Aire regulates chromatin looping by evicting CTCF from domain boundaries and favoring accumulation of cohesin on superenhancers

Kushagra Bansal[a,b,c], Daniel A. Michelson[a,b], Ricardo N. Ramirez[a,b], Aaron D. Viny[d,e,f,g], Ross L. Levine[d,e,f,g], Christophe Benoist[a,b,1], and Diane Mathis[a,b,1]

[a]Department of Immunology, Harvard Medical School, Boston, MA 02115; [b]Evergrande Center for Immunologic Diseases, Harvard Medical School and Brigham and Women's Hospital, Boston, MA 02115; [c]Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560 064, India; [d]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065; [e]Center for Hematologic Malignancies, Memorial Sloan Kettering Cancer Center, New York, NY 10065; [f]Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065; and [g]Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY 10065

Contributed by Diane Mathis, August 16, 2021 (sent for review June 14, 2021; reviewed by Nicolas Chevrier and Liang Zhou)

Aire controls immunological tolerance by driving promiscuous expression of a large swath of the genome in medullary thymic epithelial cells (mTECs). Its molecular mechanism remains enigmatic. High-resolution chromosome-conformation capture (Hi-C) experiments on ex vivo mTECs revealed Aire to have a widespread impact on higher-order chromatin structure, disfavoring architectural loops while favoring transcriptional loops. In the presence of Aire, cohesin complexes concentrated on superenhancers together with mediator complexes, while the CCCTC-binding factor (CTCF) was relatively depleted from structural domain boundaries. In particular, Aire associated with the cohesin loader, NIPBL, strengthening this factor's affiliation with cohesin's enzymatic subunits. mTEC transcripts upregulated in the presence of Aire corresponded closely to those down-regulated in the absence of one of the cohesin subunits, SA-2. A mechanistic model incorporating these findings explains many of the unusual features of Aire's impact on mTEC transcription, providing molecular insight into tolerance induction.

immune tolerance | thymus | transcription | chromatin | looping

Self-tolerance, a fundamental property of the immune system, keeps autoimmune reactions in check. Both central and peripheral mechanisms enforce T cell tolerance, depending on the particular context. Within the thymus, medullary thymic epithelial cells (mTECs) have an extraordinary ability to promote both the clonal deletion of self-reactive thymocytes and the generation of a unique—perinatal—population of regulatory T cells (1). The transcriptional regulator, Aire, is the primary driver of both of these activities (2), inducing ectopic thymic expression of a broad swath of the genome, in particular transcription of many loci encoding proteins characteristic of fully differentiated parenchymal cells in peripheral nonlymphoid organs (peripheral-tissue antigens; PTAs). Consequently, both humans and rodents with an Aire defect develop multiorgan autoimmunity.

Although Aire is known to function as a transcriptional regulator, its precise mechanism of activation remains puzzling. Unlike conventional transcription factors (TFs) that bind to consensus DNA motifs in promoters and induce the initiation of transcription, Aire preferentially localizes to "superenhancers" (3) and releases RNA polymerase II (RNA-PolII) molecules paused just downstream of transcriptional start sites (TSSs) (4–6). Superenhancers are extended stretches of chromatin that encompass clusters of enhancers and are overloaded with general and cell type–specific TFs (7). They are thought to function as depots of critical transcriptional regulators amassed for coordinate delivery to the TSSs of target loci, in particular lineage-specific genes. Correspondingly, Aire-binding superenhancers are also heavily loaded with many of its large repertoire of protein partners, such as topoisomerase 1 (TOP1), DNA protein kinase, bromodomain protein 4 (BRD4), and RNA-PolII (3, 6).

Transcriptional enhancers and superenhancers often interact with their target genes through the formation of chromatin loops (8), which are typically constrained within larger loop entities called "topologically associated domains" (TADs) or "contact domains" (CDs) (8, 9). In general, architectural proteins and transcriptional cofactors, including the cohesin and mediator complexes, facilitate enhancer or superenhancer looping with each other and with promoters, while cohesin cooperates with the CCCTC-binding factor (CTCF) in the formation of TADs and CDs. The cohesin complex, a circular structure composed of the SMC1, SMC3, RAD21, and SA-1 or SA-2 subunits, is thought to entrap DNA within its ring and to extrude it until the complex runs into CTCF molecules at the two ends. The resulting TADs are relatively insulated domains, with intra-TAD chromatin interactions being favored and inter-TAD associations disfavored. An additional level of organization entails spatial segregation of chromatin into "A" or "B" compartments,

Significance

Aire controls immunological tolerance by driving promiscuous expression of a large swath of the genome in medullary thymic epithelial cells (mTECs). Its molecular mechanism remains enigmatic. To address the hypothesis that Aire impacts chromatin organization through widespread promotion of superenhancer–promoter loops, we performed genome-wide, high-resolution chromosome-conformation capture (Hi-C) experiments on ex vivo mTECs from thymi of mice expressing Aire or not. Integration of the resulting data with inventories of the genome-wide locations of Aire, subunits of the mediator and cohesin complexes, and the chromatin-domain enforcer, CCCTC-binding factor (CTCF), indeed argued for a looping model, which was further supported by biochemical and in vivo loss-of-function analyses.

Author contributions: K.B., D.A.M., R.N.R., and D.M. designed research; K.B., D.A.M., and R.N.R. performed research; A.D.V. and R.L.L. contributed new reagents/analytic tools; K.B., D.A.M., and R.N.R. analyzed data; K.B., D.A.M., and D.M. wrote the paper; and C.B. and D.M. supervised the project.

Reviewers: N.C., University of Chicago Medical Center; and L.Z., University of Florida Health.

Competing interest statement: R.L.L. is on the supervisory board of Qiagen, and the scientific advisory boards of Loxo, Imago, Mana, Auron, C4 Therapeutics, and IsoPlexis, which include equity interest; receives research support from and consulted for Celgene and Roche, and consults for Incyte, Lilly, Janssen, Astellas, MorphoSys, and Novartis; receives research support from Prelude; and has received honoraria from Astra Zeneca, Roche, Lilly, and Amgen for invited lectures and from Gilead for grant reviews.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: cbdm@hms.harvard.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2110991118/-/DCSupplemental.

Published September 13, 2021.

usually associated with active or repressive histone marks, respectively. Genomic regions, often trapped by architectural proteins, switch between loops, TADs, and compartments, resulting in new chromatin interactions and consequent alterations in gene expression.

We hypothesized that an important element of Aire's ability to induce transcription of a large swath of the genome, especially loci encoding PTAs, is widespread promotion of superenhancer–promoter loops. To evaluate this hypothesis, we performed genome-wide, high-resolution chromosome-conformation capture (Hi-C) experiments on ex vivo mTECs from thymi of mice expressing Aire or not. Integration of the resulting data with genome-wide inventories of the genomic locations of Aire, subunits of the mediator and cohesin complexes, and CTCF indeed argued for a looping model, which was further supported by biochemical and in vivo loss-of-function analyses.

Results

Aire Strongly Influenced the Higher-Order Structure of mTEC Chromatin. To determine whether Aire influences three-dimensional (3D) chromatin structure, we performed in situ Hi-C experiments on duplicate samples of mTECs expressing high levels of major histocompatibility complex class II (MHCII) molecules (mTEC^{hi}), isolated from C57BL/6.*Aire*$^{-/-}$ vs. C57BL/6.*Aire*$^{+/+}$ littermates. Each sample was sequenced at an average depth of 300 million (M) read pairs, that is, ∼600M reads per genotype. This sequencing depth is on a par with published Hi-C data from other primary cells (*SI Appendix*, Table S1, adapted from ref. 10). Reproducibility of the data was indicated by the duplicates' clustering pattern upon principal component analysis (*SI Appendix*, Fig. S1*A*). Given the concordance of the Hi-C duplicates for each genotype, we pooled data from the two replicates for the analyses presented below in order to increase the analytical power, as performed routinely for rare primary cells (10, 11). Nonetheless, we substantiated key conclusions by also presenting data from the individual replicates.

Exemplar contact matrices permitting direct comparison of Aire^{-} and Aire^{+} mTECs are presented in Fig. 1*A* (from chromosome 12) and *SI Appendix*, Fig. S1*B* (from multiple other chromosomes). Replicate concordance of the chromosome 12 contacts is illustrated in *SI Appendix*, Fig. S2*A*. We detected a total of 150.4M contacts. Using the program HiCCUPS (9), we called 3,069 loops (false discovery rate 0.1), of which 23.9% were present in both Aire^{+} and Aire^{-} thymi, 29.6% shared one anchor in the two genotypes, 38.6% were unique to mutant mTECs, and 8.0% were unique to wild-type (WT) mTECs (*SI Appendix*, Fig. S1*C*).

Thus, Aire had a strong impact on the 3D architecture of mTEC chromatin. As will be detailed below, the Aire-dependent contact differences encompassed changes on several scales: transcriptional loops, architectural domains, A/B compartments, and interchromosomal contacts. Aire favored certain of these chromatin features, while disfavoring others.

Aire Enhanced Interactions between Superenhancers and Aire-Induced Genes. We next compared the compartment A vs. B distribution of mutant and WT mTECs to see whether Aire promoted a more active chromatin structure at the genes whose expression it upregulated. Active A compartments and repressive B compartments are defined by positive or negative values, respectively, of the first principal component (PC1) of the normalized contact correlation matrix (12). Aire-induced genes (defined in ref. 13) had higher PC1 values and preferentially transitioned from B to A compartments in Aire's presence; in contrast, a set of expression-matched Aire-neutral genes had lower PC1 values and were relatively depleted from A compartments in the presence of Aire (Fig. 1 *B*–*D*). The same observations emerged from the analysis using individual replicates (*SI Appendix*, Fig. S2 *B* and *C*).

Because Aire is known to associate with and activate superenhancers, which are defined by active histone marks and sit in A compartments, an obvious question was whether the preferential B-to-A shift of Aire-induced genes might reflect an increased interaction with superenhancers. So we performed aggregate peak analysis of superenhancers and Aire target genes, whereby we centered and averaged their contacts across the entire genome in the presence or absence of Aire. Aire enhanced interactions between superenhancers and Aire-induced genes, as evidenced by the focal increase in signal intensity at the center of the plot (Fig. 1 *E*, *Left*; center-to-lower-left [CLL] ratio 10.90, $P < 0.0001$). Individual gene tracings provide a similar picture of Aire-promoted looping from superenhancers to Aire-induced genes (examples in Fig. 1*F* and *SI Appendix*, Fig. S3*A*). In comparison, Aire decreased interactions between superenhancers and Aire-neutral genes, as indicated by the focal decrease in signal intensity at the center of the aggregate contact plot (Fig. 1 *E*, *Right*; CLL ratio 0.89, $P = 0.82$). These results are reminiscent of our observations from the A/B compartment analysis in the presence or absence of Aire (Fig. 1 *B* and *C*). We observed a much smaller increase in the interactions between conventional enhancers and Aire-induced genes, suggesting that Aire preferentially promoted ties between superenhancers and Aire-induced genes (*SI Appendix*, Fig. S3*B*; CLL ratio 4.18, $P < 0.0001$).

In brief, then, analyses performed at several scales demonstrated transition of Aire-induced genes into a more active chromatin state in the presence of Aire, concomitant with enhanced interactions between these genes and superenhancers. These findings suggest that Aire promoted transcriptional looping.

Aire Interacted with Known Orchestrators of Chromatin Conformation. We then turned our attention to the mechanism(s) underlying Aire's impact on superenhancer–promoter looping. It is by now well-established that Aire binds only weakly and nonspecifically to DNA. Rather, it exerts its influence on gene expression by partnering with chromatin-associated proteins within large multiprotein complexes (3, 6, 14). We wondered whether Aire associates with known orchestrators of chromatin conformation such as mediator, cohesin, and CTCF.

Our first approach to addressing this question was to more extensively analyze already-existing data from our published mass spectrometry (MS) experiments aimed at identifying proteins that coimmunoprecipitated (co-IPed) with Aire after incubation of nuclear extracts from *FLAG*/*Aire*–transfected HEK293T or IC6 (medullary epithelial) cells with anti-FLAG antibodies (Abs) (14). Peptides derived from subunits of the cohesin (SMC1, SMC3, RAD21) and mediator (MED8) complexes were detected in these experiments; in contrast, we never found CTCF to co-IP with Aire using this readout (Fig. 2*A*).

Secondly, we sought confirmation in biochemical experiments. Given the dearth of Aire^{+} mTEC^{hi} (∼50,000 isolable cells per mouse), we were unable to solidify these findings by Western blotting of ex vivo material co-IPed with Aire from thymic cells. Instead, as we and other investigators have routinely done for biochemical experiments, we turned to *FLAG*/*Aire*–transfected HEK293T cells, which exhibit Aire-induced gene expression and show a similar preferential colocalization of Aire and RNA-PolII on superenhancers (3). Indeed, multiple cohesin complex constituents—including SMC1, SMC3, RAD21, and SA-2—were detected on Western blots of proteins co-IPed with Aire from this model cell (Fig. 2*B*). We also found MED1 and MED12, which occupy different functional domains within the mediator complex. In contrast, we failed to detect coimmunoprecipitation of CTCF with Aire, a finding that is consistent both with the MS readout and with a published reciprocal coimmunoprecipitation with an anti-CTCF Ab (14).

A third approach was to employ ChIP-seq (chromatin immunoprecipitation followed by deep sequencing) to map mediator-, cohesin-, and CTCF-binding sites on mTEC chromatin genome-wide, evaluating, in particular, their correspondence with Aire-binding

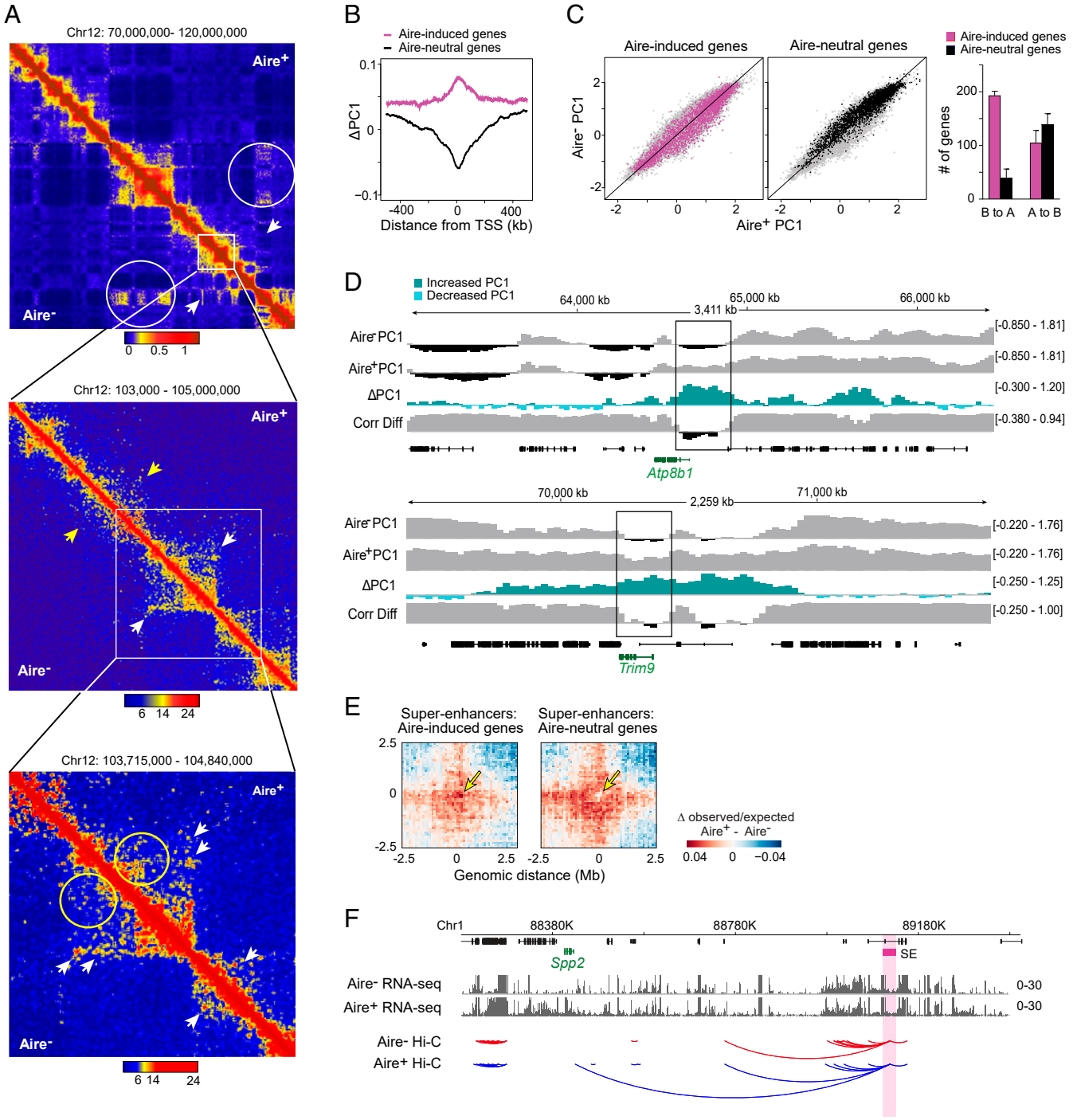

Fig. 1. Aire impacts chromatin architecture in mTECs and promotes interactions between superenhancers and Aire-induced genes. Hi-C libraries were prepared from mTEC^{hi} of B6.*Aire*$^{+/+}$ or B6.*Aire*$^{-/-}$ mice and sequenced. Data from two replicates were merged and analyzed. (*A*) Hi-C contact maps for Chr12: 70 to 120 Mb, 103 to 105 Mb, and 103.715 to 104.840 Mb of Aire$^+$ (*Upper Right*) and Aire$^-$ (*Lower Left*) mTECs. White circles and arrows indicate increased signal in Aire$^-$ mTECs, while yellow arrows and circles indicate increased signal in Aire$^+$ mTECs. Bars depict signal intensities. Interaction matrices were normalized as interactions per 100 kb^2 per 1×10^9 interactions and then used to generate Hi-C interaction heatmaps. (*B*) Line graphs of Aire-induced changes in the PC1 values 500 kb up- or downstream of the TSSs of Aire-induced genes vs. Aire-neutral genes. (*C*) Scatterplots displaying the PC1 values for Hi-C data at the TSSs of Aire-induced or Aire-neutral genes in Aire$^-$ vs. Aire$^+$ mTECs (*Left*), and quantification of compartment shifts for the two genotypes (*Right*). Each dot represents a single gene. (*D*) Genome browser tracks of Hi-C parameters for Aire$^-$ vs. Aire$^+$ mTECs. Aire-induced genes are boxed. The Hi-C data analysis in *B–D* was performed using a window size of 50 kb, sampled every 25 kb. (*E*) Heatmaps displaying modulation of Hi-C aggregate contacts between superenhancers and Aire-induced genes (*Left*) or Aire-neutral genes (*Right*) in Aire$^+$ vs. Aire$^-$ mTECs across the genome. Briefly, all such contacts were centered and averaged over a 7-Mb region, and differential observed/expected signal for Aire$^+$ vs. Aire$^-$ mTECs was plotted. Yellow arrows highlight the presence of a focal increase in the center for superenhancer–Aire-induced gene contacts vs. a decrease in superenhancer–Aire-neutral gene contacts. The analysis was performed using a 100-kb Hi-C matrix. (*F*) Genome browser view of the *Spp2* locus, displaying gain of superenhancer–TSS loops between a superenhancer (SE) (in pink) and an Aire-induced gene (in green). RNA-seq tracks highlight increased expression of *Spp2* in Aire$^+$ vs. Aire$^-$ mTECs. Data were derived from a pool of two biological replicates per genotype except for *C*, *Right*.

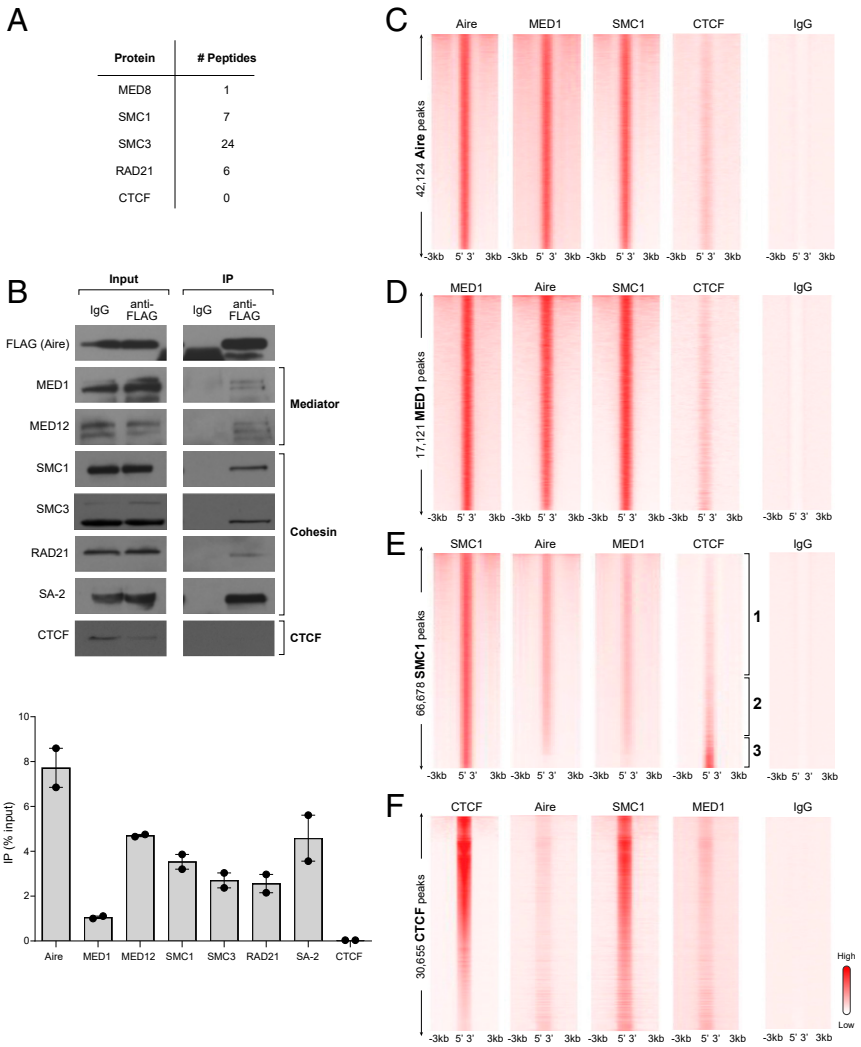

Fig. 2. Aire associates with known regulators of chromatin architecture. Nuclear extracts from *FLAG/Aire*–transfected HEK293T cells were IPed with an anti-FLAG (Aire) Ab. (*A*) Proteins were separated by sodium dodecyl sulfate–polyacrylamide-gel electrophoresis, eluted, and digested, and the resulting peptides were subjected to MS analysis. For each protein, the total number of unique peptides identified is tabulated. Data were compiled from experiments reported in refs. 3 and 14. (*B*) Coimmunoprecipitation with Aire of known regulators of chromatin architecture. The bar graph represents summary data from two experiments. Error bars represent SEM. (*C*–*F*) Heatmaps of tag density for the indicated proteins 3 kb up- or downstream of (*C*) Aire, (*D*) MED1, (*E*) SMC1, and (*F*) CTCF peaks in Aire$^+$ mTECs. Peaks are ranked according to Aire signal in *C*–*E* and CTCF signal in *F*. ChIP-seq analysis was performed on pooled data from two independent experiments. Numbers in *E* refer to the classes of CTCF complexes described in the text. Heatmaps in *C*–*F* were generated with normalized coverage vectors representing tag density information.

sites. The ChIP-seq signals so generated were robust and reproducible across biological duplicates (*SI Appendix*, Fig. S4*A* and Table S2). Given this concordance, we pooled ChIP-seq data from the two biological replicates for downstream analyses, but confirmed key results on individual replicates. At the peak-calling thresholds used, MED1, SMC1, and CTCF were bound to 17,121, 66,678, and 30,655 sites, respectively, compared with 42,124 sites bound by Aire. Like Aire, all three of these proteins were localized primarily to intergenic, TSS, and intronic regions (*SI Appendix*, Fig. S4*B*). Genome-wide, sites that bound Aire were generally co-occupied by MED1 and SMC1 but not by CTCF (Fig. 2*C* and *SI Appendix*, Fig. S5*A*). Similarly, sites occupied by MED1 also bound Aire and SMC1 but not CTCF (Fig. 2*D* and *SI Appendix*, Fig. S5*B*). In contrast, three types of sites hosting SMC1 were apparent: The majority (Fig. 2*E*, 1) were co-occupied essentially only by Aire and MED1, CTCF signals being at background levels; a smaller fraction (Fig. 2*E*, 3) cobound essentially only CTCF; and another small fraction (Fig. 2*E*, 2) was co-occupied to a similar degree by all three of the other proteins (Fig. 2*E* and *SI Appendix*, Fig. S5*C*). Accordingly, genomic sites that bound CTCF were usually co-occupied by SMC1, but rarely by Aire or MED1 (Fig. 2*F* and *SI Appendix*, Fig. S5*D*). These observations are further illustrated by Venn diagram of the ChIP-seq peaks for the various proteins (*SI Appendix*, Fig. S5*E*).

According to multiple assays, then, Aire interacted with mediator and cohesin complexes distributed across the mTEC genome. In contrast, it did not detectably interact with CTCF, even when the latter was associated with cohesin complexes.

Aire Promoted Accumulation of Cohesin and Mediator Complexes at Superenhancers. Having observed an association between Aire and protein complexes known to be involved in transcriptional looping, we then addressed the functional significance of this liaison. As mentioned above, the cohesin and mediator complexes facilitate looping of superenhancers to their target promoters; so we determined whether Aire promotes accumulation of these two complexes on superenhancers. Defining superenhancers

as genome stretches of an average size of ∼30 kb abundantly loaded with histone-3 molecules acetylated at lysine-27 (H3K27ac), a chromatin mark diagnostic of active enhancers, we previously enumerated 1,170 superenhancers in mTECs (3). The superenhancers so defined were also enriched in Aire, the mediator subunit MED1, and the cohesin subunit SMC1, but not in CTCF (Fig. 3 *A* and *B*). However, an augmentation of CTCF near the 5′ and/or 3′ superenhancer boundaries was often apparent. There were strong correlations between the Aire signal at superenhancers and those of MED1 and SMC1, suggesting that Aire might have promoted their deposition on superenhancers (Fig. 3*C*). Indeed, the abundances of MED1 and SMC1 at superenhancers were significantly higher in Aire$^+$ than Aire$^-$ mTECs (Fig. 3*D*). In addition, a comparison of Aire-induced changes in superenhancer-localized SMC1 vs. MED1 signals revealed a strong correlation (Fig. 3*E*). Tracings of a few genomic regions illustrating these observations as well as local looping patterns are provided in Fig. 3*F* and *SI Appendix*, Fig. S6.

Thus, Aire coordinately promoted preferential accumulation of MED1 and SMC1 on superenhancers. These enrichments were associated with some newly evident transcriptional loops.

Aire Decreased the Levels of CTCF at Domain Boundaries, Associated with a Reduction in the Numbers of Structural Loops. Superenhancers are often ensconced within larger chromatin domains delimited by cohesin and CTCF, namely TADs and CDs. Access of superenhancers to target promoters may require opening of these higher-order structures. Aire did seem to influence higher-order chromatin architecture as the total number of TADs decreased from 5,024 in Aire$^-$ mTECs to 4,733 in their Aire$^+$ counterparts. In addition, the median TAD length was slightly longer in Aire$^+$ than in Aire$^-$ mTECs (375 vs. 350 kb; $P = 0.02$) (Fig. 4*A*). TADs in Aire$^+$ mTECs also showed a trend toward being less insulating (median insulation score −0.52 in Aire$^+$ vs. −0.55 in Aire$^-$ mTECs, with more negative scores signifying stronger insulation; $P = 0.057$) (Fig. 4*A*). These observations are reminiscent of the above-mentioned finding that the total loop number was markedly reduced in Aire$^+$ vis-à-vis Aire$^-$ mTECs. Nearly 90% of the loop anchors contained CTCF peaks, indicating that we were primarily detecting loss of CTCF-mediated loops; no such association was observed with randomly shuffled loop anchors (*SI Appendix*, Fig. S7*A*). In addition, more interchromosomal contacts, especially around the TSSs, were made by Aire-induced than Aire-neutral genes in Aire$^+$, but not Aire$^-$, mTECs (*SI Appendix*, Fig. S7*B*).

According to the extrusion model of loop formation, DNA is forced through a cohesin ring until CTCF molecules bound to convergent CTCF-binding motifs are encountered at the two ends, forming a barrier to further extrusion, and thereby anchoring a loop (15). One mechanism by which Aire could potentially impact higher-order chromatin structure is by eviction of CTCF from architectural boundaries. To test this hypothesis, we sorted 39,138 genomic sites occupied by CTCF in mTEC chromatin and compared the CTCF ChIP-seq signals for Aire$^-$ vs. Aire$^+$ mTECs. The density of CTCF binding was significantly lower for Aire$^+$ mTECs (Fig. 4*B*). Similar results were observed with 14,146 CTCF–SMC1 co-occupied sites (*SI Appendix*, Fig. S7*C*) as well as 4,171 CTCF-bound sites at TAD boundaries (*SI Appendix*, Fig. S7*D*), arguing that Aire promoted eviction of CTCF from architectural domain boundaries. Loss of CTCF from architectural domain boundaries did not reflect decreased transcription of *Ctcf* in Aire$^+$ vs. Aire$^-$ mTECs (*SI Appendix*, Fig. S7E). Eviction of CTCF from TAD/CD boundaries in Aire$^+$ mTECs was accompanied by a reduction in CTCF–CTCF loops, as indicated by an analysis of aggregate Hi-C signals at sites bound by CTCF, both across the entire genome (Fig. 4*C*) and specifically at TAD boundaries (*SI Appendix*, Fig. S7*F*). A few representative genomic regions demonstrating the loss of CTCF–CTCF loops in Hi-C data from WT vs. mutant mTECs are presented in Fig. 4*D*, while high-resolution examples of CTCF eviction are shown in Fig. 4*E*.

These results were corroborated by coimmunoprecipitation experiments on transfected HEK293T cells, where Aire expression reduced the quantities of cohesin constituents (SMC1, SMC3, RAD21, and SA-2) co-IPed with CTCF (Fig. 4*F*). Correspondingly, expression of Aire compromised association of CTCF with the cohesin subunit SMC1 (Fig. 4*F*).

Potential Mechanisms. We then weighed several possible mechanisms related to the Aire-induced accumulation of cohesin at superenhancers and eviction of CTCF from TAD boundaries. First, we explored CTCF's interaction with its binding sites in the absence and presence of Aire. Deeper analysis of our previously published ATAC-seq (assay for transposase-accessible chromatin followed by high-throughput sequencing) data (3) failed to uncover any differences in the precise CTCF footprints for Aire$^-$ vs. Aire$^+$ mTECs (*SI Appendix*, Fig. S7*G*). Neither did we find, also using previously published data (13), major differences in methylation of the CTCF-binding site in mutant and WT mTECs (*SI Appendix*, Fig. S7*H*), relevant here because such DNA modifications can loosen CTCF binding.

Second, we examined the effect of Aire on molecules known to regulate cohesin assembly on DNA—namely the loader, NIPBL, and unloader, WAPL, elements that combine to determine overall loop lengths (16). Interestingly, the effect of removing WAPL has been reported to be quite similar to that of introducing Aire: extension of architectural chromatin loops and increased interaction between genomic loci within separate TADs (17–19). Yet, coimmunoprecipitation experiments failed to demonstrate an interaction between Aire and WAPL, nor was there evidence that Aire interfered with WAPL's association with other cohesin subunits such as SMC1, SA-1, and SA-2 (Fig. 5*A*). In contrast, Aire did interact with NIPBL, and it also enhanced NIPBL's association with all of the enzymatic cohesin subunits examined (Fig. 5*B*). Analysis of NIPBL's distribution across the mTEC genome by "cleavage under targets and tagmentation" (CUT&Tag) revealed that NIPBL preferentially localized to superenhancers (Fig. 5 *C* and *D*) and that there was a strong correlation between the Aire and NIPBL signals at superenhancers (Fig. 5*E*).

The biochemical and, in particular, the genomic data argue for a model wherein Aire promoted the eviction of CTCF from domain boundaries and the deposition of cohesin on superenhancers. Consequently, Aire expression favored superenhancer–promoter transcriptional loops while disfavoring the architectural loops delimiting CDs and TADs. Aire–NIPBL interactions, whether direct or indirect, may prove to be important drivers of these changes.

Cohesin Was Essential for Aire-Induced Gene Expression in mTECs. To determine how important cohesin was for Aire-induced gene expression in mTECs, we generated a mouse line wherein *Stag2*, the gene encoding SA-2, was deleted specifically in TECs by combining *Foxn1*^{Cre} and *Stag2*^{fl} genetic elements. Controls were of *Foxn1*$^{Cre-/-}$ genotype while experimentals were *Foxn1*$^{Cre+/-}$ heterozygotes; both groups were of genotype *Stag2*$^{fl/fl}$. (Note: It has proven impossible for us to generate appropriate *Foxn1*-$^{Cre+/+}$ homozygotes, as males carrying *Foxn1*^{Cre} show germline deletion of floxed alleles.) We confirmed that *Stag2* loss did not compromise the mTEC compartment by flow cytometric comparison of thymi from mutant and WT littermates. Thymus size was not affected by the mutation (data not shown), nor was there a detectable impact on any of the stromal (Fig. 6*A*) or thymocyte (*SI Appendix*, Fig. S8*A*) compartments examined.

To assess the transcriptional impact of SA-2, we isolated RNA from mTEC^{hi} of mice with the TEC-specific mutation of *Stag2* vs. control littermates, and performed low-input RNA-sequencing (RNA-seq) analysis. RNA-seq replicates for each genotype were well-correlated, confirming reproducibility of the data (*SI*

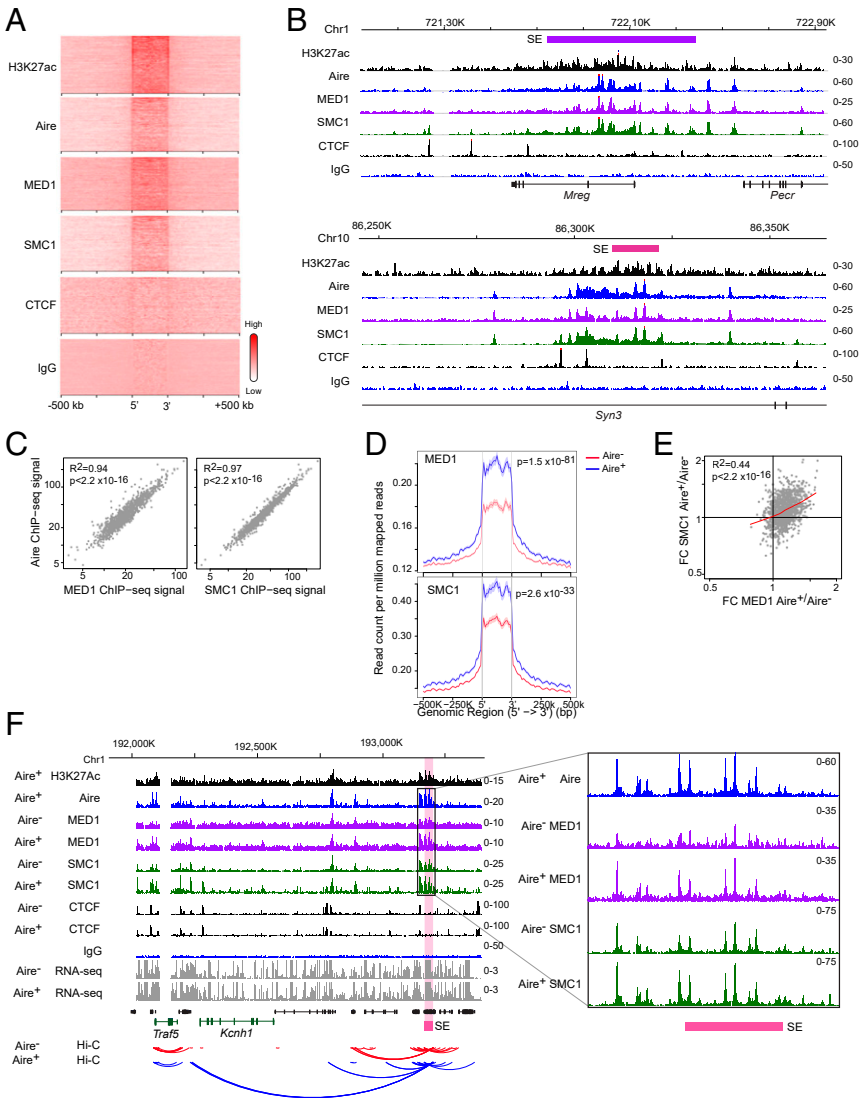

Fig. 3. Aire promotes deposition of mediator and cohesin complexes at superenhancers. (*A*) Heatmaps representing the ChIP density of the indicated proteins 500 kb up- or downstream of H3K27ac-delimited superenhancers in Aire$^+$ mTECs. Normalized coverage vectors were computed with tag density information and represented as heatmaps. (*B*) Normalized ChIP-seq profiles for the indicated proteins at exemplar superenhancers (pink and purple bars). Numbers (*Right*) indicate the ranges of normalized tag densities. (*C*) Scatterplots for the Aire vs. MED1 ChIP-seq signals (*Left*) or Aire vs. SMC1 ChIP-seq signals (*Right*) at superenhancers. (*D*) Tag density for MED1 (*Top*) or SMC1 (*Bottom*) at superenhancers in Aire$^-$ vs. Aire$^+$ mTECs. *P* values are from the Wilcoxon rank-sum test. (*E*) Scatterplot indicating the correlation between Aire-induced MED1 and SMC1 binding at superenhancers. Red line, lowess curve. FC, fold change. (*F*) Genome browser view of ChIP-seq signal for the indicated proteins in Aire$^-$ and Aire$^+$ mTECs, RNA-seq signal from Aire$^-$ and Aire$^+$ mTECs, and enrichment of superenhancer–promoter loops in Aire$^+$ mTECs at the *Kcnh1*/*Traf5* locus. Two Aire-induced genes are highlighted in green. A genomic locus corresponding to a superenhancer is highlighted in pink and expanded (*Right*) to emphasize individual ChIP-seq peaks. Analysis was performed on pooled data from two independent experiments except for *C* and *E*, where each dot represents mean signal from two replicates.

Appendix, Fig. S8*B*). *Stag2* transcripts were reduced but not ablated in the mTEC^{hi} of mutant mice, likely reflecting the *Cre* heterozygosity (*SI Appendix*, Fig. S8*C*). When expression values for the 500 genes most up-regulated by Aire in mTECs were superimposed on a volcano plot comparing mTEC gene expression for mice with or without the TEC-specific *Stag2* mutation, a strong overlap between Aire-induced and SA-2–dependent transcripts was evident (Fig. 6 *B*, *Left*). Such a striking concordance was not observed for a set of 500 expression-matched Aire-neutral genes (Fig. 6 *B*, *Right*). Examination of a heatmap of Aire-induced or Aire-neutral gene expression demonstrated that *Stag2* dependence was strongest for the most strongly Aire-induced genes, and that Aire-neutral genes showed no such dependence (*SI Appendix*, Fig. S8*D*). Importantly, the strong overlap between Aire-induced and SA-2–dependent gene transcripts was also absent when mTECs from double-deficient (Aire/SA-2) mice were analyzed (Fig. 6*C* and *SI Appendix*, Fig. S8*D*). Concordant with these findings, a strong correlation between SA-2–dependent and Aire-induced transcripts was observed when expression values for the 500 genes most down-regulated in the absence of SA-2 were overlaid on a volcano plot comparing mTEC gene expression for Aire$^+$ and Aire$^-$ mice (*SI Appendix*, Fig. S8*E*). These data indicate that the molecular circuitry of Aire and cohesin integrated in mTECs in vivo.

Discussion

The overall conclusion of this study is that Aire has a widespread impact on the 3D structure of mTEC chromosomes. It disfavors architectural loops such as those delimiting TADs and CDs while concomitantly enhancing chromatin transcribability by favoring

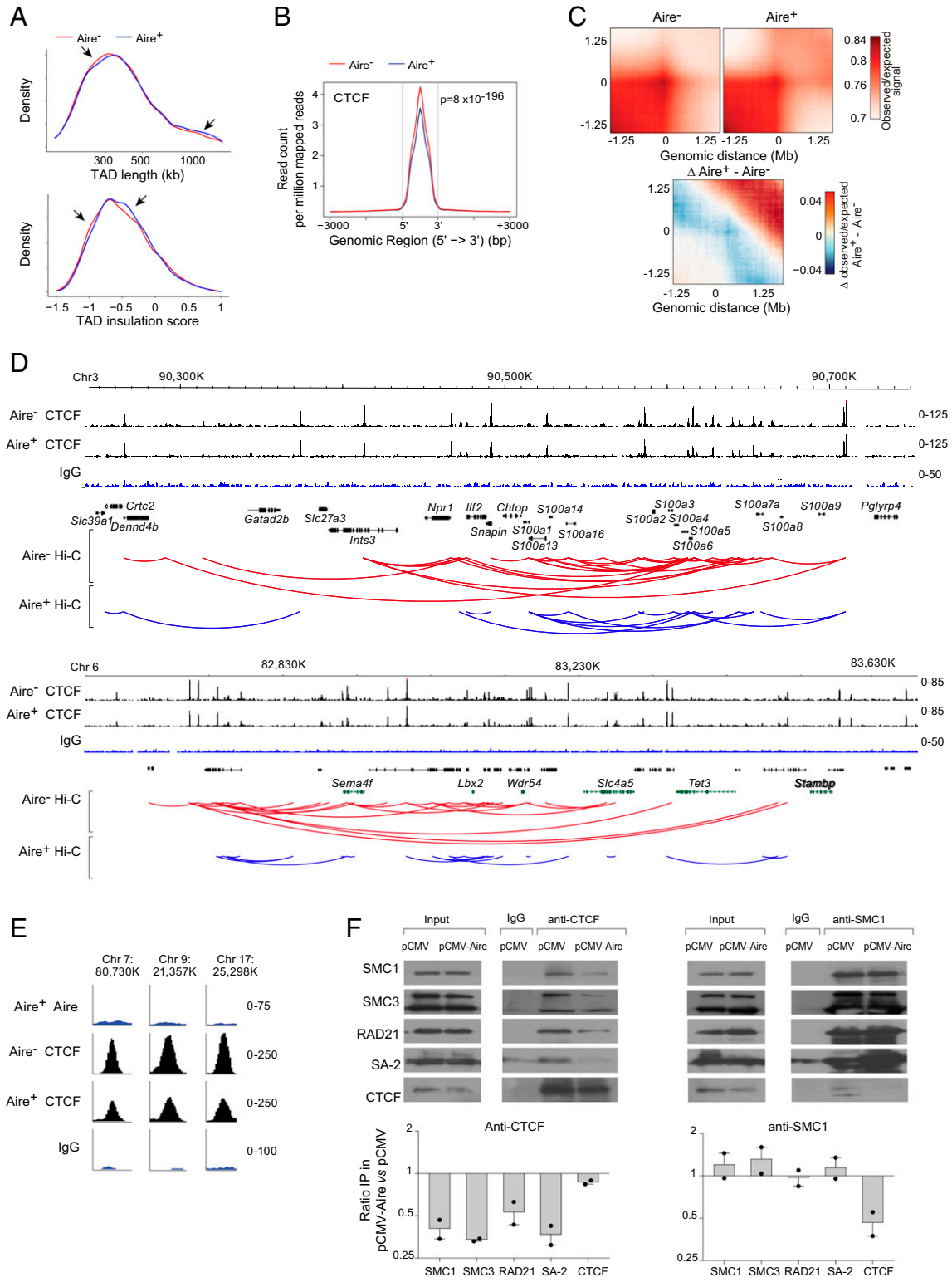

Fig. 4. Aire promotes loss of CTCF–CTCF loops by decreasing levels of CTCF at loop anchors. (*A*) Density plots showing the distribution of TAD lengths (*Top*) and insulation scores at TAD boundaries (*Bottom*) in Aire⁻ vs. Aire⁺ mTECs. Arrows highlight the difference between Aire⁻ and Aire⁺ mTECs. Data analysis was performed at 25-kb resolution. (*B*) Line plot displaying ChIP-seq tag densities for CTCF in Aire⁻ vs. Aire⁺ mTECs at CTCF sites across the genome. *P* values are from the Wilcoxon rank-sum test. (*C*) Heatmaps displaying Hi-C aggregate contacts for CTCF sites in Aire⁻ vs. Aire⁺ mTECs. Briefly, all such contacts were centered and averaged over 2-Mb regions, and the observed/expected signal in Aire⁺ vs. Aire⁻ mTECs (*Top*) as well as differential observed/expected signal for Aire⁺ vs. Aire⁻ mTECs (*Bottom*) was plotted. Data analysis was performed with a 50-kb Hi-C matrix. (*D*) Genome browser view of ChIP-seq tracks for CTCF and loops extracted from Hi-C data at peaks occupied by CTCF in Aire⁻ and Aire⁺ mTECs. (*E*) Genome browser view of loci displaying loss of CTCF in Aire⁺ vs. Aire⁻ mTECs. (*F*) Nuclear extracts from *FLAG/Aire*–transfected or control vector–transfected HEK293T cells were IPed with an anti-CTCF (*Left*) or anti-SMC1 (*Right*) Ab, and immunoblotting was performed for the indicated proteins. Summary data from two independent experiments are represented as a bar graph. Error bars represent SEM. ChIP-seq and Hi-C analysis are on pooled data from two independent experiments.

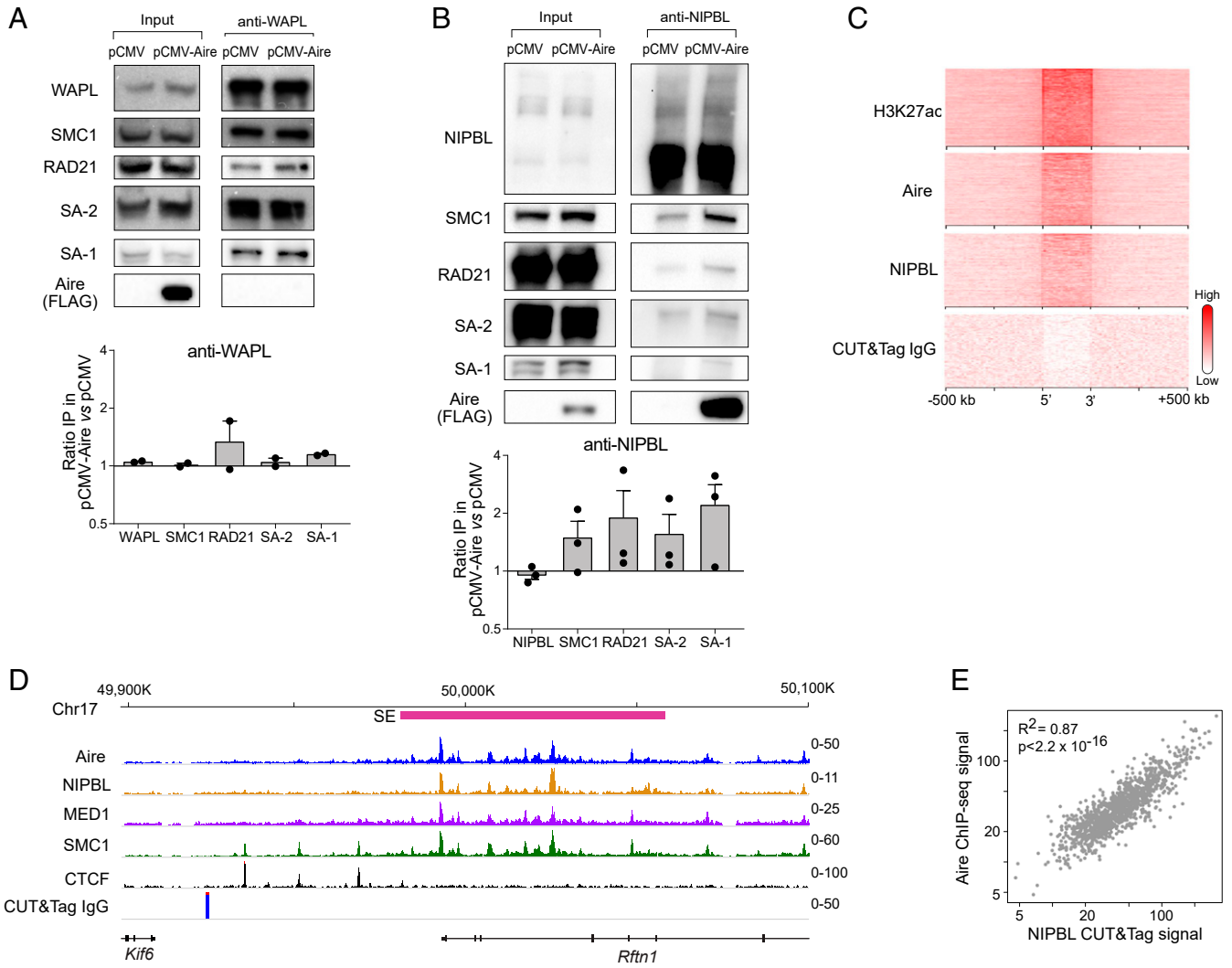

Fig. 5. Aire interacts with NIPBL, but not WAPL, in distributing cohesin to superenhancers. (*A* and *B*) HEK293T cells were transfected with *FLAG/Aire*–expressing plasmid or control vector for 48 h followed by nuclear extract preparation, IP with anti-WAPL (*A*) or anti-NIPBL (*B*) Abs, and immunoblotting for the indicated proteins. Bar graphs represent summary data for two (*A*) or three (*B*) independent experiments. Error bars represent SEM. (*C*) Heatmaps displaying the ChIP density of Aire and NIPBL 500 kb up- or downstream of H3K27ac-delimited superenhancers in Aire$^+$ mTECs. Heatmaps were plotted with normalized coverage vectors for tag density. (*D*) Genome browser track displaying occupancy of NIPBL at superenhancers. (*E*) Scatterplot for the Aire ChIP-seq vs. NIPBL CUT&Tag signal at superenhancers. ChIP-seq and CUT&Tag analysis was performed on pooled data from two (Aire) or three (NIPBL) independent experiments except for *E*, where each dot represents mean signal from two replicates of Aire ChIP-seq and three replicates of NIPBL CUT&Tag.

regulatory loops like those connecting superenhancers and promoters. Reflecting these changes, Aire promotes conversion of the inactive B chromatin state to the active A state. This constellation of effects is similar to what has been observed during differentiation of certain types of immunocytes, for example B cells or macrophages, from their precursors (20, 21).

Aire exerts its impact by promoting the accumulation of cohesin and mediator at superenhancers, thereby driving their participation in regulatory loops with promoters; concomitantly, it evicts CTCF from TAD or CD boundaries and thus antagonizes structural loops. This scenario seems superficially reminiscent of reports that nuclear factor κB or PU.1 "supervise" transcription during activation or differentiation by "stealing" TFs like BRD4, Satb1, or Runx1 from other locations (22, 23). How might Aire orchestrate CTCF eviction? According to both Western blotting and ChIP-seq data, Aire does not associate with CTCF; thus, Aire can influence loops emanating from sites where it, itself, is not stably bound. Nor does Aire seem to alter the precise CTCF contacts with its binding sites. It could, however, block cohesin access to CTCF during DNA extrusion, especially given the large multiprotein complexes in which Aire participates (14).

An important mechanistic clue for Aire's impact on cohesin distribution at superenhancers came from exploring its influences on the cohesin loader, NIPBL, and unloader, WAPL, elements whose activities combine to determine overall loop lengths (16). Aire did not bind to WAPL nor influence its binding to other components of the cohesin complex, but did interact, directly or indirectly, with NIPBL and enhanced NIPBL's associations with cohesin subunits. Moreover, Aire's binding at superenhancers was well-correlated with that of NIPBL. The potential significance of this finding emerges when integrating it with previously established aspects of Aire and NIPBL function. Aire is thought to stabilize DNA double-stranded breaks (DSBs) within superenhancers and promoters—induced by enhancer RNA transcription, nucleosome evacuation, or paused RNA-PolII—thereby provoking a local DNA-damage response (DDR) (3, 14). Several DDR elements conglomerate with Aire, and general transcription factors—such as

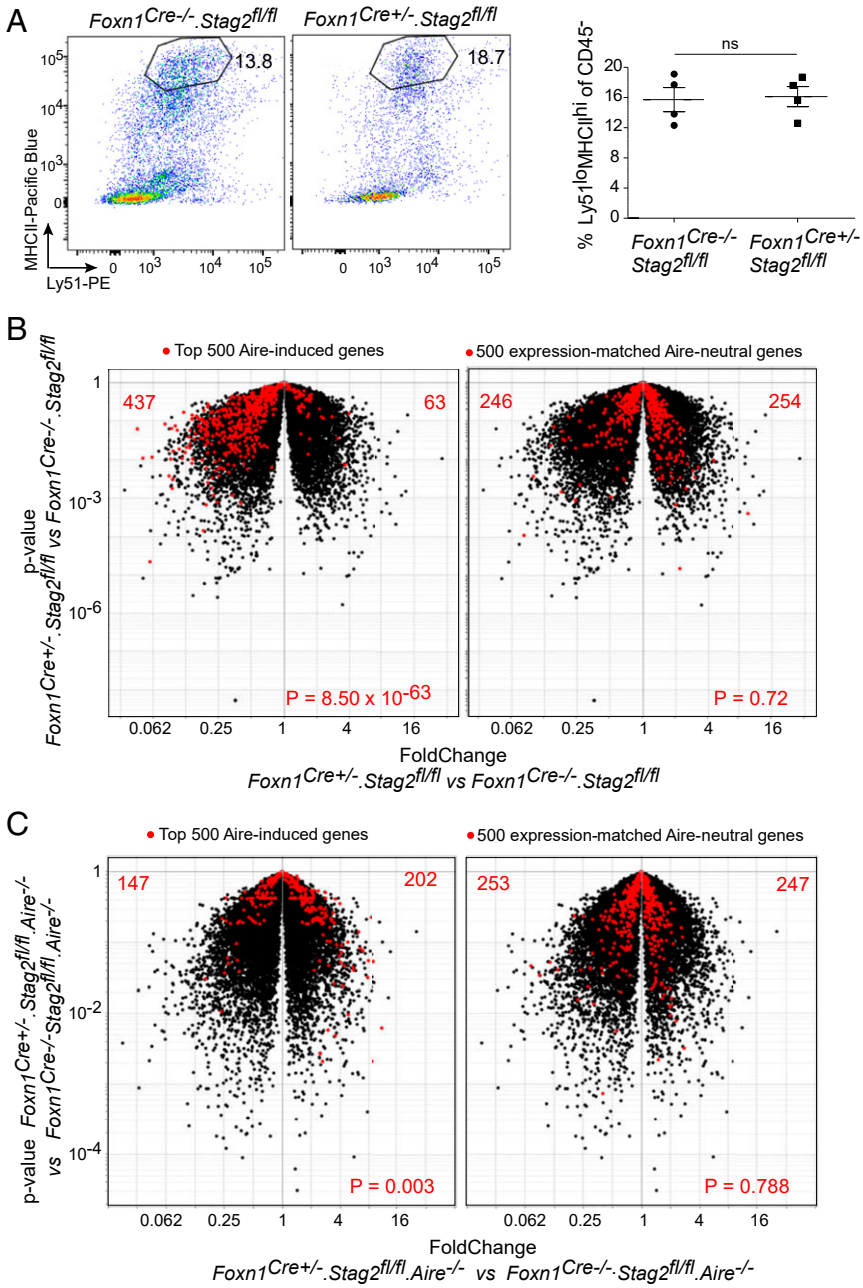

Fig. 6. Cohesin subunit SA-2 is required for Aire-induced mTEC gene transcription. (*A*) Flow cytometry analyzing the mTEC compartment from thymi of 4-wk-old mice of the indicated genotypes. Representative dot plots for $CD45^-$-gated cells (*Left*) and summary data for individual mice (*Right*). Error bars represent SEM. ns, not significant. (*B* and *C*) Volcano plots (fold change vs. *P* value) displaying transcriptome changes for mTECs of the indicated genotypes. The top 500 Aire-induced transcripts in this dataset and 500 expression-matched Aire-neutral transcripts are highlighted in red. Numbers (*Top*) refer to transcripts up- (*Right*) or down-regulated (*Left*) by conditional *Stag2* deletion. Average values are from independent triplicates or quadruplets. *P* values were calculated using a χ^2 test. Due to low expression of Aire-induced transcripts in mTECs from $Aire^{-/-}$ mice, not all of the top 500 Aire-induced transcripts are displayed in *C* (*Left*).

RNA-PolII, CBP, and BRD4—are also drawn in, resulting in superenhancers and promoters that host elevated concentrations of Aire and its critical functional collaborators. Aire–NIPBL interactions insert nicely into this scenario, potentially explaining how cohesin could accumulate on superenhancers. NIPBL is known to be attracted to DNA DSBs (24) and would thus be well-situated to load cohesin complexes onto DNA within superenhancers. Relatedly, NIPBL—like Aire—directly binds BRD4 on superenhancers and promoters (although at different sites on BRD4), driving transcription, especially at promotors with paused RNA-PolII (6, 25). Most recently, it was reported that NIPBL translocates along the DNA and is actually required for ongoing loop extrusion by cohesin (26). Nonetheless, future functional studies on NIPBL will be needed to precisely dissect its interplay with Aire.

There is mounting evidence that cohesin complexes tethering regulatory loops are structurally and functionally different from those mooring architectural loops; for example, the former contain essentially only SA-2, while the latter host both SA-1 and SA-2 (27, 28). It is possible that Aire's predation of cohesin somehow exploits this differential. For example, the strong

correspondence we observed between the genes whose expression was up-regulated by Aire and dependent on SA-2 could reflect their codedication to regulatory, rather than architectural, loops.

Regardless of the details, the model we propose in Fig. 7 can account for many previously reported particularities of Aire function. First, it explains how a single TF, by concomitantly affecting both architectural and regulatory loops, can impact the expression of a battery of genes [around 20% of the genome (13, 29)], which are otherwise expressed in different tissues, at different levels, and with different timing. Second, given that diverse cell types have divergent superenhancer–promoter and paused RNA-PolII landscapes, the model rationalizes the observation that introduction of Aire into a diversity of cell types always impacts expression of a large set of genes but the particular loci affected are highly variable (30). Third, it provides an explanation for Aire's preferential effect on genes encoding PTAs, reflecting observations that superenhancers are usually associated with cell-identity genes (31), are typically overloaded with cohesin components but not CTCF (7), and partake in loops to promoters or to each other that are constrained within TADs (8, 32). Fourth, in light of observations that CTCF (33) and CTCF-anchored chromatin loops (34) can regulate alternative splicing, the model affords a rationalization for Aire's reported effects on cotranscriptional messenger RNA splicing (14). Lastly, it nicely explains the peculiar variegated expression of Aire target genes in mTECs, namely a given PTA transcript is up-regulated in only a few percent of these cells at any one time (13, 29, 35, 36). It is now clear that chromatin conformation, in particular superenhancer–promoter looping, is highly dynamic (37–39), a conclusion solidified by recent single-cell Hi-C analyses (40, 41). Thus, at any particular instance, only a fraction of the potential loop landscape may be actualized.

Lastly, what have these findings taught us about T cell tolerance? Above all, they uncovered the root molecular mechanism by which mutation of a single transcriptional regulator in thymic stromal cells can result in broad, yet defined, autoreactivity to a set of proteins characteristic of fully differentiated cells in peripheral organs. That autoimmunity can result from defects in the regulation of long-range chromatin interactions seems to be an emerging theme. For example, Fasolino et al. recently reported that genetic loci conferring risk for type-1 diabetes NOD mice were "hyperconnected" in three dimensions in thymocytes of that strain but not diabetes-resistant C57BL/6 mice (42). These loci conglomerated the enhancers and promoters of genes related to T cell function. More broadly, many of the single-nucleotide polymorphisms causally associated with particular autoimmune diseases (e.g., type-1 diabetes, systemic lupus erythematosus) reside in noncoding regions of the genome (43); they are especially enriched in the superenhancers of disease-relevant cell types (7). Aire seems to employ a variant of this principle to broadly protect from autoimmune attack on a multiplicity of organs. By regulating superenhancer–promoter looping in mTECs, it regulates thymic expression of cell type–defining antigens, thereby promoting tolerance of cognate self-reactive T cells.

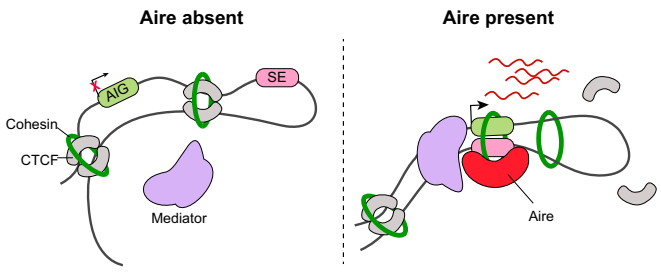

Fig. 7. Model for Aire-induced gene (AIG) transcription resulting from Aire's promotion of transcriptional (superenhancer–promoter) loops (*Right*) at the expense of architectural (TAD or CD delimiting) loops (*Left*). CTCF–cohesin complexes underlie the latter. Cohesin may be recruited from other chromatin sites (that do or do not cobind CTCF) or from the nucleoplasm.

Materials and Methods

Mice. All mice were maintained in a specific pathogen–free facility at Harvard Medical School. The breeding strategy to generate these mouse lines is detailed in *SI Appendix, Materials and Methods*. Comparisons were routinely performed on littermates. All mice were maintained in accordance with Harvard Medical School's Animal Care and Use Committee guidelines (Institutional Animal Care and Use Committee protocol IS00001257).

Staining and Sorting of Thymic Cell Populations. Thymi were dissected and connective tissue was removed with the help of forceps. Thymi were chopped with scissors and a single-cell suspension was obtained with a series of enzymatic digestions. Cells were stained with fluorochrome-labeled antibodies. Further details about enzymatic digestions, antibody staining, and cell sorting are in *SI Appendix, Materials and Methods*.

Hi-C Analysis. Hi-C was performed on Aire$^+$ and Aire$^-$ mTECs and analyzed as detailed in *SI Appendix, Materials and Methods*. Approximately three hundred million paired-end reads per library were generated on a NextSeq 500 instrument (Illumina). Analysis of Hi-C data was performed with standard programs like Juicer, Juicebox, HOMER, and Hi-C Explorer (44–46).

Cell Culture, Plasmids, Transfection, and Immunoprecipitation. HEK293T cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, L-glutamate, and penicillin/streptavidin antibiotics to about 70% confluence, and were maintained in a humidified atmosphere at 37 °C with 5% CO_2. For transfection, the cells were counted, seeded on 10-cm tissue-culture dishes, and transfected with a FLAG/*Aire* plasmid (47) using TransIT reagent (Mirus) according to the manufacturer's instructions. Transfected HEK293T cells were used for immunoprecipitation as detailed in *SI Appendix, Materials and Methods*.

ChIP-seq and CUT&Tag Analysis. ChIP-seq was performed on ~3×10^5 mTECs, while ~2×10^4 mTECs were used for CUT&Tag analysis. DNA libraries were sequenced on a NextSeq 500 (Illumina) machine to generate single-end libraries for ChIP-seq and paired-end libraries for CUT&Tag. Detailed protocols and analysis procedures are in *SI Appendix, Materials and Methods*.

ATAC-seq Footprinting. ATAC-seq footprinting was performed on data from Aire$^+$ and Aire$^-$ mTECs using a hidden Markov model approach, HINT-ATAC, as detailed in *SI Appendix, Materials and Methods*.

Methylation Analysis. Published reduced-representation bisulfite-sequencing (RRBS) data (13) were used to analyze the methylation status of the CTCF motif in Aire$^-$ and Aire$^+$ mTECs. Sequencing reads were trimmed to remove adaptor sequences with the RRBS option of the Trim Galore! wrapper script (v0.4.0) (Babraham Institute). Trimmed reads were aligned to the mm10 genome assembly with Bismark (v0.14.3) (48). The methylation status at specific genomic loci was computed using the Bioconductor package BiSeq (49).

RNA-seq Analysis. RNA-seq analysis on mTECs was performed using the ImmGen protocol, as detailed in *SI Appendix, Materials and Methods*.

Gene Expression Omnibus/Sequence Read Archive Accession Numbers. RNA-seq data from Aire$^+$ and Aire$^-$ mTECs: SRR2038194, SRR2038195, SRR2038196, and SRR2038197; RRBS data from Aire$^+$ and Aire$^-$ mTECs: SRR2038206, SRR2038210, SRR2038212, and SRR2038213; H3K27ac ChIP-seq data on mTECs: GSE74257 and GSE92597; Aire and immunoglobulin G ChIP-seq data on mTECs: GSE92597; ATAC-seq data from Aire$^+$ and Aire$^-$ mTECs: GSE92594.

Statistics. Statistical significance was assessed by a two-tailed Student's *t* test, a χ^2 test, or the Wilcoxon rank-sum test, as specified in the individual figure legends.

Data Availability. The high-throughput sequencing data reported in this paper have been deposited in the Gene Expression Omnibus repository under accession no. GSE180937. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank K. Hattori, N. Asinovski, and A. Ortiz-Lopez for experimental assistance, Dr. I. Aifantis for helpful advice, and Dr. C. Laplace for drafting final figures. This work was funded by R01AI088204 from the National Institute of Allergy and Infectious Diseases and by RO1DK060027 from the National Institute of Diabetes and Digestive and Kidney Diseases. Some of the work relied on facilities supported by P30CA008748 from the National Cancer Institute. K.B. received an American Diabetes Association Mentor-Based Postdoctoral Fellowship (7-12-MN-51) and Department of Biotechnology/Wellcome Trust India Alliance Intermediate Fellowship (IA/I/19/1/504276). D.A.M. is supported by T32GM007753 from the National Institute of General Medical Sciences.

1. J. Abramson, G. Anderson, Thymic epithelial cells. *Annu. Rev. Immunol.* **35**, 85–118 (2017).
2. G. M. Constantine, M. S. Lionakis, Lessons from primary immunodeficiencies: Autoimmune regulator and autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy. *Immunol. Rev.* **287**, 103–120 (2019).
3. K. Bansal, H. Yoshida, C. Benoist, D. Mathis, The transcriptional regulator Aire binds to and activates super-enhancers. *Nat. Immunol.* **18**, 263–273 (2017).
4. I. Oven *et al*., AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. *Mol. Cell. Biol.* **27**, 8815–8823 (2007).
5. M. Giraud *et al*., Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 535–540 (2012).
6. H. Yoshida *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.* **112**, E4448–E4457 (2015).
7. D. Hnisz *et al*., Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
8. N. Krietenstein *et al*., Ultrastructural details of mammalian chromosome architecture. *Mol. Cell* **78**, 554–565.e7 (2020).
9. S. S. Rao *et al*., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
10. W. Schwarzer *et al*., Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
11. T. M. Johanson *et al*., Transcription-factor-mediated supervision of global genome architecture maintains B cell identity. *Nat. Immunol.* **19**, 1257–1264 (2018).
12. E. Lieberman-Aiden *et al*., Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
13. M. Meredith, D. Zemmour, D. Mathis, C. Benoist, Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nat. Immunol.* **16**, 942–949 (2015).
14. J. Abramson, M. Giraud, C. Benoist, D. Mathis, Aire's partners in the molecular control of immunological tolerance. *Cell* **140**, 123–135 (2010).
15. A. L. Sanborn *et al*., Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6456–E6465 (2015).
16. S. Schoenfelder, P. Fraser, Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
17. G. A. Busslinger *et al*., Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503–507 (2017).
18. J. H. I. Haarhuis *et al*., The cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169**, 693–707.e14 (2017).
19. G. Wutz *et al*., Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).
20. D. H. Phanstiel *et al*., Static and dynamic DNA loops form AP-1-bound activation hubs during macrophage development. *Mol. Cell* **67**, 1037–1048.e6 (2017).
21. K. R. Kieffer-Kwon *et al*., Myc regulates chromatin decompaction and nuclear architecture during B cell activation. *Mol. Cell* **67**, 566–578.e10 (2017).
22. H. Hosokawa *et al*., Transcription factor PU.1 represses and activates gene expression in early T cells by redirecting partner transcription factor binding. *Immunity* **48**, 1119–1134.e7 (2018).
23. S. F. Schmidt *et al*., Acute TNF-induced repression of cell identity genes is mediated by NFκB-directed redistribution of cofactors from super-enhancers. *Genome Res.* **25**, 1281–1294 (2015).
24. C. Bot *et al*., Independent mechanisms recruit the cohesin loader protein NIPBL to sites of DNA damage. *J. Cell Sci.* **130**, 1134–1146 (2017).
25. N. Luna-Peláez *et al*., The Cornelia de Lange syndrome-associated factor NIPBL interacts with BRD4 ET domain for transcription control of a common set of genes. *Cell Death Dis.* **10**, 548 (2019).
26. I. F. Davidson *et al*., DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).
27. S. Remeseiro, A. Cuadrado, G. Gómez-López, D. G. Pisano, A. Losada, A unique role of cohesin-SA1 in gene regulation and development. *EMBO J.* **31**, 2090–2102 (2012).
28. A. Kojic *et al*., Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat. Struct. Mol. Biol.* **25**, 496–504 (2018).
29. P. Brennecke *et al*., Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. *Nat. Immunol.* **16**, 933–941 (2015).
30. M. Guerau-de-Arellano, D. Mathis, C. Benoist, Transcriptional impact of Aire varies with cell type. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14011–14016 (2008).
31. W. A. Whyte *et al*., Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
32. J. M. Dowen *et al*., Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
33. S. Shukla *et al*., CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011).
34. M. Ruiz-Velasco *et al*., CTCF-mediated chromatin loops between promoter and gene body regulate alternative splicing across individuals. *Cell Syst.* **5**, 628–637.e6 (2017).
35. J. Derbinski, S. Pinto, S. Rösch, K. Hexel, B. Kyewski, Promiscuous gene expression patterns in single medullary thymic epithelial cells argue for a stochastic mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 657–662 (2008).
36. J. Villaseñor, W. Besse, C. Benoist, D. Mathis, Ectopic expression of peripheral-tissue antigens in the thymic epithelium: Probabilistic, monoallelic, misinitiated. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15854–15859 (2008).
37. L. Giorgetti *et al*., Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157**, 950–963 (2014).
38. A. S. Hansen, I. Pustova, C. Cattoglio, R. Tjian, X. Darzacq, CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* **6**, e25776 (2017).
39. A. S. Hansen, C. Cattoglio, X. Darzacq, R. Tjian, Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20–32 (2018).
40. T. Nagano *et al*., Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
41. T. J. Stevens *et al*., 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
42. M. Fasolino *et al*.; HPAP Consortium, Genetic variation in type 1 diabetes reconfigures the 3D chromatin organization of T cells and alters gene expression. *Immunity* **52**, 257–274.e11 (2020).
43. K. K. Farh *et al*., Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
44. N. C. Durand *et al*., Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
45. S. Heinz *et al*., Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
46. J. Wolff *et al*., Galaxy HiCExplorer: A web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **46**, W11–W16 (2018).
47. S. Yang, K. Bansal, J. Lopes, C. Benoist, D. Mathis, Aire's plant homeodomain(PHD)-2 is critical for induction of immunological tolerance. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1833–1838 (2013).
48. F. Krueger, S. R. Andrews, Bismark: A flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
49. K. Hebestreit, M. Dugas, H. U. Klein, Detection of significantly differentially methylated regions in targeted bisulfite sequencing data. *Bioinformatics* **29**, 1647–1653 (2013).