

**Transcriptional and epigenetic
mechanisms of promiscuous gene
expression by thymic epithelial cells**



Fatima Dhalla

Green Templeton College

University of Oxford

A thesis submitted for the degree of
DPhil in Biomedical and Clinical Sciences

Hilary 2019

Digital Thesis Submission

A digital copy of this thesis has been deposited in the Oxford University Research Archive (ORA). The figure quality, particularly that for microscopy images, is superior within the digital copy.

Abstract

The thymus is the primary lymphoid organ responsible for the generation and selection of a T-cell repertoire that is able to respond to foreign antigens whilst remaining tolerant to self. For the purposes of T-cell central tolerance induction thymic epithelial cells (TEC) express almost all protein coding genes in a process termed promiscuous gene expression (PGE), which is partly under the control of the Autoimmune Regulator (AIRE). Many aspects of the molecular mechanisms underlying PGE in TEC remain unclear. It has not yet been conclusively established whether PGE within single mTEC is a stochastic or coordinated process. The mechanisms that target AIRE to tissue-restricted genes (TRGs) are also incompletely defined, with previous studies suggesting that interaction partners of AIRE may localise AIRE to repressive epigenetic marks associated with TRGs in TEC. The first part of this study examines the transcriptional patterns associated with PGE in single mTEC. Analysis of the transcriptomes of 6,894 single mTEC revealed 50 robustly defined co-expression groups supporting the hypothesis that PGE in individual mTEC is a highly ordered process. The second part of this study addresses the role of a putative AIRE interaction partner, Chromodomain-helicase-DNA-binding protein 4 (CHD4), in TEC biology. Using a variety of cellular and molecular tools, CHD4 was demonstrated to play important roles in TEC development, architectural organisation and function. In particular, CHD4 was found to cooperate functionally with AIRE to induce the promiscuous expression of AIRE-induced TRGs found in particularly repressive chromatin contexts and was consequently required for proper T-cell central tolerance induction.

Acknowledgements

First and foremost I would like to express my sincerest gratitude to my supervisor, Georg Hollander, for sharing his great knowledge and enthusiasm; for his guidance and even his criticism; for the many opportunities he has given me to develop my career and for helping me to begin to cultivate independence in my academic practices. I would also like to thank my clinical supervisor, Smita Patel, for her unwavering mentorship and encouragement over the past six years. I look forward to more years working alongside them both.

Heartfelt thanks also go to the past and present members of the Hollander lab that I have had the pleasure of working alongside; in particular, to Stefano, Mary, and Stanley for teaching me so many practical skills and for being my second, third, and fourth pair of hands; and to Jeanette for introducing me to computational biology and for many hours of fruitful scientific discussion. For their collaboration over the past three years, I would also like to thank Chris Ponting, Katia Georgopoulos, Jakub Abramson and Ayelet Avin, as well as the staff in the animal facility and the Wolfson Imaging Centre.

Last but not least, thanks to my family, especially to my husband, mother and grandmother for their support, nourishment, willingness of ear, and most of all, for their unconditional love and infinite patience.

Funding

This work was supported by the Wellcome Trust [109032/Z/15/Z]

Contents

Abstract	i
Acknowledgements	ii
Funding	iii
List of Abbreviations	ix

1 Introduction	1
1.1 The thymus	1
1.2 Thymic epithelial cell development.....	3
1.3 Intrathymic T-cell development	8
1.4 T-cell central tolerance	12
1.4.1 Cellular mechanisms of T-cell central tolerance induction	12
1.4.2 Promiscuous gene expression	14
1.4.3 The Autoimmune Regulator	16
1.4.4 Molecular mechanisms of AIRE-induced gene expression	21
1.4.5 Transcriptional control of PGE.....	28
1.5 Chromodomain-Helicase-DNA-binding protein 4	34
1.5.1 CHD4 structure and function	34
1.5.2 CHD4 as a putative AIRE-interaction partner	39
1.6 Aims and hypotheses	41

2 Materials and Methods	43
2.1 Mice.....	43
2.1.1 C57BL/6	43
2.1.2 BALB/c	43
2.1.3 C57BL/6 x BALB/c F1 Hybrids	43
2.1.4 Chd4 ^{fl/fl}	43
2.1.5 Foxn1Cre.....	44
2.1.6 Rag1 ^{-/-}	44
2.1.7 Timed matings	44

2.1.8 Induction of lymphopaenia	45
2.2 Genotyping	46
2.2.1 DNA extraction from ear punches	46
2.2.2 DNA extraction from tail clips	46
2.2.3 Polymerase Chain Reaction	47
2.2.4 Agarose gel electrophoresis	48
2.3 Cellular techniques	49
2.3.1 Isolation of TEC from postnatal thymi	49
2.3.2 Enrichment of TEC from postnatal thymi via depletion of CD45+ cells ..	50
2.3.3 Isolation of TEC from embryonic thymi	50
2.3.4 Isolation of thymocytes and peripheral T-cells	51
2.3.5 Flow cytometry and fluorescence activated cell sorting	51
2.3.6 Annexin V binding flow cytometry assay	56
2.3.7 Aldefluor flow cytometry assay	56
2.3.8 Fetal thymic organ culture	57
2.4 Molecular techniques	59
2.4.1 Genomic excision PCR	59
2.4.2 RNA extraction	60
2.4.3 cDNA synthesis	61
2.4.4 Quantitative real-time PCR	62
2.4.5 Bulk RNA-sequencing	65
2.4.6 Single-cell RNA-sequencing	66
2.4.7 ChIP-sequencing	67
2.4.8 Western blot	71
2.4.9 Proximity Ligation Imaging Cytometry	72
2.5 Microscopy	74
2.5.1 Haematoxylin and Eosin Staining	74
2.5.2 Immunohistochemistry	75
2.5.3 Spatial analysis of GP2+ mTEC	77
2.5.4 Indirect immunofluorescence to look for autoantibodies	78
2.6 Statistical analysis	79

3 Promiscuous Gene Expression in Single mTEC	80
3.1 Introduction, aim and hypothesis	80
3.2 Generation of RNA-sequencing libraries from single mTEC	82
3.2.1 Isolation of mTEC promiscuously expressing Tetraspanin 9 or Glycoprotein 2	82
3.2.2 Generation of large-scale RNA-sequencing data from single mTEC	84
3.3 Cell subpopulation clustering of single mTEC	86
3.3.1 Transcriptional heterogeneity within the mTEC compartment	86
3.3.2 Transcriptional heterogeneity within the main body of single cells is reflective of mTEC maturation	88
3.3.3 Seven satellite clusters lie separately from the main mTEC maturation trajectory	96
3.4 Gene module clustering of single-cell RNA-sequencing data	99
3.4.1 The presence of 50 gene co-expression groups is supportive of PGE being an ordered process	99
3.4.2 Gene membership to co-expression groups is biologically indeterminate	105
3.4.3 Random spatial distribution of GP2+ mTEC within medullary islands .	107
3.5 Discussion	109
 4 CHD4 in TEC Development, Function and AIRE-induced Gene Expression	 115
4.1 Introduction, aims and hypothesis	115
4.2 Generation of mice with CHD4 deficiency targeted to TEC	116
4.3 Heterozygous targeting of <i>Chd4</i> in TEC is insufficient to cause a phenotype	120
4.4 Homozygous <i>Chd4</i> deficiency in TEC causes thymic hypocellularity .	124
4.5 CHD4 is important for normal TEC development	125
4.5.1 CHD4 deficiency causes abnormalities in the composition of the adult TEC compartment, particularly affecting the mTEC lineage	125
4.5.2 Changes in the composition of the adult TEC compartment of CHD4 deficient thymi cannot be explained by abnormal apoptosis	127
4.5.3 Increased cell cycling may contribute to the relative increase in cTEC in CHD4 deficient thymi	129

4.5.4 The TEC defects are established in the embryonic thymus and are mostly maintained during postnatal life	130
4.5.5 E17.5 CHD4 deficient TEC express higher levels of markers associated with progenitor status	133
4.5.6 E17.5 CHD4 deficient FTOC are able to respond to exogenous Receptor activator of nuclear factor kappa-B ligand (RANK-L) but this does not rescue the cKO phenotype.....	136
4.5.7 CHD4 deficiency in TEC causes abnormal thymic morphology	138
4.6 The distorted thymic microenvironment composed of TEC devoid of CHD4 expression is unable to support regular thymocyte maturation and selection	142
4.6.1 TEC conditional CHD4 deficiency causes defective maturation of thymocytes	142
4.6.2 TEC conditional CHD4 deficiency causes defective positive selection	144
4.6.3 CHD4 deficient TEC are unable to support regular post-selectional maturation of SPCD4 thymocytes	146
4.6.4 TEC conditional CHD4 deficiency causes an increase in wave 1 and 2 negatively selected thymocytes.....	149
4.6.5 TEC conditional CHD4 deficiency causes peripheral T-cell lymphopaenia with skewing towards a memory phenotype	152
4.6.6 TEC conditional CHD4 deficiency causes peripheral expansion of T-regulatory cells	154
4.7 CHD4 deficiency causes dysregulated expression of genes belonging to several pathways important for TEC development and function	156
4.7.1 Generation of population level RNA-sequencing libraries	156
4.7.2 Deficiency of CHD4 in TEC causes dysregulated gene expression in all TEC subpopulations examined.....	160
4.7.3 CHD4 deficient cTEC show dysregulated expression of genes related to retinoic acid metabolism, insulin-like growth factor signalling, extracellular matrix homeostasis	166
4.7.4 CHD4 deficiency in TEC causes abnormal composition of the thymic extracellular matrix	167
4.7.5 Exogenous retinoic acid partially rescues the phenotype caused by TEC conditional CHD4 deficiency.....	171

4.7.6 Exogenous IGF-1 does not affect the cellularity or phenotype of either WT or cKO FTOCs	180
4.8 Altered AIRE-regulated PGE in CHD4 deficient TEC	182
4.8.1 RNA-sequencing reveals defective AIRE-regulated PGE in CHD4 deficient CD80 ⁺ CD86 ⁺ mTEC	182
4.8.2 Gene co-expression groups are lost in CHD4 deficient CD80 ⁺ CD86 ⁺ mTEC	187
4.8.3 CHD4 and AIRE do not directly physically interact	190
4.8.4 CHD4 binds chromatin at the same sites as AIRE and induces the expression of genes with a repressive chromatin context	192
4.8.5 CHD4 deficiency in TEC causes defective T-cell central tolerance induction	197
4.9 Discussion	202
 5 Final Discussion	213
 References	223
 Supplementary Figures	272
Supplementary Figure 1:	272
Supplementary Figure 2:	273
Supplementary Figure 3:	273
Supplementary Figure 4:	274

List of Abbreviations

2DG	2-deoxyguanosine
AIRE	Autoimmune regulator
AMI	Adjusted mutual information
APC	Antigen presenting cell
APECED	Autoimmune polyendocrinopathy candidiasis ectodermal dysplasia
APS1	Autoimmune polyglandular syndrome type 1
ATAC	Assay for transposase-accessible chromatin
ATP	Adenosine triphosphate
Bp	Base pairs
CARD	Caspase recruitment domain
CD	Cluster of differentiation
CHD	Chromodomain helicase DNA-binding protein
ChIP	Chromatin immunoprecipitation
cKO	Conditional knockout
CMJ	Cortico-medullary junction
cTEC	Cortical thymic epithelial cell(s)
DAPI	4',6-diamidino-2-phenylindole
DC	Dendritic cell
DDR	DNA damage response
DEAB	Diethylaminobenzaldehyde
DLL4	Delta-like 4
DN	Double negative
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DP	Double positive
DSG3	Desmoglein 3
E	Embryonic day
EDTA	Ethylenediaminetetraacetic acid
EpCAM	Epithelial cell adhesion molecule
FACS	Fluorescent activated cell sorting
FCS	Fetal calf serum
FN1	Fibronectin
FOXP1	Forkhead box N1

FOXP3	Forkhead box P3
FTOC	Fetal thymic organ culture
GP2	Glycoprotein 2
H&E	Haematoxylin and eosin
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
IGF	Insulin-like growth factor
IHC	Immunohistochemistry
IL	Interleukin
IP	Immunoprecipitation
IVL	Involucrin
K	Cytokeratin
Kb	Kilobases
LDS	Liberase digestion solution
LTβR	Lymphotoxin β receptor
MACS	Magnetic activated cell sorting
MHC	Major histocompatibility complex
mTEC	Medullary thymic epithelial cell(s)
NaCl	Sodium chloride
NLS	Nuclear localisation signal
NURD	Nucleosome remodelling deacetylase
P_{adj}	Adjusted P value
PAE	Peripheral antigen expressing
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PGE	Promiscuous gene expression
PHD	Plant homeobox domain
PLIC	Proximity ligation imaging cytometry
PRC2	Polycomb repressive complex 2
PRR	Proline-rich repeat
RA	Retinoic acid
RAG	Recombinase activating gene
RANK	Receptor activator of NFκB
RANK-L	RANK-ligand
RNA	Ribonucleic acid

rpm	Revolutions per minute
RT-qPCR	Real time quantitative PCR
S1P	Sphingosine-1-phosphate
SAND	SP100, AIRE, NUCP41/P75 and DEAF1
SCA1	Stem cell antigen 1
SDS	Sodium dodecyl sulphate
Seq	Sequencing
SIRT1	Sirtuin-1
SNN	Shared nearest neighbour
SP	Single positive
T-Reg	Regulatory T-cell
Ta	Annealing temperature
Taq	<i>Thermus aquaticus</i>
TBE	Tris-Borate-EDTA
TCR	T-cell receptor
TE	Tris-EDTA
TEC	Thymic epithelial cell
TES	Transcriptional end site
TNC	Tenascin C
TNFSF	Tumour necrosis factor superfamily
TNFSFR	Tumour necrosis factor superfamily receptor
TRA	Tissue restricted antigen
TRG	Tissue restricted gene
tSNE	t-Distributed Stochastic Neighbour Embedding
TSPAN8	Tetraspanin 8
TSS	Transcriptional start site
UEA-1	<i>Ulex europaeus</i> agglutinin I
WT	Wild type

1 Introduction

1.1 The thymus

The thymus is the primary lymphoid organ responsible for the generation and selection of a T-cell repertoire that is able to respond to foreign antigens whilst remaining tolerant to self (1, 2).

During evolution, the thymus developed alongside adaptive immunity in vertebrate phylogeny (2, 3). The hallmarks of an adaptive immune system include the ability to generate immunological memory, a process that is reliant on clonal expansion and cell differentiation, and that allows secondary encounters with the same antigen to be met with a more rapid and robust immune response (3, 4). Another key difference between the innate and adaptive immune systems lies in the receptors they use to recognise their targets. Innate immune cells use germ line encoded pattern recognition receptors to recognise broad classes of molecules and are unable to discriminate between self and non-self (4). On the other hand, although at the population level T- and B-lymphocytes of the adaptive immune system are able to recognise a vast diversity of antigens, at the single cell level each individual lymphocyte bears a receptor of fine antigen specificity and is able to discriminate between self and non-self molecules (4). The antigen receptors of T- and B-cells are not germ line encoded; instead the enormous variety is generated via somatic recombination of variable (V), diversity (D) and joining (J) gene segments during lymphocyte development (2-4). As this process is somewhat random, it has required the coordinate development of stringent

quality control or selection processes to ensure that the resulting antigen receptors are functional but not overtly autoreactive. The thymus fulfils this purpose for the T-cell arm of the adaptive immune system (2, 3).

During organogenesis the thymus develops from the pharyngeal pouches, out-pouchings of endodermal foregut. In the mouse, the thymus and the parathyroids originate from the 3rd pharyngeal pouch in a process initiated at embryonic day (E) 9.5, which results in the formation of a separate thymic anlage by E10.5 (1, 2, 5, 6). Most mammals have a single bi-lobed thymus, which migrates during development to occupy a midline position within the thoracic cavity just above the heart (2, 5).

In terms of cellular composition, the thymus is mainly comprised of developing thymocytes (5, 7). It is also structurally and functionally dependent on populations of stromal cells, the main constituent of which are endodermally derived thymic epithelial cells (TEC) (5, 8). Other components of the thymic stroma include mesenchymal cells, endothelial cells, fibroblasts, cells that form the outer capsule and septa, as well as other haematopoietic lineage cells including dendritic cells (DCs), macrophages, and B-cells (2, 7).

TEC form a specialised three-dimensional scaffold and can be categorised into cortical (c-) and medullary (m-) lineages based on their anatomical, molecular, and functional characteristics (1, 9-11). As a composite with other stromal cell types, TEC form a unique microenvironment able to commit haematopoietic

cells to a T-cell fate, support their differentiation to functional maturity and allow the selection of a T-cell receptor (TCR) repertoire bespoke for an individual organism (9-11). Importantly lymphoid precursors and their progeny also provide signals to TEC that instruct their differentiation and functional ability to support thymocyte development. This essential bidirectional interaction between thymic stromal and lymphoid cells has been designated “crosstalk” (9-13).

1.2 Thymic epithelial cell development

During embryonic thymic development, cells of the c- and mTEC lineages develop from a common bipotent progenitor, originating from the 3rd pharyngeal pouch endoderm (8, 10, 14-17). cTEC-like progenitors able to give rise to both cTEC and mTEC, and mTEC lineage restricted progenitors have also been identified leading to the proposal of the “serial progression” model of TEC development, in which bipotent TEC progenitors first acquire expression of cTEC lineage markers, such as CD205 and $\beta 5t$, before committing to either a c- or mTEC fate via lineage specific progenitors (10, 18-24). The serial progression model comprehensively integrates the findings of many studies conducted to date aiming to define TEC developmental pathways. However, to add a further layer of complexity, there additionally exists considerable evidence for distinct fetal and postnatal programmes underlying TEC development and maintenance (7, 10, 25).

During murine embryonic development the 3rd pharyngeal pouch forms at around E9.5, with the parathyroid and thymus domains being distinguishable by E10.5 via the expression of *Gcm2*, which marks the former (5, 6, 26). Early patterning of the 3rd pharyngeal pouch is under the control of a number of transcription factors, including HOXA3, EYA1, PAX1 and -9, SIX1 and -4, and TBX1 (27-34). From E11.25 the expression of *Foxn1*, the master transcriptional regulator of TEC development, can be detected within the thymus domain in the ventral aspect of the 3rd pharyngeal pouch (5, 6, 35, 36). Although bipotent TEC progenitors do develop in the absence of FOXN1, FOXN1 is required for their subsequent differentiation into the c- and mTEC lineages in the embryo, as well as for their maturation, function and maintenance in the postnatal thymus (26, 36-40). Its critical importance in thymic organogenesis is demonstrated by animal models and human disease, in which null mutations cause athymia resulting in severe T-cell immunodeficiency, together with a “nude” or hairless phenotype (41-47). The parathyroids and thymic lobes separate from the pharynx and each other starting from about E12.5, after which the two organs migrate separately towards their final anatomical positions; for the thymus this involves caudal and medial migration such that the two lobes meet in the anterior mediastinum just above the heart (5).

Direct evidence for the presence of bipotent TEC progenitors in the fetal thymus first came from a lineage tracing study in which single yellow fluorescent protein positive (YFP⁺) epithelial cells, isolated from E12 fetal

thymi, were demonstrated to develop into both c- and mTEC after injection into non-labelled E12 thymic lobes (14). These bipotent progenitor cells were found to express Placenta-expressed transcript 1 (PLET1), a marker recognised by the antibodies MTS20 and MTS24, which had previously been identified to define a subpopulation of TEC with progenitor capacity (14, 16, 17, 48). At the same time, cells with bipotent progenitor potential were identified in the postnatal thymus in a study that used a Cre-ERT2 system, allowing the reactivation of a null *Foxn1* allele in single TEC (15). This study demonstrated that single TEC progenitor cells present in the postnatal thymus could also give rise to functional c- and mTEC (15). Several other markers have been associated with cells of bipotent TEC progenitor potential including stem cell antigen 1 (SCA1) and α_6 -integrin (49, 50).

The identification of embryonic TEC expressing classical cTEC markers, such as CD205, $\beta 5t$, IL-7 and Delta-like4 (DLL4), that can give rise to both c- and mTEC lineages suggests that TEC pass through a cTEC-like stage before committing to a particular TEC fate (19, 51, 52). Downstream of this stage in TEC differentiation, there exists considerable evidence for the presence of identifiable mTEC lineage committed precursors. The identification of specific cells with progenitor status limited to the cTEC lineage has, however, been more elusive (10, 18). mTEC committed progenitors in the fetal thymus have been found to express the tight junction proteins claudin-3 and -4 (CLD3/4) together with SSEA-1, or CD24 (23, 24, 53). Although CLD3/4⁺SSEA-1⁺ mTEC have been detected in the postnatal thymus in mice at 4 weeks of age, their

capacity to generate mature mTEC is substantially reduced with respect to their fetal counterparts (24). Within the postnatal thymus, however, a further subpopulation of TEC with progenitor activity restricted to the mTEC lineage, named “junctional TEC” (jTEC) has been described; these cells reside at the cortico-medullary junction (CMJ) and express podoplanin (54).

Following commitment to the mTEC lineage, several successive stages of mTEC maturation have been identified (18, 55). Immature mTEC express low levels of MHCII and CD80 and have not yet expressed the Autoimmune regulator (AIRE), a transcriptional regulator that plays a critical role in central T-cell tolerance induction; these immature mTEC are therefore denoted $\text{MHCII}^{\text{Lo}}\text{CD80}^{\text{Lo}}\text{AIRE}^-$ (55, 56). As they mature, mTEC upregulate expression of MHCII, CD80 and AIRE ($\text{MHCII}^{\text{Hi}}\text{CD80}^{\text{Hi}}\text{AIRE}^+$) (55, 56). The expression of these key molecules for antigen presentation, T-cell co-stimulation, and self-antigen expression make mature mTEC the most functionally competent stage in the process of T-cell tolerance induction (57). Finally, mTEC undergo a process of terminal differentiation, during which they downregulate expression of MHCII, CD80 and AIRE, once again adopting an $\text{MHCII}^{\text{Lo}}\text{CD80}^{\text{Lo}}\text{AIRE}^-$ phenotype (55, 58-61). These terminally differentiated mTEC, however, can be distinguished from immature mTEC by their expression of markers associated with keratinocyte terminal differentiation, including involucrin (IVL) and cytokeratin 10 (K10), as well as reactivity to the lectin *Tetragonolobus purpureus* agglutinin (TPA) (55, 59, 60, 62).

During fetal thymic organogenesis, the developmental programmes of both TEC and thymocytes unravel in parallel and are mutually dependent upon each other. Bidirectional interactions (“crosstalk”) between developing thymocytes and stromal cells are critical for thymic organogenesis, TEC differentiation, maturation and postnatal maintenance, and consequently proper intrathymic T-cell development and selection (9-13). The development of the mTEC compartment, for example, is dependent on crosstalk signals from the Tumour necrosis factor superfamily (TNFSF) of molecules, which activate the NFκB family of transcription factors downstream of cell surface receptors expressed on mTEC (13, 63, 64). Receptor activator of NFκB (RANK)::RANK-Ligand (RANK-L) interactions are known to be important for the differentiation and maturation of mTEC, beginning in the embryonic thymus and likely extending into the postnatal thymic environment (22, 64, 65). RANK is expressed by cells of the mTEC lineage from the mTEC progenitor stage (53), and its ligand (RANK-L) is provided by various intrathymic cell types including lymphoid tissue inducer cells (22), γδ T-cells that express an invariant Vγ5 TCR (66), and positively selected αβ thymocytes (67, 68). mTEC also express CD40, another TNFSF receptor (TNFSFR) (63, 64). Although its expression by mTEC can be detected in the fetal thymus, its ligand, CD40L, is only provided within the thymic microenvironment by positively selected single positive CD4 (SPCD4) thymocytes just before birth (65, 68, 69). Accordingly, studies of mice singly and doubly deficient in RANK and CD40 have shown that whilst signalling via RANK is essential for fetal mTEC maturation, cooperation between the RANK and CD40 pathways is required for postnatal

mTEC maturation (63-65). Another TNFSFR implicated in thymic development and cross talk is the lymphotoxin β receptor (LT β R) expressed on mTEC and DCs (63, 64). Mice deficient in LT β R or its ligands, which are expressed on single positive (SP) thymocytes, have defects in thymic medullary architecture and mTEC terminal differentiation (13, 63, 64, 70, 71).

1.3 Intrathymic T-cell development

After the initial establishment of the thymic vasculature, blood-borne, bone marrow derived lymphoid progenitors enter the thymus via the vasculature at the CMJ (72). This process is mediated by interactions between P-selectin, which is expressed by thymic endothelial cells, and P-selectin glycoprotein ligand 1 (PSGL1), expressed on the surface of the lymphoid progenitors (73). Upon entry into the thymus, lymphoid progenitors are initially defined as double negative (DN) due to their overt lack of expression of the T-cell co-receptor molecules CD4 and CD8 (CD4⁻CD8⁻). This stage of thymocyte development is further subdivided in the mouse into four stages based on the expression of CD44 and CD25, namely DN1 (CD44⁺CD25⁻), DN2 (CD44⁺CD25⁺), DN3 (CD44⁻CD25⁺) and DN4 (CD44⁻CD25⁻) stages. DN thymocytes migrate from the CMJ to the subcapsular region of the thymic cortex where they progressively commit to a T-cell fate under the influence of signals mediated by the notch ligand, DLL4 (74, 75), and IL-7 (76, 77), which are expressed by cTEC (11, 52, 75). Expression of recombination-activating genes (*Rag1* and *2*) at the DN2 stage of thymocyte maturation induces the process of somatic V(D)J recombination initially at the *Tcrb*, *g* and *d* loci (78). Productive

rearrangement at the *Tcrb* locus and the assembly of TCR β with pre-TCR α chains at the DN3 stage leads to cell surface expression of the pre-TCR complex (78, 79). Ligand-independent signalling via the pre-TCR constitutes an important developmental checkpoint during thymocyte maturation, which is known as β -selection and commits the cells to an $\alpha\beta$ T-cell fate (11, 78-81). Signalling via the pre-TCR induces progression from the DN3 stage to DN4. At this stage *Rag* genes are re-expressed leading to rearrangement at the *Tcra* locus, and the TCR co-receptor molecules, CD8 and CD4, are expressed allowing the cells to progress to the double positive (DP: CD4⁺CD8⁺) stage of thymocyte development; in some mouse strains this occurs via an immature single positive CD8 (iSPCD8: CD4⁻CD8⁺TCRb^{lo}) stage (78, 79, 82).

Because the process of V(D)J recombination produces TCRs of random specificity, maturing thymocytes are subjected to rigorous sequential selection processes mediated by the recognition of self-peptide::MHC complexes presented by TEC (83). These selection processes, termed positive and negative selection, collectively provide important bespoke checkpoints sampling the utility of individual TCR specificities. Thymic selection ensures that only those cells that bear a TCR that is useful (i.e. capable of recognising antigen in the context of self-MHC), but not harmful for a given individual (i.e. not strongly autoreactive), are permitted to complete their intrathymic development and egress the thymus (9, 83).

Newly generated DP thymocytes within the thymic cortex initially express low levels of TCR $\alpha\beta$; these are first tested for their specificity via interactions with cTEC (9). Positive selection occurs when thymocytes, able to recognise self-peptide::MHC complexes via their TCRs with low avidity, are signalled to survive and continue to mature towards the SP stage of thymocyte development, via a CD4⁺CD8^{lo} stage (9, 11, 83-85). Thymocytes that fail to bind to self-peptide::MHC do not receive these survival signals and die via apoptosis in a process known as “death by neglect” (9, 11, 83). Finally, thymocytes that react too strongly to self-antigen also die via apoptosis in a first wave of negative selection that occurs within the thymic cortex (9, 11, 86).

The process of positive selection instructs commitment into either SPCD4 (CD4⁺CD8⁻) or SPCD8 (CD4⁻CD8⁺) lineages. DP thymocytes that recognise self-peptide in the context of MHCI are induced to differentiate into SPCD8 cells, whereas those that recognise self-peptide in the context of MHCII differentiate along the SPCD4 lineage (85). This is influenced by the unique antigen processing pathways operational in cTEC (87). In terms of MHCI antigen processing, the thymoproteasome, which contains a cTEC specific catalytic subunit called $\beta 5t$, produces a unique repertoire of altered self-peptides for presentation to developing thymocytes (88-90). Importantly, deletion of $\beta 5t$ results in impaired positive selection of SPCD8 thymocytes, in terms of both quantity and quality; whilst the SPCD4 compartment is unaffected (88, 89). cTEC also express distinct proteases involved in antigen processing and loading onto MHCII, such as Cathepsin L and thymus specific serine

protease (TSSP), deficiency of either of these proteases in mice causes impaired positive selection of SP^{CD4} thymocytes (87, 90-93).

Following positive selection, SP thymocytes express the chemokine receptor CCR7, which induces their movement into the thymic medulla in response to CCL19 and CCL21, produced by mTEC (9, 11, 94, 95). Once positioned in the thymic medulla, SP thymocytes are subject to additional selection processes in order to remove any overtly autoreactive cells that have escaped cortical negative selection and secure central tolerance induction; they subsequently undergo further post-selectional maturational steps before egressing to the periphery (11).

The mechanisms of central (i.e. thymic) tolerance induction are discussed in more detail below (Section 1.4). In brief, mTEC and DCs present antigens to SP thymocytes such that those that bear a high affinity TCR for self-antigens are deleted or “negatively selected”, whereas those with intermediate or low affinity are allowed to survive and differentiate into regulatory (T-Reg) and effector T-cells respectively (9, 11, 87, 96). Post-selectional maturation of SP thymocytes within the thymic medulla is permitted in those cells that have survived the checkpoints of both positive and negative selection. This process can be followed by examining the expression of several cell surface markers including CD69 and CD24, which are progressively down-regulated as the cells mature, and CD62L and Qa2 which are upregulated in the most mature SP stages (11, 97-99). Finally, mature SP thymocytes upregulate cell surface

expression of sphingosine-1-phosphate (S1P) receptor 1 and egress the thymus via the perivascular space under the influence of an S1P gradient (100-102).

Only a minority of developing thymocytes (1-3%) successfully fulfil the stringent selection criteria described above and are therefore able to exit the thymus (103, 104). This rigorous process establishes a T-cell repertoire equipped to respond to foreign antigens yet tolerant to self and thus restrained from eliciting autoimmunity (11).

1.4 T-cell central tolerance

1.4.1 Cellular mechanisms of T-cell central tolerance induction

Immunological tolerance is defined as a state of unresponsiveness to antigen, with tolerance to self-antigens being critical for the prevention of autoimmune disease (105). Mechanisms of immunological tolerance can be categorised by the cell type affected (i.e. T- or B-cell tolerance), by the site of tolerance induction (i.e. central / within primary lymphoid tissues or peripheral), and/or by the cellular mechanisms involved (105).

Central T-cell tolerance is induced during the process of intrathymic T-cell development and serves to prevent the export of autoreactive T-cell clones to the periphery where they could elicit autoimmune disease (105). As stated above, this quality control checkpoint is essential given that the process of V(D)J recombination produces TCRs of random specificities, which will include

those that are sufficiently reactive against self-antigen to cause autoimmune disease if left unchecked.

The mechanisms of tolerance induction operational within the thymus include deletional tolerance or negative selection, in which thymocytes that bear high affinity TCRs for self-antigens are clonally deleted, and clonal diversion, in which thymocytes with intermediate affinity TCRs for self are diverted to a regulatory T-cell (T-Reg) fate (9, 11, 87, 96, 106). Central T-cell tolerance induction occurs in two waves that take place within the thymic cortex and medulla respectively (9, 11, 86). Within the thymic medulla, mTEC and thymic DCs are the key cellular determinates of tolerance induction, however a possible role for intra-thymic B-cells is also recognised (87, 107-109). mTEC possess specialised transcriptional mechanisms enabling them to express and present, in the context of self MHC and alongside co-stimulation, a comprehensive representation of peripheral self-antigens to developing SP thymocytes for the purposes of central tolerance induction (110, 111). This phenomenon, designated “promiscuous gene expression” (PGE) is discussed in more detail below (see 1.4.2). The thymic DC pool consists of intrathymically generated DCs, as well as DCs that have migrated from the periphery (87, 112). Studies suggest that intrathymically generated DCs are responsible for cross-presentation of self-antigens that are expressed by mTEC and released during apoptosis or within exosomes (87, 113, 114). Migratory DCs, on the other hand, are thought to present antigens picked up in the periphery, including those not expressed by mTEC and possibly also harmless foreign

antigens, for example those derived from food or commensal organisms (87, 115). Both mTEC and thymic DCs are able to participate in negative selection as well as T-Reg induction, and signal strength is thought to be a main determinant of thymocyte fate in this context (87, 106). Whilst negative selection serves to remove from the emerging repertoire of antigen specificities those T-cell clones that bear potentially autoreactive TCRs, the intrathymic generation of CD4⁺CD25⁺FOXP3⁺ T-Regs provides a second, dominant mechanism of preventing inappropriate immune reactions towards self-antigens as these cells exercise immunoregulatory and suppressive functions over autoreactive T-cells in the periphery (106).

1.4.2 Promiscuous gene expression

The ability of the thymus to ectopically or promiscuously express genes characteristically expressed in specific peripheral organs was first noted over two decades ago in studies involving transgenic mice where the transgene expression was driven from tissue specific promoters, such as the rat insulin promoter, RIP (116-119). Indeed, thymic expression of endogenous peripheral tissue restricted genes (TRGs) was later confirmed in both human and rodent thymi (118, 120-122). Before their identity was known, the cells responsible for the promiscuous expression of tissue specific antigens in the thymus were named “peripheral antigen expressing” (PAE) cells (119). These cells were noted to have features of professional antigen presenting cells (APCs) because of their expression of MHCII (119), and initial attempts to establish the cellular identity of PAE cells variously ascribed PGE within the thymus to DCs,

macrophages and mTEC (123-126). Eventually a more comprehensive study by Derbinski *et al*, which used qPCR to determine TRG expression within FACS sorted cTEC, mTEC, intrathymic DCs and macrophages, identified mTEC as the key cell type responsible for PGE (110). Although they did detect more limited expression of TRGs within the other thymic cell types analysed, the mTEC compartment was unrivalled in terms of breadth of TRG expression and abundance of transcripts for individual TRGs (110).

Through the use of genome-wide transcriptomic analyses, it is now known that, for the process of negative selection to be comprehensive, TEC as a population express over 90% of all protein-coding genes (111, 127), including those that are tissue- or temporally-restricted in organs other than the thymus (110, 128). This uniquely broad transcriptional programme, known as “promiscuous gene expression” (PGE), provides within the thymus a molecular mirror of the body’s own antigens against which developing T-cells can be selected. The precise molecular mechanisms enabling PGE are not clear but are, in part, under the control of the Autoimmune Regulator, AIRE, which regulates the expression of thousands of TRGs in mTEC (111, 129). When AIRE is lacking from mTEC the expression of about 4000 TRGs remains unchanged indicating that there are also AIRE-independent mechanisms that control PGE, though the molecular means by which this is achieved remain largely unknown (111).

The critical importance of PGE within the thymus for central tolerance induction is exemplified by studies that have found that even the loss of expression of a single TRG within the mTEC compartment can lead to the development of organ-specific autoimmunity (130-133). Furthermore, a broader defect in PGE caused by loss-of-function mutations in murine or human *AIRE*, leads to early onset multi-organ autoimmune disease as described in more detail below (134-144).

1.4.3 The Autoimmune Regulator

The transcriptional facilitator AIRE is indispensable for the induction of central T-cell tolerance as evidenced by human diseases and animal models in which loss-of-function mutations have been shown to lead to multisystem autoimmunity and reduced PGE by mTEC (134-144).

The gene encoding human *AIRE* was first identified by positional cloning in 1997 in patients with Autoimmune Polyglandular Syndrome 1 type (APS1), which is also known as Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy (APECED) syndrome (134, 135). APS1 is characterised by early onset autoimmune disease and the presence of multiple autoantibodies (134-136, 138). The classical clinical triad consists of chronic mucocutaneous candidiasis, resulting from the neutralising serum autoantibodies against IL-17 and IL-22 (145, 146), hypoparathyroidism and primary adrenal insufficiency (also known as Addison's disease) (137, 139, 140). This presentation is often accompanied by additional autoimmune

manifestations, such as primary gonadal failure, type 1 diabetes mellitus, pernicious anaemia, autoimmune thyroid disease and autoimmune hepatitis, as well as features of ectodermal dystrophy such as tooth enamel hypoplasia and nail dystrophy, which may not have an autoimmune pathogenesis (137, 139, 140). Although APS1 was originally ascribed to autosomal recessive mutations in *AIRE* (147-153), monoallelic mutations within the molecule's PHD1 (plant homeobox 1) and SAND (SP100, AIRE, NucP41/P75 and DEAF1) domains, which exert a dominant negative effect, have recently been described in patients with attenuated phenotypes and incomplete penetrance (154-156).

The discovery of the human *AIRE* gene paved the way for the identification of the mouse homologue, *Aire* (157), and the study of experimental mouse models. AIRE deficiency in mice does not faithfully recapitulate the clinical and immunological manifestations of human APS1 (141-143). Disease manifestations in mouse models also vary depending on genetic background, with animals on a C57BL/6 background displaying a mild phenotype, whereas those on a NOD background are particularly severely affected (144). Perhaps not surprisingly these phenotypic variances have been partly attributed to differences in MHC as well as environmental factors (139, 140, 144). Despite the heterogeneity between humans and murine models, the study of AIRE deficient mice has helped to further define the role of AIRE in central tolerance induction, principally via its ability to induce PGE in mTEC (111, 142).

The function of AIRE in the induction of central T-cell tolerance is predominantly linked to its expression in a subset of mTEC (110, 142, 158, 159). Although the functional relevance is not yet clear, AIRE expression has also been described in other intrathymic cells including DCs (158) and B-cells (107), in secondary lymphoid tissues in so-called “extra-thymic AIRE-expressing cells” (ETACs) (160), as well as in the testes (161) and early during embryogenesis before gastrulation (58). Within the TEC compartment, AIRE is expressed in about 50% of mature MHCII^{Hi}CD80^{Hi} mTEC (57, 58, 162). Its expression is regulated at the transcriptional level by several mechanisms. The proximal promoter of *Aire* contains an NFκB-responsive enhancer, termed conserved noncoding sequence 1 (CNS1), which is activated by RelA containing complexes induced by RANK::RANK-L signaling in mTEC (163). The expression of *Aire* in mTEC also involves specific demethylation of the proximal promoter and exon 2, and the relief of CTCF insulation, as well as the action of a transcription factor complex containing IRF4, IRF8, TBX21 and TCF7(164). In addition, a post-transcriptional mechanism has also been described that modulates AIRE expression via the action of JMJD6. The study of mice deficient in JMJD6 has suggested that AIRE expression is regulated on the one hand by retention of intron 2, which serves to limit AIRE protein expression, and on the other hand by relief of intron 2 retention, a process mediated by the action of JMJD6 presumably involving activation of splicing regulatory factors (165).

The primary function of AIRE in mTEC is in the process of central tolerance induction via the induction of PGE in mTEC. This enables mTEC to present a molecular library of self-antigens to developing thymocytes and to negatively select autoreactive thymocytes that recognise self-antigens too strongly (166-168). Cross-presentation of self-antigens by intrathymic DCs has also been described to contribute to this process but is also reliant on the process of PGE by mTEC (169). Studies examining the transcriptomes of AIRE deficient mTEC have revealed the extent of AIRE-regulated PGE in mTEC. Sansom *et al* found that AIRE facilitates the expression of just under 4000 genes in mTEC, the majority of which are classified as TRGs (111). 533 of these were entirely dependent on AIRE for their expression (designated “AIRE-dependent”), and the expression of the remaining 3,260 was enhanced significantly in the presence of AIRE (designated “AIRE-enhanced”); they also found that the expression of a number of genes to be repressed in mTEC in the presence of AIRE, likely in an indirect fashion (111). In keeping with these findings, Meredith *et al* found that AIRE induced the expression of 2995 genes and repressed the expression of 766 genes in mTEC (129).

In addition to its role in PGE, AIRE may also augment the process of negative selection by enhancing the antigen presentation capacity of mTEC through its influence on the expression of genes involved in the antigen processing and presentation pathways (167). Another mechanism by which AIRE might contribute to T-cell tolerance is via the induction of thymic T-Regs, particularly during the perinatal period (170). Quantitative and qualitative defects of T-

Regs have been described in AIRE deficient mice and patients with APS1 respectively (170-173). AIRE has also been shown to influence the TCR repertoire of the T-Reg compartment (174-176) and to influence the development of a perinatal T-Reg population that is able to rescue the severe autoimmune phenotype in mice with neonatal T-Reg ablation (170).

Finally, there is evidence that AIRE may play a role in mTEC development. Two models have been proposed to link AIRE's roles in mTEC development and PGE. The "developmental model" proposed that the extent of PGE is progressively restricted as mTEC mature because the action of AIRE drives their differentiation from multipotent progenitors to restricted cell lineages (177). Whereas the more widely accepted and evidenced "terminal differentiation model" states that the breadth of PGE increases as mTEC mature from an MHCII^{Lo}CD80^{Lo}AIRE⁻ state to an MHCII^{Hi}CD80^{Hi}AIRE⁺ state (57, 111, 178). Studies have also suggested that AIRE plays a role in promoting mTEC apoptosis, a process which could facilitate central tolerance induction by providing a source of tissue restricted antigens (TRAs) for cross presentation by intrathymic DCs (179, 180). In contrast to this, there is substantial evidence originating from lineage tracing models of a post-AIRE terminally differentiated mTEC stage, which downregulates expression of MHCII and CD80 and expresses markers of epithelial cell terminal differentiation such as IVL, desmoglein 3 (DSG3), SPINK5, and K10 (59-62). The development of these cells is dependent on AIRE as they are lacking in

the thymi of AIRE deficient mice, and they may still contribute to tolerance induction as they continue to express TRAs (59-61).

1.4.4 Molecular mechanisms of AIRE-induced gene expression

The precise molecular mechanisms by which the transcriptional facilitator AIRE operates in TEC to enable PGE are incompletely understood (111, 139). Several studies have found AIRE to exist in the context of a large multi-protein complex, with partners that affect nuclear transport, chromatin binding and remodelling, DNA repair, gene transcription and RNA processing (139, 181-183). AIRE is therefore thought to rely on physical and functional partnerships with a number of proteins that enable and regulate its activity (181-184).

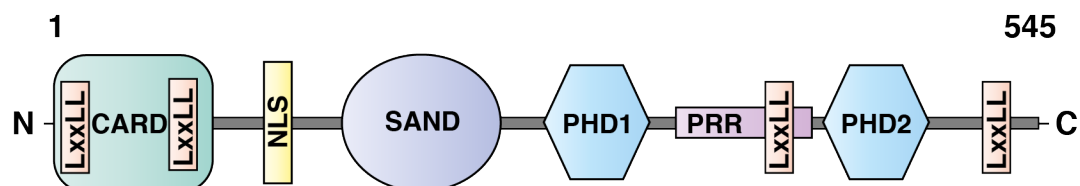


Figure 1.1 Diagram showing domain structure of the human AIRE protein

Human AIRE is 545 amino acids in length. Starting from the amino (N) terminal towards the carboxy (C) terminal, it is comprised of a caspase recruitment domain (CARD), nuclear localisation signals (NLS), a SAND (SP100, AIRE, NucP41/P75 and DEAF1) domain, and two plant homeobox domains (PHD) separated by a proline rich region (PRR). It also contains four dispersed LXXLL motifs.

AIRE is comprised of several functional domains typically found in regulators of gene transcription and chromatin binding factors (Figure 1.1). The functional importance of each of these domains is highlighted by the existence of APS1 disease causing mutations and/or experimental models disrupting their function (139, 140, 184). The caspase recruitment domain (CARD) at the

amino terminus is responsible for the homo-oligomerisation of AIRE via the formation of homotypic CARD::CARD interactions (185-189). The formation of homo-oligomers is thought to be essential to the function of AIRE, as the CARD domain is the most frequently mutated site in APS1 and because it is thought that dominant negative mutations within the SAND and PHD1 domains affect the function of AIRE homo-oligomers (150, 154-156). Along with the nuclear localization signals (NLS), the CARD domain is also responsible for the correct subcellular localisation of AIRE within nuclear speckles, which has also been shown to be essential for transcriptional activity (150, 155, 190, 191). Although associated with DNA-binding in other proteins, AIRE's SAND domain (SP100, AIRE, NUCP41/P75 and DEAF1) lacks the canonical KDWK DNA-binding motif found in other SAND family members. Nonetheless, the presence of dominant negative mutations within the SAND domain of AIRE suggests that its integrity is important for the molecule's function (154, 156). The lack of a DNA-binding motif implies that AIRE does not function as a classical transcription factor that binds a particular DNA sequence. Instead, the prevailing concept is that AIRE interacts with chromatin via a mechanism that is independent of the underlying DNA sequence (184). This capacity may be endowed, in part, by the presence of two plant homeodomain zinc fingers (PHD1 and PHD2), which typically confer an ability to read epigenetic marks. In line with this, PHD1 has been shown to bind to unmethylated H3K4 (H3K4me0) (192-194). Although the function of PHD2 is less well defined, both PHD1 and 2 are important, as mutations in either domain reduce AIRE's transcriptional ability (155, 182, 195, 196). AIRE also contains a proline rich

region (PRR) and four LXXLL motifs, both features of proteins with roles in transcriptional activation (197). The LXXLL motifs are thought to mediate homotypic interactions, so may contribute to the formation of AIRE homooligomers and/or interactions between AIRE and other proteins containing LXXLL motifs (140). Certainly, disruption of the fourth LXXLL has been shown to have a detrimental effect on the transactivation capacity of AIRE (196).

Although AIRE clearly functions as a critical transcriptional regulator or facilitator of PGE in mTEC, there are a number of features that suggest that AIRE is not a classical transcription factor (181, 183, 198). AIRE is able to regulate the expression of thousands of TRGs in mTEC that have signatures of epigenetic silencing (111, 129, 199). Furthermore, these genes are normally subject to distinct mechanisms of transcriptional regulation in their native peripheral tissues, including the use of tissue specific transcription factors and distinct transcriptional start sites (TSSs) (178). As previously mentioned, AIRE is thought to be unable to bind directly to DNA (192); indeed, a particular DNA-consensus sequence would seem unlikely in the context of the uniquely broad transcriptional programme that it is able to affect (140). Instead, AIRE is thought to recognise its target genes by recognising more general chromatin features associated with transcriptionally silenced genes, such as repressive epigenetic marks (111, 188, 192, 193, 199, 200) and stalled RNA polymerase II (Pol II) (201-203).

Evidence suggests that the AIRE complex localises to its target genes by recognising repressive epigenetic marks (111, 188, 192, 193, 199, 200). Chromatin immunoprecipitation sequencing (ChIP-seq) of primary mTEC has revealed that the TSSs of AIRE-regulated genes in mTEC are typically marked by repressive histone modifications, such as trimethylated H3K27 (H3K27me3) and H3K9me3, and devoid of permissive histone marks, such as H3K4me3 (111, 199). Although AIRE is able to bind directly to H3K4me0 via its PHD1 domain (192, 193), it is unable to bind H3K27me3 or H3K9me3 with physiological affinity (192). It is therefore possible that interaction partners could aid in localising the AIRE complex to TRGs marked by the latter marks, although this remains to be shown (111). Aside from recognition of particular histone modifications typically associated with transcriptionally silenced genes, AIRE also localises to repressed chromatin via its interaction with the ATF7ip-MBD1 complex, which recognises methylated CpG islands and associates with AIRE's SAND domain (188). Once recruited to inactive chromatin, it is not clear whether the AIRE complex is able to simply override repressive epigenetic signals and allow gene transcription, or whether a further step is required to re-programme the local epigenetic environment to a more permissive state before transcription can ensue (111).

More recently, a ChIP-seq approach has described the genomic binding sites of AIRE in primary MHCII^{Hi} mTEC (198). This revealed that AIRE binds to TSS, intronic regions, and intergenic regions; within the latter AIRE preferentially binds to and activates superenhancers, defined as long stretches of chromatin

marked by acetylated H3K27 (H3K27ac) and H3K4me1 but devoid of H3K27me3 (198). Superenhancers serve as hubs for the local accumulation of high concentrations of transcriptional regulators, which are subsequently conveyed to TSSs via chromosomal looping events in order to influence gene transcription. AIRE-bound superenhancers were also occupied by topoisomerase 1 (TOP1) and 2A (TOP2A), which have previously been identified as potential AIRE-interacting partners involved in the creation of DNA double strand breaks (181). It has further been suggested that other AIRE-interacting partners could accumulate at these sites, including DNA repair factors, CBP and BRD4. Looping events could then serve to deliver these AIRE-containing multimolecular complexes to the TSS of genes marked by repressive histone modifications and the presence of stalled Pol II (198). Not out of keeping with such a mechanism is the fact that AIRE is distributed within the nucleus within nuclear speckles. These nuclear speckles are tightly associated with the nuclear matrix, which dynamically controls chromatin organisation, and are also associated with other known AIRE-interacting partners (184, 204, 205). Furthermore, co-expressed TRGs have been shown to co-localise to the same nuclear subdomain, an observation that would be consistent with chromosomal looping (133).

Once AIRE has localised to the TSS of a gene that it regulates it is thought to induce the transcription of silenced genes by promoting the release of stalled Pol II via its interaction with pTEFb, BRD4 and HNRNPL (183, 201-203). Even in transcriptionally silenced genes, Pol II is known to initiate transcription of the

first 20-100bp before further transcription is blocked by the action of pause factors (202). For transcriptional elongation to continue this block must be released by the action of the pTEFb complex, which phosphorylates the pause factors as well as the C-terminal domain of Pol II to promote transcription (202, 203). Amongst AIRE's binding partners are the subunits that make up the pTEFb complex, CDK9 and CYCT (184, 201). pTEFb may be recruited to AIRE by the bromodomain containing protein, BRD4, which binds to post-translational modifications within the CARD domain of AIRE (203), or via the action of the mRNA processing factor HNRNPL (183).

Aside from releasing stalled Pol II, a further mechanism by which AIRE might initiate transcription of silenced genes is via the DNA damage response pathway. AIRE physically associates with a number of proteins involved in the creation of DNA double strand breaks, such as TOP1 and TOP2A, and the response to DNA damage including the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs), as well as its regulatory subunits KU70 and KU80, Poly (ADP-ribose) polymerase 1 (PARP1), the mismatch repair factor MSH2, Tripartite motif containing 28 (TRIM28), and Facilitates chromatin transcription (FACT) (140, 181, 182, 184, 198, 206). The exact mechanisms by which DNA repair factors are involved in AIRE-induced gene expression are unknown. However, transient topoisomerase-induced DNA double strand breaks and subsequent recruitment of DNA repair factors have been associated with the initiation of gene transcription in various cell types (162, 184, 207). It is therefore thought that the interaction of AIRE, TOP1, TOP2A, DNA-PK, and

perhaps other proteins involved in DNA damage response, at loci encoding AIRE-induced TRGs may play a role in promoting PGE. A role for DNA-PKcs in AIRE-induced gene expression is further supported by studies showing that mutations in the gene encoding DNA-PKcs in mouse models and human disease may cause defective expression of AIRE-regulated genes (181, 208). However, the mechanism by which DNA-PKcs participates in PGE might not be limited to its role in the DNA damage response pathway, but could additionally involve regulation of AIRE's activity via the induction of particular post-translational changes on the AIRE protein, namely by phosphorylation of particular serine and threonine residues (206).

Alongside the aforementioned transcriptional and post-transcriptional mechanisms described to regulate AIRE, its activity might further be regulated at the post-translational level by the opposing actions of sirtuin-1 (SIRT1) and CBP, which respectively deacetylate and acetylate certain lysine residues between the NLS and SAND domain of AIRE (209). The impairment of AIRE-regulated PGE and the development of autoimmune pathologies in mice with SIRT1 deficiency, as well as the association of a human *SIRT1* mutation and familial type 1 diabetes, add further support to the role for SIRT1-induced lysine deacetylation in activating the function of AIRE in vivo (209, 210).

Finally, despite the considerable advancement by studies conducted over the past 20 years, several unanswered questions remain regarding our understanding of the molecular mechanisms by which AIRE functions. With

regards to localisation to target genes, it is unclear how the AIRE protein complex is able to localise to H3K27me3 and H3K9me3, two histone modifications that have been shown to mark the TSSs of AIRE-regulated genes in mTEC (111, 192, 199). Having identified its target genes, it also remains to be elucidated whether AIRE then must remodel the local chromatin environment in order to activate expression of epigenetically silenced genes. In addition, there are several putative AIRE interaction partners that have not been studied with regards their contribution to PGE, so perhaps the study of these molecules could provide further mechanistic insight into how AIRE-induced PGE is effected and thus help in answering some of questions that remain (181-183).

1.4.5 Transcriptional control of PGE

Whilst it is accepted that at the population level TEC express almost all known protein coding genes (111, 127), precisely how this is regulated at the single cell level remains to be conclusively established. Understanding how PGE is regulated within single TEC is an important first step in understanding how these cells are able to retain their cellular identity whilst effecting such a uniquely broad transcriptional programme, essential for the generation of a TCR repertoire suitably selected not to elicit autoimmune disease.

Theoretically three possibilities exist as to how TEC as a population could achieve expression of virtually the whole transcriptome. The first possibility is that individual TEC could express all genes. The second and third possibilities

are that single TEC could each express only a subset of genes whilst still allowing TEC as a population to express the full transcriptomic repertoire, and that the genes expressed could either be chosen at random or with a particular order or pattern. If the latter scenario were to be correct, it would be interesting to investigate whether genomic location, common epigenetic signatures and/or transcriptional co-regulation by tissue- or pathway-specific transcription factors or specific binding partners of AIRE could provide a molecular mechanism for transcriptional co-expression.

Evidence against the first possibility (i.e. expression of all genes by a single TEC) comes from several studies that have shown that at single cell resolution, TRG expression is heterogeneous and mosaic with individual mature mTEC expressing only a relatively small fraction (approximately 1-3%) of TRGs at any given point during their life-span (111, 127, 129, 178, 211). A study looking at the transcriptomes of 174 single mature MHCII^{Hi} mTEC found that these cells expressed just over 5000 genes per single cell. These included an average of 3 AIRE-dependent TRGs (range 0-27), 144 AIRE-enhanced TRGs (range 46-322) and 581 AIRE-independent TRGs (range 241-969) (111). Next, examining the frequencies at which TRGs are expressed within single mTEC, they found that individual AIRE-dependent TRGs were detected in about 1%, AIRE-enhanced TRGs in about 2% and AIRE-independent in about 9% of mature mTEC (111). A theoretical advantage of such a mosaic TRG expression pattern within single mTEC is that it could serve to allow a sufficiently high density of particular self-antigen::MHC complexes on the

mTEC cell surface in order to enable strong, tolerogenic signals for negative selection of T-cells bearing cognate TCRs (178).

It still remains to be conclusively established whether TRG expression in single mTEC is stochastic or, alternatively, whether co-expression patterns exist (111, 127, 129, 133, 178, 211, 212). Efforts to answer this question have been limited by technology, approach and cell number. As such, single cell studies conducted to date have been rather divided in respect to whether PGE in mTEC is stochastic or coordinated. Three studies have argued for a stochastic mechanism (111, 178, 211) and an equal number have found evidence for co-expression patterns (127, 129, 133).

Derbinski *et al* used a single cell PCR approach to examine the expression of genes located within the casein gene locus. In contrast to mammary epithelial cells, in which the genes of the casein locus are expressed in a highly coordinated fashion, an equivalent pattern was not observed in single mTEC, with their expression of genes within the locus appearing to be entirely stochastic. However, restricting their analysis to genes within a particular locus precluded analysis of co-expression patterns between genes outside of this cluster, either on the same or even on different chromosomes. In support of the Terminal Differentiation model of PGE, Derbinski *et al* additionally found that CD80^{Hi} mTEC expressed more TRGs compared to CD80^{Lo} mTEC and that single mTEC could express genes from the casein locus alongside *Insulin 2* (*Ins2*) i.e. TRGs typical of distinct peripheral cell lineages (211). Villaseñor *et al*

also used a single cell PCR approach limited to the analysis of genes within the *S100* cluster, *Csnβ-γ*, and *Ins2*. Again arguing in favour of the Terminal Differentiation model of PGE, the authors also concluded that PGE was stochastic and not constrained by expression patterns seen in peripheral differentiated cell lineages. Specifically, they did not detect co-expression of genes within tissue-specific gene clusters and found AIRE-dependent TRGs to be mono- and bi-allelically expressed in a stochastic pattern in mTEC, with no preservation of imprinting. Furthermore, for *Ins2* expression, they showed that in comparison to pancreatic β cells, mTEC used a distinct TSS and that expression of *Ins2* in mTEC occurred independently of the transcription factors that drive its expression in the pancreas (178). Using a broader approach Sansom *et al* analysed the transcriptomes of 174 mature mTEC and also concluded that AIRE-regulated genes within single mTEC were expressed in a stochastic manner, although the authors acknowledged that their study may have been underpowered to detect co-expression patterns (111).

One of the three published studies that have found evidence for gene co-expression patterns, used mTEC preselected for the expression of cell surface-expressed TRAs by FACS sorting (133). This is likely to have added statistical power to their analysis for co-expression. Pinto *et al* used a combined approach of bulk transcriptome profiling using whole genome microarrays and single cell PCR of three human mTEC subsets defined by cell surface expression of the TRAs mucin 1 (MUC1), carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5), and sodium/glucose

cotransporter 1 (SGLT1) respectively. They found that each subset comprised a co-expression group expressing distinct but overlapping gene sets, which were enriched for TRGs. They also found that individual mTEC could cycle between distinct gene co-expression pools, a feature that could allow a diverse repertoire of self-antigens to be displayed to developing thymocytes confined within a relatively small island of the thymic medulla. Although the co-expressed gene sets observed could not be explained by functional characteristics or expression by certain peripheral tissues, they clustered within the genome and indeed also in nuclear topological domains suggesting that PGE may be regulated at the level of chromosomal structure (133). Brennecke *et al* adopted an unbiased approach in which the transcriptomes of 203 single mature MHCII^{Hi} mTEC were analysed. They found 11 gene co-expression groups, although felt this to be an underestimation due to the limited number of single cells analysed. Co-expressed genes clustered within the genome and had a more open chromatin conformation as assessed by ATAC-sequencing suggesting that local chromatin remodelling might underlie patterns of gene co-expression within mTEC. They additionally found that almost the entire protein coding genome could be detected by analysing only a few hundred mTEC and presented evidence that individual mTEC could transition between co-expression groups, again suggesting that this could serve to limit the area of medulla that a developing thymocyte is required to survey during the process of negative selection (127). Finally, Meredith *et al* used a single-cell RNA sequencing approach to identify 19 small co-expression clusters of AIRE-induced genes. However, the co-expression

groups identified differed between individual mice and genes within co-expression clusters were not related in terms of genomic location, biological function, and regulation by common transcription factors. The lack of a plausible biological mechanism that could explain membership to a particular co-expression group led them to hypothesise that co-expression patterns were the result of “epigenetic bookmarking”, in which a stochastic mechanism initially underlies expression of a set of genes within a single mTEC. These genes are then epigenetically marked for transcription and this epigenetic signal is preserved over cell divisions, thus allowing co-expression groups to be detected within clonally related mTEC (129).

Given the heterogeneity of gene expression within mTEC at the single cell level and the very low frequency at which individual TRGs are transcribed (111, 127, 129, 178, 211), it is likely that a single-cell approach examining the transcriptomes of thousands of individual mTEC would be required to provide the resolution required to faithfully describe the full complexity of the transcriptional patterns of PGE within mTEC.

1.5 Chromodomain-Helicase-DNA-binding protein 4

1.5.1 CHD4 structure and functions

Chromodomain-helicase-DNA-binding (CHD) proteins are members of the SNF2 ATP-dependent family of chromatin remodelling enzymes, which utilise energy derived from adenosine triphosphate (ATP) hydrolysis to alter DNA-nucleosome interactions and hence chromatin structure (213). The mammalian CHD family consists of nine members and is structurally defined by the presence of tandem chromodomains and an SNF2 helicase/ATPase domain. The family is further subdivided into three subfamilies based on additional structural features (213, 214).

CHD4, also known as Mi-2 β , is a ubiquitously expressed protein belonging to the second subfamily of CHD proteins, which is characterised by the additional presence of N-terminally located paired PHD zinc fingers (213, 215). It was initially identified as a dermatomyositis autoantigen (216), and is now known to have roles in various cellular processes including transcriptional regulation, resulting in both repression and activation of gene transcription in a context dependent manner (213, 217-227), as well as DNA repair (228-231) and cell cycle control (229, 230, 232).

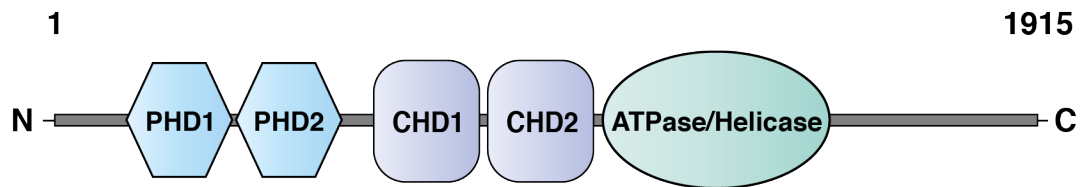


Figure 1.2 Diagram showing domain structure of the mouse CHD4 protein

Mouse CHD4 is 1,915 amino acids in length. Starting from the amino (N) terminal towards the carboxy (C) terminal, it is comprised of tandem plant homeobox domains (PHD), chromodomains (CHD), followed by an enzymatically active SNF2 ATPase/Helicase domain.

CHD4's molecular structure contains, as already alluded to, several known functional domains (Figure 1.2), including two N-terminally located PHD domains, followed by two chromodomains and an SNF2 helicase/ATPase domain (213). The tandem PHD domains mediate bivalent nucleosome binding via the recognition of H3K4me0, H3K9me3 and acetylated H3K9 (H3K9Ac), targeting either two H3 tails within the same nucleosome or on adjacent nucleosomes (233-236). The interaction with H3K4me0 is important for the function of CHD4 in transcriptional regulation and the interaction with H3K9me3 is important for targeting to heterochromatin (236). In contrast to other chromodomains, the chromodomains of CHD4 do not bind to methylated H3 tails, but instead mediate direct binding to DNA (234, 237). Finally, the catalytically active SNF2 helicase/ATPase domain provides the energy required for chromatin remodelling via nucleosome mobilisation (236). Structure-function studies have shown that the separate domains of CHD4 are highly interdependent for optimal protein function. The affinity of nucleosome binding by CHD4's PHD domains is increased by the presence of the tandem chromodomains, and this interaction in turn stimulates the catalytic activity of the ATPase domain (233, 238).

The role for CHD4 in transcriptional regulation was first described in the context of the nuclear remodelling and deacetylase (NURD) complex, a multiprotein complex with a variable subunit composition (220-222, 239). The core NURD subunits include the ATP-dependent chromatin remodelling enzymes CHD3, -4 or -5; histone deacetylases 1 and 2 (HDAC1 and 2); methyl-CpG-binding proteins MBD-1, -2, or -3; metastasis associated proteins (MTA) -1, -2 or -3; the retinoblastoma binding proteins RBBP4 or -7; and GATA2a or -2b, which may function in core histone binding and/or as structural or regulatory subunits of the NURD complex (217, 240). Although, due to its histone deacetylase activity, the NURD complex was originally thought to function solely as a transcriptional repressor, it is now appreciated that the complex contributes to a more subtle “fine-tuning” of gene expression programmes and even promotes gene activation depending on its context specific subunit composition and interactions with other proteins or protein complexes (224, 226, 241-243). CHD4 is also known to function independently of the NURD complex to influence transcriptional activation instead, forming partnerships with histone acetyltransferases (HATs), such as p300, MOZ and TAF1 (225, 244), and methyltransferases, such as ALL-1 (242). In the context of transcriptional repression, CHD4 also interacts with the PRC2 complex (227, 245). Deacetylation of H3K27 by the NURD complex was initially found to mediate recruitment of PRC2, resulting in trimethylation of H3K27 and subsequent repression of gene expression (227). Later, CHD4 was found to interact with the EZH2 component of the PRC2 complex independently of the NURD complex, localising it to the promoters of target genes for H3K27me3

deposition and transcriptional silencing (245). It was proposed that this process could involve binding of CHD4 to target genes marked by H3K9me3, with subsequent chromatin remodelling to allow access to PRC2 (245).

CHD4 has also been implicated in the maintenance of genome integrity, the DNA damage response (DDR) and control of cell cycle progression, although it is unclear how many of these processes require it to function in the context of the NURD complex (230). A role for CHD4 in the DDR and maintenance of genome integrity was suggested by studies showing that CHD4 expression is induced by UV radiation (246). Moreover, in response to DNA damage CHD4 is phosphorylated by ATR and ATM, two DDR-associated protein kinases (247, 248), and knock-down of CHD4 causes an accumulation of DNA damage and chromatin defects associated with cellular senescence (249). CHD4 has subsequently been found to be recruited to sites of DNA damage by binding to PARP1 (229, 250) or RING finger protein 8 (RNF8) (251) and is thought to remodel chromatin surrounding these sites, leading to chromatin decondensation and a permissive environment for DNA repair by allowing access to further components of the DDR machinery (215, 230, 250, 251). The contribution of CHD4 to the regulation of cell cycle progression is complex. It has been identified as a possible tumour suppressor in human cancers (230, 252, 253), although this may be linked to its role in the DDR as CHD4 is required for the G₂/M checkpoint such that cells depleted of CHD4 continue to cycle despite DNA damage induced by ionising radiation (230, 231). On the other hand, in the context of NURD complex-mediated deacetylation of p53,

which leads to a reduction in p53 activity, CHD4 plays a role in promoting progression through the G₁/S transition (229, 230).

Through the molecular mechanisms described above, CHD4 has been found to play a part in diverse but interrelated biological processes including, but not restricted to, stem cell maintenance, cell differentiation and tissue/organ development, ageing and oncogenesis (217, 223-225, 230, 240, 241, 244, 245, 249, 252-257).

CHD4 already has several described roles related to T-cell development and function (258). These range from promoting haematopoietic stem cell self-renewal and differentiation to the lymphoid lineage (256), to coordinating normal intrathymic development of T-cells (225, 244), and regulating peripheral T-cell proliferative capacity and T-helper subset polarisation (225, 259). In relation to intrathymic T-cell development, studies using mice in which *Chd4* was conditionally deleted under the control of a T-cell specific driver of Cre recombinase (Lck-Cre) have shown that CHD4 is required at different stages of thymocyte development, including the DN to DP transition and for the expression of the CD4 co-receptor in DP and CD4SP thymocytes (225, 244). A mechanism by which CHD4 controls expression of *Cd4* gene was initially proposed whereby CHD4 binds to the *Cd4* proximal enhancer and recruits the E box binding protein, HEB, and the HAT, p300, leading to hyperacetylation of H3 at the proximal enhancer and transcriptional activation (225). It was later

shown that CHD4 also acts at the *Cd4* silencer, where it antagonises the action of Ikaros to promote transcription of *Cd4* (244).

1.5.2 CHD4 as a putative AIRE interaction partner

CHD4 was identified as a putative AIRE interaction partner in an in vitro study that employed quantitative mass spectrometry to identify proteins physically interacting with chromatin bound AIRE in HEK293T cells transfected with human AIRE (182). Interestingly, the strength of interaction between AIRE and CHD4 was reduced by the presence of loss-of-function mutations affecting AIRE's PHD1, suggesting that this domain, known to mediate binding of AIRE to H3K4me0 (192, 193) and DNA-PKcs (206), may also facilitate an interaction with CHD4 (182).

The epigenetic landscape surrounding AIRE-regulated genes in mTEC is characterised by features typical of transcriptional silencing, with deposition of H3K27me3 and H3K9me3 at their TSSs, alongside depletion of H3K4me3 (111, 199). Hence, it is thought that the mechanism that allows AIRE to target TRGs for transcriptional activation involves localisation to these repressive histone post-translational modifications (111). Whilst AIRE is able to bind unmodified H3K4 via its PHD1 domain, it is unable to bind to the other histone modifications typically associated with the genes it regulates with sufficient affinity (192, 193); nor is there any convincing evidence that AIRE can directly bind to DNA (184). It has therefore been suggested that an interaction partner (or partners) of AIRE would be required for targeting to these chromatin

features (111). CHD4 is able to bind to both H3K4me0 and H3K9me3 via its PHD domains, and directly to DNA via its chromodomains (233-237). Although binding to H3K27me3 has not been shown for CHD4, it has been shown to interact with the PRC2 complex to induce the deposition of H3K27me3 at PRC2 target genes (227, 245). It has also been suggested that an interaction between AIRE and a chromatin remodeling factor might be necessary to allow opening of silenced TRG loci (111, 184). Intriguingly, as an ATP-dependent chromatin remodeling enzyme that is able to influence nucleosome mobilisation, and via its roles in DNA repair and partnerships with histone modifying enzymes, CHD4 is able to alter chromatin accessibility and promote gene transcription (213, 215, 230, 260).

In summary, a possible physical interaction with AIRE, alongside its roles in nucleosome binding and mobilization, DNA repair and transcriptional regulation, make CHD4 an attractive candidate for involvement in AIRE-induced PGE.

1.6 Aims and hypotheses

The research presented in this thesis interrogates the transcriptional and epigenetic mechanisms that regulate PGE in TEC, and in particular probes the functional role of CHD4 in this process.

Specific Aims:

1. To determine whether PGE is a random process or whether there are specific co-expression patterns associated with PGE in single mTEC.

Hypothesis: PGE in single mTEC is not stochastic, but instead defined by the presence of multiple co-expression groups. Interrogating the patterns of co-expression may offer further insight into the molecular mechanisms that enable such a uniquely broad transcriptional programme in mTEC.

2. To determine the role of CHD4 in TEC development and function, with a particular focus on AIRE-regulated PGE.

Hypothesis: CHD4 cooperates with AIRE to influence the transcription of AIRE-induced TRGs. This may occur via targeting of the AIRE protein complex to TRG loci marked by repressive histone marks and locally remodelling chromatin.

The outlined work contributes to our understanding of the molecular mechanisms that enable PGE in TEC and hence drive thymic TCR repertoire selection and central T-cell tolerance induction by providing a comprehensive mirror of immunological self. The knowledge gained may not only identify novel disease mechanisms leading to autoimmunity but could also identify targets for pharmacological intervention to enhance PGE in TEC.

2 Materials and Methods

2.1 Mice

Mice were maintained under specific pathogen-free conditions and according to institutional and United Kingdom Home Office regulations.

2.1.1 C57BL/6

Female C57BL/6 mice were obtained from Charles River and rested for at least one week before analysis at 4-5 weeks of age.

2.1.2 BALB/c

Female BALB/c mice were obtained from Charles River and rested for at least one week before analysis at 4-5 weeks of age.

2.1.3 C57BL/6 x BALB/c F1 Hybrids

Female C57BL/6 x BALB/c F1 hybrid mice were obtained from Charles River and rested for at least one week before analysis at 4-5 weeks of age.

2.1.4 Chd4^{fl/fl}

Chd4^{fl/fl} mice were a kind gift from Katia Georgopoulos. In these mice, exons 12-21 of *Chd4*, which encode the ATPase domain, are flanked by LoxP sites such that Cre-mediated recombination leads to deletion of exons 12-21 (225). It has previously been shown that a truncated protein is not expressed when this model is used to conditionally delete *Chd4* from developing thymocytes under the control of LckCre (225). Chd4^{fl/fl} mice were bred with a Foxn1Cre

transgenic strain in order to obtain mice with conditional deletion of *Chd4* in cells expressing *Foxn1*, namely the epithelial cells of the thymus and skin (261). In all experiments mice were compared to age and sex matched littermate controls. Mice with heterozygous TEC conditional *Chd4* deficiency ($Chd4^{fl/+}::Foxn1Cre$) were compared to $Chd4^{fl/+}$ controls. Whereas, mice with homozygous TEC conditional *Chd4* deficiency ($Chd4^{fl/fl}::Foxn1Cre$; designated “cKO”) were compared to $Chd4^{fl/fl}$ controls (designated “WT”).

2.1.5 Foxn1Cre

Mice expressing Cre recombinase under the control of a *Foxn1* transgene (261) were obtained from an in-house colony and crossed with $Chd4^{fl/fl}$ mice in order to generate mice with both heterozygous and homozygous *Chd4* deficiency targeted to *Foxn1* expressing cells.

2.1.6 Rag1^{-/-}

Female Rag1^{-/-} conventional knockout mice were a kind gift from Mariolina Salio. Organs from these mice were used to generate fresh frozen tissue sections used for experiments aiming to detect the presence of serum autoantibodies by indirect immunofluorescence.

2.1.7 Timed matings

For experiments using embryonic thymi at specific gestational ages, timed matings were performed. Stud males were paired with females of at least 8 weeks of age overnight and the following morning the female mice were

checked for the presence of vaginal plugs. The presence of a vaginal plug was denoted as gestational or embryonic day 0.5 (E0.5).

2.1.8 Induction of lymphopaenia

Mice on a C57BL/6 background are known to be resistant to the development of spontaneous autoimmune disease (144). Therefore, to assess whether CHD4 cKO mice had impaired self-tolerance, T-cell lymphopaenia was induced in 2 week old female CHD4 cKO and WT controls with the aim of provoking lymphopaenia-induced lymphoproliferation and the development of autoimmunity (262). In order to achieve in vivo T-cell depletion, an established protocol was used in which 2 week old mice were intraperitoneally injected with 200 μ g anti-CD4 (clone GK1.5), 100 μ g anti-CD8, (clone 53-67) and 50 μ g anti-Thy1.2 (clone T24) on three occasions at an interval of three days apart (263). All monoclonal antibodies used were generated in-house from hybridoma lines. Successful induction of lymphopaenia was assessed by quantifying the peripheral T-cell compartment 7 days post treatment. Mice rendered lymphopaenic were then aged for 25-30 weeks to allow time for autoimmune pathologies to develop before being sacrificed for analysis, at which point organs and serum were harvested in order to look for evidence of impaired self-tolerance.

2.2 Genotyping

2.2.1 DNA extraction from ear punches

For genotyping of postnatal mice, ear punches were taken at 2 weeks of age and incubated in a water bath at 55°C for 5-16 hours in 600 μ l lysis buffer containing 100mM Tris-HCl pH8 (Sigma), 5mM EDTA (Sigma), 0.2% SDS (Sigma), 200mM NaCl (VWR), and 100 μ g/ml Proteinase K (Roche). Any remaining debris was removed by pelleting the samples in a table-top centrifuge set to maximum speed for 15 minutes at room temperature and transferring the supernatant ($\approx$ 600 μ l) to a new tube. Genomic DNA was precipitated with an equal volume of isopropanol (VWR) and pelleted at maximum speed for 10 minutes at room temperature. The DNA pellet was then washed with 70% ethanol (VWR) and centrifuged at maximum speed for a further 10 minutes before being air-dried and resuspended in 50 μ l TE buffer (10mM Tris-HCl pH8, 1mM EDTA). DNA samples were kept at 4°C for up to 4 weeks or at -20°C for longer-term storage.

2.2.2 DNA extraction from tail clips

For genotyping of embryonic mice, tail clips were taken. Lysis and DNA extraction was performed using the Phire Tissue Direct polymerase chain reaction (PCR) kit (Thermo Scientific). Briefly, tail clip samples were incubated for 5 minutes at room temperature in 20 μ l Dilution Buffer supplemented with 0.5 μ l DNA Release Additive, followed by 2 minutes at 98°C. Samples were then diluted in 180 μ l of nuclease free water (Ambion) before being used directly for PCR.

2.2.3 Polymerase Chain Reaction

Reaction mixes were set up according to Table 2.1 and PCR reactions were then performed in a 96-well thermal cycler (Applied Biosystems) using the cycling conditions in Table 2.2. A non-template control, in which the DNA was replaced with nuclease-free water (Ambion), and negative and positive DNA controls were included for each primer pair analysed. The primer pairs (Sigma) used for genotyping mice used to generate CHD4 cKOs, along with their annealing temperatures (T_a) and expected product sizes are listed in Table 2.3.

Table 2.1 Genomic DNA PCR reaction mix for a single reaction tube

Reagent	Stock Concentration	Volume for 1x reaction	Final Concentration
Nuclease-free water (Ambion)	-	18.4 μ l	-
10x PCR Buffer (Sigma)	10x	2.5 μ l	1x
dNTPs (Sigma)	10mM	2 μ l	0.8mM
Forward Primer (Sigma)	10 μ M	0.5 μ l	0.2 μ M
Reverse Primer (Sigma)	10 μ M	0.5 μ l	0.2 μ M
Taq Polymerase (Sigma)	5U/ μ l	0.1 μ l	0.02U/ μ l
DNA	-	1 μ l	-
Total Volume		25μl	

Table 2.2 Cycling conditions for genotyping PCR

Step	Cycles	Temperature	Time
Initial Denaturation	1	94°C	4 minutes
Amplification	34	Denaturation: 94°C	30 seconds
		Annealing (T_a): Variable	40 seconds
		Elongation: 72°C	1 minute
Final Elongation	1	72°C	10 minutes
Hold	1	4°C	∞

Table 2.3 Primer pairs used for genotyping mice used to generate CHD4 cKO

Primer Pair	Name	Direction	Sequence		Ta	Expected Product
			5'	3'		
Chd4flox	Mi2b+F	Forward	CTCCAAGAAGAAGACGGCAGATCT		56°C	409bp (WT allele)
	Mi2bINR	Reverse	GTCCTTCCAAGAGAGAGCAAG			≈550bp (floxed allele)
Delta allele	Mi2b+F	Forward	CTCCAAGAAGAAGACGGCAGATCT		56°C	≈250bp (cKO allele)
	KG4R	Reverse	CTTCCACAGTGACGTCCAGACGCA			
iCre	iCre3961	Forward	TCTGATGAAGTCAGGAAGAACC		58°C	≈500bp (Foxn1Cre transgene)
	iCre3962	Reverse	GAGATGTCCTTCACTCTGATTC			
oIMR	oIMR8744	Forward	CAAATGTTGCTTGTCTGGTG		58°C	206bp (all samples)
	oIMR8745	Reverse	GTCAGTCGAGTGCACAGTTT			

2.2.4 Agarose gel electrophoresis

Dependent on the size of PCR product to be visualised, 1.0-1.5% agarose gels were prepared by dissolving UltraPure Agarose (Invitrogen) in 1x Tris-Borate-EDTA (TBE) buffer (Sigma) by heating in a microwave. GelRed nucleic acid dye (Biotium) was then added to achieve a final a dilution of 1:20,000 before casting in a gel tank. 10µl of each PCR product was mixed with 2µl of 6x DNA Gel Loading Dye made from 30% glycerol (Sigma), 0.25% bromophenol blue (Sigma), 0.25% xylene cyanol (Sigma). PCR products were then loaded alongside 1Kb Plus (Invitrogen) and/or 1Kb (NEB) DNA ladder. Gel electrophoresis was carried out in a tank containing 1x TBE and gels were imaged using the Gel Doc XR+ Gel Documentation System (Bio-Rad).

2.3 Cellular techniques

2.3.1 Isolation of TEC from postnatal thymi

TEC were isolated from thymic lobes via enzymatic digestion. Thymi were collected into ice-cold plain RPMI-1640 (Gibco). Adherent fat, blood and connective tissue were removed under a stereoscopic microscope (Leica). The thymic lobes were then separated and punctured several times with sharp forceps before being placed into FACS tubes and subjected to three sequential rounds of enzymatic digestion with Liberase Digestion Solution (LDS). LDS contained: 75 μ g/ml Liberase TM (Roche) and 300 μ g/ml DNaseI (PanReac AppliChem) in 1x PBS (Sigma). Up to 2 lobes were digested per tube. For the 1st round of digestion 1.5mls of pre-warmed LDS was added to each tube; after 5 minutes of incubation at 37°C in a water bath, the samples were pipetted gently in order to aid with tissue dissociation. After a further 5 minutes incubation at 37°C, the supernatant was collected into 40mls of ice-cold FACS buffer, containing 2% fetal calf serum (FCS, Sigma) in 1x PBS. Two further rounds of digestion were performed with the addition of 1ml LDS (2nd round) and then 0.8ml LDS (3rd round). The resultant cell suspensions were centrifuged at 1,500rpm for 5 minutes at 4°C, after which the cells were resuspended in 1-3mls of FACS buffer, filtered through a 70 μ m cell strainer (Greiner Bio-One) and counted. Cells from these unenriched/whole thymic digests were either prepared directly for analysis by flow cytometry (section 2.3.5) in order to allow the calculation of absolute cell counts, alternatively TEC were enriched via depletion of CD45⁺ cells as detailed below.

2.3.2 Enrichment of TEC from postnatal thymi via depletion of CD45⁺ cells

Where enrichment for TEC was required, whole thymic digest samples were resuspended in FACS buffer to achieve a cell concentration of 100 million cells/ml. Samples were labelled with 50 μ l of anti-Mouse CD45 microbeads (Miltenyi Biotec) per 100 million cells. Following incubation for 15 minutes at 4°C, cells were washed with 40mls of FACS buffer and once again resuspended at a concentration of 100 million cells/ml. DNaseI (PanReac AppliChem) was added to each sample to achieve a final concentration of 250 μ g/ml before filtering into FACS tubes using a 30 μ m cell strainer (CellTrics). Thymic stromal cells were enriched by negative selection using the “depletes” programme on the AutoMACSPro (Miltenyi Biotec) before being prepared for analysis via flow cytometry (Subsection 2.3.5).

2.3.3 Isolation of TEC from embryonic thymi

Timed matings were performed (Subsection 2.1.7) and pregnant dams were sacrificed once the embryos had reached the required gestational age. The uterine horns were removed and transferred to a sterile petri dish (Corning). Then the embryos were removed from their amniotic sacs, killed by Schedule 1 methods, and transported in ice-cold 1x PBS containing 10% heat-inactivated FCS (Hyclone). Tail clips were taken for genotyping, then individual thymic lobes were removed under a stereoscopic microscope (Leica) using fine forceps (Sigma) and cleaned of any excess connective tissue or adherent blood. Enzymatic digestion was then carried out in 500 μ l of LDS in 1.5 ml eppendorf tubes. Samples were incubated at 37°C 650rpm in a Thermomixer

(Eppendorf) for 20-40 minutes with gentle pipetting every 15 minutes to aid with dissociation. Once the samples were completely digested they were washed with ice-cold FACS buffer, filtered through a 30 μ m cell strainer (CellTrics), counted and prepared for flow cytometric analysis (Subsection 2.3.5).

2.3.4 Isolation of thymocytes and peripheral T-cells

Lymphocytes were released from thymic lobes, lymph nodes, or spleens via mechanical disruption between frosted glass slides (Thermo Scientific). Splenic samples were then incubated for 5 minutes at room temperature with red blood cell lysis buffer containing 155mM NH₄Cl (Sigma), 15mM NaHCO₃ (Sigma), 0.1mM EDTA (Sigma) in distilled water. Cells were then washed with ice-cold FACS buffer and centrifuged at 1500rpm for 5 minutes at 4°C. Cell pellets were resuspended in FACS buffer, filtered through a 70 μ m cell strainer (Greiner Bio-One) and counted before being prepared for analysis via flow cytometry (section 2.3.5).

2.3.5 Flow cytometry and fluorescence activated cell sorting

TEC, thymocytes and peripheral lymphocytes were stained using combinations of the antibodies listed in Table 2.4. Before staining, cells were resuspended at a concentration of 5-10 $\times 10^6$ /100 μ l. In general, staining for cell surface antigens was performed for 20 minutes at 4°C in the dark. In certain TEC staining panels, binding of the *Ulex europaeus* agglutinin I (UEA-1) lectin was used to identify mTEC during the cell surface staining step. This was achieved

either by incubating with UEA-1 lectin (Vector Laboratories), which had been directly labelled in-house with Cy5 (GE Healthcare), or alternatively via a three step staining protocol with purified UEA-1 lectin (Vector Laboratories, 1:500), followed by anti-UEA-1 biotin (Vector Laboratories, 1:500) and then streptavidin BV605 (BioLegend, 1:500). Where surface staining for CCR7 was required, this was performed for 30 minutes at 37°C in a water bath, and directly followed by the addition of other cell surface stains as indicated above. For the identification of intracellular antigens, the Foxp3 Transcription Factor Staining Buffer Set (eBioscience) was used according to manufacturer's instructions; intracellular stains were performed for 30-60 minutes at 4°C in the dark. Cell viability was measured using either 150ng/ml 4',6-diamidino-2-phenylindol (DAPI; Sigma) or the LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Thermo Scientific) as per the manufacturer's instructions. For assessment of cell cycle stage, DAPI staining was used to measure DNA content after an intracellular staining step for Ki67; this was achieved by incubating samples in permeabilisation buffer (eBioscience) containing 1µg/ml of DAPI for 10 minutes at room temperature. After staining, cells were acquired and sorted using a FACS Aria III (BD Biosciences) and analysed using FlowJo v10.

Table 2.4 Antibodies and streptavidin stains used for flow cytometry

Target	Clone	Species and Isotype	Conjugate	Company	Final Staining Concentration
AIRE	5H12	Rat IgG2c κ	AF488	eBioscience	1:1000
AIRE	5H12	Rat IgG2c κ	AF647	eBioscience	1:1000
Anti-mouse IgG1	Polyclonal	Goat IgG	APC-Cy7	Abcam	1:500
Anti-Rabbit IgG (H+L)	Polyclonal	Goat IgG	AF488	Jackson ImmunoResearch	1:500
Anti-Rabbit IgG (H+L)	Polyclonal	Goat F(ab') ₂ IgG	PE-Texas Red	Abcam	1:500
B220	RA3-6B2	Rat IgG2a κ	PE	BioLegend	1:500
B220	RA3-6B2	Rat IgG2a κ	BV605	BD Biosciences	1:500
CCR7	4B12	Rat IgG2a κ	BV421	BioLegend	1:200
CCR7	4B12	Rat IgG2a κ	PE	eBioscience	1:200
CD103	2E7	Armenian Hamster IgG	AF647	BioLegend	1:500
CD11b	M1/70	Rat IgG2b κ	Biotin	BioLegend	1:1000
CD11c	N418	Armenian Hamster IgG	Biotin	BioLegend	1:1000
CD19	1D3	Rat IgG2a κ	Biotin	eBioscience	1:1000
CD205	205yekta	Rat IgG2a κ	PE-eFluor610	eBioscience	1:250
CD24	M1/69	Rat IgG2b κ	FITC	BioLegend	1:1000
CD24	M1/69	Rat IgG2b κ	PerCP- eFluor710	eBioscience	1:1000
CD24	M1/69	Rat IgG2b κ	APC	BioLegend	1:500
CD25	PC61.5	Rat IgG1 λ	eFluor450	eBioscience	1:500
CD25	PC61.5	Rat IgG1 λ	Biotin	BioLegend	1:500
CD25	PC61.5	Rat IgG1 λ	BV605	BioLegend	1:500
CD25	PC61.5	Rat IgG1 λ	PerCPCy5.5	BioLegend	1:500
CD3ε	145-2C11	Armenian Hamster IgG	PerCPCy5.5	eBioscience	1:500
CD4	RM4-5	Rat IgG2a κ	APC-Cy7	BioLegend	1:500
CD4	RM4-5	Rat IgG2a κ	FITC	BioLegend	1:500
CD4	RM4-5	Rat IgG2a κ	PE-eFluor610	Invitrogen	1:500
CD44	IM7	Rat IgG2b κ	PE-eFluor610	eBioscience	1:1000

CD44	IM7	Rat IgG2b κ	PECy7	BioLegend	1:1000
CD45	30-F11	Rat IgG2b κ	AF700	BioLegend	1:1000
CD49b	DX5	Rat IgM κ	Biotin	BioLegend	1:1000
CD5	53-7.3	Rat IgG2a κ	PECy7	BioLegend	1:1000
CD5	53-7.3	Rat IgG2a κ	PerCPCy5.5	BioLegend	1:1000
CD62L	MEL-14	Rat IgG2a κ	PECy7	BioLegend	1:1000
CD69	H1.2F3	Armenian Hamster IgG	PECy5	BioLegend	1:1000
CD80	16-10A1	Armenian Hamster IgG	PECy5	BioLegend	1:1000
CD86	GL-1	Rat IgG2a κ	PECy7	BioLegend	1:1000
CD8a	53-6.7	Rat IgG2a κ	AF700	BioLegend	1:1000
CHD4	Polyclonal	Rabbit IgG	Unconjugated	Abcam ab72418	1:1000
DSG3	AK18	Mouse IgG1	Unconjugated	MBL	1:100
EpCAM	G8.8	Rat IgG2a κ	PerCPCy5.5	BioLegend	1:1000
F4/80	BM8	Rat IgG2a κ	Biotin	BioLegend	1:1000
FOXP3	FJK-16s	Rat IgG2a κ	PE-eFluor610	eBioscience	1:1000
GP2	2F11-C3	Rat IgG2a κ	AF488	MBL	1:200
GR1	RB6-8C5	Rat IgG2b κ	Biotin	BioLegend	1:1000
H2-Kb (MHC I)	AF6-88.5	Mouse IgG2a κ	PE	BioLegend	1:200
Helios	22F6	Armenian Hamster IgG	BV421	BioLegend	1:40
Helios	22F6	Armenian Hamster IgG	PE	BioLegend	1:40
I-A/I-E (MHC II)	M5/114.15.2	Rat IgG2b κ	FITC	BioLegend	1:1000
I-A/I-E (MHC II)	M5/114.15.2	Rat IgG2b κ	BV421	BioLegend	1:1000
I-A/I-E (MHC II)	M5/114.15.2	Rat IgG2b κ	APC-Cy7	BioLegend	1:1000
ICOS	C398.4A	Armenian Hamster IgG	PE	BioLegend	1:1000
Isotype control	eBR2a	Rat IgG2a κ	AF488	eBioscience	Various
Isotype control	RTK4530	Rat IgG2b κ	APC	BioLegend	Various
KI67	16A8	Rat IgG2a κ	AF647	BioLegend	1:1000
Ly51	6C3	Rat IgG2a κ	PE	BioLegend	1:200
NK1.1	PK136	Mouse IgG2a κ	Biotin	BioLegend	1:500
PD-1	RMP1-30	Rat IgG2b κ	APC	BioLegend	1:500

Podoplanin	8.1.1	Syrian Hamster IgG	APC	BioLegend	1:500
Qa2	695H1-9-9	Mouse IgG2a κ	Biotin	BioLegend	1:500
SCA1	M1/69	Rat IgG2b κ	BV510	BD Biosciences	1:500
Streptavidin			APC	BioLegend	1:1000
Streptavidin			BV605	BioLegend	1:1000
Streptavidin			APC-Cy7	BioLegend	1:500
TCRb	H57-597	Armenian Hamster IgG	PE	eBioscience	1:1000
TCRb	H57-597	Armenian Hamster IgG	APC-Cy7	BioLegend	1:1000
TCRb	H57-597	Armenian Hamster IgG	FITC	BioLegend	1:500
TCRb	H57-597	Armenian Hamster IgG	PECy5	BioLegend	1:1000
Ter119	TER-199	Rat IgG2b κ	Biotin	BioLegend	1:1000
TSPAN8	657909	Rat IgG2b	APC	R&D Systems	1:100
UEA-1	Polyclonal	Goat IgG	Biotin	Vector Laboratories	1:500
γδTCR	GL3	Armenian Hamster IgG	Biotin	eBioscience	1:1000

2.3.6 Annexin V binding flow cytometry assay

Following staining for cell surface antigens, samples were washed with 1ml Annexin V binding buffer (BioLegend), resuspended in 100 μ l of Annexin V binding buffer and then stained with 1 μ l Annexin V::FITC (BioLegend) for 15 minutes at room temperature. After washing with a further 1ml of Annexin V binding buffer, cells were stained with DAPI at a concentration of 150ng/ml for 5 minutes at room temperature and then analysed by flow cytometry. For each assay, negative and positive control tubes were included, consisting of cells unstained with Annexin V::FITC or DAPI and cells fully stained after UV light treatment for 20 minutes respectively.

2.3.7 Aldefluor flow cytometry assay

This flow cytometric assay was performed using the Aldefluor kit (Stemcell Technologies) according to the manufacturer's instructions. Briefly, following TEC isolation and enrichment (sections 2.3.1 and 2.3.2), each sample was resuspended in 2mls of Aldefluor assay buffer and subsequently split into two tubes, each containing 1ml, labelled "control" and "test" respectively. 5 μ l of activated Aldefluor reagent was added to each tube, and 10 μ l of diethylaminobenzaldehyde (DEAB), a specific inhibitor of aldehyde dehydrogenase enzymes, was immediately added to the control tubes only. After mixing, the samples were incubated at 37°C in a water bath for 45 minutes in the dark. Samples were then pelleted at 300g for 5 minutes at 4°C and resuspended in 100 μ l of Aldefluor assay buffer. Staining for cell surface antigens, subsequent wash steps, DAPI staining for viability and analysis via

flow cytometry were all subsequently performed as per Subsection 2.3.5, with the exception that Aldefluor assay buffer was used in place of FACS buffer.

2.3.8 Fetal thymic organ culture

Timed matings were performed (Subsection 2.1.7) and embryo harvesting was performed once the required gestational age had been attained (Subsection 2.3.3). All steps were performed under strict sterile conditions and in a laminar flow cell culture hood where possible. Isolated, cleaned thymic lobes were placed on autoclaved 0.8 μ m polycarbonate filter membranes (Whatman), which had been positioned shiny side facing upwards in a 24-well cell culture plate (Costar), floating in 2mls of fetal thymic organ culture (FTOC) medium containing IMDM + Glutamax (Gibco) supplemented with 10% heat inactivated FCS (Hyclone), 100 units/ml Penicillin and 100 μ g/ml Streptomycin (Sigma). FTOCs were incubated at 37°C in 10% CO₂ with the cell culture medium being replaced every 3-5 days.

The exact culture conditions varied according to specific experimental protocol:

TEC response to RANK-ligand (RANK-L): E17.5 thymic lobes were initially lymphodepleted by culturing with 1.35mmol 2-deoxyguanosine (2DG; Sigma) for 4 days. The cell culture medium was then replaced fresh FTOC medium +/- 1 μ g/ml soluble RANK-L (kind gift from Jakub Abramson) and the cultures were incubated for a further 4 days before being collected for analysis.

Thymocyte response to retinoic acid (RA): E14.5 thymic lobes were cultured for 10 days in fresh FTOC medium +/- 10^{-6} M RA (Sigma). As the stock solution of RA was dissolved in DMSO, an equivalent amount of DMSO was added to the control cultures.

TEC response to RA: E14.5 thymic lobes were lymphodepleted by culturing with 1.35mmol 2DG (Sigma) for 4 days. The cell culture medium was then replaced with fresh FTOC medium +/- 10^{-6} M RA (Sigma). As the stock solution of RA was dissolved in DMSO, an equivalent amount of DMSO was added to the control cultures. The cultures were incubated for a further 10 days before being collected for analysis.

TEC response to insulin-like growth factor 1 (IGF-1): E14.5 thymic lobes were lymphodepleted by culturing with 1.35mmol 2DG (Sigma) for 4 days. The cell culture medium was then replaced with either plain FTOC medium or FTOC medium containing 20ng/ml recombinant IGF-1 (Sigma), and the cultures were incubated for a further 7 days before being collected for analysis.

Once ready for analysis FTOCs were lifted from their membranes using fine forceps. TEC or thymocytes were isolated from FTOCs as per Subsections 2.3.3 and 2.3.4 respectively and prepared for flow cytometry (Subsection 2.3.5).

2.4 Molecular techniques

2.4.1 Genomic excision PCR

5,000 CD45⁺, cTEC, mTEC^{Lo}, and mTEC^{Hi} from CHD4 cKO and WT mice were FACS sorted into 1.5ml eppendorf tubes coated with FCS. Samples were pelleted at 500*g* for 5 minutes, then dry pellets were snap frozen in liquid nitrogen and stored at -20°C for 1-2 weeks. To extract genomic DNA, the cell pellets were lysed in 500μl lysis buffer containing 100mM Tris-HCl pH8 (Sigma), 5mM EDTA (Sigma), 0.2% SDS (Sigma), 200mM NaCl (VWR), and 100μg/ml Proteinase K (Roche) in a water bath at 55°C for 5 hours. DNA was precipitated with an equal volume of isopropanol (VWR) and pelleted at maximum speed for 10 minutes at room temperature. The DNA pellet was then washed with 70% ethanol (VWR) and centrifuged at maximum speed for a further 10 minutes before being air-dried and resuspended in 25μl TE buffer (10mM Tris-HCl pH8, 1mM EDTA). DNA samples were kept at 4°C for up to 4 weeks or at -20°C for longer-term storage. Extracted DNA was subsequently used to perform PCR reactions using 2 primers pairs (Sigma): one that amplified a product within the deleted region (“Deleted Region”); and the other that amplified a product only in the cKO (“Delta Allele”) and failed to amplify a product in the WT due to prohibitively large size (Table 2.5). Please refer to Subsections 2.2.3 and 2.2.4 for methods relating PCR and agarose gel electrophoresis respectively.

Table 2.5 Primer pairs used for *Chd4* genomic excision PCR

Primer Pair	Name	Direction	Sequence		Ta	Expected Products
			5'	3'		
Deleted Region	Chd4In14F	Forward	GGGAAGGGCTTCTTTACCAG		56°C	209bp (WT allele)
	Chd4Ex15F	Reverse	TGCTGTCTGCACAGTTTTC			
Delta allele	Mi2b+F	Forward	CTCCAAGAAGAAGACGGCAGATCT		56°C	≈250bp (cKO allele)
	KG4R	Reverse	CTTCCACAGTGACGTCCAGACGCA			

2.4.2 RNA extraction

For RNA extraction from fresh cells (i.e. non fixed/permeabilised), up to 1×10^5 cells were FACS sorted into 1.5ml DNase/RNase free eppendorf tubes containing 75µl of RLT buffer (Qiagen) and lysed by vortexing for 1 minute. The resultant cell lysates were either stored at -80°C for 1-2 weeks or processed directly. The RNeasy Micro Kit (Qiagen) was used according to the manufacturer's instructions with the inclusion of the on-column genomic DNA digestion step. During the final step, RNA was eluted in 14µl of nuclease free water, which was incubated on the column membrane for 1 minute at room temperature before the final spin step.

For RNA extraction from fixed/permeabilised cells, cells were FACS sorted into 1.5ml DNase/RNase free eppendorf tubes pre-coated with FCS and containing 150µl of 1xPBS. FACS sorted cells were subsequently washed with 1xPBS and pelleted at 500g for 10 minutes at 4°C. Dry cell pellets were then snap frozen in liquid nitrogen and stored at -80°C for up to 2 weeks. The RNeasy FFPE Kit (Qiagen) was used, according to the manufacturer's protocol for

“Purification of total RNA from microdissected FFPE tissue sections” with the inclusion of the on-column genomic DNA digestion step. During the final step, RNA was eluted in 14µl of nuclease free water, which was incubated on the column membrane for 1 minute at room temperature before the final spin step.

2.4.3 cDNA synthesis

RNA was converted to cDNA using the SensiFAST kit (Bioline). Reaction mixes were set up according to Table 2.6 and reactions were then performed on a 96-well thermal cycler (Applied Biosystems) using the cycling conditions in Table 2.7.

Table 2.6 cDNA synthesis reaction mix for a single reaction tube

Reagent	Volume for 1x reaction
RNA	6-12µl
5x TransAmp Buffer	4µl
Reverse Transcriptase	1µl
Nuclease free water	8-14µl
Total Volume	20µl

Table 2.7 Thermal cycler conditions for cDNA synthesis reaction

Step	Temperature	Time
Primer annealing	25°C	10 minutes
Reverse Transcription	42°C	15 minutes
Inactivation	85°C	5min
Hold	4°C	∞

2.4.4 Quantitative real-time PCR

Quantitative real-time PCR (RT-qPCR) was performed using the SensiMix SYBR Hi-ROX Kit (Bioline). Reaction mixes were set up according to Table 2.8. All primers (Sigma) were designed to cross exon-exon boundaries and are listed in Table 2.9. In addition, all primers were tested for specificity on first use by running PCR reaction products on an agarose gel and verifying expected fragment size. RT-qPCR reactions were performed on a StepOnePlus real-time PCR system (Applied Biosystems) using the cycling conditions in Table 2.10. The delta CT (ΔCT) method was used in order to normalise expression to that of a housekeeping gene, with relative expression values expressed as $2^{-\Delta CT}$. The delta delta CT ($\Delta\Delta CT$) method was used for analyses of fold change, which was expressed as $2^{-\Delta\Delta CT}$.

Table 2.8 RT-qPCR reaction mix for a single reaction tube

Reagent	Stock Concentration	Volume for 1x reaction	Final Concentration
Nuclease-free water (Ambion)	-	4 μ l	-
SensiMix SYBR	2x	5 μ l	1x
Forward Primer (Sigma)	10 μ M	0.25 μ l	0.25 μ M
Reverse Primer (Sigma)	10 μ M	0.25 μ l	0.25 μ M
cDNA	-	0.5 μ l	-
Total Volume		10μl	

Table 2.9 Primer pairs used for RT-qPCR

Primer Pair	Direction	Sequence	Ta	Expected Product
		5' 3'		
<i>Aire</i>	Forward	TGTGCCACGACGGAGGTGAG	60°C	144bp
	Reverse	GGTTCTGTTGGACTCTGCCCTG		
<i>bActin</i>	Forward	GTTCCGATGCCCTGAGGCTC	60°C	108bp
	Reverse	CGGATGTCAACGTCACACTTCAT		
<i>Cd40</i>	Forward	TGTGAACCCAATCAAGGGCT	60°C	96bp
	Reverse	CTTGCTGGTGCAGTGTGTGTC		
<i>Chd4 Ex5-6</i>	Forward	CCTAAGTCATCAGCCCAGCTT	60°C	131bp
	Reverse	GCAGCAATCAGGGGTCTAACA		
<i>Chd4 Ex14-15</i>	Forward	TGAGGGGTGAAGAAGGACGA	60°C	102bp
	Reverse	ACTTCACGGTGGGATCAACAG		
<i>Chd4 Ex27-28</i>	Forward	GACCGAGATTGGCAGGACGA	60°C	162bp
	Reverse	GGCCAACAGAGGAGGTAATGG		
<i>Cyt11</i>	Forward	GCTGAGAGACTTCGTGGCTT	60°C	125bp
	Reverse	AGAGGAATACCAAGTCCCGC		
<i>Dppa1</i>	Forward	TACTCTTCTCCCAGCAGTTTG	60°C	157bp
	Reverse	AGGAGGATCAGCAATGCAGC		
<i>Gapdh</i>	Forward	GGTGAAGGTCGGTGTGAACG	60°C	123bp
	Reverse	ACCATGTAGTTGAGGTCAATGAAGG		
<i>Gluca2a</i>	Forward	GTAGCCACTCTGCACTTCCA	60°C	92bp
	Reverse	CTGAGCAATGGCCTCTAGCC		
<i>Gp2</i>	Forward	CAAGAACAGATGCCCAAACCAA	60°C	214bp
	Reverse	AATGGCTGGTCTACTACTGCG		
<i>Ilfnb</i>	Forward	TGCCTTTGCCATCCAAGAGA	60°C	121bp
	Reverse	TGTCTGCTGGTGGAGTTCATC		
<i>Irf4</i>	Forward	ACAGCTCATGTGGAACCTCT	60°C	120bp
	Reverse	GCGGTGGTAATCTGGAGTGG		
<i>Knq1</i>	Forward	AAAGGAGCGTTTTCTTCCT	60°C	149bp
	Reverse	CCAAATCATCTCCTGAATAAAGGGT		
<i>Ppy</i>	Forward	CAGGCCTAGGTATGGGAAGA	60°C	159bp
	Reverse	AGAGAGGCTGCAAGTCCATT		
<i>Prap1</i>	Forward	CAGGAAACAGAGAAGGTCTGGG	60°C	99bp
	Reverse	TGCAGGCTTCTGTTTGGGTT		

<i>Rag1</i>	Forward	TGCAGACATTCTAGCACTCTGG	60°C	174bp
	Reverse	TTCAAAGGATCTCACCCCTAAACAG		
<i>SpiB</i>	Forward	GCTGTTTGTACCCAGAAGGAGT	60°C	91bp
	Reverse	CATGTAGAGTCAAGGCCCCC		
<i>Svs4</i>	Forward	TCAAAGAGAAAGAAGAGCCGGT	60°C	91bp
	Reverse	ATGGCTCCAGCATATTGGTCC		
<i>Tff3</i>	Forward	TACGTTGGCCTGTCTCCAAG	60°C	134bp
	Reverse	CAGGGCACATTTGGGATACT		
<i>Tspan8</i>	Forward	TTCAGTCGGAGTTCAAGTGCT	60°C	215bp
	Reverse	AACGGCCAGTCCAAAAGCAA		

Table 2.10 Cycling conditions for RT-qPCR

Stage	Step	Cycles	Temperature	Time	Data capture
qPCR	Initial Denaturation	1	95°C	10 minutes	
	Amplification	40	Denaturation: 95°C	15 seconds	
			Annealing: 60°C	15 seconds	
			Elongation: 72°C	15 seconds	End of elongation
Melt Curve	Denaturation	1			
	Re-annealing	1	60°C	1 minute	
	Melting	1	60-95°C	ramp rate = 0.3°C per second	With each increment
	Denaturation	1	95°C	15 seconds	End of denaturation
Hold	Hold	1	4°C	∞	

2.4.5 Bulk RNA-sequencing

2,000 cTEC, CD80⁻CD86⁻ mTEC, CD80⁺CD86⁻ mTEC and CD80⁺CD86⁺ mTEC from individual 4-5 week old female CHD4 cKO (n=3) and WT (n=3) mice were FACS sorted into 1.5ml DNase/RNase free eppendorf tubes containing 75µl of lysis buffer made from 0.2% Triton-X 100 (Sigma) and 2U/µl RNase inhibitor (Takara) in nuclease free water (Ambion). To ensure adequate lysis, samples were vortexed for 60 seconds before storing at -80°C for up to 10 days. Library preparation and sequencing was carried out by the Oxford Genomics Centre. Briefly, libraries were prepared using the SMARTer Ultra Low Input RNA kit (Clontech Laboratories) following manufacturer's instructions. Libraries were amplified on a Tetrad thermal cycler (Bio-Rad) using in-house unique dual indexing primers (264). Individual libraries were quantified using a Qubit fluorometer (Invitrogen) and the size profile was analysed on the 2200 or 4200 TapeStation (Agilent). On the basis of these measurements, molarity was calculated for each sample to allow pooling of individual libraries at an equimolar ratio. The pooled library was diluted to ~10nM for storage. The 10nM library was denatured and further diluted prior to 2x 75bp paired-end sequencing over 4 lanes of a HiSeq4000 (Illumina).

Analysis of bulk RNA-sequencing (RNA-seq) data was performed in collaboration with Dr Jeanette Baran-Gale and Professor Chris Ponting (MRC Institute of Genetics and Molecular Medicine, University of Edinburgh). Briefly, library quality was assessed using the FastQC software package (Babraham Bioinformatics) and high-quality libraries were aligned using the STAR aligner

(265). Gene counts were estimated using Salmon (266) and normalisation and analysis was performed using the DEseq2 package (267).

2.4.6 Single-cell RNA-sequencing

Cells were FACS sorted into 1.5ml DNase/RNase free eppendorf tubes pre-coated with BSA and containing 150 μ l of plain RPMI-1640. Library preparation on freshly isolated cells and sequencing was carried out by the Sanger Institute in Hinxton or the Oxford Genomics Centre. Single-cell RNA-seq libraries were prepared using the Chromium Single Cell 3' V1 kit or V2 kit (10x Genomics) and 2x 75bp paired-end sequencing was performed on a HiSeq4000 or a HiSeq2500 (Illumina).

Analysis of single-cell RNA-seq data was performed in collaboration with Dr Jeanette Baran-Gale and Professor Chris Ponting (MRC Institute of Genetics and Molecular Medicine, University of Edinburgh). Briefly, libraries were analysed using the Cell Ranger pipeline (10x Genomics) resulting in a Gene x Barcode matrix of counts for each biological sample. Secondary QC and analysis was performed using the simpleSingleCell workflow (268). Each dataset was first filtered in order to remove low quality libraries and individual samples were then sub-sampled to correct for differences in read depth between samples.

For the cell based clustering of single mTEC presented in Chapter 3, gene-by-cell matrices for each experiment were normalised, and the MNNcorrect

algorithm was used to combine the separate experiments into one corrected meta-experiment (269). Graph-based clustering was then used to assign the individual mTEC to clusters. First, a shared nearest neighbour (SNN) graph was generated from the principal components of the normalised gene-by-cell expression matrix; community structure within the resulting SNN graph was then analysed using the Louvain algorithm (270). For pseudotemporal analysis, RNA velocity was used to order the mTEC based on spliced and unspliced RNA data captured from each cell (271).

Gene module clustering was performed using a modification of the TF-IDF algorithm in order to enhance the signal from sparsely expressed genes, such as TRGs, and therefore prevent more widely expressed genes from masking the signal from these less frequently expressed genes (272). Genes detected in multiple cells were then clustered into gene modules using the HDBSCAN algorithm in order to assign genes to single gene modules that exhibited a distinctive gene co-expression profile across the mTEC dataset (273) .

2.4.7 ChIP-sequencing

ChIP-sequencing (ChIP-seq) for CHD4 was performed on 500,000 CD80⁺ mTEC per replicate (n=3). The cells were FACS sorted from a total of 24 4-6 week old female Chd4^{fl/fl} WT mice across 3 separate cell sorts. Cells were fixed and sonicated on the same day as sorting and stored at -80°C for up to 3 weeks, until all 3 replicates had been collected. ChIP samples were prepared using the iDeal ChIP-Seq Kit for Transcription Factors (Diagenode) according

to the manufacturer's instructions for scaling down for low cell numbers. 500,000 CD80⁺ mTEC were FACS sorted into 15ml Falcon tubes (Fischer Scientific) pre-coated with 2mls FCS (Sigma). After sorting, cells were washed with ice-cold 1x PBS (Sigma), pelleted at 500g for 10 minutes at 4°C, resuspended in 500µl FACS buffer and transferred to a 1.5ml low-bind eppendorf tube. Samples were fixed, according to the kit's instructions for 15 minutes at room temperature in a final concentration of 11% molecular-grade formaldehyde (Sigma) diluted in Fixation Buffer (Diagenode). The reaction was then quenched by the addition of glycine. The reaction volumes for the subsequent cell lysis and sonication steps were scaled down accordingly for 500,000 cells. Sonication was performed in 100µl of complete shearing buffer iS1b in 1.5ml TPX microtubes (Diagenode) using a Bioruptor Pico sonicator (Diagenode) cooled to 4°C. Sonication optimisation experiments were performed to establish optimal sonication conditions of 8 cycles, each consisting of 30 seconds "on", followed by 30 seconds "off". After sonication, the soluble chromatin fraction was removed and stored at -80°C until 3 replicates had been collected. Once 3 replicates had been collected, magnetic immunoprecipitation was performed according to instructions. Protein-A coated magnetic beads were incubated with antibody and blocked with BSA for 3.5 hours at 4°C on a rotator. A mixture of ChIP-grade anti-CHD4 antibodies was used for immunoprecipitation (IP) consisting of 4µg of rabbit anti-CHD4 (Abcam, polyclonal IgG, ab72418) plus 4µg of mouse anti-CHD4 (Abcam, Monoclonal IgG1, 3F2/4, ab70469) per IP. Before IP, 5% input samples were removed from each aliquot of sheared chromatin, set aside and stored at 4°C

until the elution and de-crosslinking steps. The remaining chromatin was supplemented with complete shearing buffer iS1b to make each sample up to a total volume of 250 μ l before adding to antibody conjugated protein-A beads. IPs were left at 4°C overnight on a rotator. The following day, elution from beads and decrosslinking steps were conducted as per the protocol. After a total of 4 hours of decrosslinking, DNA purification was carried out using the MicroChIP DiaPure columns (Diagenode) as per the manufacturer's instructions, with each purified DNA sample finally eluted in 12 μ l of DNA Elution Buffer. DNA concentration was quantified using the Qubit dsDNA HS Assay kit (Invitrogen) and a Qubit fluorometer (Invitrogen). Input samples were then further diluted in DNA Elution Buffer to the same concentration as their respective ChIP samples and sequencing libraries were prepared using the MicroPlex Library Preparation Kit V2 (Diagenode) according to the kit's instructions. During the library amplification step, 8 cycles of amplification were initially used for each sample. The unpurified libraries were then quantified using the 2200 TapeStation (Agilent) using high sensitivity reagents. As the ChIP'ed libraries were of lower yield compared to the input libraries they were subjected to a further 3 cycles of amplification before quantifying once again. Unpurified libraries were stored at -20°C for a few days before being purified using AMPure XP beads (Beckman Coulter) at a 1:1 ratio in order to remove any unbound adaptors. Purified libraries were eluted from the AMPure beads in 20 μ l of low TE buffer pH8, quantified and sized, as before, fluorometrically using the Qubit (Invitrogen), as well as on the 2200 TapeStation fragment analyser (Agilent), both using high sensitivity reagents. On the basis of these

measurements molarity was calculated for each sample in order to allow pooling of individual libraries at an equimolar ratio. The pooled library was diluted to ~15nM for storage and further diluted to 3.5nM prior to loading on the sequencer. 2x 75bp paired-end sequencing was carried out by the Oxford Genomics Centre over one lane of a HiSeq4000 (Illumina),

Analysis of ChIP-seq data was performed in collaboration with Dr Jeanette Baran-Gale and Professor Chris Ponting (MRC Institute of Genetics and Molecular Medicine, University of Edinburgh). Briefly, the ChIP-seq libraries were processed using the SystemPipeR workflow (274), which performs read quality filtering and trimming, before aligning the reads to the mm10 genome using Bowtie 2 (275). The aligned reads were then processed with MACS2 in broad peak mode to identify genomic regions bound by CHD4 (276). Signal relating to blacklisted regions of the genome associated with artifactual ChIP-seq signals was removed before further analysing the data to identify characteristics of CHD4 binding (277). The binding profile of CHD4 was analysed genome-wide and at specific sets of gene promoters (i.e. those relating to genes differentially expressed in the CHD4 cKO or genes regulated by AIRE) using the CEAS tool (278). Finally, deepTools was used in order to compare the binding profile of CHD4 to that of AIRE and various histone marks, as well as to assess the similarity of CHD4 binding across biological replicates (111, 198, 199, 279).

2.4.8 Western blot

750,000 CD45⁻EpCAM⁺ TEC were FACS sorted from 4-6 week old CHD4 cKO (n=16) and WT mice (n=8) in 50,000 cell batches, which were stored as dry cell pellets at -80°C until sufficient cells were collected. Cells were then pooled in 200µl of lysis buffer containing 25mM Tris-HCl pH8 (Sigma), 50mM NaCl (Sigma), 1% NP-40 (Sigma), 1mM DDT (Sigma), 10% glycerol (Sigma), and 0.2mM PMSF (Sigma) in deionised water. After vortexing, samples were incubated for 20 minutes on ice. The samples were then centrifuged at 15,000g for 10 minutes at 4°C and the supernatant, containing the cytosolic fraction, was removed and set aside. The cell pellet, containing the nuclear fraction, was incubated in 200µl of lysis buffer for a further 20 minutes on ice. 800µl of ice-cold acetone (Sigma) was added to the cytoplasmic and nuclear lysates and the samples were left at -20°C overnight in order to precipitate proteins. The following day, the samples were centrifuged at 15,000g for 10 minutes at 4°C, the supernatant was discarded and the pellet was resuspended in 15µl lysis buffer. 1µm of 14.3M β-mercaptoethanol (Sigma) and 4µl of sample buffer containing 250mM Tris-HCl pH7 (Sigma), 10% SDS (Sigma), 35% glycerol (Sigma) and 0.05% Bromophenol blue (Sigma) was subsequently added to each sample before gel electrophoresis. Proteins were separated by SDS-PAGE using a 7.5% precast gel (Biorad) loaded into a Mini-PROTEAN Tetra cell (Biorad) filled with 1x running buffer containing 25mM Tris (Sigma), 200mM glycine (Sigma), 1% SDS (Sigma). Electrophoresis was carried out at 180V for 30 minutes and samples were run alongside a protein standard (Biorad). Proteins were then transferred onto a polyvinylidene

difluoride membrane (Biorad) in a Mini-PROTEAN Tetra cell (Biorad) in 1x transfer buffer consisting of 25mM Tris (Sigma), 200mM glycine (Sigma), 20% Methanol (VWR) at 100V for 50 minutes at 4°C. The membrane was then blocked in blocking buffer consisting of 3% BSA (Sigma), in TBST buffer containing 50mM Tris-HCl pH7.5 (Sigma), 150mM NaCl (Sigma), 0.1% Tween-20 (Sigma) for 1 hour at room temperature on a shaker. The membrane was next incubated with anti-CHD4 primary antibody (Rabbit polyclonal, Abcam ab72418, used at 1:2000) for 1 hour at room temperature on a shaker. After primary antibody staining, the membrane was washed 3 times for 5 minutes each in TBS-T. Fluorescent secondary staining was performed for 30 minutes at room temperature on a shaker using an IRDye 680RD goat anti-rabbit antibody (Licor) diluted in 1:10000 blocking buffer. After secondary antibody staining, the membrane was washed 3 times for 5 minutes each in TBST before imaging on an Odyssey imaging system (Licor). As a control for protein loading, the membrane was then re-stained for EZH2 (Rabbit monoclonal, Cell Signaling, used at 1:1000) and re-imaged.

2.4.9 Proximity Ligation Imaging Cytometry

Proximity ligation imaging cytometry (PLIC) was performed in collaboration with Ayelet Avin and Jakub Abramson (Department of Immunology, Weizmann Institute of Science) as previously described (189). Full methodological details can be found within the original publication (189). To summarise, thymi from adult C57BL/6 mice were enzymatically digested using collagenase/dispase (Roche) and thymic stromal cells were enriched using Percoll (Sigma) density

gradient centrifugation. Cells were stained for cell surface antigens for 40 minutes on ice using the following antibodies: CD45::PECy7 (BioLegend, 30-F11, 1:200), EpCAM::APC-Cy7 (BioLegend, G8.8, 1:200), I-A/I-E::BV421 (BioLegend, M5/114.15.2, 1:500). The samples were fixed and permeabilised in 80% cold acetone (Sigma) for one hour at 8°C, then stained for one hour at room temperature with the appropriate combination of unconjugated antibodies against intracellular antigens according to Table 2.11, and stored at 4°C overnight.

Table 2.11 Antibodies used for PLIC

PLIC Tube	Proximity ligation targets	Details of antibodies used	Staining Concentration	Expected findings
Test Tubes	AIRE	Mouse monoclonal IgG (made in-house by Jakub Abramson)	1:100	-
	CHD4	Rabbit polyclonal IgG (Abcam ab 72418)	1:1000	
Positive Control	AIRE	Mouse monoclonal IgG (made in-house by Jakub Abramson)	1:100	PLIC signal
	SIRT1	Rabbit polyclonal IgG (Millipore)	1:200	
Negative control	CHD4 alone	Rabbit polyclonal IgG (Abcam ab 72418)	1:1000	No PLIC signal

The PLIC assay was performed the following day (189). First samples were incubated with anti-Rabbit PLUS and anti-Mouse MINUS Duolink PLA probes (Sigma). The ligation and amplification steps were then performed using Duolink In Situ Detection Reagents FarRed (Sigma), after which cells were stained with AIRE::AF488 (eBioscience, 5H12, 1:500) for 1 hour at room temperature in the dark. Samples were stored at 4°C and acquired the

following day on an ImageStream X Mark II Imaging Flow Cytometer (Amnis). Analysis was performed using IDEAS (Amnis).

2.5 Microscopy

2.5.1 Haematoxylin and Eosin Staining

Organs were harvested, cleaned under a stereoscopic microscope (Leica) and washed in 1x PBS (Sigma). They were then fixed in 10% Formalin (Sigma) for 24 hours, washed in 1x PBS for 4-5 hours and stored in 70% Ethanol (VWR) for up to 3 days before embedding. Paraffin embedding, sectioning and haematoxylin and eosin (H&E) staining were performed by the Histology Service at the Kennedy Institute of Rheumatology, University of Oxford. Briefly, samples were processed for paraffin embedding using a Tissue-Tek VIP 6 Processor (Sakura) by dehydrating in an ascending series of ethanol concentrations, clearing with xylene, and equilibrating with paraffin. The tissues were then embedded in paraffin blocks using a Tissue-Tek TEC 5 Embedding Centre (Sakura), sectioned at a thickness of 6 μ m using an SMR 2000 microtome (Sakura), mounted onto slides and dried. H&E staining was performed using a Tissue-Tek DRS 2000 Multiple Slide Stainer (Sakura). The protocol involved deparaffinisation and rehydration of sections using xylene followed by ethanol and water rinses. Haematoxylin staining was then performed by incubating sections for 6.5 minutes in Harris-modified haematoxylin solution, and then rinsing in water and 1% acid alcohol. After this sections were incubated in ammoniated water as a blueing agent, washed, stained with eosin for 1 minute and 45 seconds, and rinsed with water. The

sections were then dehydrated in ethanol and cleared in xylene before mounting in Di-n-butylphthalate mounting medium using a Tissue-Tek Glas G2 automated coverslipper (Sakura). Slides were visualised using a DM750 light microscope (Leica).

2.5.2 Immunohistochemistry

Thymi were collected into ice-cold 1x PBS (Sigma). Adherent fat, blood and connective tissue were removed under a stereoscopic microscope (Leica). Individual thymic lobes were then placed into cryomolds (Leica) and embedded in OCT compound (Tissue-Tek). Embedded samples were frozen using dry ice and butanol (Sigma), and stored at -80°C. 8µm sections were subsequently produced using a CryoStar NX70 (Thermo Scientific) cryostat, mounted onto charged glass slides (VWR) and air dried before storage at -20°C. During the staining protocol, slides were first fixed in 10% neutral buffered formalin (Sigma) and permeabilised for 10 minutes in 1x PBS, 0.3% Triton X-100 (Sigma). This was followed by blocking for one hour at room temperature in blocking buffer containing 2% goat serum (Sigma) and 0.1% Triton-X-100 in 1x PBS. Primary antibody staining was performed for 1 hour at 37°C using the antibodies listed in Table 2.12 diluted in blocking buffer. Slides were then washed three times with 1x PBS, 0.1% Triton X-100 before staining with secondary antibodies for 30 minutes at 37°C in the dark, using the antibodies listed in Table 2.12 diluted in 1x PBS, 0.1% Triton X-100. After secondary antibody staining, the slides were again washed three times with 1x PBS, 0.1% Triton X-100. For staining protocols which included the use of directly

conjugated antibodies, this step was performed after the secondary antibody staining step: staining for GP2 was performed at 4°C overnight, whereas staining for CD4 and CD8 was performed for 45 minutes at 37°C. Nuclei were stained with DAPI (Sigma), diluted to a concentration of 500ng/ml in methanol (VWR), for 10 minutes at room temperature in the dark. Finally, the slides were washed once with 1x PBS and mounted using ProLong Gold Antifade mountant (Life Technologies). Coverslips were sealed using nail polish and slides were stored at 4°C in the dark. Staining controls were included with each experiment in the form of single colour stains, isotype controls and secondary antibody only stains as appropriate. Confocal microscopy was performed on an LSM 780 inverted microscope (Zeiss) or an LSM 780 upright MP microscope and images were analysed using Fiji (280).

Table 2.12 Antibodies used for immunohistochemistry

Target	Clone	Species and Isotype	Conjugate	Company	Final Staining Concentration
AIRE	5H12	Rat IgG2c κ	-	eBioscience	1:500
CD4	RM4-5	Rat IgG2a κ	FITC	BioLegend	1:500
CD8a	53-6.7	Rat IgG2a κ	AF647	BioLegend	1:300
Cytokeratin 5 (K5)	Polyclonal	Rabbit IgG	-	Covance/BioLegend	1:500
Cytokeratin 8 (K8)	TROMA-I	Rat IgG2a κ	-	DSHB	1:500
Fibronectin (FN1)	10/Fibronectin	Mouse IgG1 κ	AF647	BD Biosciences	1:50
GP2	2F11-C3	Rat IgG2a κ	AF488	MBL	1:200
Isotype control	eBR2a	Rat IgG2a κ	AF488	eBioscience	1:200
Isotype control	MOPC-21	Mouse IgG1 κ	AF647	BioLegend	1:50
Rabbit IgG	Polyclonal	Goat	AF555	Life Technologies	1:500
Rat IgG	Polyclonal	Goat	AF488	Life Technologies	1:500
Rat IgG	Polyclonal	Goat	AF647	Life Technologies	1:500
Rat IgG	Polyclonal	Goat	AF555	Life Technologies	1:500
Tenascin C (TNC)	4C8MS	Mouse IgG1 κ	AF647	Novus Bio	1:50

2.5.3 Spatial analysis of GP2+ mTEC

Four representative 1980x1980 pixel microscope images stained with K5, GP2 and DAPI were used to identify GP2+ mTEC and areas of thymic medulla. Each colour image was processed using the EBIImage package (281). Since GP2 is only expressed by mTEC, a medullary mask was first applied to each slide by identifying areas of K5 expression. Next, GP2+ mTEC were identified and validated by visual inspection and these positions or “points” were used in the spatial analysis. To determine whether the GP2+ mTEC were clustered or

randomly dispersed within areas of thymic medulla the cumulative distribution function of the distances (G-function) between nearest neighbour GP2+ mTEC was computed and compared to those for simulated mTEC that are evenly spaced, dispersed randomly or clustered in their distribution within medullary islands.

2.5.4 Indirect immunofluorescence to look for serum autoantibodies

Organs were harvested from 8-10 week old female Rag1^{-/-} mice, including eyes, lacrimal glands, submandibular salivary glands, lungs, stomach, liver, pancreas, and ovaries. These tissues were then frozen in OCT compound, cryosectioned and stored as per Subsection 2.5.2.

Blood was taken by cardiac puncture from CHD4 cKO and WT mice previously rendered lymphopaenic (Subsection 2.1.8). After being allowed to clot at room temperature for at least an hour, samples were centrifuged at 1000g for 10 minutes at 4°C. The supernatant containing the serum was then removed using a pipette and stored in a fresh tube at -20°C.

Indirect immunofluorescent microscopy was performed as per Subsection 2.5.2 with the following adaptations: during the primary antibody staining step slides were incubated either with serum, diluted 1:50 in 1x PBS (Sigma), or with a mouse IgG antibody against alpha tubulin (Abcam, ab7291, used at 1:500), which was used as a positive control for the staining protocol. A goat polyclonal antibody against mouse IgG (Abcam, ab150117, used at 1:500) was

used for secondary antibody staining, and secondary antibody staining alone was used as a negative control.

2.6 Statistical analysis

For cell counts, FACS and RT-qPCR data, statistical analyses and graphs were produced using Microsoft Excel and GraphPad PRISM Version 7. Unless otherwise stated, graphs show mean values with error bars representing standard deviation (SD) and these values are reported within the text where appropriate as mean $\pm$ SD. Statistical significance was calculated using the Student's t-test (unpaired, two-tailed). The Bonferroni correction was applied for analyses containing multiple comparisons.

For all analyses an adjusted p value of ≤ 0.05 was considered statistically significant. In figures, p values are represented with asterisks such that ns (not significant) denotes $p > 0.05$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, and **** $p \leq 0.0001$.

3 Promiscuous Gene Expression in Single mTEC

The work detailed in this Chapter has been accepted for publication in EMBO Journal with the title “Biologically indeterminate yet ordered promiscuous gene expression in single medullary thymic epithelial cells”; at the time of submitting this thesis the manuscript was in press. In addition, a preprint has been deposited on bioRxiv (282).

3.1 Introduction, aim and hypothesis

For the process of T-cell central tolerance induction to be comprehensive, TEC, as a population, uniquely express almost the entire protein coding genome (111, 127). This exhaustive transcriptional programme, which includes genes that are tissue- (TRGs) or temporally-restricted in organs other than the thymus, is termed “promiscuous gene expression” (PGE) (110, 128). PGE serves to arm TEC with a molecular mirror of the body’s own antigens against which developing T-cells can be selected. The precise molecular mechanisms enabling PGE in TEC are not clear, but are in part under the control of AIRE, which induces the expression of just under 4000 genes in mTEC (111, 129). 533 of these genes are entirely dependent on AIRE for their expression (AIRE-dependent), and the expression of the remaining 3,260 is enhanced significantly in the presence of AIRE (AIRE-enhanced) (111). In addition, AIRE-independent mechanisms of PGE control the expression of a further 3,947 TRGs (111).

Whilst at the population level mTEC express almost all protein coding genes, at the single cell level TRG expression is heterogeneous and mosaic, with individual mature mTEC expressing only 1-3% of TRGs at a time (111, 127, 129, 178, 211). This may allow sufficiently high densities of particular self-antigen::MHC complexes to be achieved on the mTEC cell surface to elicit tolerogenic signals within excessively self-reactive thymocytes (178).

Understanding how PGE is regulated within single TEC is essential for understanding the mechanisms by which these cells achieve their uniquely broad transcriptional programme without subverting their cellular identity. However, it still remains to be conclusively established whether TRG expression in single mTEC is stochastic or, alternatively, whether co-expression patterns exist (111, 127, 129, 133, 178, 211, 212). Previous studies have been limited by low cell number, with only several hundred mTEC analysed. Therefore, the aim for this part my studies was to use large scale single-cell RNA-sequencing (RNA-seq) to characterise the transcriptional patterns associated with PGE in a large number of single mTEC, with the hypothesis that PGE is not stochastic but instead defined by the presence of multiple co-expression groups. Furthermore, interrogating patterns of co-expression may offer further insight into the molecular mechanisms underlying the uniquely broad transcriptional programme of mTEC.

3.2 Generation of RNA-sequencing libraries from single mTEC

3.2.1 Isolation of mTEC promiscuously expressing Tetraspanin 8 or Glycoprotein 2

To enrich for mTEC promiscuously expressing AIRE-regulated genes, and thus add statistical power to co-expression analyses, single-cell RNA-seq was performed using mTEC preselected for expression of two AIRE-regulated TRGs, as well as unselected mTEC. Tetraspanin 8 (TSPAN8) and Glycoprotein 2 (GP2) are TRGs that are promiscuously expressed in mTEC in an AIRE-enhanced fashion (111, 212). TSPAN8 is a transmembrane protein expressed in the gastrointestinal tract as well as in several carcinomas (283-285). GP2 is a zymogen granule membrane-associated protein whose expression in the periphery is restricted to pancreatic acinar cells and M cells of the gastrointestinal tract where it plays a role in the innate immune response to bacteria (286, 287). Loss of tolerance to GP2 has been associated with Crohn's disease and primary sclerosing cholangitis (288, 289).

It was possible to identify mTEC promiscuously expressing TSPAN8 (TSPAN8+) or GP2 (GP2+) proteins on their cell surface via flow cytometry (212), with TSPAN8+ mTEC constituting $6.62 \pm 0.78\%$ and GP2+ mTEC $2.35 \pm 0.81\%$ of the total mTEC compartment (Figure 3.1a and Supplementary Figure 1). RT-qPCR showed that FACS sorting mTEC on the basis of cell surface expression of TSPAN8 or GP2 proteins allowed for the enrichment of cells actively transcribing these AIRE-regulated TRGs (Figure 3.1b). This was important given that the intention was to add statistical power to the analyses

of PGE co-expression patterns in part by enriching for mTEC transcribing particular AIRE-regulated genes.

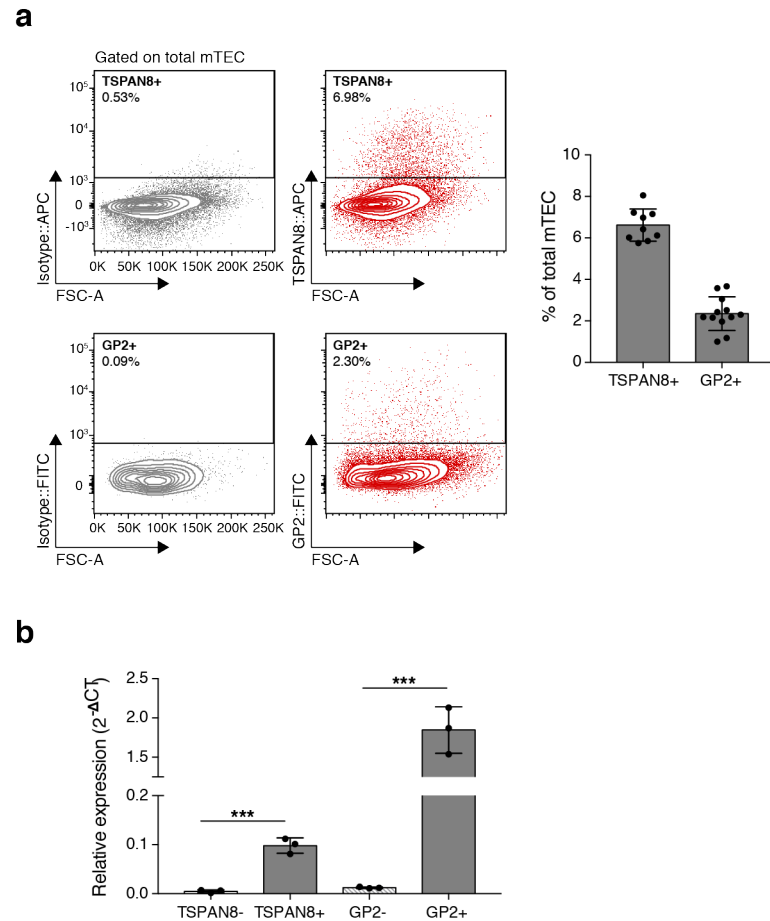


Figure 3.1 Identification of mTEC expressing the AIRE-enhanced TRGs *Tspan8* and *Gp2* via fluorescence activated cell sorting

(a) Representative FACS plots (left panels) showing staining for TSPAN8 and GP2 within total mTEC (CD45⁺ EpCAM⁺ Ly51⁻; see Supplementary Figure 1 for gating); gates for TSPAN8 and GP2 (red) were set against isotype control antibodies (grey); bar chart (right panel) showing frequency of TSPAN8⁺ cells and GP2⁺ cells within total mTEC. Results represent pooled data from 3 (TSPAN8⁺) or 4 (GP2⁺) independent experiments each containing 3 individual mice aged 4-6 weeks old. (b) Bar graph showing expression of *Tspan8* and *Gp2* mRNA by RT-qPCR on FACS sorted mTEC negative or positive for TSPAN8 protein (TSPAN8⁻ and TSPAN8⁺), or negative or positive for GP2 protein (GP2⁻ and GP2⁺); bars represent mean expression relative to *b-actin*; n=3, representative of 2 independent experiments. The fold change in mRNA expression for *Tspan8* in FACS sorted TSPAN8⁺ vs. TSPAN8⁻ mTEC was 20, and that for *Gp2* was 150.

3.2.2 Generation of large-scale RNA-sequencing data from single mTEC

Single-cell RNA-seq libraries were generated using the 10X Genomics platform from FACS sorted mTEC derived from 15 individual 4-6 week old female mice across three independent experiments (Table 3.1). To examine whether gene expression patterns differed between mouse strains, C57BL/6 (n=9) and BALB/c (n=2) mice were investigated, as well as their F1 cross, C57BL/6 x BALB/c (n=4).

In total, the transcriptomes of 6,894 single mTEC were analysed, making this the largest single-cell RNA-seq dataset investigating PGE in mTEC. The single cells analysed included 794 TSPAN8+, 935 TSPAN8-, 1,001 GP2+, 395 GP2-, and 3,769 unselected mTEC (Figure 3.2 and Table 3.1). In combination, these cells expressed 19,091/21,663 (i.e. 88%) of protein coding genes, encoded across all chromosomes, apart from Y, which was expected given that only female mice were used. These included 89% of AIRE-dependent (n = 477), 98% of AIRE-enhanced (n = 3,210) and 94% of AIRE-independent TRGs (n = 3,720) defined in a previous study (111). The transcriptomes of the single cells in this study, therefore, largely recapitulated previous TEC population-level transcriptomic analyses (111). A median of 1,830 genes was detected per cell, of which a median of 50 were AIRE-induced TRGs.

Table 3.1 Details of the 3 independent experiments contributing to single-cell RNA-seq analysis of PGE in mTEC

For each sample the experiment, mouse ID, mouse strain, cell type FACS sorted (coloured by category) and number of cells captured are listed.

Experiment	Mouse ID	Strain	Cell Type	Number of Cells
1	B6_1	C57BL/6	Unselected mTEC	885
	B6_2	C57BL/6	Unselected mTEC	859
	BC_1	BALB/c	Unselected mTEC	516
	BC_2	BALB/c	Unselected mTEC	692
	F1_1_2	C57BL/6 x BALB/c F1	GP2+ mTEC	110
	F1_3	C57BL/6 x BALB/c F1	GP2+ mTEC	243
	F1_3	C57BL/6 x BALB/c F1	GP2- mTEC	189
	F1_4	C57BL/6 x BALB/c F1	GP2+ mTEC	99
	F1_4	C57BL/6 x BALB/c F1	GP2- mTEC	206
2	B6_3	C57BL/6	Unselected mTEC	239
	B6_4	C57BL/6	Unselected mTEC	332
	B6_5	C57BL/6	Unselected mTEC	246
	B6_5	C57BL/6	GP2+ mTEC	292
	B6_6	C57BL/6	GP2+ mTEC	257
3	B6_7	C57BL/6	TSPAN8+ mTEC	77
	B6_7	C57BL/6	TSPAN8- mTEC	604
	B6_8	C57BL/6	TSPAN8+ mTEC	254
	B6_8	C57BL/6	TSPAN8- mTEC	135
	B6_9	C57BL/6	TSPAN8+ mTEC	463
	B6_9	C57BL/6	TSPAN8- mTEC	196

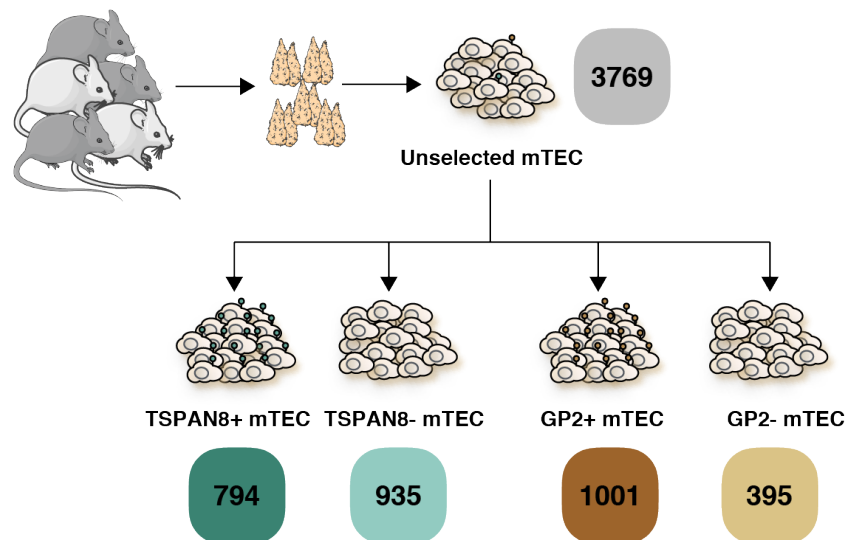


Figure 3.2 Diagram showing cell populations submitted for single-cell RNA-seq

Schematic representation summarising the cell populations sorted by flow cytometry for single-cell RNA-seq over the course of three independent experiments. Total numbers of recovered cells are indicated below and coloured by category.

3.3 Cell subpopulation clustering of single mTEC

3.3.1 Transcriptional heterogeneity within the mTEC compartment

Projection of all the 6,894 single mTEC transcriptomes into a reduced dimensional space revealed considerable transcriptional heterogeneity within the mTEC compartment. The resulting clustering structure was characterised by a large, contiguous central body of cells surrounded by a number of satellite clusters (Figure 3.3a). When analysed according to the cell surface phenotype used for FACS sorting, the central body of mTEC was characterised by a transition from TSPAN8⁻ or GP2⁻ at the lower right pole, to TSPAN8⁺ or GP2⁺ at the upper left pole (Figure 3.3a). In addition, TSPAN8⁺ and GP2⁺ mTEC each appeared to prefer distinct satellite clusters (Figure 3.3b green and brown arrows respectively), indicating that subsets of mTEC that highly express a particular TRA contain transcriptomes that differentiate them from most other mTEC, a finding that would be inconsistent with PGE being entirely stochastic within single mTEC. Despite being collected from three separate experiments spanning different strains of mice and cell surface phenotypes (Table 1), each of the conditions individually captured the diversity displayed by the combined dataset (Figure 3.3b).

Taken together, the analyses presented in this section indicate that a reproducible set of transcriptomically diverse mTEC subpopulations can be identified across independent experiments using different mouse strains, and that TRA expression biases are apparent across those subpopulations.

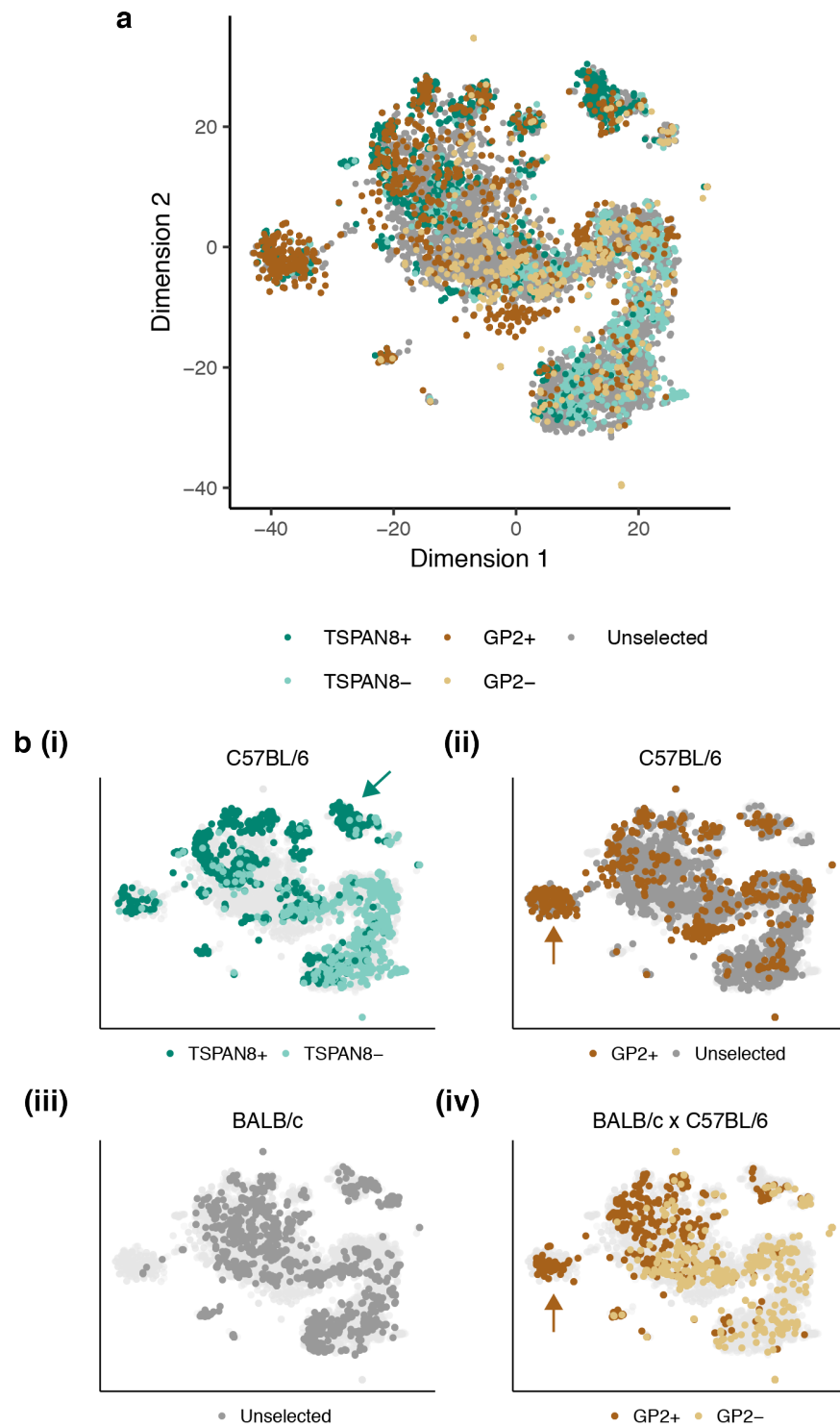


Figure 3.3 Transcriptional heterogeneity within the mTEC compartment is reproducible across different experiments, mouse strains and cell surface phenotypes

(a) tSNE visualisation of mTEC subpopulations from the combined dataset; each dot represents a single mTEC, coloured by the surface phenotype established via flow cytometry. (b) tSNE visualisations of individual datasets, overlaid on the combined dataset (light grey); each dot represents a single mTEC, coloured by the surface phenotype established via flow cytometry; green arrow identifies TSPAN8 preferred cluster and brown arrow identifies GP2 preferred cluster; the mouse strain from which the cells originate is labelled at the top of each plot.

3.3.2 Transcriptional heterogeneity within the main body of single cells is reflective of mTEC maturation

The 6,894 single mTEC resolved into 15 distinct subpopulations using an unsupervised clustering approach (Figure 3.4a). The subpopulations within the central main body of mTEC were largely defined by genes that vary in expression during mTEC maturation (Figure 3.4b-m). All 15 subpopulations were reproducible across unselected and selected mTEC, as well as across the different mouse strains used (Figure 3.5a and b). Furthermore, the structure was maintained when the cells were clustered using only TRGs, or conversely when TRGs were excluded from the analysis (Figure 3.5c), suggesting that, although the cluster structure was not purely driven by PGE (structure maintained when TRGs excluded), an ordered process may be involved in selecting TRGs to be co-expressed in single mTEC (structure maintained with only TRGs).

Clusters 1 and 2 (Figure 3.4a) contained mainly unselected, TSPAN8- or GP2- mTEC (Figure 3.3 and 3.5a) that had very little mRNA expression of *Aire* (Figure 3.4b) or AIRE-regulated TRGs, including *Tspan8* and *Gp2* (Figure 3.4c-e). Their expression of podoplanin (*Pdpn*) and *Ccl21a* (Figure 3.4f and g) was highly suggestive of an immature pre-AIRE mTEC identity. Podoplanin has been described to mark junctional TEC, which have progenitor activity restricted to the mTEC lineage and are localised to the CMJ (54); and CCL21 is a chemokine expressed by pre-AIRE mTEC, which express low levels of MHCII and CD80 (55). **Cluster 3** (Figure 3.4a) also contained mainly

unselected, TSPAN8- or GP2- mTEC (Figure 3.3 and 3.5a), however these cells did express *Aire* and low numbers of AIRE-dependent TRGs (Figure 3.4b and c). Owing to their expression of cell cycle related genes, including *Mki67* (Figure 3.4h), the cells in cluster 3 were predicted to be in the G₂/M-phase of the cell cycle and thus could represent a proliferating subpopulation or maturational stage of mTEC (290). **Cluster 4** likely represents the next stage in mTEC maturation as it contained a higher proportion of TSPAN8+ and GP2+ mTEC than the preceding clusters (Figure 3.3 and 3.5a). The cells within cluster 4 also highly expressed *Aire* and AIRE-dependent TRGs (Figure 3.4b and c). **Clusters 5 and 6** had several features that suggest they represent the most competent mature mTEC stages for central tolerance induction (57). They expressed TSPAN8 and GP2 protein and mRNA more frequently than clusters 1-4 (Figure 3.5a and 3.4d and e). They also expressed *Aire* and high numbers of AIRE-dependent TRGs (Figure 3.4a and b), as well as the co-stimulatory molecules *Cd80* and *Cd86* (Figure 3.4i and j). Finally, the cells in **clusters 7 and 8** possess characteristics of terminally differentiated post-AIRE mTEC. These mTEC expressed TSPAN8 and GP2 protein and mRNA (Figure 3.5a and 3.4d and e) and a moderate number of AIRE-dependent TRGs (Figure 3.4b), but had very little to no expression of *Aire*, *Cd80*, and *Cd86* (Figure 3.4a, i and j). They did, however, express markers of epithelial cell terminal differentiation known to mark post-AIRE mTEC and Hassall's corpuscles, including involucrin (*Iv*), cytokeratin 10 (*K10*) and *Spink5* (Figure 3.4k-m) (55, 291, 292). In support of this, TSPAN8+ and GP2+ mTEC were found to have features of terminal differentiation via flow cytometric

phenotyping (Figure 3.6a-c), including reduced expression of MHCII and CD80 (GP2+ mTEC only), reduced expression of AIRE and increased expression of Desmoglein 3 (DSG3), another marker associated with epithelial cell terminal differentiation found in Hassall's corpuscles (62, 293).

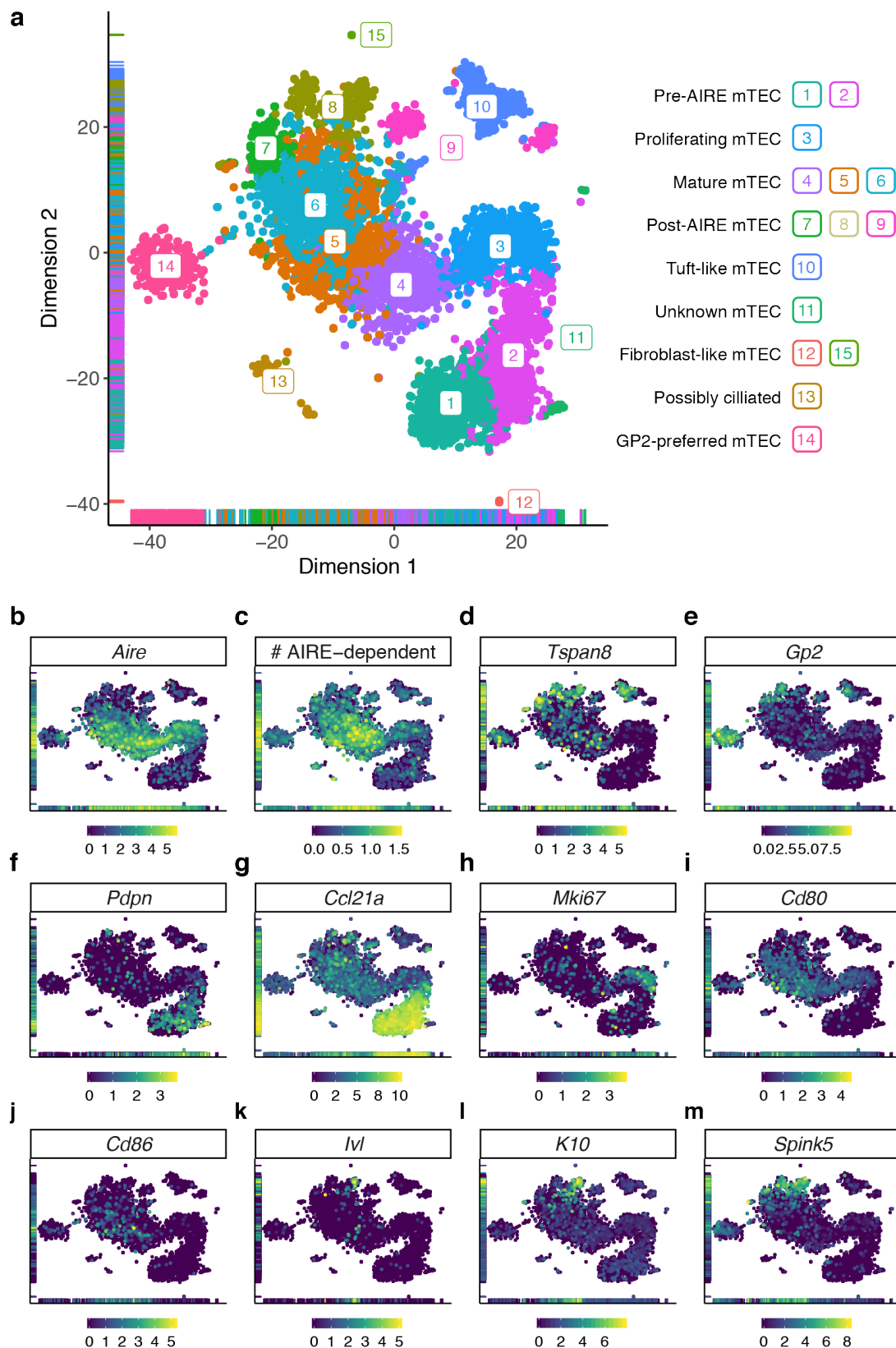


Figure 3.4 Transcriptional heterogeneity in the main body of single mTEC reflects their maturation
 (a) tSNE visualisation of mTEC subpopulations. Each dot represents a cell coloured by cluster number. The inferred identity of each cluster is provided in the legend to the right of the plot. (b and d-m) Log₂ expression level of *Aire*, *Tspan8*, *Gp2*, *Pdpr*, *Ccl21a*, *Mki67*, *Cd80*, *Cd86*, *Irf1*, *K10* and *Spink5* across the dataset. (c) Log₁₀ of the number (#) of AIRE-dependent genes expressed per cell.

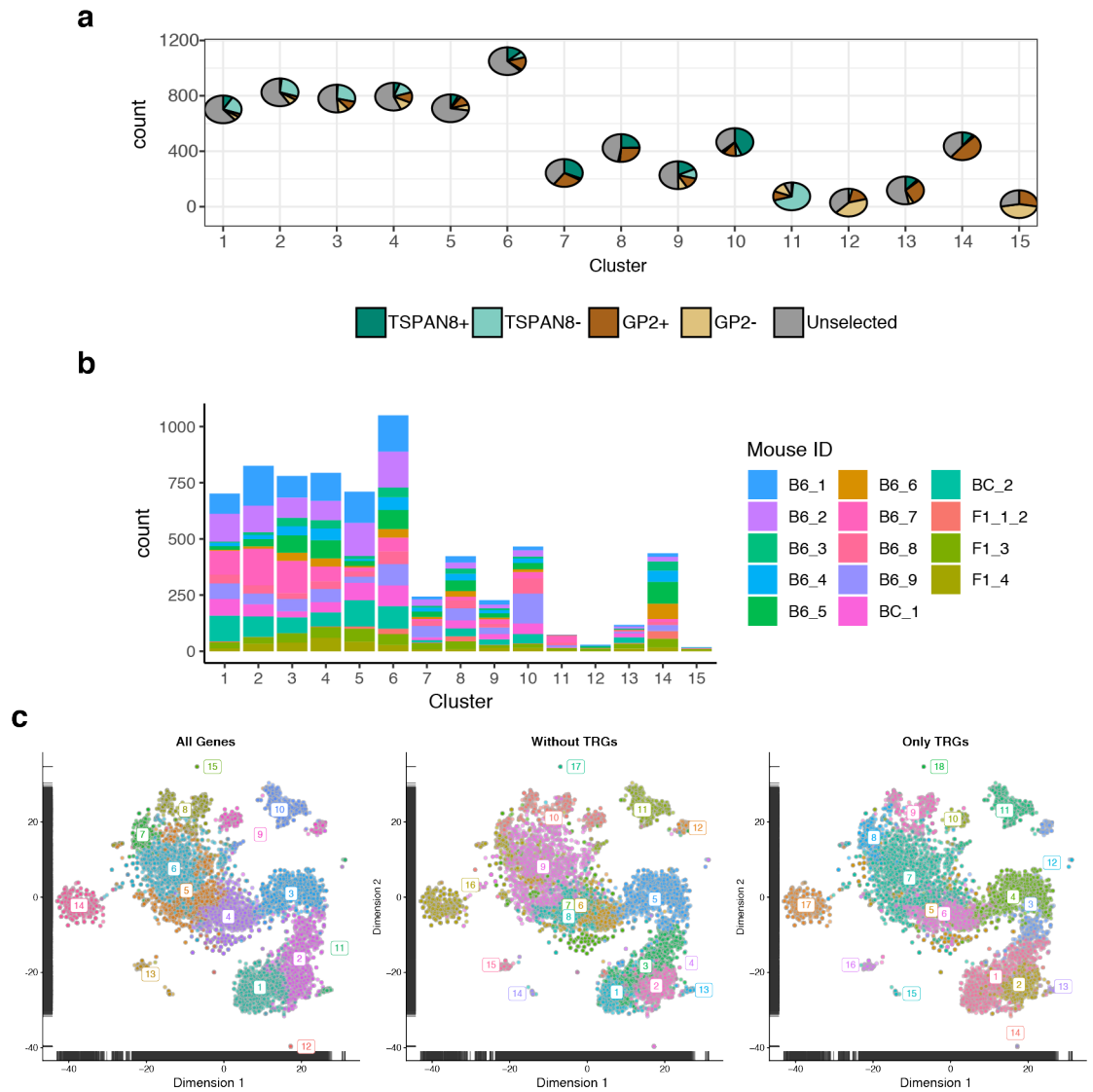


Figure 3.5 Contribution to mTEC clusters

(a) Count of cells (y-axis) within each of the 15 defined mTEC clusters (x-axis). The FACS phenotype of the mTEC within each cluster is shown as a pie chart. (B) Count of cells (y-axis) from different mice within each cluster (x-axis). (C) Clusters identified using all genes (left), without any TRGs (middle), or with only TRGs (right)

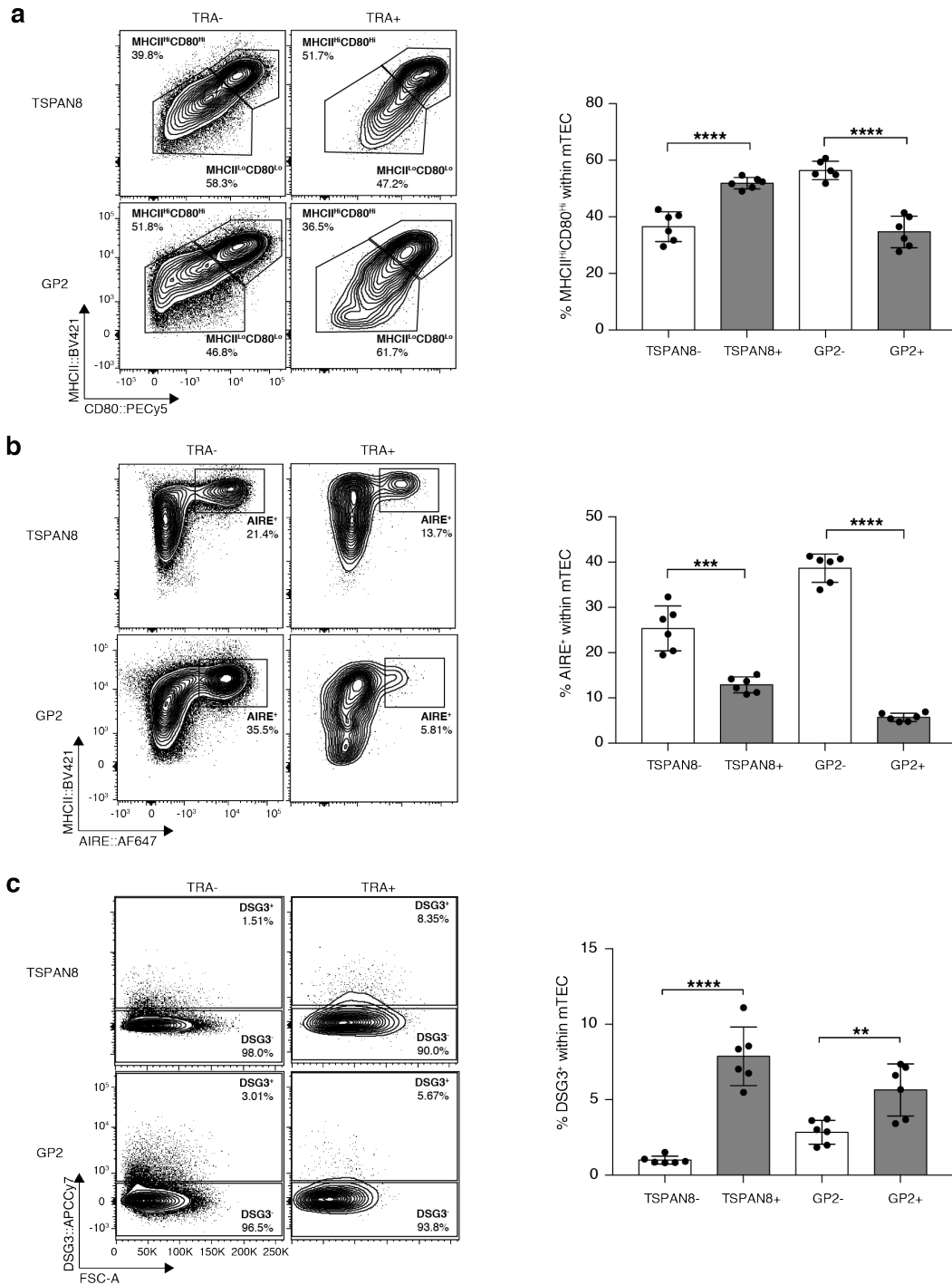


Figure 3.6 mTEC expressing the AIRE-enhanced TRAs TSPAN8 and GP2 have features of terminal differentiation

(a-c) Representative FACS plots (left) and bar charts (right) showing frequency of (a) MHCII^{Hi}CD80^{Hi}, (b) AIRE⁺ and (c) Desmoglein 3⁺ (DSG3⁺) cells within TSPAN8+ and GP2+ mTEC (TRA+) compared to their TRA- counterparts. Results represent pooled data from 2 independent experiments each containing 3 individual mice aged 4-6 weeks old.

In support of the inferred maturation trajectory presented above, in which the mTEC within the main central body of cells mature from the cells in clusters 1 and 2 at the right lower pole towards clusters 7 and 8 at the left upper pole, were findings from RNA velocity analysis (Figure 3.7). RNA velocity uses pre- and post-spliced mRNA reads to order cells, and can thus be used to produce single-cell temporal trajectories akin to traditional pseudotime analyses (271). The trajectories produced by RNA velocity analysis suggested that the *Aire*⁻ *Cd80*⁻ *Cd86*⁻ mTEC of cluster 2 preceded the proliferating mTEC in cluster 3, which in turn gave rise to the *Aire*⁺ mTEC of clusters 4-6. Finally, the *Aire*⁻ *Cd80*⁻ *Cd86*⁻ mTEC of clusters 7 and 8 appeared to originate from the *Aire*⁺ mTEC of clusters 5-6, consistent with a post-AIRE, terminally differentiated state. RNA velocity additionally suggested that the mTEC in cluster 2 also preceded those in cluster 1. The mTEC in cluster 1 highly expressed genes that have previously been associated with a quiescent mTEC progenitor subpopulation, such as *Itga6* and *Sca1* (Figure 3.8) (50). Taken together, this could suggest that the mTEC in cluster 2 represent MHCII^{Lo} immature mTEC that can either differentiate into mature AIRE⁺ and subsequently post-AIRE mTEC (via cluster 3-8) or alternatively into a pool of quiescent progenitors (cluster 1).

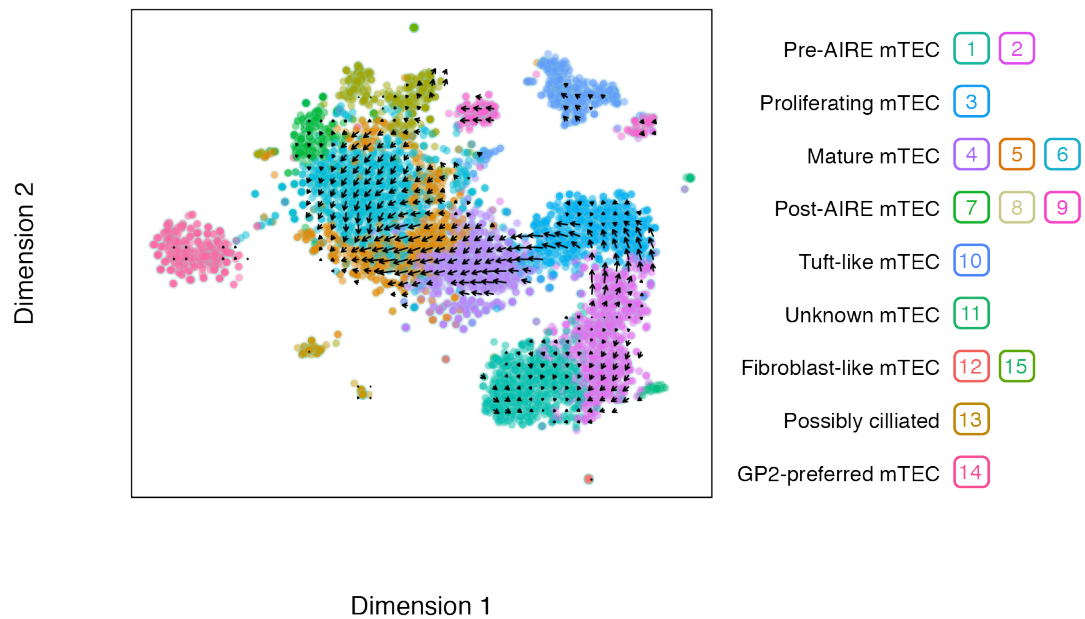


Figure 3.7 RNA velocity analysis shows trajectory of mTEC maturation

t-SNE visualisation of mTEC subpopulations with arrows showing the ordering of cells as identified by RNA velocity analysis. Each dot represents a cell coloured by cluster number. The inferred identity of each cluster is provided in the legend to the right of the plot.

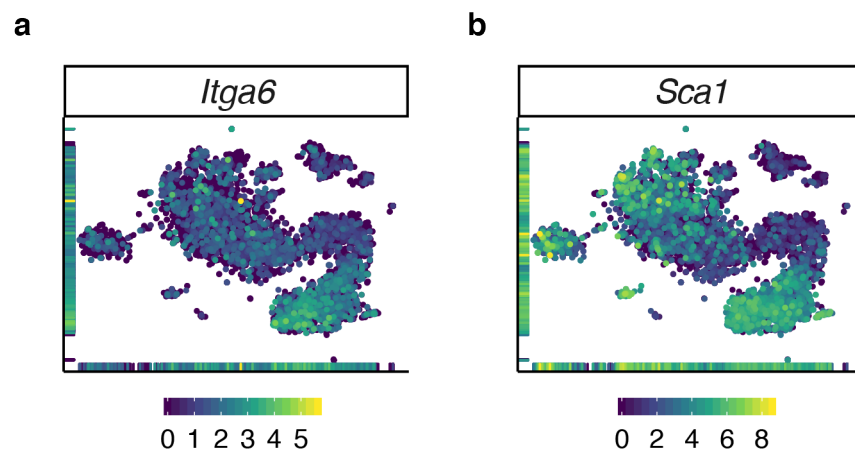


Figure 3.8 Markers consistent with a quiescent progenitor state are expressed in cluster 1

(a and b) Log₂ expression level of (a) *Itga6* and (b) *Sca1* across the dataset.

3.3.3 Seven satellite clusters lie separately from the main mTEC maturation trajectory

Seven satellite clusters surrounded the main body of mTEC (Figure 3.4a). The mTEC of **Cluster 9**, which was split into 2 sub-clusters on the tSNE visualisation (Figure 3.4a), appeared to be enriched for genes involved in several pathways, including those that deal with connexion or gap junction formation and the cellular response to stress (Figure 3.9). In addition, the mTEC of cluster 9 were a mixture of TSPAN8 and GP2 positive and negative cells (Figure 3.5a) and expressed very little *Aire* (Figure 3.4b). The transcriptome of the **cluster 10** mTEC was in keeping with a thymic tuft cell identity (294, 295). These cells highly expressed *Pou2f3* (Figure 3.10a), a transcription factor that is involved in regulating tuft cell function in the intestinal epithelium and respiratory tract, as well as its target genes involved in bitter taste sensing, including genes in the *Tas2r* family, and *Trpm5* (296, 297). Cluster 10 was also highly populated by TSPAN8+ mTEC (Figure 3.5a) and contained little-to-no *Aire* expression (Figure 3.4b). **Cluster 11** was characterised by the expression of genes related to RNA metabolism, and **clusters 12** and **15** by genes related to extracellular matrix (ECM) organisation and insulin-like growth factor (IGF) transport and binding (Figure 3.9). **Cluster 13**, which contained an over-representation of TSPAN8+ and GP2+ mTEC (Figure 3.5a), was enriched for genes involved in the response to stress and external stimuli, as well as cilium assembly (Figure 3.9). Finally, **cluster 14**, was particularly enriched for GP2+ mTEC (Figure 3.5a) that expressed low levels of *Aire* (Figure 3.4b). The mTEC of this cluster may have

a role in cell communication as they were marked by the expression of chemokine ligands, such as *Ccl6* (Figure 3.10b), *Ccl9*, and *Ccl20*, as well as the chemokine receptor *Ccr5*.

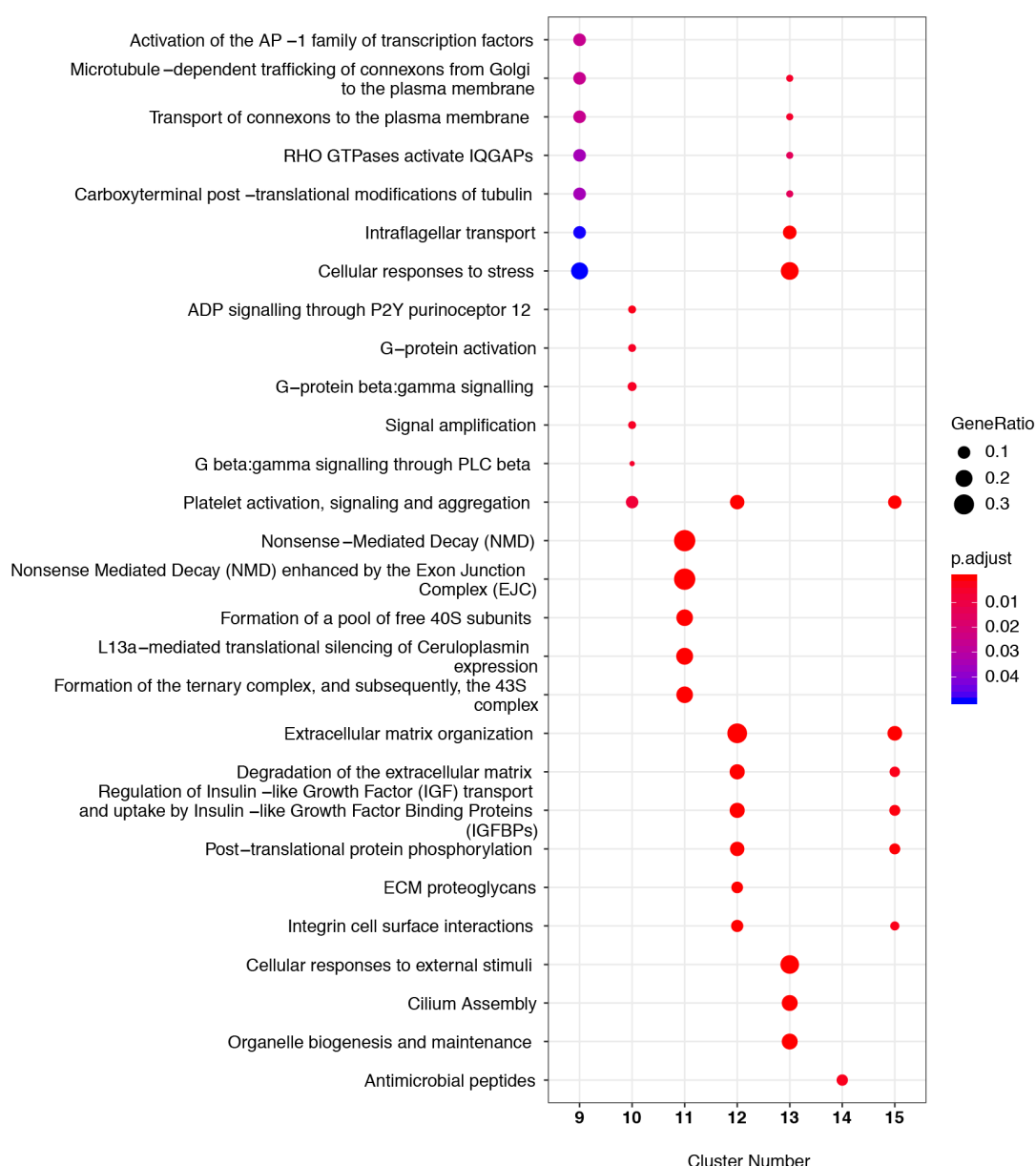


Figure 3.9 Pathway analysis for mTEC satellite clusters.

Plot showing Reactome (298) pathway analysis (y-axis) of genes expressed in each satellite cluster (x-axis). Each dot is coloured according to the adjusted p-value (p.adjust) for enrichment and its size is proportional to the gene ratio, which is the fraction of pathway genes expressed by each cluster.

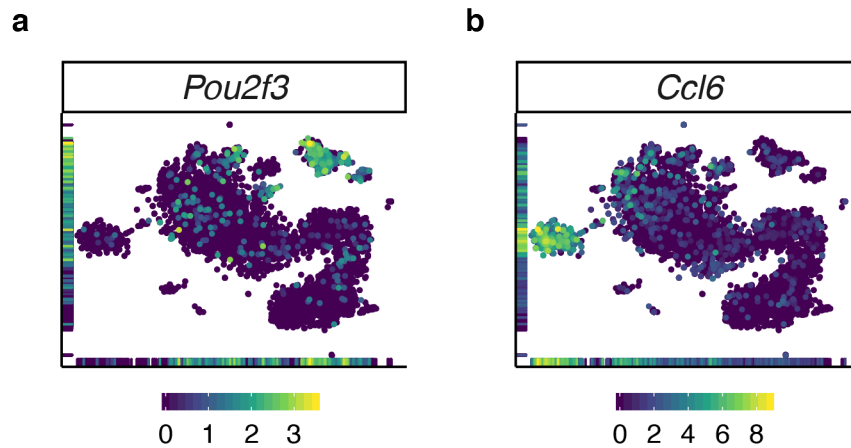


Figure 3.10 Marker genes expressed in satellite clusters 10 and 14

(a and b) Log₂ expression level of (a) *Pou2f3* (marker for cluster 10) and (b) *Ccl6* (marker for cluster 14) across the dataset.

Using FACS to enrich for mTEC expressing the AIRE-induced TRAs TSPAN8 and GP2, allowed for the investigation of a large number of rare cluster 10 and 14 cells. Almost half of the mTEC in these clusters were positive for their respective TRAs; 44% of cluster 10 cells were TSPAN8+ and 49% of cluster 14 cells were GP2+ (Figure 3.5a). Importantly, these clusters were robust to clustering of unselected mTEC alone (Figure 3.3biii). Furthermore, whilst cluster 10 appeared to represent thymic tuft cells (294, 295), cluster 14 was transcriptionally distinct and may therefore represent an as yet undescribed mTEC subpopulation, typified by the expression of chemokine ligands and receptors.

These observations argue for heterogeneous expression of TRGs across mTEC subpopulations, implying that satellite clusters show preference for

expression of particular gene subsets. This provides additional evidence against PGE being stochastic within single mTEC, and instead supports the notion that different maturational stages or classes of mTEC may activate TRG expression differentially.

3.4 Gene module clustering of single-cell RNA-sequencing data

3.4.1 The presence of 50 gene co-expression groups is supportive of PGE being an ordered process

The identification of 15 distinct mTEC clusters (Figure 3.4a) that were reproducible when only TRGs were considered in the analysis (Figure 3.5c), suggested that an ordered process exists that selects TRGs for co-expression in single mTEC. To further investigate the patterns of gene co-expression, a method of gene module clustering was applied to the single-cell RNA-seq dataset that accentuates expression similarities found among rarely observed genes and attenuates similarities found among frequently observed genes. Firstly, a TF-IDF transformation was applied in order to weight the clustering analysis away from frequently expressed genes (272). Then, the 22,819 genes detected in multiple cells were clustered using HDBSCAN in order to assign genes to single gene modules that exhibited a distinctive gene co-expression profile across the mTEC dataset (273). This analysis assigned 14,861 genes to 50 modules (Figure 3.10a); 7,958 genes, including many olfactory and vomeronasal receptor genes, could not be assigned to a module due to low detection frequency, and thus appeared as “noise” in the gene module structure (data points outside of demarcated gene modules in Figure 3.11a).

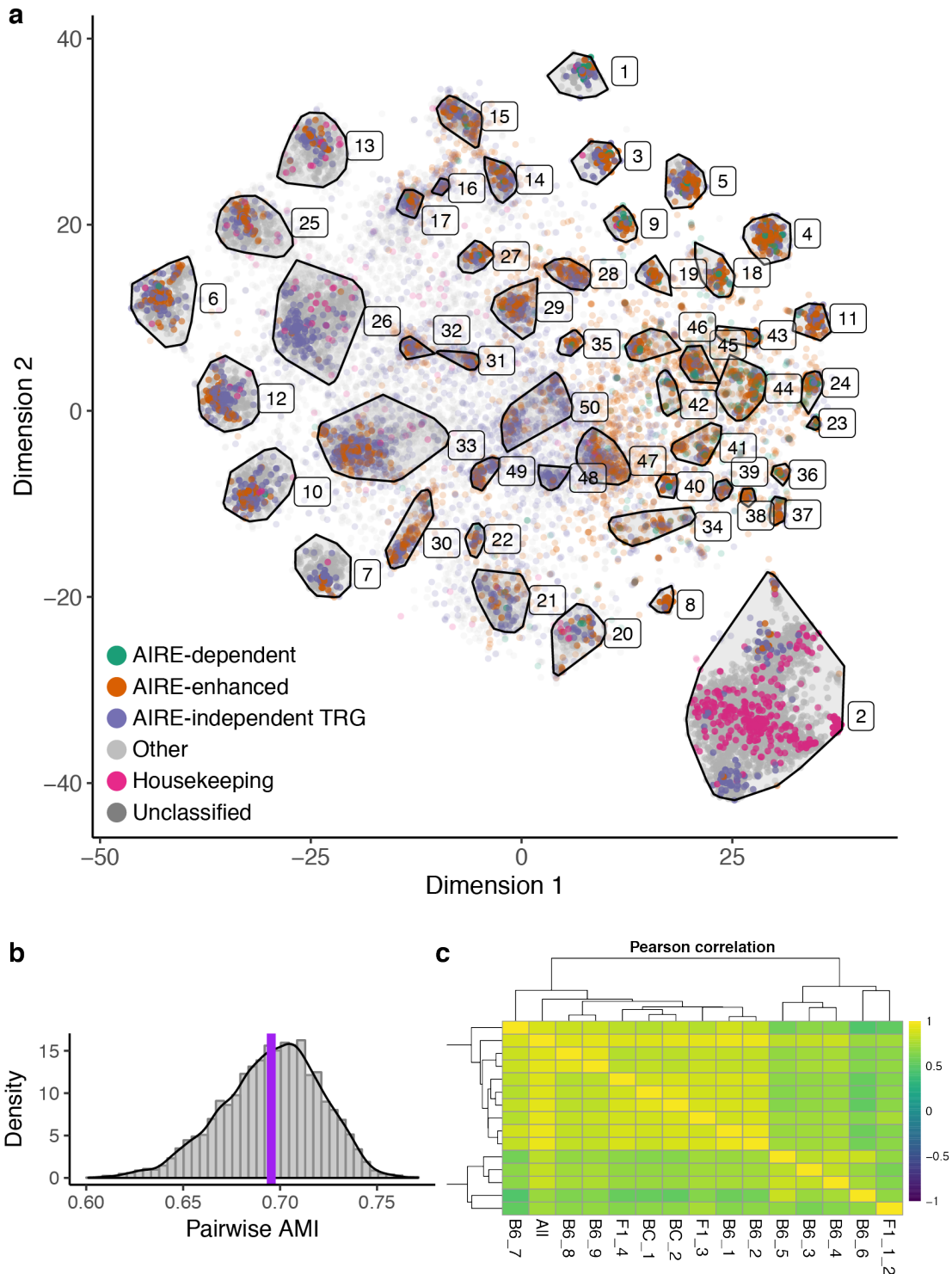


Figure 3.11 Gene module clustering reveals robustly identified gene co-expression groups indicating that TRG expression in mTEC has order

(a) tSNE visualisation of the partition of the 22,819 genes detected in multiple cells into 50 mutually exclusive gene modules. Each dot represents a gene and is coloured according category (AIRE-dependent, AIRE-enhanced, AIRE-independent TRG, Other, Housekeeping, Unclassified). The colour intensity of each dot/gene is proportional to its membership probability within each module. (b) Histogram of pairwise adjusted mutual information (AMI) between 100 random subsamples of the dataset; mean AMI value (purple line) = 0.695. (c) Pearson correlation of co-expression frequency for pairs of TRGs (AIRE-regulated and AIRE-independent) between individual mice or all mice pooled together (see Table 3.1 for mouse identifiers); mean correlation coefficient = 0.77.

The reproducibility of the 50 gene modules was assessed by re-clustering 100 random subsamples of the data. The random subsamples produced highly similar modules, with a mean pairwise adjusted mutual information (AMI) score of 0.695, where a score of 1 would denote that the compared gene modules were identical and a score of 0 that there was no overlap in gene module structure (Figure 3.11b). Whether TRG co-expression was preserved between individual mice, either of the same or different genetic backgrounds, was next assessed. For each pair of TRGs, the fraction of mTEC co-expressing both TRGs was calculated per mouse, and these fractions were then compared across all mice within the dataset. The mean Pearson correlation coefficient of these fractions across all mouse pairs was high at 0.77 indicating that the same TRG co-expression patterns were seen across individual mice, including those of different strain (Figure 3.11c).

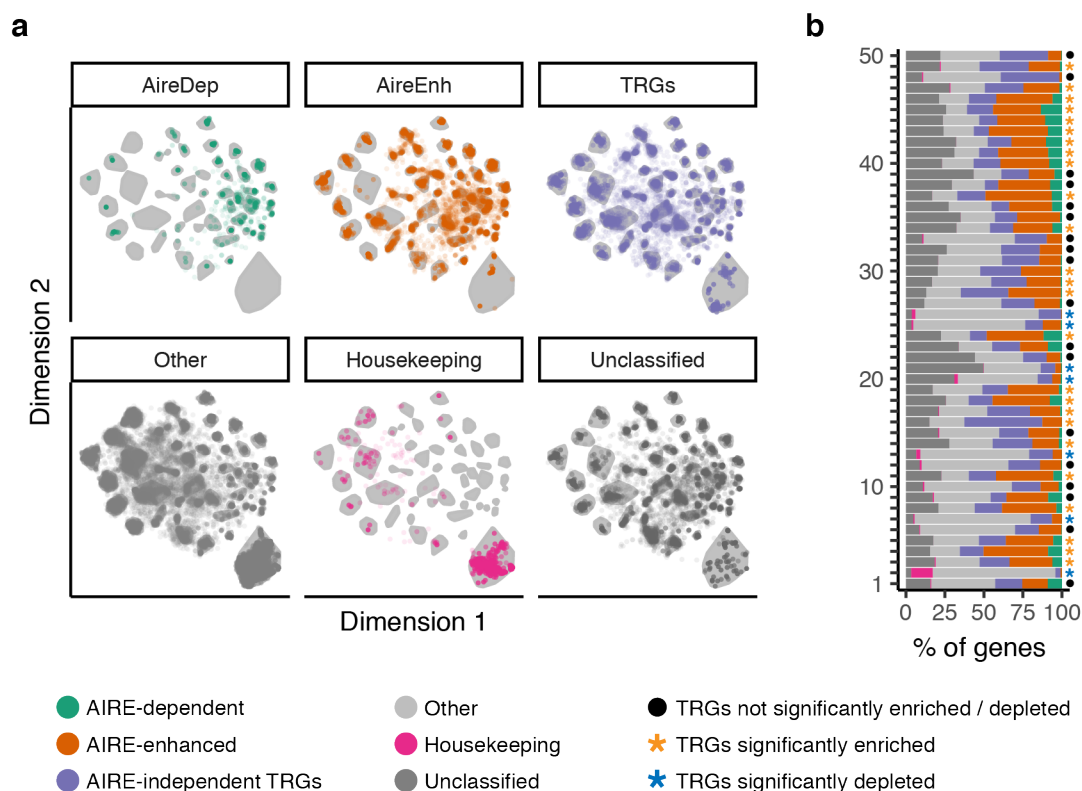


Figure 3.12 Composition of gene co-expression modules by gene category

(a) Visualisation of the composition of gene modules coloured by gene expression category (AIRE-dependent, AIRE-enhanced, AIRE-independent TRGs, Other, Housekeeping, Unclassified). (b) Composition of each gene module by gene expression category displayed as a stacked bar chart. Modules significantly enriched for TRGs (i.e. AIRE-dependent, AIRE-enhanced and AIRE-independent TRGs) are marked by an orange star; those significantly depleted for TRGs by a blue star; and those without any significant enrichment or depletion with a black dot.

All of the gene co-expression modules contained a mixture of TRGs and non-TRGs (Figure 3.12a and b). However, 25 out of the 50 gene modules were significantly enriched for TRGs, with a median TRG frequency of 53% (Figure 3.12b). By contrast, only 7 gene modules had significantly fewer TRGs than expected by chance, with a median TRG frequency of 15% (Figure 3.12b). Regarding *Gp2* and *Tspan8*, *Gp2* was a member of module 7, whose genes were most highly expressed in the post-AIRE clusters 7 and 8, and the GP2-preferred cluster 14 (Figure 3.13a). Whereas, *Tspan8* belonged to module 31,

whose genes were most highly expressed in the tuft cell-like cluster 10 (Figure 3.13b). Consequently, although both TSPAN8⁺ and GP2⁺ cells were located in some of the same cell clusters, *Tspan8* and *Gp2* were co-expressed with very different sets of TRGs.

Three of the modules that were not significantly enriched for TRGs contained genes that were frequently expressed in mTEC: module 2 contained the highest proportion of housekeeping genes (Figure 3.12b) and its member genes were detected in 45% of all mTEC; and modules 26 and 48 contained genes detected in 12% of all mTEC. Several modules contained genes highly expressed in a particular mTEC maturational state or subpopulation. For example, the genes of module 26, which included *Mki67*, cyclins and *E2f* genes, were highly expressed by the cycling mTEC of cluster 3 (Figure 3.13c). Module 48 contained genes expressed in the *Aire*-expressing mature mTEC of clusters 4 and 5 (Figure 3.13d). Modules 31 and 32 were comprised of genes that are characteristic of tuft cells and hence expressed in tuft cell-like mTEC cluster 10 (cluster 10), for example *Il10*, *Il25*, *Dclk1*, and genes regulated by the transcription factor POU2F3, such as the *Tas2r* family (Figure 3.13b and e). Module 49 contained transcripts of chemokines (e.g. *Ccl6*, *Ccl9*, *Ccl20*) and chemokine receptors (e.g. *Ccr1*, *Ccr2* and *Ccr5*) typical of cluster 14 (Figure 3.13f).

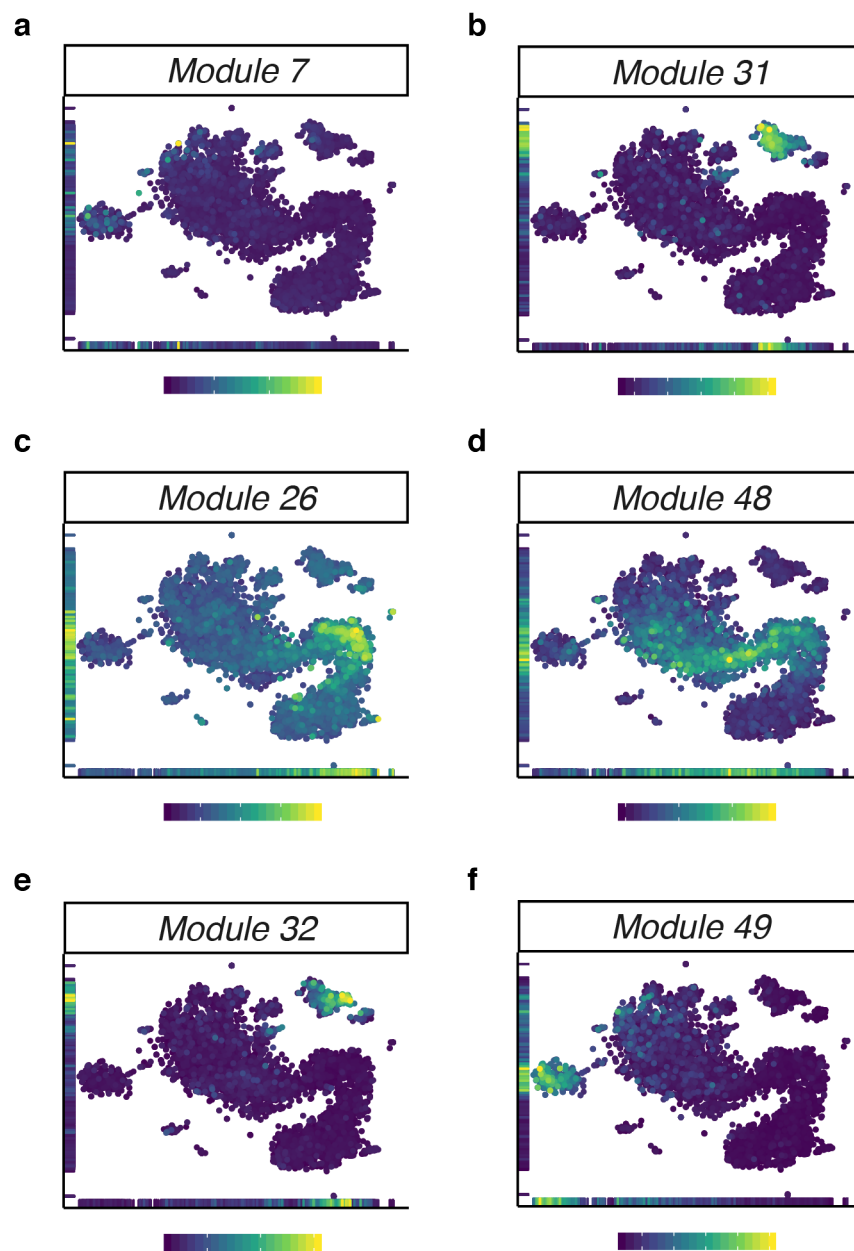


Figure 3.13 Projection of selected gene modules into mTEC clusters

(a-f) Expression of gene modules (a) 7 (b) 31, (c) 26, (d) 48, (e) 32 and (f) 49 across the dataset.

In summary, the results presented in this section demonstrate that complex gene co-expression programmes are replicated across many mTEC and are driven by TRGs as well as non-TRGs. However, the significant enrichment of TRGs in half of the gene modules implies that TRG co-expression substantially defines these modules. Furthermore, TRG co-expression is preserved across individual mice including those of different strains. Taken together, these observations argue against a stochastic mechanism for TRG expression within single mTEC and rather indicate that PGE involves ordered co-expression of TRGs.

3.4.2 Gene membership to co-expression groups is biologically indeterminate

Next, whether the observed order of TRG co-expression within single mTEC could be explained by chromosomal location, intergenic distance, tissue-, pathology- or pathway-restricted expression was considered.

The co-expression pattern identified in the vast majority of the gene modules (46 out of 50) was independent of expression by a single chromosome (Figure 3.14a). Only 4 of the 50 gene modules contained an over-representation of genes expressed from a particular chromosome (Figure 3.14a). Genes located on chromosome 3 were enriched in modules 33 and 44, and genes on chromosomes 6 and 7 were enriched in modules 32 and 35, respectively. Hence, expression by a single chromosome was not the mechanism for gene co-expression within single mTEC.

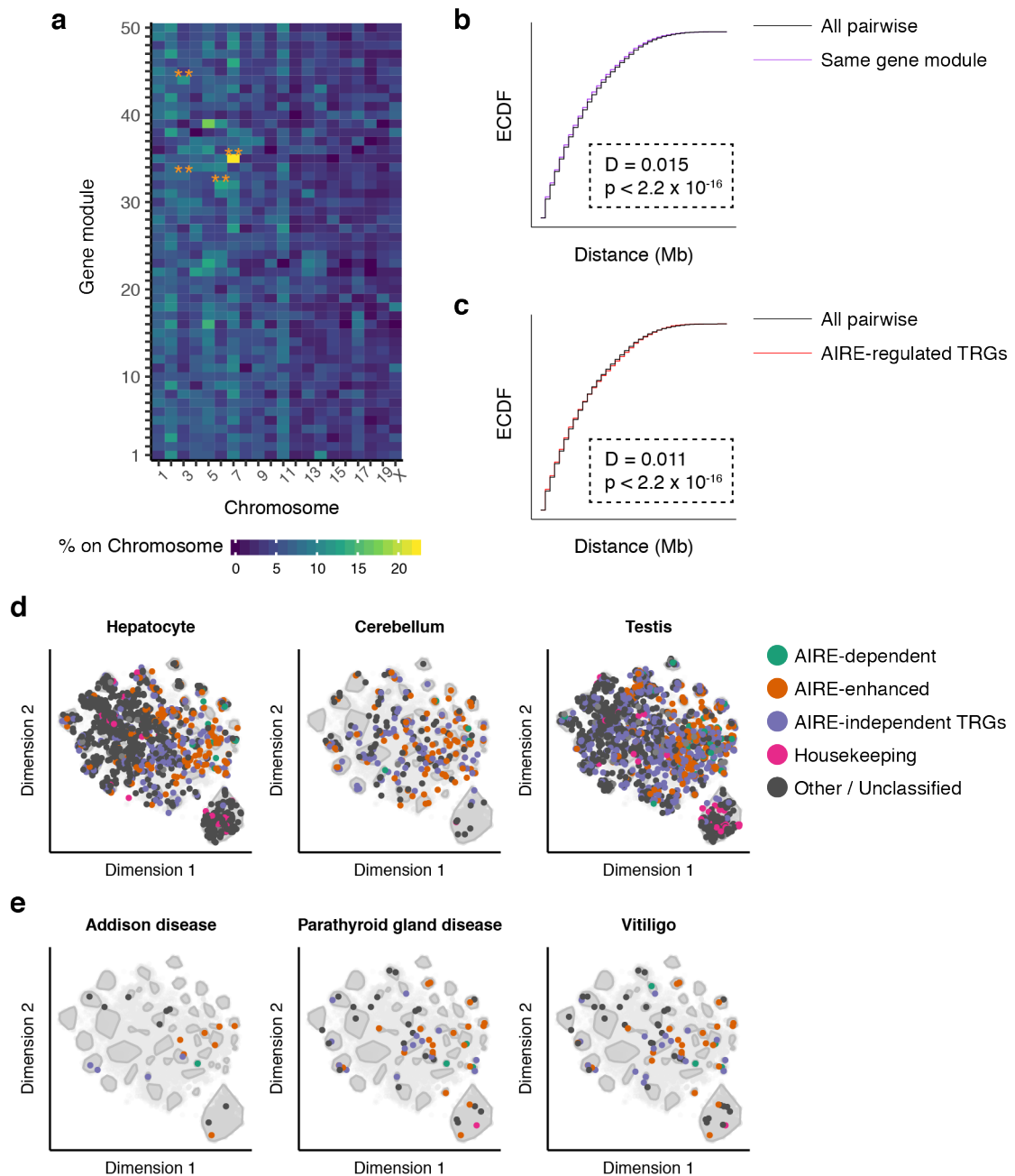


Figure 3.14 Gene membership to co-expression groups is biologically indeterminate

(a) Chromosomal locations for genes expressed within each gene module. ** Adjusted p-value < 0.01 (in orange) indicates that more genes assigned to a given module (y-axis) are located on a particular chromosome (x-axis) than expected by chance. (b and c) Empirical cumulative distribution function (ECDF) for the distribution of pairwise genomic distances between co-expressed genes. All pairwise distances (black line) are compared to the distances between genes within the same gene module (purple line) in (b) and all pairwise distances (black line) are compared to the distances between AIRE-regulated genes within the same gene module (red line) in (c). D is a measure of difference between the compared distributions, where 0 denotes that the distributions are equal and 1 denotes that the distributions compared are maximally separated. (d) Distribution of hepatocyte- cerebellar- and testes-specific genes across the gene modules; genes are depicted as a coloured dot that denotes their gene category. (g) Distribution of genes encoding auto-antigens associated with parathyroid gland disease, Addison's disease and vitiligo across the gene modules; auto-antigen genes are depicted a coloured dot that denotes their gene category.

Co-expressed genes were not clustered within the linear genome. The chromosomal distances between genes present within the same gene module were near identical to those of all other genes regardless of AIRE-dependency (Figure 3.14b and c). Finally, individual modules were not biased towards expression of TRGs belonging to individual peripheral tissues (Figure 3.14d), antigenic targets of individual organ-specific autoimmune diseases (Figure 3.14e), or genes belonging to a particular cellular pathway of Gene Ontology term (data not shown).

Together these results indicate that PGE in mTEC is a biologically indeterminate yet highly ordered process, whose order is provided by repeated co-expression of particular gene subsets. Co-expressed genes are not co-located in the linear genome and are not linked by tissue specificity, nor by biological or disease processes.

3.4.3 Random spatial distribution of GP2+ mTEC within medullary islands

Finally, the spatial distribution of mTEC expressing a given TRA was analysed. For this, thymic sections were stained for cytokeratin 5 (K5) to mark mTEC and GP2 as the TRA (Figure 3.15a). Then the cumulative distribution function of the distances (G-function) between nearest neighbour GP2+ mTEC was computed and compared to those for simulated mTEC that are evenly spaced, dispersed randomly or clustered in their distribution within medullary islands (Figure 3.15b and c). This showed that GP2+ mTEC were spatially distributed within the medulla in a manner consistent with a completely random process.

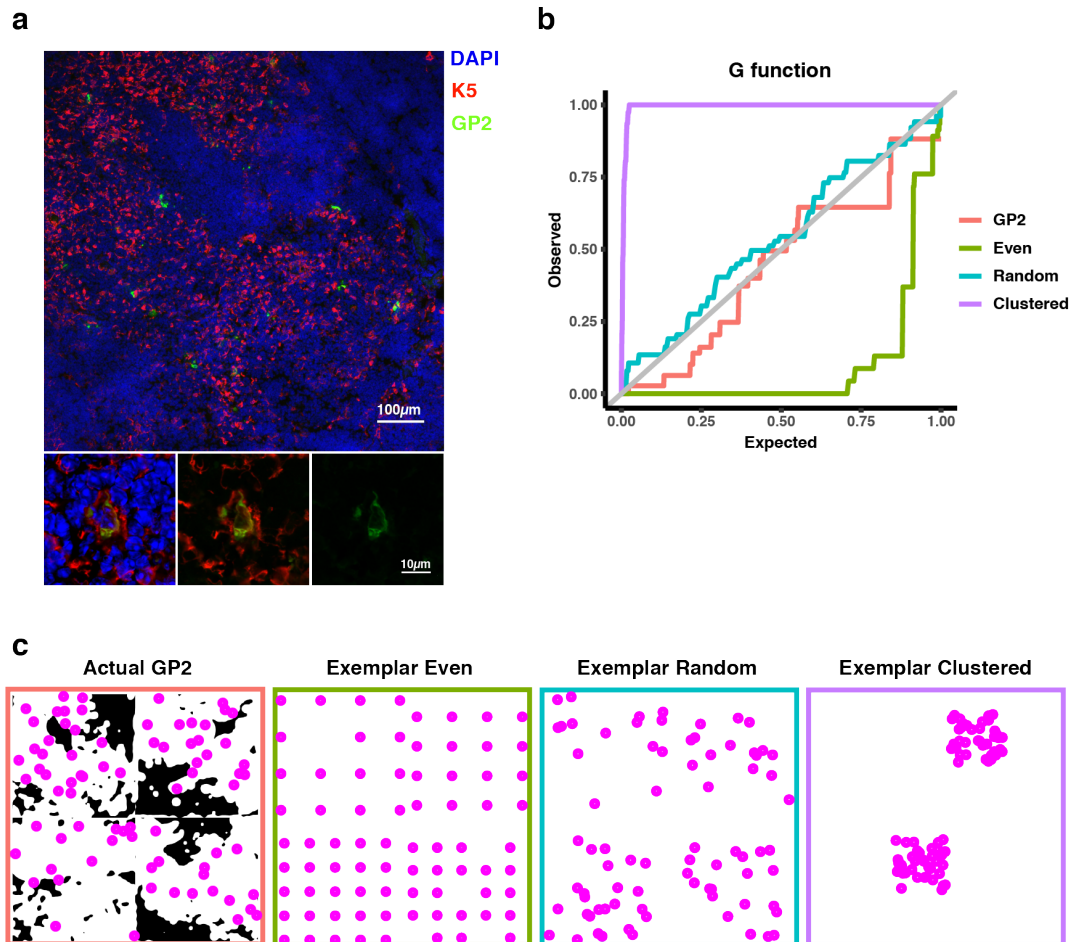


Figure 3.15 GP2+ mTEC are randomly distributed within medullary islands.

(a) Representative immunofluorescent microscopy image of a murine thymic section stained for cytokeratin 5 (K5: red) to mark mTEC, and the AIRE-enhanced TRA, GP2 (green); nuclei are stained with DAPI (blue). Zoomed images (lower panels) show the pattern of staining for a GP2+ mTEC. (b) Observed vs. expected G-function for actual GP2 spacing (red) compared with exemplar spacings. (c) Medullary regions (white) defined by K5 expression for four slides stained as in (a) with single GP2+ mTEC as magenta circles (left most panel). Exemplar distributions (3 right most panels) for comparison coloured as in (b).

3.5 Discussion

The molecular processes underlying PGE in single mTEC have remained unclear despite numerous studies addressing this topic (111, 127, 129, 133, 178, 211, 212). These studies, so far limited by the methods used and the cell numbers investigated, have not conclusively established whether TRG expression in single mTEC is a stochastic or ordered process. Amongst the studies that failed to detect co-expression patterns, two used a single-cell PCR approach limited to the detection of a very restricted set of TRGs and probed only a small number of cells (178, 211). Co-expression patterns were detectable in three studies that either preselected mTEC for the expression of cell surface-expressed TRAs or conducted single-cell RNA-sequencing of hundreds of single mTEC (127, 129, 133). Two of these studies additionally found that mTEC express a larger repertoire of TRGs as they mature and that single mTEC are not restricted to expressing TRGs belonging to particular peripheral tissues, biological pathways or controlled by the same transcription factors (129, 133).

The current study addresses the transcriptional patterns associated with PGE by examining the transcriptomes of thousands of single mTEC. This, therefore, represents the largest single-cell RNA-seq dataset investigating PGE in mTEC and its analysis revealed differentiating gene co-expression patterns that were maintained not only amongst individual mice of the same genetic background, but also in different mouse strains. In aggregate, these results strongly suggest that an ordered process regulates PGE in mTEC.

Cell-based clustering of the data was in support of previous single-cell RNA-seq studies demonstrating heterogeneity within the mTEC compartment (294, 299). This notion is corroborated by studies tracking TEC differentiation during organogenesis and regeneration. These suggest that mTEC derive from lineage committed progenitors that can be identified by the expression of CLD3/4, SSEA-1, and/or CD24 within the fetal thymus (23, 24, 53), and postnatally by the expression of podoplanin and CCL21 and anatomical localisation at the cortico-medullary junction (54). mTEC subsequently pass through several successive stages of maturation (18, 55) in which immature $\text{MHCII}^{\text{Lo}}\text{CD80}^{\text{Lo}}\text{CD86}^-\text{AIRE}^-$ mTEC transition, via intermediate stages, to an $\text{MHCII}^{\text{Hi}}\text{CD80}^{\text{Hi}}\text{CD86}^+\text{AIRE}^+$ state, which represents the most functionally mature mTEC from the perspective of PGE and central tolerance induction (55, 56). Finally, mTEC downregulate the expression of AIRE, MHCII and costimulatory molecules to enter a terminally differentiated post-AIRE state (55, 58-61).

In keeping with the above model of mTEC maturation, two clusters (clusters 1 and 2) were identified that displayed progenitor-like characteristics because they expressed genes associated with mTEC progenitor status such as *Ccl21a*, *Pdpr*, *Itga6* and *Sca1* but lacked *Aire*, *Cd80* and *Cd86* (49, 50, 54). The temporal ordering inferred from RNA velocity analysis suggested that cluster 2 were pre-AIRE immature mTEC, which transitioned to an *Aire*⁺*Cd80*⁺*Cd86*⁺ (clusters 5 and 6) phenotype through an actively cycling stage (cluster 3), and then to a post-AIRE stage (clusters 7 and 8) with low

Aire, *Cd80*, and *Cd86* and high *Krt10*, *Iti* and *Spink5* expression (55, 59, 60, 62, 291, 292).

Several of the mTEC clusters identified correlated with those identified in a recently published single-cell study by Bornstein *et al* that defined four distinct mTEC types. Clusters 1 and 2 from the current study relate to the immature pre-AIRE type I mTEC, clusters 3-6 to the mature AIRE⁺ type II mTEC, clusters 7 and 8 to type III mTEC, and cluster 10 to the tuft-like type IV mTEC identified by Bornstein *et al* (294). In addition, six novel mTEC clusters were identified, namely clusters 9, 11, 12, 13, 14 and 15. Cluster 14 was defined by high expression of chemokine ligands *Ccl6*, *Ccl9* and *Ccl20*, as well as the chemokine receptor *Ccr5*, implying a role in cell communication. This cluster was also highly populated by GP2⁺ mTEC, suggesting that the strategy of enriching for rare TRA positive mTEC using FACS sorting has revealed otherwise hidden subpopulations, and thus that analyses of mTEC expressing other TRAs would likely uncover additional satellite clusters, each with a distinct transcriptome. Cluster 9, expressed the markers *Ceacam10*, *Cd177* and *Ckm* and shared transcriptomic similarities not only with the terminally differentiated mTEC of cluster 8 but also with tuft-like mTEC of cluster 10. A third novel cluster, cluster 13, highly expressed several genes associated with cilium assembly (e.g. *Spag16*, *Wdr34* and *Bbs7*). The remaining novel clusters (11, 12 and 15) were defined by very few cells (0.3 - 1% of the collected mTEC) and therefore were difficult to characterise.

Previously published single-cell studies have been limited by low cell number; this is likely to have under-powered analyses of gene co-expression due to the low frequencies at which individual TRGs are expressed in single mTEC (111, 127, 129). Therefore, these previous studies were either unable to detect co-expression (111, 178, 211), or were able only to detect a handful of co-expression groups (127, 129, 133). The inclusion of nearly 7,000 single mTEC in the current study allowed for the detection 50 separate gene co-expression modules, the majority of which comprised TRG co-expression sets, suggesting that PGE in single mTEC follows an ordered process. Approximately one-third of the genes expressed in multiple mTEC could not be assigned to a co-expression group. Although this could reflect biological noise, it is possible that, with the inclusion of more single cells, these genes might also become assignable to specific gene modules.

A previous study that identified co-expression patterns of AIRE-induced genes, found that these differed between individual mice and suggested that gene co-expression networks were established stochastically in each mouse (129). This study investigated only 200 mTEC from two pairs of WT and *Aire* knockout mice, raising a concern that it was significantly underpowered (129). In contrast, the gene co-expression modules identified in the current study were robust as evidenced by their reproducibility across random samples of the dataset, as well as between individual mice, independent of strain. Such reproducible TRG co-expression is likely to reflect an inherent cellular property that introduces bias in which TRGs are expressed together within a cell. That

being said, it was not possible to explain membership to co-expression groups by any of the biological features examined, including gene category, chromosomal location, genomic distance, tissue specificity, autoimmune disease association, and inclusion in a specific biological pathway. The largely random distribution of genes within the same co-expression module across the genome was in keeping with observations from several previous studies (129, 133, 299). Previously published studies have also similarly concluded that PGE patterns in mTEC are dictated neither by expression patterns seen in differentiated peripheral cell types nor by co-regulation by specific transcription factors (129, 133). Whilst it was not possible to explain membership to co-expression groups by the various analyses performed, this does not preclude the existence of an underlying mechanism that explains why certain TRGs are co-expressed in single mTEC, particularly since inter-individual reproducibility was robust. Certainly there was evidence that different maturational stages or classes of mTEC activate TRG expression differentially. Furthermore, although co-expressed genes were not related by their location within the linear genome, the analyses performed do not exclude the possibility that TRGs are physically co-located on chromatin by virtue of chromatin looping, as has been suggested (133, 198).

The biologically indeterminate yet ordered process for TRG co-expression described here suggests a program that repeatedly generates mTEC with co-expressed genes that are randomly sampled with respect to disease-relevant antigens, pathways, tissues, and chromosomes. The spatial analysis for GP2+

mTEC showed that GP2 expression is randomly dispersed across the thymic medulla. This random spatial pattern of TRA presentation would provide a developing thymocyte travelling through a medullary island with the highest likelihood of encountering an mTEC expressing a given TRA against which its antigen receptor could be tested. This implies that thymocytes would only need to traverse a limited volume within the thymic medulla in order to be tested against a diverse range of TRAs.

In conclusion, reproducible order is evident among the TRGs that single mTEC express. However, the selection of the TRGs co-expressed within single mTEC is indeterminate with respect to biological processes. This degree of randomness may ensure that single TEC express a wide range of TRAs, covering diverse peripheral tissues and auto-antigens and that expression of a given TRA is spatially dispersed throughout the thymic medulla. In this way, a single thymocyte travelling through the thymic medulla may be given the greatest opportunity of encountering any given self-antigen for the purposes of central tolerance induction.

4 CHD4 in TEC Development, Function and AIRE-induced Gene Expression

4.1 Introduction, aims and hypothesis

The precise molecular mechanisms by which the transcriptional facilitator AIRE operates in TEC to enable PGE are incompletely understood (111, 139). Several studies have found AIRE to exist in the context of a large multi-protein complex affecting chromatin binding and remodelling, as well as DNA repair (181-183). The AIRE complex may localise to target genes by recognising repressive epigenetic marks such as H3K4me0, H3K27me3, and H3K9me3 (111, 199) and then remodel chromatin to allow transcription of silenced genes (111, 184). However, under in vitro experimental conditions, AIRE binds only to H3K4me0 with sufficient affinity (192). It is therefore possible that AIRE-interacting proteins recruit the complex to specific chromatin marks (111). Chromodomain-helicase-DNA-binding protein 4 (CHD4) is such a candidate molecule. CHD4 has previously been reported to physically associate with chromatin bound AIRE (182). CHD4 interacts with PRC2 to deposit H3K27me3 (227, 245) and is able to bind to H3K4me0 and H3K9me3, as well as directly to DNA (233-237). Furthermore, CHD4 is able to influence transcriptional regulation via its ability to mobilise nucleosomes and, in addition, participates in DNA repair (213, 230), a molecular process known to contribute to PGE (J. Abramson, personal communication).

The aim for this part of my studies was to determine the role of CHD4 in TEC development and function, with a particular focus on AIRE-regulated PGE, with the hypothesis that CHD4 interacts with AIRE to influence the transcription of AIRE-regulated genes, and that this may occur via targeting of the AIRE complex to TRG loci marked by repressive histone marks and locally remodelling chromatin. In order to address this I used a gene targeted mouse model in which *Chd4* is conditionally deleted in all TEC. I subsequently employed a variety of cellular and molecular tools to probe the effects of TEC conditional *Chd4* deletion on the composition of the TEC compartment, the ability of TEC to support regular thymocyte development and selection, to effect PGE, and thus to establish central T-cell tolerance induction.

4.2 Generation of mice with CHD4 deficiency targeted to TEC

CHD4 is ubiquitously expressed and early embryonic lethality has been demonstrated in mice constitutively deficient for this protein (300). To investigate the role of CHD4 in TEC biology, particularly in the process of AIRE-regulated PGE, mice were therefore created that have a conditional deletion of *Chd4* in all TEC (Figure 4.1). This genetic modification was achieved by crossing mice in which exons 12-21, encoding the ATPase domain of CHD4, are flanked by LoxP sites (*Chd4^{fl/fl}*) (225), with a transgenic strain expressing the Cre recombinase enzyme under the control of *Foxn1* (Figure 4.1a) (261). It has previously been shown that the expression of Cre recombinase under the control of this *Foxn1* transgene leads to the deletion of floxed target genes in all cells expressing *Foxn1*, namely the epithelial cells of

the thymus and skin (261). In addition, the presence of a single copy of the *Foxn1Cre* transgene in the absence of a floxed target gene has been demonstrated not to affect normal thymus development and thymocyte differentiation (261). Both mice heterozygous ($Chd4^{fl/+}::Foxn1Cre$) and homozygous for the loss of *Chd4* targeted to TEC ($Chd4^{fl/fl}::Foxn1Cre$; designated “cKO”) were generated and compared to controls with either one ($Chd4^{fl/+}$) or both *Chd4* loci harbouring LoxP sites ($Chd4^{fl/fl}$; designated “WT”).

The cKO mouse model was validated at the genomic, transcriptional and protein level (Figure 4.1 b, c and e). Correct genomic excision of *Chd4* in FACS sorted primary TEC subpopulations from cKO mice was confirmed by two PCR reactions (Figure 4.1b; and Supplementary Figure 1 for FACS gating strategy). CD45⁺ thymocytes were included as a control as these cells do not express *Foxn1* and should therefore lack Cre mediated recombination of the *Chd4* locus. The first analysis (“PCR Deleted Region”) used primers specific for a region between intron 14 and Exon 15 of *Chd4*, which lies between the LoxP sites of the targeted *Chd4* allele. Accordingly, this reaction only amplified a product in WT TEC and in CD45⁺ thymocytes from both WT and cKO mice but not in any of the TEC subpopulations isolated and analysed from cKO thymi (Figure 4.1b, upper panel). The second analysis (“PCR Delta Allele”) used a primer pair that only amplifies a product in samples where exons 12-21 of *Chd4* are deleted (225). As expected, a product was only detected in cKO TEC and not in WT TEC nor in CD45⁺ thymocytes from either genotype (Figure 4.1b, lower panel).

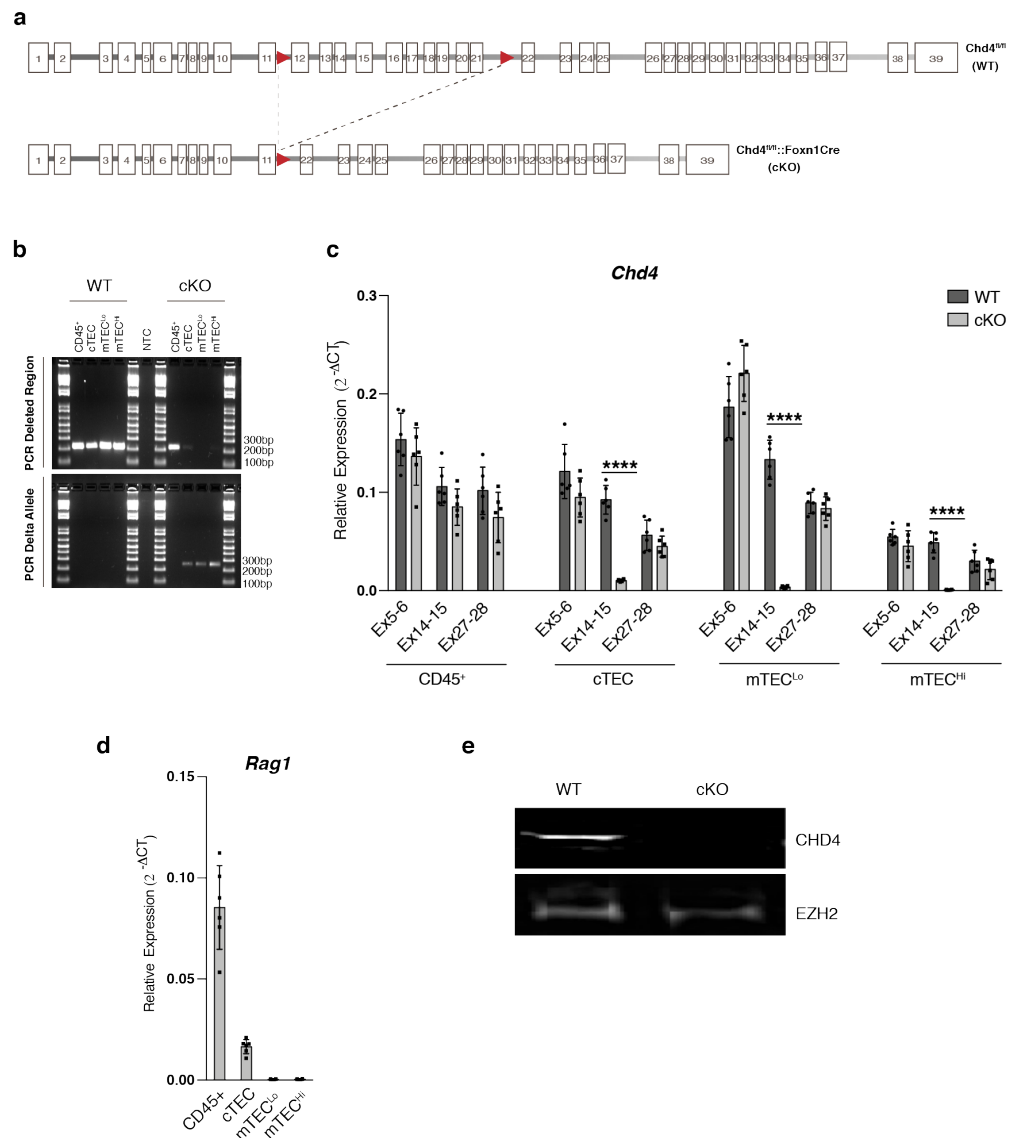


Figure 4.1 Conditional deletion of *Chd4* in TEC

(a) Schematic representation of conditional deletion strategy. Mice homozygous for a *Chd4* allele (*Chd4*^{fl/fl} or WT), where exons 12-21 are flanked by LoxP sites (red triangles), were crossed with mice expressing Foxn1Cre in order to create mice in which the floxed *Chd4* exons are conditionally excised in all TEC (*Chd4*^{fl/fl}::Foxn1Cre or cKO). (b) Genomic excision PCRs performed on FACS sorted thymocytes (CD45⁺), cTEC (CD45⁺EpCAM⁺Ly51⁺UEA-1⁻), mTEC^{Lo} (CD45⁺EpCAM⁺Ly51⁺UEA-1⁺MHCII^{Lo}CD80^{Lo}) and mTEC^{Hi} (CD45⁺EpCAM⁺Ly51⁺UEA-1⁺MHCII^{Hi}CD80^{Hi}) from WT and cKO thymi. Upper panel shows an agarose gel from a PCR using primers that amplify a product within the deleted region. Lower panel shows an agarose gel from a PCR using primers that only amplify the allele after Cre-mediated recombination (delta allele). Results are representative of 3 independent experiments. (c) RT-qPCR data showing relative expression of *Chd4* exons before (Ex5-6), within (Ex14-15), and after (Ex27-28) the deleted region in FACS sorted thymocytes (CD45⁺) and TEC subpopulations from WT and cKO thymi. Expression levels have been normalised to *Gapdh*. Graph contains data pooled from 2 independent experiments with 3 replicates each. (d) RT-qPCR data showing relative expression of *Rag1* in FACS sorted thymocytes (CD45⁺) and TEC subpopulations from cKO thymi. Expression levels have been normalised to *Gapdh*. Graph contains data pooled from 2 independent experiments with 3 replicates each. (e) Western blot for CHD4 (upper panel) using nuclear extracts generated from FACS sorted TEC (CD45⁺EpCAM⁺) from WT and cKO thymi. Detection of EZH2 (lower panel) was used as a control for protein loading. All experiments were performed using 4 week old mice.

Appropriate deletion of the floxed region of *Chd4* in cKO TEC was also confirmed at the transcript level (Figure 4.1c). RT-qPCR was performed using primers that bind proximal to the floxed region (Ex5-6), within the floxed region (Ex14-15) and distal to the floxed region (Ex27-28). Their combined use confirmed a lack of expression of the *Chd4* exons within the floxed region (Ex14-15) in mTEC^{Lo} and mTEC^{Hi} from cKO mice. A very low level of expression of Ex14-15 was detected in cKO cTEC which is most likely explained by the inclusion of thymic nurse cells within the cTEC sample as the isolation procedure typically results in the inclusion of these cells which harbour CD45⁺ thymocytes as a “contaminant” (301, 302). In keeping with this notion, *Rag1*, a gene whose expression is limited to lymphocytes during the process of antigen receptor somatic recombination, was found to be expressed in FACS sorted CD45⁺ cells and the cTEC fraction used here for analysis, but not in mTEC^{Lo} and mTEC^{Hi} from cKO mice (Figure 4.1c). Exons proximal and distal to the deleted region were expressed at comparable levels in TEC from cKO and WT mice (Figure 4.1c). Finally, western blot analysis for CHD4 protein expression in primary TEC using a polyclonal antibody raised against a peptide proximal to the deleted region failed to detect expression in cKO TEC (Figure 4.1e). This result not only verified the lack of full length CHD4 but also excluded the generation of a truncated form of the protein with a possible partial function (not shown). The expression of RNA sequences downstream of the deleted region but lack of a truncated protein are in keeping with findings in a previous study using the same mouse strain to conditionally delete *Chd4* in

thymocytes (225). Taken together the results from these experiments confirm conditional deletion of *Chd4* in TEC from cKO mice.

4.3 Heterozygous targeting of *Chd4* in TEC is insufficient to cause a phenotype

To assess whether heterozygous deletion of *Chd4* in TEC is sufficient to cause defective TEC development and/or function, thymi from 4 week old heterozygous conditional knockout mice (*Chd4*^{fl/+}::Foxn1Cre) were compared to littermate controls (*Chd4*^{fl/+}) with respect to cellularity, TEC phenotype and thymocyte maturation (Figures 4.2 and 4.3).

Total thymic cellularity was normal in mice with heterozygous TEC targeted deletion of *Chd4* (Figure 4.2a), as was the composition of the TEC compartment when probed for the relative frequency of cTEC and mTEC (Figure 4.2b), and that of distinct mTEC subpopulations as defined by the expression of MHCII, CD80, CD86 and AIRE (Figures 4.2c, d and e, and Supplementary Figure 1 for basic TEC gating strategy).

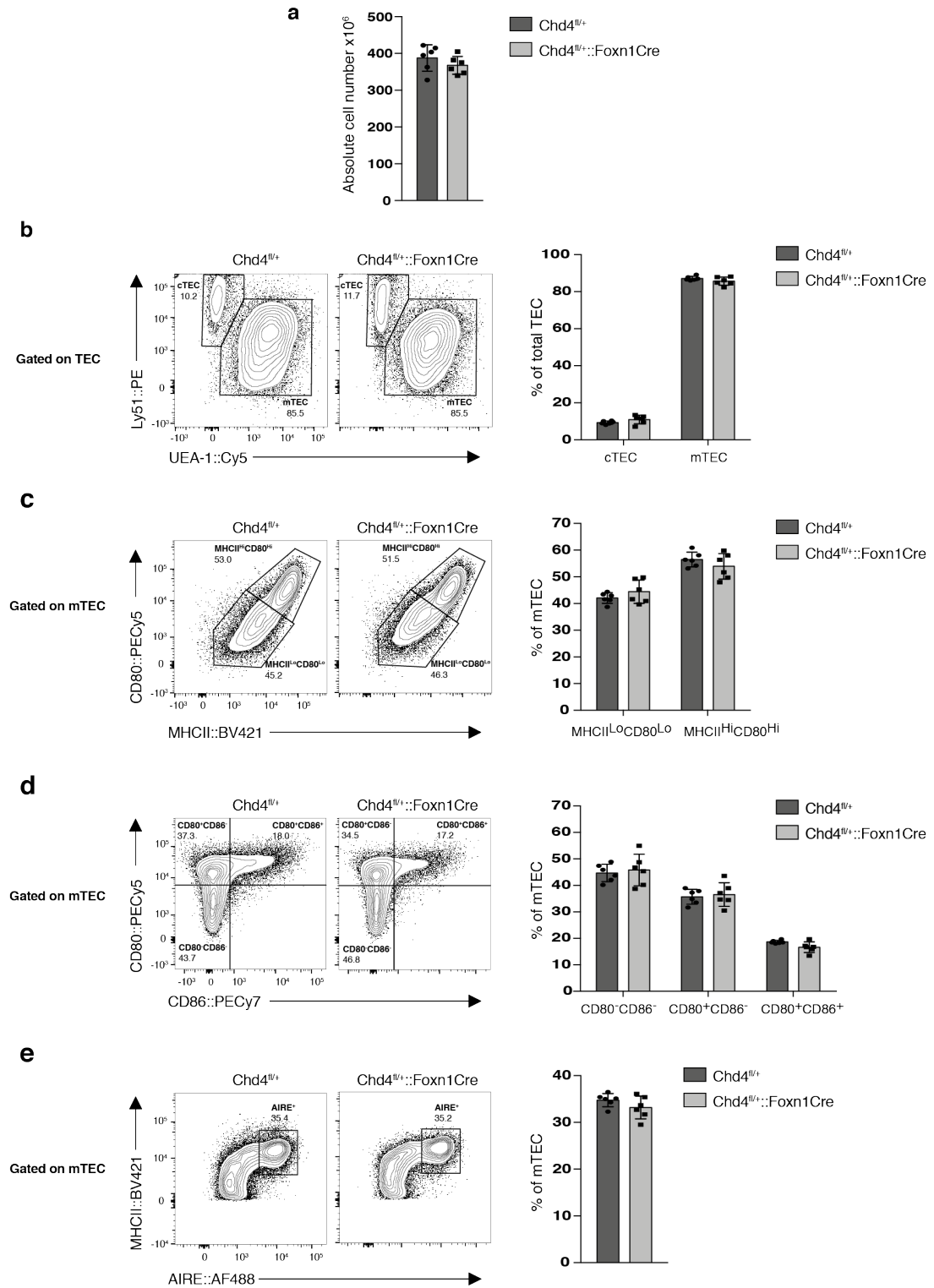


Figure 4.2 Normal thymic cellularity and TEC phenotype in mice with heterozygous TEC targeted deletion of *Chd4*

(a) Total cellularity after enzymatic digestion of thymi from mice with heterozygous TEC targeted deletion of *Chd4* (Chd4^{fl/+}::Foxn1Cre) compared to controls (Chd4^{fl/+}). (b-e) Representative FACS plots (left panels) and bar graphs (right panels) showing relative frequencies of (b) cTEC (Ly51⁺UEA-1⁻) and mTEC (Ly51⁺UEA-1⁺) within total TEC (CD45⁺EpCAM⁺), (c) MHCII^{Lo}CD80^{Lo} and MHCII^{Hi}CD80^{Hi} mTEC, (d) CD80⁺CD86⁻, CD80⁺CD86⁺ and CD80⁻CD86⁺ mTEC, and (e) AIRE⁺ mTEC in thymi from 4 week old mice. All bar graphs contain data pooled from 2 independent experiments with 3 replicates each.

Consistent with a normal TEC compartment, a basic assessment of thymocyte development also yielded normal findings in mice with heterozygous deletion of *Chd4* targeted to TEC (Figures 4.3 and Supplementary Figure 2 for basic thymocyte maturation gating strategy). The first stages of thymocyte maturation are defined by the lack of expression of markers typical for $\alpha\beta$ T-cells such as CD4, CD8 and/or TCR β . To accurately assess the earliest stages of thymocyte maturation, intrathymic cells not belonging to the $\alpha\beta$ lineage were excluded by their expression of lineage markers for myeloid cells (CD11c, F4/80, CD11b, Gr1), B lymphoid cells (CD19), NK and NKT cells (CD49b and NK1.1), $\gamma\delta$ T-cells ($\gamma\delta$ TCR) and erythroid cells (Ter119). Analysis of these intrathymic “lineage-negative” cells demonstrated that thymocyte maturation progressed normally through immature DN (i.e. CD4⁻CD8⁻), to DP (CD4⁺CD8⁺) and mature SP (CD4⁺CD8⁻ or CD4⁻CD8⁺) differentiation stages (Figure 4.3a and b). In addition, the ability of developing thymocytes to upregulate CD69 and TCR β expression in the context of heterozygous TEC conditional *Chd4* deletion was also normal, thus indicating regular positive selection in these mutant mice (Figure 4.3c) (303-305).

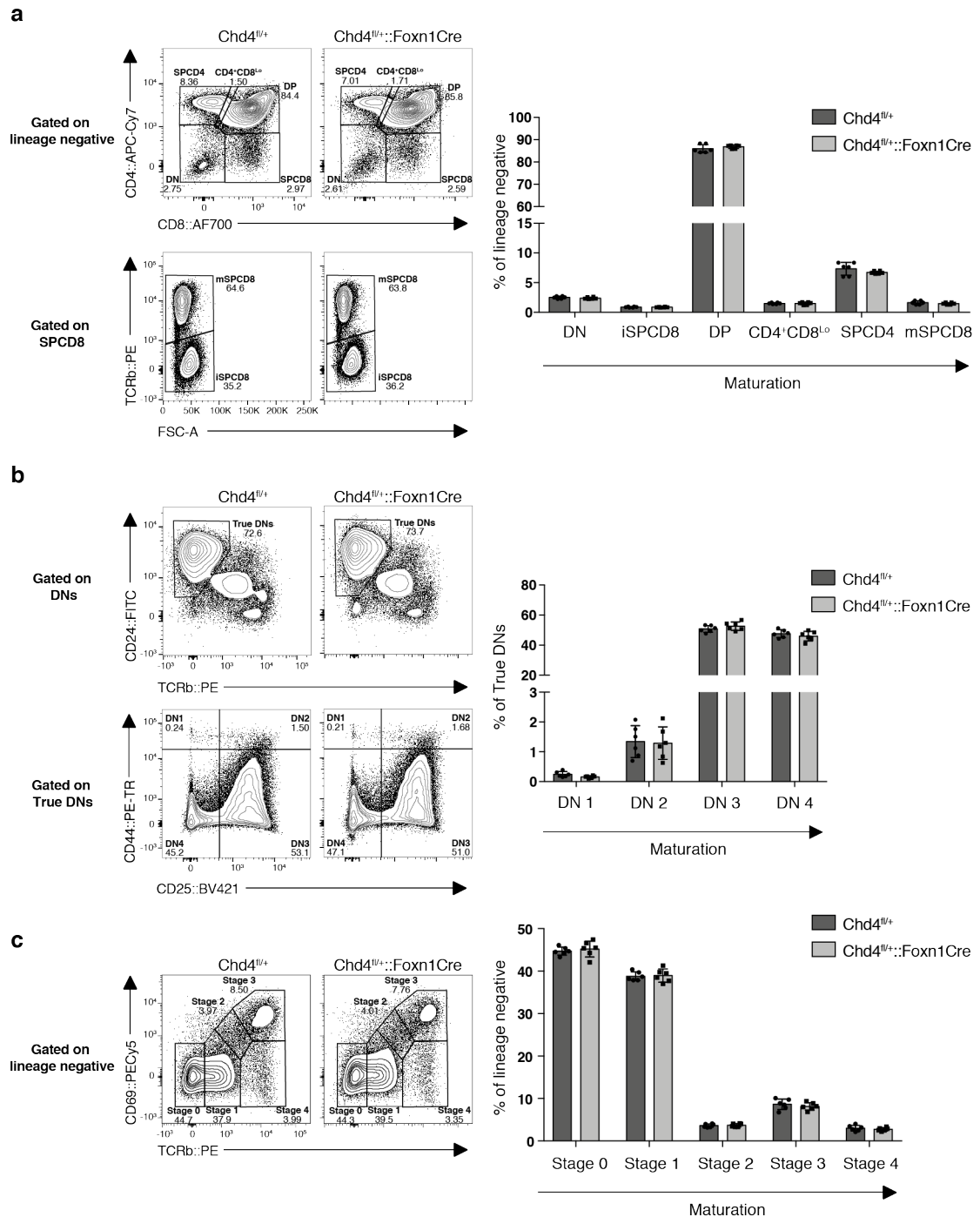


Figure 4.3 Normal thymocyte maturation in mice with heterozygous TEC targeted deletion of *Chd4*
(a-c) Representative FACS plots (left panels) and bar graphs (right panels) showing relative frequencies of (a) thymocyte maturation stages defined by CD4 and CD8 expression, progressing from double negative (DN; CD4⁻CD8⁻), to immature single positive CD8 (iSPCD8; CD4⁻CD8⁺TCRb⁻), double positive (DP; CD4⁺CD8⁺), CD4⁺CD8^{Lo}, then single positive CD4 (SPCD4; CD4⁺CD8⁻) or mature single positive CD8 (mSPCD8; CD4⁺CD8⁺TCRb⁺) (b) DN stages, defined by expression of CD44 and CD25 after selecting true DNs (CD24⁺TCRb⁻) (c) and developing thymocytes assessed by TCRb and CD69 expression. All bar graphs contain data pooled from 2 independent experiments with 3 replicates each using 4 week old mice.

In summary, TEC targeted for a heterozygous deletion of *Chd4* did not display any gross developmental or functional defects as demonstrated by normal thymic cellularity, TEC composition and maturation, and the ability to support regular thymocyte development. This, together with previous observations using both a constitutive *Chd4* knockout model and the conditional model in which *Chd4* was deleted in developing thymocytes (225, 300), suggests that there is no haploinsufficiency at the *Chd4* locus.

4.4 Homozygous *Chd4* deficiency in TEC causes thymic hypocellularity

Having established that there was no overt *Chd4* haploinsufficiency in heterozygous mice, the role for *Chd4* in TEC biology was further assessed using mice with a homozygous deletion of *Chd4* targeted to TEC (i.e. *Chd4*^{fl/fl}:Foxn1Cre; designated “cKO”). These mice were compared to *Chd4*^{fl/fl} littermates (designated “WT”), which served as the appropriate controls.

At 4 weeks of age, the analysis of cKO mice revealed a hypoplastic thymus on gross anatomical inspection when compared to age-matched WT thymi (Figure 4.4a). The absolute cell count, determined after enzymatic digestion of thymic lobes (Figure 4.4b), showed that the cellularity of cKO thymi was halved in comparison to WT thymi ($219.3 \pm 12.8 \times 10^6$ vs. $389.7 \pm 27.2 \times 10^6$). This hypocellularity proportionately affected the TEC and thymocyte compartments within the thymus (Figure 4.4c).

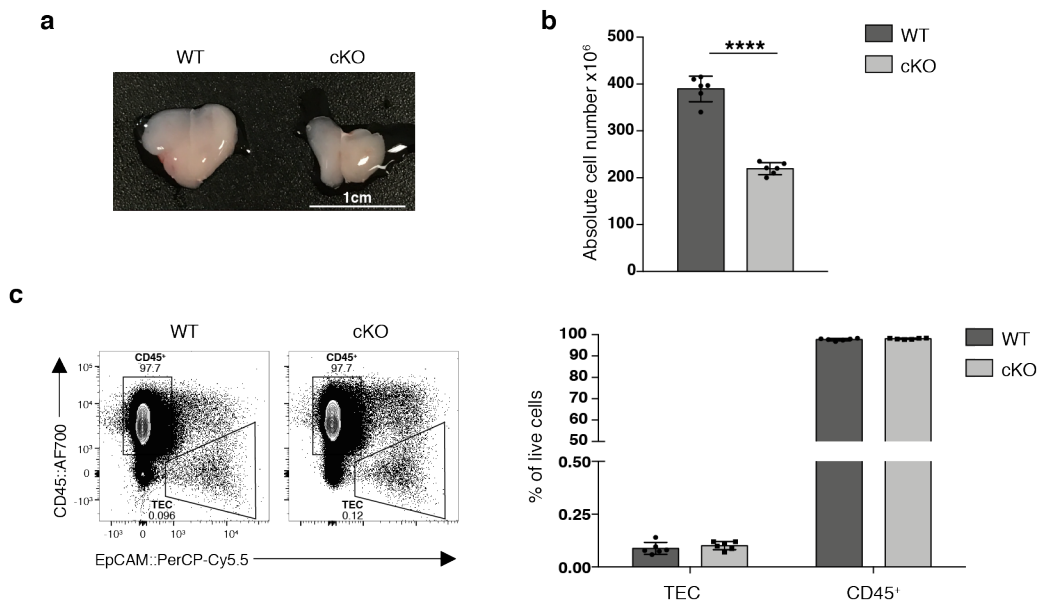


Figure 4.4 Homozygous TEC conditional *Chd4* deficiency causes thymic hypocellularity affecting both the TEC and thymocyte compartments proportionately

(a) Representative photographs showing WT and cKO thymi from 4 week old mice. (b) Bar graph showing absolute cellularity after enzymatic digestion. (c) Representative FACS plots (left panel) and bar graph (right panel) showing relative frequency of TEC (CD45⁺EpCAM⁺) and CD45⁺ cells. Bar graphs contain data pooled from 2 independent experiments with 3 replicates each, using 4 week old mice.

4.5 CHD4 is important for normal TEC development

4.5.1 CHD4 deficiency causes abnormalities in the composition of the adult TEC compartment, particularly affecting the mTEC lineage

The composition of the TEC compartment of 4 week old mice was assessed by flow cytometric phenotyping after enzymatic digestion of cKO and WT thymi, followed by depletion of CD45⁺ cells in order to enrich for the relatively rare stromal cell constituents. This approach showed that, in addition to a proportional reduction in TEC cellularity (Figure 4.4), cKO thymi had an increased frequency of cTEC ($33.1 \pm 5.5\%$ vs. $11.6 \pm 1.9\%$), and correspondingly a decreased frequency of mTEC ($64.5 \pm 5.3\%$ vs. $85.3 \pm 2.1\%$) (Figure 4.5a).

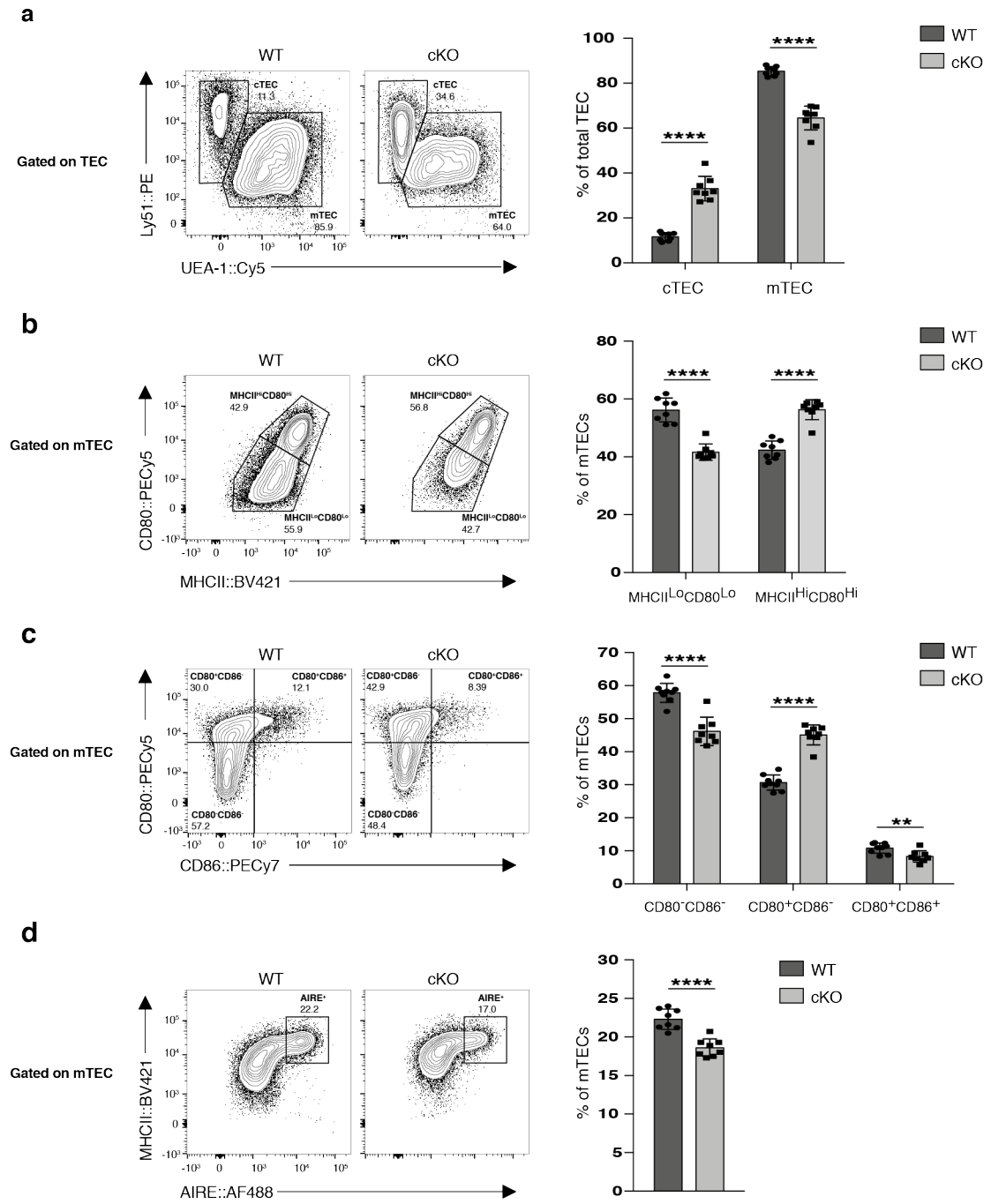


Figure 4.5 TEC conditional CHD4 deficiency causes abnormalities in the composition of the adult TEC compartment particularly affecting the mTEC lineage

(a-d) Representative FACS plots (left panels) and bar graphs (right panels) showing relative frequencies of (a) cTEC (Ly51⁺UEA-1⁺) and mTEC (Ly51⁺UEA-1⁺) within total TEC (CD45⁺EpCAM⁺), (b) MHCII^{Lo}CD80^{Lo} and MHCII^{Hi}CD80^{Hi} mTEC, (c) CD80⁻CD86⁻, CD80⁺CD86⁻ and CD80⁺CD86⁺ mTEC, and (d) AIRE⁺ mTEC in thymi from 4 week old mice. All bar graphs contain data pooled from 2 independent experiments with 4 replicates each.

Within the mTEC compartment, there were significant changes in lineage maturation. Reductions in the relative frequencies of the most mature stages including CD80⁺CD86⁺ mTEC ($8.3 \pm 1.7\%$ vs. $10.8 \pm 1.5\%$) (Figure 4.5c) and AIRE⁺ mTEC ($18.6 \pm 1.2\%$ vs. $22.3 \pm 1.3\%$) were noted (Figure 4.5d). This finding, alongside relative accumulations of CD80⁺CD86⁻ ($45.1 \pm 3.0\%$ vs. $30.6 \pm 2.3\%$) (Figure 4.5c) and MHCII^{Hi}CD80^{Hi} mTEC ($56.3 \pm 3.4\%$ vs. $42.4 \pm 3.2\%$) (Figure 4.5b) in cKO compared to WT mice, was highly suggestive of partial blocks within the mTEC developmental pathway, in which mTEC sequentially acquire the expression of MHCII, CD80, CD86 and AIRE (22, 55, 56, 179).

4.5.2 Changes in the composition of the adult TEC compartment of CHD4 deficient thymi cannot be explained by abnormal apoptosis

Whilst the findings described above were highly suggestive of partial developmental blocks in TEC differentiation, it was important to consider abnormalities in apoptosis and proliferation as potential underpinning mechanisms for the changes in mTEC frequencies observed. To quantify apoptosis within the various TEC subpopulations, a flow cytometric assay was used combining staining with Annexin V, which binds phosphatidylserine residues exposed on the plasma membrane early during apoptosis, and DAPI, which enters cells and binds DNA later during apoptosis once membrane integrity has been lost (306). There were no differences between cKO and WT in the proportions of live (AnnexinV⁻DAPI⁻), early apoptotic (AnnexinV⁺DAPI⁻) or late apoptotic (DAPI⁺) cells within the TEC subpopulations analysed (Figure

4.6), indicating that abnormal apoptosis is unlikely to explain the changes in the composition of the adult TEC compartment seen in CHD4 deficient thymi.

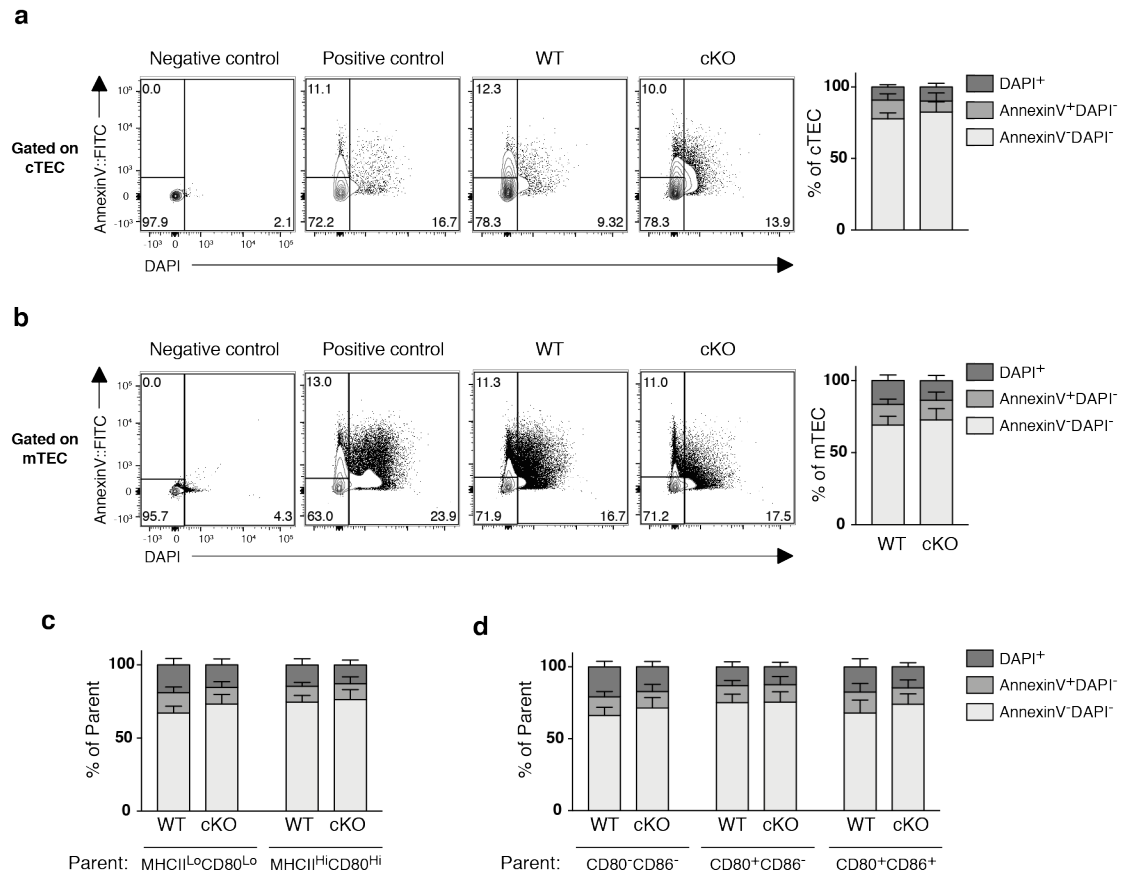


Figure 4.6 Abnormalities in apoptosis do not explain changes in the composition of the adult TEC compartment seen in CHD4 deficient thymi

(a and b) Representative FACS plots (left panels) and bar graphs (right panels) showing relative frequencies of live (AnnexinV⁻DAPI⁻), early apoptotic (AnnexinV⁺DAPI⁻) and late apoptotic (DAPI⁺) cells within (a) total cTEC (Ly51⁺UEA-1⁻) and (b) total mTEC (Ly51⁺UEA-1⁺). (c and d) Bar graphs showing relative frequencies of live (AnnexinV⁻DAPI⁻), early apoptotic (AnnexinV⁺DAPI⁻) and late apoptotic (DAPI⁺) cells within (c) MHCII^{Lo}CD80^{Lo} and MHCII^{Hi}CD80^{Hi} mTEC and (d) CD80⁻CD86⁻, CD80⁺CD86⁻ and CD80⁺CD86⁺ mTEC. Gates for AnnexinV and DAPI positivity were set on the basis of negative and positive control tubes, consisting of cells unstained for AnnexinV and DAPI, and fully stained cells after 20 minutes of UV light treatment respectively. All bar graphs contain data pooled from 2 independent experiments with 3 replicates each using 4 week old mice.

4.5.3 Increased cell cycling may contribute to the relative increase in cTEC in CHD4 deficient thymi

In order to assess whether differences in cell cycle status in cKO vs. WT could account for the observed relative differences in TEC subpopulations, the proportion of cells in each stage of the cell cycle (G_0 , G_1 , S and G_2M) was determined for both cKO and WT TEC using a flow cytometric assay incorporating staining for Ki67, a proliferation-specific marker, and cellular DNA content (307). As a measure of the proportion of cells entering cell division, the cell cycle index (i.e. $(S+G_2M)/(G_0+G_1)$) was also calculated (55) (Figure 4.7).

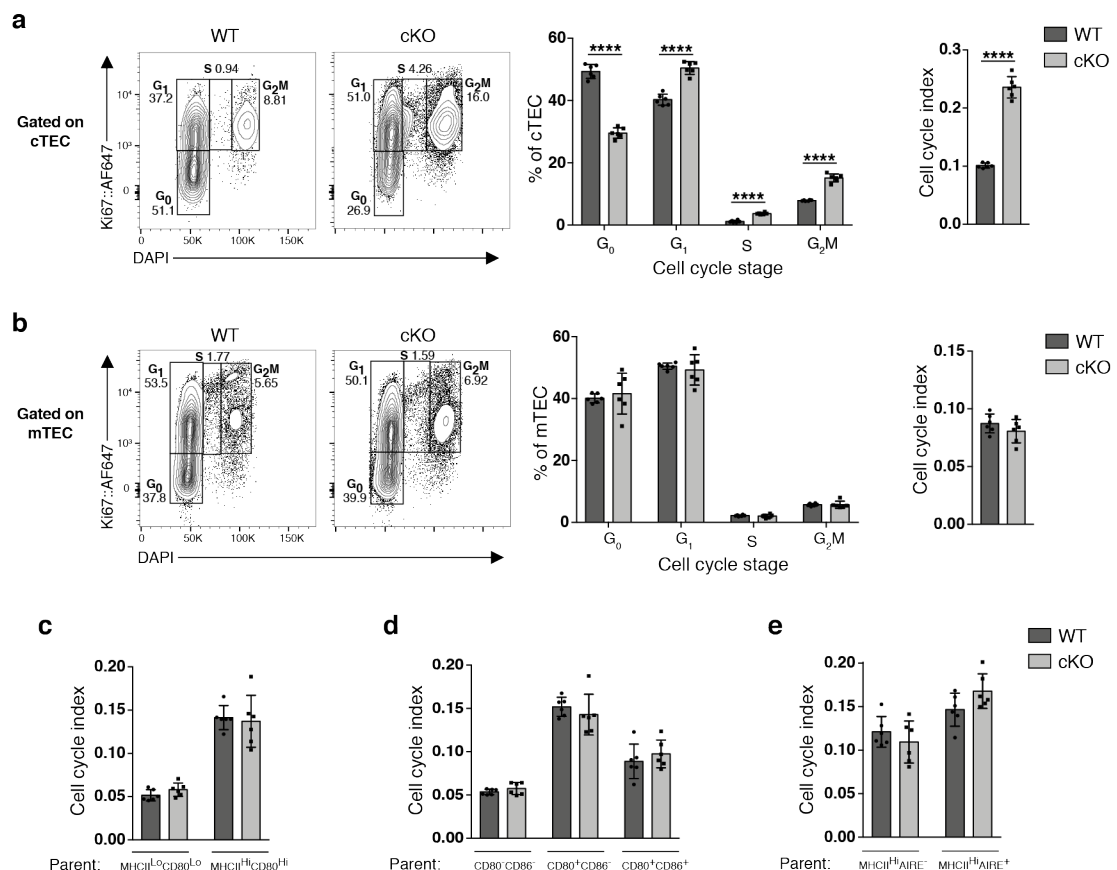


Figure 4.7 Increased cell cycle index in CHD4 deficient cTEC

(a and b) Representative FACS plots (left panels) and bar graphs (right panels) showing relative frequencies of cells in each phase of the cell cycle as well as the cell cycle index $(S+G_2M)/(G_0+G_1)$ for (a) total cTEC (Ly51⁺UEA-1⁻) and (b) total mTEC (Ly51⁺UEA-1⁺). (c and d) Bar graphs showing cell cycle index for (c) MHCII^{Lo}CD80^{Lo} and MHCII^{Hi}CD80^{Hi} mTEC, (d) CD80^{Lo}CD86^{Lo}, CD80^{Lo}CD86^{Hi} and CD80^{Hi}CD86^{Hi} mTEC and (e) MHCII^{Lo}AIRE^{Lo} and MHCII^{Hi}AIRE^{Hi} mTEC. All bar graphs contain data pooled from 2 independent experiments with 3 replicates each using 4 week old mice.

Using these measures, no cell cycle differences were observed within the mTEC compartment of cKO in comparison to WT 4 week old thymi (Figure 4.7b, c, d and e). However, cKO cTEC displayed a significantly higher frequency of cells in the G₁, S and G₂M phases of the cell cycle and a higher cell cycle index ($0.24 \pm 0.02\%$ vs. $0.10 \pm 0.00\%$) when compared to WT cTEC (Figure 4.7a). Given the relative increase in the size of the cTEC compartment and the known role of CHD4 in activating the G₂M checkpoint (230, 231), these results are compatible with the notion that a loss of CHD4 causes increased cell cycling in cTEC.

4.5.4 The TEC defects are established in the embryonic thymus and are mostly maintained during postnatal life

Total thymic cellularity, the frequency of thymocytes and TEC, and the composition of the TEC compartment were compared between thymi isolated from cKO and WT mice at embryonic day (E) 17.5 as well as at several time points after birth, i.e. week 1, 12 and 26, to assess whether the phenotype described at 4 weeks is already established earlier in life or during embryonic development. This analysis was also designed to assess whether the observed changes are stable throughout postnatal thymic development including when age-related thymic involution physiologically occurs (Figure 4.8). Whilst mTEC can be detected from E13.5 onwards (40, 308), E17.5 was selected to have a sufficient number of mTEC per mouse for a phenotypic analysis without the need to pool individual mice to obtain sufficient cells (40, 56). The embryonic mTEC compartment has a distinct composition in

comparison to that of the adult, particularly with respect to the expression of MHCII and (309), and AIRE (159). Therefore the embryonic data were only used in comparisons of thymic cellularity, thymocyte and TEC frequency, and frequency of c- and m-TEC and not for the more detailed assessments of mTEC phenotype.

Thymic hypocellularity was already present at E17.5 (Figure 4.8a), equally affecting thymocytes and TEC (Figure 4.8b). This phenotype was maintained at all time points studied thereafter. The relative overrepresentation of cTEC (Figure 4.8c) and deficiency of mTEC (Figure 4.8d) was likewise already present in E17.5 thymi and again this phenotype was maintained over time. Most of the alterations within the cKO mTEC compartment were stable across the postnatal time points analysed (Figure 4.8e-h), with the exception of CD80⁺CD86⁺ and AIRE⁺ mTEC (Figure 4.8i and j). Their frequencies, although decreased at 1 and 4 weeks, apparently normalised by 12 weeks of age and remained as expected thereafter. Taken together, the relative decreases in the most immature mTEC subpopulations, namely MHCII^{Lo}CD80^{Lo} and CD80⁻CD86⁻ mTEC, alongside accumulations of intermediate maturity CD80⁺CD86⁻ mTEC, and deficiencies in mature CD80⁺CD86⁺ and AIRE⁺ mTEC are highly suggestive of a partial developmental block in postnatal mTEC maturation, with partial normalisation over time.

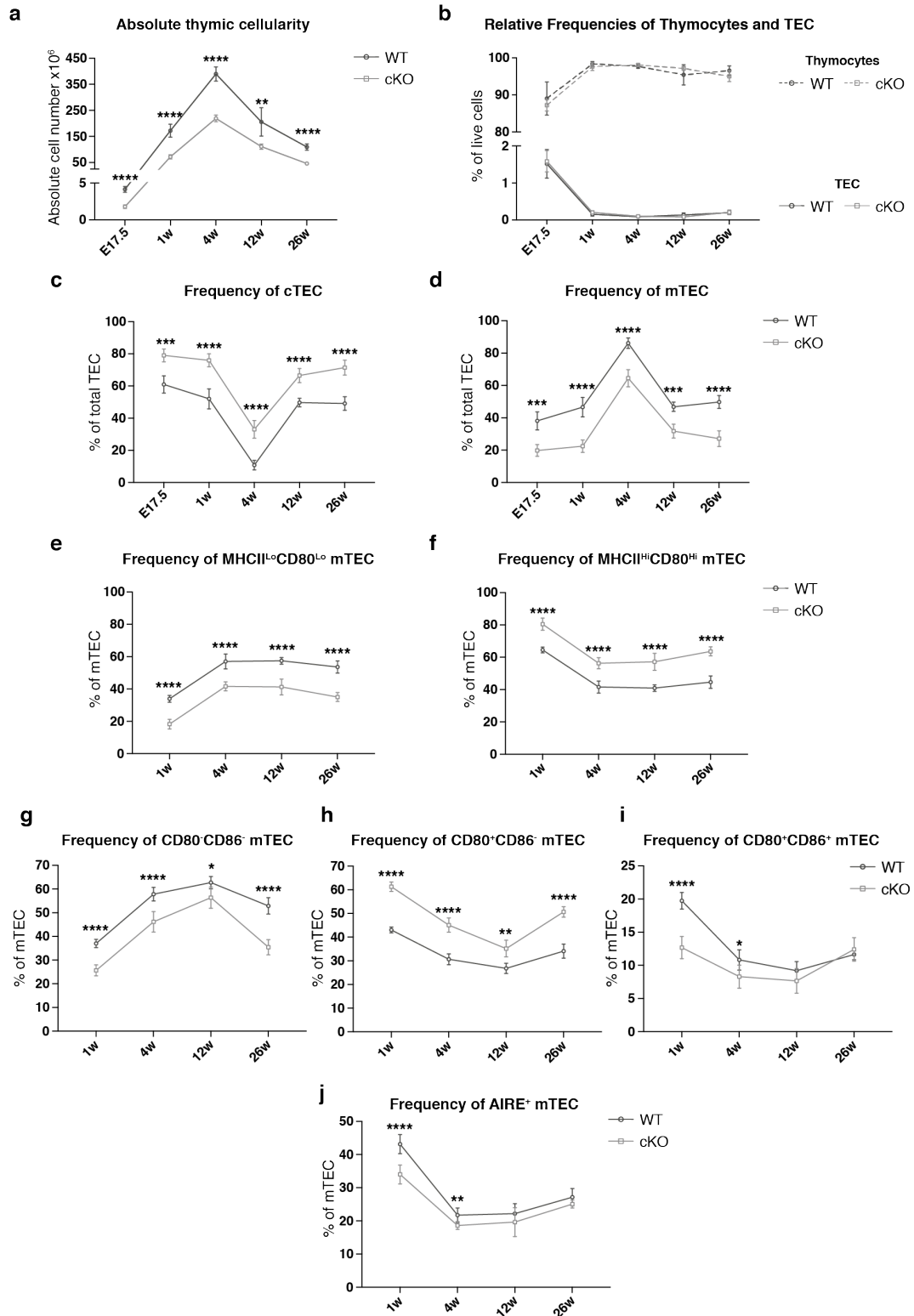


Figure 4.8 The TEC defects observed in CHD4 deficient thymi are established in the embryonic thymus and mostly maintained during ageing

(a) Total thymic cellularity after enzymatic digestion and relative frequencies of (b) thymocytes (CD45⁺EpCAM⁺) and TEC (CD45⁺EpCAM⁺) (c) cTEC (Ly51⁺UEA-1⁻), (d) mTEC (Ly51⁺UEA-1⁺), (e) MHCII^{Lo}CD80^{Lo} and (f) MHCII^{Hi}CD80^{Hi} mTEC, (g) CD80⁺CD86⁻, (h) CD80⁺CD86⁻ and (i) CD80⁺CD86⁺ mTEC, and (j) AIRE⁺ mTEC in thymi from mice at embryonic day (E) 17.5 and 1, 4, 12 and 26 weeks (W) postnatally. All points contain data pooled from 2 independent experiments with 3 replicates each.

In summary, these analyses suggest that the defects observed within the TEC compartment of cKO mice are established in the embryonic thymus, are generally maintained during ageing and occur due to partial blocks in the development and maturation of the mTEC lineage in particular.

4.5.5 E17.5 CHD4 deficient TEC express higher levels of markers associated with progenitor status

Given that abnormalities within the TEC compartment of CHD4 deficient thymi were already established at E17.5, a more detailed phenotyping of TEC at this embryonic stage was performed (Figure 4.9). Consistent with the findings in the postnatal thymus, there was also an increase in MHCII and CD80 expression by cKO E17.5 mTEC compared to WT (Figure 4.9a). MHCII expression was assessed by geometric mean fluorescence intensity (gMFI) as an MHCII^{Lo} population is not distinguishable at this stage in ontogeny (Figure 4.9a). Markers of TEC progenitor status were also assessed, including SCA1, CD24 and podoplanin. SCA1 and CD24 have been shown to identify both embryonic and adult TEC progenitors (49, 50, 53, 310-312). SCA1 has been shown to mark TEC progenitors with bipotent potential (49, 50) and CD24 is expressed on embryonic cTEC-like mTEC progenitors (53). The latter cell type appears to express cTEC markers, including Ly51, β 5t, and cytokeratin 8 (K8), as well as mTEC markers, such as K5, and has the potential to differentiate into mature AIRE⁺ mTEC in response to RANK signalling (53). Podoplanin has also been described as marker of committed mTEC progenitors, albeit in the postnatal thymus (54).

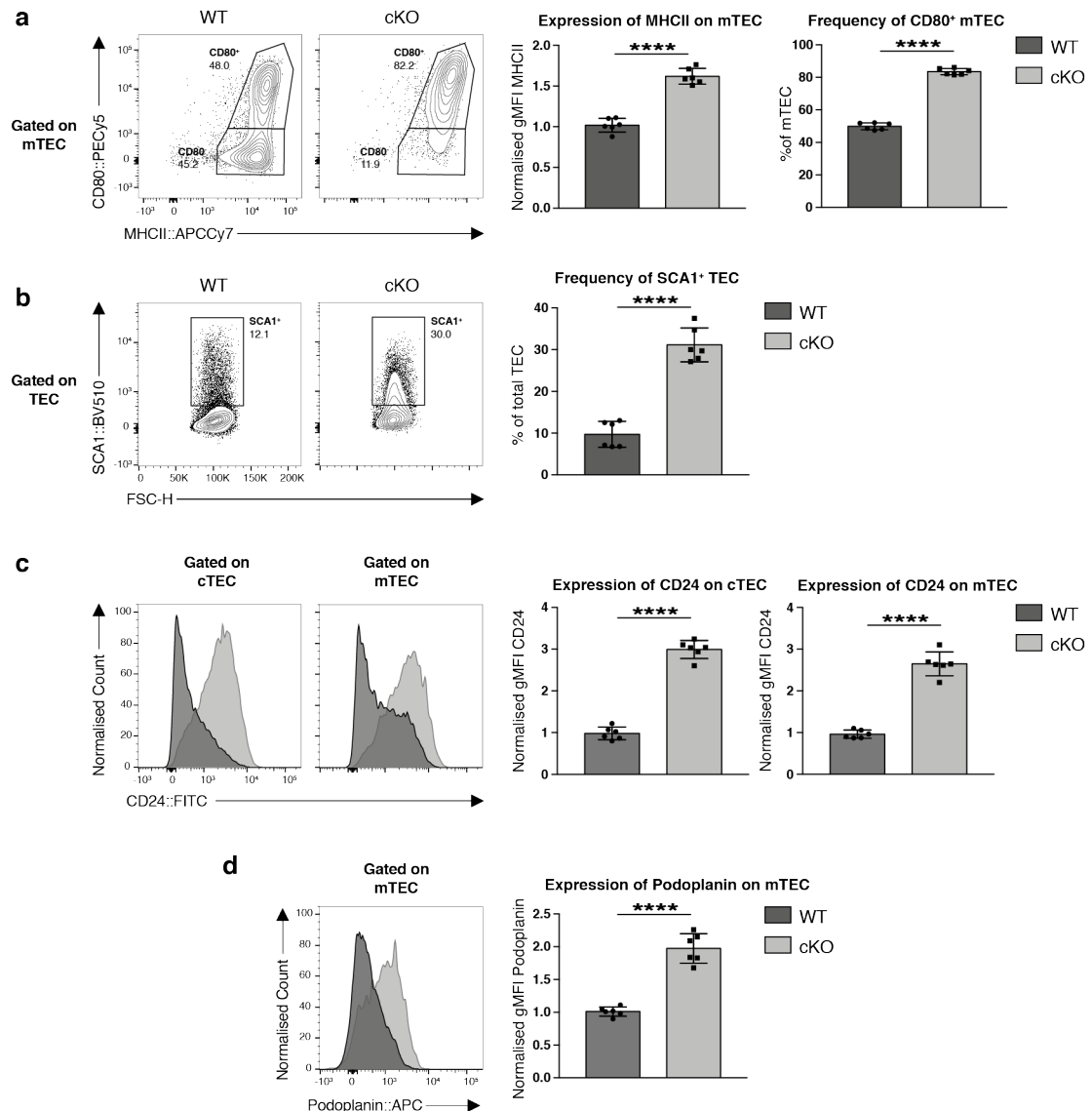


Figure 4.9 E17.5 CHD4 deficient TEC express higher levels of MHCII, CD80 and markers associated with progenitor status

(a-d) Representative FACS plots (left panels) and bar graphs (right panels) showing expression of (a) MHCII and CD80 on mTEC (Ly51⁺UEA-1⁺), (b) SCA1 on total TEC (CD45⁺EpCAM⁺), (c) CD24 on cTEC (Ly51⁺UEA-1⁺) and mTEC (Ly51⁺UEA-1⁺), and (d) Podoplanin on mTEC (Ly51⁺UEA-1⁺) from E17.5 thymi. All graphs contain data pooled from 2 independent experiments with 3 replicates each. Normalised gMFI was calculated by dividing the gMFI of each sample within an individual experiment by the average of the WT gMFIs within the same experiment.

The analyses for progenitor markers showed an increased frequency of SCA1⁺ TEC within E17.5 cKO thymi compared to WT (31.2 ± 4.1% vs. 9.7 ± 3.1%) (Figure 4.9b), increased CD24 expression on cKO cTEC and mTEC (Figure 4.9c), and increased expression of Podoplanin (Figure 4.9d) on cKO mTEC.

These results are in support of TEC defects in cKO thymi being established during embryonic TEC development with CHD4 deficient TEC being impaired in their ability to differentiate from a progenitor state to more mature stages (Figure 4.10).

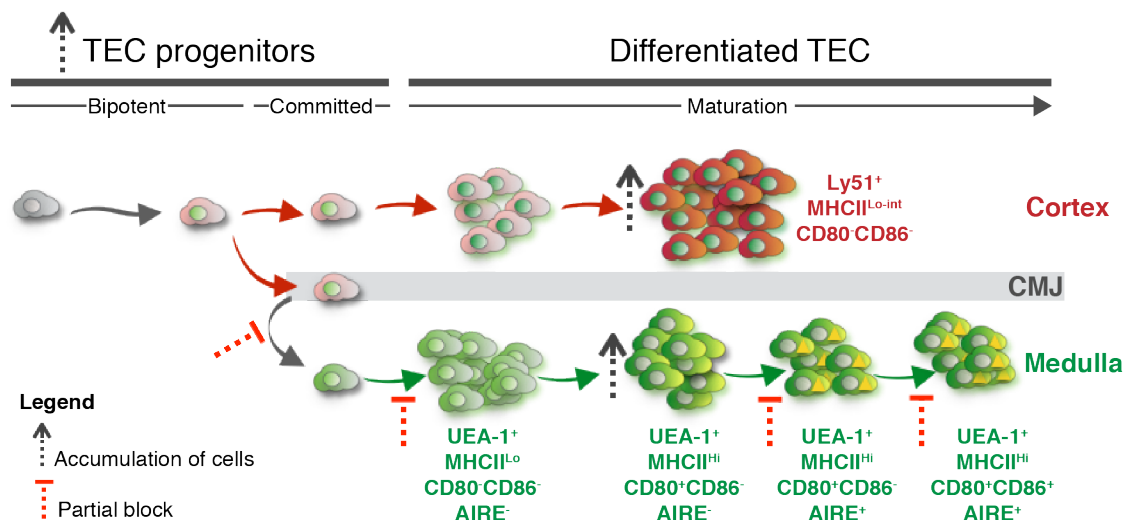


Figure 4.10 Diagram summarising the effects of CHD4 deficiency on TEC development and maturation

TEC conditional CHD4 deficiency causes defective development and maturation of TEC beginning during embryonic stages of thymic organogenesis. There are relative accumulations of TEC with a progenitor phenotype, cTEC, and immature mTEC consistent with an impaired ability of CHD4 deficient TEC to differentiate from a progenitor state to mature mTEC stages.

4.5.6 E17.5 CHD4 deficient FTOC are able to respond to exogenous Receptor activator of nuclear factor kappa-B ligand (RANK-L) but this does not rescue the cKO phenotype

RANK:RANK-L interactions are known to be important for the differentiation and maturation of mTEC, particularly within the fetal thymus (22, 65). RANK is expressed by cells of the mTEC lineage from the mTEC progenitor stage (53), and its ligand (RANK-L) is provided by various intrathymic cell types including lymphoid tissue inducer cells (22), $\gamma\delta$ T-cells that express an invariant V γ 5 TCR (66), and positively selected $\alpha\beta$ thymocytes (67, 68).

The analyses of embryonic and postnatal TEC phenotypes presented above suggest that the development and maturation of the mTEC lineage are compromised in the absence of CHD4, a finding that could be explained by with impaired RANK signalling. Interestingly, TEC from E17.5 cKO thymi also expressed more CD24 compared to WT, a marker that has been described to identify mTEC progenitors that differentiate into mature mTEC in response to RANK-L signalling (53). Therefore, the ability of cKO TEC to respond to exogenous RANK-L was tested in fetal thymic organ cultures (FTOCs). To this end, cKO and WT E17.5 thymic lobes were first depleted of lymphoid cells using 2-deoxyguanosine (2DG), as these cells may confound the analysis by providing physiological RANK-L, as well as other signals that stimulate mTEC maturation such as CD40L and LT β R (65). “Lymphodepleted” lobes were cultured as fetal thymic organ cultures in the presence or absence of

exogenous RANK-L ($1\mu\text{g/ml}$). Their phenotype and cellularity were assessed after 4 days of culture using flow cytometry (Figure 4.11).

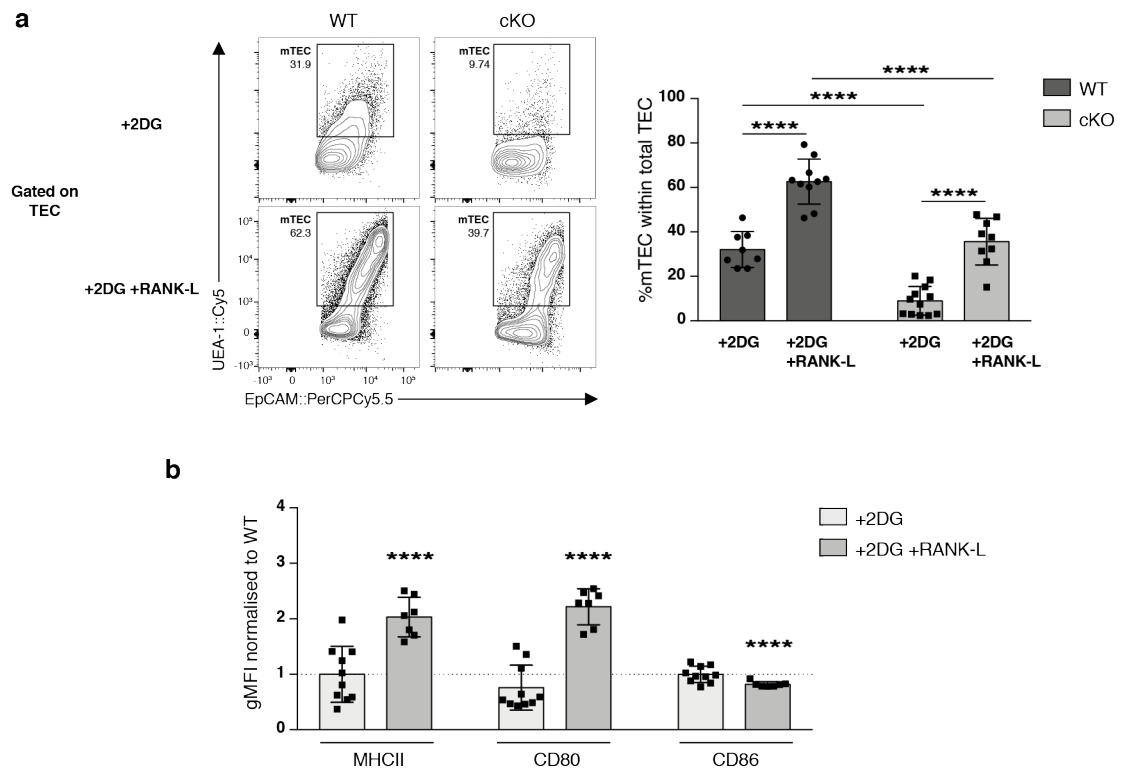


Figure 4.11 RANK-L treatment of lymphodepleted CHD4 deficient FTOCs induces mTEC development but does not rescue the defect in TEC maturation

(a) Representative FACS plots (left panels) and bar graph (right panel) showing frequency of mTEC (UEA-1⁺) within total TEC (CD45⁺EpCAM⁺) in cKO and WT E17.5 FTOCs after lymphodepletion with 2-deoxyguanosine (2DG) +/- treatment with $1\mu\text{g/ml}$ RANK-L (b) Normalised gMFI of MHCII, CD80 and CD86 in mTEC from 2DG treated cKO FTOCs +/- $1\mu\text{g/ml}$ RANK-L. gMFI was normalized to WT values by dividing the gMFI of each sample within an individual experiment by the average of the WT gMFIs within the same experiment, resulting in a mean WT value of 1 (dotted line); stars indicate results in cKO that are significantly different from WT values. All graphs contain data pooled from 2 independent experiments.

The frequency of mTEC in CHD4 deficient 2DG treated FTOCs increased from $9.0 \pm 6.5\%$ to $35.6 \pm 10.5\%$ after RANK-L treatment (Figure 4.11a), indicating that an absence of CHD4 in TEC does not abrogate the cells' ability to respond to RANK-L. In fact, the fold increase in the frequency of mTEC in RANK-L treated vs. untreated FTOCs was higher for cKO compared to WT (4-fold vs. 2-fold increase). RANK-L treatment however was not able to rescue the cKO TEC phenotype. RANK-L treated cKO FTOCs continued to display a relative reduction in mTEC compared to WT ($35.6 \pm 10.5\%$ vs. $62.7 \pm 10.1\%$) (Figure 4.11a), and the expression of MHCII, CD80 and CD86 on mTEC from cKO FTOCs treated with RANK-L recapitulated the findings described in postnatal thymi, with increased expression of MHCII and CD80 but decreased expression of CD86 compared to WT (Figure 4.11b). Taken together these results indicate that RANK responsiveness (at least under the chosen experimental conditions) remained unaffected by the absence of CHD4 expression in TEC.

4.5.7 CHD4 deficiency in TEC causes abnormal thymic morphology

Having described the composition of the TEC compartment by flow cytometry of embryonic and postnatal thymi the spatial arrangement of the thymic stroma was next determined because a normal thymic architectural organisation is required to provide distinct microenvironments specialised for each stage of thymocyte maturation (2, 7, 159). Alterations in thymic structure have accordingly been associated with immune deficiency and autoimmune disease (2, 7, 313-316).

Alongside the aforementioned reductions in thymic size and cellularity (Figure 4.4a and b, and 4.8a), microscopic examination of thymic sections from 4 week old cKO mice revealed prominent defects in thymic architectural organisation (Figure 4.12). Haematoxylin and eosin (H&E) staining showed a less complex structure to the medullary island network in cKO sections compared to WT, with fewer separate medullary islands (Figure 4.12a). The cortico-medullary disorganisation was further delineated by immunofluorescent microscopy for cytokeratins 8 (K8) and 5 (K5), which mark cTEC and mTEC, respectively (Figure 4.12b and c). This analysis showed an unstructured mixing of K8⁺ cTEC and K5⁺ mTEC in cKO thymic sections, which consequently resulted in a lack of distinct medullary islands (Figure 4.12b and c). Equally striking were the presence of large areas devoid of thymic epithelial cells and therefore the provision of the characteristic stromal scaffold (Figure 4.12 b and c), which normally fosters the required physical and functional cross talk between the stroma and developing thymocytes. Despite the lack of an epithelial cell scaffold, thymocytes accumulated within these epithelial-free areas (Figure 4.12d). The majority of the thymocytes present in these areas were cells with a DP phenotype, although SPCD4 and SPCD8 thymocytes could also be detected (Figure 4.12d).

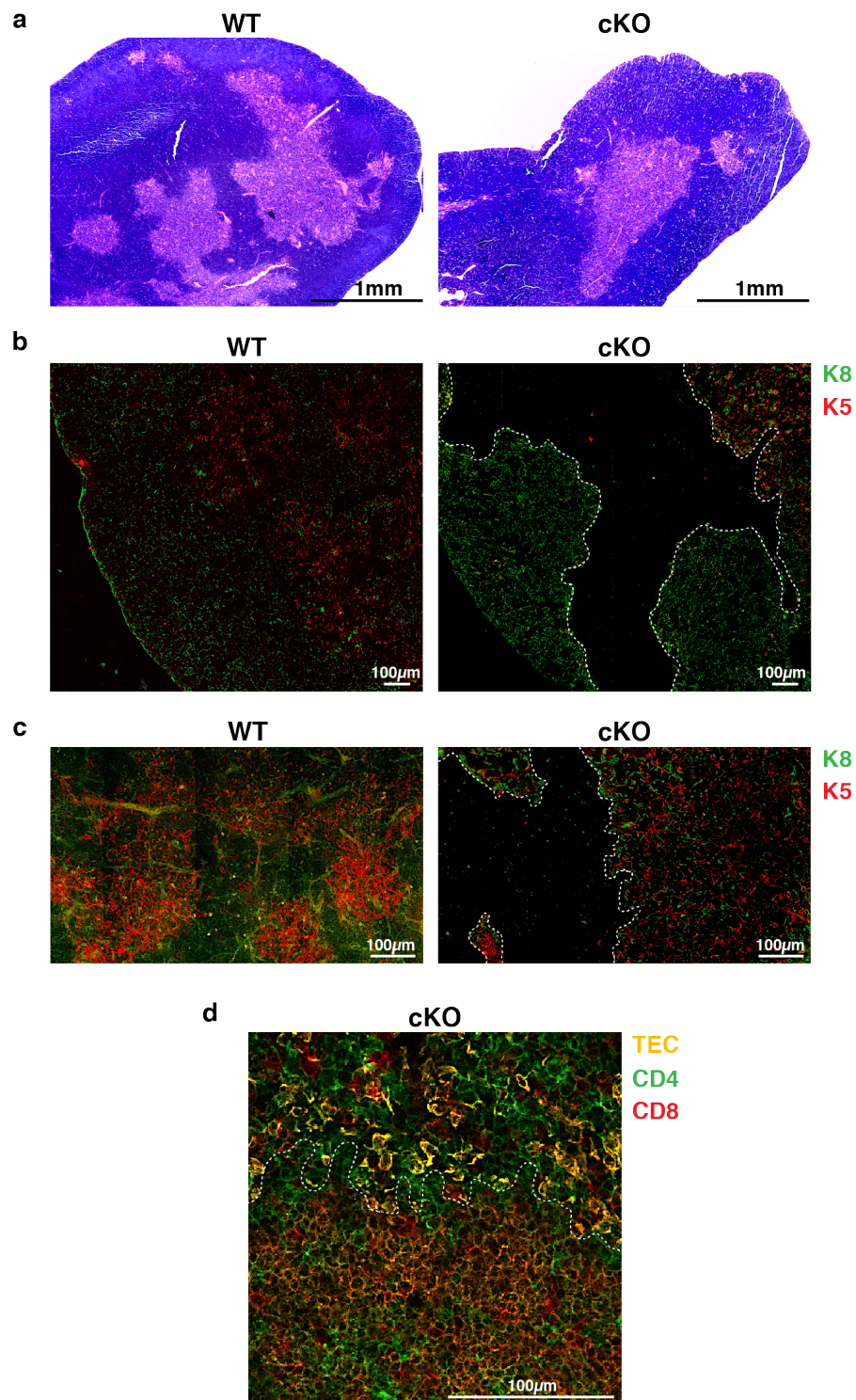


Figure 4.12 Thymic architecture is altered in thymi composed of TEC conditionally deficient in CHD4

(a-d) Microscopic examination of thymic sections from 4 week old cKO and WT thymi (a) Haematoxylin and eosin (H&E) staining (b and c) Immunofluorescence microscopy for cytokeratins 8 (K8: Green) and 5 (K5: Red); epithelial cell free areas are marked by white dotted lines (d) Immunofluorescence microscopy for TEC (Yellow), CD4 (Green) and CD8 (Red); epithelial cell free area is marked by white dotted line.

In summary, additional to the defects in TEC development and maturation identified by flow cytometry, CHD4 deficiency in TEC causes thymic structural malformations with abnormal cortico-medullary separation and accumulations of developing thymocytes in areas devoid of an epithelial scaffold. How such an aberrant thymic microenvironment, with an abnormal TEC compartment in terms of both cellular composition and architectural organisation, is able to support the regular development and selection of thymocytes forms the focus of the next section of this thesis.

4.6 The distorted thymic microenvironment composed of TEC devoid of CHD4 expression is unable to support regular thymocyte maturation and selection

4.6.1 TEC conditional CHD4 deficiency causes defective maturation of thymocytes

The ability of the distorted thymic microenvironment caused by TEC conditional CHD4 deficiency to support regular thymocyte maturation and selection was next probed in a series of flow cytometric experiments. Firstly, thymocyte maturation from the DN to SP stages was assessed (Figure 4.13). As previously shown (Figure 4.4), TEC conditional CHD4 deficiency causes hypocellularity of the thymus, with proportionate effects on the TEC and thymocyte compartments. Analysis of thymocyte maturation revealed relative accumulations of cells at the DN ($CD4^-CD8^-$) and DP ($CD4^+CD8^+$) stages, alongside relative deficiencies in the generation of mature SPCD4 ($CD4^+CD8^-$; $4.9 \pm 0.4\%$ vs. $8.0 \pm 1.4\%$) and mSPCD8 ($CD4^-CD8^+TCRb^+$; $1.4 \pm 0.2\%$ vs. $2.3 \pm 0.6\%$) thymocytes in cKO compared to WT mice (Figure 4.13a). However, maturation within the DN thymocyte compartment was unperturbed in cKO mice, as evidenced by comparable frequencies of cells in the DN1 ($CD44^+CD25^-$), DN2 ($CD44^+CD25^+$), DN3 ($CD44^-CD25^+$) and DN4 ($CD44^-CD25^-$) stages of development (Figure 4.13b). Taken together, these results are suggestive of a partial block in thymocyte maturation between the DP and SP stages. The contributions of defects in positive and negative selection, as well as post-selectional maturation of single positive thymocytes to this block in thymocyte development were next assessed.

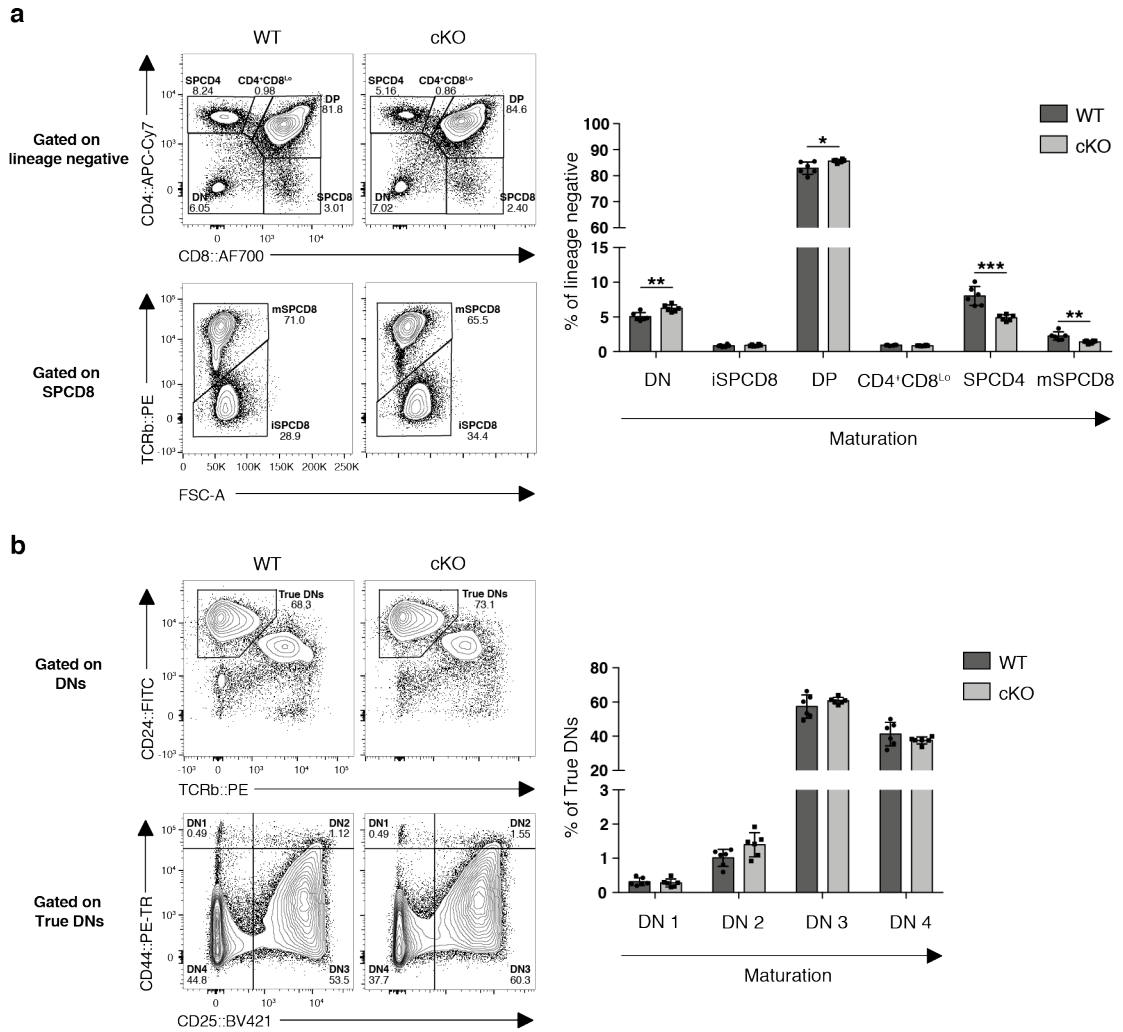


Figure 4.13 Impaired generation of mature single positive thymocytes in thymi from mice with TEC conditional CHD4 deficiency

(a and b) Representative FACS plots (left panels) and bar graphs (right panels) showing relative frequencies of (a) thymocyte maturation stages defined by CD4 and CD8 expression, progressing from double negative (DN; CD4⁻CD8⁻), to immature single positive CD8 (iSPCD8; CD4⁻CD8⁺TCRb⁺), double positive (DP; CD4⁺CD8⁺), CD4⁺CD8^{L0}, then single positive CD4 (SPCD4; CD4⁺CD8⁻) or mature single positive CD8 (mSPCD8; CD4⁻CD8⁺TCRb⁺) (b) DN stages, defined by expression of CD44 and CD25 after selecting true DNs (CD24⁺TCRb⁺). All bar graphs contain data pooled from 2 independent experiments with 3 replicates each using 4 week old mice.

4.6.2 TEC conditional CHD4 deficiency causes defective positive selection

The process of thymocyte positive selection was compared in CHD4 cKO vs. WT thymi by examining the expression of CD69 and TCR β on developing thymocytes (Figure 4.14) (303-305). Pre-selectional thymocytes can be identified by expression of undetectable to low levels of TCR β and lack of expression of the activation marker CD69 (Stages 0 and 1 in Figure 4.14); then, during the process of positive selection, the cells upregulate their expression of TCR β to an intermediate level and, as a result of activating signals mediated by the CD3/TCR complex, begin to express CD69 on their cell surface (Stage 2 in Figure 4.14) (304). Immediately post-selectional thymocytes express high levels of CD69 and TCR β (Stage 3 in Figure 4.14), and finally the cells downregulate expression of CD69 during the process of post-selectional maturation (Stage 4 in Figure 4.14) (304). The relative accumulation of pre-selectional Stage 1 thymocytes seen in cKO thymi compared to WT ($43.2 \pm 2.0\%$ vs. $38.6 \pm 1.3\%$), in combination with a reduced frequency of thymocytes at the post-selectional Stages 3 ($4.4 \pm 0.7\%$ vs. $6.2 \pm 0.8\%$) and 4 ($3.1 \pm 0.5\%$ vs. $4.7 \pm 1.4\%$) are consistent with a partial defect in positive selection (Figure 4.14). Whether aberrant post-selectional maturation or negative selection could also contribute to the deficiencies of Stage 3 and 4 thymocytes in cKO thymi is further assessed in the sections below.

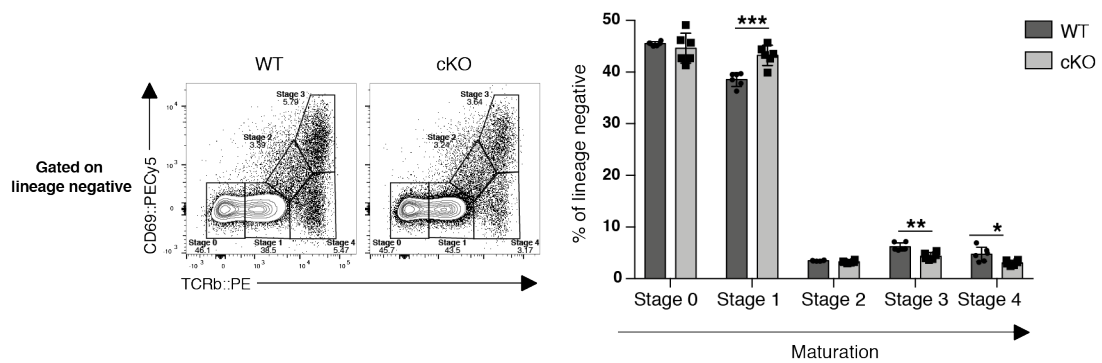


Figure 4.14 TEC conditional CHD4 deficiency causes Impaired upregulation of TCR β and CD69 in developing thymocytes consistent with defective positive selection

Representative FACS plots (left panel) and bar graph (right panel) showing relative frequencies of developing thymocytes assessed by TCR β and CD69 expression. Bar graph contains data pooled from 2 independent experiments with 3 replicates each using 4 week old mice.

4.6.3 CHD4 deficient TEC are unable to support regular post-selectional maturation of SPCD4 thymocytes

Following positive selection, maturation within the single positive thymocyte lineages can be differentiated into several stages based on the expression of various cell surface markers, which include TCR β , CCR7, CD24, CD69, and MHCI (97-99). Using this combination of markers, conventional SPCD4 thymocytes (CD5^{Hi}TCRb^{Int-Hi}CD4⁺CD8⁻CD25⁻) initially express high levels of the activation markers CD24 and CD69 (Stage A: CD24^{Hi}CD69^{Hi}TCRb^{Int}CCR7⁻). They then upregulate expression of CCR7, the chemokine receptor which regulates migration of thymocytes from the cortex to the medulla in response to CCL19 and CCL21 gradients (Stage B: CD24^{Hi}CD69^{Hi}TCRb^{Hi}CCR7⁺MHCI^{Lo}). The next stage of maturation within the conventional SPCD4 lineage involves the upregulation of MHCI (Stage C: CD24^{Hi}CD69^{Hi}TCRb^{Hi}CCR7⁺MHCI^{Hi}). Finally, mature SPCD4 thymocytes downregulate CD24 and CD69 cell surface expression (Stage D: CD24^{Lo}CD69^{Lo}) and gain the ability to egress the thymus to the periphery (Figure 4.15a for gating strategy) (97-99). Comparison of SPCD4 maturation in cKO and WT thymi revealed a relative increase in the earliest SPCD4 stage (Stage A), alongside a decreased frequency of the most mature stage (Stage D), indicative of defective maturation within the committed SPCD4 lineage following positive selection (Figure 14.5a and b).

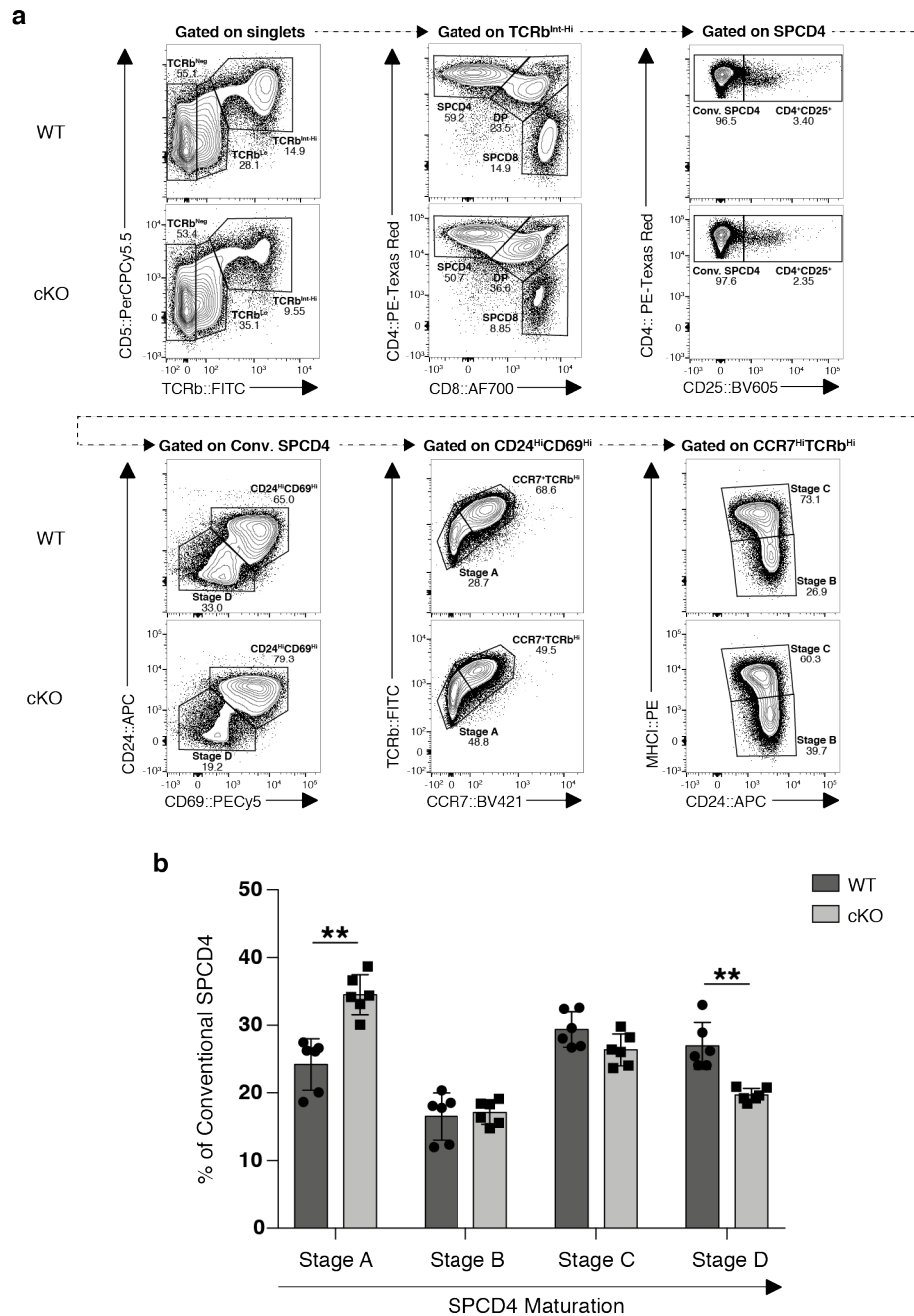


Figure 4.15 CHD4 deficient TEC are unable to support regular post-selectional maturation of SPCD4 thymocytes

(a) Representative FACS plots from WT and cKO samples demonstrating gating strategy used to analyse post-selectional maturation of the SPCD4 lineage (b) bar graph showing relative frequencies of SPCD4 maturational stages. Bar graphs contains data pooled from 2 independent experiments with 3 replicates each using 4 week old mice

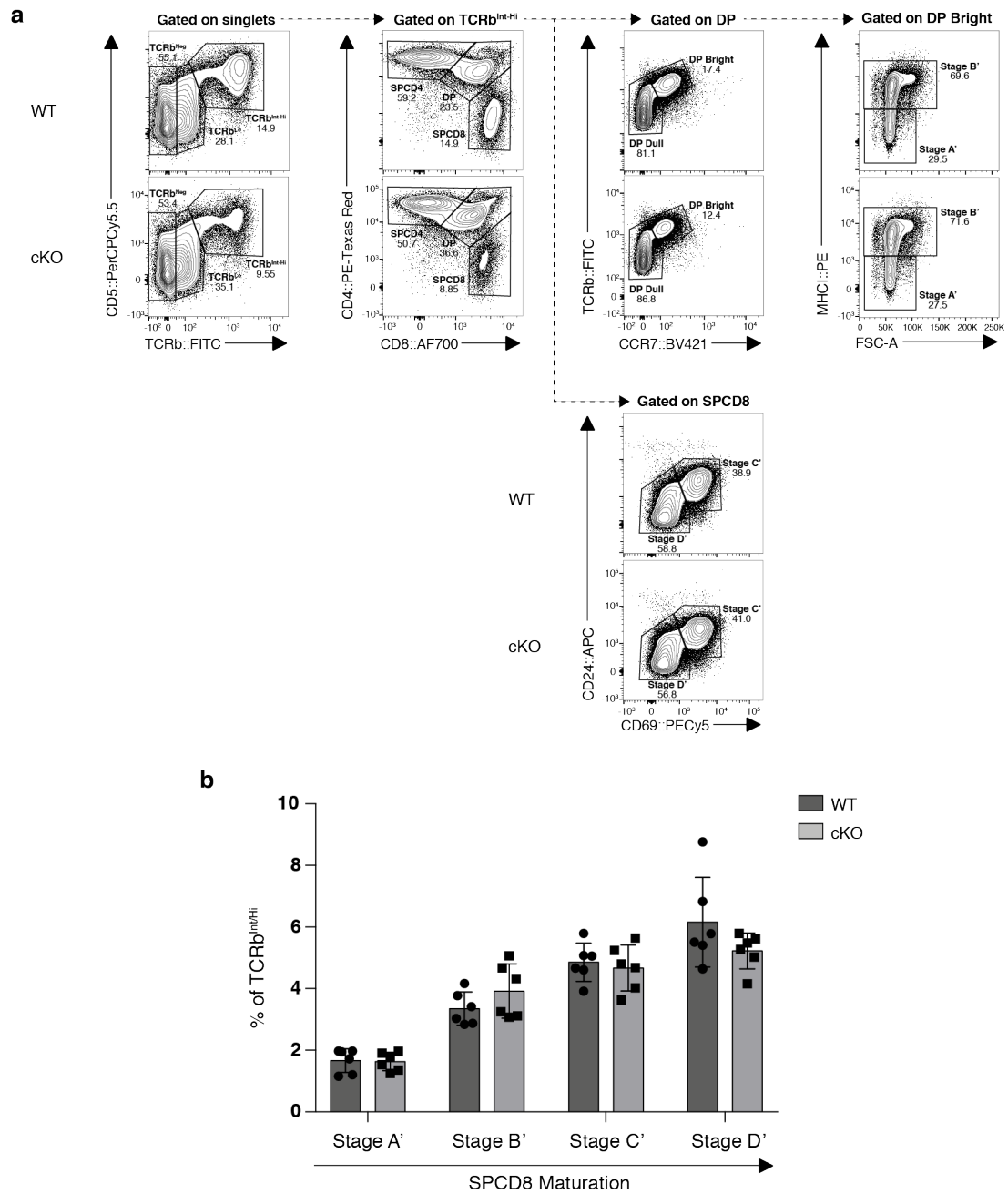


Figure 4.16 Post-selectional maturation within the committed SPCD8 lineage is grossly normal in the context of TEC conditional CHD4 deficiency

(a) Representative FACS plots from WT and cKO samples demonstrating gating strategy used to analyse post-selectional maturation of the SPCD8 lineage (b) bar graph showing relative frequencies of SPCD8 maturational stages. Bar graphs contains data pooled from 2 independent experiments with 3 replicates each using 4 week old mice

Maturation of CD8 lineage thymocytes can be followed in a fashion similar to that of SP^{CD4} thymocytes though taking into account a few key differences as follows. The earliest stage of commitment to the SP^{CD8} lineage falls within the DP gate ($CD5^{Hi}TCRb^{Hi}CD4^{+}CD8^{+}$) in cells that can be identified by expression of CCR7 and higher levels of TCR β (Stage A': $TCRb^{Hi}CCR7^{+}MHCI^{Lo}$) (317-319). These cells upregulate expression of MHCI (Stage B': $TCRb^{Hi}CCR7^{+}MHCI^{Hi}$), become single positive for CD8 expression ($CD5^{Hi}TCRb^{Int-Hi}CD4^{-}CD8^{+}$). They initially express high levels of CD24 and CD69 (Stage C': $CD24^{Hi}CD69^{Hi}$) and then finally downregulate expression of these markers (Stage D': $CD24^{Lo}CD69^{Lo}$) before egressing the thymus as mature CD8 T-cells (Figure 4.16a) (97-99). Comparison of sequential stages of SP^{CD8} maturation within cKO and WT thymi revealed no significant differences (Figure 4.16a and b), indicating that following positive selection and commitment to the SP^{CD8} lineage, maturation of SP^{CD8} thymocytes in the context of TEC conditional CHD4 deficiency is grossly normal. The small reduction in the frequency of SP^{CD8} thymocytes in cKO thymi (Figure 4.13a) may therefore be due to reduced positive selection (Figure 4.14).

4.6.4 TEC conditional CHD4 deficiency causes an increase in wave 1 and 2 negatively selected thymocytes

Expression of the transcription factor Helios can be used to identify two waves of thymocytes undergoing negative selection (86). The first wave of negative selection occurs in the thymic cortex and concerns $TCR^{+}CCR7^{-}$ thymocytes that concurrently express PD-1 and Helios in response to negatively selecting

signals (86). A second wave of negative selection can be identified in Helios⁺ medullary SPCD4 thymocytes that express CCR7 and CD69 (86). Using these phenotypic features to assess negative selection, an increased frequency of wave 1 Helios⁺PD1⁺ negatively selected cells was detected in thymi composed of CHD4 deficient TEC when compared to WT ($0.30 \pm 0.03\%$ vs. $0.17 \pm 0.04\%$), indicating either increased cortical (i.e. wave 1) negative selection or decreased clearance of thymocytes undergoing negative selection within the cortex (Figure 4.17a).

A statistically significant but smaller increase in wave 2 Helios⁺ cells was observed for SPCD4 thymocytes localised in the medulla (i.e. CCR7⁺) in cKO compared to WT ($8.8 \pm 0.8\%$ vs. $7.2 \pm 0.7\%$) (Figure 4.17b). In agreement with the analysis of SPCD4 post-selectional maturation presented above (Figure 4.15), the current experiment also showed a relative deficiency of mature CD69⁺ SPCD4 thymocytes within the cKO thymi compared to WT ($16.3 \pm 2.1\%$ vs. $22.2 \pm 3.8\%$) (Figure 4.17c). Taken together, these findings could be consistent with an increased extent of negative selection of SPCD4 thymocytes within the medulla. Increased wave 1 and wave 2 negative selection may, in part, be explained by the relative increases in cTEC and MHCII^{Hi}CD80^{Hi} mTEC (Figure 4.5), respectively, and could contribute to the deficiencies in the generation of mature SP thymocytes, particularly within the SPCD4 lineage, observed in the context of CHD4 deficiency in TEC.

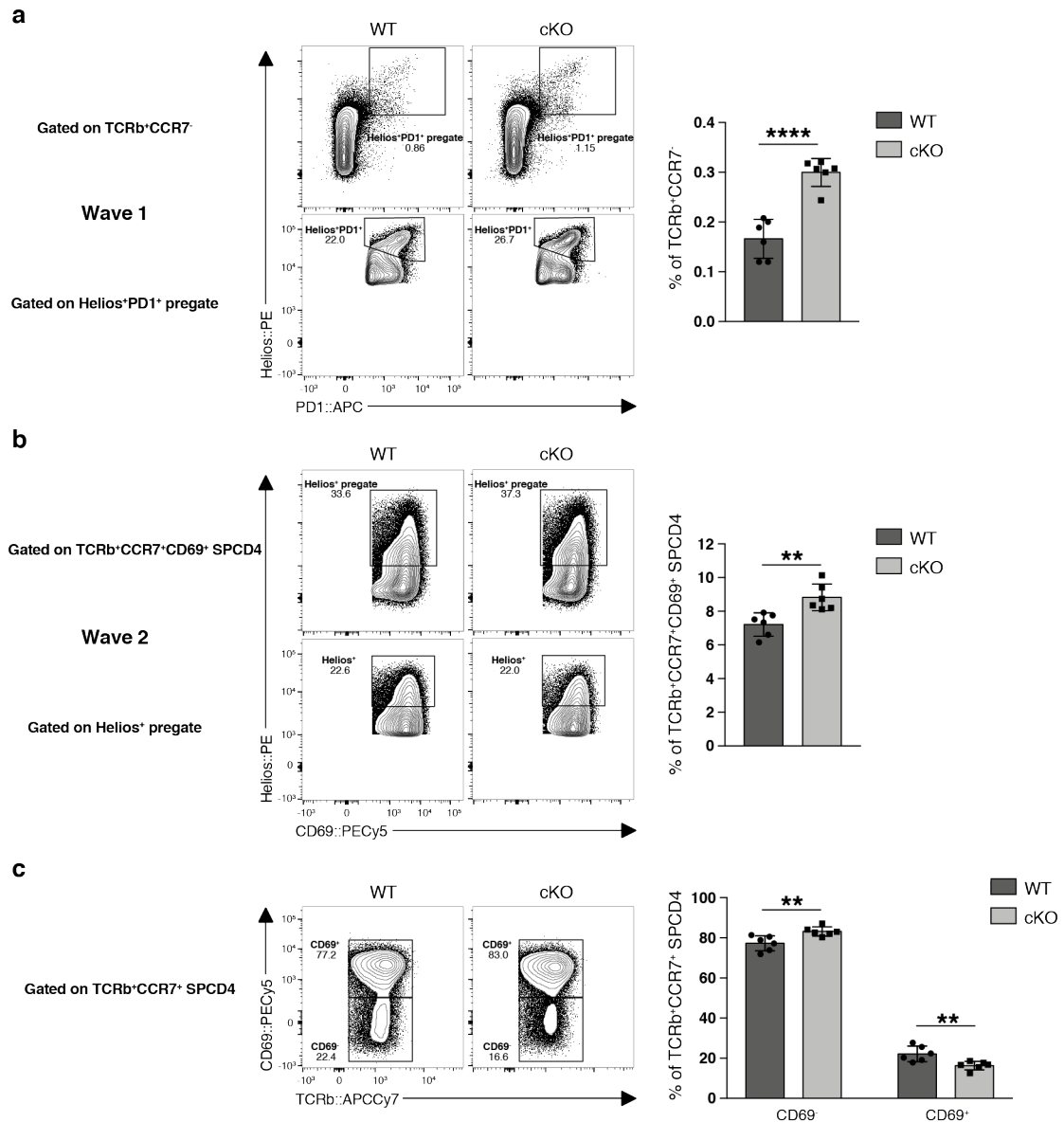


Figure 4.17 Increased wave 1 and 2 negative selection by CHD4 deficient TEC

(a-c) Representative FACS plots (left panels) and bar graphs (right panels) showing relative frequencies of thymocytes undergoing negative selection in cKO vs. WT. (a) Relative frequency of Helios⁺PD1⁺ cells within cortical thymocytes (FoxP3⁺CD25⁺TCRb⁺CCR7⁺) undergoing wave 1 of negative selection. (b) Relative frequency of Helios⁺ cells within medullary SPCD4 thymocytes (FoxP3⁺CD25⁺TCRb⁺CCR7⁺CD4⁺CD8⁺CD69⁺) undergoing wave 2 of negative selection. (c) Relative frequencies of CD69⁺ immature and CD69⁻ mature SPCD4 thymocytes within medullary SPCD4 thymocytes. Bar graphs contain data pooled from 2 independent experiments with 3 replicates each using 4 week old mice.

4.6.5 TEC conditional CHD4 deficiency causes peripheral T-cell lymphopaenia with skewing towards a memory phenotype

The peripheral T-cell compartment was next assessed in cKO and WT mice. Whole spleens were used in order to allow an assessment of absolute cellularity. In line with the thymic hypocellularity and reduced generation of SP thymocytes in cKO mice, these animals also displayed a reduction in splenic T-cells when compared to WT animals ($3.3 \pm 0.7 \times 10^6$ vs. $8.7 \pm 0.9 \times 10^6$ in WT controls (Figure 4.18a). This T-cell lymphopaenia concerned peripheral CD4⁺ and CD8⁺ T cells, although the former was more severely affected (Figure 4.18b). This result is consistent with the observation that maturation among post-selectional SP-CD4 stages was particularly affected (Figure 4.15 and 4.16) resulting in a significant reduction in the frequency of the most mature SP-CD4 thymocyte stages (Figure 4.15 and 4.17c). Within both T-cell splenic subsets in the CHD4 cKO mice there was a skewing towards a memory phenotype (Figure 4.18c and d), which is likely to be the result of homeostatic expansion driven by T-cell lymphopaenia (320-322).

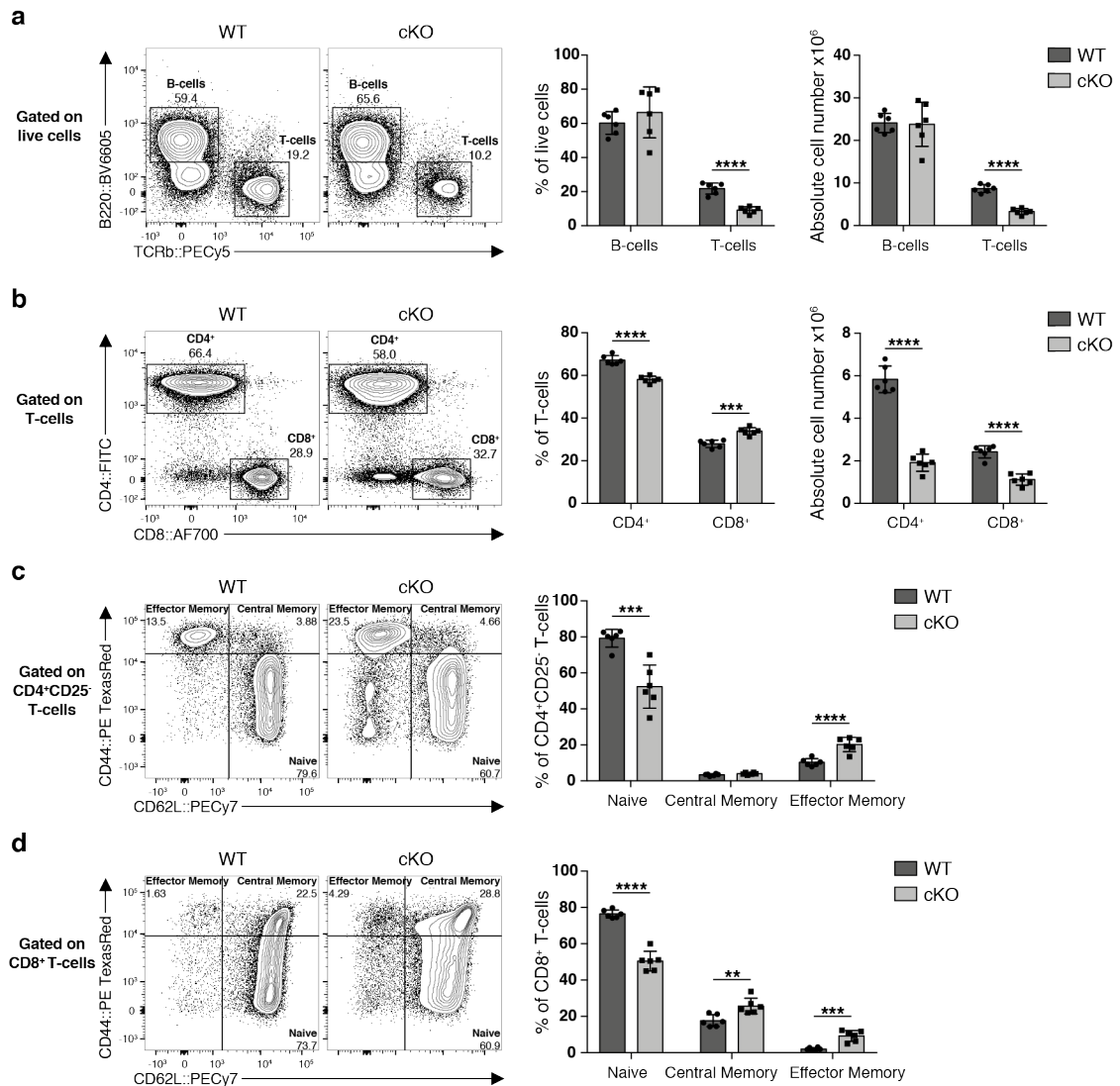


Figure 4.18 TEC conditional CHD4 deficiency causes peripheral T-cell lymphopaenia with skewing towards a memory phenotype

(a-b) Representative FACS plots (left panels) and bar graphs (right panels) showing relative frequencies and absolute cell counts of splenic (a) B-cells (B220⁺TCRb⁺) and T-cells (B220⁺TCRb⁺) and (b) CD4⁺ and CD8⁺ T-cells. (c-d) Representative FACS plots (left panels) and bar graphs (right panels) showing relative frequencies only of naïve (CD44⁺CD62L⁺), central memory (CD44⁺CD62L⁺) and effector memory (CD44⁺CD62L⁻) cells within (c) CD4⁺CD25⁺ T-cells and (d) CD8⁺ T-cells. Bar graphs contain data pooled from 2 independent experiments with 3 replicates each using 4 week old mice.

4.6.6 TEC conditional CHD4 deficiency causes peripheral expansion of T-regulatory cells

Thymic and peripheral T-regulatory cells (T-Regs) were next assessed in both thymic lobes and spleens isolated from 4 week old cKO and WT mice (Figure 4.19).

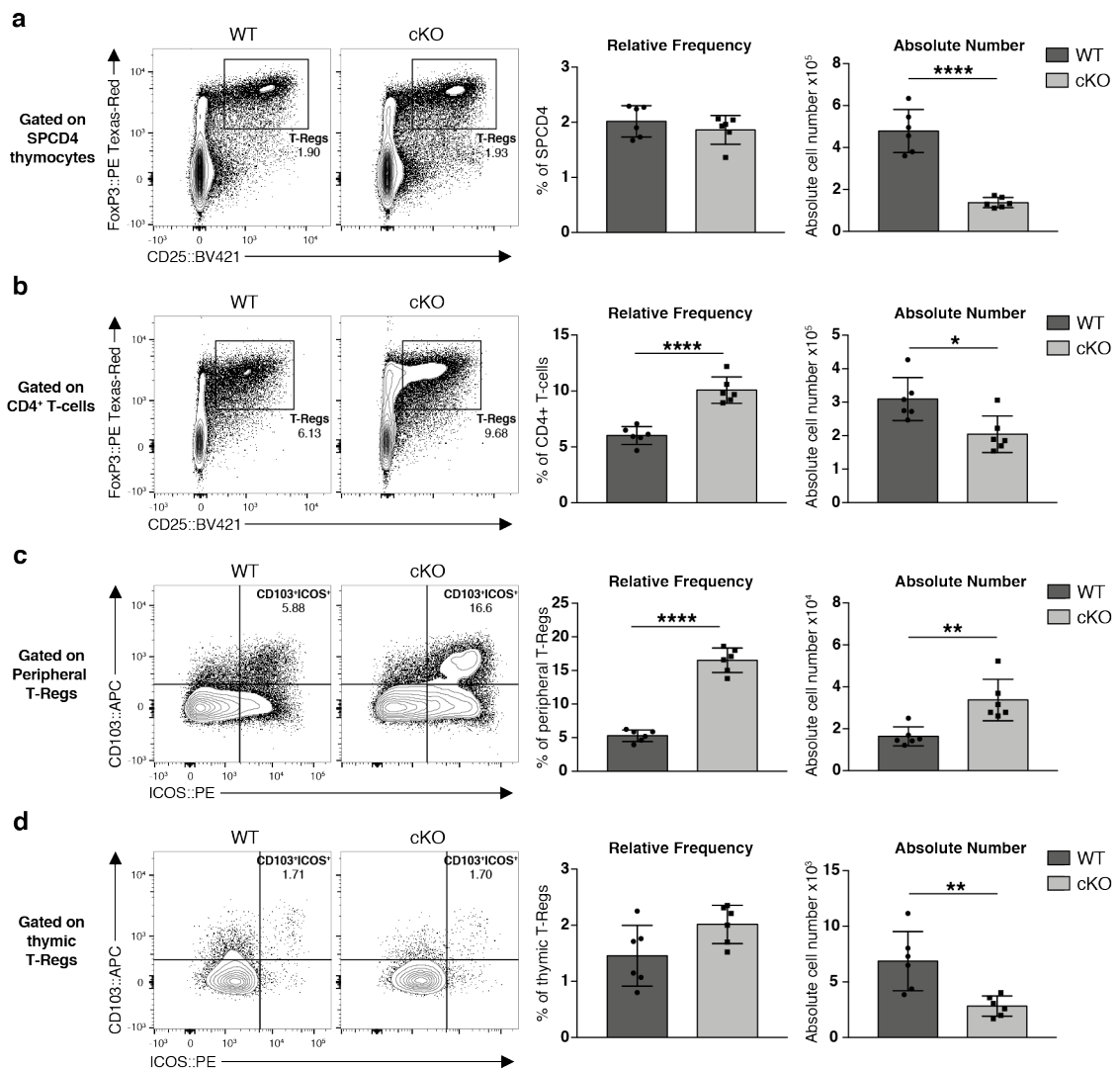


Figure 4.19 TEC conditional CHD4 deficiency causes peripheral expansion of T-Regs

(a-d) Representative FACS plots (left panels) and bar graphs (right panels) showing relative frequencies and absolute cell counts of (a) thymic T-Regs (FoxP3⁺CD25⁺) within the SPCD4 compartment, (b) splenic T-Regs (FoxP3⁺CD25⁺) within CD4⁺ T-cells, and (c) splenic and (d) thymic T-Regs. Bar graphs contain data pooled from 2 independent experiments with 3 replicates each using 4 week old mice.

Although the frequency of thymic T-Regs, defined by the co-expression of FoxP3 and CD25, within the SP^{CD4} compartment was comparable between cKO and WT, there was an absolute deficiency in the generation of thymic T-Regs in thymi composed of CHD4 deficient TEC ($1.4 \pm 0.2 \times 10^5$ vs. $4.8 \pm 1.0 \times 10^5$) (Figure 4.19a). In the periphery, however, there was an increased relative frequency of T-Regs within the CD4⁺ T-cell compartment in cKO compared to WT ($10.1 \pm 1.2\%$ vs. $6.0 \pm 0.8\%$) (Figure 4.19b). This was accounted for by an expansion of CD103⁺ICOS⁺ cells within the peripheral T-Reg compartment (Figure 4.19c) in both relative ($16.5 \pm 1.8\%$ vs. $5.3 \pm 0.9\%$) and absolute cellularity ($3.4 \pm 1.0 \times 10^5$ vs. $1.6 \pm 0.5 \times 10^5$). T-Regs that are double positive for CD103 and ICOS have previously been shown to be peripherally generated in the context of lymphopaenia and to have superior suppressive effector functions (323). Although, CD103⁺ICOS⁺ T-Regs have been shown to recirculate back to the thymus (323), this population of T-Regs was not overrepresented in the thymi of CHD4 cKO mice (Figure 4.19d)

In summary, the thymocyte and peripheral T-cell analyses presented above have shown that the abnormal thymic microenvironment caused by TEC conditional CHD4 deficiency leads to defective thymocyte maturation, particularly affecting the generation of SP from DP thymocytes and maturation within the committed SP^{CD4} lineage; as well as abnormal thymocyte selection. This translates in the periphery to T-cell lymphopaenia, with evidence for lymphopaenia-induced proliferation of CD4⁺CD25⁻ and CD8⁺ T-cells, as well as of T-Regs, as evidenced by an increased frequency of CD103⁺ICOS⁺ T-Regs.

4.7 CHD4 deficiency causes dysregulated expression of genes belonging to several pathways important for TEC development and function

To identify mechanisms accounting for the observed defects in TEC development consequent to the absence of CHD4, population level (i.e. bulk) RNA-sequencing (RNA-seq) of cKO and WT TEC was performed. Because phenotyping embryonic and postnatal cKO TEC suggested abnormal cell cycling within the cTEC compartment and partial blocks in mTEC development and maturation (Section 4.5), individual, well defined TEC subpopulations were selected for population level RNA-seq analyses.

4.7.1 Generation of population level RNA-sequencing libraries

Whole transcriptome RNA-sequencing data was generated from cTEC, CD80⁻CD86⁻ mTEC, CD80⁺CD86⁻ mTEC, and CD80⁺CD86⁺ mTEC from 4-5 week old female cKO and WT mice (3 biological replicates, 2,000 cells per replicate) using a SMARTer amplification protocol and paired-end 2x75bp sequencing of cDNA libraries.

Analysis of RNA-seq data was performed in collaboration with Dr Jeanette Baran-Gale and Professor Chris Ponting (MRC Institute of Genetics and Molecular Medicine, University of Edinburgh). Libraries were of high quality with 23-50 million reads per sample. Reads were first trimmed to remove SMARTer adapters, then aligned to the mm10 genome using the STAR aligner (265). Gene counts were quantified using Salmon (266) and finally normalised

gene expression was computed using DESeq2 (267). Genes were considered to be differentially expressed in cKO compared to WT samples if the fold change was $\geq \pm 2$ with a p value of ≤ 0.05 .

First, principal component analysis (PCA) showed high concordance between biological replicates, which clustered tightly together (Figure 4.20a). The largest source of variation within the data appeared to align with the TEC subpopulation analysed. However, a secondary source of variation was CHD4 status as cKO and WT samples clustered away from each other for each TEC subpopulation analysed.

The expression of housekeeping genes and genes that are known to differentiate the four different TEC subpopulations sequenced (i.e. cTEC, CD80⁻CD86⁻ mTEC, CD80⁺CD86⁻ mTEC, and CD80⁺CD86⁺ mTEC) were next examined as a further measure of the quality of the data generated (Figure 4.20b). There was no differential expression of several commonly used housekeeping genes (*Gapdh*, *Actb*, *Ppia*, *Eef2*, *Rpl38*) in cKO TEC compared to their WT counterparts.

Examination of known makers of the four TEC subpopulations sequenced showed the expected expression profile. The expression of *Psm11* (encoding $\beta 5t$), *Enpep* (encoding Ly51), *Dll4*, and *Ly75* (encoding CD205) were highest in cTEC samples; and there was no differential expression of these markers between cKO and WT cTEC (Figure 4.20b). mTEC were marked by

expression of the TNFSF receptors *Tnfrsf11a* (encoding RANK) and *Cd40*, with no differences in expression levels between cKO and WT mTEC (Figure 4.20b). The expression of *Cd80* and *Aire* was highest in CD80⁺CD86⁻ mTEC and CD80⁺CD86⁺ mTEC, with no differences between cKO and WT within these subsets (Figure 4.20b). Expression of *Cd86* was highest in CD80⁺CD86⁺ mTEC, with no difference between cKO and WT. Lower levels of *Cd86* mRNA were detected in CD80⁺CD86⁻ mTEC, presumably reflecting a lag between transcript and protein expression during mTEC maturation. *Cd86* expression was downregulated in cKO compared to WT CD80⁺CD86⁻ mTEC, which, in keeping with the FACS phenotyping analyses presented above (Section 4.5), may be reflective of defective maturation of cKO mTEC from CD80⁺CD86⁻ to CD80⁺CD86⁺ mTEC (Figure 4.20b). Whilst the expression of the master transcriptional regulator of TEC development, *Foxn1*, was appropriately detected in all the TEC subpopulations, its expression level was significantly higher in cTEC, CD80⁻CD86⁻ mTEC, and CD80⁺CD86⁺ mTEC derived from cKO compared to WT animals (Figure 4.20b). The increased *Foxn1* expression in cKO TEC was not, however, associated with consistently increased expression of FOXN1 target genes (Figure 4.20c) (36).

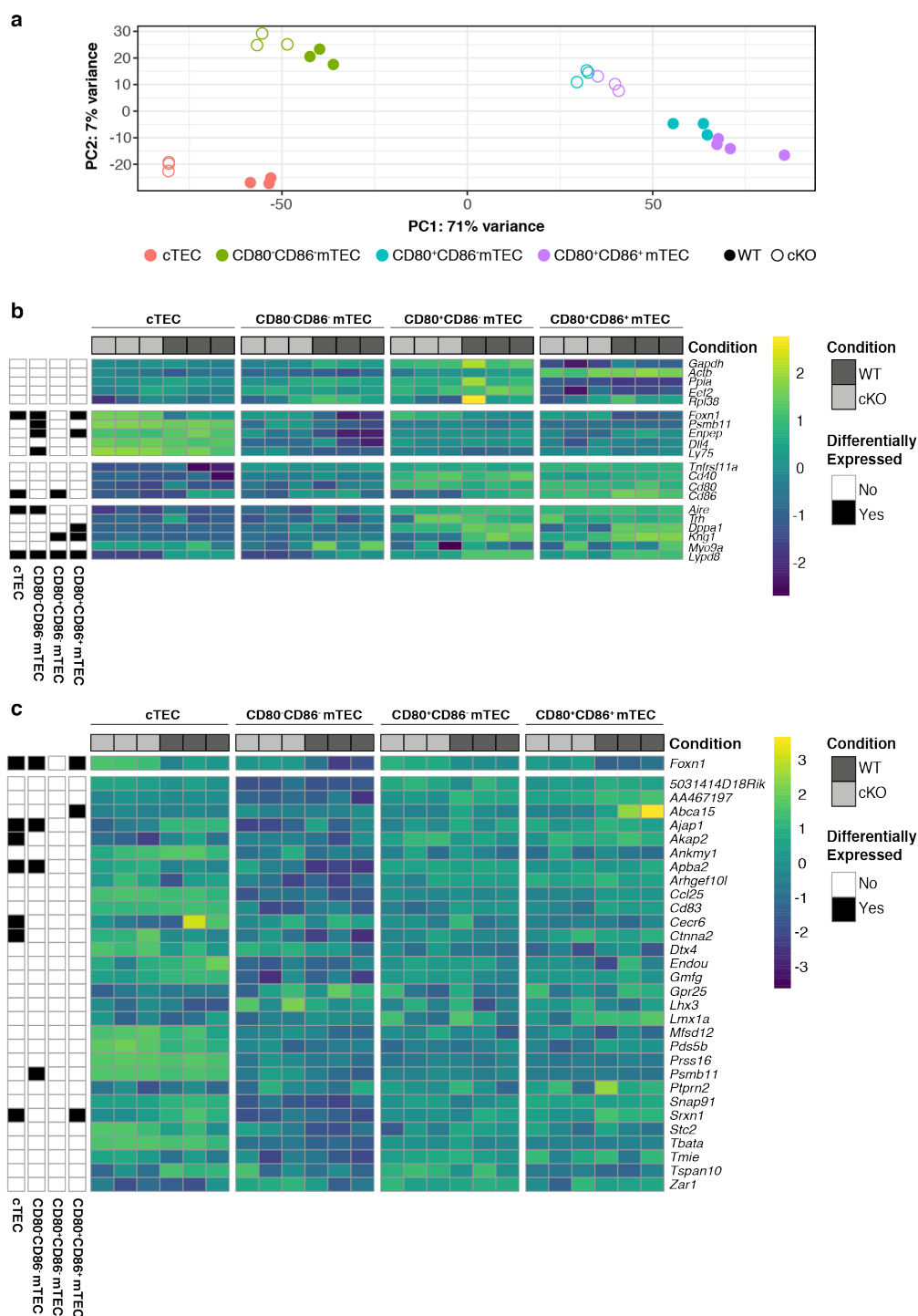


Figure 4.20 Population level RNA-sequencing of CHD4 deficient and WT TEC subpopulations

(a) Principal component analyses (PCA) of cKO (open circles) and WT (filled circles) RNA-seq samples, coloured according to TEC subpopulation. (b) Heatmap showing the expression of selected housekeeping genes, and genes typically expressed in TEC in cKO and WT RNA-seq samples; differential expression (filled black boxes) is indicated on the left. (c) Heatmap showing the expression of FOXP1 target genes in cKO and WT RNA-seq samples, differential expression (filled black boxes) is indicated on the left. Normalised expression of each gene is shown as a z-score (yellow-blue) to highlight relative differences in gene expression regardless of overall expression level. Genes were considered to be differentially expressed if the fold change in gene expression was $\geq \pm 2$ with a p value of ≤ 0.05 . For each TEC subpopulation, RNA-seq was performed on 3 biological replicates from 4 week old female mice.

Finally the expression of several TRGs was examined. The AIRE-induced genes *Trh*, *Dppa1* and *Knq1* were detected in the more mature mTEC subsets (i.e. CD80⁺CD86⁻ and CD80⁺CD86⁺ mTEC), with *Dppa1* expression being downregulated in CD80⁺CD86⁺ mTEC and *Knq1* in both CD80⁺CD86⁻ and CD80⁺CD86⁺ mTEC from cKO compared to WT mice. The AIRE-independent TRGs *Myo9a* and *Lypd8* on the other hand, were detectable in all TEC subsets, with only *Lypd8* being downregulated in all cKO TEC subpopulations compared to their WT counterparts (Figure 4.20b)

In summary, these data describe the differences in gene expression between cKO and WT cTEC, CD80⁻CD86⁻ mTEC, CD80⁺CD86⁻ mTEC, and CD80⁺CD86⁺ mTEC. The data were rich (23-50 million read-depth), biological replicates were highly concordant with each other as indicated by PCA analysis and the expression of a number of genes known to differentiate the four different TEC subpopulations sequenced segregated appropriately with cell type.

4.7.2 Deficiency of CHD4 in TEC causes dysregulated gene expression in all TEC subpopulations examined

Overall differential expression analysis returned 696 up- and 878 down-regulated genes in cTEC; 550 up- and 466 down-regulated genes in CD80⁻CD86⁻ mTEC; 385 up- and 513 down-regulated genes in CD80⁺CD86⁻ mTEC; and 540 up- and 1,259 down-regulated genes in CD80⁺CD86⁺ mTEC (Figure

4.21a-d). In the paragraphs that follow, some of the most differentially expressed genes for each TEC subpopulation analysed will be discussed.

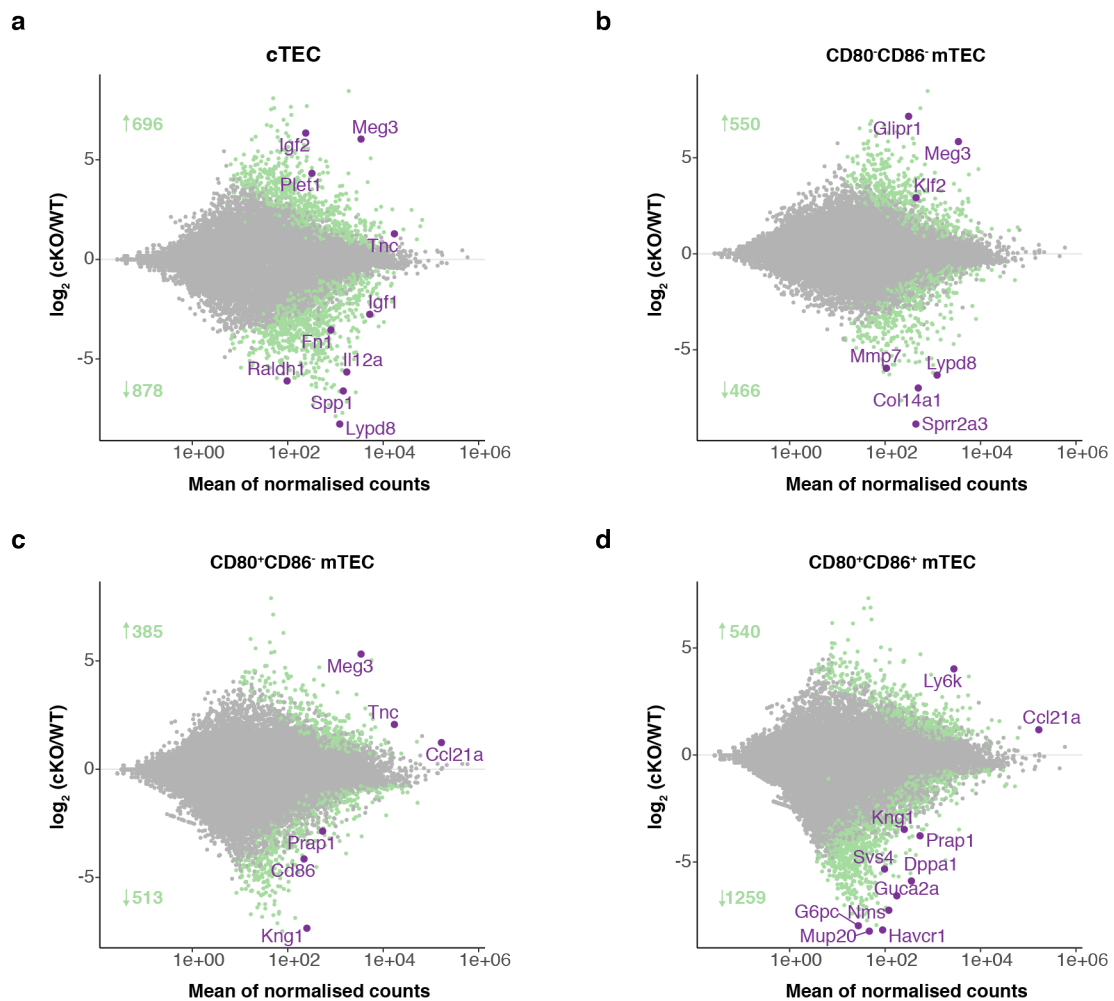


Figure 4.21 Dysregulated gene expression in CHD4 deficient TEC

(a-d) MA plots showing differential gene expression between cKO and WT population level RNA-seq samples for (a) cTEC, (b) CD80⁺CD86⁻ mTEC, (c) CD80⁺CD86⁻ mTEC, and (d) CD80⁺CD86⁺ mTEC. Each dot represents a gene and those relating to significantly differentially expressed genes are coloured green. Genes were considered to be differentially expressed if the fold change in gene expression was $\geq \pm 2$ with a p value of ≤ 0.05 . Examples of the most differentially expressed genes are annotated with gene symbols in purple. For each TEC subpopulation, RNA-seq was performed on 3 biological replicates from 4 week old female mice.

Amongst the most upregulated genes in cKO vs. WT cTEC were insulin like growth factor-2 (*Igf2*), Maternally expressed 3 (*Meg3*), Placenta Expressed Transcript 1 (*Plet1*) and Tenascin c (*Tnc*) (Figure 4.21a). Whilst *Igf2* expression was significantly upregulated in cKO cTEC, *Igf1* was significantly downregulated (Figure 4.21a). The latter is of potential interest with regards TEC conditional CHD4 deficiency as IGF-1 is known to play a role in thymic growth and recovery by promoting TEC cellularity and thymopoiesis (324, 325). *Meg3* was amongst the most upregulated genes in cKO vs. WT cTEC, CD80⁻CD86⁻ mTEC and CD80⁺CD86⁻ mTEC (Figure 4.21a, b and c). *Meg3* is a long noncoding RNA, which functions as a tumour suppressor and has been shown to promote the degradation of β -catenin in liver cancer cells (326). Precise regulation of Wnt/ β -catenin signalling has previously been shown to be important for normal thymic development and function. Constitutive activation of Wnt/ β -catenin signalling pathway in TEC causes defective thymic organogenesis, TEC proliferation and thymocyte development (261). On the other hand induced ablation of Wnt/ β -catenin signalling in adult K5⁺ TEC results in thymic atrophy with an overrepresentation of cTEC and a relative deficiency of mTEC (327), reminiscent of the phenotype seen in the CHD4 cKO. PLET1 has been previously identified as a marker associated with bipotent progenitor TEC in the fetal and postnatal thymus (48, 49). Therefore, its upregulation in cKO cTEC are consistent with the flow cytometry data presented above showing evidence of increased expression of progenitor markers in cKO TEC (Subsection 4.5.5 and Figure 4.9). *Tnc* encodes an extracellular matrix (ECM) glycoprotein, which negatively regulates cell

adhesion by binding pro-adhesive Fibronectin (*Fn1*) and preventing its interaction with syndecan-4 (328). Additionally, *Fn1* was downregulated in cKO cTEC compared to WT (Figure 4.21a). These findings together with the observation of distorted thymic architecture with large epithelial-cell free areas in CHD4 cKO thymi (Figure 4.12) make ECM homeostasis and organisation pathways of particular interest in the context of TEC conditional CHD4 deficiency. Alongside *Igf1* and *Fn1*, the most downregulated genes in cKO vs. WT cTEC included *Il12a*, retinaldehyde dehydrogenase 1 (*Raldh1*), secreted phosphoprotein 1 (*Spp1*), and *Lypd8* (Figure 4.21a). *Il12a* encodes the p35 subunit of IL12, which has been shown to inhibit thymic involution (329). *Spp1*, which is also downregulated in cKO cTEC, encodes a glycoprotein that functions both as an ECM constituent and as a cytokine that enhances the expression of Th1 cytokines including IL12 (330). *Raldh1* encodes an aldehyde dehydrogenase enzyme that converts retinaldehyde to retinoic acid. Fine balance of retinoic acid homeostasis is important for the development and function of thymus with both deficiency and excess being implicated in thymic pathology (310, 331-334). Finally, *Lypd8*, an AIRE-independent TRG, is downregulated in all cKO TEC subpopulations tested (Figure 4.20b) and is amongst the most downregulated genes in cKO cTEC and CD80⁻CD86⁻ mTEC (Figure 4.21a and b).

In addition to *Meg3*, *Glipr1* and Kruppel-like factor 2 (*Klf2*), are examples of the most upregulated genes in cKO vs. WT CD80⁻CD86⁻ mTEC (Figure 4.21b). GLIPR1 functions as a tumour suppressor (335) and KLF2 is a transcription factor with diverse biological functions including cell growth and differentiation,

maintaining epithelial integrity and inflammation (336). *Lypd8* (see above), matrix metalloproteinase-7 (*Mmp7*), *Col14a1*, and *Spr2a3* were particularly downregulated in cKO CD80⁻CD86⁻ mTEC (Figure 4.21b). Both *Mmp7* and *Col14a1* function in ECM homeostasis, which is of particular interest in the CHD4 cKO as discussed above, and *Spr2a3* encodes a small proline-rich protein that is expressed during epithelial cell terminal differentiation (337).

Genes that were particularly upregulated in cKO CD80⁺CD86⁻ mTEC included *Meg3* and *Tnc* (discussed above), as well as *Ccl21a*, which was also amongst the most upregulated genes in cKO CD80⁺CD86⁺ mTEC. CCL21 is a chemokine expressed by mTEC and is responsible, alongside CCL19, for the attraction of single positive thymocytes bearing CCR7 from the thymic cortex to the medulla (9, 11, 94). *Ccl21* has also been identified as a marker gene of junctional TEC, which reside at the cortico-medullary junction, express podoplanin and have precursor potential for the mTEC lineage. Interestingly podoplanin expression was also found to be upregulated at the protein level on cKO mTEC (Subsection 4.5.5 and Figure 4.9). Taken together with the data presented previously (Section 4.4 and 4.5) showing an absolute and relative deficiency in the generation of mature mTEC, these results could suggest maybe in keeping with CHD4 deficient TEC being impaired in their ability to differentiate from a progenitor state to more mature stages (Figure 4.10). The most downregulated genes in cKO CD80⁺CD86⁻ mTEC included *Cd86* which, in keeping with the FACS phenotyping analyses presented above (Section 4.5), may be reflective of defective maturation of cKO mTEC from CD80⁺CD86⁻

to CD80⁺CD86⁺ mTEC, as well as *Prap1* and *Knk1*, which are both AIRE-enhanced TRGs, suggesting that cKO mTEC may have a compromised capacity for AIRE-induced PGE.

The most upregulated genes in cKO vs. WT CD80⁺CD86⁺ mTEC included *Ccl21a* (discussed above), and *Ly6k*, an AIRE-independent TRG. The most downregulated genes in cKO CD80⁺CD86⁺ mTEC were almost exclusively TRGs induced by AIRE, further demonstrating that CHD4 may indeed play a role in AIRE-regulated PGE as hypothesised (discussed further in Section 4.8 below). *Svs4*, *Dppa1*, and *Nms* are all AIRE-dependent TRGs; *Prap1*, *Knk1*, *Guca2a*, *G6pc*, and *Mup20* are AIRE-enhanced TRGs and *Havcr1* is an AIRE-independent TRG.

In the sections that follow, further analyses of the RNA-seq data relating to cTEC and CD80⁺CD86⁺ mTEC are presented alongside experimental data serving to verify and functionally probe potentially interesting findings relating to the role of CHD4 in TEC development and PGE, respectively. Further analysis, validation and functional interrogation of the RNA-seq data related to CD80⁻CD86⁻ mTEC and CD80⁺CD86⁻ mTEC were beyond the scope of this project and will form part of future studies and as such will not be discussed further within this thesis.

4.7.3 CHD4 deficient cTEC show dysregulated expression of genes related to retinoic acid metabolism, insulin-like growth factor signalling, and extracellular matrix homeostasis

The data presented in Sections 4.4 and 4.5 were suggestive of partial blocks in TEC development at multiple points during the developmental trajectory including the cTEC to mTEC transition (Figure 4.10). Therefore, the differentially expressed genes in cKO vs. WT cTEC were first probed with the aim of identifying mechanisms that could account for the defects in TEC development observed in the context of CHD4 deficiency. To this end genes found to be significantly upregulated (Figure 4.22a) and downregulated (Figure 4.22b) in cKO compared to WT cTEC were assessed for enrichment of particular biological pathways or processes (298).

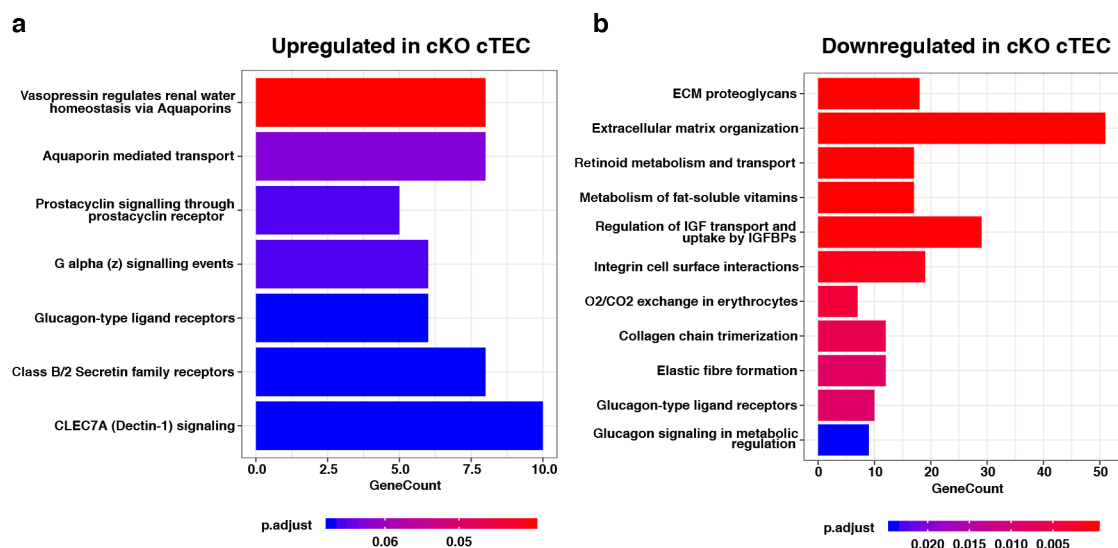


Figure 4.22 Pathway analysis of differentially expressed genes in CHD4 deficient vs. WT cTEC

(a-b) Bar plots visualizing Reactome (298) pathway analysis of genes differentially expressed between cKO and WT cTEC for (a) genes upregulated in cKO cTEC and (b) genes downregulated in cKO cTEC. Bars are coloured according to the adjusted p-value, which denotes significance level for pathway enrichment, and GeneCount on the x-axis shows the number of differentially expressed genes overlapping the pathway of interest. Only significantly differentially expressed genes were considered for the enrichment analysis i.e. fold change $\geq \pm 2$ with an adjusted p value of ≤ 0.05 . RNA-seq was performed on 3 biological replicates from 4 week old female mice.

The only biological pathway/process found to be significantly associated with genes upregulated in CHD4 deficient cTEC was that related to renal water homeostasis and therefore was not considered further (Figure 4.22a). However, the pathway analysis for genes downregulated in cKO cTEC returned several pathways with known relevance in TEC biology that warrant further consideration for their potential contribution to the TEC developmental and structural defects observed in the context of TEC conditional CHD4 deficiency. These included terms relating to ECM organisation and function, retinoid metabolism and the IGF pathway and are discussed further in the sections that follow (Figure 4.22b).

4.7.4 CHD4 deficiency in TEC causes abnormal composition of the thymic extracellular matrix

As mentioned, the pathway analysis for genes downregulated in cKO cTEC returned several terms related to ECM organisation and function. These included ECM proteoglycans, ECM organisation, integrin cell surface interactions, collagen chain trimerisation and elastic fibre formation (Figure 4.22b). ECM components, such as collagens, laminins, fibronectins and integrins, form part of the thymic microenvironment and play roles in the structure, organisation and function of the thymus (338).

In order to further investigate ECM organisation in the CHD4 cKO thymus, IHC was performed on cKO and WT thymic sections for two ECM constituents

differentially expressed by cKO TEC in the RNA-seq analysis, namely fibronectin (FN1) and tenascin C (TNC).

The mRNA expression level of *Fn1* was significantly downregulated in cKO cTEC, as well as CD80⁻CD86⁻ and CD80⁺CD86⁺ mTEC (Figure 4.23a). FN1 is a pro-adhesive constituent of the thymic ECM, produced by TEC, fibroblasts and intra-thymic APCs (339, 340). FN1 expression on TEC promotes TEC::thymocyte adhesion via interaction with its receptors VLA-4 and VLA-5 expressed on the surface of developing thymocytes. FN1 also promotes intrathymic migration of developing thymocytes, in part by binding chemokines such as CXCL12 and CCL19 (340, 341). Examination of FN1 expression via IHC confirmed the RNA-seq findings at the protein level with reduced expression of FN1 in cKO thymic sections (Figure 4.23b). Notably, there was also an even more striking deficiency of FN1 staining within the epithelial free zones of the cKO thymi.

Tenascin C (TNC) is an anti-adhesive ECM glycoprotein, which antagonises the function of FN1 (328). *Tnc* expression was upregulated at the level of transcription in all four cKO TEC subpopulations analysed (Figure 4.24a). Increased TNC expression at the protein level in cKO thymic sections was confirmed by IHC. Similar to FN1, TNC was notably absent from the epithelial cell free zones in the cKO sections (Figure 4.24b).

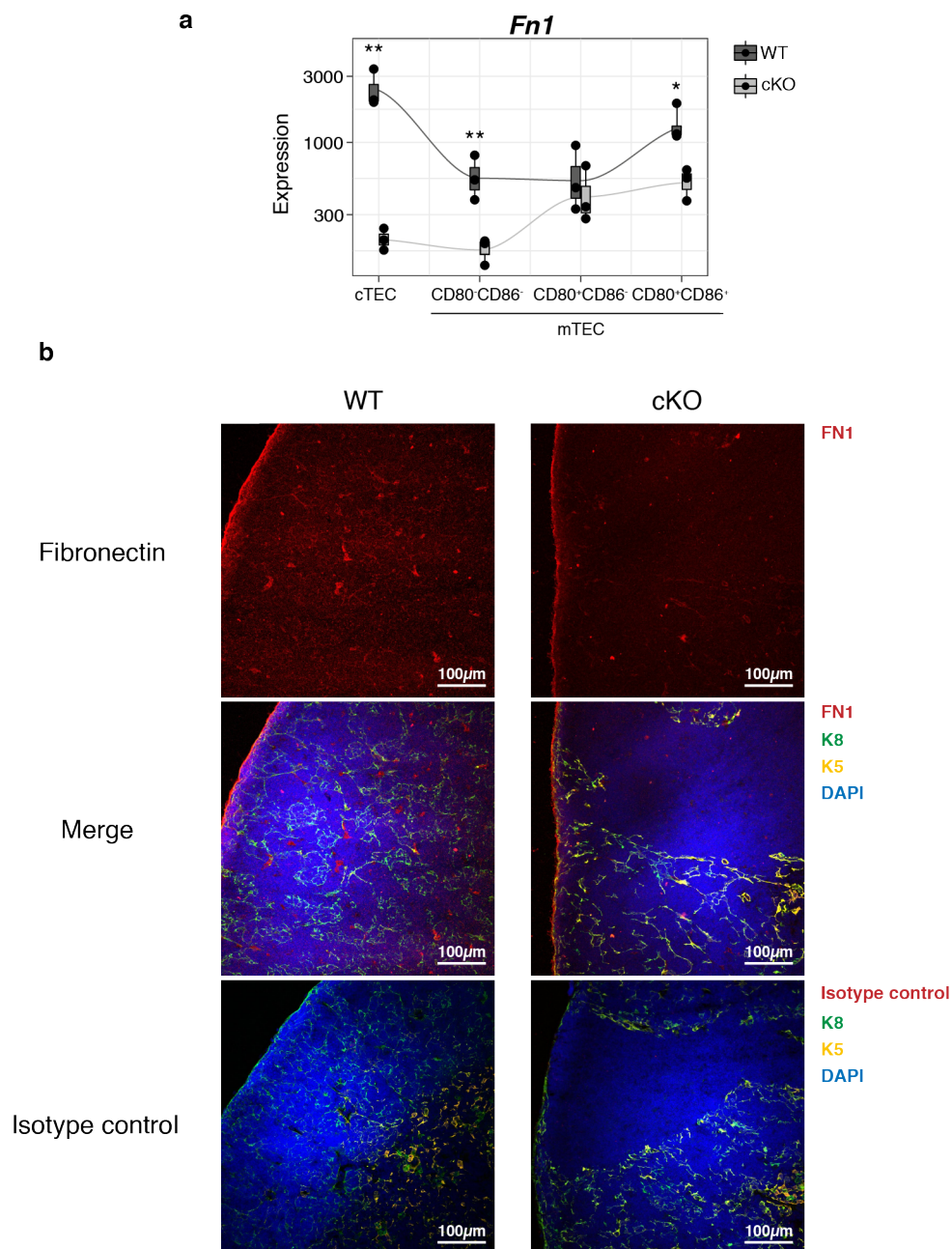


Figure 4.23 Expression of the ECM component fibronectin (FN1) is reduced in cKO thymi

(a) Box and whisker plot showing normalised expression of *Fn1* in cKO and WT RNA-seq TEC subpopulations. (b) Immunofluorescence microscopy for fibronectin (FN1: Red), cytokeratin 8 (K8: Green) and 5 (K5: Yellow) in sections from 4 week old cKO and WT thymi. Upper panel (labeled "Fibronectin") shows fibronectin/AF647 channel only; middle panel shows composite image (labeled "Merge") and lower panel shows staining pattern using an isotype control antibody in place of anti-fibronectin::AF647.

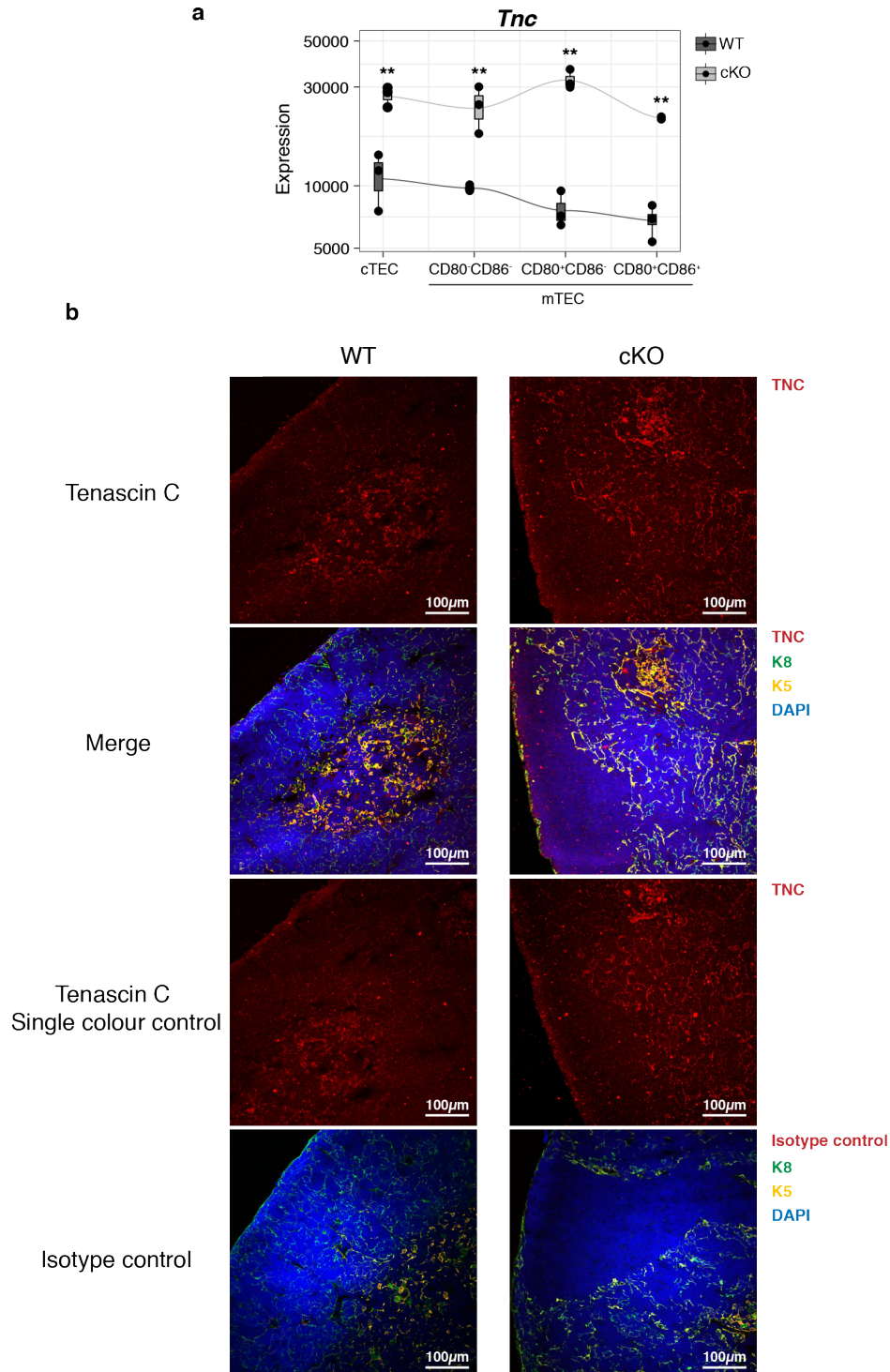


Figure 4.24 Expression of the ECM component tenascin C (TNC) is increased in cKO thymi

(a) Box and whisker plot showing normalised expression of *Tnc* in cKO and WT RNA-seq TEC subpopulations (b) Immunofluorescence microscopy for tenascin C (TNC: Red), cytokeratin 8 (K8: Green) and 5 (K5: Yellow) in sections from 4 week old cKO and WT thymi. Upper panel (labeled “Tenascin C”) shows TNC/AF647 channel only; second panel shows composite image (labeled “Merge”); third panel shows TNC single colour control, in which staining was performed only for TNC in order to demonstrate that the observed staining pattern was not due to fluorescence spill over from the K5 channel; and lower panel shows staining pattern using an isotype control antibody in place of anti-TNC::AF647.

In summary, CHD4 deficiency in TEC causes abnormal composition of the thymic ECM with reduced levels of the pro-adhesive FN1 and increased levels of the anti-adhesive TNC. Such a disruption of the thymic ECM could contribute to the abnormal architectural organisation, thymocyte-filled epithelial cell free zones, and consequently defective thymopoiesis observed in the CHD4 cKO thymus via impairment of effective FN1-mediated TEC::thymocyte interactions and thymocyte migration.

4.7.5 Exogenous retinoic acid partially rescues the phenotype caused by TEC conditional CHD4 deficiency

Pathway analysis of the downregulated genes in cKO cTEC found significant enrichment for genes involved in retinoid metabolism and transport (Figure 4.22b). Fine balance of retinoic acid (RA) homeostasis is important for the development and function of thymus with both deficiency and excess being implicated in thymic pathology, including thymic hypoplasia and atrophy (310, 331-334). Thymic epithelial cells have been shown to endogenously produce retinoids via the action of retinaldehyde dehydrogenase (RALDH) enzymes (334). Of the retinaldehyde dehydrogenase enzymes capable of converting retinaldehyde to RA, RALDH1 is produced by postnatal TEC, particularly cTEC, although its function in TEC has not been formally studied (334). A germline hypomorphic allele of *Raldh2* has been shown to result in a DiGeorge syndrome phenotype in mice, including thymic hypoplasia or atrophy, in a manner dependent on mesenchyme-derived RALDH2 with subsequent RA signalling in the pharyngeal endoderm (331, 342). Finally, *Raldh3* is expressed

in thymic mesenchymal cells as well as TEC, however, it has not been studied for a potential role in thymic organogenesis or biology (342). Although the role(s) of the RALDH enzymes have not been specifically studied within TEC, a mouse model in which RA signalling was conditionally abrogated in TEC via the expression of a dominant negative form of the RA receptor RAR α showed defective TEC development and thymopoiesis underscoring the importance of proper RA signalling in TEC (310). The absence of RA signalling in TEC caused reduced thymic cellularity, an expansion of the cTEC compartment and a reduction in mTEC, increased cTEC proliferation, increased SCA1⁺ cTEC, and a moderate reduction in generation of SPCD4 and SPCD8 thymocytes (310). This phenotype is highly reminiscent of that observed in the CHD4 cKO supporting the notion that CHD4 deficiency in TEC causes dysregulated retinoic acid signalling which negatively impacts on the cells' differentiation and functional ability to support thymopoiesis. Therefore, the role of impaired RA signalling in the CHD4 cKO was pursued.

Raldh1 was amongst the most downregulated genes in cTEC from cKO compared to WT thymi (Figure 4.21a and 4.25a). *Raldh2* mRNA was not detected at all in cKO cTEC RNA-seq samples and its expression was highly variable between WT biological replicates (Figure 4.25b). The expression of *Raldh3* was unchanged in cKO vs. WT cTEC (Figure 4.25c). In order to test the functional relevance of the changes in *Raldh* gene expression observed in the cTEC RNA-seq data the Aldefluor assay, a flow cytometric assay that measures total cellular aldehyde hydrogenase activity, was employed. In this

assay cells expressing high levels of aldehyde dehydrogenase become brightly fluorescent (designated “Aldefluor⁺”) with gates for background fluorescence being set on the basis of a control tube to which a specific inhibitor to aldehyde dehydrogenase, diethylaminobenzaldehyde (DEAB), is added. Using this assay differences were not discerned when comparing cKO and WT cTEC with regards to the frequency of Aldefluor⁺ cells (Figure 4.25d) and the fold change in Aldefluor gMFI in test vs. control samples (Figure 4.25e). This result suggested that total aldehyde dehydrogenase activity within cKO cTEC was comparable to normal.

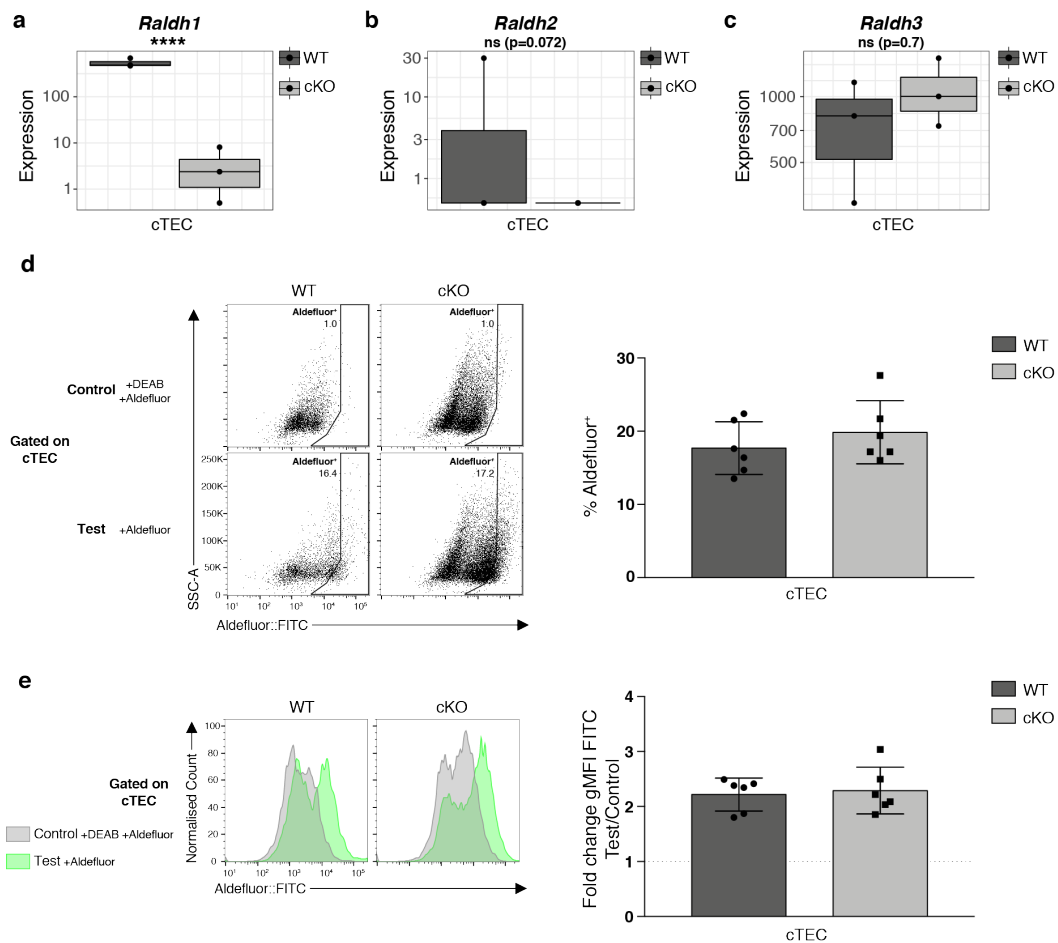


Figure 4.25 Aldehyde dehydrogenase expression and activity in CHD4 deficient TEC

(a-c) Box and whisker plots showing normalised expression of (a) *Raldh1*, (b) *Raldh2*, and (c) *Raldh3* in cKO and WT cTEC RNA-seq samples; ns= not significant (d-e) Representative FACS plots (left panels) and bar graphs (right panels) from Aldefluor assay in cKO and WT cTEC showing (d) frequency of Aldefluor⁺ cTEC and (e) fold change in gMFI of Aldefluor (FITC) in test vs. control tubes. Gates for Aldefluor positivity were set on the basis of a negative control tube containing diethylaminobenzaldehyde (DEAB). Bar graphs contain data pooled from 2 independent experiments with 3 replicates each using 4 week old mice.

The Aldefluor assay measures total cellular aldehyde dehydrogenase activity rather than the activity of specific enzymes, such as RALDH1. In addition to *Raldh1*, several other genes related to retinoid metabolism and transport had dysregulated expression in cKO TEC (Figure 4.22b). Therefore, the ability and extent to which cKO TEC were able respond to exogenous RA was tested in the context of FTOCs. To this end, cKO and WT E14.5 thymic lobes were first

depleted of lymphoid cells using 2-deoxyguanosine (2DG) as thymocytes express RA receptors and are thus able to respond to retinoid signalling independently of TEC via proliferation and/or apoptosis (334, 343, 344). Their inclusion may, therefore, confound the analysis of TEC intrinsic responses to exogenous RA. The analysis of lymphodepleted FTOCs showed that without RA treatment there were significantly fewer mTEC in cKO FTOCs compared to WT ($1.42 \pm 0.77\%$ vs. $4.42 \pm 1.25\%$) (Figure 4.26a). This is in keeping with the TEC phenotyping data presented above (Sections 4.5.1 and 4.5.4). However, after 10 days of culture with exogenous RA at a concentration of 10^{-6}M there was a significant increase in the proportion of mTEC within cKO FTOCs up to levels comparable with WT (Figure 4.26).

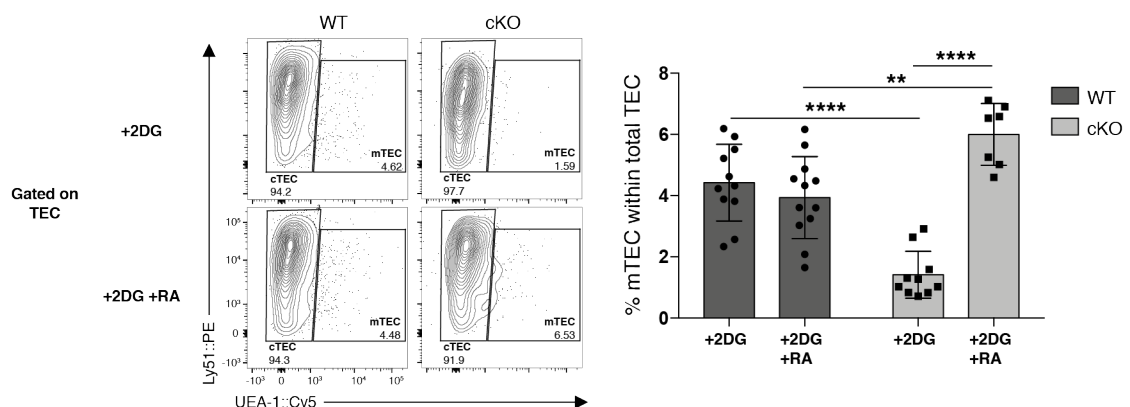


Figure 4.26 Addition of exogenous retinoic acid leads to an increased frequency of mTEC in CHD4 cKO FTOCs

Representative FACS plots (left panels) and bar graph (right panel) showing frequency of mTEC (UEA-1⁺) within total TEC (CD45⁺EpCAM⁺) in cKO and WT E14.5 FTOCs after lymphodepletion with 2-deoxyguanosine (2DG) +/- treatment with 10^{-6}M retinoic acid (RA). Bar graph contains data pooled from 2 independent experiments.

To test the impact of the relative increase in mTEC in FTOCs treated with RA (Figure 4.26) on thymocyte development, FTOCs with the addition of exogenous RA were next performed on E14.5 thymic lobes without prior

lymphodepletion with the limitation that any effects on thymocyte development seen could be due to the effect of RA on the thymocytes directly, or on TEC or on both cell types. The addition of 10^{-6} M RA to cKO and WT E14.5 FTOCs without prior lymphodepletion lead to a significant expansion of cells falling within the CD4⁺CD8⁺ DN FACS gate in both cKO ($12.22 \pm 3.10\%$ without vs. $43.07 \pm 13.80\%$ with RA) and WT FTOCs ($9.34 \pm 1.82\%$ without vs. $85.51 \pm 14.15\%$ with RA), although to a greater extent in the latter (Figure 4.27a). The majority of the cells falling within the CD4⁺CD8⁺ DN gate after culture with RA were CD69⁺TCR β ⁺ (Figure 4.27b). These cells are therefore unlikely to be thymocytes as CD69 expression on developing thymocytes is usually first seen after positively selecting signals, which are TCR dependent (303, 305). CD69 is a transmembrane C-type lectin whose expression is induced upon activation of most leucocytes, not just cells of the $\alpha\beta$ T-lymphocyte lineage, and it is also a constitutively expressed on tissue resident memory T-cells and platelets (345, 346). Therefore, it is likely that the expanded cells within CD4⁺CD8⁺ DN gate represent non- $\alpha\beta$ T-lymphocyte lineage cells that were not removed in the preceding gate, which was used with the intention to “electronically” remove non- $\alpha\beta$ T-lymphocyte lineage cells, specifically those expressing lineage markers for myeloid cells (CD11c, F4/80, CD11b, Gr1), B lymphoid cells (CD19), NK and NKT cells (CD49b and NK1.1), $\gamma\delta$ T-cells ($\gamma\delta$ TCR) and erythroid cells (Ter119) (see Section 4.3). On the basis of this, CD4⁺CD8⁺ DN cells were removed from subsequent analyses.

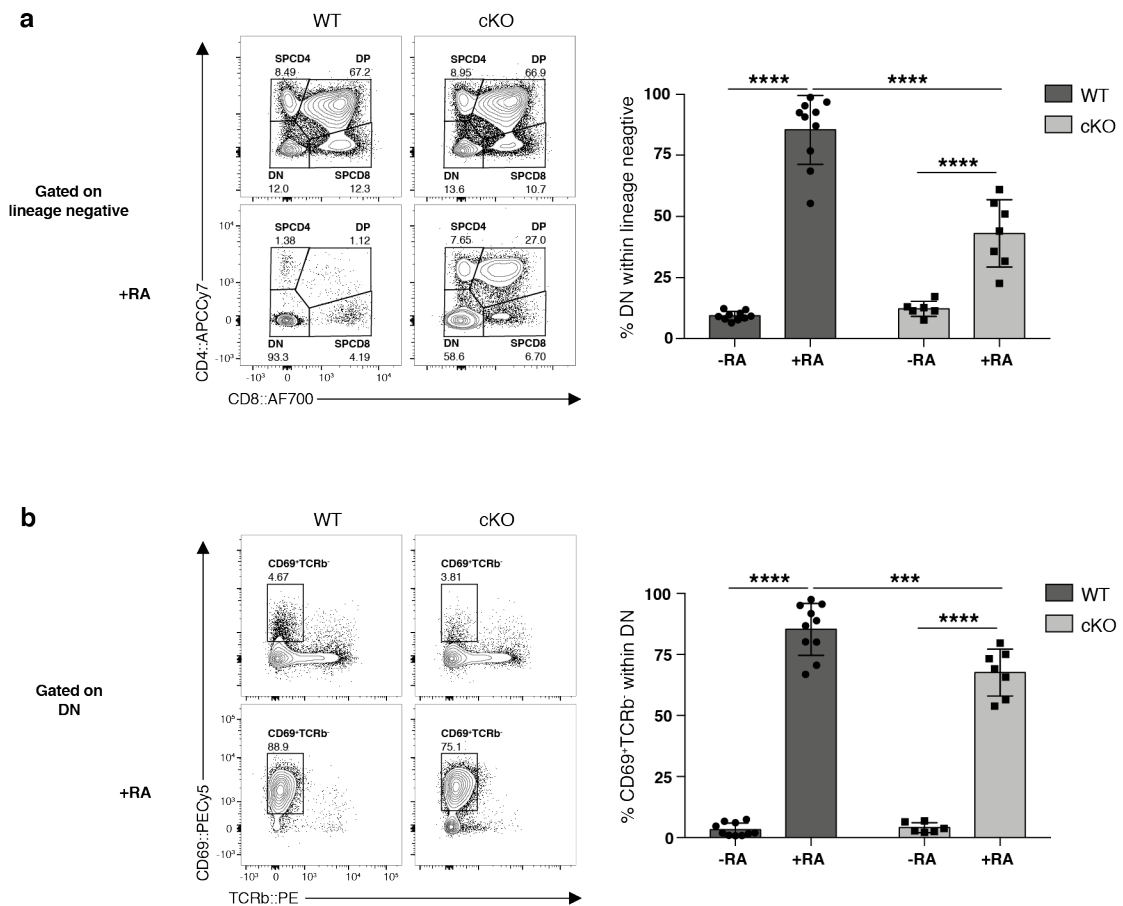


Figure 4.27 Addition of exogenous retinoic acid leads to an expansion of CD69⁺TCRb⁻CD4⁻CD8⁻ cells in WT and cKO FTOCs

(a-b) Representative FACS plots (left panels) and bar graphs (right panels) showing (a) frequency of double negative cells (DN: CD4⁻CD8⁻) within lineage negative cells; and (b) frequency of CD69⁺TCRb⁻ cells within DN cells in cKO and WT E14.5 FTOCs +/- treatment with 10⁻⁶M retinoic acid (RA). Bar graphs contain data pooled from 2 independent experiments.

After exclusion of DN cells, it would appear that WT FTOCs responded to the addition of RA with increased differentiation of SPCD4 (9.15 ± 1.30% without vs. 22.42 ± 5.50% with RA) and SPCD8 thymocytes (15.79 ± 3.13% without vs. 44.26 ± 16.56% with RA) (Figure 4.28a). In WT FTOCs there was also increased maturation of at least SPCD8 thymocytes as evidenced by a significant increase in the frequency of CD5^{Hi}TCRb^{Hi} cells within this population (20.62 ± 6.32% without vs. 52.06 ± 22.12% with RA) (Figure 4.28b). In contrast, there was no significant increase in SPCD8 thymocytes in cKO FTOCs following treatment with RA (Figure 4.28a). Although there was an

increase in the frequency of SP^{CD4} thymocytes after culture with RA ($9.83 \pm 1.30\%$ without vs. $17.70 \pm 2.48\%$ with RA) this was not to the same extent as for WT FTOCs (Figure 4.28a) and there were significantly fewer mature CD5^{Hi}TCRb^{Hi} cells within SP^{CD4} thymocytes in cKO compared to WT FTOCs after treatment with RA ($10.93 \pm 3.65\%$ vs. $49.91 \pm 21.77\%$) (Figure 4.28b).

In conclusion, the addition of RA rescued the proportion of mTEC in cKO lymphodepleted FTOCs to levels comparable with WT, providing functional evidence that the RA pathway is perturbed in TEC deficient in CHD4 and at least partly responsible for the TEC phenotype observed. Whether the increase in the proportion of mTEC translates to improved thymopoiesis is difficult to ascertain experimentally given that both TEC and thymocytes express RA receptors, rendering it impossible with this experimental design to distinguish the effect of RA on TEC from that of thymocytes.

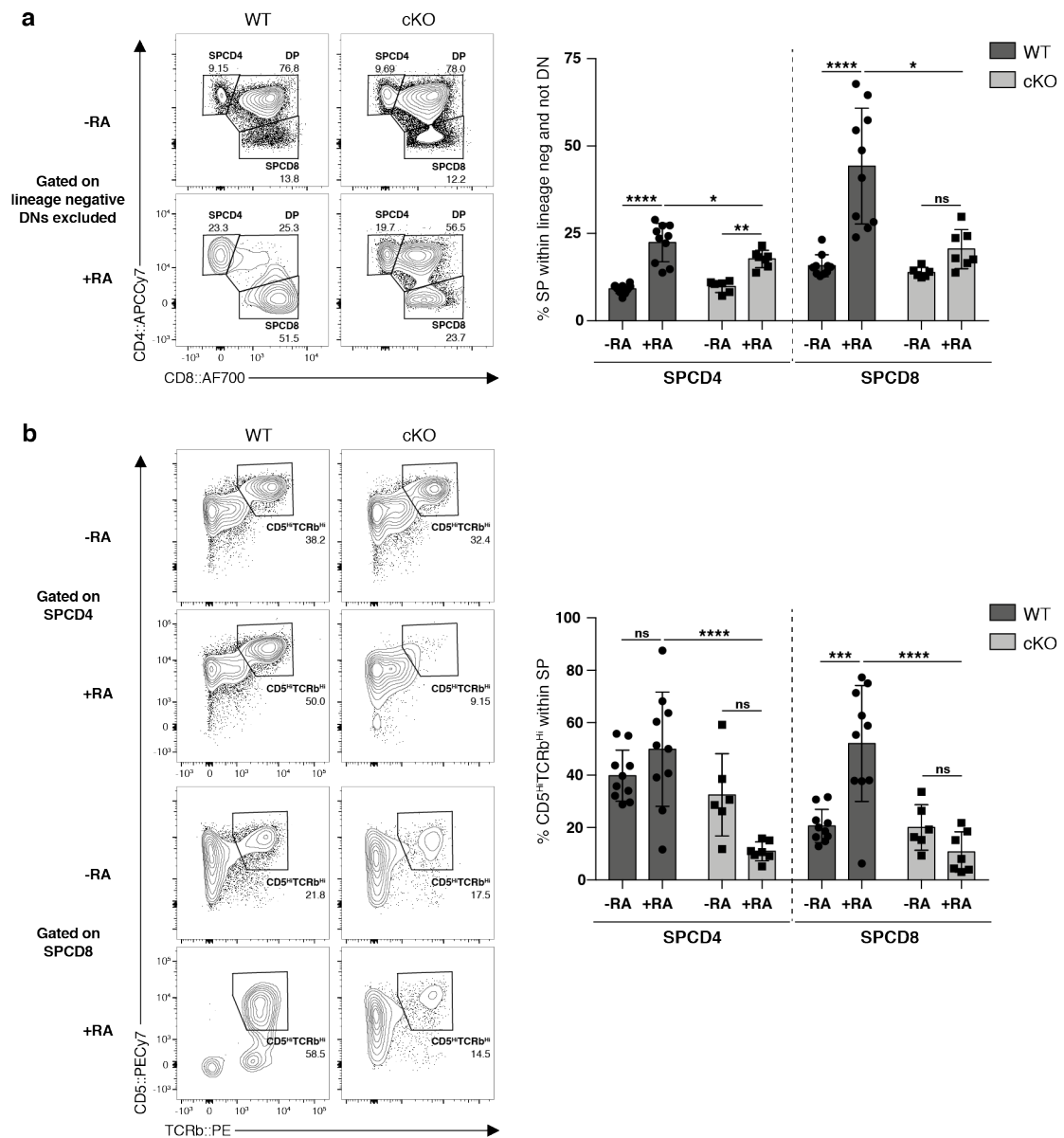


Figure 4.28 Addition of exogenous retinoic acid leads to increased production of mature SP thymocytes in WT but not cKO FTOCs

(a-b) Representative FACS plots (left panels) and bar graphs (right panels) showing (a) frequency of single positive CD4 (SPCD4) and CD8 (SPCD8) thymocytes cells within lineage negative cells and with the exclusion of any remaining CD4⁺CD8⁺ DN cells; and (b) frequency of CD5⁺TCRb⁺ cells within SPCD4 and SPCD8 thymocytes in cKO and WT E14.5 FTOCs +/- treatment with 10⁻⁶M retinoic acid (RA). Bar graphs contain data pooled from 2 independent experiments; ns = not significant.

4.7.6 Exogenous IGF-1 does not affect the cellularity or phenotype of either WT or cKO FTOCs

Genes belonging to the IGF pathway were found to be enriched amongst the genes downregulated in cKO cTEC (Figure 4.22b). *Igf1* was one of the genes within this pathway with significantly reduced expression in cKO compared to WT cTEC (Figure 4.29a). Although increased expression of *Igf2* in cKO cTEC (Figure 4.29b) could theoretically rescue any effect of reduced *Igf1* expression. Both TEC and thymocytes are known to express the receptor for IGF-1 (325). Exogenous IGF-1 has been shown in vitro to stimulate proliferation of human TEC cell lines (347), and to promote TEC:: thymocyte adhesion by enhancing TEC expression of ECM ligands and receptors including fibronectin and its receptor, VLA-5 (348, 349). Furthermore, IGF-1 has been shown in vivo to play a role in thymic growth and recovery by stimulating TEC cellularity and thymopoiesis (324, 325).

To test the functional significance of reduced *Igf1* expression in cKO cTEC exogenous IGF-1 was added to FTOCs derived from E14.5 cKO and WT thymi after lymphodepletion with 2DG. After culture for 7 days in the presence or absence 20ng/ml of recombinant IGF-1 FTOCs were assessed for TEC cellularity and subpopulation composition (Figure 4.29 c and d). This analysis revealed no differences in TEC cellularity (Figure 4.29c) nor in the frequency of mTEC within the total TEC compartment (Figure 4.29d) in either WT or cKO FTOCs treated with exogenous IGF-1 compared to untreated controls. This could either reflect the need for experimental optimisation with regards dose of

IGF-1 and/or the length of culture. Alternatively, there may be a differential requirement for IGF-1 and -2 at different stages of thymic development. Indeed, IGF-2 has previously been shown to be important during fetal thymic development (350, 351). Thus, the experiment conducted with the aim of testing the functional significance of reduced *Igf1* expression in cKO cTEC with regards to further TEC differentiation was not conclusive.

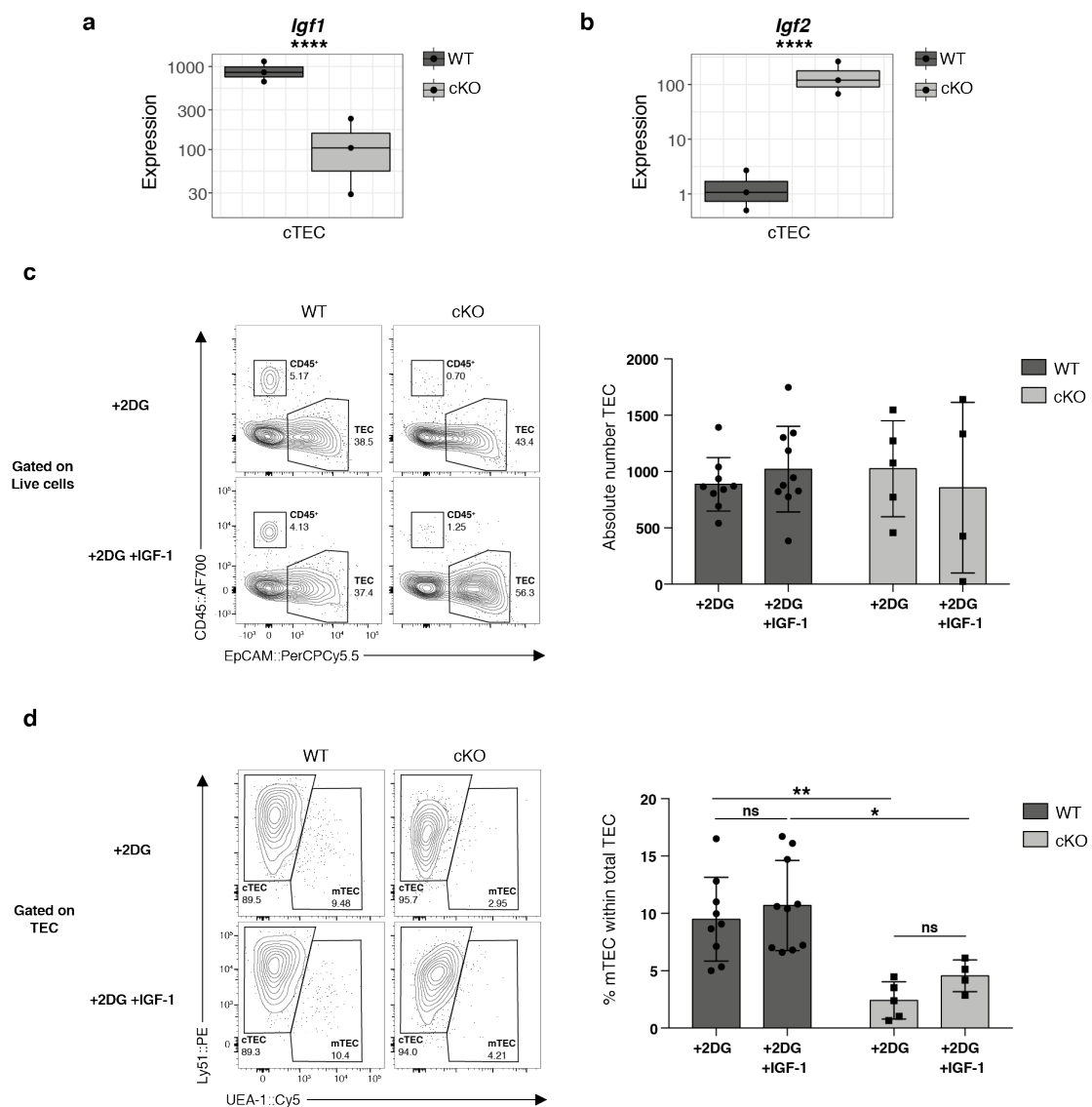


Figure 4.29 Addition of exogenous insulin-like growth factor 1 (IGF-1) has no effect on TEC cellularity or TEC subpopulation composition in either WT or cKO FTOCs

(a-b) Box and whisker plots showing normalised expression of (a) *Igf1*, and (b) *Igf2* in cKO and WT cTEC RNA-seq samples (c-d) Representative FACS plots (left panels) and bar graphs (right panels) showing (c) absolute number of TEC (CD45-EPCAM+), and (d) frequency of mTEC (UEA-1+) within total TEC in cKO and WT E14.5 FTOCs after lymphodepletion with 2-deoxyguanosine (2DG) +/- treatment with 20ng/ml IGF-1. Bar graphs contain data pooled from 2 independent experiments; ns = not significant.

4.8 Altered AIRE-regulated PGE in CHD4 deficient TEC

4.8.1 RNA-sequencing reveals defective AIRE-regulated PGE in CHD4 deficient CD80⁺CD86⁺ mTEC

CHD4 was identified as a potential interaction partner of AIRE in an in vitro overexpression system (182). In addition, its described roles in binding to the same histone modifications known to mark AIRE-induced genes in mTEC, together with its ability to directly bind to DNA and influence transcriptional regulation via nucleosome mobilisation and DNA repair (213, 227, 230, 233-237, 245), make CHD4 a particularly attractive candidate for participation in the process of AIRE-regulated gene expression.

To assess the consequences of an absence of CHD4 on AIRE-regulated PGE, RNA-seq was performed on CD80⁺CD86⁺ mTEC as these cells express the highest levels of AIRE, are the most competent for PGE and are thus thought to be the most effective in central tolerance induction (57, 128, 179). There were significant, albeit small, decreases in the frequency of AIRE⁺ cells within the total mTEC population ($18.6 \pm 1.2\%$ vs. $22.3 \pm 1.3\%$) (Figure 4.5d), and within the CD80⁺CD86⁺ subset in cKO mice when compared to WT animals ($49.6 \pm 3.7\%$ vs. $62.7 \pm 5.3\%$) (Figure 4.30a). Therefore, RNA-seq was performed at both the population and single-cell level with the intention of normalising for *Aire* expression at the mRNA level, in the case of the latter analysis, by comparing cKO and WT single cells expressing *Aire* at similar levels.

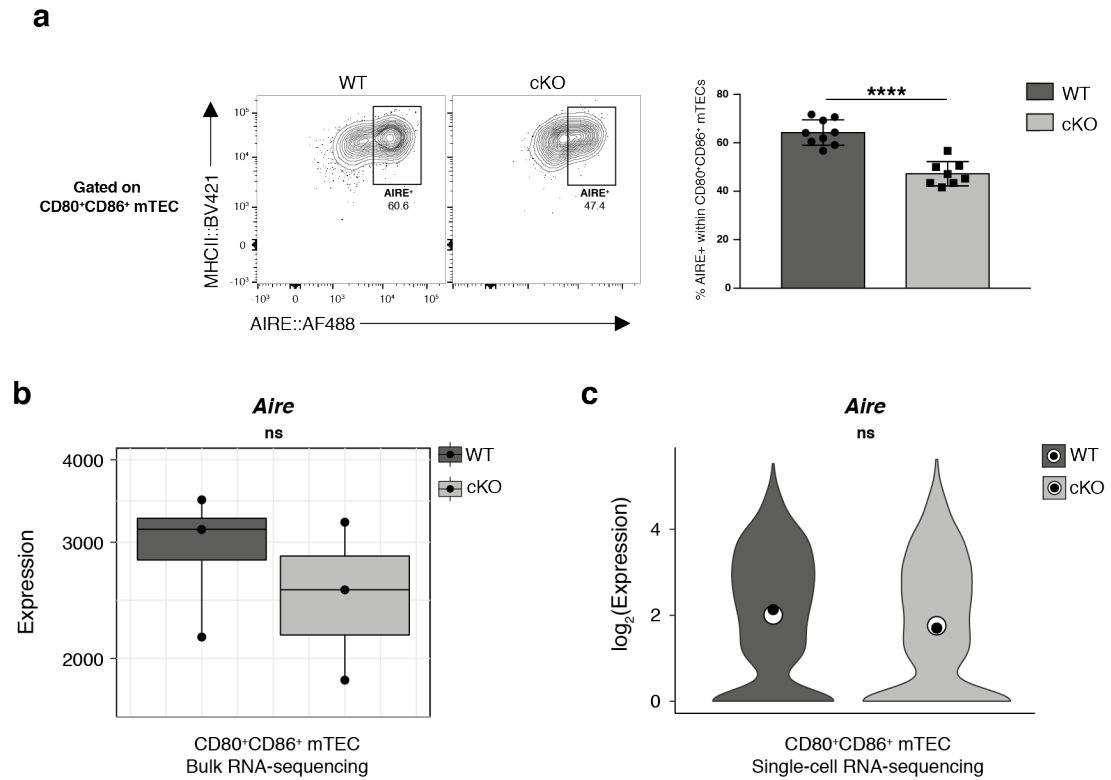


Figure 4.30 Discrepant AIRE expression at the protein and mRNA level in CD80⁺CD86⁺ mTEC from cKO and WT thymi

(a) Representative FACS plots (left panels) and bar graph (right panel) showing frequency of cells expressing AIRE protein (AIRE⁺) within CD80⁺CD86⁺ mTEC from cKO compared to WT thymi; bar graph contains data pooled from 2 independent experiments using 4-6 week old mice. (b) Box and whisker plot showing normalised expression of *Aire* mRNA in cKO and WT CD80⁺CD86⁺ bulk RNA-seq samples. (c) Violin plot showing normalised expression of *Aire* mRNA in cKO and WT CD80⁺CD86⁺ single-cell RNA-seq samples; white dot = mean, black dot = median, ns = not significant.

Bulk RNA-seq libraries were generated from 2,000 CD80⁺CD86⁺ mTEC from 4-5 week old female cKO and WT mice (3 biological replicates each) using the SMARTer amplification protocol. Single-cell RNA-seq libraries were generated from 2,000 CD80⁺CD86⁺ mTEC from 4-5 week old female cKO and WT mice (3 biological replicates) using the 10X Genomics platform. In both cases cDNA libraries were subject to paired-end 2x75bp sequencing. Surprisingly, there was no statistically significant difference in *Aire* mRNA expression, despite the

difference demonstrated at the protein level (Figure 4.30a), in either the bulk or single cell CD80⁺CD86⁺ mTEC datasets (Figure 4.30b and c). This precluded conducting the single-cell analysis after normalisation for *Aire* expression as originally planned.

Focusing on PGE, the population level (bulk) RNA-seq data revealed that cKO CD80⁺CD86⁺ mTEC transcribed significantly fewer AIRE-induced genes compared to WT cells (Figure 4.31a): a median of 265 AIRE-dependent genes were detected in cKO CD80⁺CD86⁺ mTEC compared to 374 in the WT, and for AIRE-enhanced genes a median of 2,654 genes were detected in cKO compared to 2,907 in WT. Whereas, there were no statistically significant differences in the number of AIRE-independent TRGs, housekeeping or other genes detected between cKO and WT CD80⁺CD86⁺ mTEC (Figure 4.31a). Differential expression analysis of the bulk RNA-seq data suggested that differentially expressed AIRE-dependent (68 downregulated vs. 6 upregulated) and AIRE-enhanced (388 downregulated vs. 37 upregulated) genes tended to be down- rather than upregulated in cKO CD80⁺CD86⁺ mTEC (Figure 4.31b). On the other hand, there was less of a tendency towards downregulation amongst the differentially expressed AIRE-independent TRGs (211 downregulated vs. 98 upregulated). There were very few differentially expressed housekeeping genes (1 downregulated vs. 4 upregulated) and a more even split between downregulation and upregulation within other differentially expressed genes (345 downregulated vs. 253 upregulated) (Figure 4.31b). Gene set enrichment analysis for AIRE-dependent, AIRE-

enhanced and AIRE-independent TRGs further showed that, when considered as a gene set, only AIRE-regulated TRGs, and not AIRE-independent, were significantly downregulated (Figure 4.31c). At the single-cell level cKO CD80⁺CD86⁺ expressed significantly fewer AIRE-induced genes per cell compared to their WT counterparts (median 114 vs. 133) (Figure 4.31d). Finally, in order to assess whether the defect in AIRE-induced gene expression observed in cKO CD80⁺CD86⁺ mTEC was independent of the reduced AIRE expression in the cKO, RT-qPCR for *Aire* and several AIRE-regulated genes that were found to be downregulated in the RNA-seq analyses was performed in FACS sorted AIRE⁺ mTEC from cKO and WT thymi (Figure 4.31e). Cells with a similar expression level of AIRE in cKO and WT thymi were selected for analysis by sorting cells displaying a similar gMFI for AIRE::AF488. Following FACS sorting, RNA was extracted using a kit that allows for RNA extraction from fixed cells and tissues. This analysis showed that the defect in AIRE-induced PGE observed in cKO mTEC was maintained even when AIRE expression was controlled for at the protein and mRNA level (Figure 4.31e).

In summary, in support of a role for a functional interaction between CHD4 and AIRE in the process of PGE, bulk and single-cell RNA-sequencing revealed reduced expression AIRE-induced TRGs in CHD4 deficient CD80⁺CD86⁺ mTEC.

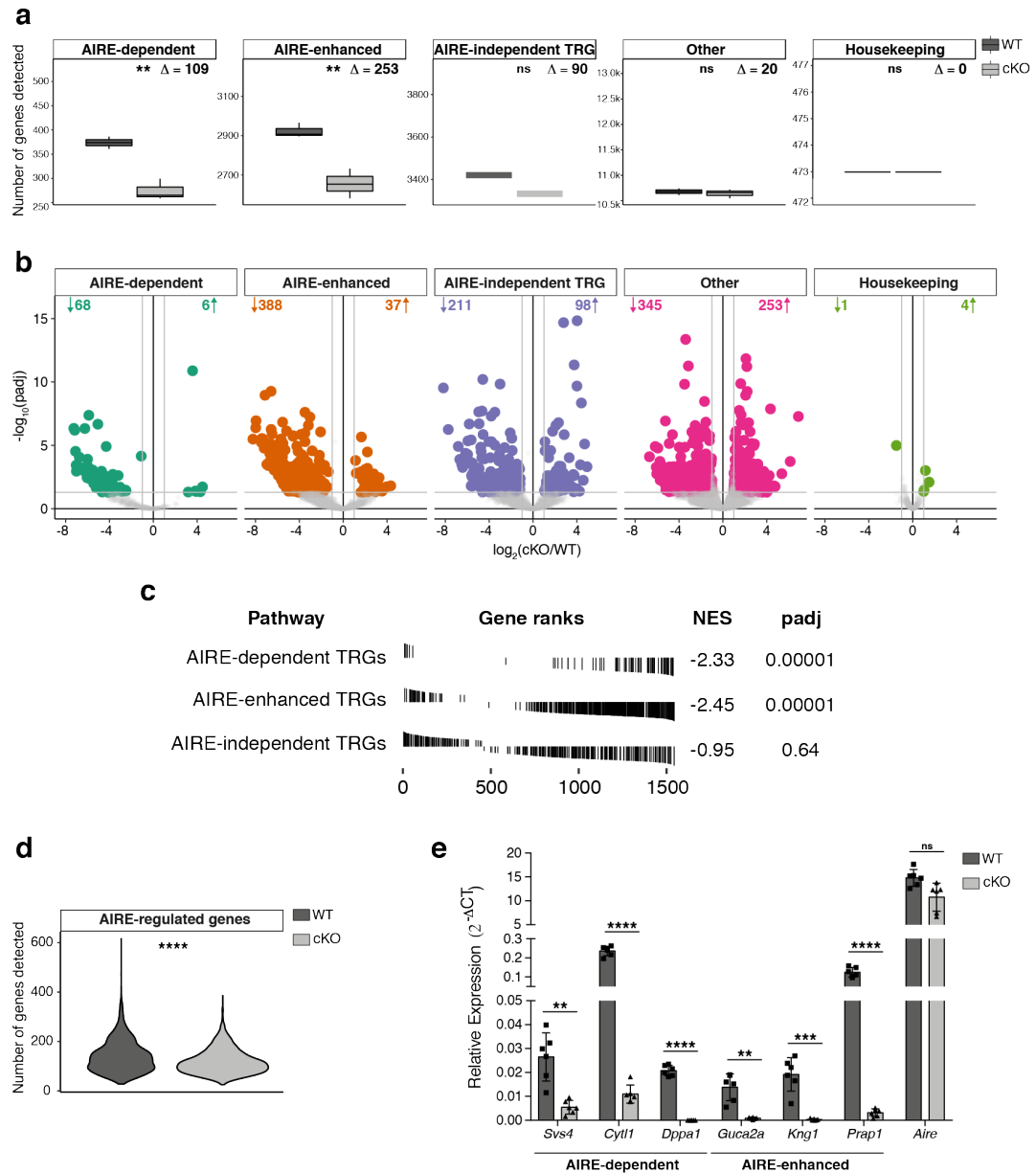


Figure 4.31 Defective AIRE-regulated PGE in CHD4 deficient CD80⁺CD86⁺ mTEC

(a) Box and whisker plot showing the number of AIRE-dependent and AIRE-enhanced genes, AIRE-independent tissue restricted genes (TRGs), other and housekeeping genes detected in bulk RNA-seq samples from cKO vs. WT CD80⁺CD86⁺ mTEC; Δ = median number of genes detected in WT - median number of genes detected in cKO. (b) Volcano plots showing transcriptional changes in AIRE-dependent, AIRE-enhanced, AIRE-independent TRGs, other and housekeeping genes in bulk RNA-seq samples from cKO vs. WT CD80⁺CD86⁺ mTEC; fold change ($\log_2$) in cKO vs. WT is displayed on the x-axis and the adjusted p value ($-\log_{10}$) is displayed on the y-axis; each dot represents a gene and coloured dots represent genes that are significantly differentially expressed (i.e. fold change ≥ 2 and adjusted p value ≤ 0.05); the number of significantly down- and upregulated genes are displayed in the upper left and right hand corners respectively. (c) Barcode plot showing gene set enrichment analysis for AIRE-dependent, AIRE-enhanced and AIRE-independent TRGs; each bar represents a differentially expressed gene sorted by log fold change, which is represented by the height of the bar. NES = normalised enrichment score, padj = adjusted p value. (d) Violin plot showing number of AIRE-induced genes detected per single cell in cKO vs. WT CD80⁺CD86⁺ mTEC. (e) RT-qPCR data showing relative expression of *Aire*, 3 AIRE-dependent and 3 AIRE-enhanced TRGs in FACS sorted AIRE⁺ mTEC from WT and cKO thymi. Expression levels have been normalised to *Gapdh*. Graph contains data pooled from 2 independent experiments with 3 replicates each from 4-6 week old mice.

4.8.2 Gene co-expression groups are lost in CHD4 deficient CD80⁺CD86⁺ mTEC

As the loss of CHD4 in mature mTEC disproportionately compromised the expression of AIRE-induced genes, the impact of CHD4 deficiency on gene co-expression groups was next assessed (See Chapter 3). The work presented in Chapter 3 of this thesis concerning gene expression patterns in single mTEC identified multiple gene co-expression modules in individual mTEC, the majority of which contained sets of TRGs (282). These co-expression modules were not determined by an identifiable biological process as genes within the same module were not related by chromosomal location, distance within the linear genome, tissue specificity, pathway, nor involvement in particular organ specific autoimmune diseases. However, the reproducibility of the gene co-expression groups, particularly between individual mice of different strains, suggested the existence of an underlying, but as yet uncharacterised, mechanism controlling the membership of genes to specific co-expression groups (282). For example, different cofactors involved in PGE could induce the expression of certain co-expression groups, perhaps via the recognition of a particular chromatin environment or epigenetic landscape, or via inducing chromosome looping to bring distantly located genes into close physical proximity to allow coordinate expression.

To assess whether CHD4 plays a role in the co-expression of TRGs in single mTEC, the single-cell RNA-seq data generated from cKO and WT CD80⁺CD86⁺ mTEC was used (see Subsection 4.8.1). To uncover the gene

module structure of single WT and CHD4 cKO CD80⁺CD86⁺ mTEC, a TF-IDF (272) transformation of the single-cell data was performed, and the resulting matrix was clustered using HDBSCAN (273). The gene modules generated by this method identified genes that were co-expressed in the combined WT and cKO single-cell datasets (Figure 4.32a). Next, by separately analysing the WT and cKO gene module structures, the amount of the overall structure that was defined by either WT or cKO CD80⁺CD86⁺ mTEC was determined. This revealed that the gene module structure generated from the combined WT and cKO datasets was very similar to that generated from the WT dataset alone having an adjusted mutual information (AMI) score of 0.736, where a score of 1 would denote that the compared gene modules are identical and a score of 0 that there is no overlap in gene module structure (Figure 4.32b). However, the cKO only gene module structure was quite different from the overall structure with an AMI of 0.442 (Figure 4.32c). Together with the data presented above (Subsection 4.8.1), this suggests that the loss of CHD4 not only resulted in a reduction in the expression of AIRE-regulated TRGs at both the population and single-cell level, but also impaired the quality of TRG representation by single mTEC by weakening gene co-expression. Overall, this may damage the TRG diversity expressed by single CHD4 deficient mTEC and impair their ability to select a self-tolerant T-cell repertoire, a possibility that is further explored in Subsection 4.8.5 below.

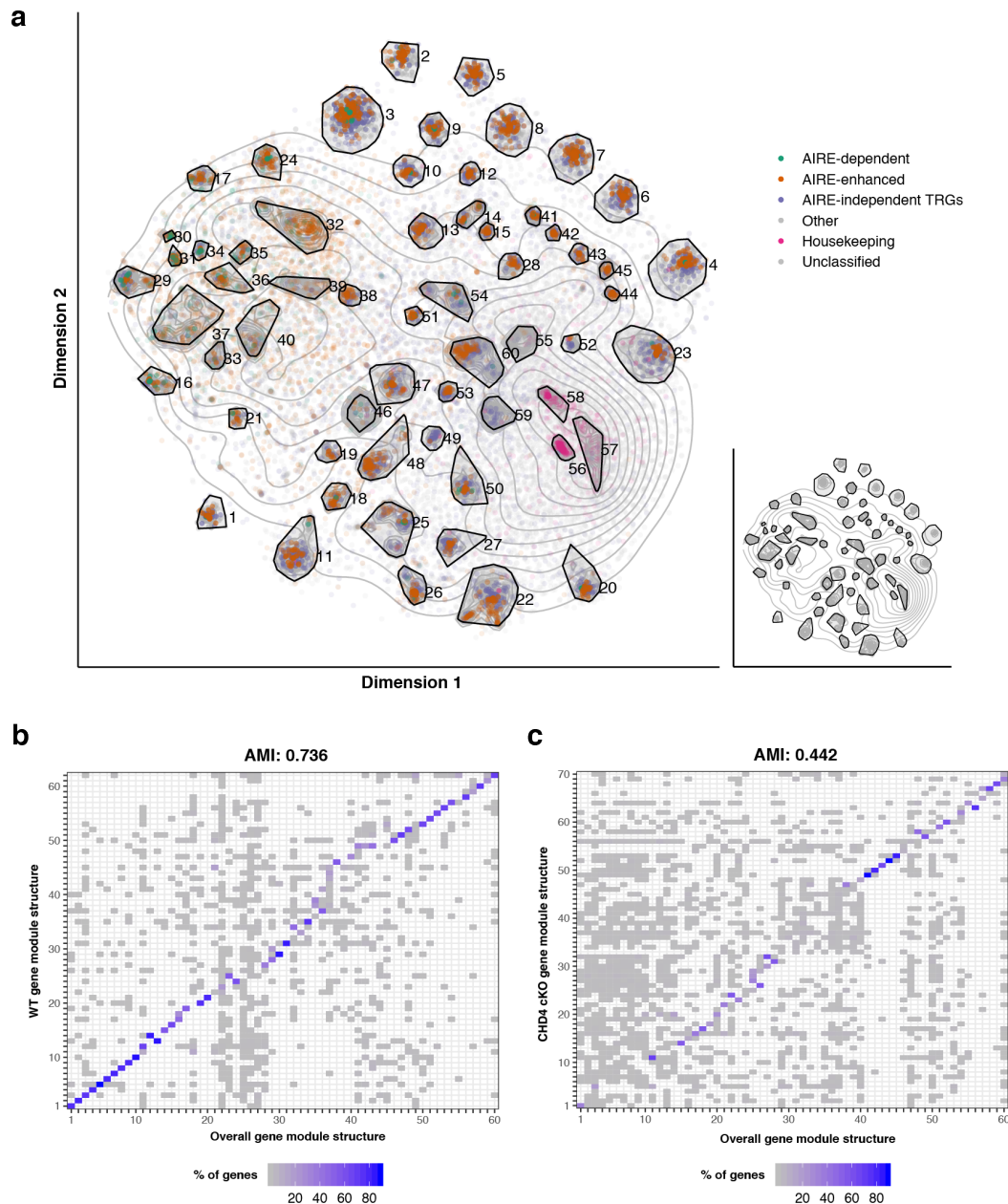


Figure 4.32 TRG co-expression groups are lost in cKO CD80⁺CD86⁺ mTEC

(a) t-SNE visualisation of the partition of the genes expressed in single CD80⁺CD86⁺ mTEC from WT and cKO data sets combined into 60 mutually exclusive gene modules containing genes expressed at similar levels and in similar subsets of CD80⁺CD86⁺ mTEC. Each dot is a gene colored by its gene category (AIRE-dependent, AIRE-enhanced, AIRE-independent TRG, other, housekeeping, unclassified). The colour intensity of each dot/gene is proportional to its membership probability within each module. Inset: The same tSNE visualisation with the 2d density shown as a contour plot (grey lines) with the gene modules outlined in black, demonstrating that gene modules are defined by high density regions with similar expression characteristics. (b) Comparison of gene module structure defined by all WT and cKO CD80⁺CD86⁺ mTEC combined (x-axis: Overall gene module structure) to that defined by WT CD80⁺CD86⁺ mTEC only (y-axis: WT gene module structure). (c) Comparison of gene module structure defined by all WT and cKO CD80⁺CD86⁺ mTEC combined (x-axis: Overall gene module structure) to that defined by cKO CD80⁺CD86⁺ mTEC only (y-axis: cKO gene module structure). The grid is colored by the percentage of genes from the x-axis that are within that square on the y-axis. The adjusted mutual information (AMI) score (range 0 to 1) compares the structures and shows that co-expression patterns from WT CD80⁺86⁺ mTEC are driving the overall co-expression structure derived from the combined WT and cKO datasets.

4.8.3 CHD4 and AIRE do not directly physically interact

CHD4 was previously reported to physically associate with chromatin bound AIRE in an in vitro study using HEK293T cells transfected with human AIRE (182). To test for the presence of a direct physical interaction between CHD4 and AIRE in primary ex vivo mTEC, proximity ligation imaging cytometry (PLIC) was employed (189). This technique combines proximity ligation assay with imaging cytometry allowing for the detection and quantitation of protein::protein interactions occurring over a maximum distance of 30-40nm at the single-cell level. PLIC has recently been successfully used in mTEC to characterise the interaction between AIRE and Sirtuin-1 (SIRT1) (189). PLIC testing for a direct physical interaction between CHD4 and AIRE was performed alongside a positive control (PLIC AIRE and SIRT1), and a negative control (PLIC CHD4 only). The gate for a positive PLIC signal was set on the basis of the technical negative control (PLIC CHD4 only). In addition, this gate was checked against CD45⁺ cells within the positive control tube testing for an interaction between AIRE and SIRT1, as these cells serve as a negative control cell population as they lack expression of AIRE. There was no PLIC signal under the conditions tested for a physical interaction between CHD4 and AIRE, in contrast to the expected signal observed with the positive control (Figure 4.33a-c and Supplementary Figure 3 for initial gating strategy). The anti-CHD4 antibody used in the assay was both functional and specific, as evidenced by imaging cytometry on primary TEC using anti-CHD4 and a secondary antibody conjugated to AF488 (Figure 4.33d), as well as by the

western blot previously shown (Figure 4.1e) and ChIP-seencing (Subsection 4.8.4).

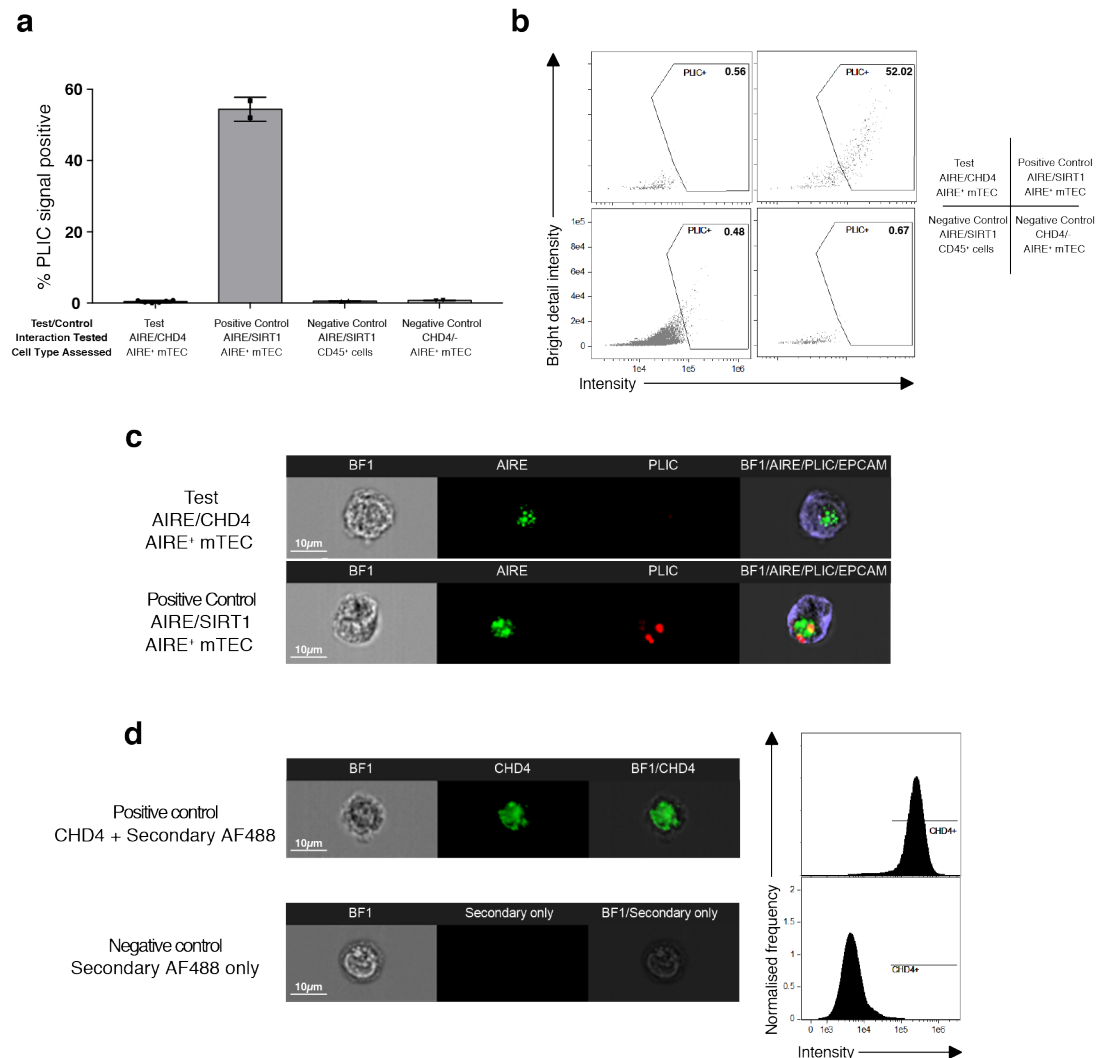


Figure 4.33 CHD4 and AIRE do not directly physically interact

(a) Bar graph showing results of PLIC testing for a direct physical interaction between CHD4 and AIRE, alongside positive and negative controls. (b) Representative plots showing PLIC signal in test and control tubes. (c) Representative images showing a lack of interaction between CHD4 and AIRE (top) vs. a positive interaction between AIRE and SIRT1 (bottom). Images from left to right: bright field (BF1); AIRE; PLIC signal; merge of BF1, AIRE, PLIC and EpCAM. (d) Representative images (left) and histograms (right) showing CHD4 staining in primary TEC (top) compared to staining with the secondary antibody only (bottom) Images from left to right: bright field (BF1); CHD4 (top) or secondary only (bottom); merge. The results shown are representative of 2 independent experiments using 4-6 week old WT C57BL/6 mice. For each independent experiment 3 biological replicates were performed for CHD4/AIRE PLIC alongside one of each of the controls.

These results therefore demonstrate that CHD4 does not directly physically interact with AIRE in primary mTEC. However, this finding does not negate a functional interaction, a notion which is supported by the RNA-seq data presented above (see Subsection 4.8.1 and Figure 4.31). Moreover, the negative PLIC result does not exclude an indirect physical interaction between CHD4 and AIRE, for example in the context of a larger chromatin associated multi-protein complex. This latter explanation would be in keeping with previous studies describing a complex AIRE interactome (181, 182).

4.8.4 CHD4 binds chromatin at the same sites as AIRE and induces the expression of genes with a repressive chromatin context

To determine the genomic context in which CHD4 binds in primary mature mTEC, ChIP for CHD4 was performed on primary CD80⁺ mTEC from 4-6 week old female WT mice (3 replicates, 500,000 cells per replicate), next generation sequencing libraries were prepared and then subjected to paired-end 2x75bp sequencing.

Analysis of CHD4 ChIP-seq data was performed in collaboration with Dr Jeanette Baran-Gale and Professor Chris Ponting (MRC Institute of Genetics and Molecular Medicine, University of Edinburgh). Libraries were of high quality with 674-1,368 million reads per sample (mean 874 million reads/sample). CHD4 ChIP-seq libraries were aligned to the mm10 genome using Bowtie2 (275), the data from each replicate was merged to provide more sensitive detection of CHD4 binding and the peaks were called with MACS2

using the broad peak calling mode (352). Next, to determine the reproducibility of the data, peaks were called separately for each replicate and the correlation of CHD4 binding between replicates was computed. The CHD4 ChIP-seq data generated in this study was analysed alongside previously published ChIP-seq data for AIRE and histone modifications (111, 198, 199) as well as ATAC-seq data from mature WT mTEC (199).

The binding profile of CHD4 in CD80⁺ mTEC revealed that, CHD4 binds in broad domains rather than narrow peaks, in a pattern remarkably similar to that observed for AIRE (198) (Figure 4.34a). Narrow peaks are characteristically found in ChIP-seq analyses for transcription factors, which bind to particular DNA sequence motifs (353), and, as such, motif analysis for CHD4 was inconclusive (data not shown). In contrast, binding to broad genomic regions is more typical of histone modifications, but has also been described for chromatin remodelling enzymes, including CHD4, in mouse mammary epithelial cells (353, 354). A total of 58,476 CHD4 binding sites were identified in CD80⁺ mTEC. CHD4 binding was highly correlated between replicates (Pearson $r = 0.98$) indicating that the data were robust and reproducible (Figure 4.34b). Correlation analysis additionally revealed a striking concordance between the genomic binding locations of CHD4 and those for AIRE over the promoter regions of all genes (± 2 kb from TSS), indicating that, in mature mTEC, CHD4 and AIRE display comparable binding patterns over gene promoters (Figure 4.34b).

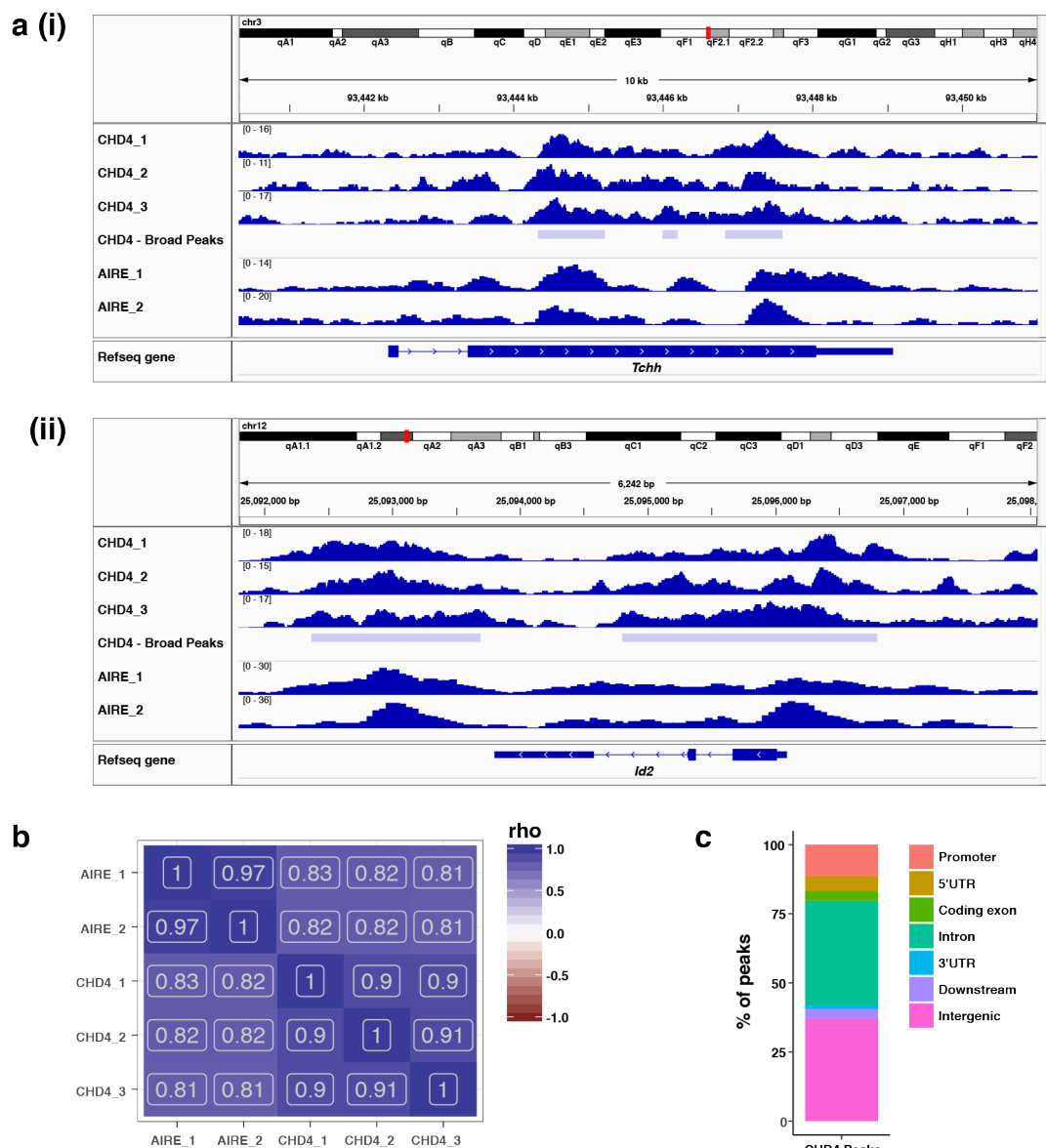


Figure 4.34 CHD4 occupies the same genomic binding sites as AIRE in primary mature mTEC

(a) Normalised ChIP-seq profiles for CHD4 (CHD4_1, _2, _3) and AIRE (AIRE_1, _2) replicates (198), with CHD4 peaks annotated (CHD4 Broad Peaks) for 2 AIRE-regulated genes that were downregulated in CHD4 cKO mature mTEC (i) *Tchh* and (ii) *Id2*. (b) Heatmap of Spearman Rank correlation coefficients (Rho) for CHD4 and AIRE ChIP-seq peaks at gene promoters (+/- 2kb from TSS). (c) Bar chart showing genome-wide distribution of CHD4 binding sites in WT CD80⁺ mTEC. CHD4 ChIP-seq was performed on primary CD80⁺ mTEC from 4-6 week old female WT mice (3 replicates, 500,000 cells per replicate), AIRE ChIP-seq data was previously published (198).

To assess whether CHD4 binds particular genomic features associated with annotated genes in CD80⁺ mTEC (e.g. promoter, 5' untranslated region (UTR), exon, intron, 3'UTR, downstream, intergenic region) the proportion of CHD4 binding sites associated with each of these features was assessed. This analysis revealed that CHD4, like AIRE (198), was broadly distributed across the genome but was particularly bound to promoters, introns and intergenic regions (Figure 4.34c).

Next, the genomic binding sites for CHD4 in mature mTEC were additionally compared to those for H3K4me3, H3K9me3 and H3K27me3 (111, 198, 199), as well as to chromatin accessibility (199). For these analyses, genes that had been found to be upregulated, downregulated and unchanged by RNA-seq in CHD4 deficient mature mTEC were considered separately (Figure 4.35). CHD4 binding was highest at the TSSs of genes. CHD4 was also broadly distributed across the genome i.e. at genes both upregulated, downregulated and unchanged in their mRNA expression in CHD4 cKO mature mTEC. Again, the binding profile for CHD4 was remarkably similar to that for AIRE (Figure 4.35). Interestingly, in comparison to genes that were not regulated by CHD4 (i.e. those not differentially expressed in CHD4 deficient compared to WT mature mTEC), genes that were positively regulated by CHD4 (i.e. those downregulated in the context of CHD4 deficiency) had a more repressive chromatin context. Specifically, genes that were positively regulated by CHD4 had significantly less H3K4me3 at their TSSs ($p < 2.2 \times 10^{-16}$), more H3K9me3

throughout their gene bodies ($p < 2.2 \times 10^{-16}$), more H3K27me3 ($p < 2.2 \times 10^{-16}$), and a more compact or closed chromatin conformation on ATAC-seq ($p = 1.98 \times 10^{-10}$) (Figure 4.35).

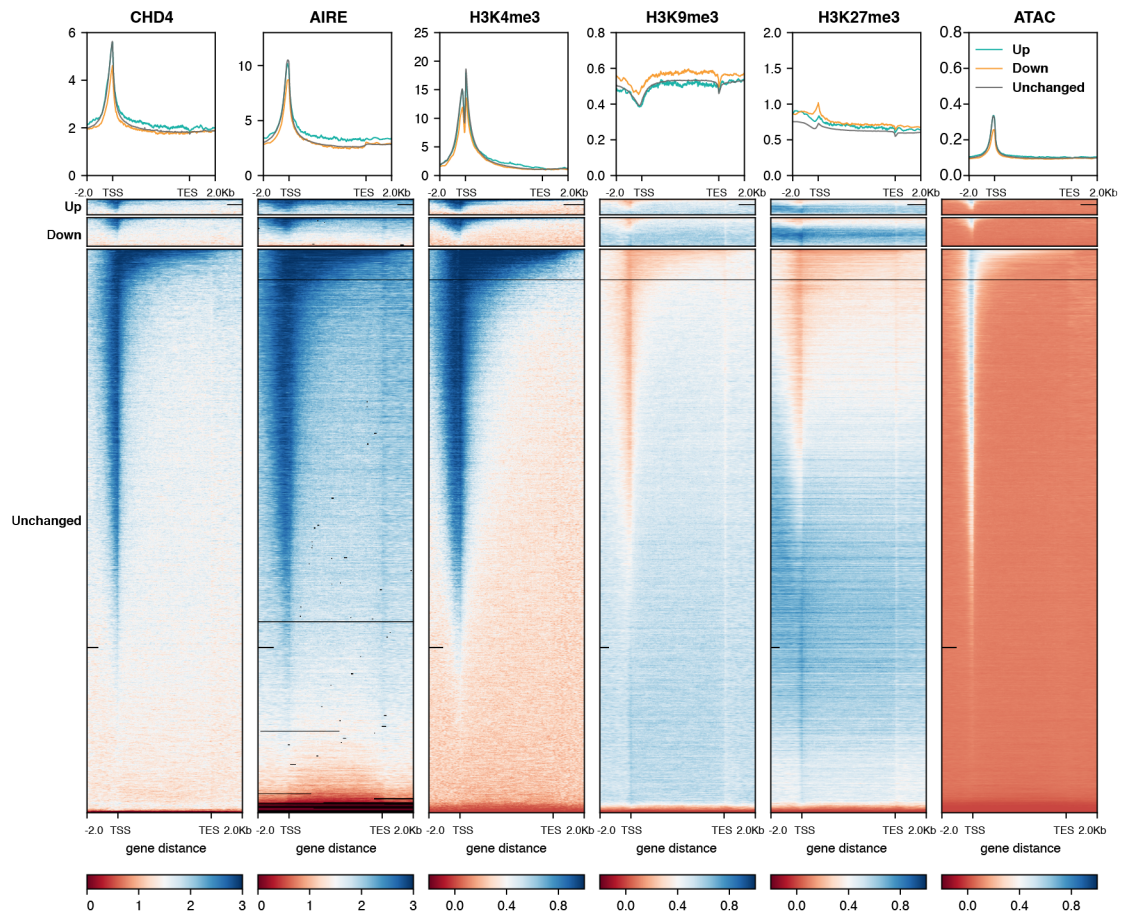


Figure 4.35 CHD4 positively regulates expression on genes with a repressive chromatin context in mature mTEC

Heatmaps (bottom) showing signals for ChIP-seq for CHD4, AIRE, H3K4me3, H3K9me3, H3K27me3 as well as ATAC-seq in primary mature mTEC for genes upregulated, downregulated and unchanged in CHD4 cKO mature mTEC. Metagene plots (top) are displayed above their respective heatmaps showing the average signal for genes upregulated (blue), downregulated (orange) and unchanged (grey) in CHD4 cKO mature mTEC. For both plots data are displayed spanning 2Kb upstream (-2.0) and 2Kb downstream (+2.0Kb) of genes, with the transcriptional start site (TSS) and transcriptional end site (TES) annotated. ChIP-seq was performed on primary CD80⁺ mTEC from 4-6 week old female WT mice (3 replicates, 500,000 cells per replicate), ATAC-seq, and ChIP-seq data for AIRE and histone modifications were previously published (111, 198, 199).

In summary, ChIP-seq revealed that CHD4 binding is broadly distributed across the genome at the same genomic locations as AIRE, supporting the idea that, in lieu of a direct physical interaction, CHD4 and AIRE interact indirectly, perhaps in the context of a larger chromatin bound multiprotein complex. Given that CHD4 binds genes that are both upregulated, downregulated and unchanged in the context of CHD4 deficiency it is possible that the function of proteins that interact with CHD4 instruct whether a gene bound by CHD4 is positively or negatively regulated. Interestingly, genes whose expression was positively regulated by CHD4 (i.e. those downregulated in CHD4 deficient mature mTEC) had a more repressive chromatin context, a finding that is consistent with a role for CHD4 in de-repressing such loci in partnership with AIRE.

4.8.5 CHD4 deficiency in TEC causes defective T-cell central tolerance induction

To determine whether the defects observed in CHD4 TEC cKO mice were sufficient to cause defective T-cell central tolerance induction, the presence of tissue-specific autoantibodies and lymphocytic organ infiltration was determined in cKO vs. WT mice. Even in the context of AIRE deficiency, mice on a C57BL/6 background are known to be almost completely resistant to the development of spontaneous autoimmune disease (144). Therefore, to assess whether CHD4 cKO mice had impaired self-tolerance, T-cell lymphopaenia was induced in 2 week old female CHD4 cKO and WT controls with the aim of provoking lymphopaenia-induced lymphoproliferation and the development of

autoimmunity (262). Mice were treated and analysed in 2 batches consisting of 4 cKO mice and 4 WT mice per batch (i.e. $n = 8$ in total for each genotype). Successful T-cell ablation was assessed by quantifying the peripheral T-cell compartment 7 days post treatment with intraperitoneal injection of a mixture of monoclonal antibodies directed against CD4, CD8 and Thy1.1 (Figure 4.36). Mice were then aged for 25-30 weeks in order to allow time for the development of autoimmune pathologies.

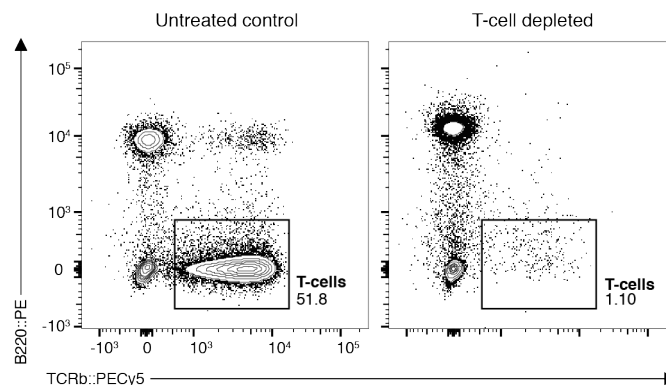


Figure 4.36 Successful T-cell ablation 7 days post injection with monoclonal antibodies against CD4, CD8, and Thy1.1

Representative FACS plots showing frequency of T-cells (TCRb⁺B220⁻) within lymph nodes harvested from an untreated WT mouse and a WT mouse 7 days after T-cell depletion with intraperitoneal anti-CD4/anti-CD8/anti-Thy1.1. This data is representative of 2 independent experiments.

Following ageing for 25-30 weeks there were no overt clinical signs of ill-health in either WT or cKO mice, nor was there any evidence of reduced weight in cKO mice compared to WT animals, a result which may have been indicative of severe organ-specific or systemic autoimmune disease (Figure 4.37).

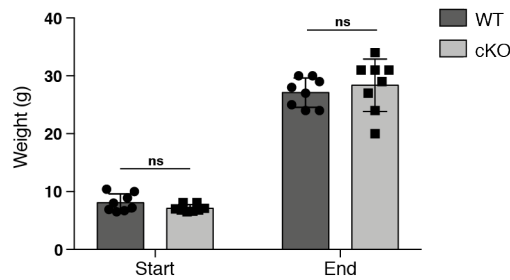


Figure 4.37 No difference in weight of cKO compared to WT mice after induction of T-cell lymphopaenia and ageing for 25-30 weeks

Bar chart showing weight in grams of cKO compared to WT mice at the age of 2 weeks old before induction of T-cell lymphopaenia (Start) and after induction of T-cell lymphopaenia and ageing for 25-30 weeks (End). Graph contains data pooled from 2 independent experiments.

Serum and organs were harvested from aged cKO and WT mice in order to look for the presence of tissue specific autoantibodies and lymphocytic tissue infiltration, respectively. Based the autoimmune pathology observed in AIRE deficient mouse models, several tissues/organs were selected for study, including retina, lacrimal gland, salivary gland, lung, stomach, liver, pancreas and ovary (144, 262). To test for the presence of serum autoantibodies, indirect immunofluorescence was performed using tissues from RAG1 deficient mice in order to avoid false positive results due to endogenous immunoglobulins present in the extracellular space of the harvested organs (Figure 4.38a). As a positive control for the indirect immunohistological staining process, a slide from each tissue was stained with a mouse IgG anti-alpha tubulin antibody followed by the same secondary antibody used to detect binding of serum autoantibodies, and as a negative control secondary antibody staining alone was used (Supplementary Figure 4). Tissue-specific serum autoantibodies directed against various organs were detected in 6/8 cKO mice analysed compared to only 1/8 of the WT mice that had reactivity against

retinal tissue only. Within the cKO group 4/8 mice had evidence of serum autoantibody binding to retinal tissue, 1/8 to lacrimal gland, 3/8 to lung, 3/8 to stomach, 1/8 to liver and 4/8 to ovary (Figure 4.38a). To test for the presence of lymphocytic organ infiltration, H&E staining was performed on tissues harvested from cKO and WT mice. All 8 cKO mice had evidence of lymphocytic infiltration in various organs, whereas only 1/8 of the WT mice had organ infiltration affecting the lacrimal gland, salivary gland and liver. Within the cKO group 1/8 mice had evidence of lymphocytic infiltration in retinal tissue, 3/8 in lacrimal gland, 3/8 in salivary gland, 5/8 in lung, 4/8 in stomach, 2/8 to liver and 3/8 to ovary (Figure 4.38b).

In summary, the presence of tissue-specific autoantibodies and lymphocytic organ infiltration is indicative of a breakdown of T-cell central tolerance induction in mice with CHD4 deficiency targeted to TEC. The detection of serum IgG autoantibodies specifically signals a breakdown of tolerance within the CD4⁺ T-cell compartment as these cells are required to provide help for class-switched antibody production. The analyses performed have not addressed the phenotype of the tissue infiltrating lymphocytes and therefore do not specifically address whether CD8⁺ T-cell tolerance is also affected.

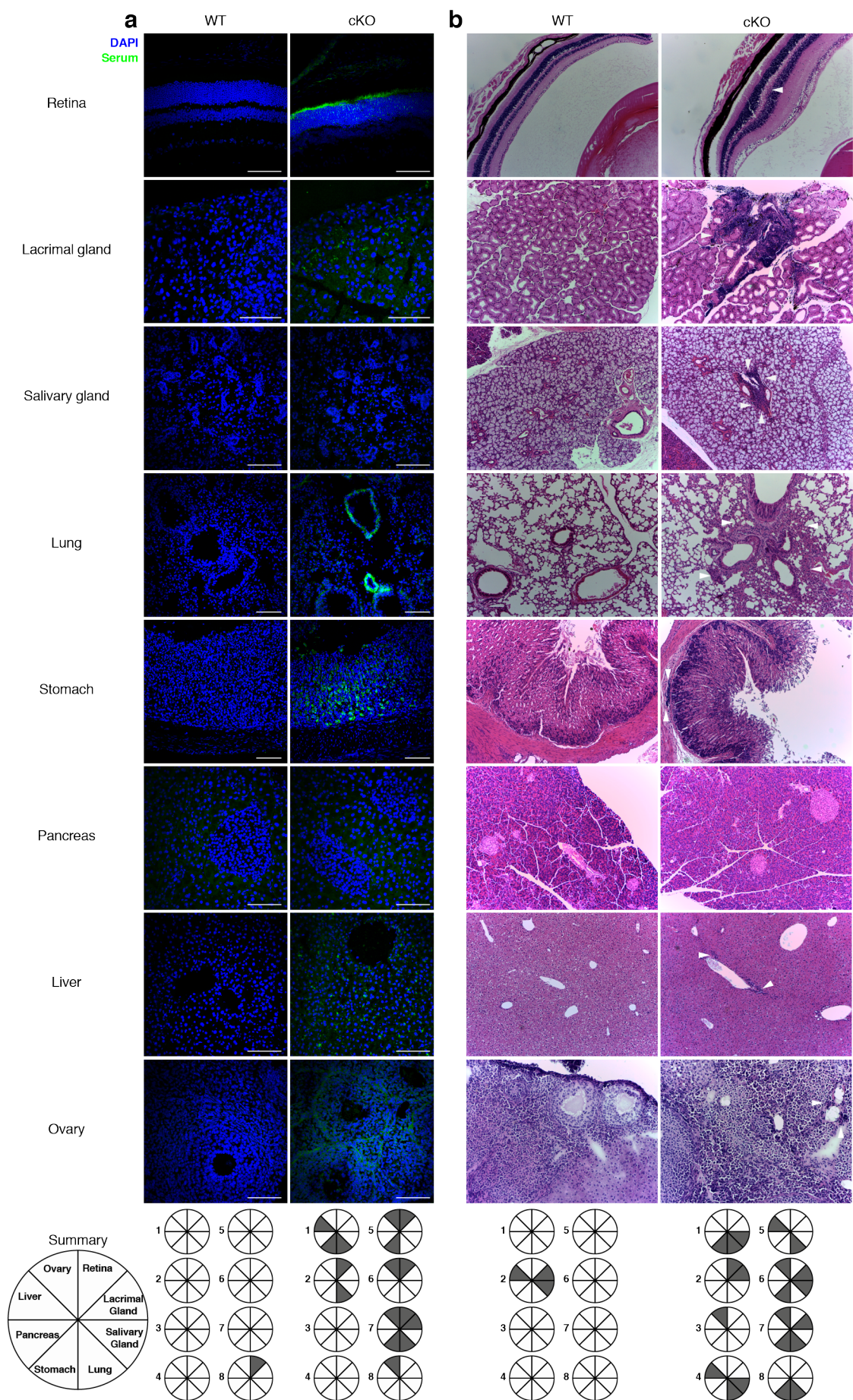


Figure 4.38 Peripheral autoimmunity in CHD4 cKO mice suggests failure of T-cell central tolerance induction

(a) Indirect immunofluorescence staining of tissues from RAG1 deficient mice using sera from WT and CHD4 cKO mice following induction of T-cell lymphopaenia and ageing for 25-30 weeks; scale bars represent 100 μ m (b) H&E staining of tissues harvested from WT and CHD4 cKO mice following induction of T-cell lymphopaenia and ageing for 25-30 weeks; white arrows signal areas of lymphocytic infiltration. For both (a) and (b) representative images are shown for each organ with results summarised for each individual mouse in the pie charts below where each segment represents an organ, with segments coloured grey to indicate the presence of (a) autoantibodies or (b) lymphocytic infiltration.

4.9 Discussion

A thymocyte intrinsic role for CHD4 in intrathymic T-cell development has previously been reported in which CHD4 is required for proper CD4 expression and therefore progression from the DN to DP, and the DP to SPCD4 stages of thymocyte maturation (225, 244). CHD4 positively regulates transcription of the *Cd4* gene via two mechanisms. Firstly, it binds and positively regulates the *Cd4* enhancer by recruiting the E-box binding transcription factor HEB and p300 resulting in local hyperacetylation of H3 and transactivation of the *Cd4* locus (225). At the same time, CHD4 antagonises the action of Ikaros to negatively regulate the activity of the *Cd4* silencer, again via a mechanism that involves the recruitment of histone acetyltransferases (HATs) and local H3 hyperacetylation (244).

The current study addresses, for the first time, the cell intrinsic role of CHD4 in TEC development and their ability to support proper intrathymic T-cell development, particularly focussing on its role in AIRE-regulated PGE and consequently T-cell central tolerance induction. To study the TEC intrinsic

function of CHD4 a conditional mouse model was used in which *Chd4* is deleted early in TEC ontogeny. This revealed that CHD4 is required for TEC development and maturation, particularly of the mTEC lineage, as well as for TEC architectural organisation of the thymic stroma. CHD4 is also necessary for TEC function; specifically, loss of CHD4 impacts the ability of TEC to support regular T-cell lymphopoiesis, PGE, particularly of AIRE-regulated TRGs, and consequently T-cell central tolerance induction.

Examination of the CHD4 deficient TEC compartment revealed hypocellularity, an increased relative frequency of cTEC and a reduced frequency of mTEC. Within the mTEC compartment there were further reductions in the frequency of the most mature developmental stages, namely those that express AIRE and CD86. These findings are consistent with CHD4 deficiency causing multiple partial developmental blocks affecting the generation and maturation of the mTEC lineage.

TEC develop from bipotent progenitors that express markers such as PLET1, α_6 -intergrin and SCA1, and are ultimately able to give rise to both the c- and m-TEC lineages (8, 10, 14-17, 49, 50). The cells are initially thought to pass through a cTEC-like stage before committing to either the c- and m-TEC lineage (19, 51). Development of the mTEC lineage then occurs via mTEC committed progenitors, marked by expression of CLD3/4, SSEA-1, CD24, and/or podoplanin, that first give rise to MHCII^{Lo}CD80^{Lo}AIRE⁻ immature mTEC (23, 24, 53-56). As they mature, mTEC sequentially upregulate expression of

key molecules for antigen presentation, T-cell co-stimulation, and self-antigen expression to adopt a MHCII^{Hi}CD80^{Hi}CD86⁺AIRE⁺ phenotype before entering the process of terminal differentiation (57-62).

In support of the notion that the abnormal TEC phenotype observed in the context of CHD4 deficiency is due to multiple partial developmental blocks affecting the generation and maturation of the mTEC lineage, the relative deficiencies in total mTEC and CD86⁺/AIRE⁺ mTEC were accompanied by accumulations of the preceding maturational stages, namely cTEC and MHCII^{Hi}CD80⁺CD86⁻ mTEC respectively. Study of various time points spanning embryonic and postnatal thymic development, as well as early thymic involution further revealed that these TEC defects were established during embryonic thymic development and were not due to increased apoptosis or decreased cell cycling of mTEC. Finally, increased expression of TEC progenitor markers including SCA1, CD24 and podoplanin was observed in embryonic TEC. Taken together these findings suggest that CHD4 is required from the earliest stages of TEC development to help unravel their differentiation program from the progenitor status to the mature mTEC stage. The effect of CHD4 deficiency on the development and maturation of the mTEC compartment was not due to lack of responsiveness to RANK-L as CHD4 deficient TEC responded to exogenous RANK-L by differentiating into mTEC.

In addition to the defects in mTEC development and maturation described above, CHD4 deficiency in TEC also causes thymic structural malformations, including abnormal cortico-medullary separation and the presence of epithelial cell free areas that harbour accumulations of developing thymocytes.

The aberrant thymic microenvironment caused by CHD4 deficiency in TEC in terms of both cellular composition and architectural organisation is unable to support the normal development and selection of thymocytes. CHD4 cKO thymi displayed hypocellularity within the thymocyte compartment, reduced generation of SP thymocytes and defective maturation within the committed SP^{CD4} lineage. The deficiency of SP thymocytes may, in part, be due to decreased positive selection, which was evidenced by an accumulation of pre-selectional CD69⁺TCRb^{-/lo} thymocytes alongside a deficiency of post-selectional CD69⁺TCRb^{Hi} thymocytes (304). There was also a small relative increase in the proportion of cells undergoing cortical (wave 1) and medullary (wave 2) negative selection detectable within the CHD4 cKO thymus (86). This could reflect a real increase in negative selection possibly due to the increased proportion of cTEC and MHCII^{Hi}CD80^{Hi} mTEC, in relation to wave 1 and wave 2, respectively. Alternatively, the increased frequency of Helios⁺ thymocytes could simply reflect less efficient clearance of negatively selected cells within the CHD4 deficient thymic microenvironment rather than truly increased negative selection per se, particularly as cKO mice were susceptible to autoimmunity. In aggregate, the defects in thymopoiesis were reflected in the

periphery by T-cell lymphopaenia with evidence for lymphopaenia-induced proliferation within both the effector T-cell and regulatory T-cell compartments.

RNA-sequencing of adult TEC revealed that CHD4 regulates the expression of thousands of genes in postnatal TEC suggesting that the TEC developmental, organisational and functional defects observed in the CHD4 TEC cKO are likely to have a complex aetiology, not necessarily ascribable to the dysregulation of a single biological pathway. Indeed, pathway analysis of differentially expressed genes in CHD4 deficient cTEC identified several pathways with known relevance in TEC biology including, but not limited to, retinoid metabolism and ECM organisation and function.

Tight regulation of retinoic acid signalling is important for thymic development and function (310, 331-334). When retinoic acid signalling is abrogated in TEC the resulting phenotype is highly reminiscent of that observed in the CHD4 cKO, with thymic hypocellularity, an expansion of the cTEC compartment and a reduction in mTEC, increased cTEC proliferation, increased SCA1⁺ cTEC, and a moderate reduction in generation of SPCD4 and SPCD8 thymocytes (310). In the present study, functional evidence that the retinoic acid pathway is perturbed in CHD4 deficient TEC came from experiments in which FTOCs were supplemented with exogenous retinoic acid, showing that the addition of retinoic acid rescued the proportion of mTEC in CHD4 deficient FTOCs to levels comparable with WT. Taken together, the phenotyping, RNA-seq and functional FTOC data presented in this study, alongside previously published

data examining the role of retinoic acid signalling in TEC, are highly suggestive that dysregulated retinoic acid signalling is at least partly responsible for the TEC developmental defects observed when CHD4 is deficient in TEC.

The ECM forms an important part of the thymic microenvironment contributing not only to the structural organisation of the thymus, but also its function (338). Indeed, the use of native decellularised thymic ECM, as opposed to synthetic scaffold, leads to improved TEC maturation and thymopoiesis in thymic organoids (355, 356). Aside from its effect on TEC maturation, the thymic ECM contributes to thymopoiesis by promoting TEC::thymocyte adhesion as well as thymocyte migration (340, 341, 357, 358). Following on from the pathway analysis, which found genes relating to ECM organisation and function to be particularly dysregulated in CHD4 deficient cTEC, IHC of cKO postnatal thymic sections revealed abnormal composition of the thymic ECM with reduced levels of the pro-adhesive FN1 and increased levels of TNC, an anti-adhesive ECM glycoprotein, which antagonises the function of FN1 (328). FN1, produced by TEC, has been shown to promote adhesion to thymocytes via binding to VLA-4 and VLA-5 expressed on the surface of developing thymocytes (338, 339). The critical importance for TEC::thymocyte interactions for the maintenance of the TEC scaffold has been demonstrated in other mouse models in which disruptions of such interactions have been implicated in the development of intrathymic epithelial free areas (359, 360). FN1 is also able to bind chemokines such as CXCL12 and CCL19 and thus promote intrathymic migration of developing thymocytes (340, 341). Therefore,

disruption of the thymic ECM could contribute substantially to the abnormal architectural organisation, thymocyte-filled epithelial cell free zones, and defective thymopoiesis observed in the CHD4 cKO thymus. Although not formally tested, the mechanism could, at least in part, involve impairment of FN1-mediated TEC::thymocyte interactions and thymocyte migration. To functionally test the contribution of an abnormal ECM to the TEC developmental and thymopoietic defects seen in the CHD4 cKO it would be interesting to see whether the provision of native WT decellularised scaffold could rescue the phenotype observed in the cKO. Such experiments are highly technically specialised and time consuming, and were, therefore, beyond the scope of this study.

A complete assessment of the possible mechanisms underlying the TEC developmental defects observed in the absence of CHD4 will additionally require careful probing of the RNA-seq data generated from $CD80^-CD86^-$, $CD80^+CD86^-$ and $CD80^+CD86^+$ mTEC derived from cKO and WT postnatal thymi. These analyses will hopefully highlight pathways for further functional studies investigating the relevance of CHD4 for mTEC lineage maturation. Given that the TEC developmental defects were already established in the embryonic thymus by E17.5, it would also be important to study in more detail the consequences of CHD4 deficiency on embryonic TEC biology. Therefore, future studies will include detailed phenotyping at various embryonic time points prior to E17.5 and stage specific RNA-seq analyses of embryonic TEC.

The primary reason for studying the role of CHD4 in TEC was to establish whether CHD4 plays a role in AIRE-induced PGE. CHD4 had previously been identified as a potential interaction partner of chromatin bound AIRE in an in vitro overexpression system (182). This, together with its roles in nucleosome binding and mobilization, DNA repair and transcriptional regulation, made CHD4 a particularly attractive candidate as a cofactor in AIRE-induced PGE (213, 215, 230). Indeed, a biologically relevant role for CHD4 in AIRE-regulated PGE was supported by transcriptomic analysis of mature CD80⁺CD86⁺ mTEC, which showed that the loss of CHD4 in mature mTEC disproportionately affected the expression of AIRE-induced TRGs in comparison to AIRE-independent TRGs or non-TRGs. Furthermore, a failure of proper T-cell central tolerance induction was demonstrated in CHD4 cKO mice via the detection of autoreactive antibodies and lymphocytic tissue infiltration in mice aged for 6 months post T-cell ablation.

RNA-seq revealed that a number of AIRE-induced TRGs were not detected in CHD4 deficient CD80⁺CD86⁺ mTEC, suggesting that their expression is entirely dependent on the presence of CHD4. A further subset of AIRE-induced TRGs was significantly downregulated in the absence of CHD4, suggesting that CHD4 is required for their full expression. Although, CHD4 deficiency did affect the expression of AIRE itself within mTEC, this is unlikely to account for the extent of the reduction in AIRE-induced PGE observed. This is because the most PGE competent subpopulation of mTEC, that is those that co-express the co-stimulatory molecules CD80 and CD86 together with the highest levels

of AIRE (57), was chosen for the bulk and single cell RNA-seq analyses used to assess PGE. The difference in the frequency of AIRE⁺ cells between cKO and WT CD80⁺CD86⁺ mTEC was modest and unlikely to be of biological consequence with respect to PGE. In support of this, RT-qPCR for AIRE-regulated TRGs in FACS sorted AIRE⁺ mTEC, demonstrated the reduced expression of AIRE-regulated TRGs in CHD4 deficient mTEC was preserved even after controlling for AIRE expression at the protein level.

Comparison of the gene module structure in cKO and WT single CD80⁺CD86⁺ mTEC revealed a substantial loss of gene co-expression groups in cKO cells. Building upon the observations and conclusions drawn from Chapter 3 of this thesis, which concerned the transcriptomic patterns underlying PGE in single mTEC, this loss of gene co-expression could impair the ability of CHD4 deficient mTEC to induce self-tolerance by damaging the TRG diversity expressed by single cells. The analyses conducted in Chapter 3 of this thesis could not identify a biological process underlying TRG co-expression in mTEC. Co-expressed genes were not related by chromosomal location, distance within the linear genome, tissue specificity, pathway, nor involvement in particular organ specific autoimmune diseases (282). The loss of gene co-expression groups in the context of CHD4 deficiency could support the hypothesis that different cofactors involved in PGE induce the expression of specific co-expression groups. It would therefore be interesting to study whether the loss of other functional partners of AIRE also affects TRG co-expression structure in single mTEC.

The use of *Foxn1* as the driver for Cre recombinase expression somewhat complicated the analyses aiming to address whether CHD4 functionally interacts with AIRE. This is because, although interesting from the perspective of TEC development, the abnormal TEC differentiation and architectural organisation in the cKO make it difficult to ascribe the defects in PGE and central tolerance induction observed solely to the loss of cooperation between CHD4 and AIRE. Indeed, an abnormal maturational state and/or microenvironment could negatively impact on the ability of CHD4 deficient mTEC to promiscuously express TRGs and to induce T-cell tolerance independent of any functional interaction between CHD4 and AIRE. Although the use of the CD80⁺CD86⁺ mTEC, which constitute the most mature and PGE competent mTEC (57), for the transcriptome analysis of PGE somewhat mitigates against this, future studies are planned in which the effect of CHD4 deficiency on PGE will be examined in the context of normal thymic development. To this end, a mouse model in which CHD4 is knocked out only after proper TEC development has occurred will be used. This will be achieved either in the context of an inducible conditional system or alternatively via the use of a Cre recombinase driver that is only expressed in mature mTEC.

Future work will also seek to further define the mechanisms by which CHD4 cooperates with AIRE in the process of PGE. This does not appear to involve a direct physical interaction between CHD4 and AIRE as demonstrated by PLIC, although an indirect physical interaction, perhaps in the context of a larger

chromatin bound multiprotein complex, remains a possibility and is supported by the ChIP-seq analysis, which found that CHD4 binds the same genomic locations as AIRE. Genes whose expression is positively regulated by CHD4 tend to have a more repressive chromatin context, with decreased H3K4me3 and increased H3K27me3 at their start sites, increased H3K9me3 in their gene bodies and a more closed chromatin conformation; all features that have independently been identified as characteristic of AIRE-induced loci (111, 199). This is consistent with the original hypothesis that CHD4 targets TRG loci marked by repressive histone marks and de-represses such loci in partnership with AIRE. Whether the mechanism of de-repression involves chromatin remodelling and/or the DNA damage response pathway awaits further investigation.

In conclusion, the data presented in this Chapter demonstrates, for the first time, the important and pleiotropic role that CHD4 plays in TEC development, architectural organisation and function. In particular, CHD4 regulates the differentiation of the mTEC lineage from TEC precursors in a mechanism that involves retinoic acid signaling and likely also the correct composition and function of the ECM. Additionally, CHD4 functionally cooperates with AIRE to induce the promiscuous expression of AIRE-induced TRGs found in particularly repressive chromatin contexts and is consequently required for proper T-cell central tolerance induction.

5 Final Discussion

The work presented in this thesis sets out to address several unresolved questions in relation to the transcriptional and epigenetic mechanisms that regulate PGE in TEC, specifically, whether PGE in single mTEC is stochastic or whether co-expression patterns exist, and whether CHD4 cooperates with AIRE to influence the transcription of AIRE-induced TRGs by targeting repressive epigenetic marks.

Whilst previous studies have shown that TEC, as a population, express almost all known protein coding genes (111, 127), precisely how this is regulated at the single cell level remains to be conclusively established (111, 127, 129, 133, 178, 211). Understanding how PGE is regulated within single TEC is an important first step in understanding how these cells are able to retain their cellular identity whilst effecting such a uniquely broad transcriptional programme, which is essential for the generation of a TCR repertoire suitably selected not to elicit autoimmune disease. Furthermore, a thorough knowledge of this process is essential before the system can be therapeutically modulated in order to stimulate thymic regeneration for example after exogenous insults or in the context of age-related thymic involution, as well as for the treatment of autoimmune diseases or alternatively for promoting anti-tumour immunity (100, 361, 362).

Given the heterogeneity of gene expression within mTEC at the single cell level and the very low frequency at which individual TRGs are transcribed

(111, 127, 129, 178, 211), a single-cell approach examining the transcriptomes of thousands of individual mTEC was necessary to provide the resolution and statistical power required to faithfully describe the full complexity of the transcriptional patterns of PGE within mTEC. As such, analysis of the transcriptomes of 6,894 single mTEC revealed 50 robustly defined co-expression groups supporting the hypothesis that PGE in individual mTEC is a highly ordered process. These co-expression groups were reproducible across random samples of the data set, as well as between individual mice independent of their strain. Such reproducible TRG co-expression is likely to reflect an inherent cellular property that introduces bias in which TRGs are expressed together within a cell. Nonetheless, the selection of the TRGs co-expressed within single mTEC was indeterminate with respect to several biological features tested, including gene category, chromosomal location, distance within the linear genome, tissue specificity, autoimmune disease association, and inclusion in a specific biological pathway.

Whilst it was not possible to explain membership to co-expression groups by the various analyses performed, this does not preclude the existence of an underlying mechanism that explains why certain TRGs are co-expressed in single mTEC, particularly since inter-individual reproducibility was robust. Certainly there was evidence that different maturational stages or classes of mTEC activate TRG expression differentially. Furthermore, although co-expressed genes were not related by their location within the linear genome, it is possible that chromatin looping might constitute a mechanism underlying

TRG co-expression in single mTEC. This notion is supported by previous studies showing, via DNA-FISH, that co-expressed TRGs are co-localised within the same nuclear subdomains (133) and, via ChIP-seq, that AIRE acts at superenhancers, which are known to localise to the genes they regulate via looping (198). Future studies examining chromatin conformation, employing techniques such as Hi-C, will help to resolve 3D chromatin organisation and establish conclusively whether chromatin looping could underlie TRG co-expression patterns within single mTEC. Another potential, but not mutually exclusive, mechanism that could explain TRG co-expression in mTEC is that different cofactors involved in PGE could induce the expression of specific co-expression groups. In support of this, the work presented in Chapter 4 of this thesis, revealed substantial loss of gene co-expression groups in single CD80⁺ mTEC deficient in CHD4. It would therefore be interesting to study whether the loss of other functional partners of AIRE also affect TRG co-expression structure in single mTEC.

Spatial analysis of mTEC expressing the AIRE-regulated TRA, GP2, revealed that GP2⁺ mTEC are randomly dispersed within thymic medullary islands. Should the observed random spatial distribution of GP2 presentation be generalisable to other TRAs, this would imply that thymocytes would only need to traverse a relatively limited volume within the thymic medulla in order to be tested against a given TRA.

In aggregate, the analyses examining patterns of PGE by mTEC presented in Chapter 3 of this thesis suggest that, whilst there is reproducible order in the TRGs that single mTEC express, the randomness in their selection and spatial distribution may ensure that a single thymocyte travelling through the thymic medulla may be given the greatest opportunity of encountering any given self-antigen for the purposes of central tolerance induction.

Cell-based clustering of the data presented in Chapter 3 supports and expands significantly upon previous observations that the mTEC compartment is highly heterogeneous, constituting different maturation stages and classes or subtypes of mTEC (294, 299). The transcriptomes of Clusters 1 and 2 are consistent with progenitor-like and pre-AIRE mTEC, respectively (49, 50, 54); Clusters 4, 5 and 6 with AIRE⁺ mature mTEC; Clusters 7 and 8 with post-AIRE mTEC (55, 59, 60, 62, 291, 292); and Cluster 10 with that of the recently described tuft-like mTEC (294, 295). In addition to these known TEC maturational stages or subtypes, several novel mTEC clusters were identified, some of which might represent novel mTEC subpopulations that warrant further experimental characterisation as part of future studies. Cluster 14 constitutes the largest newly described subpopulation (436 observed mTEC, 4.6% of unselected mTEC), enriched for mTEC expressing GP2 and defined by high expression of chemokine ligands *Ccl6*, *Ccl9* and *Ccl20*, as well as the chemokine receptor *Ccr5*. The second largest novel cluster is Cluster 9 (227 cells), which expresses the markers *Ceacam10*, *Cd177* and *Ckm* and shares transcriptomic similarities not only with the terminally differentiated mTEC of

cluster 8 but also with the tuft-like mTEC of cluster 10. A third novel cluster, Cluster 13 (117 cells), highly expresses several genes associated with cilium assembly (*Spag16*, *Wdr34* and *Bbs7*). The remaining novel clusters (Clusters 11, 12 and 15) are defined by few cells (0.3 - 1% of the collected mTEC) and remain largely uncharacterised. The cells in these clusters retained a strong signature of expression of core mTEC genes (gene module 2), and while it is not possible to completely rule out contamination by other classes of TEC (e.g. cortical TEC), other cell types (e.g. fibroblasts) or technical artefacts (e.g. doublets), the repeated observation of these cells across multiple experiments combined with their similarity to other TEC suggested that they represented a rare subpopulation of TEC rather than a contaminant. The ontogeny, spatial distribution and function of these novel mTEC subsets will be characterised as part of further studies.

The work presented Chapter 4 of this thesis concerns the study of mice with a genetically targeted TEC conditional deficiency of CHD4, and revealed, for the first time, the cell intrinsic role of CHD4 in multiple aspects of TEC biology including their development, architectural organisation and function. This has particular clinical relevance as CHD4 has been identified as a therapeutic target in acute myeloid leukaemia and potentially other cancers (363-365). In this context, the data presented in this thesis suggest that defective T-cell development and impaired central tolerance could be potential adverse effects of therapeutic CHD4 inhibition.

CHD4 deficiency in TEC was found to negatively impact on the cells' development from the embryonic time point studied (E17.5) through to postnatal stages. The increased expression of markers associated with TEC progenitor status, together with the observed reduction in the frequency of mTEC, particularly those with a mature phenotype (i.e. AIRE⁺ and CD86⁺), was highly suggestive of a role for CHD4 in regulating the differentiation of the mTEC lineage from TEC precursors. In addition, CHD4 deficiency in TEC was found to result in an aberrant thymic structure, including abnormal cortico-medullary separation and the presence of areas, which appeared to lack the normal epithelial cell scaffold, defined by the presence of K8 or K5 expression, in which developing thymocytes accumulated.

RNA-sequencing of adult TEC revealed that CHD4 regulates the expression of thousands of genes in postnatal TEC suggesting that CHD4 has a pleiotropic role in TEC biology and further that the TEC developmental and organisational defects described are likely to have a complex aetiology, which is unlikely to be consequent on the dysregulation of a single biological pathway. With the aim of more comprehensively describing the TEC developmental and structural defects, which occur when CHD4 is absent, future work will include the analysis of further embryonic time points, starting from the earliest stages of thymic organogenesis. The development and composition of the epithelial cell free areas present in the postnatal thymus will be further delineated via the study of embryonic sections at various stages and the inclusion of additional stains, such as for alternative markers for TEC (e.g. EpCAM, CD205, K14),

fibroblasts, endothelial cells, DCs and B-cells. In addition to detailed phenotyping and structural studies, stage specific RNA-sequencing of embryonic TEC will help to uncover dysregulated expression of genes belonging to biological pathways, which could then be probed functionally in order to define in more detail the mechanisms by which CHD4 contributes to TEC development and structural organisation.

Pathway analysis of the most differentially expressed genes in CHD4 deficient adult cTEC revealed enrichment for genes associated with retinoic acid signalling and the correct organisation and function of the ECM. The maintenance of retinoic acid homeostasis is known to be important for proper thymic development and function (310, 331-334). Furthermore, when retinoic acid signalling is abrogated in TEC via the expression of a dominant negative form of the retinoic acid receptor, RAR α , the resulting phenotype is highly similar to that observed in the CHD4 cKO, with thymic hypocellularity, an expansion of the cTEC compartment with increased proliferation and SCA1 expression, a reduction in mTEC, and a moderate reduction in generation of SP $CD4$ and SP $CD8$ thymocytes (310). Whilst this is highly suggestive that dysregulated retinoic acid signalling plays a role in the TEC developmental defects observed in the context of CHD4 deficiency further functional studies are needed. The FTOC experiments detailed in Subsection 4.7.5 of Chapter 4 will be expanded upon to include the provision of an extrinsic stimulus for mTEC development and maturation, namely via the addition of recombinant RANK-L. These analyses will additionally include a more detailed phenotypic

analysis of TEC, so that the effect of exogenous retinoic acid on cTEC numbers and proliferation, as well as on mTEC numbers and maturation, can be determined.

One of the main functional readouts of TEC function in this study was their ability to support regular thymocyte development. Detailed phenotyping of the thymocyte and peripheral T-cell compartments of mice with TEC conditional CHD4 deficiency revealed multiple defects, the majority of which were small albeit statistically significant. For example, there were small reductions in the generation of SP thymocytes and defective maturation within the committed SPCHD4 lineage. The observed accumulation of CD69⁺TCRb^{-/lo} thymocytes alongside a deficiency of CD69⁺TCRb^{hi} thymocytes was suggestive of decreased positive selection (304), and the small increase of Helios⁺ thymocytes of increased negative selection (86). Although, the latter could also be reflective of decreased clearance of negatively selected cells within the CHD4 deficient thymic microenvironment rather than truly increased negative selection per se. Therefore, it will be important in future studies to probe thymocyte selection in this model, for example via the use of a TCR transgenic system such the HY TCR-transgenic mouse model, which confers reactivity to the male HY antigen presented on a particular MHC class I or II antigen (90, 304). In addition, whether reduced thymic egress or defective homeostatic proliferation contribute to the peripheral T-cell lymphopaenia observed in the CHD4 TEC cKO model will be also addressed.

The primary reason for studying CHD4 in TEC was to establish its role in AIRE-regulated PGE as CHD4 has previously been identified as a putative member of the AIRE containing multimolecular complex (182). CHD4 was found to cooperate functionally with AIRE to induce the promiscuous expression of AIRE-induced TRGs and instruct proper T-cell central tolerance induction. The loss of gene co-expression groups in CHD4 deficient mTEC supports the hypothesis that co-factors of AIRE influence TRG co-expression patterns in single mTEC. Although, CHD4 and AIRE were not found to physically interact via PLIC, this assay is only able to detect interactions over a maximal distance of 30-40nm (189). It is therefore remains possible that CHD4 and AIRE interact indirectly in the context of a large chromatin bound multiprotein complex. In support of this notion, the genomic binding sites for CHD4 and AIRE determined by ChIP-seq were highly correlated, however, further work is needed in the form in order to define the physical interactomes of CHD4 and AIRE in primary mTEC.

Consistent with the original hypothesis that CHD4 targets TRG loci marked by repressive histone marks, CHD4 was specifically found to induce the expression of genes found in particularly repressive chromatin contexts, although the precise mechanisms by which CHD4 de-represses such loci in partnership with AIRE remain to be established. To this end, chromatin accessibility and DNA damage repair will be assessed in the CHD4 cKO TEC compared to their WT counterparts.

Finally, further studies are planned in order to be able to confidently ascribe the defects in AIRE-regulated PGE observed to the cooperation between CHD4 and AIRE in mTEC rather than to the distorted thymic microenvironment, which results from knocking out CHD4 at such an early stage in TEC development. To this end, mice expressing Cre recombinase under the control of *Csnb*, an AIRE-independent TRG expressed in the thymus only in mature mTEC, will be crossed to the *Chd4^{fl/fl}* strain and a reporter strain (294). In this model, thymic organogenesis and TEC maturation would be expected to proceed normally, with CHD4 knocked out of approximately 50% of mature mTECs, which will be fluorescently marked. These cells will then be probed for their expression of AIRE-induced TRGs.

Taken together, the work presented in this thesis contributes to our understanding of TEC development and function, particularly in the context of the molecular mechanisms that drive PGE in TEC and consequently thymic TCR repertoire selection and immunological tolerance by providing a comprehensive mirror of immunological self. The knowledge gained provides insights into the molecular mechanisms within TEC that prevent autoimmunity and may also aid in identifying targets for pharmacological intervention to modulate PGE in TEC with relevance to both autoimmune disease and anti-tumour immunity.

References

1. Barthlott T, Keller MP, Krenger W, Hollander GA. A short primer on early molecular and cellular events in thymus organogenesis and replacement. *Swiss Medical Weekly*. 2006;136(23-24):365-9.
2. Rodewald H-R. Thymus Organogenesis. *Annual Review of Immunology*. 2008;26(1):355-88.
3. Boehm T, Swann JB. Origin and evolution of adaptive immunity. *Annu Rev Anim Biosci*. 2014;2:259-83.
4. Bonilla FA, Oettgen HC. Adaptive immunity. *J Allergy Clin Immunol*. 2010;125(2 Suppl 2):S33-40.
5. Gordon J, Manley NR. Mechanisms of thymus organogenesis and morphogenesis. *Development*. 2011;138(18):3865-78.
6. Gordon J, BAR, Blackburn C. C. and Manley N. R. Gcm2 and Foxn1 mark early parathyroid- and thymus-specific domains in the developing third pharyngeal pouch. *Mechanisms of Development*. 2001;102:3.
7. Manley N RE, Blackburn C, Condie B, Sage J. Structure and function of the thymic microenvironment. *Frontiers in Bioscience*. 2011;16(1):2461.
8. Gordon J, Wilson VA, Blair NF, Sheridan J, Farley A, Wilson L, et al. Functional evidence for a single endodermal origin for the thymic epithelium. *Nature Immunology*. 2004;5(5):546-53.
9. Anderson G, Takahama Y. Thymic epithelial cells: working class heroes for T cell development and repertoire selection. *Trends Immunol*. 2012;33(6):256-63.

10. Abramson J, Anderson G. Thymic Epithelial Cells. *Annu Rev Immunol.* 2017;35:85-118.
11. Takahama Y. Journey through the thymus: stromal guides for T-cell development and selection. *Nature Reviews Immunology.* 2006;6(2):127-35.
12. van Ewijk W, Shores EW, Singer A. Crosstalk in the mouse thymus. *Immunology Today.* 1994;15(5):214-7.
13. Lopes N, Serge A, Ferrier P, Irla M. Thymic Crosstalk Coordinates Medulla Organization and T-Cell Tolerance Induction. *Front Immunol.* 2015;6:365.
14. Rossi SW, Jenkinson WE, Anderson G, Jenkinson EJ. Clonal analysis reveals a common progenitor for thymic cortical and medullary epithelium. *Nature.* 2006;441(7096):988-91.
15. Bleul CC, Corbeaux T, Reuter A, Fisch P, Monting JS, Boehm T. Formation of a functional thymus initiated by a postnatal epithelial progenitor cell. *Nature.* 2006;441(7096):992-6.
16. Bennett AR, Farley A, Blair NF, Gordon J, Sharp L, Blackburn CC. Identification and Characterization of Thymic Epithelial Progenitor Cells. *Immunity.* 2002;16(6):803-14.
17. Gill J, Malin M, Holländer GA, Boyd R. Generation of a complete thymic microenvironment by MTS24+ thymic epithelial cells. *Nature Immunology.* 2002;3:635.
18. Alves NL, Takahama Y, Ohigashi I, Ribeiro AR, Baik S, Anderson G, et al. Serial progression of cortical and medullary thymic epithelial microenvironments. *Eur J Immunol.* 2014;44(1):16-22.

19. Baik S, Jenkinson EJ, Lane PJ, Anderson G, Jenkinson WE. Generation of both cortical and Aire(+) medullary thymic epithelial compartments from CD205(+) progenitors. *Eur J Immunol*. 2013;43(3):589-94.
20. Ohigashi I, Zuklys S, Sakata M, Mayer CE, Hamazaki Y, Minato N, et al. Adult Thymic Medullary Epithelium Is Maintained and Regenerated by Lineage-Restricted Cells Rather Than Bipotent Progenitors. *Cell Rep*. 2015;13(7):1432-43.
21. Mayer CE, Zuklys S, Zhanybekova S, Ohigashi I, Teh HY, Sansom SN, et al. Dynamic spatio-temporal contribution of single beta5t+ cortical epithelial precursors to the thymus medulla. *Eur J Immunol*. 2016;46(4):846-56.
22. Rossi SW, Kim M-Y, Leibbrandt A, Parnell SM, Jenkinson WE, Glanville SH, et al. RANK signals from CD4+3-inducer cells regulate development of Aire-expressing epithelial cells in the thymic medulla. *The Journal of Experimental Medicine*. 2007;204(6):1267-72.
23. Hamazaki Y, Fujita H, Kobayashi T, Choi Y, Scott HS, Matsumoto M, et al. Medullary thymic epithelial cells expressing Aire represent a unique lineage derived from cells expressing claudin. *Nat Immunol*. 2007;8(3):304-11.
24. Sekai M, Hamazaki Y, Minato N. Medullary thymic epithelial stem cells maintain a functional thymus to ensure lifelong central T cell tolerance. *Immunity*. 2014;41(5):753-61.
25. Hamazaki Y. Adult thymic epithelial cell (TEC) progenitors and TEC stem cells: Models and mechanisms for TEC development and maintenance. *Eur J Immunol*. 2015;45(11):2985-93.

26. Vaidya HJ, Briones Leon A, Blackburn CC. FOXP1 in thymus organogenesis and development. *European Journal of Immunology*. 2016;46(8):1826-37.
27. Manley NR, Capecchi MR. The role of Hoxa-3 in mouse thymus and thyroid development. *Development*. 1995;121(7):1989.
28. Manley NR, Capecchi MR. Hox group 3 paralogs regulate the development and migration of the thymus, thyroid, and parathyroid glands. *Dev Biol*. 1998;195(1):1-15.
29. Su D, Ellis S, Napier A, Lee K, Manley NR. Hoxa3 and pax1 regulate epithelial cell death and proliferation during thymus and parathyroid organogenesis. *Dev Biol*. 2001;236(2):316-29.
30. Xu P-X, Zheng W, Laclef C, Maire P, Maas RL, Peters H, et al. Eya1 is required for the morphogenesis of mammalian thymus, parathyroid and thyroid. *Development*. 2002;129(13):3033.
31. Wallin J, Eibel H, Neubuser A, Wilting J, Koseki H, Balling R. Pax1 is expressed during development of the thymus epithelium and is required for normal T-cell maturation. *Development*. 1996;122(1):23.
32. Hetzer-Egger C, Schorpp M, Haas-Assenbaum A, Balling R, Peters H, Boehm T. Thymopoiesis requires Pax9 function in thymic epithelial cells. *European Journal of Immunology*. 2002;32(4):1175-81.
33. Zou D, Silvius D, Davenport J, Grifone R, Maire P, Xu PX. Patterning of the third pharyngeal pouch into thymus/parathyroid by Six and Eya1. *Dev Biol*. 2006;293(2):499-512.

34. Jerome LA, Papaioannou VE. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nat Genet.* 2001;27(3):286-91.
35. Romano R, Palamaro L, Fusco A, Giardino G, Gallo V, Del Vecchio L, et al. FOXP1: A Master Regulator Gene of Thymic Epithelial Development Program. *Front Immunol.* 2013;4:187.
36. Žuklys S, Handel A, Zhanybekova S, Govani F, Keller M, Maio S, et al. Foxp1 regulates key target genes essential for T cell development in postnatal thymic epithelial cells. *Nature Immunology.* 2016;17:1206.
37. Blackburn CC, Augustine CL, Li R, Harvey RP, Malin MA, Boyd RL, et al. The nu gene acts cell-autonomously and is required for differentiation of thymic epithelial progenitors. *Proceedings of the National Academy of Sciences.* 1996;93(12):5742.
38. Nowell CS, Bredenkamp N, Tetélin S, Jin X, Tischner C, Vaidya H, et al. Foxp1 Regulates Lineage Progression in Cortical and Medullary Thymic Epithelial Cells But Is Dispensable for Medullary Sublineage Divergence. *PLOS Genetics.* 2011;7(11):e1002348.
39. Chen L, Xiao S, Manley NR. Foxp1 is required to maintain the postnatal thymic microenvironment in a dosage-sensitive manner. *Blood.* 2009;113(3):567.
40. O'Neill KE, Bredenkamp N, Tischner C, Vaidya HJ, Stenhouse FH, Peddie CD, et al. Foxp1 Is Dynamically Regulated in Thymic Epithelial Cells during Embryogenesis and at the Onset of Thymic Involution. *PLoS One.* 2016;11(3):e0151666.

41. Nehls M, Pfeifer D, Schorpp M, Hedrich H, Boehm T. New member of the winged-helix protein family disrupted in mouse and rat nude mutations. *Nature*. 1994;372(6501):103-7.
42. Flanagan SP. 'Nude', a new hairless gene with pleiotropic effects in the mouse. *Genetical Research*. 2009;8(3):295-309.
43. Wortis HH, Nehlsen S, Owen JJ. ABNORMAL DEVELOPMENT OF THE THYMUS IN NUDE MICE. *The Journal of Experimental Medicine*. 1971;134(3):681.
44. Pignata C, Gaetaniello L, Masci AM, Frank J, Christiano A, Matrecano E, et al. Human equivalent of the mouse Nude/SCID phenotype: long-term evaluation of immunologic reconstitution after bone marrow transplantation. *Blood*. 2001;97(4):880-5.
45. Pignata C, Fiore M, Guzzetta V, Castaldo A, Sebastio G, Porta F, et al. Congenital Alopecia and nail dystrophy associated with severe functional T-cell immunodeficiency in two sibs. *Am J Med Genet*. 1996;65(2):167-70.
46. Frank J, Pignata C, Panteleyev AA, Prowse DM, Baden H, Weiner L, et al. Exposing the human nude phenotype. *Nature*. 1999;398(6727):473-4.
47. Rota IA, Dhalla F. FOXP1 deficient nude severe combined immunodeficiency. *Orphanet J Rare Dis*. 2017;12(1):6.
48. DePreter MG, Blair NF, Gaskell TL, Nowell CS, Davern K, Pagliocca A, et al. Identification of Plet-1 as a specific marker of early thymic epithelial progenitor cells. *Proc Natl Acad Sci U S A*. 2008;105(3):961-6.

49. Ulyanchenko S, O'Neill KE, Medley T, Farley AM, Vaidya HJ, Cook AM, et al. Identification of a Bipotent Epithelial Progenitor Population in the Adult Thymus. *Cell Rep.* 2016;14(12):2819-32.
50. Wong K, Lister NL, Barsanti M, Lim JM, Hammett MV, Khong DM, et al. Multilineage potential and self-renewal define an epithelial progenitor cell population in the adult thymus. *Cell Rep.* 2014;8(4):1198-209.
51. Ohigashi I, Zuklys S, Sakata M, Mayer CE, Zhanybekova S, Murata S, et al. Aire-expressing thymic medullary epithelial cells originate from beta5t-expressing progenitor cells. *Proc Natl Acad Sci U S A.* 2013;110(24):9885-90.
52. Ribeiro AR, Rodrigues PM, Meireles C, Di Santo JP, Alves NL. Thymocyte Selection Regulates the Homeostasis of IL-7-Expressing Thymic Cortical Epithelial Cells In Vivo. *The Journal of Immunology.* 2013;191(3):1200-9.
53. Akiyama N, Takizawa N, Miyauchi M, Yanai H, Tateishi R, Shinzawa M, et al. Identification of embryonic precursor cells that differentiate into thymic epithelial cells expressing autoimmune regulator. *J Exp Med.* 2016;213(8):1441-58.
54. Onder L, Nindl V, Scandella E, Chai Q, Cheng HW, Caviezel-Firner S, et al. Alternative NF-kappaB signaling regulates mTEC differentiation from podoplanin-expressing precursors in the cortico-medullary junction. *Eur J Immunol.* 2015;45(8):2218-31.
55. Michel C, Miller CN, Kuchler R, Brors B, Anderson MS, Kyewski B, et al. Revisiting the Road Map of Medullary Thymic Epithelial Cell Differentiation. *J Immunol.* 2017;199(10):3488-503.

56. Gabler J, Arnold J, Kyewski B. Promiscuous gene expression and the developmental dynamics of medullary thymic epithelial cells. *Eur J Immunol.* 2007;37(12):3363-72.
57. Derbinski J, Gabler J, Brors B, Tierling S, Jonnakuty S, Hergenhausen M, et al. Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. *J Exp Med.* 2005;202(1):33-45.
58. Nishikawa Y, Hirota F, Yano M, Kitajima H, Miyazaki J, Kawamoto H, et al. Biphaseic Aire expression in early embryos and in medullary thymic epithelial cells before end-stage terminal differentiation. *J Exp Med.* 2010;207(5):963-71.
59. Yano M, Kuroda N, Han H, Meguro-Horike M, Nishikawa Y, Kiyonari H, et al. Aire controls the differentiation program of thymic epithelial cells in the medulla for the establishment of self-tolerance. *J Exp Med.* 2008;205(12):2827-38.
60. Metzger TC, Khan IS, Gardner JM, Mouchess ML, Johannes KP, Krawisz AK, et al. Lineage tracing and cell ablation identify a post-Aire-expressing thymic epithelial cell population. *Cell Rep.* 2013;5(1):166-79.
61. Nishikawa Y, Nishijima H, Matsumoto M, Morimoto J, Hirota F, Takahashi S, et al. Temporal lineage tracing of Aire-expressing cells reveals a requirement for Aire in their maturation program. *J Immunol.* 2014;192(6):2585-92.
62. Wang X, Laan M, Bichele R, Kisand K, Scott HS, Peterson P. Post-Aire maturation of thymic medullary epithelial cells involves selective expression of keratinocyte-specific autoantigens. *Front Immunol.* 2012;3(March):19.

63. Akiyama T, Shinzawa M, Akiyama N. TNF receptor family signaling in the development and functions of medullary thymic epithelial cells. *Front Immunol*. 2012;3:278.
64. Lucas B, McCarthy NI, Baik S, Cosway E, James KD, Parnell SM, et al. Control of the thymic medulla and its influence on alphabetaT-cell development. *Immunol Rev*. 2016;271(1):23-37.
65. Akiyama T, Shimo Y, Yanai H, Qin J, Ohshima D, Maruyama Y, et al. The Tumor Necrosis Factor Family Receptors RANK and CD40 Cooperatively Establish the Thymic Medullary Microenvironment and Self-Tolerance. *Immunity*. 2008;29(3):423-37.
66. Roberts NA, White AJ, Jenkinson WE, Turchinovich G, Nakamura K, Withers DR, et al. Rank signaling links the development of invariant gammadelta T cell progenitors and Aire(+) medullary epithelium. *Immunity*. 2012;36(3):427-37.
67. Desanti GE, Cowan JE, Baik S, Parnell SM, White AJ, Penninger JM, et al. Developmentally regulated availability of RANKL and CD40 ligand reveals distinct mechanisms of fetal and adult cross-talk in the thymus medulla. *J Immunol*. 2012;189(12):5519-26.
68. Hikosaka Y, Nitta T, Ohigashi I, Yano K, Ishimaru N, Hayashi Y, et al. The Cytokine RANKL Produced by Positively Selected Thymocytes Fosters Medullary Thymic Epithelial Cells that Express Autoimmune Regulator. *Immunity*. 2008;29(3):438-50.

69. Irla M, Hugues S, Gill J, Nitta T, Hikosaka Y, Williams IR, et al. Autoantigen-specific interactions with CD4⁺ thymocytes control mature medullary thymic epithelial cell cellularity. *Immunity*. 2008;29(3):451-63.
70. White AJ, Nakamura K, Jenkinson WE, Saini M, Sinclair C, Seddon B, et al. Lymphotoxin signals from positively selected thymocytes regulate the terminal differentiation of medullary thymic epithelial cells. *J Immunol*. 2010;185(8):4769-76.
71. Boehm T, Scheu S, Pfeffer K, Bleul CC. Thymic medullary epithelial cell differentiation, thymocyte emigration, and the control of autoimmunity require lympho-epithelial cross talk via LTbetaR. *J Exp Med*. 2003;198(5):757-69.
72. Lind EF, Prockop SE, Porritt HE, Petrie HT. Mapping Precursor Movement through the Postnatal Thymus Reveals Specific Microenvironments Supporting Defined Stages of Early Lymphoid Development. *The Journal of Experimental Medicine*. 2001;194(2):127-34.
73. Rossi FMV, Corbel SY, Merzaban JS, Carlow DA, Gossens K, Duenas J, et al. Recruitment of adult thymic progenitors is regulated by P-selectin and its ligand PSGL-1. *Nature Immunology*. 2005;6(6):626-34.
74. Hozumi K, Mailhos C, Negishi N, Hirano K-i, Yahata T, Ando K, et al. Delta-like 4 is indispensable in thymic environment specific for T cell development. *The Journal of Experimental Medicine*. 2008;205(11):2507-13.
75. Koch U, Fiorini E, Benedito R, Besseyrias V, Schuster-Gossler K, Pierres M, et al. Delta-like 4 is the essential, nonredundant ligand for Notch1 during thymic T cell lineage commitment. *The Journal of Experimental Medicine*. 2008;205(11):2515-23.

76. Peschon JJ MP, Grabstein KH, Ramsdell FJ, Maraskovsky E, Gliniak BC, Park LS, Ziegler SF, Williams DE, Ware CB, Meyer JD, Davison BL. Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice. *Journal of Experimental Medicine*. 1994;180(5):6.
77. Shitara S, Hara T, Liang B, Wagatsuma K, Zuklys S, Hollander GA, et al. IL-7 Produced by Thymic Epithelial Cells Plays a Major Role in the Development of Thymocytes and TCR + Intraepithelial Lymphocytes. *The Journal of Immunology*. 2013;190(12):6173-9.
78. Koch U, Radtke F. Mechanisms of T cell development and transformation. *Annu Rev Cell Dev Biol*. 2011;27:539-62.
79. von Boehmer H. Unique features of the pre-T-cell receptor α -chain: not just a surrogate. *Nature Reviews Immunology*. 2005;5:571.
80. Yamasaki S, Saito T. Molecular basis for pre-TCR-mediated autonomous signaling. *Trends Immunol*. 2007;28(1):39-43.
81. MacDonald HR, Radtke F, Wilson A. T cell fate specification and alphabeta/gammadelta lineage commitment. *Curr Opin Immunol*. 2001;13(2):219-24.
82. Xiong J, Armato MA, Yankee TM. Immature single-positive CD8+ thymocytes represent the transition from Notch-dependent to Notch-independent T-cell development. *Int Immunol*. 2011;23(1):55-64.
83. Starr TK, Jameson SC, Hogquist KA. Positive and negative selection of T cells. *Annual Review of Immunology*. 2003;21(1):139-76.

84. Lucas B, Germain RN. Unexpectedly Complex Regulation of CD4/CD8 Coreceptor Expression Supports a Revised Model for CD4+CD8+ Thymocyte Differentiation. *Immunity*. 1996;5(5):461-77.
85. Germain RN. T-cell development and the CD4–CD8 lineage decision. *Nature Reviews Immunology*. 2002;2:309.
86. Daley SR, Hu DY, Goodnow CC. Helios marks strongly autoreactive CD4+ T cells in two major waves of thymic deletion distinguished by induction of PD-1 or NF-kappaB. *J Exp Med*. 2013;210(2):269-85.
87. Klein L, Hinterberger M, Wirnsberger G, Kyewski B. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nature Reviews Immunology*. 2009;9(12):833-44.
88. Murata S, Sasaki K, Kishimoto T, Niwa S, Hayashi H, Takahama Y, et al. Regulation of CD8+ T cell development by thymus-specific proteasomes. *Science*. 2007;316(5829):1349-53.
89. Nitta T, Murata S, Sasaki K, Fujii H, Ripen AM, Ishimaru N, et al. Thymoproteasome shapes immunocompetent repertoire of CD8+ T cells. *Immunity*. 2010;32(1):29-40.
90. Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol*. 2014;14(6):377-91.
91. Nakagawa T, Roth W, Wong P, Nelson A, Farr A, Deussing J, et al. Cathepsin L: Critical Role in li Degradation and CD4 T Cell Selection in the Thymus. *Science*. 1998;280(5362):450.

92. Gommeaux J, Gregoire C, Nguessan P, Richelme M, Malissen M, Guerder S, et al. Thymus-specific serine protease regulates positive selection of a subset of CD4⁺ thymocytes. *Eur J Immunol*. 2009;39(4):956-64.
93. Honey K, Nakagawa T, Peters C, Rudensky A. Cathepsin L regulates CD4⁺ T cell selection independently of its effect on invariant chain: a role in the generation of positively selecting peptide ligands. *The Journal of experimental medicine*. 2002;195(10):1349-58.
94. Ueno T, Saito F, Gray DHD, Kuse S, Hieshima K, Nakano H, et al. CCR7 Signals Are Essential for Cortex–Medulla Migration of Developing Thymocytes. *The Journal of Experimental Medicine*. 2004;200(4):493-505.
95. Nitta T, Nitta S, Lei Y, Lipp M, Takahama Y. CCR7-mediated migration of developing thymocytes to the medulla is essential for negative selection to tissue-restricted antigens. *Proc Natl Acad Sci U S A*. 2009;106(40):17129-33.
96. Hogquist KA, Jameson SC. The self-obsession of T cells: how TCR signaling thresholds affect fate 'decisions' and effector function. *Nat Immunol*. 2014;15(9):815-23.
97. Hogquist KA, Xing Y, Hsu FC, Shapiro VS. T Cell Adolescence: Maturation Events Beyond Positive Selection. *J Immunol*. 2015;195(4):1351-7.
98. Xu X, Zhang S, Li P, Lu J, Xuan Q, Ge Q. Maturation and emigration of single-positive thymocytes. *Clin Dev Immunol*. 2013;2013:282870.
99. Xing Y, Wang X, Jameson SC, Hogquist KA. Late stages of T cell maturation in the thymus involve NF-kappaB and tonic type I interferon signaling. *Nat Immunol*. 2016;17(5):565-73.

100. Matloubian M, Lo CG, Cinamon G, Lesneski MJ, Xu Y, Brinkmann V, et al. Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature*. 2004;427(6972):355-60.
101. Allende ML, Dreier JL, Mandala S, Proia RL. Expression of the sphingosine 1-phosphate receptor, S1P1, on T-cells controls thymic emigration. *J Biol Chem*. 2004;279(15):15396-401.
102. James KD, Jenkinson WE, Anderson G. T-cell egress from the thymus: Should I stay or should I go? *Journal of Leukocyte Biology*. 2018;104(2):275-84.
103. Egerton M SR, Shortman K. Kinetics of mature T-cell development in the thymus. *Proc Natl Acad Sci U S A*. 1990;87(7):4.
104. Scollay RG, Butcher EC, Weissman IL. Thymus cell migration: Quantitative aspects of cellular traffic from the thymus to the periphery in mice. *European Journal of Immunology*. 1980;10(3):210-8.
105. Sprent J, Kishimoto H. The thymus and central tolerance. *Philos Trans R Soc Lond B Biol Sci*. 2001;356(1409):609-16.
106. Klein L, Robey EA, Hsieh CS. Central CD4(+) T cell tolerance: deletion versus regulatory T cell differentiation. *Nat Rev Immunol*. 2018.
107. Yamano T, Nedic J, Hinterberger M, Steinert M, Koser S, Pinto S, et al. Thymic B Cells Are Licensed to Present Self Antigens for Central T Cell Tolerance Induction. *Immunity*. 2015;42(6):1048-61.
108. Yamano T, Steinert M, Klein L. Thymic B Cells and Central T Cell Tolerance. *Front Immunol*. 2015;6:376.

109. Cosway EJ, Lucas B, James KD, Parnell SM, Carvalho-Gaspar M, White AJ, et al. Redefining thymus medulla specialization for central tolerance. *J Exp Med*. 2017;214(11):3183-95.
110. Derbinski J, Schulte A, Kyewski B, Klein L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nature Immunology*. 2001;2(11):1032-9.
111. Sansom SN, Shikama-Dorn N, Zhanybekova S, Nusspaumer G, Macaulay IC, Deadman ME, 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*. 2014;24(12):1918-31.
112. Li J, Park J, Foss D, Goldschneider I. Thymus-homing peripheral dendritic cells constitute two of the three major subsets of dendritic cells in the steady-state thymus. *J Exp Med*. 2009;206(3):607-22.
113. Klein L, Hinterberger M, von Rohrscheidt J, Aichinger M. Autonomous versus dendritic cell-dependent contributions of medullary thymic epithelial cells to central tolerance. *Trends Immunol*. 2011;32(5):188-93.
114. Skogberg G, Telemo E, Ekwall O. Exosomes in the Thymus: Antigen Transfer and Vesicles. *Front Immunol*. 2015;6:366.
115. Hadeiba H, Lahl K, Edalati A, Oderup C, Habtezion A, Pachynski R, et al. Plasmacytoid dendritic cells transport peripheral antigens to the thymus to promote central tolerance. *Immunity*. 2012;36(3):438-50.
116. Heath WR AJ, Hoffmann MW, Schonrich G, Hammerling G, Arnold B MJ. Autoimmune diabetes as a consequence of locally produced interleukin-2.

. Nature. 1992;359:2.

117. von Herrath MG, Dockter J, Oldstone MBA. How virus induces a rapid or slow onset insulin-dependent diabetes mellitus in a transgenic model. Immunity. 1994;1(3):231-42.

118. Jolicoeur C, Hanahan D, Smith KM. T-cell tolerance toward a transgenic beta-cell antigen and transcription of endogenous pancreatic genes in thymus. Proceedings of the National Academy of Sciences. 1994;91(14):6707-11.

119. Hanahan D. Peripheral-antigen-expressing cells in thymic medulla: factors in self-tolerance and autoimmunity. Curr Opin Immunol. 1998;10(6):656-62.

120. Pugliese A, Zeller M, Fernandez A, Jr., Zalcberg LJ, Bartlett RJ, Ricordi C, et al. The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. Nat Genet. 1997;15(3):293-7.

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

122. Heath VL, Moore NC, Parnell SM, Mason DW. Intrathymic expression of genes involved in organ specific autoimmune disease. J Autoimmun. 1998;11(4):309-18.

123. Smith K OD, Hirose R, Hanahan D. Pancreatic gene expression in rare cells of thymic medulla: evidence for functional contribution to T cell tolerance. International Immunology. 1997;9(9):1355-65.

124. Throsby M H-DF, Chevenne D, Goya R, Dardenne M, Pleau J. Pancreatic Hormone Expression in the Murine Thymus: Localization in Dendritic Cells and Macrophages. *Endocrinology*. 1998;139(5):2399-406.
125. Klein L KT, Ruther U, Kyewski B. CD4 T Cell Tolerance to Human C-reactive Protein, an Inducible Serum Protein, Is Mediated by Medullary Thymic Epithelium. *Journal of Experimental Medicine*. 1998;188(1):11.
126. Klein L RB, Kyewski B. Sampling of complementing self-antigen pools by thymic stromal cells maximizes the scope of central T cell tolerance. *European Journal of Immunology*. 2001;31(8):10.
127. Brennecke P, Reyes A, Pinto S, Rattay K, Nguyen M, Küchler R, et al. Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. *Nature Immunology*. 2015;16(9):933-41.
128. Kyewski B, Klein L. A Central Role for Central Tolerance. *Annual Review of Immunology*. 2006;24(1):571-606.
129. Meredith M, Zemmour D, Mathis D, Benoist C. Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nat Immunol*. 2015;16(9):942-9.
130. DeVoss J, Hou Y, Johannes K, Lu W, Liou GI, Rinn J, et al. Spontaneous autoimmunity prevented by thymic expression of a single self-antigen. *J Exp Med*. 2006;203(12):2727-35.
131. Gavanescu I, Kessler B, Ploegh H, Benoist C, Mathis D. Loss of Aire-dependent thymic expression of a peripheral tissue antigen renders it a target of autoimmunity. *Proc Natl Acad Sci U S A*. 2007;104(11):4583-7.

132. Fan Y, Rudert WA, Grupillo M, He J, Sisino G, Trucco M. Thymus-specific deletion of insulin induces autoimmune diabetes. *EMBO J*. 2009;28(18):2812-24.
133. Pinto S, Michel C, Schmidt-Glenewinkel H, Harder N, Rohr K, Wild S, et al. Overlapping gene coexpression patterns in human medullary thymic epithelial cells generate self-antigen diversity. *Proceedings of the National Academy of Sciences*. 2013;110(37):E3497-E505.
134. Aaltonen J, Björnses P, Perheentupa J, Horelli-Kuitunen N, Palotie A, Peltonen L, et al. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nature Genetics*. 1997;17(4):399-403.
135. Nagamine K, Peterson P, Scott HS, Kudoh J, Minoshima S, Heino M, et al. Positional cloning of the APECED gene. *Nature Genetics*. 1997;17(4):393-8.
136. Peterson P, Nagamine K, Scott H, Heino M, Kudoh J, Shimizu N, et al. APECED: a monogenic autoimmune disease providing new clues to self-tolerance. *Immunology Today*. 1998;19(9):384-6.
137. Ahonen P, Myllarniemi S, Sipila I, Perheentupa J. Clinical variation of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) in a series of 68 patients. *N Engl J Med*. 1990;322(26):1829-36.
138. Kisand K, Peterson P. Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy: known and novel aspects of the syndrome. *Annals of the New York Academy of Sciences*. 2011;1246(1):77-91.

139. Mathis D, Benoist C. Aire. *Annual Review of Immunology*. 2009;27(1):287-312.
140. Abramson J, Husebye ES. Autoimmune regulator and self-tolerance - molecular and clinical aspects. *Immunol Rev*. 2016;271(1):127-40.
141. Ramsey C WO, Puhakka L, Halonen M, Moro A, Kämpe O, Eskelin P, Peltto-Huikko M, Peltonen L. Aire deficient mice develop multiple features of APECED phenotype and show altered immune response. *Human Molecular Genetics*. 2002;11(4):397-409.
142. Anderson MS. Projection of an Immunological Self Shadow Within the Thymus by the Aire Protein. *Science*. 2002;298(5597):1395-401.
143. Pontynen N, Miettinen A, Arstila TP, Kampe O, Alimohammadi M, Vaarala O, et al. Aire deficient mice do not develop the same profile of tissue-specific autoantibodies as APECED patients. *J Autoimmun*. 2006;27(2):96-104.
144. Jiang W, Anderson MS, Bronson R, Mathis D, Benoist C. Modifier loci condition autoimmunity provoked by Aire deficiency. *J Exp Med*. 2005;202(6):805-15.
145. Puel A, Doffinger R, Natividad A, Chrabieh M, Barcenas-Morales G, Picard C, et al. Autoantibodies against IL-17A, IL-17F, and IL-22 in patients with chronic mucocutaneous candidiasis and autoimmune polyendocrine syndrome type I. *J Exp Med*. 2010;207(2):291-7.
146. Kisand K, Boe Wolff AS, Podkrajsek KT, Tserel L, Link M, Kisand KV, et al. Chronic mucocutaneous candidiasis in APECED or thymoma patients

correlates with autoimmunity to Th17-associated cytokines. *J Exp Med*. 2010;207(2):299-308.

147. Bjorses P, Aaltonen J, Vikman A, Perheentupa J, Ben-Zion G, Chiumello G, et al. Genetic homogeneity of autoimmune polyglandular disease type I. *Am J Hum Genet*. 1996;59(4):879-86.

148. Pearce SH, Cheetham T, Imrie H, Vaidya B, Barnes ND, Bilous RW, et al. A common and recurrent 13-bp deletion in the autoimmune regulator gene in British kindreds with autoimmune polyendocrinopathy type 1. *Am J Hum Genet*. 1998;63(6):1675-84.

149. Rosatelli MC, Meloni A, Meloni A, Devoto M, Cao A, Scott HS, et al. A common mutation in Sardinian autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy patients. *Hum Genet*. 1998;103(4):428-34.

150. Bjorses P, Halonen M, Palvimo JJ, Kolmer M, Aaltonen J, Ellonen P, et al. Mutations in the AIRE gene: effects on subcellular location and transactivation function of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy protein. *Am J Hum Genet*. 2000;66(2):378-92.

151. Stolarski B, Pronicka E, Korniszewski L, Pollak A, Kostrzewa G, Rowinska E, et al. Molecular background of polyendocrinopathy-candidiasis-ectodermal dystrophy syndrome in a Polish population: novel AIRE mutations and an estimate of disease prevalence. *Clin Genet*. 2006;70(4):348-54.

152. Wolff AS, Erichsen MM, Meager A, Magitta NF, Myhre AG, Bollerslev J, et al. Autoimmune polyendocrine syndrome type 1 in Norway: phenotypic variation, autoantibodies, and novel mutations in the autoimmune regulator gene. *J Clin Endocrinol Metab*. 2007;92(2):595-603.

153. Orlova EM, Bukina AM, Kuznetsova ES, Kareva MA, Zakharova EU, Peterkova VA, et al. Autoimmune polyglandular syndrome type 1 in Russian patients: clinical variants and autoimmune regulator mutations. *Horm Res Paediatr*. 2010;73(6):449-57.
154. Cetani F, Barbesino G, Borsari S, Pardi E, Cianferotti L, Pinchera A, et al. A novel mutation of the autoimmune regulator gene in an Italian kindred with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy, acting in a dominant fashion and strongly cosegregating with hypothyroid autoimmune thyroiditis. *J Clin Endocrinol Metab*. 2001;86(10):4747-52.
155. Oftedal BE, Hellesen A, Erichsen MM, Bratland E, Vardi A, Perheentupa J, et al. Dominant Mutations in the Autoimmune Regulator AIRE Are Associated with Common Organ-Specific Autoimmune Diseases. *Immunity*. 2015;42(6):1185-96.
156. Abbott JK, Huoh YS, Reynolds PR, Yu L, Rewers M, Reddy M, et al. Dominant-negative loss of function arises from a second, more frequent variant within the SAND domain of autoimmune regulator (AIRE). *J Autoimmun*. 2018;88:114-20.
157. Wang C-Y, Shi J-D, Davoodi-Semiromi A, She J-X. Cloning of Aire, the Mouse Homologue of the Autoimmune Regulator (AIRE) Gene Responsible for Autoimmune Polyglandular Syndrome Type 1 (APS1). *Genomics*. 1999;55(3):322-6.
158. Heino M, Peterson P, Kudoh J, Nagamine K, Lagerstedt A, Ovod V, et al. Autoimmune regulator is expressed in the cells regulating immune tolerance in thymus medulla. *Biochem Biophys Res Commun*. 1999;257(3):821-5.

159. Zuklys S, Balciunaite G, Agarwal A, Fasler-Kan E, Palmer E, Hollander GA. 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*. 2000;165(4):1976-83.
160. Gardner JM, Devoss JJ, Friedman RS, Wong DJ, Tan YX, Zhou X, et al. Deletional tolerance mediated by extrathymic Aire-expressing cells. *Science*. 2008;321(5890):843-7.
161. Schaller CE, Wang CL, Beck-Engeser G, Goss L, Scott HS, Anderson MS, et al. Expression of Aire and the Early Wave of Apoptosis in Spermatogenesis. *The Journal of Immunology*. 2008;180(3):1338-43.
162. Anderson MS, Su MA. AIRE expands: new roles in immune tolerance and beyond. *Nat Rev Immunol*. 2016;16(4):247-58.
163. Haljasorg U, Bichele R, Saare M, Guha M, Maslovskaja J, Kond K, et al. A highly conserved NF-kappaB-responsive enhancer is critical for thymic expression of Aire in mice. *Eur J Immunol*. 2015;45(12):3246-56.
164. Herzig Y, Nevo S, Bornstein C, Brezis MR, Ben-Hur S, Shkedy A, et al. Transcriptional programs that control expression of the autoimmune regulator gene Aire. *Nat Immunol*. 2017;18(2):161-72.
165. Yanagihara T, Sanematsu F, Sato T, Uruno T, Duan X, Tomino T, et al. Intronic regulation of Aire expression by Jmjd6 for self-tolerance induction in the thymus. *Nat Commun*. 2015;6:8820.
166. Liston A, Lesage S, Wilson J, Peltonen L, Goodnow CC. Aire regulates negative selection of organ-specific T cells. *Nat Immunol*. 2003;4(4):350-4.

167. Anderson MS, Venzani ES, Chen Z, Berzins SP, Benoist C, Mathis D. The cellular mechanism of Aire control of T cell tolerance. *Immunity*. 2005;23(2):227-39.
168. Taniguchi RT, DeVoss JJ, Moon JJ, Sidney J, Sette A, Jenkins MK, et al. Detection of an autoreactive T-cell population within the polyclonal repertoire that undergoes distinct autoimmune regulator (Aire)-mediated selection. *Proc Natl Acad Sci U S A*. 2012;109(20):7847-52.
169. Hubert FX, Kinkel SA, Davey GM, Phipson B, Mueller SN, Liston A, et al. Aire regulates the transfer of antigen from mTECs to dendritic cells for induction of thymic tolerance. *Blood*. 2011;118(9):2462-72.
170. Yang S, Fujikado N, Kolodin D, Benoist C, Mathis D. Immune tolerance. Regulatory T cells generated early in life play a distinct role in maintaining self-tolerance. *Science*. 2015;348(6234):589-94.
171. Kekalainen E, Tuovinen H, Joensuu J, Gylling M, Franssila R, Pontynen N, et al. A Defect of Regulatory T Cells in Patients with Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy. *The Journal of Immunology*. 2007;178(2):1208-15.
172. Laakso SM, Laurinolli TT, Rossi LH, Lehtoviita A, Sairanen H, Perheentupa J, et al. Regulatory T cell defect in APECED patients is associated with loss of naive FOXP3(+) precursors and impaired activated population. *J Autoimmun*. 2010;35(4):351-7.
173. Aricha R, Feferman T, Scott HS, Souroujon MC, Berrih-Aknin S, Fuchs S. The susceptibility of Aire(-/-) mice to experimental myasthenia gravis involves alterations in regulatory T cells. *J Autoimmun*. 2011;36(1):16-24.

174. Aschenbrenner K, D'Cruz LM, Vollmann EH, Hinterberger M, Emmerich J, Swee LK, et al. Selection of Foxp3⁺ regulatory T cells specific for self antigen expressed and presented by Aire⁺ medullary thymic epithelial cells. *Nat Immunol*. 2007;8(4):351-8.
175. Perry JSA, Lio CJ, Kau AL, Nutsch K, Yang Z, Gordon JI, et al. Distinct contributions of Aire and antigen-presenting-cell subsets to the generation of self-tolerance in the thymus. *Immunity*. 2014;41(3):414-26.
176. Malchow S, Leventhal DS, Nishi S, Fischer BI, Shen L, Paner GP, et al. Aire-dependent thymic development of tumor-associated regulatory T cells. *Science*. 2013;339(6124):1219-24.
177. Gillard GO, Farr AG. Contrasting models of promiscuous gene expression by thymic epithelium. *J Exp Med*. 2005;202(1):15-9.
178. Villasenor 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*. 2008;105(41):15854-9.
179. Gray D, Abramson J, Benoist C, Mathis D. Proliferative arrest and rapid turnover of thymic epithelial cells expressing Aire. *J Exp Med*. 2007;204(11):2521-8.
180. Liiv I, Haljasorg U, Kisand K, Maslovskaja J, Laan M, Peterson P. AIRE-induced apoptosis is associated with nuclear translocation of stress sensor protein GAPDH. *Biochem Biophys Res Commun*. 2012;423(1):32-7.
181. Abramson J, Giraud M, Benoist C, Mathis D. Aire's Partners in the Molecular Control of Immunological Tolerance. *Cell*. 2010;140(1):123-35.

182. Gaetani M, Matafora V, Saare M, Spiliotopoulos D, Mollica L, Quilici G, et al. AIRE-PHD fingers are structural hubs to maintain the integrity of chromatin-associated interactome. *Nucleic Acids Research*. 2012;40(22):11756-68.
183. Giraud M, Jmari N, Du L, Carallis F, Nieland TJ, Perez-Campo FM, et al. An RNAi screen for Aire cofactors reveals a role for Hnnp1 in polymerase release and Aire-activated ectopic transcription. *Proc Natl Acad Sci U S A*. 2014;111(4):1491-6.
184. Abramson J, Goldfarb Y. AIRE: From promiscuous molecular partnerships to promiscuous gene expression. *Eur J Immunol*. 2016;46(1):22-33.
185. Ferguson BJ, Alexander C, Rossi SW, Liiv I, Rebane A, Worth CL, et al. AIRE's CARD revealed, a new structure for central tolerance provokes transcriptional plasticity. *J Biol Chem*. 2008;283(3):1723-31.
186. Pitkanen J, Doucas V, Sternsdorf T, Nakajima T, Aratani S, Jensen K, et al. The autoimmune regulator protein has transcriptional transactivating properties and interacts with the common coactivator CREB-binding protein. *J Biol Chem*. 2000;275(22):16802-9.
187. Kumar PG, Laloraya M, Wang CY, Ruan QG, Davoodi-Semiromi A, Kao KJ, et al. The autoimmune regulator (AIRE) is a DNA-binding protein. *J Biol Chem*. 2001;276(44):41357-64.
188. Waterfield M, Khan IS, Cortez JT, Fan U, Metzger T, Greer A, et al. The transcriptional regulator Aire coopts the repressive ATF7ip-MBD1 complex for the induction of immunotolerance. *Nature Immunology*. 2014;15(3):258-65.

189. Avin A, Levy M, Porat Z, Abramson J. Quantitative analysis of protein-protein interactions and post-translational modifications in rare immune populations. *Nat Commun.* 2017;8(1):1524.
190. Ramsey C BA, Peltonen L. Systematic mutagenesis of the functional domains of AIRE reveals their role in intracellular targeting. *Human Molecular Genetics.* 2002;11(26):3299-308.
191. Halonen M, Kangas H, Ruppell T, Ilmarinen T, Ollila J, Kolmer M, et al. APECED-causing mutations in AIRE reveal the functional domains of the protein. *Hum Mutat.* 2004;23(3):245-57.
192. Koh AS, Kuo AJ, Park SY, Cheung P, Abramson J, Bua D, et al. Aire employs a histone-binding module to mediate immunological tolerance, linking chromatin regulation with organ-specific autoimmunity. *Proceedings of the National Academy of Sciences.* 2008;105(41):15878-83.
193. Org T, Chignola F, Hetényi C, Gaetani M, Rebane A, Liiv I, et al. The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO reports.* 2008;9(4):370-6.
194. Chignola F, Gaetani M, Rebane A, Org T, Mollica L, Zucchelli C, et al. The solution structure of the first PHD finger of autoimmune regulator in complex with non-modified histone H3 tail reveals the antagonistic role of H3R2 methylation. *Nucleic Acids Res.* 2009;37(9):2951-61.
195. Yang S, Bansal K, Lopes J, Benoist C, Mathis D. Aire's plant homeodomain(PHD)-2 is critical for induction of immunological tolerance. *Proc Natl Acad Sci U S A.* 2013;110(5):1833-8.

196. Meloni A, Incani F, Corda D, Cao A, Rosatelli MC. Role of PHD fingers and COOH-terminal 30 amino acids in AIRE transactivation activity. *Mol Immunol*. 2008;45(3):805-9.
197. Peterson P, Org T, Rebane A. Transcriptional regulation by AIRE: molecular mechanisms of central tolerance. *Nat Rev Immunol*. 2008;8(12):948-57.
198. Bansal K, Yoshida H, Benoist C, Mathis D. The transcriptional regulator Aire binds to and activates super-enhancers. *Nat Immunol*. 2017;18(3):263-73.
199. Handel AE, Shikama-Dorn N, Zhanybekova S, Maio S, Graedel AN, Zuklys S, et al. Comprehensively Profiling the Chromatin Architecture of Tissue Restricted Antigen Expression in Thymic Epithelial Cells Over Development. *Front Immunol*. 2018;9:2120.
200. Org T, Rebane A, Kisand K, Laan M, Haljasorg U, Andreson R, et al. AIRE activated tissue specific genes have histone modifications associated with inactive chromatin. *Hum Mol Genet*. 2009;18(24):4699-710.
201. Oven I, Brdickova N, Kohoutek J, Vaupotic T, Narat M, Peterlin BM. AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. *Mol Cell Biol*. 2007;27(24):8815-23.
202. Giraud M, Yoshida H, Abramson J, Rahl PB, Young RA, Mathis D, et al. Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. *Proc Natl Acad Sci U S A*. 2012;109(2):535-40.
203. Yoshida H, Bansal K, Schaefer U, Chapman T, Rioja I, Proekt I, et al. Brd4 bridges the transcriptional regulators, Aire and P-TEFb, to promote

elongation of peripheral-tissue antigen transcripts in thymic stromal cells. *Proc Natl Acad Sci U S A*. 2015;112(32):E4448-57.

204. Akiyoshi H, Hatakeyama S, Pitkanen J, Mouri Y, Doucas V, Kudoh J, et al. Subcellular expression of autoimmune regulator is organized in a spatiotemporal manner. *J Biol Chem*. 2004;279(32):33984-91.

205. Tao Y, Kupfer R, Stewart BJ, Williams-Skipp C, Crowell CK, Patel DD, et al. AIRE recruits multiple transcriptional components to specific genomic regions through tethering to nuclear matrix. *Mol Immunol*. 2006;43(4):335-45.

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

207. Ju BG, Lunyak VV, Perissi V, Garcia-Bassets I, Rose DW, Glass CK, et al. A topoisomerase II β -mediated dsDNA break required for regulated transcription. *Science*. 2006;312(5781):1798-802.

208. Mathieu AL, Verronese E, Rice GI, Fouyssac F, Bertrand Y, Picard C, et al. PRKDC mutations associated with immunodeficiency, granuloma, and autoimmune regulator-dependent autoimmunity. *J Allergy Clin Immunol*. 2015;135(6):1578-88 e5.

209. Chuprin A, Avin A, Goldfarb Y, Herzig Y, Levi B, Jacob A, et al. The deacetylase Sirt1 is an essential regulator of Aire-mediated induction of central immunological tolerance. *Nat Immunol*. 2015;16(7):737-45.

210. Biason-Lauber A, Boni-Schnetzler M, Hubbard BP, Bouzakri K, Brunner A, Cavelti-Weder C, et al. Identification of a SIRT1 mutation in a family with type 1 diabetes. *Cell Metab*. 2013;17(3):448-55.

211. Derbinski J, Pinto S, Rosch 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*. 2008;105(2):657-62.
212. Rattay K, Meyer HV, Herrmann C, Brors B, Kyewski B. Evolutionary conserved gene co-expression drives generation of self-antigen diversity in medullary thymic epithelial cells. *Journal of Autoimmunity*. 2016;67:65-75.
213. Marfella CG, Imbalzano AN. The Chd family of chromatin remodelers. *Mutat Res*. 2007;618(1-2):30-40.
214. Yap KL, Zhou MM. Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. *Biochemistry*. 2011;50(12):1966-80.
215. Stanley FK, Moore S, Goodarzi AA. CHD chromatin remodelling enzymes and the DNA damage response. *Mutat Res*. 2013;750(1-2):31-44.
216. Seelig HP, Moosbrugger I, Ehrfeld H, Fink T, Renz M, Genth E. The major dermatomyositis-specific mi-2 autoantigen is a presumed helicase involved in transcriptional activation. *Arthritis & Rheumatism*. 1995;38(10):1389-99.
217. Denslow SA, Wade PA. The human Mi-2/NuRD complex and gene regulation. *Oncogene*. 2007;26(37):5433-8.
218. Hoffmeister H, Fuchs A, Erdel F, Pinz S, Grobner-Ferreira R, Bruckmann A, et al. CHD3 and CHD4 form distinct NuRD complexes with different yet overlapping functionality. *Nucleic Acids Res*. 2017;45(18):10534-54.

219. Cai Y, Geutjes EJ, de Lint K, Roepman P, Bruurs L, Yu LR, et al. The NuRD complex cooperates with DNMTs to maintain silencing of key colorectal tumor suppressor genes. *Oncogene*. 2014;33(17):2157-68.
220. Wade PA, Geggion A, Jones PL, Ballestar E, Aubry F, Wolffe AP. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat Genet*. 1999;23(1):62-6.
221. Zhang Y, LeRoy G, Seelig H-P, Lane WS, Reinberg D. The Dermatomyositis-Specific Autoantigen Mi2 Is a Component of a Complex Containing Histone Deacetylase and Nucleosome Remodeling Activities. *Cell*. 1998;95(2):279-89.
222. Xue Y, Wong J, Moreno GT, Young MK, Côté J, Wang W. NURD, a Novel Complex with Both ATP-Dependent Chromatin-Remodeling and Histone Deacetylase Activities. *Molecular Cell*. 1998;2(6):851-61.
223. McDonel P, Costello I, Hendrich B. Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development. *Int J Biochem Cell Biol*. 2009;41(1):108-16.
224. Miccio A, Wang Y, Hong W, Gregory GD, Wang H, Yu X, et al. NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development. *EMBO J*. 2010;29(2):442-56.
225. Williams CJ, Naito T, Arco PG-d, Seavitt JR, Cashman SM, De Souza B, et al. The Chromatin Remodeler Mi-2 β Is Required for CD4 Expression and T Cell Development. *Immunity*. 2004;20(6):719-33.
226. Shimono Y, Murakami H, Kawai K, Wade PA, Shimokata K, Takahashi M. Mi-2 beta associates with BRG1 and RET finger protein at the distinct

regions with transcriptional activating and repressing abilities. *J Biol Chem.* 2003;278(51):51638-45.

227. Reynolds N, Salmon-Divon M, Dvinge H, Hynes-Allen A, Balasooriya G, Leaford D, et al. NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *EMBO J.* 2012;31(3):593-605.

228. Schmidt DR, Schreiber SL. Molecular Association between ATR and Two Components of the Nucleosome Remodeling and Deacetylating Complex, HDAC2 and CHD4†. *Biochemistry.* 1999;38(44):14711-7.

229. Polo SE, Kaidi A, Baskcomb L, Galanty Y, Jackson SP. Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. *EMBO J.* 2010;29(18):3130-9.

230. O'Shaughnessy A, Hendrich B. CHD4 in the DNA-damage response and cell cycle progression: not so NuRDy now. *Biochem Soc Trans.* 2013;41(3):777-82.

231. Smeenk G, Wiegant WW, Vrolijk H, Solari AP, Pastink A, van Attikum H. The NuRD chromatin-remodeling complex regulates signaling and repair of DNA damage. *J Cell Biol.* 2010;190(5):741-9.

232. Sims JK, Wade PA. Mi-2/NuRD complex function is required for normal S phase progression and assembly of pericentric heterochromatin. *Mol Biol Cell.* 2011;22(17):3094-102.

233. Morra R, Lee BM, Shaw H, Tuma R, Mancini EJ. Concerted action of the PHD, chromo and motor domains regulates the human chromatin remodelling ATPase CHD4. *FEBS Lett.* 2012;586(16):2513-21.

234. Musselman CA, Mansfield RE, Garske AL, Davrazou F, Kwan AH, Oliver SS, et al. Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem J*. 2009;423(2):179-87.
235. Mansfield RE, Musselman CA, Kwan AH, Oliver SS, Garske AL, Davrazou F, et al. Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *J Biol Chem*. 2011;286(13):11779-91.
236. Musselman CA, Ramirez J, Sims JK, Mansfield RE, Oliver SS, Denu JM, et al. Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression. *Proc Natl Acad Sci U S A*. 2012;109(3):787-92.
237. Bouazoune K, Mitterweger A, Langst G, Imhof A, Akhtar A, Becker PB, et al. The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. *EMBO J*. 2002;21(10):2430-40.
238. Watson AA, Mahajan P, Mertens HD, Deery MJ, Zhang W, Pham P, et al. The PHD and chromo domains regulate the ATPase activity of the human chromatin remodeler CHD4. *J Mol Biol*. 2012;422(1):3-17.
239. Tong JK, Hassig CA, Schnitzler GR, Kingston RE, Schreiber SL. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature*. 1998;395(6705):917-21.
240. Nitarska J, Smith JG, Sherlock WT, Hillege MM, Nott A, Barshop WD, et al. A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. *Cell Rep*. 2016;17(6):1683-98.

241. Ingram KG, Curtis CD, Silasi-Mansat R, Lupu F, Griffin CT. The NuRD chromatin-remodeling enzyme CHD4 promotes embryonic vascular integrity by transcriptionally regulating extracellular matrix proteolysis. *PLoS Genet.* 2013;9(12):e1004031.
242. Nakamura T, Mori T, Tada S, Krajewski W, Rozovskaia T, Wassell R, et al. ALL-1 Is a Histone Methyltransferase that Assembles a Supercomplex of Proteins Involved in Transcriptional Regulation. *Molecular Cell.* 2002;10(5):1119-28.
243. Bornelov S, Reynolds N, Xenophontos M, Gharbi S, Johnstone E, Floyd R, et al. The Nucleosome Remodeling and Deacetylation Complex Modulates Chromatin Structure at Sites of Active Transcription to Fine-Tune Gene Expression. *Mol Cell.* 2018;71(1):56-72 e4.
244. Naito T, Gomez-Del Arco P, Williams CJ, Georgopoulos K. Antagonistic interactions between Ikaros and the chromatin remodeler Mi-2beta determine silencer activity and Cd4 gene expression. *Immunity.* 2007;27(5):723-34.
245. Sparmann A, Xie Y, Verhoeven E, Vermeulen M, Lancini C, Gargiulo G, et al. The chromodomain helicase Chd4 is required for Polycomb-mediated inhibition of astroglial differentiation. *EMBO J.* 2013;32(11):1598-612.
246. Burd CJ, Kinyamu HK, Miller FW, Archer TK. UV radiation regulates Mi-2 through protein translation and stability. *J Biol Chem.* 2008;283(50):34976-82.
247. Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER, 3rd, Hurov KE, Luo J, et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science.* 2007;316(5828):1160-6.

248. Stokes MP, Rush J, Macneill J, Ren JM, Sprott K, Nardone J, et al. Profiling of UV-induced ATM/ATR signaling pathways. *Proc Natl Acad Sci U S A*. 2007;104(50):19855-60.
249. Pegoraro G, Kubben N, Wickert U, Gohler H, Hoffmann K, Misteli T. Ageing-related chromatin defects through loss of the NURD complex. *Nat Cell Biol*. 2009;11(10):1261-7.
250. Smith R, Sellou H, Chapuis C, Huet S, Timinszky G. CHD3 and CHD4 recruitment and chromatin remodeling activity at DNA breaks is promoted by early poly(ADP-ribose)-dependent chromatin relaxation. *Nucleic Acids Res*. 2018;46(12):6087-98.
251. Luijsterburg MS, Acs K, Ackermann L, Wiegant WW, Bekker-Jensen S, Larsen DH, et al. A new non-catalytic role for ubiquitin ligase RNF8 in unfolding higher-order chromatin structure. *EMBO J*. 2012;31(11):2511-27.
252. Kim MS, Chung NG, Kang MR, Yoo NJ, Lee SH. Genetic and expressional alterations of CHD genes in gastric and colorectal cancers. *Histopathology*. 2011;58(5):660-8.
253. Le Gallo M, O'Hara AJ, Rudd ML, Urick ME, Hansen NF, O'Neil NJ, et al. Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. *Nat Genet*. 2012;44(12):1310-5.
254. Zhang Y. Biology of the Mi-2/NuRD Complex in SLAC (Stemness, Longevity/Ageing, and Cancer). *Gene Regul Syst Bio*. 2011;5:1-26.
255. Zhao H, Han Z, Liu X, Gu J, Tang F, Wei G, et al. The chromatin remodeler Chd4 maintains embryonic stem cell identity by controlling

pluripotency- and differentiation-associated genes. *J Biol Chem.* 2017;292(20):8507-19.

256. Yoshida T, Hazan I, Zhang J, Ng SY, Naito T, Snippert HJ, et al. The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation. *Genes Dev.* 2008;22(9):1174-89.

257. Gomez-Del Arco P, Perdiguero E, Yunes-Leites PS, Acin-Perez R, Zeini M, Garcia-Gomez A, et al. The Chromatin Remodeling Complex Chd4/NuRD Controls Striated Muscle Identity and Metabolic Homeostasis. *Cell Metab.* 2016;23(5):881-92.

258. Dege C, Hagman J. Mi-2/NuRD chromatin remodeling complexes regulate B and T-lymphocyte development and function. *Immunol Rev.* 2014;261(1):126-40.

259. Hosokawa H, Tanaka T, Suzuki Y, Iwamura C, Ohkubo S, Endoh K, et al. Functionally distinct Gata3/Chd4 complexes coordinately establish T helper 2 (Th2) cell identity. *Proc Natl Acad Sci U S A.* 2013;110(12):4691-6.

260. Curtis CD, Griffin CT. The chromatin-remodeling enzymes BRG1 and CHD4 antagonistically regulate vascular Wnt signaling. *Mol Cell Biol.* 2012;32(7):1312-20.

261. Zuklys S, Gill J, Keller MP, Hauri-Hohl M, Zhanybekova S, Balciunaite G, et al. Stabilized beta-catenin in thymic epithelial cells blocks thymus development and function. *J Immunol.* 2009;182(5):2997-3007.

262. Kisand K, Peterson P, Laan M. Lymphopenia-induced proliferation in *aire*-deficient mice helps to explain their autoimmunity and differences from human patients. *Front Immunol.* 2014;5:51.

263. Zuklys S, Mayer CE, Zhanybekova S, Stefanski HE, Nusspaumer G, Gill J, et al. MicroRNAs control the maintenance of thymic epithelia and their competence for T lineage commitment and thymocyte selection. *J Immunol*. 2012;189(8):3894-904.
264. Lambie S, Batty E, Attar M, Buck D, Bowden R, Lunter G, et al. Improved workflows for high throughput library preparation using the transposome-based Nextera system. *BMC Biotechnol*. 2013;13:104.
265. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
266. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417-9.
267. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
268. Aaron Lun DM, John Marioni. simpleSingleCell: A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. R package version 1212018.
269. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol*. 2018;36(5):421-7.
270. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008;2008(10):P10008.

271. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. *Nature*. 2018;560(7719):494-8.
272. Manning CD, Raghavan, P. and Schutze, H. *Introduction to Information Retrieval*. : Cambridge University Press; 2008.
273. McInnes L, Healy J. Accelerated Hierarchical Density Based Clustering. 2017:33-42.
274. TW HB, Girke T. systemPipeR: NGS workflow and report generation environment. *BMC Bioinformatics*. 2016;17:388.
275. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-9.
276. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nature Protocols*. 2012;7:1728.
277. Carroll TS, Liang Z, Salama R, Stark R, de Santiago I. Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Front Genet*. 2014;5:75.
278. Shin H, Liu T, Manrai AK, Liu XS. CEAS: cis-regulatory element annotation system. *Bioinformatics*. 2009;25(19):2605-6.
279. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42(Web Server issue):W187-91.
280. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676-82.

281. Pau G, Fuchs F, Sklyar O, Boutros M, Huber W. EBIImage--an R package for image processing with applications to cellular phenotypes. *Bioinformatics*. 2010;26(7):979-81.
282. Dhalla F, Baran-Gale J, Maio S, Chappell L, Hollander GA, Ponting C. Biologically indeterminate yet ordered promiscuous gene expression in single medullary thymic epithelial cells. *bioRxiv*. 2019:554899.
283. Agaesse G, Barbolat-Boutrand L, El Kharbili M, Berthier-Vergnes O, Masse I. p53 targets TSPAN8 to prevent invasion in melanoma cells. *Oncogenesis*. 2017;6(4):e309.
284. Zhu Y, Ailane N, Sala-Valdes M, Haghighi-Rad F, Billard M, Nguyen V, et al. Multi-factorial modulation of colorectal carcinoma cells motility - partial coordination by the tetraspanin Co-029/tspan8. *Oncotarget*. 2017;8(16):27454-70.
285. Zhao K, Erb U, Hackert T, Zoller M, Yue S. Distorted leukocyte migration, angiogenesis, wound repair and metastasis in Tspan8 and Tspan8/CD151 double knockout mice indicate complementary activities of Tspan8 and CD51. *Biochim Biophys Acta Mol Cell Res*. 2018;1865(2):379-91.
286. Cogger KF, Sinha A, Sarangi F, McGaugh EC, Saunders D, Dorrell C, et al. Glycoprotein 2 is a specific cell surface marker of human pancreatic progenitors. *Nat Commun*. 2017;8(1):331.
287. Ohno H, Hase K. Glycoprotein 2 (GP2): grabbing the FimH bacteria into M cells for mucosal immunity. *Gut Microbes*. 2010;1(6):407-10.
288. Werner L, Sturm A, Roggenbuck D, Yahav L, Zion T, Meirowitz E, et al. Antibodies against glycoprotein 2 are novel markers of intestinal

inflammation in patients with an ileal pouch. *J Crohns Colitis*. 2013;7(11):e522-32.

289. Tornai T, Tornai D, Sipeki N, Tornai I, Alsulaimani R, Fechner K, et al. Loss of tolerance to gut immunity protein, glycoprotein 2 (GP2) is associated with progressive disease course in primary sclerosing cholangitis. *Sci Rep*. 2018;8(1):399.

290. Scialdone A, Natarajan KN, Saraiva LR, Proserpio V, Teichmann SA, Stegle O, et al. Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods*. 2015;85:54-61.

291. Galliano MF, Roccasecca RM, Descargues P, Micheloni A, Levy E, Zambruno G, et al. Characterization and expression analysis of the Spink5 gene, the mouse ortholog of the defective gene in Netherton syndrome. *Genomics*. 2005;85(4):483-92.

292. Bitoun E, Micheloni A, Lamant L, Bonnart C, Tartaglia-Polcini A, Cobbold C, et al. LEKTI proteolytic processing in human primary keratinocytes, tissue distribution and defective expression in Netherton syndrome. *Hum Mol Genet*. 2003;12(19):2417-30.

293. Wada N, Nishifuji K, Yamada T, Kudoh J, Shimizu N, Matsumoto M, et al. Aire-dependent thymic expression of desmoglein 3, the autoantigen in pemphigus vulgaris, and its role in T-cell tolerance. *J Invest Dermatol*. 2011;131(2):410-7.

294. Bornstein C, Nevo S, Giladi A, Kadouri N, Pouzolles M, Gerbe F, et al. Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. *Nature*. 2018;559(7715):622-6.

295. Miller CN, Proekt I, von Moltke J, Wells KL, Rajpurkar AR, Wang H, et al. Thymic tuft cells promote an IL-4-enriched medulla and shape thymocyte development. *Nature*. 2018;559(7715):627-31.
296. Reid L, Meyrick B, Antony VB, Chang LY, Crapo JD, Reynolds HY. The mysterious pulmonary brush cell: a cell in search of a function. *Am J Respir Crit Care Med*. 2005;172(1):136-9.
297. Yamashita J, Ohmoto M, Yamaguchi T, Matsumoto I, Hirota J. Skn-1a/Pou2f3 functions as a master regulator to generate Trpm5-expressing chemosensory cells in mice. *PLoS One*. 2017;12(12):e0189340.
298. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res*. 2018;46(D1):D649-D55.
299. Miragaia RJ, Zhang X, Gomes T, Svensson V, Ilicic T, Henriksson J, et al. Single-cell RNA-sequencing resolves self-antigen expression during mTEC development. *Scientific Reports*. 2018;8(1):685.
300. O'Shaughnessy-Kirwan A, Signolet J, Costello I, Gharbi S, Hendrich B. Constraint of gene expression by the chromatin remodelling protein CHD4 facilitates lineage specification. *Development*. 2015;142(15):2586-97.
301. Gotter J, Brors B, Hergenhausen M, Kyewski B. Medullary epithelial cells of the human thymus express a highly diverse selection of tissue-specific genes colocalized in chromosomal clusters. *J Exp Med*. 2004;199(2):155-66.
302. Nakagawa Y, Ohigashi I, Nitta T, Sakata M, Tanaka K, Murata S, et al. Thymic nurse cells provide microenvironment for secondary T cell receptor

- alpha rearrangement in cortical thymocytes. *Proc Natl Acad Sci U S A*. 2012;109(50):20572-7.
303. Swat W, Dessing M, von Boehmer H, Kisielow P. CD69 expression during selection and maturation of CD4+8+ thymocytes. *Eur J Immunol*. 1993;23(3):739-46.
304. Hu Q, Nicol SA, Suen AY, Baldwin TA. Examination of thymic positive and negative selection by flow cytometry. *J Vis Exp*. 2012(68).
305. Yamashita I, Nagata T, Tada T, Nakayama T. CD69 cell surface expression identifies developing thymocytes which audition for T cell antigen receptor-mediated positive selection. *International Immunology*. 1993;5(9):1139-50.
306. Cossarizza A, Chang HD, Radbruch A, Akdis M, Andra I, Annunziato F, et al. Guidelines for the use of flow cytometry and cell sorting in immunological studies. *Eur J Immunol*. 2017;47(10):1584-797.
307. Kim KH, Sederstrom JM. Assaying Cell Cycle Status Using Flow Cytometry. *Curr Protoc Mol Biol*. 2015;111:28 6 1-11.
308. Van Vliet E, Jenkinson EJ, Kingston R, Owen JJ, Van Ewijk W. Stromal cell types in the developing thymus of the normal and nude mouse embryo. *Eur J Immunol*. 1985;15(7):675-81.
309. Gray DH, Seach N, Ueno T, Milton MK, Liston A, Lew AM, et al. Developmental kinetics, turnover, and stimulatory capacity of thymic epithelial cells. *Blood*. 2006;108(12):3777-85.

310. Wendland K, Niss K, Kotarsky K, Wu NYH, White AJ, Jendholm J, et al. Retinoic Acid Signaling in Thymic Epithelial Cells Regulates Thymopoiesis. *J Immunol*. 2018;201(2):524-32.
311. Ucar O, Li K, Dvornikov D, Kreutz C, Timmer J, Matt S, et al. A Thymic Epithelial Stem Cell Pool Persists throughout Ontogeny and Is Modulated by TGF-beta. *Cell Rep*. 2016;17(2):448-57.
312. Meireles C, Ribeiro AR, Pinto RD, Leitao C, Rodrigues PM, Alves NL. Thymic crosstalk restrains the pool of cortical thymic epithelial cells with progenitor properties. *Eur J Immunol*. 2017;47(6):958-69.
313. Takeoka Y, Chen S-Y, Boyd RL, Tsuneyama K, Taguchi N, Morita S, et al. A Comparative Analysis of the Murine Thymic Microenvironment in Normal, Autoimmune, and Immunodeficiency States. *Developmental Immunology*. 1997;5(2):79-89.
314. Akiyama T, Maeda S, Yamane S, Ogino K, Kasai M, Kajiura F, et al. Dependence of self-tolerance on TRAF6-directed development of thymic stroma. *Science*. 2005;308(5719):248-51.
315. Osada M, Ito E, Fermin HA, Vazquez-Cintron E, Venkatesh T, Friedel RH, et al. The Wnt signaling antagonist Kremen1 is required for development of thymic architecture. *Clin Dev Immunol*. 2006;13(2-4):299-319.
316. Cejalvo T, Munoz JJ, Tobajas E, Fanlo L, Alfaro D, Garcia-Ceca J, et al. Ephrin-B-dependent thymic epithelial cell-thymocyte interactions are necessary for correct T cell differentiation and thymus histology organization: relevance for thymic cortex development. *J Immunol*. 2013;190(6):2670-81.

317. Saini M, Sinclair C, Marshall D, Tolaini M, Sakaguchi S, Seddon B. Regulation of Zap70 expression during thymocyte development enables temporal separation of CD4 and CD8 repertoire selection at different signaling thresholds. *Sci Signal*. 2010;3(114):ra23.
318. Sinclair C, Bains I, Yates AJ, Seddon B. Asymmetric thymocyte death underlies the CD4:CD8 T-cell ratio in the adaptive immune system. *Proc Natl Acad Sci U S A*. 2013;110(31):E2905-14.
319. Xinye Yin EL, Shiao Wei Chan, Ou Li, Nigel Killeen, Dietmar J. Kappes and Ellen A. Robey. CCR7 Expression in Developing Thymocytes Is Linked to the CD4 versus CD8 Lineage Decision. *Journal of Immunology*. 2007;179(11):7.
320. Bryan K. Cho VPR, Qing Ge, Herman N. Eisen, Jianzhu Chena. Homeostasis-Stimulated Proliferation Drives Naive T Cells to Differentiate Directly into Memory T Cells. *Journal of Experimental Medicine*. 2000;192(4):8.
321. Charles D. Surh JS. Homeostatic T cell proliferation: how far can T cells be activated to self-ligands? *Journal of Experimental Medicine*. 2000;192(4):5.
322. Surh CD, Sprent J. Homeostasis of naive and memory T cells. *Immunity*. 2008;29(6):848-62.
323. Barthlott T, Bosch AJ, Berkemeier C, Nogales-Cadenas R, Jeker LT, Keller MP, et al. A subpopulation of CD103(pos) ICOS(pos) Treg cells occurs at high frequency in lymphopenic mice and represents a lymph node specific differentiation stage. *Eur J Immunol*. 2015;45(6):1760-71.
324. Hollander GA, Krenger W, Blazar BR. Emerging strategies to boost thymic function. *Curr Opin Pharmacol*. 2010;10(4):443-53.

325. Chu YW, Schmitz S, Choudhury B, Telford W, Kapoor V, Garfield S, et al. Exogenous insulin-like growth factor 1 enhances thymopoiesis predominantly through thymic epithelial cell expansion. *Blood*. 2008;112(7):2836-46.
326. Zheng Q, Lin Z, Xu J, Lu Y, Meng Q, Wang C, et al. Long noncoding RNA MEG3 suppresses liver cancer cells growth through inhibiting beta-catenin by activating PKM2 and inactivating PTEN. *Cell Death Dis*. 2018;9(3):253.
327. Liang CC, You LR, Yen JJ, Liao NS, Yang-Yen HF, Chen CM. Thymic epithelial beta-catenin is required for adult thymic homeostasis and function. *Immunol Cell Biol*. 2013;91(8):511-23.
328. Radwanska A, Grall D, Schaub S, Divonne SB-dIF, Ciais D, Rekima S, et al. Counterbalancing anti-adhesive effects of Tenascin-C through fibronectin expression in endothelial cells. *Scientific Reports*. 2017;7(1):12762.
329. Li L, Hsu H-C, Stockard CR, Yang P, Zhou J, Wu Q, et al. IL-12 Inhibits Thymic Involution by Enhancing IL-7- and IL-2-Induced Thymocyte Proliferation. *The Journal of Immunology*. 2004;172(5):2909.
330. Lund SA, Giachelli CM, Scatena M. The role of osteopontin in inflammatory processes. *J Cell Commun Signal*. 2009;3(3-4):311-22.
331. Vermot J, Niederreither K, Garnier JM, Chambon P, Dolle P. Decreased embryonic retinoic acid synthesis results in a DiGeorge syndrome phenotype in newborn mice. *Proc Natl Acad Sci U S A*. 2003;100(4):1763-8.
332. West KP, Jr., Howard GR, Sommer A. Vitamin A and infection: public health implications. *Annu Rev Nutr*. 1989;9:63-86.

333. Dennert G, Lotan R. Effects of retinoic acid on the immune system: stimulation of T killer cell induction. *Eur J Immunol.* 1978;8(1):23-9.
334. Kiss I, Ruhl R, Szegezdi E, Fritzsche B, Toth B, Pongracz J, et al. Retinoid receptor-activating ligands are produced within the mouse thymus during postnatal development. *Eur J Immunol.* 2008;38(1):147-55.
335. Thompson TC. Glioma pathogenesis-related protein 1: tumor-suppressor activities and therapeutic potential. *Yonsei Med J.* 2010;51(4):479-83.
336. McConnell BB, Yang VW. Mammalian Kruppel-like factors in health and diseases. *Physiol Rev.* 2010;90(4):1337-81.
337. Song HJ, Poy G, Darwiche N, Lichti U, Kuroki T, Steinert PM, et al. Mouse *Spr2* genes: a clustered family of genes showing differential expression in epithelial tissues. *Genomics.* 1999;55(1):28-42.
338. Pezzi N, Assis AF, Cotrim-Sousa LC, Lopes GS, Mosella MS, Lima DS, et al. Aire knockdown in medullary thymic epithelial cells affects Aire protein, deregulates cell adhesion genes and decreases thymocyte interaction. *Mol Immunol.* 2016;77:157-73.
339. Berrih S, Savino W, Cohen S. Extracellular matrix of the human thymus: immunofluorescence studies on frozen sections and cultured epithelial cells. *J Histochem Cytochem.* 1985;33(7):655-64.
340. Savino W, Mendes-Da-Cruz DA, Smaniotto S, Silva-Monteiro E, Villa-Verde DM. Molecular mechanisms governing thymocyte migration: combined role of chemokines and extracellular matrix. *J Leukoc Biol.* 2004;75(6):951-61.

341. Savino W, Mendes-da-Cruz DA, Silva JS, Dardenne M, Cotta-de-Almeida Vc. Intrathymic T-cell migration: a combinatorial interplay of extracellular matrix and chemokines? *Trends in Immunology*. 2002;23(6):305-13.
342. Sitnik KM, Kotarsky K, White AJ, Jenkinson WE, Anderson G, Agace WW. Mesenchymal cells regulate retinoic acid receptor-dependent cortical thymic epithelial cell homeostasis. *J Immunol*. 2012;188(10):4801-9.
343. Szondy Z, Reichert U, Bernardon J-M, Michel S, Tóth R, Ancian P, et al. Induction of Apoptosis by Retinoids and Retinoic Acid Receptor γ -Selective Compounds in Mouse Thymocytes through a Novel Apoptosis Pathway. *Molecular Pharmacology*. 1997;51(6):972-82.
344. Dillehay DL, Cornay WJ, Walia AS, Lamon EW. Effects of retinoids on human thymus-dependent and thymus-independent mitogenesis. *Clinical Immunology and Immunopathology*. 1989;50(1):100-8.
345. Cibrian D, Sanchez-Madrid F. CD69: from activation marker to metabolic gatekeeper. *Eur J Immunol*. 2017;47(6):946-53.
346. Ziegler SF, Ramsdell F, Alderson MR. The activation antigen CD69. *Stem Cells*. 1994;12(5):456-65.
347. Timsit J, Savino W, Safieh B, Chanson P, Gagnerault MC, Bach JF, et al. Growth hormone and insulin-like growth factor-I stimulate hormonal function and proliferation of thymic epithelial cells. *J Clin Endocrinol Metab*. 1992;75(1):183-8.

348. de Mello-Coelho V, Villa-Verde DMS, Dardenne M, Savino W. Pituitary hormones modulate cell–cell interactions between thymocytes and thymic epithelial cells. *Journal of Neuroimmunology*. 1997;76(1-2):39-49.
349. de Mello Coelho V, Villa-Verde DM, Farias-de-Oliveira DA, de Brito JM, Dardenne M, Savino W. Functional insulin-like growth factor-1/insulin-like growth factor-1 receptor-mediated circuit in human and murine thymic epithelial cells. *Neuroendocrinology*. 2002;75(2):139-50.
350. Kermani H, Goffinet L, Mottet M, Bodart G, Morrhaye G, Dardenne O, et al. Expression of the growth hormone/insulin-like growth factor axis during Balb/c thymus ontogeny and effects of growth hormone upon ex vivo T cell differentiation. *Neuroimmunomodulation*. 2012;19(3):137-47.
351. Kecha O, Brilot F, Martens H, Franchimont N, Renard C, Greimers R, et al. Involvement of insulin-like growth factors in early T cell development: a study using fetal thymic organ cultures. *Endocrinology*. 2000;141(3):1209-17.
352. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
353. Starmer J, Magnuson T. Detecting broad domains and narrow peaks in ChIP-seq data with hiddenDomains. *BMC Bioinformatics*. 2016;17:144.
354. Morris SA, Baek S, Sung MH, John S, Wiench M, Johnson TA, et al. Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. *Nat Struct Mol Biol*. 2014;21(1):73-81.
355. Hun M, Barsanti M, Wong K, Ramshaw J, Werkmeister J, Chidgey AP. Native thymic extracellular matrix improves in vivo thymic organoid T cell

output, and drives in vitro thymic epithelial cell differentiation. *Biomaterials*. 2017;118:1-15.

356. Fan Y, Tajima A, Goh SK, Geng X, Gualtierotti G, Grupillo M, et al. Bioengineering Thymus Organoids to Restore Thymic Function and Induce Donor-Specific Immune Tolerance to Allografts. *Mol Ther*. 2015;23(7):1262-77.

357. Gameiro J, Nagib P, Verinaud L. The thymus microenvironment in regulating thymocyte differentiation. *Cell Adhesion & Migration*. 2014;4(3):382-90.

358. Savino W, Mendes-da-Cruz DA, Golbert DC, Riederer I, Cotta-de-Almeida V. Laminin-Mediated Interactions in Thymocyte Migration and Development. *Front Immunol*. 2015;6:579.

359. Garcia-Ceca J, Montero-Herradon S, Alfaro D, Zapata AG. Increased epithelial-free areas in thymuses with altered EphB-mediated thymocyte-thymic epithelial cell interactions. *Histochem Cell Biol*. 2017;148(4):381-94.

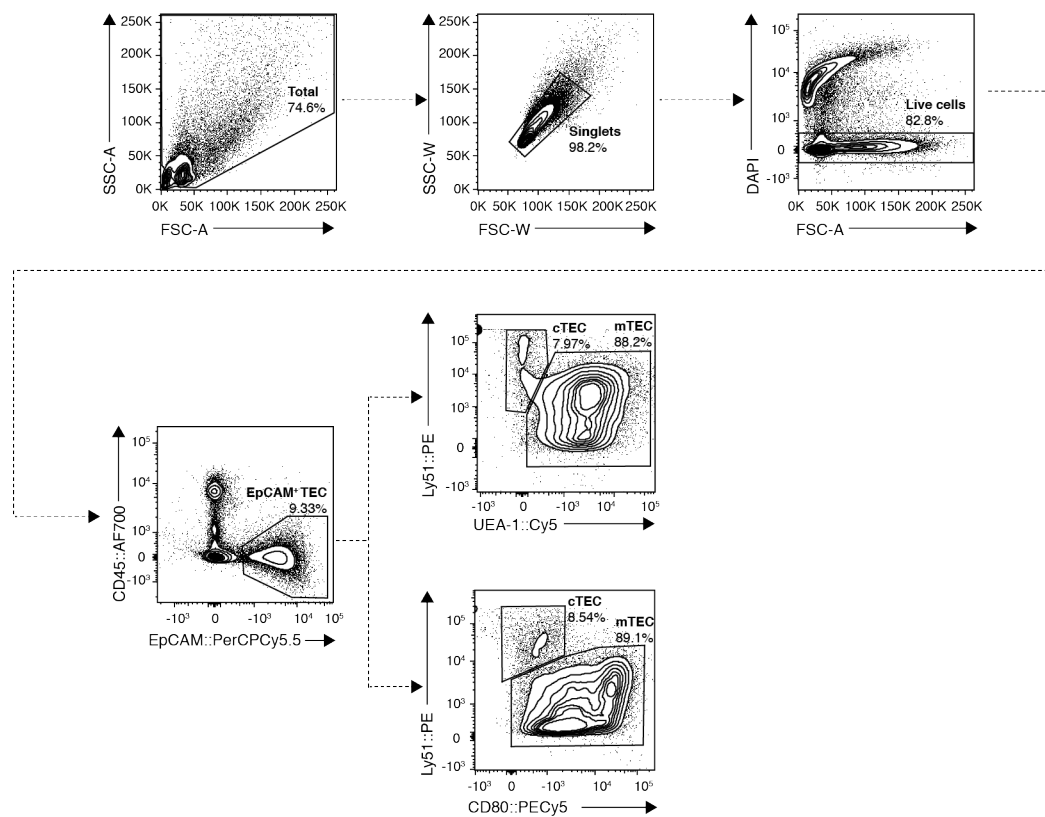
360. Garcia-Ceca J, Alfaro D, Montero-Herradon S, Tobajas E, Munoz JJ, Zapata AG. Eph/Ephrins-Mediated Thymocyte-Thymic Epithelial Cell Interactions Control Numerous Processes of Thymus Biology. *Front Immunol*. 2015;6:333.

361. Chaudhry MS, Velardi E, Dudakov JA, van den Brink MR. Thymus: the next (re)generation. *Immunol Rev*. 2016;271(1):56-71.

362. Su MA, Anderson MS. Pulling RANK on Cancer: Blocking Aire-Mediated Central Tolerance to Enhance Immunotherapy. *Cancer Immunol Res*. 2019;7(6):854-9.

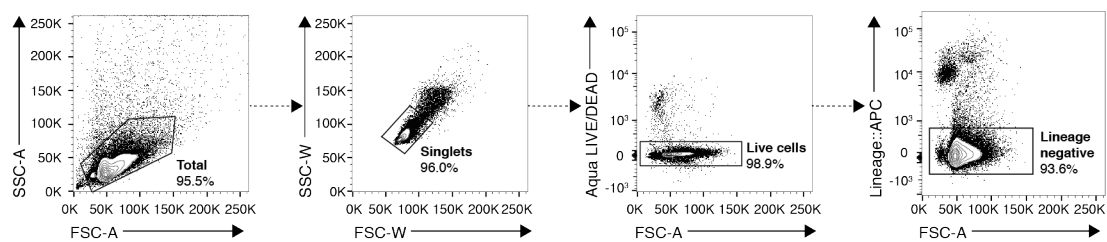
363. Sperlazza J, Rahmani M, Beckta J, Aust M, Hawkins E, Wang SZ, et al. Depletion of the chromatin remodeler CHD4 sensitizes AML blasts to genotoxic agents and reduces tumor formation. *Blood*. 2015;126(12):1462-72.
364. Heshmati Y, Turkoz G, Harisankar A, Kharazi S, Bostrom J, Dolatabadi EK, et al. The chromatin-remodeling factor CHD4 is required for maintenance of childhood acute myeloid leukemia. *Haematologica*. 2018;103(7):1169-81.
365. Xia L, Huang W, Bellani M, Seidman MM, Wu K, Fan D, et al. CHD4 Has Oncogenic Functions in Initiating and Maintaining Epigenetic Suppression of Multiple Tumor Suppressor Genes. *Cancer Cell*. 2017;31(5):653-68 e7.

Supplementary Figures



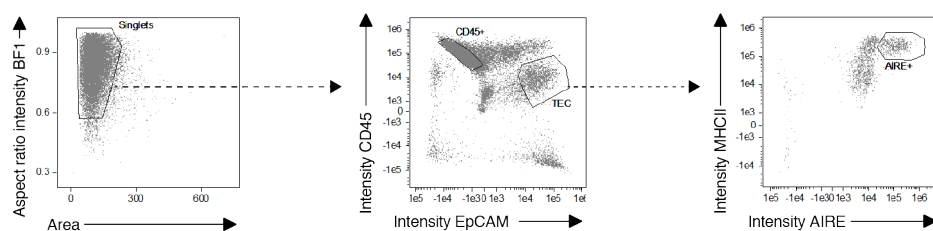
Supplementary Figure 1 Basic FACS gating strategy for TEC

Representative FACS plots showing basic gating strategy for TEC. Total cells were first selected, followed by singlets, and then live cells; TEC were next identified as CD45⁺EpCAM⁺; within the total TEC compartment cTEC were identified on the basis of Ly51 expression (Ly51⁺) and mTEC were identified either as Ly51⁻UEA-1⁺ or as Ly51⁻CD80^{Lo/Hi}; analysis of TEC was then conducted as per the figures presented in Chapters 3 and 4.



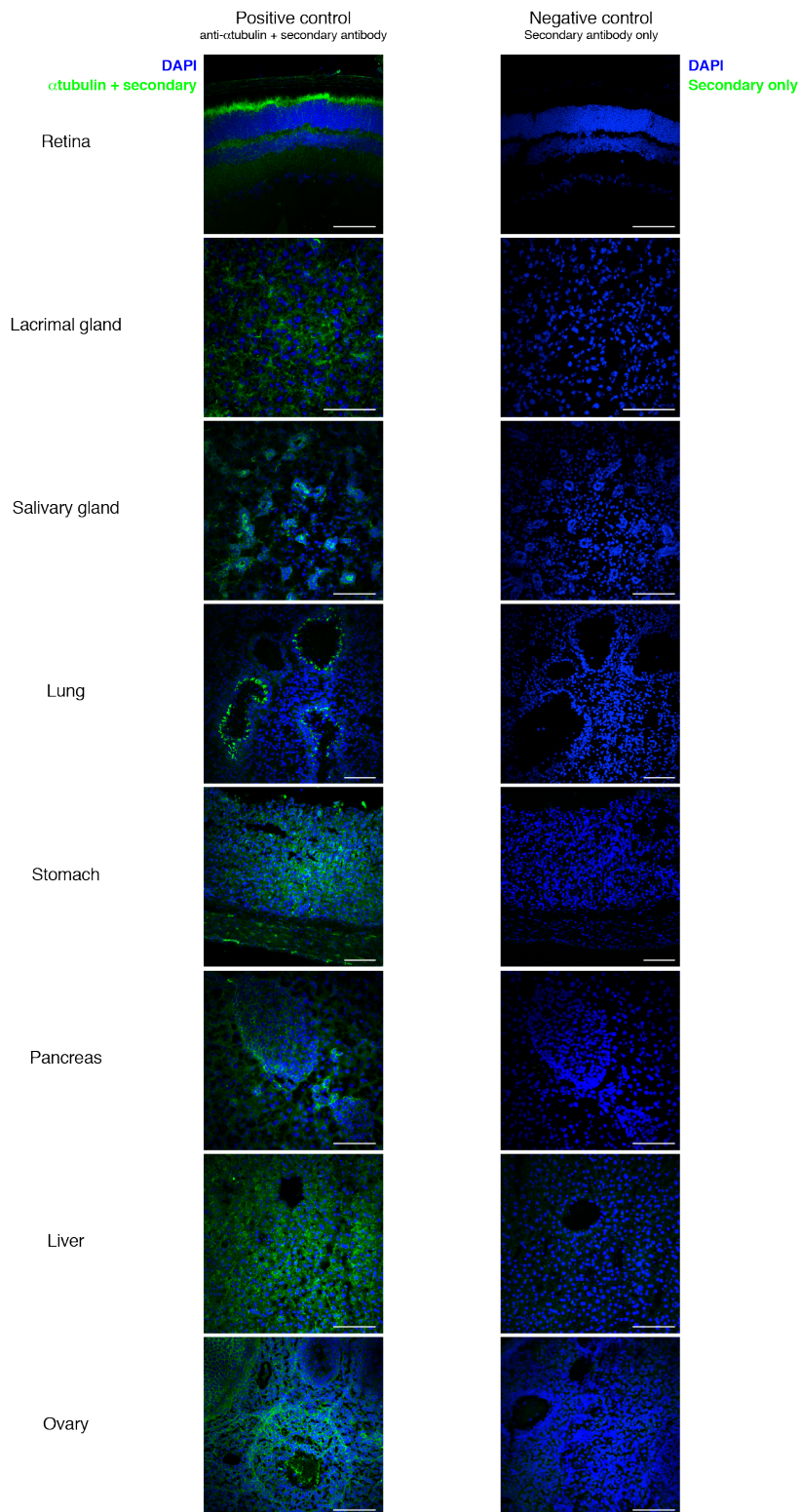
Supplementary Figure 2 Basic FACS gating strategy for thymocytes

Representative FACS plots showing basic gating strategy for Thymocytes. Total cells were first selected, followed by singlets, and then live cells. Next, intrathymic cells not belonging to the $\alpha\beta$ lineage were excluded by their expression of lineage markers for myeloid cells (CD11c, F4/80, CD11b, Gr1), B lymphoid cells (CD19), NK and NKT cells (CD49b and NK1.1), $\gamma\delta$ T-cells ($\gamma\delta$ TCR) and erythroid cells (Ter119); analysis of “lineage negative” thymocytes was then conducted as per the figures presented in Chapter 4.



Supplementary Figure 3 Basic gating strategy for Proximity Ligation Imaging Cytometry (PLIC)

Representative image stream plots showing basic gating strategy PLIC. Singlets were first selected, followed by TEC (CD45⁺EpCAM⁺); AIRE expressing mTEC were next identified on the basis of positivity for AIRE and expression of high levels of MHCII.



Supplementary Figure 4 Positive and negative control stains for indirect immunofluorescence staining

Immunofluorescence microscopy images showing positive (left panels) and negative (right panels) controls for the indirect immunohistological staining process used to detect serum autoantibodies in CHD4 cKO and WT mice. As a positive control (left panels) a slide from each tissue was stained with a mouse IgG anti- α tubulin antibody followed by the same secondary antibody used to detect binding of serum autoantibodies; and as a negative control (right panels) secondary antibody staining alone was used; scale bars represent 100 μ m; tissues were derived from RAG deficient mice.