HHS Public Access

Author manuscript

Nat Immunol. Author manuscript; available in PMC 2017 July 30.

Published in final edited form as:

Nat Immunol. 2017 March ; 18(3): 263–273. doi:10.1038/ni.3675.

The transcriptional regulator Aire binds to and activates super-enhancers

Kushagra Bansal, **Hideyuki Yoshida**, **Christophe Benoist***, and **Diane Mathis***

Division of Immunology, Department of Microbiology and Immunobiology, Harvard Medical School, Boston MA 02115

Abstract

Aire is a transcription factor that controls T cell tolerance by inducing the expression of a large repertoire of genes specifically in thymic stromal cells. It interacts with scores of protein partners of diverse functional classes. Here we showed that Aire and some of its partners, notably those implicated in the DNA-damage response, preferentially localized within and activated long chromatin stretches overloaded with transcriptional regulators, known as “super-enhancers”. We also identified topoisomerase 1 as a cardinal Aire partner that co-localized on super-enhancers and was required for the interaction of Aire with all of its other associates. We propose a model that entails looping of super-enhancers to efficiently deliver Aire-containing complexes to local and distal transcriptional start sites.

Medullary thymic epithelial cells (mTECs) are involved in both negative selection of effector T cells and positive selection of regulatory T cells[1]. A unique feature of mTECs, critical for their roles in tolerance induction, is expression of a large fraction of the genome, in particular scores of loci encoding antigens characteristic of fully differentiated parenchymal cells (peripheral-tissue antigens or PTAs)[2–4]. Much of this transcription is driven by Aire[5]. mTECs from Aire-deficient mice and humans show severely compromised PTA expression, causing these individuals to develop autoimmune infiltrates and autoantibodies targeting multiple peripheral tissues[6,7].

Several observations argue that Aire is a transcriptional regulator that operates differently from conventional transcription factors[6,7]. First, unlike traditional factors, the transcriptional impact of Aire on mTECs involves a large, though still select, portion of the genome[2–4]. Experimental introduction of Aire into extra-thymic cells induces expression of large sets of transcripts, which differ from cell-type to cell-type and also diverge from those induced in

*Address correspondence to: Christophe Benoist and Diane Mathis, Division of Immunology, Department of Microbiology and Immunobiology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, cbdm@hms.harvard.edu, Phone: (617) 432-7741, Fax: (617) 432-7744.

AUTHOR CONTRIBUTIONS

K.B. and H.Y. performed the experiments; K.B., C.B. and D.M. designed the study, analyzed/interpreted the data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

mTECs[8]. Second, Aire-induced gene expression has a strong element of stochasticity, with individual mTECs transcribing only a small subset of the total repertoire of induced PTA transcripts[3,4,9,10]. The subsets of transcripts induced in individual cells exhibit both intra- and inter-chromosomal clustering[3,4]. Third, Aire appears not to bind to a particular promoter or enhancer motif, exhibiting only a low, non-discriminatory affinity for DNA[11]. Rather, it seems to recognize generic features of transcriptionally quiescent sites, like chromatin marks typical of silenced loci, e.g. unmethylated lysine 4 of histone 3 (H3K4me0)[11,12], and promoters with stalled RNA polymerase II (RNA-PolII)[13,14].

Screening approaches have uncovered a large cast of structural and functional Aire partners, which fall into multiple functional classes, notably nuclear transport, chromatin structure and/or binding, transcription (including the DNA-damage response) and pre-mRNA processing[15,16]. However, we remain ignorant of the genomic location, architecture, biogenesis and function of the resulting Aire-containing complexes. We have addressed these issues by exploiting recent advances in genome-wide chromatin mapping techniques, which now permit analysis of the small mTEC numbers available *ex vivo*[17], and by applying diverse biochemical approaches. Aire was located on and activated super-enhancers, defined as chromatin regions that host exceptionally high concentrations of transcriptional regulators. The topoisomerase, TOP1, emerged as a cardinal Aire partner that co-localized on super-enhancers and was required for Aire interaction with all of its other partners.

RESULTS

Aire is located on mTEC super-enhancers

Our first goal was to map the genome-wide distribution of Aire *ex vivo* in mTECs – in particular, its relationship to diagnostic histone marks – using chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq). Aire bound to 42,124 sites scattered throughout the genome of $Ly51^{lo}MHCII^{hi}$ mTECs (called $mTEC^{hi}$ hereafter) from 4–6-week-old female C57BL/6(B6).$Aire^{+/+}$ mice, in comparison with the <12,000 sites detected in *Aire*-transfected cell lines[14,18]. The Aire signal was robust and reproducible, with >75% of the binding sites common to the two biological replicates routinely examined (Supplementary Fig. 1 and Supplementary Table 1). Genome-wide, Aire bound primarily to intergenic, transcriptional start-site (TSS) and intronic regions (Fig. 1a). Aire was situated along both Aire-induced and Aire-neutral genes but there was a substantially higher ratio of intergenic/TSS localization in the former compared with the latter set of genes (7.3/1 vs 1.2/1). The density of Aire around TSSs was demonstrably lower for Aire-induced than Aire-neutral genes, while its representation on intergenic and intronic regions did not vary much with the gene-induction status (Fig. 1b).

Super-enhancers are chromatin elements that serve as extended and overloaded depots for a multiplicity of general and cell-type-specific transcription factors[19,20]. Because they are often associated with, and regulate, genes diagnostic of fully differentiated cell-types (i.e. PTA genes), we investigated whether such elements might preferentially harbor Aire in mTECs. Following standard practice[21], we defined mTEC super-enhancers as stretches >3.9kb of histone 3 acetylated at lysine 27 (H3K27ac), a chromatin mark diagnostic of active enhancers. By this definition, $mTEC^{hi}$ had 1170 super-enhancers, of ~30kb average

size, scattered over the chromosomes (Fig. 1c and Supplementary Fig. 2a). H3K4me1, another active chromatin mark, was selectively included in these super-enhancers, while H3K27me3, diagnostic of inactive chromatin was excluded (Fig. 1d,e and Supplementary Fig. 2b), supporting the validity of our mTEC^{hi} super-enhancer designations. Aire was also preferentially bound to these super-enhancers, along with its structural and functional associate, RNA-PolII (Fig. 1d,e and Supplementary Fig. 2b). To determine whether the super-enhancers we had delineated were merely a conglomerate of typical enhancers, we bioinformatically sorted all of the super-enhancers and conventional enhancers from the "hockey plot" derived from ROSE algorithm[21] (Fig. 1c), focused on individual H3K27ac peaks mapping within the two sets of enhancers, and compared their Aire densities according to the ChIP-seq data. Super-enhancers hosted higher densities of Aire than did conventional enhancers (Fig. 1f).

To address the functional significance of Aire localization to super-enhancers, we compared ChIP-seq data obtained in parallel on mTEC^{hi} isolated from B6.*Aire*$^{+/+}$ and B6.*Aire*$^{-/-}$ littermates. The density of H3K27ac marks on super-enhancers was significantly lower in *Aire*$^{-/-}$ mTEC^{hi} (Fig. 1g). In addition, a re-analysis of published ChIP-seq datasets[22], indicated a lower density of H3K27ac marks on super-enhancers from immature Aire^{-} mTECs compared with mature Aire^{+} mTECs from wild-type mice (Fig. 1g).

To localize the transcriptional impact of Aire in relation to super-enhancers, we compared gene expression in regions stretching 500kb 5′ or 3′ of super-enhancers in Aire^{+} and Aire^{-} mTEC^{hi}. Aire induced transcription in genomic regions extending in both directions from, but not overlapping, super-enhancers (Fig. 1h), while similar Aire-dependent transcriptional changes did not occur up- or down-stream of random, size-matched genome stretches (Fig. 1h).

In addition, we performed ATAC-seq (assay of transposase-accessible chromatin followed by high-throughput sequencing[23]) to provide a genome-wide view of chromatin accessibility in *ex vivo* mTEC^{hi}. The chromatin stretches delineated above to be mTEC^{hi} super-enhancers had elevated ATAC signal densities (compared with other chromatin regions) in Aire^{+} mTEC^{hi}, but not in control earskin fibroblasts (Fig. 2a), indicating that they were preferentially open in the former case. As we had seen for H3K27ac signals, ATAC signals were substantially higher for individual peaks within super-enhancers than within conventional enhancers (Fig. 2b). In addition, comparison of ATAC signals in mTEC^{hi} of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice revealed that Aire induced the accessibility of super-enhancers (Fig. 2c), and to a lesser extent that of conventional enhancers (not shown). This effect is perhaps most evident from a plot of Aire ChIP-seq signal against the fold-change in ATAC signals in Aire^{+} versus Aire^{-} mTEC^{hi}. This ATAC signal differential was higher in super-enhancers than in random size-matched chromatin stretches (Fig. 2d), as indicated by the preferential concentration of ATAC signal on mTEC^{hi} super-enhancers in Aire^{+} mTEC^{hi} in comparison with Aire^{-} mTEC^{hi} and control earskin fibroblasts (Fig. 2e). These observations indicate that Aire was preferentially associated with and activated super-enhancers, resulting in transcriptional induction upstream and downstream.

Aire participates in multiple multi-protein complexes

As the architecture and dynamics of Aire-containing multi-protein complexes are ill defined, we next investigated how Aire-containing complexes assemble to promote transcription. Because the low numbers of mTEC^{hi} that can be isolated from mouse thymi precludes this type of study, we used HEK293T cells transfected with a construct encoding Aire FLAG-tagged at the amino-terminus (FLAG-Aire). A plot of genome-wide H3K27ac densities from published ChIP-seq data[24] revealed that Aire and RNA-PolII were preferentially located on super-enhancers in HEK293T cells (Supplementary Fig. 3).

Gel filtration chromatography provides information on protein complex number and composition. Because Aire is known to form molecular conglomerates of >670 kDa[25], we applied nuclear extracts from FLAG-Aire-HEK293T cells to a Superose-6 column and immunoblotted eluted fractions with a FLAG antibody (Ab). Aire was broadly distributed, spanning fractions 9–14 (669kDa-2,000kDa) (Fig. 3a). We then pooled fractions 9–11 and fractions 12–14, immunoprecipitated Aire-containing complexes with a FLAG Ab, and immunoblotted the precipitated material with antibodies recognizing a panel of Aire partners. Based on their co-elution profiles, Aire partners divided into three groups: those maximally eluted in fractions 9–11 (SFRS3, DDX5), in fractions 12–14 (DNA-PK$_{cs}$, Ku80, PARP-1, DSIF, CDK9, BRD4) and in all of the Aire-containing fractions (TOP2A, RNA-PolII) (Fig. 3b). These findings indicate that Aire participated in at least two multi-protein complexes.

Next, we performed antibody pre-clearing experiments to reveal the extent to which designated Aire partners co-reside within complexes. DNA-PK$_{cs}$ is one of the Aire-interacting proteins most consistently detected[15]. Co-immunoprecitation experiments showed that DNA-PK$_{cs}$ associated with all Aire partners implicated in transcription (Ku80, PARP-1, BRD4, CDK9, TOP2A, DSIF, RNA-PolII), but none of those involved in pre-mRNA processing (DDX5, SFR53) (Fig. 3c). Pre-clearing of nuclear extracts with a DNA-PK$_{cs}$ Ab removed the Aire-containing complexes that also hosted DNA-PK$_{cs}$ and, to a differential degree, Aire-containing complexes that hosted its other partners (Fig. 3c). Plotting for each Aire partner its propensity to associate with DNA-PK$_{cs}$ versus its degree of interaction with Aire after DNA-PK$_{cs}$ depletion (Fig. 3d), showed three classes of Aire-partner-interacting proteins: those not co-residing with DNA-PK$_{cs}$ in Aire-containing complexes (DDX5, SFRS3), those largely (>80%) co-residing (Ku80, BRD4, DSIF, RNA-PolII), and these partially (40–60%) co-residing (PARP-1, TOP2A, CDK9). These results also suggested the existence of 2–3 distinct Aire-containing complexes.

To address how removal of particular protein partners influenced the formation of Aire-containing complexes, we co-transfected FLAG-Aire-HEK293T cells with one of four pre-validated cognate shRNAs for individual Aire partners (expressed in the pLKO.1 vector), and assessed the ability of Aire to interact with its remaining partners in immunoprecipitation experiments. Because knock-down of BRD4 was not efficient using this approach, we used the small molecule inhibitor, I-BET151[26]. We focused on the Aire partners implicated in transcriptional regulation, because the partners involved in transcription and pre-mRNA processing are known to behave independently in this assay[15]. Knock-down of DNA-PK$_{cs}$ or PARP-1 expression in HEK293T cells abolished the

interaction of Aire with all of the Aire partners examined (Fig. 3e and Supplementary Fig. 4), indicating that these factors are essential for the assembly of the Aire-containing complexes. Inhibition of BRD4 and CDK9 inhibited the interactions between Aire and its associates CDK9, TOP2A and DSIF (Fig. 3c and Supplementary Fig. 4), known to be involved in transcriptional elongation[27,28], but not those involved in the DNA-damage response[29]. Knock-down of TOP2A and DSIF compromised the ability of Aire to interact with only DSIF and RNA-PolII (Fig. 3c and Supplementary Fig. 4). Together, these observations indicate that Aire participates in at least two, potentially three, multi-protein complexes.

TOP1 is a primary Aire-partner

We previously proposed that TOP2A is an early Aire partner, suggesting a scenario wherein Aire "freezes" the enzymatic activity of TOP2A, thereby stabilizing double-stranded breaks (DSBs) and inciting the DNA-damage response via DNA-PK activation[15]. Because shRNA-mediated knock-down of TOP2A had a limited effect and only inhibited the association of Aire with DSIF and RNA-PolII (Fig. 3e, Supplementary Fig. 4) we investigated whether the ability of Aire to promote DSBs[15] reflected an early interaction with other topoisomerases. First, we revisited mass-spectrometry (MS) data from several published or unpublished experiments aimed at identifying proteins that co-immunoprecipitate with Aire in HEK293T cells ([15] and data not shown). TOP2A peptides were detected in most of these experiments, but peptides from TOP2B and TOP1 were also found (Table 1). All three enzymes were detected on immunoblots of proteins co-immunoprecipitated with Aire in FLAG-Aire-HEK293T cells (Fig 4a). shRNA-mediated knock-down of TOP2B in these same cells significantly decreased Aire interactions with only DSIF and RNA-PolII (Fig 4b, Supplementary Fig. 5a), indicating that TOP2B had a restricted effect on the association of Aire with its partners, similar to what was observed above after dampening TOP2A expression. In contrast, knock-down of TOP1 strongly inhibited the interaction of Aire with all partners tested, except DDX5 and SFRS3, which are involved in pre-mRNA processing (Fig. 4c and Supplementary Fig. 5b), a pattern reminiscent of that seen above upon dampening DNA-PK_{cs} or PARP-1 expression. Notably, TOP1 knock-down reduced the association of Aire with TOP2B and TOP2A (Fig. 4d), while TOP2B knock-down inhibited Aire interaction with only TOP2A, and TOP2A knock-down failed to impact the association of Aire with either TOP2B or TOP1 (Fig. 4d). The antibody pre-clearing assay showed that removal of TOP1-containing complexes strongly compromised the interaction of Aire with all partners implicated in transcription (Fig. 4e, Supplementary Fig. 5c), while not affecting associations with partners involved in pre-mRNA processing. Addition of the DNA intercalator ethidium bromide during the pull-down indicated that the Aire-containing complexes were not preformed, but required DNA binding[30] (Supplementary Fig. 5d).

To evaluate the scenario whereby TOP1 induced the DSBs that initiated the formation of Aire-containing multi-protein complexes, while TOP2A and TOP2B were involved in downstream events, we performed ChIP-seq analysis on *ex vivo* $mTEC^{hi}$, comparing the distribution of TOP1 and TOP2A (for which reliable ChIP-seq antibodies were available) with those of H3K27ac (which defines super-enhancers), Aire and γH2AX (which delineates regions adjacent to DSBs[29]). TOP1 and γH2AX co-localized highly preferentially

with Aire at super-enhancer regions (Fig. 5a, b and Supplementary Fig. 6), while the distribution of TOP2A was more dispersed, spreading beyond super-enhancers locally and far-distally (Fig. 5a, b and Supplementary Fig. 6). In addition, a comparison of the Aire-induced changes in super-enhancer-localized γH2AX ChIP-seq signals with Aire-induced changes in topoisomerase ChIP-seq signals showed a strong correlation for TOP1 but not TOP2A (Fig. 5c). Thus, Aire coordinately affected the localizations of TOP1 and DSBs.

TOP2A was less concentrated than was TOP1 at mTEC^{hi} super-enhancers while the overall genomic distribution – specifically the partitioning between intergenic, TSS and intronic regions – of the two topoisomerases was similar, with relatively little TOP2A and TOP1 localizing to exonic regions (Fig. 5d). Focusing on the statistically significant TOP1 peaks annotated to Aire-induced genes, we found co-association of Aire, RNA-PolII and γH2AX at intergenic, TSS and intronic stretches; as expected, the γH2AX peaks were broad, especially at the TSSs (Fig. 5e). We detected little TOP2A within the TOP1 peaks, regardless of the genome element examined (Fig. 5e). In contrast, the statistically significant TOP2A peaks showed co-association with Aire, RNA-PolII, γH2AX and TOP1, but almost exclusively at the TSSs (Fig. 5f). These findings also suggest that TOP1 and TOP2A played divergent roles in Aire-induced transcription in mTECs, TOP1 primarily during initial complex assembly at super-enhancers and TOP2A mainly during downstream events.

TOP1 and TOP2 are required for Aire induction of gene expression

We used the topoisomerase inhibitors topotecan and etoposide, which block TOP1 and TOP2, respectively, to assess the importance of these topisomerases *in vivo*. These small-molecule inhibitors stabilize the enzymes when covalently bound to the DNA they just clipped, thereby inhibiting religation[28]. We injected 4-week-old B6.*Aire*$^{+/+}$ female mice intraperitoneally with vehicle (DMSO), topotecan, etoposide or both drugs every day for 3 days, and then sorted the mTEC^{hi} fraction. None of the drugs altered the proportions of the major thymocyte or stromal cell compartments, as determined by flow cytometry (Supplementary Fig. 7). Neither did topotecan or etoposide treatment detectably influence mTEC^{hi} expression of Aire or any of the topoisomerases examined, as determined by flow cytometry (Supplementary Fig. 8a, b). Microarray-based gene-expression profiling, performed in biological triplicate on RNA from sorted mTEC^{hi} from drug-treated versus vehicle-treated mice, indicated that topotecan and etoposide, and especially the two together, preferentially repressed expression of the set of genes normally upregulated by Aire in mTEC^{hi} (Fig. 6a). While the number of genes inhibited by each of the three treatments was similar (625–700), the reduction was less strong for etoposide compared with topotecan, while using both drugs resulted in the strongest reduction (Fig 6a).

Feature-level analysis of gene-expression profiling data was previously used to show that the impact of Aire on transcription is minimal just after the TSS, but increases after about 200 nucleotides, the reversal of RNA-PolII pausing[14,26]. Plotting the ratio of Aire^{-} to Aire^{+} mTEC^{hi} expression values versus distance from the TSS illustrates the impact of Aire on RNA-PolII pausing (Fig. 6b) as previously described[14,26]. Feature-level analysis of the mTEC^{hi} gene-expression data from mice treated with etoposide + topotecan versus vehicle alone suggested that these drugs have preferential effect on distal features (>200 nucleotides

from TSS), and weaker preference with the single-drug treatments (Fig. 6b). Thus, topoisomerase inhibitors, especially in combination, operated at least in part by potentiating RNA-PolII pausing.

Co-immunoprecipitation studies on FLAG-Aire-HEK293T cells showed that topotecan treatment blocked the ability of Aire to interact with all three topoisomerases, while etoposide compromised the association of Aire with TOP2, but not TOP1 (Fig. 6c). Together, these data indicate that both TOP1 and TOP2 played a substantial role in Aire-induced gene expression in mTECs. However, their roles were different, TOP1 appearing to exert a stronger, earlier effect.

TOP1 and TOP2 are necessary for imprinting tolerance

Finally, we examined the immunologic consequences of disrupting the interaction between Aire and TOP1 or TOP2 by evaluating the clonal deletion of self-reactive thymocytes. We quantified the fraction of $CD4^+$ T cells that recognize peptide P2 (294–306) of the retinal protein IRBP. Clonal deletion of these cells is known to be Aire-dependent[31]. Four-week-old B6.$Aire^{+/+}$ mice were injected intraperitoneally with topotecan, etoposide or just vehicle every third day for 3 weeks, followed by intraperitoneal inoculation with the P2 peptide in complete Freund's adjuvant (CFA). Compared with vehicle, treatment with etoposide or topotecan reduced the expression of *Irbp* mRNA in whole-thymus homogenates ten days after P2 injection (Supplementary Fig. 8c). In addition, significantly more P2-specific $CD4^+$ T cells were found in the peripheral lymphoid organs of drug- versus vehicle-treated mice (Fig. 7a), as determined by staining of pooled lymph node and spleen leukocytes with the A^b:P2 tetramer. Together, these observations indicate that topotecan and etoposide reduced negative selection of IRBP P2-specific T cells. In the same assay, treatment with etoposide or topotecan did not reduce the negative selection of $CD4^+$ T cells that bound the A^b:P7 tetramer, containing peptide 786–797 of IRBP (Fig. 7b), which is selected in the thymus independent of Aire[31].

To further elucidate the effect of topoisomerase inhibition on immunologic tolerance, we quantified the leuckocyte cell infiltration in various organs targeted by inflammatory infiltrates in $Aire^{-/-}$ mice on a NOD genetic background. These mice were selected for analysis because the autoimmunity that typically occurs in the absence of Aire is prominent in NOD mice[32]. We intraperitoneally injected NOD.$Aire^{+/+}$ pups with topotecan or etoposide at days 2, 4 and 8 after birth, and histologically assessed various organs for leukocytic infiltration at 15 weeks of age. Inhibition of TOP1 and TOP2 with topotecan and etoposide significantly augmented leukocyte attack on the retina and lung at 15 weeks of age compared with vehicle-treatment, but we did not see an increase in infiltration in stomach, lacrimal gland or salivary gland (Fig. 7c). The inflammatory attack on the retina and lung was not seen when the same experiment was performed on $Aire^{-/-}$ mice (Supplementary Fig. 8d, e), suggesting it was not due to non-specific drug toxicity. In $Aire^{-/-}$ mice, treatment with TOP inhibitors protected the mice from tissue pathology (not shown), as expected from prior data showing that etoposide and Aire induce the same set of transcripts in Aire-deficient cells[15]. These results indicate that inhibition of TOP1 or TOP2 results in a break in immunological tolerance akin to, but not a precise mimic of, that characteristic of mice

lacking Aire. Considering these insights on Aire-containing multi-protein complexes and chromosomal localization, we propose a model of the molecular mechanism of Aire that involves preferential localization of Aire on super-enhancers for efficient delivery of Aire-containing complexes to the TSS of its target genes (Supplementary Fig. 9).

DISCUSSION

This study employed advanced genomic and biochemical approaches to investigate the molecular mechanism of Aire action. Aire was found on both Aire-induced and Aire-neutral genes, particularly along chromatin stretches overloaded with H3K27ac and H3K4me1 and underloaded with H3K27me3, a profile of histone marks routinely used to delineate super-enhancers[19,21,33]. Furthermore, the super-enhancers of mature *Aire*$^{+/+}$ mTEC^{hi} were enriched in H3K27ac marks vis-à-vis both mature *Aire*$^{-/-}$ mTEC^{hi} and immature Aire^{-} mTEC^{hi} from *Aire*$^{+/+}$ mice indicating that Aire activates super-enhancers. In addition, multiple topoisomerases played an important role in Aire induction of mTEC gene expression. TOP1 was a primary Aire partner, co-concentrated on super-enhancers and critical for Aire association with all of its other partners, while TOP2 was more involved at later stages of transcription.

Super-enhancers are defined as exceptionally long chromatin stretches hosting exceptionally high densities of general and cell-type-specific transcription factors[19–21]. They are thought to serve as depots for effective collection of relevant transcriptional regulators to enable their efficient and coordinate delivery to TSSs via intra-chromosomal looping or inter-chromosomal interactions. Super-enhancers are preferentially associated with genes that set the identity of and control the activities of fully differentiated cell types, or that are rapidly induced upon environmental or physiologic stimulation. Localization of Aire within super-enhancers could explain several of its unusual features. First, Aire has a huge impact on mTEC transcription, regulating around 20% of the genes expressed in this cell type[3,4]. High-concentration depots of Aire-containing multi-protein complexes could drive this prodigious activity, perhaps within the nuclear speckles reported to host Aire and certain of its critical partners (e.g. CBP)[34,35]. Second, while Aire exerts a strong influence on the mTEC transcriptome at the population level, its effect on an individual cell is much more restrained[3,4,9,10]. The repertoire of Aire-induced transcripts within single mTECs exhibits both intra- and inter-chromasomal clustering[3,4]. Dynamic looping of super-enhancers along a chromosome or engagement of a super-enhancer and TSS on different chromosomes could provide a potential framework for understanding such cell-by-cell variation. Third, the fact that super-enhancers are characteristically associated with genes mobilized during terminal differentiation of parenchymal cells might explain the preferential influence of Aire on loci encoding PTAs. Lastly, there is a strong correspondence between those genes induced by Aire and those repressed by a small-molecule inhibitor of BET proteins, amongst which is BRD4, a critical Aire-partner[26]. Super-enhancers, which are overloaded with BRD4, are also known to be particularly sensitive to BET protein inhibitors[36,37]. Indeed, the preferential localization of BRD4 on upstream intergenic regions and its relative depletion from TSSs[26] anticipated the partitioning of Aire revealed in the present studies.

There is a growing body of evidence that topoisomerases play important roles in gene transcription[38,39]. These enzymes cleave and rapidly reseal one (TOP1) or both (TOP2A/B) DNA strands, thereby generating a transient break through which topological changes are effected. Failure to complete the enzymatic reaction leads to trapping of a covalent DNA-topoisomerase intermediate, resulting in a single-stranded nick in the case of TOP1 and a double-stranded break in the case of TOP2. Confrontation of TOP1-induced nicks by the replication or transcription machinery often culminates in DSBs[40,41]. Several processes involved in the transcription of protein-coding genes generate topological strain that is relieved by these topoisomerases: chromatin remodeling[42,43]; synthesis of enhancer RNAs (eRNAs) and nucleosome depletion within enhancers [44]; DNA contortions at TSSs due to nucleosome depletion, transcription factor binding or RNA-PolII pausing[45,46]; and transcriptional elongation, which induces positive supercoils upstream of the polymerase and negative supercoils downstream of it[28,47,48].

We previously suggested that Aire stabilizes TOP2A-induced DSBs at and downstream of TSSs[15], thereby provoking recruitment of a histone-eviction complex composed of DNA-PK_{cs}, Ku80, TOP2, PARP-1 and FACT that facilitates transcriptional elongation[49]. This notion was based on several lines of evidence: co-immunoprecipitation of TOP2A, DNA-PK_{cs} and the other members of the eviction complex with Aire; the ability of Aire to promote DSBs *in vitro* and *in vivo*; and the strong correspondence between mTEC genes induced by Aire and by treatment of Aire-deficient cells with the TOP2 poison, etoposide. However, results reported here argue that this scenario is incorrect, or at least incomplete. Instead, TOP1 emerged as a primary Aire partner, seeding the formation of Aire-containing multi-protein complexes, notably at super-enhancers, through recruitment of elements of the DNA-damage response such as γH2AX, DNA-PK_{cs}, Ku80, PARP-1. Both TOP2A and TOP2B also had required roles in the induction of gene expression by Aire, but rather in downstream events, a function that might still be performed by the histone-eviction complex. Findings on other systems support several elements of this revised scenario: that eRNAs and TOP1-mediated DNA breaks are linked[44]; that the DNA breaks induced by TOP1 can mobilize the DNA-damage response[44]; and that TOP1 and TOP2 cooperate to optimize transcription[28,43,47].

On the basis of these and previous[11,14,15,26] observations, we propose a simplified model of Aire-induced gene expression. Attracted by hypomethylated H3K4me0 and/or repelled by hypermethylated H3K4me3[11], Aire localizes to mTEC super-enhancers and interacts with TOP1, stabilizing DNA DSBs promoted by eRNA transcription and nucleosome depletion[15] and thereby promoting recruitment of γH2AX, DNA-PK_{cs} and other elements of the DNA-damage response, including Ku80 and PARP-1[15]. General transcription factors, such as RNA-PolII, CBP and BRD4[26] are also drawn in, resulting in super-enhancers that host high concentrations of Aire-containing multi-protein complexes. Via chromatin looping, the super-enhancers serve as transcription factor depots for regional TSSs, in particular those contorted by paused RNA-PolII[14], which, itself, can promote TOP2-induced DSBs, and thus independently seed the formation of some Aire-containing complexes. BRD4 within the complexes recruits pTEFb (composed of CycT1+CDK9 subunits)[26], which phosphorylates DSIF, thereby lifting transcriptional pausing and promoting elongation. TOP1 and especially TOP2 ride along with RNA-PolII to relieve the torsional stresses introduced behind and in

front of it. Additional factors (e.g. DDX5, SFRS3) are independently incorporated into Aire-containing complexes, serving to link the transcription and splicing machineries.

Further validation of this model will probably require significant evolution of existing genome-scale methods. The cell-to-cell variability of the impact of Aire, best appreciated from single-cell RNA-seq, may ultimately demand single-cell chromatin-capture approaches.

METHODS

Maintenance, generation and treatment of mice

Mice were housed and bred under specific-pathogen-free conditions at the Harvard Medical School Center for Animal Resources and Comparative Medicine (Institutional Animal Care and Use Committee protocol #02954). C57BL/6 (B6).*Aire*$^{+/-}$ mice[5] were bred to generate *Aire*$^{+/+}$ and *Aire*$^{-/-}$ littermates for experiments. Igrp-Gfp (Adig) reporter mice were provided by Dr. Mark Anderson, and were appropriately bred to yield *Aire*$^{+/+}$ and *Aire*$^{-/-}$ littermates. Unless specified otherwise, females were used.

For the microarray experiments, 4-week-old B6.*Aire*$^{+/+}$ and B6.*Aire*$^{-/-}$ mice were injected intraperitoneally with 5 mg/kg topotecan (Sigma-Aldrich), etoposide (Sigma-Aldrich) or both drugs, dissolved in dimethylsulfoxide (DMSO), once a day for 3 consecutive days. For the tetramer-staining experiments, mice of the same types were administered 1.25 mg/kg topotecan or etoposide once a day every 3rd day for 3 weeks. To analyze effects on autoimmunity, we intraperitoneally injected *Aire*$^{+/+}$ and *Aire*$^{-/-}$ NOD/LtJ pups with 0.675 mg/kg topotecan or 1.25 mg/kg etoposide once a day on the 2nd, 4th and 8th days after birth.

Isolation, sorting and analysis of thymic and dermal cells

Thymus tissue from individual 4–6-week-old *Aire*$^{+/+}$ or *Aire*$^{-/-}$ mice was minced with scissors to release thymocytes and the fragments were digested with collegenase (Roche) and DNase (Sigma-Aldrich) for 15 min, then with collagenase/dispase (Roche) for 30 min, as previously described[14]. The released cells were stained with primary antibodies (Abs) (MHCII-APC; Ly51-PE; CD45-PE/Cy5), and CD45^{+} cells were depleted by MACS separation with anti-PE beads (Miltenyi). DAPI^{-}CD45^{-}Ly51^{lo}MHCII^{hi} mTECs were sorted on a MoFlo instrument (Cytomation) into Trizol for RNA preparation (for microarray) or into Fetal Bovine Serum (FBS) (Gibco) for ChIP-seq library preparation, while GFP^{+}(Aire^{+}) mTECs were sorted in FACS buffer [phosphate-buffered saline (PBS), 0.5% bovine serum albumin, 2mM EDTA] for ATAC-seq library preparation.

For isolation of dermal fibroblasts, earskin tissue was minced with scissors and digested with collagenase type IV (Gibco) and DNase (Sigma-Aldrich) for 60 min. The single-cell suspension was stained with primary Abs (CD45-APC; EpCAM-APC; CD31-FITC; Ter-119-FITC; Sca-1-PE), and DAPI^{-}CD45^{-}EpCAM^{-}CD31^{-}Ter-119^{-}Sca-1^{+} dermal fibroblasts were sorted in FACS buffer for preparation of ATAC-seq libraries.

For flow cytometric sorting or analysis, the following Abs were used: Ly51-PE (108308, BioLegend); CD45-PE/Cy5 (103110, BioLegend); MHCII-APC (107614, BioLegend); Aire

(14-5934-80, eBioscience); TOP1 (ab85038, Abcam); TOP2A (ab12318, Abcam); TOP2B (ab72334, Abcam); CD3ε-PE (100308, BioLegend), CD4-PerCP/Cy5.5 (100434, BioLegend), CD8a-APC/Cy7 (100714, BioLegend), CD19-PE/Cy7 (115520, BioLegend), B220-PE/Cy7 (103222, BioLegend), CD11b-PE/Cy7 (101216, BioLegend), CD11c-PB (117322, BioLegend), F4/80-PE/Cy7 (123114, BioLegend). CD45-APC (103112, BioLegend), EpCAM-APC (118214, BioLegend), CD31-FITC (1625-02, SouthernBiotech), Ter-119-FITC (116206, BioLegend), Sca-1-PE (12-5981-83, eBioscience). Anti-rat IgG secondary Abs conjugated with FITC were from SouthernBiotech, while anti-rabbit IgG-Alexa Fluor® 647 Abs were purchased from Jackson Immunoresearch.

ChIP-seq analysis

1.5×10^5 mTEC^{hi} from 4–6-week-old female B6 mice were used for each ChIP-seq sample, adapting published protocols[17,50]. Briefly, mTECs were cross-linked with 1% formaldehyde for 8 min, sorted and lysed for 10 min on ice in RadioImmunoPrecipitation Assay (RIPA) buffer [10mM Tris-HCl (pH 8.0), 1mM EDTA (pH 8.0), 140mM NaCl, 1% Triton X-100, 0.1% sodium dodecyl sulphate (SDS) and 0.1% sodium deoxycholate] supplemented with complete protease inhibitor cocktail (Roche). Chromatin was sheared using an AFA™ Focused-ultrasonicator (Covaris) for 15 min (duty cycle 2%, intensity 3, cycle/burst 200) and the sheared material was cleared by a 10 min centrifugation at 13,000 rpm at 4°C. The cleared material was immunoprecipitated overnight at 4°C with Abs conjugated to magnetic Protein-G beads (Life Technologies, Dynabeads), followed by extensive washing of the beads with ice-cold RIPA, high-salt RIPA [10mM Tris-HCl (pH 8.0), 1mM EDTA (pH 8.0), 500mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate], LiCl [10mM Tris-HCl (pH 8.0), 1mM EDTA (pH 8.0), 250mM LiCl, 0.5% NP-40 and 0.5% sodium deoxycholate] and TE [10mM Tris-HCl (pH 8.0) and 1mM EDTA (pH 8.0)]. Chromatin derived from 1.5×10^5 cells immunoprecipitated with specific Abs was eluted from the beads, treated with 1μg DNase-free RNase (Roche) for 30 min at 37°C and with Proteinase K (Roche) for 2 hrs at 37°C followed by reverse cross-linking by leaving the plate at 65°C overnight. DNA from reverse cross-linked material was purified with SPRI beads (Agencourt AMPure XP beads, Beckman Coulter); and sequential steps of end-repair, A-base addition, adaptor-ligation and PCR amplification (15 cycles) were performed to prepare the ChIP-seq library for each sample, as described previously[17]. ChIP-seq for H3K27me3 was performed as previously reported[51].

Individual ChIP-seq libraries were size-selected for 200–500bp fragments with SPRI beads. Equivalent amounts of barcoded libraries were pooled and sequenced using HiSeq 2500 or NextSeq 500 (Illumina) instruments. To control for background noise, we immunoprecipitated sheared chromatin with purified rabbit IgG Abs, and a ChIP-seq library was prepared and sequenced as described above.

For ChIP-seq analysis, the following Abs were used: Aire (14-5934-80, eBioscience); TOP1 (ab3825 and ab85038, Abcam); TOP2A (WH0007153M1, Sigma-Aldrich); RNA-PolII (MMS-128P, Covance); γH2AX (05-636, Millipore); H3K4me1 (ab8895, Abcam and 07-436, Millipore); H3K27ac (ab4729, Abcam) and H3K27me3 (ab6002, Abcam).

Short reads (50bp, single end) were aligned to the mouse reference genome (mm10) using bowtie aligner version 2.2.4[52]. Reads with multiple alignments were removed with samtools (v1.1) and de-duplicated with picard (v1.130). To identify peaks from ChIP-seq reads, we used the HOMER package makeTagDirectory followed by the findPeaks command with the 'histone' parameter[53]. Peaks displaying 4-fold enrichment and poison *P*-value of 1×10^{-4} against background IgG ChIP were considered significant and were used for further analysis. To visualize individual ChIP-seq data on Integrative Genomics Viewer (IGV)[54], we converted bam output files from picard into normalized bigwig format using the bamCoverage function in deepTools (v1.6) with options – fragmentLength 200 – normalizeUsingRPKM[55]. HOMER-generated peak files for H3K27ac were used for the identification of super-enhancers, using the ROSE algorithm described previously[21], wherein enhancer peaks are stitched together if they are located within 12.5 kb of each other and if they do not have multiple active promoters in between; enhancers were then ranked according to increasing H3K27ac signal intensity. Heatmaps as in Fig. 1d or line plots as in Fig. 1b were generated using program ngs.plot[56].

ATAC-seq analysis

1×10^4 mTECs or fibroblasts from 4–6-week-old female Adig mice were used for preparation of ATAC-seq libraries, adapting published protocols[23,57]. Briefly, cells were suspended in 100μl of cold hypotonic lysis buffer [10mM Tris-HCl (pH 7.5), 10mM NaCl, 3mM MgCl2 and 0.1% NP40], followed by immediate centrifugation at 550g for 30 min. The pellet was re-suspended in 5μl of transposition reaction mix [1μl of Tagment DNA Enzyme and 2.5μL of Tagment DNA Buffer from Nextera DNA Sample Prep Kit (Illumina), 1.5μl H_2O], and was incubated for 60 min at 37°C for DNA to be fragmented and tagged. For library preparation, two sequential 7-cycles of PCR were performed in order to enrich small tagmented DNA fragments. After the first PCR, the libraries were selected for small fragments (less than 600bp) using SPRI beads followed by a second round of PCR with the same conditions in order to obtain the final library. Libraries were sequenced on the NextSeq 500 instrument to generate paired-end short reads (50bp, forward; 34bp, reverse). Data were processed essentially as per ChIP-seq analysis, except reads mapping to mitochondrial DNA (1~7%) were removed before analysis and peaks were identified using the 'factor' parameter in the findPeaks command of the Homer package.

Cell culture and transfection

HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, L-glutamate and penicillin/streptomycin antibiotics, and were maintained in a humidified atmosphere at 37°C with 5% CO_2. For transfection, the cells were seeded in 10cm tissue-culture plates, and were transfected with the specified plasmids using TransIT reagent (Mirus) according to the manufacturer's instructions. A plasmid driving expression of wild-type mouse *Aire* FLAG-tagged at the amino-terminus (*FLAG-Aire*) was constructed by in-frame insertion between the BglII and SalI sites of the pCMV-tag1 vector (Clontech), as originally described in[58].

Gel filtration chromatography, immunoprecipitations and mass spectrometry

HEK293T cells were transfected with empty or *FLAG-Aire*-containing pCMV-tag1 vector. 48 hours later, the cells were harvested and lysed in a hypotonic lysis buffer [0.05% NP-40, 10mM HEPES, 1.5mM $MgCl_2$, 10mM KCl, 5mM EDTA (3×10^7 cells/ml)] plus complete protease inhibitor cocktail (Roche), pH 7.4, followed by incubation on ice for 15 min. Nuclei were separated from the cytosolic fraction by centrifugation at 800g for 10 min, and were incubated at 4°C for 1 hr in a native nuclear extraction buffer [50mM Bis-Tris, 750mM 6-aminocaproic acid, 3mM $CaCl_2$, 10% Glycerol, EDTA-free complete protease inhibitor cocktail (Roche) and micrococcal nuclease (Nuclease S7; Roche), pH 7.4, (6×10^7 cells/ml)].

For gel filtration chromatography, ~1.5mg of nuclear extract was injected into a Superose-6 10/300 GL column (GE Healthcare Life Sciences), and was separated by fast-protein liquid chromatography (FPLC) using the elution buffer 20mM HEPES, 150mM NaCl, 20mM KCl, 0.5 mM $MgCl_2$, 3mM $CaCl_2$, EDTA-free complete protease inhibitor cocktail, pH 7.4. Twenty-seven fractions of 1ml each were collected. The indicated fractions, used directly or pooled, were concentrated via filter centrifugation (Amicon Ultra, 10 kDa cutoff, Millipore) to 500μl.

For immunoprecipitations, concentrated, pooled chromatographic fractions (from the above-mentioned chromatography experiment) or nuclear extracts (for standard immunoprecipitation experiments) were incubated with the indicated Abs conjugated to Protein-G Sepharose beads (Life Technologies) overnight with rotation at 4°C. To identify DNA-dependent protein interactions, nuclear extracts were treated with ethidium bromide (100μg/ml) for 15 min at 4°C followed by 10 min centrifugation at 13,000 rpm at 4°C and supernatants were used for immunoprecipitation as above. Ethidium bromide was also included in all the later washing steps for this experiment. Beads were washed thrice with ice-cold PBS containing 0.05% NP-40, and once with ice-cold PBS. Bound proteins were eluted by boiling the beads in sample buffer for 15 min, separated by SDS-PAGE, electro-transferred to polyvinylidene difluoride (PVDF) membranes (Bio-Rad), blocked for 60 min with 5% non-fat dried milk solution in PBST [PBS (pH 7.4), 0.05% Tween 20], and were probed with primary Abs overnight at 4°C. After a wash with PBST, membranes were incubated with secondary Abs linked to horseradish peroxidase. The blots were then developed with an enhanced chemiluminescence detection system (Thermo Scientific) as per the manufacturer's instructions. For quantification, the chemiluminescent images were processed with Multi Gauge v2.3 (Fujifilm).

For immunoprecipitation studies, Abs recognizing the following proteins were used: FLAG-tag (M2 mouse mAb, Sigma-Aldrich); DNA-PKcs (MS-423-P1, Thermo Scientific); Ku80 (ab55408, Abcam); PARP-1 (9542, Cell Signaling); TOP1 (ab85038, Abcam); TOP2A (ab12318, Abcam); TOP2B (ab72334, Abcam); RNA-PolII (sc-899, Santa Cruz); SPT5 (sc-28678, Santa Cruz); CDK9 (sc-484, Santa Cruz); BRD4 (ab84776, Abcam); SFRS3 (H00006428-M08, Abnova) and DDX5 (sc-166167, Santa Cruz). Anti-mouse and anti-rabbit IgG secondary Abs conjugated with horseradish peroxidase were purchased from Jackson Immunoresearch.

Mass-spectrometry technique for analysis of FLAG-Aire immunoprecipitates has been detailed earlier[15]. Briefly, nuclear extracts from FLAG-Aire-HEK293T cells were incubated with 20μl Protein-G Sepharose beads conjugated to anti-FLAG Abs overnight with rotation at 4°C. Beads were washed three times with ice-cold PBS containing 0.05% NP-40, and once with ice-cold PBS. Immunoprecipitated proteins were eluted by boiling in sample buffer for 15 min, and were separated by 10% SDS-PAGE. Gels were stained with Coomassie G-250, and tryptic digests of individual lanes were analyzed by LC-MS/MS using an LTQ mass spectrometer. Analysis of the MS/MS data was performed using the SEQUEST algorithm as described previously[15].

shRNA-mediated knock-down of Aire-partners

Knock-down of Aire-partners in HEK293T cells was accomplished by expression of cognate shRNAs (4 per partner) in the lentiviral vector pLKO.1, procured from the RNAi Consortium of the Broad Institute. shRNAs targeting LacZ served as controls. We transduced HEK293T cells with individual shRNA-containing lentivirus particles, selected them using Puromycin (Gibco), then transfected them with the *FLAG-Aire* plasmid. 48h later, we performed immunoprecipitations with anti-FLAG Abs, as described above. The densities of immunoprecipitated protein bands were quantified with Multi Gauge v2.3 (Fujifilm). The densities for all immunoprecipitated protein bands, after transduction of LacZ (two hairpins) or cognate shRNAs (four hairpins), were averaged for two independent experiments and scaled considering immunoprecipitation after LacZ transduction as 100%.

Microarray and quantitative PCR analyses

RNA was prepared from mTEC^{hi} of individual mice treated with vehicle (DMSO) alone, topotecan, etoposide or both drugs followed by amplification and cDNA preparation as previously described[58]. cDNA was either hybridized to Affymetrix ST1.0 microarrays or used for quantitative PCR analysis of *Irbp* expression. Quantitative PCR was performed using Power SYBR Green master mix (Thermo Scientific) and the StepOnePlus real-time PCR system (Applied Biosystems). Primer sequences were *Irbp*-forward, CTACAACCGGCCCAATGACT; *Irbp*-reverse, AAGTAAATTCCTCGGCGGCA; *Hprt*-forward, TGCCGAGGATTTGGAAAAAGTG; *Hprt*-reverse, TGGCCTCCCATCTCCTTCAT.

Microarray data were processed using the robust multiarray average (RMA) algorithm for probe-level normalization and analyzed by the multiplot module of GenePattern (Broad Institute, Cambridge, USA). The feature-level analysis of microarray data was performed as described previously, with slight modifications[14]. Briefly, we processed the raw probe-level data files (.CEL) from Affymetrix ST1.0 microarrays with the RMA algorithm to generate normalized exon-level and gene-level data files for each sample. The genome-wide locations of microarray probes on “mm10 (mouse) build” were extracted from the Affymetrix website. Exon-level expression values for Aire-induced genes ($Aire^{+/+}/Aire^{-/-}$ >2) were taken for further analysis if the gene displayed exon1 imbalance (i.e. the ratio between exon $Aire^{+/+}/Aire^{-/-}$ fold change to transcript $Aire^{+/+}/Aire^{-/-}$ fold change was >2 or <0.5). For Aire-induced genes flagged for exon1 imbalance, expression levels of the exons were plotted against their distance from the TSSs.

Tetramer analyses

Inhibitor- or vehicle-treated mice were immunized with 100μg P2 peptide (IRBP271–290) or P7 peptide (IRBP771-790) emulsified in CFA, as described previously [31]. APC-conjugated A^b:P2 and A^b:P7 tetramers were generated by the National Institutes of Health Tetramer Core Facility. For tetramer staining: 10 days after immunization, peripheral lymph node and spleen cells were pooled and stained for 1 hr at room temperature, followed by magnetic-bead purification using anti-APC beads to enrich for tetramer-positive cells. The selected cells were stained with anti-CD3ε, -CD4, -CD8, -B220, -CD19, -F4/80, -CD11b, and -CD11c. Stained cells were analyzed on an LSRII (BD Biosciences), and tetramer-reactive cells were gated as $CD3^+CD4^+CD8^-CD11b^-CD11c^-F4/80^-B220^-CD19^-$ using FlowJo software (TreeStar). Tetramer-positive cells were enumerated by counting the total number of cells by MACSQuant (Miltenyi Biotech), and determining the fraction of tetramer-reactive cells on FlowJo.

Autoimmune disease monitoring

Inhibitor- or DMSO-treated *Aire*$^{+/+}$ mice were sacrificed at 15 weeks of age, while similarly treated *Aire*$^{-/-}$ were sacrificed at 12 weeks of age or when they had lost 15–20% body weight relative to that of littermates. The designated tissues were removed, fixed in 10% formalin and embedded in paraffin. Tissue sections were stained with hematoxylin and eosin (H+E), and infiltration of various organs was scored. In general, scores of 0, 0.5, 1, 2, 3 and 4 indicate no, trace, mild, moderate, or severe lymphocytic infiltration, and complete destruction, respectively. For retinal degeneration, 0 = lesion present without any photoreceptor layer lost; 1 = lesion present, but less than half of the photoreceptor layer lost; 2 = more than half of the photoreceptor layer lost; 3 = entire photoreceptor layer lost without or with mild outer nuclear layer attack; and 4 = the entire photoreceptor layer and most of the outer nuclear layer destroyed. All samples were scored blindly and independently by two investigators.

Statistical analysis

Data were routinely presented as mean ± s.d. or s.e.m. Statistical significance was assessed by Student's *t*-test, $\chi 2$ test or the Wilcoxon rank-sum test, as specified in individual figure legends.

Data availability

The ATAC-seq, ChIP-seq and microarray datasets reported in this manuscript can be accessed in GEO with accession codes GSEXXXXX and GSE92509. Other referenced publically available datasets: B6.*Aire*$^{+/+}$ and B6.*Aire*$^{-/-}$ RNA-seq, SRR2038194, SRR2038195, SRR2038196 and SRR2038197; H3K27ac ChIP-seq for $Aire^+$ and $Aire^-$ mTECs, GSE74257; H3K27ac and H3K4me1 ChIP-seq for HEK293T cells, GSE51633. Aire, RNA-PolII and IgG ChIP-seq data in Aire transfected HEK293T cells came from[14].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank G. Buruzula, K. Rothamel, A. Rhoads, K. Hattori, Dr. A. Lopez, G. Gopalan, K. Waraska, M. Thorsen for experimental assistance and C. Laplace for help with manuscript preparation. The NIH Tetramer Core Facility (contract HHSN272201300006C) kindly provided tetramers. This work was supported by the grant NIH R01 DK060027. K.B. was supported by American Diabetes Association Mentor-Based Postdoctoral Fellowship #7-12-MN-51 to D.M.

References

1. Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). Nat Rev Immunol. 2014; 14:377–391. [PubMed: 24830344]
2. Sansom SN, et al. Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. Genome Res. 2014; 24:1918–1931. [PubMed: 25224068]
3. Meredith M, Zemmour D, Mathis D, Benoist C. Aire controls gene expression in the thymic epithelium with ordered stochasticity. Nat Immunol. 2015; 16:942–949. [PubMed: 26237550]
4. Brennecke P, et al. Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. Nat Immunol. 2015; 16:933–941. [PubMed: 26237553]
5. Anderson MS, et al. Projection of an immunological self shadow within the thymus by the aire protein. Science. 2002; 298:1395–1401. [PubMed: 12376594]
6. Peterson P, Org T, Rebane A. Transcriptional regulation by AIRE: molecular mechanisms of central tolerance. Nat Rev Immunol. 2008; 8:948–957. [PubMed: 19008896]
7. Mathis D, Benoist C. AIRE. Ann Rev Immunol. 2009; 27:287–312. [PubMed: 19302042]
8. Guerau-de-Arellano M, Mathis D, Benoist C. Transcriptional impact of Aire varies with cell type. Proc Natl Acad Sci U S A. 2008; 105:14011–14016. [PubMed: 18780794]
9. Derbinski J, et al. Promiscuous gene expression patterns in single medullary thymic epithelial cells argue for a stochastic mechanism. Proc Natl Acad Sci U S A. 2008; 105:657–662. [PubMed: 18180458]
10. Villasenor J, Besse W, Benoist C, Mathis D. Ectopic expression of peripheral-tissue antigens in the thymic epithelium: probabilistic, monoallelic, misinitiated. Proc Natl Acad Sci U S A. 2008; 105:15854–15859. [PubMed: 18836079]
11. Koh AS, et al. Aire employs a histone-binding module to mediate immunological tolerance, linking chromatin regulation with organ-specific autoimmunity. Proc Natl Acad Sci U S A. 2008; 105:15878–15883. [PubMed: 18840680]
12. Org T, et al. The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. EMBO Rep. 2008; 9:370–376. [PubMed: 18292755]
13. Oven I, et al. AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. Mol Cell Biol. 2007; 27:8815–8823. [PubMed: 17938200]
14. Giraud M, et al. Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. Proc Natl Acad Sci U S A. 2012; 109:535–540. [PubMed: 22203960]
15. Abramson J, Giraud M, Benoist C, Mathis D. Aire's partners in the molecular control of immunological tolerance. Cell. 2010; 140:123–135. [PubMed: 20085707]
16. Giraud M, et al. An RNAi screen for Aire cofactors reveals a role for Hnrnpl in polymerase release and Aire-activated ectopic transcription. Proc Natl Acad Sci U S A. 2014; 111:1491–1496. [PubMed: 24434558]
17. Blecher-Gonen R, et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nat Protoc. 2013; 8:539–554. [PubMed: 23429716]
18. Org T, et al. AIRE activated tissue specific genes have histone modifications associated with inactive chromatin. Hum Mol Genet. 2009; 18:4699–4710. [PubMed: 19744957]

19. Hnisz D, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013; 155:934–947. [PubMed: 24119843]
20. Witte S, O'Shea JJ, Vahedi G. Super-enhancers: Asset management in immune cell genomes. Trends Immunol. 2015; 36:519–526. [PubMed: 26277449]
21. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–319. [PubMed: 23582322]
22. LaFlam TN, et al. Identification of a novel cis-regulatory element essential for immune tolerance. J Exp Med. 2015; 212:1993–2002. [PubMed: 26527800]
23. Buenrostro JD, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013; 10:1213–1218. [PubMed: 24097267]
24. Liu W, et al. Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. Cell. 2013; 155:1581–1595. [PubMed: 24360279]
25. Halonen M, et al. APECED-causing mutations in AIRE reveal the functional domains of the protein. Hum Mutat. 2004; 23:245–257. [PubMed: 14974083]
26. Yoshida H, et al. Brd4 bridges the transcriptional regulators, Aire and P-TEFb, to promote elongation of peripheral-tissue antigen transcripts in thymic stromal cells. Proc Natl Acad Sci U S A. 2015; 112:E4448–E4457. [PubMed: 26216992]
27. Saunders A, Core LJ, Lis JT. Breaking barriers to transcription elongation. Nat Rev Mol Cell Biol. 2006; 7:557–567. [PubMed: 16936696]
28. King IF, et al. Topoisomerases facilitate transcription of long genes linked to autism. Nature. 2013; 501:58–62. [PubMed: 23995680]
29. Polo SE, Jackson SP. Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. Genes Dev. 2011; 25:409–433. [PubMed: 21363960]
30. Lai JS, Herr W. Ethidium bromide provides a simple tool for identifying genuine DNA-independent protein associations. Proc Natl Acad Sci U S A. 1992; 89:6958–6962. [PubMed: 1495986]
31. Taniguchi RT, et al. Detection of an autoreactive T-cell population within the polyclonal repertoire that undergoes distinct autoimmune regulator (Aire)-mediated selection. Proc Natl Acad Sci U S A. 2012; 109:7847–7852. [PubMed: 22552229]
32. Jiang W, et al. Modifier loci condition autoimmunity provoked by Aire deficiency. J Exp Med. 2005; 202:805–815. [PubMed: 16172259]
33. Dowen JM, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell. 2014; 159:374–387. [PubMed: 25303531]
34. Bjorses P, et al. Localization of the APECED protein in distinct nuclear structures. Hum Mol Genet. 1998; 8:259–266.
35. Saare M, et al. Autoimmune regulator is acetylated by transcription coactivator CBP/p300. Exp Cell Res. 2012; 318:1767–1778. [PubMed: 22659170]
36. Chapuy B, et al. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell. 2013; 24:777–790. [PubMed: 24332044]
37. Loven J, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013; 153:320–334. [PubMed: 23582323]
38. Ashour ME, Atteya R, El-Khamisy SF. Topoisomerase-mediated chromosomal break repair: an emerging player in many games. Nat Rev Cancer. 2015; 15:137–151. [PubMed: 25693836]
39. Baranello L, Kouzine F, Levens D. DNA topoisomerases beyond the standard role. Transcription. 2013; 4:232–237. [PubMed: 24135702]
40. Strumberg D, et al. Conversion of topoisomerase I cleavage complexes on the leading strand of ribosomal DNA into 5′-phosphorylated DNA double-strand breaks by replication runoff. Mol Cell Biol. 2000; 20:3977–3987. [PubMed: 10805740]
41. Sordet O, et al. Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. EMBO Rep. 2009; 10:887–893. [PubMed: 19557000]

42. Trotter KW, King HA, Archer TK. Glucocorticoid receptor transcriptional activation via the BRG1-dependent recruitment of TOP2beta and Ku70/86. Mol Cell Biol. 2015; 35:2799–2817. [PubMed: 26055322]
43. Husain A, et al. Chromatin remodeller SMARCA4 recruits topoisomerase 1 and suppresses transcription-associated genomic instability. Nat Commun. 2016; 7:10549. [PubMed: 26842758]
44. Puc J, et al. Ligand-dependent enhancer activation regulated by topoisomerase-I activity. Cell. 2015; 160:367–380. [PubMed: 25619691]
45. Baranello L, et al. RNA polymerase II regulates topoisomerase 1 activity to favor efficient transcription. Cell. 2016; 165:357–371. [PubMed: 27058666]
46. Madabhushi R, et al. Activity-induced DNA breaks govern the expression of neuronal early-response genes. Cell. 2015; 161:1592–1605. [PubMed: 26052046]
47. Teves SS, Henikoff S. Transcription-generated torsional stress destabilizes nucleosomes. Nat Struct Mol Biol. 2014; 21:88–94. [PubMed: 24317489]
48. Bunch H, et al. Transcriptional elongation requires DNA break-induced signalling. Nat Commun. 2015; 6:10191. [PubMed: 26671524]
49. Heo K, et al. FACT-mediated exchange of histone variant H2AX regulated by phosphorylation of H2AX and ADP-ribosylation of Spt16. Mol Cell. 2008; 30:86–97. [PubMed: 18406329]
50. Garber M, et al. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. Mol Cell. 2012; 47:810–822. [PubMed: 22940246]
51. Brind'Amour J, et al. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. Nat Commun. 2015; 6:6033. [PubMed: 25607992]
52. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. [PubMed: 22388286]
53. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010; 38:576–589. [PubMed: 20513432]
54. Robinson JT, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29:24–26. [PubMed: 21221095]
55. Ramirez F, et al. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014; 42:W187–W191. [PubMed: 24799436]
56. Shen L, Shao N, Liu X, Nestler E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics. 2014; 15:284. [PubMed: 24735413]
57. Lara-Astiaso D, et al. Chromatin state dynamics during blood formation. Science. 2014; 345:943–949. [PubMed: 25103404]
58. Yang S, et al. Aire's plant homeodomain(PHD)-2 is critical for induction of immunological tolerance. Proc Natl Acad Sci U S A. 2013; 110:1833–1838. [PubMed: 23319629]

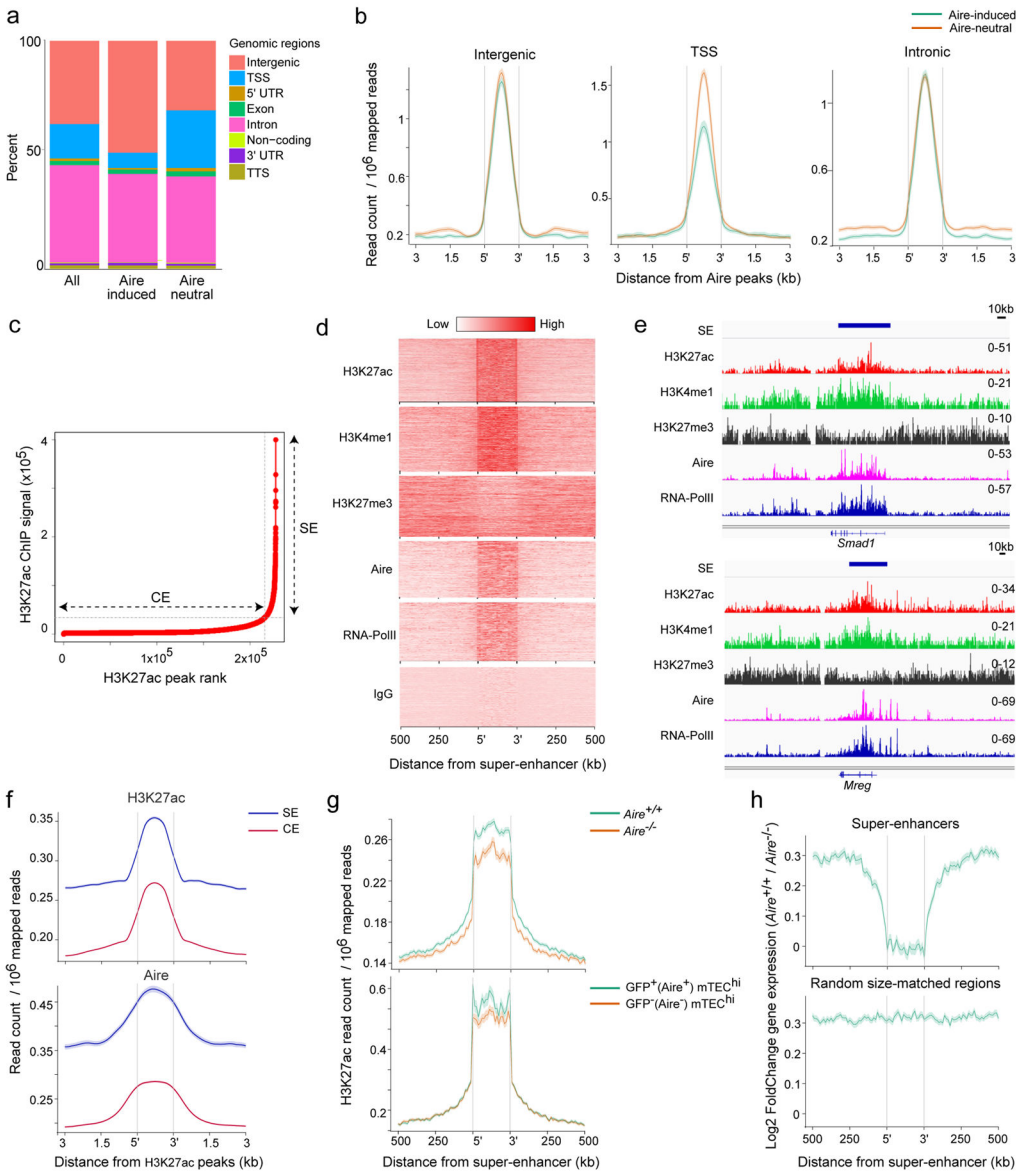

Fig. 1. Concurrence of Aire and super-enhancers in mTECs

a) Distribution of 42,124 Aire-binding sites, derived from merged Aire ChIP-seq data from two independent experiments, along the entire genome in B6.*Aire*$^{+/+}$ mTEC^{hi} (“All”); 3,588 sites annotated to Aire-induced genes; and 6,238 sites annotated to expression-matched Aire-neutral genes. Aire-induction status was defined as *Aire*$^{+/+}$/*Aire*$^{-/-}$ >2 for Aire-induced and as *Aire*$^{+/+}$/*Aire*$^{-/-}$ >0.9 and <1.1 for Aire-neutral genes. TSS, transcription start site; TTS, transcription termination site; UTR, untranslated region. Genomic architecture came from RefSeq. (b) Aire tag density 3kb up- or down-stream of Aire binding sites annotated to the TSS, intronic or intergenic regions of Aire-induced or -neutral genes in B6.*Aire*$^{+/+}$ mTEC^{hi}. $P = 10^{-32}$ for TSS (Wilcoxon rank-sum test) c) Delineation of super-enhancers, based on H3K27ac overloading, in B6.*Aire*$^{+/+}$ mTEC^{hi} using the ROSE algorithm[21]. SE, super-enhancer; CE, conventional enhancer. d) Heatmaps of tag density for the indicated proteins 500kb up- or down-stream of H3K27ac-delimited super-enhancers in B6.*Aire*$^{+/+}$

mTEC^{hi}. e) Normalized ChIP-seq profiles for the indicated proteins at exemplar super-enhancers in B6.*Aire*$^{+/+}$ mTEC^{hi}. Numbers to the right indicate the ranges of normalized tag densities. The genes indicated are Aire-neutral. SE, super-enhancer. f) Tag density for H3K27ac (upper panel) and Aire (lower panel) at individual H3K27ac peaks extracted from super-enhancers (SE) *vs* conventional enhancers (CE), as defined in c. g) Tag density for H3K27ac in B6.*Aire*$^{+/+}$ *vs Aire*$^{-/-}$ mTEC^{hi} (upper panel) and Aire-positive mTEC^{hi} *vs* Aire-negative mTEC^{hi} from Adig mice (lower panel). $P = 4.1 \times 10^{-37}$, upper panel; $P = 2.7 \times 10^{-10}$, lower panel (Wilcoxon rank-sum test). h) Aire-induced transcripts (*Aire*$^{+/+}$ *vs Aire*$^{-/-}$ mTEC^{hi}) 500kb up- or down-stream of H3K27ac-delimited super-enhancers (upper panel), compared with random, size-matched chromatin stretches (lower panel). Data are representative of two independent experiments. Data for one (c–f) or both (g, lower panel) H3K27ac ChIP-seq experiments came from[22] while RNA-seq data in (h) came from[3].

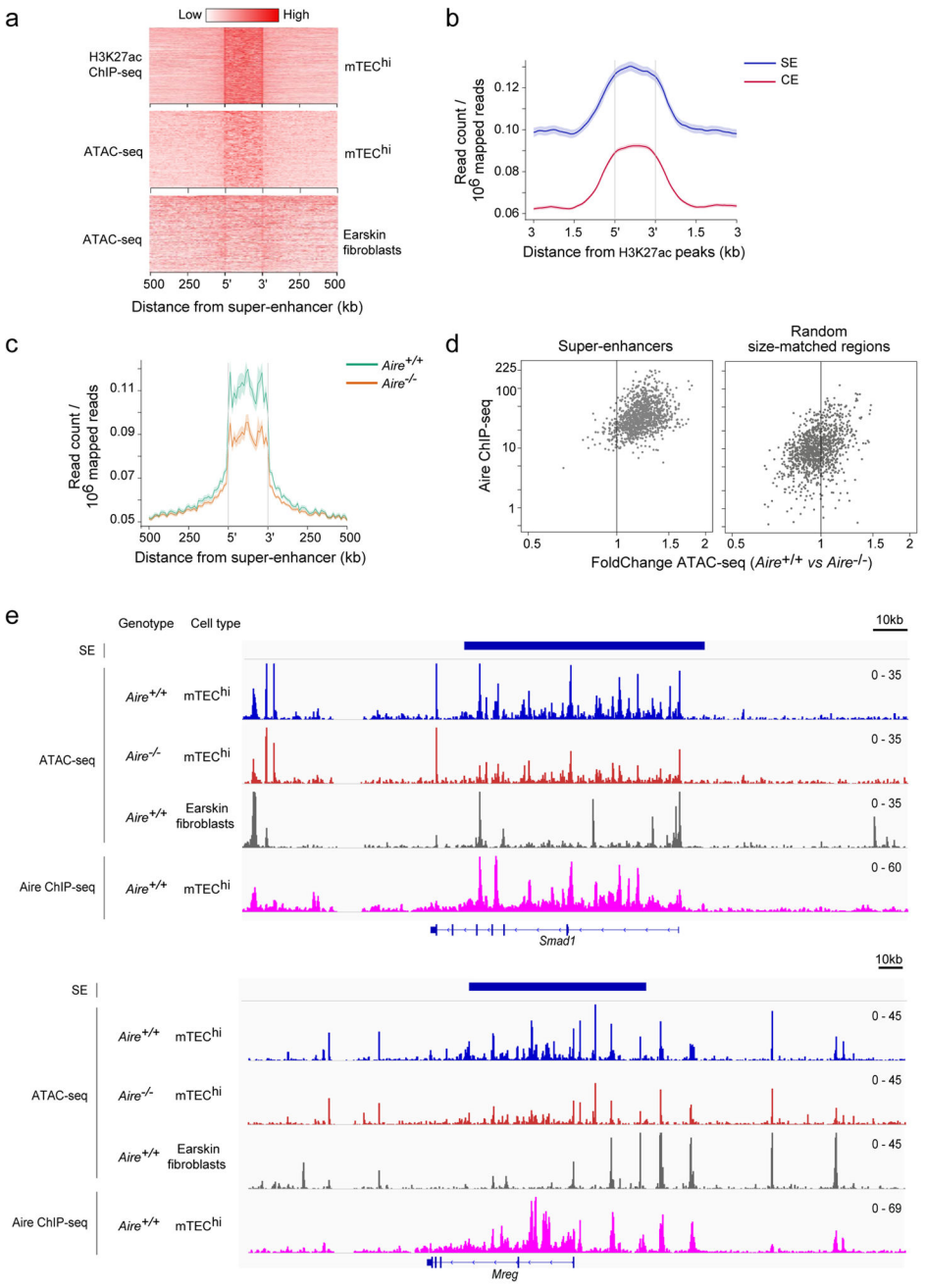

Fig. 2. Aire controls chromatin accessibility of super-enhancers

a) Heatmaps displaying tag density for H3K27ac ChIP-seq in B6.*Aire*$^{+/+}$ mTEC^{hi} and ATAC-seq signal in *Aire*$^{+/+}$ mTEC^{hi} *vs* earskin fibroblasts at mTEC^{hi} super-enhancers defined in Fig. 1c. b) Histogram for ATAC-seq tag density at individual H3K27ac peaks in super-enhancers (SE) *vs* conventional enhancers (CE) in *Aire*$^{+/+}$ mTEC^{hi}. c) Line-plot displaying ATAC-seq tag density in *Aire*$^{+/+}$ *vs* *Aire*$^{-/-}$ mTEC^{hi} at mTEC^{hi} super-enhancers. $P = 6.6 \times 10^{-52}$ (Wilcoxon rank-sum test). d) Scatter-plot for Aire-induced (*Aire*$^{+/+}$ *vs* *Aire*$^{-/-}$) chromatin accessibility *vs* Aire ChIP-seq signal at super-enhancers and random, size-matched genomic stretches as in Fig. 1h (lower panel). e) Normalized ATAC-seq and

Aire ChIP-seq profile at exemplar super-enhancers in various cell populations. Numbers to the right indicate the ranges of normalized tag densities. The indicated genes are Aire-neutral. SE, super-enhancer. Data are representative of two independent experiments. Data for one of the H3K27ac ChIP-seq experiments in (a) came from[22].

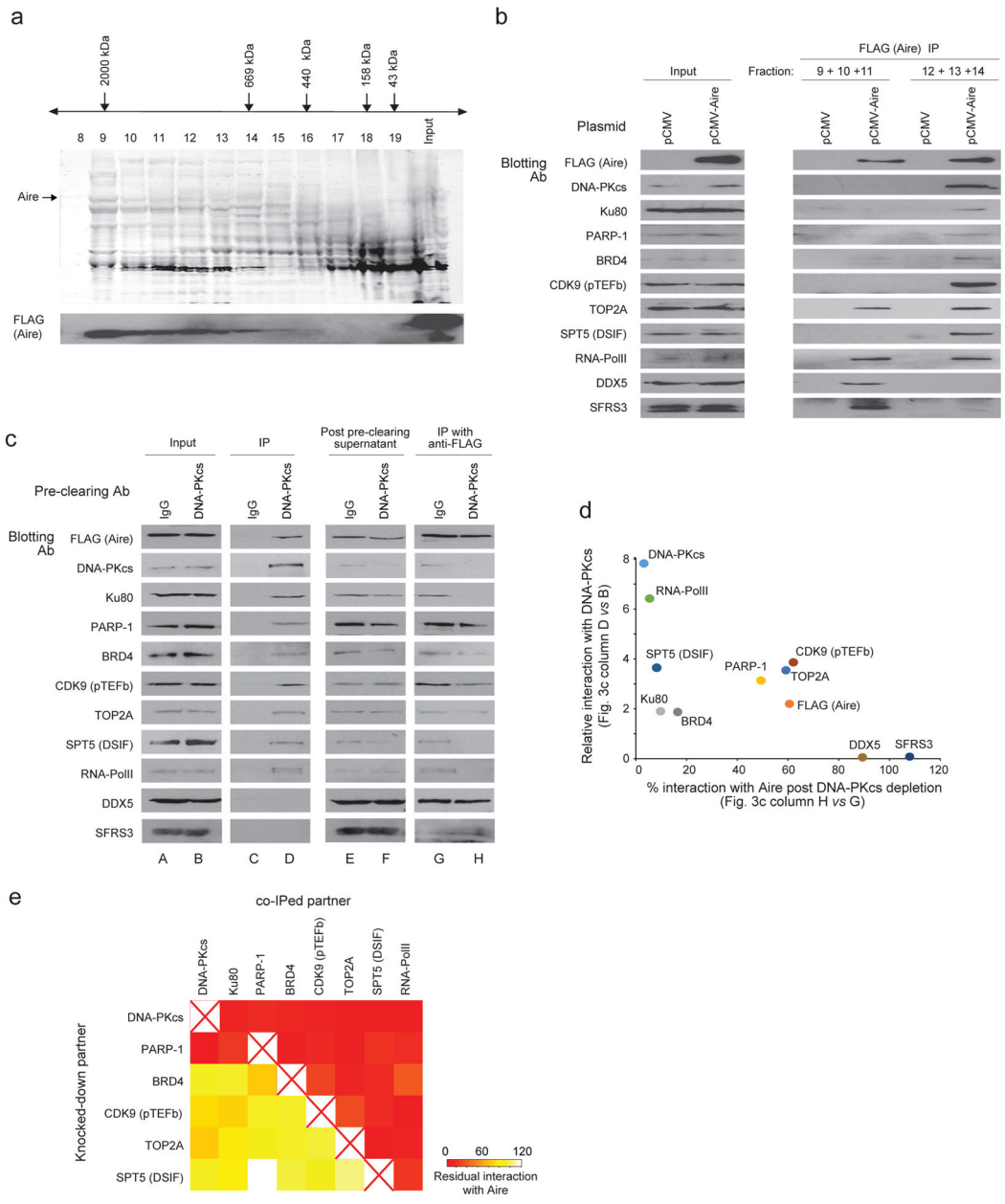

Fig. 3. Association of Aire with multiple multi-protein complexes

a) Coomasie-stained SDS-PAGE gel and anti-FLAG (Aire) immunoblot for gel filtration fractions (8–19) of HEK293T cells transfected with *FLAG-Aire*-containing expression plasmid b) Immunoblotting for interaction of Aire with its various partners (left margin) in pooled gel filtration fractions 9–11 and 12–14 of HEK293T cells transfected with empty plasmid (pCMV) or *FLAG-Aire*-containing expression plasmid (pCMV-Aire). IP, immunoprecipitation. c) Detection of interaction of various Aire partners (left margin) with DNA-PK_{cs} (column C–D) or with FLAG-Aire after DNA-PK_{cs} depletion with anti-DNA-PK_{cs} antibody (column G–H) in *FLAG-Aire*-transfected HEK293T cells. IP, immunoprecipitation d) Scatter-plot displaying the relative interaction of various Aire

partners with DNA-PK_{cs} (as % input, i.e. column D *vs* B of c) versus their % interaction with Aire post-DNA-PK_{cs} depletion compared with post-IgG (control) treatment (i.e. column H *vs* G of c). e) Heatmap displaying % interaction of the various partners with Aire post shRNA-mediated partner knock-down (normalized on LacZ-shRNA (control) transduction). HEK293T cells were transduced with shRNA against indicated Aire partners (except BRD4, which was inhibited by iBET-151), followed by transfection with a *FLAG-Aire*-containing expression plasmid, immunoprecipitation with an anti-FLAG (Aire) antibody and immunoblotting for the indicated proteins. Representative primary immunoblot data for e can be found in Supplementary Fig. 4. IP, immunoprecipitation. Data are representative of two (a, b, e) or three (c, d) independent experiments with similar results.

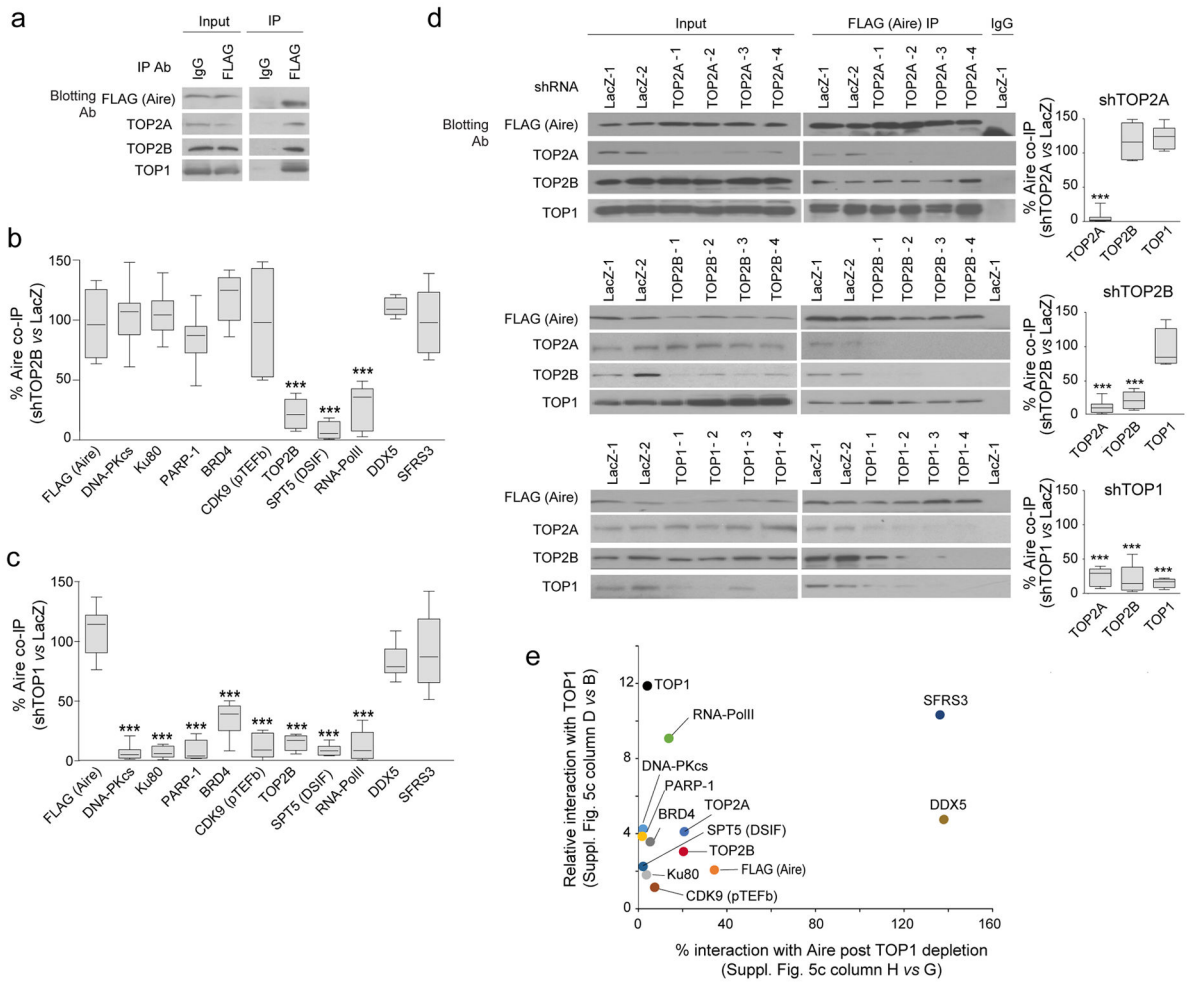

Fig. 4. Primacy of TOP1 as an Aire-partner

a) Immunoblots displaying interaction of Aire (FLAG) with TOP2A, TOP2B and TOP1 in *FLAG-Aire* transfected HEK293T cells. b–d) Immunoblots (d) and quantitative graphs (b–d) displaying % interaction of Aire with various partners in *FLAG-Aire* expressing HEK293T cells transduced with shRNA against *TOP2B* (b, d), *TOP1* (c, d) or *TOP2A* (d) relative to HEK293T cells transduced with LacZ-shRNA (control). ***$P < 0.001$ *vs* shLacZ (unpaired Student's *t*-test). Representative primary immunoblot data for b and c can be found in Supplementary Fig. 5a, b. IP, immunoprecipitation. e) Scatter-plot displaying the relative interaction of indicated Aire partners with TOP1 (as % input) versus their % interaction with Aire after anti-TOP1 antibody mediated TOP1 depletion from nuclear extracts of *FLAG-Aire* expressing HEK293T cells relative to control rabbit IgG mediated depletion. Primary immunoblot data supporting for e can be found in Supplementary Fig. 5c. Data are representative of two independent experiments with similar results (standard box-and-whisker graphs in (b–d) from $n = 8$ measurements pooled from two experiments).

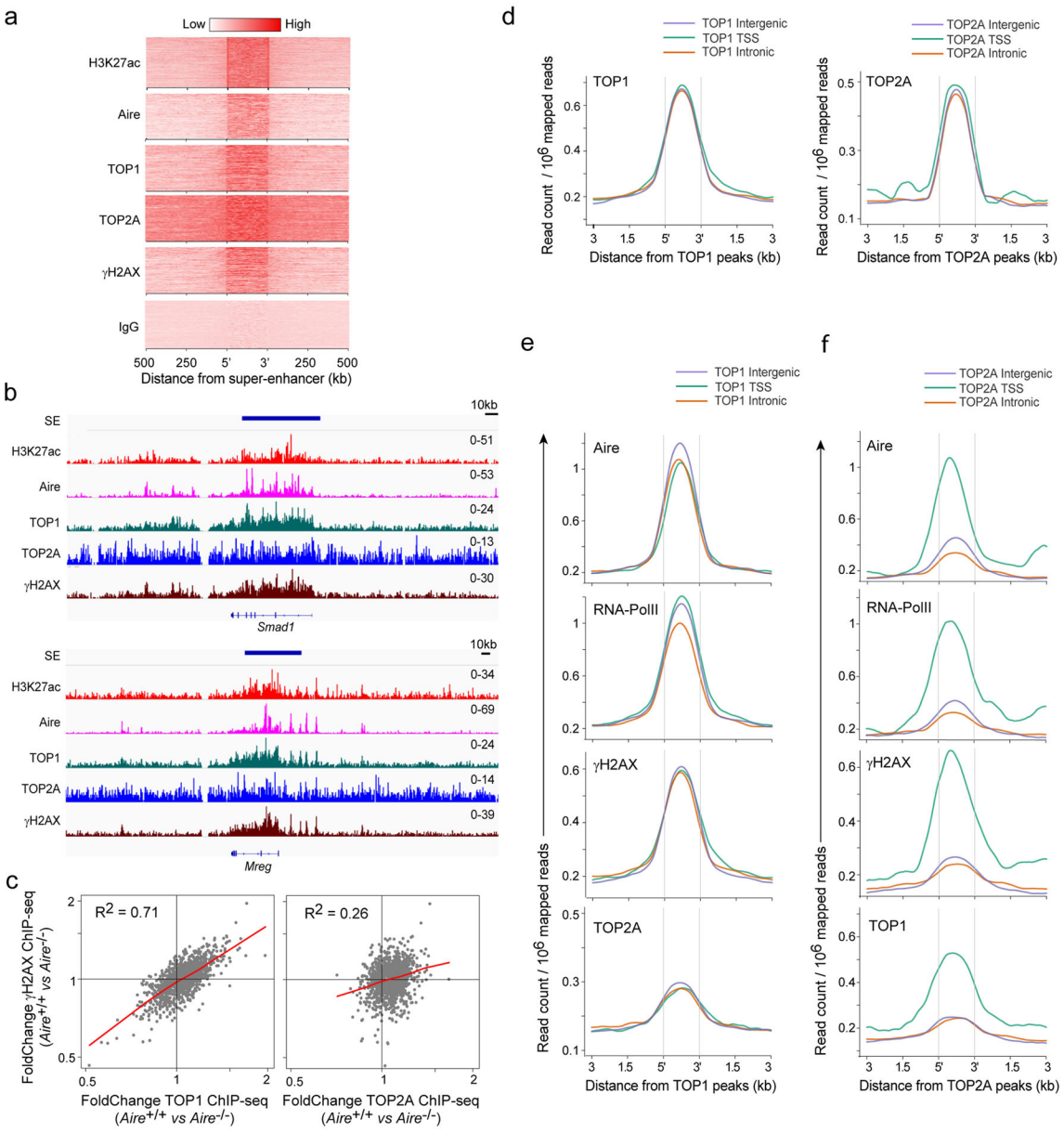

Fig. 5. Co-localization of TOP1 and Aire at super-enhancers

a) Heatmaps of binding of the indicated proteins up- and down-stream of mTEC^{hi} super-enhancers delineated in Fig. 1c. b) Normalized ChIP-seq profiles for indicated proteins at exemplar super-enhancers in B6.*Aire*$^{+/+}$ mTEC^{hi}. Numbers to the right indicate the ranges of normalized tag densities. The genes indicated are Aire-neutral. SE, super-enhancer. c) Scatter-plot for correlation between Aire-induced (*Aire*$^{+/+}$ *vs Aire*$^{-/-}$) topoisomerase binding (left, TOP1; right, TOP2A) and Aire-induced DSBs (γH2AX binding) at mTEC^{hi} super-enhancers. Red line, lowess curve; R^2, Pearson's correlation coefficient. TOP1, $P < 2.2\times10^{-16}$; TOP2A, $P < 2.2\times10^{-16}$ (for correlation coefficient from Student's *t*-test). d) Histograms of ChIP-seq tag density for TOP1 (left panel) or TOP2A (right panel) 3kb up- or down-stream of their binding sites annotated to the TSS, intronic or intergenic regions of

Aire-induced genes in B6.*Aire*$^{+/+}$ mTEC^{hi}. e, f) Histograms of tag densities of indicated proteins (left top corner) 3kb up- or down-stream of TOP1 (e) and TOP2A (f) binding sites annotated to the TSS, intronic or intergenic regions of Aire-induced genes in B6.*Aire*$^{+/+}$ mTEC^{hi}. Aire, $P = 0.01$; RNA-PolII, $P = 0.003$; γH2AX, $P = 0.001$; TOP1, $P = 7.4 \times 10^{-5}$ (for TSS *vs* intergenic TOP2A peaks from Wilcoxon rank-sum test). Aire, $P = 0.003$; RNA-PolII, $P = 0.002$; γH2AX, $P = 0.002$; TOP1, $P = 8.9 \times 10^{-5}$ (for TSS *vs* intronic TOP2A peaks from Wilcoxon rank-sum test). Data are representative of two independent experiments. Data for one of the H3K27ac ChIP-seq experiments in (a, b) came from [22].

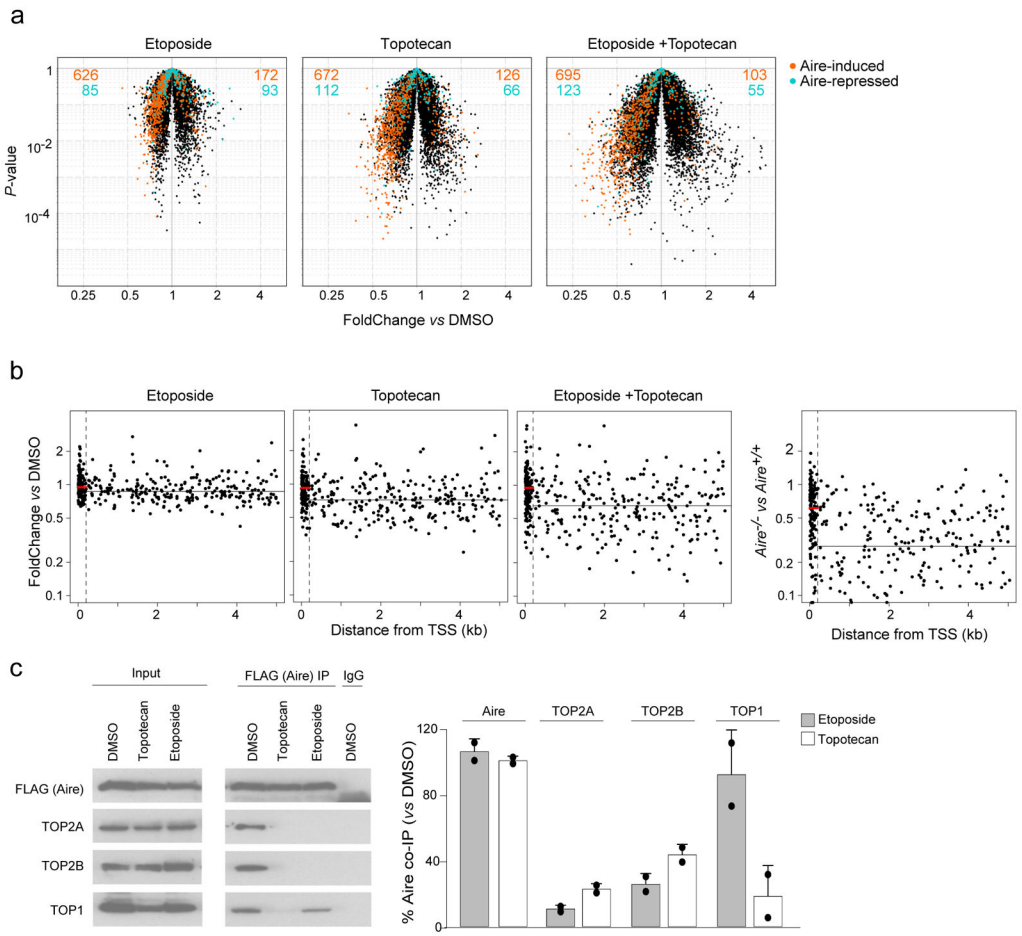

Fig. 6. Requirement of TOP1 and TOP2 for Aire-induced mTEC gene transcription

a) Volcano plots (fold-change *vs P*-value) displaying microarray based analysis of transcriptome changes in mTEC^{hi} of B6.*Aire*$^{+/+}$ mice treated with the indicated inhibitor/s or just vehicle (DMSO) every day for 3 days. Orange: transcripts increased >2-fold in vehicle-treated *Aire*$^{+/+}$ *vs Aire*$^{-/-}$ mice; green: transcripts decreased >2-fold. Numbers refer to transcripts up- (right) or down- (left) regulated by the indicated inhibitor/s. Etoposide, $P = 4 \times 10^{-58}$; Topotecan, $P = 3 \times 10^{-83}$; Etoposide + Topotecan, $P = 2 \times 10^{-107}$ (*P*-values for Aire-induced genes from $\chi 2$ test). b) Feature-level analysis (as per [14]) of the transcriptome data depicted in a. Expression-value fold-change for each feature (exon) of Aire-induced transcripts that exhibited an exon1 imbalance in inhibitor *vs* vehicle-only treatment (left-most three panels) in *Aire*$^{+/+}$ mice or in *Aire*$^{-/-}$ *vs Aire*$^{+/+}$ mice (right-most panel) was plotted relative to the feature's distance from the TSS. Red horizontal line, median of fold-change for proximal features (<200 bp from TSS); black horizontal line, median fold-change for distal features (>200 bp from TSS). Etoposide, $P = 4 \times 10^{-4}$; Topotecan, $P = 2.4 \times 10^{-9}$; Etoposide + Topotecan, $P = 2.2 \times 10^{-10}$; *Aire*$^{-/-}$, $P = 2.8 \times 10^{-16}$ (Wilcoxon rank-sum test). c) Immunoblots for interaction of FLAG-Aire with TOP2A, TOP2B and TOP1 in *FLAG-Aire* transfected HEK293T cells treated with indicated inhibitor for 24h. Left, representative immunoblots; right, summary quantitative data. IP, immunoprecipitation. $*P < 0.05$ and $**P < 0.01$ *vs* DMSO (unpaired Student's *t*-test). Data are representative of three (a, b) or two (c)

independent experiments (error bars (c), mean ± s.d. for $n = 2$ measurements pooled from two experiments).

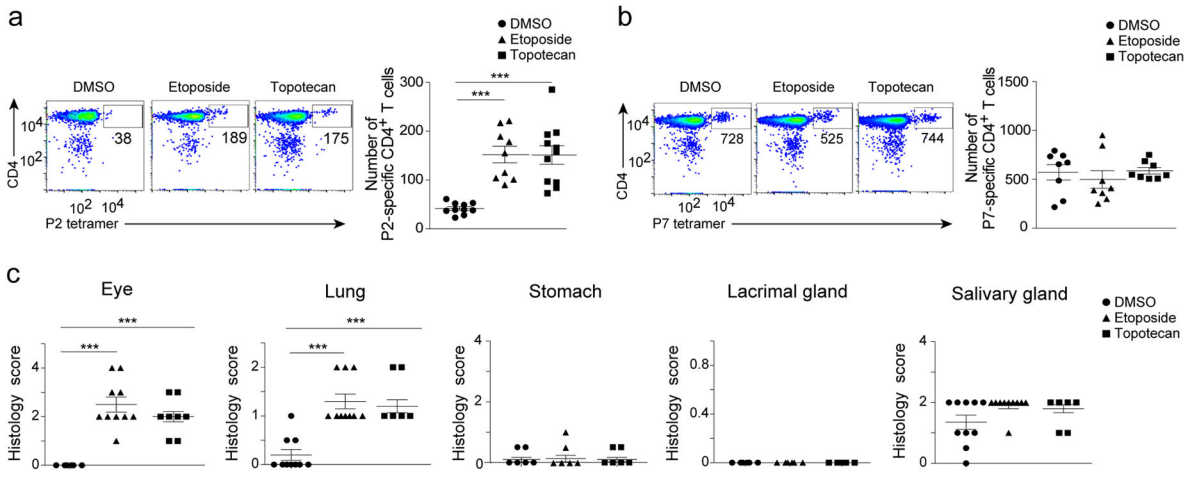

Fig. 7. Requirement of TOP1 and TOP2 for imprinting of immunological tolerance

a, b) Cytofluormetric dot-plots (left) and summary quantitative data (right) for A^b:P2 (a) or A^b:P7 (b) tetramer stained $CD4^+$ T cells from pooled spleen and lymph nodes of 4-week-old B6.*Aire*$^{+/+}$ mice treated with topotecan, etoposide or just vehicle (DMSO) every third day for 3 weeks. $CD3^+CD8^-CD11b^-CD11c^-F4/80^-B220^-CD19^-$ gated cells are displayed. Values on cytofluormetric dot-plots refer to the number of cells in the indicated gate. ***$P <$ 0.001 (unpaired Student's *t*-test). c) Organ histology scores at 15 weeks of age for NOD/LtJ pups (*Aire*$^{+/+}$) treated with topotecan, etoposide or just vehicle (DMSO) on the 2nd, 4th and 8th day after birth. Scores reflect the scale described in Methods. ***$P < 0.001$ (unpaired Student's *t*-test). Data is representative of three experiments with similar results (error bars (a–c), mean ± s.e.m. from $n \geq 9$ (a), $n = 8$ (b), $n = 10$ (c) measurements).

Table 1

Mass-spectrometry based analysis of FLAG-Aire immunoprecipitates from *FLAG-Aire* transfected HEK293T cells. For each topoisomerase, the number of unique peptides identified in each experiment is tabulated. Experiments 4 and 5 were reported in [15]. Exp, experiment.

Protein	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Average number of hits
TOP2A	61	1	34	1	2	20
TOP2B	28	1	34	0	0	13
TOP1	0	0	6	7	0	3