EPIGENETIC ANALYSIS OF PROMISCUOUS GENE EXPRESSION IN CENTRAL TOLERANCE

By

JENNIFER NOELLE HEATH

A thesis submitted to

The University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

School of Immunity & Infection
Institute of Biomedical Research
University of Birmingham
October 2009

ABSTRACT

The autoimmune regulator (AIRE), a key player in negative selection of developing thymocytes, acts as a transcriptional regulator, inducing the expression of tissue restricted antigens (TRA) within medullary thymic epithelial cells in a process known as promiscuous gene expression (PGE). Here we demonstrate how AIRE influences PGE through a direct impact on post-translational modifications of core histones which are associated with the regulation of transcription. Through native chromatin immunoprecipitation on thymic epithelial cells transfected with AIRE, we show how *in vitro*, TRA are enriched in active acetylation and methylation of core histones, yet retain silencing modifications. Furthermore, across a cluster of AIRE-regulated genes, histone modifications were deposited across the entire domain, dependent upon the expression profile of each gene, suggesting a role for domain-wide epigenetic regulation by AIRE. Extension of these studies *in vivo*, utilising the recently developed carrier ChIP technique, allowed examination of the epigenetic status of TRA throughout the thymic developmental pathway. We report how poised TRA are marked with combinations of active and silent modifications early in thymic development and that the chromatin signatures re-organise as the cells differentiate. The epigenetic patterning differs on a gene-by-gene basis, however their significance is implied upon disruption to normal development as the predictive pattern of modifications is lost.

For my fantastic parents, to whom I owe so much.

Without you I wouldn't be who I am today.

For Vicky, my wonderful sister, my greatest friend.

Thank you for always making me smile.

And for Ben.

Words are not enough.

I love you all past the sky.

ACKNOWLEDGEMENTS

I would like to express my eternal gratitude to my two fantastic PhD supervisors Dr Laura O'Neill and Prof. Graham Anderson whose continued support, both academically and emotionally, has kept me going through the past three years. I am especially indebted to Laura for all the guidance you have given me. Without your technical and theoretical brilliance, and friendship, none of this would have been possible.

I would also like to thank all members of the Anderson Lab, University of Birmingham, in particular Dr Andrea White for the isolation of primary cell populations from foetal thymic organ cultures, and Sonia Parnell for all your help with quantitative PCR. I am constantly amazed by your selflessness and willingness to lend a hand and for this I am truly grateful.

Thank you also to members of the Chromatin and Gene Expression Group, University of Birmingham, both past and present. I would never have survived without the daily assistance and encouragement from all of you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

As human beings, we are under relentless attack from our environment; constantly in danger of infection and disease caused by pathogenic organisms. On occasion, these invasions are successful, however they are dealt with swiftly and effectively by our immune system; a highly complex arrangement of cells and molecules, with a series of sophisticated mechanisms in place to identify and eradicate invaders. There are two major types of immune responses; innate immunity which provides an immediate reaction to tissue damage and infection by a wide range of potentially infectious agents; and adaptive immunity which is activated should innate immunity fail. The adaptive immune system is organised around two basic branches; humoural and cell-mediated responses, carried out by B-lymphocytes (B-cells) and T-lymphocytes (T-cells) respectively (Beck and Habicht 1996; Walker and Abbas 2002). It is predicted that the T-cell repertoire alone can provide protection against a vast number of foreign pathogens, with a capacity to recognise around 25 million antigen specificities (Walker and Abbas 2002). This is achieved through a combination of plasticity of antigen recognition and random somatic gene mutation of the T-cell receptor (TCR) genes, during the development of T-cells within the thymus.

1.1. T-Lymphocytes: the Main Players in Adaptive Immunity

Each day, approximately two million new T-cells are released from the thymus into the circulation, generated from an average of 10-100 entering haematopoietic precursors (Kyewski and Klein 2006; Nedjic *et al.* 2008). However, these two million

mature T-cells are the lucky ones; for at its peak, an average of 50 million thymocytes are produced within the thymus from the minimal number of precursor cells, all of which must go through an intense selection process, ultimately resulting in the death of 96% of developing cells daily (Nedjic *et al.* 2008). Each individual T-cell is encoded for a certain antigen-specificity on its TCR, thus the generation of such a large number of novel thymocytes requires great flexibility in TCR genes and random gene rearrangement. However, this extremely stochastic process unavoidably results in the generation of either inert or potentially damaging T-cells. Inert T-cells are simply of no use to the immune system as, although capable of recognising foreign antigens, they are unable to bind to them when presented via self-major histocompatibility complex (MHC) on the surface of antigen-presenting cells (Nedjic *et al.* 2008). The alternative; T-cells capable of binding self-MHC, but in combination with self-antigens originating from our body's own proteins, have the potential to cause much damage should they be released, and many autoimmune diseases result from a failure of the selection processes within the thymus allowing these self-reactive T-cells to slip through the net. Fortunately, healthy individuals are able to purge the system of these T-cells before they can do harm in a process known as tolerance, which can take place within the thymus and the periphery.

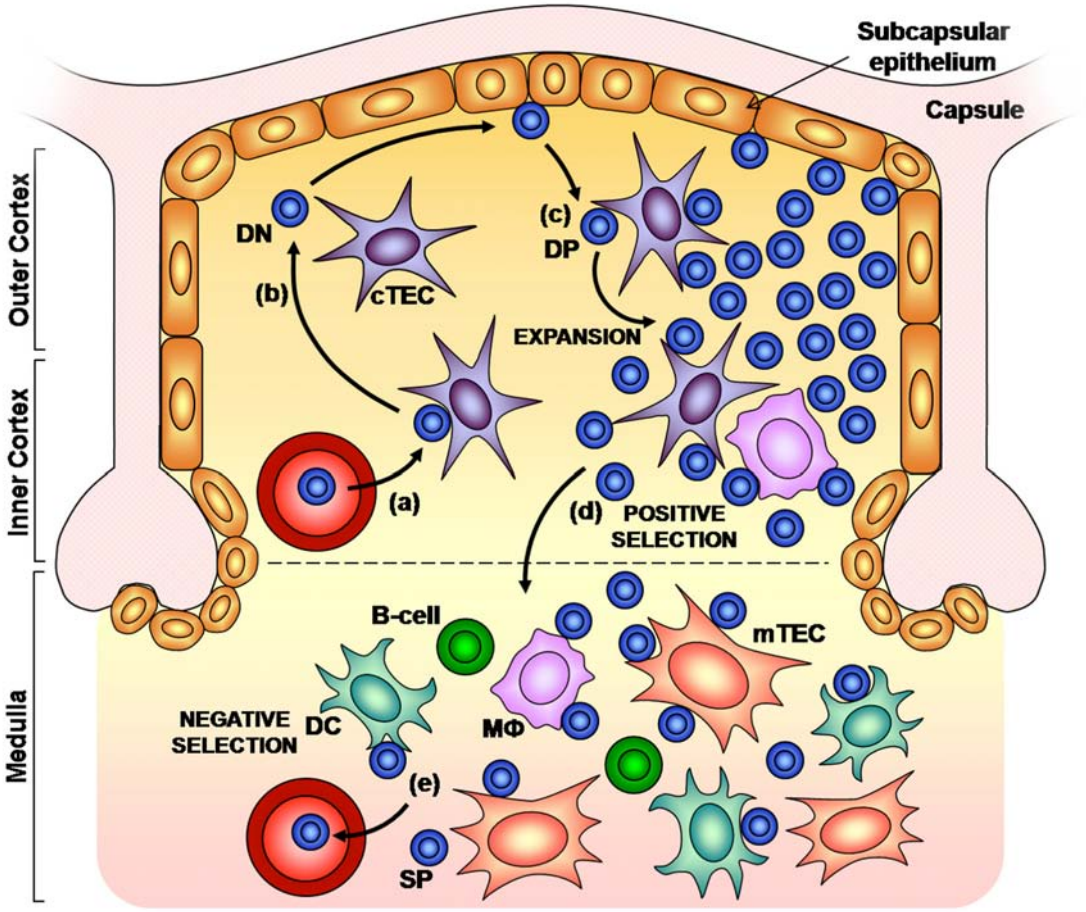
1.2. T-Lymphocyte Development and Central Tolerance

The process of central tolerance is an integral part of T-cell development within the thymus and is therefore crucial for the generation of an effective immune system. The thymus represents a specialised microenvironment whose primary function is to

generate a repertoire of diverse but safe T-cells. Stringent selection processes are orchestrated by thymic stromal cells and other accessory cells, which communicate with the developing T-cells as they pass through the thymus. The thymic cells signal through chemokines, cytokines, selectins and transcription factors, influencing thymocyte migration, proliferation and differentiation during the complex stages of T-cell development (Figure 1.1) (Rodewald 2006; Takahama 2006; Wu 2006). In addition, it is not only the maturing thymocytes that benefit from these interactions as it has been shown that during late thymic organogenesis, the development of the thymic stromal cells, specifically maintenance of specialised thymic epithelial lineages, relies on thymocyte-derived signals in a process termed thymic crosstalk (Anderson *et al.* 1993; Anderson and Jenkinson 1995; Hollander *et al.* 1995; Nehls *et al.* 1996; Manley 2000; Klug *et al.* 2002; Jenkinson *et al.* 2005; Rothenberg *et al.* 2008; White *et al.* 2008; Palmer and Naeher 2009; Shakib *et al.* 2009).

Anatomically, the thymus can be compartmentalised into cortical and medullary regions, both of which are responsible for separate phases of T-cell development (Shakib *et al.* 2009). The main processes in T-cell development can be divided into a number of key events, with the initial stages taking place in the outer cortex of the thymus following the influx of bone-marrow-derived lymphoid progenitor cells, which transmigrate across the cortico-medullary junction to the thymic parenchyma from the blood (Figure 1.1) (Takahama 2006). These immature cells then begin the process of development, aided by Notch-mediated and interleukin-7 (IL-7) signals from cortical stromal cells (Takahama 2006). Thymocyte progression through this stage is exemplified by the expression profiles of cell surface markers CD25, CD44 and c-kit

Figure 1.1 - The Development of T-Lymphocytes



The main stages in T-lymphocyte development. (a) Transmigration of immature lymphoid progenitor cells into the thymus across the cortico-medullary junction to the thymic parenchyma from the blood (b) development of these cells, aided by Notch-mediated and interleukin-7 (IL-7) signals from cortical thymic epithelial cells (cTECs) results in the generation of $CD4^{-}CD8^{-}$ double negative (DN) thymocytes and then (c) $CD4^{+}CD8^{+}$ double positive (DP) thymocytes, which occurs in the outer cortex of the thymus. Massive expansion of DP thymocytes occurs and the cells subsequently migrate to the inner cortex for (d) positive selection, a process brought about by cTECs and macrophages (MΦ). The DP cells then transfer to the medulla for negative selection, aided by signals from medullary thymic epithelial cells (mTECs), B-lymphocytes (B-cells) and dendritic cells (DCs), generating single positive (SP) $CD4^{+}$ or $CD8^{+}$ T-cells, which are finally (e) exported to the periphery as mature T-cells (Iwasaki and Akashi 2006; Rodewald 2006; Wu 2006). Figure adapted from (Kyewski and Klein 2006; Takahama 2006).

(Iwasaki and Akashi 2006; Takahama 2006; Wu 2006). The cells alter their expression of these molecules as they differentiate and proliferate, passing from $CD4^-CD8^-CD25^-CD44^+c\text{-}kit^+$ double negative (DN) 1 thymocytes to $CD4^-CD8^-CD25^+CD44^+c\text{-}kit^+$ DN2, and eventually to $CD4^-CD8^-CD25^+CD44^-c\text{-}kit^-$ DN3 thymocytes (Iwasaki and Akashi 2006; Takahama 2006; Wu 2006).

For progression along the developmental pathway, DN3 cells must successfully complete in-frame rearrangement of VDJ TCRβ-chain and pre-TCRα genes, thus allowing formation of a cell-surface pre-TCR complex (Iwasaki and Akashi 2006; Takahama 2006). The thymocytes are then able to migrate through the cortex towards the subcapsular zone where expression of CD4 and CD8 can occur, generating double positive (DP) thymocytes, a process which is negatively regulated by high levels of transforming growth factor-β 1 (TGFβ1) (Figure 1.1) (Rubtsov and Rudensky 2007).

It is at this point that the first blueprint of central tolerance is imprinted into the developing T-cell repertoire. The first major checkpoint ensures a useful T-cell repertoire in a process termed positive selection (Figure 1.1). Through presentation of peptide/MHC ligands on the surface of cortical thymic stromal cells and dendritic cells (DCs), and subsequent interactions with TCRαβ complexes on the developing DP thymocytes, only those capable of recognising and binding to self-MHC, receive positive signals to survive and continue differentiation (Kyewski and Klein 2006; Shakib *et al.* 2009).

For progression along the pathway of differentiation, positively-selected DP thymocytes must lose expression of either CD4 or CD8 to become single positive (SP) thymocytes. These cells are then able to migrate towards the medulla, where maturation can occur, along with the final phase of tolerance induction; negative selection (Figure 1.1) (Takahama 2006). Negative selection occurs mainly, although not exclusively, within the thymic medulla, and results in the clonal deletion of highly auto-reactive thymocytes, or commitment of mildly self-reactive T-cells to the forkhead box P3 (Foxp3)-expressing regulatory T-cell (T_{reg}) lineage, before the release of mature T-cells into the circulation (Kyewski and Klein 2006; Aschenbrenner *et al.* 2007; Hamazaki *et al.* 2007; Rubtsov and Rudensky 2007; Venanzi *et al.* 2007; Lars-Oliver Tykocinski 2008; McCaughtry *et al.* 2008; Shakib *et al.* 2009).

The three-dimensional meshwork of thymic epithelial cells (TECs) provide a constant input to all stages of T-cell development, and the impact these cells have in moulding the T-cell repertoire has been well documented (Anderson *et al.* 1993; Anderson and Jenkinson 1995; Hollander *et al.* 1995; Nehls *et al.* 1996; Manley 2000; Jenkinson *et al.* 2005). This is highlighted by the fact that TEC frequency is directly proportional to the efficacy of thymic T-cell output and that age-related thymic atrophy leads to a decline in T-cell production, to the extent that by 40-50 years thymic output has fallen to less than 10% of its maximum potential, a process again attributed to a decline in TEC frequency and/or function (Chidgey *et al.* 2007; Jenkinson *et al.* 2007; Shakib *et al.* 2009).

1.3. PROMISCUOUS GENE EXPRESSION IN THE THYMUS AND NEGATIVE SELECTION

For decades, immunologists have debated the relative importance of central versus peripheral tolerance, an enduring argument being that the majority of the body's proteome would not be visible to developing T-cells within the sterile thymic microenvironment, hence central tolerance to self would be practically impossible. Added to this the re-emergence of the concept of T-cell subsets with regulatory characteristics, peripheral tolerance mechanisms have dominated the field (Walker and Abbas 2002; Mathis and Benoist 2004). Within the periphery, a large number of systems are in place to prevent the activation of mature autoimmune T-cells and induce peripheral tolerance including T-cell-intrinsic mechanisms; clonal ignorance, anergy, phenotypic skewing and apoptosis, and T-cell-extrinsic mechanisms; which rely on auxiliary cells such as T_{reg} and tolerogenic dendritic cells (Walker and Abbas 2002; Siggs *et al.* 2006; Aschenbrenner *et al.* 2007). Yet despite this, there was no denying the fact that such large numbers of immature T-cells are produced and then immediately deleted within the thymus. Furthermore, the importance of TECs upon the prevention of self-reactive T-cell specificities is evidenced by the resulting development of autoimmune deficiencies when the TEC developmental pathway is disrupted (Shakib *et al.* 2009). This challenged the classic views of a dominant role for peripheral tolerance mechanisms and fuelled further investigation into central tolerance (Mathis and Benoist 2004; Kyewski and Klein 2006). The first major insight into how intra-thymic tolerance to self could realistically be achieved, revealed that a number of diverse tissues were represented at the RNA level in distinct areas of the thymus including the pancreas (insulin), and the central nervous

system (myelin basic protein and proteolipid protein) (Gotter *et al.* 2004; Mathis and Benoist 2004; Kyewski and Klein 2006). Numerous reports followed in support of this, and with the development of a meticulous purification protocol for thymic stromal cell populations, the theory of intra-thymic expression of self-antigens, and the significance of the central tolerance concept, was again brought to the forefront, challenging the general rules of cell-type-specific patterns of gene expression (Jolicoeur *et al.* 1994; Klein *et al.* 1998; Mathis and Benoist 2004; Kyewski and Klein 2006).

Further studies into this phenomenon have built upon these findings, showing that essentially all organs, including genes thought to be restricted in their spatial, developmental, sex-dependent and temporal expression are represented in the thymic microenvironment (Mathis and Benoist 2009). A predicted pool of up to 3000 tissue-restricted antigens (TRA) are expressed; representing a significant portion of the genome, which include transcripts for, among others; structural proteins, hormones, secreted proteins and transcription factors (Anderson *et al.* 2002; Kyewski and Derbinski 2004; Kyewski and Klein 2006; Mathis and Benoist 2009). These self-antigens can be generated in a number of ways including; expression within bone-marrow derived antigen presenting cells (APC) of the circulatory system (dendritic cells, macrophages and B-cells); import of a limited number of antigens to the thymus via APCs from the periphery, or directly through the bloodstream; however the principal way in which these TRA are produced was found to be from the proteome of resident thymic stromal cells (Koble and Kyewski 2009). This process, termed promiscuous gene expression (PGE), was found to be predominantly restricted to a

very select subset of thymic epithelial cells, primarily located in the medulla and cortico-medullary junction; medullary thymic epithelial cells (mTECs) (Anderson and Jenkinson 2001; Derbinski *et al.* 2001; Kyewski and Derbinski 2004; Mathis and Benoist 2004; McCaughtry *et al.* 2008). This small fraction of cells make up only approximately 0.005% of the whole thymus, however their role is crucial for preventing autoimmunity, highlighted by the fact that the majority of autoantigens, either known or suspected to cause specific autoimmune diseases, can be detected in murine mTECs including, but not limited to; thyroglobulin and thyroid peroxidise (TPO), causative autoantigens of Hashimoto's thyroiditis; insulin, glutamic acid decarboylase-67 (GAD67) and IA-2, major autoantigens of type I diabetes; and the rheumatoid arthritis antigen, collagen II (Gotter *et al.* 2004). In addition, mTECs displayed many cancer-germline group antigens such as members of the melanoma antigen (MAGE)-A group or NY-ESO-1, along with differentiation antigens including tyrosinase and MART-1 (melanoma antigen recognized by T-cells), known targets of circulating T-cells (Gotter *et al.* 2004). Thus, PGE within mTECs acts to expose developing T-cells to an extensive array of TRA, imprinting them with tolerance to self.

1.4. Molecular Regulation of Promiscuous Gene Expression by AIRE

The cellular and molecular intricacy of central tolerance, although not thoroughly understood, was brought one step closer to comprehension through the discovery of a molecular determinant of PGE, following studies into a relatively rare autoimmune disease. Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy

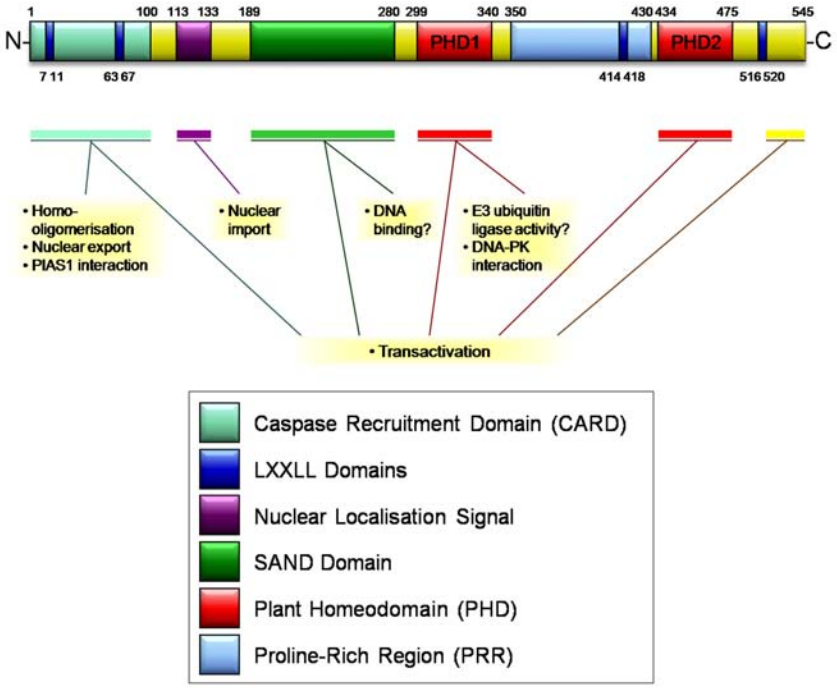
(APECED), also known as Autoimmune Polyendocrinopathy Syndrome Type 1 (APS1), is a devastating primary autoimmune disease which manifests itself as organ-specific autoimmunity to multiple organs due to the presence of circulating tissue-specific auto-antibodies, which target and attack various organs (Aaltonen *et al.* 1994; Liston 2006; Devoss and Anderson 2007; Hubert *et al.* 2008). Sufferers classically had to present with at least two of the three characteristic features; chronic mucocutaneous candidiasis, hypoparathyroidism, and primary adrenal insufficiency, however, more recently, diagnosis extends to more atypical symptoms, following identification of the causal genetic lesion, including; chronic diarrhoea, keratitis, autoimmune hepatitis, vitiligo, alopecia, periodic rash with fever, severe constipation and enamel hypoplasia (Husebye *et al.* 2009; Mathis and Benoist 2009). Unlike the majority of autoimmune diseases, APECED is curiously uncomplicated in its genetic cause; found in its most common form to be an autosomal-recessive disorder (Aaltonen *et al.* 1994; Nagamine *et al.* 1997; Heino *et al.* 2001; Liston *et al.* 2003; Halonen *et al.* 2004; Liston 2006; Devoss and Anderson 2007; Mathis and Benoist 2009). It was therefore possible to map the disease locus to a single chromosomal location; human chromosome 21q22.3 which was found to encode a novel 2027 base pair (bp) gene; the Autoimmune Regulator (AIRE) (Aaltonen *et al.* 1994; Nagamine *et al.* 1997; Heino *et al.* 2001; Mathis and Benoist 2009).

1.5. The Many Domains of AIRE: Linking Structure to Function

The AIRE gene encodes a large 545 amino acid (aa) protein, with a molecular weight of approximately 58kDa, whose structure is characteristic of known transcription

factors and contains regions well-conserved across vertebrates (Figure 1.2) (Aaltonen *et al.* 1994; Nagamine *et al.* 1997; Uchida *et al.* 2004; Saltis *et al.* 2008; Mathis and Benoist 2009). To date, more than 60 APECED-causing mutations have been uncovered across the AIRE gene, distributed throughout the locus and therefore present within all structural domains of the protein, highlighting the importance of each of these regions for AIRE's function (Heino *et al.* 2001; Halonen *et al.* 2004; Mathis and Benoist 2009). The most common, however, are a cytosine to tyrosine transition at position 769bp, and a 13bp deletion in exon 8 (967-979bp), with either of these two mutations occurring in approximately 95% of patients (Husebye *et al.* 2009). The key domains of the AIRE protein, represented in Figure 1.2 include; a SAND (Sp100, AIRE, NucP41/75, and DEAF-1) domain; found within many transcriptional modifiers and with predicted DNA-binding properties; two plant homeodomain (PHD)-zinc fingers, essential for AIRE's function; four LXXLL motifs (where L is leucine and X is any amino acid), thought to mediate protein-protein interactions; a homogeneously staining region which has recently been identified as a caspase recruitment domain (CARD), the site of many mutations in APECED patients and a potential impetus for AIRE homo- or hetero-dimerisation, or alternatively interactions with other proteins with roles in transcriptional control (Gibson *et al.* 1998; Heino *et al.* 2001; Kumar *et al.* 2001; Bottomley *et al.* 2005; Goldrath and Hedrick 2005; Ferguson *et al.* 2007; Meloni *et al.* 2008; Mathis and Benoist 2009). The protein also contains two functional nuclear localisation signals (NLS); a classical importin-α and –β pathway import signal; and a CRM1-dependent

Figure 1.2 - Structure of the Autoimmune Regulator Protein

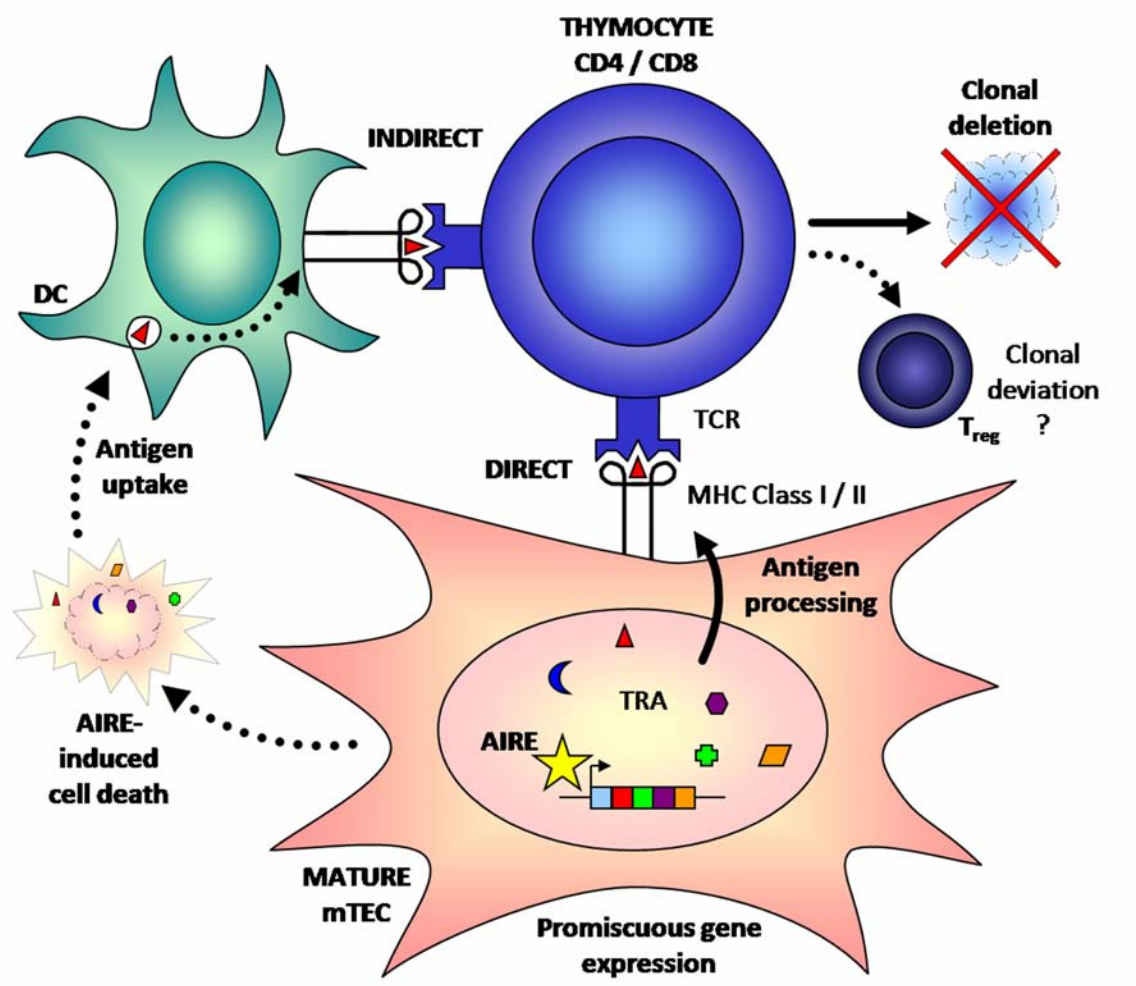


Schematic representation of the functional domains of the 545 amino acid (aa) human autoimmune regulator (AIRE) protein and their identified roles. The N-terminal caspase recruitment domain (CARD) is believed to be required for homo-oligomerisation and for the formation of AIRE nuclear bodies. This domain is known to interact with protein inhibitor of activated STAT-1 (PIAS1); a transcriptional co-regulator known to associate with the nuclear matrix (Ilmarinen *et al.* 2008). AIRE has four LXXLL motifs (aa7-11, aa63-67, aa414-418 and aa516-520) which are putative sites for nuclear receptor binding, and a conserved bipartite nuclear localisation signal (NLS), present in the N-terminal, has been implicated in nuclear import. AIRE's unconfirmed DNA-binding properties are thought to be co-ordinated through the SAND (Sp100, AIRE, NucP41/75, and DEAF-1) domain (Gibson *et al.* 1998; Kumar *et al.* 2001). DNA-dependent protein kinase (DNA-PK) is able to phosphorylate AIRE and aids in transcriptional transactivation (Liiv *et al.* 2008). This interaction occurs via the first of AIRE's two cysteine-rich plant homeodomain (PHD)-zinc finger domains. PHD1 also displays E3-ubiquitin ligase activity *in vitro*, although this result in controversial (Uchida *et al.* 2004). The CARD and SAND domain, PHD1 and 2, and the C-terminal region are all essential for AIRE's ability to influence transcription, and mutations in any of these domains disrupt this functional property (Heino *et al.* 2001; Halonen *et al.* 2004). AIRE has also demonstrated associations with; the transcriptional co-activator CREB-binding protein (CBP); with positive transcription elongation factor-b (P-TEFb), a key player in transcription elongation; and also with the nuclear matrix, although to date the regions required for these activities are not well defined (Pitkanen *et al.* 2000; Akiyoshi *et al.* 2004; Pitkanen *et al.* 2005; Ferguson *et al.* 2007; Oven *et al.* 2007; Peterson *et al.* 2008). Figure adapted from (Su and Anderson 2004; Peterson *et al.* 2008).

export signal, located in the amino terminal region of AIRE (Pitkanen *et al.* 2001; Ilmarinen *et al.* 2006). These features give important clues to AIRE's molecular role, however there is still controversy over AIRE's true function.

The expression of AIRE is detected most prominently within the lymphoid organs, primarily localised to the thymic medulla and correlating significantly with the promiscuous expression of self-antigens within mTECs (Figure 1.3) (Nagamine *et al.* 1997; Heino *et al.* 1999; Liston *et al.* 2004; Klamp *et al.* 2006; Mathis and Benoist 2009). The protein has been detected in the nucleus of mTECs, where staining with anti-AIRE antibodies reveals a punctate, perinuclear speckled pattern similar to promyelocytic leukaemia (PML) nuclear bodies, in which the AIRE homologues Sp100, Sp140 and Lysp100 can be found (Bjorses *et al.* 1999; Heino *et al.* 1999; Pitkanen *et al.* 2000; Halonen *et al.* 2001; Pitkanen *et al.* 2001; Akiyoshi *et al.* 2004; Cavadini *et al.* 2005; Hubert *et al.* 2008). However, from co-immunofluorescence staining, it has been shown that AIRE does not co-localise with these PML structures, but in fact resides within unique subnuclear structures (Akiyoshi *et al.* 2004). AIRE nuclear bodies are distinct from, but located adjacent to nuclear speckles, in which pre-mRNA splicing components, factors required for 3` end RNA processing and proteins involved in transcription are found (Su *et al.* 2008). The formation of AIRE nuclear bodies is dependent upon structurally sound AIRE protein, as mutations have demonstrated a block in their formation (Pitkanen *et al.* 2001; Halonen *et al.* 2004; Ferguson *et al.* 2007; Peterson *et al.* 2008; Su *et al.* 2008). These AIRE complexes were shown to be excluded from nucleoli and to have an association with the nuclear matrix, thus indicating a possible interaction with

Figure 1.3 - Autoimmune Regulator Protein in the Control of Promiscuous Gene Expression and the Induction of Central Tolerance



Diagrammatic representation of the role of autoimmune regulator (AIRE) in promiscuous gene expression (PGE) and central tolerance. AIRE controls the expression of a catalogue of self-proteins characteristic of peripheral organs termed tissue-restricted antigens (TRA). Exclusively within medullary thymic epithelial cells (mTECs), TRA are then processed and loaded onto cell-surface-displayed MHC molecules for direct antigen presentation. The rapid turnover rate of mTECs, upon AIRE activation, results in uptake of TRA for indirect antigen presentation by thymic dendritic cells (DCs) (Gray *et al.* 2007). Developing $CD4^+$ or $CD8^+$ single-positive thymocytes migrate through the medulla and interactions between self-peptide-MHC complexes and T-cell receptors (TCRs) dictate their fate. TCRs capable of recognising these TRA-MHC within a given affinity / avidity, would primarily be removed by clonal deletion, although an alternative suggestion is that some potentially auto-reactive T-cells may survive by clonal deviation to a more regulatory role (Walker *et al.* 2003; Anderson *et al.* 2005; Kuroda *et al.* 2005). Figure adapted from (Mathis and Benoist 2009).

transcriptional machinery (Pitkanen *et al.* 2001; Devoss and Anderson 2007; Hubert *et al.* 2008). At a molecular level, the AIRE protein displays characteristics of a transcriptional regulator (Pitkanen *et al.* 2000; Meloni *et al.* 2008). It is capable of transcriptional transactivation of numerous reporter and endogenous promoters, has demonstrated the potential to directly bind DNA and has been shown to co-localise with a number of protein partners, resulting in the integration of AIRE into large biological complexes greater than 650kDa *in vitro* (Pitkanen *et al.* 2000; Halonen *et al.* 2004; Pitkanen *et al.* 2005; Purohit *et al.* 2005; Oven *et al.* 2007; Ruan *et al.* 2007; Liiv *et al.* 2008; Mathis and Benoist 2009).

Shortly after the identification of the AIRE gene, investigators were eager to generate a knockout mouse model, to further study AIRE's function (Venanzi *et al.* 2004). The AIRE-knockout mice displayed an autoimmune syndrome, comparable to APECED in humans, with auto-antibodies directed against multiple organ systems (Kyewski 2008). However, rather than the whole organ being targeted, it was individual and highly unique regions within, to which the damage was directed, for example; the photoreceptor layer of the retina in the eye; parietal cells of the stomach; and oocytes in the ovaries (Anderson *et al.* 2002). Through isolation and gene expression profiling of mTECs from AIRE-deficient mice, it was revealed that the promiscuous expression of TRA genes was significantly depressed, which provided the first real clues as to AIRE's role; in preventing autoimmunity (Figure 1.3) (Su and Anderson 2004; Venanzi *et al.* 2004). Studies into this phenomenon revealed the importance of this promiscuous gene expression with regards to autoimmune disease susceptibility, however, it wasn't until more recently that the breakdown of thymic

expression of individual self-antigens was demonstrated to be sufficient for the development of selective organ-specific autoimmunity (Klein *et al.* 2000; Derbinski *et al.* 2001; Avichezer *et al.* 2003; Gotter *et al.* 2004; DeVoss *et al.* 2006; Gavanescu *et al.* 2007). Through the use of AIRE-deficient mice, DeVoss *et al* (2006) found that spontaneous eye-specific autoimmune disease occurred as a direct result of the loss of expression of a highly specific eye antigen; interphotoreceptor retinoid-binding protein (IRBP), within the thymus, due to the absence of AIRE (DeVoss *et al.* 2006). Following this, Gavanescu *et al* (2007) reported an analogous result; with an absence of the AIRE-regulated stomach antigen mucin 6, triggering gastritis (Gavanescu *et al.* 2007). Thus AIRE appeared to play an essential role in central tolerance, as a transcriptional regulator promoting the expression of TRA within mTECs for subsequent direct antigen presentation to developing thymocytes (Figure 1.3). However, given the very low numbers of AIRE-positive mTECs, in relation to the millions of developing thymocytes, it would seem unlikely that the entire T-cell repertoire could be checked for self-reactivity (Kyewski and Derbinski 2004). Thus, the finding that thymocytes spend up to 5 days in the medulla and are highly motile may counteract this problem (Kyewski and Derbinski 2004). Added to this, AIRE-positive mTECs have demonstrated very high turnover rates, surviving only a few days after the induction of AIRE expression, which therefore may also alleviate the question of how so few mTECs could process and present such a massive catalogue of AIRE-regulated TRA (Gray *et al.* 2007). Through rapid death and renewal of the mTEC subset, the gamut of TRAs on display to developing T-cells would be constantly updated, and could even be maximised through uptake of apoptotic mTECs by thymic dendritic cells for indirect presentation of TRA leading to the

deletion of self-reactive thymocytes (Figure 1.3) (Gray *et al.* 2007; Ferguson *et al.* 2008). An alternative mechanism is that AIRE may be involved in the positive selection of T_{reg} cells, although this is not believed to be AIRE's major role as the numbers of $CD4^+CD25^+$ T_{reg} cells are comparable to wild type in AIRE-deficient mice (Walker *et al.* 2003; Anderson *et al.* 2005; Kuroda *et al.* 2005).

Interestingly, AIRE's role does not seem to be limited to central tolerance mechanisms as numerous groups detected AIRE mRNA within a wide range of peripheral tissues including kidneys, testis, ovaries, adrenal glands, pancreas and liver (Halonen *et al.* 2001; Adamson *et al.* 2004). However it would appear that these small quantities of mRNA are not sufficient for translation to occur as immunofluorescence and *in situ* hybridisation revealed peripheral AIRE protein restricted to immunologically relevant sites, such as the spleen, lymph nodes and bone marrow, although one group, using a novel monoclonal antibody specific for murine AIRE, recently demonstrated the presence of $AIRE^+$ cells in the thymic medulla only (Heino *et al.* 1999; Maarit Heino 2000; Halonen *et al.* 2001; Anderson *et al.* 2002; Adamson *et al.* 2004; Klamp *et al.* 2006; Lee *et al.* 2007). Peripheral AIRE has been predicted to be required for the maintenance of tolerance, for antigen presentation and even for T-cell-independent B-cell responses, and indeed, the finding that AIRE expression is restricted to cells of the haematopoietic lineage and peripheral $AIRE^+$ stromal cells certainly supports these theories (Sillanpaa *et al.* 2004; Lee *et al.* 2007; Gardner *et al.* 2008; Lindh *et al.* 2008; Pontynen *et al.* 2008; Mathis and Benoist 2009). An array of dendritic cell types, macrophages and monocytes were found to express AIRE and that loss of this expression disrupted the

antigen-presenting capabilities and transcriptional programmes of the cells (Sillanpaa *et al.* 2004; Pontynen *et al.* 2008). However, the transcript levels of AIRE observed, were always at considerably lower levels than those within mTECs and the protein was undetectable by flow cytometry, yet it may be possible that AIRE expression in peripheral haematopoietic cells, could have a role to play in antigen presentation, particularly under inflammatory conditions (Anderson *et al.* 2002; Hubert *et al.* 2008; Mathis and Benoist 2009). More recently, two groups have identified populations of lymph node stromal cells analogous to mTECs, able to express AIRE and a limited range of TRAs (Lee *et al.* 2007; Gardner *et al.* 2008). Lee *et al* (2007) showed that $CD45^-$ cortical lymph node stromal cells which constitutively expressed certain TRAs were able to present these self-antigens to naive $CD8^+$ T-cells, resulting in their primary activation and subsequent tolerance (Lee *et al.* 2007). Gardner *et al* (2008) then went on to describe a separate population of stromal cells also capable of deleting auto-reactive T-cells, residing ubiquitously in lymphoid organs including the mesenteric lymph nodes, Peyer's patches and tertiary lymphoid structures, which were $CD45^-$MHC $II^+EpCAM1^+$ and expressed TRAs, but displayed no co-stimulatory molecules (CD80 or CD86) unlike mTECs (Gardner *et al.* 2008). It is essential that self-tolerance be maintained once T-cells leave the thymus, thus the finding of AIRE expression outside the thymus may point to a more secondary role for AIRE-regulated peripheral tolerance, as a back-up to central tolerance. This hypothesis is strengthened by the finding that the repertoire of TRA generated by $AIRE^+$ stromal cells is much more limited than those within mTECs, and that there was little overlap between the two expression profiles, meaning that any peripheral self-antigens not displayed to the developing thymocytes within the thymus could theoretically be

present within the lymph nodes or the spleen, ensuring deletion of any escaped auto-reactive cells (Lee *et al.* 2007; Gardner *et al.* 2008; Kyewski 2008). In addition, transplantation of AIRE-deficient thymi into wild-type recipient mice still results in autoimmune manifestations, highlighting the fact that AIRE and TRA expression within the periphery is unable to prevent autoimmune attack, although it may still play a role (Anderson *et al.* 2002; Liston *et al.* 2003; Kuroda *et al.* 2005).

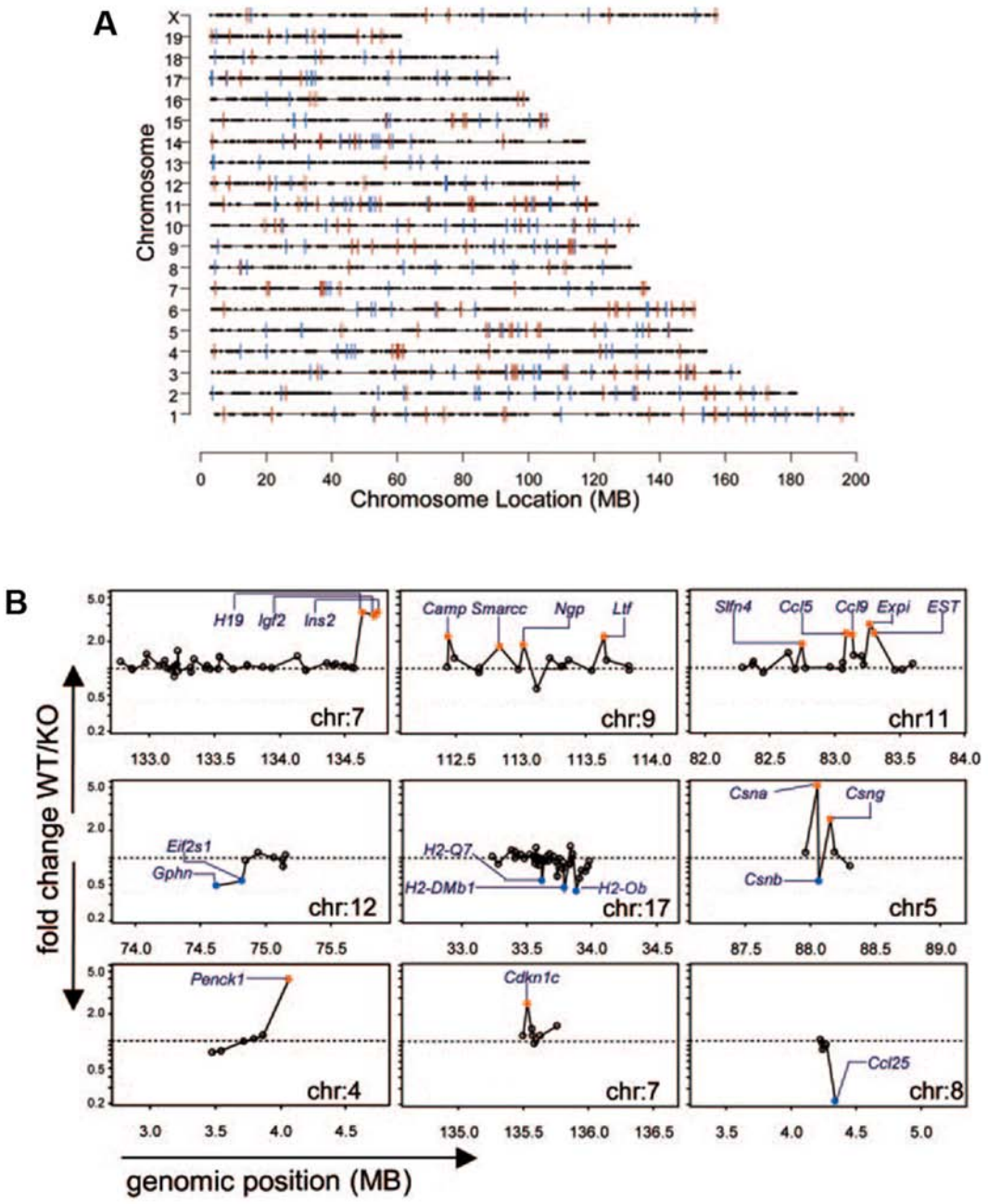
1.6. AIRE-Regulated Genes: Clustering on Chromosomes

This work also illustrated the scope of promiscuously expressed genes under the control of AIRE; the fact that AIRE's target genes extended far beyond the pool of autoantigens in APECED (Anderson *et al.* 2002; Gotter *et al.* 2004). It was noted that these genes did not have any obvious association in terms of structure or function, and even more strikingly, the usual constraints of sex-specific or developmental-specific patterns of gene expression appeared to be over-ridden by AIRE (Gotter *et al.* 2004). The mammary gland casein proteins, normally restricted to expression in late-pregnancy, along with the testis-specific antigen sperm-associated antigen 6 (SPAG6) and the placental hormone placental lactogen (PL), showed equal levels of expression in male and female mTECs, which offered a potential explanation for the enigma of how tolerance to self-antigens, which only arise in adulthood, was brought about during the development of the immune system, when these proteins would not be visible to the developing T-cell repertoire (Gotter *et al.* 2004; Derbinski *et al.* 2005; Johnnidis *et al.* 2005). AIRE-regulated genes showed a broad diversity, thus, in order to elucidate any common features regarding the

localisation of these genes throughout the genome, comprehensive analysis of their expression was undertaken, which revealed a global distribution throughout the genome, with no significant under- or over-representation of particular chromosomes (Figure 1.4) (Gotter *et al.* 2004; Johnnidis *et al.* 2005). Instead, across individual chromosomes, a considerable clustering of co-regulated genes was revealed, with 53 of 200 AIRE-regulated genes analysed occurring alongside a similarly AIRE-activated neighbour within a region of up to 200kbp, which led to the proposal that AIRE could direct gene expression through a broad activity on chromatin conformation (Figure 1.4) (Johnnidis *et al.* 2005). Gene clustering is a feature with a number of evolutionary advantages, including co-ordinated gene expression in single cells and sharing of enhancer or locus-control regions, with clustered genes generally switched on *en masse*, a highly efficient process as the transcriptional machinery would be able to access and induce the expression of co-expressed genes more easily if they were neighbours, than if they were distributed randomly across the chromosomes (Caron *et al.* 2001; Boutanaev *et al.* 2002; Roy *et al.* 2002; Gotter *et al.* 2004; Gierman *et al.* 2007; Soshnikova and Duboule 2009).

For a number of AIRE-regulated clusters, this did in fact appear to be the case (Johnnidis *et al.* 2005). For example, the two strict TRA salivary proteins-1 and -2, found on *Mus musculus* chromosome 15, are located within 130kbp of each other and both are switched on in the presence of AIRE (Johnnidis *et al.* 2005). Similarly, gephyrin (*Gphn*), eukaryotic translation initiation factor 2, subunit 1α (*Eif2s1*) and arginase type II (*Arg2*), clustered on *Mus musculus* chromosome 12 are also affected by AIRE *en masse*, although instead of an increase in expression, all three genes

Figure 1.4 – Chromosomal Clustering of Genes under the Transcriptional Control of the Autoimmune Regulator



Genomic distribution of gene targets of the *Mus musculus* autoimmune regulator (AIRE). *A*, AIRE-regulated genes appear to distribute evenly throughout all chromosomes. The top 200 AIRE-activated genes are represented in red, the top 200 AIRE-repressed genes are in blue. *B*, The impact of AIRE upon expression from each cluster can vary; with AIRE targets either being located adjacent to each other, or interspersed with AIRE-independent genes. Certain clusters also show divergent regulation, with adjacent loci showing opposing levels of expression in the presence of AIRE. Data taken from (Johnnidis *et al.* 2005).

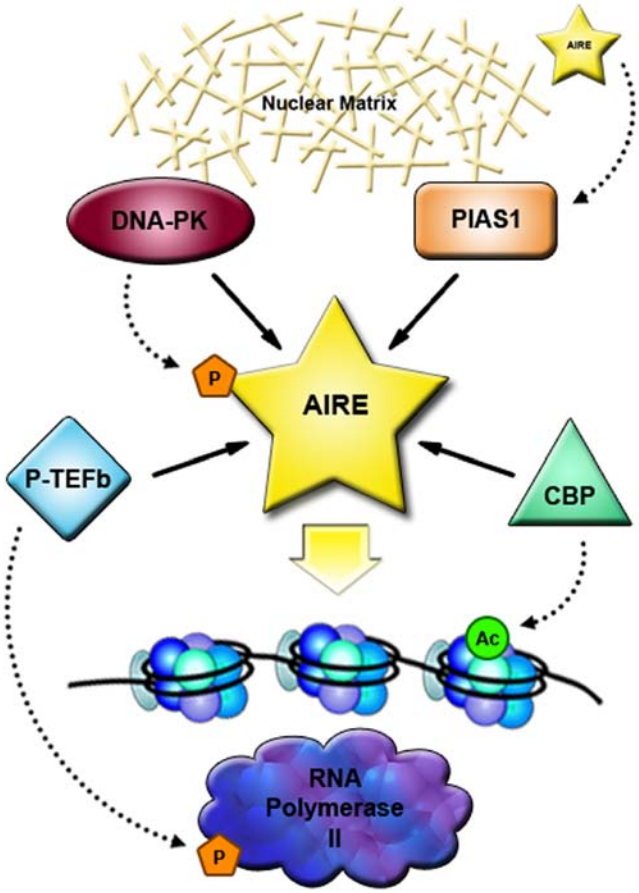
were shown to be negatively regulated by AIRE, suggesting that AIRE's role as a transcriptional regulator is decidedly more complex (Figure 1.4) (Johnnidis *et al.* 2005). Interestingly, however, the impact of AIRE on the control of transcription within its target clusters was not a uniform on / off switch as a significant number of clusters showed a rather punctate expression profile; with the induction in expression of only two or three AIRE-dependent genes, while neighbouring genes in the same loci were not induced, or more surprisingly, their expression decreased (Figure 1.4) (Johnnidis *et al.* 2005). The epidermal differentiation complex on *Mus musculus* chromosome 3, had previously been mapped following studies into the development of epidermal tissue, yet this cluster of genes was also found to be expressed in mTECs, under the transcriptional control of AIRE, with some loci being repressed including; synaptosomal-associated protein 25bp (*Snap25bp*); S100 calcium binding protein (*S100*) A 13; *S100a3*; *S100a4*; and *S100a6*, and others being activated (*S100a8*) (Johnnidis *et al.* 2005). A number of clusters were also identified which contained members of the keratin family (Johnnidis *et al.* 2005). These proteins have been shown to be specific to certain cortical and medullary stromal cell subsets, and may therefore be required for the development of the thymic architecture (Klug *et al.* 1998). The revelation that the expression of keratin 4, located on *Mus musculus* chromosome 5, increased in the presence of AIRE and that keratin 18, downstream of keratin 4, showed lower transcript levels in the presence of AIRE, may explain the differential expression patterns of these structural proteins in the thymus (Johnnidis *et al.* 2005).

In their thorough analysis of PGE within different subsets of thymic stromal cells (DCs, cTECs and mature and immature mTECs), Derbinski *et al* (2005) demonstrated four distinct pools of TRA; (i) those that show equal levels of expression in the TEC lineage; (ii) genes restricted to mTECs; (iii) AIRE-regulated genes only expressed in mature mTECs; and (iv) those that are switched on upon mTEC maturation, but that are in-dependent of AIRE (Derbinski *et al.* 2005). This group chose to focus on the casein cluster on *Mus musculus* chromosome 5 which was found to house genes from a number of these different gene categories, with casein-β and -κ showing AIRE-independent expression; being up-regulated in mature mTECs in both wild-type and AIRE-deficient mice, yet all other genes in the cluster; casein-α, -γ and –δ, highly dependent on AIRE in mature mTECs (Figure 1.4) (Derbinski *et al.* 2005). However, the distribution of these genes is such that AIRE-independent and AIRE-dependent genes are interspersed along the 1Mbp cluster, often lying directly adjacent to each other (Derbinski *et al.* 2005). Thus an alternative mode for AIRE's control of promiscuous gene expression may involve two possible levels of regulation; a large-scale control of whole gene clusters indicative of epigenetic mechanisms, and a more specific targeting of individual genes (Derbinski *et al.* 2005; Johnnidis *et al.* 2005).

1.7. Protein Partners of AIRE

The interaction of AIRE with other highly active proteins complements this theory, suggesting a role for AIRE as a co-activator in large transcriptional complexes (Figure 1.5). CREB-binding protein (KAT3A/CBP), a common transcriptional co-

Figure 1.5 – Autoimmune Regulator Binding Partners



A model of autoimmune regulator (AIRE)-containing complexes in the control of promiscuous gene expression. AIRE has been shown to interact with several protein partners with defined functions, suggesting a role in AIRE's transcriptional regulation. DNA-dependent protein kinase (DNA-PK), a protein involved in non-homologous end joining, may aid in AIRE recruitment to the nucleus through tethering to nuclear matrix (Liiv *et al.* 2008). DNA-PK is able to phosphorylate two sites in AIRE, which enhances AIRE's transcriptional activity (Liiv *et al.* 2008). Protein inhibitor of activated STAT-1 (PIAS1); an inhibitor of STAT (signal transducer and activator of transcription)-mediated cytokine signalling, which functions as a transcriptional co-regulator, is also implicated in the recruitment of AIRE to the nuclear matrix (Ilmarinen *et al.* 2008). CREB-binding protein (KAT3A/CBP), a common transcriptional co-activator, co-localises with AIRE in nuclear bodies (Pitkanen *et al.* 2000; Akiyoshi *et al.* 2004; Pitkanen *et al.* 2005; Ferguson *et al.* 2007). With intrinsic histone acetyltransferase activity, KAT3A/CBP may facilitate AIRE-mediated expression through acetylation (Ac) of histone tail residues. AIRE has been found to bind positive transcription elongation factor-b (P-TEFb), a key player in transcription elongation (Oven *et al.* 2007). P-TEFb can phosphorylate the serine residues of RNA polymerase II (RNA pol II), converting stalled RNA pol II to the elongating form, triggering active expression (Oven *et al.* 2007). The formation of these large nuclear complexes may contribute to AIRE's control of promiscuous gene expression. Figure adapted from (Peterson *et al.* 2008).

activator with intrinsic histone acetyltransferase activity for a wide range of transcription factors, has been shown to co-localise with AIRE in nuclear bodies and together can bring about the activation of a number of genes within cultured cells (Pitkanen *et al.* 2000; Akiyoshi *et al.* 2004; Pitkanen *et al.* 2005; Ferguson *et al.* 2007). Protein inhibitor of activated STAT-1 (PIAS1); an inhibitor of STAT (signal transducer and activator of transcription)-mediated cytokine signalling, functions as a transcriptional co-regulator and is also located in nuclear bodies and known to associate with the nuclear matrix (Ilmarinen *et al.* 2008). However, AIRE nuclear bodies do not co-localise but are instead found neighbouring PIAS1, indicating that the binding of AIRE and PIAS1 may be through additional nuclear matrix components (Ilmarinen *et al.* 2008). Another partner of AIRE; DNA-dependent protein kinase (DNA-PK) is a protein shown to be involved in non-homologous end joining and also to be linked to nuclear matrix (Liiv *et al.* 2008). This protein has demonstrated a key role in AIRE's function as it phosphorylates two sites in the AIRE protein, mutation of which has a negative impact on AIRE's transcriptional activity (Liiv *et al.* 2008). In addition, AIRE has been found to bind positive transcription elongation factor-b (P-TEFb), a key player in transcription elongation, resulting in recruitment to RNA polymerase II (RNA pol II) found at the promoters of AIRE-target genes (Oven *et al.* 2007; Peterson *et al.* 2008). Thus it is possible that AIRE, along with its many protein partners may control PGE via an influence on the organisation of DNA indirectly; recruiting members of a transcriptional regulation complex to AIRE-dependent gene clusters scattered throughout the genome. Through interactions with the nuclear matrix, AIRE may then be able to recruit factors required for

transcriptional elongation, thus resulting in the induction of PGE (Peterson *et al.* 2008).

1.8. Epigenetics and the Control of Gene Expression

Transcriptional regulators, including AIRE, with influence over multiple gene expression programmes, are all faced with a number of complexities. Once transported through the nuclear envelope, these proteins encounter the initial problem of locating their target genes. The nucleus is a highly heterogeneous structure, consisting of many sub-compartments and components, with the DNA molecule making up only approximately 6% of the total nuclear volume, and of this only 1-2% represents functional genes in humans (Lander *et al.* 2001). The DNA itself is highly dynamic and is divided up into large functional units called chromosomes, the distribution of which is far from random within the nucleus, with particular chromosomes localising to preferred sites, referred to as chromosome territories, which correlate with gene activity (Kurz *et al.* 1996; Mahy *et al.* 2002; Parada and Misteli 2002; Simonis *et al.* 2006; Dostie *et al.* 2007; van Berkum and Dekker 2009). However, the DNA within these territories is not a static structure as the spatial distribution of certain genes with respect to their chromosome territory has been shown to correlate with their activity (Kurz *et al.* 1996; Mahy *et al.* 2002; Fraser and Bickmore 2007; Sexton *et al.* 2007). In some cases their position ensures maximal expression and alternatively it can result in repression (Gierman *et al.* 2007; Sexton *et al.* 2007).

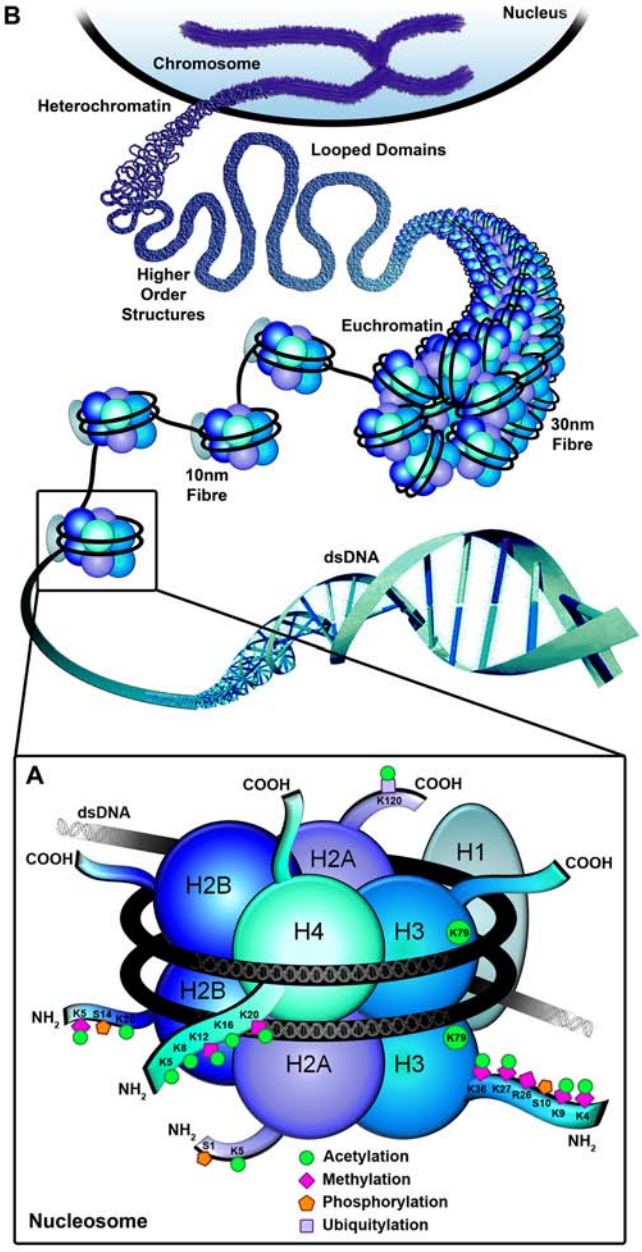
In mammalian nuclei, active transcription was found to occur at meta-stable foci, distributed throughout the nucleus and containing high concentrations of RNA pol II, called transcription factories (Jackson *et al.* 1993; Wansink *et al.* 1993; Iborra *et al.* 1996; Pombo *et al.* 1999; Osborne *et al.* 2004). In contrast, silenced genes mapped to discrete and more dynamic nuclear sub-compartments known as Polycomb bodies due to the presence of Polycomb group proteins (PcG); originally shown to be essential for switching off the developmental homeobox (*Hox*) genes in *Drosophila melanogaster* (Saurin *et al.* 1998; Boyer *et al.* 2006; Grimaud *et al.* 2006; Lee *et al.* 2006; Sexton *et al.* 2007). The number of both PcG bodies and transcription factories within a single cell is limiting and hence silenced and active genes respectively are thought to share these sites, allowing co-ordinated control of multiple genes (Osborne *et al.* 2004; Schwartz *et al.* 2006; Osborne *et al.* 2007; Sexton *et al.* 2007). These transcriptions factories may in fact aid transcriptional regulators such as AIRE in not only locating and inducing expression of target genes, but also in co-ordinating the expression of multiple genes, distributed throughout the genome across multiple chromosomes, as it has been revealed that not only genes separated by tens of Mbps on the same chromosome, but also those located on different chromosomes can associate and share these sites (Osborne *et al.* 2004; Osborne *et al.* 2007; Sexton *et al.* 2007).

However, once the initial trial of locating individual target genes from the vast network of DNA has been overcome, regulatory proteins and transcription factors will then be confronted with the next challenge; gaining access to their recognition sequence, since all DNA in the nucleus is packaged into chromatin; a complex of proteins and

DNA. It was initially believed that these proteins, referred to as histones, around which the DNA is tightly bound, acted as a simple, inert scaffold, there to aid in the compaction of DNA, however it was found that histones are subject to a plethora of modifications that correlated with gene activity. This led to the idea that histones are dynamic players in essential DNA-based processes including DNA replication, repair, recombination and transcription. In order to initiate gene expression, transcriptional regulators, including AIRE, will need to interact either directly or indirectly with their specific DNA binding sites. However, most of these sequences will be buried inside the compact chromatin and therefore activation of a gene requires selective disruption of the folded structures and exposure of naked DNA.

The principal packaging element of DNA is the nucleosome; a disc-shaped octamer containing two copies of each histone H2A, H2B, H3 and H4, which form four histone fold heterodimers of H3-H4 and H2A-H2B, around which 147bp of DNA are wrapped 1.67 times and stabilised by linker histone H1 (Figure 1.6) (Davie 1997; Luger *et al.* 1997; Barski *et al.* 2007; Turner 2007; Li and Shogren-Knaak 2008; Probst *et al.* 2009). Nucleosomes tend to position themselves on the DNA at fixed intervals, a process known as phasing, with short stretches of linker DNA (typically 30-60bp) connecting adjacent nucleosomes, the length of which can differ considerably depending on the organism and the compaction state of the DNA (Tremethick 2007; Bassett *et al.* 2009; Jiang and Pugh 2009). The repeating nucleosomal 'beads on a string' array is however the primary structure of chromatin, which can further assemble into higher-order secondary and even tertiary structures of increasing

Figure 1.6 – Structure of the Nucleosome Core Particle and Higher-Order Packaging of Chromatin



A, The basic building block of chromatin is the nucleosome; a histone octamer comprising two H2A-H2B and two H3-H4 heterodimers, around which 147 base pairs (bp) of double-stranded DNA (dsDNA) are wrapped 1.67 times and stabilised by linker histone H1 (inset) (Luger *et al.* 1997). Each histone is composed of a globular domain and amino- (NH_2) and carboxyl- (COOH) terminal tails which are all subject to multiple post-translational modifications. *B*, Compaction of the 10nm nucleosomal array generates a two-start double helical 30nm fibre, which is then looped and further compacted into heterochromatic higher order structures (Tremethick 2007). These eventually form the most compact structure of chromatin; the mitotic chromosomes. Figure adapted from (Schones and Zhao 2008; Probst *et al.* 2009).

complexity, the extreme being condensed metaphase chromosomes (Figure 1.6) (Sterner and Berger 2000).

In non-replicating cells, generally chromatin is compartmentalised into defined domains, distributed throughout the nuclear environment in more condensed regions (heterochromatin) and more open or decondensed regions (euchromatin) (Figure 1.6). Euchromatic regions of chromosomes are generally more flexible and open, due to irregular spacing of nucleosomes (Henikoff 2008; Schones *et al.* 2008; Jiang and Pugh 2009). This therefore increases the accessibility for transcriptional regulators and DNA repair and replication complexes, meaning genes in euchromatic areas are generally considered active, and that these stretches of DNA are replicated early in S-phase (Kouzarides 2007). In contrast, heterochromatin has few genes and is rich in repeating units of certain nucleotide bases. It is frequently transcriptionally silent and replicates late in S-phase as regularly spaced nucleosomes decrease the accessibility of the DNA (Richards and Elgin 2002). It is therefore the nucleosome that shapes the DNA molecule, from the atomic level; bending the DNA strand around the histone core, through to the large-scale influence on whole genes; compacting the DNA into higher-order helices (Luger *et al.* 1997).

1.9. Structure of the Core Histone Proteins

Core histone proteins are one of the most highly conserved amongst eukaryotes (Felsenfeld and Groudine 2003). From the determination of a high resolution crystal structure in 1997, Luger *et al* revealed in detail the numerous histone-DNA and

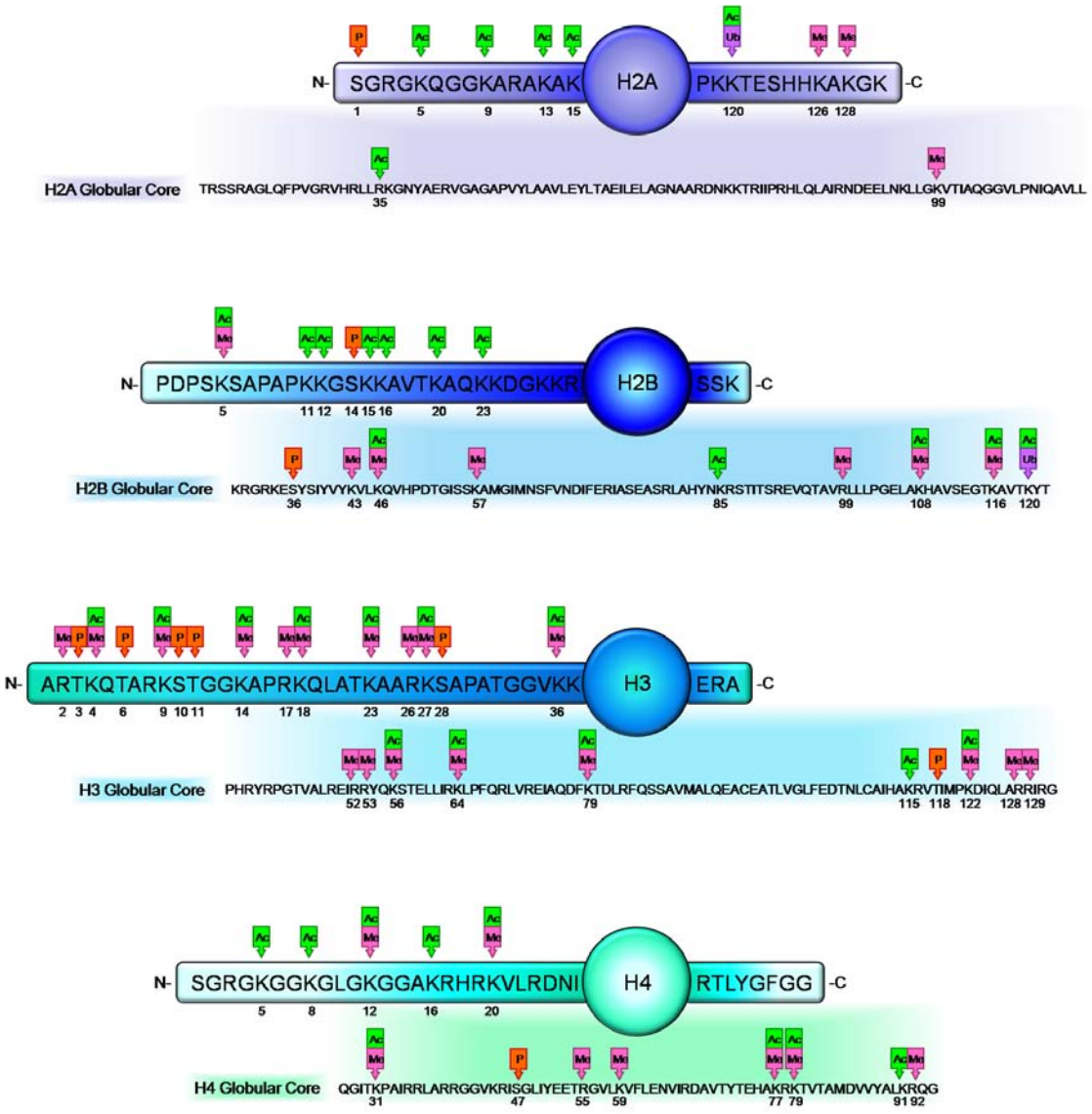
histone-histone interactions within the nucleosome (Luger *et al.* 1997). Individual histones are composed of a structural α-helical globular domain, forming the scaffold for the DNA, however, each histone also possesses highly basic, flexible amino (NH_2-) and carboxyl (COOH-) terminal tail sequences which make up approximately 28% of their mass (Luger *et al.* 1997). These histone tails contribute significantly to the structure and function of the nucleosome as they pass through the gyres of DNA into the nuclear space (Luger *et al.* 1997). In particular, amino acids 16-25 of the H4 tail interact with the exposed face of the H2A-H2B dimer of the adjacent nucleosome (Luger *et al.* 1997). The histone tails, along with the globular core domains, are subject to a vast array of covalent and non-covalent post-translational modifications including; acetylation of lysine residues; methylation of lysines and arginines; phosphorylation of serines and threonines; ubiquitylation of lysines; sumoylation of lysines; and ADP ribosylation of glutamic acid residues, which can all influence the way in which the underlying DNA is controlled (Santos-Rosa *et al.* 2002; Kouzarides 2007).

Research into the vast range of post-translational histone modifications highlighted how these marks impact upon the identity of each cell in a variety of ways. This diversity of modifications, coupled with the finding that several correlated with the expression level of the corresponding genes, indicated an enormous potential for these histone modifications to act as gateways, regulating access to DNA through complex combinations of post-translational codes, and it is for this reason that much work has focused on mapping of these marks, with the hope to unveil their specific roles in regulation of gene expression and chromatin structure.

1.10. POST-TRANSLATIONAL HISTONE MODIFICATIONS

The first clues into a role for post-translational modification of histone side chains came in 1964, when Allfrey and Mirsky noted that when histones were acetylated, the rate of RNA synthesis increased, however, as investigation into the replication and transcription of DNA then took centre stage, it was not until much later that the key link between histone modifications and control of transcription was again brought to the forefront (Allfrey *et al.* 1964; Allfrey 1966; Fuchs *et al.* 2009). Two significant reports; the finding that a human homologue of yeast KAT2/Gcn5 (general control nonderepressible 5), a well-characterised enzyme known to be involved in transcriptional activation, was able to acetylate histones directly; and the discovery of a human protein capable of removing acetylation groups from histones, linking this activity to transcriptional repression, provided the first real biological interpretation for Allfrey and Mirsky's observations in the 1960's (Brownell *et al.* 1996; Taunton *et al.* 1996; Nagy and Tora 2007; Fuchs *et al.* 2009). These works kick-started an extensive body of research, resulting in the identification of numerous post-translational modifications occurring at specific residues both along the histone tails and within the globular cores (Figure 1.7), the majority of which have been found to exert considerable influence over the control of DNA-based processes. Since the advancement of antibodies specific to individual histone marks, techniques which capture a snapshot of chromatin have enabled the study of the levels of these modifications across specific gene regions. Chromatin immunoprecipitation (ChIP); a technique used to analyse the genomic location of chromatin-associated proteins or modified histones, has been combined with DNA microarrays (ChIP-chip), SAGE

Figure 1.7 – The Conserved Location of Key Post-Translational Histone Modifications



The four histone proteins, H2A, H2B, H3 and H4, all contain a structural globular core domain and flexible N- and C-terminal tail regions, which protrude through the gyres of DNA on the surface of the nucleosome. A highly diverse array of covalent post-translational histone modifications can be deposited on specific, conserved residues across all histone domains. These modifications include acetylation (Ac) of lysine residues, methylation (Me) of lysines and arginines and phosphorylation (P) of serines and threonines, ubiquitylation (Ub) of lysines and glutamic acid residues, along with sumoylation of lysines and ADP ribosylation of glutamic acid residues (not displayed). The positions of a range of histone modifications are indicated across the flexible tails and globular cores of all four histones (Zhang and Reinberg 2001; Kouzarides 2007). Figure adapted from (Cosgrove 2007).

(ChIP-SAGE), and more recently with massively parallel sequencing (ChIP-Seq), thus allowing the profiling of multiple global histone modifications (O'Neill and Turner 2003; Kouzarides 2007; Schones and Zhao 2008). The results are often varied, highlighting their dynamic nature and non-uniform distribution, however, certain key commonalities have arisen, particularly with regards to acetylation and methylation, the two most extensively studied modifications (Kouzarides 2007).

1.11. Acetylation and Deacetylation of Histones

Histone acetylation is one of the most prevalent post-translational histone modifications, yet it is restricted to conserved lysine (K) residues across the core histones (Figure 1.7). The addition of an acetyl moiety to the ε-amino group of lysine from the coenzyme acetyl-CoA has a significant effect upon the amino acid chemical properties, neutralising the basic charge and generating ε-*N*-acetyllysine (Fuchs *et al.* 2009). This has been predicted to affect the structure of the nucleosome and its interaction with the DNA, as unmodified, positively-charged lysines bind and therefore stabilise negatively-charged DNA. Disruption of the positive charge may directly influence the compaction of chromatin fibres through alterations to the way in which the DNA molecule is wrapped around the nucleosome core, making the DNA more accessible to ATP-dependent nucleosome remodelling factors and DNA-binding proteins involved in transcription, DNA repair and replication (Krajewski and Becker 1998; Akhtar and Becker 2000; Sterner and Berger 2000; Carrozza *et al.* 2003; Nagy and Tora 2007; Bassett *et al.* 2009). Acetylated lysines themselves can also facilitate this process as many remodelling factors possess bromodomains,

which recognise and bind to this modification (Fuchs *et al.* 2009). In addition, the more recent discovery that the core histone globular domains are also subject to an array of modifications, supports these theories. Several of these modified residues are positioned on the lateral surface of the nucleosome and therefore able to disturb histone-DNA interactions [reviewed in Cosgrove (2007)]. One such example involves acetylation of lysine 56, a residue found in the globular domain of histone H3 (H3K56ac). As one of the key positively-charged amino acids electrostatically bound to the DNA, mutation of this residue to prevent acetylation results in decreased gene accessibility through a reduction in nucleosome mobility (Masumoto *et al.* 2005; Chen *et al.* 2008). Nucleosome-nucleosome contacts are also affected greatly by acetylation. Specifically, H4K16ac has been shown to impact directly on the formation of higher-order structures, through the disruption of the interaction between this residue with the acidic patch of the H2A-H2B dimer on the adjacent nucleosome (Luger *et al.* 1997; Shogren-Knaak *et al.* 2006). In this regard, acetylation is generally considered a mark of open, active chromatin domains and in line with this, actively-transcribed genes are typically enriched within promoter regions and at the 5`end of their coding regions with high levels of acetylation (Kouzarides 2007). However, it would appear that this cannot be explained through a bulk acetylation of the histone proteins, but instead by acetylation of specific lysine residues, which, on the N-terminal tail of H4 for example, have been shown to be acetylated preferentially in a defined order, with H4K16 being the first to be modified, followed by K12, K8 and finally K5 (Turner and Fellows 1989; Turner *et al.* 1989). In a study by Roh *et al*, whole-genome mapping of histone tail modifications H3K9ac and H3K14ac in human T-cells revealed more than 46,000 sites across the genome which carried these

marks (Roh *et al.* 2005). These 'acetylation islands' were found to correlate with transcriptional start sites (TSS) of active genes, CpG islands and functional regulatory elements, particularly within gene-rich regions and interestingly, upon T-cell activation and TCR signalling, 4045 additional acetylation loci were induced, reflecting the increase in gene activation (Roh *et al.* 2005). However, acetylation of H4K16, K12, K8 and K5 within euchromatic genes in a lymphoblastoid cell line, has also been shown to bear no influence on actual levels of expression, but instead defines chromatin states; distinguishing enriched euchromatic regions from depleted heterochromatin (O'Neill and Turner 1995; Johnson *et al.* 1998).

Histone acetylation is a highly dynamic process, with a very rapid turnover rate and a half-life of just minutes, therefore its influence on the DNA is transient (de Ruijter *et al.* 2003; Liu *et al.* 2005). The overall levels of acetylation across chromatin are determined by histone acetyltransferases (HATs), which catalyse the addition of acetyl moieties to lysine residues and histone deacetylase (HDAC) enzymes, which counteract this by removing the acetyl group. Many HATs and HDACs have been identified to date, the majority of which have demonstrated the ability to act on multiple lysine residues across all four histones, and also on non-histone proteins (Nagy and Tora 2007; Haberland *et al.* 2009). HATs can be classified into five families, which include; the GCN5-related *N*-acetyltransferases (GNATs); p300/CBP HATs; the MYST (MOZ, Ybf2/Sas3, Sas2 and Tip60)-related HATs; the general transcription factor HATs; and the nuclear hormone-related HATs (Table 1.1) (Nagy and Tora 2007). Members of each of these groups have demonstrated numerous activities, however, the common theme that arises is of a clear and direct connection

Table 1.1 – Summary of Common Histone Acetyltransferase Enzymes

HAT	ORGANISM	HISTONE SPECIFICITY
GNAT SUPERFAMILY		
KAT1/HAT1	Various (yeast to humans)	H4 (K5, K12)
KAT2/GCN5	Various (yeast to humans)	H3 (K9, K14, K18) / H2B
KAT2B/PCAF	Humans / Mice	H3 (K9, K14, K18) / H2B
KAT9/ELP3	Various (yeast to humans)	H3
KAT10/Hap2	Yeast	H3 (K14) / H4
P300/CBP HATS		
KAT3A/CBP	Various (multicellular)	H2A / H2B / H3 / H4
KAT3B/p300	Various (multicellular)	H2A / H2B / H3 / H4
MYST-RELATED HATS		
KAT5/TIP60	Various (yeast to humans)	H2A (yeast K4, K7; chicken K5, K9, K13, K15) / H4 (K5, K8, K12, K16)
KAT6/Sas3	Yeast	H3 (K14, K23)
KAT6A/MOZ	Various (yeast to humans)	H3 (K14)
KAT6B/MORF	Humans	H3 (K14)
KAT7/HBO1	Various (yeast to humans)	H4 (K5, K8, K12) > H3
KAT8/HMOF	Various (yeast to humans)	H4 (K16)
GENERAL TRANSCRIPTION FACTOR HATS		
KAT4/TAF	Various (yeast to humans)	H3 > H4
KAT12/TFIIIC90	Humans	H3 (K9, K14, K18)
NUCLEAR HORMONE-RELATED HATS		
KAT13A/SRC1	Humans / Mice	H3 / H4
KAT13B/ACTR	Humans / Mice	H3 / H4
KAT13C/P160	Humans / Mice	H3 / H4

A selection of common histone acetyltransferases (HATs), divided into their appropriate family, the typical organisms in which they are found, and the specific histone residues upon which they act. Novel nomenclature is displayed prior to original names. Devised from (Sterner and Berger 2000; Allis *et al.* 2007; Kouzarides 2007).

between acetylation and transcriptional activation (Sterner and Berger 2000). Currently, 11 classical HDACs have been identified in humans (Table 1.2), which can be separated into four major families according to sequence similarity; class I HDACs (HDAC1, 2, 3, 8); class IIa (HDAC4, 5, 7, 9); class IIb (HDAC6, 10); and class IV (HDAC11) (Haberland *et al.* 2009). In addition, a further group of non-classical NAD-dependent HDACs, the sirtuins, are often referred to as class III HDACs (Gallinari *et al.* 2007; Haberland *et al.* 2009). Consistent with the role of acetylation in transcriptional activation, deacetylation, particularly at gene promoter regions, is generally associated with gene silencing, nucleosome stabilisation and diminished nucleosome-remodelling (Fuchs *et al.* 2009; Haberland *et al.* 2009).

1.12. Histone Methylation and Demethylation

In contrast to acetylation, methylation of histones is significantly more intricate. Firstly, methylation can occur both on conserved lysine residues, and arginines (R) across all four histone proteins, and secondly, as an added layer of complexity, up to three methyl moieties can be applied to the lysine ε-amino group (mono-, di- and tri-methylation), while arginines can be mono- or di-methylated, with the latter being either symmetric or asymmetric in its distribution (Zhang and Reinberg 2001; Fuchs *et al.* 2009). Each degree of methylation can potentially represent a different biological outcome, depending on the specific lysine or arginine methylated, and its location in a gene (Zhang and Reinberg 2001; Fuchs *et al.* 2009). The role histone methylation plays in the control of DNA-based processes is extensive, however, this is presumably orchestrated solely through recruitment of additional regulatory factors,

Table 1.2 – Summary of Known Histone Deacetylase Enzymes

HDAC	LOCALISATION	TISSUE DISTRIBUTION
CLASS I		
HDAC1	Nucleus	Ubiquitous
HDAC2	Nucleus	Ubiquitous
HDAC3	Nucleus	Ubiquitous
HDAC8	Nucleus	Ubiquitous? Smooth muscle differentiation
CLASS IIA		
HDAC4	Nucleus / Cytoplasm	Heart / Skeletal Muscle / Brain
HDAC5	Nucleus / Cytoplasm	Heart / Skeletal Muscle / Brain
HDAC7	Nucleus / Cytoplasm	Heart / Placenta / Pancreas / Skeletal Muscle
HDAC9	Nucleus / Cytoplasm	Heart / Skeletal Muscle / Brain
CLASS IIB		
HDAC6	Mostly cytoplasm	Heart / Liver / Kidney /Pancreas
HDAC10	Mostly cytoplasm	Liver / Spleen /Kidney
CLASS IV		
HDAC11	Nucleus / Cytoplasm	Brain / Heart / Smooth Muscle / Kidney
SIRTUINS (CLASS III)		
SIRT1	Nucleus → H4K16 specific	Brain / Smooth Muscle / Kidney / Heart
SIRT2	Cytoplasm → H4K16 specific	Brain / Smooth Muscle / Kidney / Heart / Liver
SIRT3	Mitochondria	Ubiquitous
SIRT5	Mitochondria	Ubiquitous

A representation of all histone deacetylase (HDAC) enzymes known to date, which can act upon multiple histone residues. HDACs are divided into their appropriate class, and their typical localisation within cells and tissue distribution are displayed. Devised from (Michishita *et al.* 2005; Haigis and Guarente 2006; Dokmanovic *et al.* 2007).

since methylation of lysines and arginines does not affect the charge of the residue and thus has no direct impact on histone-histone or histone-DNA binding (Fuchs *et al.* 2009). Yet histone methylation is still believed to play a role in the formation and maintenance of higher-order chromatin structures, with a plethora of methylation marks being reported as essential for establishment of euchromatic or heterochromatic regions of the genome, for example the requirement of tri-methylation (me3) of H3K9 in the formation of silent heterochromatin (Noma *et al.* 2001; Barski *et al.* 2007; Kouzarides 2007).

Histone methyltransferases (HMTs) are enzymes responsible for the addition of methyl groups to either lysine or arginine residues. Methylation of arginines is catalysed by a family of protein arginine methyltransferases (PRMTs), of which at least 11 are known to exist in mammals, some of which show great specificity, however, only relatively little is known about this modification (Smith and Denu 2009; Wolf 2009). PRMTs can be allocated into two main groups; those which generate asymmetric di-methylation (type I PRMTs), and those which bring about symmetric di-methylation (type II PRMTs) (Table 1.3) (Zhang and Reinberg 2001; Smith and Denu 2009; Wolf 2009). In contrast, histone lysine methylation is very well characterised with regards to both its roles in regulating DNA-based processes and the enzymes responsible for addition and removal of this moiety. Lysine HMTs can be divided into two main groups; those containing a SET domain; and those with a DOT1 domain, which frequently show exclusivity; preferentially methylating a limited number, or often only one specific lysine residue, to only one particular degree

Table 1.3 – Summary of Known Protein Arginine Methyltransferase Enzymes

PRMT	LOCALISATION	HISTONE SPECIFICITY
TYPE I		
PRMT1	Nucleus / Cytoplasm	H4 (R3)
PRMT4	Nucleus	H3 (R2, R17, R26) / H4 (R3)
PRMT6	Nucleus	H3
PRMT8	Plasma Membrane	H4
TYPE II		
PRMT5	Cytoplasm	H2A / H3 (R8) / H4 (R3)
PRMT7	Nucleus / Cytoplasm	H2A / H4
PRMT9	Nucleus / Cytoplasm	H2A / H4

A representation of all protein arginine methyltransferase (PRMT) enzymes known to date, divided into their appropriate types, their typical localisation within cells and the specific histone residues upon which they act. Devised from (Pahlich *et al.* 2006; Wolf 2009).

(mono-, di- or tri-), and also within a restricted genomic location (Table 1.4) (Thomas *et al.* 2008; Smith and Denu 2009).

Originally it was believed that methylation represented a more stable, permanent histone modification, as the global turnover rate of this mark was low in comparison to acetylation (Ng *et al.* 2009). However, the relatively recent discovery of families of enzymes capable of removing methyl groups from histones has since shown that this mark is equally as dynamic. These fall into a number of different groups, of which there are two major classes; those typified by lysine specific demethylase (LSD1, recently re-named KDM1), which can only remove mono- and di-methyl marks; and a family of *Jumonji-C* (JmjC) domain-containing histone demethylases (JHDMs), which can de-methylate all three levels of histone lysine methylation (Table 1.5) (Cloos *et al.* 2008; Ng *et al.* 2009).

1.13. The Role of Post-Translational Histone Modifications in Transcriptional Regulation

Even in its most uncondensed form, nucleosomal DNA is essentially repressive for transcription, hence nucleosomes must be moved and it has been shown that histones are able to *cis*-translocate or slide along the DNA fragment considerable distances and also *trans*-displace, being entirely removed from the DNA (Pennings *et al.* 1991; Whitehouse *et al.* 1999). This process requires considerable energy input as the DNA molecule is tightly bound to the core histone proteins, with more than 120 direct atomic interactions (Pennings *et al.* 1991; Whitehouse *et al.* 1999). It is

Table 1.4 – Summary of Common Histone Lysine Methyltransferase Enzymes

KMT	ORGANISM	HISTONE SPECIFICITY
SET DOMAIN KMTS		
KMT1/Su(Var)3–9	Flies / Yeast	H3K9
KMT1A/SUV39H1	Humans / Mice	H3 (K9me3)
KMT1B/SUV39H1	Humans / Mice	H3 (K9me3)
KMT1C/G9a	Humans	H3 (K9me1, me2)
KMT1E/ESET/SETDB1	Humans	H3 (K9me3)
KMT2/Set1/COMPASS	Yeast	H3 (K4me1, me2, me3)
KMT2A/MLL1/Trx	Humans / Flies	H3 (K4me1, me2, me3)
KMT2B/MLL2/Trx	Humans / Flies	H3 (K4me1, me2, me3)
KMT2C/MLL3/Trr	Humans / Flies	H3 (K4me1, me2, me3)
KMT2D/MLL4/Trr	Humans / Flies	H3 (K4me1, me2, me3)
KMT2E/MLL5	Humans	H3 (K4me1, me2, me3)
KMT2F/hSET1A	Humans	H3 (K4me1, me2, me3)
KMT2G/hSET1B	Humans	H3 (K4me1, me2, me3)
KMT3A/SET2	Humans	H3 (K36)
KMT5/Set9	Yeast	H4 (K20)
KMT6/Ezh2/E(Z)	Humans / Flies	H3 (K27)
DOT DOMAIN KMTS		
KMT4/DOT1L/Dot1	Humans / Yeast	H3 (K79)

A selection of common histone lysine methyltransferases (KMTs), divided into their appropriate groups, the typical organisms in which they are found, and the specific histone residues upon which they act. Novel nomenclature is displayed prior to original names. Devised from (Zhang and Reinberg 2001; Allis *et al.* 2007; Shilatifard 2008; Thomas *et al.* 2008; Smith and Denu 2009).

Table 1.5 – Summary of Common Histone Demethylase Enzymes

HDM	ORGANISM	HISTONE SPECIFICITY
LSD1 KDMs		
KDM1/LSD1/BHC110	Various (yeast to humans)	H3 (K4me1, me2, K9me1, me2)
JHDM1 KDMs		
KDM2/Jhd1	Yeast	H3 (K36me1, me2)
KDM2A/JHDM1a/FBXL11	Humans	H3 (K36me1, me2)
JHDM2 KDMs		
KDM3A/JHDM2a	Humans	H3 (K9me1, me2)
KDM3B/JHDM2b	Humans	H3 (K9)
JMJD2 KDMs		
KDM4/Rph1	Yeast	H3 (K9, K36me2, me3)
KDM4C/JMJD2C/GASC1	Humans	H3 (K9, K36me2, me3)
KDM4D/JMJD2D	Humans	H3 (K9me2, me3)
JARID1 KDMs		
KDM5/Lid/Jhd2/Jmj2	Flies / Yeast	H3 (K4me2, me3)
KDM5A/JARID1A/RBP2	Humans	H3 (K4me2, me3)
KDM5B/JARID1B/PLU-1	Humans	H3 (K4me1, me2, me3)
KDM5C/JARID1C/SMCX	Humans	H3 (K4me2, me3)
KDM5D/JARID1D/SMCY	Humans	H3 (K4me2, me3)
UTX/JMJD3 KDMs		
KDM6A/UTX	Humans	H3 (K27me2, me3)
KDM6B/JMJD3	Humans	H3 (K27me2, me3)
ARGININE-SPECIFIC HDM		
JMJD6	Multicellular	H3 (R2me2) / H4 (R3me2)

A selection of common histone lysine demethylases (KDMs) and the recently discovered arginine-specific methyltransferase. Demethylases are divided into their appropriate groups, and displayed are the typical organisms in which they are found, and the specific histone residues upon which they act. Novel nomenclature is displayed prior to original names. Devised from (Allis *et al.* 2007; Chang *et al.* 2007; Cloos *et al.* 2008; Smith and Denu 2009; Wolf 2009).

therefore catalysed by a group of molecules containing intrinsic ATP-ase activity; ATP-dependent nucleosome remodelling complexes, which work in combination with the modification of histone proteins to allow displacement of the nucleosome and exposure of the raw DNA (Pennings *et al.* 1991; Whitehouse *et al.* 1999; Soutoglou and Talianidis 2002). Through the development of genome-wide nucleosomal maps, coupled with DNA sequencing and microarray hybridisation, the position of every nucleosome across a genome could be determined (Barski *et al.* 2007; Mavrich *et al.* 2008; Schones *et al.* 2008; Jiang and Pugh 2009). These works showed that the majority of genes conformed to a common, stable theme of organisation, with long stretches of raw DNA at the 3` and 5` ends of genes, referred to as the nucleosome free region (NFR) (Barski *et al.* 2007; Mavrich *et al.* 2008; Schones *et al.* 2008; Jiang and Pugh 2009). The discovery of these NFR challenged the view that the promoter regions of silenced genes would be occluded by nucleosomes, thus preventing random transcription as RNA pol II is unable to bind nucleosomal DNA (Bondarenko *et al.* 2006; Jiang and Pugh 2009). However, the finding that these open promoter states occur even in genes with very low transcription rates, which could essentially be classed as turned off, suggests that these NFR are not sufficient to induce transcription (Jiang and Pugh 2009).

The regulation of euchromatin gene expression therefore requires the recruitment of DNA-bound transcription factors to gene promoters. The array of specific post-translational histone modifications of core histones, present on the nucleosome surface, can serve as recruitment signals for protein effectors able to exert functional

effects. This in turn brings about alterations to the patterns of modifications across the locus and ultimately leads to either transcriptional activation or repression.

1.14. Marks of Gene Activation

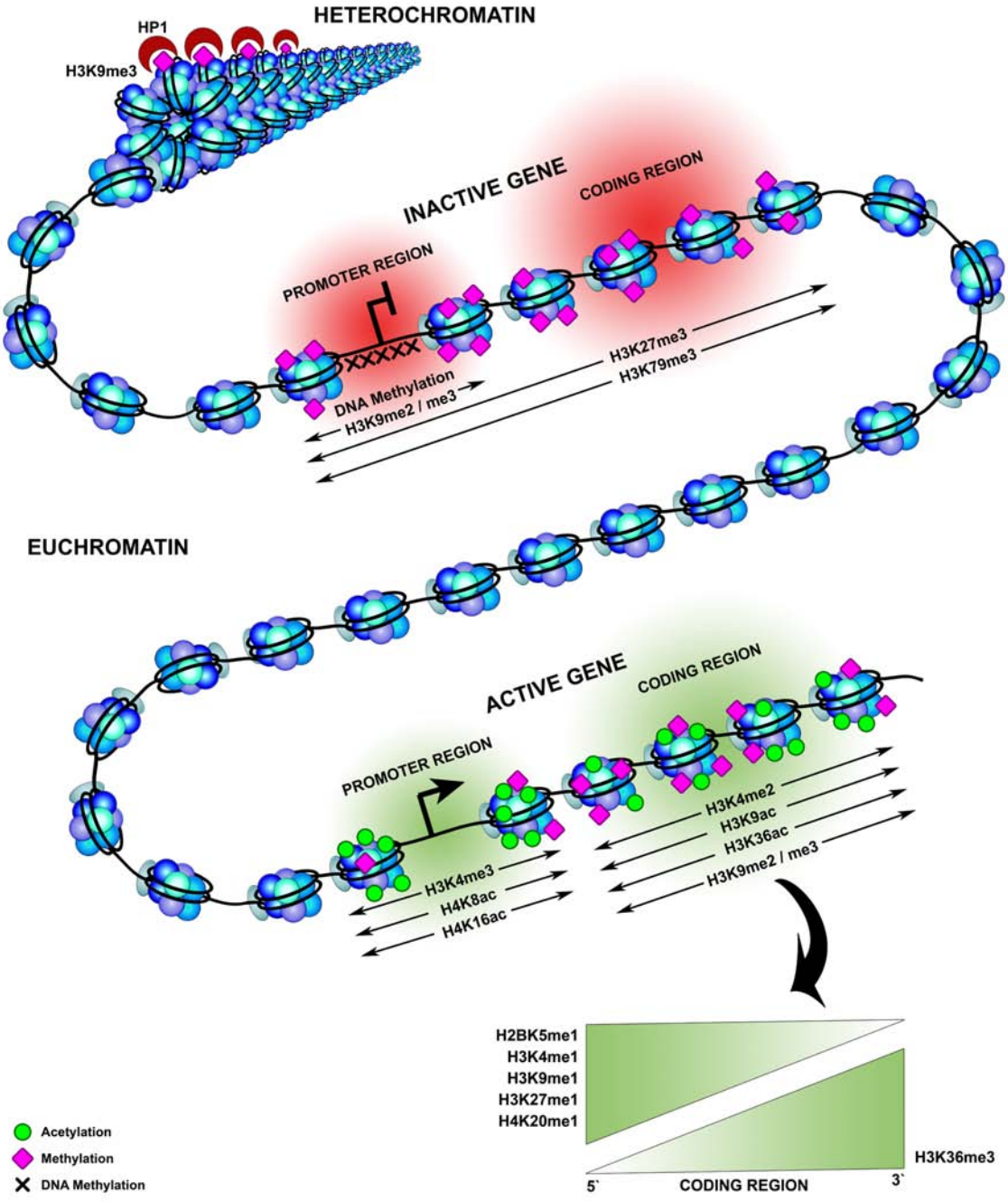
Many histone modifications have been implicated in the activation of gene expression programmes, due to their presence within transcribed loci (Figure 1.8). In general, acetylation of lysine residues is considered a mark of active expression, and many transcriptional regulators and transcription factors such as KAT3A/CBP and the basal transcription factor TFIID, possess intrinsic HAT activity (Sterner and Berger 2000). However, many HATs have been shown to associate preferentially to defined gene regions, leading to the deposition of specific acetylation marks in different locations (Wang *et al.* 2008). For example, the area surrounding transcriptional start sites (TSS) has been shown to be enriched for H2AK9ac, H2BK5ac, H3K9ac, H3K36ac and H4K91ac, whereas the promoter and coding regions demonstrated H2BK20ac, H3K4ac, H4K8ac and H4K16ac, among others (Wang *et al.* 2008).

In contrast, histone methylation is not as straightforward, with certain methyl marks having the potential to influence gene expression in opposing ways, dependent on different conditions. One of the most extensively defined activating modification is H3K4me3, which has been shown to be present at the 5` end of open reading frames (ORFs) along with the serine 5 phosphorylated, initiating form of RNA pol II, as genes are induced, it is known to promote transcription elongation, and also plays a role in the regulation of RNA processing, which may be due in part to the recruitment of

Figure 1.8 - Post-Translational Histone Modifications in the Regulation of Gene Expression

Post-translational modification to histone N- and C-terminal tails and globular cores impact directly on the genome. A selection of common modifications and their predicted roles in the regulation of transcriptional processes are illustrated. The presence of H3K9me3 within heterochromatic regions aids in the recruitment of heterochromatin protein 1 (HP1) (Motamedi *et al.* 2008). However, this modification also plays a role in euchromatin along with H3K9me2; being found in silenced genes and the coding regions of actively transcribed genes (Vakoc *et al.* 2005). Mono-methylation of H3K9 also marks silent genes at their promoters, yet is deposited in the coding regions of active genes (Vakoc *et al.* 2006). Inactive genes are also marked by broad domains of H3K27me3 and H3K79me3 across the entire gene. Methylation of H3K4 is generally perceived as an active modification, with me1 and me2 appearing across the body of transcribed genes, but me3 occurring preferentially at active promoters, where un-methylated CpG islands can also be found (Barski *et al.* 2007). Acetylation modifications such as H4K8ac and H4K16ac are often found peaking in the promoter regions of active genes, whereas H3K9ac and H3K36ac can appear more broadly across the 3` ends of expressed loci, in combination with mono-methylation of H2BK5, H3K4, H3K9, H3K27 and H4K20, which become progressively less enriched towards the 3` end of active genes (Li *et al.* 2007; Schones and Zhao 2008; Wang *et al.* 2008). This is inversely proportional to levels of H3K36me3, which have been found to peak at the 3` end (Barski *et al.* 2007). Figure adapted from (Li *et al.* 2007; Schones and Zhao 2008).

Figure 1.8 - Post-Translational Histone Modifications in the Regulation of Gene Expression



protein effectors known to positively influence transcription (Santos-Rosa *et al.* 2002; Bernstein *et al.* 2005; Bernstein *et al.* 2006; Berger 2007; Sims *et al.* 2007; Shilatifard 2008). Proteins can interact with the H3K4me3 mark through the presence of one of a number of binding modules which include; PHD domains, like those found in AIRE; WD40 domains; and proteins with Royal family domains (chromo-, Tudor- and MBT-domains) (Berger 2007). For example, a PHD finger of BPTF (bromodomain and PHD finger transcription factor); a subunit within the ATP-dependent chromatin remodelling complex NURF, has been shown to bind H3K4me3, leading to targeting of the locus for activation (Wysocka *et al.* 2006). A direct interaction has also been demonstrated between H3K4me3 and the PHD-domain of TAF_{III}, a subunit of TFIID, which enhances gene expression (Vermeulen *et al.* 2007; van Ingen *et al.* 2008). Within the gene body, active loci are typically marked with H2BK5me1, H3K4me1 and me2, H3K9me1, H3K27me1 and H4K20me1, levels of which decrease the further you move from the TSS (Barski *et al.* 2007; Li *et al.* 2007; Schones and Zhao 2008; Wang *et al.* 2008). In addition, the coding regions of active genes also show H3K36me3, however, its distribution contrasts with the previous modifications; peaking instead at the 3` end (Figure 1.8) (Carrozza *et al.* 2005; Joshi and Struhl 2005; Keogh *et al.* 2005). This is thought to be due to an interaction with the elongating form of RNA pol II, phosphorylated at serine 2, within the 3` end of active genes, which, through H3K36me3-induced recruitment of Eaf3, a component of the yeast HDAC Rpd3(S), can lead to preferential removal of any acetylation deposited within the coding region during transcription, thus re-stabilising nucleosomes, restoring the chromatin to its repressive state to prevent aberrant intragenic

transcription (Carrozza *et al.* 2005; Joshi and Struhl 2005; Keogh *et al.* 2005; Berger 2007; Kouzarides 2007).

However, as investigation continues, these general classifications of histone modifications as strictly activating are constantly being challenged by the finding that often, marks which were originally believed to be 'active', can be detected within silent genes (Berger 2007). For instance, certain protein complexes associated with transcriptional repression have been shown to bind to the classically activating modification H3K4me3 including a factor found in the Sin3-HDAC1 deacetylation complex, thus removing acetyl marks and shutting down transcription (Shi *et al.* 2006). In addition, H3K4me3 has been shown to act as a recognition site for the H3K9me3 and H3K36me3-sepcific demethylase KDM4A/JMJD2A, which is found in repressor complexes such as N-CoR (nuclear hormone co-repressor complex) and thus implicates H3K4me3 in gene silencing (Huang *et al.* 2006; Shi *et al.* 2006; Berger 2007). Together, these two interactions present evidence for a link between H3K4me3 and active gene repression, a function which could be dictated by both the timing and location of this mark; with the recruitment of positive-acting effector complexes during transcriptional initiation or elongation, immediately proceeded by binding of negative-acting factors to terminate expression (Berger 2007). Interestingly, the impact H3K36me3 has on gene activity also adheres to no strict rules and has been shown to be highly dependent on its precise distribution pattern. When found at the 3` end of genes, active expression is observed, however, if deposited at promoter regions, H3K36me3 exerts a negative effect upon transcription

(Strahl *et al.* 2002; Bannister *et al.* 2005; Barski *et al.* 2007; Berger 2007; Kouzarides 2007).

1.15. Marks of Gene Repression

Transcriptional repression has always believed to be brought about by H3K9me2 and me3, H3K27me3 and H4K20me3, for example (Barski *et al.* 2007). H3K9me2 and me3 have been detected in the promoter regions of inactive genes and is found associated with heterochromatin protein 1 (HP1), linking this modification with the formation of heterochromatic domains (Noma *et al.* 2001; Barski *et al.* 2007; Motamedi *et al.* 2008). H3K27me3 is also implicated in gene silencing, via unique interactions with Polycomb group proteins (PcG); originally shown to be essential for switching off the developmental homeobox (*Hox*) genes in *Drosophila melanogaster* (Breiling *et al.* 2004; Schwartz and Pirrotta 2007; Henikoff 2008). The PcG system is an extensively studied example of maintenance of gene silencing throughout development, which work alongside *trithorax*-group proteins (trxG), to counteract the repression, should gene expression be required (Breiling *et al.* 2004; Schuettengruber *et al.* 2009). The mammalian Polycomb repressive complex 2 (PRC2) contains a H3K27-specific methyltransferase KMT6/Ezh2, which, in combination with its co-factors Eed (embryonic ectoderm development), Suz12 (Suppressor of *zeste* 12), and MTF2 (metal response element-binding transcription factor), acts together with PRC1 and Pleiohomeotic (PHO) complexes at Polycomb response elements (PREs), to deposit the silencing modification H3K27me3, thus keeping genes off during development, even after the PRC and PHO complexes are

removed (Breiling *et al.* 2004; Cao and Zhang 2004; Dellino *et al.* 2004; Schwartz and Pirrotta 2007; Henikoff 2008; Kondo *et al.* 2008).

The classification of these modifications has also been questioned. For example, the typically silent methylation of H3K9 can also be detected in the coding regions of active genes (Vakoc *et al.* 2005; Squazzo *et al.* 2006; Vakoc *et al.* 2006; Kouzarides 2007). Although this seemingly contradictory distribution may relate to the level of methylation; H3K9me2 and me3 limited to heterochromatic regions and H3K9me1 restricted to active domains, a trend that was confirmed following the profiling of histone methylation across the human genome, instead Vakoc *et al* (2005) demonstrated H3K9me2 and me3 mapping specifically to the transcribed regions of active mammalian genes, including the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*), suggesting that the presence of these marks reflected steady-state transcription rates (Vakoc *et al.* 2005). These modifications were detected in combination with RNA pol II and the γ form of HP1, which had previously been found localised to both heterochromatin and euchromatic areas (Vakoc *et al.* 2005; Vakoc *et al.* 2006). This scenario may, in a similar way to H3K36me3, be due to the recruitment of negative-acting factors to transcribed genes leading to the re-formation of chromatin, preventing irregular transcription, thus H3K9me2 / me3 could be acting as a safety feature to ensure genes are switched off following expression.

1.16. Chromatin Complexities and the Histone Code

In addition to these histone modifications, a number of alternative mechanisms are in place to contribute to the management of the genome, and it is these combined forces that generate the great diversity of cell types that make up a multicellular organism; of which the vast majority share identical genotypes yet have well-defined, individual and stable profiles of gene expression (Goldberg *et al.* 2007). One of the best characterised mechanisms is DNA methylation and its role in the regulation of genes and of chromatin organisation is well defined, particularly during embryogenesis and gametogenesis (Goll and Bestor 2005; Goldberg *et al.* 2007). Methylation of cytosine residues is a common feature of large-genome eukaryotes, occurring mainly in regions of DNA rich in CpG islands; where cytosine nucleotides neighbour guanines along the DNA, linked by a phosphate group (Goll and Bestor 2005). Methyl groups are put in place on the DNA by a group of highly conserved proteins called DNA (cytosine-5) methyltransferases (DNMTs), of which two general classes exist (Freitag and Selker 2005; Goll and Bestor 2005). *De novo* DNMTs act after DNA replication to set up methylation patterns early in development, and maintenance DNMTs preserve these patterns throughout cell division and semi-conservative DNA replication by the addition of methyl groups to hemi-methylated CpG dinucleotides (Goll and Bestor 2005; Goldberg *et al.* 2007). DNA methylation influences a number of different cellular processes including silencing of regions of repetitive DNA and centromeric sequences, the inactivation of the second X chromosome in females, mammalian imprinting, where genes are expressed either from the allele inherited from the mother or from the father, and transcriptional

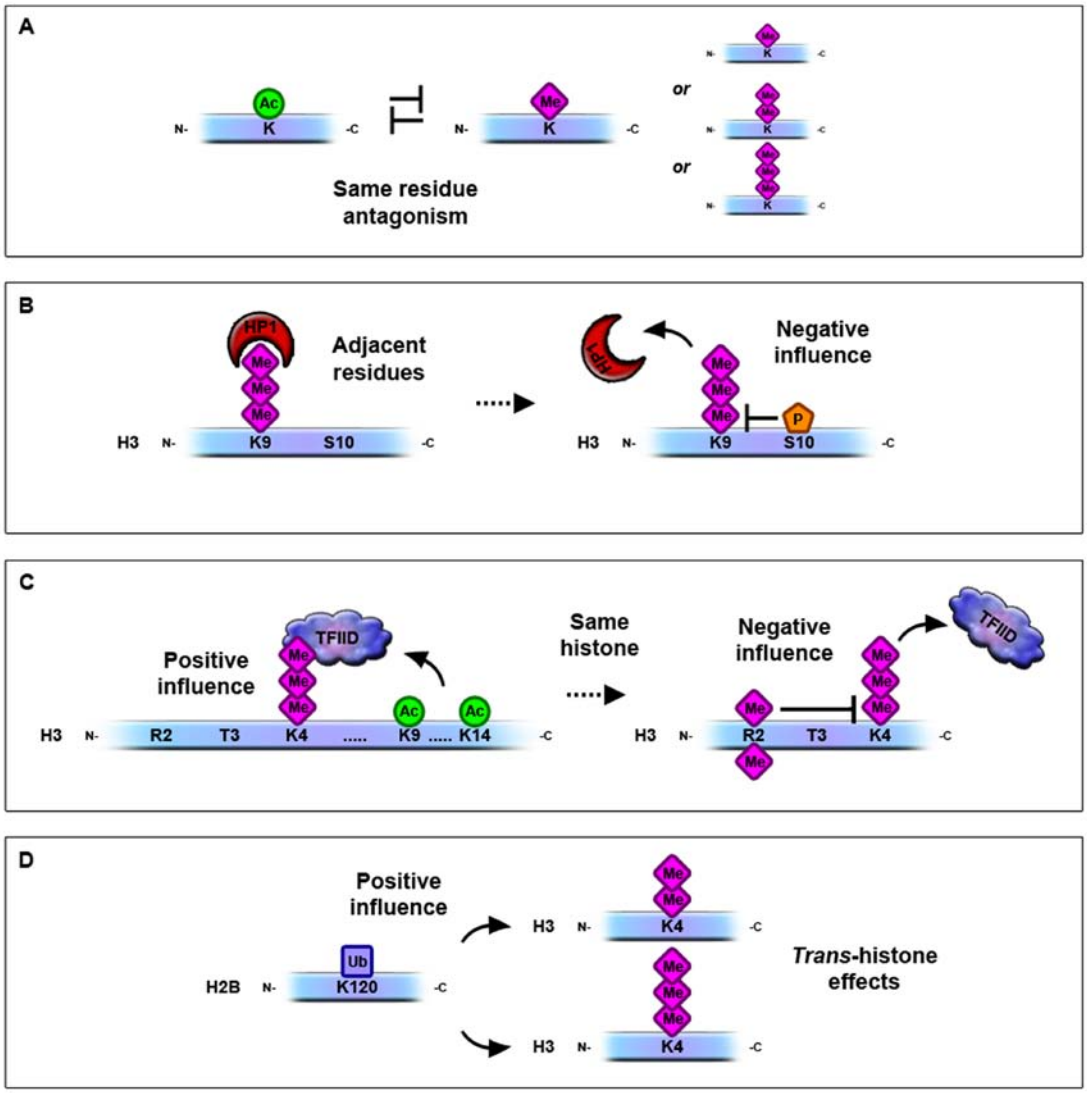
regulation (Reik and Walter 2001; Goldberg *et al.* 2007; Li and Shogren-Knaak 2008). Approximately 76% of human promoters contain high concentrations of CpG islands, and the methylation of these regions correlates significantly with transcriptional repression (Goll and Bestor 2005; Goldberg *et al.* 2007). Non-coding RNAs such as repeat-associated small-interfering (si)RNAs, small RNAs in yeast, micro RNAs and *Xist* RNA also contribute to variation in the chromatin template through the introduction of stable silencing of genes and repetitive DNA sequences, that can be inherited during cell division (Goldberg *et al.* 2007; Marks *et al.* 2009).

DNA methylation, in combination with non-coding RNAs and post-translational histone modifications, which have evolved to alter the phenotype of a cell without any alteration to the underlying DNA sequence, are collectively known as epigenetics. Often, these mechanisms do not act alone to alter the chromatin landscape, and significant 'crosstalk' between epigenetic pathways has been observed (Goldberg *et al.* 2007). One prominent example is X chromosome inactivation, where non-coding *Xist* RNA coating the inactive X, DNA methylation imprinting, and histone modification alterations, act in concert to shut down the inactive X (Okamoto *et al.* 2004; Goldberg *et al.* 2007; Marks *et al.* 2009). High-density genomic tiling arrays have also recently been employed to show that imprinted gene clusters in mice are marked with high levels of DNA methylation, in combination with overlapping H3K4me3 and H3K9me3 domains (Dindot *et al.* 2009). Epigenetic crosstalk frequently occurs between DNMTs and chromatin-modifying enzymes which deposit or remove histone tail modifications (Goldberg *et al.* 2007). The DNMT-like protein DNMT3L is one such example of a protein able to amalgamate epigenetic signals

(Deplus *et al.* 2002; Freitag and Selker 2005; Goldberg *et al.* 2007). DNMT3L is an accessory protein essential to the germline *de novo* DNMTs, which through interaction with HDAC1, brings about gene repression (Deplus *et al.* 2002; Freitag and Selker 2005; Goldberg *et al.* 2007).

The traditional view of epigenetic histone modifications as static on/off switches in the control of gene expression is also now being replaced with the idea that these marks are more dynamic, acting in concert to direct gene expression through organisation of the chromatin structure and recruitment of key effector complexes, a fundamental concept of the histone code hypothesis. In support of this, certain modifications are also able to crosstalk; with one mark influencing the deposition or removal of another, or modulating the binding of effector proteins (Nightingale *et al.* 2006; Kouzarides 2007; Suganuma and Workman 2008). This could be achieved through a number of different ways (Figure 1.9). The simplest form of crosstalk is the antagonism between differing modification states on lysine and arginine residues, where addition of one moiety can block further modification, which is the case for methylation and acetylation, and also for the mutually exclusive mono-, di- and tri-methylation states, thus if a lysine residue is methylated, it cannot also be acetylated, and similarly, only one level of methylation can be present; either me1, me2 or me3 (Latham and Dent 2007). The next level of crosstalk involves recruitment or removal of a protein complex by an adjacent modification, as is the case for phosphorylation of serine 10 on H3 which is necessary for the displacement of H3K9me3-bound HP1 during M-phase of the cell cycle (Figure 1.9) (Fischle *et al.* 2005). Alternatively, one enzyme complex can be affected by multiple histone modifications, which has been shown to

Figure 1.9 – Epigenetic Crosstalk



Broad arrays of epigenetic marks are able to act in concert to influence the deposition or removal of separate modifications or of effector proteins in a process termed 'crosstalk'. *A*, Different modification states on the same lysine (K) residue can antagonise each other thereby addition of an acetyl mark (Ac) can block methylation (Me), furthermore only one level of methylation can be present; either mono- (me1), di- (me2) or tri- (me3) (Latham and Dent 2007). *B*, Adjacent histone modifications can affect the recruitment or removal of a protein complex *in cis*, for example phosphorylation (P) of serine S10 on histone H3 disrupts the binding of heterochromatin protein 1 (HP1) to H3K9me3 (Fischle *et al.* 2005). *C*, Alternatively, one enzyme complex can be affected by multiple histone modifications, for example H3K9ac and H3K14ac enhance the binding of the general transcription factor TFIID to H3K4me3, but asymmetric di-methylation of H3 arginine R2 prevents it (Vermeulen *et al.* 2007). *D*, Crosstalk can also occur *in trans*, for example mono-ubiquitylation (Ub) of H2BK120 is required for H3K4me2 and H3K4me3 deposition (Li *et al.* 2007; Shilatifard 2008).

be the case for the binding of TFIID to H3K4me3; with asymmetric di-methylation of arginine 2 on histone H3 (H3R2me2a) impacting negatively on this interaction, but H3K9ac and H3K14ac enhancing it (Figure 1.9) (Johnson *et al.* 1998). Interestingly, H3R2me2a has been shown to prevent tri-methylation of H3K4 through inhibition of the HMT KMT2/Set1p required to add methyl groups to H3K4, which may explain the negative effect on H3K4me3 binding to the general transcription factor (Allis *et al.* 2007; Guccione *et al.* 2007; Kirmizis *et al.* 2007; Vermeulen *et al.* 2007). Crosstalk can even involve *trans*-histone effects, where modifications on different histone proteins can regulate each other, such as mono-ubiquitylation of K120 of histone H2B (H2BK120ub1) which is required for di- and tri-methylation of H3K4 (Li *et al.* 2007; Shilatifard 2008). These new insights into the cross-regulation of histone modifications, coupled with the identification of novel marks, challenge the paradigms of known modifications and highlight their enormous regulatory potential, both individually and in combination with each other. Genome-wide analysis of histone modifications has allowed the study of a large number of marks and has revealed some emerging themes with respect to particular genes or gene regions. For example, in a study by Wang *et al* (2008), a panel of 16 modifications were found to preferentially co-localise across the human genome at the level of individual nucleosomes, and that the promoters harbouring this 'backbone' of marks, which represent 25% of human promoters, usually displayed higher levels of gene expression (Wang *et al.* 2008). Of these 16 modifications, acetylation was the most prominent (H2BK5ac, H2BK12ac, H2BK20ac, H2BK120ac, H3K4ac, H3K9ac, H3K18ac, H3K27ac, H3K36ac, H4K5ac, H4K8ac and H4K91ac), with methylation also playing a role (H3K4me1, H3K4me2, H3K4me3 and H3K9me1) (Wang *et al.*

2008). However, as this pattern was only reflected in a quarter of genes, it is likely that alternative combinations exist, and that the modifications will fluctuate throughout the course of development to direct transcriptional events.

1.17. Histone Modifications in the Control of Development

A prominent demonstration of how histone modifications are able to impact on developmental gene expression was revealed following investigations into the epigenome of mouse embryonic stem (ES) cells and pluripotent cell populations (Bernstein *et al.* 2005; Azuara *et al.* 2006; Bernstein *et al.* 2006; Bernstein *et al.* 2007; Kouzarides 2007). It was found that the chromatin of these undifferentiated cells harbours domains of the classical silencing modification H3K27me3, coinciding with more punctate regions of H3K4me3, a mark which at the time, was considered to be present only within active genes (Azuara *et al.* 2006; Bernstein *et al.* 2006). These regions, aptly named 'bivalent domains' due to the occurrence of opposing histone modifications at the some locus on the same chromosome, were mainly found at genes encoding transcription factors essential for development. Within the pluripotent ES cells, these genes were essentially silent, showing only very low levels of 'leaky' gene expression, yet they assumed an early-replicating open chromatin structure, usually indicative of highly expressed euchromatic loci, thus it was anticipated that these genes, although silenced in the undifferentiated state, were poised for future gene expression upon development (Azuara *et al.* 2006; Bernstein *et al.* 2006). This was confirmed through induced differentiation of the cells, showing that genes which switched on retained the active H3K4me3 mark and lost the

silencing H3K27me3, and for genes which remained off, the reverse was seen, leaving relatively few loci with the bivalent characteristic (Bernstein *et al.* 2006). One issue that these studies did not address referred to how the fate of each bivalent gene was decided upon commitment of the cells to a particular lineage, however, a recent study has shed further light on this. By monitoring genome-wide alterations to histone modifications throughout development of hematopoietic stem cells (HSC) into erythrocyte precursors, Cui *et al* (2009) found that those genes which subsequently lose the H3K27me3 mark to become active were associated with high levels of H3K4me1, H3K9me1, H4K20me1 and RNA pol II within the undifferentiated state, and that those which remained silent and lost H3K4me3, displayed none of these additional chromatin marks within the HSCs (Cui *et al.* 2009). Thus it would appear that once again, it is the signature of combinatorial histone marks across genomic loci that determine the outcomes of individual genes during development. Yet even this bivalent domain concept has been questioned, with a recent report by Golebiewska *et al* (2009) suggesting a more dominant role for H3K9me2 in the silencing of certain bivalent genes in human ES cells, rather than H3K27me3 (Golebiewska *et al.* 2009). It was not until the first stages of lineage commitment that the H3K27me3 mark was seen to increase, suggesting that H3K9me2 may be required for initiation of silencing for some lineage-specific genes (Golebiewska *et al.* 2009).

Clearly, the true nature of every modification is not yet known, although many common patterns are emerging; from the relative simplicity of bivalent domains, to the vast array of modifications marking promoters (Figure 1.8) (Bernstein *et al.* 2006;

Wang *et al.* 2008). What is becoming most apparent however, is the importance of context; different combinations of modifications working together with other epigenetic phenomena and protein complexes such as chromatin-modifying enzymes and transcriptional regulators, to orchestrate transcription and the management of the genome. Undeniably, histone modifications have a major role to play in these processes, which is why investigation into their distribution, not only on a genome-wide scale, but also at a localised gene-specific level is crucial to develop our understanding of these fundamentally significant chemical marks.

1.18. Histone Modifications in the Control of Central Tolerance

In line with this, investigation into the patterns of histone modifications across promiscuously expressed genes under the transcriptional control of AIRE may provide insight not only into the way AIRE controls their expression, but also into the modifications themselves and how they distribute across these highly specific loci when expressed outside of their target organ within medullary thymic epithelial cells. Transcriptional regulators such as AIRE, often have the capacity to distinguish between, and bind to individual histone modification states, providing a direct readout of the activity of the gene, and it has been suggested that the first PHD finger of AIRE may mediate an association with un-modified H3K4 (me0) (Koh *et al.* 2008; Org *et al.* 2008; Chignola *et al.* 2009). The presence of this un-methylated residue is typically a sign of silent or low levels of expression, thus AIRE may be able to recognise its target genes through their modification patterning, bind to them through H3K4me0-PHD1 interactions and either induce or suppress transcription (Matsumoto

2009). Thus, this novel role for AIRE as a histone-binding effector protein may suggest the involvement of epigenetic mechanisms for the regulation of so many thousands of AIRE-regulated target genes (Koh *et al.* 2008; Org *et al.* 2008; Matsumoto 2009).

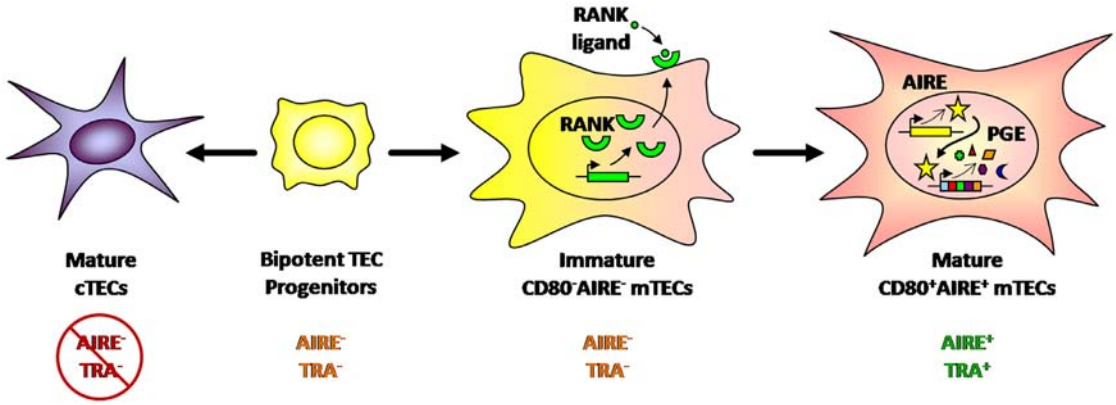
1.19. The Pathway of Thymic Epithelial Development

The recent categorization of the sequential phases in the embryonic TEC developmental pathway has uncovered some interesting and distinct sub-populations of cells, and provides an opportunity to investigate how AIRE may function as a transcriptional regulator of promiscuous gene expression as mTECs mature *in vivo*. In the early embryo, at embryonic day (E) 12, the murine thymic rudiment consists simply of undifferentiated epithelial cells and a mesenchyme capsule (Manley 2000). Isolation and culture of these undeveloped structures revealed a capacity for the generation of normal thymic structures including cortical and medullary regions, suggesting the pre-programmed nature of the undifferentiated cells even at such early stages (Rossi *et al.* 2006). This scenario was clarified by Rossi *et al* (2006), who, through attempts to isolate sub-populations of epithelial progenitor cells at E12, discovered that the thymic rudiment at this stage is phenotypically homogeneous; expressing the cell surface markers EpCAM1, cytokeratins 5 and 8, and MTS24 (Rossi *et al.* 2006). Using individually isolated enhanced yellow fluorescent protein (eYFP)-positive E12 thymic precursor cells, transferred to wild-type thymic lobes and grafted to the kidney capsule, this group were able to demonstrate the development of normal thymic architecture (Rossi *et al.* 2006). Furthermore, after sufficient

growth, it was revealed that both cortical and medullary regions of the newly formed thymic lobes exhibited eYFP signals, which led to the conclusion that a single bipotent TEC progenitor cell can generate both cortical and medullary lineages (Figure 1.10) (Rossi *et al.* 2006). This group then went on to further characterise the developmental pathway of mTECs, showing that the two known sub-populations of $Ly51^{-}EpCAM1^{+}$ mTECs which are either $CD80^{-}AIRE^{-}$ or $CD80^{+}AIRE^{+}$, represent different maturational states rather than distinct lineages, with mature $CD80^{+}AIRE^{+}$ mTECs developing from an immature $CD80^{-}AIRE^{-}$ mTEC progenitor population (Figure 1.10) (Rossi *et al.* 2007). In this same report, the signal responsible for this maturation was also examined, which was found to emanate from $CD4^{+}CD3^{-}$ lymphoid tissue inducer cells (LTi) (Rossi *et al.* 2007). Receptor activator of NF-κB (RANK), a cell surface marker expressed at high levels by $CD80^{-}AIRE^{-}$ mTECs, responds to LTi-derived RANK ligand (RANKL), or alternatively to $CD4^{+}$-derived signals, and these RANK-RANKL signals, in addition to other molecular interactions such as CD40-CD40L, are the trigger for the up-regulation of AIRE and subsequent promiscuous gene expression (Rossi *et al.* 2007; Akiyama *et al.* 2008; White *et al.* 2008; Zhu and Fu 2008).

Two alternative models have arisen following investigation into the role of AIRE in thymic embryogenesis, proposed to explain the high levels of PGE within the thymus. The developmental / progressive restriction model suggests that immature mTEC progenitors exhibit a wide catalogue of TRA, but that as differentiation occurs,

Figure 1.10 – The Pathway of Embryonic Thymic Development



Within the developing embryo, thymic organogenesis involves a well defined pathway of thymic epithelial cell (TEC) differentiation. The two distinct lineages of TECs; medullary (mTECs) and cortical (cTECs), are derived from a common bipotent TEC progenitor present in the murine thymus at embryonic day E12 (Rossi *et al.* 2006; Rossi *et al.* 2007). These progenitor cells do not express AIRE or tissue-restricted antigens (TRA), however, these genes will be poised for activation should the cells develop into mTECs. An immature $CD80^{-}AIRE^{-}$ mTEC progenitor population exists, which do not express AIRE or TRA but produce a cell surface receptor; RANK (Rossi *et al.* 2007; Akiyama *et al.* 2008; White *et al.* 2008; Zhu and Fu 2008). Stimulation from RANK ligand induces AIRE up-regulation and promiscuous gene expression, generating mature $CD80^{+}AIRE^{+}$ mTECs (Rossi *et al.* 2007; Akiyama *et al.* 2008; White *et al.* 2008; Zhu and Fu 2008). Bipotent TEC progenitors can also differentiate into mature cTECs, which will never express AIRE or TRA (Rossi *et al.* 2006; Rossi *et al.* 2007). The four main subpopulations of cells are displayed.

transcriptional promiscuity is progressively restricted (Farr and Rudensky 1998; Farr *et al.* 2002). This would ultimately generate a mosaic of fully mature mTECs throughout the medulla which display only a limited number of related TRAs, thus masquerading as terminally differentiated peripheral tissue cells such as hepatocytes or oligodendrocytes, hence the medulla would resemble a patchwork of different tissues (Farr *et al.* 2002; Kyewski and Derbinski 2004; Kyewski and Klein 2006). Alternatively, the terminal differentiation model assumes PGE to be an autonomous trait of TECs, and that a random de-repression or dysregulation of transcriptional control occurs as mTECs continuously differentiate and turnover, allowing the expression of genes only typically expressed in terminally differentiated cells (Farr *et al.* 2002; Kyewski *et al.* 2002; Gotter and Kyewski 2004; Devoss and Anderson 2007). Therefore, the discovery of an immature $CD80^{-}AIRE^{-}$ mTEC population, from which $CD80^{+}AIRE^{+}$ mature mTECs develop provides evidence in favour of the terminal differentiation model, and the recent finding that $AIRE^{+}$ mTECs only acquire the machinery for PGE upon maturation, with AIRE controlling the differentiation programme of these cells, also supports this theory (Rossi *et al.* 2007; Yano *et al.* 2008).

The breakdown of the mTEC developmental pathway provides an obvious foundation for the analysis of the epigenetic control of AIRE-regulated PGE. Bipotent TEC progenitors, although AIRE- and TRA-negative themselves, have the potential to develop into either fully mature cTECs, which will never express AIRE or its target genes, or alternatively into mTECs that will ultimately become AIRE-positive and therefore switch on TRA (Figure 1.10). Hence, could TRAs within the bipotent TEC

progenitors be marked in some way for future gene expression, possibly by poised bivalent domains? The more recent discovery of the intermediate population of immature mTECs adds to this story, allowing observation of any changes to the distributions of histone modifications as the cells progress through the developmental programme. As AIRE is switched on, which in turn leads to the expression of TRAs, the distribution of histone modifications may alter to reflect the imminent up-regulation of the AIRE-regulated genes. Due to the large number of AIRE target genes, and their significant clustering across chromosomes, it could be possible that alteration to chromatin on a domain-wide or global scale could allow for the up-regulation of so many TRA. As already discussed, acetylation is typically a mark of active gene expression, however it is also often a pre-requisite, allowing protein effectors access to their target promoters or binding sites prior to transcriptional activation (Ferguson *et al.* 2008). Histone acetylation across AIRE target genes, particularly those located in clusters could therefore be required to relax and open up the chromatin structure in either immature or newly formed mature mTECs. The interaction between AIRE and the histone acetyltransferase KAT3A/CBP is consistent with this hypothesis (Ferguson *et al.* 2008). Although these large-scale changes to chromatin have been shown to result in cellular stress, which can lead to DNA damage and ultimately apoptosis, this could potentially explain the finding that $AIRE^{+}$ mTECs have very rapid turnover rates, representing a short-lived fraction of stromal cells which are destined to die (Konishi *et al.* 2003; Gray *et al.* 2007; Ferguson *et al.* 2008). Furthermore, given that PGE is a feature of a terminally differentiated cell population, the formation of a permissive chromatin background would potentially facilitate transcriptional regulation by AIRE, and being inherently

cytotoxic would simply allow for the development of the next generation of mTECs and hence a new array of TRA, or promote cross-presentation of TRA by thymic antigen-presenting cells (Figure 1.3) (Gray *et al.* 2007; Ferguson *et al.* 2008).

Modelling of mature $AIRE^{+}$ mTECs and $AIRE^{-}$ TECs can be accomplished through application of a thymic epithelial (TEP) cell line stably transfected with a MIg virus construct containing AIRE and GFP, in parallel with a negative control cell line transformed with MIg-GFP alone, verified for both wild-type AIRE expression and PGE. Investigation into the epigenetic status of TRA requires the use of native chromatin immunoprecipitation (NChIP), a procedure which utilises unfixed chromatin, prepared by nuclease digestion. This process maintains the strong associations between histone proteins and DNA, yet allows greater resolution across defined DNA domains and higher levels of immunoprecipitation when antibodies to individual histone modifications are applied (O'Neill and Turner 2003). Large cell numbers can be generated through culture systems and greater amounts of chromatin can be isolated and analysed with unlimited numbers of antibodies, and hence NChIP offers an important insight into the regulation of PGE, restricted to the comparison of fully differentiated cell lines; those expressing AIRE and those not. To truly understand the role of AIRE in PGE and its impact throughout development, it would be invaluable to be able to examine the distribution of histone modifications across TRA within each step of the TEC developmental pathway.

A technique that has been used successfully to study the individual cell populations within the developing thymus and tease apart the T-cell developmental programme

in full; foetal thymic organ cultures (FTOC) allows maintenance of the *in vivo* characteristics of the embryonic thymus (Jenkinson and Anderson 1994). The complex three-dimensional architecture of the developing thymus, and phenotype of thymic stromal cells are preserved through this technique and therefore the defined programmes of cell development, in particular that of TECs, can occur as normal (Jenkinson and Anderson 1994; Anderson and Jenkinson 1995; Anderson and Jenkinson 2008). Through FTOC and FACS sorting based upon specific cell surface markers, the four cell subsets in the TEC developmental pathway can be isolated and pure populations of the EpCAM1^{+} bipotent TEC progenitor, immature CD45^{-}EpCAM1^{+}Ly51^{-}CD80^{-}, and mature CD45^{-}EpCAM1^{+}Ly51^{-}CD80^{+} mTECs, and mature CD45^{-}EpCAM1^{+}Ly51^{+} cTECs can be obtained and individually investigated (Rossi *et al.* 2007).

Chromatin immunoprecipitation on cultured thymic epithelial cells *in vitro* and on cells isolated from the TEC developmental pathway *in vivo* should give insight into how AIRE may impact on post-translational modifications of core histones across its target genes. Thus, the influence of epigenetics upon promiscuous expression of tissue-restricted antigens and the co-ordination of central tolerance processes can be examined.

1.20. AIMS

- To reproduce AIRE *in vitro* in a thymic cellular environment, ensuring wild-type AIRE localisation and function. A *Mus musculus* thymic epithelial (TEP) cell

line transfected with MIg virus bicistronic constructs containing either AIRE-GFP (TEP-AIRE), or MIg-GFP alone (TEP-GFP) will be employed for modelling mature $AIRE^{+}$ mTECs and $AIRE^{-}$ TECs.

- An investigation into whether promiscuous gene expression is associated with any epigenetic changes, through examining a panel of activating and repressive post-translational histone modifications and the binding status of certain histone-modifying enzymes, across a catalogue of AIRE-regulated target loci and AIRE-independent control genes.
- To further these studies and establish the role histone modifications play throughout the TEC developmental pathway *in vivo*, the distribution of these marks across TRA within each step of TEC development will be examined through carrier chromatin immunoprecipitation (CChIP) with small numbers of FACS sorted primary *Mus musculus* thymic stromal cells, generated through foetal thymic organ cultures (FTOC).
- To assay and compare the epigenetic status of TRA in TEC from athymic FoxN1-deficient nude mice, within which normal TEC development is disrupted, with wild-type primary *Mus musculus* TEC, to highlight the importance these modifications have in the establishment of promiscuous gene expression *in vivo*.

2. MATERIALS AND METHODS

2.1. CELL CULTURE AND PREPARATION OF PRIMARY CELL POPULATIONS

2.1.1. *Mus musculus* Thymic Epithelial Cell Culture

Mus musculus thymic epithelial (TEP) cells were a gift from Dr D. Kioussis of the National Institute of Medical Research. TEP cells, retrovirally transfected with the MIg virus, containing either GFP and AIRE, (TEP-AIRE cells), or GFP alone as a control line (TEP-GFP cells), were maintained in monolayers at 37°C in a 5% CO_2 humidified atmosphere. Cells were grown in 1xDulbecco's Modified Eagle's Medium (DMEM, Gibco) supplemented with 10% heat inactivated foetal bovine serum (FBS, Gibco), 1xMEM non-essential amino acids (without glutamine) (Gibco), 10mM Hepes (Sigma), $5x10^{-5}$M mercaptoethanol (Gibco), 4mM L-glutamine, 5000 IV/ml Penicillin (Gibco) and 5000µg/ml Streptomycin (Gibco). Cells were cultured in T75 (Sarstedt) tissue culture flasks. Passage by trypsinisation (0.05% trypsin-EDTA, Gibco) was performed once cells reached 70% confluence. Cells were harvested by trypsinisation at 37^oC and washed three times in ice-cold Ca_2 / Mg_2-free 1xphosphate buffered saline (1xPBS) containing 5mM sodium butyrate and centrifuged at 1200rpm (MSE 3000) for 5 minutes.

2.1.2. *Mus musculus* 3T3 Fibroblast Cell Culture

Mus musculus fibroblast cells (3T3s) were also transfected with MIg-GFP-AIRE (3T3-AIRE) or MIg-GFP (3T3-GFP) and grown following the same protocol as for TEP cell lines.

2.1.3. *Drosophila melanogaster* SL2 Cell Culture

Drosophila melanogaster SL2 cells were grown anaerobically in suspension at 26^{o}C in Schneider's medium (Gibco) supplemented with 8% FBS and antibiotics (50 units/ml Penicillin and 50µg/ml Streptomycin).

2.1.4. Mouse Husbandry and Breeding

Wild-type BALB/c (haplotype H-2^{d}) and FoxN1-deficient BALB/c nude (nu/nu) mice used in this study were housed and maintained in the Biomedical Sciences Unit (BMSU), University of Birmingham, in accordance to home office regulations.

2.1.5. Foetal Thymic Organ Culture (FTOC)

FTOC preparation and the sorting of thymic epithelial subsets were performed by Dr A. White (Anderson lab, University of Birmingham) (Jenkinson *et al.* 1992; Anderson *et al.* 1993). For reference, protocols for these methods are given below.

Thymic lobes were removed from mice embryos at the required stage of gestation (typically embryonic day E15) and placed on 0.8µm sterile nucleopore filters (Millipore) on sterile artiwrap sponges (Medipost Ltd) in DMEM in 90mm, single vent sterile Petri dishes (Sterilin). Organ cultures were grown in DMEM containing 10% FBS for seven days at 37°C, 10% CO_2.

2.1.5.1. Isolation of Primary Cell Populations

Cultured embryonic thymic lobes were washed three times in Ca_2 / Mg_2-free 1xphosphate buffered saline (1xPBS) (Sigma) and incubated in 600µl 0.25% trypsin,

0.02% EDTA (Sigma) for 15-20 minutes, 37°C (precise timing depends on the stage of thymus development and days in organ culture). Trypsinisation was stopped by the addition of RPMI-1640 Hepes (RF10-H), and stromal cell suspensions were generated from disaggregated lobes through gentle pipetting. An equal volume of RF10-H medium was added and cells were centrifuged for 10 minutes at 1000rpm and the supernatant removed. The cell pellet was resuspended in 1ml RF10-H and the cells counted in a haemocytometer.

An immunomagnetic separation technique was employed using Dynabeads (Dynal) for the depletion of any residual haematopoietic cells. Anti-rat IgG magnetic beads (Dynal), coated with anti-mouse CD45 (clone M1-9; ATCC) were added to cell suspensions in 200µl RF10-H at an approximate ratio of 10:1. Cells and beads were centrifuged for 10 minutes at 1000rpm, 4°C and allowed to bind. Once positive binding was detected, unbound cells were isolated from cell-bound beads using an Eppendorf (1.5ml) Dynal Magnetic Particle Concentrator (Dynal). Supernatants contained thymic stromal cell populations that were used for sorting of thymic epithelial subsets.

2.1.5.2. Flow Cytometric High Speed Sorting

EpCAM1^{+} E12 bipotent TEC progenitors and EpCAM1^{+} E12 FoxN1-deficient nude TEC were isolated from suspensions of E12 thymus lobes on the basis of expression of EpCAM1, as described (Rossi *et al.* 2006). To isolate TEC subsets, we used the following phenotypes: immature CD45^{-}EpCAM1^{+}Ly51^{-}CD80^{-} mTECs, mature CD45^{-}EpCAM1^{+}Ly51^{-}CD80^{+} mTECs, and mature CD45^{-}EpCAM1^{+}Ly51^{+} cTECs (Derbinski

et al. 2001; Rossi *et al.* 2007). In brief, cell lymphocyte-depleted suspensions were immunolabelled with the appropriate antibodies. Cells were resuspended in 50µl of the relevant primary antibody (Table 2.1) and incubated for 30 minutes on ice. Appropriate single colour controls and negative controls (secondary antibody only) were also set up. Cells were washed in 1ml 1xPBS and centrifuged. Where primary antibodies were not directly conjugated to a fluorochrome, relevant secondary antibodies (Table 2.1) were added to cells and incubated for 30 minutes on ice. Following immunolabelling, cells were washed and resuspended in RF10-H. Single colour controls, of approximately 100,000 cells, were resuspended in 200µl RF10-H. Experimental samples of no less than $1x10^6$ cells were resuspended in 1ml RF10-H. Sorting was performed on a MoFlo high speed sorter (Dako Cytomations) with forward and side scatter gates set to exclude nonviable cells. Sorted cells were counted and snap frozen for storage at -80°C until use. Sorting performed by Dr R. Bird (University of Birmingham).

2.1.6. Snap Freezing of Cell Populations

Cell populations were transferred into 1.5ml RNase-free microcentrifuge tubes and centrifuged at 1000rpm (MSE 3000) for 10 minutes. Supernatants were removed to leave a dry pellet before immersion in liquid nitrogen and storage at -80°C.

Table 2.1 – Antibodies for Isolation of Primary Cell Populations

ANTIBODY SPECIFICITY	ORIGIN
PRIMARY ANTIBODIES	
Anti-CD45 biotin (clone 30F11)	eBioscience
Anti-EpCAM1 (G8.8) conjugated to Alexa 647	Gift from Andy Farr
Anti-Ly51 (clone BP-1) PE	eBioscience
Anti-CD80 (clone 16-10A1) FITC	eBioscience
SECONDARY ANTIBODIES	
Streptavidin PeCy7	eBioscience
Streptavidin APC	BD
Streptavidin PE	eBioscience

Foetal thymic organ culture was employed for the investigation of the epigenetic status of tissue-restricted antigens across the embryonic thymic epithelial developmental pathway. Fluorescent Activated Cell Sorting (FACS) was used to isolate four cell populations; $EpCAM1^{+}$ bipotent TEC progenitors, immature $CD45^{-}EpCAM1^{+}Ly51^{-}CD80^{-}$, and mature $CD45^{-}EpCAM1^{+}Ly51^{-}CD80^{+}$ mTECs, and mature $CD45^{-}EpCAM1^{+}Ly51^{+}$ cTECs. Primary and secondary antibodies used in FACS along with their origin are displayed.

2.2. Analysis of the Integrity of the TEP and 3T3 Model Systems

2.2.1. Immunofluorescence Labelling

TEP-GFP, TEP-AIRE, 3T3-GFP and 3T3-AIRE cells were grown to 70% confluence on sterile 22mm x 22mm glass coverslips in 6-well plates, each well containing 2ml DMEM. Coverslips were transferred to fresh acetone for a 6 minute fixation then dried and washed for 5 minutes in 1xPBS with 0.1% Tween20 (Sigma) (PBST). To each coverslip, 50µl fluorescein isothiocyanate (FITC)-conjugated primary anti-AIRE antibody (rat anti-Aire (B1/02-5H12-2), a gift from H. Scott), diluted 1:100 in PBST, was added, and incubated for 1hr at room temperature in the dark. Following incubation, coverslips were washed for 5 minutes in PBST. Slides were prepared with 10µl mounting medium with 4',6-diamidino-2-phenylindole (DAPI) (Vectashield). Coverslips were dried, placed cell-side down onto the mounting medium and sealed. Images were captured using a Zeiss Axioplan microscope and SmartCaptureX Software (Digital Scientific).

2.2.2. Fluorescence-Activated Cell Sorting (FACS) Analysis for GFP Levels

For analysis of the efficiency of retroviral transfection, harvested TEP and 3T3 lines were washed in 1xPBS and transferred to 12.5ml polystyrene FACS tubes (Becton Dickinson) for FACS analysis. Acquisition was performed using a Becton Dickinson LSR flow cytometer and subsequent analysis was carried out using FloJo software. Freshly isolated thymocytes were used as negative control sample in order to set negative peaks for GFP. Where possible at least 100,000 events were recorded per experimental sample, with forward and side scatter gates to exclude non-viable cells.

2.2.3. Analysis of the Expression Levels of Promiscuous Genes

2.2.3.1. High Purity cDNA Extraction

For the analysis of promiscuous gene expression within cultured TEP and 3T3 lines or MoFlo-sorted primary cell populations, high purity cDNA was obtained from purified mRNA using the μMACs™ mRNA Isolation Kit (Miltenyi Biotec) and μMACs™ cDNA Synthesis Module (Miltenyi Biotec) according to the manufacturers instructions. The whole procedure was carried out in a laminar flow cabinet. Lysis/binding and wash buffers were allowed to equilibrate to room temperature prior to use. Frozen primary cell populations obtained from MoFlo cell sorting were suspended in 900μl lysis/binding buffer and vigorously vortexed for 3 minutes to ensure total cell lysis. Cultured TEP and 3T3 cells were snap frozen (section 2.1.6.), suspended in 900μl lysis/binding buffer and lysed using 21-gauge needles.

Cell lysates were then centrifuged at 13000rpm (MSE microcentaur) for 3 minutes in LysateClear columns. To each lysate, 50μl oligo microbeads were added and gently mixed. Lysates were transferred to μMACS™ columns; previously primed with 100μl lysis/binding buffer. Magnetically labelled mRNA remains bound to the column. Columns were rinsed twice with 200μl lysis/binding buffer to remove proteins and DNA, followed by four 100μl rinses with wash buffer to remove ribosomal RNA and DNA. For direct cDNA synthesis, bound mRNA was not eluted from the column. Instead, columns were washed twice with 100μl equilibration/wash buffer. Lyophilised enzyme mix was dissolved in 20μl resuspension buffer and applied on top of the column followed by 1μl of sealing solution, applied directly to the top of the

column to prevent evaporation. The thermoMACS^{TM} Separator was then turned on to 42°C and left for 1 hour 15 minutes for reverse transcription to take place. Columns were rinsed twice with 100µl equilibration/wash buffer. To release cDNA, 20µl release solution was added and incubated for 30 minutes at 42°C. Synthesised cDNA was eluted with 50µl elution buffer, collected into 1.5ml RNase-free microcentrifuge tubes and stored at -20°C. High purity cDNA extraction and expression analysis of primary cell populations performed by S. Parnell (Anderson Lab, University of Birmingham).

2.2.3.2. Quantitative Real-Time Polymerase Chain Reaction (qPCR) Analysis of Expression

Quantitative real-time PCR (qPCR) was performed on the Rotor-Gene™ RG-3000 (Corbett Research) using SYBR Green with expression primers specific for various genes of interest (Table 2.2 and 2.3). For sample normalisation prior to amplifying target genes, β-actin was used as the housekeeping gene. Primers were made by Invitrogen and primer pairs were designed using the aid of Primer3 software (http://frodo.wi.mit.edu/primer3/).

PCR reactions were performed as described previously (Shakib *et al.* 2009). Reactions were carried out in triplicates or duplicates in 15µl volumes in 2xSensiMix (Quantace) reaction buffer containing 50xSYBR© Green1 Solution (Quantace) and 200nM forward and reverse primers. After an initial denaturation step (95°C, 10 minutes), cycling was performed at 95°C for 15 seconds, 59-62°C (depending on primer pair) for 20 seconds and 72°C for 5 seconds (39 cycles). The fluorescent

Table 2.2 – Sequences of the *Mus musculus* Expression Primer Sets for Tissue-Restricted Antigens Used in Quantitative Real-Time Polymerase Chain Reaction (PCR) with cDNA

PRIMER NAME		NCBI REFERENCE	AMPLICON LENGTH	Tm
Beta-Actin [Actb]		**NM_007393**	**100bp**	**60°C**
Forward:	5` -CGTGAAAAGATGACCCAGATCA- 3`			
Reverse:	5` -TGGTACGACCAGAGGCATACAG- 3`			
Autoimmune Regulator [AIRE]		**NM_009646**	**186bp**	**62°C**
Forward:	5` -TGCATAGCATCCTGGACGGCTTCC- 3`			
Reverse:	5` -CCTGGGCTGGAGACGCTCTTTGAG- 3`			
Casein-Alpha [Csn1s1]		**NM_007784**	**190bp**	**60°C**
Forward:	5` -CATCATCCAAGACTGAGCCAG- 3`			
Reverse:	5` -CCTGTGGAAAGTAAGCCCAAAG- 3`			
Salivary Protein-1 [Spt1]		**NM_009267**	**116bp**	**59°C**
Forward:	5` -AGCAGTGTTGGTATCATCAGTG- 3`			
Reverse:	5` -CTGGTGAAAATACTGGCTCTGAA- 3`			
Salivary Protein-2 [Spt2] (Renamed Mucin-like 1 [Mucl1] 14.04.09)		**NM_009268**	**122bp**	**60°C**
Forward:	5` -TCAGACCAAAGTGGGTGACA- 3`			
Reverse:	5` -CCTCTTGTTTCTCATTGGAGGT- 3`			
Selection and Upkeep of Intraepithelial T-cells 1 [Skint1]		**NM_001102662**	**143bp**	**60°C**
Forward:	5` -TTCAGATGGTCACAGCAAGC- 3`			
Reverse:	5` -GAACCAGCGAATCTCCATGT- 3`			
Proteasome Subunit Beta-Type 11 [Psmβ11]		**NM_175204**	**231bp**	**60°C**
Forward:	5` -ATCGCTGCGGCTGATACTC- 3`			
Reverse:	5` -GCAGGACATCATAGCTGCCAA- 3`			

For the analysis of promiscuous gene expression within cultured TEP and 3T3 lines or FACS-sorted primary cell populations, high purity cDNA was obtained from purified mRNA and quantitative real-time PCR (qPCR) performed with expression primers specific for AIRE and the AIRE-regulated tissue-restricted antigens (*Csn1s1*, *Spt1*, *Spt2*) and control genes (*Skint1*, *Psmβ11*) listed. For sample normalisation prior to amplifying target genes, β-actin was used as the housekeeping gene. Illustrated are the NCBI reference code, the length of amplicon generated and the annealing temperature (Tm) of each primer set.

Table 2.3 – Sequences of the *Mus musculus* Expression Primer Sets for the Keratin Cluster Used in Quantitative Real-Time Polymerase Chain Reaction (PCR) with cDNA

PRIMER NAME		NCBI REFERENCE	AMPLICON LENGTH	Tm
Keratin 4 [Krt4]		**NM_008475**	**188bp**	**60°C**
Forward:	5` -GAGCATCTCGGTAGTTGGCG- 3`			
Reverse:	5` -GAGCACCAGGAAGACTGGAG- 3`			
Keratin 79 [Krt79]		**NM_146063**	**166bp**	**60°C**
Forward:	5` -GGAGCTGAGGAACGTACAGG- 3`			
Reverse:	5` -TGTCAAGCTGTCCACTTTGC- 3`			
Keratin 78 [Krt78]		**NM_212487**	**228bp**	**60°C**
Forward:	5` -GCCTCAGGAAGCAGAATGAC- 3`			
Reverse:	5` -CCTCACTCTCCAGCAACCTC- 3`			
Keratin 8 [Krt8]		**NM_031170**	**151bp**	**62°C**
Forward:	5` -ATCGAGATCACCACCTACCG- 3`			
Reverse:	5` -TGAAGCCAGGGCTAGTGAGT- 3`			
Keratin 18 [Krt18]		**NM_010664**	**161bp**	**60°C**
Forward:	5` -TCCTTCTGCATCTGGAG- 3`			
Reverse:	5` -ATCGTTGAGACTGAAATC- 3`			
Eukaryotic Translation Initiation Factor 4b [Eif4b]		**NM_145625**	**222bp**	**60°C**
Forward:	5` -GTTGCTGATCAAGCACAGGA- 3`			
Reverse:	5` -GTCCCGATATCCGTCCCTAT- 3`			
Tensin like C1 Domain-containing Phosphatase [Tenc1]		**NM_153533**	**199bp**	**60°C**
Forward:	5` -CAGGACCCTTGGCTTCTACA- 3`			
Reverse:	5` -GGAGACCTGGTGGTGTCTTG- 3`			
SPRY Domain-containing 3 [Spryd3]		**NM_001033277**	**240bp**	**60°C**
Forward:	5` -GCCCAGATCTTCTTCACCAA- 3`			
Reverse:	5` -CTTCCCCAAATACTCCAGCA- 3`			

For the analysis of gene expression across the keratin cluster within cultured TEP lines, cDNA was obtained from purified mRNA and quantitative real-time PCR (qPCR) performed with expression primers specific for the genes listed. For sample normalisation prior to amplifying target genes, β-actin was used as the housekeeping gene. Illustrated are the NCBI reference code, the length of amplicon generated and the annealing temperature (Tm) of each primer set.

signal produced from the amplicon was acquired at the end of each polymerisation step. Specific amplification of target genes was verified by melt curve analysis (72-99°C, hold 30 seconds on 1st step then 5 seconds on next steps) and by fractionation of PCR products on a 2% agarose gels, which were identified by their fragment size. Reaction amplification efficiency and Ct values were obtained using Rotor-Gene^{TM} 6.0 software (Corbett Research). Standard curves with reaction efficiencies no greater than 1 and R^2 values no less than 0.98 were generated for each primer set using serial dilutions of cDNA from E15 BALB/c foetal thymic organ cultures (FTOCs), mammary gland cells and salivary gland cells. For calculation of the relative expression values for samples normalised to β-actin, the Pfaffl model that takes gene-dependent differences in the amplification efficiency into account was employed (Pfaffl 2001).

2.3. Analysis of Histone Proteins

2.3.1. Affinity-Purified Antibodies

Rabbit polyclonal antisera to H3K4me1, H3K4me2, H3K4me3, H3K9ac, H4K8ac and H4K16ac were raised by immunization with synthetic peptides conjugated to ovalbumin as previously described (Turner *et al.* 1989; O'Neill and Turner 1995; White *et al.* 1999). Rabbit anti-H3K27me3 (07-449) and anti-H3K9me2 (07-212) were purchased from Millipore. Specificity was assayed by inhibition ELISA for all in-house and commercial antisera used and checked by Western blotting. For all antisera, cross-reaction with epitopes other than that against which the antiserum was raised was insignificant (O'Neill *et al.* 2006).

2.3.2. Histone Acid Extraction from Cultured Thymic Epithelial Cells

For the analysis of histone proteins in the two cell lines, TEP-GFP and TEP-AIRE cells were harvested, washed in ice-cold 1xPBS and cell numbers determined using an Improved Neybauer Haemocytometer Counting Chamber. Cells were resuspended at $1x10^7$ cells/ml in Triton extraction buffer (TEB; 10% Triton X-100, 0.1M PMSF, 2% Na azide in 1xPBS/5mM Na butyrate), incubated on ice for 10 minutes and pelleted at 1200rpm (MSE 3000) for 10 minutes at 4°C. Histones were extracted in 0.4N HCl ($2x10^6$ cells / 50µl) at 4°C overnight. Histones were isolated by centrifugation at 13000rpm (MSE microcentaur) for 1 minute, and histone-containing supernatants removed and stored at 4°C. Protein concentration of histone samples was determined using the Pierce assay. Samples were diluted 50x in Pierce-3 Agent (ThermoScientific) and left on ice for the colour change to develop. The absorbance (595nm) of histones was then compared against a standard curve generated using a dilution series of bovine serum antigen (1-0.2mg/ml) to determine their concentrations.

2.3.3. Sodium Dodecyl Sulphate (SDS) Polyacrylamide Gel Electrophoresis (PAGE)

SDS-PAGE was performed as described in Laemmli (1970) with the isolated histone fractions (Laemmli 1970). For histone separation, resolving gels were prepared (15% acrylamide, 0.4% NN'bisacrylamide, 375mM Tris-HCl pH 8.8, 0.1% SDS polymerised with 300µl 10% w/v ammonium persulphate and 30µl TEMED per 30ml gel solution). These were overlaid with iso-butanol during polymerisation to prevent evaporation which was then washed off before addition of the stacking gel (3%

acrylamide, 0.16% N,N'bisacrylamide, 125mM Tris-HCl pH 6.9, 0.1% SDS, polymerised with 100µl 10% w/v ammonium persulphate and 10µl TEMED per 10ml of gel solution).

Samples were prepared in 100µl volumes with 10µg histone proteins, 50% glycerol, 0.02% bromophenol blue and 10x standard dissociating buffer (SDB; 1M Tris-HCl pH 7.2, 10mM Na_2EDTA, 10% SDS, 1.432M 2-mercaptoethanol). Proteins were denatured for 10 minutes at 95°C and incubated on ice for 5 minutes prior to being loaded onto the gel. Electrophoresis was performed at 400 volts, 30 mA for at least 2 hours in Mini Protean II™ (BIO-RAD) apparatus and 1xSDS running buffer (50mM Tris base, 380mM Glycine, 0.1% SDS).

2.3.4. Analysis of Global Levels of Histone Modifications by Western Blot

Histone proteins were transferred onto Hybond C-Extra nitrocellulose paper (Amersham) as described (Towbin *et al.* 1979). Briefly, Hybond C, pre-soaked in transfer buffer (25M Tris-HCl, 192mM Glycine, 20% Methanol) was overlaid on top of the SDS-PAGE gel. Gels were sandwiched between Whatman No1 filter paper and Scotch brite sponge pads. All air bubbles were removed and the gel, plus pads, placed in a cassette and slotted into Trans Blot Cell apparatus (BIO-RAD). Protein transfer was carried out for 3 hours (180V, 300mA, 20w). Staining with Ponceau Red prior to antibody binding ensured an equal transfer of histones. Membranes were incubated for 1 hour at room temperature in blocking solution (5% milk in 1xPBS with 0.1% Tween20 (PBST)) in order to prevent non-specific antibody binding. Primary antibody (diluted in blocker; Table 2.4) was added and incubated at room

Table 2.4 – Primary and Secondary Antibodies Used for Western Blot and Immunofluorescence Analysis of the Global Levels of Histone Modifications

ANTIBODY SPECIFICITY	ORIGIN	DILUTION (IN PBST)
PRIMARY ANTIBODIES		
Rabbit anti-H3K4me1	In-house R204	1:200
Rabbit anti-H3K4me2	In-house R148	1:400
Rabbit anti-H3K4me3	In-house R614	1:400
Rabbit anti-H3K9ac	In-house R609	1:100
Rabbit anti-H4K8ac	In-house R404	1:1000
Rabbit anti-H4K16ac	In-house R252	1:400
Rabbit anti-H3K9me2	In-house R616	1:400
Rabbit anti-H3K27me3	Millipore (07-449)	1:800
SECONDARY ANTIBODIES		
Peroxidase Goat anti-Rabbit IgG	Sigma	1:1500

For the analysis of global levels of histone modifications in cultured TEP lines, Western blot analysis of bulk histones and immunofluorescence staining of metaphase chromosomes was performed. The names and specificities of primary and secondary antibodies used in these procedures are displayed along with their origin and the dilution factor in 1xPBS with 0.1% Tween20 (PBST).

temperature for 1 hour, followed by washing of the membranes three times in PBST. Secondary antibody peroxidase goat anti-rabbit IgG was applied and the filters left for a further hour at room temperature. After washing the filters three times with PBST, antibody binding was detected using enhanced chemiluminescence (ECL) Detection Reagents (Amersham) as per the manufacturers instruction.

2.3.5. Immunofluorescence Labelling of Metaphase Chromosomes from Thymic Epithelial Cells

For visualisation of the effects of AIRE upon the global distribution of histone modifications across metaphase chromosomes, TEP-GFP and TEP-AIRE cells were cultured with 10µl/ml Colcemid for 3 hours, and then harvested and washed twice in 1xPBS. A cell count was performed and samples diluted to $2x10^5$ cells/ml in 0.1M KCl and incubated for 10 minutes at room temperature. Cytospins were performed with 200µl/chamber ($4x10^4$ cells/slide) in a Shandon cytospin® cytocentrifuge (ThermoFisher Scientific) at 1800rpm for 10 minutes. Cell areas were marked with a diamond pen and immediately immersed in KCM buffer (120mM KCl, 20mM NaCl, 10mM Tris-HCl pH 8.0, 0.5mM EDTA, 0.1% v/v Triton X-100) for 8 minutes at room temperature. Primary antibodies (Table 2.4) were diluted in KCM buffer with 0.1% BSA, then 40µl were applied to each slide, covered with parafilm and left at 4^oC for 1 hour in humid chamber. Slides were washed twice in KCM, 10 minutes each wash, then incubated with 40µl FITC-conjugated secondary antibody (Table 2.4) for 1 hour, 4^oC. Slides were again washed twice in KCM, 10 minutes each wash, then fixed in 4% formaldehyde in KCM for 10 minutes, and finally washed in distilled water then mounted in 7.5µl DAPI. Performed by R. Muraleedharan.

2.4. CHROMATIN IMMUNOPRECIPITATION

2.4.1. Chromatin Isolation

2.4.1.1. Preparation of Unfixed Chromatin from Cultured Thymic Epithelial Cells for Native Chromatin Immunoprecipitation

Extraction and subsequent digestion of chromatin was performed as previously described (O'Neill and Turner 1995). TEP-GFP and TEP-AIRE cultures were trypsinised and washed in ice-cold 1xPBS/5mM Na butyrate (PBS-NaB) three times. Cell counts were performed then cells were resuspended in 1xTBS (15mM NaCl, 10mM Tris-HCl pH 7.5, 3mM $CaCl_2$, 2mM $MgCl_2$, 5mM Na butyrate) to a density of $2x10^7$ cells/ml. An equal volume of 1% Tween40/TBS and $1/200^{th}$ volume 0.1M PMSF was added to the suspension and the cells were stirred on ice for 1 hour in order to puncture the cell membranes for release of the nuclei. Nuclei were released by homogenisation, on ice, using 10 strokes with a Dounce all-glass homogeniser with a “tight” pestle, which resulted in a 75-80% yield of intact nuclei, verified by microscopy. The nuclei suspension was then centrifuged (2000rpm, MSE 3000, 4°C, 10 minutes) and pellets were resuspended in 5ml 5% sucrose/TBS. After centrifugation (3000rpm, MSE 3000, 4°C, 10 minutes), nuclei pellets were resuspended in 5ml digestion buffer (0.32M Sucrose, 50mM Tris-HCl pH 7.5, 4mM $MgCl_2$, 1mM $CaCl_2$, 0.1mM PMSF, 5mM Na butyrate) and a rough estimate of concentration determined using $A_{260/280}$ absorbance. Samples were then centrifuged (2000rpm, MSE 3000, 4°C, 10 minutes) and pellets were resuspended in digestion buffer to a concentration of 0.5mg/ml.

For analysis of chromatin, a chromatin ladder rich in mono-, di-, tri-, tetra-, and penta-nucleosomes is required. Chromatin was released from the nuclear preparations by digestion at 37°C for 5 minutes with 50U micrococcal nuclease (Amersham) per 0.5mg chromatin. Digestion was terminated by addition of 0.5M EDTA to a final concentration of 5mM and placed on ice for 5 minutes. The samples were centrifuged at 13000rpm (MSE microcentaur) and the first supernatant (S1) was removed and stored at 4^{o}C. Pellets were resuspended in 500µl lysis buffer (1mM Tris-HCl pH 7.4, 0.2mM Na_2EDTA, 0.2mM PMSF, 5mM Na butyrate) and dialysed overnight against 2l lysis buffer at 4^{o}C. Samples were centrifuged (2000rpm, MSE 3000, 4^{o}C, 10 minutes) and the second supernatant (S2) collected and the insoluble pelleted material (P) resuspended in 200µl lysis buffer before the concentration ($A_{260/280}$) of DNA within S1, S2 and P samples determined. For each sample, 2µg (with 0.3% SDS), were loaded onto 1.2% agarose gels for agarose gel electrophoresis (AGE), followed by ethidium bromide staining to determine the extent of micrococcal nuclease digestion. S1 and S2 fractions were routinely combined, representing approximately 90% total chromatin, and the DNA concentration ($A_{260/280}$) determined.

2.4.1.2. Preparation of Unfixed Chromatin from Primary Thymic Epithelial Cells for Carrier Chromatin Immunoprecipitation

For the CChIP procedure, extraction and subsequent digestion of chromatin was performed as previously described (O'Neill *et al.* 2006). *Drosophila melanogaster* SL2 cells were pelleted and washed three times in ice-cold PBS-NaB. Cells were resuspended to $5x10^7$ cells in 1ml ice-cold NB buffer (15mM Tris-HCL pH 7.4, 60mM

KCl, 15mM NaCl, 5mM $MgCl_2$, 0.1mM EGTA, 0.5mM 2-mercaptoethanol, 0.1mM PMSF, 5mM Na butyrate) then snap frozen and stored at -80^{o}C. For the CChIP procedure, SL2 cells were resuspended in 400µl ice-cold NB buffer. Small numbers of *Mus musculus* thymic epithelial cell populations obtained from MoFlo cell sorting (section 2.1.5.2.), which had previously been frozen, were resuspended in 100µl ice-cold NB buffer, transferred to each aliquot of 5x10^{7} SL2 cells and an equal volume 1% Tween40/NB buffer/0.1mM PMSF added, bringing the final volume to 1ml. For a 75-80% yield of intact nuclei, homogenisation was performed on ice with seven strokes using a Dounce all-glass homogeniser with a "tight" pestle. Nuclei were pelleted (2000rpm, MSE 3000, 4^{o}C, 10 minutes), resuspended in 5ml NB buffer, 5% (v/v) sucrose and pelleted (2000rpm, MSE 3000, 4^{o}C, 10 minutes). Nuclei were resuspended in 2ml digestion buffer and the DNA concentration determined, samples were centrifuged (2000rpm, MSE 3000, 4^{o}C, 10 minutes) then routinely resuspended in 500µl digestion buffer. Micrococcal nuclease (25U) was added to each aliquot and incubated for 7 minutes 30 seconds at 28^{o}C. The S1, S2 and P fractions were isolated and the DNA concentration ($A_{260/280}$) determined followed by analysis on 1.2% AGE. S1 and S2 fractions were routinely combined.

2.4.1.3. Preparation of Fixed Chromatin from Cultured Thymic Epithelial Cells for Cross-linked Chromatin Immunoprecipitation

Formaldehyde cross-linked chromatin was prepared essentially by the method outlined previously (Orlando and Paro 1993). Essentially, TEP cells were harvested, washed three times in PBS-NaB and resuspended at a concentration of 1x10^{6} cells/ml. Cells were cross-linked in 1% paraformaldehyde for 8 minutes at room

temperature. The reaction was stopped by addition of glycine to a final concentration of 150mM. Cross-linked cells were washed twice with PBS-NaB and resuspended in 130µl XChIP lysis buffer (50mM Tris-HCl, 10mM EDTA, 1% SDS, 5mM Na butyrate). Cells were sonicated using the Diagnode Biorupter for 10 minutes on medium at 4^{o}C. An aliquot of cross-linked chromatin was taken for reversal of the cross-links by proteinase K digestion at a concentration of 50µg/ml at 68^{o}C, 300rpm for 2 hours in an Eppendorf Thermomixer. DNA was extracted by two phenol/chloroform extractions and one chloroform extraction. DNA was precipitated by centrifuging at 13000rpm (MSE microcentaur) and resuspended in water. DNA was ran out on a 1% agarose gel to check the size of the fragments. Fragments were typically between 300 and 600bp.

2.4.2. Immunoprecipitation

2.4.2.1. Immunoprecipitation from Unfixed Chromatin (NChIP and CChIP)

From this point onwards, the use of siliconised microcentrifuge tubes, Pasteur pipettes and 15ml centrifuge tubes maximised DNA recovery throughout the immunoprecipitation. To 100-200µg freshly prepared, unfixed chromatin, 100-200µl (50-100µg Ig) affinity purified antibody (Table 2.5) was added in fresh pre-lubricated siliconised 1.5ml centrifuge tubes, and the volume adjusted to 1ml with incubation buffer (50mM NaCl, 20mM Tris–HCl pH 7.5, 20mM Na butyrate, 5mM, Na_2EDTA, 0.1mM PMSF). After overnight incubation on a slowly rotating platform at 4°C, 200µl protein A-sepharose (50% w/v, Pharmacia) was added and the incubation continued at room temperature for a further 3 hours on a fast rotating turntable. After

Table 2.5 – Affinity Purified Antibodies Used for Chromatin Immunoprecipitation

ANTIBODY SPECIFICITY	ORIGIN
NCHIP / CCHIP ANTIBODIES	
Rabbit anti-H3K4me1	In-house R204
Rabbit anti-H3K4me2	In-house R149
Rabbit anti-H3K4me3	In-house R612
Rabbit anti-H3K9ac	In-house R607
Rabbit anti-H4K8ac	In-house R403
Rabbit anti-H4K16ac	In-house R252/R232
Rabbit anti-H3K9me2	Millipore (07-212)
Rabbit anti-H3K27me3	Millipore (07-449)
Rabbit Pre-immune	In-house
XCHIP ANTIBODIES	
Rat anti-AIRE (B1/02-5H12-2)	A gift from H. Scott
Rabbit anti-RNA Polymerase II	Abcam (ab26721)
Rabbit anti-KMT2A/MLL1	Abcam (ab25735)
Rabbit anti-KMT6/Ezh2	Abcam (ab3748)
Rabbit anti-Eed	Abcam (ab4469)

For the quantitation of levels of histone modifications across AIRE-regulated genes, native chromatin immunoprecipitation (NChIP), carrier CChIP and formaldehyde-fixed XChIP were employed. Immunoprecipitation was performed with the affinity-purified antibodies listed. The origin of each is also displayed.

centrifugation (13000rpm, MSE microcentaur, 10 minutes) the supernatant was removed and stored on ice (unbound fraction, UB) and the protein A-sepharose pellet was washed three times in 10ml wash buffer (50mM Tris-HCl pH 7.5, 10mM EDTA, 5mM sodium butyrate, 150mM NaCl), with centrifugation at 2000rpm (MSE 3000, 4°C, 7 minutes). For the elution of the immunoprecipitated (bound, B) material from the protein A-sepharose, pellets were resuspended in 250µl 1% SDS/incubation buffer and incubated for 15 minutes on a fast rotating turntable at room temperature. After centrifugation (13000rpm, MSE microcentaur, 10 minutes) the supernatant (B) was removed and stored on ice. This was then repeated with a further 250µl 1% SDS/incubation buffer. The two extracts were combined and an equal volume of incubation buffer added to reduce the concentration of SDS to 0.5% (B).

For the isolation of DNA, both the B and UB fractions were treated to two phenol/chloroform washes and one chloroform extraction. DNA was ethanol precipitated with 1/10th volume (100µl) 4M LiCl, 50µg glycogen as a carrier, and 4x volume of ice-cold ethanol. The samples were vortexed thoroughly and allowed to precipitate overnight at -80°C. Precipitated DNA was centrifuged (3000rpm, MSE 3000, 4°C, 25 minutes) then pellets were re-dissolved in 300µl UltraPure d.H_2O (NChIP samples) or 40-80µl UltraPure d.H_2O (CChIP samples).

2.4.2.2. Immunoprecipitation from Fixed Chromatin (XChIP)

Antibody-bead complexes were formed by first washing Dynabeads protein A (Invitrogen) four times in RIPA buffer (10mM Tris-HCl pH 7.5, 1mM EDTA, 0.5mM EGTA, 1% Triton, 0.1% SDS, 0.1% Na deoxycholate, 150mM NaCl) and then

incubating with 2.5µg antibody (Table 2.5) overnight at 4^{o}C. To each antibody/bead complex, 25µg cross-linked chromatin was added and rotated at room temperature for 2 hours on a fast turntable. Beads were washed with 200µl RIPA buffer five times before washing once with TE (1mM EDTA, 10mM Tris). Antibody-bound DNA was eluted by addition of Elution buffer (20mM Tris-Hcl pH 7.5, 5mM EDTA, 5mM Na butyrate, 50mM NaCl, 1% SDS). Cross-links were reversed by proteinase K digestion at a concentration of 50µg/ml at 68^{o}C, 300rpm for 2 hours in an Eppendorf Thermomixer. DNA was ethanol precipitated following two phenol/chloroform extractions and one chloroform extraction. DNA was recovered by centrifuging at 13000rpm (MSE microcentaur) and resuspended in 20µl UltraPure d.H_2O.

2.4.3. PicoGreen Assay of ChIP DNA

For the determination of the yield of DNA in both the B and UB samples following immunoprecipitation, 2µl DNA was diluted 1:100 in Quant-iT PicoGreen reagent (Invitrogen). The percentage pull down (B/UBx100) for each antibody was then calculated and NChIP UB samples were diluted in d.H_2O to the concentration of the appropriate B sample concentration, to ensure that equal amounts of DNA were analysed. CChIP UB samples were diluted 1:2 in UltraPure d.H_2O.

2.4.4. Quantitative Real-Time Polymerase Chain Reaction (qPCR) Analysis of NChIP and XChIP DNA

For the analysis of the relative levels of histone modifications across specific promiscuously-expressed genes following NChIP or XChIP, quantitative real-time PCR (qPCR) was performed on the Rotor-Gene™ RG-3000 (Corbett Research) or

ABI 7900HT (Applied Biosystems) using SYBR Green with primers specific for various genes of interest (Table 2.6, 2.7 and 2.8). Primers were made by Invitrogen and primer pairs were designed using the aid of Primer3 software (http://frodo.wi.mit.edu/primer3/).

PCR reactions were carried out in triplicates in 10-15µl volumes in 1xQuantiTect SYBR Green PCR Master Mix (Qiagen) and 500nM forward and reverse primers, with equal concentrations of either B or UB or input ChIP DNA. After an initial denaturation step (95°C, 15 minutes), cycling was performed at 94°C for 15 seconds, 57-60°C (depending on primer pair) for 30 seconds and 68-72°C for 15 seconds (44 cycles). The fluorescent signal produced from the amplicon was acquired at the end of each polymerisation step. Specific amplification of target genes was verified by melt curve analysis (72-99°C, hold 30 seconds on 1st step then 5 seconds on next steps) and by fractionation of PCR products on 2% agarose gels, which were identified by their fragment size. Reaction amplification efficiency and Ct values were obtained using Rotor-Gene™ 6.0 software (Corbett Research). Standard curves with reaction efficiencies no greater than 1 and R^2 values no less than 0.98 were generated for each primer set using serial dilutions of *Mus musculus* genomic DNA (gDNA), and used to generate concentrations for each PCR product. B:UB ratios were then calculated. XChIP data was a ratio of B:input and was normalised to results for glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*).

Table 2.6 – Sequences of *Mus musculus* Genomic DNA Tissue-Restricted Antigen Primer Sets Used in Quantitative Real-Time Polymerase Chain Reaction (PCR) with ChIP DNA

PRIMER NAME		NCBI REFERENCE	AMPLICON LENGTH	Tm
Glyceraldehyde-3-phosphate Dehydrogenase Pseudogene (Gm12033) [Gapdh]		**100042746**	**149bp**	**60°C**
Promoter Region Forward:	5` -TGTGGCCAAGCACTTGTATAAC- 3`			
Promoter Region Reverse:	5` -TATGTCTGACCAGAGGAGAGCA- 3`			
Casein-Alpha [Csn1s1] (Extension at 68°C)		**NM_007784**	**150bp**	**57°C**
Promoter Region Forward:	5` -CCCTACTCTTGGGTTCAAGG- 3`			
Promoter Region Reverse:	5` -GCTCTTAGCGTACTGGAACAAA- 3`			
Salivary Protein-1 [Spt1]		**NM_009267**	**120bp**	**60°C**
Promoter Region Forward:	5` -TGGCTGTGTGGTTGATTCTC- 3`			
Promoter Region Reverse:	5` -CAGGGTTCCACATCAAGGAC- 3`			
Salivary Protein-1 [Spt1]		**NM_009267**	**108bp**	**60°C**
Coding Region Forward:	5` -GGTGGGACCAATAACATTCC- 3`			
Coding Region Reverse:	5` -GCCTGAGTTTCAGAGCCAGT- 3`			
Salivary Protein-2 [Spt2] (Renamed Mucin-like 1 [Mucl1] 14.04.09)		**NM_009268**	**140bp**	**60°C**
Promoter Region Forward:	5` -TAATTGGCCTCTGGCTGTGT- 3`			
Promoter Region Reverse:	5` -TCTTGACACCAGGGTTCCAC- 3`			
Salivary Protein-2 [Spt2] (Renamed Mucin-like 1 [Mucl1] 14.04.09)		**NM_009268**	**200bp**	**60°C**
Coding Region Forward:	5` -GGTGGGACCAATAACATTCC- 3`			
Coding Region Reverse:	5` -GCCTGAGTTTCAGAGCCAGT- 3`			
Recombination Activating Gene 1 [Rag1]		**NM_009019**	**154bp**	**60°C**
Coding Region Forward:	5` -AACTCAGGCTAGGGTCAGCA- 3`			
Coding Region Reverse:	5` -GGGATCAGCCAGAATGTGTT- 3`			
Selection and Upkeep of Intraepithelial T-cells 1 [Skint1]		**NM_001102662**	**143bp**	**60°C**
Promoter Region Forward:	5` -CAATGGGATCCACAGGACTA- 3`			
Promoter Region Reverse:	5` -TGCTCCTAACTATTCCCAACAAA- 3`			
Proteasome Subunit Beta-Type 11 [Psmβ11]		**NM_175204**	**197bp**	**60°C**
Promoter Region Forward:	5` -GTTCTCTGAGGTGGGTGGAG- 3`			
Promoter Region Reverse:	5` -GCTGAGTGAGAATCGGAAGG- 3`			

For the quantitation of levels of histone modifications across AIRE-regulated and control genes following chromatin immunoprecipitation, quantitative real-time PCR (qPCR) was performed with primers specific for the genes listed. Illustrated are the NCBI reference code, the length of amplicon generated and the annealing temperature (Tm) of each primer set.

Table 2.7 – Sequences of *Mus musculus* Genomic DNA Keratin Cluster Primer Sets Used in Quantitative Real-Time Polymerase Chain Reaction (PCR) with ChIP DNA

PRIMER NAME		NCBI REFERENCE	AMPLICON LENGTH	Tm
Keratin 4 [Krt4]		**NM_008475**	**181bp**	**60°C**
Promoter Region Forward:	5` -AGCTCCCATCCAAGATCACA- 3`			
Promoter Region Reverse:	5` -CAGACCCTGGAACCTGAGAG- 3`			
Keratin 79 [Krt79]		**NM_146063**	**173bp**	**60°C**
Coding Region Forward:	5` -GTGGTGGCAGCTGCTCTTAT- 3`			
Coding Region Reverse:	5` -GGTACCAGGACTCAGCCTCA- 3`			
Keratin 78 [Krt78]		**NM_212487**	**225bp**	**60°C**
Promoter Region Forward:	5` -CAGCAAGTGGCAGACACAGT- 3`			
Promoter Region Reverse:	5` -AAAGGCAGAACACGCTGAGT- 3`			
Keratin 8 [Krt8]		**NM_031170**	**187bp**	**60°C**
Promoter Region Forward:	5` -GTTAGGCCCTGCCCTCTAGT- 3`			
Promoter Region Reverse:	5` -TGGACATGGTGAAGTCTGGA- 3`			
Keratin 18 [Krt18]		**NM_010664**	**169bp**	**60°C**
Promoter Region Forward:	5` -CTCCCAAGTGCTGGGATAAA- 3`			
Promoter Region Reverse:	5` -AGCATACCTGCCATCCTCAC- 3`			
Eukaryotic Translation Initiation Factor 4b [Eif4b]		**NM_145625**	**134bp**	**60°C**
Promoter Region Forward:	5` -AAAAGCCCATGGTTCAAATG- 3`			
Promoter Region Reverse:	5` -GGGTGTGCCACAATTGATTT- 3`			
Tensin like C1 Domain-containing Phosphatase [Tenc1]		**NM_153533**	**206bp**	**60°C**
Promoter Region Forward:	5` -CCTGCACTTCCCTCCACTT- 3`			
Promoter Region Reverse:	5` -GCTCTGAGCAACCTTTCCAC- 3`			
SPRY Domain-containing 3 [Spryd3]		**NM_001033277**	**226bp**	**60°C**
Promoter Region Forward:	5` -GCTGAGAGGCCTATGGATGA- 3`			
Promoter Region Reverse:	5` -ACCTGTCAATGAGGCTACGC- 3`			

For the quantitation of levels of histone modifications across AIRE-regulated genes of the keratin cluster following chromatin immunoprecipitation, quantitative real-time PCR (qPCR) was performed with primers specific for the genes listed. Illustrated are the NCBI reference code, the length of amplicon generated and the annealing temperature (Tm) of each primer set.

Table 2.8 – Sequences of *Mus musculus* Genomic DNA Keratin Cluster Intergenic Region Primer Sets Used in Quantitative Real-Time Polymerase Chain Reaction (PCR) with ChIP DNA

PRIMER NAME		AMPLICON LENGTH	Tm
Krt4→Krt79 Intergenic Region [INT1]		**166bp**	**60°C**
Forward:	5` -CCCTAGCTCCCAGGGTAGAG- 3`		
Reverse:	5` -ATGTCAGGCAGTGCTGTGAG- 3`		
Krt79→Krt78 Intergenic Region [INT2]		**234bp**	**60°C**
Forward:	5` -TTGGGCTTAATTCCTGAACG- 3`		
Reverse:	5` -CAAAATATCCCCCACCACAG- 3`		
Krt78→Krt8 Intergenic Region A [INT3A]		**187bp**	**60°C**
Forward:	5` -TCTTTTCAGGACTGGGGATG- 3`		
Reverse:	5` -CCAATTCCAGGGATCGAGTA- 3`		
Krt78→Krt8 Intergenic Region B [INT3B]		**179bp**	**60°C**
Forward:	5` -CATGCTCCGTGCAAACTAGA- 3`		
Reverse:	5` -CCTCCACACAGACCTGGAAT- 3`		
Krt78→Krt8 Intergenic Region C [INT3C]		**196bp**	**60°C**
Forward:	5` -GTGGCTTTGGATTTGAGAGC- 3`		
Reverse:	5` -AATGGTGACCTGAGGCAAAC- 3`		
Krt8→Krt18 Intergenic Region [INT4]		**234bp**	**60°C**
Forward:	5` -CTGATGATGGTGACCTGGTG- 3`		
Reverse:	5` -TGCCTCCTAGCCGCTATTTA- 3`		
Krt18→Eif4b Intergenic Region A [INT5A]		**206bp**	**60°C**
Forward:	5` -TGCTGGGGTGGTATTGTGTA- 3`		
Reverse:	5` -GGCCTGCCTCTTAGGATCTC- 3`		
Krt18→Eif4b Intergenic Region B [INT5B]		**192bp**	**60°C**
Forward:	5` -CCAGACCCTGTCTCAGGAAA- 3`		
Reverse:	5` -TCACCACATGGAGGTCAGAA- 3`		
Krt18→Eif4b Intergenic Region C [INT5C]		**188bp**	**60°C**
Forward:	5` -GTGGTCAGGCTGAAAACCAT- 3`		
Reverse:	5` -TCACCGTGCTTGGTAGATTG- 3`		
Eif4b → Tenc1 Intergenic Region [INT6]		**243bp**	**60°C**
Forward:	5` -CCTAGCCTTTGCCAGAACAG- 3`		
Reverse:	5` -CAGCGTTCCTTAGCAGATCC- 3`		
Tenc1 → Spryd3 Intergenic Region [INT7]		**118bp**	**60°C**
Forward:	5` -GCCGCCGAGTGTTAGGTA- 3`		
Reverse:	5` -CCGCTAAATCTGACCAATCC- 3`		

For the quantitation of levels of histone modifications across AIRE-regulated genes of the keratin cluster following chromatin immunoprecipitation, quantitative real-time PCR (qPCR) was performed with primers specific for the genes listed. Illustrated are the length of amplicon generated and the annealing temperature (Tm) of each primer set.

2.4.5. Radioactive Polymerase Chain Reaction (PCR) Analysis of CChIP DNA

To assay histone modifications in primary cell populations following CChIP, radioactive PCR was performed with B and UB CChIP DNA, alongside *Mus musculus* and *Drosophila melanogaster* gDNA controls to monitor cross-hybridization, using primers specific for various genes of interest (Table 2.6). Primers were made by Invitrogen and primer pairs were designed using the aid of Primer3 software (http://frodo.wi.mit.edu/primer3/).

PCR reactions were carried out in 50µl volumes in Reddy Mix PCR Master Mix (AB Gene) and 500nM forward and reverse primers, with 2µl neat B and 2µl 1:2 diluted UB DNA. To each reaction, 0.1µCi of dCTP radiolabelled with α-^{32}P (Perkin Elmer) was added. Cycling was performed at 94°C for 60 seconds, 57-60°C (depending on primer pair) for 60 seconds and 68-72°C for 90 seconds (41 cycles).

As standard, 7µl aliquots were removed after 38 and 41 cycles, loaded onto 5% polyacrylamide gels and electrophoresed at 250 volts, 120 mA for 30 minutes. Gels were dried onto filter paper (SpeedGel System, Thermo Savant) for a minimum of 2 hours. Filters were exposed to a phosphor screen overnight and scanned with a PhosphorImager (Typhoon 9200, Amersham). Intensity values for each PCR product were analysed with 'Image Quant TL' software (Molecular Dynamics), the ratio of B to UB signal was calculated from the percentage pull-down values of each immunoprecipitation.

3. RESULTS

3.1. CHARACTERISATION OF THE THYMIC EPITHELIAL CELL MODEL SYSTEM

To date, promiscuous gene expression (PGE) of tissue-restricted antigens (TRA) within the thymus is very well characterised in terms of the panel of TRA regulated by AIRE, the timing of AIRE expression *in vivo*, and the specific cell type in which this process occurs. However, many questions remain unanswered about this phenomenon (Derbinski *et al.* 2001; Anderson *et al.* 2002; Rossi *et al.* 2007). Although AIRE has demonstrated transcriptional transactivational properties, is found within large multi-protein complexes at distinct foci throughout the nucleus, along with the histone acetyltransferase CREB-binding protein (KAT3A/CBP) and is able to bind unmethylated histone H3 lysine 4, the fundamental aspects of how this protein functions as a transcriptional regulator remain unknown (Pitkanen *et al.* 2000; Pitkanen *et al.* 2005; Ruan *et al.* 2007; Koh *et al.* 2008; Org *et al.* 2008). To decipher the molecular complexities of AIRE's control of PGE, we undertook an analysis of the epigenetic status of TRA under the transcriptional control of AIRE. Investigating the role AIRE plays *in vivo* is particularly challenging as the number of mature $CD80^{+}AIRE^{+}$ mTECs is extremely low, thus a *Mus musculus* thymic epithelial (TEP) cell model system was employed, in order to replicate AIRE function *in vivo*. The TEP cell line, which normally lacks AIRE, was transfected with MIg virus bicistronic constructs containing both AIRE and GFP (TEP-AIRE), or GFP alone as a control (TEP-GFP). Both parent lines were FACS sorted based on GFP expression, and purified populations clonally expanded prior to experimental analysis to ensure

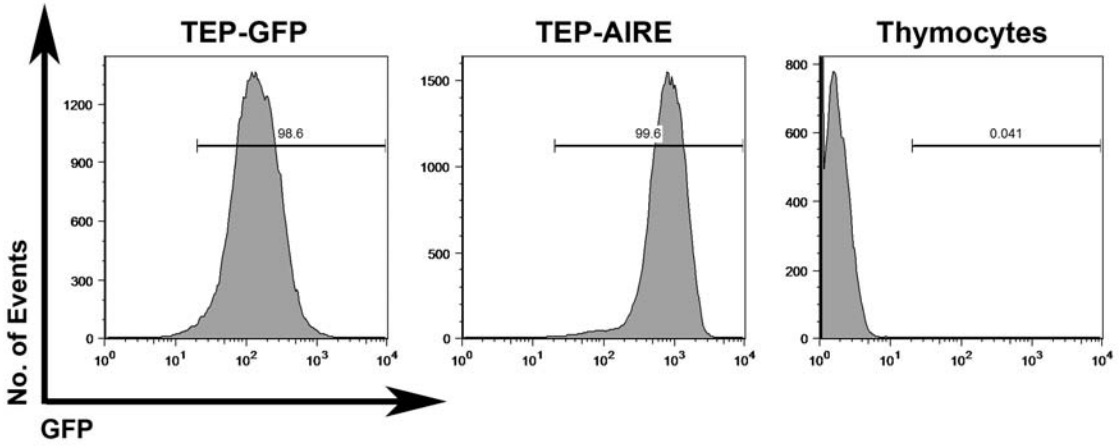
GFP^+AIRE^+ cells were expressing AIRE at 100% efficiency which was verified by flow cytometric analysis (Figure 3.1).

3.1.1. Localisation and Expression of AIRE and Tissue-Restricted Antigens in the Thymic Epithelial Cell Line

The up-regulation of AIRE upon RANK-RANKL stimulation is known to result in the induced expression of a wide range of TRA within mature $CD80^+AIRE^+$ mTECs, and this event has been successfully modelled in mammalian cell culture previously, by numerous groups (Pitkanen *et al.* 2001; Halonen *et al.* 2004; Pitkanen *et al.* 2005; Org *et al.* 2008). Correct localisation and activity of AIRE within the TEP cell line was verified through immunofluorescence microscopy and expression analysis (Figure 3.2).

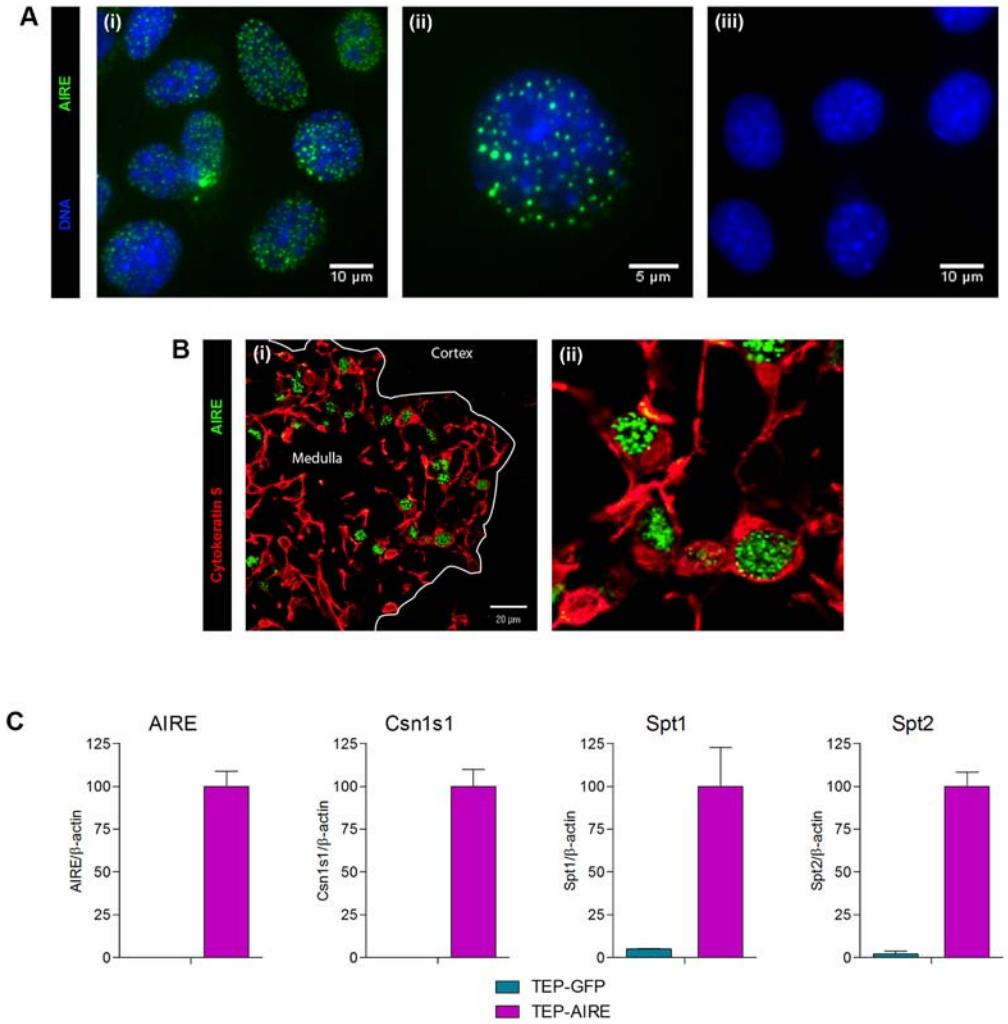
AIRE displays a predominantly nuclear localisation pattern, forming distinct, punctate nuclear speckles *in vivo*, although associations with microtubular structures within the cytoplasm have also been observed in transfected cell lines (Pitkanen *et al.* 2001; Hubert *et al.* 2008). Cultured TEP-AIRE and TEP-GFP cells were immunostained for AIRE and counterstained with DAPI for DNA, which revealed a speckled pattern of AIRE, limited to the nucleus, in the majority of AIRE-positive cells, not seen in control TEP-GFP cells (Figure 3.2 A). Comparison to AIRE in wild-type BALB/c (haplotype H-2d) mTEC *in vivo*, revealed an analogous distribution (Figure 3.2 B). Quantitative real-time PCR (qPCR) with cDNA, prepared from TEP-AIRE and TEP-GFP mRNA, was performed for AIRE and three known TRA, positively influenced by AIRE; casein-α (*Csn1s1*), salivary protein-1 (*Spt1*) and

Figure 3.1 – Characterisation of Thymic Epithelial Cell Lines: Transfection Efficiency



Mus musculus thymic epithelial (TEP) cell lines were transfected with MIg virus bicistronic constructs containing either AIRE-GFP (TEP-AIRE), or GFP alone (TEP-GFP) as a control. Transfected populations were clonally expanded and purified. For verification of the efficiency of transfection, flow cytometric analysis was carried out, gated on GFP, with freshly isolated thymocytes as a negative control. TEP-AIRE and TEP-GFP lines were seen to express GFP at greater than 96% efficiency.

Figure 3.2 – Characterisation of Thymic Epithelial Cell Lines: Subcellular Distribution and Expression of the Autoimmune Regulator and Tissue-Restricted Antigens



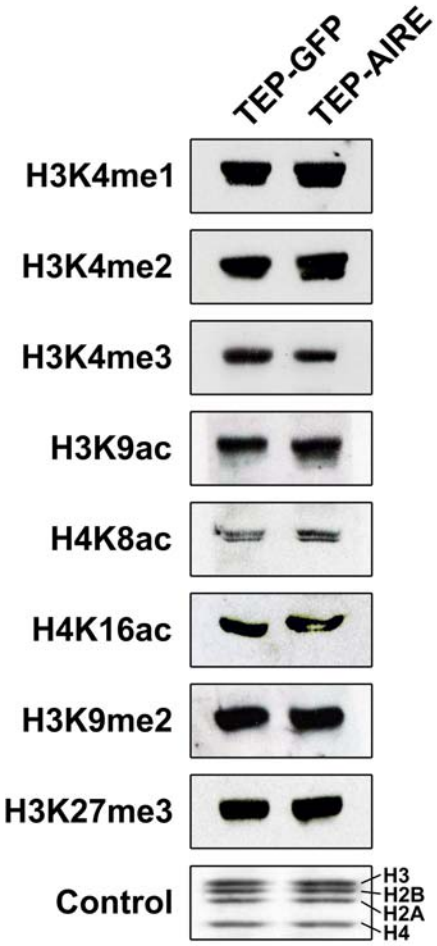
A, Immunofluorescence of *Mus musculus* thymic epithelial (TEP) cell lines, revealed a subcellular nuclear localisation of AIRE equivalent to AIRE in mTECs *in vivo*. TEP cells transfected with either MIg-AIRE-GFP (TEP-AIRE), or MIg-GFP (TEP-GFP), were cultured and stained with FITC anti-AIRE antibody (green) and DNA counterstained with DAPI (blue). Punctate AIRE nuclear bodies were observed in TEP-AIRE cells (i, ii), but not in TEP-GFP (iii). *B*, Subcellular localisation of AIRE (green) within thymic medulla (i, ii), for comparison. Medullary TECs denoted by cytokeratins 5 staining (red). Punctate AIRE nuclear staining of individual mTECs shown in (ii). *C*, Real-time quantitative PCR was used to compare the relative mRNA expression levels of AIRE and three tissue-restricted antigens; casein-α (*Csn1s1*); salivary protein-1 (*Spt1*); and salivary protein-2 (*Spt2*), as indicated, showing active expression of AIRE and its target genes. Turquoise columns represent data from TEP-GFP cells and pink columns represent results from TEP-AIRE cells. Data was normalised to β-actin expression levels as standard. Data are the mean ±SEM from technical triplicate reactions, and are representative of at least two distinct cDNA preparations.

salivary protein-2 (*Spt2*), and as standard expression levels were normalised to β-actin, which remained constant for the two cell populations. High levels of AIRE, *Csn1s1*, *Spt1* and *Spt2* transcript were detected in TEP-AIRE cells, which were not observed for TEP-GFP cells (Figure 3.2 C). In conclusion, these findings indicate that the TEP cell line expresses AIRE in the appropriate sub-cellular compartment, leading to AIRE-dependent TRA expression. On this basis, this system was used in initial experiments to study the epigenetic status of AIRE-dependent genes in the presence and absence of AIRE.

3.1.2. Global Analysis of the Relative Levels of Post-translational Histone Modifications in AIRE-Positive and AIRE-Negative Thymic Epithelial Cells

Since AIRE has been reported to complex with the transcriptional co-activator KAT3A/CBP, a known histone acetyltransferase, and given that AIRE has been shown to directly regulate expression of a large number of genes found within clusters along chromosomes of mature mTECs, we wanted to investigate whether alterations to the global levels of histone modifications could account for PGE. Histones were extracted from TEP-AIRE and TEP-GFP cells and equal concentrations of protein were separated by SDS-PAGE, Western blotted and labelled with antibodies specific to acetylation (H3K9ac, H4K8ac and H4K16ac) and methylation (H3K4me1, me2 and me3, H3K9me2 and H3K27me3) post-translational histone modifications (Figure 3.3). The functional relevance of these histone modifications is displayed in Table 3.1. This revealed an equal distribution of all eight marks across the two TEP populations, suggesting that, at the protein level, AIRE does not impact on histone modifications on a global scale. However, this

Figure 3.3 – Characterisation of Thymic Epithelial Cell Lines: Global Levels of Post-Translational Histone Modifications



SDS-PAGE and Western blot analysis of global post-translational histone modifications in *Mus musculus* thymic epithelial (TEP) cell lines revealed no change in the presence and absence of AIRE. Core histones were acid extracted from cultured TEP-AIRE and TEP-GFP populations, resolved by SDS-PAGE and transferred to membranes for Western blot analysis. Equivalent loading was determined by Ponceau red staining of bulk histones (control). Blots were probed with antibodies directed against acetylation and methylation of specific lysine residues on histones H3 and H4 (H3K4me1, H3K4me2, H3K4me3, H3K9ac, H4K8ac, H4K16ac, H3K9me2 and H3K27me3). Immunocomplexes were detected following secondary antibody peroxidase goat anti-rabbit IgG binding, by enhanced chemiluminescence.

Table 3.1 – Typical Functions of the Histone Modifications Investigated Through Native and Carrier Chromatin Immunoprecipitation

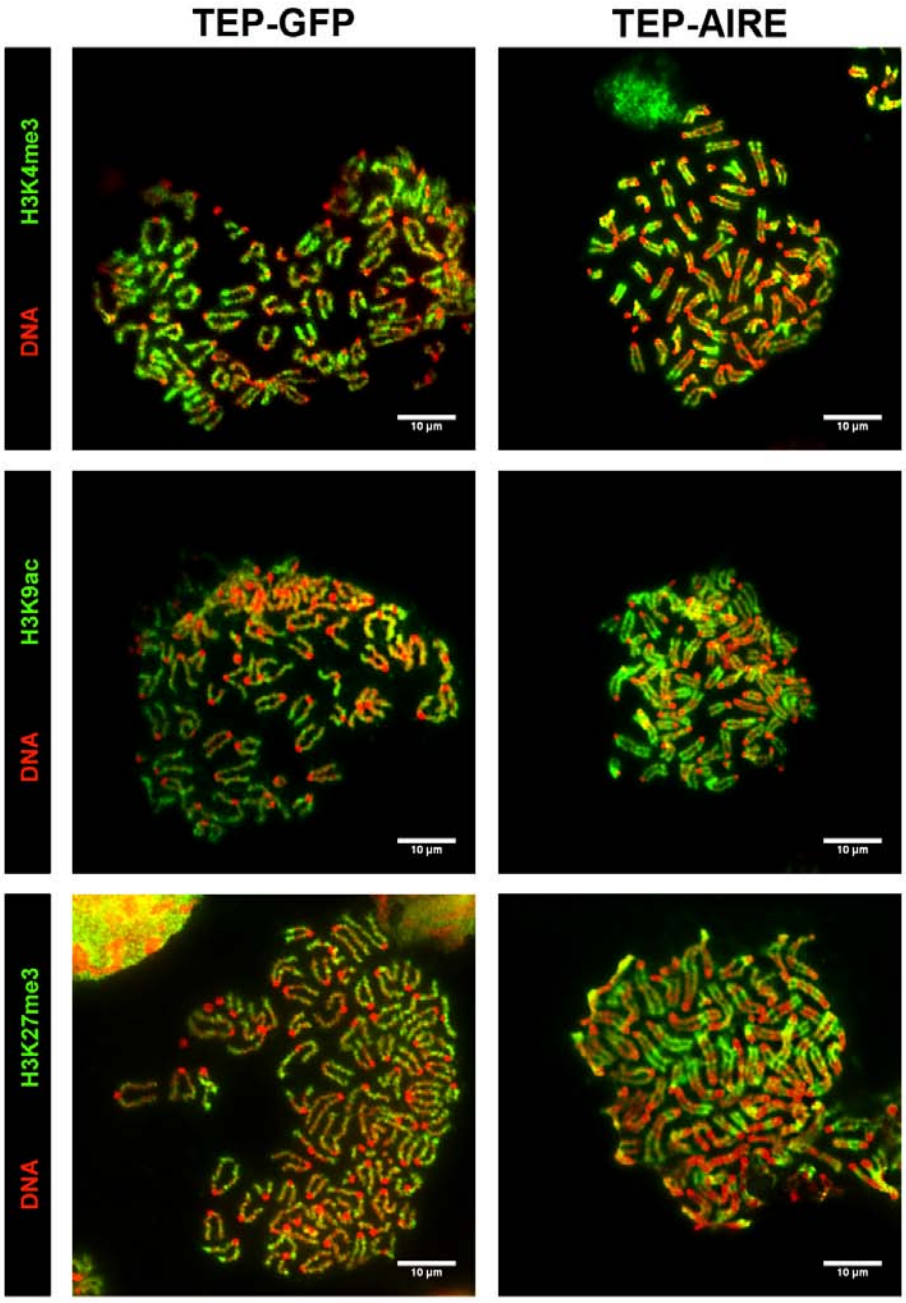
ChIP ANTIBODIES AND THEIR FUNCTIONS		
ANTIBODY	SPECIFICITY	FUNCTION
R204	H3K4me1	ACTIVE
R149	H3K4me2	ACTIVE
R612	H3K4me3	ACTIVE
R607	H3K9ac	ACTIVE
R403	H4K8ac	ACTIVE
R252	H4K16ac	ACTIVE
07-212	H3K9me2	INACTIVE
07-449	H3K27me3	INACTIVE
PI	Pre-immune	CONTROL

The name of each antibody used in native chromatin immunoprecipitation with cultured AIRE-negative and AIRE-positive thymic epithelial cell lines is displayed along with the histone modification against which they act. For reference, the general role these marks play in the control of transcription is represented. Pre-immune is a no antibody control.

procedure does not fully address the issue of whether AIRE imparts its function through a genome-wide effect on epigenetic marks. We therefore employed a technique which has been used frequently to assess the global distribution of histone modifications across metaphase chromosome spreads, revealing highly defined banding patterns along the individual chromosome arms, and thus making any large-scale changes easily identifiable (O'Neill *et al.* 1999; O'Neill *et al.* 2003). Unfixed metaphase chromosomes from TEP-AIRE and TEP-GFP cells were immunolabelled with antibodies to an acetylation mark H3K9ac, an active methyl mark H3K4me3 and a silent methyl mark H3K27me3, and counterstained with DAPI for DNA (Figure 3.4).

The spread and intensity of acetylation and methylation was consistently comparable to the control cells across the genome in the presence of AIRE, with no loss of the characteristic banding patterns. It should be noted that the TEP cell line is tetraploid, with double the expected number of chromosomes, however, this trait is common to both the TEP-AIRE and TEP-GFP cells. These data suggest that AIRE does not impart its function through any detectable genome-wide alterations to acetylation or methylation levels on the histone protein tails, despite the global distribution of its thousands of target genes, and given its known association with a histone acetyltransferase and localisation within PML-like bodies within the nucleus.

Figure 3.4 – Characterisation of Thymic Epithelial Cell Lines: Acetylation and Methylation Levels across Metaphase Chromosomes



Metaphase chromosome spreads were prepared from *Mus musculus* thymic epithelial (TEP) cells transfected with either MIg-AIRE-GFP (TEP-AIRE), or MIg-GFP (TEP-GFP). Chromosomes were immunostained with primary antibodies (green) to H3K9ac, H3K4me3 and H3K27me3, with FITC goat anti-rabbit secondary antibody, as indicated. DNA was counterstained with DAPI (false coloured red). Spreads were captured at a high magnification (x100 objective) and show equivalent banding patterns characteristic of each modification, in both the presence and absence of AIRE. Both cell populations also present with a tetraploid genome.

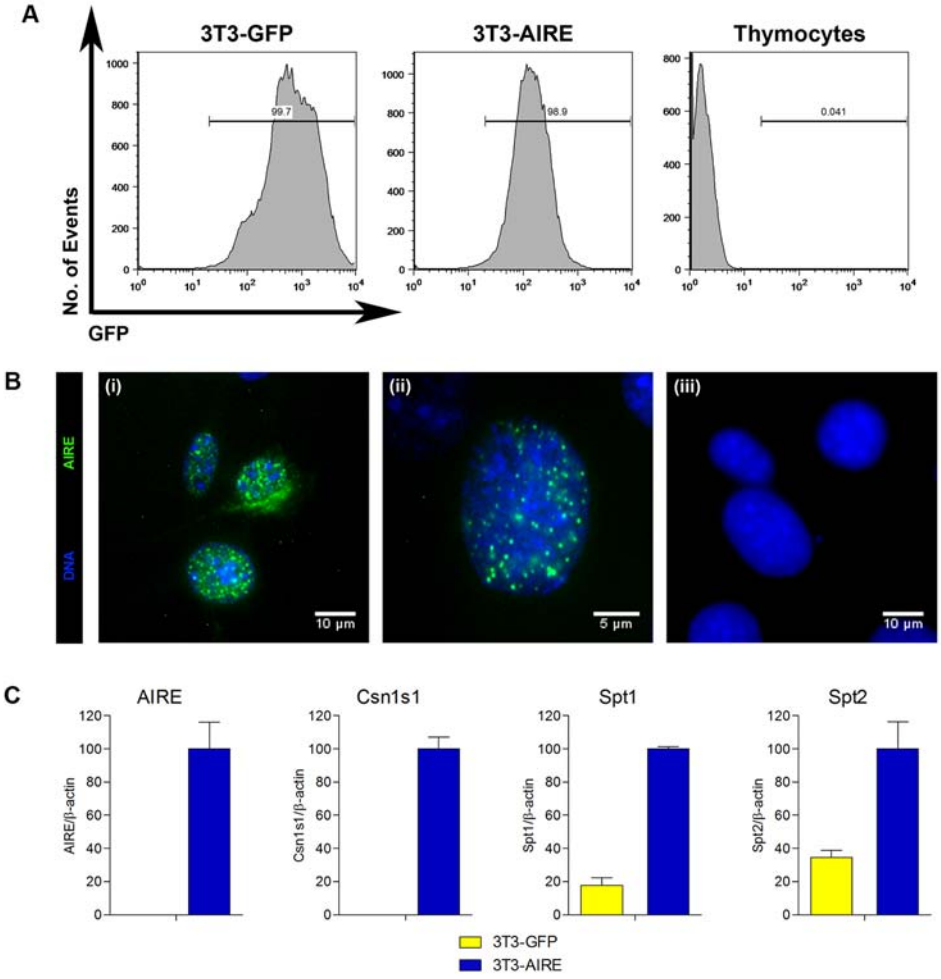
3.2. Analysis of the Effects of AIRE in a Non-Thymic Cell Line

3.2.1. Localisation and Expression of AIRE and Tissue-Restricted Antigens in a Non-Thymic 3T3 Fibroblast Cell Line

In a thymic epithelial cell background, we have demonstrated that AIRE is able to assemble and localise to nuclear sub-structures and exert its transcriptional activity, mirroring its behaviour *in vivo*. However, in order to determine whether AIRE alone is sufficient for promiscuous gene expression, or whether factors unique to thymic-derived cells are required for correct functioning, a *Mus musculus* fibroblast 3T3 line was stably transfected with either MIg-GFP-AIRE (3T3-AIRE), or MIg-GFP alone (3T3-GFP) as a control. As for the TEP model system, both 3T3 populations were FACS sorted to ensure 100% transfection efficiency, verified through flow cytometric analysis (Figure 3.5 A). Immunofluorescence staining with anti-AIRE antibody revealed a sub-cellular localisation similar to that shown for AIRE in the thymic epithelial cell background, with AIRE-positive 3T3 cells displaying a punctate nuclear localisation of AIRE not observed in the 3T3-GFP control cells (Figure 3.5 B).

We next sought to determine whether the presence of AIRE within these cells was sufficient to trigger promiscuous gene expression. The degree of expression of AIRE and the three TRA *Csn1s1*, *Spt1* and *Spt2* was analysed with 3T3-AIRE and 3T3-GFP cDNA, and compared to β-actin, levels of which were equal for both 3T3 populations. Transcript levels of AIRE, *Csn1s1*, *Spt1* and *Spt2* were detected at elevated levels in 3T3-AIRE cells, while the 3T3-GFP control cells showed little or no

Figure 3.5 – Analysis of the Effects of AIRE in a Non-Thymic Cell Line: Transfection Efficiency, Subcellular Distribution and Expression of the Autoimmune Regulator and Tissue-Restricted Antigens



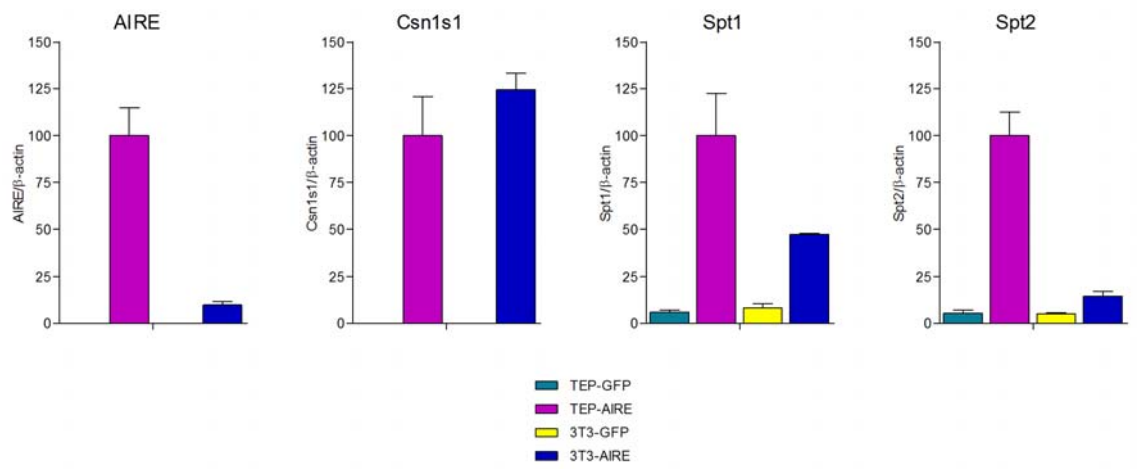
A, *Mus musculus* 3T3 fibroblast (3T3) cell lines were transfected with MIg virus bicistronic constructs containing either AIRE-GFP (3T3-AIRE), or GFP alone (3T3-GFP) as a control. Transfected populations were clonally expanded and purified. For verification of the efficiency of transfection, flow cytometric analysis was carried out, gated on GFP, with freshly isolated thymocytes as a negative control. 3T3-AIRE and 3T3-GFP lines were seen to express GFP at 100% efficiency. *B*, Immunofluorescence of 3T3 cells, staining with FITC anti-AIRE antibody (green) and DNA counterstained with DAPI (blue). Punctate AIRE nuclear bodies were observed in 3T3-AIRE cells (i, ii), but not in 3T3-GFP (iii). *C*, Real-time quantitative PCR was used to compare the relative mRNA expression levels of AIRE and three tissue-restricted antigens; casein-α (*Csn1s1*); salivary protein-1 (*Spt1*); and salivary protein-2 (*Spt2*), as indicated, showing active expression of AIRE and its target genes. Yellow columns represent data from 3T3-GFP cells and blue columns represent results from 3T3-AIRE cells. Data was normalised to β-actin expression levels as standard. Data are the mean ±SEM from technical triplicate reactions, and are representative of at least two distinct cDNA preparations.

amount of these genes (Figure 3.5 C). These results clearly illustrate that AIRE is able to exert its function independently of a thymic cell environment; localising to nuclear speckles and inducing the expression of TRA, in a manner directly comparable to that within wild-type mTECs and the TEP model cell line. These findings also confirm that AIRE is capable of up-regulating TRA without additional molecular requirements and that the presence of AIRE is itself sufficient for the initiation of promiscuous gene expression. However, the absolute transcript levels detected for AIRE, *Spt1* and *Spt2* in the 3T3-AIRE cells were considerably lower compared to the levels within TEP-AIRE cells (Figure 3.6). This was not a result of transfection efficiency since both parent lines were FACS sorted and purified prior to experimental analysis and frequent analysis of AIRE expression confirmed the maintenance of this transfection. Therefore this decreased expression level of AIRE and the two salivary proteins in the 3T3-AIRE cells may imply that AIRE requires certain unidentified thymic factors to function optimally.

3.3. Patterns of Histone Modifications across Tissue-Restricted Antigens in the Thymic Epithelial Cell Model System as Revealed by Native Chromatin Immunoprecipitation

AIRE, with its SAND domain, two PHD-zinc fingers and LXXLL motifs, shows transcriptional transactivating properties and is predicted to be a transcriptional regulator, contributing to central tolerance through the control of promiscuous gene expression within the thymus (Pitkanen *et al.* 2000; Kumar *et al.* 2001; Su and Anderson 2004; Devoss and Anderson 2007). Current technologies enable the

Figure 3.6 - Analysis of the Transcriptional Effects of AIRE: Comparative Expression Levels of the Autoimmune Regulator and Tissue-Restricted Antigens within Thymic and Non-Thymic Backgrounds



Real-time quantitative PCR was used to compare the relative mRNA expression levels of AIRE and three tissue-restricted antigens in *Mus musculus* thymic epithelial (TEP) and 3T3 fibroblast (3T3) cell lines, transfected with MIg virus bicistronic constructs containing either AIRE-GFP (TEP-AIRE / 3T3-AIRE), or GFP alone (TEP-GFP / 3T3-GFP) as a control. Data displayed is a comparison of the levels of AIRE expression, and its three target genes; casein-α (*Csn1s1*); salivary protein-1 (*Spt1*); and salivary protein-2 (*Spt2*), as indicated. Although AIRE-positive populations were sorted to ensure 100% transfection efficiency, expression of AIRE, *Spt1* and *Spt2* was more efficient in the thymic background. *Csn1s1* levels appear unaffected. Turquoise columns represent data from TEP-GFP cells, pink columns; results from TEP-AIRE cells, yellow columns; from 3T3-GFP cells and blue columns; from 3T3-AIRE cells. Data was normalised to β-actin expression levels as standard and all expression levels were set relative to those of TEP-AIRE for comparison. Data are the mean ±SEM from technical triplicate reactions, and are representative of at least two distinct cDNA preparations.

comprehensive study of gene expression, and with the advance of knowledge regarding the role nucleosomal histone proteins play in these processes, the function of effector molecules involved in transcriptional activation, such as AIRE, can be examined more thoroughly.

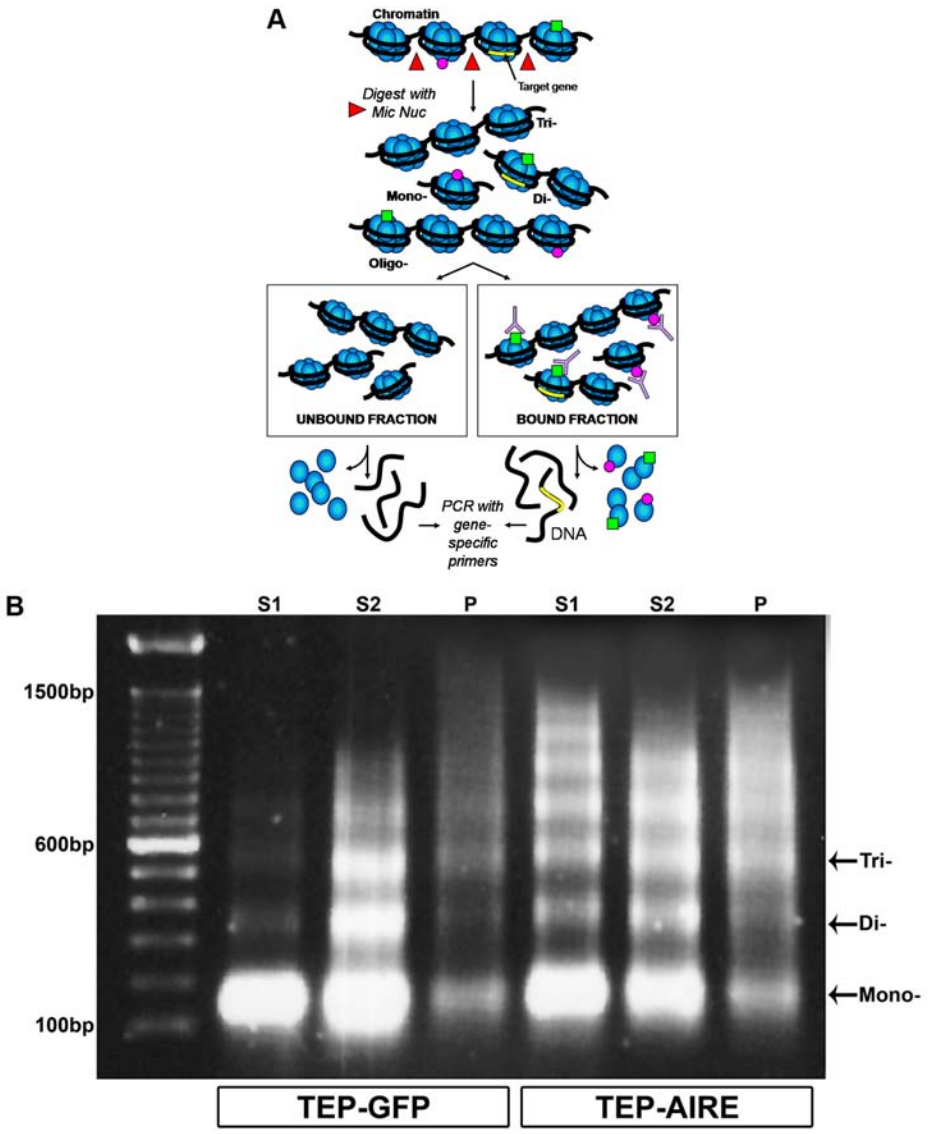
Post-translational modifications to the tails of histone proteins, such as acetylation and methylation, are important epigenetic marks, playing a role in the control of eukaryotic gene expression programmes by influencing chromatin structure (Kouzarides 2007). Whilst immunofluorescence microscopy and Western blotting have proved invaluable for investigating the distribution of histone modifications on a genome-wide scale, they are unable to directly address the epigenetic status of chromatin at the single gene level. Through the development of native chromatin immunoprecipitation (NChIP), investigation into the relative levels of a variety of histone tail modifications at specific gene regions has been made possible (O'Neill and Turner 2003).

Given AIRE's ability to switch on so many TRA, we sought to analyse the histone modifications associated with these AIRE-regulated genes through NChIP. Cultured TEP-AIRE and TEP-GFP cells were harvested in the absence of a cross-linker and chromatin was isolated and prepared through mild micrococcal nuclease digestion. Subsequent analysis on 1.2% agarose gels consistently revealed classical oligonucleosomal ladders, with the first supernatant (S1) fractions showing a high proportion of mononucleosomes, the S2 fractions displaying the di-, tri-, tetra- and pentanucleosomes, and the pellet (P) samples showing the higher molecular

insoluble material. For immunoprecipitation, an input sample representative of the starting material was required and therefore supernatants S1 and S2 were routinely combined, which together approximate to 90% of the DNA. Input chromatin was then incubated with antibodies directed against a panel of activating and silencing histone modifications (Table 3.1), global levels of which were previously shown to be unaffected by AIRE (Figures 3.3 and 3.4). Immunocomplexes were then isolated through addition of protein A-sepharose, generating bound (B) and unbound (UB) fractions following elution. The success of each IP was routinely monitored by calculation of the percentage pull-down of DNA in the bound fraction and visually by 1.2% AGE (Figure 3.7), and typically antibody pull-downs were in the range of 1 to 28% (Table 3.2).

For the determination of the relative amounts of histone modification across specific promiscuous gene regions, qPCR was employed. To ensure accurate measurement and interpretation of the data generated for each gene region, a standard curve was generated for each primer set with a range of *Mus musculus* genomic DNA standards, to which comparisons were made following each qPCR run. Equal concentrations of bound and unbound NChIP DNA were used for qPCR analysis, and the ratio of bound to unbound signal was calculated. For each histone modification, a bound:unbound ratio less than 1 suggests a depletion of that specific mark, whereas a ratio greater than 1 equates to enrichment over that specific gene region.

Figure 3.7 – Native Chromatin Immunoprecipitation: Isolation and Micrococcal Nuclease Digestion of Chromatin from Unfixed Thymic Epithelial Cells



Outline of the native chromatin immunoprecipitation (NChIP) protocol. *A*, Diagrammatic representation of the NChIP procedure. Chromatin from cultured cell nuclei is isolated and digested with micrococcal nuclease (Mic Nuc) then immunoprecipitated with antibodies directed against post-translational histone modifications. DNA from antibody-bound and –unbound fractions is then purified for analysis of the relative levels of each modification. *B*, Routine analysis of chromatin fractions by 1.2% agarose gel electrophoresis. Chromatin from a minimum of $1x10^7$ unfixed AIRE-positive and AIRE-negative *Mus musculus* thymic epithelial (TEP) cells was isolated and digested with micrococcal nuclease, generating a ladder rich in mono-, di-, tri-, tetra-, and pentanucleosomes. Equal amounts of the first supernatant (S1), soluble (S2) and insoluble (P) fractions were analysed by electrophoresis and visualised with ethidium bromide. Size was determined by reference to a 100 base pair molecular marker (Invitrogen). Mono-, di- and tri-nucleosomes are indicated.

Table 3.2 – Efficiency of Pull-Down for Each Histone Modification Following Native Chromatin Immunoprecipitation

NChIP PERCENTAGE PULL-DOWNS				
ANTIBODY	SPECIFICITY	TEP-GFP	TEP-AIRE	REPLICATES
R204	H3K4me1	0.70-6.06%	1.17-10.68%	3
R149	H3K4me2	0.76-5.15%	0.90-10.34%	3
R612	H3K4me3	1.28-5.05%	5.19-12.38%	3
R607	H3K9ac	6.86-9.53%	0.49-6.92%	3
R403	H4K8ac	0.70-12.55%	1.49-28.15%	4
R252	H4K16ac	0.65-4.25%	1.56-28.67%	4
07-212	H3K9me2	1.64-1.77%	1.08-1.18%	3
07-449	H3K27me3	3.18-7.53%	2.19-15.24%	3
PI	Pre-immune	0.46-0.61%	0.23-0.61%	4

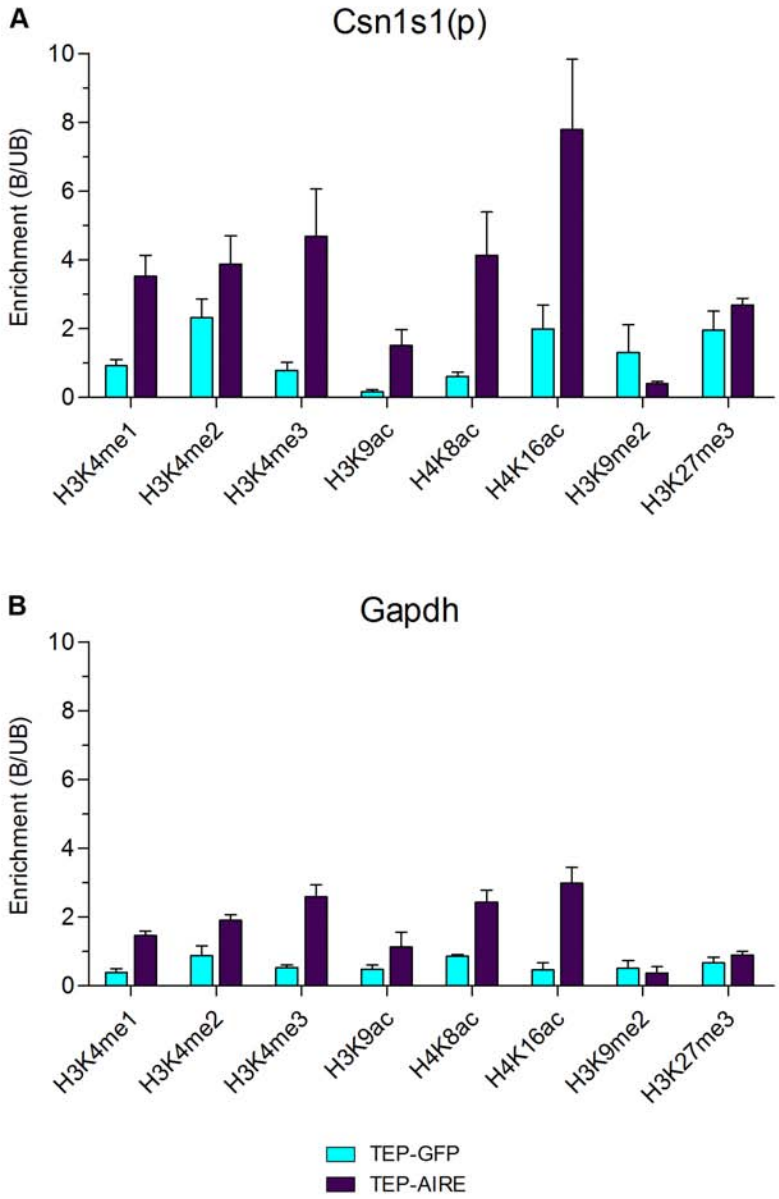
Native chromatin immunoprecipitation was performed with cultured AIRE-negative (TEP-GFP) and AIRE-positive (TEP-AIRE) thymic epithelial cell lines, using antibodies directed against the post-translational histone modifications displayed above. DNA from antibody-bound (B) and –unbound (UB) fractions was purified and the percentage pull down (B/UBx100) for each antibody was then calculated. Pre-immune no antibody controls were included which show comparatively low pull-down efficiencies. The number of biological replicates for each modification is displayed.

3.3.1. Pattern of Histone Modifications at the Casein-α Promoter Region and the Glyceraldehyde-3-phosphate Dehydrogenase Locus in the Thymic Epithelial Cell Model System

The expression of the TRA *Csn1s1* has been reliably demonstrated to be heavily dependent on AIRE, both within the thymic medulla, and in the presence of AIRE within our TEP model system and thus represents an ideal candidate gene (Anderson *et al.* 2002; Derbinski *et al.* 2005; Derbinski *et al.* 2008). The panel of histone modifications examined reflects a diversity of potential functional outcomes, with three acetylation marks; H3K9ac, H4K8ac and H4K16ac, which are generally associated with active euchromatic genes, and three active methylation marks; H3K4me1, me2, me3, and two silent methylation marks H3K9me2 and H3K27me3 (Table 3.1) (Kouzarides 2007). Verification that the global levels of these modifications were unaffected by AIRE means that any changes observed on a gene-by-gene basis can be attributed to AIRE's control of PGE.

For TEP-AIRE cells, in which *Csn1s1* is actively expressed, the promoter region of this gene is marked by significantly elevated levels of H3 and H4 acetylation, modifications usually indicative of active gene expression (Figure 3.8 A) (Fuchs *et al.* 2009). All three forms of H3K4 methylation are also enriched, in particular, high levels of H3K4me3 (bound:unbound ratio 4.68) were detected. H3K4 methylation is generally considered to be associated with active euchromatic genes, as revealed by whole-genome ChIP-Seq approaches (Barski *et al.* 2007; Kouzarides 2007). H3K4me3 is often found localised to the promoter region of transcriptionally active genes, whereas me1 and me2 can occur in a more widespread pattern downstream

Figure 3.8 - Quantitation of the Relative Levels of Histone Modifications across the Promoter Region of Casein-α and Glyceraldehyde-3-phosphate Dehydrogenase by Native Chromatin Immunoprecipitation



Quantitation by native chromatin immunoprecipitation (NChIP) of levels of histone modifications at the promoter regions of casein-α (*Csn1s1, A*) and glyceraldehyde-3-phosphate dehydrogenase (*Gapdh, B*) in cultured *Mus musculus* thymic epithelial (TEP) cell lines. Turquoise columns represent results from AIRE-negative TEP-GFP cells. Purple columns are from TEP cells transfected with AIRE. Immunoprecipitation performed with affinity-purified antibodies to H3K4me1, H3K4me2, H3K4me3, H3K9ac, H4K8ac, H4K16ac, H3K9me2 and H3K27me3, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA by quantitative real-time PCR. Data are the mean ±SEM from at least two separate NChIP experiments.

of transcriptional start sites (TSS), with the marks becoming progressively more restricted to the TSS as the degree of methylation increases from me1 to me2 to me3 (Barski *et al.* 2007). In contrast, lower levels of H3K9me2 and H3K27me3 were seen in the presence of AIRE, with H3K9me2 showing a very low bound:unbound ratio of 0.40, yet H3K27me3 was not depleted, with a bound:unbound ratio of 2.69. For control TEP-GFP cells, the promoter region of *Csn1s1* is marked by a contrasting pattern of modifications with a general depletion of active marks including H3K4me1, H3K4me3, H3K9ac and H4K8ac. However, the promoter of this silent gene was not devoid of active modifications as H3K4me2 and H4K16ac were detected, although these levels were matched by enrichment of silencing modifications H3K9me2 and H3K27me3 (bound:unbound ratio 1.31 and 1.96 respectively). The presence of these marks at the promoter of a non-expressed gene is typical as these histone modifications are frequently associated with transcriptional repression; H3K9me2 through recruitment of heterochromatin protein 1 (HP1); and H3K27me3 via interactions with PcG proteins such as PRC1 (Cao and Zhang 2004; Dellino *et al.* 2004; Kondo *et al.* 2008). Overall, we see how the presence of AIRE leads to heightened levels of active marks, a loss of H3K9me2 but a maintenance of H3K27me3.

Changes to the levels of histone modifications at the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*) were notably more moderate across the two cell populations in comparison to those seen for *Csn1s1* (Figure 3.8 B). It should be noted that normalisation to *Gapdh*, or any other gene, was not carried out as this would assume maintenance of a constant epigenetic state, which

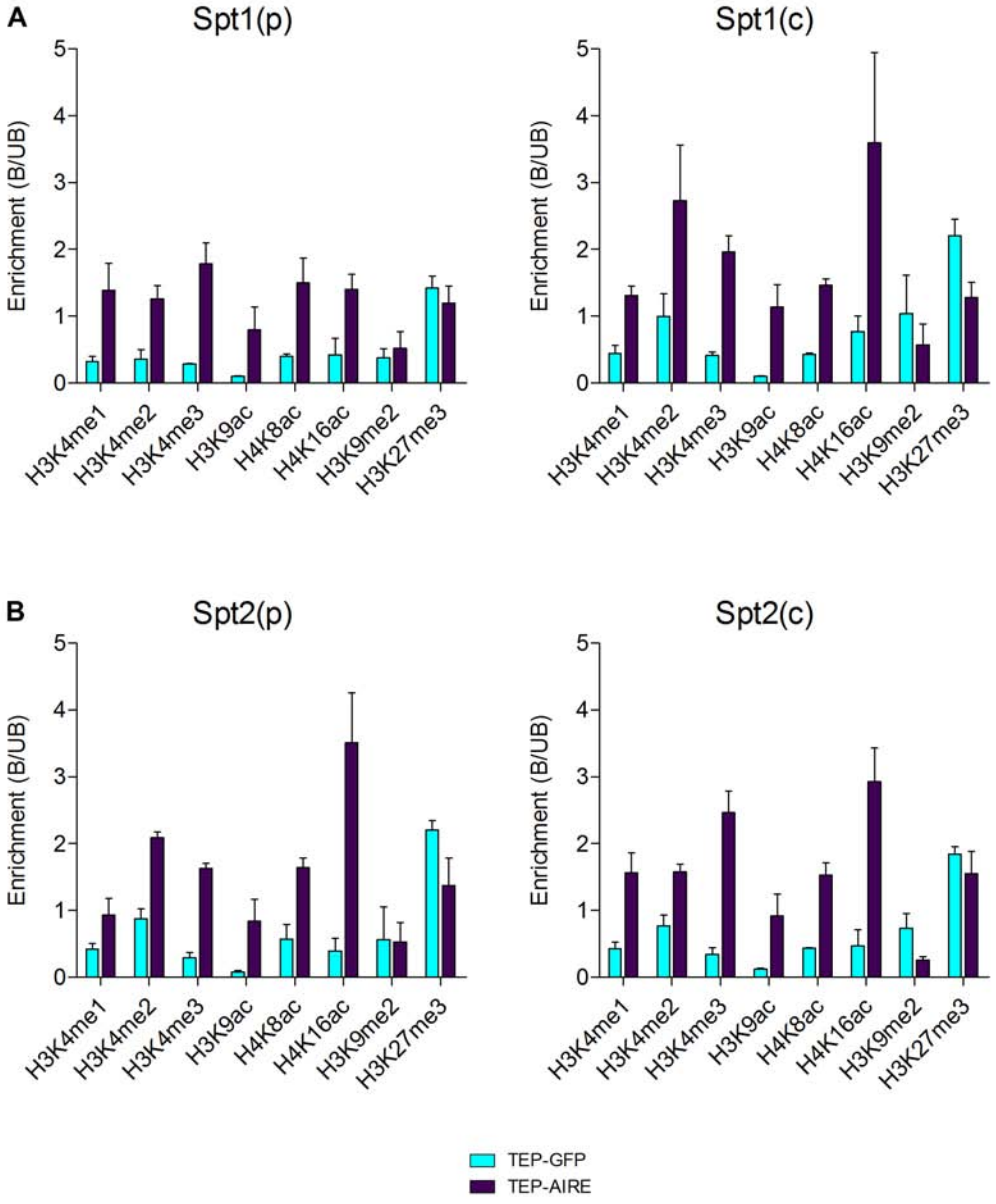
is not always the case (Valls *et al.* 2005; O'Neill *et al.* 2006). Most importantly, this analysis showed a depletion of the marks characteristic of silent promoters, H3K9me2 and H3K27me3 in TEP-AIRE and TEP-GFP cells for this actively-expressed gene, with both cell populations displaying similar levels of the two modifications. Noticeably, there appears to be an overall depletion in all histone modifications for TEP-GFP, with only H4K8ac and H3K4me2 detected at levels greater than H3K9me2 and H3K27me3. In addition, H3K9 acetylation, which is frequently enriched in active genes, is approximately 3-fold greater at *Gapdh* when compared to the *Csn1s1* promoter (Roh *et al.* 2005). However, AIRE does appear to influence the distribution of marks across this housekeeping gene, as TEP-AIRE cells show an enrichment of the active methylation and acetylation marks on histones H3 and H4, while TEP-GFP cells show a general depletion of these marks.

3.3.2. Analysis of the Effects of AIRE upon the Salivary Protein Genes on *Mus musculus* Chromosome 15 in the Thymic Epithelial Cell Model System

The TRA *Spt1* and *Spt2* have been well documented in terms of their AIRE-dependence by numerous groups, and showed induced expression in the presence of AIRE within our TEP model system (Anderson *et al.* 2002; Rossi *et al.* 2007; Kont *et al.* 2008). These genes are located adjacent to each other on *Mus musculus* chromosome 15 and we therefore wanted to investigate the pattern of histone modifications across the promoter (p) and coding (c) regions of both salivary proteins to determine AIRE's affect on neighbouring genes.

The distribution of modifications for both salivary proteins essentially reflected those seen for *Csn1s1*, however, in general, all eight modifications were less enriched in both cell populations across the two salivary proteins, when compared to the levels seen for *Csn1s1*. When we compare the pattern of modifications across *Csn1s1* and the two salivary proteins it would appear that a common theme is arising for TEP-GFP cells, with an overall depletion of H3 and H4 acetylation and mono-, di- and tri-methylation at histone H3K4; levels consistently showing me3 < me1 < me2 preference (Figures 3.8 and 3.9). For the salivary proteins, TEP-GFP cells show a depletion of all but H3K27me3, levels of which are maintained across the promoter and coding regions of both *Spt1* and *Spt2* at around a bound:unbound ratio of 2.00 (Figure 3.9). In contrast, H3K9me2 appears to show a preferential localisation at the 3` end of both salivary proteins (*Spt1*(c) 1.04, *Spt2*(c) 0.74), with comparatively lower levels at the promoters (*Spt1*(p) 0.38, *Spt2*(p) 0.54). Although the levels of H3K9me2 are surprisingly low, given this marks recognised localisation in silenced genes, they are consistently higher than the archetypal active modifications including H3K4me3 and H3K9ac in the TEP-GFP cells (Barski *et al.* 2007). Therefore the dominance of H3K27me3 and H3K9me2 at these silent gene regions in the control cells may dictate the silencing of these genes. TEP-AIRE cells also show some commonalities across all three TRA *Csn1s1*, *Spt1* and *Spt2* including hyperacetylation of histone H4, elevated levels of methylation of histone H3K4, and a depletion of H3K9me2, yet detectable levels of H3K27me3, often equalling the background TEP-GFP levels (Figures 3.8 and 3.9). As for *Csn1s1*, the trend observed for the salivary proteins in the presence of AIRE is that of enriched levels

Figure 3.9 - Quantitation of the Relative Levels of Histone Modifications across Salivary Protein-1 and Salivary Protein-2 by Native Chromatin Immunoprecipitation



Quantitation by native chromatin immunoprecipitation (NChIP) of levels of histone modifications at the promoter and coding regions of salivary protein-1 (*Spt1, A*) and salivary protein-2 (*Spt2, B*) in cultured *Mus musculus* thymic epithelial (TEP) cell lines. Turquoise columns represent results from AIRE-negative TEP-GFP cells. Purple columns are from TEP cells transfected with AIRE. Immunoprecipitation performed with affinity-purified antibodies to H3K4me1, H3K4me2, H3K4me3, H3K9ac, H4K8ac, H4K16ac, H3K9me2 and H3K27me3, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA by quantitative real-time PCR. Data are the mean ±SEM from at least two separate NChIP experiments.

of active marks, most noticeably for H3K9ac and H3K4me3, with little change to the silencing modifications.

Distinct differences not only present themselves between the two salivary proteins themselves, but also between the promoter and coding regions. TEP-AIRE cells display subtly diverse patterns of histone modifications which fluctuate as you move from the 5` to the 3` of both *Spt1* and *Spt2*. Acetylation, for example, appears to be generally constant across both loci, with H4K8ac levels remaining relatively constant at a bound:unbound ratio of around 1.50, as does H3K9ac (≈1.00). H4K16ac also shows a domain-wide enrichment, starting at a ratio of 1.40 in *Spt1*(p) then increasing to around 3.00 in the coding regions, a level which is then maintained across *Spt2*. Methylation however, shows a less consistent distribution. For *Spt1*(p), the highest level of enrichment is for H3K4me3, however, as you move to the coding region of this gene, H3K4me2 levels peak, whereas the reverse is seen for *Spt2*.

Analysis and comparison of histone modifications across three TRA under the transcriptional control of AIRE; *Csn1s1*, *Spt1* and *Spt2*, has revealed some interesting and distinct differences. The arrangement of epigenetic marks in TEP-AIRE cells varies greatly from what appears to be a set pattern for these three genes within the TEP-GFP control population, particularly with regards to H3K4 methylation, which shows significant enrichment in the presence of AIRE. This investigation has also highlighted the specificity of histone modifications, with unique arrangements occurring even when two genes are adjacent to each other on a

chromosome, possibly suggesting individual regulation of each gene. These distinctions become even more apparent when comparisons are made between genes located on different chromosomes (*Spt1* / *Spt2*; *Mus musculus* chromosome 15, *Csn1s1*; *Mus musculus* chromosome 5). However, some of our findings do suggest that the regulation of these genes may occur through a domain-wide deposition of histone modifications, as evidenced by the high levels of H4K8 and H4K16 acetylation across both salivary proteins.

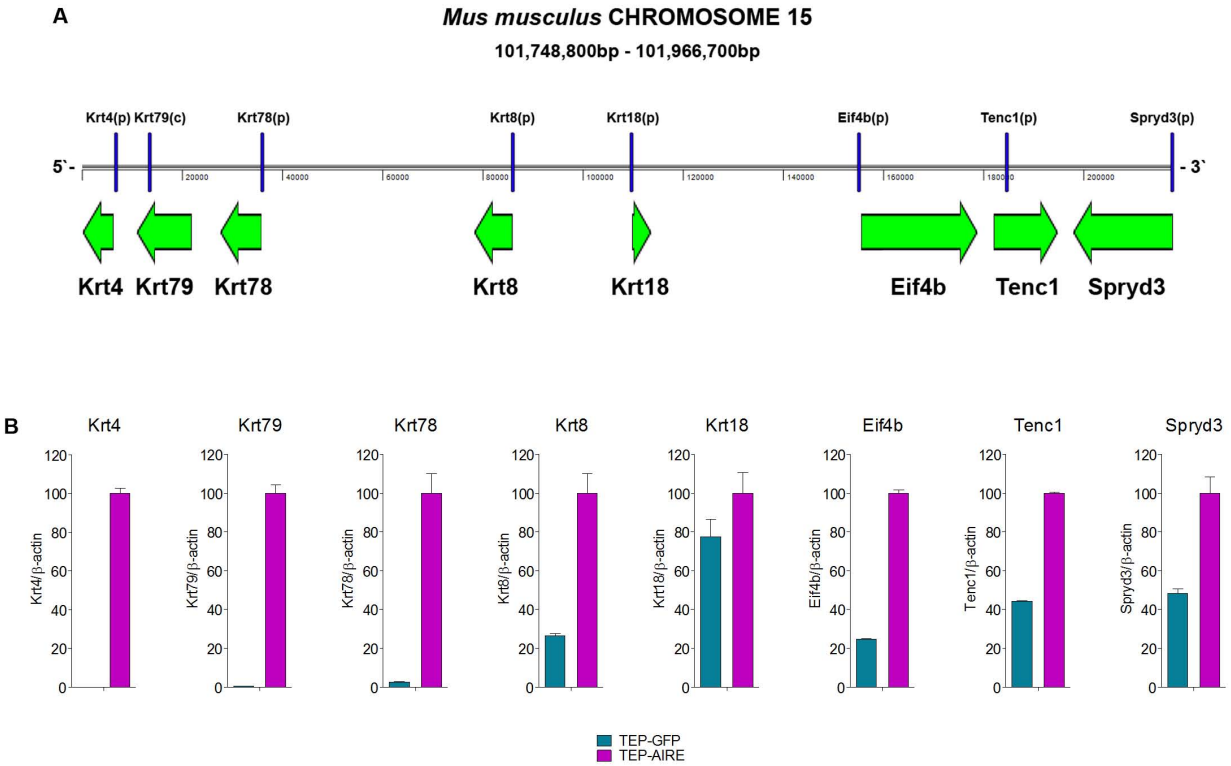
3.3.3. Analysis of the Effects of AIRE upon a Cluster of AIRE-Regulated Genes; the Keratin Cluster on *Mus musculus* Chromosome 15 in the Thymic Epithelial Cell Model System

Genome-wide analysis of genes under the control of AIRE highlighted the significant impact this molecule has on transcriptional programmes within mTECs (Anderson *et al.* 2002; Gotter *et al.* 2004). The many thousands of TRAs influenced by AIRE were found to cluster along individual chromosomes and currently the issue of how AIRE controls the expression of each individual gene remains elusive (Gotter *et al.* 2004; Derbinski *et al.* 2005; Johnnidis *et al.* 2005). AIRE is not believed to induce gene expression on a gene-by-gene basis, but instead is thought to regulate transcription through a domain-wide process (Gotter *et al.* 2004; Derbinski *et al.* 2005; Johnnidis *et al.* 2005). Interestingly however, the impact of AIRE on each gene within a cluster is not a simple on or off switch of all loci, but rather disperse; some genes increasing in expression, while neighbouring genes were either unaffected or their expression decreased (Gotter *et al.* 2004; Derbinski *et al.* 2005; Johnnidis *et al.* 2005). This punctate expression pattern of certain AIRE-regulated clusters tends to argue

against a domain-wide control such as deposition of activating histone modifications across the whole locus, but instead may suggest that individual genes within the cluster are affected in distinct ways.

The keratin cluster is one such example of an AIRE-regulated gene cluster (Johnnidis *et al.* 2005). This cluster, located on *Mus musculus* chromosome 15, contains five keratin genes (*Krt4*, *Krt79*, *Krt78*, *Krt8* and *Krt18*) at its 5` end, and eukaryotic translation initiation factor 4b (*Eif4b*), tensin-like C1 domain-containing phosphatase (*Tenc1*) and SPRY domain-containing 3 (*Spryd3*) further downstream. Expression profiling had been performed with a number of the genes in this cluster, revealing differential expression levels in the presence of AIRE, however, *Krt79*, *Krt78*, *Krt8*, *Tenc1* and *Spryd3* had not been investigated (Johnnidis *et al.* 2005). We therefore completed this analysis within our TEP model system for all genes across the cluster. Interestingly, transcript levels of all genes showed up-regulation in the presence of AIRE (Figure 3.10). We observed particularly strong signals for keratin 4, in agreement with published expression data, but also for keratin 79 and 78, with little or no detection in TEP-GFP cells when normalised to β-actin. We found that the next gene in the cluster, keratin 8 was expressed in the control cells, however, TEP-AIRE cells displayed an approximate 4-fold up-regulation. In addition, keratin 18, which had shown to be negatively-regulated by AIRE, displayed a consistent level of expression in both cell populations, suggesting that this gene is not under the transcriptional control of AIRE (Johnnidis *et al.* 2005). We found that for *Eif4b*, which had previously shown decreased transcription in the presence of AIRE, expression levels were instead positively influenced by AIRE in our system,

Figure 3.10 – Impact of AIRE on a Cluster of Genes: Expression Levels across the Keratin Gene Cluster



A, Diagrammatic representation of the keratin cluster on *Mus musculus* chromosome 15 indicating the position of all genes and primer sets at promoter (p) and coding (c) regions. *B*, Real-time quantitative PCR was used to compare the relative mRNA expression levels for all neighbouring genes across the keratin cluster, as indicated, in thymic epithelial (TEP) cells showing fluctuating expression levels across the cluster. Turquoise columns represent data from TEP-GFP cells and pink columns represent results from TEP-AIRE cells. Data was normalised to β-actin expression levels as standard. Data are the mean ±SEM from technical triplicate reactions, and are representative of at least two distinct cDNA preparations.

increasing approximately 4-fold in TEP-AIRE cells (Johnnidis *et al.* 2005). Transcript levels of *Tenc1* and *Spryd3* were consistently 2-fold higher in the presence of AIRE, however, robust expression was occurring in the TEP-GFP cells for all three 3` genes. Collectively, this confirms that the keratin cluster represents a true AIRE-regulated cluster, with AIRE acting on a gene-by-gene basis to either induce transcription or, as is the case for keratin 18, exert no influence.

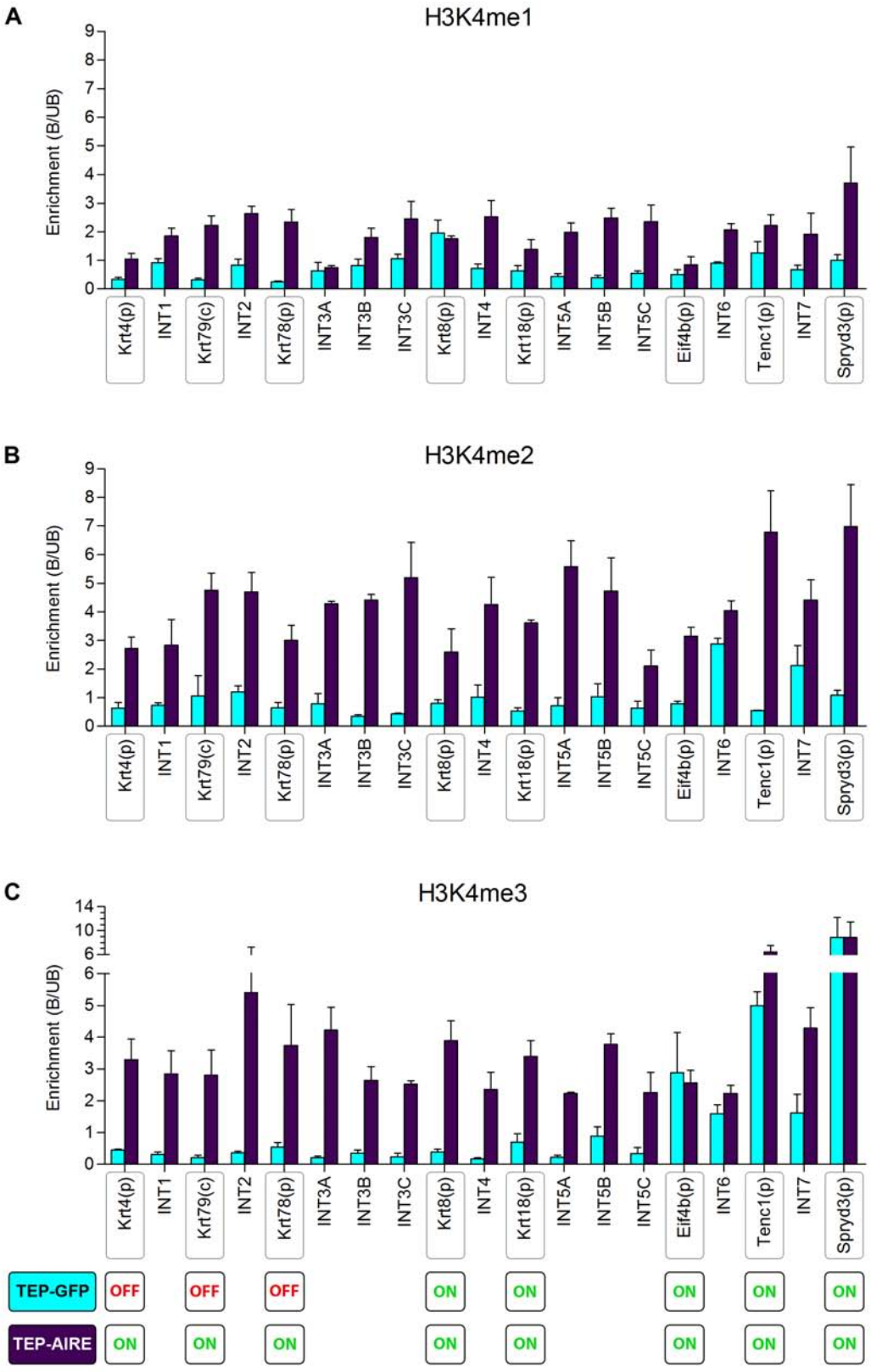
3.3.3.1. Levels of Methylation at Histone H3 Lysine 4 across the Keratin Cluster in the Thymic Epithelial Cell Model System

NChIP was used to assay levels of mono-, di- and tri-methylation of histone H3K4 across the promoter or coding regions of the genes in the keratin cluster, in addition to intergenic (INT) regions, in the TEP model system (Figure 3.11). The most striking observation was that all three forms of methylation at this residue were enriched in the presence of AIRE, while TEP-GFP cells showed markedly lower levels. Across the entire cluster, the degree of methylation progressively increased in TEP-AIRE cells, the lowest amounts being for H3K4me1 and the highest for H3K4me3, which may be expected as expression of all genes is high in the presence of AIRE, and the three forms of methylation are known to compete with each other for the same lysine (Zhang and Reinberg 2001; Fuchs *et al.* 2009). Despite these initial similarities, some well defined differences do emerge between H3K4 methylation levels. For example, TEP-AIRE H3K4me1 (Figure 3.11 A) rises and falls in waves along the cluster, with peaks at the intergenic regions INT2, 3C, 4 and 5B and finally at *Spryd3*. In contrast, TEP-GFP cells show depletion of H3K4me1 at the promoter regions of keratins 4, 79 and 78 which are silent, with

Figure 3.11 - Impact of AIRE on a Cluster of Genes: Quantitation of the Relative Levels of Histone H3 Lysine 4 Methylation across the Keratin Cluster by Native Chromatin Immunoprecipitation

Quantitation by native chromatin immunoprecipitation (NChIP) of levels of histone modifications across the keratin cluster on *Mus musculus* chromosome 15 in cultured thymic epithelial (TEP) cell lines. Turquoise columns represent results from AIRE-negative TEP-GFP cells. Purple columns are from TEP cells transfected with AIRE. Immunoprecipitation performed with affinity-purified antibodies to; *A*, H3K4me1; *B*, H3K4me2; and *C*, H3K4me3, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA by quantitative real-time PCR. Data are the mean ±SEM from at least two separate NChIP experiments.

Figure 3.11 - Impact of AIRE on a Cluster of Genes: Quantitation of the Relative Levels of Histone H3 Lysine 4 Methylation across the Keratin Cluster by Native Chromatin Immunoprecipitation



correspondingly higher levels at the intergenic regions. However, this trend is lost at keratin 8, a gene actively expressed by these cells (albeit at a level 4–fold lower than in the presence of AIRE), when there is an obvious peak of H3K4me1, enrichment rising to a bound:unbound ratio of 1.95, equalling that of TEP-AIRE for this locus. TEP-GFP H3K4me1 levels then fall and remain low, with minor peaks occurring at *Tenc1* and *Spryd3* promoter regions, which correspond with the active expression of these genes in the control population.

The levels of H3K4me2 (Figure 3.11 B) displayed an overall increase across the cluster when compared to H3K4me1. TEP-AIRE H3K4me2 showed a similar distribution to me1 in the first half of the cluster, with waves of enrichment peaking at the intergenic regions INT2, 3C and 4. However, at the 3` end of the cluster, unlike the mono-methyl mark, H3K4me2 appears to show a domain of higher levels of enrichment, from *Tenc1* (ratio 6.78) and remaining high until the end of the cluster. TEP-GFP cells show an overall depletion of H3K4me2, similar to the mono-methyl mark, however, on a gene-by-gene basis some subtle contrasts were observed. For example, in general there was little distinction between TEP-GFP H3K4me2 levels at genes and intergenic regions in the 5` end of the cluster, whereas H3K4me1 showed more distinct peaks in the 5` intergenic regions. However, at the 3` end of the cluster, H3K4me2 peaks do appear in the intergenic regions, with correspondingly lower levels in the gene regions of *Eif4b*, *Tenc1* and *Spryd3* which show active transcription in the TEP-GFP cells, again a reversal of the pattern of H3K4me1 for TEP-GFP cells, which marked the 3` gene promoter regions preferentially.

The distribution of H3K4me3 (Figure 3.11 C) in TEP-AIRE cells, showed a more uniform pattern, especially in the first half of the cluster, in comparison to the waves of H3K4me1 and me2. H3K4me3 remained consistently high across all gene and intergenic regions, with particularly increased levels at the extreme 3` end of the cluster at *Tenc1* (bound:unbound ratio 6.42) and *Spryd3* (bound:unbound ratio 8.84). However, it is in the control population where the most distinctive pattern is observed, which appears to contrast with those of H3K4me1 and me2 for these cells. An almost total depletion of H3K4me3 occurred across the first half of the cluster until keratin 18 when levels start to rise. This corresponds directly with the TEP-GFP expression profile; transcription not taking place at significantly high levels until keratin 18. Interestingly, for keratin 8 where we do detect low levels of transcript in the control cells, H3K4me3 remains depleted. A definite domain of H3K4me3 is then identified in the downstream region, with peaks occurring specifically at the promoter regions of *Eif4b*, *Tenc1* and *Spryd3* at levels equal to those detected in the presence of AIRE, which again correlates well with the active transcription of these genes in the control cells. Taken together, the variations between the arrangements of H3K4 methylation across the cluster are particularly interesting and suggest a complex regulation of these genes, although each level of H3K4 methylation does appear to be affected by AIRE in a distinct way. Further analysis of alternative histone modifications is needed however, to clarify the role AIRE is playing in the regulation of the genes.

3.3.3.2. Levels of Acetylation at Histones H3 and H4 across the Keratin Cluster in the Thymic Epithelial Cell Model System

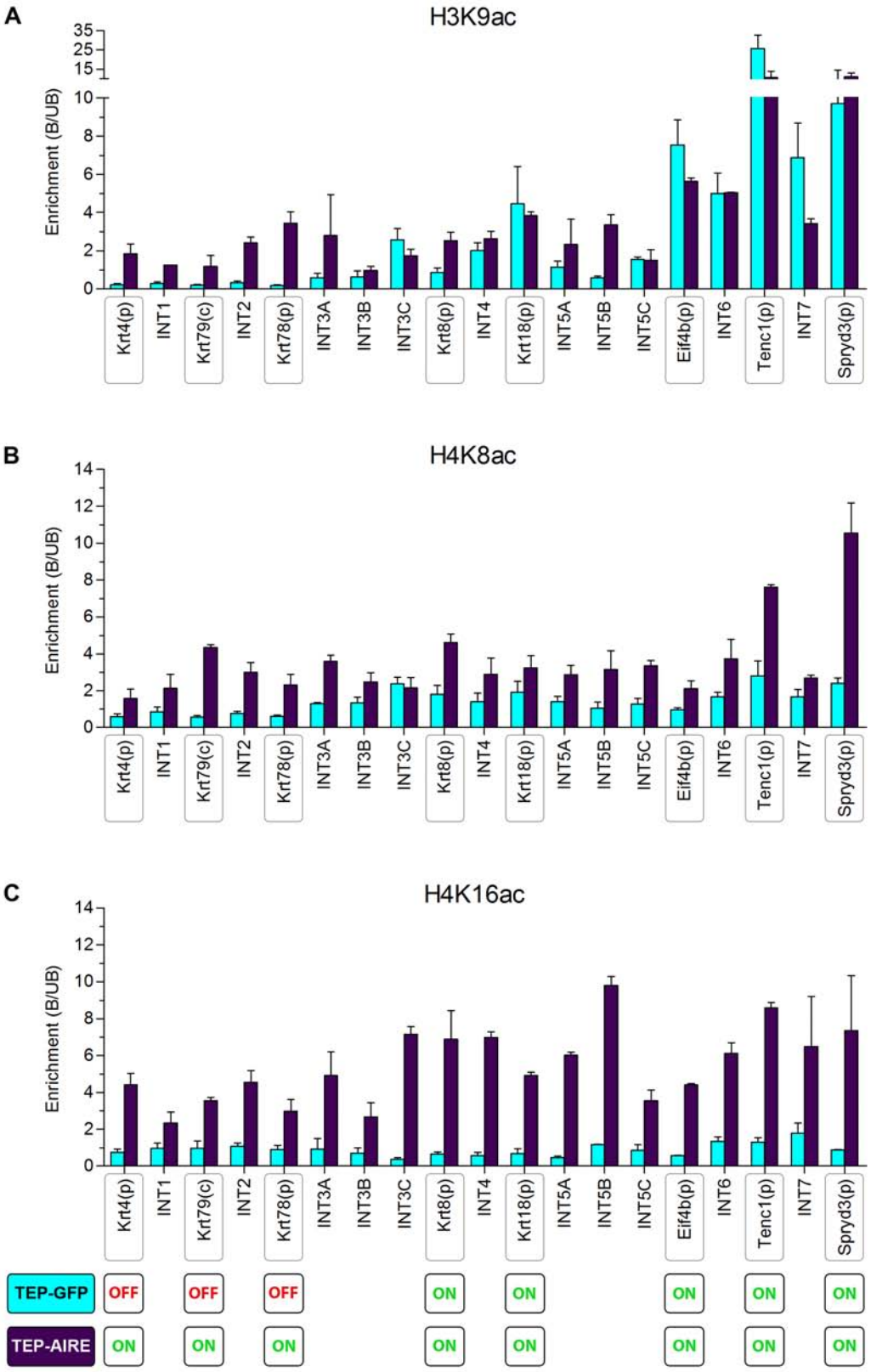
Three acetylation marks were analysed across the keratin cluster; H3K9, H4K8 and H4K16 (Figure 3.12). On first inspection, the patterns of acetylation vary widely for both TEP-AIRE and TEP-GFP cells, and between the three acetylation marks themselves, with a prominently defined domain of H3K9 acetylation detected downstream of the cluster for both cell lines, yet a more widespread distribution of H4 acetylation across all loci.

The pronounced domain of H3K9 acetylation (Figure 3.12 A) in the latter half of the cluster mirrors that observed for the other classically activating histone mark H3K4me3. This resemblance is particularly significant as these two modifications have been shown to be associated with active transcription (Barski *et al.* 2007). Both TEP-AIRE and TEP-GFP cells show a similar trend for H3K9ac, with a pronounced hyperacetylated domain in the 3` region of the cluster. Peaks of this modification occur mainly in the gene regions for both cell populations and correlate with the expression profile. TEP-AIRE cells show enrichment of H3K9ac at all genes across the cluster which is indicative of their active status. For TEP-GFP cells however, active expression is not detected in the first half of the cluster, where levels of H3K9ac remain very low, however from INT3C, levels of acetylation increase; correlating with the up-regulation of the 3` genes. By keratin 18; a gene whose expression is equally high in both the presence and absence of AIRE, we observed a high level of H3K9ac in the TEP-GFP cells (bound:unbound ratio 4.45); equalling the levels seen for this gene in the TEP-AIRE cells. As you continue along the

Figure 3.12 - Impact of AIRE on a Cluster of Genes: Quantitation of the Relative Levels of Histone Acetylation across the Keratin Cluster by Native Chromatin Immunoprecipitation

Quantitation by native chromatin immunoprecipitation (NChIP) of levels of histone modifications across the keratin cluster on *Mus musculus* chromosome 15 in cultured thymic epithelial (TEP) cell lines. Turquoise columns represent results from AIRE-negative TEP-GFP cells. Purple columns are from TEP cells transfected with AIRE. Immunoprecipitation performed with affinity-purified antibodies to; *A*, H3K9ac; *B*, H4K8ac; and *C*, H4K16ac, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA by quantitative real-time PCR. Data are the mean ±SEM from at least two separate NChIP experiments.

Figure 3.12 - Impact of AIRE on a Cluster of Genes: Quantitation of the Relative Levels of Histone Acetylation across the Keratin Cluster by Native Chromatin Immunoprecipitation



cluster, H3K9ac levels briefly fall in the intergenic INT5 region of TEP-GFP cells, but then increase dramatically for the final three genes, with enrichment matching or surpassing the values observed for TEP-AIRE cells; in particular for *Eif4b* (1.34-fold higher) and *Tenc1* (2.44-fold higher), despite the increased levels of expression in the presence of AIRE.

For acetylation of H4K8, an analogous distribution to H3K9ac is seen for the TEP-AIRE cells, with peaks at *Tenc1* and *Spryd3*, although H4K8ac is less dynamic, showing a more even distribution across the cluster (Figure 3.12 B). However, for the control TEP-GFP population, a striking contrast between H3K9 and H4K8 acetylation is observed with no indication of a 3` domain of H4K8 hyperacetylation. Instead, TEP-GFP H4K8ac shows a small domain of enrichment spreading from INT3A onwards, remaining relatively unchanged despite the altering levels of gene expression. INT3A is the intergenic region proceeding keratin 8, which is the first gene in the cluster to show expression in TEP-GFP cells, though at a very low level. Unlike for H3K9ac, this spread of H4K8ac does not show dramatic gene-specific peaks at *Eif4b*, *Tenc1* and *Spryd3* in the control cells, instead marking both promoters and intergenic regions equally, whereas H3K9ac shows a more specific preference for actively expressed genes.

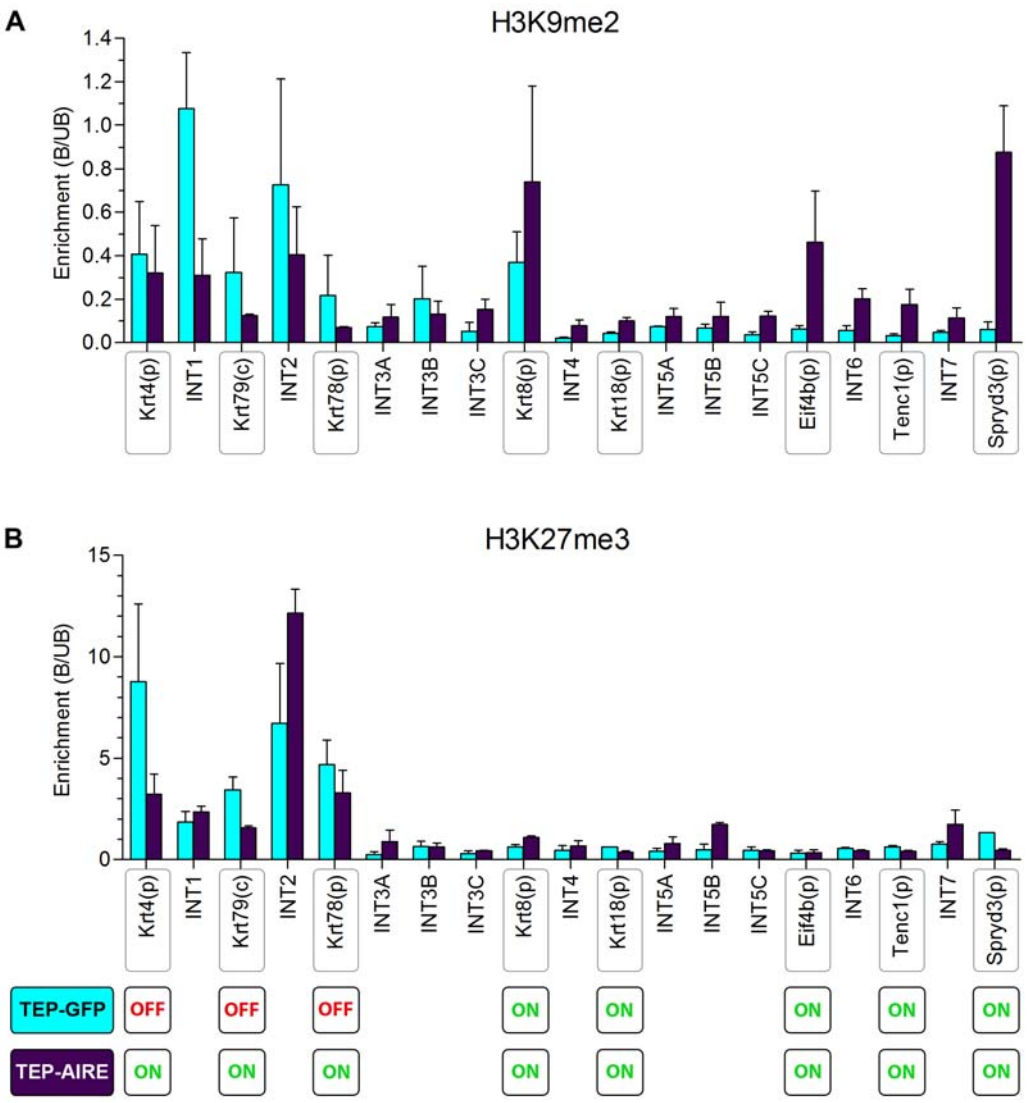
The pattern of H4K16 acetylation, in stark contrast to the previous marks, differs greatly for the TEP-AIRE and TEP-GFP populations (Figure 3.12 C). AIRE seems to induce a blanket of H4K16 hyperacetylation across the entire cluster, while the control cells present with very low levels of this modification, regardless of

transcriptional activity. This is particularly significant for keratin 18, whose expression is equal in both cell populations. Hence, this modification shows no specification for actively-transcribed genes, unlike the other acetylation marks. Lysine 16 of histone H4 is the first residue to be post-translationally modified and the addition of an acetyl moiety is known to disrupt the bridges between neighbouring nucleosomes (Luger *et al.* 1997; Shogren-Knaak *et al.* 2006). This mark may therefore represent a domain-wide opening of the chromatin, orchestrated by AIRE, thus facilitating the higher levels of gene expression across the cluster in the presence of AIRE.

3.3.3.3. Levels of Methylation at Histone H3 Lysine 9 and Lysine 27 across the Keratin Cluster in the Thymic Epithelial Cell Model System

Given the fluctuating levels of expression across the keratin cluster, analysis of two marks; H3K9me2 and H3K27me3, whose presence within a gene is usually indicative of repression was carried out (Figure 3.13) (Barski *et al.* 2007). The most striking observation is the strong enrichment of these two histone modifications at the 5` end of the cluster; an almost perfect reversal of the trend observed for the archetypal active histone modifications H3K4me3 and H3K9ac. Levels of the two silent marks correlated significantly with transcription levels in the control TEP-GFP population, with the highest levels present in the genes for which no transcription was detected. It should be noted that levels of H3K9me2 were much lower than those of H3K27me3, however it is clear that this modification still plays a role in the silencing of the genes in TEP-GFP cells (Figure 3.13 A). In the TEP-GFP control population, the four initial genes in the cluster (keratin 4, 79, 78 and 8) are either not

Figure 3.13 - Impact of AIRE on a Cluster of Genes: Quantitation of the Relative Levels of Histone H3 Lysine 9 and Histone H3 Lysine 27 Methylation across the Keratin Cluster by Native Chromatin Immunoprecipitation



Quantitation by native chromatin immunoprecipitation (NChIP) of levels of histone modifications across the keratin cluster on *Mus musculus* chromosome 15 in cultured thymic epithelial (TEP) cell lines. Turquoise columns represent results from AIRE-negative TEP-GFP cells. Purple columns are from TEP cells transfected with AIRE. Immunoprecipitation performed with affinity-purified antibodies to; *A*, H3K9me2; and *B*, H3K27me3, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA by quantitative real-time PCR. Data are the mean ±SEM from at least two separate NChIP experiments.

expressed, or expressed at a very low level and therefore increased levels of both inactive histone modifications within these gene regions echoes their silent nature. Further along the cluster, both modifications become dramatically depleted, H3K9me2 from keratin 18 onwards and H3K27me3 from INT3A as gene expression begins in the TEP-GFP cells. Although TEP-AIRE H3K9me2 levels were similarly very low, the specific 5` domain-wide enrichment was not detected to as great a level than in the control population. Peaks were seen at keratin 8 in the middle of the cluster and at *Eif4b* and *Spryd3* in the 3` end, which are all expressed approximately 4-fold higher in the presence of AIRE. For H3K27me3 however, these downstream peaks were not seen and a more well-defined distribution of this silent mark was observed for both TEP-AIRE and TEP-GFP populations, with a region of H3K27 hypermethylation at the first three genes (keratin 4, 79 and 78), followed by a depletion of this mark for the remainder of the cluster.

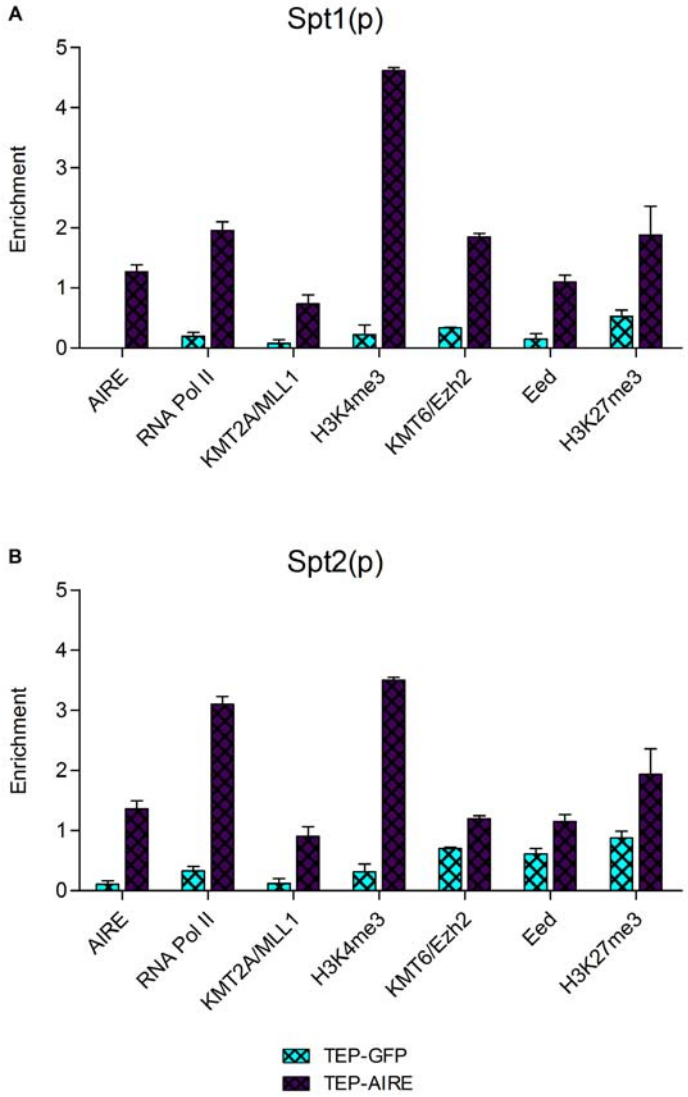
As a further control we wanted to analyse a gene silent in both TEP populations, for which recombination activating gene-1 (*Rag1*) was chosen. *Rag1*, in combination with *Rag2,* controls the assembly and rearrangement of antigen receptor genes during the early stages of B- and T-cell development in a process known as V(D)J recombination. They are highly lymphoid-specific and hence permanently silenced within thymic epithelial cells. We were unable to detect signal in either the bound or unbound material for *Rag1*, thus highlighting this gene's inactive status within these cells. The aberrant expression of this gene within thymic stromal cells may result in serious consequences, thus the absence of active histone modifications may reflect a more permanent silencing of *Rag1*, possibly suggesting a heterochromatic

compaction of this gene leading to its retention in the insoluble pellet following chromatin isolation.

3.3.4. Analysis of the Binding Status of AIRE, RNA Polymerase II and Histone Methyltransferases within AIRE-Regulated Gene Regions in the Thymic Epithelial Cell Model System

Given that AIRE has demonstrated putative associations with DNA, mediated through its SAND domain, and can bind un-modified H3K4 via its first PHD-zinc finger, we wanted to determine whether AIRE was able to interact directly with its target genes (Gibson *et al.* 1998; Kumar *et al.* 2001; Koh *et al.* 2008; Org *et al.* 2008; Chignola *et al.* 2009). AIRE has also been shown to associate with positive transcription elongation factor-b (P-TEFb) at the promoters of TRA, where Oven *et al* (2007) found RNA pol II already engaged, thus we also sought to establish the binding status of RNA pol II (Oven *et al.* 2007). Consequently, we performed conventional cross-linked chromatin immunoprecipitation (XChIP) with cultured TEP-AIRE and TEP-GFP lines. Cells were fixed with 1% paraformaldehyde, and chromatin was isolated and prepared through sonication, generating 300-600bp fragments. Input chromatin was then immunoprecipitated with antibodies directed against AIRE and RNA pol II. Relative amounts of these complexes across the promoter regions of the TRA *Spt1* and *Spt2* were determined by qPCR, through normalisation to the levels across *Gapdh*. This analysis revealed the presence of AIRE at the *Spt1* and *Spt2* promoter regions, which was absent or negligible in the TEP-GFP cells, suggesting that AIRE exerts its action through a direct association at each of its target genes (Figure 3.14 A and B). We also found increased levels of

Figure 3.14 – The Epigenetic Binding Status of Genes under the Transcriptional Control of the Autoimmune Regulator: Quantitation of the Relative Levels of AIRE, RNA Polymerase II and Chromatin-Modifying Enzymes by Cross-Linked Chromatin Immunoprecipitation



Quantitation by cross-linked chromatin immunoprecipitation (XChIP) of levels of AIRE, RNA polymerase II, KMT2A/MLL1, H3K4me3, KMT6/Ezh2, Eed and H3K27me3 across the promoter regions (p) of the tissue-restricted antigens salivary protein-1 (*Spt1*, *A*) and salivary protein-2 (*Spt2*, *B*) in cultured *Mus musculus* thymic epithelial (TEP) cell lines. Turquoise columns represent results from AIRE-negative TEP-GFP cells. Purple columns are from TEP cells transfected with AIRE. Immunoprecipitation performed with affinity-purified antibodies to AIRE, RNA polymerase II, KMT2A/MLL1, H3K4me3, KMT6/Ezh2, Eed and H3K27me3, as indicated. Relative levels (enrichment) of each protein/ modification were calculated from immunoprecipitated (bound) and input DNA by quantitative real-time PCR. Data was normalised to relative levels of enrichment at glyceraldehydes-3-phosphade dehydrogenase (*Gapdh*) as standard. Data are the mean ±SEM from at least two separate XChIP experiments.

RNA pol II in the presence of AIRE, with only low levels in the control cells. This seems to contrast with the findings of Oven *et al* (2007), who reported the presence of stalled RNA pol II at TRA in the absence of AIRE, however in a report by Org *et al* (2009), RNA pol II was only detectable on approximately 4% of AIRE target promoters in the absence of AIRE, thus our data clearly supports their findings (Org *et al.* 2009). In its role as a transcriptional regulator, AIRE may also recruit additional protein complexes to the local chromatin environment to facilitate the reading and transcribing of each locus. Our results thus far have indicated that AIRE is able to modulate certain histone modifications both on a gene-by-gene basis and on a greater scale. For example, we have shown increases in H3K4me3 and maintenance of H3K27me3 levels across many AIRE-regulated genes, modifications which are put in place by very well-defined histone methyltransferases and we therefore wanted to determine the binding status of these enzymes in the presence and absence of AIRE. We again employed XChIP to ascertain whether the H3K4 methylation-specific methyltransferase KMT2A/MLL1, and the two PcG proteins; KMT6/Ezh2, an H3K27 methylation-specific methyltransferase; and its co-factor Eed, were present at the promoters of TRA (Figure 3.14 A and B). We also verified our findings from unfixed NChIP analysis, using antibodies directed against H3K4me3 and H3K27me3 (Figure 3.14 A and B). This showed that in the control cells, levels of KMT2A/MLL1 are very low at AIRE-regulated TRA, along with its product H3K4me3, which confirmed our previous observations. In the presence of AIRE, higher levels of KMT2A/MLL1 were seen, however levels of H3K4me3 increased dramatically. This could suggest that AIRE engagement at the locus results in the recruitment of KMT2A/MLL1 for the deposition of H3K4me3. At the *Spt1* promoter (Figure 3.14 A),

moderate levels of KMT6/Ezh2 were detected in the TEP-GFP population, along with H3K27me3, with low amounts of Eed, however levels of these were higher in the control cells for *Spt2* (Figure 3.14 B). Interestingly, levels of all three increased in the presence of AIRE, which may account for the maintained or heightened levels of H3K27me3 revealed within the TEP-AIRE population through NChIP. Taken together, this implies that AIRE is able to bind chromatin at its individual target genes, leading to the recruitment of RNA pol II and KMT2A/MLL1 for deposition of positive-acting H3K4 methylation, but also of negative-acting PcG proteins KMT6/Ezh2 and Eed, which results in a maintenance or increase in H3K27me3.

In conclusion, it would appear that AIRE is able to interact directly with TRA and influence histone modifications not only on a gene-by-gene basis, as demonstrated by the distinct differences between individual TRA; *Csn1s1* (Figure 3.8) and the two salivary proteins (Figure 3.9), but also on a more large-scale, seen for the domain-wide hyperacetylation at H4K16 (Figure 3.12).

3.4. Elucidation of AIRE's Control of Promiscuous Gene Expression within the Thymus *in vivo*

Within the thymus, AIRE is expressed exclusively within stromal cells of the medulla; the site of PGE, generating TRA which promote the self-tolerance of developing thymocytes (Anderson *et al.* 2002). However, AIRE expression is only carried out by fully differentiated mTECs and recently a complex developmental network of thymic epithelial cells has been elucidated, revealing four main subpopulations of cells

(Rossi *et al.* 2006; Rossi *et al.* 2007). It was found that the two distinct medullary and cortical lineages of thymic stromal cells, are derived from a common bipotent TEC progenitor present in the embryonic murine thymus (Rossi *et al.* 2006; Rossi *et al.* 2007). Additionally, an immature $CD80^{-}AIRE^{-}$ mTEC progenitor population exists, which do not express AIRE but produce a cell surface receptor; RANK (Rossi *et al.* 2007; Akiyama *et al.* 2008; White *et al.* 2008; Zhu and Fu 2008). Upon stimulation from lymphoid tissue inducer cell-derived RANK ligand, these immature mTECs switch on AIRE and promiscuous gene expression begins, giving rise to mature $CD80^{+}AIRE^{+}$ mTECs (Rossi *et al.* 2007; Akiyama *et al.* 2008; White *et al.* 2008; Zhu and Fu 2008). Through foetal thymic organ culture and FACS sorting based upon specific cell surface markers (Table 3.3), the four cell subsets in the TEC developmental pathway (bipotent TEC progenitors, immature and mature mTECs and mature cTECs) can be isolated and pure populations obtained.

Although clearly successful for elucidation of AIRE's effects upon histone modifications associated with key TRA within a thymic epithelial cell background, NChIP is limited in its application as it requires a minimum of $1x10^{7}$ cells. This is easily achievable with cultured cell systems, however these are often not a true reflection of the *in vivo* situation. Typically the number of cells in each population of the TEC developmental pathway following FACS are very low; in the range of 10^{3}-10^{4} cells and for this reason investigation into histone modifications associated with the activation or silencing of TRA across the TEP developmental pathway has until now not been possible. It is therefore necessary to employ a technique recently developed in our laboratory termed carrier chromatin immunoprecipitation (CChIP)

Table 3.3 – The Fluorescent Activated Cell Sorting Parameters Used to Isolate Primary Cell Populations from *Mus musculus* Foetal Thymic Organ Cultures

***Mus musculus* EMBRYONIC THYMIC CELL POPULATIONS**		
POPULATION	FACS PARAMETERS	CELL NUMBERS PER IP
E12 Bipotent TEC Progenitors	$EpCAM1^+$	$3.4x10^4$
E15 Immature $CD80^-AIRE^-$ mTEC	$CD45^-$	$2.5x10^3$
	$EpCAM1^+$	
	$Ly51^-$	
	$CD80^-$	
E15 Mature $CD80^+AIRE^+$ mTEC	$CD45^-$	$2.5x10^3$
	$EpCAM1^+$	
	$Ly51^-$	
	$CD80^+$	
E15 Mature cTEC	$CD45^-$	$2.5x10^4$
	$EpCAM1^+$	
	$Ly51^+$	
E12 FoxN1-Deficient Nude TEC	$EpCAM1^+$	$5x10^3$

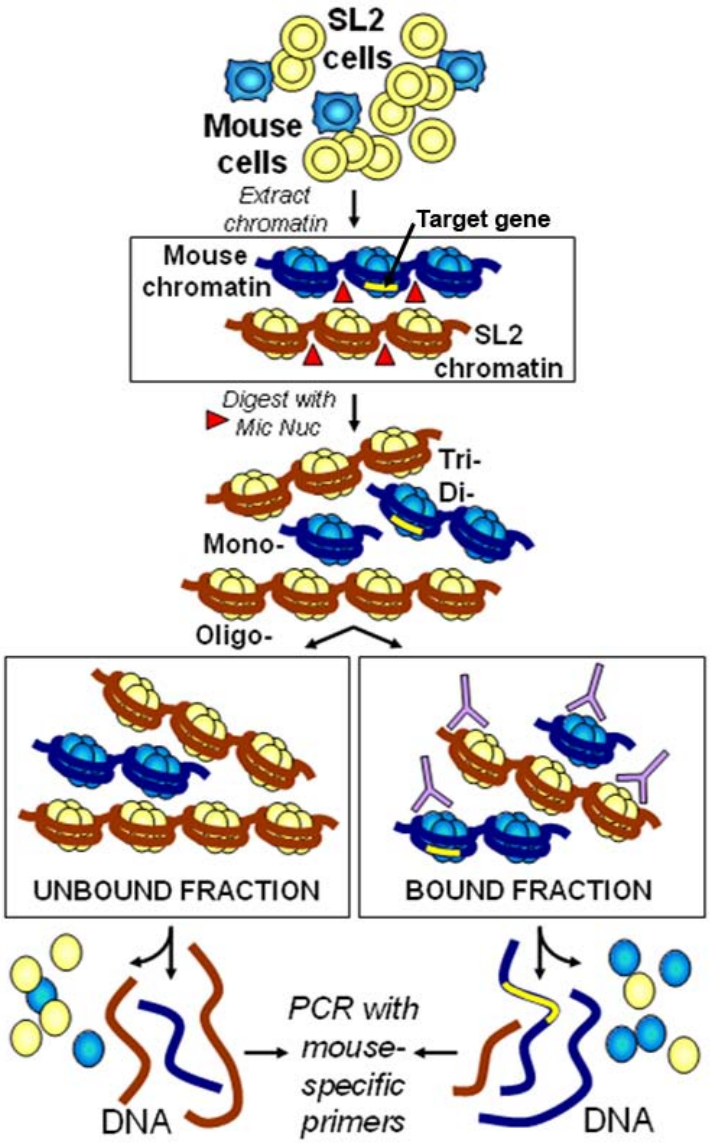
Foetal thymic organ culture was employed for the investigation of the epigenetic status of tissue-restricted antigens across the embryonic thymic epithelial developmental pathway. Fluorescent Activated Cell Sorting (FACS) was used to isolate four cell populations; $EpCAM1^+$ bipotent TEC progenitors, immature $CD45^-EpCAM1^+Ly51^-CD80^-$, and mature $CD45^-EpCAM1^+Ly51^-CD80^+$ mTECs, and mature $CD45^-EpCAM1^+Ly51^+$ cTECs, with the parameters displayed. Typical cell numbers following FACS, used subsequently for each individual immunoprecipitation (IP) by carrier chromatin IP, are also illustrated.

(O'Neill *et al.* 2006). The CChIP method is an adaptation of the more conventional NChIP procedure, drawing on the same basic principles and methodology, and published data shows a close correlation between CChIP and NChIP results (O'Neill *et al.* 2006). The CChIP technique involves mixing of the very low numbers of target *Mus musculus* cells with a 'carrier' *Drosophila melanogaster* cell line (SL2) and in doing so protects the target cells from the harsh isolation conditions required to obtain clean chromatin (Figure 3.15). Preparation of nuclei and chromatin is then performed as for NChIP, however due to the very low numbers of target cells, the volumes are kept low to minimise losses. Standard immunoprecipitation is then carried out but strict species-specific PCR is required for detection of target *Mus musculus* DNA. All primers were tested to ensure species specificity, as cross-reactivity with SL2 DNA would result in an over-representation of the signal (Figure 3.16).

3.4.1. Analysis of the Effects of AIRE upon Tissue-Restricted Antigens *in vivo*

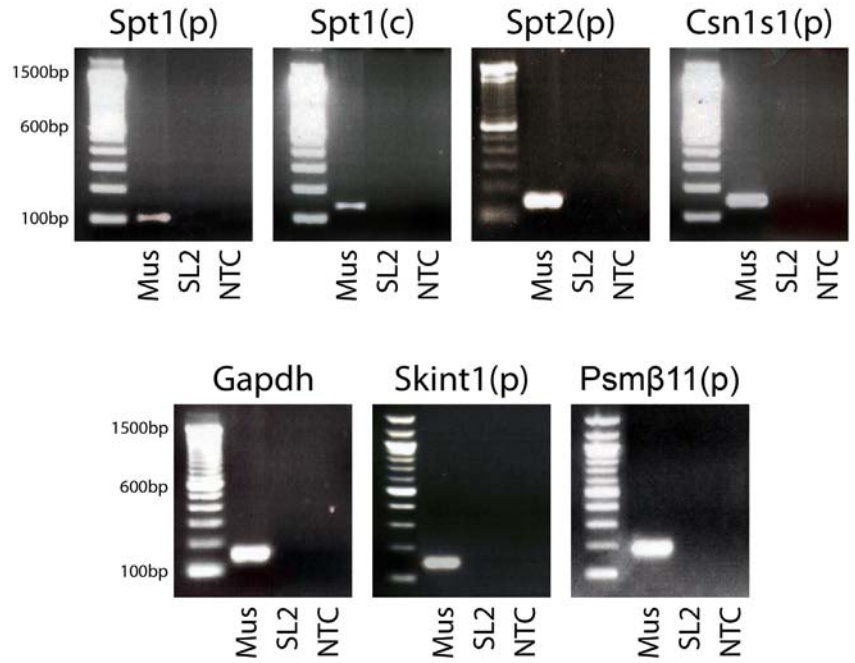
In order to elucidate the distribution of histone modifications throughout the TEC developmental pathway, purified bipotent TEC progenitors, immature and mature mTECs and mature cTECs were mixed separately with SL2 cells for CChIP analysis. Due to the limited number of target FACS sorted cells, a restricted number of histone modifications could be analysed. We therefore chose to investigate four modifications (H4K8ac, H3K4me3, H3K27me3 and H3K9me2) which, in the TEP model system, were shown to be heavily influenced by AIRE. The presence of the functionally opposing modifications H3K4me3 and H3K27me3 have been shown to co-localise across key developmental genes, priming them for future transcription

Figure 3.15 - Carrier Chromatin Immunoprecipitation



Diagrammatic representation of the carrier chromatin immunoprecipitation (CChIP) protocol. The CChIP technique involves mixing of small numbers of primary cell populations; *Mus musculus* cells from the thymic epithelial cell developmental pathway in this case, with *Drosophila melanogaster* cells (SL2) which act as a 'carrier', protecting the target cells throughout the procedure. Chromatin is isolated from the target cell/SL2 combination and digested with micrococcal nuclease (Mic Nuc) then immunoprecipitated with antibodies directed against post-translational histone modifications. DNA from antibody-bound and –unbound fractions is then purified for analysis of the relative levels of each modification, through strict species-specific PCR.

Figure 3.16 – Verification of the Species Specificity of Primer Sets Utilised for the Analysis of DNA from Carrier Chromatin Immunoprecipitation



For the verification of strict *Mus musculus* species specificity of primers, ensuring no cross-reactivity with *Drosophila melanogaster* (SL2) DNA, fixed point polymerase chain reaction (PCR) was carried out with *Mus musculus* (Mus) and *Drosophila melanogaster* (SL2) genomic DNA, along with d.H_2O no template controls (NTC), with each primer set. Primer sets were targeted to the promoter (p) and coding (c) region of salivary protein-1 (*Spt1*), and the promoter regions of salivary protein-2 (*Spt2*), casein-α (*Csn1s1*), glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*), selection and upkeep of intraepithelial T-cells 1 (*Skint1*) and proteasome subunit β-type 11 (*Psmβ11*), as indicated. Size was determined by reference to a 100 base pair molecular marker (Invitrogen).

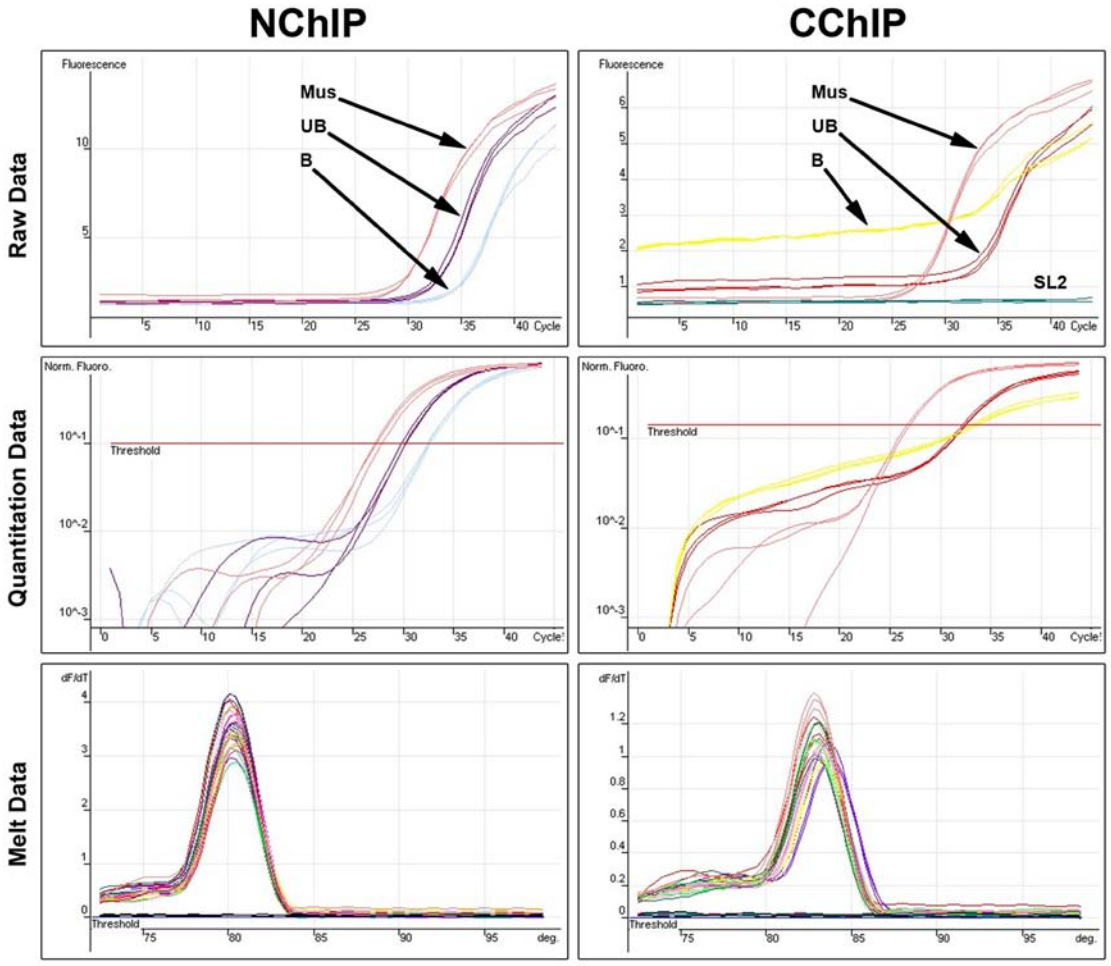
(Azuara *et al.* 2006; Bernstein *et al.* 2006). We wanted to monitor whether genes regulated by AIRE carried these bivalent modifications in the bipotent progenitor which changed upon activation of gene transcription as the cells differentiated. In addition, we investigated H3K9me2, whose presence within euchromatic genes has been linked to their silencing, although the reported association with active genes is intriguing, and from our analysis in the TEP model system, this mark clearly has a role to play (Vakoc *et al.* 2005; Squazzo *et al.* 2006; Vakoc *et al.* 2006; Gierman *et al.* 2007). Following CChIP, the concentration of DNA within antibody-bound and unbound fractions was analysed and the percentage pull-down of each immunoprecipitation calculated (Table 3.4). For the determination of levels of enrichment of histone modifications across TRA within the precipitated *Mus musculus* DNA, we initially attempted quantitative real-time PCR (qPCR), using the same analysis technique as for NChIP with the TEP cell line. Unfortunately, we were met with numerous technical difficulties. The samples presented with a background noise of SYBR Green signal (Figure 3.17) which may have been due to the large concentration of SL2 DNA within each sample, quenching the signal. Attempts to reduce the concentration of SL2 DNA through dilution of the samples did not alleviate the problem, making accurate analysis through qPCR impossible, especially considering the limited sample sizes (typically bound samples were resuspended in 40µl, unbound in 80µl to avoid over-dilution of the *Mus musculus* signal). Amplification curves were unreliable and we therefore turned to the analysis technique adopted in the original CChIP method (O'Neill *et al.* 2006). Radioactive PCR incorporating α-^{32}P dCTP into the reaction which, although more time consuming, reliably produced specific mouse products without quenching from the

Table 3.4 – Efficiency of Pull-Down for Each Histone Modification Following Carrier Chromatin Immunoprecipitation

CChIP PERCENTAGE PULL-DOWNS				
POPULATION	ANTIBODY	SPECIFICITY	PULL-DOWN	REPLICATES
E12 Bipotent TEC Progenitors	R232	H4K8ac	1.85-4.83%	2
	R612	H3K4me3	2.23-5.88%	2
	07-449	H3K27me3	7.36-7.74%	2
	07-212	H3K9me2	3.67-4.93%	2
	PI	Pre-immune	0.62-2.61%	2
E15 Immature $CD80^{-}AIRE^{-}$ mTEC	R232	H4K8ac	7.96-12.50%	2
	R612	H3K4me3	5.16-5.94%	2
	07-449	H3K27me3	6.49-28.17%	2
	07-212	H3K9me2	8.39-24.78%	2
	PI	Pre-immune	1.44-2.55%	2
E15 Mature $CD80^{+}AIRE^{+}$ mTEC	R232	H4K8ac	7.21-16.63%	2
	R612	H3K4me3	6.90-8.45%	2
	07-449	H3K27me3	10.80-14.88%	2
	07-212	H3K9me2	5.10-14.89%	2
	PI	Pre-immune	0.24-4.90%	2
E15 Mature cTEC	R232	H4K8ac	2.06-2.67%	2
	R612	H3K4me3	3.61-7.75%	2
	07-449	H3K27me3	4.50-5.66%	2
	07-212	H3K9me2	1.12-1.86%	2
	PI	Pre-immune	0.34-1.47%	2
E12 FoxN1-Deficient Nude TEC	R232	H4K8ac	4.10-5.75%	2
	R612	H3K4me3	5.91-20.70%	2
	07-449	H3K27me3	16.21-24.27%	2
	07-212	H3K9me2	7.67-11.99%	2
	PI	Pre-immune	2.03-2.04%	2

Carrier chromatin immunoprecipitation was performed with *Mus musculus* embryonic thymic cell populations, using antibodies directed against the post-translational histone modifications displayed above. DNA from antibody-bound (B) and –unbound (UB) fractions was purified and the percentage pull down (B/UBx100) for each antibody was then calculated. Pre-immune no antibody controls were included which show comparatively low pull-down efficiencies. The number of biological replicates for each modification is displayed.

Figure 3.17 – Comparison of Quantitative Real-Time Polymerase Chain Reaction with DNA from Native and Carrier Chromatin Immunoprecipitation



Quantitative real-time polymerase chain reaction (qPCR) on DNA isolated from antibody-bound (B) and –unbound (UB) fractions following native chromatin immunoprecipitation (NChIP) with cultured *Mus musculus* thymic epithelial (TEP) cell lines or carrier chromatin immunoprecipitation (CChIP) with *Drosophila melanogaster* and primary *Mus musculus* target cells. Immunoprecipitation with anti-H3K27me3 antibody and qPCR with primer set salivary protein-2 promoter are displayed as an example. *Mus musculus* (Mus) and *Drosophila melanogaster* (SL2) genomic DNA were included as controls. Raw data (fluorescence per cycle), quantitation data (log fluorescence per cycle) and melt data (dF/dT per degree) are illustrated to show efficiency and specificity of primers.

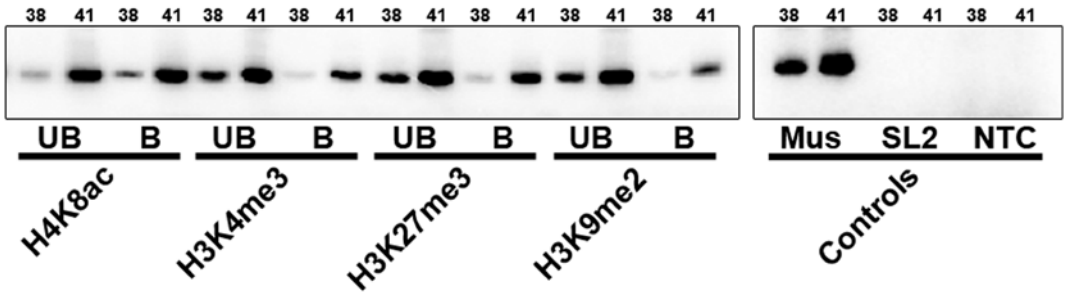
carrier SL2 DNA, and was therefore used in all subsequent analysis. Through polyacrylamide gel electrophoresis (PAGE) and phosphorimaging, the specificity of each amplification can be monitored routinely, and levels of bound and unbound signal be determined. Figure 3.18 shows a typical radioactive PAGE gel (using *Spt2*(p) primer set), highlighting the specificity of primers through the absence of signal in SL2 negative controls.

Examination of the three TRA; *Csn1s1*, *Spt1*, *Spt2* and genes of the keratin cluster in the TEP cell line, indicated that AIRE is able to induce alterations to the distribution of histone modifications, forming clear and distinct patterns, unique to each mark. Positively-regulated TRAs have a very well-defined expression profile throughout the TEC developmental pathway. In the bipotent TEC progenitor cells TRAs are silent and remain so in the immature mTECs, however, following stimulation by RANK ligand and subsequent AIRE up-regulation in mature mTECs, these genes are turned on. Differentiation into mature cTECs however, means the genes will never be expressed (Figure 3.19). To determine whether TRA are marked with specific epigenetic patterns throughout development, we chose to assay the four histone modifications at each stage of the pathway through CChIP.

3.4.1.1. Pattern of Histone Modifications for Salivary Protein-1 in vivo

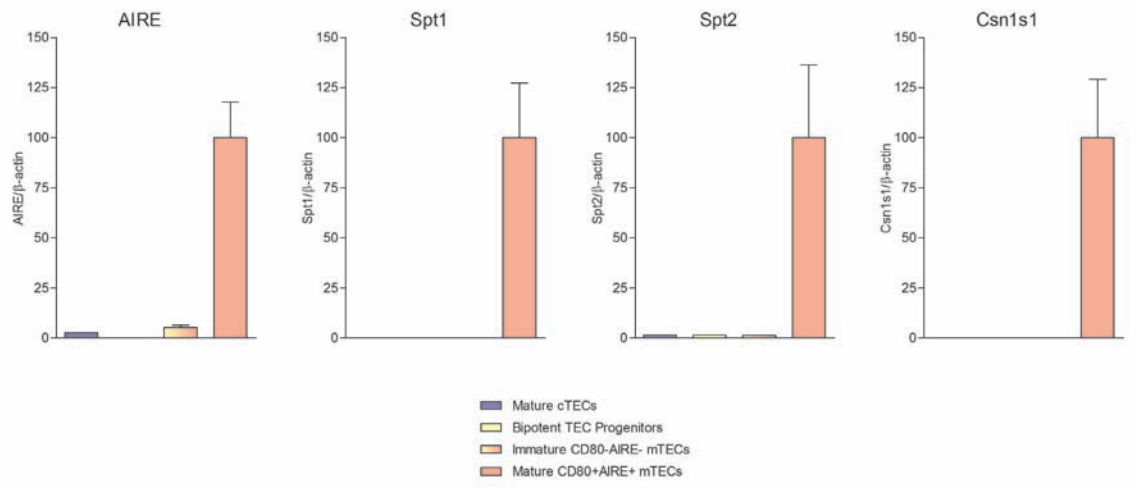
Bipotent TEC progenitors are an undifferentiated population of cells within which all TRA such as *Spt1* are silenced. Our observation of silencing modifications in this population at the *Spt1* promoter supports this with the presence of enriched H3K27me3 and H3K9me2 (Figure 3.20 B). However, we also detect equally high

Figure 3.18 – Analysis of DNA from Carrier Chromatin Immunoprecipitation: Radioactive Polymerase Chain Reaction



Radioactive PCR incorporating α-^{32}P dCTP was employed for quantitation of relative levels of histone modifications following carrier chromatin immunoprecipitation (CChIP). Immunoprecipitation was with affinity-purified antibodies to H4K8ac, H3K4me3, H3K27me3 and H3K9me2, as indicated. Radioactive PCR reactions were carried out with 2µl neat bound (B) and 2µl 1:2 diluted unbound (UB) DNA. Samples were taken after 38 and 41 cycles and visualised through 5% polyacrylamide gel electrophoresis and phosphorimaging. Relative levels of histone modifications (bound:unbound) were generated from intensity values for each PCR product and calculated from the percentage pull-down values of each immunoprecipitation. Specificity of each amplification was monitored through inclusion of *Mus musculus* (Mus) and *Drosophila melanogaster* (SL2) genomic DNA, and d.H_2O no template controls (NTC). Data displayed is from radioactive PCR with primer set salivary protein-2 promoter as an example.

Figure 3.19 - Analysis of AIRE's Role *in vivo*: Expression of the Autoimmune Regulator and Tissue-Restricted Antigens throughout the TEC Developmental Pathway

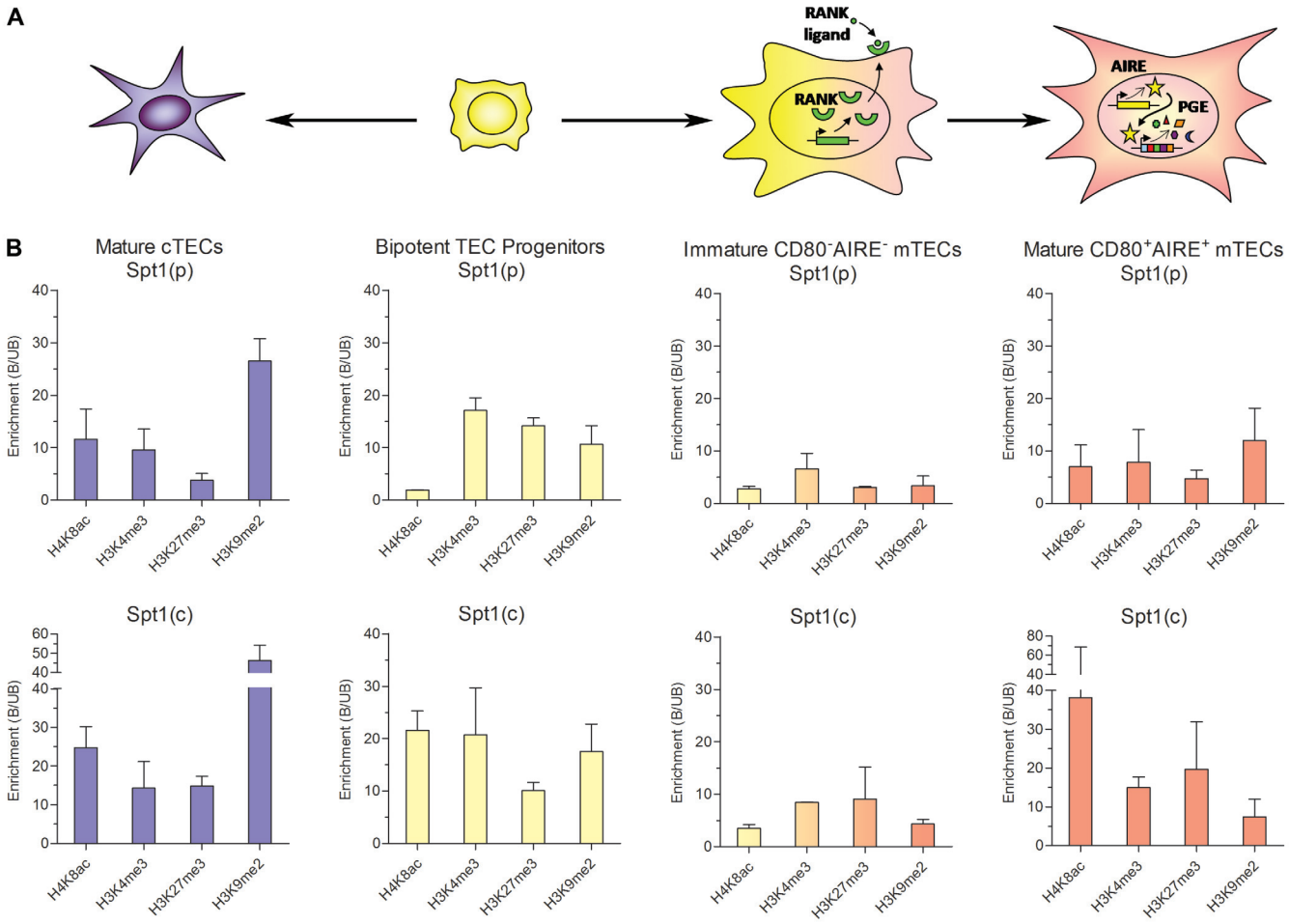


AIRE-regulated promiscuous gene expression of tissue-restricted antigens throughout the developmental pathway of thymic epithelial cells (TECs). Foetal thymic organ culture was employed, followed by FACS to isolate the four *Mus musculus* cell populations indicated, based on surface EpCAM1, CD45, Ly51 and CD80 expression. Real-time quantitative PCR was used to compare the relative mRNA expression levels of AIRE and three tissue-restricted antigens; casein-α (*Csn1s1*), salivary protein-1 (*Spt1*) and salivary protein-2 (*Spt2*), as indicated, for each cell type. Data was normalised to β-actin expression levels as standard. Data are the mean ±SEM from technical triplicate reactions, and are representative of at least two distinct cDNA preparations.

Figure 3.20 - Analysis of AIRE's Role *in vivo*: Quantitation of the Relative Levels of Histone Modifications across Salivary Protein-1 throughout the TEC Developmental Pathway

Quantitation by carrier chromatin immunoprecipitation (CChIP) of levels of histone modifications across salivary protein-1 (*Spt1*) promoter (p) and coding region (c), throughout the developmental pathway of thymic epithelial cells (TECs). *A*, Schematic overview of the four main stages of the TEC developmental pathway. *B*, Foetal thymic organ culture was employed, followed by FACS to isolate the four *Mus musculus* cell populations indicated, based on surface EpCAM1, CD45, Ly51 and CD80 expression. Immunoprecipitation performed with affinity-purified antibodies to H4K8ac, H3K4me3, H3K27me3 and H3K9me2, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA following radioactive PCR. Data are the mean ±SEM from at least two separate CChIP experiments.

Figure 3.20 - Analysis of AIRE's Role *in vivo*: Quantitation of the Relative Levels of Histone Modifications across Salivary Protein-1 throughout the TEC Developmental Pathway



levels of the active mark H3K4me3, which is interesting as the status of the *Spt1* gene within these progenitors is essentially in a poised state awaiting the correct signals to be either silenced or up-regulated depending on whether the bipotent population commit to mTECs where *Spt1* will be switched on, or cTECs in which *Spt1* expression will never occur. Hence, H3K4me3 within the progenitor cells may be acting as a predictive mark for future expression of this gene within the mTEC lineage, with H3K27me3 and H3K9me2 keeping *Spt1* silent in the progenitors. As development continues through to the immature mTEC stage, the cells move one step closer towards the imminent expression of *Spt1* and this saw a dramatic depression of all four histone modifications. Upon AIRE activation in the mature mTEC population, where *Spt1* becomes active, we saw an increase in the active marks H4K8ac and H3K4me3 at the promoter, compared to the immature mTECs, correlating with the active transcriptional status of this gene. Levels of H3K27me3 remained stable and an increase in the levels of H3K9me2 was also observed, their presence possibly representing a safety feature for this gene, allowing for tight regulation of its expression. This cell population is represented in our TEP model system by TEP-AIRE cells, which are permanently expressing AIRE. Comparisons with the results from the AIRE-positive cell line highlighted some differences, for example; in the cell line the *Spt1* promoter is marked predominantly by active histone modifications such as H3K4me3 and H4K8ac (Figure 3.9). Interestingly, of the two silent modifications investigated, higher levels of H3K27me3 were frequently present in the TEP line, whereas in the primary mature mTECs this was not the case for the *Spt1*(p). Analysis of the distribution of histone modifications in the reverse scenario whereby bipotent TEC progenitors differentiate to mature cTECs reveals a

contrasting re-organisation of chromatin to that seen in the mTEC lineage. Mature cTECs, which will never express AIRE or *Spt1*, show a decrease in H3K4me3 and H3K27me3 along with a small increase in H4K8ac, however, the most striking change is the dramatic increase in H3K9me2 from a bound:unbound ratio of 10.69 within the bipotent TEC progenitors, to a ratio of 26.50 within the mature cTECs. This possibly signifies the more permanent silencing of this gene within these cortical cells. The mature cTECs could be compared to TEP-GFP control cells, where again we observed contrasting results. While the cultured cell line displayed very high levels of H3K27me3, this is not observed *in vivo*. These differences between the two model systems may possibly point to a more strict H3K9me2-mediated regulation of the genes under the transcriptional control of AIRE *in vivo* as their expression within an incorrect cell type or at an inappropriate stage may be detrimental to the development of a functional, self-tolerant T-cell repertoire.

Analysis of an area within the coding region of *Spt1* shows a similar pattern to the promoter region of this gene, in particular for the distribution of H3K4me3 which follows the same trend throughout the four cell populations (Figure 3.20 B). However, comparisons between the promoter region and the gene body do show some differences. For example, in the bipotent TEC progenitor cells, there appears to be a switch in the silencing modifications, with the promoter showing higher H3K27me3 but in the coding regions H3K9me2 predominates. In addition, these progenitor cells display particularly high levels of H4K8ac in the gene body, which were not seen at the promoter. Thus, although the classical H3K4me3 / H3K27me3 bivalency was not observed in the *Spt1* coding region, it is clear that this locus does

display a poised chromatin signature with the presence of equally high levels of active and silent marks. Within the immature mTECs, again we saw a dramatic reduction in the histone modifications, similar to the promoter region, the only change being an approximate 3-fold increase in H3K27me3 from the 5` to the 3` region of *Spt1*. As the mTECs mature, we again found a re-organisation of the epigenetic marks with levels of H4K8ac and H3K27me3 increasing from the immature mTECs. In comparison to the *Spt1* promoter region in mature mTECs, the coding region displays very high H4K8ac, however there is a reversal of the silencing modifications, with H3K27me3 predominating in the exonic region. A significant finding for the coding region of *Spt1* is the almost identical distribution of modifications within the mature cTEC population when compared to the promoter region, with the low levels of H3K4me3 and H3K27me3, slightly higher enrichments of H4K8ac, but most importantly, a dramatic increase in H3K9me2; to a bound:unbound ratio of 26.5 at the promoter region and 46.2 in the coding region, thus confirming our observations.

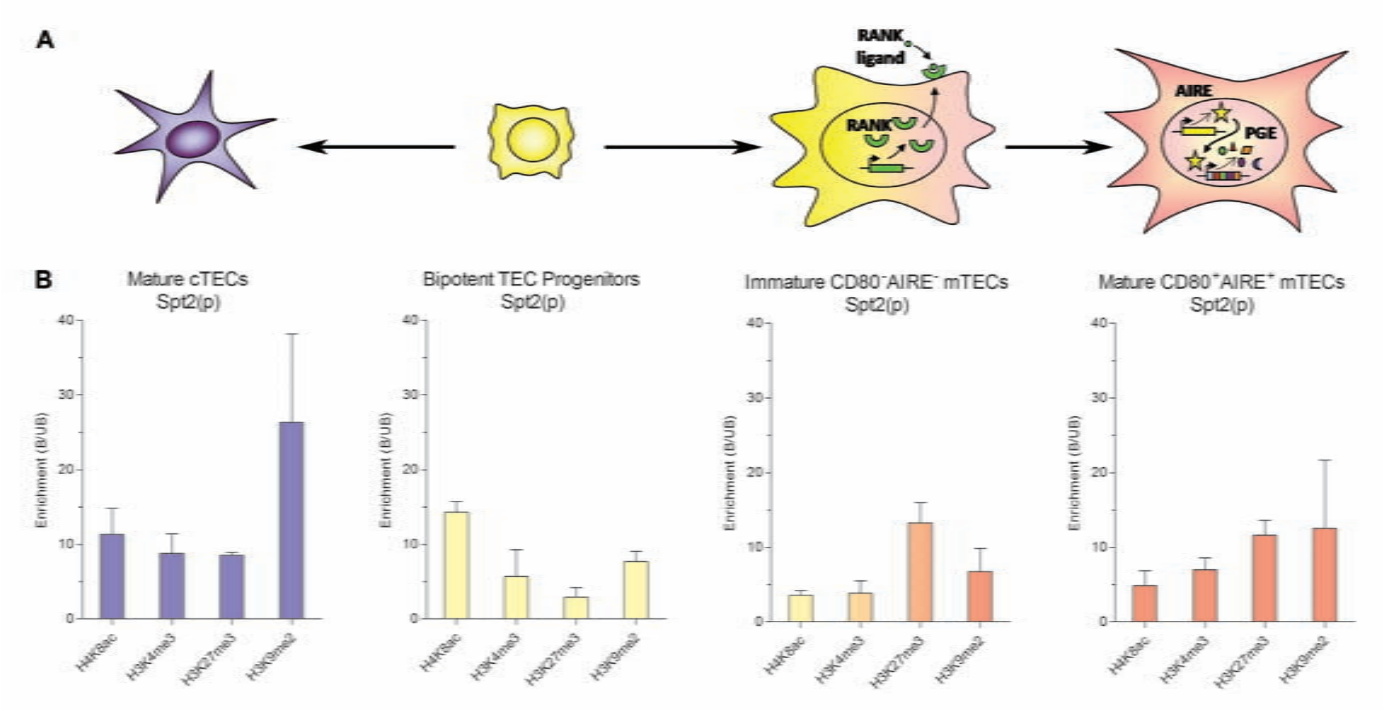
These results give a good indication of how AIRE-regulated genes are controlled throughout thymic development. TRA within the bipotent TEC progenitor cells are marked with a combination of active and silent modifications, keeping them in a poised but 'off' state. Chromatin is then reorganised as the cells differentiate, with a fall in all four marks in immature mTECs. As *Spt1* is up-regulated, elevated levels of the activating acetylation and methylation marks are induced, in combination with high levels of H3K9me2 and H3K27me3, potentially regulating expression. However as the bipotent progenitor cells differentiate to mature cTECs, a strong pattern of

modifications is displayed, with strikingly high levels of H3K9me2 possibly keeping *Spt1* more permanently silenced.

3.4.1.2. Pattern of Histone Modifications at the Salivary Protein-2 Promoter Region in vivo

Histone modifications across the promoter region of the second salivary protein show a distribution distinct from *Spt1* (Figure 3.21 B). Although the classical H3K4me3 / H3K27me3 bivalent chromatin structure is not observed at *Spt2*, the bipotent TEC progenitors do present with a bivalency of sorts; with predominance instead of acetylation at H4K8, which may be acting as an alternative predictive mark, in combination with H3K9me2, which may aid in the silencing of this gene within the progenitor population. Similar to findings for this subset at *Spt1*, we detected a general depression in all modifications as the cells develop into immature mTECs, with the exception of H3K27me3, levels of which rise from the bipotent TEC progenitors for *Spt2* thus shifting the balance of silencing modifications from H3K9me2 to H3K27me3. Maturation of the mTECs and subsequent AIRE-induced up-regulation of *Spt2*, brought about only a minor alteration to histone modifications from the immature mTECs; maintenance of H4K8ac, H3K4me3 and H3K27me3 and a 1.8-fold increase in H3K9me2, whereas for *Spt1* dramatic changes to the epigenetic markings were observed as maturity was reached, particularly in the coding regions. Most striking is the distribution of marks within the cTECs, where *Spt2* is silenced, results for which mirror those seen at both the promoter and coding regions of *Spt1*. The lowest levels being H3K4me3 and H3K27me3, minimal amounts of H4K8ac and a prominent enrichment of H3K9me2; increasing 3.4-fold

Figure 3.21 - Analysis of AIRE's Role *in vivo*: Quantitation of the Relative Levels of Histone Modifications across Salivary Protein-2 throughout the TEC Developmental Pathway



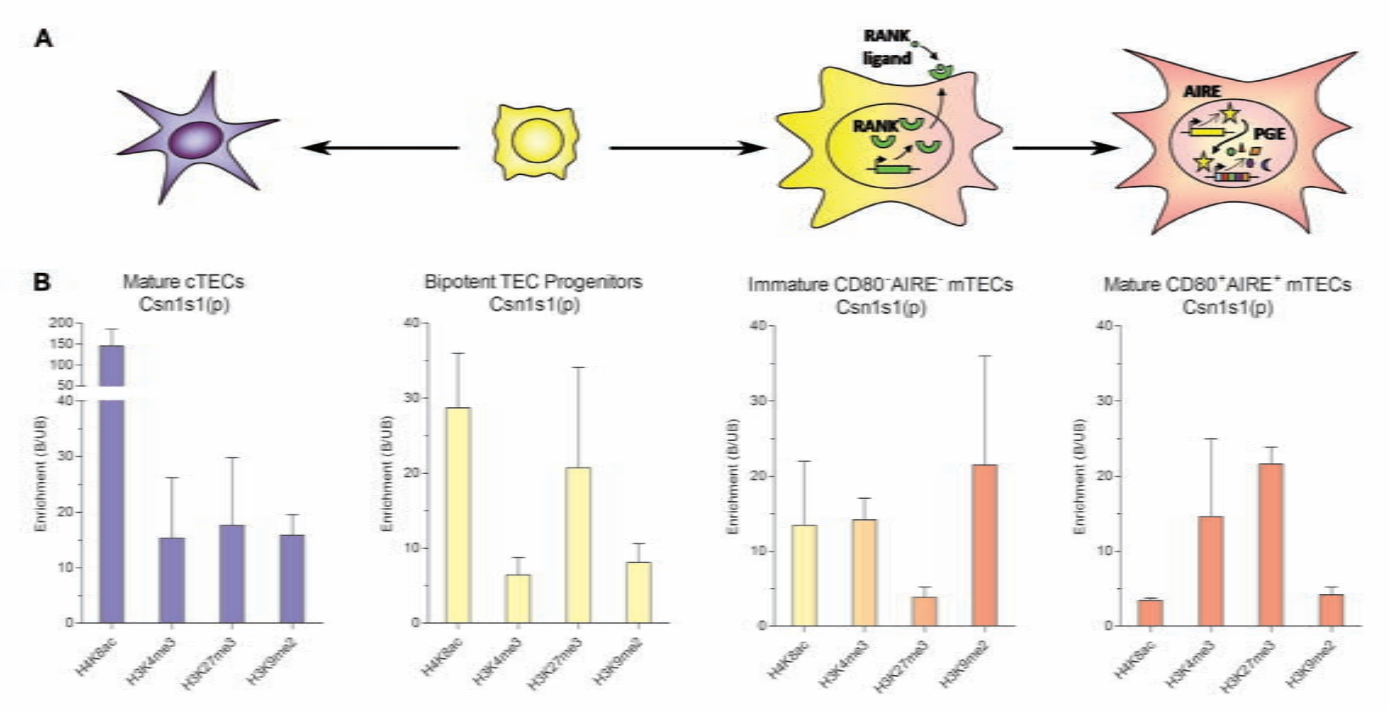
Quantitation by carrier chromatin immunoprecipitation (CChIP) of levels of histone modifications across salivary protein-2 (*Spt2*) promoter (p) region, throughout the developmental pathway of thymic epithelial cells (TECs). *A*, Schematic overview of the four main stages of the TEC developmental pathway. *B*, Foetal thymic organ culture was employed, followed by FACS to isolate the four *Mus musculus* cell populations indicated, based on surface EpCAM1, CD45, Ly51 and CD80 expression. Immunoprecipitation performed with affinity-purified antibodies to H4K8ac, H3K4me3, H3K27me3 and H3K9me2, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA following radioactive PCR. Data are the mean ±SEM from at least two separate CChIP experiments.

from the bipotent TEC progenitors at *Spt2*. This distribution of marks within cTECs across both salivary proteins is particularly intriguing, especially considering the variation observed between the bipotent TEC progenitors for these genes, and may suggest that this arrangement of chromatin is essential for their silencing within the cortical cells.

3.4.1.3. Pattern of Histone Modifications at the Casein-α Promoter Region in vivo

The epigenetic status of the promoter region of *Csn1s1* contrasts remarkably with the data for the two salivary proteins, which may be suggestive of alternative modes of regulation for these genes (Figure 3.22 B). The histone modifications assayed fluctuate dramatically as the cells progress through development, for example the levels of H4K8ac show an inverse correlation with gene expression. A clear pattern is seen throughout the developmental pathway, marking the progenitors with high levels, which fall as the cells mature to mTECs, but increase in the cTEC population, resulting in a cTEC > progenitors > immature mTECs > mature mTECs distribution. For *Csn1s1*, the bipotent TEC progenitors again displayed a mix of active and silent modifications, however the precise combination of modifications at this locus was unique, with enrichment of H3K27me3 and H4K8ac, and correspondingly low levels of H3K4me3 and H3K9me2, reflecting the poised nature of this gene, but a pattern not observed in the salivary proteins. Progression down the mTEC lineage to the immature mTECs again shows a unique pattern of histone marks. In addition to the decline in H4K8ac, we observed increased H3K9me2 and H3K4me3 in combination with a dramatic 5.3-fold decrease in H3K27me3 from the progenitor cells. Maturity

Figure 3.22 - Analysis of AIRE's Role *in vivo*: Quantitation of the Relative Levels of Histone Modifications across Casein-α throughout the TEC Developmental Pathway



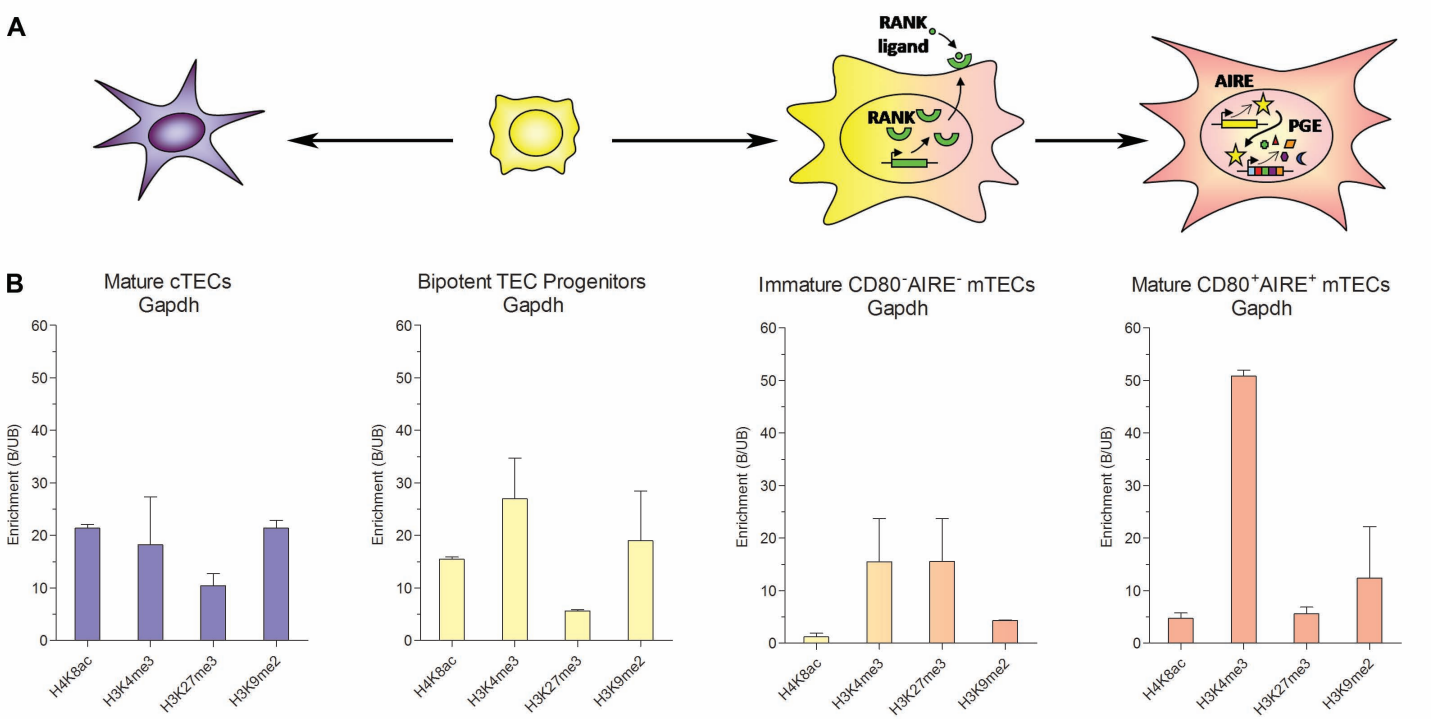
Quantitation by carrier chromatin immunoprecipitation (CChIP) of levels of histone modifications across casein-α (*Csn1s1*) promoter (p) region, throughout the developmental pathway of thymic epithelial cells (TECs). *A*, Schematic overview of the four main stages of the TEC developmental pathway. *B*, Foetal thymic organ culture was employed, followed by FACS to isolate the four *Mus musculus* cell populations indicated, based on surface EpCAM1, CD45, Ly51 and CD80 expression. Immunoprecipitation performed with affinity-purified antibodies to H4K8ac, H3K4me3, H3K27me3 and H3K9me2, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA following radioactive PCR. Data are the mean ±SEM from at least two separate CChIP experiments.

of the mTECs and subsequent up-regulation of *Csn1s1* again reveals a unique re-distribution of modifications, ultimately leading to an almost total loss of H3K9me2 and H4K8ac (bound:unbound ratio 4.22 and 3.45 respectively). Yet, high levels of H3K4me3 and H3K27me3 (bound:unbound ratio 14.63 and 21.63 respectively) are maintained. However, it is the development through to mature cTECs where the most interesting differences were observed, which have until now displayed an almost identical pattern across both salivary proteins. Whereas for the salivary proteins, the predominant mark was H3K9me2, for *Csn1s1* the cTEC population displays very high levels of H4K8ac, with correspondingly lower amounts of the three methylation marks. The dramatic and characteristic peak in H3K9me2 was not seen, however there was a 2.0-fold increase of this mark from the progenitor cells. Thus it would seem that the distribution of modifications across the TEC developmental pathway differs dramatically on a gene-by-gene basis.

3.4.1.4. Pattern of Histone Modifications at the Glyceraldehyde-3-phosphate Dehydrogenase Locus in vivo

As a control, we chose to assay the housekeeping gene *Gapdh* which is ubiquitously expressed throughout the TEC developmental pathway. The profile of histone modifications across each of the four stages of thymic development, does reflect this (Figure 3.23 B), a common theme being the dominance of the active histone modifications H4K8ac and H3K4me3. However, some variations between cell populations are observed. While the bipotent progenitors and mature cTECs are marked by high H4K8ac and H3K9me2, mTECs display low levels of these modifications. Thus, although these results confirm the active status of this gene

Figure 3.23 - Analysis of AIRE's Role *in vivo*: Quantitation of the Relative Levels of Histone Modifications across Glyceraldehyde-3-phosphate Dehydrogenase throughout the TEC Developmental Pathway



Quantitation by carrier chromatin immunoprecipitation (CChIP) of levels of histone modifications across glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*), throughout the developmental pathway of thymic epithelial cells (TECs). *A*, Schematic overview of the four main stages of the TEC developmental pathway. *B*, Foetal thymic organ culture was employed, followed by FACS to isolate the four *Mus musculus* cell populations indicated, based on surface EpCAM1, CD45, Ly51 and CD80 expression. Immunoprecipitation performed with affinity-purified antibodies to H4K8ac, H3K4me3, H3K27me3 and H3K9me2, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA following radioactive PCR. Data are the mean ±SEM from at least two separate CChIP experiments.

across the TEC developmental pathway, they also highlight the complexities of epigenetic modifications, with no single defined combination of marks signifying the activity of a given gene.

3.4.1.5. Pattern of Histone Modifications at the Selection and Upkeep of Intraepithelial T-cells 1 Promoter Region in vivo

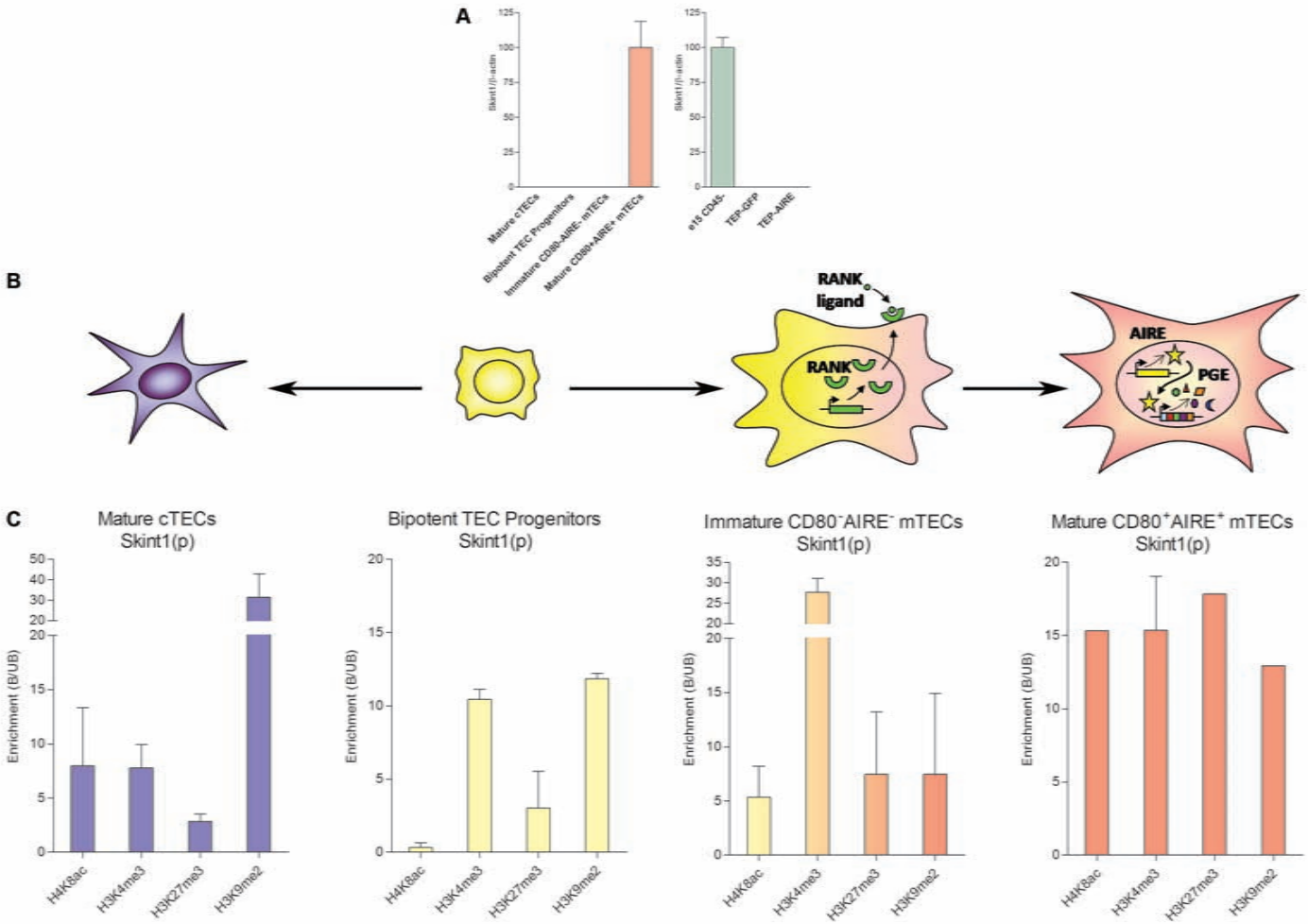
As a further control we wanted to analyse the distribution of histone modifications across a gene not under the transcriptional control of AIRE, whose expression pattern correlates with TRAs. Selection and upkeep of intraepithelial T-cells 1 (*Skint1*) is a newly identified gene whose expression within the thymus has been shown to be involved in the previously undefined selection of γδ T-cells, in particular, those which express the Vγ5Vδ1 T-cell receptor (Boyden *et al.* 2008). Expression analysis of *Skint1* (Figure 3.24 A), demonstrated the presence of this gene within the mature mTEC population, but not within the bipotent TEC progenitors, the immature mTECs or the mature cTEC lineage. Despite this pattern of regulation akin to an AIRE-regulated gene, *Skint1* is not under the transcriptional control of AIRE, as its expression is undetected in both the presence and absence of AIRE in the TEP cell line (Figure 3.24 A).

Bipotent TEC progenitors are marked with an enrichment of H3K4me3, which is matched by equal levels of H3K9me2 indicating the poised nature of *Skint1* (Figure 3.24 C). Increases in both active marks H4K8ac and H3K4me3 are observed in the immature mTEC population, along with maintenance of H3K9me2 and a 2.5–fold increase in H3K27me3. For the mature mTECs, where *Skint1* is switched on, we

Figure 3.24 - Analysis of AIRE's Role *in vivo*: Quantitation of Expression Levels and the Relative Levels of Histone Modifications across Selection and Upkeep of Intraepithelial T-Cells 1 throughout the TEC Developmental Pathway

Expression and relative levels of histone modifications associated with the selection and upkeep of intraepithelial T-cells 1 (*Skint1*) promoter region (p), throughout the developmental pathway of thymic epithelial cells (TECs). *A*, Foetal thymic organ culture (FTOC) was employed, followed by FACS to isolate the four *Mus musculus* cell populations indicated, based on surface EpCAM1, CD45, Ly51 and CD80 expression. Real-time quantitative PCR was used to compare the relative mRNA expression levels of *Skint1* for each cell type. Expression analysis was also carried out with mRNA from cultured AIRE-negative and AIRE-positive thymic epithelial (TEP) cell lines, alongside whole FTOC mRNA for comparison. Data was normalised to β-actin expression levels as standard. Data are the mean ±SEM from technical triplicate reactions, and are representative of at least two distinct cDNA preparations. *B*, Schematic overview of the four main stages of the TEC developmental pathway. *C*, Carrier chromatin immunoprecipitation was performed with affinity-purified antibodies to H4K8ac, H3K4me3, H3K27me3 and H3K9me2, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA following radioactive PCR. Data are the mean ±SEM from at least two separate CChIP experiments.

Figure 3.24 - Analysis of AIRE's Role *in vivo*: Quantitation of Expression Levels and the Relative Levels of Histone Modifications across Selection and Upkeep of Intraepithelial T-Cells 1 throughout the TEC Developmental Pathway



observed a domain-wide blanket of all marks, with comparatively high levels of active and repressive modifications. The cTEC lineage, where *Skint1* is not expressed, displays a significant enrichment of H3K9me2; at least 3.9 times greater than the other three marks, resulting in a pattern equivalent to those seen for the cTECs across the AIRE-regulated genes. This reflects the more permanent silent nature of this gene within the cTEC population. Comparisons between the findings for the *Skint1* promoter region and TRA, showed a divergent pattern of histone modifications at this non-AIRE-regulated gene and this may reflect the fact that this gene is mTEC-specific, unlike the TRA whose native expression is within peripheral tissues outside of the thymus.

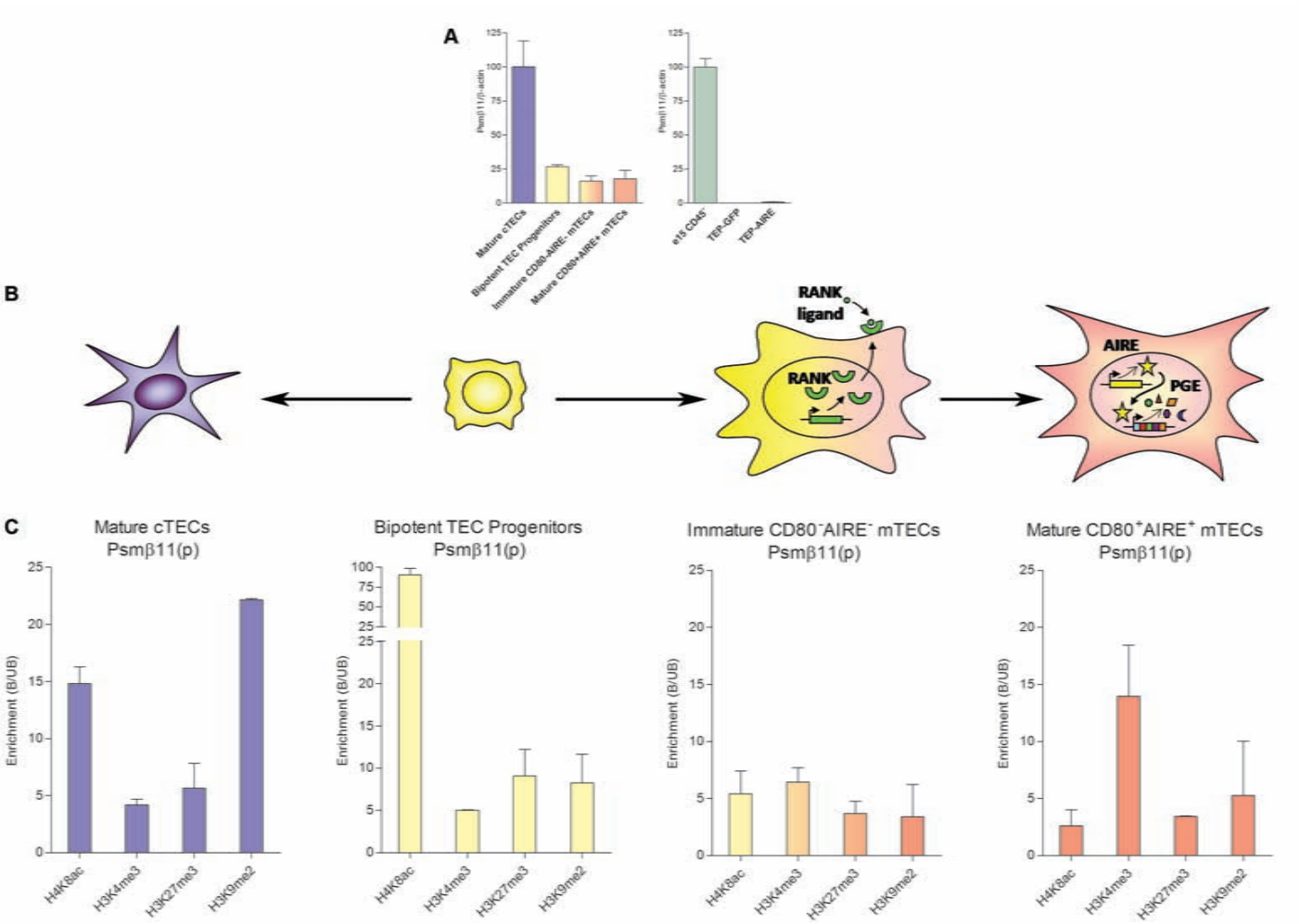
3.4.1.6. Pattern of Histone Modifications at the Proteasome Subunit β-Type 11 Promoter Region in vivo

For a final control, we wanted to analyse the epigenetic status of a gene whose expression is limited to the cTEC lineage, and therefore not influenced by AIRE in any way. The discovery of a previously unrecognised catalytic subunit of the proteasome, a structure responsible for cleaving polypeptides into smaller fragments for presentation by the class I MHC, provided us with a useful gene for this analysis. Proteasome subunit β-type 11 (*Psmβ11*) was found to be essential for the positive selection of MHC class I-restricted $CD8^+$ T-cell repertoire, and is therefore exclusively expressed by cortical thymic epithelial cells (Murata *et al.* 2007). This was further confirmed when expression analysis was carried out (Figure 3.25 A) showing high levels of expression in cTECs and low or no expression in the immature and mature mTECs and progenitor cells. *Psmβ11* was also shown to be

Figure 3.25 - Analysis of AIRE's Role *in vivo*: Quantitation of Expression Levels and the Relative Levels of Histone Modifications across Proteasome Subunit β-Type 11 throughout the TEC Developmental Pathway

Expression and relative levels of histone modifications associated with the proteasome subunit β-type 11 (*Psmβ11*) promoter region (p), throughout the developmental pathway of thymic epithelial cells (TECs). *A*, Foetal thymic organ culture (FTOC) was employed, followed by FACS to isolate the four *Mus musculus* cell populations indicated, based on surface EpCAM1, CD45, Ly51 and CD80 expression. Real-time quantitative PCR was used to compare the relative mRNA expression levels of *Psmβ11* for each cell type. Expression analysis was also carried out with mRNA from cultured AIRE-negative and AIRE-positive thymic epithelial (TEP) cell lines, alongside whole FTOC mRNA for comparison. Data was normalised to β-actin expression levels as standard. Data are the mean ±SEM from technical triplicate reactions, and are representative of at least two distinct cDNA preparations. *B*, Schematic overview of the four main stages of the TEC developmental pathway. *C*, Carrier chromatin immunoprecipitation was performed with affinity-purified antibodies to H4K8ac, H3K4me3, H3K27me3 and H3K9me2, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA following radioactive PCR. Data are the mean ±SEM from at least two separate CChIP experiments.

Figure 3.25 - Analysis of AIRE's Role *in vivo*: Quantitation of Expression Levels and the Relative Levels of Histone Modifications across Proteasome Subunit β-Type 11 throughout the TEC Developmental Pathway



unaffected by AIRE as its expression within the TEP model system is not detected in both cell populations (Figure 3.25 A). What was instantly evident following CChIP analysis for *Psmβ11* was the significant enrichment of acetylation in the bipotent TEC progenitors, coupled with lower levels of methylation. This acetylation mark is not however balanced by the presence of a silencing mark, yet it could signify a mark within the bipotent cells for future expression of *Psmβ11* in cTECs (Figure 3.25 C).

For the immature mTECs, there was little change to the three methyl marks, yet a dramatic 16.7-fold fall in H4K8ac at the *Psmβ11* promoter resulting in a distribution of marks akin to those seen for *Spt1*, which is surprising as this gene is silent within the mTEC lineage. Continuation to fully mature mTECs again reveals unexpected results for *Psmβ11*, with a prevalence of H3K4me3, a mark usually associated with active gene transcription, resulting in a distribution of marks similar to that seen in this cell population for the ubiquitously-expressed *Gapdh*. Finally, for mature cTECs, where *Psmβ11* is expressed exclusively, a high level of H4K8ac was observed which, although lower than the level of this mark in the progenitor population, represents one of the highest of the modifications in the cTECs, matched only by H3K9me2. This predominance of H4K8ac is not observed in the mTECs and may therefore be responsible for the active status of this gene in cTECs. Interestingly, there was also an enrichment of H3K9me2, as observed for the genes which are silent within mature cTECs, however, for the TRA, with the exception of *Csn1s1*, the level of H3K9me2 is significantly higher than all other marks, whereas

for *Psmβ11*, this is not the case due to the elevated amounts of the active acetylation mark.

3.5. Examination of the Distribution of Histone Modifications across Tissue-Restricted Antigens within FoxN1-Deficient Nude Thymic Epithelial Cells

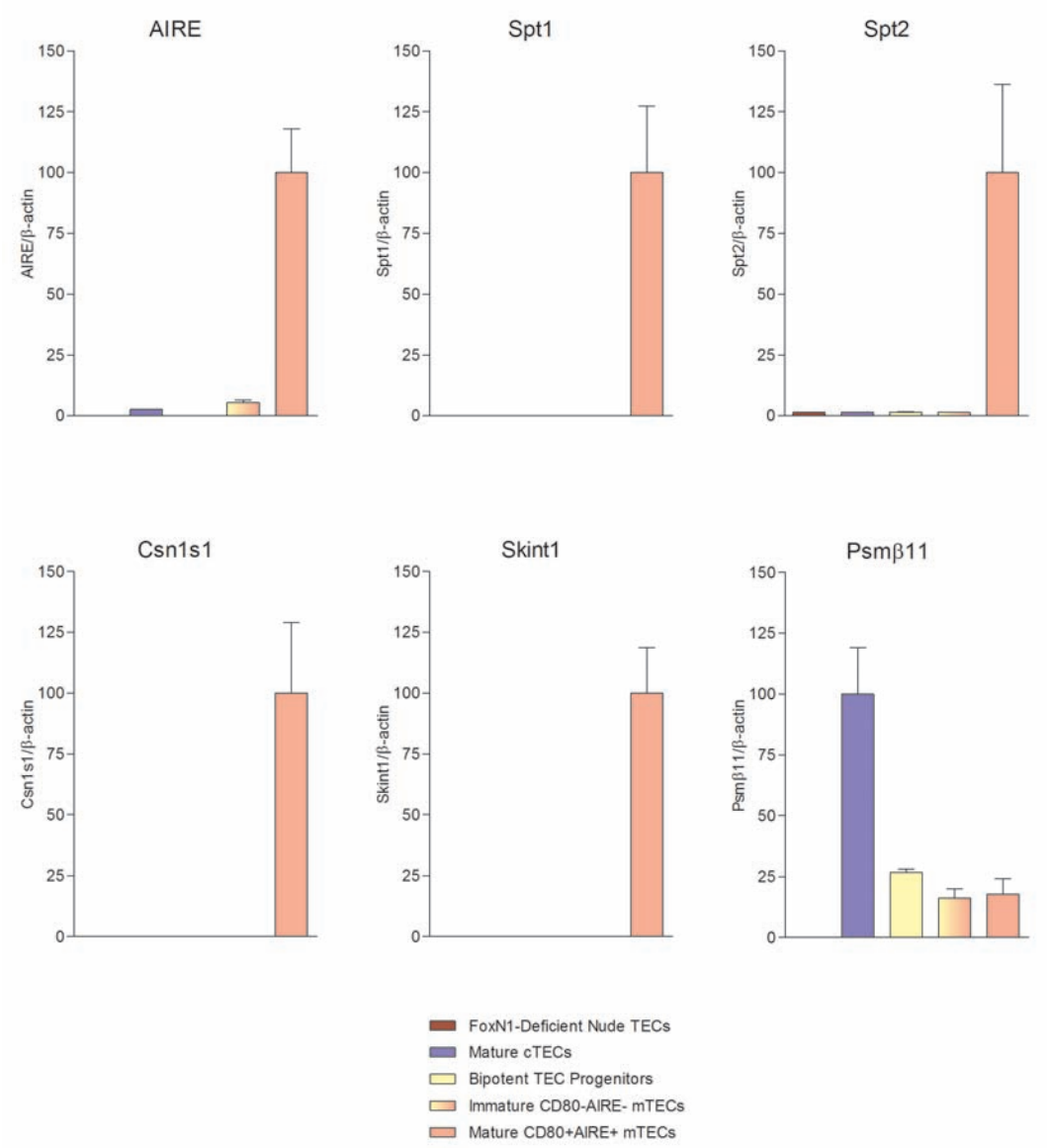
To more thoroughly assess the role histone modifications play in PGE, we used CChIP to analyse the epigenetic status of TRA when thymus embryogenesis is disrupted. For this we turned to the athymic nude mouse model which is deficient in the forkhead transcription factor FoxN1. FoxN1 is required in a cell-autonomous fashion for initial TEC differentiation (Jenkinson *et al.* 2008; Chen *et al.* 2009). These mice present with a non-functional cystic thymic rudiment due to the block in normal TEC development at an early stage (Blackburn *et al.* 1996; Anderson and Jenkinson 2001; Jenkinson *et al.* 2008). Interestingly however, FoxN1 does not appear to be required for the preliminary stages of thymic organogenesis, in particular the initial colonisation and growth of bipotent thymic epithelial progenitors (Anderson and Jenkinson 2001; Jenkinson *et al.* 2008). It is the subsequent differentiation of these progenitor cells into cortical or medullary lineages that is stalled due to the lack of FoxN1 (Anderson *et al.* 2007; Jenkinson *et al.* 2008). We therefore sought to analyse the pattern of histone modifications across *Spt1*, *Spt2* and *Csn1s1* within these TEC progenitors from foetal FoxN1-deficient BALB/c nude (Nu/Nu) mice. The development of these FoxN1-deficient nude TEC is blocked and hence expression of these genes will not occur. Comparisons back to the results already obtained in bipotent TEC progenitor cells from normal mice, would highlight whether the nude

TEC progenitor equivalents lose the predictive pattern of histone modifications. We also chose to examine the housekeeping gene *Gapdh*, whose expression is ubiquitous in the nude cell populations, the non-AIRE-regulated gene *Skint1* whose expression pattern correlates with TRA, and the cTEC-specific *Psmβ11*. The silent nature of *Spt1*, *Spt2*, *Csn1s1*, *Skint1* and *Psmβ11* within the FoxN1-deficient nude TEC was first confirmed through expression analysis from FACS sorted nude cells (Figure 3.26). Again, the cell yields following FACS of FoxN1-deficient nude TEC are very low and therefore in order to assay histone modifications in this rare population of cells, CChIP was required.

3.5.1. Pattern of Histone Modifications for Salivary Protein-1, Salivary Protein-2 and Casein-α within FoxN1-Deficient Nude Thymic Epithelial Cells

To allow for comparisons with the distribution of modifications in the equivalent wild-type bipotent TEC progenitors and also mature cTECs, where these genes are silenced, FoxN1-deficient nude TEC results are displayed alongside those from the wild-type cells in Figures 3.27 and 3.28. FoxN1-deficient nude TEC, which lack the capacity to express TRA, are marked by divergent histone modifications that differ greatly from the equivalent wild-type cells. Also evident are the distinct variations between the gene regions analysed, with each gene responding in an individual way to the change in cell fate. However, some commonalities do present themselves such as the predominance of H4K8ac within the nude cells and the correspondingly lower levels of H3K4me3 and H3K27me3.

Figure 3.26 – The Epigenetic Patterning of FoxN1-Deficient Nude Thymic Epithelial Cells: Expression of the Autoimmune Regulator and Tissue-Restricted Antigens upon Disruption of Normal TEC Development



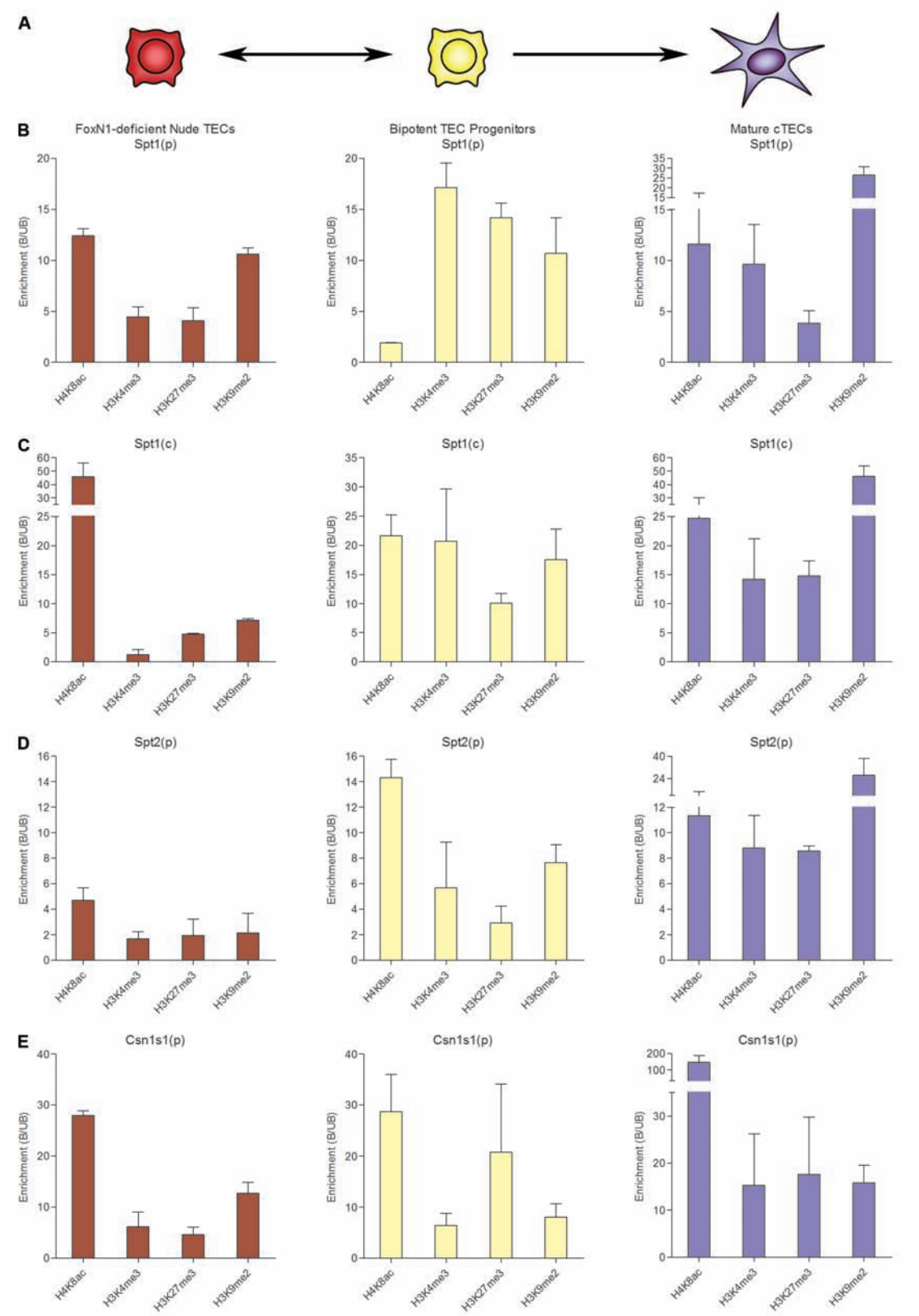
Expression profile of AIRE-regulated tissue-restricted antigens and control genes in athymic FoxN1-deficient nude mice. Foetal thymic organ culture was employed, followed by FACS to isolate *Mus musculus* EpCAM1^{+} FoxN1-deficient TEC and real-time quantitative PCR was used to compare the relative mRNA expression levels of AIRE and three tissue-restricted antigens; casein-α (*Csn1s1*); salivary protein-1 (*Spt1*); and salivary protein-2 (*Spt2*), in addition to the mTEC-specific selection and upkeep of intraepithelial T-cells 1 (*Skint1*) and the cTEC-specific proteasome subunit β-type 11 (*Psmβ11*) as indicated. For comparison, expression levels of these genes within normal bipotent TEC progenitors, mTECs and cTECs are displayed. Data was normalised to β-actin expression levels as standard. Data are the mean ±SEM from technical triplicate reactions, and are representative of at least two distinct cDNA preparations.

Comparisons between the bipotent TEC progenitors and the equivalent cells within the nude mice across *Spt1*, revealed an exciting shift in the histone modifications. Where the wild-type bipotent TEC progenitors showed a clear bivalency at the promoter region of *Spt1* (Figure 3.27 B), with equally high levels of H3K4me3 and H3K27me3, enrichment of these modifications was much lower in the nude population. Instead, these cells displayed a predominance of H4K8ac and the silencing modification H3K9me2. For the *Spt1* coding region (Figure 3.27 C), a different pattern of histone marks was seen for the nude mouse cells, with a very high level of acetylation at H4K8; at an intensity approximately 2-fold greater than the normal progenitors, along with an almost total loss of H3K4me3 and low levels of H3K27me3 and H3K9me2. However, what was common for both the promoter and coding region of *Spt1* was the subordinance of the bivalent marks in the nude cells. These alterations to the histone modifications across *Spt1*, a gene which could potentially be up-regulated in the wild-type cells, but whose expression will never occur within the nude mice, were especially exciting as they were more akin to those seen in wild-type mature cTECs, where *Spt1* is also off (displayed in Figure 3.27 as indicated). This trend is possibly indicative of the fact that *Spt1* is more permanently silenced in the nude mice due to their halted differentiation. However, one noticeable discrepancy is that the significant enrichment of H3K9me2 seen for the mature cTECs is not observed to the same extent in the nude mice, and this may be a result of the stunted developmental programme of these cells. FoxN1-deficient nude TEC are fundamentally representative of the undifferentiated bipotent TEC progenitors, however, they are effectively ‘in limbo’ as no further differentiation will

Figure 3.27 – The Epigenetic Patterning of FoxN1-Deficient Nude Thymic Epithelial Cells: Quantitation of the Relative Levels of Histone Modifications across Tissue-Restricted Antigens upon Disruption of Normal TEC Development

Quantitation by carrier chromatin immunoprecipitation (CChIP) of levels of histone modifications across the tissue-restricted antigens; salivary protein-1 (*Spt1*) promoter (p) and coding region (c); salivary protein-2 (*Spt2*) promoter; and casein-α (*Csn1s1*) promoter in athymic FoxN1-deficient nude mice. *A*, Schematic overview of FoxN1-deficient thymic epithelial cells (TEC). *B*, Foetal thymic organ culture was employed, followed by FACS to isolate *Mus musculus* $EpCAM1^{+}$ FoxN1-deficient TEC and CChIP was carried out with affinity-purified antibodies to H4K8ac, H3K4me3, H3K27me3 and H3K9me2, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA following radioactive PCR. As a comparison, data from normal bipotent TEC progenitors and mature cTECs are displayed. Data are the mean ±SEM from at least two separate CChIP experiments.

Figure 3.27 – The Epigenetic Patterning of FoxN1-Deficient Nude Thymic Epithelial Cells: Quantitation of the Relative Levels of Histone Modifications across Tissue-Restricted Antigens upon Disruption of Normal TEC Development



occur. These cells may therefore retain the acetylation mark, keeping the chromatin open at these gene regions, should the developmental pathway continue.

For the promoter region of *Spt2* (Figure 3.27 D), the nude bipotent TEC progenitor cells again show an altered distribution of marks to the wild-type progenitor cells. In comparison to the normal progenitors, nude cells displayed a reduction of the active histone modification H3K4me3, an approximate 3–fold decrease in H4K8ac and H3K9me2, but maintenance of the level of H3K27me3. Thus when the TEC developmental pathway is disrupted, a locus-wide depression of acetylation and methylation was observed, which, unlike for *Spt1*, does not reflect the epigenetic status of mature cTECs.

The pattern of histone modifications for *Csn1s1* (Figure 3.27 E) in the FoxN1-deficient population, also varied from that seen in the wild-type progenitors. Despite no change to the overall amounts of the active marks H4K8ac and H3K4me3, the level of H3K27me3 was 4.6-fold lower in the nude cells, coupled with a small increase in the level of H3K9me2. This resulted in a distribution of histone modifications similar to those seen across *Spt1* in the nude TEC population (Figure 3.27 B and C), and also within the mature cTECs for other TRA. This pattern was not however, observed for the mature cTECs at the *Csn1s1* promoter region, which in fact more closely resembles the pattern at the coding region of *Spt1* (Figure 3.27 C) for the nude mice cells, with high H4K8ac and low methylation.

The observed differences between the histone modifications across wild-type bipotent and FoxN1-deficient TEC progenitors for three genes under the transcriptional control of AIRE, shows how significant these marks are as indicators of future transcription. With the loss of FoxN1, the TEC developmental pathway is blocked prior to the critical lineage choice of the progenitor cells into either $AIRE^-$ cortical or $AIRE^+$ medullary TECs and thus the programme of transcription of these TRA is also halted. These genes will never be expressed, as the nude cells cannot progress through to mature mTECs. From our results it would appear that the future of these genes is reflected in the chromatin, with the alterations to the histone marks within the nude mice often showing similarities to mature cTECs where the genes are most definitely off. Yet despite this, nude cells do not mirror cTECs entirely, showing instead consistently high H4K8ac levels. This retention of acetylation is interesting and possibly suggests a 'memory' of the normal programme of development, marking these genes for potential expression should normal differentiation resume, even though it never will. The genes are therefore not shut down totally, as in the mature cTECs, hence the differences observed between the two populations.

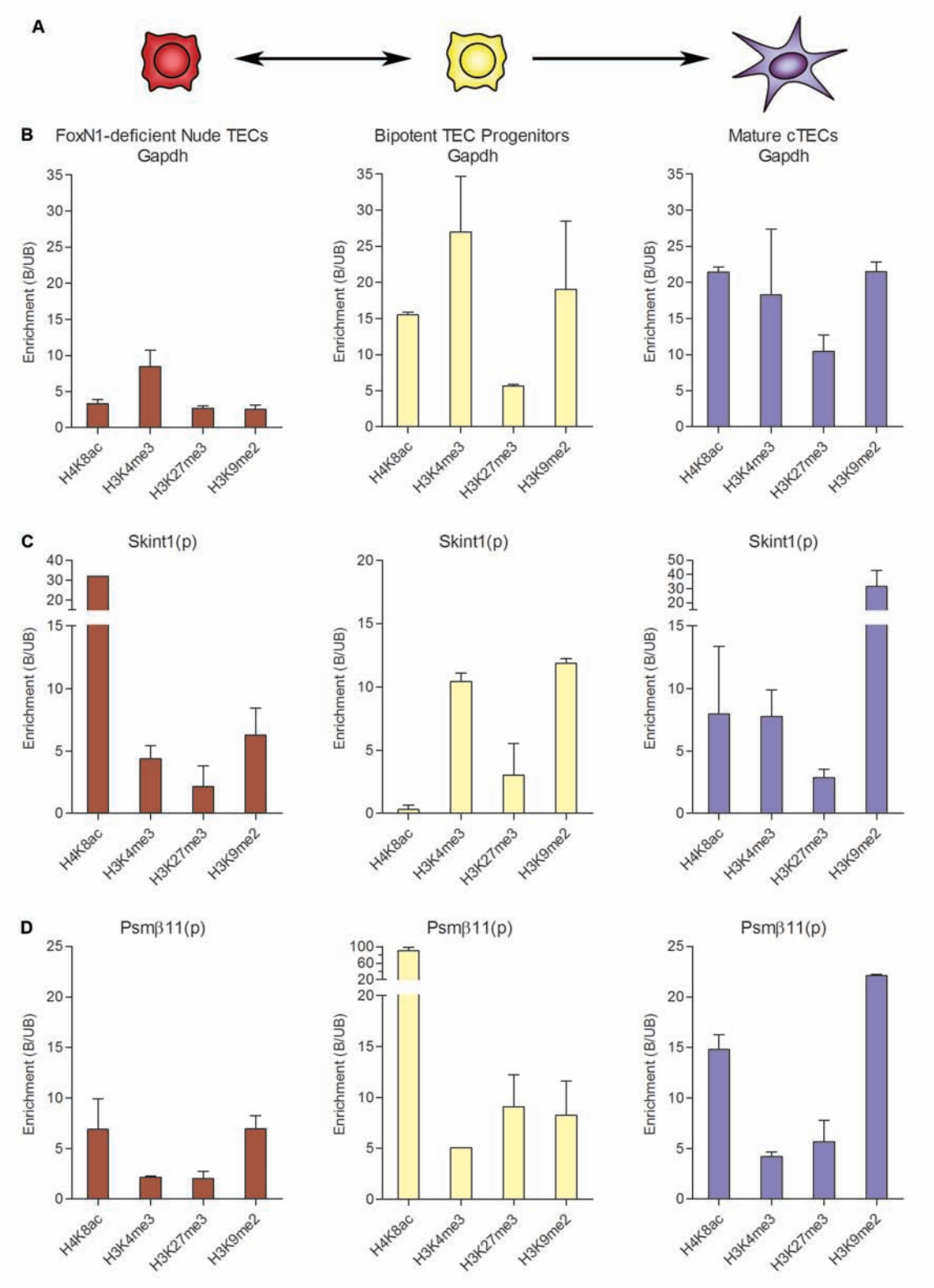
3.5.2. Pattern of Histone Modifications for Glyceraldehyde-3-phosphate, Selection and Upkeep of Intraepithelial T-cells 1 and Proteasome Subunit β-Type 11 within FoxN1-Deficient Nude Thymic Epithelial Cells

When the histone modifications across the three control genes *Gapdh*, *Skint1* and *Psmβ11* were analysed, we again observed divergence from the wild-type (Figure 3.28). This was particularly interesting considering the unchanged active status of

Figure 3.28 – The Epigenetic Patterning of FoxN1-Deficient Nude Thymic Epithelial Cells: Quantitation of the Relative Levels of Histone Modifications across Glyceraldehyde-3-phosphate Dehydrogenase, Selection and Upkeep of Intraepithelial T-Cells 1 and Proteasome Subunit β-Type 11 Upon Disruption of Normal TEC Development

Quantitation by carrier chromatin immunoprecipitation (CChIP) of levels of histone modifications across the control genes; glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*); mTEC-specific selection and upkeep of intraepithelial T-cells 1 (*Skint1*) promoter (p) and cTEC-specific proteasome subunit β-type 11 (*Psmβ11*) promoter in athymic FoxN1-deficient nude mice. *A*, Schematic overview of FoxN1-deficient thymic epithelial cells (TEC). *B*, Foetal thymic organ culture was employed, followed by FACS to isolate *Mus musculus* $EpCAM1^{+}$ FoxN1-deficient TEC and CChIP was carried out with affinity-purified antibodies to H4K8ac, H3K4me3, H3K27me3 and H3K9me2, as indicated. Relative levels (bound:unbound) of histone modifications were calculated from immunoprecipitated (bound) and unprecipitated (unbound) DNA following radioactive PCR. As a comparison, data from normal bipotent TEC progenitors and mature cTECs are displayed. Data are the mean ±SEM from at least two separate CChIP experiments.

Figure 3.28 – The Epigenetic Patterning of FoxN1-Deficient Nude Thymic Epithelial Cells: Quantitation of the Relative Levels of Histone Modifications across Glyceraldehyde-3-phosphate Dehydrogenase, Selection and Upkeep of Intraepithelial T-Cells 1 and Proteasome Subunit β-Type 11 Upon Disruption of Normal TEC Development



Gapdh within the nude cells. In addition, noticeable contrasts between the findings for the TRA *Spt1*, *Spt2* and *Csn1s1* and the control genes arose. For *Gapdh*, (Figure 3.28 B) we observed a shift in the histone modifications in the nude cells in comparison to the equivalent bipotent TEC progenitors, however the predominance of H3K4me3 points to the active status of this gene within the nude cells. Thus it is clear that even the epigenetic status of a ubiquitously-expressed gene is affected by the loss of FoxN1.

For the promoter region of the non-AIRE-regulated *Skint1* (Figure 3.28 C), the nude bipotent TEC progenitor population again showed a remarkable re-distribution of marks from the equivalent wild-type cells. The most obvious change was the significant gain in acetylation at H4K8; a trend which has been consistent for the majority of genes analysed thus far in the nude cells. The overall arrangement of the three methylation marks (H3K4me3, H3K27me3 and H3K9me2) for *Skint1* in the nude progenitor cells does not differ greatly from the wild-type, thus the resultant distribution of all four modifications surprisingly reflects those seen for *Spt1*(c) and *Csn1s1* in nude cells (Figure 3.27 C and E). Comparisons between *Skint1* in nude cells and normal mature cTECs showed little correlation, with the significant enrichment of H3K9me2 detected in the cTECs, not observed in the nude population.

For the FoxN1-deficient nude TEC, the cTEC-specific *Psmβ11* promoter (Figure 3.28 D) is marked with much lower levels of H4K8ac when compared to the wild-type progenitor population, which was surprising as for the majority of genes

analysed, levels of acetylation have generally shown an increase in the nude cells. What is also apparent is the almost total loss of H3K4me3 and H3K27me3, resulting in an arrangement of modifications similar to that seen for the nude cells at the promoter region of *Spt1* (Figure 3.27 B), where low levels of these marks were detected along with comparatively higher levels of H4K8ac and H3K9me2. In comparison with the mature cTECs, where *Psmβ11* is switched on, we observed lower levels of the two active marks H4K8ac and H3K4me3, however the patterns seemed to resemble each other, despite the contrasting activities of this gene within these two populations.

Together, these findings seem to indicate a default epigenetic patterning in the FoxN1-deficient nude TECs, with a predominance of acetylation coupled with a depletion of H3K4me3 and H3K27me3, which may reflect the blocked differentiation programme in these cells.

4. DISCUSSION

The discovery of the transcriptional regulator, AIRE, arguably represents one of the most significant milestones in T-cell biology in the past decade. Its finding reintroduced the concept of central tolerance as a major force in the education of the T-cell repertoire, highlighting this protein's essential role in the prevention of autoimmunity. Since the identification of AIRE, many groups set out to characterise this protein and its importance in T-cell development, and to date much is known; about AIRE's pattern of expression within thymic stromal cells of the medulla; of the ever-increasing list of pathological mutations; genes under the transcriptional control of AIRE; and the proteins with which AIRE has shown an interaction with in the nucleus (Pitkanen *et al.* 2000; Derbinski *et al.* 2001; Anderson *et al.* 2002; Pitkanen *et al.* 2005; Rossi *et al.* 2007; Ruan *et al.* 2007; Koh *et al.* 2008; Org *et al.* 2008). However, the way in which AIRE functions at a molecular level in the control of promiscuous gene expression remains elusive. We have now demonstrated that AIRE is able to bind its individual target genes, triggering a cascade of post-translational histone modifications, leading to the up-regulation of the loci.

4.1. AIRE FUNCTIONS AS A TRANSCRIPTIONAL REGULATOR IN RETROVIRALLY-TRANSFECTED TEP AND 3T3 CELLS

Within the thymus, AIRE is limited to a small subset of $CD80^{+}$ mTECs, making large-scale analysis of the epigenetic status of tissue-restricted antigens under the transcriptional control of AIRE particularly challenging. Thus, in order to mimic the *in*

vivo situation, we employed a *Mus musculus* thymic epithelial (TEP) cell model system, transfected with MIg virus bicistronic constructs containing both AIRE and GFP (TEP-AIRE), or GFP alone as a control (TEP-GFP). Although induction of AIRE in mTECs and its over-expression within a primary mTEC-derived 1C6 cell line is reported to lead to apoptosis, we were able to maintain our TEP lines indefinitely, with no loss of either AIRE or TRA expression (Gray *et al.* 2007). Our data revealed that AIRE was able to up-regulate a number of previously identified TRA; casein-α (*Csn1s1*), and salivary protein-1 and -2 (*Spt1*, *Spt2*), whose expression was either not detected or at a very low level in the control line, with no corresponding impact on housekeeping genes (Anderson *et al.* 2002). Stable transfection of a non-thymic 3T3 fibroblast line with MIg-AIRE-GFP also reliably reproduced PGE, suggesting that AIRE is able to exert its action regardless of the cellular environment, which has been demonstrated in a number of alternative cell lines previously, such as African Green Monkey SV40-transfected kidney fibroblast (COS7) cells, human embryonic kidney (HEK293) cells, pancreatic islet β cells, rhabdomyosarcoma (RD) cells and human leukemic monocyte lymphoma (U937) cells (Halonen *et al.* 2001; Pitkanen *et al.* 2001; Halonen *et al.* 2004; Pitkanen *et al.* 2005; Guerau-de-Arellano *et al.* 2008; Org *et al.* 2008; Org *et al.* 2009). However, what we did find was that the expression of TRA within the 3T3 line was at a much lower intensity than that within the TEP line, thus it would appear that, although the presence of AIRE alone is sufficient for the initiation of PGE, its optimum transcriptional potential is only reached within a thymic epithelial background. This observation could also be due to an alternative repertoire of genes under the transcriptional control of AIRE within different cell populations, which has been previously reported (Gardner *et al.* 2008; Guerau-de-

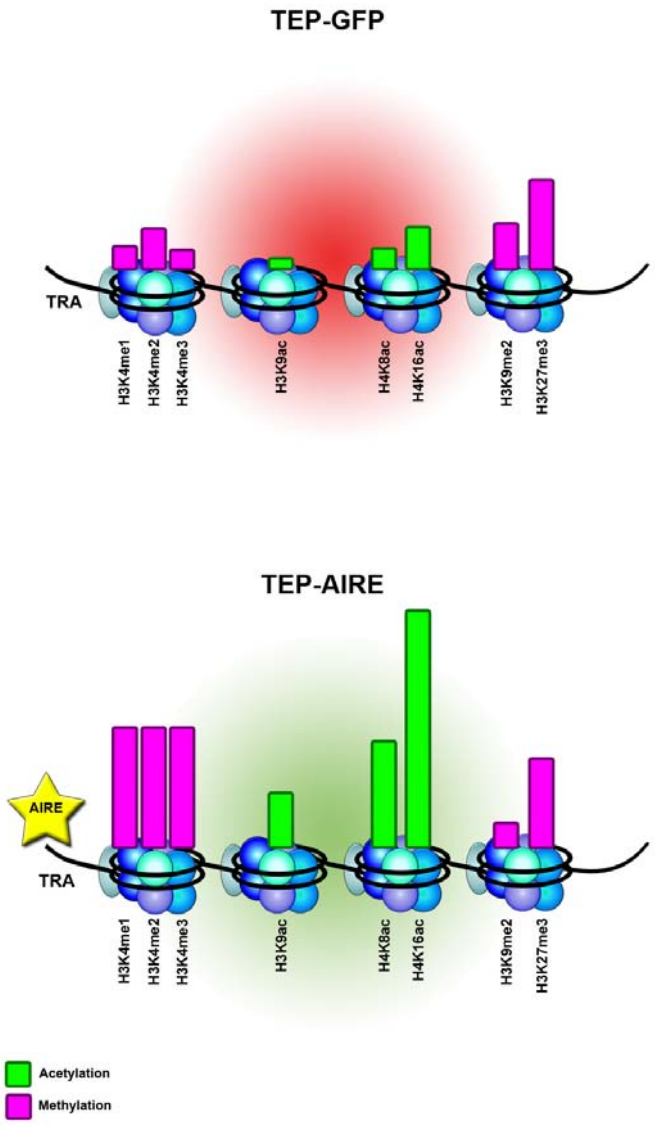
Arellano *et al.* 2008; Org *et al.* 2009). Guerau-de-Arellano *et al* (2008) showed through transfection of a non-thymic, unrelated epithelial cell line; pancreatic islet β cells, that AIRE's general mode of transcriptional regulation can be replicated in tissues other than mTECs, as we saw with the 3T3 line, but that the cellular environment and underlying programme of gene expression ultimately defines which genes will be influenced by AIRE (Guerau-de-Arellano *et al.* 2008). Gardner *et al* (2008) then went on to show that the same is true *in vivo*, with mTECs and extra-thymic AIRE-expressing cells within secondary lymphoid organs expressing distinct catalogues of TRA (Gardner *et al.* 2008). However, this group noted that both the total number of genes and fold-change of expression were much greater in mTECs than in the periphery (Gardner *et al.* 2008). Thus it would appear that cellular environmental cues have a major impact upon the plethora of genes regulated by AIRE. Our TEP line efficiently expressed the TRA; *Csn1s1*, *Spt1* and *Spt2*, all of which are also detected in $CD80^{+}AIRE^{+}$ mTECs in the thymus and was therefore an ideal system for modelling AIRE's mechanism of action within mTECs *in vivo* (Anderson *et al.* 2002).

4.2. AIRE Induces the Enrichment of Active Histone Modifications across Individual Tissue-Restricted Antigens

The development of native chromatin immunoprecipitation (NChIP), which takes advantage of unfixed chromatin to maximise the efficiency and resolution of analysis, has allowed thorough investigation into the distribution of post-translational histone modifications across the genome (O'Neill and Turner 2003). Through the application

of NChIP to the TEP model system, we have assayed eight key histone modifications, which have demonstrated dynamic roles in transcriptional activation and repression. We found that in the absence of AIRE, individual TRA; *Csn1s1*, *Spt1* and *Spt2*, were marked with a relatively consistent pattern of modifications (Figure 4.1). In general, the TEP-GFP cells displayed low levels of the archetypal active marks, and correspondingly higher levels of silent marks across these non-expressed TRA. In particular, a set pattern of H3K4 methylation was observed in the TEP-GFP population, which conformed to a preferential me2 > me1 > me3 distribution, the very low levels of H3K4me3 reflecting the silent nature of these genes. Although H3K4me1 and H3K4me3 were depleted in the TEP-GFP cells at TRA, marks which are typically found in active genes, H3K4me2 was often enriched, however the relationship between H3K4me2 levels and transcription is less well defined and its presence within silenced genes has been noted (Orford, *et al.* 2008). We observed a depletion of H3K9 acetylation across all TRA in the absence of AIRE (Figure 4.1), again correlating with their inactive status, and this modification has been found specifically within active genes at 'acetylation islands', in combination with H3K14ac (Roh *et al.* 2005). Yet acetylation was not absent in the TEP-GFP population at these loci, with fluctuating levels of enrichment detected for acetyl H4 (Figure 4.1), which is in agreement with previous studies showing that the impact of H4K16 and K8 acetylation on the genome is not linked to the control of ongoing transcription, but may instead define euchromatic genes as a mark of potential expression (O'Neill and Turner 1995; Johnson *et al.* 1998). In nucleosomal arrays, H4K16ac has been shown to relax the chromatin structure and hence may prime the genes for transcription (Luger *et al.* 1997; Shogren-Knaak *et al.* 2006). This may be

Figure 4.1 – Model for the Epigenetic Patterning of Tissue-Restricted Antigens under the Transcriptional Control of AIRE in Thymic Epithelial Cell Lines



Establishment of the typical epigenetic marking of AIRE-regulated tissue-restricted antigens (TRA) in *Mus musculus* thymic epithelial (TEP) cell lines. Through native chromatin immunoprecipitation, levels of histone modifications at the TRA casein-α and salivary protein-1 and -2 were determined. A relatively stable pattern of modifications was observed in the absence of AIRE (TEP-GFP) where these genes are silent, with low levels of the archetypal active marks H3K4me3 and H3K9ac in combination with higher levels of the silencing marks H3K9me2 and H3K27me3. In the presence of AIRE (TEP-AIRE), TRA were actively transcribed, thus concordant rises in the active marks; H3K4 methylation, H3K9ac were observed along with a maintenance of H3K27me3, possibly allowing for shut down of the genes after transcription. Although H4 acetylation was present in the control cells, upon AIRE-induced gene activation these marks were seen to increase.

a feature of the TEP background as these genes would never ordinarily be expressed outside of their tissue-restricted niche, however within the thymus, they have the potential to be transcribed in the presence of AIRE. The predominant mark within TEP-GFP cells was the archetypal silencing modification H3K27me3 (Figure 4.1), which has been shown to be essential in the repression of developmental homeobox (*Hox*) genes in *Drosophila melanogaster*, through interactions with Polycomb group proteins (PcG), thus its localisation at TRA is indicative of their off state and possibly implicates PcG complexes in the silencing of these genes (Breiling *et al.* 2004; Schwartz and Pirrotta 2007; Henikoff 2008). This was further confirmed through our analysis of the binding status of PcG proteins; the H3K27me-specific methyltransferase KMT6/Ezh2 and its co-factor Eed at the *Spt1* and *Spt2* promoters, via conventional formaldehyde cross-linked XChIP, showing detectable levels of these proteins in TEP-GFP cells. We also identified the presence of the alternative silencing modification H3K9me2 (Figure 4.1), albeit at lower levels that those of H3K27me3, suggesting a dominance of PcG-mediated silencing mechanisms in the TEP cell line.

The observed distributions of histone modifications in the control TEP-GFP population reflect the silent nature of TRA in these cells, and their relatively static levels are indicative of a locus-specific chromatin signature, patterning each gene with a background of epigenetic signals which could denote these tissue-specific genes, acting as a post-code, facilitating the recruitment of AIRE. Recently, the first PHD finger of AIRE has demonstrated the potential to interact with un-modified H3K4 (me0), an association which was lost upon methylation of this residue (Koh *et al.*

2008; Org *et al.* 2008; Chignola *et al.* 2009). However, from our results, we can conclude that the presence of methylation at H3K4 at TRA does not impede AIRE's actions, as detectable levels of me1, me2 and me3 are found across *Csn1s1*, *Spt1* and *Spt2* in the absence of AIRE and that despite this, AIRE is still recruited and these genes are switched on, although this does result in substantial increases in methylation at this residue. One possible explanation for these contrasting findings could be the different cell lines and techniques utilised for each study, the majority of previous work being carried out in HEK293 cells and through *in vitro* biochemical analysis (Koh *et al.* 2008; Org *et al.* 2008; Chignola *et al.* 2009). Given our observation that AIRE's function is influenced significantly by the cellular background within which it is operating, it may be that the different model systems employed could account for the discrepancies. HEK293 cells are an epithelial line derived from human embryonic kidney and their continued use as a model system, enabling the study of AIRE's mechanism of action, has provided valuable insight into how AIRE functions, however many inconsistencies have also arisen in terms of AIRE's localisation, showing cytoplasmic distributions alongside the *in vivo* punctate nuclear staining, and to the catalogue of genes AIRE influences (Halonen *et al.* 2001; Pitkanen *et al.* 2005; Org *et al.* 2008; Org *et al.* 2009). We observed similar differences with our 3T3 line which, although sorted for 100% transfection efficiency, expressed AIRE and the two salivary proteins at a much lower efficiency that the TEP line. Thus as a representation of AIRE activity within its primary niche, retrovirally transfected TEP cells are more reliable for the epigenetic analysis of TRA. Alternatively, AIRE may require the combinational patterns of histone modifications we observed for its recruitment, reading the chromatin environment at its target

genes and identifying where in the genome it should bind. Due to the regular spacing of nucleosomes, there will be many copies of each histone protein across a single gene and therefore H3K4me0 may serve as a docking site for AIRE, but could function in combination with the modified histones we observed to facilitate AIRE's transcriptional activity. Un-modified H3K4 could be located adjacent to methylated H3K4 on a neighbouring histone or nucleosome, providing a crosstalk. In support of this hypothesis, one group noted, through the use of *in vitro* peptide binding assays, that di-methylation of H3R2 abrogated AIRE's potential to bind H3K4me0 (Chignola *et al.* 2009). However, the complex language of chromatin, including the enormous plethora of post-translational histone modifications and the vast network of histone-modifying enzymes, is difficult to replicate with *in vitro* arrays, hence the true impact of histone modifications on AIRE-regulated expression can only really be modelled through native ChIP analysis of unfixed chromatin in a TEC-like cell line.

In the presence of AIRE we observed highly dynamic chromatin reorganisation with dramatic changes to the distribution of histone modifications across the TRA *Csn1s1*, *Spt1* and *Spt2*, signifying the change in transcriptional status of each loci (Figure 4.1). The most striking transformations were to the active histone marks, always resulting in significantly higher levels of enrichment in the TEP-AIRE cells. For the majority of loci we observed a shift in the balance of H3K4 methylation for the TEP-AIRE population. While the background levels of H3K4 methylation were predominantly di-methyl, in the TEP-AIRE cells tri-methylation dominated; a mark which is found to peak at the promoters of active genes (Barski *et al.* 2007). All three forms of H3K4 methylation were shown to increase in the presence of AIRE, linking

their enrichment to the induction of active transcription. Through XChIP analysis, we observed heightened levels of the H3K4 methyltransferase KMT2A/MLL1 bound to the promoters of *Spt1* and *Spt2* in the presence of AIRE, which could account for the increases in H3K4 methylation across TRA. AIRE also impacts considerably on histone acetylation. Given the near depletion of H3K9ac in the control cells, AIRE-induced up-regulation of the loci leads to its enrichment in TEP-AIRE cells, in line with work showing heightened levels of this modification upon gene activation (Roh *et al.* 2005; Wang *et al.* 2008). Furthermore, despite the presence of H4K8ac and H4K16ac in the TEP-GFP population, these marks also intensify in the presence of AIRE and these modifications are frequently found localised to active loci (Barski *et al.* 2007; Kouzarides 2007). Interestingly, the activation of the loci did not result in dramatic reductions in the silencing modifications H3K27me3 and H3K9me2 from the control TEP-GFP cells (Figure 4.1). However, from our XChIP analysis of the binding status of PcG proteins at *Spt1* and *Spt2*, enrichments of KMT6/Ezh2 and Eed were observed in the presence of AIRE, which may account for the maintenance of H3K27me3 in the TEP-AIRE cells. The presence of inactive modifications in active gene regions has regularly been reported and may point to a system put in place to regulate expression, allowing for the shutdown of the loci after transcription (Vakoc *et al.* 2005; Squazzo *et al.* 2006; Vakoc *et al.* 2006; Barski *et al.* 2007; Kouzarides 2007; Shilatifard 2008). The marks may also prevent over-expression by AIRE due to a lack of normal cellular signalling pathways which would ordinarily operate to regulate transcription within each TRA's natural peripheral environment.

Together, these results show that AIRE is able to induce major alterations to the distribution of histone modifications across individual TRA. Active acetylation and methylation marks increase significantly from the background signature of modifications patterning these genes in the absence of AIRE, whereas silent marks often remain unchanged. Yet these changes are subtly different between TRA, which may account for the altering levels of expression of these genes.

4.3. AIRE's Transcriptional Control of a Cluster of Genes Involves Domain-Wide Alterations to Histone Modifications

As a transcriptional regulator, AIRE must presumably bind each target gene, or a regulatory region of DNA to either activate or repress the loci. Yet AIRE is faced with a number of complexities as it not only controls the expression of many thousands of target genes, they are distributed randomly across all chromosomes (Gotter *et al.* 2004). However, the finding that many of these loci are clustered along the chromosome arms, points to a more domain-wide regulation by AIRE (Derbinski *et al.* 2005; Johnnidis *et al.* 2005). Although our Western blot analysis and immunofluorescence of metaphase chromosome spreads indicated that AIRE does not exert its transcriptional activity through genome-wide acetylation or methylation, our observation of AIRE-induced increases in H3K4 methylation and H3 and H4 acetylation across the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*), with no change to the transcription levels, may support this theory; suggesting a general deposition of H3K4 methylation, mediated by AIRE. Our analysis of a cluster of genes which show differential expression profiles in the

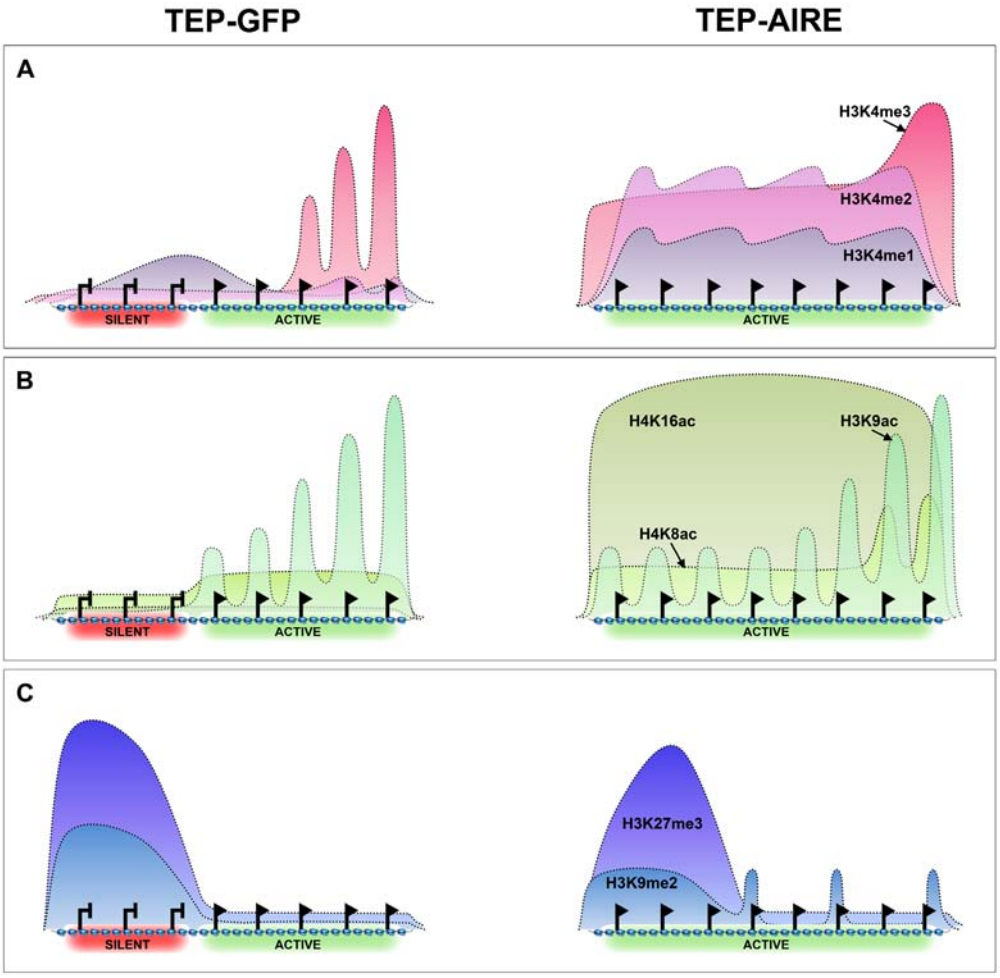
presence and absence of AIRE, does however suggest that AIRE may influence the distribution of certain histone modifications in a domain-wide manner.

Previous analysis into AIRE's control of transcription showed fluctuating levels of output from each target gene and that those located in clusters did not respond equally, often with neighbouring genes acting in opposing ways (Gotter *et al.* 2004; Derbinski *et al.* 2005; Johnnidis *et al.* 2005). Preliminary studies by Johnnidis *et al* (2005) into the clustering of AIRE-regulated genes, demonstrated that one such cluster on *Mus musculus* chromosome 15 contained three genes which were differentially regulated by AIRE (Johnnidis *et al.* 2005). They found that keratin 4 increased in expression, while keratin 18 and eukaryotic translation initiation factor 4b (*Eif4b*) showed a decrease in the presence of AIRE (Johnnidis *et al.* 2005). We extended these studies, to cover the entire keratin cluster, which actually encompasses five keratin genes (*Krt4*, *Krt79*, *Krt78*, *Krt8* and *Krt18*), and three further genes (*Eif4b*, tensin-like C1 domain-containing phosphatase (*Tenc1*) and SPRY domain-containing 3 (*Spryd3*)) within a 217kbp region. We instead found that AIRE exerted a positive influence over all genes, with the exception of keratin 18, which did not appear to be affected by AIRE as approximately equal transcript levels were detected in TEP-AIRE and TEP-GFP cells. Our method of analysis and model system differed from those of the original study, which may explain the observed discrepancies. While Johnnidis *et al* (2005) opted for a large-scale bioinformatic analysis of published microarray data on medullary RNA transcripts from AIRE-deficient versus wild-type mice, we performed quantitative real-time PCR (qPCR) with our TEP cell lines (Johnnidis *et al.* 2005). Validation of the expression of each

individual gene following microarray is an arduous and unrealistic task, however, qPCR is often a more reliable representation of the true situation, which is why we chose to analyse the transcriptional profile across the entire cluster using this technique. The ectopic expression of AIRE-regulated genes within the thymus is known to occur at considerably lower levels when compared to their expression in the relevant peripheral tissues, hence any fluctuations in transcript levels will also be small, and could be lost or misinterpreted following genome-wide analyses (Johnnidis *et al.* 2005). In addition, as mentioned previously, certain variation can arise between different cell types, we therefore wanted to verify the situation in our TEP model system.

Through the use of NChIP we were able to, for the first time, assay the distribution of histone modifications across each of the genes in the AIRE-regulated keratin cluster, in addition to the intergenic regions between the loci. This gave us a thorough analysis of how AIRE influences transcription on a broader scale. Our results provide a direct demonstration that AIRE is able to modulate histone modifications, impacting on each individual mark in a distinct way (Figure 4.2). In general, AIRE-induced up-regulation of the keratin cluster genes resulted in heightened levels of active methylation and acetylation marks, with broad domains of these modifications occurring across all genetic regions of the cluster, often irrespective of the ultimate level of expression when compared to the basal levels in TEP-GFP cells (Figure 4.2 A, B). However, we also observed domains of certain modifications in the control cells, reflecting the expression profile of the genes, particularly in the 3` end of the cluster where active transcription is occurring. In particular, TEP-GFP cells harbour

Figure 4.2 – Model for the Domain-Wide Epigenetic Patterning of a Cluster of AIRE-Regulated Genes in Thymic Epithelial Cell Lines



The impact of AIRE on domain-wide distributions of histone modifications across a cluster of AIRE-regulated genes in *Mus musculus* thymic epithelial (TEP) cell lines. Through native chromatin immunoprecipitation, levels of histone modifications across the keratin cluster on *Mus musculus* chromosome 15 were determined. Histone modifications investigated included; *A*, H3K4 methylation (me1, me2, me3), *B*, H3K9, H4K8 and H4K16 acetylation, and *C*, silencing modifications H3K9me2 and H3K27me3. In the absence of AIRE, genes in the keratin cluster showing active expression in TEP-GFP cells were typically marked with peaks of the archetypal active modifications H3K4me3 and H3K9ac. This pattern was mirrored in the presence of AIRE, suggesting AIRE has limited influence over these marks. Mono-and di-methylation of H4 however were consistently higher in TEP-AIRE cells. In the presence of AIRE, H4 acetylation occurred in a blanket across the cluster, not seen in the active genes in the TEP-GFP cells, suggestive of a role for AIRE in the enrichment of these marks. Silencing modifications were seen to peak across silent genes in TEP-GFP cells, however, this patterning was not lost in the presence of AIRE despite active gene expression, although overall levels were reduced. Peaks of H3K9me2 were introduced in the TEP-AIRE cells at active genes across the cluster and AIRE may play a role in their deposition.

increased levels of the two classical active marks H3K4me3 and H3K9ac at the promoters of active genes (Figure 4.2 A, B) (Roh *et al.* 2005; Shilatifard 2008). Interestingly, the enrichment levels of these two histone modifications, show little change at these genes in the presence of AIRE, when active expression is occurring in both cell populations, despite the approximate 4-fold increases in expression in the TEP-AIRE cells. This may suggest that AIRE does not impact significantly on these modifications, and that their presence within the transcriptionally active genes merely reflects their active status in both the presence and absence of AIRE. In contrast, some modifications show major changes in response to AIRE (Figure 4.2). For example, AIRE appears to modulate acetylation of H4K8 and H4K16, which both appear to show little correlation with transcriptional activity in the control cells in that only minor enrichments were observed in the active 3` genes. Although a small domain-wide increase in H4K8ac was observed for the actively-expressed genes and their corresponding intergenic regions in TEP-GFP cells, the intensity at each active region is heightened in the presence of AIRE. This may account for the increased levels of transcription for the majority of genes in the TEP-AIRE cells. The same is seen for H4K16ac yet to an even greater extent (Figure 4.2 B). In the control cells, there was no correlation between this modification and the expression profile across the cluster, with a consistently low level in both the silent and active loci in TEP-GFP cells, however a broad domain of acetylation spread across the entire cluster in the presence of AIRE. This observation is in agreement with the seminal report by O'Neill and Turner (1995), who discovered that euchromatic coding regions showing ongoing transcriptional activity, did not coincide with induced hyperacetylation, but instead H4 acetylation marked large euchromatic domains, distinguishing them from

the hypoacetylated heterochromatin (O'Neill and Turner 1995). With this in mind, our demonstration of constant background levels of H4 acetylation in TEP-GFP cells and AIRE-induced H4 hyperacetylation may suggest a role for AIRE in the direct or indirect deposition of these marks. This may allow an opening of the locus, facilitating higher levels of expression across the keratin cluster mediated by AIRE. This is particularly interesting considering the co-localisation of AIRE and KAT3A/CBP, a ubiquitous transcriptional activator with intrinsic histone and non-histone acetyltransferase activity, an association shown to enhance AIRE's transactivating potential (Pitkanen *et al.* 2000; Akiyoshi *et al.* 2004; Pitkanen *et al.* 2005; Ferguson *et al.* 2007). Although KAT3A/CBP participates in hundreds of different transcriptional programmes, binding transcription factors including members of the general transcription machinery such as TFIID, TFIIB and RNA polymerase II, its nuclear co-localisation within mTECs has been shown to be specifically dependent upon AIRE (Bannister and Kouzarides 1996; Vo and Goodman 2001; Ferguson *et al.* 2007). Through synchronous AIRE induction by anti-RANK stimulation of foetal thymic organ cultures, Ferguson *et al* (2007) showed that prior to the up-regulation of AIRE, KAT3A/CBP remained cytoplasmic, only to translocate to the nucleus upon AIRE production, leading to the co-accumulation of AIRE and KAT3A/CBP within focal nuclear puncta (Ferguson *et al.* 2007). The heightened levels of H4 acetylation we see in TEP-AIRE cells across the entire keratin cluster may be brought about as a result of this interaction. Given that depletion of KAT3A/CBP suppresses the rate of transcription, as demonstrated for IFNβ gene expression, the interaction between AIRE and KAT3A/CBP may facilitate the

enhanced levels of transcription across the majority of the genes in the cluster (Yie *et al.* 1999).

Our results also imply that AIRE influences the modifications typically associated with silent genes; H3K9me2 and H3K27me3 (Figure 4.2 C). Both of these marks are concentrated towards the 5` end of the cluster and are then lost further downstream which, in the control cells, correlates perfectly with the expression profile across the cluster. Interestingly, TEP-AIRE cells also show this domain of silencing modifications, despite active expression occurring for every gene. However, while TEP-GFP cells show peaks in silent promoters, TEP-AIRE cells show higher levels in the 5` intergenic regions. Further along the cluster, peaks of H3K9me2 do occur, but only in the TEP-AIRE cells, specifically at promoter regions thus, whereas the levels of H3K9me2 correlate perfectly with the expression profile in the TEP-GFP cells, in the presence of AIRE we see a deposition of this mark at actively expressed genes, hence it would appear that AIRE is able to manipulate the levels of H3K9me2. The presence of silent marks in active genes is in agreement with our findings for the TRA *Csn1s1*, *Spt1* and *Spt2* and also with published data, strengthening the idea that these marks could be indicative of an internal safety feature for AIRE-regulated transcription; preventing aberrant expression (Vakoc *et al.* 2005; Squazzo *et al.* 2006; Vakoc *et al.* 2006; Kouzarides 2007; Shilatifard 2008). The fact that the levels of these marks fluctuate frequently in the presence and absence of AIRE shows how there is no strict rule for the functional outcome of each individual modification, but that it is the overall combination of marks that dictates a genes activity.

Collectively, these results show that, in its role as a transcriptional regulator, AIRE is able to command both subtle and domain-wide changes in the epigenetic status of its target genes. Our data also highlight the significant bearing post-translational histone modifications impart upon the control of gene expression and of chromatin in general.

4.4. Tissue-Restricted Antigens are Marked with Dynamic Histone Modifications Which Rapidly Rearrange Upon Differentiation Throughout the TEC Developmental Pathway

The use of cultured cell lines as model systems of the *in vivo* situation are however limited in their application. In the case of the TEP line, the transfected cells ultimately represent the end points of a well established developmental pathway which begins in the embryo and continues throughout adulthood, with mature $CD80^{+}AIRE^{+}$ mTECs constantly turning over and replenishing the population (Gray *et al.* 2007; Ferguson *et al.* 2008). Although our TEP system replicated AIRE's role with regards to its expression, subcellular localisation, and its control of PGE, we have already demonstrated AIRE's requirement for a thymic cell background to function optimally. The perfect scenario would therefore be to look within the embryonic thymus, in the cells in which AIRE is expressed *in vivo*, ideally immediately following AIRE's up-regulation, thus giving a snapshot into the epigenetic state of the cells prior to the onset of AIRE-induced replicative cell death. The major steps in the pathway of TEC development have recently been uncovered in two influential reports by Rossi *et al* (2006, 2007), showing that cortical and medullary TECs stem from a common

bipotent TEC progenitor (Rossi *et al.* 2006; Rossi *et al.* 2007). They also defined an additional layer of differentiation, which exists for the mTEC lineage, with mature $CD80^{+}AIRE^{+}$ mTECs passing through an immature $CD80^{-}AIRE^{-}$ phase, only completing development upon stimulation with RANK ligand (Rossi *et al.* 2007). These important findings provided an insight into how thymic development is controlled; finally resolving the long-standing issue of whether cortical and medullary TECs derive from individual stem cell lineages or a single progenitor. However, this work also clarified how and when AIRE is switched on in the thymus, thus offering evidence in support of the terminal differentiation model; assuming PGE to be an autonomous trait of mTECs, with the number and complexity of genes expressed increasing as the cells mature (Farr *et al.* 2002; Kyewski *et al.* 2002; Gotter and Kyewski 2004; Devoss and Anderson 2007).

Foetal thymic organ culture preserves the three-dimensional architecture of the developing thymus, and the phenotype of each stromal cell subset (Jenkinson and Anderson 1994). Thus regular programmes of cell development, in particular that of TECs, can occur as normal (Jenkinson and Anderson 1994; Anderson and Jenkinson 1995; Anderson and Jenkinson 2008). Through FACS sorting, pure populations of each of the four phases in the TEC developmental pathway (bipotent TEC progenitors, immature and mature mTECs and mature cTECs) can be isolated. Due to the limited cell numbers in each population, we were unable to analyse the distribution of histone modifications across TRA by native ChIP, which requires a minimum of $1x10^{7}$ cells, thus we turned to the novel technique of carrier ChIP, developed within our lab for the detailed analysis of small numbers of cells (O'Neill *et*

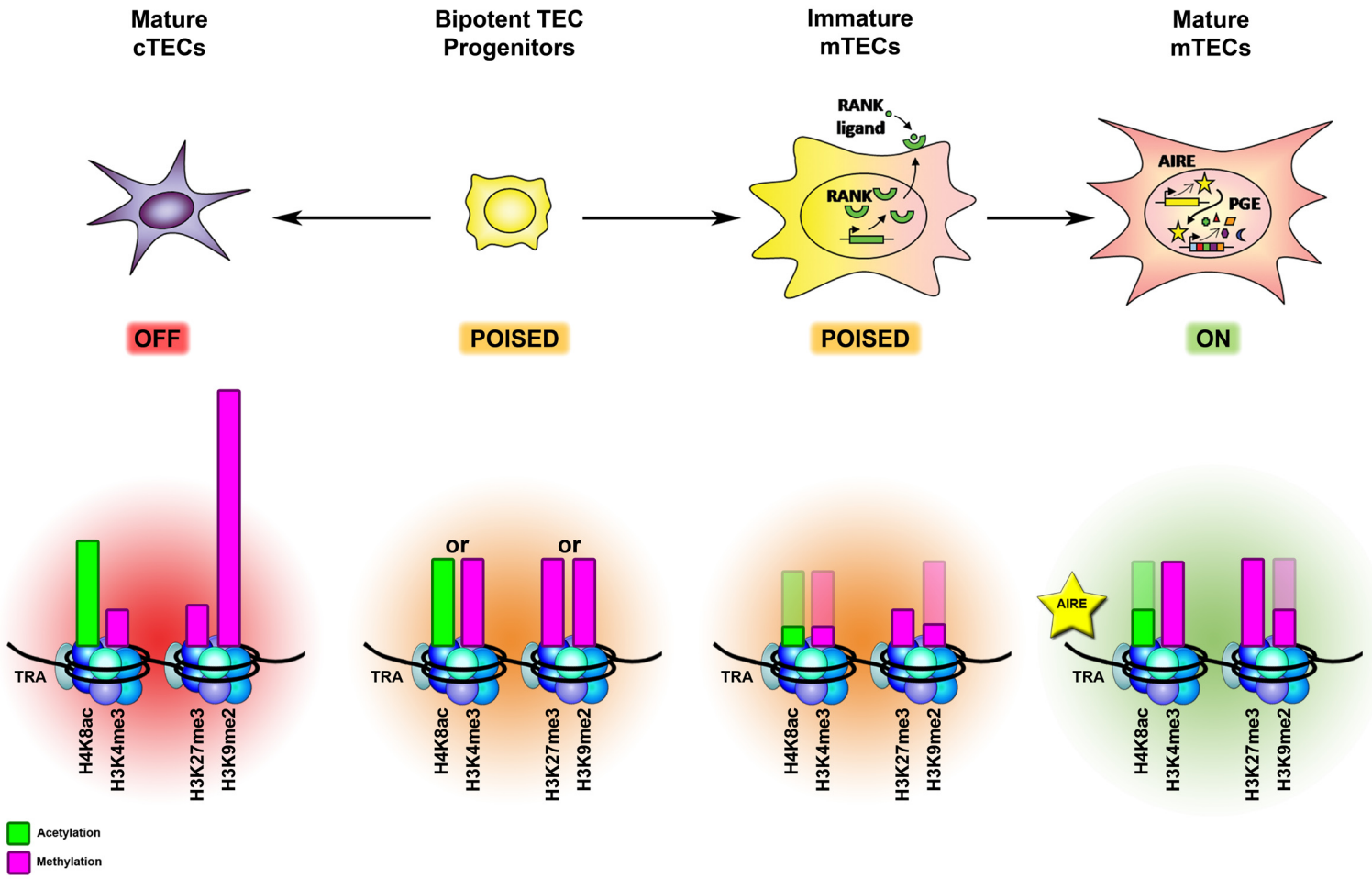
al. 2006). This procedure takes advantage of the high efficiency of native ChIP with unfixed chromatin and, by reducing wash volumes and keeping losses at each stage to a minimum, results in immunoprecipitation recoveries of up to 28%, although the number of possible PCR assays is severely limited.

We have successfully employed CChIP to assess the patterns of four histone modifications (H4K8ac, H3K4me3, H3K27me3 and H3K9me2) across TRA and control genes throughout the four main stages of the embryonic TEC developmental pathway (Figure 4.3). Within the TEC developmental pathway, bipotent TEC progenitors essentially represent an undifferentiated population and we therefore chose to analyse the bivalent modifications H3K4me3 and H3K27me3, which have been shown to prime key developmental genes in pluripotent embryonic stem cells for future transcription (Azuara *et al.* 2006; Bernstein *et al.* 2006). Although we repeatedly saw a co-localisation of active and silent marks for TRA in the undifferentiated cells, this was not always the traditional H3K4me3 / H3K27me3 bivalency (Figure 4.3). For example, across *Spt1* we observed conventional bivalency at the promoter region, however the *Spt2* promoter showed a predominance of H4K8ac and H3K9me2 and the *Csn1s1* promoter displayed an H4K8ac / H3K27me3 patterning. Thus it would appear that the established H3K4me3 / H3K27me3 bivalency does not always play an essential role in the priming of TRA throughout the TEC developmental pathway, which may be due to the more developed nature of these progenitor cells, being bipotent rather than pluripotent embryonic stem cells. However, what is apparent is that these TRA are marked with a poised patterning of histone modifications, frequently displaying a

Figure 4.3 – Model for the Epigenetic Patterning of Tissue-Restricted Antigens under the Transcriptional Control of AIRE throughout the TEC Developmental Pathway

Model of the typical epigenetic marking of AIRE-regulated tissue-restricted antigens (TRA) across the four main stages of the *Mus musculus* thymic epithelial cell (TEC) developmental pathway *in vivo*. Through carrier chromatin immunoprecipitation of FACS sorted primary cell populations from foetal thymic organ culture, levels of histone modifications at the TRA casein-α and salivary protein-1 and -2 were determined. In general, the distribution of histone modifications were dynamic, showing highly gene-specific patterning, however some key commonalities do occur. A bivalent chromatin structure was always found in the bipotent TEC progenitors, in which TRA are poised for future expression within mTECs. Although generally not the classical bivalent marks H3K4me3 / H3K27me3, progenitor TRA frequently showed equal levels of an active mark (H4K8ac or H3K4me3) and a silencing mark (H3K9me2 or H3K27me3) or a combination. Immature mTECs showed dramatic re-organisation to their chromatin, with the two salivary proteins displaying very low levels of all four modifications, although this was not always the case as casein-α had high levels of the active marks and high H3K9me2. Activation of the loci in mature mTECs again showed variations to the distribution of modifications, although typically genes were marked with the presence of both active and silent marks with either a blanket of all four modifications, or H4K8ac and H3K9me2 were low. Mature cTECs, where the loci were more permanently silenced, showed a relatively stable pattern of modifications, with low H3K4me3 and H3K27me3, moderate H4K8ac and very high H3K9me2, keeping the genes silent.

Figure 4.3 – Model for the Epigenetic Patterning of Tissue-Restricted Antigens under the Transcriptional Control of AIRE throughout the TEC Developmental Pathway



combination of an active and a silent mark (Figure 4.3). The presence of an active mark such as H4K8ac or H3K4me3 marking the genes for expression should these bipotent TEC progenitors receive signals to differentiate into mTECs, would maintain a relaxed chromatin structure to allow for future transcription, whilst the silent marks present keep the genes off during development and would enable rapid shutdown of the loci upon differentiation into cTECs. This epigenetic priming is highlighted when the distribution of marks are evaluated across the entire developmental pathway, particularly when comparisons are made to the mature cTEC population where TRA expression is switched off more permanently. Typically mature cTEC TRA were marked with very high levels of H3K9me2, signifying the silent nature of these genes in the cortex. Interestingly, even when H3K27me3 was the predominant silencing mark in the bipotent TEC progenitor population, cTECs still presented with higher H3K9me2 than H3K27me3, suggesting that H3K9me2 is possibly a stronger silencing signal on these genes. Mature cTECs also showed low levels of H3K4me3 and H3K27me3 along with moderate H4K8ac and this pattern of modifications was repeated in the majority of TRA. This distribution was also mirrored in the cTECs at the mTEC-specific, but AIRE-independent selection and upkeep of intraepithelial T-cells 1 (*Skint1*), whose expression profile matches that of the AIRE-regulated TRA. This may suggest that this combination of marks is essential for the more permanent silencing of these genes in the mature cTECs, possibly reflecting a default or stable chromatin state, as opposed to the poised nature of these genes throughout the mTEC lineage, hence their more dynamic histone modifications. This pattern did not always follow however, as *Csn1s1* was instead marked with high levels of H4K8ac in the mature cTEC population (Figure 4.3). This was interesting given our observation

of AIRE-induced domain-wide H4 hyperacetylation and the presence of H4 acetylation across TRA within the control TEP-GFP cells, indicating that these marks show little correlation with active, on-going transcription. This high H4 acetylation could be due to an alternative repertoire of histone acetyltransferases or deacetylases at the different loci, leading to diverse turnover rates of acetylation. However, what is obvious is that this high H4K8ac is not able to override the silencing signals. This is also reflected in the mTEC lineage where again large and highly dynamic chromatin re-organisations are observed as the cells develop (Figure 4.3). Typically, upon lineage commitment to immature mTECs, we saw a fall in the levels of all four modifications. But again this was not ubiquitous as *Csn1s1* showed higher levels of H4K8ac, H3K4me3 and H3K9me2. This re-distribution of marks within the immature mTECs reflects the change in phenotype and development, with the TRA epigenetic patterning altering as these genes move one step closer to their ultimate expression, despite remaining in a poised state. Maturation of the cells and the initiation of AIRE-induced TRA expression sees further interesting shifts in histone modifications, which again are unique for each gene region analysed. In general however, we often recorded increases or maintenance of the levels of the active marks H4K8ac and H3K4me3 from the immature mTECs, with only *Csn1s1* showing a decrease in H4K8ac. We frequently observed increases in H3K27me3 and H3K9me2, again with the exception of *Csn1s1* where H3K9me2 levels fell in the mature mTECs. The presence of the silencing marks within the mTEC lineage and bipotent progenitors, and frequent predominance of H3K9me2, may imply a more strict regulation of gene expression within the embryo, maintaining a high level of the mark across the progenitors and immature mTECs to prevent aberrant gene

expression, and within the mature mTECs to allow for rapid shut down of the locus when required. Alternatively, H3K9me2 could be acting as a signal for the recruitment of AIRE, marking its target genes as transcriptionally silent and working in combination with the active marks present across the loci to permit transcription. These findings emphasize how individual genes respond in different ways throughout a developmental pathway, supporting the hypothesis that combinations of marks act in synergy to direct transcriptional programmes. Our results have also highlighted the importance of *in vivo* investigations, which could not have been possible without the development of the CChIP technique. We often observed contrasting findings between the primary TEC developmental pathway and the TEP model system, for example, the clear cut and recurring enrichment of active marks in combination with lower levels of silent marks seen in the TEP line, were less obvious across the four TEC populations. Instead, high levels of silencing marks, in particular H3K9me2 were observed for the majority of TRA across all four cell populations *in vivo*, often predominating over the active marks even when gene expression was occurring. Although enrichment of silencing modifications was also seen in the TEP lines in active loci, the dominant mark was typically H3K27me3, which may imply a less strict regulation of expression within the cell lines.

Interestingly, the control gene *Gapdh* exhibited fluctuating levels of histone modifications, although the common theme throughout each stage of the developmental pathway for this gene was the enrichment of active modifications. However these were frequently matched by equally high levels of silencing marks as observed for the TRA. In the study by Vakoc *et al* (2005), which showed H3K9me2

and me3, along with HP1γ, within the transcribed regions of active mammalian genes, *Gapdh* was also found to display these marks (Vakoc *et al.* 2005). The authors concluded that the presence of these silencing modifications at the ubiquitously expressed *Gapdh*, in addition to the increases they observed within genes which were up-regulated in their model system, reflected more steady-state transcription rates, rather than genes undergoing dynamic fluctuations in expression (Vakoc *et al.* 2005). They proposed that the specific location of H3K9 methylation may determine its functional outcome; associating with transcriptional repressors at the promoter region, yet aiding transcriptional elongation within transcribed regions (Vakoc *et al.* 2005). Our results however, do not support this theory as we observed high levels at both the promoter and coding regions of *Spt1* throughout all stages of the developmental pathway when the gene is both on and off. Instead our findings substantiate this group's alternative theory of H3K9 methylation, in combination with HP1γ, exerting an attenuating effect on elongation by RNA pol II, thus preventing aberrant expression (Vakoc *et al.* 2005).

Analysis of the distribution of modifications across each stage of the TEC developmental pathway has allowed an insight into how these marks contribute to PGE prior to, and immediately following AIRE's up-regulation. In a recent report by Org *et al* (2009), cross-linked ChIP was used to analyse H3K4me3, bulk acetylated (Ac)H3 and H3K27me3 in AIRE-transfected and control HEK293 cells, and in $CD80^{-}$ $AIRE^{-}$ and $CD80^{+}AIRE^{+}$ mTEC populations (Org *et al.* 2009). This group, who also reported on AIRE's ability to bind H3K4me0, showed that in the absence of AIRE within cultured cells, the TRA *GHR* and *LPL* showed low levels of active marks

H3K4me3 and AcH3, which increased in the AIRE-transfected cells, from which they concluded that the absence of these marks is a signal for AIRE-recruitment and binding to H3K4me0 (Org *et al.* 2009). This was not always the case however as they also found high levels of H3K4me3 in the AIRE-regulated TRA *INV*, *INS* and *S100A8* in their AIRE-negative control line, which were maintained when AIRE was present (Org *et al.* 2009). When they looked *in vivo* in the mouse, they saw low levels of H3K4me3 in $CD80^{-}AIRE^{-}$ mTECs across *S100a8* and *Ins2* in comparison to the levels of this mark in the corresponding peripheral tissues, which contrasted with their cell line data for these genes (Org *et al.* 2009). Upon maturation of the cells to $CD80^{+}AIRE^{+}$ mTECs, levels of H3K4me3 increased for *S100a8*, but remained low for *Ins2*, again contrasting with the high levels of this mark seen in their HEK293 lines for this gene (Org *et al.* 2009). They do not address these discrepancies in their report, however this is in agreement with our data, showing that histone modifications can fluctuate between model systems and between individual genes. Given that their data is generated from non-thymic HEK293 cells, this may account for their contrasting findings. In our observations the two salivary proteins for example displayed a substantial decrease in all four modifications we assayed as the cells progressed from bipotent TEC progenitors to immature $CD80^{-}AIRE^{-}$ mTECs, levels of which then rose as the cells matured. However, for *Csn1s1*, we observed a contrasting scenario, with high levels of enrichment of H4K8ac, H3K4me3 and H3K9me2 in the immature cells, that were either maintained (H3K4me3), or fell (H4K8ac and H3K9me2) as AIRE was switched on. These dynamic fluctuations throughout each stage of the TEC developmental pathway were not addressed by Org *et al* (2009) as their analysis did not stretch to the four definable TEC

populations and their range of marks assayed was limited, therefore our findings offer a more comprehensive view of the role of histone modifications in the control of promiscuous gene expression *in vivo*.

Throughout the TEC developmental pathway, we observed considerable changes to the distribution of histone modifications at each phase of differentiation that varied for every gene analysed, highlighting the dynamic nature of these epigenetic phenomena. Our data suggest that instead of individual marks working as static on/off switches, histone modifications act in concert to direct gene expression and that combinatorial patterns dictate whether a gene is on or off.

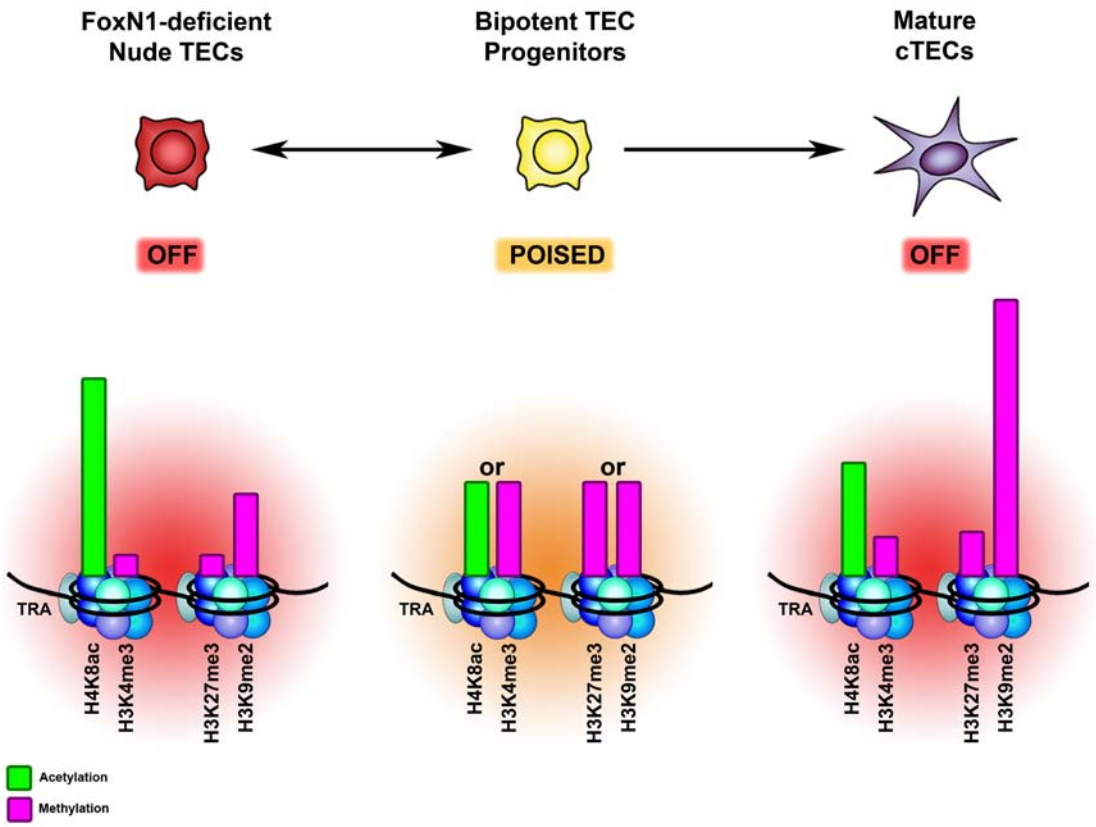
4.5. The Epigenetic Priming of AIRE-Regulated Genes is Lost Upon Disruption to the Pathway of TEC Development

For the first time, a comprehensive analysis of histone modifications throughout a cellular developmental pathway, covering a population of bipotent progenitors, along with their two end points of differentiation, has been made possible through the application of the novel and complex CChIP technique. This has allowed an insight into the roles these modifications play in priming genes for future expression, in addition to the ultimate outcomes when the genes are either expressed or silenced. Yet it does not answer the question of whether these marks are essential for development. The only way to address this is to disrupt or block the pathway of development and then analyse the distribution of modifications once normal differentiation is prevented. To this end, the athymic nude mouse model which is

deficient in the forkhead transcription factor FoxN1, provides an invaluable tool for these investigations. The lack of FoxN1 leads to the generation of a non-functional cystic thymic rudiment due to an early block in normal TEC development (Blackburn *et al.* 1996; Anderson and Jenkinson 2001; Jenkinson *et al.* 2008). Both medullary and cortical lineages cannot develop, however the preliminary stages of thymic organogenesis in which bipotent TEC progenitors are generated, does occur (Anderson and Jenkinson 2001; Anderson *et al.* 2007; Chen *et al.* 2008; Jenkinson *et al.* 2008). These small numbers of FoxN1-deficient nude TEC, representing the wild-type bipotent TEC population, were FACS sorted and used subsequently in CChIP to analyse the epigenetic status of TRA when normal thymus embryogenesis is blocked.

Our investigation into the patterns of histone modifications across the TRA *Spt1*, *Spt2* and *Csn1s1* within these TEC progenitors from foetal FoxN1-deficient nude mice has provided direct evidence for their importance in the establishment of future gene expression profiles. Due to the disruption to FoxN1-deficient nude TEC development, these genes lose the potential to be expressed within mature mTECs, and this was reflected in the chromatin, with often major shifts in the distribution of modifications (Figure 4.4). While the normal bipotent TEC progenitors occasionally displayed a bivalent chromatin signature, not necessarily with the classic bivalent marks H3K4me3 and H3K27me3, but a mixture of active and silent marks, the FoxN1-deficient nude TEC displayed particularly low levels of H3K4me3 and H3K27me3. Thus it would appear that the nude TEC progenitor equivalents lose the pattern of histone modifications set up in the wild-type bipotent TEC population to

Figure 4.4 – Model for the Epigenetic Patterning of Tissue-Restricted Antigens under the Transcriptional Control of AIRE upon Disruption to the TEC Developmental Pathway



Model of the typical epigenetic marking of AIRE-regulated tissue-restricted antigens (TRA) within *Mus musculus* FoxN1-deficient nude thymic epithelial cells (TEC), following disruption to normal TEC development. The nude mouse model, athymic due to a deficiency in the forkhead transcription factor FoxN1, can generate FoxN1-deficient TEC progenitors, but medullary and cortical TECs cannot develop (Blackburn *et al.* 1996; Anderson and Jenkinson 2001; Anderson *et al.* 2007; Chen *et al.* 2008; Jenkinson *et al.* 2008). Through carrier chromatin immunoprecipitation of FACS sorted primary cell populations from foetal thymic organ culture, levels of histone modifications at the TRA casein-α and salivary protein-1 and -2 were determined. Although FoxN1-deficient nude TEC represent bipotent TEC progenitors, the poised status of TRA within these cells is lost as mTECs can never develop and this was reflected in the chromatin. FoxN1-deficient nude TEC were instead marked with a relatively stable pattern of modifications, typically with low H3K4me3 and H3K27me3, moderate H3K9me2 keeping the loci silent and very high H4K8ac, possible representing a 'memory' of normal development. This pattern was similar to that seen in mature cTECs where the loci are permanently silenced, however the level of H3K9me2 was not as high in the nude cells, indicating that the loci may not be switched off as strictly.

prime the genes for future transcription. Instead, these cells were marked predominantly with high levels of H4K8ac, along with modest H3K9me2 for the majority of TRA and control genes (Figure 4.4). Although subtle differences did present themselves between individual genes, what was consistent was the significant divergence between the patterns seen in the nude cells when compared to the equivalent wild-type cells, suggesting that because of the halt in development, these genes are instead marked for an alternative fate of permanent silencing. In support of this, the patterns observed in the FoxN1-deficient nude TEC often reflected those seen in mature cTECs, where these genes are off (Figure 4.4). However, one major difference distinguishing the nude cells from the cTECs is the dominance of H4K8ac over H3K9me2; the mark which was pronounced in the cTECs, denoting the silencing of the TRA in these cells. The retention of a mark of active transcription in the nude TEC may represent the preservation of an open chromatin structure, as a 'memory' of the developmental pathway. Prior to the block in differentiation, these cells could have been 'unaware' of their fate, hence beginning the process of establishing the epigenetic status of the TRA. Of the four modifications we investigated, H4K8ac may be deposited across these genes early in development as a predictive mark, allowing an opening of the domain and preventing the formation of repressive heterochromatin across the TRA that would ordinarily be expressed further along the differentiation pathway. It is known that the presence of H3K4me3 and H3K27me3 across certain genes in pluripotent embryonic stem cells renders them in a poised state (Azuara *et al.* 2006; Bernstein *et al.* 2006). Cui *et al* (2009) have added to this story, showing that those genes which go on to become active, losing the repressive H3K27me3 mark, uniquely carried high levels of

H3K4me1, H3K9me1, H4K20me1 and RNA pol II in the undifferentiated cells, not observed in the genes which subsequently remained silent (Cui *et al.* 2009). Alternatively, Golebiewska *et al* (2009) have demonstrated that, rather than H3K27me3, H3K9me2 in combination with H3K4me3, could play a more dominant role in the poising of certain embryonic stem cell genes (Golebiewska *et al.* 2009). They did observe heightened H3K27me3, but not until the first stages of lineage commitment had been completed (Golebiewska *et al.* 2009). Thus some genes are often 'conscious' of their futures, which is reflected in their chromatin. Our findings, showing a default epigenetic patterning in the FoxN1-deficient nude TECs with enrichment of H4K8ac and low H3K4me3 and H3K27me3 across TRA and the control genes *Skint1* and proteasome subunit β-type 11 (*Psmβ11*), may also represent a 'knowledge' of the destiny of these genes. They will never be expressed, however, the retention of acetylation keeps the chromatin in an open and accessible form, so that should AIRE and other transcriptional regulators be switched on, the genes would be prepared and not completely shut down as is the case for TRA within the cTEC lineage. In support of Golebiewska *et al* (2009), we observed higher levels of H3K9me2 than H3K27me3, however, this was matched with consistently low levels of H3K4me3 (Golebiewska *et al.* 2009). This could suggest that the active mark H3K4me3, which is known to peak at active gene promoters, may be a more dynamic modification, being rapidly removed in response to the block in the developmental pathway, thus preventing aberrant expression due to the loss of developmental cues for either silencing or activation of these genes (Vermeulen *et al.* 2007).

4.6. CONCLUSION

Histone modifications are notoriously complicated in both their regulation and their functional outcomes. The preliminary data that we have demonstrated here gives an insight into the role these marks play in the control of PGE, and how AIRE may be able to manipulate their positioning and intensity. We have shown how simply the introduction of AIRE into an AIRE-negative background is sufficient for the onset of PGE, suggesting that AIRE is able to function as a transcriptional regulator without the need for additional mTEC-specific factors, but that this activity is dependent upon the cellular environment. Our results indicate that AIRE is recruited to the loci of TRA along with RNA pol II and chromatin modifying complexes including KMT2A/MLL1 and KMT6/Ezh2 with its co-factor Eed, which leads to the redistribution of histone modification and the induction of expression. We saw how AIRE was able to increase the levels of active acetylation and methylation marks, in particular gene-specific H3K4 methylation, yet silencing modifications were often maintained as a theoretical safety net, preventing unnecessary or un-regulated expression. AIRE was also able to mediate domain-wide changes to histone modifications across the keratin cluster to allow for the heightened levels of expression seen for the majority of loci, with hyperacetylated H4 potentially relaxing the chromatin structure, facilitating AIRE's actions. Through FTOC and CChIP we have been able to track the epigenetic control of PGE throughout the recently defined TEC developmental pathway *in vivo* which has provided further clues into the function of post-translational histone modifications in the management of developmental programmes of gene expression. This has revealed how AIRE-regulated TRA are marked or

primed for future transcription with both active and silent modifications co-existing throughout TEC development in bipotent TEC progenitors and immature mTECs, although typically not the traditional H3K4me3 / H3K27me3 bivalent marks. Furthermore the pattern of modifications changes dramatically as the cells progress through differentiation, highlighting the dynamic nature of post-translational histone marks. However, a more stable epigenetic signature is required for the permanent silencing of the loci in mature cTECs, where expression would be detrimental. The importance of these unique, gene-specific patterns of modifications were confirmed through disruption to normal TEC development, when the alternative transcriptional fate of the TRA was reflected in the chromatin.

Collectively our results show how significant epigenetic patterning is to the control of expression of individual genes throughout a developmental pathway, and how these patterns can differ between model systems both on a gene-by-gene basis and on a more broad domain-wide scale. CChIP has made the investigation of post-translational histone modifications finally applicable to small numbers of primary cells, however because of this the scope of analysis is always restricted. We chose to study four key histone modifications throughout TEC development which has given invaluable insight onto the control of PGE *in vivo*, yet from our preliminary results in cultured TEP cells it is clear that a great plethora of modifications act in concert to direct transcriptional programmes. Thus, determining the definitive contribution epigenetic marks have in the regulation of PGE will require further analysis with additional modifications. Clarification of the direct impact of AIRE and chromatin-modifying complexes on the expression of these genes *in vivo* would provide insight

into how PGE is directed on an epigenetic level, however at present a protocol for miniaturised XChIP has not been developed. Alternatively, a model system in which AIRE expression is inducible, such as a thymic epithelial line stably transfected with RANK whereby anti-RANK stimulation would switch on biologically relevant levels of AIRE, could allow a more accurate reproduction of the mTEC lineage. Through inhibition of certain histone modifying enzymes such as histone deacetylase enzymes, the significance of the chromatin signature of each TRA could be determined, bringing us one step closer to a clarification of the roles of the various histone modifications both individually and in combination.

5. REFERENCES

Aaltonen, J., P. Bjorses, L. Sandkuijl, J. Perheentupa and L. Peltonen (1994). "An autosomal locus causing autoimmune disease: autoimmune polyglandular disease type I assigned to chromosome 21." Nat Genet **8**(1): 83-7.

Adamson, K. A., S. H. Pearce, J. R. Lamb, J. R. Seckl and S. E. Howie (2004). "A comparative study of mRNA and protein expression of the autoimmune regulator gene (Aire) in embryonic and adult murine tissues." J Pathol **202**(2): 180-7.

Akhtar, A. and P. B. Becker (2000). "Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila." Mol Cell **5**(2): 367-75.

Akiyama, T., Y. Shimo, H. Yanai, J. Qin, D. Ohshima, Y. Maruyama, Y. Asaumi, J. Kitazawa, H. Takayanagi, J. M. Penninger, for full list of author see journal. (2008). "The Tumor Necrosis Factor Family Receptors RANK and CD40 Cooperatively Establish the Thymic Medullary Microenvironment and Self-Tolerance." Immunity **29**(3): 423-437.

Akiyoshi, H., S. Hatakeyama, J. Pitkanen, Y. Mouri, V. Doucas, J. Kudoh, K. Tsurugaya, D. Uchida, A. Matsushima, K. Oshikawa, for full list of author see journal. (2004). "Subcellular expression of autoimmune regulator is organized in a spatiotemporal manner." J Biol Chem **279**(32): 33984-91.

Allfrey, V. G. (1966). "Structural modifications of histones and their possible role in the regulation of ribonucleic acid synthesis." Proc Can Cancer Conf **6**: 313-35.

Allfrey, V. G., R. Faulkner and A. E. Mirsky (1964). "Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis." Proc Natl Acad Sci U S A **51**: 786-94.

Allis, C. D., S. L. Berger, J. Cote, S. Dent, T. Jenuwien, T. Kouzarides, L. Pillus, D. Reinberg, Y. Shi, R. Shiekhattar, for full list of author see journal. (2007). "New nomenclature for chromatin-modifying enzymes." Cell **131**(4): 633-6.

Anderson, G. and E. J. Jenkinson (1995). "The role of the thymus during T-lymphocyte development in vitro." Seminars in Immunology **7**(3): 177-183.

Anderson, G. and E. J. Jenkinson (2001). "Lymphostromal interactions in thymic development and function." Nat Rev Immunol **1**(1): 31-40.

Anderson, G. and E. J. Jenkinson (2008). "Bringing the Thymus to the Bench." J Immunol **181**(11): 7435-7436.

Anderson, G., E. J. Jenkinson, N. C. Moore and J. J. T. Owen (1993). "MHC class II-positive epithelium and mesenchyme cells are both required for T-cell development in the thymus." Nature **362**(6415): 70-73.

Anderson, G., P. J. Lane and E. J. Jenkinson (2007). "Generating intrathymic microenvironments to establish T-cell tolerance." Nat Rev Immunol **7**(12): 954-63.

Anderson, M. S., E. S. Venanzi, Z. Chen, S. P. Berzins, C. Benoist and D. Mathis (2005). "The cellular mechanism of Aire control of T cell tolerance." Immunity **23**(2): 227-39.

Anderson, M. S., E. S. Venanzi, L. Klein, Z. Chen, S. P. Berzins, S. J. Turley, H. von Boehmer, R. Bronson, A. Dierich, C. Benoist, for full list of author see journal. (2002). "Projection of an immunological self shadow within the thymus by the aire protein." Science **298**(5597): 1395-401.

Aschenbrenner, K., L. M. D'Cruz, E. H. Vollmann, M. Hinterberger, J. Emmerich, L. K. Swee, A. Rolink and L. Klein (2007). "Selection of Foxp3+ regulatory T cells specific for self antigen expressed and presented by Aire+ medullary thymic epithelial cells." Nat Immunol **8**(4): 351-8.

Avichezer, D., R. S. Grajewski, C. C. Chan, M. J. Mattapallil, P. B. Silver, J. A. Raber, G. I. Liou, B. Wiggert, G. M. Lewis, L. A. Donoso, for full list of author see journal. (2003). "An immunologically privileged retinal antigen elicits tolerance: major role for central selection mechanisms." J Exp Med **198**(11): 1665-76.

Azuara, V., P. Perry, S. Sauer, M. Spivakov, H. F. Jorgensen, R. M. John, M. Gouti, M. Casanova, G. Warnes, M. Merkenschlager, for full list of author see journal. (2006). "Chromatin signatures of pluripotent cell lines." Nat Cell Biol **8**(5): 532-8.

Bannister, A. J. and T. Kouzarides (1996). "The CBP co-activator is a histone acetyltransferase." Nature **384**(6610): 641-3.

Bannister, A. J., R. Schneider, F. A. Myers, A. W. Thorne, C. Crane-Robinson and T. Kouzarides (2005). "Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes." J Biol Chem **280**(18): 17732-6.

Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev and K. Zhao (2007). "High-resolution profiling of histone methylations in the human genome." Cell **129**(4): 823-37.

Bassett, A., S. Cooper, C. Wu and A. Travers (2009). "The folding and unfolding of eukaryotic chromatin." Curr Opin Genet Dev **19**(2): 159-65.

Beck, G. and G. S. Habicht (1996). "Immunity and the invertebrates." Sci Am **275**(5): 60-3, 66.

Berger, S. L. (2007). "The complex language of chromatin regulation during transcription." Nature **447**(7143): 407-12.

Bernstein, B. E., M. Kamal, K. Lindblad-Toh, S. Bekiranov, D. K. Bailey, D. J. Huebert, S. McMahon, E. K. Karlsson, E. J. Kulbokas, 3rd, T. R. Gingeras, for full list of author see journal. (2005). "Genomic maps and comparative analysis of histone modifications in human and mouse." Cell **120**(2): 169-81.

Bernstein, B. E., A. Meissner and E. S. Lander (2007). "The mammalian epigenome." Cell **128**(4): 669-81.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, for full list of author see journal. (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-26.

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

Blackburn, C. C., C. L. Augustine, R. Li, R. P. Harvey, M. A. Malin, R. L. Boyd, J. F. Miller and G. Morahan (1996). "The nu gene acts cell-autonomously and is required for differentiation of thymic epithelial progenitors." Proc Natl Acad Sci U S A **93**(12): 5742-6.

Bondarenko, V. A., L. M. Steele, A. Ujvari, D. A. Gaykalova, O. I. Kulaeva, Y. S. Polikanov, D. S. Luse and V. M. Studitsky (2006). "Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II." Mol Cell **24**(3): 469-79.

Bottomley, M. J., G. Stier, D. Pennacchini, G. Legube, B. Simon, A. Akhtar, M. Sattler and G. Musco (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.

Boutanaev, A. M., A. I. Kalmykova, Y. Y. Shevelyov and D. I. Nurminsky (2002). "Large clusters of co-expressed genes in the Drosophila genome." Nature **420**(6916): 666-9.

Boyden, L. M., J. M. Lewis, S. D. Barbee, A. Bas, M. Girardi, A. C. Hayday, R. E. Tigelaar and R. P. Lifton (2008). "Skint1, the prototype of a newly identified immunoglobulin superfamily gene cluster, positively selects epidermal [gamma][delta] T cells." Nat Genet **40**(5): 656-662.

Boyer, L. A., K. Plath, J. Zeitlinger, T. Brambrink, L. A. Medeiros, T. I. Lee, S. S. Levine, M. Wernig, A. Tajonar, M. K. Ray, for full list of author see journal. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature **441**(7091): 349-53.

Breiling, A., L. P. O'Neill, D. D'Eliseo, B. M. Turner and V. Orlando (2004). "Epigenome changes in active and inactive polycomb-group-controlled regions." EMBO Rep **5**(10): 976-82.

Brownell, J. E., J. Zhou, T. Ranalli, R. Kobayashi, D. G. Edmondson, S. Y. Roth and C. D. Allis (1996). "Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation." Cell **84**(6): 843-51.

Cao, R. and Y. Zhang (2004). "The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3." Curr Opin Genet Dev **14**(2): 155-64.

Caron, H., B. van Schaik, M. van der Mee, F. Baas, G. Riggins, P. van Sluis, M. C. Hermus, R. van Asperen, K. Boon, P. A. Voute, for full list of author see journal. (2001). "The human transcriptome map: clustering of highly expressed genes in chromosomal domains." Science **291**(5507): 1289-92.

Carrozza, M. J., B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W. J. Shia, S. Anderson, J. Yates, M. P. Washburn, for full list of author see journal. (2005). "Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription." Cell **123**(4): 581-92.

Carrozza, M. J., R. T. Utley, J. L. Workman and J. Cote (2003). "The diverse functions of histone acetyltransferase complexes." Trends Genet **19**(6): 321-9.

Cavadini, P., W. Vermi, F. Facchetti, S. Fontana, S. Nagafuchi, E. Mazzolari, A. Sediva, V. Marrella, A. Villa, A. Fischer, for full list of author see journal. (2005). "AIRE deficiency in thymus of 2 patients with Omenn syndrome." J Clin Invest **115**(3): 728-32.

Chang, B., Y. Chen, Y. Zhao and R. K. Bruick (2007). "JMJD6 is a histone arginine demethylase." Science **318**(5849): 444-7.

Chen, C. C., J. J. Carson, J. Feser, B. Tamburini, S. Zabaronick, J. Linger and J. K. Tyler (2008). "Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair." Cell **134**(2): 231-43.

Chen, L., S. Xiao and N. R. Manley (2008). "Foxn1 is required to maintain the postnatal thymic microenvironment in a dosage-sensitive manner." Blood: blood-2008-05-156265.

Chen, L., S. Xiao and N. R. Manley (2009). "Foxn1 is required to maintain the postnatal thymic microenvironment in a dosage-sensitive manner." Blood **113**(3): 567-74.

Chidgey, A., J. Dudakov, N. Seach and R. Boyd (2007). "Impact of niche aging on thymic regeneration and immune reconstitution." Seminars in Immunology **19**(5): 331-340.

Chignola, F., M. Gaetani, A. Rebane, T. Org, L. Mollica, C. Zucchelli, A. Spitaleri, V. Mannella, P. Peterson and G. Musco (2009). "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 **37**(9): 2951-61.

Cloos, P. A., J. Christensen, K. Agger and K. Helin (2008). "Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease." Genes Dev **22**(9): 1115-40.

Cui, K., C. Zang, T. Y. Roh, D. E. Schones, R. W. Childs, W. Peng and K. Zhao (2009). "Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation." Cell Stem Cell **4**(1): 80-93.

Davie, J. R. (1997). "Nuclear matrix, dynamic histone acetylation and transcriptionally active chromatin." Mol Biol Rep **24**(3): 197-207.

de Ruijter, A. J., A. H. van Gennip, H. N. Caron, S. Kemp and A. B. van Kuilenburg (2003). "Histone deacetylases (HDACs): characterization of the classical HDAC family." Biochem J **370**(Pt 3): 737-49.

Dellino, G. I., Y. B. Schwartz, G. Farkas, D. McCabe, S. C. Elgin and V. Pirrotta (2004). "Polycomb silencing blocks transcription initiation." Mol Cell **13**(6): 887-93.

Deplus, R., C. Brenner, W. A. Burgers, P. Putmans, T. Kouzarides, Y. de Launoit and F. Fuks (2002). "Dnmt3L is a transcriptional repressor that recruits histone deacetylase." Nucleic Acids Res **30**(17): 3831-8.

Derbinski, J., J. Gabler, B. Brors, S. Tierling, S. Jonnakuty, M. Hergenhahn, L. Peltonen, J. Walter and B. Kyewski (2005). "Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels." J Exp Med **202**(1): 33-45.

Derbinski, J., S. Pinto, S. Rosch, K. Hexel and B. Kyewski (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.

Derbinski, J., A. Schulte, B. Kyewski and L. Klein (2001). "Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self." Nat Immunol **2**(11): 1032-9.

DeVoss, J., Y. Hou, K. Johannes, W. Lu, G. I. Liou, J. Rinn, H. Chang, R. R. Caspi, L. Fong and M. S. Anderson (2006). "Spontaneous autoimmunity prevented by thymic expression of a single self-antigen." J Exp Med **203**(12): 2727-35.

Devoss, J. J. and M. S. Anderson (2007). "Lessons on immune tolerance from the monogenic disease APS1." Curr Opin Genet Dev **17**(3): 193-200.

Dindot, S. V., R. Person, M. Strivens, R. Garcia and A. L. Beaudet (2009). "Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions." Genome Res **19**(8): 1374-83.

Dokmanovic, M., C. Clarke and P. A. Marks (2007). "Histone deacetylase inhibitors: overview and perspectives." Mol Cancer Res **5**(10): 981-9.

Dostie, J., Y. Zhan and J. Dekker (2007). "Chromosome conformation capture carbon copy technology." Curr Protoc Mol Biol **Chapter 21**: Unit 21 14.

Farr, A. G., J. L. Dooley and M. Erickson (2002). "Organization of thymic medullary epithelial heterogeneity: implications for mechanisms of epithelial differentiation." Immunol Rev **189**: 20-7.

Farr, A. G. and A. Rudensky (1998). "Medullary thymic epithelium: a mosaic of epithelial "self"?" J Exp Med **188**(1): 1-4.

Felsenfeld, G. and M. Groudine (2003). "Controlling the double helix." Nature **421**(6921): 448-53.

Ferguson, B. J., C. E. Alexander, S. W. Rossi, I. Liiv, A. Rebane, C. L. Worth, J. Wong, M. Laan, P. Peterson, E. J. Jenkinson, for full list of author see journal. (2007). "AIRE's card revealed; A new structure for central tolerance provokes transcriptional plasticity." J. Biol. Chem.: M707211200.

Ferguson, B. J., A. Cooke, P. Peterson and T. Rich (2008). "Death in the AIRE." Trends in Immunology **29**(7): 306-312.

Fischle, W., B. S. Tseng, H. L. Dormann, B. M. Ueberheide, B. A. Garcia, J. Shabanowitz, D. F. Hunt, H. Funabiki and C. D. Allis (2005). "Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation." Nature **438**(7071): 1116-22.

Fraser, P. and W. Bickmore (2007). "Nuclear organization of the genome and the potential for gene regulation." Nature **447**(7143): 413-7.

Freitag, M. and E. U. Selker (2005). "Controlling DNA methylation: many roads to one modification." Curr Opin Genet Dev **15**(2): 191-9.

Fuchs, S. M., R. N. Laribee and B. D. Strahl (2009). "Protein modifications in transcription elongation." Biochim Biophys Acta **1789**(1): 26-36.

Gallinari, P., S. Di Marco, P. Jones, M. Pallaoro and C. Steinkuhler (2007). "HDACs, histone deacetylation and gene transcription: from molecular biology to cancer therapeutics." Cell Res **17**(3): 195-211.

Gardner, J. M., J. J. DeVoss, R. S. Friedman, D. J. Wong, Y. X. Tan, X. Zhou, K. P. Johannes, M. A. Su, H. Y. Chang, M. F. Krummel, for full list of author see journal. (2008). "Deletional Tolerance Mediated by Extrathymic Aire-Expressing Cells." Science **321**(5890): 843-847.

Gavanescu, I., B. Kessler, H. Ploegh, C. Benoist and D. Mathis (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.

Gibson, T. J., C. Ramu, C. Gemund and R. Aasland (1998). "The APECED polyglandular autoimmune syndrome protein, AIRE-1, contains the SAND domain and is probably a transcription factor." Trends Biochem Sci **23**(7): 242-4.

Gierman, H. J., M. H. Indemans, J. Koster, S. Goetze, J. Seppen, D. Geerts, R. van Driel and R. Versteeg (2007). "Domain-wide regulation of gene expression in the human genome." Genome Res **17**(9): 1286-95.

Goldberg, A. D., C. D. Allis and E. Bernstein (2007). "Epigenetics: a landscape takes shape." Cell **128**(4): 635-8.

Goldrath, A. W. and S. M. Hedrick (2005). "Central tolerance matters." Immunity **23**(2): 113-4.

Golebiewska, A., S. P. Atkinson, M. Lako and L. Armstrong (2009). "Epigenetic landscaping during hESC differentiation to neural cells." Stem Cells **27**(6): 1298-308.

Goll, M. G. and T. H. Bestor (2005). "Eukaryotic cytosine methyltransferases." Annu Rev Biochem **74**: 481-514.

Gotter, J., B. Brors, M. Hergenhahn and B. Kyewski (2004). "Medullary epithelial cells of the human thymus express a highly diverse selection of tissue-specific genes colocalized in chromosomal clusters." J Exp Med **199**(2): 155-66.

Gotter, J. and B. Kyewski (2004). "Regulating self-tolerance by deregulating gene expression." Curr Opin Immunol **16**(6): 741-5.

Gray, D., J. Abramson, C. Benoist and D. Mathis (2007). "Proliferative arrest and rapid turnover of thymic epithelial cells expressing Aire." J. Exp. Med. **204**(11): 2521-2528.

Grimaud, C., F. Bantignies, M. Pal-Bhadra, P. Ghana, U. Bhadra and G. Cavalli (2006). "RNAi components are required for nuclear clustering of Polycomb group response elements." Cell **124**(5): 957-71.

Guccione, E., C. Bassi, F. Casadio, F. Martinato, M. Cesaroni, H. Schuchlautz, B. Luscher and B. Amati (2007). "Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive." Nature **449**(7164): 933-937.

Guerau-de-Arellano, M., D. Mathis and C. Benoist (2008). "Transcriptional impact of Aire varies with cell type." Proceedings of the National Academy of Sciences: -.

Haberland, M., R. L. Montgomery and E. N. Olson (2009). "The many roles of histone deacetylases in development and physiology: implications for disease and therapy." Nat Rev Genet **10**(1): 32-42.

Haigis, M. C. and L. P. Guarente (2006). "Mammalian sirtuins--emerging roles in physiology, aging, and calorie restriction." Genes Dev **20**(21): 2913-21.

Halonen, M., H. Kangas, T. Ruppell, T. Ilmarinen, J. Ollila, M. Kolmer, M. Vihinen, J. Palvimo, J. Saarela, I. Ulmanen, for full list of author see journal. (2004). "APECED-causing mutations in AIRE reveal the functional domains of the protein." Hum Mutat **23**(3): 245-57.

Halonen, M., M. Pelto-Huikko, P. Eskelin, L. Peltonen, I. Ulmanen and M. Kolmer (2001). "Subcellular location and expression pattern of autoimmune regulator

(Aire), the mouse orthologue for human gene defective in autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED)." J Histochem Cytochem **49**(2): 197-208.

Hamazaki, Y., H. Fujita, T. Kobayashi, Y. Choi, H. S. Scott, M. Matsumoto and N. Minato (2007). "Medullary thymic epithelial cells expressing Aire represent a unique lineage derived from cells expressing claudin." Nat Immunol **8**(3): 304-11.

Heino, M., P. Peterson, J. Kudoh, K. Nagamine, A. Lagerstedt, V. Ovod, A. Ranki, I. Rantala, M. Nieminen, J. Tuukkanen, for full list of author see journal. (1999). "Autoimmune regulator is expressed in the cells regulating immune tolerance in thymus medulla." Biochem Biophys Res Commun **257**(3): 821-5.

Heino, M., P. Peterson, J. Kudoh, N. Shimizu, S. E. Antonarakis, H. S. Scott and K. Krohn (2001). "APECED mutations in the autoimmune regulator (AIRE) gene." Hum Mutat **18**(3): 205-11.

Henikoff, S. (2008). "Nucleosome destabilization in the epigenetic regulation of gene expression." Nat Rev Genet **9**(1): 15-26.

Hollander, G. A., B. Wang, A. Nichogiannopoulou, P. P. Platenburg, W. van Ewijk, S. J. Burakoff, J. C. Gutierrez-Ramos and C. Terhorst (1995). "Developmental control point in induction of thymic cortex regulated by a subpopulation of prothymocytes." Nature **373**(6512): 350-3.

Huang, Y., J. Fang, M. T. Bedford, Y. Zhang and R. M. Xu (2006). "Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A." Science **312**(5774): 748-51.

Hubert, F.-X., S. A. Kinkel, K. E. Webster, P. Cannon, P. E. Crewther, A. I. Proeitto, L. Wu, W. R. Heath and H. S. Scott (2008). "A Specific Anti-Aire Antibody Reveals Aire Expression Is Restricted to Medullary Thymic Epithelial Cells and Not Expressed in Periphery." J Immunol **180**(6): 3824-3832.

Husebye, E. S., J. Perheentupa, R. Rautemaa and O. Kampe (2009). "Clinical manifestations and management of patients with autoimmune polyendocrine syndrome type I." J Intern Med **265**(5): 514-29.

Iborra, F. J., A. Pombo, D. A. Jackson and P. R. Cook (1996). "Active RNA polymerases are localized within discrete transcription "factories' in human nuclei." J Cell Sci **109 (Pt 6)**: 1427-36.

Ilmarinen, T., H. Kangas, T. Kytomaa, P. Eskelin, J. Saharinen, J. S. Seeler, K. Tanhuanpaa, F. Y. Chan, R. M. Slattery, K. Alakurtti, for full list of author see

journal. (2008). "Functional interaction of AIRE with PIAS1 in transcriptional regulation." Mol Immunol **45**(7): 1847-62.

Ilmarinen, T., K. Melen, H. Kangas, I. Julkunen, I. Ulmanen and P. Eskelin (2006). "The monopartite nuclear localization signal of autoimmune regulator mediates its nuclear import and interaction with multiple importin alpha molecules." Febs J **273**(2): 315-24.

Iwasaki, H. and K. Akashi (2006). "Thymus Exclusivity: All the Right Conditions for T Cells." Immunity **25**(5): 697-700.

Jackson, D. A., A. B. Hassan, R. J. Errington and P. R. Cook (1993). "Visualization of focal sites of transcription within human nuclei." Embo J **12**(3): 1059-65.

Jenkinson, E. J. and G. Anderson (1994). "Fetal thymic organ cultures." Current Opinion in Immunology **6**(2): 293-297.

Jenkinson, E. J., G. Anderson and J. J. Owen (1992). "Studies on T cell maturation on defined thymic stromal cell populations in vitro." J Exp Med **176**(3): 845-53.

Jenkinson, W. E., A. Bacon, A. J. White, G. Anderson and E. J. Jenkinson (2008). "An Epithelial Progenitor Pool Regulates Thymus Growth." J Immunol **181**(9): 6101-6108.

Jenkinson, W. E., S. W. Rossi, E. J. Jenkinson and G. Anderson (2005). "Development of functional thymic epithelial cells occurs independently of lymphostromal interactions." Mechanisms of Development **122**(12): 1294-1299.

Jenkinson, W. E., S. W. Rossi, S. M. Parnell, E. J. Jenkinson and G. Anderson (2007). "PDGFR{alpha}-expressing mesenchyme regulates thymus growth and the availability of intrathymic niches." Blood **109**(3): 954-960.

Jiang, C. and B. F. Pugh (2009). "Nucleosome positioning and gene regulation: advances through genomics." Nat Rev Genet **10**(3): 161-172.

Johnnidis, J. B., E. S. Venanzi, D. J. Taxman, J. P. Ting, C. O. Benoist and D. J. Mathis (2005). "Chromosomal clustering of genes controlled by the aire transcription factor." Proc Natl Acad Sci U S A **102**(20): 7233-8.

Johnson, C. A., L. P. O'Neill, A. Mitchell and B. M. Turner (1998). "Distinctive patterns of histone H4 acetylation are associated with defined sequence elements within both heterochromatic and euchromatic regions of the human genome." Nucl. Acids Res. **26**(4): 994-1001.

Jolicoeur, C., D. Hanahan and K. M. Smith (1994). "T-cell tolerance toward a transgenic beta-cell antigen and transcription of endogenous pancreatic genes in thymus." Proc Natl Acad Sci U S A **91**(14): 6707-11.

Joshi, A. A. and K. Struhl (2005). "Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation." Mol Cell **20**(6): 971-8.

Keogh, M. C., S. K. Kurdistani, S. A. Morris, S. H. Ahn, V. Podolny, S. R. Collins, M. Schuldiner, K. Chin, T. Punna, N. J. Thompson, for full list of author see journal. (2005). "Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex." Cell **123**(4): 593-605.

Kirmizis, A., H. Santos-Rosa, C. J. Penkett, M. A. Singer, M. Vermeulen, M. Mann, J. Bahler, R. D. Green and T. Kouzarides (2007). "Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation." Nature **449**(7164): 928-932.

Klamp, T., U. Sahin, B. Kyewski, J. Schwendemann, K. Dhaene and O. Tureci (2006). "Expression profiling of autoimmune regulator AIRE mRNA in a comprehensive set of human normal and neoplastic tissues." Immunol Lett **106**(2): 172-9.

Klein, L., T. Klein, U. Ruther and B. Kyewski (1998). "CD4 T cell tolerance to human C-reactive protein, an inducible serum protein, is mediated by medullary thymic epithelium." J Exp Med **188**(1): 5-16.

Klein, L., M. Klugmann, K. A. Nave, V. K. Tuohy and B. Kyewski (2000). "Shaping of the autoreactive T-cell repertoire by a splice variant of self protein expressed in thymic epithelial cells." Nat Med **6**(1): 56-61.

Klug, D. B., C. Carter, E. Crouch, D. Roop, C. J. Conti and E. R. Richie (1998). "Interdependence of cortical thymic epithelial cell differentiation and T-lineage commitment." Proc Natl Acad Sci U S A **95**(20): 11822-7.

Klug, D. B., C. Carter, I. B. Gimenez-Conti and E. R. Richie (2002). "Cutting Edge: Thymocyte-Independent and Thymocyte-Dependent Phases of Epithelial Patterning in the Fetal Thymus." J Immunol **169**(6): 2842-2845.

Koble, C. and B. Kyewski (2009). "The thymic medulla: a unique microenvironment for intercellular self-antigen transfer." J Exp Med **206**(7): 1505-13.

Koh, A. S., A. J. Kuo, S. Y. Park, P. Cheung, J. Abramson, D. Bua, D. Carney, S. E. Shoelson, O. Gozani, R. E. Kingston, for full list of author see journal. (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.

Kondo, Y., L. Shen, A. S. Cheng, S. Ahmed, Y. Boumber, C. Charo, T. Yamochi, T. Urano, K. Furukawa, B. Kwabi-Addo, for full list of author see journal. (2008). "Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation." Nat Genet **40**(6): 741-750.

Konishi, A., S. Shimizu, J. Hirota, T. Takao, Y. Fan, Y. Matsuoka, L. Zhang, Y. Yoneda, Y. Fujii, A. I. Skoultchi, for full list of author see journal. (2003). "Involvement of histone H1.2 in apoptosis induced by DNA double-strand breaks." Cell **114**(6): 673-88.

Kont, V., M. Laan, K. Kisand, A. Merits, H. S. Scott and P. Peterson (2008). "Modulation of Aire regulates the expression of tissue-restricted antigens." Mol Immunol **45**(1): 25-33.

Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.

Krajewski, W. A. and P. B. Becker (1998). "Reconstitution of hyperacetylated, DNase I-sensitive chromatin characterized by high conformational flexibility of nucleosomal DNA." Proc Natl Acad Sci U S A **95**(4): 1540-5.

Kumar, P. G., M. Laloraya, C. Y. Wang, Q. G. Ruan, A. Davoodi-Semiromi, K. J. Kao and J. X. She (2001). "The autoimmune regulator (AIRE) is a DNA-binding protein." J Biol Chem **276**(44): 41357-64.

Kuroda, N., T. Mitani, N. Takeda, N. Ishimaru, R. Arakaki, Y. Hayashi, Y. Bando, K. Izumi, T. Takahashi, T. Nomura, for full list of author see journal. (2005). "Development of autoimmunity against transcriptionally unrepressed target antigen in the thymus of Aire-deficient mice." J Immunol **174**(4): 1862-70.

Kurz, A., S. Lampel, J. E. Nickolenko, J. Bradl, A. Benner, R. M. Zirbel, T. Cremer and P. Lichter (1996). "Active and inactive genes localize preferentially in the periphery of chromosome territories." J Cell Biol **135**(5): 1195-205.

Kyewski, B. (2008). "IMMUNOLOGY: A Breath of Aire for the Periphery." Science **321**(5890): 776-777.

Kyewski, B. and J. Derbinski (2004). "Self-representation in the thymus: an extended view." Nat Rev Immunol **4**(9): 688-98.

Kyewski, B., J. Derbinski, J. Gotter and L. Klein (2002). "Promiscuous gene expression and central T-cell tolerance: more than meets the eye." Trends Immunol **23**(7): 364-71.

Kyewski, B. and L. Klein (2006). "A Central Role for Central Tolerance." Annual Review of Immunology **24**(1): 571-606.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-5.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, for full list of author see journal. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lars-Oliver Tykocinski, A. S. B. K. (2008). "The Thymus Medulla Slowly Yields Its Secrets." Annals of the New York Academy of Sciences **1143**(The Year in Immunology 2008): 105-122.

Latham, J. A. and S. Y. R. Dent (2007). "Cross-regulation of histone modifications." Nat Struct Mol Biol **14**(11): 1017-1024.

Lee, J. W., M. Epardaud, J. Sun, J. E. Becker, A. C. Cheng, A. R. Yonekura, J. K. Heath and S. J. Turley (2007). "Peripheral antigen display by lymph node stroma promotes T cell tolerance to intestinal self." Nat Immunol **8**(2): 181-90.

Lee, T. I., R. G. Jenner, L. A. Boyer, M. G. Guenther, S. S. Levine, R. M. Kumar, B. Chevalier, S. E. Johnstone, M. F. Cole, K. Isono, for full list of author see journal. (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell **125**(2): 301-13.

Li, B., M. Carey and J. L. Workman (2007). "The role of chromatin during transcription." Cell **128**(4): 707-19.

Li, S. and M. A. Shogren-Knaak (2008). "Cross-talk between histone H3 tails produces cooperative nucleosome acetylation." Proc Natl Acad Sci U S A **105**(47): 18243-8.

Liiv, I., A. Rebane, T. Org, M. Saare, J. Maslovskaja, K. Kisand, E. Juronen, L. Valmu, M. J. Bottomley, N. Kalkkinen, for full list of author see journal. (2008). "DNA-PK contributes to the phosphorylation of AIRE: importance in transcriptional activity." Biochim Biophys Acta **1783**(1): 74-83.

Lindh, E., S. M. Lind, E. Lindmark, S. HÃ¤ssler, J. Perheentupa, L. Peltonen, O. Winqvist and M. C. I. Karlsson (2008). "AIRE regulates T-cell-independent B-

cell responses through BAFF." Proceedings of the National Academy of Sciences **105**(47): 18466-18471.

Liston, A. (2006). "There and back again: Autoimmune Polyendocrinopathy Syndrome Type I and the Aire knockout mouse." Drug Discovery Today: Disease Models **3**(1): 33-40.

Liston, A., D. H. Gray, S. Lesage, A. L. Fletcher, J. Wilson, K. E. Webster, H. S. Scott, R. L. Boyd, L. Peltonen and C. C. Goodnow (2004). "Gene dosage--limiting role of Aire in thymic expression, clonal deletion, and organ-specific autoimmunity." J Exp Med **200**(8): 1015-26.

Liston, A., S. Lesage, J. Wilson, L. Peltonen and C. C. Goodnow (2003). "Aire regulates negative selection of organ-specific T cells." Nat Immunol **4**(4): 350-4.

Liu, C. L., T. Kaplan, M. Kim, S. Buratowski, S. L. Schreiber, N. Friedman and O. J. Rando (2005). "Single-nucleosome mapping of histone modifications in S. cerevisiae." PLoS Biol **3**(10): e328.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-60.

Maarit Heino, P. P., Niko Sillanpää, Sandrine Guérin, Li Wu, Graham Anderson, Hamish S. Scott, Stylianos E. Antonarakis, Jun Kudoh, Nobuyoshi Shimizu, Eric J. Jenkinson, Philippe Naquet, Kai J. E. Krohn (2000). "RNA and protein expression of the murine autoimmune regulator gene (Aire) in normal, RelB-deficient and in NOD mouse." European Journal of Immunology **30**(7): 1884-1893.

Mahy, N. L., P. E. Perry, S. Gilchrist, R. A. Baldock and W. A. Bickmore (2002). "Spatial organization of active and inactive genes and noncoding DNA within chromosome territories." J Cell Biol **157**(4): 579-89.

Manley, N. R. (2000). "Thymus organogenesis and molecular mechanisms of thymic epithelial cell differentiation." Seminars in Immunology **12**(5): 421-428.

Marks, H., J. C. Chow, S. Denissov, K. J. Francoijs, N. Brockdorff, E. Heard and H. G. Stunnenberg (2009). "High-resolution analysis of epigenetic changes associated with X inactivation." Genome Res **19**(8): 1361-73.

Masumoto, H., D. Hawke, R. Kobayashi and A. Verreault (2005). "A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response." Nature **436**(7048): 294-8.

Mathis, D. and C. Benoist (2004). "Back to central tolerance." Immunity **20**(5): 509-16.

Mathis, D. and C. Benoist (2009). "Aire." Annu Rev Immunol **27**: 287-312.

Matsumoto, M. (2009). "The role of autoimmune regulator (Aire) in the development of the immune system." Microbes Infect **11**(12): 928-34.

Mavrich, T. N., I. P. Ioshikhes, B. J. Venters, C. Jiang, L. P. Tomsho, J. Qi, S. C. Schuster, I. Albert and B. F. Pugh (2008). "A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome." Genome Res **18**(7): 1073-83.

McCaughtry, T. M., T. A. Baldwin, M. S. Wilken and K. A. Hogquist (2008). "Clonal deletion of thymocytes can occur in the cortex with no involvement of the medulla." J. Exp. Med. **205**(11): 2575-2584.

Meloni, A., F. Incani, D. Corda, A. Cao and M. C. Rosatelli (2008). "Role of PHD fingers and COOH-terminal 30 amino acids in AIRE transactivation activity." Mol Immunol **45**(3): 805-9.

Michishita, E., J. Y. Park, J. M. Burneskis, J. C. Barrett and I. Horikawa (2005). "Evolutionarily conserved and nonconserved cellular localizations and functions of human SIRT proteins." Mol Biol Cell **16**(10): 4623-35.

Motamedi, M. R., E.-J. E. Hong, X. Li, S. Gerber, C. Denison, S. Gygi and D. Moazed (2008). "HP1 Proteins Form Distinct Complexes and Mediate Heterochromatic Gene Silencing by Nonoverlapping Mechanisms." Molecular Cell **32**(6): 778-790.

Murata, S., K. Sasaki, T. Kishimoto, S.-i. Niwa, H. Hayashi, Y. Takahama and K. Tanaka (2007). "Regulation of CD8+ T Cell Development by Thymus-Specific Proteasomes." Science **316**(5829): 1349-1353.

Nagamine, K., P. Peterson, H. S. Scott, J. Kudoh, S. Minoshima, M. Heino, K. J. Krohn, M. D. Lalioti, P. E. Mullis, S. E. Antonarakis, for full list of author see journal. (1997). "Positional cloning of the APECED gene." Nat Genet **17**(4): 393-8.

Nagy, Z. and L. Tora (2007). "Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation." Oncogene **26**(37): 5341-57.

Nedjic, J., M. Aichinger, J. Emmerich, N. Mizushima and L. Klein (2008). "Autophagy in thymic epithelium shapes the T-cell repertoire and is essential for tolerance." Nature **455**(7211): 396-400.

Nehls, M., B. Kyewski, M. Messerle, R. Waldschutz, K. Schuddekopf, A. J. H. Smith and T. Boehm (1996). "Two Genetically Separable Steps in the Differentiation of Thymic Epithelium." Science **272**(5263): 886-889.

Ng, S. S., W. W. Yue, U. Oppermann and R. J. Klose (2009). "Dynamic protein methylation in chromatin biology." Cell Mol Life Sci **66**(3): 407-22.

Nightingale, K. P., L. P. O'Neill and B. M. Turner (2006). "Histone modifications: signalling receptors and potential elements of a heritable epigenetic code." Curr Opin Genet Dev **16**(2): 125-36.

Noma, K., C. D. Allis and S. I. Grewal (2001). "Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries." Science **293**(5532): 1150-5.

O'Neill, L. P., A. M. Keohane, J. S. Lavender, V. McCabe, E. Heard, P. Avner, N. Brockdorff and B. M. Turner (1999). "A developmental switch in H4 acetylation upstream of Xist plays a role in X chromosome inactivation." Embo J **18**(10): 2897-907.

O'Neill, L. P., T. E. Randall, J. Lavender, H. T. Spotswood, J. T. Lee and B. M. Turner (2003). "X-linked genes in female embryonic stem cells carry an epigenetic mark prior to the onset of X inactivation." Hum Mol Genet **12**(15): 1783-90.

O'Neill, L. P. and B. M. Turner (1995). "Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner." Embo J **14**(16): 3946-57.

O'Neill, L. P. and B. M. Turner (2003). "Immunoprecipitation of native chromatin: NChIP." Methods **31**(1): 76-82.

O'Neill, L. P., M. D. VerMilyea and B. M. Turner (2006). "Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations." Nat Genet **38**(7): 835-41.

Okamoto, I., A. P. Otte, C. D. Allis, D. Reinberg and E. Heard (2004). "Epigenetic dynamics of imprinted X inactivation during early mouse development." Science **303**(5658): 644-9.

Orford, K., P. Kharchenko, W. Lai, M. C. Dao, D. J. Worhunsky, A. Ferro, V. Janzen, P. J. Park and D. T. Scadden (2008). "Differential H3K4 Methylation Identifies Developmentally Poised Hematopoietic Genes." Developmental Cell **14**(5): 798-809.

Org, T., F. Chignola, C. Hetenyi, M. Gaetani, A. Rebane, I. Liiv, U. Maran, L. Mollica, M. J. Bottomley, G. Musco, for full list of author see journal. (2008). "The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression." EMBO Rep **9**(4): 370-6.

Org, T., A. Rebane, K. Kisand, M. Laan, U. Haljasorg, R. Andreson and P. Peterson (2009). "AIRE activated tissue specific genes have histone modifications associated with inactive chromatin." Hum Mol Genet.

Orlando, V. and R. Paro (1993). "Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin." Cell **75**(6): 1187-98.

Osborne, C. S., L. Chakalova, K. E. Brown, D. Carter, A. Horton, E. Debrand, B. Goyenechea, J. A. Mitchell, S. Lopes, W. Reik, for full list of author see journal. (2004). "Active genes dynamically colocalize to shared sites of ongoing transcription." Nat Genet **36**(10): 1065-71.

Osborne, C. S., L. Chakalova, J. A. Mitchell, A. Horton, A. L. Wood, D. J. Bolland, A. E. Corcoran and P. Fraser (2007). "Myc dynamically and preferentially relocates to a transcription factory occupied by Igh." PLoS Biol **5**(8): e192.

Oven, I., N. Brdickova, J. Kohoutek, T. Vaupotic, M. Narat and B. M. Peterlin (2007). "AIRE Recruits P-TEFb for Transcriptional Elongation of Target Genes in Medullary Thymic Epithelial Cells." Mol. Cell. Biol. **27**(24): 8815-8823.

Pahlich, S., R. P. Zakaryan and H. Gehring (2006). "Protein arginine methylation: Cellular functions and methods of analysis." Biochim Biophys Acta **1764**(12): 1890-903.

Palmer, E. and D. Naeher (2009). "Affinity threshold for thymic selection through a T-cell receptor-co-receptor zipper." Nat Rev Immunol **9**(3): 207-213.

Parada, L. and T. Misteli (2002). "Chromosome positioning in the interphase nucleus." Trends Cell Biol **12**(9): 425-32.

Pennings, S., G. Meersseman and E. M. Bradbury (1991). "Mobility of positioned nucleosomes on 5 S rDNA." J Mol Biol **220**(1): 101-10.

Peterson, P., T. Org and A. Rebane (2008). "Transcriptional regulation by AIRE: molecular mechanisms of central tolerance." Nat Rev Immunol **8**(12): 948-57.

Pfaffl, M. W. (2001). "A new mathematical model for relative quantification in real-time RT-PCR." Nucleic Acids Res **29**(9): e45.

Pitkanen, J., V. Doucas, T. Sternsdorf, T. Nakajima, S. Aratani, K. Jensen, H. Will, P. Vahamurto, J. Ollila, M. Vihinen, for full list of author see journal. (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.

Pitkanen, J., A. Rebane, J. Rowell, A. Murumagi, P. Strobel, K. Moll, M. Saare, J. Heikkila, V. Doucas, A. Marx, for full list of author see journal. (2005). "Cooperative activation of transcription by autoimmune regulator AIRE and CBP." Biochem Biophys Res Commun **333**(3): 944-53.

Pitkanen, J., P. Vahamurto, K. Krohn and P. Peterson (2001). "Subcellular localization of the autoimmune regulator protein. characterization of nuclear targeting and transcriptional activation domain." J Biol Chem **276**(22): 19597-602.

Pombo, A., D. A. Jackson, M. Hollinshead, Z. Wang, R. G. Roeder and P. R. Cook (1999). "Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III." Embo J **18**(8): 2241-53.

Pontynen, N., M. Strengell, N. Sillanpaa, J. Saharinen, I. Ulmanen, I. Julkunen and L. Peltonen (2008). "Critical immunological pathways are downregulated in APECED patient dendritic cells." J Mol Med **86**(10): 1139-52.

Probst, A. V., E. Dunleavy and G. Almouzni (2009). "Epigenetic inheritance during the cell cycle." Nat Rev Mol Cell Biol **10**(3): 192-206.

Purohit, S., P. G. Kumar, M. Laloraya and J. X. She (2005). "Mapping DNA-binding domains of the autoimmune regulator protein." Biochem Biophys Res Commun **327**(3): 939-44.

Reik, W. and J. Walter (2001). "Genomic imprinting: parental influence on the genome." Nat Rev Genet **2**(1): 21-32.

Richards, E. J. and S. C. Elgin (2002). "Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects." Cell **108**(4): 489-500.

Rodewald, H. R. (2006). "Immunology: a second chance for the thymus." Nature **441**(7096): 942-3.

Roh, T. Y., S. Cuddapah and K. Zhao (2005). "Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping." Genes Dev **19**(5): 542-52.

Rossi, S. W., W. E. Jenkinson, G. Anderson and E. J. Jenkinson (2006). "Clonal analysis reveals a common progenitor for thymic cortical and medullary epithelium." Nature **441**(7096): 988-991.

Rossi, S. W., M.-Y. Kim, A. Leibbrandt, S. M. Parnell, W. E. Jenkinson, S. H. Glanville, F. M. McConnell, H. S. Scott, J. M. Penninger, E. J. Jenkinson, for full list of author see journal. (2007). "RANK signals from CD4+3- inducer cells regulate development of Aire-expressing epithelial cells in the thymic medulla." Journal of Experimental Medicine **204**(6): 1267-72.

Rothenberg, E. V., J. E. Moore and M. A. Yui (2008). "Launching the T-cell-lineage developmental programme." Nat Rev Immunol **8**(1): 9-21.

Roy, P. J., J. M. Stuart, J. Lund and S. K. Kim (2002). "Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans." Nature **418**(6901): 975-9.

Ruan, Q. G., K. Tung, D. Eisenman, Y. Setiady, S. Eckenrode, B. Yi, S. Purohit, W. P. Zheng, Y. Zhang, L. Peltonen, for full list of author see journal. (2007). "The autoimmune regulator directly controls the expression of genes critical for thymic epithelial function." Journal of Immunology **178**(11): 7173-80.

Rubtsov, Y. P. and A. Y. Rudensky (2007). "TGFbeta signalling in control of T-cell-mediated self-reactivity." Nat Rev Immunol **7**(6): 443-53.

Saltis, M., M. F. Criscitiello, Y. Ohta, M. Keefe, N. S. Trede, R. Goitsuka and M. F. Flajnik (2008). "Evolutionarily conserved and divergent regions of the autoimmune regulator (Aire) gene: a comparative analysis." Immunogenetics **60**(2): 105-14.

Santos-Rosa, H., R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. Emre, S. L. Schreiber, J. Mellor and T. Kouzarides (2002). "Active genes are tri-methylated at K4 of histone H3." Nature **419**(6905): 407-11.

Saurin, A. J., C. Shiels, J. Williamson, D. P. Satijn, A. P. Otte, D. Sheer and P. S. Freemont (1998). "The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain." J Cell Biol **142**(4): 887-98.

Schones, D. E., K. Cui, S. Cuddapah, T. Y. Roh, A. Barski, Z. Wang, G. Wei and K. Zhao (2008). "Dynamic regulation of nucleosome positioning in the human genome." Cell **132**(5): 887-98.

Schones, D. E. and K. Zhao (2008). "Genome-wide approaches to studying chromatin modifications." Nat Rev Genet **9**(3): 179-91.

Schuettengruber, B., M. Ganapathi, B. Leblanc, M. Portoso, R. Jaschek, B. Tolhuis, M. van Lohuizen, A. Tanay and G. Cavalli (2009). "Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos." PLoS Biol **7**(1): e13.

Schwartz, Y. B., T. G. Kahn, D. A. Nix, X. Y. Li, R. Bourgon, M. Biggin and V. Pirrotta (2006). "Genome-wide analysis of Polycomb targets in Drosophila melanogaster." Nat Genet **38**(6): 700-5.

Schwartz, Y. B. and V. Pirrotta (2007). "Polycomb silencing mechanisms and the management of genomic programmes." Nat Rev Genet **8**(1): 9-22.

Sexton, T., H. Schober, P. Fraser and S. M. Gasser (2007). "Gene regulation through nuclear organization." Nat Struct Mol Biol **14**(11): 1049-1055.

Shakib, S., G. E. Desanti, W. E. Jenkinson, S. M. Parnell, E. J. Jenkinson and G. Anderson (2009). "Checkpoints in the Development of Thymic Cortical Epithelial Cells." J Immunol **182**(1): 130-137.

Shi, X., T. Hong, K. L. Walter, M. Ewalt, E. Michishita, T. Hung, D. Carney, P. Pena, F. Lan, M. R. Kaadige, for full list of author see journal. (2006). "ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression." Nature **442**(7098): 96-9.

Shilatifard, A. (2008). "Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation." Curr Opin Cell Biol **20**(3): 341-8.

Shogren-Knaak, M., H. Ishii, J. M. Sun, M. J. Pazin, J. R. Davie and C. L. Peterson (2006). "Histone H4-K16 acetylation controls chromatin structure and protein interactions." Science **311**(5762): 844-7.

Siggs, O. M., L. E. Makaroff and A. Liston (2006). "The why and how of thymocyte negative selection." Curr Opin Immunol **18**(2): 175-83.

Sillanpaa, N., C. G. Magureanu, A. Murumagi, A. Reinikainen, A. West, A. Manninen, M. Lahti, A. Ranki, K. Saksela, K. Krohn, for full list of author see journal. (2004). "Autoimmune regulator induced changes in the gene expression

profile of human monocyte-dendritic cell-lineage." Mol Immunol **41**(12): 1185-98.

Simonis, M., P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B. van Steensel and W. de Laat (2006). "Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C)." Nat Genet **38**(11): 1348-54.

Sims, R. J., 3rd, S. Millhouse, C. F. Chen, B. A. Lewis, H. Erdjument-Bromage, P. Tempst, J. L. Manley and D. Reinberg (2007). "Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing." Mol Cell **28**(4): 665-76.

Smith, B. C. and J. M. Denu (2009). "Chemical mechanisms of histone lysine and arginine modifications." Biochim Biophys Acta **1789**(1): 45-57.

Soshnikova, N. and D. Duboule (2009). "Epigenetic temporal control of mouse Hox genes in vivo." Science **324**(5932): 1320-3.

Soutoglou, E. and I. Talianidis (2002). "Coordination of PIC assembly and chromatin remodeling during differentiation-induced gene activation." Science **295**(5561): 1901-4.

Squazzo, S. L., H. O'Geen, V. M. Komashko, S. R. Krig, V. X. Jin, S. W. Jang, R. Margueron, D. Reinberg, R. Green and P. J. Farnham (2006). "Suz12 binds to silenced regions of the genome in a cell-type-specific manner." Genome Res **16**(7): 890-900.

Sterner, D. E. and S. L. Berger (2000). "Acetylation of histones and transcription-related factors." Microbiol Mol Biol Rev **64**(2): 435-59.

Strahl, B. D., P. A. Grant, S. D. Briggs, Z. W. Sun, J. R. Bone, J. A. Caldwell, S. Mollah, R. G. Cook, J. Shabanowitz, D. F. Hunt, for full list of author see journal. (2002). "Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression." Mol Cell Biol **22**(5): 1298-306.

Su, M. A. and M. S. Anderson (2004). "Aire: an update." Curr Opin Immunol **16**(6): 746-52.

Su, M. A., K. Giang, K. Zumer, H. Jiang, I. Oven, J. L. Rinn, J. J. Devoss, K. P. Johannes, W. Lu, J. Gardner, for full list of author see journal. (2008). "Mechanisms of an autoimmunity syndrome in mice caused by a dominant mutation in Aire." J Clin Invest **118**(5): 1712-26.

Suganuma, T. and J. L. Workman (2008). "Crosstalk among Histone Modifications." Cell **135**(4): 604-7.

Takahama, Y. (2006). "Journey through the thymus: stromal guides for T-cell development and selection." Nat Rev Immunol **6**(2): 127-35.

Taunton, J., C. A. Hassig and S. L. Schreiber (1996). "A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p." Science **272**(5260): 408-11.

Thomas, L. R., H. Miyashita, R. M. Cobb, S. Pierce, M. Tachibana, E. Hobeika, M. Reth, Y. Shinkai and E. M. Oltz (2008). "Functional Analysis of Histone Methyltransferase G9a in B and T Lymphocytes." J Immunol **181**(1): 485-493.

Towbin, H., T. Staehelin and J. Gordon (1979). "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications." Proc Natl Acad Sci U S A **76**(9): 4350-4.

Tremethick, D. J. (2007). "Higher-order structures of chromatin: the elusive 30 nm fiber." Cell **128**(4): 651-4.

Turner, B. M. (2007). "Defining an epigenetic code." Nat Cell Biol **9**(1): 2-6.

Turner, B. M. and G. Fellows (1989). "Specific antibodies reveal ordered and cell-cycle-related use of histone-H4 acetylation sites in mammalian cells." Eur J Biochem **179**(1): 131-9.

Turner, B. M., L. P. O'Neill and I. M. Allan (1989). "Histone H4 acetylation in human cells Frequency of acetylation at different sites defined by immunolabeling with site-specific antibodies." FEBS Letters **253**(1-2): 141-145.

Uchida, D., S. Hatakeyama, A. Matsushima, H. Han, S. Ishido, H. Hotta, J. Kudoh, N. Shimizu, V. Doucas, K. I. Nakayama, for full list of author see journal. (2004). "AIRE functions as an E3 ubiquitin ligase." J Exp Med **199**(2): 167-72.

Vakoc, C. R., S. A. Mandat, B. A. Olenchock and G. A. Blobel (2005). "Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin." Mol Cell **19**(3): 381-91.

Vakoc, C. R., M. M. Sachdeva, H. Wang and G. A. Blobel (2006). "Profile of histone lysine methylation across transcribed mammalian chromatin." Mol Cell Biol **26**(24): 9185-95.

Valls, E., S. Sanchez-Molina and M. A. Martinez-Balbas (2005). "Role of histone modifications in marking and activating genes through mitosis." J Biol Chem **280**(52): 42592-600.

van Berkum, N. L. and J. Dekker (2009). "Determining Spatial Chromatin Organization of Large Genomic Regions Using 5C Technology." Methods Mol Biol **567**: 189-213.

van Ingen, H., F. M. A. van Schaik, H. Wienk, J. Ballering, H. Rehmann, A. C. Dechesne, J. A. W. Kruijzer, R. M. J. Liskamp, H. T. M. Timmers and R. Boelens (2008). "Structural Insight into the Recognition of the H3K4me3 Mark by the TFIID Subunit TAF3." Structure **16**(8): 1245-1256.

Venanzi, E. S., C. Benoist and D. Mathis (2004). "Good riddance: Thymocyte clonal deletion prevents autoimmunity." Curr Opin Immunol **16**(2): 197-202.

Venanzi, E. S., D. H. D. Gray, C. Benoist and D. Mathis (2007). "Lymphotoxin Pathway and Aire Influences on Thymic Medullary Epithelial Cells Are Unconnected." J Immunol **179**(9): 5693-5700.

Vermeulen, M., K. W. Mulder, S. Denissov, W. W. Pijnappel, F. M. van Schaik, R. A. Varier, M. P. Baltissen, H. G. Stunnenberg, M. Mann and H. T. Timmers (2007). "Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4." Cell **131**(1): 58-69.

Vo, N. and R. H. Goodman (2001). "CREB-binding protein and p300 in transcriptional regulation." J Biol Chem **276**(17): 13505-8.

Walker, L. S., A. Chodos, M. Eggena, H. Dooms and A. K. Abbas (2003). "Antigen-dependent proliferation of CD4+ CD25+ regulatory T cells in vivo." J Exp Med **198**(2): 249-58.

Walker, L. S. K. and A. K. Abbas (2002). "The Enemy Within: Keeping Self-reactive T Cells at Bay in the Periphery." Nature Reviews Immunology **2**(1): 11-19.

Wang, Z., C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T.-Y. Roh, W. Peng, M. Q. Zhang, for full list of author see journal. (2008). "Combinatorial patterns of histone acetylations and methylations in the human genome." Nat Genet **40**(7): 897-903.

Wansink, D. G., W. Schul, I. van der Kraan, B. van Steensel, R. van Driel and L. de Jong (1993). "Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus." J Cell Biol **122**(2): 283-93.

White, A. J., D. R. Withers, S. M. Parnell, H. S. Scott, D. Finke, P. J. Lane, E. J. Jenkinson and G. Anderson (2008). "Sequential phases in the development of Aire-expressing medullary thymic epithelial cells involve distinct cellular input." Eur J Immunol **38**(4): 942-7.

White, D. A., N. D. Belyaev and B. M. Turner (1999). "Preparation of site-specific antibodies to acetylated histones." Methods **19**(3): 417-24.

Whitehouse, I., A. Flaus, B. R. Cairns, M. F. White, J. L. Workman and T. Owen-Hughes (1999). "Nucleosome mobilization catalysed by the yeast SWI/SNF complex." Nature **400**(6746): 784-7.

Wolf, S. S. (2009). "The protein arginine methyltransferase family: an update about function, new perspectives and the physiological role in humans." Cell Mol Life Sci **66**(13): 2109-21.

Wu, L. (2006). "T lineage progenitors: the earliest steps en route to T lymphocytes." Current Opinion in Immunology **18**(2): 121-126.

Wysocka, J., T. Swigut, H. Xiao, T. A. Milne, S. Y. Kwon, J. Landry, M. Kauer, A. J. Tackett, B. T. Chait, P. Badenhorst, for full list of author see journal. (2006). "A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling." Nature **442**(7098): 86-90.

Yano, M., N. Kuroda, H. Han, M. Meguro-Horike, Y. Nishikawa, H. Kiyonari, K. Maemura, Y. Yanagawa, K. Obata, S. Takahashi, for full list of author see journal. (2008). "Aire controls the differentiation program of thymic epithelial cells in the medulla for the establishment of self-tolerance." J. Exp. Med. **205**(12): 2827-2838.

Yie, J., K. Senger and D. Thanos (1999). "Mechanism by which the IFN-beta enhanceosome activates transcription." Proc Natl Acad Sci U S A **96**(23): 13108-13.

Zhang, Y. and D. Reinberg (2001). "Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails." Genes Dev **15**(18): 2343-60.

Zhu, M. and Y.-X. Fu (2008). "Coordinating Development of Medullary Thymic Epithelial Cells." Immunity **29**(3): 386-388.