DISSERTATIONES MEDICINIAE UNIVERSITATIS TARTUENSIS
167

DISSERTATIONES MEDICINIAE UNIVERSITATIS TARTUENSIS
167

TÕNIS ORG

Molecular function of the first PHD finger domain of Autoimmune Regulator protein

TARTU UNIVERSITY
PRESS

Molecular Pathology, Institute of General and Molecular Pathology, University of Tartu

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (medicine) on the 17th of February, 2010 by the Council of the Faculty of Medicine, University of Tartu, Estonia.

Supervisor: Professor Pärt Peterson, PhD
Molecular Pathology, Institute of General Molecular Pathology, University of Tartu, Estonia

Reviewed by: Anti Kalda, MD, PhD
Institute of Pharmacology, University of Tartu, Estonia
Arnold Kristjuhan, PhD
Institute of Molecular and Cell Biology, University of Tartu, Estonia

Opponent: Professor Rein Aasland, PhD
Department of Molecular Biology, University of Bergen, Norway

Commencement: 15.04.2010

Publication of this dissertation is granted by University of Tartu

This research was supported by the European Regional Development Fund

ISSN 1024–395x
ISBN 978–9949–19–325–7 (trükis)
ISBN 978–9949–19–326–4 (PDF)

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 106

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

I **Org T**, Chignola F, Hetényi C, Gaetani M, Rebane A, Liiv I, Maran U, Mollica L, Bottomley MJ, Musco G, Peterson P. The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. EMBO Rep. 2008 Apr;9(4):370–6.

II Chignola F, Gaetani M, Rebane A, **Org T**, Mollica L, Zucchelli C, Spitaleri A, Mannella V, Peterson P, Musco G. 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 May;37(9):2951–61.

III **Org T**, Rebane A, Kisand K, Laan M, Haljasorg U, Andreson R, Peterson P. AIRE activated tissue-specific genes have histone modifications associated with inactive chromatin. Hum Mol Gen. 2009 Dec 15;18(24):4699–710

IV Peterson P, **Org T**, Rebane A. Transcriptional regulation by AIRE: molecular mechanisms of central tolerance. Nat Rev Immunol. 2008 Dec;8(12):948–57. Review.

Contribution of Tõnis Org to original publications:

Study I: participation in study design; performing the experiments, except for NMR, ITC and fluorescence titration assays; writing the paper

Study II: performing the peptide binding and Western blot experiments; manuscript preparation

Study III: participation in study design, performing the experiments and data analysis, writing the paper

Study IV: participation in preparing the manuscript and figures

ABBREVIATIONS

aa	amino acid
ACF1	ATP-dependent chromatin assembly factor 1
AIRE	Autoimmune Regulator
AL	alfin1-like (AL) protein
ALPS	autoimmune lymphoproliferative syndrome
APCs	antigen presenting cells
APECED	Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy
APS-1	Autoimmune Polyendocrine Syndrome type-1
AtING	*Arabidopsis thaliana* ING (inhibitor of growth) homologues
ADP	adenosine diphosphate
ATP	adenosine triphosphate
ATRX	Alpha Thalassaemia and Mental Retardation Syndrome, X-linked
AtSCE1	*Arabidopsis thaliana* sumo conjugation enzyme 1
AtSIZ1	*Arabidopsis thaliana* SAP and MIZ1
BCL9	B-cell CLL/lymphoma 9
BPTF	bromodomain and PHD finger-containing transcription factor
BRCT	BRCA1 C-terminal
CARD	caspase recruitment domain
CBP	CREB-binding protein
CD	cluster of differentiation
CDH12	cadherin 12
CDK9	cycline dependent kinase 9
CEAS	cis-regulatory element annotation system
CHD1	chromodomain-helicase-DNA-binding protein 1
ChIP	chromatin immunoprecipitation
CLDN1	claudin 1
DNA	deoxyribonucleic acid
cRNA	complementary RNA
CYCT1	cyclin T1
CYP	cytochrome
DC	dendritic cell
DMEM	Dulbecco's modified Eagle's medium
DN	double-negative
DNA-PK	DNA-dependent protein kinase
DP	double positive
DSB	double stranded brake
EFHC2	EF-hand domain C-terminal containing 2
FACT	facilitates chromatin transcription
FDR	false discovery rate
GAD2	glutamate decarboxylase 2
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
Gcn5p	general control of amino-acid synthesis 5p

GEO	Gene Expression Omnibus
GHR	growth hormone receptor
GST	glutathione S-transferase
GTE3	global transcription factor group E3
HAT	histone acetyltransferase
HBG2	hemoglobin gamma
HCP5	HLA complex protein P5
HD1	homology domain 1
HDAC	histone deacetylase
HDM	histone demethylase
HEK293	human embryonic kidney 293 cells
HP1	heterochromatin protein 1
HSBP8	human selenium binding protein 8
HSR	homologously staining region
ICBP90	inverted-CCAAT-box-binding protein of 90 kDa
IFI16	interferon, gamma-inducible protein 16
IL-2Rα	interleukin 2 receptor α
ING2	inhibitor of growth-2
INS	insulin
INV	involucrin
IPEX	immune dysregulation, polyendocrinopathy, enteropathy, X-linked
ITC	isothermal titration calorimetry
JMJD2A	jumonji domain 2
KAP1	kinase anchor protein 1
KPNB1	karyopherin subunit beta-1
L3MBTL	lethal(3)malignant brain tumor-like protein
LOR	loricrin
LPL	lipoprotein lipase
MAL2	MAL proteolipid protein 2
MBT	malignant brain tumor
MDC1	mediator of DNA damage checkpoint protein 1
MHC	major histocompatibility complex
mRNA	messenger RNA
MS	mass spectrometry
Msc1	meiotic sister-chromatid recombination 1
MSH2/6	mutS homolog 2/6
mTECs	medullary thymic epithelial cells
MYRIP	myosin VIIA and Rab interacting protein
NC	negative control
NLS	nuclear-localization signal
NMR	nuclear magnetic resonance
NSD1	nuclear receptor-binding SET domain-containing 1
NUP93	nucleoporin 93 kDa
NURF	nucleosomal remodeling factor

OAS3	2'-5'-oligoadenylate synthetase 3
OMIM	Online Mendelian Inheritance in Man
PAK3	p21-activated kinase 3
PARP1	poly (ADP-ribose) polymerase 1
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein database
PHD	plant homeodomain
PIAS1	protein inhibitor of activated STAT1
PLXDC2	plexin domain containing 2
PML	promyelocytic leukemia
PMSF	phenylmethyl sulphonyl fluoride
polII	RNA polymerase II
PRMT6	protein arginine methyltransferase 6
PRR	proline rich region
PSMD4	proteasome 26S non-ATPase regulatory subunit 4
P-TEFb	positive transcription elongation factor b
Q^2ChIP	quick and quantitative ChIP
qPCR	quantitative PCR
RAG2	recombination activating gene 2
RANK	receptor activator of NF-κB
Rco1	essential subunit of the histone deacetylase Rpd3S complex 1
REEF	regionally enriched features
RING	really interesting new gene
RNA	ribonucleic acid
Rpd3S	reduced potassium dependency 3S
RPMI	Roswell Park Memorial Institute
RT-PCR	reverse transcription PCR
RUVBL2	ruvB-like 1
SAND	Sp100, AIRE-1, NucP41/75 and DEAF-1/suppressin
SD	standard deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SMC1	structural maintenance of chromosomes 1
SP	single-positive
SPRR2F	small proline-rich protein 2F
SRA	SET- and RING-associated
STAT1	signal transducers and activators of transcription
SUMO	small ubiquitin-like modifier
SUSD4	sushi domain containing 4
TAF1	TATA box binding protein (TBP)-associated factor
TCR	T-cell receptor
TECs	thymic epithelial cells
TG	thyroglobulin
THBS2	thrombospondin 2
THP-1	human acute monocytic leukemia cell line

TiGER	tissue-specific gene expression and regulation
TOP2a	DNA topoisomerase II alpha
TOX3	thymus high mobility group box family member 3
TPO	thyroid peroxidase
TRIM28	tripartite motif-containing 28
TSS	transcription start site
WDR5	WD repeat-containing protein 5
WT	wild type
XPO1	exportin 1
YFP	yellow fluorescent protein
ZBTB7B	zinc finger and BTB domain containing 7B

1. INTRODUCTION

The immune system has evolved to protect the host from diseases, which can be caused by pathogens or tumor cells. It consists of specialized molecules, cells, tissues and organs. In order to function properly, the immune system has to identify and mount immune responses to foreign molecules while remaining tolerant to molecules produced by the host. How the immunological tolerance to self antigens is achieved has intrigued scientists for years and it is far from being fully understood. The breakdown of the immunological tolerance to self may lead to autoimmune diseases. The causal mechanisms of majority of autoimmune diseases are unknown, because they are complex traits influenced by different genetic and environmental factors.

In rare cases, mutations in a single gene are sufficient to cause an autoimmune disease. These rare diseases provide a unique model to better understand the mechanisms that lead to the development or break-down of self-tolerance. One such example is APECED (Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy) syndrome caused by mutations in *AIRE* (Autoimmune Regulator) gene. APECED is characterized by a breakdown of self-tolerance leading to destructive autoimmune reactions in several, mainly endocrine, organs. AIRE is predominantly expressed in thymic medullary epithelial cells (mTECs) where it promotes the promiscuous expression of tissue-specific antigens, which can then be presented to developing thymocytes. This results in proper negative selection of autoreactive T-cells and is therefore crucial for the establishment of central tolerance.

The molecular mechanisms behind AIRE mediated gene expression activation are not clear. One factor that influences gene expression is chromatin structure, which is shaped by several means, including histone posttranslational modifications. Certain protein domains, like PHD (plant homeodomain) zinc fingers, have been shown to recognize specific histone posttranslational modifications. Proteins with such domains can selectively interact with chromatin and thereafter launch downstream events.

Current thesis focuses on studying the molecular mechanisms behind AIRE promoted promiscuous expression of tissue-specific antigens, concentrating on the function of the first PHD finger domain of AIRE.

2. REVIEW OF LITERATURE

2.1. Central tolerance and AIRE

2.1.1. Immune tolerance

Immune or immunological tolerance is a property of the immune system not to react against specific antigens. Although immune tolerance can be generated against foreign antigens, it is especially important in maintaining unresponsiveness to self constituents. The mechanisms that lead to self-immune tolerance can be broadly divided into two: central tolerance and peripheral tolerance. Central tolerance occurs in the thymus during lymphocyte development and leads to the elimination of majority of T-cells that are potentially dangerous to the host (Kyewski and Klein, 2006). Peripheral tolerance operates outside of thymus and is needed to keep these self-reactive T-cells under control that have escaped the mechanisms of central tolerance (Mueller, 2010).

2.1.2. Thymus as a place for central tolerance

T-cells originate from the common lymphoid progenitors residing in the bone marrow but for their maturation and selection the T-cell progenitors migrate to the thymus. Thymus consists of two lobes that are organized into numerous lobules each consisting of separate outer cortex, which contains the majority of immature thymocytes, and inner medulla, which contains more mature T-cells. The thymus is hold together by supporting network of stromal cells including thymic epithelial cells (TECs), dendritic cells (DC), macrophages and fibroblasts. Thymocytes and stromal cells interact tightly with each other. Thymic stromal cells are needed for the T-cell development and selection. Thymocytes in turn coach the stromal cells to provide the appropriate microenvironment to promote further thymocyte development. This bidirectional lympho-stromal crosstalk creates a unique environment that cannot be recapitulated by any other tissue in the body (Nitta *et al.*, 2008).

Within thymus, thymocytes undergo a series of maturation stages, which require their dynamic relocation and can be classified by the surface expression of CD4 and CD8 molecules (Figure 1). After entering the thymus at the corticomedullary junction, $CD4^-CD8^-$ double negative (DN) thymocytes first migrate through the cortex outwards to the sub-capsular zone, then move back inwards, become double positive (DP) and finally, once they reach medulla, become single positive (SP). During this journey, thymocytes rearrange their T-cell receptor (TCR) gene sequences, which lead to TCRs with a huge repertoire of antigen specificities. As the generation of TCR diversity is random, many of the resulting TCRs are either unfunctional or recognize self-proteins. The generation of T-cell repertoire that is functional and self-tolerant is achieved when thymocytes are presented with different self-peptides bound to the major histocompatibility (MHC) complex on the surface of antigen-presenting cells (APCs). T cells that do not recognize MHC molecules with high enough avidity

undergo apoptosis. This process, known as a positive selection, takes place in the cortex and ensures the elimination of unfunctional T-cells. In order to maintain tolerance to self-antigens, autoreactive T-cells that have passed positive selection go through further screening. Thymocytes with TCRs that bind to MHC-peptide complex with high avidity are eliminated by apoptosis. This process, called negative selection, occurs mainly in medulla and ensures that most of the self-reactive T-cells are eliminated before they migrate out of the thymus. Positive and negative selection together eliminate more than 95% of all thymocytes and only the small minority of remaining thymocytes can finish their differentiation and migrate out of the thymus (Kyewski and Klein, 2006; Klein *et al.*, 2009).

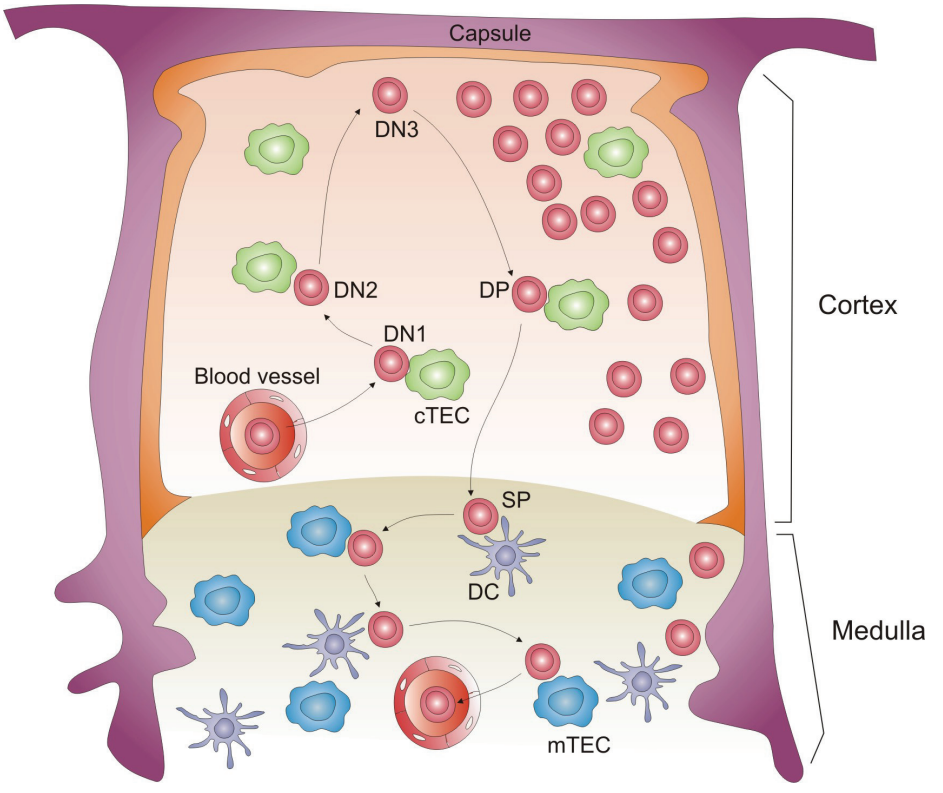

Figure 1. The thymocyte migratory route in the thymus. Developmental stages (double-negative (DN), double-positive (DP) and single-positive (SP)) and interactions of thymocytes with stromal cells (cortical thymic epithelial cells (cTECs), medullary TECs (mTECs) and dendritic cells (DCs)), that are needed for positive and negative selection, are indicated. Adapted from (Klein *et al.*, 2009).

While dendritic cells are the most efficient antigen presenting cells in the thymus, mTECs have unique role in maintaining self-tolerance since they are able to express large number of genes that are usually expressed only in specific tissues (Derbinski *et al.*, 2001). This ectopic or promiscuous expression of tissue-specific genes significantly broadens the repertoire of self-antigens that can be presented to developing T-cells and is therefore crucial for establishing central self-tolerance (Tykocinski *et al.*, 2008).

2.1.3. Autoimmunity

The failure of an organism to recognize its own constituents as self leads to autoimmunity, which results in an immune response against its own cells and tissues and may lead to an autoimmune disease. Autoimmune disorders fall into two general types: those that damage many organs (systemic) and those where only a single organ or tissue is directly damaged by the autoimmune process (organ-specific). Some of the most common types of autoimmune disorders include: type 1 diabetes, celiac disease, systemic lupus erythematosus, rheumatoid arthritis. The cause of majority of the autoimmune diseases is unknown, but in many cases, in addition to environmental factors, there is an inherited genetic predisposition to develop an autoimmune disease (Gregersen and Olsson, 2009). In rare cases mutations in a single gene are sufficient to cause an autoimmune disease. These diseases provide unique model systems to study different aspects of self-tolerance development. Currently there are four examples of monogenic autoimmune diseases: IPEX (immune dysregulation, polyendocrinopathy, enteropathy, X-linked) (OMIM 304930), ALPS (autoimmune lymphoproliferative syndrome) (OMIM 601859), IL-2Rα deficiency (OMIM 601859), and APECED (OMIM 240300) (Lehman and Ballow, 2008).

2.1.4. Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy

Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy (APECED) also known as Autoimmune Polyendocrine Syndrome Type 1 (APS-1) is a monogenic autoimmune disease inherited in autosomal recessive manner (Husebye *et al.*, 2009). APECED is a rare disease worldwide, with higher frequency among certain populations such as Iranian Jews (1:9000) (Zlotogora and Shapiro, 1992), Sardinians (1:14000) (Rosatelli *et al.*, 1998) and Finns (1:25000) (Ahonen *et al.*, 1990). The disease is characterized by breakdown of self-tolerance leading to autoimmunity that targets multiple, most commonly endocrine organs. The classic triad of clinical manifestations includes hypoparathyroidism, primary adrenocortical insufficiency and chronic mucocutaneous candidiasis. In addition, many other autoimmune disease components occur less frequently in various combinations (Husebye *et al.*, 2009). Interestingly, the pattern of autoimmune disease in this syndrome shows a tendency for organ-specific autoimmunity, because systemic autoimmune diseases, such

as systemic lupus erythematosus and rheumatoid arthritis, are not usually described in APECED patients (DeVoss and Anderson, 2007). The diagnosis of APECED has been simplified by the finding of high titre autoantibodies against interferon-ω and -α2 at early stage of the disease, which provides a relatively simple diagnostic test (Meager *et al.*, 2006). After the identification of the disease-causing gene *AIRE* in 1997, APECED has become a major disease model for autoimmunity, which has lead to a new understanding of central immunological tolerance (Finnish-German_APECED_Consortium, 1997; Nagamine *et al.*, 1997).

2.1.5. AIRE

The Autoimmune Regulator (*AIRE)* gene locates on human chromosome 21q22.3, is 13 kb in size, consists of 14 exons and encodes for a 545 amino acid protein with a molecular mass of approximately 58 kDa. The AIRE protein contains several domains characteristic for transcriptional regulators and chromatin-binding proteins (Figure 2).

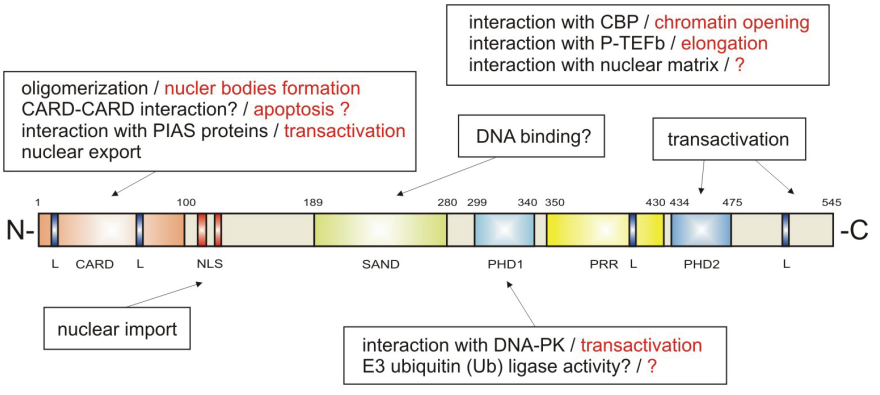

Figure 2. Schematic representation of human AIRE protein. The domains and functional elements of the AIRE protein are shown in different colors; CARD, SAND, PHD, PRR, L, NLS. Functions of AIRE are shown in boxes and indicated by arrows pointing to the corresponding domain, if known.

The N-terminal region of AIRE contains homologously staining region (HSR) domain, which is structurally highly similar to caspase recruitment domain (CARD) (Ferguson *et al.*, 2008) and has been shown to be responsible for the dimerization of the protein (Pitkanen *et al.*, 2001). CARD domain is followed by SAND domain, named after Sp100, AIRE-1, NucP41/75 and DEAF-1/suppressin, that has been shown to be responsible for DNA binding in other proteins (Gibson *et al.*, 1998). The DNA-binding specificity of SAND domain in other proteins seems to be rather flexible; comprising short recognition motives that contain CG nucleotides and is therefore methylation sensitive

(Burnett *et al.*, 2001; Isaac *et al.*, 2006). The KD/NWK motif needed for DNA binding, however, is not conserved in AIRE. The C-terminal region of AIRE contains two plant homeodomain (PHD) zinc fingers, which are separated by proline rich region (PRR). Mutations in AIRE PHD fingers markedly reduce the transcriptional activation capacity of the protein (Bjorses *et al.*, 2000; Pitkanen *et al.*, 2000; Halonen *et al.*, 2004; Meloni *et al.*, 2008). The first PHD finger has been suggested to have E3 ubiquitin ligase activity (Uchida *et al.*, 2004), this has, however, not been supported by other studies (Bottomley *et al.*, 2005). No specific function has been found for the second PHD finger of AIRE. In addition, AIRE contains nuclear localization signal (NLS) and four LXXLL (L) motives, which occur in transcriptional co-activator proteins such as CBP (CREB-binding protein) (Pitkanen *et al.*, 2001).

Much of the information available for AIRE has become from the studies with mice. Human and mouse AIRE are highly similar. The overall identity between the mouse and human AIRE proteins is 71%. Mouse *Aire* localizes to chromosome 10, is split into 14 exons over 13 kb and encodes for a protein of 552 amino acids (Mittaz *et al.*, 1999). Similar to APECED patients, mice deficient for Aire develop autoimmune phenotype with lymphocytic infiltrations in several peripheral tissues, most commonly in salivary glands, eyes, stomach and liver (Anderson *et al.*, 2002; Ramsey *et al.*, 2002). Although there are several differences in autoantigens and disease manifestations between mice and human, Aire deficient mice still provide a good model for studying AIRE function and the mechanisms of central tolerance (Mathis and Benoist, 2009).

2.1.6. AIRE function in the thymus

AIRE is expressed in thymus mainly by a subpopulation of mTECs, which are characterized by the high expression of MHC class II molecules and are positive for the co-stimulatory molecules CD80 and CD86 (Heino *et al.*, 1999; Derbinski *et al.*, 2005). These are considered to be mature, terminally differentiated mTECs (Derbinski *et al.*, 2005; Gray *et al.*, 2007). Intracellularly, AIRE localizes to nucleus, where it forms discrete dot-like structures, which are distinct from promyelocytic leukemia (PML) nuclear bodies (Halonen *et al.*, 2001; Pitkanen *et al.*, 2001). The function of the AIRE nuclear bodies is not known, but it seems that they are not the sites of active transcription (Pitkanen *et al.*, 2005).

Gene expression analysis of mTECs from Aire deficient and WT mice thymus revealed that Aire is needed for the expression of large repertoire of transcripts (200-1200), many of which are characterized by tissue-specific expression pattern (Anderson *et al.*, 2002). Further experiments showed that the thymic expression of peripheral tissue-specific antigens is needed for the proper negative selection of self-reactive T-cells (Liston *et al.*, 2003; Anderson *et al.*, 2005). Therefore, according to current knowledge, the mechanism by which AIRE ensures proper negative selection involves the promotion of tissue-specific antigen expression in mTECs (Figure 3).

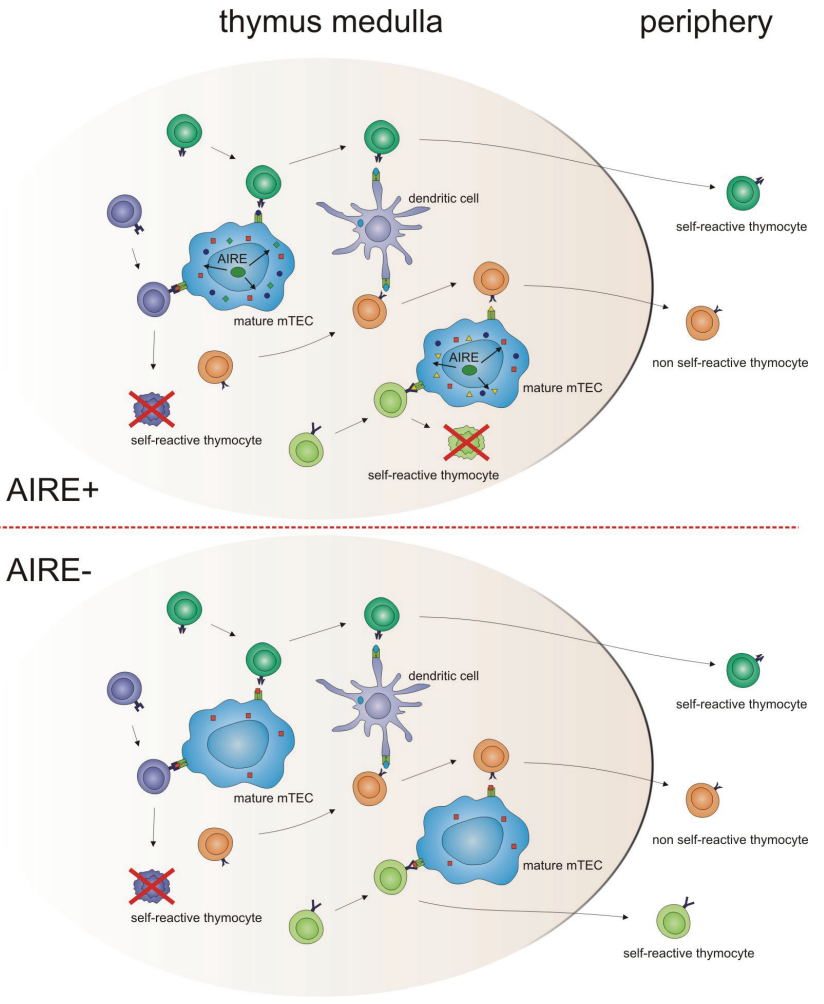

Figure 3. AIRE enhances the scope of negative selection. Thymocytes expressing various TCR specificities migrate through the thymic medulla and make contact with mTECs and dendritic cells, which present self-peptide-MHC complexes on their surface. The avidity of interaction between the TCR and peptide-MHC complex determines the survival of the thymocyte, where self-reactive thymocytes are deleted by apoptosis (red cross). In the presence of AIRE, mTECs express a large number of tissue-specific antigens (green rhombs, blue circles, and yellow triangles inside mTEC). Although single mTECs do not express all the tissue-specific antigens, the pool of mTECs encompasses self antigens that cover essentially all tissues of the body. The presentation of these antigens to developing thymocytes results in the efficient negative selection of self-reactive thymocytes (blue and green) so that mostly non-self-reactive thymocytes exit to the periphery (orange). In the absence of AIRE, a large subset of potentially self-reactive thymocytes (green) is not deleted, because of aberrant expression and presentation of certain tissue-specific antigens, and these self-reactive cells escape to the periphery.

The importance of intrathymic expression of tissue-specific antigens is underlined by studies which show that lack of thymic expression of a single antigen can be sufficient to elicit an autoimmune attack against the corresponding peripheral tissue (DeVoss *et al.*, 2006; Gavanescu *et al.*, 2007; Fan *et al.*, 2009). Therefore, to ensure immune tolerance, the expression of tissue-specific antigens in the thymus must be carefully regulated. In addition to promoting the expression of tissue-specific genes, there is evidence that AIRE might also influence other processes, such as processing and/or presentation of self-antigens (Kuroda *et al.*, 2005), differentiation of mTECs (Yano *et al.*, 2008) or induction of apoptosis (Gray *et al.*, 2007).

2.1.7. AIRE as a transcriptional activator

Already the domain structure of the AIRE protein suggests that it might function as transcriptional activator. Indeed, in addition to mTECs AIRE is capable of activating gene expression in various cell models. The region responsible for transcriptional activation was first mapped to the two PHD zinc fingers (Bjorses *et al.*, 2000; Pitkanen *et al.*, 2000). Later studies showed that other parts of the protein and several mutations in CARD and SAND domain also influence the transactivation ability to different degrees (Halonen *et al.*, 2004; Pitkanen *et al.*, 2005; Ferguson *et al.*, 2008; Meloni *et al.*, 2008). Therefore, it seems that different regions of AIRE have distinct functions, which depending on the experimental model used are required for the transactivating ability.

The molecular mechanisms behind AIRE mediated transcriptional activation are not well understood. Although there are studies suggesting that AIRE might directly bind to certain DNA sequences (Kumar *et al.*, 2001; Purohit *et al.*, 2005; Ruan *et al.*, 2007), conclusive data showing that AIRE binds to defined *cis*-acting sequences are still lacking. Instead, recent data suggest that AIRE might bind DNA in a rather unspecific fashion (Koh *et al.*, 2008). Data that AIRE is capable of activating large number of genes, which vary according to the cell type also suggest that it might have more general function in transcription (Gardner *et al.*, 2008; Guerau-de-Arellano *et al.*, 2008). The variation, or stochastic nature of AIRE regulated gene expression, is further extended by studies which used single cell PCR to show that the expression of a given AIRE dependent tissue-specific gene in mTECs is probabilistic (Derbinski *et al.*, 2008; Villasenor *et al.*, 2008). In addition, the expression can be monoallelic, use different start sites compared to peripheral tissues and is independent of transcriptional regulators required for the expression of these genes in the periphery (Villasenor *et al.*, 2008). Furthermore, the promiscuous gene expression of tissue-specific genes, including those that are controlled by AIRE, show high intra-individual variability, both in human and in mice (Taubert *et al.*, 2007; Venanzi *et al.*, 2008).

Most likely AIRE cooperates with other proteins in order to exert its functions. In principle, these other proteins could work upstream, downstream or concurrently with AIRE. Indeed, AIRE has been shown to interact with

several proteins. CBP, a common transcriptional co-activator was the first protein shown to interact with AIRE (Pitkanen *et al.*, 2000). CBP is a histone and non-histone acetyltransferase that synergizes with AIRE to activate transcription of several reporter constructs and endogenous genes (Akiyoshi *et al.*, 2004; Pitkanen *et al.*, 2005; Ferguson *et al.*, 2008). In case of low PML protein levels, AIRE and CBP preferentially colocalize to nuclear bodies (Pitkanen *et al.*, 2000). Recent data show that upregulation of AIRE through the RANK (receptor activator of NF-κB) signaling in fetal thymic organ cultures is accompanied by migration of CBP from cytoplasm to nucleus where it can cooperate with AIRE in transcriptional activation (Ferguson *et al.*, 2008).

AIRE has been shown to interact with DNA-dependent protein kinase (DNA-PK) complex, which is a serine/threonine kinase, composed of two regulatory subunits (Ku70 and Ku80) and a large catalytic subunit DNA-PKcs (Liiv *et al.*, 2008). DNA-PK is involved in DNA repair and in the phosphorylation of many proteins that are implicated in transcription and regulation of cell cycle. DNA-PK can phosphorylate AIRE *in vitro* and this influences AIRE-mediated transcriptional regulation, as mutations in the putative DNA-PK phosphorylation sites in AIRE markedly decrease its transcriptional activation ability (Liiv *et al.*, 2008).

Protein inhibitor of activated STAT1 (PIAS1) is another interaction partner of AIRE (Ilmarinen *et al.*, 2008). Although PIAS proteins function as E3 SUMO ligases, AIRE is not sumoylated. AIRE and PIAS1 localize to adjacent nuclear bodies and the simultaneous expression of PIAS1 and AIRE results in the coactivation of an insulin-promoter-driven reporter, and this co-activation is dependent on the SP-RING domain of PIAS1 (Ilmarinen *et al.*, 2008).

Both of these AIRE partner proteins (DNA-PK and PIAS1) have been shown to bind nuclear-matrix-associated DNA sequences (Mauldin *et al.*, 2002; Okubo *et al.*, 2004). Nuclear matrix serves as a platform for the organization of macromolecular structures, such as chromatin and nuclear bodies (Hancock, 2004; Rowat *et al.*, 2008). In agreement with this, AIRE has been also shown to be tightly bound to nuclear matrix (Tao *et al.*, 2006). However, the functional details of these interactions with AIRE and their relevance in nuclear matrix mediated transcriptional regulation remain to be studied.

A study by Oven et. al showed that AIRE promotes transcriptional elongation by binding and recruiting positive transcription elongation factor b (P-TEFb) complex to target gene promoters with stalled DNA polymerase II (Oven *et al.*, 2007). P-TEFb is a heterodimer of CDK9 and CYCT1 proteins that is a key factor in promoting transcriptional elongation. Mammalian gene transcription is initiated by the binding of RNA polymerase II to gene promoters and the formation of the pre-initiation complex. Recent studies have shown that large number of inactive eukaryotic genes actually contain pre-initiation complexes on their promoter regions, that produce short mRNAs but are unable to proceed further (Guenther *et al.*, 2007; Muse *et al.*, 2007). The phosphorylation of RNA polymerase II by P-TEFb results in dissociation of the negative elongation factors and enable the transcription to proceed to elongation phase (Peterlin and Price, 2006).

In a very recent study, Abramson et. al performed a broad screen to identify AIRE interacting proteins by co-immunoprecipitations followed by mass spectrometry (MS) (Abramson *et al.*, 2010). They managed to identify many proteins in complex with AIRE, which could be divided into four functional categories: proteins involved in nuclear transport (e.g. XPO1, NUP93, KPNB1), which were shown to mediate the shuttling of AIRE into or out of the nucleus; chromatin binding/structure (e.g. histones, the cohesin complex, etc); post-initiation RNA polII mediated transcription events (e.g. a putative complex of DNA-PK/PARP1/TOP2a/FACT) and pre-mRNA processing (including several splicing factors and RNA helicases). They also evaluated the functional relevance of candidate AIRE-associated proteins by a number of RNA interference based and other assays focusing on proteins in the two latter functional categories. In a series of experiments, they showed that AIRE is involved in promotion of TOP2a-initiated double stranded brakes (DSBs), which are able in turn to recruit and activate multiple members of this complex, such as DNA-PK, Ku80, PARP1 and histone H2AX, as well as several other AIRE-associated molecules, including SMC1, TRIM28, MSH2/6, and RUVBL2. TOP2a-introduced DNA breaks ease superhelical tensions generated by advancing RNA polII during transcriptional elongation, which helps it to proceed more smoothly by unwinding the DNA helix, resulting in more efficient mRNA synthesis. In another series of experiments they showed that, while AIRE controlled the levels of fully spliced mRNAs from a number of genes, it had strikingly little effect on the corresponding unspliced pre-mRNAs suggesting its role in mRNA processing. Considering that in eukaryotes, transcript elongation and pre-mRNA processing occur simultaneously, and that they are tightly coupled spatially (Moore and Proudfoot, 2009), the above findings are not surprising. However, the exact role of AIRE in modulating the above mentioned processes requires more research.

2.2. Transcriptional control of gene expression

2.2.1. Chromatin

For the expression of a gene, the underlying DNA needs to be accessible to the transcriptional machinery. Eukaryotic DNA is not free in the cell nucleus but is tightly associated with proteins forming chromatin. The basic subunits of chromatin are nucleosomes, which consist of histone octamers wrapped almost twice around by 147 bp of DNA (Figure 4). Histone octamers consist of four small basic proteins – histones H2A, H2B, H3, and H4, two copies of each. At active genes or at genes that are poised for activation, histones H2A and H3 are replaced by the histone variants H2A.Z and H3.3 (Jiang and Pugh, 2009). Nucleosomes are separated by 10–80 bp of linker DNA and histone H1. Non-condensed nucleosomes without the linker histone form "beads on a string" structure and histone H1 is involved with the packing of these sub-structures into a higher order structures (Happel and Doenecke, 2009). Chromatin does not only have structural role, allowing compaction of DNA within the nucleus but it

also regulates a number of processes essential for normal cellular functions, including gene transcription, DNA replication, repair, and recombination (Li *et al.*, 2007). The dynamic access to packed DNA required by these processes is mediated by various means, including chromatin remodeling complexes, which use the energy of ATP hydrolysis to move, destabilize, eject or restructure nucleosomes (Clapier and Cairns, 2009).

2.2.2. Histone posttranslational modifications

In addition to chromatin remodeling complexes, the access to DNA is regulated by the posttranslational modification of certain amino acids on the histones. These include the acetylation and methylation of lysines (K) and arginines (R) phosporylation of serines (S) and threonines (T), ubiquitinylation, sumolyation and biotinylation of lysines as well as ADP ribosylation (Figure 4). Moreover, the methylation can occur at different degrees – up to three methyl groups can be added to lysines and two methyl groups (symmetric or asymmetric) to arginines. More than 60 different modification sites have been described, which are found more often at the N-terminal tails of histones that protrude from the nucleosome core and are thus more accessible to the enzymes that catalyze the addition or removal of the covalent modifications (Kouzarides, 2007).

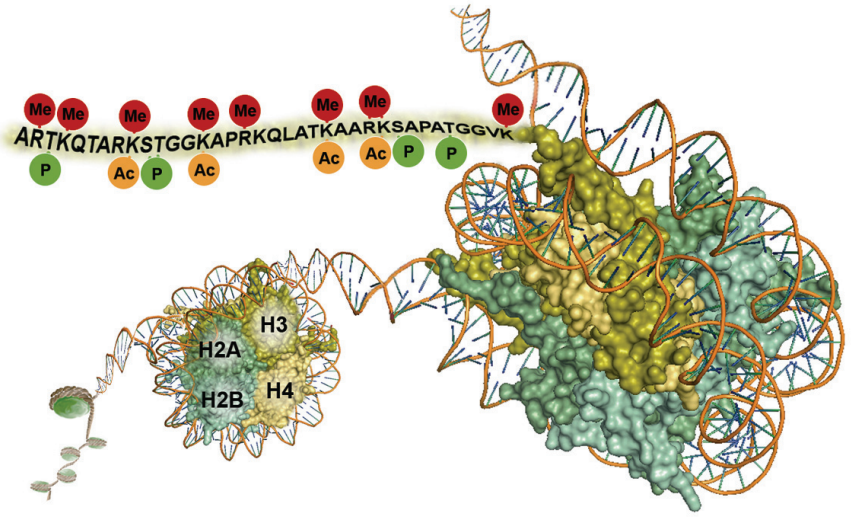

Figure 4. Nucleosomal structure of chromatin and posttranslational modifications on histone H3 N-terminal tail (Me – methylation, Ac – acetylation, P – phosphorylation).

Histone modifications are dynamic and their addition or removal is mediated by different enzymes. The first histone modifying enzymes found were histone deacetylase (HDAC) in yeast (Taunton *et al.*, 1996) and histone acetyltrans-

ferase (HAT) from *Tetrahymena thermophilia* (Brownell *et al.*, 1996). Since then numerous classes of histone modifying enzymes with different specificities have been described including kinases, lysine and arginine specific methyltransferases, arginine deiminases, ubiquitinases, deubiquitinases and lysine- and arginine specific demethylases (HDMs) (Marmorstein and Trievel, 2009).

Different modifications have different impact on chromosomal processes. For example the methylation of histone H3 lysine 4 is associated with transcriptional activation whereas trimethylation of histone H3 lysine 27 leads to gene silencing (Kouzarides, 2007). Moreover, there is a considerable amount of cross-talk between the different modifications. Addition of one modification to a residue can chemically block additional modification of that amino acid. For example, lysine methylation blocks subsequent acetylation and vice versa. Modifications on different amino acids can be mutually exclusive, such as the dimethylation of arginine 2 and trimethylation of lysine 4 on histone H3 (Guccione *et al.*, 2007; Kirmizis *et al.*, 2007). Some modifications, on the opposite, occur preferentially together, such as phosphorylation of histone H3 serine 10 and acetylation at lysine 14 (Cheung *et al.*, 2000; Lo *et al.*, 2000). Crosstalk can also occur *in-trans*, where one histone and its modifications affect the modification of another histone. For example H2B lysine 123 ubiquitination is required to trigger histone H3 lysine 4 and lysine 36 methylation in *Saccharomyces cerevisiae* (Weake and Workman, 2008).

The combinatorial nature of histone amino-terminal modifications has been proposed to form a "histone code" that together with DNA methylation constitutes an epigenetic marking system, which can be inheritable from one cell generation to another and has an impact on most, if not all, chromatin-templated processes (Jenuwein and Allis, 2001).

2.2.3. Histone binding domains

The biological outcomes of the different histone modifications or combinations of modifications largely depend on the proteins that can read the „histone code" and thereafter facilitate meaningful downstream events. Specific evolutionally conserved protein domains have been shown to mediate the recognition of different histone tail modifications (Table 1). The recognition depends on modification state, position within a histone sequence and modification status of other nearby amino acid residues (Taverna *et al.*, 2007).

Bromodomains are the sole protein modules that have been shown to bind acetylated lysines. They are found in many chromatin associated proteins, especially in histone acetyltransferases such as CBP/p300 and components of certain chromatin remodeling complexes (Mujtaba *et al.*, 2007). Bromodomains in different proteins target acetylated lysines in variety of sequence context. For example, the bromodomain of Gcn5p interacts with histone H4 acetylated at lysine 16 (Owen *et al.*, 2000) while double bromodomains of TAF1 are capable of recognizing multiple acetylated lysines on histone H4 (Jacobson *et al.*, 2000).

Many serine and threonine histone residues can be modified by phosphorylation. Currently there are two examples of phosphohistone binders. Phosphorylated serine 10 and serine 28 of histone H3 are recognized by 14-3-3 proteins (Macdonald *et al.*, 2005) and the phosphorylated serine 139 of histone H2A variant γH2AX is recognized by tandem BRCT domains of MDC1 protein (Lee *et al.*, 2005; Stucki *et al.*, 2005).

Table 1. Protein domains shown to recognize histone modifications. Modified from (Taverna *et al.*, 2007)

	Protein domain	**Histone modifications**
	Bromodomain	many histone Kac
Royal family	Chromodomain	H3K9me2/3, H3K27me2/3,
	Double chromodomain	H3K4me1/2/3
	Chromo barrel	H3K36me2/3
	Tudor	Rme2s
	Double/Tandem tudor	H3K4me3, H4K20me3, H4K20me1/2, Kme2
	MBT repeats	H4K20me1/2, H1K26me1/2, H3K4me1, H3K9me1/2
	WD40 repeat	H3R2/K4me2
	14-3-3	H3S10ph, H3S28ph
	BRCT domain	γH2AXS139ph
	PHD-like ADD domain	H3K4me0
	PHD finger	H3K4me0, H3K4me3, H3K9me3, H3K9ac, H3K36me3

Methylated lysines are recognized by variety of different protein domains including WD40 repeats and Royal superfamily of chromodomains, double chromodomains, tudor, double/tandem tudor, MBT and chromo barrel protein domains. Some of these, like chromodomains, recognize higher lysine methylation states (di- and trimethylation). The chromodomain of HP1 protein binds to histone H3 trimethylated at lysine 9, whereas double chromodomains of CHD1 and double tudor domains of JMJD2A target histone H3 trimethylated lysine 4. Some other domains have preference for lower lysine methylation states. Tandem tudor domains of 53BP1 bind to mono- or dimethylated lysine 20 of histone H4 (Botuyan *et al.*, 2006) and MBT repeats of L3MBTL1 protein are capable of binding mono- and dimethyllysine in a partly sequence independent manner (Li *et al.*, 2007; Trojer *et al.*, 2007). The WD40 repeat protein WDR5 preferentially binds to histone H3 dimethylated at lysine 4 (Couture *et al.*, 2006; Ruthenburg *et al.*, 2006; Schuetz *et al.*, 2006).

2.2.4. PHD zinc fingers

The most recent addition to the list of known methylated histone binding modules is the plant homeodomain zinc finger. The PHD fingers are about 60 amino acid domains characterized by a C4HC3 signature (four cysteines, one histidine, three cysteines) that binds two zinc ions in a cross-brace coordination scheme (Aasland *et al.*, 1995). There are about 150 PHD fingers containing proteins in the human genome (Bienz, 2006).

Due to the resemblance to the RING domain, which typically has C3HC4 signature that binds two zinc ions, it was proposed that PHD fingers might also function as E3 ubiquitin ligases. The literature has been controversial about this issue (Capili *et al.*, 2001; Scheel and Hofmann, 2003; Uchida *et al.*, 2004; Bottomley *et al.*, 2005). Recently, the plant homeodomain fingers of *fission yeast* Msc1 were shown to exhibit E3 ubiquitin ligase activity (Dul and Walworth, 2007) and the PHD domain of the KAP1 corepressor was shown to function as an intramolecular E3 ligase for sumoylation of the adjacent bromodomain (Ivanov *et al.*, 2007). In another recent study it was found that the PHD domain of plant PIAS protein AtSIZ1 mediates sumoylation of AtSCE1 and GTE3 (Garcia-Dominguez *et al.*, 2008). These examples show that some PHD fingers can indeed function as E3 ubiquitin ligases.

The vast majority of PHD fingers are found in nuclear proteins and it was proposed soon after their discovery that they likely have chromatin related function (Aasland *et al.*, 1995). Indeed, the isolated PHD finger of p300 was shown to bind acetylated nucleosomes in electromobility shift assays. In a more stringent nucleosome retention assay, this interaction, however, required also the adjacent bromodomain (Ragvin *et al.*, 2004). Another study showed that the two PHD fingers of ACF1, a subunit of an ATP dependent chromatin remodeling complex interact with the central domains of all four core histones and this results in increased nucleosome sliding efficiency (Eberharter *et al.*, 2004). Similar result was obtained with RAG2 protein, which C-terminus containing PHD finger was found to interact with all core histones and most stably with histone H3 (West *et al.*, 2005).

Four papers published in *Nature* thereafter showed that the PHD fingers of ING2 (inhibitor of growth-2) protein and BPTF (Bromodomain and PHD finger-containing transcription factor), bind directly to histone H3 tails trimethylated at lysine 4 (Li *et al.*, 2006; Shi *et al.*, 2006; Wysocka *et al.*, 2006). ING2 is a native member of the repressive Sin3-histone deacetylase complex. In response to DNA damage, recognition of H3K4me3 by the ING2 PHD domain stabilizes the mSin3a-HDAC1 complex at the promoters of proliferation genes leading to gene repression (Shi *et al.*, 2006). BPTF is the largest subunit of the nucleosomal remodeling factor (NURF) ATP dependent chromatin remodeling complex and the interaction of BPTF with histone H3K4me3 is required for the transcription of key developmental genes in *Xenopus laevis* (Wysocka *et al.*, 2006). Thus the recognition of the same histone modification by PHD finger of different proteins may lead to opposite results. These studies were followed by an array of publications describing proteins which PHD fingers recognize histone H3 with different modifications, most commonly having preference towards some trimethylated lysine (Table 2).

Table 2. Proteins containing PHD fingers that bind to histones

Protein name	Histone modification of preferential binding	References
BPTF	H3K4me3	(Li *et al.*, 2006) (Wysocka *et al.*, 2006)
ING2	H3K4me3	(Pena *et al.*, 2006) (Shi *et al.*, 2006) (Pena *et al.*, 2009)
Yng1p	H3K4me3	(Martin *et al.*, 2006) (Taverna *et al.*, 2006)
ING4	H3K4me3	(Palacios *et al.*, 2006) (Palacios *et al.*, 2008) (Hung *et al.*, 2009)
Yng2 Pho23 Bye1 Cti6 Jhd1 Spp1 Set3 Ecm5 Nto1-1	H3K4me3 H3K36me3	(Shi *et al.*, 2007)
TAF3	H3K4me3	(Vermeulen *et al.*, 2007) (van Ingen *et al.*, 2008)
RAG2	H3K4me3	(Liu *et al.*, 2007) (Matthews *et al.*, 2007) (Ramon-Maiques *et al.*, 2007)
BHC80	Unmodified H3	(Lan *et al.*, 2007)
SMCX	H3K9me3	(Iwase *et al.*, 2007)
SpRAG2L	H3K4me2	(Wilson *et al.*, 2008)
ING1	H3K4me3	(Pena *et al.*, 2008)
ING5	H3K4me3	(Champagne *et al.*, 2008)
DPF3	acH3, acH4, H3K4me2	(Lange *et al.*, 2008)
AtING AL	H3K4me3	(Lee *et al.*, 2009)
ORC1	H3K4me3	(de la Paz Sanchez and Gutierrez, 2009)
JADE1	H3K36me3	(Saksouk *et al.*, 2009)
CHD4	H3K9ac, H3K9me3	(Musselman *et al.*, 2009)

The PHD finger binding to methylated histones occurs in a "surface-groove" mode of recognition. The N-terminal tail of histone H3 docks onto the surface of the PHD finger forming a third antiparallel β-sheet. The specificity towards methylated lysines is determined by a binding pocket composed of aromatic amino acid residues (Li *et al.*, 2006; Pena *et al.*, 2006).

Some PHD fingers cooperate with other domains to specifically interact with certain histone modifications. Specific binding of ICBP90 to histone H3 methyl K9 is determined by PHD (plant homeodomain) finger that defines the binding specificity and a SRA (SET- and RING-associated) domain that promotes binding activity (Karagianni *et al.*, 2008). Combined action of PHD and chromo domains of the Rco1 subunit of the Rpd3S HDAC complex is needed for methyl H3K36 recognition (Li *et al.*, 2007). Pygo PHD fingers associate with their cognate HD1 domains from BCL9/Legless to bind specifically to the histone H3 tail methylated at lysine 4 (Fiedler *et al.*, 2008).

In addition to histone binding and E3 ubiquitin ligase activity, PHD fingers can have other functions. For example, the PHD finger of ING2, which binds to histone H3 trimetylated lysine 4 has also been shown to bind to phosphoinositides (Gozani *et al.*, 2003). Thus, one PHD finger may bind to different ligands.

The importance of PHD fingers is underlined by the fact that mutations or translocations that target the PHD fingers of many genes such as recombination activating gene 2 (*RAG2*), Inhibitor of Growth (*ING*), nuclear receptor-binding SET domain-containing 1 (*NSD1*) and Alpha Thalassaemia and Mental Retardation Syndrome, X-linked (*ATRX*) have been associated with a many human diseases including immunological disorders, cancers, and neurological diseases (Baker *et al.*, 2008).

3. AIMS OF THE STUDY

The general aim of the current thesis was to study the molecular mechanisms behind AIRE mediated promiscuous gene expression of tissue-specific antigens in the thymus concentrating on the function of the first PHD finger domain of AIRE.

More specific aims were:

1. To test AIRE PHD fingers for interaction with histones.
2. To characterize AIRE binding to histone H3 N-terminal tails.
3. To test if AIRE histone binding is needed for AIRE mediated transcriptional activation.
4. To characterize AIRE target genes and posttranslational histone modification status of their promoters in HEK293 cell-line and in mouse mTECs.

4. MATERIAL AND METHODS

4.1. Plasmid construction and protein expression

The truncated AIRE fragments aa 178-482, aa 290-482, PHD1 (aa 290-349) and PHD2 (aa 425-482) were PCR-amplified and cloned into EcoRI/XhoI sites of pGex-1ZT-SH3 vector (a gift from Dr. K. Saksela, University of Tampere, Finland). AIRE-PHD1 mutant constructs D297A, D312A, C302P and C311Y (all aa 290-349) were generated by altering the corresponding amino acids by PCR-based site-directed mutagenesis and cloned into EcoRI/XhoI sites of pGex-1ZT-SH3 vector. Full-length AIRE mutant constructs were generated by altering the corresponding amino acids by PCR-based site-directed mutagenesis and cloned into EcoRI/HindIII sites of pcDNA3.1-myc-his vector (D297A, D312A) or into EcoRI/XhoI sites of pGex-1ZT-SH3 vector (C302P, C311Y). All constructs were verified by DNA sequencing. All the previously constructed plasmids used in this thesis are listed in the Table 3 together with a reference to original paper describing the construction of the plasmid or the company from where the plasmid was purchased.

Table 3. Previously constructed plasmids used in the study with corresponding reference

Plasmid	**Reference**
pHCE-H2B	(Tanaka *et al.*, 2004)
pHCE-H3	(Tanaka *et al.*, 2004)
Flag-ING2	(Shi *et al.*, 2006)
GST-AIRE (aa 1-138)	(Pitkanen *et al.*, 2001)
GST-AIRE (aa 1-293)	(Pitkanen *et al.*, 2001)
GST-AIRE (aa 1-348)	(Pitkanen *et al.*, 2001)
GST-AIRE-PHD1 (aa 293-354)	(Bottomley *et al.*, 2005)
GST-AIRE-PHD2 (aa 426-485)	(Bottomley *et al.*, 2005)
pcAIRE	(Heino *et al.*, 1999)
pVAX-PRMT6	(Guccione *et al.*, 2007)
pd2EYFP-N1	Clontech, Mountain View, CA, USA
pcDNA3.1B-myc/his	Invitrogen, Carlsbad, CA, USA

4.2. Protein expression

Recombinant proteins from each construct were expressed in BL21 *Escherichia coli* strain by addition of 0.8 mM IPTG (and 0.2 mM $ZnCl_2$ to PHD finger-containing constructs) and purified using glutathione Sepharose 4B (Amersham Biosciences, Piscataway, NJ, USA) as described earlier (Pitkanen et al., 2000). If necessary, the bound proteins were eluted from the sepharose beads according to the manufacturer's instructions. The preparation of histone H3 and histone

H2B recombinant proteins was carried out as described previously (Tanaka et al., 2004). TALON Metal affinity resin (BD Biosciences, San Jose, CA, USA) was used for protein purification according to manufacturer's instructions.

4.3. Antibodies and peptides

Antibodies used in Western Blot, ChIP and Q^2ChIP together with the source reference are listed in the Table 4.

Table 4. Antibodies used in the study and a corresponding reference

Antibody	**Reference**
αH3-1	(a gift from Dr. A. Kristjuhan, University of Tartu, Tartu, Estonia)
αH3-2 (ab1791)	Abcam, Cambridge, UK
αAcH3 (06-599)	Upstate Biotechnology, Billerica, MA, USA
αH3K4me1 (07-436)	Upstate Biotechnology, Billerica, MA, USA
αH3K9me3 (07-442)	Upstate Biotechnology, Billerica, MA, USA
αH3K4me3 (ab8580)	Abcam, Cambridge, UK
αH3K27me3 (07-449)	Upstate Biotechnology, Billerica, MA, USA
αH2B (ab1790)	Abcam, Cambridge, UK
α-rabbit polyclonal IgG (sc-2027)	Santa Cruz Biotechnology, Santa Cruz Biotechnology, Santa Cruz, CA, USA
AIRE-1 (sc-17985)	Santa Cruz Biotechnology, Santa Cruz Biotechnology, Santa Cruz, CA, USA
AIRE 6.1	(Pitkanen *et al.*, 2001)
αGST	(a gift from Dr. V. Ovod, FIT Biotech, Finland)
αMyc (9E10)	Sigma-Aldrich, St. Louis, MO, USA
αFlag (F1804)	Sigma-Aldrich, St. Louis, MO, USA
αH3R2me2 (07-585)	Upstate Biotechnology, Billerica, MA, USA
αH3R2me1 (ab15584)	Abcam, Cambridge, UK
αPol II (sc-56767)	Santa Cruz Biotechnology, Santa Cruz Biotechnology, Santa Cruz, CA,USA

Biotinylated peptides used in the study were (ARTKQTARKSTGGKAPR KQLA-GGK-Biotin with corresponding modifications) H3, H3K4me1, H3K4me3, H3K9me3, H3S10P (Upstate Biotechnology, Billerica, MA, USA), H3R2me2s, H3R2me2a (Biopeptide Co Inc., San Diego, CA, USA).

4.4. Preparation of native mononucleosomes

The preparation of ^{3}H labeled native mononucleosomes was performed from THP-1 cells as described previously (Ragvin et al., 2004).

4.5. *In vitro* binding assays

Sepharose bound proteins (2-10 µg) were incubated with 10 µg of calf thymus total histones (Sigma-Aldrich, St. Louis, MO, USA) in binding buffer I (50 mM Tris-HCl pH 7.5, 1 M NaCl, 1 % NP-40, 0.5 mM EDTA, 1 mM phenylmethyl sulphonyl fluoride (PMSF) plus 1µg/ml leupeptin and aprotinin) at 4 °C for 4 h, followed by eight washes with binding buffer I. The bound proteins were visualized by separation on 15 % SDS-PAGE followed by Coomassie staining or Western blot (αH3-1, αH2B, αH3K4me1, αH3K4me3, αH3K9me3). To disrupt the PHD finger structure, 20 µg of sepharose bound AIRE-PHD1 protein was incubated overnight at 4 °C with rotation in binding buffer I containing 50 mM EDTA. The sepharose bound AIRE-PHD1 was then washed with binding buffer I and used in histone binding assays as described above. The histone binding experiments were repeated with acid extracted histones from several cell-lines (HEK293, THP-1) with similar results (data not shown).

For recombinant histone binding, the sepharose bound proteins were incubated with 10 µg purified recombinant histones (H3, H2B) in 0,5 M NaCl containing binding buffer I at 4 °C for 4 h. After extensive washing, bound proteins were analyzed by SDS-PAGE and Coomassie staining.

Mononucleosome binding was carried out overnight at 4 °C in binding buffer II (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% NP-40, 1 µM DTT, 10 µM $ZnCl_2$ and 1 mM PMSF plus 1 µg/ml leupeptin and aprotinin). After extensive washing, bound proteins were analyzed by SDS-PAGE followed by Western blotting with αH3-1 and αH2B antibodies. Bound ^{3}H labeled DNA was analyzed with liquid scintillation counting.

For histone peptide binding, 1 µg of biotinylated histone peptides was bound to the Streptavidin beads (Thermo Scientific, Rockford, IL, USA) according to manufacturer's instructions. Peptides bound to streptavidin beads and 10 µg of eluted protein were incubated in binding buffer III (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 0.1% NP-40, 1 mM PMSF plus 1µg/ml leupeptin and aprotinin) overnight at 4 °C with rotation. After extensive washing, bound proteins were analyzed by SDS-PAGE and Western blotting with αGST antibody.

4.6. Cell-lines, transfection and establishment of stable cell-lines

HEK293 cells were grown in Dulbecco's modified Eagle's medium (DMEM) (PAA Laboratories GmbH, Pasching, Austria) supplemented with 10% fetal calf serum, and 100 U/ml penicillin/streptomycin. THP-1 cells were grown in RPMI 1640 medium (PAA Laboratories GmbH, Pasching, Austria) supplemented with 2 mM L-glutamine, 10% fetal calf serum and 100 U/ml penicillin/streptomycin.

To establish HEK-AIRE (AIRE1, AIRE2), HEK-AIRE-D297A, HEK-AIRE-D312A (AIRE-D312A) and HEK-control (NC) cell-lines, approximately $2x10^6$ HEK293 cells were transfected either with pcAIRE, pcAIRE containing D297A or D312A mutations or pd2EYFP-N1 using ExGen500 *in vitro* reagent (Fermentas, Vilnius, Lithuania) according to the manufacturer's protocol. Selection with media containing 800 µg/ml of G418 (Sigma-Aldrich, St. Louis, MO, USA) was started 48 hours later. After 14 days of culture in selection media, G418-resistant clones were isolated. The YFP signals from the HEK-control cells were determined with fluorescence microscopy, the HEK-AIRE and the mutant clones were analyzed by Western blotting and immunofluorescence. The representative clones were further expanded.

For transient transfections, HEK293 cells were transfected either with pcAIRE myc-tagged pVAX-PRMT6 (a gift from Dr. B. Amati, DNAX Research Institute, CA, USA), pd2EYFP-N1 or pcDNA3.1B-myc/his using ExGen500 *in vitro* reagent (Fermentas, Vilnius, Lithuania) according to the manufacturer's protocol. pd2EYFP-N1 was added to compensate the total DNA amount if AIRE and PRMT6 were separately transfected, equal amount of pcAIRE and pVAX-PRMT6 were used in each transfection. To detect transfected protein levels, the following antibodies were used in Western analysis: αH3-1, AIRE 6.1, αH3R2me2 and αH3R2me1.

4.7. Expression analyses

RNA purification and expression analysis was essentially performed as previously described (Pitkanen *et al.*, 2005). Briefly, cytoplasmic RNA was purified using Trizol (Invitrogen, Carlsbad, CA, USA) and contaminating traces of genomic DNA were removed using TURBO DNase (Ambion, Austin, TX, USA) according to the manufacturer's protocols. Quantitative RT-PCR was performed with the ABI Prism 7900HT instrument (Applied Biosystems, Foster City, CA, USA) using qPCR SYBR Green Core Kit (Eurogentec, Liege, Belgium). The amplification program included an initial denaturation step at 95 °C for 10 min, followed by denaturation at 94 °C for 15 s, annealing at 58 °C for 30 s, and extension at 60 °C for 1 min, repeated 50 cycles. SYBR Green fluorescence was measured after each extension step; the specificity of amplification was subjected to melting curve analysis. The relative gene expression levels were calculated

using the comparative Ct method (according to Applied Biosystems), where the relative expression is calculated as $2^{-\Delta\Delta Ct}$, and Ct represents the threshold cycle. Every sample was run at least three times in three parallel reactions. Quantitative RT-PCR primer sequences used in the study are listed in section 4.13.

4.8. Chromatin immunoprecipitation (ChIP) and Q^{2}ChIP

Protein ChIP was performed as described (Ricke and Bielinsky, 2005). DNA ChIP was performed essentially according to Upstate Chromatin Immunoprecipitation Assay protocol (Upstate Biotechnology, Billerica, MA, USA) using antibodies (αH3-2, αAcH3, αH3K4me3, αH3K27me3, αPol II, α-rabbit polyclonal IgG). Q^{2}ChIP (quick and quantitative ChIP) using antibodies (αH3-2, αH3K4me3, αH3K27me3, α-rabbit polyclonal IgG) was performed essentially as described in (Dahl and Collas, 2007) except, instead of Protein A-coated paramagnetic beads, protein G-sepharose beads (GE Healthcare, WI, USA) preblocked with 100µg/ml BSA and 500µg/ml salmon sperm DNA were used. ChIP and Q^{2}ChIP samples were analyzed in triplicate by quantitative PCR using qPCR SYBR Green Core Kit (Eurogentec, Liege, Belgium) on ABI Prism 7900HT (Applied Biosystems, Foster City, CA, USA). The results were analyzed using comparative Ct method (according to Applied Biosystems). Datasets of each primer pairs were normalized to ChIP input values and the relative amount of immunoprecipitated material compared to IgG (Study I) or H3 (Study III) was calculated. Primer sequences used for ChIP and Q^{2}ChIP analysis with qPCR are listed in section 4.13.

4.9. Gene expression arrays

The RNA from AIRE1, AIRE2, AIRE-D312A and NC was extracted with Trizol reagent (Invitrogen, Carlsbad, CA, USA), treated 2x30 min with Turbo DNAse (Ambion, Austin, TX, USA) and purified with QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). Labeled cRNA was prepared using Illumina TotalPrep RNA Amplification Kit (Ambion, Austin, TX, USA) and hybridized to Illumina HumanWG-6_V2_0_R2 BeadChip expression arrays (lllumina, San Diego, CA, USA). The quality of the RNA and labeled cRNA was controlled using 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). With all cell-lines, three independent experiments were performed. The raw data was analyzed with BeadStudio Gene Expression Module v3.3.7 (Illumina, San Diego, CA, USA) using Illumina's custom rank invariant method. Genes with differential expression values >13 (corresponding to p-value <0.05) were considered as differentially expressed. Multi Experiment Viewer version 4.0 was used for unsupervised hierarchical clustering (using Manhattan distance and complete linkage) and for the visualization of the results. The microarray data can be accessed through Gene Expression Repository under GEO accession number GSE16877.

4.10. Characterization of AIRE target genes

REEF program was used (http://telethon.bio.unipd.it/bioinfo/reef/index.html) to find significantly clustered AIRE target genes (Coppe *et al.*, 2006). Analysis was performed with 200 kb window size and 10 kb shift at false discovery rate of 0.05. Chi-square test was used to determine the significance of the observed to expected deviation.

TiGER database (http://bioinfo.wilmer.jhu.edu/tiger/) (Liu *et al.*, 2008) was used to determine the proportion of tissue-specific genes among AIRE upregulated, downregulated and all Refseq genes on Illumina BeadChip expression arrays. Chi-square test was used to determine the significance of the observed to expected deviation.

The distribution comparison of AIRE regulated genes with that of the whole genome according to size, exon number and expression levels was based on all known mapped genes (RefSeq transcripts) of the human genome (NCBI RefSeq 19,360; build 36.2).

4.11. ChIP-on-Chip

The custom made Nimblegen tiling arrays contained non-repetitive sequences throughout the chosen regions represented by 386748 probes with median spacing 72 bp. Array contained several control and 51 AIRE target gene regions that were chosen based on the results with Illumina Beadchip expression arrays. The chosen genes had to be significantly upregulated by AIRE (p-value <0.05) and have fold change greater than 5. The arrays covered the whole gene and 200 kb up- and downstream of each gene. In addition, a 5.2 mb region was chosen from chromosome 1, which contained the epidermal differentiation cluster. Chromatin immunoprecipitation was performed with αH3K4me3, αAcH3, αH3K27me3, αH3-2, αPol II and αAIRE-1 from AIRE1 and NC cell-lines as described in section 4.5. The immunoprecipitated and input material were amplified using WGA2 and WGA1 GenomePlex Whole Genome Amplification Kits (Sigma-Aldrich, St. Louis, MO, USA). Sample labeling and hybridization to custom made tiling arrays was performed by Nimblegen (NimbleGen Systems of Iceland, Reykjavík, Iceland). For each probe, a scaled log2-ratio was calculated. Scaling was performed by subtracting the bi-weight mean for the log2-ratio values for all probes on the array from each log2-ratio value. Peaks were detected using NimbleScan by searching for 4 or more probes whose signals are above specified cut-off values, ranging from 15% to 90%, using a 500 bp sliding window. The cut-off values are percentage of hypothetical maximum, which is mean + 6 standard deviation. Each peak is assigned a false discovery rate (FDR) based on 20 times randomization of ratio data. For histone modifications FDR cut-off at 0.01 level was used.

For the correlation of expression levels with the presence of histone posttranslational modifications on promoters, genes were first sorted according to the signal intensities on Illumina Beadchip expression arrays. A sliding window with the size of 50 probes and step of 1 probe was then used to

calculate the percent of promoters having corresponding posttranslational histone modification (Supplementary Figure 7 in Study III). For the calculation of average levels of histone posttranslational modifications on the promoters of AIRE target genes, the data for each antibody was normalized against H3 and the average signal intensity was calculated for each gene in a region of +/- 2 kb of transcription start sites (Supplementary Figure 8 in Study III). For the generation of graphs displaying regions with average levels of histone posttranslational modifications, average signal intensities for each probe were sorted according to their position along the corresponding chromosome. A sliding window analysis, consisting of 50 probes and step of 1 probe, was then used to calculate average signal intensities (Figure 4; Supplementary Figure 9, 10 in Study III). Original ChIP-on-chip data can be accessed from the GEO database under the accession number GSE17216.

4.12. Mice

C57BL/6J mice used in the study were maintained at the mouse facility of the Institute of Molecular and Cell Biology, University of Tartu.

4.13. Cell sorting

Thymi from 6-8 week old mice were dissected and collected into RPMI. Small cuts were made into the capsules of thymi and the thymocytes were released by repetitive pipeting. The remaining thymic fragments were incubated in 0.5 mg/ml dispase/collagenase (Roche, Basel, Switzerland) and 5 µg/ml DNase I (AppliChem, Darmstadt, Germany) in PBS at 37 °C for 20 min, with gentle agitation. The released cells were collected to separate fractions and fresh enzyme solution was added 4 times. Each cell fraction was counted and were pooled starting from the end fractions to gain 200 x 10^6 cells. A negative depletion was performed to enrich for CD45- cells using CD45 microbeads (Miltenyi Biotec, Hilden, Germany) and the AutoMACS system (Miltenyi Biotec, Hilden, Germany), according to the manufacturer's instructions. The negative fraction was stained with anti-G8.8-FITC (anti-EpCAM, generated from a G8.8 hybridoma cell-line), anti-Ly51-PE, anti-CD45-PerCP-Cy5.5 and anti-CD80 biotin (all from BD Biosciences, San Jose, CA, USA) followed by second-stage staining with Streptavidin-PE-Cy7 (Serotec Ltd, Oxford, UK). Cell sorting and analysis was performed on a FACSAria (BD Biosciences, San Jose, CA, USA) instrument to get the fractions of mTECs ($CD45^-$, G8.8high, Ly51low) and cTECs ($CD45^-$, G8.8low, Ly51high). According to the CD80 expression, the mTEC fraction was further divided into the CD80high and CD80low mTECs.

Neutrophils were extracted from spleen by mechanical disruption of the tissue followed by positive selection with Ly6G microbeads (Miltenyi Biotec, Hilden, Germany) and the AutoMACS system (Miltenyi Biotec, Hilden, Germany). To enrich the pancreatic tissue for Langerhans islets, the tissue was

first cut into small pieces, which was then incubated with 0.5 mg/ml dispase/ collagenase (Roche, Basel, Switzerland) and 5 µg/ml DNase I (AppliChem, Darmstadt, Germany) in PBS at 37 °C for 20 min. Discontinuous Optiprep™ (Axis Shield PoC AS, Oslo, Norway) density gradient (1.108; 1.096; 1.069; 1.037) centrifugation was carried out for 10 min at 1800 rpm and cells from two upper layers were used for further experiments.

4.14. Primer sequences

The sequences of all primers (TAG Copenhagen, Copenhagen, Denmark) used in the study together with additional information are listed in Table 5.

Table 5. Primers used in the study together with corresponding sequence, species, gene and application used in

Oligo name	Sequence	Species	Gene	Application
SA10ch2 for	ATGCAGAGCGCTCCAAGCCA	human	*S100A10*	ChIP
SA10ch2 rev	AAGCTCCCGCAGAGACGGAACT	human	*S100A10*	ChIP
chINV F5	CCAATCCTTTAGATATGGTACACAG	human	*INV*	ChIP
chINV R5	TCCCCAGGTCTCTGGTTCTT	human	*INV*	ChIP
chSA82 for	CTGGGCTGCTGGCATCCACT	human	*S100A8*	ChIP
chSA82 rev	GGCCTGACCACCAATGCAGGG	human	*S100A8*	ChIP
hGAPDH_prom_F	TACTAGCGGTTTTACGGGCG	human	*GAPDH*	ChIP
hGAPDH_prom_R	ACAGGAGGAGCAGAGAGCGA	human	*GAPDH*	ChIP
chINS2 rev	CTGTGGGAAGATCTCCTTGG	human	*INS*	ChIP
chINS2 for	CTGTGAGCAGGGACAGGTCT	human	*INS*	ChIP
h_PSMD4_Ch3'_F	GATAGTCCCGGGTTACCAC	human	*PSMD4*	ChIP
h_PSMD4_Ch3'_R	TGTAGCTAAAGACAGACCCG	human	*PSMD4*	ChIP
h_PSMD4_Ch5'_F	CCTTGGTGAGCAAATGTTGA	human	*PSMD4*	ChIP
h_PSMD4_Ch5'_R	TCTAGGACTATCCAGCCTGA	human	*PSMD4*	ChIP
h_CDH12_Ch3'_F	TGTTGTGTGAAACTTCCCTG	human	*CDH12*	ChIP
h_CDH12_Ch3'_R	GTATTGCCACAATGATGCCT	human	*CDH12*	ChIP
h_CDH12_Ch5'_F	TGCTTTGAATGAATCCGAGG	human	*CDH12*	ChIP
h_CDH12_Ch5'_R	GTGCTATCACTTTCTCAGCG	human	*CDH12*	ChIP
h_CLDN1_Ch3'_F	GGCCAAGGGAATACCTGATA	human	*CLDN1*	ChIP
h_CLDN1_Ch3'_R	TGGGAAATTGAGATTGGGTTG	human	*CLDN1*	ChIP
h_CLDN1_Ch5'_F	CTCTCTGCTCTCTTCTGTCA	human	*CLDN1*	ChIP
h_CLDN1_Ch5'_R	GGTCTGTTTCCCAATCTGTAG	human	*CLDN1*	ChIP
h_EFHC2_Ch3'_F	CTGGAATTACAGACGTGAACTG	human	*EFHC2*	ChIP
h_EFHC2_Ch3'_R	AAGAAACAGTAACCAATCAGCC	human	*EFHC2*	ChIP
h_EFHC2_Ch5'_F	TCCTGAGGAGTCTTCTCTCT	human	*EFHC2*	ChIP
h_EFHC2_Ch5'_R	ACAAGAAAGGCTTTGAGTTCT	human	*EFHC2*	ChIP
h_GHR_Ch3'_F	TGCAATTCTGGCCTTAAAAGT	human	*GHR*	ChIP

Oligo name	Sequence	Species	Gene	Application
h_GHR_Ch3'_R	AACAGAAGGAAAAGGCAGTT	human	*GHR*	ChIP
h_GHR_Ch5'_F	ATTTTCCTCTAGGAGGAGCC	human	*GHR*	ChIP
h_GHR_Ch5'_R	GAGTTGCTGGGATGAAGTCT	human	*GHR*	ChIP
h_HBG2_Ch3'_F	CTGAAGACAACCATGTGTGA	human	*HBG2*	ChIP
h_HBG2_Ch3'_R	CCATGATGCAGAGCTTTCAA	human	*HBG2*	ChIP
h_HBG2_Ch5'_F	CACCTTCTTGTCTGTAGCTC	human	*HBG2*	ChIP
h_HBG2_Ch5'_R	CCCATCAAAAATCCTGGACC	human	*HBG2*	ChIP
h_LPL_Ch3'_F	CCTCCAAAAATGATGCACCT	human	*LPL*	ChIP
h_LPL_Ch3'_R	AATGGATCAGAGAATGGCAG	human	*LPL*	ChIP
h_LPL_Ch5'_F	GAGTGGAACCCCTTAAGCTA	human	*LPL*	ChIP
h_LPL_Ch5'_R	GTAATGAGGATCAACCCTGC	human	*LPL*	ChIP
h_PAK3_Ch3'_F	AAACTGGATTTTTGCCGTCT	human	*PAK3*	ChIP
h_PAK3_Ch3'_R	GGAGCTCTGGAGTTCCATTA	human	*PAK3*	ChIP
h_PAK3_Ch5'_F	TTTCCCCCGAATAAAGTGTG	human	*PAK3*	ChIP
h_PAK3_Ch5'_R	TTCCTTAATTGCTCCCCTCA	human	*PAK3*	ChIP
h_SUSD4_Ch3'_F	CAGGATACAGCTCCCAAGAT	human	*SUSD4*	ChIP
h_SUSD4_Ch3'_R	TTTCTCCAGGTTCCACAAGA	human	*SUSD4*	ChIP
h_SUSD4_Ch5'_F	CCACAGGAAAGTCTAGAGCC	human	*SUSD4*	ChIP
h_SUSD4_Ch5'_R	CCACATTGTGAGATTCCCAG	human	*SUSD4*	ChIP
h_THBS2_Ch3'_F	TGTCCTATGCTCACTCCCTA	human	*THBS2*	ChIP
h_THBS2_Ch3'_R	CTGGGGTTACAACTCAACAC	human	*THBS2*	ChIP
h_THBS2_Ch5'_F	TTATATGCCGCTTCTCCTGT	human	*THBS2*	ChIP
h_THBS2_Ch5'_R	TACTCGGAGCAAATGGTGAG	human	*THBS2*	ChIP
h_ZBTB7B_Ch3'_F	AGTGTACCCCACTTTTGTCT	human	*ZBTB7B*	ChIP
h_ZBTB7B_Ch3'_R	ATCTGCATGCGTGTGTATCT	human	*ZBTB7B*	ChIP
h_ZBTB7B_Ch5'_F	CACAGCTACGACCTCAAGAA	human	*ZBTB7B*	ChIP
h_ZBTB7B_Ch5'_R	CTCCTTGGCGAAAGCCTTG	human	*ZBTB7B*	ChIP
mGapdh_pr2_F	AGGGCTGCAGTCCGTATTTA	mouse	*Gapdh*	Q^2ChIP
mGapdh_pr2_R	CCCTTGAGCTAGGACTGGAT	mouse	*Gapdh*	Q^2ChIP
mGad2_pr_F	GCAGACCTGGAGCGCCTGTG	mouse	*Gad2*	Q^2ChIP
mGad2_pr_R	ATGGAGGCGCTGGGTCGTGA	mouse	*Gad2*	Q^2ChIP
mS100A8_pr_F	GGGCTGGGTCCCTGGATATGTCT	mouse	*S100a8*	Q^2ChIP
mS100A8_pr_R	GGACTGGACCATGGAGAACAGTAAGGAC	mouse	*S100a8*	Q^2ChIP
mAIRE_pr_F	GCGTTCCCTTCTCCCTTCCTTGG	mouse	*Aire*	Q^2ChIP
mAIRE_pr_R	CAAACGTCGCCTTTCCGCCG	mouse	*Aire*	Q^2ChIP
mIns_pr_F	AGGGCTGGTGGTTACTGGGTCC	mouse	*Ins*	Q^2ChIP
mIns_pr_R	CCTACCCCACCTGGAGCCCTT	mouse	*Ins*	Q^2ChIP
h_PSMD4_expr_F	GAAGGTGAAAGAGACTCAGAC	human	*PSMD4*	qPCR
h_PSMD4_expr_R	GTCATACTGCTTAGGTCAGG	human	*PSMD4*	qPCR
h_BAT2_expr_F	CCAGAGCAAATCTTACCCAG	human	*BAT2*	qPCR
h_BAT2_expr_R	TACTGTAACCTGCTGGAGAG	human	*BAT2*	qPCR
h_CDH12_expr_F	TGACTTCAGAAACAACACAG	human	*CDH12*	qPCR
h_CDH12_expr_R	TGTCTTCTATTACAACAGGGAG	human	*CDH12*	qPCR

Oligo name	Sequence	Species	Gene	Application
h_CLDN1_expr_F	AATTCTATGACCCTATGACCC	human	*CLDN1*	qPCR
h_CLDN1_expr_R	GACAGGAACAGCAAAGTAGG	human	*CLDN1*	qPCR
h_EFHC2_expr_F	AGAGATGTGAAGATGTTTGG	human	*EFHC2*	qPCR
h_EFHC2_expr_R	GGTTATTCCTCCTCTAAGCC	human	*EFHC2*	qPCR
h_GHR_expr_F	ATGAAATAGTGCAACCAGATCC	human	*GHR*	qPCR
h_GHR_expr_R	CCCATCTCACTTGGATATCTG	human	*GHR*	qPCR
h_HCP5_expr_F	GGCAGATTACAATTACAATCAAGG	human	*HCP5*	qPCR
h_HCP5_expr_R	TTGCATCTCAGTCTATTGCC	human	*HCP5*	qPCR
h_MAL2_expr_F	ACCAGTATAACATAAACGTAGCAG	human	*MAL2*	qPCR
h_MAL2_expr_R	CCATCTTCGTAAAGCCAGAC	human	*MAL2*	qPCR
h_OAS3_expr_F	CAACGGGACTTCATCATCTC	human	*OAS3*	qPCR
h_OAS3_expr_R	CTTGGAGATCTTGGTACACTG	human	*OAS3*	qPCR
h_PHF11_expr_F	GACAAAGTTCATTCAATTCCAG	human	*PHF11*	qPCR
h_PHF11_expr_R	TCTTCGAACAATCTACAGTC	human	*PHF11*	qPCR
h_SUSD4_expr_F	CCGGATTTAAACTTGATGGG	human	*SUSD4*	qPCR
h_SUSD4_expr_R	AAAGATGTTCAGGTCTTCCTC	human	*SUSD4*	qPCR
h_TNMD_expr_F	CAAGTGAGGAAGAACTTCCA	human	*TNMD*	qPCR
h_TNMD_expr_R	GGCAGTAAATACAACAATAACC	human	*TNMD*	qPCR
h_TOX3_expr_F	GTATTATCGCAGGCTATACCT	human	*TOX3*	qPCR
h_TOX3_expr_R	CTTCTGGAGAAATAAGGCCA	human	*TOX3*	qPCR
hHBG2_exp_F	CATAAAGCACCTGGATGATCTC	human	*HBG2*	qPCR
hHBG2_exp_R	CAGGAGCTTGAAGTTCTCAG	human	*HBG2*	qPCR
hIFI16_exp_F	CTGTGAGGAAGGAGATAAACTG	human	*IFI16*	qPCR
hIFI16_exp_R	TCTTGATGACCTTGATGTGAC	human	*IFI16*	qPCR
hLPL_exp_F	TGCTTGAGTTGTAGAAAGAACC	human	*LPL*	qPCR
hLPL_exp_R	TTGGTAATGGAAGACTTTGTAGG	human	*LPL*	qPCR
hLY6G6D_exp_F	TGGAACAGATCAAGCTACCT	human	*LY6G6D*	qPCR
hLY6G6D_exp_R	CTCTGTCTCCACTTGATTGC	human	*LY6G6D*	qPCR
hPAK3_exp_F	GGATGAAGGACAGATAGCAG	human	*PAK3*	qPCR
hPAK3_exp_R	CCCAAAGTCAGTCAATTTAACAG	human	*PAK3*	qPCR
hTHBS2_exp_F	GCTACATCAGAGTCTTAGTGC	human	*TBS2*	qPCR
hTHBS2_exp_R	AGAGAAGACAAATAGACCCAG	human	*TBS2*	qPCR
hOAS2f	CACCAGCTCCAATCAGCGAG	human	*OAS2*	qPCR
hOAS2r	TCAGCCATTGCCAGCATATTTTATC	human	*OAS3*	qPCR
hINS K5	GCAGCCTTTGTGAACCAACA	human	*insulin*	qPCR
hINS K3	GTGTGTAGAAGAAGCCTCGTTCC	human	*insulin*	qPCR
hAIRE1F	CTGCCAAGGATGACACTGCCA	human	*AIRE*	qPCR
hAIRE1R	CGAAGGTGTGCTCGCTCAGAAG	human	*AIRE*	qPCR
S100A8-for	CTCAGTATATCAGGAAAAAGGGTGCAGAC	human	*S100A8*	qPCR
S100A8-rev	CACGCCCATCTTTATCACCAGAATGAG	human	*S100A8*	qPCR
S100A10-for	TCCCTGGATTTTTGGAAAATCAAAAAGACC	human	*S100A10*	qPCR
S100A10-rev	CCCGCAATTAGGGAAAAGAAGCTCTGGAA	human	*S100A10*	qPCR
HPRTexon6_F	GACTTTGCTTTCCTTGGTCAGG	human	*HPRT*	qPCR

Oligo name	Sequence	Species	Gene	Application
HPRTexon7_R	AGTCTGGCTTATATCCAACACTTCG	human	*HPRT*	qPCR
hINV FOR	GCCTTACTGTGAGTCTGGTTGACA	human	*INV*	qPCR
hINV REV	GGAGGAACAGTCTTGAGGAGCT	human	*INV*	qPCR
G3PDH5	TCCACCACCCTGTTGCTGTAG	human	*GAPDH*	qPCR
G3PDH3	GACCACAGTCCATGCCATCACT	human	*GAPDH*	qPCR

5. RESULTS AND DISCUSSION

5.1. Possible mechanisms of AIRE mediated gene expression activation

Several studies have shown that the expression of AIRE in a given cell type leads to the upregulation of many genes (Anderson *et al.*, 2002; Gardner *et al.*, 2008; Guerau-de-Arellano *et al.*, 2008). Broadly, such gene expression activation could be achieved either indirectly or directly. In the first case AIRE activates one or few mediators that thereafter upregulate the expression of many other genes. The structure of the AIRE protein, which has many features of transcriptional regulators, experiments which show that AIRE is capable of direct gene expression activation in various cell models and the fact that, no common transcription regulators that are activated by AIRE in different cell types have been found, strongly support the second option that AIRE activates its target genes directly. The direct activation could also be achieved in various ways. It could be mediated either at the transcriptional level or by enhancing some of the aspects of post-translational gene expression regulation, such as mRNA processing or stability. When acting at the transcriptional level, AIRE could act locally on the promoters of its target genes as well as mediate the activation through enhancer elements that could be far away from the gene. Since AIRE target genes have been shown to localize in clusters in the genome, it is also possible that AIRE somehow affects the chromatin structure of certain chromosomal regions making them more accessible to the transcriptional machinery. Most of these possibilities assume that AIRE somehow interacts with chromatin. Many transcription factors interact directly with DNA in a sequence specific manner. Despite of some supporting data (Kumar *et al.*, 2001; Purohit *et al.*, 2005), there is no conclusive evidence that AIRE interacts with DNA in a sequence specific manner (Mathis and Benoist, 2009). Lack of sequence specificity is also supported by the fact that there is little overlap between the genes upregulated by AIRE in various cell types (Guerau-de-Arellano *et al.*, 2008). Besides direct interaction with DNA, it is also possible that AIRE is tethered to chromatin through the interaction with other chromatin bound proteins of which the most abundant are histones.

5.2. AIRE interacts with histone H3 through its first PHD finger

Since the PHD fingers of ING2 and BPTF proteins were shown to directly bind histone H3 N-terminal tails and since AIRE contains two PHD fingers, we tested whether AIRE might also interact with histones. We first expressed full-length GST-AIRE protein in *Escherichia coli*. After the purification by binding to gluthatione sepharose beads, the protein was used in an interaction assay with whole histones from calf thymus. The results of the experiment showed that AIRE is indeed capable of interacting with one of the histones, which was

confirmed to be histone H3 by Western blotting (Figure 5A). Next, we wanted to know which part of the AIRE protein is responsible for the interaction. Using different GST-AIRE fragments in an interaction assay with calf histones, we demonstrated that only fragments containing the two PHD fingers and not N-terminal parts of the protein or GST alone were capable of interacting with histone H3 (Supplementary Figure S1B in Study I). We next asked whether one or both PHD fingers are required for the interaction with histones. Experiments using GST-fusions of the two AIRE-PHD fingers separately showed that only the first PHD finger (AIRE-PHD1) is required and sufficient for the interaction with histone H3 (Figure 5B).

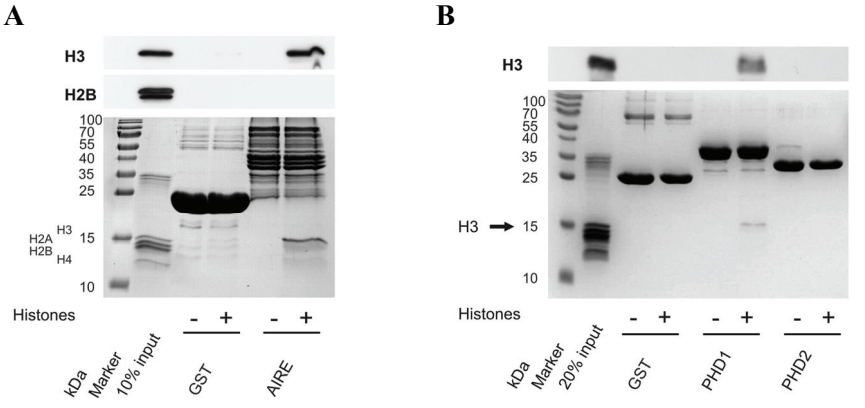

Figure 5. The first PHD finger of AIRE interacts with histone H3. (A) GST and GST-AIRE interaction with whole histones visualized by Coomassie staining (bottom) and detected by Western blot using αH3-1 and αH2B antibodies (top and middle). (B) The same assay as (A) using GST, GST-AIRE-PHD1, GST-AIRE-PHD2 proteins.

Since the first set of experiments used whole histone extracts from calf thymus it is possible that the interaction is mediated by some other proteins present in the extract. To test if AIRE and its first PHD finger are capable of interacting directly with histone H3, we expressed recombinant histidine tagged histones H3 and H2B in *Escherichia coli* and used TALON Metal Affinity Resin to purify the proteins. The binding assays with recombinant histones confirmed that AIRE is able to directly interact with histone H3 and not with histone H2B through its first PHD zinc finger (Figure 1D in Study I; Supplementary Figure S1C in Study I).

It has been established that the binding of two zinc ions is required for the correct folding of PHD fingers (Bottomley *et al.*, 2005). Therefore, the removal of the zinc ions should lead to the destruction of the domain structure and would be expected to abolish its function, in our case, the interaction with histone H3. Accordingly, following the incubation of AIRE-PHD1 with 50 mM EDTA, a known divalent metal cation chelator, the interaction with histone H3 was clearly diminished (Supplementary Figure S1D in Study I). Furthermore, we

used mutated AIRE proteins in which cysteine 302 or cysteine 311 of AIRE-PHD1 was substituted with proline or tyrosine, respectively. These cysteines have been shown to coordinate zinc ions within the AIRE-PHD1 finger and are therefore crucial in maintaining the correct fold of the domain. The binding experiments with calf thymus histones showed that the C302P and C311Y mutations in AIRE-PHD1 abolish the interaction with histone H3 (Supplementary Figure S1E, F in Study I), confirming that the correct fold of the PHD finger is required for the interaction.

This set of experiments collectively showed that, similar to some other proteins with PHD fingers, AIRE is able to interact directly and specifically with histone H3 through its first PHD zinc finger domain.

5.3. AIRE binds to mononucleosomes

In cell nucleus chromatin binding proteins operate in the context of nucleosomes. We therefore studied if AIRE-PHD1 can interact with nucleosomes. Binding experiments with isolated ^{3}H labeled mononucleosomes showed that AIRE-PHD1 also interacts with a small fraction of native mononucleosomes, as assessed by Western blot against histones H3 and H2B (Figure 6A) and by analyzing bound DNA (Figure 6B). This result showed that AIRE is able to interact with nucleosomes and in addition to sequence specific DNA binding, provided an alternative mechanism how AIRE might be tethered to chromatin.

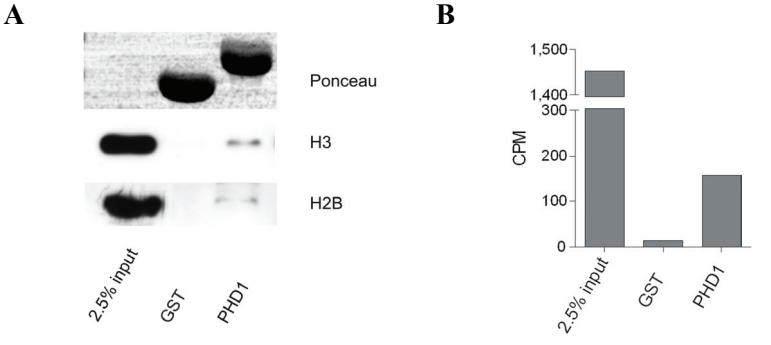

Figure 6. AIRE binds to the nucleosomes. GST-AIRE-PHD1, but not GST alone, interacts with native mononucleosomes detected by (A) Western blot with αH3-1 (middle) and αH2B (bottom) antibody or (B) measuring the GST-AIRE-PHD1 pulled down ^{3}H-labeled DNA by radioactivity (CPM by liquid scintillation counting). Equal input of GST proteins is shown with Ponceau red staining (A, top).

5.4. Histone H3 N- terminal posttranslational modifications hinder AIRE binding

Since PHD fingers were shown to work as read out modules for lysine methylation events in the histone code (Li *et al.*, 2006; Pena *et al.*, 2006), we sought whether AIRE-PHD1 also binds preferentially to certain histone lysine modifications. To explore this possibility, we first analyzed AIRE-PHD1 bound histones with Western blot using antibodies recognizing methylated forms of H3: H3K4me1, H3K4me3 and H3K9me3. Although detectable bands were seen with all antibodies, we repeatedly observed only a very weak signal with αH3K4me3 antibody (Figure 1F in Study I). This result suggested the possibility that methylation of H3K4 hinders AIRE-PHD1 binding to H3, or that AIRE-PHD1 is able to bind H3 tails differentially modified at several positions.

The results obtained with the modification-specific antibodies can be influenced by the presence of multiple modifications on the same histone tails. To determine more precisely whether the AIRE-PHD1 has a preference towards certain lysine methylations, we performed binding experiments with biotinylated N-terminal histone peptides with or without different modifications (unmodified H3, H3R2me2s, H3R2me2as, H3K4me1, H3K4me3, H3K4me9, H3S10P) bound to streptavidin coated beads and purified GST-AIRE-PHD1 protein eluted from the sepharose beads. The Western blot results with αGST antibody confirmed that AIRE-PHD1 is capable of interacting with histone H3 and that the 20 N-terminal amino acids of histone H3 are sufficient for the binding (Figure 7).

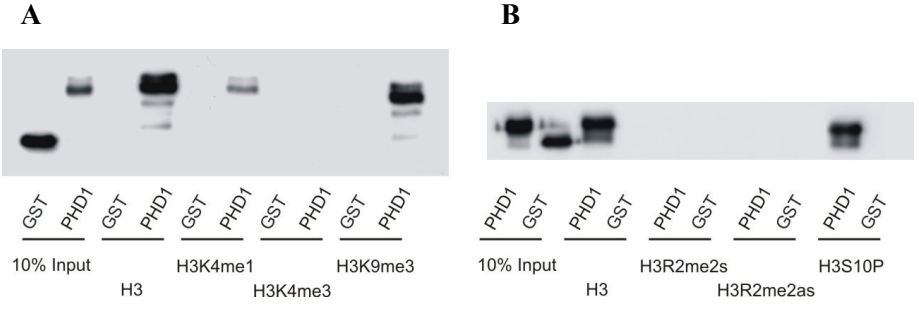

Figure 7. Influence of different histone H3 modifications to interaction with AIRE-PHD1. Interaction between GST and GST-AIRE-PHD1 fusion proteins and N-terminal histone H3 peptides: H3, H3K4me1, H3K4me3, H3K9me3 (A), H3, H3R2me2s, H3R2me2as, H3S10P (B), all detected by αGST antibody.

The intensity of the signals revealed that the binding is influenced by several histone H3 N-terminal posttranslational modifications. While unmodified H3, H3K9me3 (Figure 7A) and H3S10P (Figure 7B) histone peptides bound AIRE-PHD1 with similarly strong affinity, the binding was diminished depending on the degree of K4 methylation. A weak band was detected with H3K4me1 peptide and no band was visible with H3K4me3 peptide (Figure 7A). Similarly to the results with H3K4me3 peptide, no band was seen with H3R2me2s or

H3R2me2as peptides (Figure 7B). In conclusion, these results show that AIRE-PHD1 interacts with the N-terminal tail of histone H3 and several post-translational modifications interfere with the binding.

Importantly, our results were confirmed by our collaborators, Dr. Giovanna Musco and her colleagues from Dulbecco Telethon Institute c/o S. Raffaele Scientific Institute, Milan, who used two-dimensional 1H-15N NMR (nuclear magnetic resonance), tryptophan fluorescence spectroscopy and isothermal titration calorimetry (ITC) to assess the binding of AIRE-PHD1 to N-terminal histone H3 peptides with different modifications. The use of the two latter methods allowed determining the interaction affinity Kd for different mutants and peptides with different modifications (Table 1 in Study I; Table 2 in Study II). The Kd for AIRE-PHD1 and unmodified histone H3 N-terminal peptide interaction was measured to be ~5μM, which is in the same range as measured for other PHD fingers for their preferential histone peptide suggesting that the interaction is physiologically relevant.

5.5. AIRE interacts with histone H3 *in vivo* in HEK293 cells

Next, we used protein chromatin immunoprecipitation (ChIP) assays to study AIRE and histone H3 interaction *in vivo* in HEK293 cells stably transfected with AIRE. As a positive control we studied the histone binding of ING2 protein, which has been shown to preferentially interact with H3K4me3. The results of the protein ChIP showed that AIRE is found in complexes with a small fraction of histone H3 but not with H3K4me3. By contrast, binding of ING2, was detected for both H3 and H3K4me3 (Figure 8). This result shows that AIRE selectively interacts with histones that lack H3K4me3 also *in vivo*.

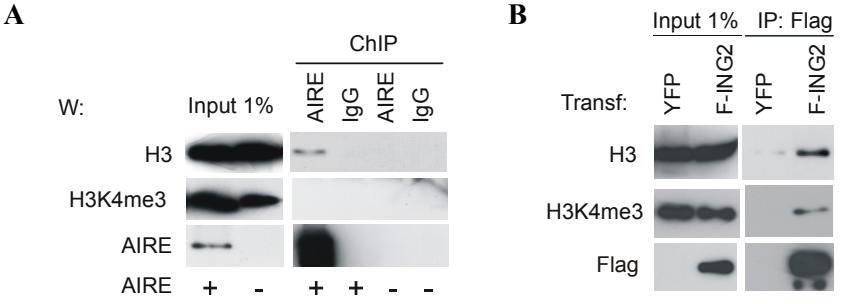

Figure 8. AIRE *in vivo* interaction with histone H3 in HEK-AIRE cell-line. (A) Protein ChIP was carried out with AIRE-1 antibody and IgG from HEK-AIRE (+) or control cells (-). The bound proteins were detected by αH3-1, αH3K4me3 and AIRE 6.1 antibodies. (B) ING2 *in vivo* interaction with histone H3K4me3. Protein ChIP was carried out with anti-Flag agarose beads from HEK293 cells, transfected either with Flag-ING2 (F-ING2) or pd2EYFP-N1 (YFP). The bound proteins were detected by αH3-1, αH3K4me3 and αFlag antibodies.

5.6. The structure of the AIRE-PHD1 in complex with histone H3 peptide

In collaboration with Drs. Csaba Hetenyi and Uko Maran from the Institute of Chemical Physics, we used the published PHD finger structures from the human NURF BPTF PHD finger in complex with H3K4me3 peptide (PDB code 2fuu) and AIRE1-PHD128 (PDB code 1xwh), to model the interaction of AIRE-PHD1 with unmodified histone H3 peptide and performed molecular dynamics calculations for 10 ns (Figure 3 in Study I). The molecular dynamics calculations were in agreement with the experimental data showing that lysine 4 methylation disrupts the interaction (Supplementary Figure S4 in Study I). Later on, our collaborators, Dr Giovanna Musco and her colleagues from Dulbecco Telethon Institute c/o S. Raffaele Scientific Institute, Milan used heteronuclear and isotope-filtered NMR experiments to determine the three-dimensional solution structure of AIRE-PHD1 in complex with unmodified histone H3 peptide (PDB code 2ke1) (Figure 9A, B). Superimposing the two structures showed that predicted model was in a good agreement with the measured structure (data not shown).

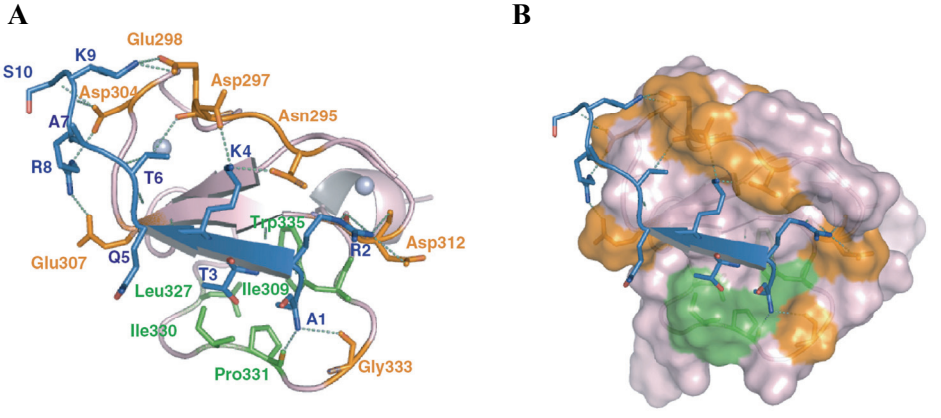

Figure 9. Solution structure of AIRE-PHD1 in complex with unmodified histone H3 peptide. Cartoon representation (A) and surface plot (B) of the lowest energy complex structure. Cyan-dashed lines represent intermolecular hydrogen bonds. Colour coding: pink, AIRE-PHD1; blue, unmodified histone H3; orange, protein residues forming specific polar contacts with unmodified histone H3; green, protein residues forming hydrophobic contacts with unmodified histone H3; grey, Zn2+ ions.

The overall interaction mode of AIRE-PHD1 and unmodified histone H3 N-terminal peptide is similar to that observed for other PHD:H3 complexes. The binding interface localizes to one negatively charged side of the PHD domain, involving residues in the N-terminus of the PHD finger, the first β-strand, and the loop connecting the first and the second β-strands. AIRE-PHD1 constitutes a stable pre-formed binding platform as formation of the complex does not induce

significant conformational changes in the protein backbone. Histone H3 N-terminal tail docks on to the surface via an induced fit mechanism by forming a third antiparallel β-strand that pairs with the protein's two-stranded antiparallel β-sheet. The side chains of the first four H3 tail residues fit inside the pre-formed AIRE-PHD1 binding groove. Charge complementarities and hydrophilic interactions appear to be the driving force of the binding. The methyl groups of A1 and T3 of histone H3 insert into hydrophobic pockets formed by AIRE Pro331, Leu308, Ile330 and Trp335 while the R2 and K4 of histone H3 occupy two adjacent negatively charged cavities formed by AIRE Cys311, Asp312 and Asn295, Asp297 respectively. Because of narrow space of the negatively charged cavities, methylation of the side chains results in a steric exclusion, which explains the binding preference for unmodified H3 N-terminal tails.

To verify this model we mutated amino acid aspartic acid 297 to alanine (D297A) in AIRE-PHD1, which according to the structure is needed for proper interaction. As expected, the mutation considerably weakened AIRE-PHD1 interaction in binding assays with calf thymus histones as well as with histone H3 peptides (Figure 4A, B in Study I). The importance of D297 and D312 as well as several other amino acids of AIRE-PHD1 in histone binding was also confirmed by our Italian collaborators (Table 1 in Study I; Table 2 in Study II).

Our results, showing that AIRE specifically interacts with histone H3 N-terminal tails and demonstrating that several posttranslational modifications interfere with the binding, are novel in many aspects. At the time when we carried out the binding experiments, there were no examples in the literature showing that PHD fingers can preferentially interact with unmodified histone H3 N-terminal tails. During the preparation of Study I, a paper published in *Nature* showed that the PHD finger of BHC80 can preferentially interact with unmodified histone H3 N-terminal tails (Lan *et al.*, 2007). Our finding adds another example to this novel family of PHD fingers that preferentially interact with unmodified H3 N-terminal tails. This set of PHD fingers exemplifies that not only different histone post-translational modifications but also the lack of posttranslational modifications can be part of "histone code". In addition, these results reveal new important aspects about the molecular mechanisms how AIRE might function in the thymus. Typically, genes with high expression are marked by high levels of H3K4me3 on their promoter whereas tissue-specific genes, which by default are not expressed in the thymus, lack H3K4me3 on their promoter. Therefore, AIRE binding to unmodified histone H3 N-terminal tails might be one of the mechanisms how AIRE is recruited to certain regions in the chromatin, leading to the preferential activation of tissue-specific genes in mTECs.

5.7. AIRE acts as a transcriptional activator in HEK293 cells

In order to study further how the histone modifications influence the AIRE mediated activation of gene expression, we used HEK293 cell-lines stably expressing AIRE. This model system was used because AIRE is expressed only in a small number of mature mTECs, and so far no cell-line with endogenous

AIRE expression is available. In a first set of experiments, we used one of the AIRE expressing (AIRE1) and yellow fluorescent protein (YFP) expressing negative control (NC) HEK293 cell-lines to study the expression of AIRE target genes *INS*, *INV* and *S100A8* by quantitative RT-PCR. and detected higher mRNA levels in the presence of AIRE (Figure 5A in Study I). This result suggested that HEK293 cells have all the necessary factors needed for AIRE mediated gene expression activation and thus provide a feasible model for studying the molecular mechanisms of ectopic expression.

To identify more AIRE target genes in HEK293 cell-lines we performed experiments with Illumina whole genome BeadChip arrays. In these experiments we used two different AIRE expressing cell-lines (AIRE1, AIRE2), a cell-line expressing AIRE with the D312A mutation in the first PHD finger (AIRE-D312A; this mutant is not capable of interacting with histone H3 N-terminal tails) and a NC cell-line as a control. Comparing the gene expression profiles of AIRE expressing cell-lines to the NC cell-line yielded 140 significantly ($p<0.05$) up- and 154 significantly ($p<0.05$) downregulated genes (Figure 10; Supplementary Figure 1A, B in Study III).

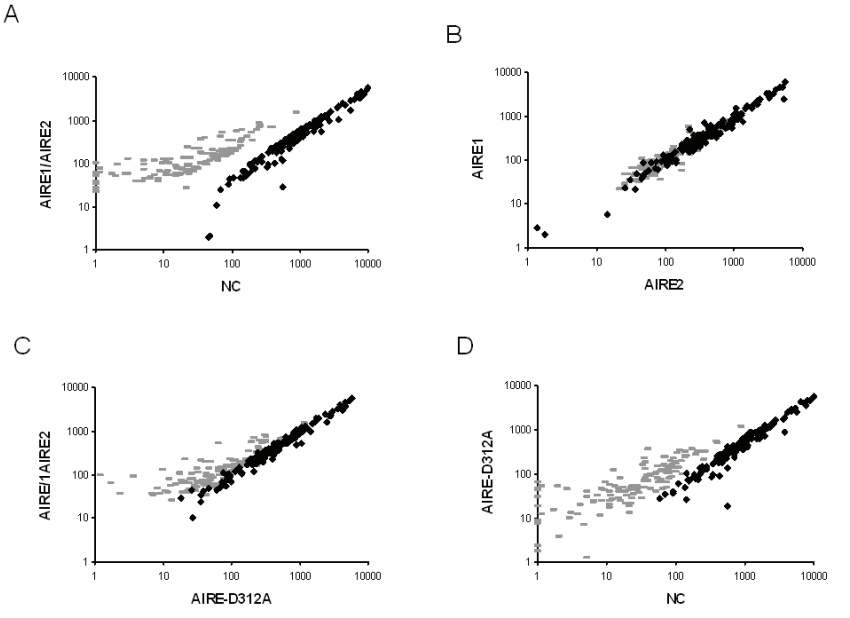

Figure 10. AIRE changes the expression of many genes in HEK293 cell-line. Signal intensity scatter-plots of 140 significantly up- (gray lines) and 154 downregulated (black spots) genes in combined AIRE1 and AIRE2 (AIRE1/AIRE2) datasets compared to the NC cell-line (A), AIRE1 cell-line compared to AIRE2 cell-line (B), AIRE1/ AIRE2 compared to AIRE-D312A cell-line (C) and AIRE-D312A compared to NC cell-line (D).

The number of AIRE upregulated genes identified with Illumina BeadChip arrays was probably underestimated since none of our initially identified AIRE target genes (*INS*, *INV*, *S100A8*) turned up as upregulated. Later on, we used quantitative RT-PCR to verify the BeadChip results and also confirmed the upregulation of *INS*, *INV*, *S100A8* (Figure 2A, B and Supplementary Figure 6 in Study III) and many other genes such as *S100A5*, *S100A6*, *S100A7*, *S100A9*, *LOR*, *SPRR2F*, which were not detected as differentially expressed by the BeadChip arrays (data not shown). These data collectively show that similar to mTECs, AIRE is able to increase the gene expression of many genes in HEK293 cells. Other groups have also observed, that expression arrays are not sensitive enough to pick up all AIRE induced gene expression changes, leading to the conclusion that AIRE may influence more genes than previously thought – several thousands rather than several hundreds (Mathis and Benoist, 2009).

The comparison of upregulated gene expression profiles of the two AIRE expressing cell-lines showed that they are highly similar (Figure 10B; Supplementary Figure 1A in Study III). This result shows that although there are differences in genes regulated by AIRE in different cell types, it tends to regulate very similar list of genes on the same cellular background. The results with AIRE-D312A mutant showed that while it was clearly less potent activator of gene expression, the repressive effect seemed not to be affected (Figure 10C, D; Supplementary Figure 1A in Study III) suggesting that histone binding is not essential for AIRE mediated gene expression repression.

Quantitative RT-PCR also confirmed that the activation of some of the AIRE target genes depended completely (*INS*, *INV*, *S100A8*, *HBG2*, *IFI16*) or partially (*OAS3*, *HCP5*, *CLDN1*) on the intact PHD finger domain, while the activation of other genes such as *GHR*, *TOX3*, *PAK3*, *SUSD4*, *CDH12* was not affected by the PHD mutation (Figure 11; Figure 6A in Study I; Supplementary Figure 6 in Study III).

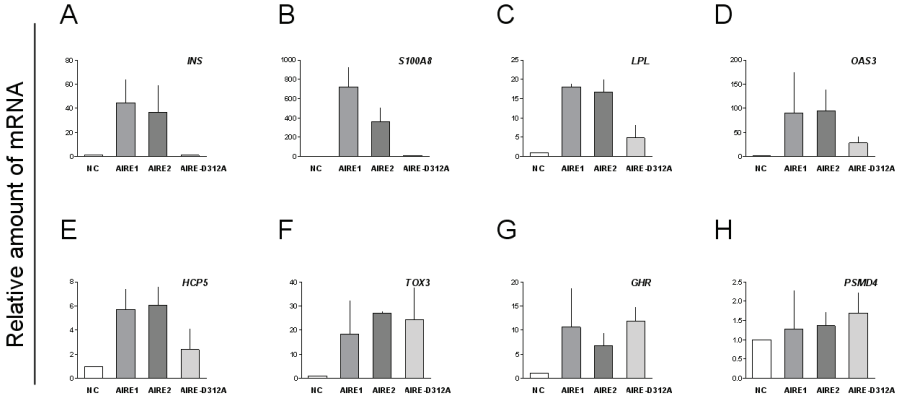

Figure 11. The extent of transcription activation induced by AIRE varies widely. RT-PCR determined mRNA levels of AIRE target genes, *INS* (A), *S100A8* (B), *LPL* (C), *OAS3* (D), *HCP5* (E), *TOX3* (F), *GHR* (G) and AIRE independent *PSMD4* gene (H) in NC, AIRE1, AIRE2 and AIRE-D312A cell-lines are shown relative to the respective mRNA levels in NC cell-line(=1). Average data of three independent experiments with SD are shown.

These data show that the intact first PHD finger domain of AIRE is needed for the activation of full range of AIRE target genes. This seems to be also relevant in physiological context, as a loss of thymic expression of even a single antigen can be sufficient to cause autoimmunity (DeVoss *et al.*, 2006; Gavanescu *et al.*, 2007; Fan *et al.*, 2009).

5.8. Characterization of AIRE target genes in HEK293 cells

The list of genes upregulated by AIRE seemed to consist of random genes. Although we also analyzed the expression of well-characterized APECED autoantigen genes (*CYP17A*, *CYP21A2*, *CYP11A*, *TG*, *TPO*), these were not activated by AIRE in HEK293 cells. This prompted us to ask whether there is anything else in common for AIRE activated genes in HEK293 cells?

It has been shown that regardless of the cellular environment, AIRE preferentially activates genes that are characterized by tissue-specific expression pattern (Derbinski *et al.*, 2005; Gardner *et al.*, 2008). To determine whether AIRE upregulated genes are enriched for tissue-specific genes in HEK293 cells, we determined the proportion of tissue-specific genes among AIRE upregulated genes and all RefSeq genes present on Illumina BeadChip arrays using the tissue-specific gene expression and regulation (TiGER) database (http://bioinfo.wilmer.jhu.edu/tiger/) (Liu *et al.*, 2008). The comparison of the observed versus expected frequencies of tissue-specific genes among the AIRE upregulated genes revealed that the proportion of tissue-specific genes was indeed significantly higher than would have been expected by chance (chi-square test p-value 2.7×10^{-8}). Furthermore, as previous studies with mTECs have shown (Derbinski *et al.*, 2005), the range of tissues expressing the AIRE activated genes was extremely wide, covering essentially the whole body (Supplementary Figure 2 in Study III).

Previous studies have shown that AIRE-regulated genes tend to cluster in the genome (Derbinski *et al.*, 2005; Johnnidis *et al.*, 2005). To determine whether AIRE-regulated genes are clustered in HEK293 cells, we used a program called REEF (http://telethon.bio.unipd.it/bioinfo/reef/index.html) that compares the frequency of selected features (AIRE-regulated genes) against reference features (genes of the whole genome) by using a sliding window approach (Coppe *et al.*, 2006). The analysis of gene density with a 200 kb window size at a false discovery rate <0.05 identified 9 clusters with AIRE-upregulated, 6 with AIRE-downregulated and 1 containing both up- and downregulated genes (Supplementary Figure 3 in Study III). The repeated use of a random set of 140 genes resulted in average of 3 significant clusters. This shows that AIRE-regulated genes tend to be more clustered than would be expected (chi-square test p-value 3.6×10^{-6}), not only in mTECs but also in our HEK293 cell model.

The scatter plots comparing the gene expression profiles of up- and downregulated genes (Figure 9A) suggested that, compared to downregulated genes, AIRE upregulated genes have lower initial expression levels. Looking at

the distribution of genes by their signal intensities on the Illumina BeadChips confirmed that AIRE upregulated genes have low initial expression level and AIRE downregulated genes have high initial expression level (Figure 12).

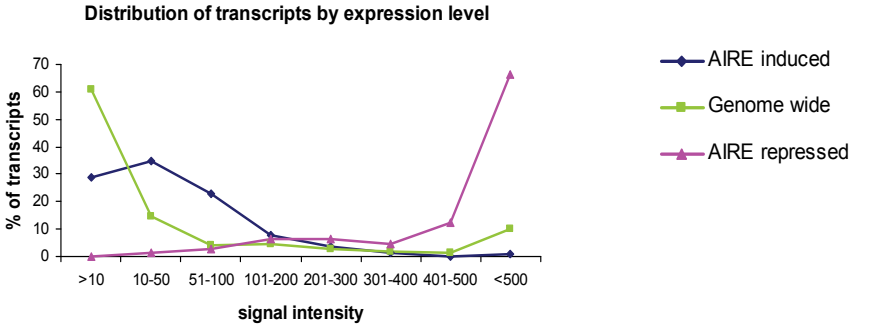

Figure 12. The distribution of AIRE induced and repressed genes compared to that of the whole genome by signal intensity in NC cell-line.

This finding was further supported by comparison of the average signal intensity of all genes on the Illumina arrays (average signal intensity 290) with AIRE-upregulated (57) or downregulated (1357) average gene signal intensities. Moreover, the average increase of AIRE-upregulated genes was far higher (12.4-fold) than the average decrease (2.6-fold) of repressed genes. In addition, there was a significant negative correlation between initial expression levels in NC cells and the fold of induction (Pearson correlation, two-tailed p-value= 0.0009), so that genes with low initial expression levels had higher induction rates (Supplementary Figure 5A in Study III). We also analyzed the distribution of AIRE-regulated genes by size and exon number, but overall, we found no difference compared to genes of the whole genome (Supplementary Figure 4B, C in Study III).

Collectively, these results show that similar to the other cell types, AIRE preferentially activates wide range of tissue-specific genes also in HEK cells. Furthermore, these genes tend to be clustered in the genome as shown for AIRE regulated genes in mTECs. A possible explanation for the clustering is provided by a recent study, which showed that AIRE is in the same complex with TOP2 and induces similar gene expression changes as etoposide, a chemical that is known to stabilize TOP2-introduced single-stranded breaks (Abramson *et al.*, 2010). Interestingly, it has been previously shown that in contrast to genes regulated by c-myc oncogene, the etoposide regulated genes locate much closer to each other than would have been expected by random distribution of genes in genome (Reymann and Borlak, 2008). Such clustering suggests that, besides being controlled individually, AIRE target genes may also be subject to regulation according to their location within the genome.

The fact that AIRE preferentially activates tissue-specific genes with low initial expression seems to have physiological importance as the purpose of promiscuous gene expression in the thymus is to widen the scope of proteins that could be presented to developing T-cells and with that to assure effective negative selection. AIRE seems to raise the expression level to certain extent, more for the genes with low initial expression level and less for genes with higher expression. Since these proteins do not perform their usual function in the thymus, they are only needed at certain expression level, which is enough to ensure the proper negative selection of self-reactive T-cells. Similarly, others have noted that AIRE does not operate as a simple on/off switch but seems to enhance the transcription to certain extent (Mathis and Benoist, 2009). The data is also compatible with the threshold model, which defines certain parameters that need to be fulfilled for the proper central tolerance. These parameters include: the number of mTECs expressing a given tissue-specific antigen; the level of expression of tissue-specific genes per mTEC; proper processing and presentation of a given antigen; the proper three dimensional organization of mTECs; lifespan of a mTECs and a residence time of the thymocytes in the medulla. Deviation in any of these parameters, including reduction in AIRE defined transcript levels, can lead to the breakdown of central tolerance (Tykocinski *et al.*, 2008).

5.9. Histone modifications at AIRE target genes

After the identification of AIRE target genes in HEK293 cells, we next aimed to study chromatin structure nearby these genes on a broader scale using ChIP assay with AIRE1 and NC cell-lines followed by hybridization of the immunoprecipitated material to Nimblegen custom made tiling arrays (ChIP-on-Chip). The arrays contained genomic regions of several control genes and 51 AIRE target genes identified previously by expression analysis in HEK293 cells. To distinguish between possibilities of proximal promoter-related or more distant regulation mechanisms, the arrays were designed to contain genomic regions that covered the 200 kb up- and downstream of the target gene. To distinguish between different chromatin states, we used antibodies against acetylated histone H3 (AcH3) and H3K4me3, which together should mark open chromatin and actively transcribed genes. In addition, we were interested in studying H3K4me3 because of its function to block the AIRE binding to the chromatin. For the identification of transcriptionally silent chromatin, we used antibody against H3K27me3 modification and, as a control, we also used antibody against general histone H3.

First, we verified the overall performance of the arrays using the previous knowledge that H3K4me3 and AcH3 mark transcriptionally active genes and H3K27me3 marks transcriptionally inactive genes (Kouzarides, 2007). Using the Nimblegen peak finding algorithm to identify the presence of studied histone modifications, we saw a clear trend for genes with higher expression levels to preferentially have H3K4me3 and AcH3 marks on their promoters (Supplementary Figure 7A, B in Study III). In contrast, genes with low

expression level tended to have H3K27me3 more often on their promoters (Supplementary Figure 7C in Study III).

Having seen that the histone posttranslational modification behaved as expected, we next analyzed the presence of chromatin modifications on the promoters (2 kb up- and downstream of TSS) of AIRE target genes. As expected from the low basal expression levels, 85% AIRE upregulated gene promoters in NC cell-line were devoid of transcriptionally active gene marks H3K4me3 and AcH3 (Figure 13). A proportion (18%) of the genes contained H3K27me3 mark and majority of the genes (70%) lacked all studied modifications. When compared to random genes with high or low expression, the AIRE target genes are more similar to the latter (Figure 13; Table 1 in Study III).

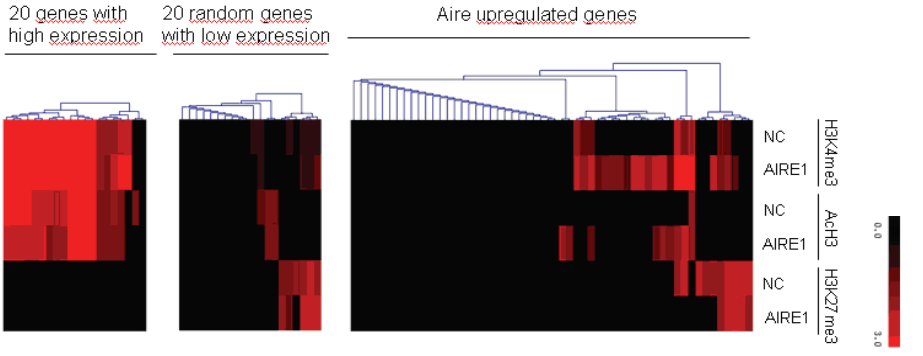

Figure 13. AIRE mediated transcriptional activation is accompanied by changes in histone posttranslational modifications. The presence of histone modifications (H3K4me3, AcH3, H3K27me3) on the promoters of AIRE upregulated genes, and on the promoters (2 kb up- and downstream of transcription start sites) of genes with high and low expression levels, in AIRE1 and NC cell-lines. Red and black colours denote the presence and lack of corresponding modifications, respectively. The shades of red represent peak scores, which translate into the amounts of corresponding modifications.

Compared to NC cell-line, we found several changes in the posttranslational modifications on the promoters of AIRE target genes in AIRE1 cell-line. There was clear increase in the number of genes having H3K4me3 (36% in AIRE1 vs. 15% in NC) or AcH3 (16% in AIRE1 vs. 2% in NC) mark and the amount of the modification also increased as assessed by the peak score. As expected from the increase of modifications that characterize active genes, the number of genes with H3K27me3 mark was decreased (9% in AIRE1 vs. 18% in NC). Such changes were not seen neither among random genes with low expression nor among genes with high expression (Figure 13; Table 1 in Study III).

As an alternative to the use of the peak finding algorithm of the SignalMap program (Nimblegen), we normalized the results against H3 and calculated the average signal intensity for each histone modification in a region of 2 kb up- and downstream of AIRE target genes transcription start sites. Such an analysis also

showed clearly that in AIRE1 cell-line, there is an increase in H3K4me3, AcH3 and decrease in H3K27me3 levels on majority of the AIRE target genes (Supplementary Figure 8 in Study III). In addition, this analysis revealed that genes having higher levels of H3K4me3 on their promoters also had higher levels of H3K27me3 on their promoters in NC cell-line. Most of these genes remained bivalent in AIRE1 cell-line but the ratio of the modifications shifted towards H3K4me3 (Supplementary Figure 8 in Study III). Whole genome analysis of histone modifications have identified that a number of genes have bivalent modifications (H3K4me3 and H3K27me3) on their promoters and these genes are usually inactive but poised for transcription (Mikkelsen *et al.*, 2007).

Analysis of individual genes more closely showed that AIRE independent genes, like *GAPDH*, had similar levels of studied histone posttranslational modifications in NC and AIRE1 cell-lines whereas some of the AIRE upregulated genes, like *GHR*, had several changes in histone posttranslational modifications that reflect transcriptional activation (Figure 14). While H3 levels were similar in NC and AIRE1 cell-line throughout the whole region, there was an increase in H3K4me3 and AcH3 levels exactly adjacent to the transcriptional start site. H3K27 levels, however, were decreased in a ~200kb region that covered the whole gene (Figure 14). Similar changes were seen with *OAS3*, *LPL*, *TOX3* and *HSBP8* genes (Supplementary Figure 9A, B, C, D in Study III).

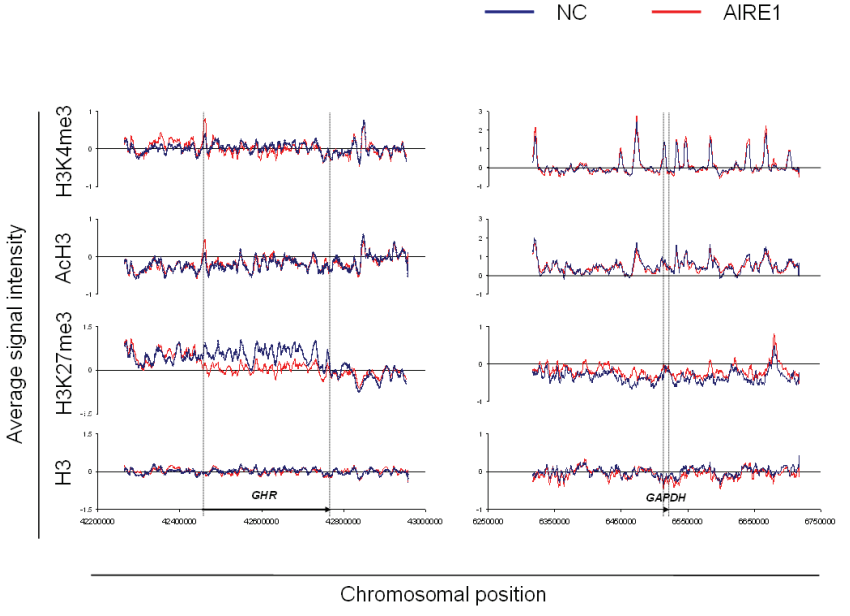

Figure 14. AIRE mediated gene activation is accompanied by changes in histone posttranslational modifications. Average levels of histone posttranslational modifications (H3K4me3, AcH3, H3K27me3) and general H3 in a region of 200 kb of up- and downstream of AIRE upregulated *GHR* and AIRE independent *GAPDH* genes (arrows) in NC (blue line) and AIRE1 cell-lines (red line).

Other genes had various combinations of changes. For example, *PLXDC2*, *MAL2* had increased H3K4me3 and AcH3 levels on the promoters, but comparable levels of H3K27me3 throughout the gene (data not shown). *PLD5* and *MYRIP* had only H3K27me3 decrease inside of the gene, but no H3K4me3 and AcH3 changes around transcription start site (data not shown). Analysis of other genes showed that many of the AIRE target genes had largely comparable levels of studied histone posttranslational modifications in both cell-lines.

To verify that the changes seen in histone posttranslational modifications by ChIP-on-chip approach were reproducible, we performed conventional chromatin immunoprecipitation assays with αH3K4me3, αH3K27me3, αAcH3 and αH3-2 antibodies using NC and AIRE1 cell-lines. As observed with ChIP-on-chip results, the promoter regions of *GHR* and *LPL* genes contained more H3K4me3 and AcH3 marks and on their promoter regions in AIRE1 cell-line, whereas H3K27me3 levels were reduced in the promoter region and also in the 3' region of the genes (Figure 5 in Study III). Such changes were not found with *INV*, *INS* and *S100A8* genes as well as with AIRE independent *PSMD4* and *GAPDH* genes (Figure 5, Supplementary Figure 10 in Study III). Small rise in H3K4me3 levels were also seen on the promoters of *CHD12*, *PAK3*, *SUSD4*, *CLDN1* genes whereas *EFHC2*, *HBG2*, *THBS2*, *ZBTB7B* gene promoters had comparable levels of H3K4me3 (Supplementary Figure 10 in Study III).

In conclusion, these results show that the gene expression activation by AIRE can be accompanied by various changes in histone posttranslational modifications on the promoter regions and/or within the target genes. Although interesting, these changes are most likely not a direct effect of AIRE. Rather, the wide variation of changes seen in the histone posttranslational modifications among AIRE target genes suggest that the final outcome of gene activation is determined by the local environment – by chromatin state and the presence of other (histone modifying) proteins.

5.10. Posttranslational histone modifications in AIRE regulated gene clusters

As AIRE regulated genes tend to cluster in the genome, it has been proposed that AIRE might induce structural changes of larger chromatin regions (Derbinski *et al.*, 2005; Johnnidis *et al.*, 2005). Therefore, we examined an 1.8 mb genomic region on chromosome 1, which includes an epidermal differentiation cluster that contains several AIRE target genes. In general, the overall distribution of the studied posttranslational histone modifications in AIRE1 and NC cell-lines was very similar (Supplementary Figure 11 in Study III). This result suggests that at least in this 1.8 mb region AIRE does not mediate large epigenetic rearrangements. We also analyzed other AIRE regulated gene clusters but found no changes in histone modifications that covered more than one gene region (data not shown). These results suggest that most likely AIRE does not mediate large rearrangements in chromatin structure.

5.11. The presence of RNA polymerase II on AIRE target gene promoters

Following a study by Oven et al., which showed that AIRE promotes transcriptional elongation by recruiting P-TEFb to target gene promoters (Oven et al., 2007), we also studied the presence of RNA polymerase II on the promoters of AIRE target genes in NC and AIRE1 cell-lines by ChIP-on-chip and for the confirmation of the results, we also performed conventional ChIP. The scenario in which AIRE propagates elongation assumes the presence of stalled RNA polymerase II on the target gene promoters. We, however, found RNA polymerase II only on 4% of AIRE target gene promoters in NC cell-line. In agreement with the increased expression, RNA polymerase II was found on 15 (26,8%) AIRE upregulated gene promoters in AIRE1 cell-line. The lack of RNA polymerase II in our experiments could be explained by low detection efficiency, which is in part supported by the data that many of the AIRE target genes contained bivalent modifications characteristic for genes being poised for transcription (Supplementary Figure 8 in Study III). However, we observed an increase in RNA polymerase II levels on several AIRE target genes in AIRE1 cell-line by ChIP-on-chip (Data not shown) and this was also confirmed by conventional ChIP experiments (Figure 5 in Study III). These results suggest that AIRE mediated increase in mRNA levels is, at least to some extent, due to enhanced transcription and involves the recruitment of RNA polymerase II to the target gene promoters.

5.12. AIRE binding to chromatin

Several studies have shown that AIRE interacts with chromatin on the promoters of its target genes (Oven *et al.*, 2007; Ruan *et al.*, 2007). We therefore also studied the presence of AIRE on its target gene promoters. Initial ChIP experiments showed the presence of AIRE on the promoters of its target genes *INS*, *INV* and *S100A8* (Figure 5B in Study I) while these interactions were not seen in the cell-lines expressing AIRE with a D297A or D312A mutation (Figure 6B in Study I). In addition, less binding was observed on the promoter of housekeeping gene *GAPDH* and on the promoter of AIRE independent *S100A10* gene.

We also studied the binding of AIRE to the chromatin by ChIP-on-chip. The immunoprecipitation results with AIRE-1 antibody from AIRE1 cell-line were normalized against AIRE immunoprecipitation results from NC cell-line. From the normalized data, the Nimblegen algorithm managed to identify 1854 peaks, which were at first sight distributed rather randomly throughout the studied regions. Analysis of the regions with CEAS system (http://ceas.cbi.pku.edu.cn/) (Ji *et al.*, 2006) showed that compared to the background (α-AIRE immunoprecipitation results from NC cell-line), the AIRE binding sites localized more to gene promoters and exons (Table 6).

Table 6. The genomic distribution of AIRE binding sites (AIRE) compared to the background (NC)

	Exon %	Intron %	5'UTR %	3'UTR %	Proximal promoter %	Immediate downstream %	Enhancer %
AIRE	14.7	32.8	0.05	1.6	4.4	2.6	43.9
NC	6.1	34.0	0.02	0.7	1.4	2.3	55.5

The Nimblegen arrays contained 704 genes and out of these, in a region of 2 kb up- and downstream of transcription start sites, AIRE binding peaks were found on 217 (30.8%) genes including on one AIRE downregulated (out of 13 on the array; 7.7%) and 17 upregulated (out of 51 on the array; 33.3%) genes. In addition, the AIRE binding sites were found inside of 30 (55.6%) AIRE up-regulated and 3 (23%) downregulated genes. Combining the two analyses, the AIRE binding sites were found within or nearby 34 (63%) AIRE upregulated and 3 (23%) downregulated genes. These results suggest that AIRE binds to chromatin in a rather random fashion and/or the binding is dynamic. The random or temporal binding of AIRE to the chromatin is in agreement with studies, which show that at a single cell level mTECs express only a limited number of tissue-specific genes that are regulated by AIRE at the cell population level (Derbinski *et al.*, 2008; Villasenor *et al.*, 2008). However, these studies are restricted by technical limitations and are compatible with scenarios were individual mTECs express different sets of tissue-specific genes or the overall set of AIRE regulated genes expressed in individual mTECs is the same but all the genes are not expressed simultaneously, therefore, in a given time point, only certain transcripts can be detected (Figure 3 in Study IV). At the cell population level both scenarios lead to the similar list of AIRE regulated genes and both scenarios are also compatible with our data.

5.13. Histone H3 N-terminal modifications influence AIRE mediated transcriptional activation

Our experiments with differentially modified histone H3 N-terminal peptides showed that posttranslational modifications, such as H3K4 methylation, H3T3 phosphorylation and H3R2 methylation disrupt AIRE binding to histone H3 N-terminal ends. This is in agreement with our initial AIRE ChIP data, however, the rather random binding observed by ChIP-on-chip prompted us to search for further evidence that histone posttranslational modifications influence AIRE mediated transcriptional activation.

Our gene expression analysis showed that the activation of some of the AIRE target genes was not affected by the D312A mutation. Since this mutant is not capable of interacting with histones AIRE must use other mechanisms than histone binding to interact with chromatin to activate its target genes. Accordingly, the H3K4me3 levels on the promoters of these genes should not

influence AIRE binding. The activation of another subset of AIRE target genes, however, was decreased by the AIRE D312A mutation. The activation of these genes relies on the interaction with histones and therefore the promoters of these genes should have lower levels of H3K4me3 on their promoters since this mark blocks AIRE binding. To address this question, we compared 15 genes most or least affected by the D312A mutation and found that affected genes have significantly lower H3K4me3 levels on their promoters, both in NC and AIRE1 cell-lines (Student's two tailed t-test p-values 0.01 and 0.001 respectively) (Figure 15).

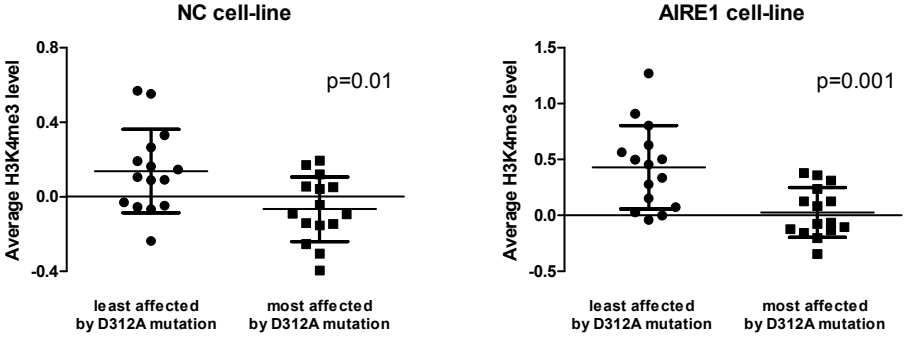

Figure 15. H3K4me3 levels on AIRE target gene promoters influence AIRE-D312A mutant ability to activate gene expression. 2 kb up- and downstream of transcription start sites are included to calculate average levels of H3K4me3 levels on the promoters. Student's two tailed t-test p-values were 0.01 and 0.001 in NC and AIRE1 cell-lines, respectively.

This result further supports the idea that AIRE interaction with histones is needed for the activation of genes with low level of H3K4me3 on their promoters. It seems that in the physiological context the activation of genes that lack H3K4me3 is especially important since these genes have a low expression level, which needs to be increased over a certain threshold to ensure proper negative selection.

We also studied the effect of increased H3R2 methylation on AIRE mediated gene expression activation in HEK293 cells. We used the over-expression of histone H3R2 methyltransferase PRMT6 to increase the dimethylation levels of R2. Indeed, PRMT6 overexpression lead to strong R2 dimethylation increase both when transfected with control plasmid or with AIRE (Figure 16A). The activation of AIRE target genes *S100A8* and *INV* was remarkably reduced in the presence of higher R2 methylation level when PRMT6 was co-transfected with AIRE (Figure 16A, B). Importantly, no reduction of studied mRNA levels was observed when PRMT6 was transfected alone, indicating that R2 methylation antagonizes the AIRE ability to activate gene expression.

Three out of first eight N-terminal amino acids of histone H3 can be post-translationally modified (R2 methylation, T3 phosphorylation and K4 methy-

lation) and all these three modifications interfere with AIRE binding. Not much is known about the function of histone H3T3 phosporylation. In *Chlamydomonas* the mark has been associated with silencing of gene expression (Casas-Mollano *et al.*, 2008). More is known about histone H3 R2 methylation. Recent studies have shown that H3K4me3 and H3R2me2 marks are mutually exclusive. Accordingly, R2 dimethylation localizes more to the 3' region of the genes and is depleted on the promoters of active genes, which have high levels of H3K4me3. The highest levels of R2 dimethylation were associated with heterochromatin (Guccione *et al.*, 2007; Kirmizis *et al.*, 2007). Another genome-wide ChIP-seq study found that H3R2me1 and H3R2me2 are not enriched on the promoters of neither active nor inactive genes (Barski *et al.*, 2007). H3K4me3 has been shown to preferentially locate on the promoters of active genes (Bernstein *et al.*, 2005). The data suggest that these modifications occupy different genomic regions and therefore independently hold the potential to restrict the locations in chromatin where AIRE can bind to. However, considering the low level of H3R2 methylation, compared to H3K4 methylation, the latter seems to be the major negative determinator for AIRE interaction with histones.

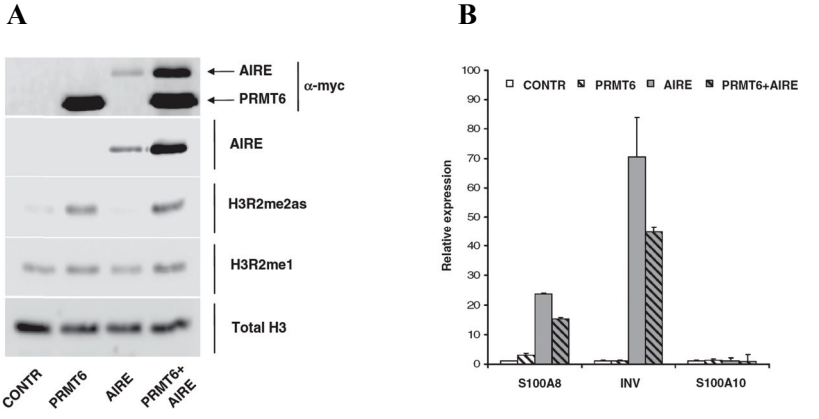

Figure 16. R2 methylation hinders AIRE ability to activate target genes. (A) Western blot analysis of transfected HEK293 cells. Transfections were carried out as indicated below, antibodies and/or detected proteins are indicated on the right. (B) Relative expression level of AIRE regulated *INV* and *S100A8* genes and AIRE independent *S100A10* gene. Activations, measured with quantitative RT–PCR, are shown as relative quantity of mRNA compared to control transfection (= 1) of each mRNA. The data are the averages of two independent transfections.

5.14. AIRE target gene posttranslational histone modifications in mouse mTECs and in the peripheral tissues

Our next aim was to study histone posttranslational modifications on the promoters of selected tissue-specific genes in mouse mTECs and their corresponding peripheral tissues. According to our hypothesis, Aire dependent tissue-specific genes should have low levels of H3K4me3 on their promoters in mTECs and higher levels of H3K4me3 in their peripheral tissues, where they are highly expressed. In the thymus, immature mTECs characterized by low expression of co-stimulatory molecule CD80 are negative for AIRE. During their differentiation into mature mTECs, they upregulate CD80 expression and also start to express AIRE. To identify the differences between AIRE negative and positive mTECs, we studied immature CD80low and mature CD80high populations separately. The major obstacle in studying histone modifications from primary cells has been the small quantity of cellular material available. Recent advances in methodology, however, enable researchers to use as few as 100 cells for chromatin immunoprecipitation (Dahl and Collas, 2007). In our experiments, we used the Q^2ChIP method, which enables to study histone modifications from a small number (~20 000) of cells.

The Q^2ChIP results showed that housekeeping gene *Gapdh* promoter contained similarly high levels of H3K4me3 and low levels of H3K27me3 in all studied tissues: mTECs (CD80low, CD80high), pancreas and neutrophils (Figure 17).

As expected, compared to other tissues, the H3K4me3 levels on the promoters were higher in the tissues where each of the studied genes is preferentially expressed: *Ins2* and *Gad2* in pancreas and *S100a8* in neutrophils. H3K27me3 levels were low almost on all studied promoters in all tissues except for slightly higher levels on *S100a8* promoter in pancreas. In CD80low mTECs, the H3K4me3 levels were low on the promoters of AIRE dependent tissue-specific antigens *Ins2* and *S100a8*, but much higher on the promoter of AIRE independent tissue-specific gene, *Gad2* promoter. In CD80high compared to CD80low mTECs, a rise in H3K4me3 levels was detected on the promoters of *Gad2* and *S100a8*, while the promoter of *Ins2* remained unchanged. Comparing CD80low and CD80high subsets, H3K4me3 levels were also higher on the *Aire* promoter correlating well with the increase in Aire expression (Figure 17).

Together, these results are in agreement with data acquired from HEK cell-lines showing that AIRE target genes have initially low levels of H3K4me3 on their promoters, which then depending on the gene, can be enhanced in the presence of AIRE.

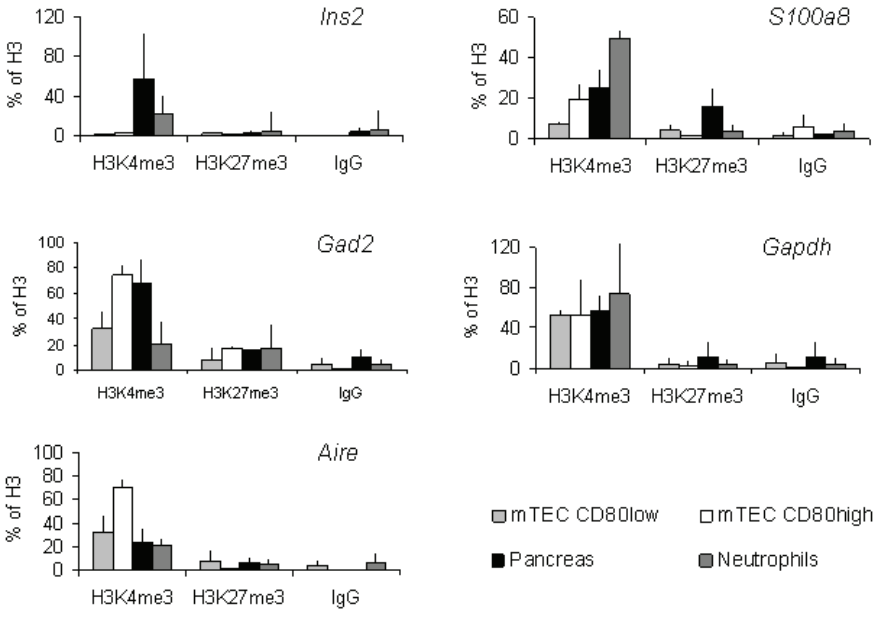

Figure 17. Aire target genes have low levels of H3K4me3 on their promoters in mouse mTECs. Q^2ChIP analysis of the promoters of Aire dependent tissue-specific antigens, *Ins2* and *S100A8* as well as Aire independent tissue-specific antigen *Gad2*, housekeeping gene *Gapdh* and *Aire* in CD80high, CD80low mTECs, pancreas and neutrophils. Average data of two independent experiments with SD are shown.

6. GENERAL DISCUSSION

The potential capacity of the immune system to cause a disease was first recognized at the turn of the 20th century. Although Paul Ehrlich's "horror autotoxicus" postulate that autoimmune diseases could not occur was misleading and dominated the field for about 50 years, it also clearly referred to the existence of mechanisms that would ensure immunological tolerance against self. Yet, it is clear now that these mechanisms are not always perfect, which is reflected by the increasing number of diseases that are suspected to have an autoimmune origin. The understanding of any autoimmune disease process requires the detailed characterization of the mechanisms that contribute to the maintenance and break-down of self-tolerance.

Negative selection of self-reactive thymocytes in the thymus has emerged as a fundamental process to ensure central tolerance and it has been shown that the expression of the AIRE protein is needed for a proper negative selection. The basis of negative selection is the presentation self-peptide–MHC complexes by the APCs and subsequent thymocyte TCR binding avidity to the MHC:peptide complex. The scope of the negative selection is determined by the peptides available for the presentation in the thymus. In ideal case, each developing thymocyte should be screened against every self protein. While this is probably not true, mTECs do express a large number of tissue-specific genes and the presence of the AIRE protein expands the number of genes expressed even further (Anderson *et al.*, 2002). Therefore, the main mechanism by which AIRE has been proposed to be responsible for a proper negative selection is the promotion of the expression of tissue specific genes in mTECs thus making them available for a presentation. Compatible with this model, studies published so far indicate that regardless of the cell type, AIRE activates the expression of many genes, however, the set of activated genes is specific for each cell type (Anderson *et al.*, 2002; Gardner *et al.*, 2008; Guerau-de-Arellano *et al.*, 2008). In addition, although different sets of genes are activated, they are enriched for genes having tissue specific expression pattern and tend to localize in clusters in the genome. This was also true for the HEK293 cells used in this study. These data suggest that the gene expression activation mechanism used by AIRE must be rather general to function in different cell types. Yet, the mechanism utilizes something that is common for tissue specific genes as well as something that is specific for a given cell type, which determines the final set of activated genes.

Our results, showing that AIRE utilizes its first PHD zinc finger for the interaction with histone H3 are in a good agreement with these data. Histones are by far the most abundant chromatin bound proteins present in eukaryotic cells. However, the chromatin structure and histone posttranslational modifications that play an important role in determining which genes are accessible to the transcriptional machinery vary between cell types. Moreover, the ability of AIRE to recognize different histone H3 N-terminal modifications, especially discriminating against active transcription mark H3K4me3, could be the key to the notion that AIRE preferentially activates genes which have low initial expression level and are enriched for tissue specific genes. Tissue-specific genes, which are expressed only in a few specific tissues, lack H3K4me3 in

other tissues, including thymus, and are thus potential targets of AIRE (Supplementary Figure S8 in Study I).

Our gene expression array data suggest that in addition to histone H3 binding, AIRE uses some other means, for example binding to DNA or other chromatin bound proteins to interact with chromatin. In some cases, these other interactions seem to be sufficient to assure AIRE mediated gene expression activation. The use of multiple domains to interact with chromatin is probably widespread and helps to increase the affinity and specificity of the interaction. The importance of AIRE interaction with other, not necessarily directly chromatin bound proteins, is further underlined by the fact that the recognition of the same histone modifications can lead to opposing biological readouts. The interaction with unmodified H3 N-terminal tails in case of AIRE is needed for gene expression activation, whereas in case of BHC80 leads to LSD1 mediated gene expression repression (Lan *et al.*, 2007).

Some of the recent studies shed light to some of the other interactions and processes that, in addition to histone H3 binding, are needed for AIRE mediated gene expression activation (Figure 18; Figure 2 in Study IV) (Oven *et al.*, 2007; Abramson *et al.*, 2010). Accordingly, using its first PHD zinc finger AIRE selectively interacts with unmodified histone H3 N-terminal tails and is therefore recruited to chromatin regions which lack histone H3 N-terminal modifications. These regions exclude the promoters of actively transcribed genes but include the promoters of inactive genes, which are enriched for genes characterized by tissue-specific expression pattern. In addition to histone binding, interaction with DNA or other proteins might be necessary to guide AIRE to specific regions. On target gene regulatory regions, AIRE seems to promote the formation of protein complexes that favor transcript elongation and mRNA processing. AIRE interaction with TOP2 leads to the relaxation of the underling chromatin by introducing DSBs to the DNA.

The DSBs would mobilize H2AX, and activate DNA-PK/Ku80, PARP-1 and other proteins found to associate with AIRE. The accumulation of DSBs would also explain the notion that AIRE promotes apoptosis of mTECs (Gray *et al.*, 2007). AIRE further enhances the transcription by recruitment of the positive transcription elongation factor b (P-TEFb) complex, which phosphorylates the serine residues of stalled RNA polymerase II. This phosphorylation converts the polymerase from a paused to elongating form and results in the activation of gene expression. AIRE interaction with other proteins involved in mRNA processing further promotes the accumulation of fully mature mRNAs. The interaction of AIRE with CBP, PIAS1 and nuclear matrix could also be involved in modulating the AIRE mediated transcriptional activation. Although AIRE seems to interact with many proteins that are involved in different stages of transcription, the exact order and nature of the underlying processes requires far more research.

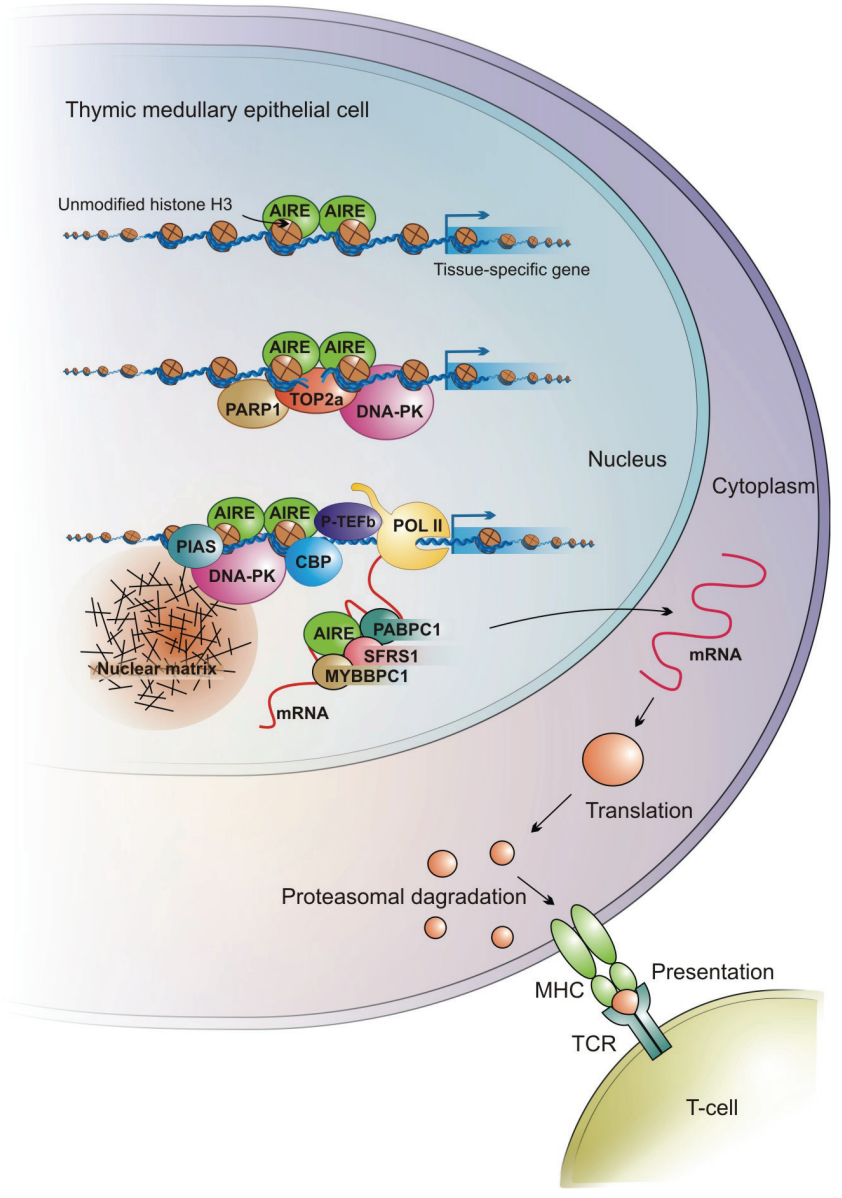

Figure 18. Proposed mechanism of AIRE-mediated preferential activation of tissue-specific genes in the thymus, making them available for presentation to T-cells. AIRE preferential binding to unmodified histone H3 N-terminal tails is needed to direct AIRE to certain regions in chromatin including the regulatory regions of tissue-specific genes where AIRE seems to promote the formation of protein complexes that favor transcript elongation and mRNA processing. Adapted from (Peterson *et al.*, 2008; Kyewski and Peterson, 2010).

Although the complete understanding of the mechanism leading to the activation of tissue-specific antigens in the thymus remains to be achieved, the importance of AIRE histone binding in transcriptional activation is also confirmed by other studies where other research groups independently of us have got essentially the same results (Koh *et al.*, 2008). In the physiological context, the relevance of AIRE histone binding is underlined by the fact that it is most needed for the activation of genes that have low levels of H3K4me3 on their promoters and have therefore a low expression level. These are the genes which expression needs to be increased the most in order to make them available for presentation and with that assure proper negative selection and self-tolerance.

7. CONCLUSIONS

1. We identified that similar to other proteins with PHD fingers, AIRE can directly interact with histone H3 through its first PHD zinc finger domain, thus providing a new mechanism how AIRE might be tethered to chromatin.
2. We showed that AIRE interaction with histone H3 is disrupted by several histone H3 N-terminal modifications, such as H3K4me3 and H3R2me2. Therefore, AIRE preferentially interacts with histone H3 N-terminal tails that have no posttranslational modifications. In collaboration with Dr. Giovanna Musco and her colleagues from Dulbecco Telethon Institute c/o S. Raffaele Scientific Institute, Milan, the interaction has been characterized in detail with NMR experiments. The recognition of unmodified histone H3 tails by AIRE-PHD1 is in addition to the BHC80 protein an important example of unmodified histone H3 N-terminal tail readers. This exemplifies that also lack of histone posttranslational modifications is a part of histone code needed to regulate different biological processes. We also showed that AIRE can bind to mononucleosomes *in vitro* and that it selectively interacts with histones *in vivo*, therefore providing evidence how AIRE can be recruited to certain locations in chromatin, which results in preferential activation of tissue-specific genes
3. Using a model system, HEK293 cell-lines stably expressing AIRE, we showed that AIRE indeed preferentially activates genes that are tissue-specific and characterized by low levels of initial expression. The activation of full set of AIRE target genes is dependent on the intact first PHD finger of AIRE and therefore on the AIRE interaction with histones.
4. We also identified that AIRE regulated genes lack active chromatin marks, such as H3K4me3 and AcH3, on their promoters. During activation by AIRE, some of the target genes acquire histone H3 posttranslational modifications associated with transcription and RNA polymerase II, suggesting that AIRE mediated increase in mRNA levels is mediated, at least in part, at the transcriptional level. Furthermore, we provided evidence that although AIRE binding to chromatin seemed to be temporal or stochastic, certain histone posttranslational modifications, such as H3K4me3 and H3R2me2, influence AIRE mediated transcriptional activation. Finally, we verified our model that AIRE target genes have low levels of H3K4me3 on their promoters *in vivo* by studying histone posttranslational modifications in mouse mTECs and peripheral tissues.

In summary, our studies have lead to the identification of new aspects on the molecular mechanisms of AIRE mediated preferential activation of tissue-specific genes, thus, providing new important insight into the mechanisms how central tolerance is achieved in the thymus.

REFERENCES

1. Aasland R, Gibson TJ and Stewart AF (1995). The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem Sci 20(2): 56–9.
2. Abramson J, Giraud M, Benoist C and Mathis D (2010). Aire's partners in the molecular control of immunological tolerance. Cell 140(1): 123–35.
3. Ahonen P, Myllarniemi S, Sipila I and Perheentupa J (1990). Clinical variation of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) in a series of 68 patients. N Engl J Med 322(26): 1829–36.
4. Akiyoshi H, Hatakeyama S, Pitkanen J, Mouri Y, Doucas V, Kudoh J, Tsurugaya K, Uchida D, Matsushima A, Oshikawa K, Nakayama KI, Shimizu N, Peterson P and Matsumoto M (2004). Subcellular expression of autoimmune regulator is organized in a spatiotemporal manner. J Biol Chem 279(32): 33984–91.
5. Anderson MS, Venanzi ES, Chen Z, Berzins SP, Benoist C and Mathis D (2005). The cellular mechanism of Aire control of T cell tolerance. Immunity 23(2): 227–39.
6. Anderson MS, Venanzi ES, Klein L, Chen Z, Berzins SP, Turley SJ, von Boehmer H, Bronson R, Dierich A, Benoist C and Mathis D (2002). Projection of an immunological self shadow within the thymus by the aire protein. Science 298(5597): 1395–401.
7. Baker LA, Allis CD and Wang GG (2008). PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. Mutat Res 647(1–2): 3–12.
8. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I and Zhao K (2007). High-resolution profiling of histone methylations in the human genome. Cell 129(4): 823–37.
9. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ, 3rd, Gingeras TR, Schreiber SL and Lander ES (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120(2): 169–81.
10. Bienz M (2006). The PHD finger, a nuclear protein-interaction domain. Trends Biochem Sci 31(1): 35–40.
11. Bjorses P, Halonen M, Palvimo JJ, Kolmer M, Aaltonen J, Ellonen P, Perheentupa J, Ulmanen I and Peltonen L (2000). Mutations in the AIRE gene: effects on subcellular location and transactivation function of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy protein. Am J Hum Genet 66(2): 378–92.
12. Bottomley MJ, Stier G, Pennacchini D, Legube G, Simon B, Akhtar A, Sattler M and Musco G (2005). NMR structure of the first PHD finger of autoimmune regulator protein (AIRE1). Insights into autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) disease. J Biol Chem 280(12): 11505–12.
13. Botuyan MV, Lee J, Ward IM, Kim JE, Thompson JR, Chen J and Mer G (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell 127(7): 1361–73.
14. Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY and Allis CD (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell 84(6): 843–51.
15. Burnett E, Christensen J and Tattersall P (2001). A consensus DNA recognition motif for two KDWK transcription factors identifies flexible-length, CpG-

methylation sensitive cognate binding sites in the majority of human promoters. J Mol Biol 314(5): 1029–39.
16. Capili AD, Schultz DC, Rauscher IF and Borden KL (2001). Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains. Embo J 20(1–2): 165–77.
17. Casas-Mollano JA, Jeong BR, Xu J, Moriyama H and Cerutti H (2008). The MUT9p kinase phosphorylates histone H3 threonine 3 and is necessary for heritable epigenetic silencing in Chlamydomonas. Proc Natl Acad Sci U S A 105(17): 6486–91.
18. Champagne KS, Saksouk N, Pena PV, Johnson K, Ullah M, Yang XJ, Cote J and Kutateladze TG (2008). The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide. Proteins 72(4): 1371–6.
19. Cheung P, Tanner KG, Cheung WL, Sassone-Corsi P, Denu JM and Allis CD (2000). Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. Mol Cell 5(6): 905–15.
20. Clapier CR and Cairns BR (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem 78: 273–304.
21. Coppe A, Danieli GA and Bortoluzzi S (2006). REEF: searching REgionally Enriched Features in genomes. BMC Bioinformatics 7: 453.
22. Couture JF, Collazo E and Trievel RC (2006). Molecular recognition of histone H3 by the WD40 protein WDR5. Nat Struct Mol Biol 13(8): 698–703.
23. Dahl JA and Collas P (2007). A quick and quantitative chromatin immunoprecipitation assay for small cell samples. Front Biosci 12: 4925–31.
24. de la Paz Sanchez M and Gutierrez C (2009). Arabidopsis ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. Proc Natl Acad Sci U S A 106(6): 2065–70.
25. Derbinski J, Gabler J, Brors B, Tierling S, Jonnakuty S, Hergenhahn M, Peltonen L, Walter J and Kyewski B (2005). Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. J Exp Med 202(1): 33–45.
26. Derbinski J, Pinto S, Rosch S, Hexel K and Kyewski B (2008). Promiscuous gene expression patterns in single medullary thymic epithelial cells argue for a stochastic mechanism. Proc Natl Acad Sci U S A 105(2): 657–62.
27. Derbinski J, Schulte A, Kyewski B and Klein L (2001). Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. Nat Immunol 2(11): 1032–9.
28. DeVoss J, Hou Y, Johannes K, Lu W, Liou GI, Rinn J, Chang H, Caspi RR, Fong L and Anderson MS (2006). Spontaneous autoimmunity prevented by thymic expression of a single self-antigen. J Exp Med 203(12): 2727–35.
29. DeVoss JJ and Anderson MS (2007). Lessons on immune tolerance from the monogenic disease APS1. Curr Opin Genet Dev 17(3): 193–200.
30. Dul BE and Walworth NC (2007). The plant homeodomain fingers of fission yeast Msc1 exhibit E3 ubiquitin ligase activity. J Biol Chem 282(25): 18397–406.
31. Eberharter A, Vetter I, Ferreira R and Becker PB (2004). ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts. Embo J 23(20): 4029–39.
32. Fan Y, Rudert WA, Grupillo M, He J, Sisino G and Trucco M (2009). Thymus-specific deletion of insulin induces autoimmune diabetes. Embo J 28(18): 2812–24.
33. Ferguson BJ, Alexander C, Rossi SW, Liiv I, Rebane A, Worth CL, Wong J, Laan M, Peterson P, Jenkinson EJ, Anderson G, Scott HS, Cooke A and Rich T (2008).

AIRE's CARD revealed, a new structure for central tolerance provokes transcriptional plasticity. J Biol Chem 283(3): 1723–31.
34. Fiedler M, Sanchez-Barrena MJ, Nekrasov M, Mieszczanek J, Rybin V, Muller J, Evans P and Bienz M (2008). Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex. Mol Cell 30(4): 507–18.
35. Finnish-German_APECED_Consortium (1997). An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. Nat Genet 17(4): 399–403.
36. Garcia-Dominguez M, March-Diaz R and Reyes JC (2008). The PHD domain of plant PIAS proteins mediates sumoylation of bromodomain GTE proteins. J Biol Chem 283(31): 21469–77.
37. Gardner JM, Devoss JJ, Friedman RS, Wong DJ, Tan YX, Zhou X, Johannes KP, Su MA, Chang HY, Krummel MF and Anderson MS (2008). Deletional tolerance mediated by extrathymic Aire-expressing cells. Science 321(5890): 843–7.
38. Gavanescu I, Kessler B, Ploegh H, Benoist C and Mathis D (2007). Loss of Aire-dependent thymic expression of a peripheral tissue antigen renders it a target of autoimmunity. Proc Natl Acad Sci U S A 104(11): 4583–7.
39. Gibson TJ, Ramu C, Gemund C and Aasland R (1998). The APECED poly-glandular autoimmune syndrome protein, AIRE-1, contains the SAND domain and is probably a transcription factor. Trends Biochem Sci 23(7): 242–4.
40. Gozani O, Karuman P, Jones DR, Ivanov D, Cha J, Lugovskoy AA, Baird CL, Zhu H, Field SJ, Lessnick SL, Villasenor J, Mehrotra B, Chen J, Rao VR, Brugge JS, Ferguson CG, Payrastre B, Myszka DG, Cantley LC, Wagner G, Divecha N, Prestwich GD and Yuan J (2003). The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. Cell 114(1): 99–111.
41. Gray D, Abramson J, Benoist C and Mathis D (2007). Proliferative arrest and rapid turnover of thymic epithelial cells expressing Aire. J Exp Med 204(11): 2521–8.
42. Gregersen PK and Olsson LM (2009). Recent advances in the genetics of autoimmune disease. Annu Rev Immunol 27: 363–91.
43. Guccione E, Bassi C, Casadio F, Martinato F, Cesaroni M, Schuchlautz H, Luscher B and Amati B (2007). Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. Nature 449(7164): 933–7.
44. Guenther MG, Levine SS, Boyer LA, Jaenisch R and Young RA (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130(1): 77–88.
45. Guerau-de-Arellano M, Mathis D and Benoist C (2008). Transcriptional impact of Aire varies with cell type. Proc Natl Acad Sci U S A 105(37): 14011–6.
46. Halonen M, Kangas H, Ruppell T, Ilmarinen T, Ollila J, Kolmer M, Vihinen M, Palvimo J, Saarela J, Ulmanen I and Eskelin P (2004). APECED-causing mutations in AIRE reveal the functional domains of the protein. Hum Mutat 23(3): 245–57.
47. Halonen M, Pelto-Huikko M, Eskelin P, Peltonen L, Ulmanen I and Kolmer M (2001). Subcellular location and expression pattern of autoimmune regulator (Aire), the mouse orthologue for human gene defective in autoimmune polyendo-crinopathy candidiasis ectodermal dystrophy (APECED). J Histochem Cytochem 49(2): 197–208.
48. Hancock R (2004). Internal organisation of the nucleus: assembly of compartments by macromolecular crowding and the nuclear matrix model. Biol Cell 96(8): 595–601.

49. Happel N and Doenecke D (2009). Histone H1 and its isoforms: contribution to chromatin structure and function. Gene 431(1–2): 1–12.
50. Heino M, Peterson P, Kudoh J, Nagamine K, Lagerstedt A, Ovod V, Ranki A, Rantala I, Nieminen M, Tuukkanen J, Scott HS, Antonarakis SE, Shimizu N and Krohn K (1999). Autoimmune regulator is expressed in the cells regulating immune tolerance in thymus medulla. Biochem Biophys Res Commun 257(3): 821–5.
51. Hung T, Binda O, Champagne KS, Kuo AJ, Johnson K, Chang HY, Simon MD, Kutateladze TG and Gozani O (2009). ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation. Mol Cell 33(2): 248–56.
52. Husebye ES, Perheentupa J, Rautemaa R and Kampe O (2009). Clinical manifestations and management of patients with autoimmune polyendocrine syndrome type I. J Intern Med 265(5): 514–29.
53. Ilmarinen T, Kangas H, Kytomaa T, Eskelin P, Saharinen J, Seeler JS, Tanhuanpaa K, Chan FY, Slattery RM, Alakurtti K, Palvimo JJ and Ulmanen I (2008). Functional interaction of AIRE with PIAS1 in transcriptional regulation. Mol Immunol 45(7): 1847–62.
54. Isaac A, Wilcox KW and Taylor JL (2006). SP100B, a repressor of gene expression preferentially binds to DNA with unmethylated CpGs. J Cell Biochem 98(5): 1106–22.
55. Ivanov AV, Peng H, Yurchenko V, Yap KL, Negorev DG, Schultz DC, Psulkowski E, Fredericks WJ, White DE, Maul GG, Sadofsky MJ, Zhou MM and Rauscher FJ, 3rd (2007). PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. Mol Cell 28(5): 823–37.
56. Iwase S, Lan F, Bayliss P, de la Torre-Ubieta L, Huarte M, Qi HH, Whetstine JR, Bonni A, Roberts TM and Shi Y (2007). The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. Cell 128(6): 1077–88.
57. Jacobson RH, Ladurner AG, King DS and Tjian R (2000). Structure and function of a human TAFII250 double bromodomain module. Science 288(5470): 1422–5.
58. Jenuwein T and Allis CD (2001). Translating the histone code. Science 293(5532): 1074–80.
59. Ji X, Li W, Song J, Wei L and Liu XS (2006). CEAS: cis-regulatory element annotation system. Nucleic Acids Res 34(Web Server issue): W551–4.
60. Jiang C and Pugh BF (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet 10(3): 161–72.
61. Johnnidis JB, Venanzi ES, Taxman DJ, Ting JP, Benoist CO and Mathis DJ (2005). Chromosomal clustering of genes controlled by the aire transcription factor. Proc Natl Acad Sci U S A 102(20): 7233–8.
62. Karagianni P, Amazit L, Qin J and Wong J (2008). ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. Mol Cell Biol 28(2): 705–17.
63. Kirmizis A, Santos-Rosa H, Penkett CJ, Singer MA, Vermeulen M, Mann M, Bahler J, Green RD and Kouzarides T (2007). Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. Nature 449(7164): 928–32.
64. Klein L, Hinterberger M, Wirnsberger G and Kyewski B (2009). Antigen presentation in the thymus for positive selection and central tolerance induction. Nat Rev Immunol 9(12): 833–44.

65. Koh AS, Kuo AJ, Park SY, Cheung P, Abramson J, Bua D, Carney D, Shoelson SE, Gozani O, Kingston RE, Benoist C and Mathis D (2008). Aire employs a histone-binding module to mediate immunological tolerance, linking chromatin regulation with organ-specific autoimmunity. Proc Natl Acad Sci U S A 105(41): 15878–83.
66. Kouzarides T (2007). Chromatin modifications and their function. Cell 128(4): 693–705.
67. Kumar PG, Laloraya M, Wang CY, Ruan QG, Davoodi-Semiromi A, Kao KJ and She JX (2001). The autoimmune regulator (AIRE) is a DNA-binding protein. J Biol Chem 276(44): 41357–64.
68. Kuroda N, Mitani T, Takeda N, Ishimaru N, Arakaki R, Hayashi Y, Bando Y, Izumi K, Takahashi T, Nomura T, Sakaguchi S, Ueno T, Takahama Y, Uchida D, Sun S, Kajiura F, Mouri Y, Han H, Matsushima A, Yamada G and Matsumoto M (2005). Development of autoimmunity against transcriptionally unrepressed target antigen in the thymus of Aire-deficient mice. J Immunol 174(4): 1862–70.
69. Kyewski B and Klein L (2006). A central role for central tolerance. Annu Rev Immunol 24: 571–606.
70. Kyewski B and Peterson P (2010). Aire, master of many trades. Cell 140(1): 24–6.
71. Lan F, Collins RE, De Cegli R, Alpatov R, Horton JR, Shi X, Gozani O, Cheng X and Shi Y (2007). Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. Nature 448(7154): 718–22.
72. Lange M, Kaynak B, Forster UB, Tonjes M, Fischer JJ, Grimm C, Schlesinger J, Just S, Dunkel I, Krueger T, Mebus S, Lehrach H, Lurz R, Gobom J, Rottbauer W, Abdelilah-Seyfried S and Sperling S (2008). Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. Genes Dev 22(17): 2370–84.
73. Lee MS, Edwards RA, Thede GL and Glover JN (2005). Structure of the BRCT repeat domain of MDC1 and its specificity for the free COOH-terminal end of the gamma-H2AX histone tail. J Biol Chem 280(37): 32053–6.
74. Lee WY, Lee D, Chung WI and Kwon CS (2009). Arabidopsis ING and Alfin1-like protein families localize to the nucleus and bind to H3K4me3/2 via plant homeodomain fingers. Plant J 58(3): 511–24.
75. Lehman HK and Ballow M (2008). Immune deficiency disorders with auto-immunity and abnormalities in immune regulation-monogenic autoimmune diseases. Clin Rev Allergy Immunol 34(2): 141–5.
76. Li B, Carey M and Workman JL (2007). The role of chromatin during transcription. Cell 128(4): 707–19.
77. Li B, Gogol M, Carey M, Lee D, Seidel C and Workman JL (2007). Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. Science 316(5827): 1050–4.
78. Li H, Fischle W, Wang W, Duncan EM, Liang L, Murakami-Ishibe S, Allis CD and Patel DJ (2007). Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. Mol Cell 28(4): 677–91.
79. Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD and Patel DJ (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature 442(7098): 91–5.
80. Liiv I, Rebane A, Org T, Saare M, Maslovskaja J, Kisand K, Juronen E, Valmu L, Bottomley MJ, Kalkkinen N and Peterson P (2008). DNA-PK contributes to the

phosphorylation of AIRE: importance in transcriptional activity. Biochim Biophys Acta 1783(1): 74–83.
81. Liston A, Lesage S, Wilson J, Peltonen L and Goodnow CC (2003). Aire regulates negative selection of organ-specific T cells. Nat Immunol 4(4): 350–4.
82. Liu X, Yu X, Zack DJ, Zhu H and Qian J (2008). TiGER: a database for tissue-specific gene expression and regulation. BMC Bioinformatics 9: 271.
83. Liu Y, Subrahmanyam R, Chakraborty T, Sen R and Desiderio S (2007). A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. Immunity 27(4): 561–71.
84. Lo WS, Trievel RC, Rojas JR, Duggan L, Hsu JY, Allis CD, Marmorstein R and Berger SL (2000). Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. Mol Cell 5(6): 917–26.
85. Macdonald N, Welburn JP, Noble ME, Nguyen A, Yaffe MB, Clynes D, Moggs JG, Orphanides G, Thomson S, Edmunds JW, Clayton AL, Endicott JA and Mahadevan LC (2005). Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14–3–3. Mol Cell 20(2): 199–211.
86. Marmorstein R and Trievel RC (2009). Histone modifying enzymes: structures, mechanisms, and specificities. Biochim Biophys Acta 1789(1): 58–68.
87. Martin DG, Baetz K, Shi X, Walter KL, MacDonald VE, Wlodarski MJ, Gozani O, Hieter P and Howe L (2006). The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3. Mol Cell Biol 26(21): 7871–9.
88. Mathis D and Benoist C (2009). Aire. Annu Rev Immunol 27: 287–312.
89. Matthews AG, Kuo AJ, Ramon-Maiques S, Han S, Champagne KS, Ivanov D, Gallardo M, Carney D, Cheung P, Ciccone DN, Walter KL, Utz PJ, Shi Y, Kutateladze TG, Yang W, Gozani O and Oettinger MA (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature 450(7172): 1106–10.
90. Mauldin SK, Getts RC, Liu W and Stamato TD (2002). DNA-PK-dependent binding of DNA ends to plasmids containing nuclear matrix attachment region DNA sequences: evidence for assembly of a repair complex. Nucleic Acids Res 30(18): 4075–87.
91. Meager A, Visvalingam K, Peterson P, Moll K, Murumagi A, Krohn K, Eskelin P, Perheentupa J, Husebye E, Kadota Y and Willcox N (2006). Anti-interferon autoantibodies in autoimmune polyendocrinopathy syndrome type 1. PLoS Med 3(7): e289.
92. Meloni A, Incani F, Corda D, Cao A and Rosatelli MC (2008). Role of PHD fingers and COOH-terminal 30 amino acids in AIRE transactivation activity. Mol Immunol 45(3): 805–9.
93. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES and Bernstein BE (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448(7153): 553–60.
94. Mittaz L, Rossier C, Heino M, Peterson P, Krohn KJ, Gos A, Morris MA, Kudoh J, Shimizu N, Antonarakis SE and Scott HS (1999). Isolation and characterization of the mouse Aire gene. Biochem Biophys Res Commun 255(2): 483–90.

95. Moore MJ and Proudfoot NJ (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell 136(4): 688–700.
96. Mueller DL (2010). Mechanisms maintaining peripheral tolerance. Nat Immunol 11(1): 21–7.
97. Mujtaba S, Zeng L and Zhou MM (2007). Structure and acetyl-lysine recognition of the bromodomain. Oncogene 26(37): 5521–7.
98. Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J and Adelman K (2007). RNA polymerase is poised for activation across the genome. Nat Genet 39(12): 1507–11.
99. Musselman CA, Mansfield RE, Garske AL, Davrazou F, Kwan AH, Oliver SS, Oleary H, Denu JM, Mackay JP and Kutateladze TG (2009). Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. Biochem J.
100. Nagamine K, Peterson P, Scott HS, Kudoh J, Minoshima S, Heino M, Krohn KJ, Lalioti MD, Mullis PE, Antonarakis SE, Kawasaki K, Asakawa S, Ito F and Shimizu N (1997). Positional cloning of the APECED gene. Nat Genet 17(4): 393–8.
101. Nitta T, Murata S, Ueno T, Tanaka K and Takahama Y (2008). Thymic microenvironments for T-cell repertoire formation. Adv Immunol 99: 59–94.
102. Okubo S, Hara F, Tsuchida Y, Shimotakahara S, Suzuki S, Hatanaka H, Yokoyama S, Tanaka H, Yasuda H and Shindo H (2004). NMR structure of the N-terminal domain of SUMO ligase PIAS1 and its interaction with tumor suppressor p53 and A/T-rich DNA oligomers. J Biol Chem 279(30): 31455–61.
103. Owen DJ, Ornaghi P, Yang JC, Lowe N, Evans PR, Ballario P, Neuhaus D, Filetici P and Travers AA (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. Embo J 19(22): 6141–9.
104. Oven I, Brdickova N, Kohoutek J, Vaupotic T, Narat M and Peterlin BM (2007). AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. Mol Cell Biol 27(24): 8815–23.
105. Palacios A, Garcia P, Padro D, Lopez-Hernandez E, Martin I and Blanco FJ (2006). Solution structure and NMR characterization of the binding to methylated histone tails of the plant homeodomain finger of the tumour suppressor ING4. FEBS Lett 580(30): 6903–8.
106. Palacios A, Munoz IG, Pantoja-Uceda D, Marcaida MJ, Torres D, Martin-Garcia JM, Luque I, Montoya G and Blanco FJ (2008). Molecular basis of histone H3K4me3 recognition by ING4. J Biol Chem 283(23): 15956–64.
107. Pena PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R and Kutateladze TG (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature 442(7098): 100–3.
108. Pena PV, Hom RA, Hung T, Lin H, Kuo AJ, Wong RP, Subach OM, Champagne KS, Zhao R, Verkhusha VV, Li G, Gozani O and Kutateladze TG (2008). Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. J Mol Biol 380(2): 303–12.
109. Pena PV, Musselman CA, Kuo AJ, Gozani O and Kutateladze TG (2009). NMR assignments and histone specificity of the ING2 PHD finger. Magn Reson Chem 47(4): 352–8.
110. Peterlin BM and Price DH (2006). Controlling the elongation phase of transcription with P-TEFb. Mol Cell 23(3): 297–305.
111. Peterson P, Org T and Rebane A (2008). Transcriptional regulation by AIRE: molecular mechanisms of central tolerance. Nat Rev Immunol 8(12): 948–57.

112. Pitkanen J, Doucas V, Sternsdorf T, Nakajima T, Aratani S, Jensen K, Will H, Vahamurto P, Ollila J, Vihinen M, Scott HS, Antonarakis SE, Kudoh J, Shimizu N, Krohn K and Peterson P (2000). The autoimmune regulator protein has transcriptional transactivating properties and interacts with the common coactivator CREB-binding protein. J Biol Chem 275(22): 16802–9.
113. Pitkanen J, Rebane A, Rowell J, Murumagi A, Strobel P, Moll K, Saare M, Heikkila J, Doucas V, Marx A and Peterson P (2005). Cooperative activation of transcription by autoimmune regulator AIRE and CBP. Biochem Biophys Res Commun 333(3): 944–53.
114. Pitkanen J, Vahamurto P, Krohn K and Peterson P (2001). Subcellular localization of the autoimmune regulator protein. characterization of nuclear targeting and transcriptional activation domain. J Biol Chem 276(22): 19597–602.
115. Purohit S, Kumar PG, Laloraya M and She JX (2005). Mapping DNA-binding domains of the autoimmune regulator protein. Biochem Biophys Res Commun 327(3): 939–44.
116. Ragvin A, Valvatne H, Erdal S, Arskog V, Tufteland KR, Breen K, AM OY, Eberharter A, Gibson TJ, Becker PB and Aasland R (2004). Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300. J Mol Biol 337(4): 773–88.
117. Ramon-Maiques S, Kuo AJ, Carney D, Matthews AG, Oettinger MA, Gozani O and Yang W (2007). The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. Proc Natl Acad Sci U S A 104(48): 18993–8.
118. Ramsey C, Winqvist O, Puhakka L, Halonen M, Moro A, Kampe O, Eskelin P, Pelto-Huikko M and Peltonen L (2002). Aire deficient mice develop multiple features of APECED phenotype and show altered immune response. Hum Mol Genet 11(4): 397–409.
119. Reymann S and Borlak J (2008). Topoisomerase II inhibition involves characteristic chromosomal expression patterns. BMC Genomics 9: 324.
120. Ricke RM and Bielinsky AK (2005). Easy detection of chromatin binding proteins by the Histone Association Assay. Biol Proced Online 7: 60–9.
121. Rosatelli MC, Meloni A, Meloni A, Devoto M, Cao A, Scott HS, Peterson P, Heino M, Krohn KJ, Nagamine K, Kudoh J, Shimizu N and Antonarakis SE (1998). A common mutation in Sardinian autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy patients. Hum Genet 103(4): 428–34.
122. Rowat AC, Lammerding J, Herrmann H and Aebi U (2008). Towards an integrated understanding of the structure and mechanics of the cell nucleus. Bioessays 30(3): 226–36.
123. Ruan QG, Tung K, Eisenman D, Setiady Y, Eckenrode S, Yi B, Purohit S, Zheng WP, Zhang Y, Peltonen L and She JX (2007). The autoimmune regulator directly controls the expression of genes critical for thymic epithelial function. J Immunol 178(11): 7173–80.
124. Ruthenburg AJ, Wang W, Graybosch DM, Li H, Allis CD, Patel DJ and Verdine GL (2006). Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. Nat Struct Mol Biol 13(8): 704–12.
125. Saksouk N, Avvakumov N, Champagne KS, Hung T, Doyon Y, Cayrou C, Paquet E, Ullah M, Landry AJ, Cote V, Yang XJ, Gozani O, Kutateladze TG and Cote J (2009). HBO1 HAT complexes target chromatin throughout gene coding regions via multiple PHD finger interactions with histone H3 tail. Mol Cell 33(2): 257–65.

126. Scheel H and Hofmann K (2003). No evidence for PHD fingers as ubiquitin ligases. Trends Cell Biol 13(6): 285–7; author reply 7–8.
127. Schuetz A, Allali-Hassani A, Martin F, Loppnau P, Vedadi M, Bochkarev A, Plotnikov AN, Arrowsmith CH and Min J (2006). Structural basis for molecular recognition and presentation of histone H3 by WDR5. Embo J 25(18): 4245–52.
128. Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, Carney D, Pena P, Lan F, Kaadige MR, Lacoste N, Cayrou C, Davrazou F, Saha A, Cairns BR, Ayer DE, Kutateladze TG, Shi Y, Cote J, Chua KF and Gozani O (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature 442(7098): 96–9.
129. Shi X, Kachirskaia I, Walter KL, Kuo JH, Lake A, Davrazou F, Chan SM, Martin DG, Fingerman IM, Briggs SD, Howe L, Utz PJ, Kutateladze TG, Lugovskoy AA, Bedford MT and Gozani O (2007). Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. J Biol Chem 282(4): 2450–5.
130. Stucki M, Clapperton JA, Mohammad D, Yaffe MB, Smerdon SJ and Jackson SP (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell 123(7): 1213–26.
131. Zlotogora J and Shapiro MS (1992). Polyglandular autoimmune syndrome type I among Iranian Jews. J Med Genet 29(11): 824–6.
132. Tanaka Y, Tawaramoto-Sasanuma M, Kawaguchi S, Ohta T, Yoda K, Kurumizaka H and Yokoyama S (2004). Expression and purification of recombinant human histones. Methods 33(1): 3–11.
133. Tao Y, Kupfer R, Stewart BJ, Williams-Skipp C, Crowell CK, Patel DD, Sain S and Scheinman RI (2006). AIRE recruits multiple transcriptional components to specific genomic regions through tethering to nuclear matrix. Mol Immunol 43(4): 335–45.
134. Taubert R, Schwendemann J and Kyewski B (2007). Highly variable expression of tissue-restricted self-antigens in human thymus: implications for self-tolerance and autoimmunity. Eur J Immunol 37(3): 838–48.
135. Taunton J, Hassig CA and Schreiber SL (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science 272(5260): 408–11.
136. Taverna SD, Ilin S, Rogers RS, Tanny JC, Lavender H, Li H, Baker L, Boyle J, Blair LP, Chait BT, Patel DJ, Aitchison JD, Tackett AJ and Allis CD (2006). Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. Mol Cell 24(5): 785–96.
137. Taverna SD, Li H, Ruthenburg AJ, Allis CD and Patel DJ (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol 14(11): 1025–40.
138. Trojer P, Li G, Sims RJ, 3rd, Vaquero A, Kalakonda N, Boccuni P, Lee D, Erdjument-Bromage H, Tempst P, Nimer SD, Wang YH and Reinberg D (2007). L3MBTL1, a histone-methylation-dependent chromatin lock. Cell 129(5): 915–28.
139. Tykocinski LO, Sinemus A and Kyewski B (2008). The thymus medulla slowly yields its secrets. Ann N Y Acad Sci 1143: 105–22.
140. Uchida D, Hatakeyama S, Matsushima A, Han H, Ishido S, Hotta H, Kudoh J, Shimizu N, Doucas V, Nakayama KI, Kuroda N and Matsumoto M (2004). AIRE functions as an E3 ubiquitin ligase. J Exp Med 199(2): 167–72.
141. van Ingen H, van Schaik FM, Wienk H, Ballering J, Rehmann H, Dechesne AC, Kruijzer JA, Liskamp RM, Timmers HT and Boelens R (2008). Structural insight

into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. Structure 16(8): 1245–56.
142. Weake VM and Workman JL (2008). Histone ubiquitination: triggering gene activity. Mol Cell 29(6): 653–63.
143. Venanzi ES, Melamed R, Mathis D and Benoist C (2008). The variable immunological self: genetic variation and nongenetic noise in Aire-regulated transcription. Proc Natl Acad Sci U S A 105(41): 15860–5.
144. Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M and Timmers HT (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell 131(1): 58–69.
145. West KL, Singha NC, De Ioannes P, Lacomis L, Erdjument-Bromage H, Tempst P and Cortes P (2005). A direct interaction between the RAG2 C terminus and the core histones is required for efficient V(D)J recombination. Immunity 23(2): 203–12.
146. Villasenor J, Besse W, Benoist C and Mathis D (2008). Ectopic expression of peripheral-tissue antigens in the thymic epithelium: probabilistic, monoallelic, misinitiated. Proc Natl Acad Sci U S A 105(41): 15854–9.
147. Wilson DR, Norton DD and Fugmann SD (2008). The PHD domain of the sea urchin RAG2 homolog, SpRAG2L, recognizes dimethylated lysine 4 in histone H3 tails. Dev Comp Immunol 32(10): 1221–30.
148. Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P, Wu C and Allis CD (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature 442(7098): 86–90.
149. Yano M, Kuroda N, Han H, Meguro-Horike M, Nishikawa Y, Kiyonari H, Maemura K, Yanagawa Y, Obata K, Takahashi S, Ikawa T, Satoh R, Kawamoto H, Mouri Y and Matsumoto M (2008). Aire controls the differentiation program of thymic epithelial cells in the medulla for the establishment of self-tolerance. J Exp Med 205(12): 2827–38.

SUMMARY IN ESTONIAN

Autoimmuunregulaator valgu esimese PHD domeeni molekulaarne funktsioon

Normaalselt funktsioneeriv immuunsüsteem on võimeline organismi kaitsma mitmesuguste haigust tekitavate patogeenide eest ja samal ajal mitte ründama keha enda koostisosasid. Sellist immuunsüsteemi omadust mitte reageerida keha enda koostisosadega nimetatakse immunoloogiliseks tolerantsiks ja vead selle kujunemisel võivad viia autoimmuunhaiguste tekkimiseni. Immunoloogilise tolerantsi oluliseks kujunemiskohaks on tüümus, kus toimub T-lüfotsüütide ehk T-rakkude areng. T-rakkude arengu käigus toimub T-raku retseptori geeni segmentide juhuslik ümberkorraldamine. Selle tulemusena tekib tohutul hulgal erinevad T-raku retseptoreid, mis on võimelised ära tundma erinevaid MHC molekulidega kompleksis olevaid peptiide. Sellega tagatakse, et organism oleks võimeline ära tundma võimalikult suurt hulka võõrvalke. Kuna aga T-raku retseptorite kujunemine on juhuslik, siis tekib ka palju selliseid T-raku retseptoreid (TCR), mis on võimelised ära tundma keha enda valke ja T-rakud, mis selliseid TCR kannavad on organismile ohtlikud. Et autoreaktiivsed T-rakud ei pääseks organismis kahju tekitama toimub tüümuses arenevate autoreaktiivsete T-rakkude selektiivne elimineerimine, mida nimetatakse negatiivseks selektsiooniks. Selles protsessis osalevad antigeene presenteerivad rakud, dendriitrakud ja tüümuse medullaarsed epiteelirakud (mTECid). Nendest eriti olulised on mTECid, kuna need rakud on võimelised ekspresseerima suurel hulgal geene, mis muidu on ekspresseeritud ainult teatud spetsiifilistes kudedes. Nii näiteks on mTECides ekspresseerunud insuliin, mida muidu ekspresseerivad spetsiifiliselt ainult pankrease Langerhansi saarekestes olevad beetarakud. Kasutades MHC molekule presenteerivad mTECid arenevatele T-rakkudele erinevaid kehaomaseid peptiide ja need T-rakud, mille TCR seostub MHC-peptiid kompleksiga suure afiinsusega suunatakse apoptoosi. Immunoloogilise tolerantsi tagamiseks on seetõttu äärmiselt oluline mTECides aset leidev koespetsiifiliste geenide ekspressioon, mida nimetatakse ka avatud geenidekspressiooniks.

On näidatud, et suure osa koespetsiifiliste geenide ekspressiooni eest tüümuses vastutab AIRE (Autoimmuunregulaator) valk. Kui AIRE valk on vigane, siis suurt hulka koespetsiifilisi geene ei ekspresseerita ja T-rakkude negatiivne selektsioon ei toimu korralikult, mille tõttu pääsevad autoreaktiivsed T-rakud tüümusest välja ja põhjustavad autoimmuunseid reaktsioone. Mutatsioonid *AIRE* geenis põhjustavad inimesel autoimmuunset retsessiivset sündroomi nimega APECED (Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy). APECED patsientidel on häirunud immunoloogiline tolerants, mille tõttu tekivad autoimmuunsed reaktsioonid erinevate, kõige sagedasemini endokriin organite vastu. *Aire* geeni väljalülitamine hiirel põhjustab samuti autoimmuunset fenotüüpi, mida iseloomustavad autoantikehad ja lümfotsüütide infiltraadid erinevates organites. Kuigi on näidatud, et AIRE valk on oluline erinevate koespetsiifiliste geenide ekspressiooniks tüümuses, ei ole mehhanismid, kuidas see molekulaarsel tasemel toimub kuigi hästi teada.

Selleks, et geenidekspressiooni reguleerida, peab vastav geen olema transkriptsiooni läbiviidavatele valkudele kättesaadav. Raku tuumas olev DNA ei ole vaba, vaid on kokku pakitud spetsiifiliste valkude poolt, mida nimetatakse histoonideks. Histooni valgud moodustavad nukleosoomi, mille ümber DNA on keritud. Histooni valkude N-terminaalsed otsad ulatuvad nukleosoomist välja ja teatud aminohapetele nendes on võimalik ensüümide abiga lisada erinevaid modifikatsioone, millel on erinev funktsioon. Näiteks aktiivsetel geenidel on histoonil H3 trimetüleeritud neljandas positsioonis asuv aminohape lüsiin (H3K4me3), samas kui inaktiivsetel geenidel see märgis puudub ja trimetüleeritud on hoopis lüsiin 27. Erinevate rakusiseste protsesside reguleerimisel on olulised valgud, mis on võimelised läbi oma spetsiifiliste domeenide teatud histooni modifikatsioone ära tundma ja seeläbi selektiivselt teatud kromatiini piirkondadega interakteeruma. On välja pakutud, et selline histoonide posttranslatsiooniline modifitseerimine moodustab omaette „histooni koodi“, millel on väga oluline roll erinevate rakusiseste protsesside reguleerimisel.

Käesolev dissertatsioon koosneb töödest milles on uuritud, kuidas AIRE valk tõstab just koespetsiifiliste geenide ekspressiooni tüümuses. Töös on keskendutud AIRE valgu esimese PHD domeeni (AIRE-PHD1) funktsiooni uurimisele, kuna hiljuti näidati, et antud domeenid teistes valkudes on võimelised ära tundma histoon H3 teatud trimetüleeritud lüsiine.

Oma töödes näitasime, et sarnaselt teistele PHD domeene sisaldavatele valkudele on AIRE valk võimeline läbi oma PHD1 domeeni interakteeruma histoon H3-ga *in vitro* ja et erinevad N-terminaalsed posttranslatsioonilised modifikatsioonid nagu arginiin 2 dimetüleerimine (H3R2me2) ja lüsiin 4 trimetüleerimine segavad seda seondumist. Seetõttu interakteerub AIRE valk eelistatud histoon H3 N-terminaalsete otsadega, millel puuduvad posttranslatsioonilised modifikatsioonid. Sellise spetsiifikaga histoonidega seondumine AIRE valgu poolt on lisaks BHC80-le teine näide valkudest, mille PHD domeen seostub eelistatult just mittemodifitseeritud histoon H3-ga N-terminaalsete otsadega, mis näitab, et lisaks erinevatele modifikatsioonidele võib ka modifikatsioonide puudumine histoonidel olla spetsiifiliseks signaaliks.

Lisaks näitasime, et AIRE valk on võimeline interakteeruma mononukleosoomidega ja selektiivselt histoonidega millel puudub H3K4me3 *in vivo*. Selline võime interakteeruda just nende histoon H3 molekulidega, millel puudub aktiivsete geenide märgis H3K4me3 viitab võimalusele, et AIRE valk suunatakse spetsiifiliselt inaktiivsetele geenidele, milledest paljud on mTECides koespetsiifilised geenid. Seetõttu aktiveerib AIRE mTECides eelistatult just koespetsiifiliste geenide ekspressiooni, mis võimaldab läbi viia korrektset T-rakkude negatiivset selektsiooni, mis omakorda tagab tsentraalse immunoloogilise tolerantsi tekke.

Antud hüpoteesi edasiseks uurimiseks kasutasime mudelsüsteemi, milles AIRE valku ekspresseeritakse püsivalt HEK293 rakkudes. Kasutades kogu genoomi katvaid ekspressioonikiipe tuvastasime, et AIRE valk on võimeline HEK rakkudes aktiveerima suure hulga geenide ekspressiooni, millest paljud on koespetsiifilised ja algselt madala ekspressiooniga. AIRE märklaudgeenide korrektne aktivatsioon on sõltuv tervest PHD domeenist ja seega AIRE interaktsioonist histoonidega. Kasutades kromatiini immuunosadestamist ja

saadud materjali hübridiseerimist genoomsetele kiipidele näitasime, et AIRE märklaud geenide promootoritel puuduvad aktiivse kromatiini märgised H3K4me3 ja histoon H3 atsetüleerimine. AIRE poolt vahendatud aktivatsiooni järgselt toimuvad mitmetel tema märklaudgeenidel transkriptsiooni aktivatsiooniga kaasnevad muutused histoonide posttranslatsioonilistes modifikatsioonides ja RNA polümeraas II taseme tõus. See näitab, et AIRE vahendatud geenikspressiooni aktivatsioon toimub, vähemalt osaliselt, transkriptsioonilisel tasemel. Kuigi AIRE enda seondumine kromatiinile tundub olevat stohhastiline või ajas muutuv, mõjutavad AIRE poolset geenikspressiooni aktivatsiooni erinevad histoon H3 posttranslatsioonilised modifikatsioonid nagu H3K4me3 ja H3R2me2. Histoonide posttranslatsiooniliste modifikatsioone uurimine *in vivo* hiire mTECides ja perifeersetes kudedes kinnitas, et AIRE märklaudgeenide promootoritel tüümuses on madal H3K4me3 tase, olles seega kooskõlas meie välja pakutud hüpoteesiga.

Kokkuvõtteks võib öelda, et meie uurimistöö tulemused on viinud uue molekulaarse mehhanismi avastamiseni, kuidas AIRE eelistatult aktiveerib just koespetsiifilisi geene, mis on oluliseks osaks seletamaks mehhanisme, kuidas tüümuses tagatakse tsentraalne immunoloogiline tolerants.

ACKNOWLEDGEMENTS

This study was carried out at the Molecular Pathology study group, Department of General and Molecular Pathology, University of Tartu. There are many people who I would like to thank:

– My supervisor prof. Pärt Peterson for the support and guidance throughout the years. It has been a privilege to have such an excellent scientist for a supervisor.

– All the coauthors of the papers from our group and outside for their contribution, in particular Ana Rebane who has really helped me a lot!

– Arnold Kristjuhan and Anti Kalda for reviewing the dissertation, for the discussion and comments made.

– All our lab members for a great atmosphere; it has been a pleasure to work with you!

– My friends, especially Priit, Veiko, Sten, Reidar, Oliver, Viljo – the off lab time with you is always fun and has helped me to stay on track.

– My family, my parents and grandmother for their love and support and finally my dear wife Elin and my children, you make me complete!

PUBLICATIONS

CURRICULUM VITAE

Tõnis Org

Date and place of birth:	06.11.1979, Tartu, Estonia
Citizenship:	Estonian
Marital status:	Married, two children
Address:	Molecular Pathology, Institute of General and Molecular Pathology, University of Tartu, Ravila 19, 50411, Tartu, Estonia
Telephone:	+372 521 1484
e-mail:	toniso@ut.ee

Education

1995–1998	Hugo Treffner Grammar School
1998–2002	BSc in molecular diagnostics, Department of Biotechnology, Institute of Molecular and Cell Biology, Faculty of Biology and Geography, University of Tartu
2002–2004	MSc in molecular diagnostics, Department of Biotechnology, Institute of Molecular and Cell Biology , Faculty of Biology and Geography, University of Tartu
2004–	PhD student, Molecular Pathology, Institute of General and Molecular Pathology, Faculty of Medicine, University of Tartu

Professional employment

2002–2003	Asper Biotech
2003–2004	GSF isnstitute, Munich, Germany
2004–2009	University of Tartu, Institute of General and Molecular Pathology, specialist
2009–	University of Tartu, Institute of General and Molecular Pathology, scientist

Scientific work

Main research interests:

- Central tolerance induction in the thymus. Molecular mechanisms of AIRE mediated promiscuous gene expression of tissue-specific genes in the thymus.
- Seven scientific publications.

ELULOOKIRJELDUS

Tõnis Org

Sünniaeg ja koht:	06.11.1979, Tartu, Eesti
Kodakondsus:	Eesti
Perekonnaseis:	Abielus, kaks last
Aadress:	Molekulaarpatoloogia, Üld- ja Molekulaarpatoloogia Instituut, Tartu Ülikool, Ravila 19, 50411, Tartu, Eesti
Telefon:	+3725211484
e-mail:	toniso@ut.ee

Haridus

1995–1998	Hugo Treffneri Gümnaasium
1998–2002	BSc molekulaardiagnostika erialal, Biotehnoloogia õppetool, Molakulaar- ja Rakubioloogia Instituut, Bioloogia-Geograafia Teaduskond, Tartu Ülikool
2002–2004	MSc molekulaardiagnostika erialal, Biotehnoloogia õppetool, Molakulaar- ja Rakubioloogia Instituut, Bioloogia-Geograafia Teaduskond, Tartu Ülikool
2004–	Doktorantuur, Molekulaarpatoloogia, Üld- ja Molekulaarpatoloogia Instituut, Arstiteaduskond, Tartu Ülikool

Teenistuskäik

2002–2003	Asper Biotech
2003–2004	GSF instituut, Münhen, Saksamaa
2004–2009	Tartu Ülikool, Üld- ja Molekulaarpatoloogia Instituut, spetsialist
2009–	Tartu Ülikool, Üld- ja Molekulaarpatoloogia Instituut, teadur

Teadustöö

Peamised uurimisvaldkonnad:

- Tsentraalse tolerantsi tekkemehhanismid tüümuses. Molekulaarsed mehhanismid, kuidas AIRE valk reguleerib koespetsiifiliste antigeenide ekspressiooni tüümuses.
- Seitse avaldatud teaduspublikatsiooni.

DISSERTATIONES MEDICINAE UNIVERSITATIS TARTUENSIS

1. **Heidi-Ingrid Maaroos.** The natural course of gastric ulcer in connection with chronic gastritis and *Helicobacter pylori*. Tartu, 1991.
2. **Mihkel Zilmer.** Na-pump in normal and tumorous brain tissues: Structural, functional and tumorigenesis aspects. Tartu, 1991.
3. **Eero Vasar.** Role of cholecystokinin receptors in the regulation of behaviour and in the action of haloperidol and diazepam. Tartu, 1992.
4. **Tiina Talvik.** Hypoxic-ischaemic brain damage in neonates (clinical, biochemical and brain computed tomographical investigation). Tartu, 1992.
5. **Ants Peetsalu.** Vagotomy in duodenal ulcer disease: A study of gastric acidity, serum pepsinogen I, gastric mucosal histology and *Helicobacter pylori*. Tartu, 1992.
6. **Marika Mikelsaar.** Evaluation of the gastrointestinal microbial ecosystem in health and disease. Tartu, 1992.
7. **Hele Everaus.** Immuno-hormonal interactions in chronic lymphocytic leukaemia and multiple myeloma. Tartu, 1993.
8. **Ruth Mikelsaar.** Etiological factors of diseases in genetically consulted children and newborn screening: dissertation for the commencement of the degree of doctor of medical sciences. Tartu, 1993.
9. **Agu Tamm.** On metabolic action of intestinal microflora: clinical aspects. Tartu, 1993.
10. **Katrin Gross.** Multiple sclerosis in South-Estonia (epidemiological and computed tomographical investigations). Tartu, 1993.
11. **Oivi Uibo.** Childhood coeliac disease in Estonia: occurrence, screening, diagnosis and clinical characterization. Tartu, 1994.
12. **Viiu Tuulik.** The functional disorders of central nervous system of chemistry workers. Tartu, 1994.
13. **Margus Viigimaa.** Primary haemostasis, antiaggregative and anticoagulant treatment of acute myocardial infarction. Tartu, 1994.
14. **Rein Kolk.** Atrial versus ventricular pacing in patients with sick sinus syndrome. Tartu, 1994.
15. **Toomas Podar.** Incidence of childhood onset type 1 diabetes mellitus in Estonia. Tartu, 1994.
16. **Kiira Subi.** The laboratory surveillance of the acute respiratory viral infections in Estonia. Tartu, 1995.
17. **Irja Lutsar.** Infections of the central nervous system in children (epidemiologic, diagnostic and therapeutic aspects, long term outcome). Tartu, 1995.
18. **Aavo Lang.** The role of dopamine, 5-hydroxytryptamine, sigma and NMDA receptors in the action of antipsychotic drugs. Tartu, 1995.
19. **Andrus Arak.** Factors influencing the survival of patients after radical surgery for gastric cancer. Tartu, 1996.
20. **Tõnis Karki.** Quantitative composition of the human lactoflora and method for its examination. Tartu, 1996.

21. **Reet Mändar.** Vaginal microflora during pregnancy and its transmission to newborn. Tartu, 1996.
22. **Triin Remmel.** Primary biliary cirrhosis in Estonia: epidemiology, clinical characterization and prognostication of the course of the disease. Tartu, 1996.
23. **Toomas Kivastik.** Mechanisms of drug addiction: focus on positive reinforcing properties of morphine. Tartu, 1996.
24. **Paavo Pokk.** Stress due to sleep deprivation: focus on $GABA_A$ receptor-chloride ionophore complex. Tartu, 1996.
25. **Kristina Allikmets.** Renin system activity in essential hypertension. Associations with atherothrombogenic cardiovascular risk factors and with the efficacy of calcium antagonist treatment. Tartu, 1996.
26. **Triin Parik.** Oxidative stress in essential hypertension: Associations with metabolic disturbances and the effects of calcium antagonist treatment. Tartu, 1996.
27. **Svetlana Päi.** Factors promoting heterogeneity of the course of rheumatoid arthritis. Tartu, 1997.
28. **Maarike Sallo.** Studies on habitual physical activity and aerobic fitness in 4 to 10 years old children. Tartu, 1997.
29. **Paul Naaber.** *Clostridium difficile* infection and intestinal microbial ecology. Tartu, 1997.
30. **Rein Pähkla.** Studies in pinoline pharmacology. Tartu, 1997.
31. **Andrus Juhan Voitk.** Outpatient laparoscopic cholecystectomy. Tartu, 1997.
32. **Joel Starkopf.** Oxidative stress and ischaemia-reperfusion of the heart. Tartu, 1997.
33. **Janika Kõrv.** Incidence, case-fatality and outcome of stroke. Tartu, 1998.
34. **Ülla Linnamägi.** Changes in local cerebral blood flow and lipid peroxidation following lead exposure in experiment. Tartu, 1998.
35. **Ave Minajeva.** Sarcoplasmic reticulum function: comparison of atrial and ventricular myocardium. Tartu, 1998.
36. **Oleg Milenin.** Reconstruction of cervical part of esophagus by revascularised ileal autografts in dogs. A new complex multistage method. Tartu, 1998.
37. **Sergei Pakriev.** Prevalence of depression, harmful use of alcohol and alcohol dependence among rural population in Udmurtia. Tartu, 1998.
38. **Allen Kaasik.** Thyroid hormone control over β-adrenergic signalling system in rat atria. Tartu, 1998.
39. **Vallo Matto.** Pharmacological studies on anxiogenic and antiaggressive properties of antidepressants. Tartu, 1998.
40. **Maire Vasar.** Allergic diseases and bronchial hyperreactivity in Estonian children in relation to environmental influences. Tartu, 1998.
41. **Kaja Julge.** Humoral immune responses to allergens in early childhood. Tartu, 1998.
42. **Heli Grünberg.** The cardiovascular risk of Estonian schoolchildren. A cross-sectional study of 9-, 12- and 15-year-old children. Tartu, 1998.

43. **Epp Sepp.** Formation of intestinal microbial ecosystem in children. Tartu, 1998.
44. **Mai Ots.** Characteristics of the progression of human and experimental glomerulopathies. Tartu, 1998.
45. **Tiina Ristimäe.** Heart rate variability in patients with coronary artery disease. Tartu, 1998.
46. **Leho Kõiv.** Reaction of the sympatho-adrenal and hypothalamo-pituitary-adrenocortical system in the acute stage of head injury. Tartu, 1998.
47. **Bela Adojaan.** Immune and genetic factors of childhood onset IDDM in Estonia. An epidemiological study. Tartu, 1999.
48. **Jakov Shlik.** Psychophysiological effects of cholecystokinin in humans. Tartu, 1999.
49. **Kai Kisand.** Autoantibodies against dehydrogenases of α-ketoacids. Tartu, 1999.
50. **Toomas Marandi.** Drug treatment of depression in Estonia. Tartu, 1999.
51. **Ants Kask.** Behavioural studies on neuropeptide Y. Tartu, 1999.
52. **Ello-Rahel Karelson.** Modulation of adenylate cyclase activity in the rat hippocampus by neuropeptide galanin and its chimeric analogs. Tartu, 1999.
53. **Tanel Laisaar.** Treatment of pleural empyema — special reference to intrapleural therapy with streptokinase and surgical treatment modalities. Tartu, 1999.
54. **Eve Pihl.** Cardiovascular risk factors in middle-aged former athletes. Tartu, 1999.
55. **Katrin Õunap.** Phenylketonuria in Estonia: incidence, newborn screening, diagnosis, clinical characterization and genotype/phenotype correlation. Tartu, 1999.
56. **Siiri Kõljalg.** *Acinetobacter* – an important nosocomial pathogen. Tartu, 1999.
57. **Helle Karro.** Reproductive health and pregnancy outcome in Estonia: association with different factors. Tartu, 1999.
58. **Heili Varendi.** Behavioral effects observed in human newborns during exposure to naturally occurring odors. Tartu, 1999.
59. **Anneli Beilmann.** Epidemiology of epilepsy in children and adolescents in Estonia. Prevalence, incidence, and clinical characteristics. Tartu, 1999.
60. **Vallo Volke.** Pharmacological and biochemical studies on nitric oxide in the regulation of behaviour. Tartu, 1999.
61. **Pilvi Ilves.** Hypoxic-ischaemic encephalopathy in asphyxiated term infants. A prospective clinical, biochemical, ultrasonographical study. Tartu, 1999.
62. **Anti Kalda.** Oxygen-glucose deprivation-induced neuronal death and its pharmacological prevention in cerebellar granule cells. Tartu, 1999.
63. **Eve-Irene Lepist.** Oral peptide prodrugs – studies on stability and absorption. Tartu, 2000.
64. **Jana Kivastik.** Lung function in Estonian schoolchildren: relationship with anthropometric indices and respiratory symptomas, reference values for dynamic spirometry. Tartu, 2000.

65. **Karin Kull.** Inflammatory bowel disease: an immunogenetic study. Tartu, 2000.
66. **Kaire Innos.** Epidemiological resources in Estonia: data sources, their quality and feasibility of cohort studies. Tartu, 2000.
67. **Tamara Vorobjova.** Immune response to *Helicobacter pylori* and its association with dynamics of chronic gastritis and epithelial cell turnover in antrum and corpus. Tartu, 2001.
68. **Ruth Kalda.** Structure and outcome of family practice quality in the changing health care system of Estonia. Tartu, 2001.
69. **Annika Krüüner.** *Mycobacterium tuberculosis* – spread and drug resistance in Estonia. Tartu, 2001.
70. **Marlit Veldi.** Obstructive Sleep Apnoea: Computerized Endopharyngeal Myotonometry of the Soft Palate and Lingual Musculature. Tartu, 2001.
71. **Anneli Uusküla.** Epidemiology of sexually transmitted diseases in Estonia in 1990–2000. Tartu, 2001.
72. **Ade Kallas.** Characterization of antibodies to coagulation factor VIII. Tartu, 2002.
73. **Heidi Annuk.** Selection of medicinal plants and intestinal lactobacilli as antimicrobil components for functional foods. Tartu, 2002.
74. **Aet Lukmann**. Early rehabilitation of patients with ischaemic heart disease after surgical revascularization of the myocardium: assessment of health-related quality of life, cardiopulmonary reserve and oxidative stress. A clinical study. Tartu, 2002.
75. **Maigi Eisen.** Pathogenesis of Contact Dermatitis: participation of Oxidative Stress. A clinical – biochemical study. Tartu, 2002.
76. **Piret Hussar.** Histology of the post-traumatic bone repair in rats. Elaboration and use of a new standardized experimental model – bicortical perforation of tibia compared to internal fracture and resection osteotomy. Tartu, 2002.
77. **Tõnu Rätsep.** Aneurysmal subarachnoid haemorrhage: Noninvasive monitoring of cerebral haemodynamics. Tartu, 2002.
78. **Marju Herodes.** Quality of life of people with epilepsy in Estonia. Tartu, 2003.
79. **Katre Maasalu.** Changes in bone quality due to age and genetic disorders and their clinical expressions in Estonia. Tartu, 2003.
80. **Toomas Sillakivi.** Perforated peptic ulcer in Estonia: epidemiology, risk factors and relations with *Helicobacter pylori*. Tartu, 2003.
81. **Leena Puksa.** Late responses in motor nerve conduction studies. F and A waves in normal subjects and patients with neuropathies. Tartu, 2003.
82. **Krista Lõivukene**. *Helicobacter pylori* in gastric microbial ecology and its antimicrobial susceptibility pattern. Tartu, 2003.
83. **Helgi Kolk.** Dyspepsia and *Helicobacter pylori* infection: the diagnostic value of symptoms, treatment and follow-up of patients referred for upper gastrointestinal endoscopy by family physicians. Tartu, 2003.

84. **Helena Soomer.** Validation of identification and age estimation methods in forensic odontology. Tartu, 2003.
85. **Kersti Oselin.** Studies on the human MDR1, MRP1, and MRP2 ABC transporters: functional relevance of the genetic polymorphisms in the *MDR1* and *MRP1* gene. Tartu, 2003.
86. **Jaan Soplepmann.** Peptic ulcer haemorrhage in Estonia: epidemiology, prognostic factors, treatment and outcome. Tartu, 2003.
87. **Margot Peetsalu.** Long-term follow-up after vagotomy in duodenal ulcer disease: recurrent ulcer, changes in the function, morphology and *Helicobacter pylori* colonisation of the gastric mucosa. Tartu, 2003.
88. **Kersti Klaamas.** Humoral immune response to *Helicobacter pylori* a study of host-dependent and microbial factors. Tartu, 2003.
89. **Pille Taba.** Epidemiology of Parkinson's disease in Tartu, Estonia. Prevalence, incidence, clinical characteristics, and pharmacoepidemiology. Tartu, 2003.
90. **Alar Veraksitš**. Characterization of behavioural and biochemical phenotype of cholecystokinin-2 receptor deficient mice: changes in the function of the dopamine and endopioidergic system. Tartu, 2003.
91. **Ingrid Kalev.** CC-chemokine receptor 5 (CCR5) gene polymorphism in Estonians and in patients with Type I and Type II diabetes mellitus. Tartu, 2003.
92. **Lumme Kadaja.** Molecular approach to the regulation of mitochondrial function in oxidative muscle cells. Tartu, 2003.
93. **Aive Liigant**. Epidemiology of primary central nervous system tumours in Estonia from 1986 to 1996. Clinical characteristics, incidence, survival and prognostic factors. Tartu, 2004.
94. **Andres, Kulla.** Molecular characteristics of mesenchymal stroma in human astrocytic gliomas. Tartu, 2004.
95. **Mari Järvelaid.** Health damaging risk behaviours in adolescence. Tartu, 2004.
96. **Ülle Pechter.** Progression prevention strategies in chronic renal failure and hypertension. An experimental and clinical study. Tartu, 2004.
97. **Gunnar Tasa.** Polymorphic glutathione S-transferases – biology and role in modifying genetic susceptibility to senile cataract and primary open angle glaucoma. Tartu, 2004.
98. **Tuuli Käämbre.** Intracellular energetic unit: structural and functional aspects. Tartu, 2004.
99. **Vitali Vassiljev.** Influence of nitric oxide syntase inhibitors on the effects of ethanol after acute and chronic ethanol administration and withdrawal. Tartu, 2004.
100. **Aune Rehema.** Assessment of nonhaem ferrous iron and glutathione redox ratio as markers of pathogeneticity of oxidative stress in different clinical groups. Tartu, 2004.
101. **Evelin Seppet.** Interaction of mitochondria and ATPases in oxidative muscle cells in normal and pathological conditions. Tartu, 2004.

102. **Eduard Maron.** Serotonin function in panic disorder: from clinical experiments to brain imaging and genetics. Tartu, 2004.
103. **Marje Oona.** *Helicobacter pylori* infection in children: epidemiological and therapeutic aspects. Tartu, 2004.
104. **Kersti Kokk.** Regulation of active and passive molecular transport in the testis. Tartu, 2005.
105. **Vladimir Järv.** Cross-sectional imaging for pretreatment evaluation and follow-up of pelvic malignant tumours. Tartu, 2005.
106. **Andre Õun.** Epidemiology of adult epilepsy in Tartu, Estonia. Incidence, prevalence and medical treatment. Tartu, 2005.
107. **Piibe Muda.** Homocysteine and hypertension: associations between homocysteine and essential hypertension in treated and untreated hypertensive patients with and without coronary artery disease. Tartu, 2005.
108. **Külli Kingo.** The interleukin-10 family cytokines gene polymorphisms in plaque psoriasis. Tartu, 2005.
109. **Mati Merila.** Anatomy and clinical relevance of the glenohumeral joint capsule and ligaments. Tartu, 2005.
110. **Epp Songisepp**. Evaluation of technological and functional properties of the new probiotic *Lactobacillus fermentum* ME-3. Tartu, 2005.
111. **Tiia Ainla.** Acute myocardial infarction in Estonia: clinical characteristics, management and outcome. Tartu, 2005.
112. **Andres Sell.** Determining the minimum local anaesthetic requirements for hip replacement surgery under spinal anaesthesia – a study employing a spinal catheter. Tartu, 2005.
113. **Tiia Tamme.** Epidemiology of odontogenic tumours in Estonia. Pathogenesis and clinical behaviour of ameloblastoma. Tartu, 2005.
114. **Triine Annus**. Allergy in Estonian schoolchildren: time trends and characteristics. Tartu, 2005.
115. **Tiia Voor.** Microorganisms in infancy and development of allergy: comparison of Estonian and Swedish children. Tartu, 2005.
116. **Priit Kasenõmm.** Indicators for tonsillectomy in adults with recurrent tonsillitis – clinical, microbiological and pathomorphological investigations. Tartu, 2005.
117. **Eva Zusinaite.** Hepatitis C virus: genotype identification and interactions between viral proteases. Tartu, 2005.
118. **Piret Kõll.** Oral lactoflora in chronic periodontitis and periodontal health. Tartu, 2006.
119. **Tiina Stelmach.** Epidemiology of cerebral palsy and unfavourable neurodevelopmental outcome in child population of Tartu city and county, Estonia Prevalence, clinical features and risk factors. Tartu, 2006.
120. **Katrin Pudersell.** Tropane alkaloid production and riboflavine excretion in the field and tissue cultures of henbane (*Hyoscyamus niger* L.). Tartu, 2006.
121. **Külli Jaako.** Studies on the role of neurogenesis in brain plasticity. Tartu, 2006.

122. **Aare Märtson.** Lower limb lengthening: experimental studies of bone regeneration and long-term clinical results. Tartu, 2006.
123. **Heli Tähepõld.** Patient consultation in family medicine. Tartu, 2006.
124. **Stanislav Liskmann.** Peri-implant disease: pathogenesis, diagnosis and treatment in view of both inflammation and oxidative stress profiling. Tartu, 2006.
125. **Ruth Rudissaar.** Neuropharmacology of atypical antipsychotics and an animal model of psychosis. Tartu, 2006.
126. **Helena Andreson.** Diversity of *Helicobacter pylori* genotypes in Estonian patients with chronic inflammatory gastric diseases. Tartu, 2006.
127. **Katrin Pruus.** Mechanism of action of antidepressants: aspects of serotoninergic system and its interaction with glutamate. Tartu, 2006.
128. **Priit Põder.** Clinical and experimental investigation: relationship of ischaemia/reperfusion injury with oxidative stress in abdominal aortic aneurysm repair and in extracranial brain artery endarterectomy and possibilities of protection against ischaemia using a glutathione analogue in a rat model of global brain ischaemia. Tartu, 2006.
129. **Marika Tammaru.** Patient-reported outcome measurement in rheumatoid arthritis. Tartu, 2006.
130. **Tiia Reimand.** Down syndrome in Estonia. Tartu, 2006.
131. **Diva Eensoo.** Risk-taking in traffic and Markers of Risk-Taking Behaviour in Schoolchildren and Car Drivers. Tartu, 2007.
132. **Riina Vibo.** The third stroke registry in Tartu, Estonia from 2001 to 2003: incidence, case-fatality, risk factors and long-term outcome. Tartu, 2007.
133. **Chris Pruunsild.** Juvenile idiopathic arthritis in children in Estonia. Tartu, 2007.
134. **Eve Õiglane-Šlik.** Angelman and Prader-Willi syndromes in Estonia. Tartu, 2007.
135. **Kadri Haller.** Antibodies to follicle stimulating hormone. Significance in female infertility. Tartu, 2007.
136. **Pille Ööpik.** Management of depression in family medicine. Tartu, 2007.
137. **Jaak Kals.** Endothelial function and arterial stiffness in patients with atherosclerosis and in healthy subjects. Tartu, 2007.
138. **Priit Kampus.** Impact of inflammation, oxidative stress and age on arterial stiffness and carotid artery intima-media thickness. Tartu, 2007.
139. **Margus Punab.** Male fertility and its risk factors in Estonia. Tartu, 2007.
140. **Alar Toom**. Heterotopic ossification after total hip arthroplasty: clinical and pathogenetic investigation. Tartu, 2007.
141. **Lea Pehme.** Epidemiology of tuberculosis in Estonia 1991–2003 with special regard to extrapulmonary tuberculosis and delay in diagnosis of pulmonary tuberculosis. Tartu, 2007.
142. **Juri Karjagin.** The pharmacokinetics of metronidazole and meropenem in septic shock. Tartu, 2007.
143. **Inga Talvik.** Inflicted traumatic brain injury shaken baby syndrome in Estonia – epidemiology and outcome. Tartu, 2007.

144. **Tarvo Rajasalu.** Autoimmune diabetes: an immunological study of type 1 diabetes in humans and in a model of experimental diabetes (in RIP-B7.1 mice). Tartu, 2007.
145. **Inga Karu.** Ischaemia-reperfusion injury of the heart during coronary surgery: a clinical study investigating the effect of hyperoxia. Tartu, 2007.
146. **Peeter Padrik.** Renal cell carcinoma: Changes in natural history and treatment of metastatic disease. Tartu, 2007.
147. **Neve Vendt.** Iron deficiency and iron deficiency anaemia in infants aged 9 to 12 months in Estonia. Tartu, 2008.
148. **Lenne-Triin Heidmets.** The effects of neurotoxins on brain plasticity: focus on neural Cell Adhesion Molecule. Tartu, 2008.
149. **Paul Korrovits.** Asymptomatic inflammatory prostatitis: prevalence, etiological factors, diagnostic tools. Tartu, 2008.
150. **Annika Reintam.** Gastrointestinal failure in intensive care patients. Tartu, 2008.
151. **Kristiina Roots.** Cationic regulation of Na-pump in the normal, Alzheimer's and CCK_2 receptor-deficient brain. Tartu, 2008.
152. **Helen Puusepp.** The genetic causes of mental retardation in Estonia: fragile X syndrome and creatine transporter defect. Tartu, 2009.
153. **Kristiina Rull.** Human chorionic gonadotropin beta genes and recurrent miscarriage: expression and variation study. Tartu, 2009.
154. **Margus Eimre.** Organization of energy transfer and feedback regulation in oxidative muscle cells. Tartu, 2009.
155. **Maire Link.** Transcription factors FoxP3 and AIRE: autoantibody associations. Tartu, 2009.
156. **Kai Haldre.** Sexual health and behaviour of young women in Estonia. Tartu, 2009.
157. **Kaur Liivak.** Classical form of congenital adrenal hyperplasia due to 21-hydroxylase deficiency in Estonia: incidence, genotype and phenotype with special attention to short-term growth and 24-hour blood pressure. Tartu, 2009.
158. **Kersti Ehrlich.** Antioxidative glutathione analogues (UPF peptides) – molecular design, structure-activity relationships and testing the protective properties. Tartu, 2009.
159. **Anneli Rätsep.** Type 2 diabetes care in family medicine. Tartu, 2009.
160. **Silver Türk.** Etiopathogenetic aspects of chronic prostatitis: role of mycoplasmas, coryneform bacteria and oxidative stress. Tartu, 2009.
161. **Kaire Heilman.** Risk markers for cardiovascular disease and low bone mineral density in children with type 1 diabetes. Tartu, 2009.
162. **Kristi Rüütel.** HIV-epidemic in Estonia: injecting drug use and quality of life of people living with HIV. Tartu, 2009.
163. **Triin Eller.** Immune markers in major depression and in antidepressive treatment. Tartu, 2009.

164. **Siim Suutre.** The role of TGF-β isoforms and osteoprogenitor cells in the pathogenesis of heterotopic ossification. An experimental and clinical study of hip arthroplasty. Tartu, 2010.
165. **Kai Kliiman.** Highly drug-resistant tuberculosis in Estonia: Risk factors and predictors of poor treatment outcome. Tartu, 2010.
166. **Inga Villa.** Cardiovascular health-related nutrition, physical activity and fitness in Estonia. Tartu, 2010.