labeled secondary antibody at room temperature for 30 min. Samples were then incubated with pAG-RTase for an additional 30 min. A reverse transcription reaction mixture was prepared by mixing 2 uM adapter-RT primer (5-AGACGTGTGCTCTTCC-GATCT-10 N-3), 0.05 mM biotin-16-dUTP (deoxyuridine triphosphate) (Jena Bioscience), 0.05 mM biotin-16-dCTP (deoxycytidine triphosphate) (Jena Bioscience), 0.05 mM deoxythymidine triphosphate (Thermo Fisher Scientific), 0.05 mM dCTP (Thermo Fisher Scientific), 0.1 mM deoxyadenosine triphosphate (Thermo Fisher Scientific), 0.1 mM deoxyguanosine triphosphate (Thermo Fisher Scientific), RNaseOUT (1 U/ul, Thermo Fisher Scientific) in 50 ul of buffer of DPBS supplemented with 3 mM MgCl₂. In situ reverse transcription was performed by adding RT reaction mixture to cells and incubating at 37°C for 30 min and then quenched by adding 20 mM EDTA and 10 mM EGTA. To check the success of in situ reverse transcription, cells were stained with biotin monoclonal antibody (BK-1/39, RRID:AB10598675) conjugated with Alexa Fluor 488 (AF488, Thermo Fisher Scientific), and imaged by a Leica SP8 laser confocal microscope. The fluorescence intensity distribution on a line was quantified by Fiji (version 2.3.0) (67). After imaging, cells were digested with proteinase K (Thermo Fisher Scientific), and the nucleic acids, including the generated biotinylated cDNA, were recovered by phenol-chloroform extraction and concentrated by ethanol precipitation. RNA was digested with RNase H (NEB) and RNase A/T1 (Thermo Fisher Scientific) at 37°C for 1 hour, followed by biotinylated cDNA enrichment using Dynabeads MyOne Streptavidin C1 (Thermo Fisher Scientific). The 3' cDNA adapter (5Phos-8 N-AGATCGGAAGAG-CGTCGTGT-3SpC3) was ligated by T4

RNA ligase 1 (NEB) by incubating at 25°C for 16 hours, and cDNA was recovered with the elution buffer of 95% (v/v) formamide and 10 mM EDTA (pH 8.0) by boiling at 95°C for 10 min, followed by ethanol precipitation. The library can be obtained by PCR amplification with next-generation sequencing primer and gel purification of size between 180 and 400 bp. Sequencing was performed at the University of Chicago Genomics Facility on an Illumina NovaSeq 6000 platform in single-end mode with 100 bp.

CHAPTER 5:

DISCUSSION

5.1 Summary

Collectively, these studies illuminate molecular mechanisms that govern the expression of tissue-specific genes in thymic medullary epithelial cells. Overall, the work identifies that the regulation of immune tolerance is not governed by a single pathway, but instead emerges from the integration of multiple layers of control, including chromatin state, transcription factor activity, and post-transcriptional gene regulation. These regulatory systems act in a highly dynamic and stage-dependent manner during thymic epithelial cell development, ensuring that self-antigen expression is appropriately timed and distributed. The findings further suggest that transcription factor networks associated with developmental and signaling pathways are reorganized across maturation stages to enable context-specific gene expression programs. In parallel, RNA-based regulatory mechanisms contribute an additional layer of control by modulating the availability and activity of key transcriptional regulators. Together, this work supports a model in which immune tolerance arises from the coordinated interaction of epigenetic, transcriptional, and post-transcriptional processes. Disruption of this coordination can lead to defective self-antigen expression and breakdown of immune tolerance, underscoring the importance of multi-level regulatory integration in immune system development.

In Chapter 2, we demonstrate that the range of NF- κ B activity is a key determinant of the diversity of TSGs expressed in medullary thymic epithelial cells. Rather than simply promoting transcription, NF- κ B operates within a constrained activity window that shapes which genomic regions are engaged across the cell population. We identify the Vitamin D receptor as a critical regulator of this range, functioning to establish an upper bound on NF- κ B activity by limiting access to chromatin regions enriched for high-density NF- κ B motifs—putative high-affinity binding sites. Mechanistically, VDR appears to restrict

NF- κ B occupancy at these sites, effectively redistributing NF- κ B molecules away from dominant, high-affinity regions and toward a broader set of lower-affinity targets. This constraint promotes a more even distribution of NF- κ B binding events across the genome, thereby increasing cell-to-cell heterogeneity in transcriptional outputs. In this way, VDR does not simply repress NF- κ B activity, but instead tunes its genomic engagement to favor diversity over dominance. Loss of VDR disrupts this balance, allowing NF- κ B to preferentially occupy high-affinity, motif-dense regions. This leads to a skewed transcriptional program in which a subset of TSGs is expressed at higher frequency across the mTEC population, at the expense of broader antigen representation. As a consequence, over 1,000 TSGs exhibit reduced expression, reflecting a collapse in transcriptional diversity rather than a global loss of NF- κ B activity. Functionally, this narrowing of the self-antigen repertoire impairs negative selection and results in the development of multi-organ autoimmunity. Together, these findings establish VDR as a key modulator of transcription factor “bandwidth,” demonstrating that constraining the upper limits of signaling activity is essential for maintaining the breadth of self-antigen expression required for immune tolerance.

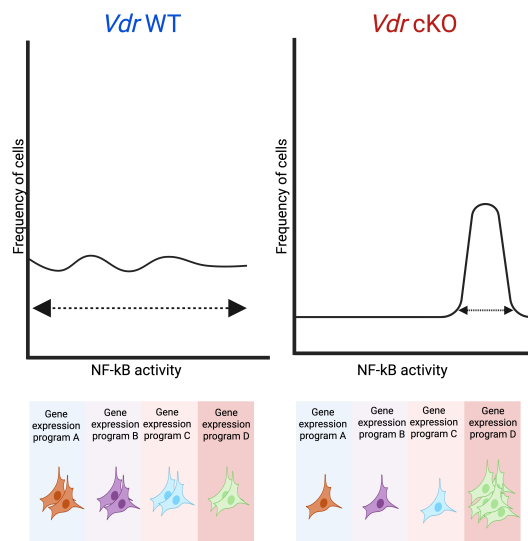


Figure 5.1 | *Vdr* limits the amplitude and increases the range of NF- κ B activity in mature mTECs

In Chapter 3, we explore how multiple layers of regulation coordinate transcription factor network transitions during mTEC development. We identify the AP-1 and NF- κ B families as key regulators exhibiting dynamic and stage-specific activity, with both factors strongly enriched at genomic regions associated with TSGs. Notably, the activities of AP-1 and NF- κ B appear inversely related across developmental stages, with a pronounced reduction in AP-1 activity during the transit-amplifying phase. Focusing on the AP-1 component FOS, we observe a striking redistribution of binding from mTEC_{lo} to mTEC_{hi} cells, with substantial enrichment at TSG loci in mature mTECs. These findings support a model in which NF- κ B, particularly the subunit RELA, primes the chromatin landscape early in development, establishing accessibility at regulatory regions that are later occupied by AP-1. Concurrently, the m6A reader and mRNA decay factor YTHDF2 promotes the clearance of *Fos* transcripts during intermediate stages, enabling dynamic control of AP-1 activity and facilitating its subsequent redistribution. Disruption of this coordinated process impairs TSG expression. Loss of either RELA or YTHDF2 leads to increased chromatin accessibility at sites preferentially bound by FOS in mTEC_{lo} cells, coupled with a failure to establish proper AP-1 occupancy at TSG-associated loci in mature cells. Interestingly, the genes most affected are those associated with high FOS activity, suggesting that precise temporal regulation of AP-1 is required for their expression. Failure of this redistribution mechanism results in a breakdown of central tolerance and the development of multi-organ autoimmunity. Together, these findings demonstrate how transcriptional, chromatin-based, and post-transcriptional regulatory mechanisms converge to orchestrate transcription factor network transitions, enabling the dynamic reprogramming required for tissue-specific gene expression and immune tolerance.

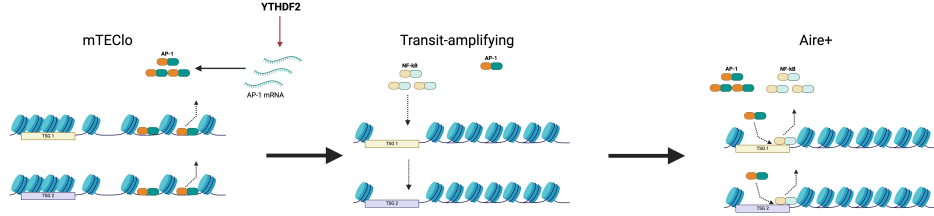


Figure 5.2 | AP-1 redistribution mediated by YTHDF2 clearance of Fos transcripts and chromatin landscape changes by RelA

5.2 Discussion

5.2.1 Regulation of TF networks for a particular fate

Together, these findings reveal a multi-layered regulatory architecture in which transcription factor activity is controlled through distinct but coordinated mechanisms to achieve a common goal: maximizing cellular plasticity while maintaining tight control over cell state. Achieving this balance requires stratified regulation across chromatin accessibility, transcription factor dynamics, and post-transcriptional control, thereby preventing aberrant activation of dominant programs. In both contexts described, the directed localization of transcription factors—mediated in part by the chromatin landscape—emerges as a central mechanism for shaping cellular behavior and ensuring appropriate gene expression outcomes.

Within this framework, RELA appears to function as a key upstream regulator of the chromatin landscape on which other transcription factors act. RELA plays a prominent role in establishing chromatin accessibility, potentially through cooperation with chromatin

remodeling complexes such as BAF³⁶. Given the central importance of NF- κ B signaling in mTECs, it is not surprising that this pathway is tightly regulated. We show that the VDR constrains the range of RELA activity by limiting its access to chromatin regions enriched for high-density NF- κ B motifs. In doing so, VDR effectively restricts RELA occupancy at high-affinity sites, redistributing NF- κ B activity across a broader set of genomic regions. Conceptually, this highlights an important regulatory principle: the activity of a transcription factor can be tuned not only by modulating its abundance or activation state, but also by constraining where it is permitted to bind.

A similar layering of control is evident in the relationship between RELA and AP-1. RELA not only regulates its own genomic distribution but also shapes the chromatin landscape in a manner that influences subsequent binding by other factors, including FOS. During mTEC development, RELA-dependent chromatin remodeling in earlier stages appears to reset accessibility patterns, enabling the later redistribution of FOS to TSG-associated loci in mature cells. This process is further refined by post-transcriptional regulation: the clearance of *Fos* mRNA by YTHDF2 transiently reduces AP-1 activity, creating a window in which the chromatin landscape can be reconfigured before FOS re-engages. As FOS levels recover, it is poised to bind a newly established set of regulatory elements, potentially with different cofactors, thereby supporting the transition to a new transcriptional network.

Together, these mechanisms illustrate shared principles of transcriptional regulation—namely, the control of factor localization, timing, and activity range—while operating through distinct molecular layers. In combination, they enable dynamic yet controlled reprogramming of gene expression, ensuring that mTECs can achieve the diversity of tissue-specific gene expression required for immune tolerance without sacrificing regulatory stability.

5.2.2 Reframing central tolerance as a problem of controlled heterogeneity

TSG expression and T cell affinity for antigen are often conceptualized in binary terms—genes are either expressed or not, and T cells are either self-reactive or not. In

reality, both processes are inherently graded. Tissue-specific genes are expressed across a spectrum of frequencies and amplitudes, while T cell receptors span a continuous range of affinities for self peptide–MHC complexes. Consequently, the efficacy of central tolerance depends not simply on whether self-antigens are presented, but on how they are distributed across the mTEC population. Both the probability of expression (frequency) and the level of expression (amplitude) shape the landscape of self-antigen presentation.

This reframes central tolerance as a problem of controlled heterogeneity. Uniform low-level expression of all self-antigens would fail to reach the thresholds required to eliminate high-affinity autoreactive clones. Conversely, high-level expression of a limited subset of antigens would efficiently purge T cells specific to those targets while leaving large regions of the self-antigen space unrepresented. In both cases, breakdown arises not from a lack of antigen presentation, but from insufficient diversity in how antigens are deployed across cells.

Here, we isolate one axis of this system—regulation of expression frequency—and show that it is a key determinant of thymic selection. Disrupting the distribution of TSG expression across cells is sufficient to impair tolerance and promote autoimmunity.

Effective central tolerance therefore requires a balance: enough heterogeneity to ensure broad coverage of the self-antigen repertoire, coupled with sufficient expression to enforce deletion of high-affinity clones. In this view, thymic selection is fundamentally a population-level process, where diversity in antigen presentation—not uniformity—underpins robust self-tolerance.

5.2.3 Defining an expanded role of NF- κ B in mTECs

In this work, we expand current models of NF- κ B function in mTECs by identifying a previously underappreciated role for the canonical factor RELA in priming cells for tissue-specific gene TSG expression. RELB has long been established as a central regulator of mTEC development: disruption of upstream signaling components or deletion of RELB

itself results in a profound loss of mature mTECs. Although RELA and RELB have been thought to play distinct roles within the NF- κ B pathway, the functional basis of this division has remained unclear.

Here, we show that RELA primarily acts to shape the chromatin landscape, pre-disposing mTECs to express specific subsets of TSGs. RELA activity within individual cells correlates with their subsequent TSG expression profiles, suggesting that it functions upstream of transcriptional output by establishing permissive chromatin states. This role is fundamentally distinct from the canonical pro-inflammatory function of RELA, where it directly drives transcription of cytokines and other immune effector genes. In mTECs, our data instead support a model in which RELA regulates chromatin accessibility—potentially in coordination with chromatin remodeling complexes—to enable selective and heterogeneous TSG expression.

5.2.4 Transcription factor redistribution as a driver of TSG expression

Here, we demonstrate that the AP-1 component FOS undergoes a pronounced redistribution during mTEC maturation, shifting from binding at mTEC_{lo}-associated regulatory regions to loci associated with tissue-specific genes in mature cells. This redistribution provides a compelling mechanism for enabling the broad and heterogeneous expression of TSGs required for effective central tolerance. Given the extensive diversity of the self-antigen repertoire, regulatory factors must be deployed in a manner that supports both flexibility and specificity in gene activation.

AP-1 factors are well positioned to fulfill this role. They are known to interact with a wide range of transcriptional partners and co-factors, enabling context-dependent gene regulation across diverse genomic environments.^{172,173} Notably, AP-1 and NF- κ B motifs frequently co-occur within regulatory regions, suggesting potential coordination or competition between these pathways at shared genomic sites. This raises the possibility that AP-1 redistribution is not an isolated event, but is mechanistically linked to earlier NF- κ B activity.

Our data support a model in which NF- κ B, particularly RELA, establishes a primed chromatin landscape early in mTEC development, which is subsequently interpreted by AP-1 factors. Within this framework, at least two non-mutually exclusive scenarios can be envisioned. In the first, RELA remains bound at a subset of regulatory elements, where it cooperates with FOS to stabilize enhancer activity and promote transcriptional activation of TSGs, potentially through recruitment of additional co-factors. In the second, FOS binding displaces RELA at previously primed sites, representing a handoff mechanism in which early NF- κ B activity licenses later AP-1-dependent transcriptional programs.

While the precise balance between these mechanisms remains to be determined, both models are consistent with a sequential and coordinated reorganization of transcription factor occupancy across development. In either case, the redistribution of AP-1 represents a critical step in translating a primed chromatin state into active and diverse gene expression, thereby contributing to the establishment of a broad self-antigen repertoire in mature mTECs.

5.3 Future Directions

5.3.1 *Consequences of modulation of individual TSG expression frequency or amplitude*

Limited expression of individual TSGs has been a concern in the field for many years, as low antigen abundance may restrict the efficiency of central tolerance induction.¹⁷⁴ In contrast, the consequences of excessive expression of individual TSGs have received comparatively little attention. Our findings in the *Vdr* cKO model demonstrate that disruption of thymic epithelial cell regulatory networks can substantially increase expression of specific TSGs, raising important questions regarding how elevated self-antigen abundance influences thymocyte selection.

The effects of increased expression of individual TSGs on thymic selection remain largely unexplored, particularly when considered independently of broader expression trade-offs. Because developing T cells exhibit a spectrum of affinities for multiple self-peptide-MHC complexes, increasing the abundance of a particular self-antigen could shift

the thresholds governing thymocyte fate decisions. One potential consequence is enhanced deletion of T cell clones that recognize the overexpressed antigen. While this may strengthen tolerance to that specific self-antigen, it could also remove T cell receptors that would otherwise contribute to protective immune responses against pathogens through cross-reactivity. In addition to clonal deletion, strong self-reactivity can promote diversion into the Treg lineage when TCR signaling falls within a particular affinity window. Whether increased expression of individual TSGs alters the abundance, diversity, or antigen specificity of Treg populations remains unknown. Addressing this question will require characterization of the TCR repertoires of both conventional and regulatory T cells, ideally through TCR sequencing approaches capable of linking TCR specificity to cell fate. Future studies that experimentally modulate the expression levels of individual TSGs in thymic epithelial cells will therefore be critical for defining how quantitative variation in self-antigen presentation shapes central tolerance. Such work could establish whether optimal ranges of TSG expression exist and provide new insight into how dysregulation of antigen abundance contributes to autoimmune susceptibility or immune homeostasis.

5.3.2 Determining the role of AP-1 in TSG expression

This work supports a model in which AP-1 transcription factors are redistributed during mTEC maturation through the combined actions of YTHDF2-mediated degradation of AP-1 mRNAs and RELA-dependent remodeling of the chromatin landscape. Within this framework, AP-1 activity is progressively redirected away from early maturation programs and toward loci associated with TSG expression. Because AP-1 factors frequently function in cooperation with other transcription factors, this redistribution may facilitate interactions with lineage-specific regulatory partners that promote the activation of distinct TSG programs in mature mTECs.

While several observations are consistent with this model, important mechanistic questions remain unresolved. Chromatin accessibility analyses demonstrate enrichment of

AP-1 motifs at TSG-associated regulatory regions, and CUT&RUN data indicate binding of FOS family members at many TSG loci. However, these findings are correlative and do not establish a direct causal role for AP-1 in regulating TSG expression. Determining whether AP-1 activity is necessary or sufficient for TSG induction will require targeted perturbation experiments coupled with transcriptional profiling. For example, stage-specific loss- or gain-of-function approaches could be used to assess how alterations in AP-1 activity affect both global transcriptional programs and expression of individual TSGs.

A major challenge in pursuing these studies is the broad biological role of AP-1 factors. AP-1 proteins regulate diverse processes including proliferation, differentiation, stress responses, and cell survival, making it likely that conventional perturbations would generate widespread secondary effects. Consequently, observed changes in TSG expression could reflect indirect consequences of altered cellular state rather than direct transcriptional regulation. Future experimental designs should therefore prioritize approaches that restrict AP-1 perturbation to defined stages of mTEC maturation and minimize disruption of unrelated cellular processes. Inducible genetic models, temporally controlled perturbations, or selective targeting of individual AP-1 family members may help disentangle direct effects on TSG regulation from broader alterations in mTEC development.

Beyond establishing causality, identifying the specific AP-1 complexes and co-factors involved in TSG regulation represents an important future direction. Different AP-1 dimers possess distinct DNA-binding preferences and regulatory functions, raising the possibility that only a subset of AP-1 family members participates in TSG induction. Clearly defining these interactions and their genomic targets would provide a more complete understanding of how transcription factor redistribution contributes to the establishment of promiscuous gene expression and central tolerance in the thymus.

5.3.3 *Reassessing the weight of thymic VDR function in immune reactivity*

Genetic variation in *Vdr* has been associated with susceptibility to numerous autoimmune diseases, as discussed in Chapter 1. These associations are typically interpreted through the lens of peripheral immune regulation, where VDR signaling influences T cell activation, cytokine production, and regulatory T cell function. However, the findings presented in this dissertation raise the possibility that at least some disease-associated effects of altered VDR activity may originate much earlier during immune development. Specifically, disruption of VDR function in mTECs alters TSG expression, suggesting that defects in central tolerance could contribute to the generation of autoreactive T cell repertoires. An important future direction will therefore be to determine whether autoimmune phenotypes associated with impaired VDR signaling are driven, in part, by altered thymic selection rather than exclusively by defects in peripheral immune cell function.

A related question concerns the role of VDR ligands in regulating thymic epithelial cell biology. Previous work demonstrated that deletion of *Cyp27b1*, which encodes the 1α -hydroxylase responsible for generating the active vitamin D metabolite calcitriol, results in altered thymic cellularity and changes in TSG expression.¹⁷⁵ However, interpretation of these findings is complicated by the use of a global knockout model. Because *Cyp27b1* is expressed in multiple tissues, the observed thymic phenotypes may arise indirectly through systemic effects or through loss of vitamin D signaling in non-thymic cell types. As a result, direct administration of calcitriol may provide a more controlled approach for investigating how ligand availability influences VDR-dependent regulatory programs within the thymus.

Such studies would also help distinguish ligand-dependent and ligand-independent functions of VDR. The model proposed in this dissertation suggests that VDR restrains NF- κ B-driven transcriptional programs in mTECs. However, nuclear receptors can mediate transcriptional repression through interactions with co-repressor complexes such as NCOR in a ligand-independent manner.¹⁷⁶ If repression of NF- κ B activity by VDR primarily reflects this ligand-independent mechanism, increasing calcitriol availability may have little

effect on the pathways identified here. Conversely, if ligand binding enhances or modifies VDR-mediated regulation of chromatin accessibility, transcription factor activity, or TSG expression, calcitriol treatment could reveal additional layers of regulation not captured by genetic deletion models.

Future studies combining calcitriol administration with thymus-specific genetic models will be particularly valuable for dissecting these possibilities. Such experiments could determine whether vitamin D availability directly influences central tolerance mechanisms and clarify the extent to which thymic versus peripheral functions of VDR contribute to autoimmune disease susceptibility. More broadly, these studies may help establish whether manipulation of vitamin D signaling can be leveraged to modulate thymic selection and immune tolerance in a therapeutic context.

5.4 Conclusion

Together, this work defines a unifying framework for how central tolerance is established through the coordinated regulation of transcriptional diversity in medullary thymic epithelial cells. Rather than being driven by the activity of individual factors or pathways, tolerance emerges from the precise integration of signaling dynamics, chromatin accessibility, and RNA-mediated control mechanisms that collectively shape the breadth and distribution of tissue-specific gene expression.

A central theme that emerges is that the range and timing of regulatory activity are as important as their presence. In Chapter 2, we show that constraining NF- κ B signaling through VDR is essential not to limit transcription per se, but to preserve diversity by preventing dominance of high-affinity genomic targets. In Chapter 3, we extend this concept by demonstrating that transcription factor activity must also be dynamically reconfigured over developmental time. RELA-dependent chromatin priming establishes a permissive landscape that is subsequently interpreted by AP-1, while post-transcriptional regulation via YTHDF2 ensures that these transitions occur with proper timing and fidelity. Together,

these mechanisms illustrate that transcription factor networks are not static drivers of gene expression, but are actively redistributed and repurposed across developmental stages to enable context-specific transcriptional programs.

These findings support a model in which central tolerance is fundamentally a systems-level property, arising from the controlled heterogeneity of self-antigen presentation across the mTEC population. Disruption at any level—whether through altered signaling bandwidth, misconfigured chromatin landscapes, or failures in post-transcriptional control—leads to a collapse in transcriptional diversity, narrowing the self-antigen repertoire and compromising negative selection. In this way, autoimmunity can emerge not only from the loss of specific genes or pathways, but from failures in the coordination that maintains diversity itself.

More broadly, this work highlights a general principle in gene regulation: that limiting the dominance of strong signals and enforcing dynamic transitions between regulatory states are critical for generating complex and diverse transcriptional outputs. In the thymus, this principle ensures robust self-tolerance. In other contexts, similar strategies may underlie the generation of cellular diversity and the stability of differentiated states.

REFERENCES

1. 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. *Proceedings of the National Academy of Sciences* **105**, 657–662 10.1073/pnas.0707486105 (2008).
2. Sansom, S. N. *et al.* Population and single-cell genomics reveal the *Aire* dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. *Genome Research* **24**, 1918–1931 10.1101/gr.171645.113 (2014).
3. Brennecke, P. *et al.* Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. *Nature Immunology* **16**, 933–941 10.1038/ni.3246 (2015).
4. Miragaia, R. J. *et al.* Single-cell RNA-sequencing resolves self-antigen expression during mTEC development. *Scientific Reports* **8**, 685 10.1038/s41598-017-19100-4 (2018).
5. Murphy, J. B. & Rous, P. THE BEHAVIOR OF CHICKEN SARCOMA IMPLANTED IN THE DEVELOPING EMBRYO. *Journal of Experimental Medicine* **15**, 119–132 10.1084/jem.15.2.119 (1912).
6. Owen, R. D. Immunogenetic Consequences of Vascular Anastomoses Between Bovine Twins. *Science* **102**, 400–401 10.1126/science.102.2651.400 (1945).
7. Billingham, R. E., Brent, L. & Medawar, P. B. ‘Actively Acquired Tolerance’ of Foreign Cells. *Nature* **172**, 603–606 10.1038/172603a0 (1953).
8. Burnet, F. M. A Modification of Jerne’s Theory of Antibody Production using the Concept of Clonal Selection. *CA: A Cancer Journal for Clinicians* **26**, 119–121 10.3322/canjclin.26.2.119 (1976).
9. Marshall, J. S. *et al.* Introduction to immunology and immune disorders. *Allergy, Asthma & Clinical Immunology* **20**, 69 10.1186/s13223-024-00932-5 (2024).
10. Kaech, S. M. & Cui, W. Transcriptional control of effector and memory CD8+ T cell differentiation. *Nature Reviews Immunology* **12**, 749–761 10.1038/nri3307 (2012).
11. Zhu, J., Yamane, H. & Paul, W. E. Differentiation of Effector CD4 T Cell Populations. *Annual Review of Immunology* **28**, 445–489 10.1146/annurev-immunol-030409-101212 (2010).
12. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402 10.1038/334395a0 (1988).
13. Mason, D. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunology Today* **19**, 395–404 10.1016/S0167-5699(98)01299-7 (1998).
14. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575–581 10.1038/302575a0 (1983).

15. Uehara, S., Grinberg, A., Farber, J. M. & Love, P. E. A Role for CCR9 in T Lymphocyte Development and Migration. *The Journal of Immunology* **168**, 2811–2819 10.4049/jimmunol.168.6.2811 (2002).
16. Ueno, T. *et al.* CCR7 Signals Are Essential for Cortex–Medulla Migration of Developing Thymocytes. *The Journal of Experimental Medicine* **200**, 493–505 10.1084/jem.20040643 (2004).
17. Rothenberg, E. V. & Taghon, T. MOLECULAR GENETICS OF T CELL DEVELOPMENT. *Annual Review of Immunology* **23**, 601–649 10.1146/annurev.immunol.23.021704.115737 (2005).
18. Schatz, D. G. & Swanson, P. C. V(D)J Recombination: Mechanisms of Initiation. *Annual Review of Genetics* **45**, 167–202 10.1146/annurev-genet-110410-132552 (2011).
19. Arstila, T. P. *et al.* A Direct Estimate of the Human $\alpha\beta$ T Cell Receptor Diversity. *Science* **286**, 958–961 10.1126/science.286.5441.958 (1999).
20. Starr, T. K., Jameson, S. C. & Hogquist, K. A. Positive and Negative Selection of T Cells. *Annual Review of Immunology* **21**, 139–176 10.1146/annurev.immunol.21.120601.141107 (2003).
21. Dorfman, J. R. & Germain, R. N. MHC-dependent survival of naïve T cells? A complicated answer to a simple question. *Microbes and Infection* **4**, 547–554 10.1016/S1286-4579(02)01571-X (2002).
22. Ashby, K. M. & Hogquist, K. A. A guide to thymic selection of T cells. *Nature Reviews Immunology* **24**, 103–117 10.1038/s41577-023-00911-8 (2024).
23. Klein, L., Kyewski, B., Allen, P. M. & Hogquist, K. A. Positive and negative selection of the T cell repertoire: what thymocytes see (and don’t see). *Nature Reviews Immunology* **14**, 377–391 10.1038/nri3667 (2014).
24. Vriskoop, N., Monteiro, J. P., Mandl, J. N. & Germain, R. N. Revisiting Thymic Positive Selection and the Mature T Cell Repertoire for Antigen. *Immunity* **41**, 181–190 10.1016/j.immuni.2014.07.007 (2014).
25. Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J. & Schreiber, R. D. Cancer immunoeediting: from immunosurveillance to tumor escape. *Nature Immunology* **3**, 991–998 10.1038/ni1102-991 (2002).
26. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 10.1126/science.aaa4971 (2015).
27. Schmidt, J. *et al.* Neoantigen-specific CD8 T cells with high structural avidity preferentially reside in and eliminate tumors. *Nature Communications* **14**, 3188 10.1038/s41467-023-38946-z (2023).
28. Punt, J. A., Suzuki, H., Granger, L. G., Sharrow, S. O. & Singer, A. Lineage Commitment in the Thymus: Only the Most Differentiated (TCRhibcl-2hi) Subset of CD4+CD8+Thymocytes Has Selectively Terminated CD4 or CD8 Synthesis. *The Journal of Experimental Medicine* **184**, 2091–2100 10.1084/jem.184.6.2091 (1996).

29. Derbinski, J., Schulte, A., Kyewski, B. & Klein, L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nature Immunology* **2**, 1032–1039 10.1038/ni723 (2001).
30. Peterson, P. *et al.* APECED: a monogenic autoimmune disease providing new clues to self-tolerance. *Immunology Today* **19**, 384–386 10.1016/S0167-5699(98)01293-6 (1998).
31. Bjorses, P. Gene defect behind APECED: a new clue to autoimmunity. *Human Molecular Genetics* **7**, 1547–1553 10.1093/hmg/7.10.1547 (1998).
32. Anderson, M. S. *et al.* Projection of an Immunological Self Shadow Within the Thymus by the Aire Protein. *Science* **298**, 1395–1401 10.1126/science.1075958 (2002).
33. Perniola, R. & Musco, G. The biophysical and biochemical properties of the autoimmune regulator (AIRE) protein. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1842**, 326–337 10.1016/j.bbadis.2013.11.020 (2014).
34. Halonen, M. *et al.* APECED-causing mutations in AIRE reveal the functional domains of the protein. *Human Mutation* **23**, 245–257 10.1002/humu.20003 (2004).
35. Koh, A. S. *et al.* Aire employs a histone-binding module to mediate immunological tolerance, linking chromatin regulation with organ-specific autoimmunity. *Proceedings of the National Academy of Sciences* **105**, 15878–15883 10.1073/pnas.0808470105 (2008).
36. Koh, A. S. *et al.* Rapid chromatin repression by Aire provides precise control of immune tolerance. *Nature Immunology* **19**, 162–172 10.1038/s41590-017-0032-8 (2018).
37. Giraud, M. *et al.* Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. *Proceedings of the National Academy of Sciences* **109**, 535–540 10.1073/pnas.1119351109 (2012).
38. Žumer, K., Plemenitaš, A., Saksela, K. & Peterlin, B. M. Patient mutation in AIRE disrupts P-TEFb binding and target gene transcription. *Nucleic Acids Research* **39**, 7908–7919 10.1093/nar/gkr527 (2011).
39. Yoshida, H. *et al.* Brd4 bridges the transcriptional regulators, Aire and P-TEFb, to promote elongation of peripheral-tissue antigen transcripts in thymic stromal cells. *Proceedings of the National Academy of Sciences* **112** 10.1073/pnas.1512081112 (2015).
40. Bansal, K., Yoshida, H., Benoist, C. & Mathis, D. The transcriptional regulator Aire binds to and activates super-enhancers. *Nature Immunology* **18**, 263–273 10.1038/ni.3675 (2017).
41. Bansal, K. *et al.* Aire regulates chromatin looping by evicting CTCF from domain boundaries and favoring accumulation of cohesin on superenhancers. *Proceedings of the National Academy of Sciences* **118**, e2110991118 10.1073/pnas.2110991118 (2021).
42. Hayden, M. S. & Ghosh, S. NF- κ B, the first quarter-century: remarkable progress and outstanding questions. *Genes & Development* **26**, 203–234 10.1101/gad.183434.111 (2012).

43. Shih, V. F.-S., Tsui, R., Caldwell, A. & Hoffmann, A. A single NF κ B system for both canonical and non-canonical signaling. *Cell Research* **21**, 86–102 10.1038/cr.2010.161 (2011).
44. Burkly, L. *et al.* Expression of relB is required for the development of thymic medulla and dendritic cells. *Nature* **373**, 531–536 10.1038/373531a0 (1995).
45. Zuklys, S. *et al.* Normal Thymic Architecture and Negative Selection Are Associated with *Aire* Expression, the Gene Defective in the Autoimmune-Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy (APECED). *The Journal of Immunology* **165**, 1976–1983 10.4049/jimmunol.165.4.1976 (2000).
46. Weih, F. *et al.* Multiorgan inflammation and hematopoietic abnormalities in mice with a targeted disruption of RelB, a member of the NF- κ B/Rel family. *Cell* **80**, 331–340 10.1016/0092-8674(95)90416-6 (1995).
47. Boehm, T., Scheu, S., Pfeffer, K. & Bleul, C. C. Thymic Medullary Epithelial Cell Differentiation, Thymocyte Emigration, and the Control of Autoimmunity Require Lympho–Epithelial Cross Talk via LT β R. *The Journal of Experimental Medicine* **198**, 757–769 10.1084/jem.20030794 (2003).
48. Akiyama, T. *et al.* Dependence of Self-Tolerance on TRAF6-Directed Development of Thymic Stroma. *Science* **308**, 248–251 10.1126/science.1105677 (2005).
49. Rossi, S. W. *et al.* RANK signals from CD4+3- inducer cells regulate development of Aire-expressing epithelial cells in the thymic medulla. *The Journal of Experimental Medicine* **204**, 1267–1272 10.1084/jem.20062497 (2007).
50. Akiyama, T. *et al.* The Tumor Necrosis Factor Family Receptors RANK and CD40 Cooperatively Establish the Thymic Medullary Microenvironment and Self-Tolerance. *Immunity* **29**, 423–437 10.1016/j.immuni.2008.06.015 (2008).
51. Riemann, M. *et al.* Central immune tolerance depends on crosstalk between the classical and alternative NF- κ B pathways in medullary thymic epithelial cells. *Journal of Autoimmunity* **81**, 56–67 10.1016/j.jaut.2017.03.007 (2017).
52. Baik, S., Jenkinson, E. J., Lane, P. J. L., Anderson, G. & Jenkinson, W. E. Generation of both cortical and Aire^+ medullary thymic epithelial compartments from $\text{CD}4^+205^+$ progenitors. *European Journal of Immunology* **43**, 589–594 10.1002/eji.201243209 (2013).
53. Lkhagvasuren, E., Sakata, M., Ohigashi, I. & Takahama, Y. Lymphotoxin β Receptor Regulates the Development of CCL21-Expressing Subset of Postnatal Medullary Thymic Epithelial Cells. *The Journal of Immunology* **190**, 5110–5117 10.4049/jimmunol.1203203 (2013).
54. Akiyama, T., Shinzawa, M., Qin, J. & Akiyama, N. Regulations of Gene Expression in Medullary Thymic Epithelial Cells Required for Preventing the Onset of Autoimmune Diseases. *Frontiers in Immunology* **4** 10.3389/fimmu.2013.00249 (2013).

55. Metzger, T. C. *et al.* Lineage Tracing and Cell Ablation Identify a Post-Aire-Expressing Thymic Epithelial Cell Population. *Cell Reports* **5**, 166–179 10.1016/j.celrep.2013.08.038 (2013).
56. Michelson, D. A., Hase, K., Kaisho, T., Benoist, C. & Mathis, D. Thymic epithelial cells co-opt lineage-defining transcription factors to eliminate autoreactive T cells. *Cell* **185**, 2542–2558.e18 10.1016/j.cell.2022.05.018 (2022).
57. Nishijima, H. *et al.* Aire Controls Heterogeneity of Medullary Thymic Epithelial Cells for the Expression of Self-Antigens. *The Journal of Immunology* **208**, 303–320 10.4049/jimmunol.2100692 (2022).
58. Gamble, N. *et al.* Thymic epithelial cells amplify epigenetic noise to promote immune tolerance. *Nature* **646**, 724–733 10.1038/s41586-025-09424-x (2025).
59. Whitmarsh, A. J. & Davis, R. J. A central control for cell growth. *Nature* **403**, 255–256 10.1038/35002220 (2000).
60. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature Genetics* **43**, 264–268 10.1038/ng.759 (2011).
61. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147–151 10.1038/nature01763 (2003).
62. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–577 10.1038/386569a0 (1997).
63. Su, W., Jackson, S., Tjian, R. & Echols, H. DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. *Genes & Development* **5**, 820–826 10.1101/gad.5.5.820 (1991).
64. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 10.1016/0092-8674(81)90413-X (1981).
65. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**, 1190–1195 10.1126/science.1222794 (2012).
66. Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics* **20**, 437–455 10.1038/s41576-019-0128-0 (2019).
67. Lim, F. *et al.* Affinity-optimizing enhancer variants disrupt development. *Nature* **626**, 151–159 10.1038/s41586-023-06922-8 (2024).
68. Levine, M., Cattoglio, C. & Tjian, R. Looping Back to Leap Forward: Transcription Enters a New Era. *Cell* **157**, 13–25 10.1016/j.cell.2014.02.009 (2014).
69. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nature Genetics* **51**, 1664–1669 10.1038/s41588-019-0538-0 (2019).
70. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10**, 1213–1218 10.1038/nmeth.2688 (2013).

71. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics* **20**, 207–220 10.1038/s41576-018-0089-8 (2019).
72. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 10.1038/nature11232 (2012).
73. Kornberg, R. D. & Lorch, Y. Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. *Cell* **98**, 285–294 10.1016/S0092-8674(00)81958-3 (1999).
74. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 10.1016/j.cell.2007.02.005 (2007).
75. Clapier, C. R. & Cairns, B. R. The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry* **78**, 273–304 10.1146/annurev.biochem.77.062706.153223 (2009).
76. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 10.1038/nature14590 (2015).
77. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes & Development* **25**, 2227–2241 10.1101/gad.176826.111 (2011).
78. Davidson, E. & Levin, M. Gene regulatory networks. *Proceedings of the National Academy of Sciences* **102**, 4935–4935 10.1073/pnas.0502024102 (2005).
79. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics* **13**, 613–626 10.1038/nrg3207 (2012).
80. Whyte, W. A. *et al.* Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* **153**, 307–319 10.1016/j.cell.2013.03.035 (2013).
81. Hu, S. *et al.* Transcription factor antagonism regulates heterogeneity in embryonic stem cell states. *Molecular Cell* **82**, 4410–4427.e12 10.1016/j.molcel.2022.10.022 (2022).
82. Csete, M. & Doyle, J. Bow ties, metabolism and disease. *Trends in Biotechnology* **22**, 446–450 10.1016/j.tibtech.2004.07.007 (2004).
83. Werner, S. L., Barken, D. & Hoffmann, A. Stimulus Specificity of Gene Expression Programs Determined by Temporal Control of IKK Activity. *Science* **309**, 1857–1861 10.1126/science.1113319 (2005).
84. Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The I κ B-NF- κ B Signaling Module: Temporal Control and Selective Gene Activation. *Science* **298**, 1241–1245 10.1126/science.1071914 (2002).
85. Adelaja, A. *et al.* Six distinct NF κ B signaling codons convey discrete information to distinguish stimuli and enable appropriate macrophage responses. *Immunity* **54**, 916–930.e7 10.1016/j.immuni.2021.04.011 (2021).

86. Baud, V. & Karin, M. Is NF- κ B a good target for cancer therapy? Hopes and pitfalls. *Nature Reviews Drug Discovery* **8**, 33–40 10.1038/nrd2781 (2009).
87. Sweeney, K. & McClean, M. N. Transcription factor localization dynamics and DNA binding drive distinct promoter interpretations. *Cell Reports* **42**, 112426 10.1016/j.celrep.2023.112426 (2023).
88. Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455**, 485–490 10.1038/nature07292 (2008).
89. Yissachar, N. *et al.* Dynamic Response Diversity of NFAT Isoforms in Individual Living Cells. *Molecular Cell* **49**, 322–330 10.1016/j.molcel.2012.11.003 (2013).
90. Batchelor, E., Loewer, A., Mock, C. & Lahav, G. Stimulus-dependent dynamics of p53 in single cells. *Molecular Systems Biology* **7**, 488 10.1038/msb.2011.20 (2011).
91. Evans, R. M. The Steroid and Thyroid Hormone Receptor Superfamily. *Science* **240**, 889–895 10.1126/science.3283939 (1988).
92. Wurtz, J.-M. *et al.* A canonical structure for the ligand-binding domain of nuclear receptors. *Nature Structural Biology* **3**, 87–94 10.1038/nsb0196-87 (1996).
93. Weikum, E. R., Liu, X. & Ortlund, E. A. The nuclear receptor superfamily: A structural perspective. *Protein Science* **27**, 1876–1892 10.1002/pro.3496 (2018).
94. Evans, R. M. & Mangelsdorf, D. J. Nuclear Receptors, RXR, and the Big Bang. *Cell* **157**, 255–266 10.1016/j.cell.2014.03.012 (2014).
95. Pawlak, M., Lefebvre, P. & Staels, B. General Molecular Biology and Architecture of Nuclear Receptors. *Current Topics in Medicinal Chemistry* **12**, 486–504 10.2174/156802612799436641 (2012).
96. Jin, P. *et al.* Nuclear receptors in health and disease: signaling pathways, biological functions and pharmaceutical interventions. *Signal Transduction and Targeted Therapy* **10**, 228 10.1038/s41392-025-02270-3 (2025).
97. Oakley, R. H. & Cidlowski, J. A. The biology of the glucocorticoid receptor: New signaling mechanisms in health and disease. *Journal of Allergy and Clinical Immunology* **132**, 1033–1044 10.1016/j.jaci.2013.09.007 (2013).
98. Wang, Y.-X. PPARs: diverse regulators in energy metabolism and metabolic diseases. *Cell Research* **20**, 124–137 10.1038/cr.2010.13 (2010).
99. Brown, G. Retinoic acid receptor regulation of decision-making for cell differentiation. *Frontiers in Cell and Developmental Biology* **11**, 1182204 10.3389/fcell.2023.1182204 (2023).
100. Haussler, M. R. *et al.* The Nuclear Vitamin D Receptor: Biological and Molecular Regulatory Properties Revealed. *Journal of Bone and Mineral Research* **13**, 325–349 10.1359/jbmr.1998.13.3.325 (1998).
101. Carlberg, C. & Molnar, F. Current Status of Vitamin D Signaling and Its Therapeutic Applications. *Current Topics in Medicinal Chemistry* **12**, 528–547 10.2174/156802612799436623 (2012).

102. Zhang, Y.-g., Wu, S. & Sun, J. Vitamin D, vitamin D receptor and tissue barriers. *Tissue Barriers* **1**, e23118 10.4161/tisb.23118 (2013).
103. Bikle, D. D. Vitamin D Metabolism, Mechanism of Action, and Clinical Applications. *Chemistry & Biology* **21**, 319–329 10.1016/j.chembiol.2013.12.016 (2014).
104. Haussler, M. R., Jurutka, P. W., Mizwicki, M. & Norman, A. W. Vitamin D receptor (VDR)-mediated actions of $1\alpha,25(\text{OH})_2$ vitamin D₃: Genomic and non-genomic mechanisms. *Best Practice & Research Clinical Endocrinology & Metabolism* **25**, 543–559 10.1016/j.beem.2011.05.010 (2011).
105. Bouillon, R. *et al.* Vitamin D and Human Health: Lessons from Vitamin D Receptor Null Mice. *Endocrine Reviews* **29**, 726–776 10.1210/er.2008-0004 (2008).
106. Norman, A. W. From vitamin D to hormone D: fundamentals of the vitamin D endocrine system essential for good health. *The American Journal of Clinical Nutrition* **88**, 491S–499S 10.1093/ajcn/88.2.491S (2008).
107. Hoenderop, J. G. J., Nilius, B. & Bindels, R. J. M. Calcium Absorption Across Epithelia. *Physiological Reviews* **85**, 373–422 10.1152/physrev.00003.2004 (2005).
108. McCann, L. M. & Beto, J. Roles of Calcium-Sensing Receptor and Vitamin D Receptor in the Pathophysiology of Secondary Hyperparathyroidism. *Journal of Renal Nutrition* **20**, 141–150 10.1053/j.jrn.2010.01.004 (2010).
109. Hewison, M. Vitamin D and immune function: an overview. *Proceedings of the Nutrition Society* **71**, 50–61 10.1017/S0029665111001650 (2012).
110. Penna, G. & Adorini, L. $1\alpha,25$ -Dihydroxyvitamin D₃ Inhibits Differentiation, Maturation, Activation, and Survival of Dendritic Cells Leading to Impaired Alloreactive T Cell Activation. *The Journal of Immunology* **164**, 2405–2411 10.4049/jimmunol.164.5.2405 (2000).
111. Liu, P. T. *et al.* Toll-Like Receptor Triggering of a Vitamin D-Mediated Human Antimicrobial Response. *Science* **311**, 1770–1773 10.1126/science.1123933 (2006).
112. Kongsbak, M., Levring, T. B., Geisler, C. & Von Essen, M. R. The Vitamin D Receptor and T Cell Function. *Frontiers in Immunology* **4** 10.3389/fimmu.2013.00148 (2013).
113. Morgan, J. W., Kouttab, N., Ford, D. & Maizel, A. L. Vitamin D-Mediated Gene Regulation in Phenotypically Defined Human B Cell Subpopulations*. *Endocrinology* **141**, 3225–3234 10.1210/endo.141.9.7666 (2000).
114. Van Belle, T. L. *et al.* $1,25$ -Dihydroxyvitamin D₃ and Its Analog TX527 Promote a Stable Regulatory T Cell Phenotype in T Cells from Type 1 Diabetes Patients. *PLoS ONE* **9** (ed Von Herrath, M. G.) e109194 10.1371/journal.pone.0109194 (2014).
115. Ysmail-Dahlouk, L., Nouari, W. & Aribi, M. $1,25$ -dihydroxyvitamin D₃ downmodulates the production of proinflammatory cytokines and nitric oxide and enhances the phosphorylation of monocyte-expressed STAT6 at the recent-onset type 1 diabetes. *Immunology Letters* **179**, 122–130 10.1016/j.imlet.2016.10.002 (2016).
116. Boutitah-Benyaich, I. *et al.* Multiple sclerosis: molecular pathogenesis and therapeutic intervention. *Signal Transduction and Targeted Therapy* **10**, 324 10.1038/s41392-025-02415-4 (2025).

117. Imani, D., Razi, B., Motallebnezhad, M. & Rezaei, R. Association between vitamin D receptor (VDR) polymorphisms and the risk of multiple sclerosis (MS): an updated meta-analysis. *BMC Neurology* **19**, 339 10.1186/s12883-019-1577-y (2019).
118. Oeckinghaus, A. & Ghosh, S. The NF- κ B Family of Transcription Factors and Its Regulation. *Cold Spring Harbor Perspectives in Biology* **1**, a000034–a000034 10.1101/cshperspect.a000034 (2009).
119. Liu, T., Zhang, L., Joo, D. & Sun, S.-C. NF- κ B signaling in inflammation. *Signal Transduction and Targeted Therapy* **2**, 17023 10.1038/sigtrans.2017.23 (2017).
120. Chen, Y. *et al.* Vitamin D Receptor Inhibits Nuclear Factor κ B Activation by Interacting with I κ B Kinase β Protein. *Journal of Biological Chemistry* **288**, 19450–19458 10.1074/jbc.M113.467670 (2013).
121. Szeto, F. L. *et al.* Involvement of the vitamin D receptor in the regulation of NF- κ B activity in fibroblasts. *The Journal of Steroid Biochemistry and Molecular Biology* **103**, 563–566 10.1016/j.jsbmb.2006.12.092 (2007).
122. Villaseñor, J., Besse, W., Benoist, C. & Mathis, D. Ectopic expression of peripheral-tissue antigens in the thymic epithelium: Probabilistic, monoallelic, misinitiated. *Proceedings of the National Academy of Sciences* **105**, 15854–15859 10.1073/pnas.0808069105 (2008).
123. Waddington, C. H. The Epigenotype. *International Journal of Epidemiology* **41**, 10–13 10.1093/ije/dyr184 (2012).
124. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315–326 10.1016/j.cell.2006.02.041 (2006).
125. Lara-Astiaso, D. *et al.* Chromatin state dynamics during blood formation. *Science* **345**, 943–949 10.1126/science.1256271 (2014).
126. Ghisletti, S. *et al.* Cooperative NCoR/SMRT interactions establish a corepressor-based strategy for integration of inflammatory and anti-inflammatory signaling pathways. *Genes & Development* **23**, 681–693 10.1101/gad.1773109 (2009).
127. Chen, T., Boyer, L. A. & Agarwal, D. Assessment of dispersion metrics for estimating single-cell transcriptional variability. *PLOS Computational Biology* **22** (ed Lin, Z.) e1014030 10.1371/journal.pcbi.1014030 (2026).
128. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biology* **11**, R106 10.1186/gb-2010-11-10-r106 (2010).
129. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology* **33**, 495–502 10.1038/nbt.3192 (2015).
130. Turner, S. J., Doherty, P. C., McCluskey, J. & Rossjohn, J. Structural determinants of T-cell receptor bias in immunity. *Nature Reviews Immunology* **6**, 883–894 10.1038/nri1977 (2006).
131. Wucherpfennig, K. W. & Strominger, J. L. Molecular mimicry in T cell-mediated autoimmunity: Viral peptides activate human T cell clones specific for myelin basic protein. *Cell* **80**, 695–705 10.1016/0092-8674(95)90348-8 (1995).

132. Sewell, A. K. Why must T cells be cross-reactive? *Nature Reviews Immunology* **12**, 669–677 10.1038/nri3279 (2012).
133. Derbinski, J. *et al.* Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. *The Journal of Experimental Medicine* **202**, 33–45 10.1084/jem.20050471 (2005).
134. St-Pierre, C., Trofimov, A., Brochu, S., Lemieux, S. & Perreault, C. Differential Features of AIRE-Induced and AIRE-Independent Promiscuous Gene Expression in Thymic Epithelial Cells. *The Journal of Immunology* **195**, 498–506 10.4049/jimmunol.1500558 (2015).
135. Romano, R. *et al.* FOXP1: A Master Regulator Gene of Thymic Epithelial Development Program. *Frontiers in Immunology* **4** 10.3389/fimmu.2013.00187 (2013).
136. Zhou, Q., Chipperfield, H., Melton, D. A. & Wong, W. H. A gene regulatory network in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences* **104**, 16438–16443 10.1073/pnas.0701014104 (2007).
137. Baskar, R. *et al.* Integrating transcription-factor abundance with chromatin accessibility in human erythroid lineage commitment. *Cell Reports Methods* **2**, 100188 10.1016/j.crmeth.2022.100188 (2022).
138. Chambon, P. A decade of molecular biology of retinoic acid receptors. *The FASEB Journal* **10**, 940–954 10.1096/fasebj.10.9.8801176 (1996).
139. Glass, C. K. & Saijo, K. Nuclear receptor transrepression pathways that regulate inflammation in macrophages and T cells. *Nature Reviews Immunology* **10**, 365–376 10.1038/nri2748 (2010).
140. Perissi, V. & Rosenfeld, M. G. Controlling nuclear receptors: the circular logic of cofactor cycles. *Nature Reviews Molecular Cell Biology* **6**, 542–554 10.1038/nrm1680 (2005).
141. Helsen, C. *et al.* Structural basis for nuclear hormone receptor DNA binding. *Molecular and Cellular Endocrinology* **348**, 411–417 10.1016/j.mce.2011.07.025 (2012).
142. Yoshida, H. *et al.* The cis-Regulatory Atlas of the Mouse Immune System. *Cell* **176**, 897–912.e20 10.1016/j.cell.2018.12.036 (2019).
143. Koidl, S. & Timmers, H. T. M. greenCUT&RUN: Efficient Genomic Profiling of GFP-Tagged Transcription Factors and Chromatin Regulators. *Current Protocols* **1**, e266 10.1002/cpz1.266 (2021).
144. Huxford, T. & Ghosh, G. A Structural Guide to Proteins of the NF- κ B Signaling Module. *Cold Spring Harbor Perspectives in Biology* **1**, a000075–a000075 10.1101/cshperspect.a000075 (2009).
145. Voss, T. C. & Hager, G. L. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature Reviews Genetics* **15**, 69–81 10.1038/nrg3623 (2014).
146. Meyer, M. B. & Pike, J. W. Corepressors (NCoR and SMRT) as well as coactivators are recruited to positively regulated 1 α ,25-dihydroxyvitamin D₃-responsive genes. *The Journal of Steroid Biochemistry and Molecular Biology* **136**, 120–124 10.1016/j.jsbmb.2012.08.006 (2013).

147. Yan, F., Powell, D. R., Curtis, D. J. & Wong, N. C. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biology* **21**, 22 10.1186/s13059-020-1929-3 (2020).
148. Aubin, A.-M. *et al.* The NOD Mouse Beyond Autoimmune Diabetes. *Frontiers in Immunology* **13**, 874769 10.3389/fimmu.2022.874769 (2022).
149. Niki, S. Alteration of intra-pancreatic target-organ specificity by abrogation of Aire in NOD mice. *Journal of Clinical Investigation* **116**, 1292–1301 10.1172/JCI26971 (2006).
150. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 10.1016/j.cell.2018.01.029 (2018).
151. Boyer, L. A. *et al.* Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell* **122**, 947–956 10.1016/j.cell.2005.08.020 (2005).
152. Zhang, P. *et al.* Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proceedings of the National Academy of Sciences* **96**, 8705–8710 10.1073/pnas.96.15.8705 (1999).
153. Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nature Immunology* **4**, 330–336 10.1038/ni904 (2003).
154. Shaffer, A. *et al.* Blimp-1 Orchestrates Plasma Cell Differentiation by Extinguishing the Mature B Cell Gene Expression Program. *Immunity* **17**, 51–62 10.1016/S1074-7613(02)00335-7 (2002).
155. Sacconi, S., Marazzi, I., Beg, A. A. & Natoli, G. Degradation of Promoter-bound p65/RelA Is Essential for the Prompt Termination of the Nuclear Factor κ B Response. *The Journal of Experimental Medicine* **200**, 107–113 10.1084/jem.20040196 (2004).
156. Li, Z. *et al.* Suppression of m6A reader Ythdf2 promotes hematopoietic stem cell expansion. *Cell Research* **28**, 904–917 10.1038/s41422-018-0072-0 (2018).
157. Patrick, R. *et al.* The activity of early-life gene regulatory elements is hijacked in aging through pervasive AP-1-linked chromatin opening. *Cell Metabolism* **36**, 1858–1881.e23 10.1016/j.cmet.2024.06.006 (2024).
158. Stein, B., Cogswell, P. C. & Baldwin, A. S. Functional and physical associations between NF-kappa B and C/EBP family members: a Rel domain-bZIP interaction. *Molecular and Cellular Biology* **13**, 3964–3974 10.1128/MCB.13.7.3964 (1993).
159. Fujioka, S. *et al.* NF- κ B and AP-1 Connection: Mechanism of NF- κ B-Dependent Regulation of AP-1 Activity. *Molecular and Cellular Biology* **24**, 7806–7819 10.1128/MCB.24.17.7806-7819.2004 (2004).
160. Rauscher, F. J., Voulalas, P. J., Franza, B. R. & Curran, T. Fos and Jun bind cooperatively to the AP-1 site: reconstitution in vitro. *Genes & Development* **2**, 1687–1699 10.1101/gad.2.12b.1687 (1988).
161. Hess, J., Angel, P. & Schorpp-Kistner, M. AP-1 subunits: quarrel and harmony among siblings. *Journal of Cell Science* **117**, 5965–5973 10.1242/jcs.01589 (2004).

162. Greenberg, M. E. & Ziff, E. B. Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. *Nature* **311**, 433–438 10.1038/311433a0 (1984).
163. Kallies, A. & Nutt, S. L. Transcription Factor Theft—PU.1 Caught Red-Handed. *Immunity* **48**, 1063–1065 10.1016/j.immuni.2018.05.007 (2018).
164. Hosokawa, H. *et al.* Transcription Factor PU.1 Represses and Activates Gene Expression in Early T Cells by Redirecting Partner Transcription Factor Binding. *Immunity* **48**, 1119–1134.e7 10.1016/j.immuni.2018.04.024 (2018).
165. Houseley, J. & Tollervey, D. The Many Pathways of RNA Degradation. *Cell* **136**, 763–776 10.1016/j.cell.2009.01.019 (2009).
166. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 10.1038/nature12730 (2014).
167. Liao, J. *et al.* Insight into the structure, physiological function, and role in cancer of m6A readers—YTH domain-containing proteins. *Cell Death Discovery* **8**, 137 10.1038/s41420-022-00947-0 (2022).
168. Meyer, K. D. & Jaffrey, S. R. Rethinking m⁶ A Readers, Writers, and Erasers. *Annual Review of Cell and Developmental Biology* **33**, 319–342 10.1146/annurev-cellbio-100616-060758 (2017).
169. Cheng, Y. *et al.* m6A RNA Methylation Maintains Hematopoietic Stem Cell Identity and Symmetric Commitment. *Cell Reports* **28**, 1703–1716.e6 10.1016/j.celrep.2019.07.032 (2019).
170. Xiao, Y. *et al.* Profiling of RNA-binding protein binding sites by in situ reverse transcription-based sequencing. *Nature Methods* **21**, 247–258 10.1038/s41592-023-02146-w (2024).
171. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 10.1016/j.cell.2019.05.031 (2019).
172. Eferl, R. & Wagner, E. F. AP-1: a double-edged sword in tumorigenesis. *Nature Reviews Cancer* **3**, 859–868 10.1038/nrc1209 (2003).
173. Shaulian, E. & Karin, M. AP-1 as a regulator of cell life and death. *Nature Cell Biology* **4**, E131–E136 10.1038/ncb0502-e131 (2002).
174. Ucar, O. & Rattay, K. Promiscuous Gene Expression in the Thymus: A Matter of Epigenetics, miRNA, and More? *Frontiers in Immunology* **6** 10.3389/fimmu.2015.00093 (2015).
175. Artusa, P. *et al.* Skewed epithelial cell differentiation and premature aging of the thymus in the absence of vitamin D signaling. *Science Advances* **10**, eadm9582 10.1126/sciadv.adm9582 (2024).
176. Hörlein, A. J. *et al.* Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* **377**, 397–404 10.1038/377397a0 (1995).