Human Molecular Genetics, 2009, Vol. 18, No. 24 ***4699–4710***
doi:10.1093/hmg/ddp433
Advance Access published on September 10, 2009

AIRE activated tissue specific genes have histone modifications associated with inactive chromatin

Tõnis Org[1], Ana Rebane[1], Kai Kisand[1], Martti Laan[1], Uku Haljasorg[1], Reidar Andreson[2] and Pärt Peterson[1,*]

[1]Department of Molecular Pathology, Institute of General and Molecular Pathology and [2]Department of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

Received July 8, 2009; Revised and Accepted September 8, 2009

The Autoimmune Regulator (AIRE) protein is expressed in thymic medullary epithelial cells, where it promotes the ectopic expression of tissue-restricted antigens needed for efficient negative selection of developing thymocytes. Mutations in AIRE cause APECED syndrome, which is characterized by a breakdown of self-tolerance. The molecular mechanism by which AIRE increases the expression of a variety of different genes remains unknown. Here, we studied AIRE-regulated genes using whole genome expression analysis and chromatin immunoprecipitation. We show that AIRE preferentially activates genes that are tissue-specific and characterized by low levels of initial expression in stably transfected HEK293 cell model and mouse thymic medullary epithelial cells. In addition, the AIRE-regulated genes lack active chromatin marks, such as histone H3 trimethylation (H3K4me3) and acetylation (AcH3), on their promoters. We also show that during activation by AIRE, the target genes acquire histone H3 modifications associated with transcription and RNA polymerase II. In conclusion, our data show that AIRE is able to promote ectopic gene expression from chromatin associated with histone modifications characteristic to inactive genes.

INTRODUCTION

The ability to maintain self-tolerance is one of the fundamental properties of the normal immune system. During T-cell development in the thymus, the random genomic rearrangement of T-cell receptor (TCR) gene segments create a large number of different TCRs, many of which can recognize self-proteins. In order to maintain tolerance to self-antigens, auto-reactive T-cells undergo negative selection in the thymus medulla, where self-peptides bound to the MHC complex on the surface of antigen-presenting cells are presented to thymocytes. Thymocytes with high TCR affinity to the MHC-peptide complex are eliminated by apoptosis (1). Efficient negative selection is therefore dependent on the availability of self-proteins in the thymus.

Thymic medullary epithelial cells (mTECs) have an important role in maintaining self-tolerance, as they are capable of expressing large numbers of genes that are usually expressed only in specific tissues (2). This so-called ectopic or promiscuous gene expression significantly broadens the repertoire of self-antigens that can be presented to developing T-cells. In mice, the expression of many tissue-specific antigens has been shown to be dependent on the Autoimmune Regulator (AIRE) protein expressed in mature mTECs (3). The absence of AIRE leads to the escape of auto-reactive T-cells to the periphery and subsequent autoimmunity (4,5). In humans, mutated AIRE causes an autosomal recessive disease, Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy (APECED), which is characterized by a breakdown of self-tolerance, leading to destructive autoimmune reactions in several, mainly endocrine, organs (6).

Lack of thymic expression of a single antigen can be sufficient to elicit an autoimmune attack against the corresponding tissue (7). Therefore, to ensure immune tolerance, the expression of tissue-specific antigens in the thymus must be carefully regulated. Although the ability of AIRE to promote gene expression in cell models and in the thymus has been described by several groups (3,8,9), the exact molecular mech-

*To whom correspondence should be addressed at: Molecular Pathology, University of Tartu, Tartu 50411, Estonia. Tel: +372 7374202; Fax: +372 7374207; Email: part.peterson@ut.ee

anisms by which AIRE enhances the expression of tissue-specific genes in mTECs are still not clear.

One important factor that influences the transcriptional status of a gene is chromatin structure. Post-translational modifications of the N-terminal ends of histones have an important role in this process, as they determine whether proteins with specific domains can interact with chromatin and can predict whether genes are active. For example, the promoters of transcriptionally active genes are usually enriched for histone H3 trimethylation at lysine 4 (H3K4me3), whereas silenced gene promoters often have histone H3 trimethylated at lysine 27 (H3K27me3) (10). We and others have recently shown that the PHD finger of AIRE interacts with the unmodified N-terminal ends of histone H3, and this interaction is required for AIRE-dependent transcriptional activation (11–13). Since AIRE is able to sense histone H3 N-terminal modifications, it is likely that chromatin structure plays an important role in AIRE-mediated gene activation (14). Accordingly, post-translational histone H3 modifications may direct AIRE to specific locations on chromatin to activate target genes. Likewise, it has been proposed that AIRE may be involved in structural changes of larger chromatin regions (15).

To address the role of chromatin structure and histone binding in AIRE-promoted ectopic expression, we used genome-wide expression analysis, chromatin immunoprecipitation followed by array analysis (ChIP-on-chip) and a bioinformatics approach. Due to the scarcity of AIRE-positive cells in the thymus, we took advantage of a model system of a stably transfected human embryonic epithelial HEK293 cell line. Earlier studies suggested that the basic mechanism by which AIRE activates its broad range of target genes is independent of the mTECs' cellular environment (16,17). Indeed, similar to results of studies with mTECs, we have shown that AIRE efficiently activates several genes, including *insulin*, in HEK293 cells (13). Thus, these cells have all the necessary factors for AIRE-mediated gene activation and provide a feasible model for studying the molecular mechanisms of ectopic expression.

In this study, using the epithelial cell model, we show that AIRE is able to activate ectopic expression of genes, which are tissue-specific, characterized by low levels of initial expression and lack transcriptionally active chromatin marks, such as H3K4me3 and AcH3, on their promoters. During their activation, AIRE-regulated genes acquire active histone modifications and RNA polymerase II. Importantly, we also confirm the lack of active chromatin marks on AIRE target genes *in vivo* in mouse mTECs.

RESULTS

AIRE acts as a transcriptional activator in HEK293 cells

We recently showed that AIRE increases the expression of its target genes, *S100A8*, *involucrin* and *insulin*, in stably transfected HEK293 cells (13). To further identify AIRE-regulated genes in this cell model, we used Illumina HumanWG-6_V2_0_R2 BeadChip expression arrays that cover more than 48 000 transcripts. The experiments were performed in triplicate with cell lines expressing AIRE (two independent AIRE lines: AIRE1 and AIRE2), AIRE with a D312A mutation (AIRE-D312A) and yellow fluorescence protein as a negative control (NC). The D312A mutation resides in the first PHD zinc finger (PHD1) of AIRE, and structural studies have shown that it prevents zinc finger binding to histone H3 N-terminal tails (13).

The array analysis results demonstrated the ability of AIRE to induce ectopic gene expression in HEK293 cells. The comparison of the combined data sets of AIRE1 and AIRE2 with NC yielded 140 and 154 genes that were significantly ($P < 0.05$) up- and downregulated, respectively (Fig. 1A and Supplementary Material, Fig. S1A, B). The gene expression scatter plots and heat map of 140 upregulated genes showed that cells with AIRE1 and AIRE2 had almost identical expression profiles (Fig. 1B), whereas expression in AIRE-D312A cells was more similar to that of the NC (Fig. 1C and D; Supplementary Material, Fig. S1A). The latter result shows that the intact structure of PHD1 involved in histone binding is crucial for the efficient activation of a majority of AIRE-regulated genes. However, histone H3 binding does not seem to be essential for regulation of a portion of the AIRE targets (62 genes), as these were still upregulated in cells expressing AIRE-D312A. Nevertheless, when compared with AIRE wild-type lines, the AIRE-D312A line was clearly less potent (Fig. 1C and D; Supplementary Material, Fig. S1A) and resulted in lower activation levels. Indeed, the induction levels of 41 out of 140 genes, upregulated in the AIRE1/AIRE2 cell lines, were significantly lower in AIRE-D312A when compared with AIRE1/AIRE2. In contrast, analysis of 154 genes indicated that their downregulation was not dependent on intact PHD1 structure (Fig. 1C and D; Supplementary Material, Fig. S1B), suggesting that histone binding is not essential for AIRE-mediated gene repression.

We noticed that several previously identified AIRE targets (*S100A8*, *involucrin*, *insulin*) were not differentially expressed in our array analysis. For example, although *S100A8* expression increased almost 5-fold and ranked 311th among activated genes, its differential *P*-value was 0.42. Therefore, the number of AIRE-regulated genes identified by expression array was most likely underestimated. The smaller number of genes in our microarray approach was probably due to rigorous statistical analysis. Analysis of our data with the commonly used arbitrary 2-fold change threshold yielded 540 up- and 529 downregulated genes, which are in the same range as those of previously published microarray results of transgenic expression of AIRE in pancreatic beta cells (17). Interestingly, although we specifically analyzed the expression of well-characterized APECED autoantigen genes (*CYP17A*, *CYP21A2*, *CYP11A*, *TG*, *TPO*), these were not activated by AIRE. Thus, these results are in agreement with previous data showing that AIRE promotes ectopic expression by a generic mechanism not specific to mTECs, whereas the cellular environment defines which genes are regulated by AIRE (16,17).

AIRE target genes have tissue-specific expression patterns and tend to cluster in the genome

In mouse mTECs and in a rare subset of stromal cells in the lymph nodes, AIRE preferentially activates tissue-specific genes (16,18). To determine whether AIRE-upregulated genes in HEK293 cells can be considered tissue-specific, we searched the *ti*ssue-specific *g*ene *e*xpression and *r*egulation (TiGER)

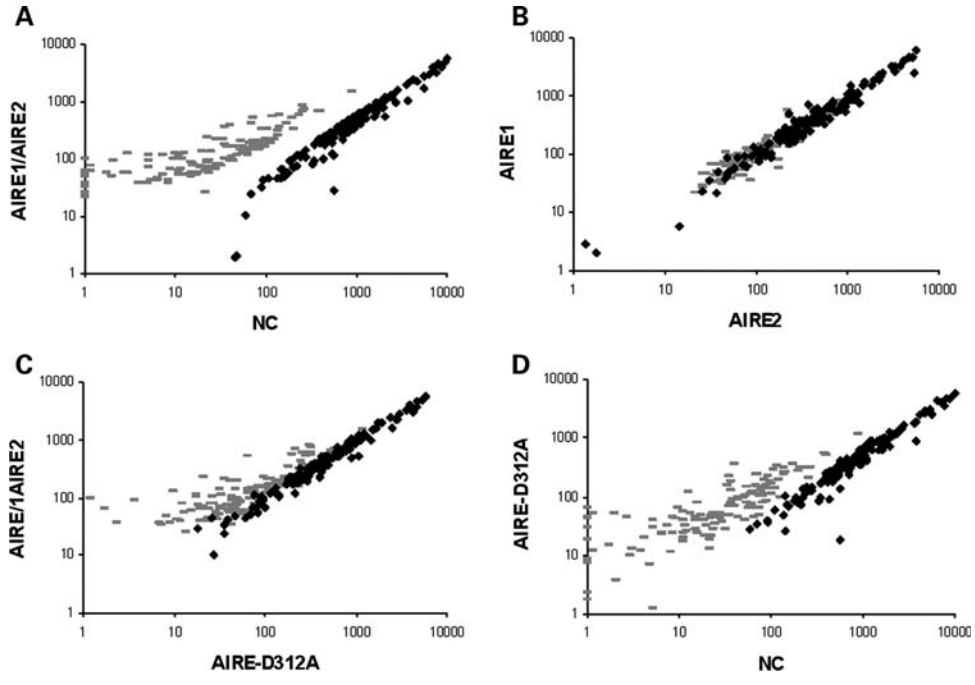

Figure 1. AIRE changes the expression of many genes in HEK293 cell line. Signal intensity scatter-plots of 140 significantly up- (gray lines) and 154 down-regulated (black spots) genes in combined AIRE1 and AIRE2 (AIRE1/AIRE2) data sets compared with the NC cell line (**A**), AIRE1 cell line compared with AIRE2 cell line (**B**), AIRE1/AIRE2 compared with AIRE-D312A cell line (**C**) and AIRE-D312A compared with NC cell line (**D**). While the expression profiles of AIRE1 and AIRE2 cell lines are very similar, AIRE-D312A cell line shows decreased activation of several genes when compared with wild-type AIRE cell lines.

database (http://bioinfo.wilmer.jhu.edu/tiger/) (19). Of the 140 AIRE-upregulated genes, the TiGER database contained 121, and 59 of these were classified as having tissue-specific expression patterns. We next used the database to identify the total number of tissue-specific genes present on Illumina arrays (6005) and compared the observed versus expected frequencies of tissue-specific genes among the AIRE target genes. Strikingly, the proportion of tissue-specific genes was indeed significantly higher among AIRE-upregulated genes than would have been expected by chance (chi-square test P-value 2.7×10^{-8}). Furthermore, as previous studies with mTECs have shown (18), the range of tissues expressing the genes activated by AIRE was extremely wide, covering essentially the whole body (Supplementary Material, Fig. S2). This association, however, was not found when the analysis was performed on genes downregulated by AIRE.

Previous studies have shown that AIRE-regulated genes tend to cluster in the genome (18,20). We also observed that compared with random set, AIRE regulated genes in HEK293 cells were found to be more clustered than would be expected (chi-square test P-value 3.6×10^{-6}) (Supplementary Material, Fig. S3).

AIRE targets have low initial expression levels

We next analyzed the distribution of AIRE-regulated genes by expression level. Strikingly, we found that AIRE preferentially activates genes with low initial expression levels, whereas in contrast, the repressed genes had high initial expression levels (Supplementary Material, Fig. S4A). This finding was further supported by comparison of the average signal intensity of all genes on the Illumina arrays (average signal intensity 290) with AIRE-upregulated (57) or downregulated (1357) average gene signal intensities. Moreover, the average increase of AIRE-upregulated genes was far higher (12.4-fold) than the average decrease (2.6-fold) of repressed genes. In addition, there was a significant negative correlation between initial expression levels in NC cells and the fold of induction (Pearson correlation, two-tailed P-value $= 0.0009$), so that genes with low initial expression levels had higher induction rates (Supplementary Material, Fig. S5A). No such correlation between expression level and fold of repression was present in genes downregulated by AIRE (Supplementary Material, Fig. S5B). We also analyzed the distribution of AIRE-regulated genes by size and exon number, but overall, we found no difference compared with genes of the whole genome (Supplementary Material, Fig. S4B and C).

To verify the microarray results, we performed qPCR-based expression analysis on samples from AIRE1/AIRE2, AIRE-D312A and NC cells using primers specific to several identified AIRE target genes. In general, the qPCR results confirmed the expression differences found using the Illumina BeadChips (Fig. 2 and Supplementary Material, Fig. S6). However, the qPCR approach was by far more sensitive than the microarray and also confirmed strong upregulation of previously identified target genes with low expression levels (such as *insulin* and *S100A8*; Fig. 2) and several other genes, such as *S100A5*, *S100A6*, *S100A7*, *S100A9*, *LOR*, *SPRR2F* in epidermal differentiation cluster on chromosome 1 (data not shown). It should be noted that the fold of induction varied widely between individual genes. For example, the induction of *S100A8, HBG2* and *IFI16* genes was more than

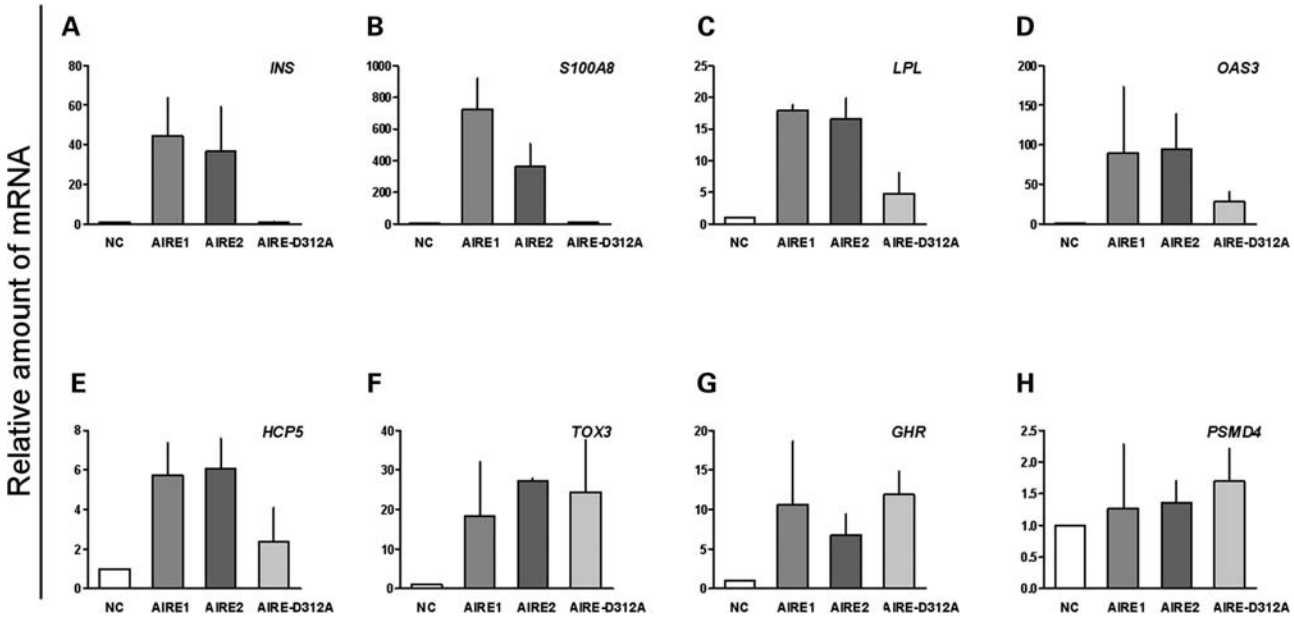

Figure 2. The extent of transcription activation induced by AIRE varies widely. qPCR determined mRNA levels of AIRE target genes, *INS* (**A**), *S100A8* (**B**), *LPL* (**C**), *OAS3* (**D**), *HCP5* (**E**), *TOX3* (**F**), *GHR* (**G**) and AIRE independent *PSMD4* gene (**H**) in NC, AIRE1, AIRE2 and AIRE-D312A cell lines are shown relative to the respective mRNA levels in NC cell line (=1). Average data of three independent experiments with SD are shown.

200-fold, whereas *LPL, PHF11, THBS2* and *LY6G6D* were upregulated less than 20-fold (Supplementary Material, Fig. S6). Again, larger fold changes were seen with genes with low initial expression levels (data not shown). The qPCR results with AIRE-D312A also confirmed the array results, demonstrating that the expression of a subset of genes was partially (*OAS3, HCP5, CLDN1*) or completely (*GHR, TOX3, PAK3, SUSD4, CDH12*) independent of the PHD1 finger (Fig. 2 and Supplementary Material, Fig. S6). The expression levels of AIRE-independent control genes (*PSMD4* and *BAT2*) were not changed (Fig. 2 and Supplementary Material, Fig. S6).

Taken together, these results show that the target genes activated by AIRE are expressed at lower levels compared with the average expression level.

AIRE target gene promoters have low H3K4me3 and AcH3 modification levels

We next aimed to study the status of chromatin near AIRE-regulated genes. For this purpose, we used a chromatin immunoprecipitation assay with AIRE1 and NC cells, followed by hybridization of the immunoprecipitated material to custom designed tiling arrays (ChIP-on-chip). We included AIRE target and control gene regions to the tiling array, which covered 36 Mb of the genomic DNA sequence. To detect histone modifications, we used specific antibodies against H3K4me3 and AcH3 modifications, which mark a transcriptionally active state, and against H3K27me3, which is characteristic to silenced chromatin (10), as well as an antibody against unmodified histone H3 as a control.

We first verified the overall performance of the arrays. As can be seen in Supplementary Material, Figure S7A and B, we observed a clear trend for genes with high expression levels to have H3K4me3 and AcH3 marks at their promoter regions. In contrast, the genes with low expression levels tended to have the H3K27me3 modification at their promoter regions (Supplementary Material, Fig. S7C).

After confirmation that the histone modifications correlated with gene expression levels as expected, we next analyzed the presence of chromatin modifications on the promoters (2 kb up- and downstream of transcription start site) of AIRE target genes. Remarkably, AIRE upregulated gene promoters were largely devoid of the transcriptionally active gene marks H3K4me3 and AcH3 (Table 1; Fig. 3). When we compared AIRE target gene histone modification to 200 random genes with low or high expression, the AIRE target genes were clearly more similar to genes with low levels of expression (Table 1; Fig. 3). We conclude that the majority of AIRE target gene promoters have low levels of H3K4me3 and AcH3, the posttranslational histone modifications that characterize the active genes.

Table 1. The distribution histone modifications on AIRE upregulated gene promoters and on the promoters of 200 random genes with low or high expression

	H3K4me3		AcH3		H3K27me3		No mod.	
	NC	AIRE1	NC	AIRE1	NC	AIRE1	NC	AIRE1
AIRE upregulated	15	36	2	16	18	9	70	52
200 low	16	19	20	21	17	12	64	67
200 high	74	75	69	67	2	1	23	22

Percent of genes lacking all (No mod.) or having corresponding modification (H3K4me3, AcH3, H3K27me3) on their promoters ($\pm$ 2 kb of TSS) is shown.

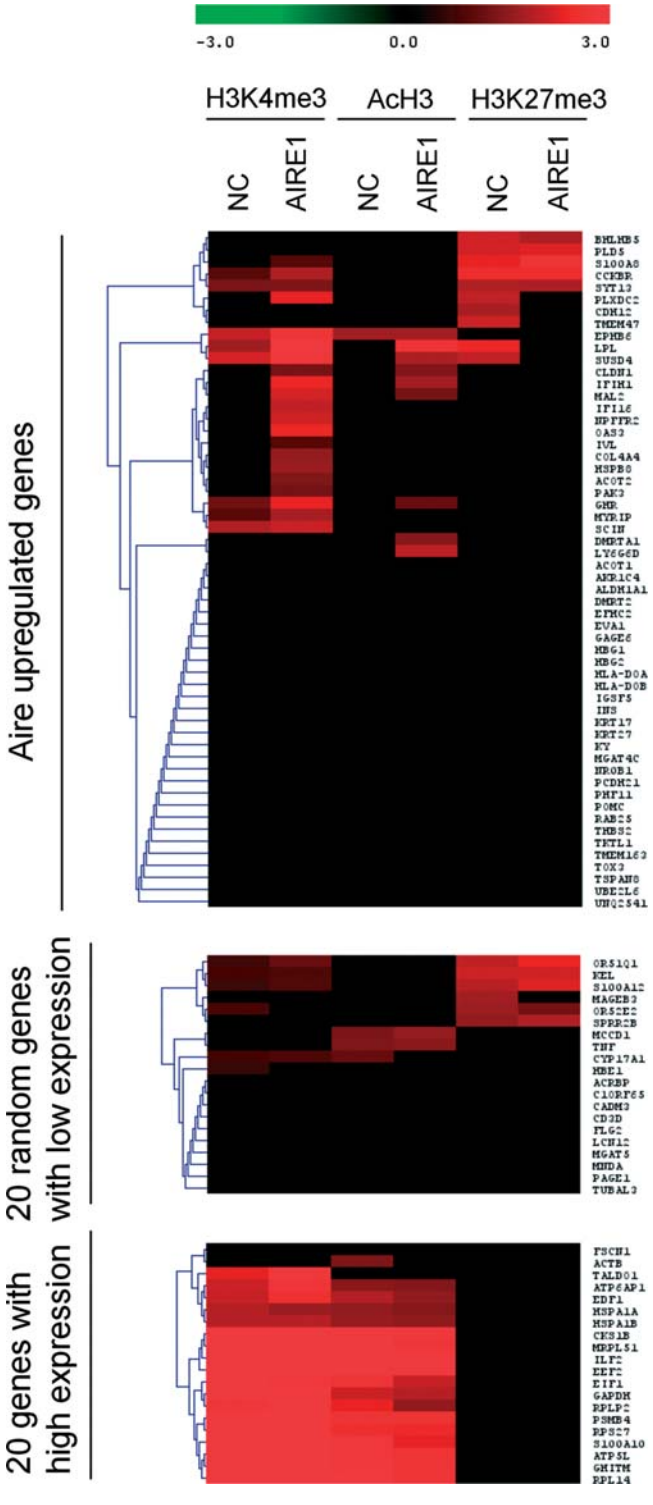

Figure 3. AIRE-mediated transcriptional activation is accompanied by changes in histone modifications. The presence of histone modifications (H3K4me3, AcH3, H3K27me3) on AIRE upregulated, on the promoters (2 kb up- and downstream of transcriptional start sites) of genes with low or high expression (20 out of 200 genes are shown) in NC and AIRE1 cell lines. Red color denotes presence and black color lack of corresponding modification. The shades of red represent peak scores, which translate into the amounts of corresponding modifications.

AIRE induces active histone modifications and RNA polymerase recruitment

Interestingly, we noticed that in AIRE-expressing cells compared with NC cells, the histone modifications in several AIRE target gene promoters were changed (Table 1). We found a higher proportion of genes having H3K4me3 (36% in AIRE1 versus 15% in NC) and AcH3 (16% in AIRE1 versus 2% in NC) marks, and the number of corresponding modifications was also increased, as assessed by the peak scores. This correlated well with the decreased number of genes having the H3K27me3 modification (9% in AIRE1 versus 18% in NC). Importantly, such dramatic changes were not seen in 200 random genes with either low or high expression levels (Table 1; Fig. 3). To confirm these findings, we used an alternative approach where we normalized the histone modification results to H3 and calculated the average signal intensity for each histone modification in a region 2 kb up- and downstream of the AIRE target gene transcription start sites in NC and AIRE1 cells. This analysis also clearly indicated the presence of increased H3K4me3 and AcH3 levels and decreased H3K27me3 levels in a majority of the AIRE target genes in AIRE1 cells (Supplementary Material, Fig. S8).

We then analyzed more closely individual genes from our ChIP-on-chip array data. One such example of AIRE-upregulated genes, *GHR*, had increased levels of histone modifications characteristic for transcriptional activation in cells expressing AIRE (Fig. 4). The increase in H3K4me3 and AcH3 levels occurred adjacent to the transcriptional start site, whereas the H3K27me3 levels were decreased throughout the gene (Fig. 4). Similar changes were also seen with other genes, such as *OAS3*, *LPL, TOX3* and *HSBP8* (Supplementary Material, Fig. S9A–D). In contrast, in AIRE-independent genes, like *GAPDH*, the levels of posttranslational modifications in AIRE1 and NC cells were similar.

A study by Oven *et al*. has shown that AIRE promotes transcriptional elongation by recruiting P-TEFb to target gene promoters (21). This scenario would assume the presence of RNA polymerase II on the promoters of AIRE target genes. In this context, we performed ChIP-on-chip experiments to study engagement of RNA polymerase II with AIRE target genes, with and without AIRE expression. In NC cells, we found RNA polymerase II on only ~4% of AIRE target promoters. However, in agreement with increased expression, RNA polymerase II was present on 27% of the target gene promoters in AIRE-positive cells.

To verify the changes of histone modifications seen in ChIP-on-chip experiments, we performed conventional ChIP assays with H3K4me3, H3K27me3, AcH3 and RNA polymerase II antibodies in AIRE1 and NC cells. As expected, the promoter regions of *GHR* and *LPL* genes contained higher H3K4me3, AcH3 and RNA polymerase II levels in AIRE1 cells, whereas H3K27me3 levels were reduced (Fig. 5). A small increase in H3K4me3 levels was also observable on other target gene (*CDH12*, *PAK3*, *SUSD4*, *CLDN1*) promoters but not in all target genes, including, for example, *INV*, *INS* and *S100A8*, and in AIRE-independent negative control genes (*PSMD4* and *GAPDH*) (Fig. 5; Supplementary Material, Fig. S10).

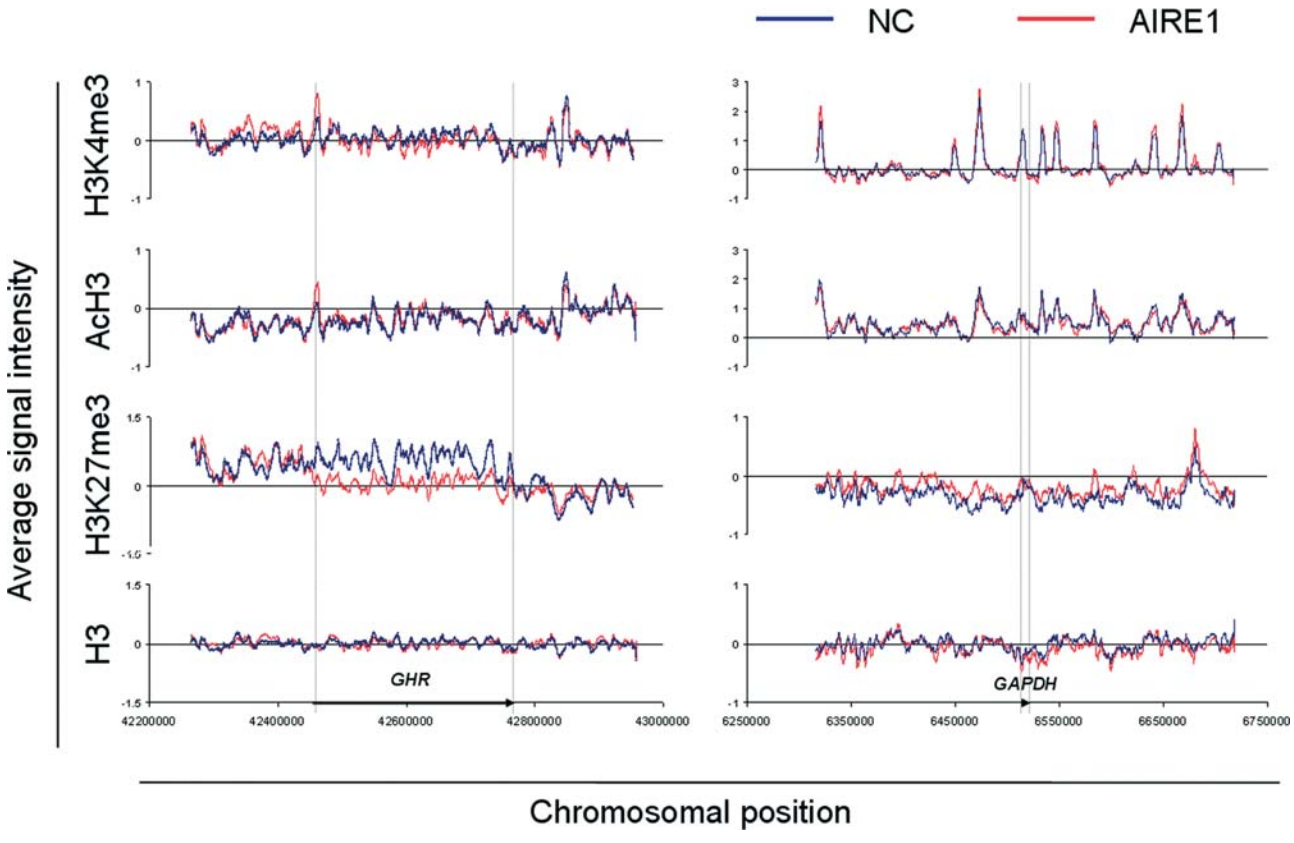

Figure 4. AIRE-mediated gene activation is accompanied by changes in histone modifications. Average levels of histone modifications (H3K4me3, AcH3, H3K27me3) and general H3 in a region of 200 kb of up- and downstream of AIRE upregulated GHR and AIRE independent GAPDH genes (arrows) in NC (blue line) and AIRE1 cell lines (red line).

Taken together, our results show that gene activation by AIRE is accompanied by an increase in the number of the transcriptionally active histone posttranslational modifications H3K4me3 and AcH3 and that AIRE is able to initiate RNA polymerase II recruitment to target gene promoter regions.

Posttranslational histone modifications in AIRE-regulated gene clusters

As AIRE-regulated genes tend to cluster in genome, AIRE's role in the induction of structural changes of larger chromatin regions has been proposed (18,20). We examined a 1.8 Mb region on chromosome 1 that included the epidermal differentiation cluster, which contains several AIRE target genes. In general, the overall distribution of the studied posttranslational histone modifications in AIRE1 and NC cells was very similar (Supplementary Material, Fig. S11). We also analyzed other AIRE regulated gene clusters but found no changes in chromatin modifications that covered more than one gene region (data not shown). This result suggests that, AIRE does not mediate large epigenetic rearrangements.

AIRE binding to chromatin

We also studied AIRE binding to chromatin by ChIP-on-chip assays. The ChIP-on-chip assay, using an anti-AIRE antibody in AIRE cells, identified 1854 peaks, which were distributed rather randomly throughout the studied 36 Mb chromatin region. Analysis of these regions with the CEAS system (http://ceas.cbi.pku.edu.cn) (22) showed that, compared with the background in NC cells, the AIRE-binding sites localized more to gene promoters and exons (Table 2). In a region of 2 kb up- and downstream of transcription start sites, AIRE-binding peaks were found in 30% of AIRE-upregulated genes present on the array. These results suggest that AIRE binds to chromatin in a random fashion and/or that the binding is dynamic.

AIRE activation of promoters with low H3K4me3 is dependent on PHD1 domain

Previous studies have shown that H3K4 methylation hinders AIRE binding to histone H3 N-terminal tails (12,13). Our gene expression analysis showed that the activation of some of the AIRE target genes is not affected by D312A mutation and should be thus less-dependent on the interaction with histones and H3K4me3 levels. To test this hypothesis, we performed an analysis to determine if there is a difference in the promoter H3K4 methylation levels between the genes which expression is either affected or not by D312A mutation. The comparison of 15 genes most or least affected by the D312A mutation revealed that genes which are affected show significantly lower levels of H3K4me3 levels on their promoters, both in NC and AIRE1 cell lines (Student's

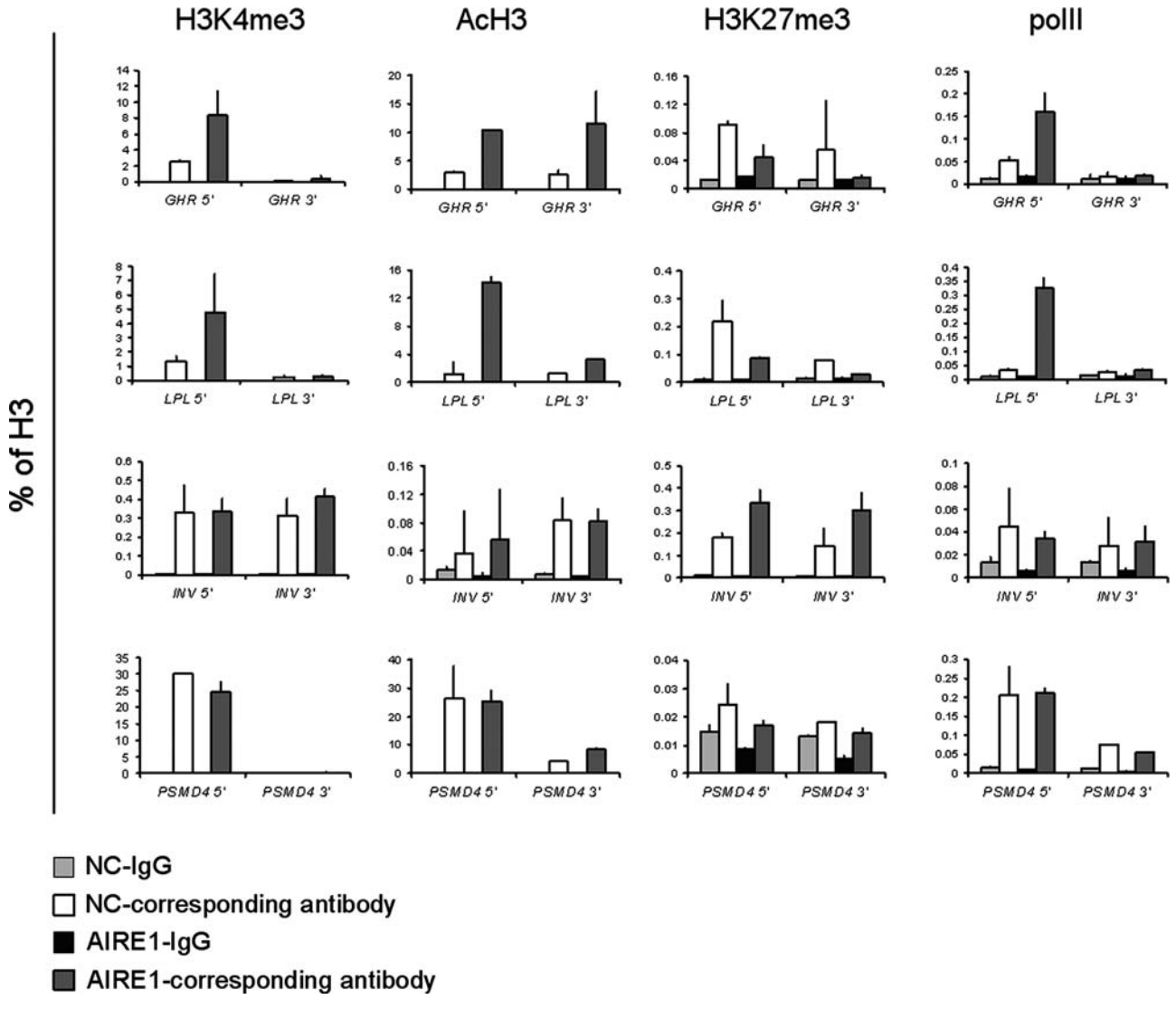

Figure 5. Verification of ChIP-on-chip data. Immunoprecipitated DNA was analyzed using primers specific to the AIRE target genes *GHR*, *LPL* and *INV*, and AIRE-independent *PSMD4* gene to determine the levels of H3K4me3, AcH3, H3K27me3 and RNA polymerase II (polII) on the promoters (5′) and 3′ regions in NC and AIRE1 cell lines. Average data of two independent experiments with SD are shown. The SD derived from two experiments is not appropriate but is included merely to demonstrate that the two experiments gave similar results.

two-tailed *t*-test *P*-values 0.01 and 0.001, respectively) (Supplementary Material, Fig. S12). This result further supports the idea that AIRE interaction with histones is needed for activation of gene promoters with low level of H3K4me3.

AIRE target genes have low levels of transcriptionally active histone modifications *in vivo*

We next studied the levels of histone modifications of selected tissue-specific antigens in mouse mTECs and in their corresponding peripheral tissues. Immature mTECs are negative for AIRE and CD80 expression, which are activated when the cells differentiate into mature mTECs. To identify the differences between AIRE negative and positive mTECs, we studied immature $CD80^{low}$ and mature $CD80^{high}$ populations separately. According to our hypothesis, the AIRE tissue-specific target promoters should have low levels of H3K4me3 and AcH3 in mTECs, whereas these levels should be higher in corresponding peripheral tissues, where these genes are highly expressed.

The major obstacle in studying histone modifications from primary cells is the relatively small quantity of cellular material available. Here, we used the Q^2Chip method, which enables the study of histone modifications from a small number (ca. 20 000) of cells (23).

We first found that the housekeeping gene GAPDH promoter contained similarly high levels of H3K4me3 and low levels of H3K27me3 in all studied tissues: mTECs ($CD80^{low}$ and $CD80^{high}$), pancreas and neutrophils. As expected, compared with other tissues, the H3K4me3 levels were high on promoters in tissues where each of the genes was preferentially expressed: *Ins2* and *Gad2* in the pancreas and *S100a8* in neutrophils. H3K27me3 levels were low almost on all of the studied promoters in all tissues, except for the slightly

Table 2. AIRE binding sites localize preferentially to gene promoters and exons

	Exon, %	Intron, %	5′-UTR, %	3′-UTR, %	Proximal promoter, %	Immediate downstream, %	Enhancer, %
AIRE	14.67	32.79	0.05	1.62	4.37	2.64	43.85
NC	6.1	33.96	0.02	0.71	1.43	2.27	55.51

The genomic distribution of AIRE binding sites (AIRE) compared with the background (NC).

higher levels observed on the *S100a8* promoter in the pancreas. When we studied $CD80^{low}$ mTECs, we found that, in agreement with our cell line data, the H3K4me3 levels were low on the promoters of the AIRE-dependent tissue-specific antigens *Ins2* and *S100a8*, but were much higher on the promoter of the AIRE-independent tissue-specific gene, *Gad2* (Fig. 6). When comparing $CD80^{high}$ with $CD80^{low}$ cells, we observed that the H3K4me3 levels on the promoters of *Gad2* and *S100a8* were increased, while the promoter of *Ins2* remained unchanged. Similar to some AIRE target genes, we also saw an increase in H3K4me3 levels on the AIRE promoter during mTEC differentiation. Together, these results are in agreement with our data in the AIRE-expressing cell model and show that AIRE target genes in mTECs have initially low levels of histone H3 active chromatin modification on their promoters, which, depending on the gene, can be enhanced in the presence of AIRE.

DISCUSSION

In this study, we show that AIRE promotes the expression of genes that can be characterized by their low expression levels and lack of transcriptionally active epigenetic modifications. These features are characteristic for tissue-specific genes outside of their host tissue and, indeed, the expression of many AIRE target genes activated in our epithelial cell model was restricted to a limited number of tissues. We also show that on target gene promoters, AIRE is able to increase H3K4me3 and AcH3 levels, the histone modifications associated with active chromatin and to engage RNA polymerase II. Our results support the idea that AIRE induces the expression of target genes through transcriptional activation.

AIRE is needed for proper negative selection of self-reactive thymocytes in the thymus (4). Although there is evidence that AIRE might also influence other processes, such as processing and/or presentation of self-antigens (24) or differentiation of mTECs (25), the main mechanism by which AIRE propagates negative selection involves promotion of tissue-specific antigen expression in mTECs. In addition, it has been reported that over-expression of AIRE induces apoptosis in 1C6 cell line (26); however, we did not observe this in HEK293 cells. This discrepancy is probably due to the different cell-lines used; the 1C6 cells were derived from primary mTECs (27) and therefore might be more susceptible to apoptosis than the transformed HEK293 cells used in this study. Our data show that even in an embryonic epithelial cell model, AIRE is able to induce ectopic expression of multiple genes, resembling the situation in mature mTECs. Importantly, although the repertoire of genes that was activated in this cell model was different from that in mTECs, many of these genes were classified as tissue-specific.

Considering that AIRE activates multiple genes, it was reasonable to hypothesize that in order to regulate gene expression, AIRE should be recruited to chromatin. We and others recently have shown that the AIRE PHD1 finger is able to interact with histone H3 N-terminal tails (11–13). Several posttranslational modifications on histone H3 affect this interaction, and, in particular, H3K4me3, which is enriched on the promoters of actively transcribed genes, hinders AIRE binding. According to this scenario, AIRE is able to bind chromatin regions with low levels of H3K4me3. While tissue-specific genes are highly expressed only in a limited number of tissues, they have low expression levels and low levels of H3K4me3 on their promoters in other tissues (28). This is consistent with our results, which show that AIRE target gene promoters have low initial expression levels and low levels of H3K4me3 and AcH3. Importantly, we confirmed our cell model results by showing the presence of low levels of H3K4me3 in mouse $CD80^{low}$ mTECs on the AIRE target genes *Ins2* and *S100a8*, whereas the AIRE-independent genes *Gad2* and *Gapdh* had high levels of H3K4me3.

Our findings provide several clues of AIRE's function in mTECs. It provides a mechanistic explanation of how, based on epigenetic histone marks, AIRE is guided to its target genes in the thymus. This implies that AIRE is not recruited to tissue-specific antigen genes *per se*, but instead, AIRE identifies its target genes by a specific histone code, which can often be found on tissue-specific gene promoters outside of the tissues in which they are expressed. Besides the AIRE–PHD1 interaction with histone H3, however, other factors may be involved in AIRE recruitment to chromatin, as the expression levels of some genes were not influenced by the D312A mutation. For example, AIRE harbors a putative DNA binding domain called SAND, which in the case of the Sp100 protein has been shown to preferentially bind unmethylated CpG (29). Indeed, earlier studies have proposed a direct AIRE interaction with specific DNA sequence motifs (30–32). In contrast, we (our unpublished data) and others (15) have found that AIRE interaction with DNA, when studied *in vitro*, occurs in a non-specific manner. In addition, although our initial ChIP results (13) showed that very high H3K4me3 levels might prevent AIRE binding to chromatin, our AIRE ChIP-on-chip data indicate that the AIRE protein has a fairly random pattern of chromatin binding. It is possible that although H3K4me3 modification blocks AIRE PHD1 binding, the other AIRE domains might still interact with DNA. AIRE's relatively stochastic binding to chromatin is

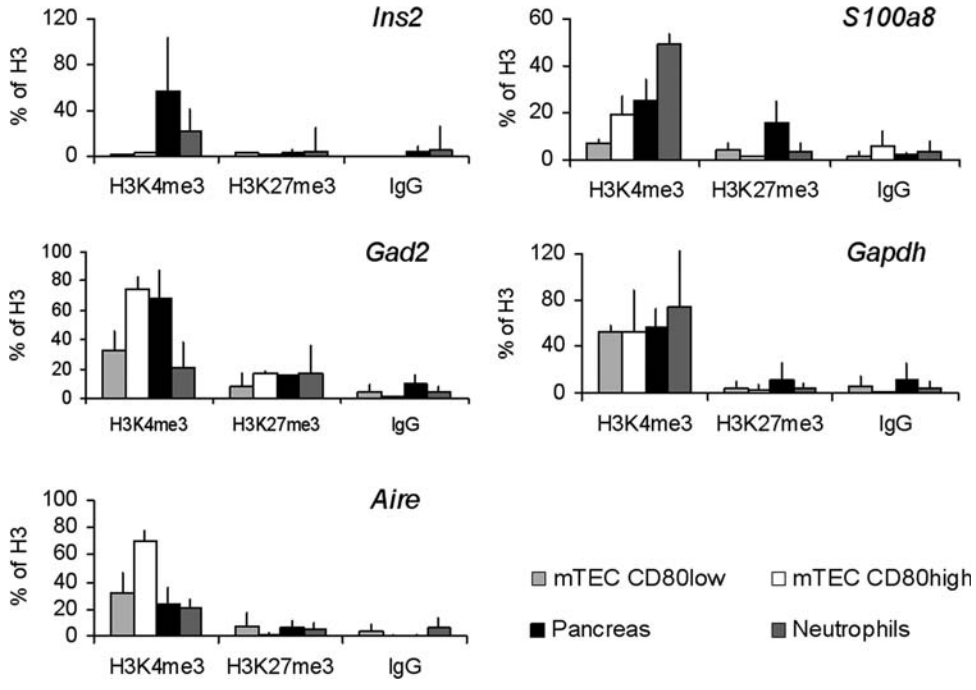

Figure 6. Aire target genes have low levels of H3K4me3 on their promoters in mouse mTECs. Q^2Chip analysis of the promoters of Aire dependent tissue specific antigens, *Ins2* and *S100A8* as well as Aire independent tissue-specific antigen *Gad2*, housekeeping gene *Gapdh* and *Aire* in $CD80^{high}$, $CD80^{low}$ mTECs, pancreas and neutrophils. Average data of two independent experiments with SD are shown. The SD derived from two experiments is not appropriate but is included merely to demonstrate that the two experiments gave similar results.

also in agreement with other studies reporting that the AIRE target genes vary in different genetic backgrounds and cell types, and that even different single cells from the same tissue may have variable patterns of gene expression (17,33,34).

After recruitment to chromatin, AIRE activates gene expression. Oven *et al*. (21) have proposed that AIRE promotes transcriptional elongation by recruiting elongation factor P-TEFb to target gene promoters. This AIRE-mediated elongation mechanism assumes the presence of stalled RNA polymerase on the target gene promoters. However, in NC cells, we did not detect RNA polymerase II on the AIRE target promoters. Even though this lack of RNA polymerase in our experiments could be explained by low detection efficiency, we still observed an increase in RNA polymerase II levels on several target genes after activation by AIRE. This was also confirmed by ChIP experiments using initiating RNA polymerase II (data not shown). AIRE's role in processing primary transcripts has been proposed as another possible mechanism (15). Our data, which show activation of AIRE target genes through changes in their histone modifications and RNA polymerase II recruitment, argue that the increase seen in mRNA levels is, at least to some extent, due to enhanced transcription. In this context, the variations in target gene selection between different cellular environments can be explained by the local milieu, such as accessibility of chromatin and the availability of transcriptional proteins.

In conclusion, our data show that AIRE is able to promote ectopic expression of tissue-specific genes from chromatin lacking transcriptionally active histone H3 modifications. Our data indicate that activation by AIRE occurs via transcriptional regulation and is accompanied by an increase in H3K4me3 and AcH3 levels on target gene promoters.

MATERIALS AND METHODS

Antibodies

Antibodies used in the study were: αH3 (ab1791), αH3K4me3 (ab8580) (Abcam); αH3K27me3 (Upstate Biotechnology); α-rabbit polyclonal IgG (sc-2027) and αAIRE (sc-17985) (Santa Cruz).

Cell lines

The generation of HEK293 cell lines expressing AIRE1, AIRE2, AIRE D312A mutant and YFP expressing negative control (NC) has been described earlier (13).

Gene expression arrays

The RNA from AIRE1, AIRE2, AIRE-D312A and NC was extracted with Trizol reagent (Invitrogen), treated 2 × 30 min with Turbo DNAse (Ambion) and purified with QIAquick PCR Purification Kit (Qiagen). Labelled cRNA was prepared using Illumina TotalPrep RNA Amplification Kit (Ambion) and hybridized to Illumina HumanWG-6_V2_0_R2 BeadChip expression arrays (lllumina). The quality of the RNA and labeled cRNA was controlled using 2100 Bioanalyzer (Agilent). With all cell lines, three independent experiments were performed. The raw data was analyzed with BeadStudio Gene Expression Module v3.3.7 (Illumina) using

Illumina's custom rank invariant method. Genes with differential expression values >13 (corresponding to P-value < 0.05) were considered as differentially expressed. Multi Experiment Viewer version 4.0 was used for unsupervised hierarchical clustering (using Manhattan distance and complete linkage) and for the visualization of the results. The microarray data can be accessed through Gene Expression Repository under GEO accession number GSE16877.

Clustering analysis of AIRE-regulated genes

REEF program was used (http://telethon.bio.unipd.it/bioinfo/reef/index.html) to find significantly clustered AIRE target genes (35). Analysis was performed with 200 kb window size and 10 K shift at false discovery rate of 0.05. Chi-square test was used to determine the significance of the observed to expected deviation.

Determination of tissue specificity

TiGER database (http://bioinfo.wilmer.jhu.edu/tiger/) (19) was used to determine the proportion of tissue specific genes among AIRE upregulated, downregulated and all Refseq genes on Illumina BeadChip expression arrays. Chi-square test was used to determine the significance of the observed to expected deviation.

Expression analysis

Expression analysis with qPCR was performed as described previously (36).

Chip-on-chip

The custom made Nimblegen tiling arrays contained non-repetitive sequences throughout the chosen regions represented by 386748 probes with median spacing 72 bp. Array contained several control and 51 AIRE target gene regions that were chosen based on the results with Illumina Beadchip expression arrays. The chosen genes had to be significantly upregulated by AIRE (P-value < 0.05) and have fold change greater than 5. The arrays covered the whole gene and 200 kb up- and downstream of each gene. In addition, a 5.2 Mb region was chosen from chromosome 1, which contained the epidermal differentiation cluster. Chromatin immunoprecipitation was performed with αH3K4me3, αAcH3, αH3K27me3, αH3 and αAIRE from AIRE1 and NC cell lines as described previously (13). The immunoprecipitated and input material were amplified using WGA2 and WGA1 GenomePlex Whole Genome Amplification Kits (Sigma). Sample labeling and hybridization to custom made tiling arrays were performed by Nimblegen (Roche). For each probe, a scaled $\log_2$-ratio was calculated. Scaling was performed by subtracting the bi-weight mean for the $\log_2$-ratio values for all probes on the array from each $\log_2$-ratio value. Peaks were detected using NimbleScan by searching for four or more probes whose signals are above specified cutoff values, ranging from 15 to 90%, using a 500 bp sliding window. The cutoff values are percentage of hypothetical maximum, which is mean+6 standard deviation. Each peak is assigned a false discovery rate (FDR) based on 20 times randomization of ratio data. For histone modifications, FDR cutoff at 0.01 level was used.

For Supplementary Material, Figure S7, genes were first sorted according to the signal intensities on Illumina Beadchip expression arrays. A sliding window with the size of 50 probes and step of 1 probe was then used to calculate the percent of promoters having corresponding posttranslational histone modification. For Supplementary Material, Figure S8, the data for each antibody were normalized against H3 and the average signal intensity was calculated for each gene in a region of ±2 kb of transcription start sites. For Figure 4 and Supplementary Material, Figures S9 and 10 signal intensities for each probe were sorted according to their position along the corresponding chromosome. A sliding window analysis, consisting of 50 probes and step of 1 probe, was then used to calculate average signal intensities. Original ChIP-on-chip data can be accessed from the GEO database under the accession number GSE17216.

Mice

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

Cell sorting

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

Neutrophils were extracted from spleen by mechanical disruption of the tissue followed by positive selection with Ly6G microbeads (Miltenyi Biotec) and the AutoMACS system (Miltenyi Biotec). To enrich the pancreatic tissue for Langerhans islets, the tissue was first cut into small pieces, which was then incubated with 0.5 mg/ml dispase/collagenase (Roche) and 5 μg/ml DNase I (AppliChem) in PBS at 37°C for 20 min. Discontinuous Optiprep™ (Axis Shield) density gradient (1.108; 1.096; 1.069; 1.037) centrifugation was carried

out for 10 min at 1800 rpm and cells from two upper layers were used for further experiments.

Chip and Q^{2}chip

Chip experiments were carried out as described earlier (13). Q^{2}chip was performed essentially as described in (23) except, instead of Protein A-coated paramagnetic beads, protein G-sepharose beads (GE Healthcare) pre-blocked with 100 μg/ml BSA and 500 μg/ml salmon sperm DNA were used. Sequences of the PCR primers used in the study are listed in Supplementary materials.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

Maire Pihlap and Annika Häling are acknowledged for technical help.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by The Wellcome Trust Senior Fellowship grant; EU FP6 project Thymaide; the European Regional Fund and Archimedes Foundation; the Estonian Science Foundation (6663, 7559 and 7197), EU European Regional Development Fund through the Estonian Centre of Excellence in Genomics to R.A. and grant from Estonian Ministry of Education and Research (SF0180026s09 to R.A.).

REFERENCES

1. Palmer, E. (2003) Negative selection—clearing out the bad apples from the T-cell repertoire. *Nat. Rev. Immunol.*, **3**, 383–391.
2. Derbinski, J., Schulte, A., Kyewski, B. and Klein, L. (2001) Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nat. Immunol.*, **2**, 1032–1039.
3. Anderson, M.S., Venanzi, E.S., Klein, L., Chen, Z., Berzins, S.P., Turley, S.J., von Boehmer, H., Bronson, R., Dierich, A., Benoist, C. *et al.* (2002) Projection of an immunological self shadow within the thymus by the aire protein. *Science*, **298**, 1395–1401.
4. Liston, A., Lesage, S., Wilson, J., Peltonen, L. and Goodnow, C.C. (2003) Aire regulates negative selection of organ-specific T cells. *Nat. Immunol.*, **4**, 350–354.
5. Ramsey, C., Winqvist, O., Puhakka, L., Halonen, M., Moro, A., Kampe, O., Eskelin, P., Pelto-Huikko, M. and Peltonen, L. (2002) Aire deficient mice develop multiple features of APECED phenotype and show altered immune response. *Hum. Mol. Genet.*, **11**, 397–409.
6. Husebye, E.S., Perheentupa, J., Rautemaa, R. and Kampe, O. (2009) Clinical manifestations and management of patients with autoimmune polyendocrine syndrome type I. *J. Intern. Med.*, **265**, 514–529.
7. DeVoss, J., Hou, Y., Johannes, K., Lu, W., Liou, G.I., Rinn, J., Chang, H., Caspi, R.R., Fong, L. and Anderson, M.S. (2006) Spontaneous autoimmunity prevented by thymic expression of a single self-antigen. *J. Exp. Med.*, **203**, 2727–2735.
8. Pitkanen, J., Doucas, V., Sternsdorf, T., Nakajima, T., Aratani, S., Jensen, K., Will, H., Vahamurto, P., Ollila, J., Vihinen, M. *et al.* (2000) The autoimmune regulator protein has transcriptional transactivating properties and interacts with the common coactivator CREB-binding protein. *J. Biol. Chem.*, **275**, 16802–16809.
9. Bjorses, P., Halonen, M., Palvimo, J.J., Kolmer, M., Aaltonen, J., Ellonen, P., Perheentupa, J., Ulmanen, I. and Peltonen, L. (2000) Mutations in the AIRE gene: effects on subcellular location and transactivation function of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy protein. *Am. J. Hum. Genet.*, **66**, 378–392.
10. Kouzarides, T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.
11. Chignola, F., Gaetani, M., Rebane, A., Org, T., Mollica, L., Zucchelli, C., Spitaleri, A., Mannella, V., Peterson, P. and Musco, G. (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**, 2951–2961.
12. Koh, A.S., Kuo, A.J., Park, S.Y., Cheung, P., Abramson, J., Bua, D., Carney, D., Shoelson, S.E., Gozani, O., Kingston, R.E. *et al.* (2008) Aire employs a histone-binding module to mediate immunological tolerance, linking chromatin regulation with organ-specific autoimmunity. *Proc. Natl Acad. Sci. USA*, **105**, 15878–15883.
13. Org, T., Chignola, F., Hetenyi, C., Gaetani, M., Rebane, A., Liiv, I., Maran, U., Mollica, L., Bottomley, M.J., Musco, G. *et al.* (2008) The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO Rep.*, **9**, 370–376.
14. Peterson, P., Org, T. and Rebane, A. (2008) Transcriptional regulation by AIRE: molecular mechanisms of central tolerance. *Nat. Rev. Immunol.*, **8**, 948–957.
15. Mathis, D. and Benoist, C. (2009) Aire. *Annu. Rev. Immunol.*, **27**, 287–312.
16. Gardner, J.M., Devoss, J.J., Friedman, R.S., Wong, D.J., Tan, Y.X., Zhou, X., Johannes, K.P., Su, M.A., Chang, H.Y., Krummel, M.F. *et al.* (2008) Deletional tolerance mediated by extrathymic Aire-expressing cells. *Science*, **321**, 843–847.
17. Guerau-de-Arellano, M., Mathis, D. and Benoist, C. (2008) Transcriptional impact of Aire varies with cell type. *Proc. Natl Acad. Sci. USA*, **105**, 14011–14016.
18. Derbinski, J., Gabler, J., Brors, B., Tierling, S., Jonnakuty, S., Hergenhahn, M., Peltonen, L., Walter, J. and Kyewski, B. (2005) Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. *J. Exp. Med.*, **202**, 33–45.
19. Liu, X., Yu, X., Zack, D.J., Zhu, H. and Qian, J. (2008) TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformat.*, **9**, 271.
20. Johnnidis, J.B., Venanzi, E.S., Taxman, D.J., Ting, J.P., Benoist, C.O. and Mathis, D.J. (2005) Chromosomal clustering of genes controlled by the aire transcription factor. *Proc. Natl Acad. Sci. USA*, **102**, 7233–7238.
21. Oven, I., Brdickova, N., Kohoutek, J., Vaupotic, T., Narat, M. and Peterlin, B.M. (2007) AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. *Mol. Cell Biol.*, **27**, 8815–8823.
22. Ji, X., Li, W., Song, J., Wei, L. and Liu, X.S. (2006) CEAS: cis-regulatory element annotation system. *Nucleic Acids Res.*, **34**, W551–W554.
23. Dahl, J.A. and Collas, P. (2007) A quick and quantitative chromatin immunoprecipitation assay for small cell samples. *Front. Biosci.*, **12**, 4925–4931.
24. Kuroda, N., Mitani, T., Takeda, N., Ishimaru, N., Arakaki, R., Hayashi, Y., Bando, Y., Izumi, K., Takahashi, T., Nomura, T. *et al.* (2005) Development of autoimmunity against transcriptionally unrepressed target antigen in the thymus of Aire-deficient mice. *J. Immunol.*, **174**, 1862–1870.
25. Yano, M., Kuroda, N., Han, H., Meguro-Horike, M., Nishikawa, Y., Kiyonari, H., Maemura, K., Yanagawa, Y., Obata, K., Takahashi, S. *et al.* (2008) Aire controls the differentiation program of thymic epithelial cells in the medulla for the establishment of self-tolerance. *J. Exp. Med.*, **205**, 2827–2838.
26. Gray, D., Abramson, J., Benoist, C. and Mathis, D. (2007) Proliferative arrest and rapid turnover of thymic epithelial cells expressing Aire. *J. Exp. Med.*, **204**, 2521–2528.
27. Mizuochi, T., Kasai, M., Kokuho, T., Kakiuchi, T. and Hirokawa, K. (1992) Medullary but not cortical thymic epithelial cells present soluble antigens to helper T cells. *J. Exp. Med.*, **175**, 1601–1605.
28. Barrera, L.O., Li, Z., Smith, A.D., Arden, K.C., Cavenee, W.K., Zhang, M.Q., Green, R.D. and Ren, B. (2008) Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome Res.*, **18**, 46–59.

29. Isaac, A., Wilcox, K.W. and Taylor, J.L. (2006) SP100B, a repressor of gene expression preferentially binds to DNA with unmethylated CpGs. *J. Cell. Biochem.*, **98**, 1106–1122.
30. Kumar, P.G., Laloraya, M., Wang, C.Y., Ruan, Q.G., Davoodi-Semiromi, A., Kao, K.J. and She, J.X. (2001) The autoimmune regulator (AIRE) is a DNA-binding protein. *J. Biol. Chem.*, **276**, 41357–41364.
31. Purohit, S., Kumar, P.G., Laloraya, M. and She, J.X. (2005) Mapping DNA-binding domains of the autoimmune regulator protein. *Biochem. Biophys. Res. Commun.*, **327**, 939–944.
32. Ruan, Q.G., Tung, K., Eisenman, D., Setiady, Y., Eckenrode, S., Yi, B., Purohit, S., Zheng, W.P., Zhang, Y., Peltonen, L. *et al.* (2007) The autoimmune regulator directly controls the expression of genes critical for thymic epithelial function. *J. Immunol.*, **178**, 7173–7180.
33. Venanzi, E.S., Melamed, R., Mathis, D. and Benoist, C. (2008) The variable immunological self: genetic variation and nongenetic noise in Aire-regulated transcription. *Proc. Natl Acad. Sci. USA*, **105**, 15860–15865.
34. Taubert, R., Schwendemann, J. and Kyewski, B. (2007) Highly variable expression of tissue-restricted self-antigens in human thymus: implications for self-tolerance and autoimmunity. *Eur. J. Immunol.*, **37**, 838–848.
35. Coppe, A., Danieli, G.A. and Bortoluzzi, S. (2006) REEF: searching REgionally Enriched Features in genomes. *BMC Bioinformat.*, **7**, 453.
36. Pitkanen, J., Rebane, A., Rowell, J., Murumagi, A., Strobel, P., Moll, K., Saare, M., Heikkila, J., Doucas, V., Marx, A. *et al.* (2005) Cooperative activation of transcription by autoimmune regulator AIRE and CBP. *Biochem. Biophys. Res. Commun.*, **333**, 944–953.