Transcriptional programs that control expression of the autoimmune regulator gene Aire

Yonatan Herzig¹, Shir Nevo¹, Chamutal Bornstein¹, Miriam R Brezis¹, Sharon Ben-Hur¹, Aya Shkedý¹, Michal Eisenberg-Bord³, Ben Levi¹, Michael Delacher², Yael Goldfarb¹, Eyal David¹, Leehay Weinberger³, Sergey Viukov³, Shifra Ben-Dor⁴, Matthieu Giraud⁵, Jacob H Hanna³, Achim Breiling⁶, Frank Lyko⁶, Ido Amit¹, Markus Feuerer² & Jakub Abramson¹

Aire is a transcriptional regulator that induces promiscuous expression of thousands of genes encoding tissue-restricted antigens (TRAs) in medullary thymic epithelial cells (mTECs). While the target genes of Aire are well characterized, the transcriptional programs that regulate its own expression have remained elusive. Here we comprehensively analyzed both cis-acting and trans-acting regulatory mechanisms and found that the Aire locus was insulated by the global chromatin organizer CTCF and was hypermethylated in cells and tissues that did not express Aire. In mTECs, however, Aire expression was facilitated by concurrent eviction of CTCF, specific demethylation of exon 2 and the proximal promoter, and the coordinated action of several transcription activators, including Irf4, Irf8, Tbx21, Tcf7 and Ctcfl, which acted on mTEC-specific accessible regions in the Aire locus.

Central tolerance is shaped in the thymus, a primary lymphoid organ, where immature T lymphocytes are ‘educated’ to recognize foreign antigens while tolerating the body’s own components¹. Medullary thymic epithelial cells (mTECs) act as the key mediators of this process² by facilitating both the negative selection of self-reactive T cells and the generation of thymic T regulatory cells³⁻⁵. Specifically, mTECs promiscuously express and subsequently present essentially all self antigens, including those that are normally highly tissue restricted⁶⁻⁷. Expression of genes encoding these tissue-restricted antigens (TRAs) by mTECs is essential for the establishment of immunological tolerance to the same self antigens that T cells would encounter once released to the periphery⁸⁻⁹. Expression of TRA-encoding genes is thought to depend predominantly on a single transcriptional regulator: Aire¹⁰. Aire-deficient mice, whose mTECs express only a fraction of the whole TRA repertoire¹⁰, develop autoantibodies and immune-cell infiltrates directed against multiple peripheral tissues, which resembles Aire deficiency in human patients with autoimmune polyendocrine syndrome type 1 (refs. 10,11).

Although the genes that are targets of Aire in mTECs are now well characterized¹⁰⁻¹², the transcriptional programs that regulate the expression of Aire have remained almost completely unknown. A body of evidence has demonstrated that Aire expression is restricted almost exclusively to terminally differentiated mTECs characterized by high expression of major histocompatibility complex class II (MHCII) and the co-stimulatory molecule CD80 (called ‘mTEC³ cells’ here)⁶⁻¹⁰. Nevertheless, studies of Aire-GFP (Adig B6.Aire-Igrp-GFP) reporter mice¹⁵ and Aire-Cre reporter mice¹⁶, which express green fluorescent protein (GFP) or Cre recombinase, respectively, under the control of the Aire promoter, have allowed the identification of additional Aire-expressing cell populations, including rare MHCII⁺EpCAM⁺CD45⁺ cells in the lymph nodes called ‘extrathymic Aire-expressing cells’¹⁵, CD19⁺Aire⁺ B cells in the thymus¹⁷ and embryonic cells before emergence of the three germ-cell layers¹⁷. However, the expression of Aire, as well as the extent of Aire’s effect on the expression of TRA-encoding genes, is much lower in these non-mTEC populations than in mTEC³ cells.

As for the molecular control of Aire expression, published studies have indicated involvement of the NF-kB signaling pathway and the highly conserved noncoding sequence CNS1 located in the proximal Aire promoter in the control of Aire expression in the mTEC³ population¹⁴⁻¹⁸. Although NF-kB signaling seems to be essential for Aire expression in mTECs, it is clearly not sufficient, and additional transcriptional regulators and molecular mechanisms are probably involved.

In this study, we sought to delineate the molecular architecture underlying Aire expression in mTECs. We found that the Aire locus in the mTEC³ population was characterized by a unique methylation profile and was physically accessible at two specific regions. By combining the gene-expression profiles of Aire-expressing cells (whether mTECs or thymic B cells) with various complementary Aire-promoter-driven reporter assays, we identified over a dozen key putative regulators of Aire expression, and subsequent functional and
biochemical analyses identified several transcription factors, including Irf4, Irf8, Tbx21 and Tc7, required for efficient Aire expression in both HEK 293-FT cells (human embryonic kidney cells engineered to be fast growing and to express the SV40 large T antigen) and in true mTECs through coordinated binding to the Aire proximal promoter. Finally, we found that the Aire locus was actively insulated by CCCTC-binding factor (CTCF) in cells and tissues that did not express Aire and that removal of CTCF from its binding sites at the Aire locus was essential for proper Aire expression in mTECs. Thus, our study provides insight into the complex molecular mechanisms that control the mTEC-specific expression of Aire in the thymus.

RESULTS
DNA-methylation profiles of the Aire locus
For gene transcription to occur, cis-regulatory elements, including the gene promoter and the enhancer regions, must become physically accessible to the trans-acting transcription-regulatory machinery. The accessibility of the cis-regulatory elements and the subsequent efficacy of gene expression is often regulated by methylation of cytosine residues within cytosine-guanosine (CG)-rich regions. To define the putative cis-regulatory elements that control the expression of Aire, we first determined sequence conservation as well as DNA-methylation patterns across the Aire locus (Fig. 1a). Sequence comparison indicated high conservation of the mouse Aire and human AIRE loci within the exonic regions and within a short sequence immediately (0–250 base pairs (bp)) upstream of the transcription start site (TSS). Very little sequence conservation was present further upstream, with the exception of the short conserved noncoding sequence CNS1 located 3–4 kilobases (kb) upstream of the TSS. By employing bioinformatics prediction tools, we identified a distinct and conserved CpG-dinucleotide island (region 1) in the vicinity of the TSS and three additional CpG islands (regions 2–4) in the gene body, spanning exons 2, 5 and 12, respectively. No apparent CpG-rich island was present further upstream of the TSS, including the CNS1 region.

Next, to assess the DNA-methylation status of those specific regions, we isolated genomic DNA from various Aire-expressing cells obtained either from the C57BL/6 (B6) strain of mice or from Aire-GFP mice. Specifically, we isolated Aire-expressing CD45EpCAMhiLy51negloMHCIIhi (mTECbi) cells and CD19Aire-GFP transgenic B cells, as well as cells that did not express Aire, such as CD45EpCAMhiLy51negloMHCIIhi (mTECbi) cells and CD19Aire-GFP thymic B cells, as well as cells that did not express Aire, such as CD45EpCAMhiLy51negloMHCIIhi (mTECbi) cells, CD45EpCAMhiLy51neglo cortical TECs (cTECs), CD3CD4+ splenic T cells and CD19Aire-GFP splenic B cells (Supplementary Fig. 1a,b), and performed locus-specific deep bisulfite sequencing to determine the degree of methylation at each CpG position (Fig. 1b and Supplementary Fig. 1c). In line with published findings, the region immediately adjacent to the TSS (i.e., most of region 1) was predominantly hypomethylated in all cell populations analyzed, including mTECbi cells (Fig. 1b). However, high-resolution analysis indicated that region 1 was hypomethylated on several CG residues in its distal 5′ part in all TEC subsets analyzed, but not in T cells or B cells (Fig. 1b). Moreover, within this region, two specific CG pairs were hypomethylated uniquely in the immature mTECbi population but not in any other population analyzed, including mTECbi cells or cTECs (Fig. 1a), which suggested that these specific residues might become transiently demethylated before Aire expression and might therefore serve as primers for Aire expression. Furthermore, region 2, which spans exons 2 and its vicinity, was hypomethylated specifically in TECs, in particular in the mTECbi cell population, relative to its methylation in all other cell types (Fig. 1b), which suggested that this region might function as a possible regulatory element for Aire expression depending on its methylation status.

To gain additional insight into the functional importance of DNA (de)methylation in the regulation of Aire expression in vivo, we sought to inactivate key components of the DNA-demethylation machinery in the thymic epithelium. We generated triple-transgenic mice with loxP sites flanking individual alleles encoding the three main methylcytosine dioxygenases: Tet1, Tet2 and Tet3 (ref. 23). We crossed those Tet1fl/flTet2fl/flTet3fl/fl mice with Foxn1-Cre mice24 to obtain conditional inactivation of the Tet-enzyme-dependent DNA-demethylation machinery in the thymic epithelium via Foxn1-Cre-mediated recombination. Flow cytometry of thymic cells isolated from 6- to 8-week old mice revealed relatively normal frequencies of individual TEC populations (including Aire+ mTECbi cells) in Foxn1-CreTet1fl/flTet2fl/flTet3fl/fl mice relative to the frequency of such cells in their Cre−(Tet1fl/flTet2fl/flTet3fl/fl) counterparts (Fig. 1c,d), which suggested a limited role for Tet enzymes in the development of Aire+ mTECs. In contrast, however, the intensity of Aire protein was significantly lower in mTECbi cells isolated from Foxn1-CreTet1fl/flTet2fl/flTet3fl/fl mice than in such cells from their Cre−littermates (Fig. 1e), which suggested that Tet-enzyme-mediated DNA demethylation was critical for efficient Aire expression. In agreement with that hypothesis, bisulfite deep sequencing of both mTECbi cell populations and mTECbi cell populations isolated from Foxn1-CreTet1fl/flTet2fl/flTet3fl/fl mice revealed hypermethylation in region 2 (i.e., exon 2) and in the 5′ area of region 1 of Aire relative to the methylation of those regions in such cells from their Cre− counterparts (Fig. 1f). In contrast, deficiency in Tet1, Tet2 and Tet3 did not alter the methylation status of the CpG residues around the TSS, which was uniformly demethylated in all cell types assessed (Fig. 1b). Collectively, these results suggested that active demethylation of the Aire locus at key residues immediately upstream and downstream (i.e., exon 2) of the TSS was required for optimal expression of Aire in mTECs.

Transcription-factor signature of Aire-expressing cells
We hypothesized that the expression of transcription regulators of Aire, would be, like Aire itself, upregulated in mature, Aire-expressing mTECs. To identify such candidates, we employed gene microarrays to analyze the expression profiles of the ~2,300 known transcriptional regulators in sorted mTECbi cell, mTECbi cell and cTEC populations (identified throughout the study as described above, unless specified otherwise). These analyses showed substantial upregulation of the expression of genes encoding about 50 transcriptional regulators in the mTECbi cell population relative to their expression in the immature mTECbi cell and/or cTEC populations (Fig. 2a, Supplementary Fig. 2 and Supplementary Table 1). Next we investigated whether the signature of the 50 mTECbi cell transcriptional regulators noted above was also present in other Aire-expressing cells (specifically, in Aire+ thymic B cells)17. For this, we sorted CD19Aire-GFP+ thymic B cells and CD19Aire-GFP splenic B cells by flow cytometry and subjected them to gene microarray analysis. We assessed whether expression of the 50 mTECbi-cell-specific transcriptional regulators was higher in the Aire-GFP+ thymic B cells than in Aire-GFP+ splenic B cell populations. The results indicated that a large fraction (19) of the 50 genes encoding mTECbi-cell-specific transcriptional regulators (for example, Aire, Tc7, Tbx21 and Myb) had higher expression in Aire+ thymic B cells than in splenic B cells (P < 0.001 (χ2 test); Fig. 2b and Supplementary Table 2). Furthermore, genes encoding a large part of the mTECbi-cell-specific transcriptional regulators (16 of 50), including factors such as SpiB and Irf8, were expressed by both splenic B cells and thymic B cells (Fig. 2b and
We also analyzed the expression of genes encoding the 50 mTEC\textsuperscript{hi}-cell-specific transcriptional regulators in previously published gene microarray data sets (available from the Gene Expression Omnibus (GEO) database) of various populations of immune cells, including B cells, T cells, dendritic cells, macrophages and natural killer cells. Clustering analysis indicated that the signature...
of these 50 transcriptional regulators was unique to mTEC\textsuperscript{hi} cells and overlapped that of the mTEC\textsuperscript{lo} cell and cTEC populations, as well as that of Aire-GFP\textsuperscript{+} B cells, the most (Fig. 2c). Collectively, these data showed that the mTEC\textsuperscript{hi} cell population was characterized by a specific signature of transcriptional regulators that also had high expression in Aire-expressing thymic CD19\textsuperscript{+} B cells.

Figure 2 Specific transcriptional-regulator signature of Aire-expressing cell. (a) Microarray analysis of gene expression in mTEC\textsuperscript{hi} cells relative to that in mTEC\textsuperscript{lo} cells (horizontal axis) or cTECs (vertical axis): gray vertical and horizontal lines indicate no difference in expression; blue lines demarcate a range of a difference in expression of 0.5-fold to 2-fold; gray diagonal line indicates linear correlation; top right, ~50 genes encoding transcriptional regulators (among the ~2,300 known (blue)) upregulated over twofold in mTEC\textsuperscript{hi} cells in both comparisons (arrow, expression coordinates of Aire).

(b) Microarray analysis of the expression of the ~2,300 genes encoding known transcriptional regulators in Aire-GFP\textsuperscript{+} CD19\textsuperscript{+} thymic B cells relative their expression in splenic CD19\textsuperscript{+} B cells, plotted against mean expression in those cells; colors other than gray indicate genes encoding ~50 mTEC\textsuperscript{hi}-cell-specific transcriptional regulators in these populations with low basal expression (yellow; cutoff, <100), or with high basal expression (cutoff, >100) and substantial (red; cutoff, >1.5-fold), moderate (pink; cutoff, >1-fold to 1.5-fold) or negative (blue; cutoff, <1-fold) ‘enrichment’ for expression in Aire-GFP\textsuperscript{+} thymic B cell relative to their expression in splenic B cells. (c) Expression profiles of genes encoding the ~50 mTEC\textsuperscript{hi}-cell-specific transcriptional regulators in various immune-cell populations (above plot), presented (in arbitrary units (AU)) as signal values (key) normalized to the value of the cell with the highest expression of that gene and clustered (left margin) on the basis of their expression profiles. GC, germinal center; DC CD4, CD4\textsuperscript{+} dendritic cells; DC CD8, CD8\textsuperscript{+} dendritic cells; T\textsubscript{reg} cells, regulatory T cells; DP, double positive (CD4\textsuperscript{+}CD8\textsuperscript{+}); NK, natural killer.

Data are representative of one experiment with four biological replicates (a), two experiments with two biological replicates (b) or one experiment with three (publicly available) biological replicates (c).
To further refine the list of putative \textit{Aire} regulators, we screened the 50 mTEC-specific candidates identified above (\textit{Supplementary Tables 1} and \textit{2}) according to their ability to induce \textit{Aire} expression in \textit{Aire}-promoter-based reporter assays. Vectors encoding the individual candidate transcriptional regulators tagged with Flag, V5 or hemagglutinin (HA) tag peptide (\textit{Supplementary Table 3}) were expressed in HEK 293-FT cells along with a luciferase reporter plasmid expressing the entire mouse DNA sequence (~4 kb) encompassing all conserved regions upstream of the \textit{Aire} TSS. The expression of individual transcription factors was confirmed by immunoblot analysis (\textit{Supplementary Table 3}). While most of the candidate regulators had no significant effect on luciferase expression relative to the effect exerted by an empty luciferase vector, one third of the candidate regulators either upregulated (for example, Ctcfl, Myb, Tox4, etc.) or downregulated (for example, Fezf2, Foxp4 or Tbx15) luciferase expression (Fig. 3a,b), which suggested that they acted as positive regulators or negative regulators, respectively, of \textit{Aire} transcription.

We also employed an alternative and complementary reporter system strategy that is sensitive to possible cooperative effects among the individual factors. Specifically, we co-transfected HEK 293-FT cells with a red fluorescence protein (RFP) reporter plasmid containing the promoter sequence of \textit{Aire} (\textit{Supplementary Fig. 3a,b}) plus a ‘cocktail’ of plasmids expressing all of the ~50 candidate factors, to produce random expression of each of the 50 candidate factors. Sorted RFP\textsuperscript{hi} or RFP\textsuperscript{lo} HEK 293-FT cells were analyzed by reverse-transcription quantitative PCR for higher expression of the individual candidate factors in the RFP\textsuperscript{hi} population than in the RFP\textsuperscript{lo} population (Fig. 3c,d and \textit{Supplementary Fig. 3b}). The RFP\textsuperscript{hi} population showed ‘enrichment’ for about one third of the screened factors (with a cutoff of a difference in expression of 1.5-fold) (Fig. 3e), and these largely overlapped the factors identified in the luciferase-based screen. In addition, the RFP screen assay identified additional candidates (i.e., Mxd1, Irf8, Irf4 and Utf1) that were not identified by the luciferase screen (Fig. 3a,c), which suggested these might act as co-factors in regulating \textit{Aire} expression.

Next, by combining the results of the gene-expression profiling and the \textit{Aire}-reporter analyses, we reduced the number of prospective \textit{Aire} regulators to 15 (Fig. 3e). Clustering analysis of these 15 candidates indicated that they were expressed in a cell-specific (i.e., not ubiquitous) manner throughout the immune system (Fig. 3f) and that they clustered into three main groups according to their expression. Cluster 1 included factors (Cdx1, Ctcfl, Foxn1 and Hnf4g) that, like \textit{Aire}, were expressed either uniquely or predominantly in the mTEC\textsuperscript{hi} cell and/or TEC populations and were typically not expressed in the other immune-cell populations; cluster 2 included factors (Irf4, Irf8, Spib, Rbpj, etc.) that were expressed predominantly in antigen-presenting cells; and cluster 3 included factors (Tbx21, Tcf7 and Myb) that were restricted mainly to T cells and/or natural killer cells (Fig. 3f). Moreover, all the factors in clusters 2 and 3 had high and uniform expression in both mTEC\textsuperscript{hi} cell and Aire-GFP\textsuperscript{lo} thymic B cell populations but not in other cell populations analyzed (\textit{Supplementary Fig. 3c}). These results suggested that \textit{Aire} expression probably depended on a unique combination of various transcription factors co-expressed specifically in mTECs (and, to some extent, in Aire\textsuperscript{lo} thymic B cells).

\section*{Induction of endogenous \textit{AIRE} by transcriptional regulators}

We next sought to determine whether ectopic expression of any of the 15 candidates could induce the expression of endogenous \textit{AIRE} mRNA in cells with negligible basal \textit{AIRE} expression. For this, we transfected HEK 293-FT cells with expression vectors encoding the individual candidates and analyzed the expression of endogenous \textit{AIRE} mRNA by quantitative PCR. Because we had found that the \textit{Aire} promoter was demethylated at multiple CG residues uniquely in mTECs relative to its methylation in the other cells that did not express \textit{Aire}, we pre-treated the HEK 293-FT cells with the DNA-methylation inhibitor 5-azacytidine (5-Aza) before transfection. Eight of the top fifteen candidates enhanced the endogenous expression of \textit{AIRE} by >2.5-fold, while three specific factors (Ctcfl, Irf8 and Tox4) enhanced the expression of \textit{AIRE} by more than tenfold (Fig. 4a,b), all relative to its expression in mock-transfected cells. Moreover, treatment with 5-Aza greatly enhanced the ability of these factors to induce \textit{AIRE} expression in the HEK 293-FT cells relative to its expression in similarly transfected but untreated (control) cells (\textit{Supplementary Fig. 4a}), which indicated that demethylation of the \textit{AIRE} locus was critical for efficient \textit{AIRE} expression.

Because multiple factors induced endogenous \textit{AIRE} expression, we next investigated whether the combined action of some of these candidate factors further potentiated its expression. We selected the candidates that elevated endogenous \textit{AIRE} expression by >2.5-fold in HEK 293-FT cells, co-transfected HEK 293-FT cells with vectors encoding those candidates in various paired combinations and analyzed their combined ability to activate the \textit{Aire} promoter in the Aire-RFP reporter system. Pairs composed of co-expression of Ctcfl or Irf8 or Irf4 with most of the other candidates exceeded the additive effect of the given combination on RFP induction relative to the sum of their individual effects (Fig. 4c,d). On the basis of those data, we selected Ctcfl and Irf8 as the strongest candidates for synergistic activation of the \textit{Aire} promoter and thus analyzed their combined ability to induce the endogenous expression of \textit{AIRE} mRNA in HEK 293-FT cells. Co-transfection of expression vectors encoding Irf8 and Ctcfl into 5-Aza-treated HEK 293-FT cells increased the expression of endogenous \textit{AIRE} mRNA by 1,000- to 2,000 fold relative to its expression in their mock-transfected (control) counterparts (Fig. 4e), which suggested that Irf8 and Ctcfl acted in synergy with each other.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure3.png}
\caption{In-vitro reporter assays highlighting putative activators of the \textit{Aire} promoter. (a) Luciferase activity in HEK 293-FT cells co-transfected with an \textit{Aire}-expressing luciferase reporter vector and individual expression vectors encoding candidate regulators (horizontal axis), plus a positive (constitutive promoter) or negative (no promoter) control plasmid, for determination of the range of signal intensity; transfection efficiency was normalized to that of a renilla reporter vector, and results were normalized to those of cells transfected with reporter plasmid only (dashed line, twofold cutoff). (b) Distribution of candidate regulators in a with an inducing effect (>2-fold), repressing effect (<0.5-fold) or no effect (key) on luciferase expression. (c) Quantitative PCR analysis of genes encoding candidate regulators (horizontal axis) in HEK 293-FT cells transfected with an \textit{Aire}-RFP reporter plasmid and with a ‘cocktail’ of plasmids encoding those regulators and sorted by flow cytometry as RFP\textsuperscript{hi} or RFP\textsuperscript{lo}. Results are presented as those in RFP\textsuperscript{hi} cells relative to those in RFP\textsuperscript{lo} cells (RFP\textsuperscript{hi}/RFP\textsuperscript{lo}) (dashed line, 1.5-fold cutoff). (d) Distribution of candidate regulators in \textit{c} showing relative ‘enrichment’ (Inducing; >1.5-fold cutoff), ‘under-representation’ (Repressing; <0.5-fold cutoff) or no difference (No effect) in expression in sorted RFP\textsuperscript{hi} cells relative to their expression in RFP\textsuperscript{lo} cells. (e) Overlap of ‘inducing’ candidate regulators from \textit{b} (those that induced the \textit{Aire} luciferase reporter (\textit{Aire}-Luc\textsuperscript{+}); blue and/or \textit{d} (‘enriched’ in RFP\textsuperscript{hi} cells; red) and/or those expressed in both mTEC\textsuperscript{hi} cells and Aire-GFP\textsuperscript{lo} CD19\textsuperscript{-} thymic B cells (green). (f) Expression profiles of genes encoding ‘inducing’ candidate regulators from \textit{b} or \textit{d} in various immune-cell populations (above plots), presented as signal values (key; as in Fig. 2c) and clustered through pairwise average linkage (top and left margin). Data are representative of three independent experiments (error bars (\textit{a,c}), s.e.m. of \textit{n} = 3 biological replicates).}
\end{figure}
Moreover, co-expression of Irf8 and Ctcfl with Hnf4g or Tbx21 or Tcf7 further enhanced the endogenous expression of AIRE mRNA by an additional 1.2- to 2-fold relative to its expression after co-expression of only Irf8 plus Ctcfl (Fig. 4f and Supplementary Fig. 4b). Those triple combinations, in turn, resulted in enhanced expression of Aire-dependent genes, including KRT14 and ALOX12 mRNA, relative to their expression in mock-transfected cells (Supplementary Fig. 4c–e). Other factor combinations assessed that did not include Irf8 or Ctcfl induced endogenous AIRE mRNA expression to a lesser extent than that seen with co-expression of Irf8 and Ctcfl (Supplementary Fig. 4f–i).
Collectively, these data demonstrated that a large fraction of candidate regulators predicted by the Aire reporter systems were able to induce endogenous AIRE expression in HEK 293-FT cells. Moreover, that ability was substantially potentiated by the combined action of some of these candidate factors, in particular Irf8 and Ctcfl.

**Regulation of Aire expression in vivo by Irf4, Irf8, Tbx21 and Tcf7**

Next, to determine whether the candidate factors identified above were critical for the regulation of Aire in vivo, we studied Tbx21−/−, Tcf7−/−, Foxn1-Cre Irf4fl/fl and Foxn1-Cre Irf8fl/fl mice, which bear germline or TEC-specific24 deletion of each candidate factor-encoding gene. Intracellular staining of Aire followed by flow cytometry of the CD45+ EpCAM+ thymic epithelium in the strains above indicated no significant change in the frequency of mTECΔ or Aire−mTECΔ populations relative to their frequency in the wild-type or Cre− littermates of those mice (Fig. 5a,b). This suggested that deletion of these transcription factors did not affect the development of mTECΔ cells. However, the mean fluorescence intensity of Aire in the Aire+ fraction of the mTECΔ cell population was significantly lower in most of the mutant strains than in its counterpart in their wild-type

![Figure 4](https://example.com/fig4.png)

**Figure 4** Effect of candidate transcriptional regulators on the induction of endogenous AIRE. (a) Quantitative PCR analysis of AIRE in 5-Aza-treated HEK 293-FT cells transfected with expression vectors encoding various candidate factors (horizontal axis); results were normalized to those of the control gene HPRT and are presented relative to that of cells transfected with empty vector (+). (b) Distribution of candidate factors in a that induced (Upregulating; >2.5-fold cutoff) or had no effect (key) on AIRE expression. (c) Frequency of RFP+ HEK 293-FT cells (log2; values; key) following co-transfection of the Aire-RFP reporter plasmid with dual combinations of vectors encoding candidates inducing AIRE expression in a (top and right margins). (d) RFP intensity of RFP+ HEK 293-FT cells treated as in c, presented as calculated observed versus predicted intensity (omega value (Ω); key), showing intensity higher than expected (yellow) or lower than expected (blue) resulting from co-transfection compared with that resulting from single transfection. (e,f) Quantitative PCR analysis of AIRE mRNA in 5-Aza-treated HEK 293-FT cells transfected with empty vector or expression vectors encoding Irf8 and/or Ctcfl (horizontal axis) (e) or co-transfected with a combination of plasmids encoding Irf8 and Ctcfl plus a given candidate factor (horizontal axis) (f); results normalized as in a) are presented relative to that of cells transfected with empty vector (e) or Irf8 plus Ctcfl only (set as 100%; f). *P < 0.001 (Student’s t-test). Data are representative of at least three independent experiments with n ≥ 2 biological replicates (error bars a,e,f, s.e.m.).
or Cre\textsuperscript{-} littermates (Fig. 5c), which indicated that they were required for Aire expression in mTEC\textsuperscript{hi} cells. Specifically, inactivation of Tbx21, Tcf7 or Irf4 alone reduced the expression of Aire in mTEC\textsuperscript{hi} cells by 30–50% relative to its expression in mTEC\textsuperscript{hi} cells from the wild-type or Cre\textsuperscript{-} littermates of the mutant mice, while inactivation of Irf8 had no significant effect on Aire expression in mTEC\textsuperscript{hi} cells (Fig. 5c). Because Irf4 and Irf8 compensate for each other's loss of function in various contexts, we next analyzed Aire expression in mTECs isolated from Foxn1-Cre\textsuperscript{+} Irf8\textsuperscript{fl/fl}Irf4\textsuperscript{fl/fl} mice, which lack expression of both transcription factors in TECs. The mean fluorescence intensity of Aire in mTEC\textsuperscript{hi} cells from Foxn1-Cre\textsuperscript{+} Irf8\textsuperscript{fl/fl}Irf4\textsuperscript{fl/fl} mice was reduced by \sim 45% relative to its expression in mTEC\textsuperscript{hi} cells from their Cre\textsuperscript{-} littermates (Fig. 5c), which represented a \sim 35% reduction in Aire expression relative to that achieved by loss of Irf4 alone. We did not observe a reduction in TRA expression in the mutant strains above (data not shown), in accord with published data showing that a 50% decrease in Aire expression, such as that in Aire\textsuperscript{+/−} mice\textsuperscript{26} (Supplementary Fig. 5), is still sufficient for normal expression of TRA-encoding genes and maintenance of self-tolerance. Together these results indicated that Irf4, Irf8, Tbx21 and Tcf7 were required for efficient Aire expression in vivo.

Binding of Tbx21-Tcf7-Irf4-Irf8 to the Aire promoter

Next, to identify the chromatin regions in the Aire locus that become accessible in mature mTECs but not in other cells, we performed ATAC-Seq (assay for transposase-accessible chromatin with high-throughput

**Figure 5** Candidate transcriptional regulators control Aire expression in vivo. (a) Flow cytometry of CD45\textsuperscript{-}EpCAM\textsuperscript{+}Ly51\textsuperscript{−}mTECs isolated from the thymus of 6- to 7-week old Tbx21\textsuperscript{−/−} (Tbx21-KO) mice (n = 3), Tcf7\textsuperscript{−/−} (Tcf7-KO) mice (n = 3), Foxn1-Cre\textsuperscript{+} Irf8\textsuperscript{fl/fl}Irf4\textsuperscript{fl/fl} (Irf4-cKO) mice (n = 6), Foxn1-Cre\textsuperscript{+} Irf8\textsuperscript{fl/fl}Irf4\textsuperscript{fl/fl} (Irf4-Irf8-dKO) mice (n = 2) (right column) and their wild-type (WT) or Cre\textsuperscript{-} littermates (n \geq 3 mice per genotype) (left column). Numbers in outlined areas indicate percent Aire\textsuperscript{+} cells in the total EpCAM\textsuperscript{+} fraction. (b) Frequency of Aire\textsuperscript{+} cells in the mTEC\textsuperscript{hi} cell fraction (gated on EpCAM\textsuperscript{+}MHCII\textsuperscript{hi}) of mice as in a (key). (c) Median fluorescence intensity of Aire staining in Aire\textsuperscript{+} mTEC\textsuperscript{hi} cells from mice as in a (key), presented relative to that of their wild-type or Cre\textsuperscript{-} littermates, set as 100%. Each symbol (b,c) represents an individual mouse; small horizontal lines indicate the mean (± s.e.m.). *P < 0.05, **P < 0.01 and ***P < 0.001 (Student's t-test). Data are representative of at least three independent experiments.
sequencing) of CD45−EpCAM+Ly51neg−lo mTEC\(^b\) and mTEC\(^b\) cells, CD45−EpCAM+Ly51hi cTECs and splenic CD4\(^+\) T and CD19\(^+\) B cells sorted from B6 mice. Essentially no reads were mapped to the Aire locus in any population analyzed except for mTEC\(^b\) cells (Fig. 6a), in which a large number of reads were mapped to several positions within the Aire locus, including the TSS, CNS1 and the gene body (Fig. 6a). These data therefore suggested that while Aire locus was blocked in the cells that did not express Aire, it became accessible for putative binding by transcriptional regulators at TSS and CNS1 regions.
as well as the upstream CNS1 region. All of the candidate regulators analyzed showed substantial enrichment in a region (from −400 bp to +400 bp) spanning the TSS of AIRE, relative to results obtained with IgG (Fig. 6b). None of the transcription factors showed any apparent enrichment in more distal regions of the AIRE locus, including the upstream CNS1 region, relative to results obtained with IgG (Fig. 6c). Furthermore, binding of all candidates to the AIRE promoter was comparable with or without 5-Aza treatment, except for Ctcfl, whose binding was most efficient after treatment with 5-Aza (Supplementary Fig. 6); this suggested that Irf4, Irf8, Hnf4g, Tbx21 and Tcf7 bound to the AIRE locus independently of DNA demethylation.

To confirm that the genomic interactions noted above also occurred in mTECs, we next used an ‘indexing-first’ ChIP protocol, which allows highly sensitive ChIP of rare cell populations, to assess occupancy of Irf8 in the genome of mTEC<sup>hi</sup> cells and mTEC<sup>lo</sup> cells sorted from B6 thymi. Irf8 showed enrichment at the TSS and the CNS1 regions of the Aire locus in mTEC<sup>hi</sup> cells relative to its abundance at these regions in mTEC<sup>lo</sup> cells (Fig. 6d). Because these regions were identified by ATAC-seq as being accessible for binding specifically in mTEC<sup>hi</sup> cells, these data suggested that all of the candidate factors analyzed (Ctcfl, Irf4, Irf8, Hnf4g, Tbx21 and Tcf7) were able to bind, either directly or indirectly, to the conserved proximal promoter sequence of Aire.

Next we assessed whether Ctcfl, Irf4, Irf8, Hnf4g, Tbx21 and Tcf7 physically interact with each other; for this we used co-immunoprecipitation analysis of HEK 293-FT cells transfected with various pairs of expression plasmids encoding the six candidate factors noted above, labeled with Flag, V5 or HA. Immunoblot analysis of the co-immunoprecipitated samples showed that while Irf4, Irf8, Hnf4g, Tbx21 and Tcf7 physically interacted with each other, Ctcfl showed no apparent binding to any of the other candidates (Fig. 6e). We then performed co-immunoprecipitation analysis of mouse samples enriched for TECs, obtained by magnetic enrichment of thymic fractions for EpCAM<sup>+</sup> cells, or of mouse liver tissue (as a control); for these assays, we used Irf4<sup>−/−</sup>, Irf8<sup>−/−</sup>, Tbx21<sup>−/−</sup> or Tcf7<sup>−/−</sup> specific antibodies or IgG (as a control antibody). Immunoblot analysis of the co-immunoprecipitated samples indicated that all four regulators physically paired with each other in the thymic epithelium (Fig. 6f), which suggested that they form a thymus-specific multi-molecular complex. Together these data suggested that Irf4, Irf8, Tbx21 and Tcf7 bind to the Aire promoter and regulate its expression in concert, possibly as a multi-protein complex. The lack of binding of Ctcfl to the other candidates suggested that Ctcfl is not part of the complex but instead operates through a parallel mechanism.

Eviction of CTCF from the Aire locus is critical for Aire expression

Because Ctcfl has effects known to be antagonistic to its ubiquitously expressed paralog, CTCF, which often operates as a general insulator in the genome, we tested whether Ctcfl had a direct role in regulating Aire expression. We first used HEK 293-FT cells treated with 5-Aza or not. Chromatin from HEK 293-FT cells transfected with expression vectors encoding HA-tagged Ctcfl, Hnf4g, Irf8, Irf4, Tbx21 or Tcf7 was immunoprecipitated with a ChIP-grade HA-specific antibody, and the enrichment for each individual candidate (relative to its abundance obtained by ChIP with an immunoglobulin G (IgG) control antibody) was assessed by quantitative PCR with primers spanning the AIRE locus, 1 kb upstream and 1 kb downstream of the TSS, and are presented relative to those of cells transfected to N.T. siRNA (relative) or for CTCF in putative binding sites at the AIRE locus independently of DNA demethylation.

Next we analyzed the ability of the individual transcription factors to physically associate with the open chromatin regions within the Aire proximal promoter. Because chromatin immunoprecipitation (ChIP) targeting transcription factors in mTECs, which are very rare, is still technically very challenging, we first used HEK 293-FT cells treated with 5-Aza or not. Chromatin from HEK 293-FT cells transfected with expression vectors encoding HA-tagged Ctcfl, Hnf4g, Irf8, Irf4, Tbx21 or Tcf7 was immunoprecipitated with a ChIP-grade HA-specific antibody, and the enrichment for each individual candidate (relative to its abundance obtained by ChIP with an immunoglobulin G (IgG) control antibody) was assessed by quantitative PCR with primers spanning the AIRE locus, 1 kb upstream and 1 kb downstream of the TSS, and are presented relative to those of cells transfected to N.T. siRNA (relative) or for CTCF in putative binding sites at the AIRE locus independently of DNA demethylation.
which suggested that the Ctcfl-induced expression of Aire was probably mediated through disruption of the binding of CTCF to the AIRE locus. Moreover, the Ctcfl-induced eviction of CTCF in HEK 293-FT cells was accompanied by enhanced binding of Irf4-Irf8-Tbx21-Tcfl to the AIRE TSS, as revealed by ChIP assessing enrichment for the individual candidates in the presence or absence of Ctcfl expression (Supplementary Fig. 7a); this suggested that CTCF inhibited the binding of those factors to the AIRE locus. Co-expression of Irf4, Irf8, Tbx21 and Tcfl in HEK 293-FT cells resulted in eviction of endogenous CTCF, in contrast to its lack of eviction in mock-transfected control cells (Supplementary Fig. 7b). This suggested a bi-directional and dynamic mode of regulation in which the transcription factors noted above potentiated eviction of CTCF from the Aire locus, while their own binding was in turn potentiated by eviction of CTCF.

Next, to validate the hypothesis that binding of CTCF to the AIRE locus was able to repress its expression, we used CTCF-specific small interfering RNA to downregulate CTCF expression in 5-Aza-treated HEK 293-FT cells (Supplementary Fig. 7c). Knockdown of CTCF significantly increased AIRE mRNA expression in 5-Aza-treated HEK 293-FT cells without a significant effect on the endogenous expression of any of the other candidate regulators (Fig. 7e and Supplementary Fig. 7d–h). Moreover, the CTCF-induced increase in AIRE mRNA expression was comparable to that mediated by transfection of Ctcfl-encoding vector into HEK 293-FT cells (Fig. 7e), which suggested that decreased expression of CTCF or its eviction was required for maximal AIRE expression.

Finally, to determine whether CTCF was evicted from the Aire locus in the mTEC hi cell population, we performed ChIP followed by PCR to assess occupancy by CTCF at various genomic positions in sorted CD45−EpCAM+Ly51lo−mid mTEC hi cells and splenic CD4+ T and CD19+ B cells from B6 mice. The binding of CTCF to its two binding sites in the Aire TSS and the downstream 3′ site, but not its binding to sites in other genomic regions, was much lower in the mTEC hi cell population than in either the B cells or the T cells (Fig. 7f). These results indicated that the CTCF-mediated insulation of Aire was removed specifically in mTEC hi cells. In summary, our data revealed that Aire expression depended on a rather complex and layered mode of regulation that involved both cis-acting and trans-acting transcriptional repressors and activators (Supplementary Fig. 8).

DISCUSSION

The emergence of Aire as the key mediator of the expression of TRA-encoding genes in the thymus more than a decade ago revolutionized the understanding of how tolerance to non-uniform self antigens is established. The unique ability of Aire to induce or boost the expression of ~20% of the genome positions it as one of the most ‘aggressive’ transcriptional regulators. Here we used bioinformatics, molecular, biochemical and in vivo functional studies to investigate the molecular mechanisms and the transcriptional programs that control Aire expression. Our analysis identified several CpG residues located ~300 bp upstream the TSS and in exon 2 of the Aire locus that were demethylated exclusively in mTECs but not in other cells, including thymic Aire+ B cells. Such demethylation took place as early as in the mTEC lo population, which is composed mainly of immature mTECs, suggestive of potential priming of the Aire locus before expression of Aire and/or most of the candidate Aire regulators. The importance of demethylation of Aire at the promoter and exon 2 regions was also highlighted by the finding that TEC-specific ablation of the three key members of the DNA-demethylation machinery (Tet1, Tet2 and Tet3) resulted in hypermethylation of these specific regions and a subsequent reduction in the expression of Aire. We note that exon 2 is missing in Aire-GFP reporter mice, which have been instrumental in the identification of several Aire-expressing non-mTEC populations, such as extrathymic Aire-expressing cells and thymic Aire+ B cells. Indeed, these reporter mice are known to have a more permissive expression of the GFP-reporter-encoding transgene than of actual endogenous Aire protein, which suggests that either diminished methylation of exon 2 or its complete elimination potentiates Aire expression. This suggests that exon 2 serves as an important regulatory element in the control of Aire expression, possibly through recruitment of putative CpG-binding repression factors.

Our data also demonstrated insulation of the Aire locus by CTCF at two prominent locations, the TSS and the 3′ end, in cells and tissues that did not express Aire. CTCF has a critical role in organizing global chromatin architecture through the formation of genomic loops, which typically block interaction between gene enhancers with their corresponding TSSs. Notably, CTCF binds only to demethylated DNA, which is well in line with our data demonstrating that the Aire locus was mostly demethylated around its TSS (i.e., one of the CTCF-binding sites) in both cells that expressed Aire and those that did not express Aire. The critical role of CTCF in insulating the Aire locus was amply indicated by CTCF-specific knockdown or its eviction by Ctcfl from the AIRE locus, which each resulted in significantly higher AIRE expression in HEK 293-FT cells. More notably, our in vivo data confirmed that CTCF was indeed evicted from both the 5′ end and the 3′ end of the Aire locus in mTEC hi cells.

Moreover, our ATAC-Seq revealed that the Aire locus was generally ‘locked’ in cells that did not express Aire, including mTEC lo cells. In contrast, in mTEC hi cell population, the locus became accessible at two key positions spanning the TSS and the enhancer (CNS1) region. Those accessible regions might in turn serve as platforms for the binding of various transcriptional regulators that are critical for the transactivation of Aire. Indeed, many of the transcription factors identified here as potential activators of Aire (Irf4, Irf8, Tbx21 and Tcfl) were found to form a multi-molecular complex that bound to the Aire TSS. It is noteworthy that except for mTEC hi cells (and, to a large extent, also thymic Aire+ B cells), other cells or tissues did not co-express most of these transcription factors. Notably, our in vivo results demonstrated that Aire expression was 30–50% lower in mTECs deficient in some of the key Aire regulators, including Tbx21, Tcfl, Irf4 or Irf8. Although that reduction was significant, it was not reflected by a similar reduction in the expression of Aire-dependent TRA-encoding genes. Such results, however, are not unexpected, given that Aire−/− mice show normal TRA expression and no autoimmune phenotype despite their 50% lower expression of Aire. Similarly, people with diverse recessive heterozygous mutations in AIRE show no signs of autoimmunity, which suggests that a reduction of even 50% in Aire is sufficient for normal expression of TRA-encoding genes and maintenance of self-tolerance.

Finally, it is likely that additional transcription factors or regulators that were either not studied in depth here (for example, Tox4 or Mxd1) or were not highlighted in our study might also participate in regulating Aire expression. This might in particular involve members of the NF-kB signaling pathway, which has been linked to the regulation of Aire expression via the CNS1 enhancer region located ~3 kb upstream of the Aire TSS.
Identification of a novel cis-regulatory element essential for Biphasic Aire expression in early embryos and in medullary thymic epithelial cells

Acknowledgments
We thank M. Anderson (University of California at San Francisco) for Aire-Igrp-GFP mice; D. Graf (University of Zurich) for B6.Foxn1-Cre mice (obtained with the consent of N. Manley (University of Georgia)); H. Clevers (Hubrecht Institute) for Tcf7-fl/fl mice; and Y. Peleg, G. Yona and V. Krupalnik for experimental expertise and help. Supported by the Israel Science Foundation (1825/10 and 1376/13), the Sy Syms Foundation, the Dr. Ceila Zwillenberg-Fridman and Dr. Lutz Fridman Career Development Chair (J.A.), the Weizmann-German Cancer Research Center PhD fellowship program (Y.H., M.D., J.A., M.F.), the German Cancer Research Center–Israel Ministry of Science and Technology foundation for German-Israeli co-operation (2431 to J.A. and M.F.), the Agence Nationale de Recherche (2011-CHEX-001-R12004KK to M.G.) and the European Federation of Immunological Societies fellowship program (M.R.B.).

Author Contributions
Y.H. and J.A. designed the study and wrote the manuscript; Y.H. performed most of the experimental work; S.N. performed several experiments, including ChIP, protein immunoprecipitation and Aire intracellular staining; C.B. performed the ‘indexing-first’ ChIP and ATAC-Seq experiments; M.R.B set up and conducted luciferase-based assays and analyzed the Tbx21-mutant mice; L.W. and S.V. performed the Tcf7-fl/fl mice; D. Graf (University of Zurich) for B6.129S7-Grox2m1Cby/To mice; and Y. Peleg, G. Yona and V. Krupalnik for experimental expertise and help. Supported by the Israel Science Foundation (1825/10 and 1376/13), the Sy Syms Foundation, the Dr. Ceila Zwillenberg-Fridman and Dr. Lutz Fridman Career Development Chair (J.A.), the Weizmann-German Cancer Research Center PhD fellowship program (Y.H., M.D., J.A., M.F.), the German Cancer Research Center–Israel Ministry of Science and Technology foundation for German-Israeli co-operation (2431 to J.A. and M.F.), the Agence Nationale de Recherche (2011-CHEX-001-R12004KK to M.G.) and the European Federation of Immunological Societies fellowship program (M.R.B.).

Competing Financial Interests
The authors declare no competing financial interests.
ONLINE METHODS

**Mice.** All mice used in this study were maintained under specific pathogen-free conditions at the Weizmann Institute's animal facility and were handled in accordance with the guidelines of the Institutional Animal Care and Use Committee (0136031-1). Wild-type C57BL/6 (B6) mice were from Harlan Laboratories. B6.Aire-Igrp-GFP reporter (Adig) mice 

**Flanking cell lineages**

For genomic DNA bisulfite sequencing, equimolar amounts of all amplicons were combined in a single tube and processed on a GS Junior sequencer (Roche) according to the manufacturer's instructions. Sequence reads were aligned and average methylation levels per covered CpG were displayed as color-coded heat maps.

**Gene-expression profiling.** The mTEC^lo^ and mTEC^hi^ cell populations were sorted from thymic suspensions of the appropriate mouse strains on a FACS Aria III cell sorter (BD) as detailed above. Typically, cells from three mice were pooled to obtain 30 × 10^6 to 60 × 10^6 sorted mTECs. Similarly, Aire^−/−^ thymic B cells were isolated from the thymi of Adig mice by first magnetically enriching for B cells using anti-CD19 MACS beads (130-052-201, Miltenyl Biotec) followed by cell sorting to obtain 30 × 10^6 to 60 × 10^6 sorted GFP^+^ cells. All cell populations were then analyzed by gene-expression profiling. Specifically, total RNA was extracted from ~30 × 10^6 pooled sorted cells with Trizol. Purified total RNA was then amplified with a MessageAmp RNA kit (Ambion). Biotinylated cRNA was then hybridized to Affymetrix Mouse Gene 1-ST arrays by the genomics core at the Weizmann Institute. Raw data were processed with the RMA algorithm for probe-level normalization and were analyzed with GenePattern software. Analysis and presentation of expression profiles of genes inspected in various cell populations of the immune system was done with GenePattern server using profiles as described above, alongside existing data sets available from the GEO database (accession code, GSE15907).

**RNA and real-time PCR analysis.** RNA was extracted from cells with Trizol reagent according to the manufacturer's instructions (Life Technologies), and the purified RNA was then used for cDNA synthesis with the High-Capacity cDNA Reverse-Transcription kit (Life Technologies) using random primers. The subsequent quantitative PCR analysis was performed with Fast SYBR Green Master Mix (Life technologies) and specific primers (complete list, Supplementary Table 6). Quantification of AIRE expression was performed by TaqMan assay (Hs0030829_m1 for AIRE and Hs99999909_m1 for HPRT) with the TaqMan Fast Universal PCR Master Mix (Life Technologies) according to the manufacturer's instructions. All samples were analyzed on the ViiA7 real-time PCR instrument (Life Technologies) and differential expression was calculated according to the change-in-cycling-threshold (ΔΔCT) method.

**Cloning.** An expression vector library composed of ~50 genes encoding candidate transcriptional regulators was generated by amplifying PCR fragments of each cDNA from total thymic epithelial cDNA generated using Superscript III (Life Technologies). Each PCR amplicon was flanked by either att B sequences to allow Gateway restriction-free based cloning (Life Technologies) or NotI and NsiI sites, to allow restriction based cloning. Each cDNA was subsequently cloned either into the Gateway pLenti6-DEST expression plasmid (Life Technologies) or pEF1a-IRESPuro expression plasmid (purchased from Addgene). pLenti6/V5-Dest was purchased from Life Technologies. pLenti6/V5-Flag-Dest was generated by replacing the V5 tag with a Flag tag flanked by XhoI and AgeI restriction sites sequences, and the vector was subsequently transformed into ccdB-survival competent bacteria (Life Technologies) to enable vector propagation. Thus, each cloned cDNA was C-terminally tagged with a V5, Flag or HA epitope. Protein expression of all plasmids was confirmed by immunoblot analysis.

Construction of reporter vectors under the control of the Aire promoter, DNA sequence encompassing 4.2 kb of the Aire promoter was amplified from a BAC clone (CH29-218, CHORI) using LongAmp Taq DNA Polymerase (NEB). To generate the pG3L Aire.Luciferase reporter, the amplicon was subsequently inserted into the pGL3-Basic Vector (Promega) using XhoI and HindIII sequences. To generate the pDsRed Aire.REP reporter, the amplicon was inserted into the pDsRed-Monomer-Hygro-C1 vector (Clontech) using NotI sequences, replacing the pre-existing CMV promoter. Site-directed deletion of predicted transcriptional regulator binding sites in the pDsRed Aire. REP vector was conducted by using reverse-complement primers, that have 5' phosphorylation modification, and that amplify the entire vector save the DNA region of choice. The PCR product was subsequently ligated to itself using T4 DNA Ligase (M0202S NEB).

A complete list of cloning primers is in Supplementary Table 7.

**Cell culture and transfection.** Cell cultures were kept at 37 °C in a humidified atmosphere with 5% CO₂, HEK 293-F cells (Life Technologies) were cultured...
with Dulbecco’s modified Eagle’s medium (DMEM) with 4.5 g/l glucose and t-glutamate (Life technologies) complemented with 10% FBS (FBS) (Life technologies) and either with or without 1μM 5-azacytidine (A2385 Sigma). Cells were transfected 24 h later with the relevant plasmid using TransIT-293 Transfection Reagent (MIR 2700, Mirus) or PEI (764965, Sigma) according to manufacturer’s protocol. Cells were harvested 48 h after transfection and used for the intended assay. For transfection of small interfering RNA, cells were transfected upon plating with the appropriate small interfering RNA set (CTCF: L-020165; Non-targeting: D-001810-1; Dharmacon ON-TARGET plus, Thermo Scientific) at a final concentration of 50 nM using jetPRIME (Polyplus) reagent.

Luciferase reporter assay. HEK 293-FT cells were seeded onto 24-well plates and incubated overnight. Each well was transfected with 400 ng of the pGL3 AireLuciferase reporter plasmid or with an empty and/or constitutive luciferase control vectors. To that were added 30 ng of one of the candidate expression vectors and 20 ng of renilla plasmid, to control for variability in transfection efficiency. Luciferase activity was measured 48 h after transfection. Cells were lysed with Passive Lysis Buffer, 5× (Promega) and luciferase or renilla signal was measured in a black 96-well plate in Infinite 200 PRO multimode reader (Tecan). The Luciferin substrate and ATP buffer were injected to the plate (30 ml of 100 mM Tris acetate, pH 7.8, 50 ml of 100 mM magnesium acetate, 1 ml of 0.5 M EDTA, Luciferin (ornat) and ATP (Sigma). The Renilla substrate buffer was injected as control (400 ml of ddH2O, 40 ml of 1 M K2HPO4, 10 ml of 1 M KH2PO4, and Renilla substrate (Coelenterazine, Promega)).

ChIP and deep sequencing. ChIP was done as previously described32. 1.5 x 10^6 HEK 293-FT cells or 3 x 10^5 primary cells (mTECs, B cells and T cells) were used for each ChIP experiment. Cells were fixed for 10 min with 1% formaldehyde (Thermo Scientific, PI-28908), quenched with glycine and washed with ice-cold PBS. Pellets were resuspended in RIPA lysis buffer supplemented with protease inhibitor. Cells were lysed for 10 min on ice and the chromatin was sheared. The sonicated cell lysate was cleared by centrifugation and mixed with 75 μl of protein G magnetic dynabeads (Life Technologies) coupled to 5 μl of anti-HA (HA7, Sigma), anti-CTCF (Upstate, 07-729), anti-Irf8 (BL12058, Bethyl) or IgG control antibody (Upstate, 12-370) per 200 μl of cell lysate and incubated for 4 h at 4°C. Using 96-well magnets, unbound cell lysate was removed and samples were washed five times with cold RIPA, twice with high-salt RIPA, twice with LiCl buffer, once with TE, and then eluted in 50 μl of direct elution buffer. The eluate was then treated sequentially with 2 μl of RNaseA for 30 min and 2.5 μl of proteinase K for 2 h and was then reverse-crosslinked at 65 °C overnight. Solid-phase reversible immobilization (SPRI) cleanup steps were performed in eight-well strips. 90 μl protein-G-coated Sepharose beads were added to the reverse-crosslinked samples, mixed and incubated for 2 min. Supernatant was separated from the beads using a 96-well magnet for 4 min. The DNA was eluted in 50 μl EB buffer. The relative enrichment of the different genomic regions along the Aire promoter in each of the samples was measured by real-time PCR.

‘Indexing-first’ ChIP of cell populations sorted by flow cytometry was conducted as previously described37. Reads were then aligned to the Mouse Genome version mm9 (NCBI37).

Assay for transposable-accessible chromatin using sequencing (ATAC-Seq) on 1 x 10^6 flow-cytometry-sorted cells from indicated populations was conducted as previously described38 and is presented alongside existing data sets available from the GEO database (accession code, GSE59636).

Bioinformatics and in-silico analysis. Aire sequence conservation was analyzed and is presented using the VISTA comparative genomics tool34. Prediction of CpG islands along the Aire promoter and gene body was performed with the DBCAT tool15. Expected vs. Observed (Omega value) of the frequency of RFP cells resulting from co-transfection of two candidate regulators was calculated by subtraction of the observed value from the multiplication of the value of each individual candidate transfected alone. Visualization of ATAC-seq results or of Irf8-binding sites in the Aire locus was done using ‘Integrative Genomic Viewer’ (IGV)36. Visualization of CTCF binding sites flanking the Aire gene was done in the UCSC genome browser37 using the LCR TFBS track38 and CTCF ChIP-seq data sets available from GEO.

Statistical analysis. Determination of the statistical significance of differences between two experimental groups in all experiments conducted in this study was done by a two-tailed Student’s t-test analysis. Determination of significant difference between groups in non-parametric comparisons (Fig. 2b) was done by the χ2 test.

Data availability. The gene-expression microarrays, ‘indexing-first’ ChIP and ATAC-seq data are available in the GEO database with the accession codes GSE89965 (expression microarrays) and GSE90050 (‘indexing-first’ ChIP and ATAC-Seq). Other publicly available data sets used in this manuscript are fully described in subsections above.

31. Dawlaty, M.M. et al. Tet1 is dispensable for maintaining pluripotency and its genomic accessibility, which is critical for the generation of induced pluripotent stem cells. Nat. Biotechnol. 34, 166–175 (2016).