CMI

Check for updates

ARTICLE

Insm1 regulates mTEC development and immune tolerance

Weihua Tao[1,2,3,4,7], Zhihuan Ye[1,2,3,7], Yiqiu Wei[1,2,3,7], Jianxue Wang[1], Weixin Yang[1], Guoxing Yu[1,2,3], Jieyi Xiong [5,6✉] and Shiqi Jia [1,2,3,4✉]

The expression of self-antigens in medullary thymic epithelial cells (mTECs) is essential for the establishment of immune tolerance, but the regulatory network that controls the generation and maintenance of the multitude of cell populations expressing self-antigens is poorly understood. Here, we show that Insm1, a zinc finger protein with known functions in neuroendocrine and neuronal cells, is broadly coexpressed with an autoimmune regulator (Aire) in mTECs. *Insm1* expression is undetectable in most mimetic cell populations derived from mTECs but persists in neuroendocrine mimetic cells. Mutation of *Insm1* in mice downregulated *Aire* expression, dysregulated the gene expression program of mTECs, and altered mTEC subpopulations and the expression of tissue-restricted antigens. Consistent with these findings, loss of Insm1 resulted in autoimmune responses in multiple peripheral tissues. We found that Insm1 regulates gene expression in mTECs by binding to chromatin. Interestingly, the majority of the Insm1 binding sites are co-occupied by Aire and enriched in superenhancer regions. Together, our data demonstrate the important role of Insm1 in the regulation of the repertoire of self-antigens needed to establish immune tolerance.

INTRODUCTION

The thymus is a primary lymphoid organ where T-cell progenitors undergo maturation and selection to become functional T cells. Immature T cells are positively selected for their ability to recognize the histocompatibility complex by cortical thymic epithelial cells (cTECs). A second negative selection step by medullary thymic epithelial cells (mTECs) eliminates T cells that recognize self-antigens [1, 2]. mTECs express thousands of peripheral tissue-restricted antigens (TRAs) in a mosaic fashion, which is believed to be key for negative selection and the establishment of self-tolerance [3]. The autoimmune regulator (Aire) is an essential transcription factor that directly regulates TRAs expression and autoimmunity [4, 5].

Recent studies using single-cell RNA sequencing (scRNA-seq) identified different stages of mTEC development, e.g., progenitors, immature TECs, transit-amplifying TECs, Aire-expressing mTECs, and highly differentiated mTEC subtypes that have downregulated expression of Aire (post-Aire cells) [6]. The heterogeneous post-Aire cells resemble peripheral cell types in their molecular characteristics and are, therefore, also called mimetic cells [7–10]. Interestingly, these mimetic cells express transcription factors that control the differentiation of the peripheral cells that they mimic, and the same factors are also needed for the formation of the particular subtype of mimetic cells [7]. Among the mimetic cells, FoxA-positive neuroendocrine cells, Hnf4a-positive enterocyte/

hepatocyte cells, Sox8/SpiB-positive microfold cells, Pou2f3-positive tuft cells, FoxJ-positive ciliated cells, Grhl-positive keratinocytes and Myog-positive muscle cells were defined [7, 11]. Thus, the cellular and molecular bases of central tolerance are established by the expression of TRAs in Aire-expressing cells and post-Aire mimetic cells [11].

In the last two decades, important insights into the molecular control of TRA expression by Aire have been obtained [12–17]. In particular, Aire regulates transcription and chromatin accessibility [12, 14, 18]. Furthermore, Aire was shown to bind to super-enhancers and to promote the interaction between enhancers and TRA genes [19]. Studies on Aire have also revealed that correct TRA expression in the perinatal but not adult stage is essential for the establishment of self-resistance [20–22], but the mechanism underlying this process is not fully understood [23]. Differences in the perinatal and adult cell populations of the thymus are known to exist and may have functional implications for establishing self-tolerance during early life. For instance, TEC progenitors are more abundant in the perinatal stage, and their number dramatically decreases with aging [23]. Similarly, muscle mTECs are more abundant in the perinatal than adult stages [7].

Herzig et al. screened potential regulators of Aire expression and identified Insulinoma-associated protein 1 (Insm1) as a transcription factor expressed in mTECs and as a candidate regulator of Aire. However, no functional analysis of Insm1 was

performed in the thymus [24]. Here, we present a genetic and molecular analysis of *Insm1*, demonstrating that Insm1 functions in both mTECs and neuroendocrine mimetic cells for the expression of self-antigens.

RESULTS
Insm1 is expressed in Aire-expressing mTECs and neuroendocrine mimetic cells
To investigate the expression of *Insm1* in the thymus, we performed immunofluorescence analysis. For this, we used *Insm1*[+/lacZ] and *Insm1*[lacZ/lacZ] mice in which one or two alleles of the *Insm1* codon sequence were replaced by *lacZ*, which encodes beta-galactosidase (β-gal) [25, 26]. Therefore, the expression of *Insm1* can be monitored by β-gal. We compared signals obtained using antibodies against β-gal and Insm1 and observed that β-gal was expressed in the medulla of the thymus in both *Insm1*[+/lacZ] and *Insm1*[lacZ/lacZ] mice (Fig. S1A). Insm1 immunoreactivity overlapped with β-gal immunoreactivity in the thymus of *Insm1*[+/lacZ] mice but was absent in *Insm1*[lacZ/lacZ] mice (Fig. 1A and Fig. S1A). Thus, the Insm1 antibody specifically detected endogenous Insm1 expression. In contrast to the exclusive nuclear location in neuroendocrine cells [27, 28], Insm1 protein was detected in both the nuclei and the cytoplasm of thymic cells (Fig. 1B).

The thymic epithelial compartment is formed by cTECs and mTECs and contains lymphocytes and dendritic cells [29]. The expression of Insm1 in thymic cells was previously screened using gene microarray datasets [24]. To further define the thymic cell types that express *Insm1*, we performed immunofluorescence analysis at both the fetal (E18.5) and adult stages. Insm1 was coexpressed with the mTEC marker keratin 5 (Krt5) but was not detected in keratin 8+ (Krt8) cTECs, CD45+ lymphocytes, or Cd11b + /Cd11c+ dendritic cells in the thymuses of fetal and adult mice (Fig. 1C, D and Fig. S1B, C). The majority (70%) of Insm1+ cells coexpressed Aire and vice versa at E18.5 (Fig. 1E). Coexpression of Insm1 and Aire was somewhat less pronounced at the adult stage, in which 39% of Insm1+ cells coexpressed Aire and 56% of Aire+ cells coexpressed Insm1 (Fig. 1E). Moreover, Aire-dependent TRAs such as insulin and somatostatin were detected in Insm1+ cells in adult thymuses (Fig. S1D, E). To systematically investigate the expression of *Insm1* in mTECs of the developing and adult thymus, we reanalyzed a published single-cell RNAseq dataset that relied on post-Aire mTECs [7] (Fig. S2A). *Insm1* expression was significantly enriched in neuroendocrine mimetic cells of both perinatal and adult mice (one-tailed Wilcoxon test, $p < 0.001$; Fig. 1F, Fig. S1A–C and Table S1). This analysis also showed that Insm1 was expressed in transit-amplifying cells and Aire-expressing mTECs, cell types that express Aire, at both the perinatal and adult stages (Fig. 1F, G). In addition, the expression of Insm1 in neuroendocrine mimetic cells, which express Foxa2 and chromogranin A, was verified by immunohistology in adult thymuses (Fig. 2A, B). In summary, Insm1 is broadly coexpressed with Aire and is also found in post-Aire neuroendocrine mimetic cells.

Insm1 regulates the expression of *Aire*
To investigate the function of Insm1 in the thymus, we analyzed thymic phenotypes using conditional mutagenesis, e.g., *Foxn1-Cre;Insm1*[flox/flox] (also called *Insm1-thycKO* or *Insm1* mutants; *Insm1*[flox/flox] were used as controls). Flow cytometry demonstrated comparable numbers of EpCAM+ thymic epithelial cells (CD45−EpCAM+) and UEA1+ mTECs (CD45−EpCAM+UEA1+) in the *Insm1* mutants and controls (Fig. S2D–F). However, the proportions of Aire-positive mTECs (CD45−EpCAM+UEA1+Aire+) were decreased (Fig. 2C). Furthermore, the immunofluorescence intensity indicated that the Aire protein level was moderately but significantly decreased in Aire-expressing mTECs from the *Insm1-*

thycKO mice (Fig. 2D). Thus, the *Insm1* mutation results in a reduced number of Aire-expressing cells and decreased expression of Aire in mTECs.

We overexpressed *Insm1* (*Insm1OE*) in thymic epithelial cells in vivo using genetic tools (Fig. S3A–C), e.g., a mouse line that harbors a *loxp-STOP-loxp-Insm1* cassette in the *Rosa26* locus, which combined with the *Foxn1Cre* allele for thymus-specific overexpression. Significantly increased *Insm1 and Aire* transcription levels were detected when the RNA of the entire thymus but not of isolated mTECs was analyzed (Fig. S3D, E). Flow cytometry analysis identified that the numbers of UEA1+ mTECs and Aire+ cells were increased (Fig. 2E, F). This finding was further verified by immunohistology analysis, which showed increased numbers of Insm1+ cells and Aire+ cells and enlarged Krt5+ mTEC areas in the *Insm1OE* thymus (Fig. 2G). Thus, *Aire* is expressed more broadly in mTECs after *Insm1* overexpression, and conversely, *Aire* expression is reduced after ablation of *Insm1*. We conclude that correct Aire expression depends on Insm1.

Insm1 regulates the differentiation of neuroendocrine mimetic cells
To determine whether Insm1 regulates mimetic cell differentiation, we examined cell type-specific marker gene expression [7] in the thymus of *Insm1-thycKO* mice. Among the tested genes that mark the identity of mimetic cells (Fig. 3A upper panel), those expressed in neuroendocrine and enterocyte/hepatocyte cells were downregulated (Fig. 3A lower panel). We further found a reduced number of ChgA+ neuroendocrine mimetic cells using flow cytometry analysis in *Insm1-thycKO* mice (Fig. 3B). Thus, Insm1 regulates the generation of neuroendocrine mimetic cells. Enterocyte/hepatocyte mimetic cells that do not express Insm1 were also affected by *Insm1* mutation, and we hypothesized that this finding might be because *Insm1* is expressed in transit-amplifying and Aire-expressing cells, which were suggested to give rise to various mimetic cell types, including enterocyte/hepatocyte mimetic cells [6]. To test this directly, we used the *Insm1*[+/lacZ] strain that expresses the product of the *lacZ* gene, β-galactosidase, under the control of the Insm1 promoter. β-galactosidase has a long half-life and, therefore can be used for lineage tracing. Hfn4α is a marker for enterocyte/hepatocyte mimetic cells. We detected coexpression of Hfn4α and β-galactosidase but not coexpression of Hfn4α and Insm1 (Fig. 3C). Thus, enterocyte/hepatocyte mimetic cells expressed Insm1 during their development but downregulated the gene in the mature stage.

Using *Insm1OE* mice, we investigated the expression of genes that mark neuroendocrine and enterocyte/hepatocyte mimetic cells. Increased levels of neuroendocrine but not enterocyte/hepatocyte marker genes were detected when RNA isolated from mTECs or the entire thymus was analyzed (Fig. 3D and Fig. S3F). Consistent with the increased expression of marker genes, the numbers of neuroendocrine mimetic cells were increased in *Insm1OE* mice (Fig. 3E). Thus, overexpression of Insm1 promotes the generation of neuroendocrine mimetic cells.

Altered TRAs levels in Insm1 mutant mTECs
Next, we systematically analyzed gene expression, particularly the expression of TRAs, in mTECs of fetal *Insm1-null* (*Insm1*[lacZ/lacZ]) and adult *Insm1-thycKO* mice by RNAseq. Gene set enrichment analysis (GSEA) showed that the downregulated genes in the fetal *Insm1-null* mice were enriched in the development of multiple tissue types, such as keratinocyte differentiation (normalized enrichment score, $NES = -2.17$; $q = 0.009$) and striated muscle cell development ($NES = -2.07$, $q = 0.038$) (Table S2), while the downregulated genes in the adult *Insm1-thycKO* mice were enriched in T-cell activation involved in the immune response ($NES = -2.10$, $q = 0.014$), cell recognition ($NES = -2.09$, $q = 0.004$) and neuroepithelial cell differentiation ($NES = -2.00$, $q = 0.009$) (Table S3).
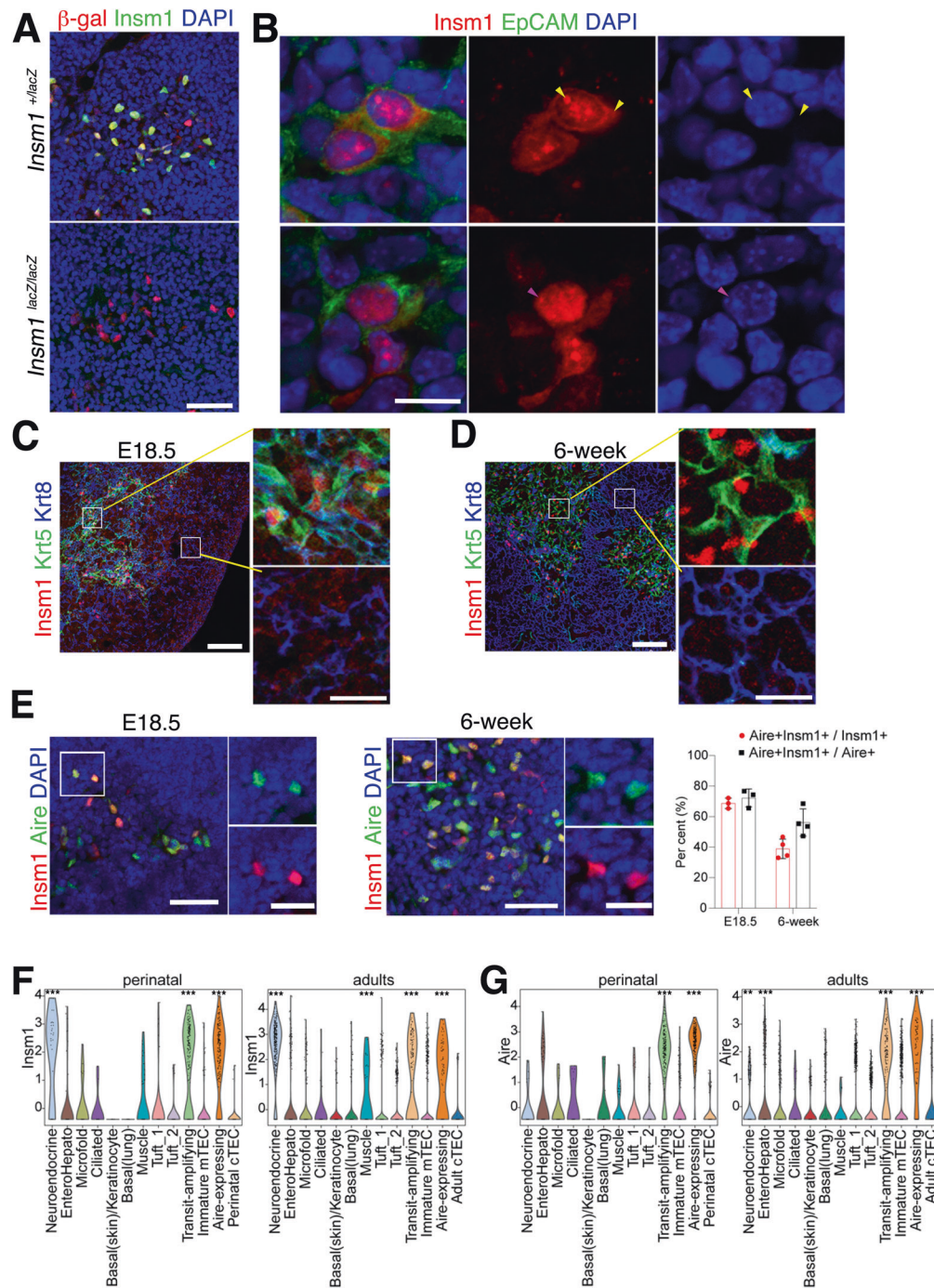
Fig. 1 *Insm1* is expressed in thymic epithelial cells. **A** Immunofluorescence analysis using antibodies against β-gal (red) and Insm1 (green); thymuses of E18.5 *Insm1*$^{+/lacZ}$ and *Insm1*$^{lacZ/lacZ}$ mice were analyzed. DAPI was used as a counterstain. Scale bar = 40 μm. **B** Immunofluorescence analysis of thymuses at E18.5 using antibodies against Insm1 (red), EpCAM (green) and DAPI. EpCAM is expressed in thymic epithelial cells and labels the cell membrane and cytoplasm. Yellow arrowheads indicate Insm1 located in both the cytoplasm and nucleus. The magenta arrowhead indicates a nucleus located in Insm1. Scale bar = 10 μm. Immunofluorescence analysis of thymuses of E18.5 (**C**) and 6-week-old (**D**) mice using antibodies against Insm1 (red), Krt5 (green) and Krt8. Magnifications are shown in the panels on the right. Scale bar shows 100 μm in the main and 20 μm in the magnified panels. **E** Immunofluorescence analysis detected Insm1 (green) and Aire (red) in the thymuses of E18.5 and 6-week-old mice. Scale bar shows 40 μm and 20 μm in the main and magnified panels, respectively. Quantification of single- and double-positive cells is shown on the right, and the data are shown as the mean ± SD (animal number $n = 3$ for E18.5 and $n = 4$ for adults; approximately 300 cells/animal/antibody were counted). **F** *Insm1* expression in post-Aire mTECs analyzed using published perinatal (left) and adult (right) scRNA-seq data [7]. **G** *Aire* expression in post-Aire mTECs analyzed using published perinatal (left) and adult (right) scRNA-seq data [7]. In (**F**) and (**G**), cells were sorted using two markers that were specifically downregulated in post-Aire mTECs, podoplanin (Pdpn) and integrin β4 (CD104), i.e., Pdpn$^-$CD104$^-$ cells from CD45$^-$EpCAM$^+$MHCII^{lo}Ly51$^-$ mTECs [7]. The isolated cells contained mostly post-Aire mTECs but also cells of immature mTECs, transit-amplifying mTECs (the direct progenitor of Aire-expressing mTECs), Aire-expressing mTECs, and cTECs. The significance of whether a subtype has a higher gene expression level than all other cells (one-tailed Wilcoxon test) is shown above each violin: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$
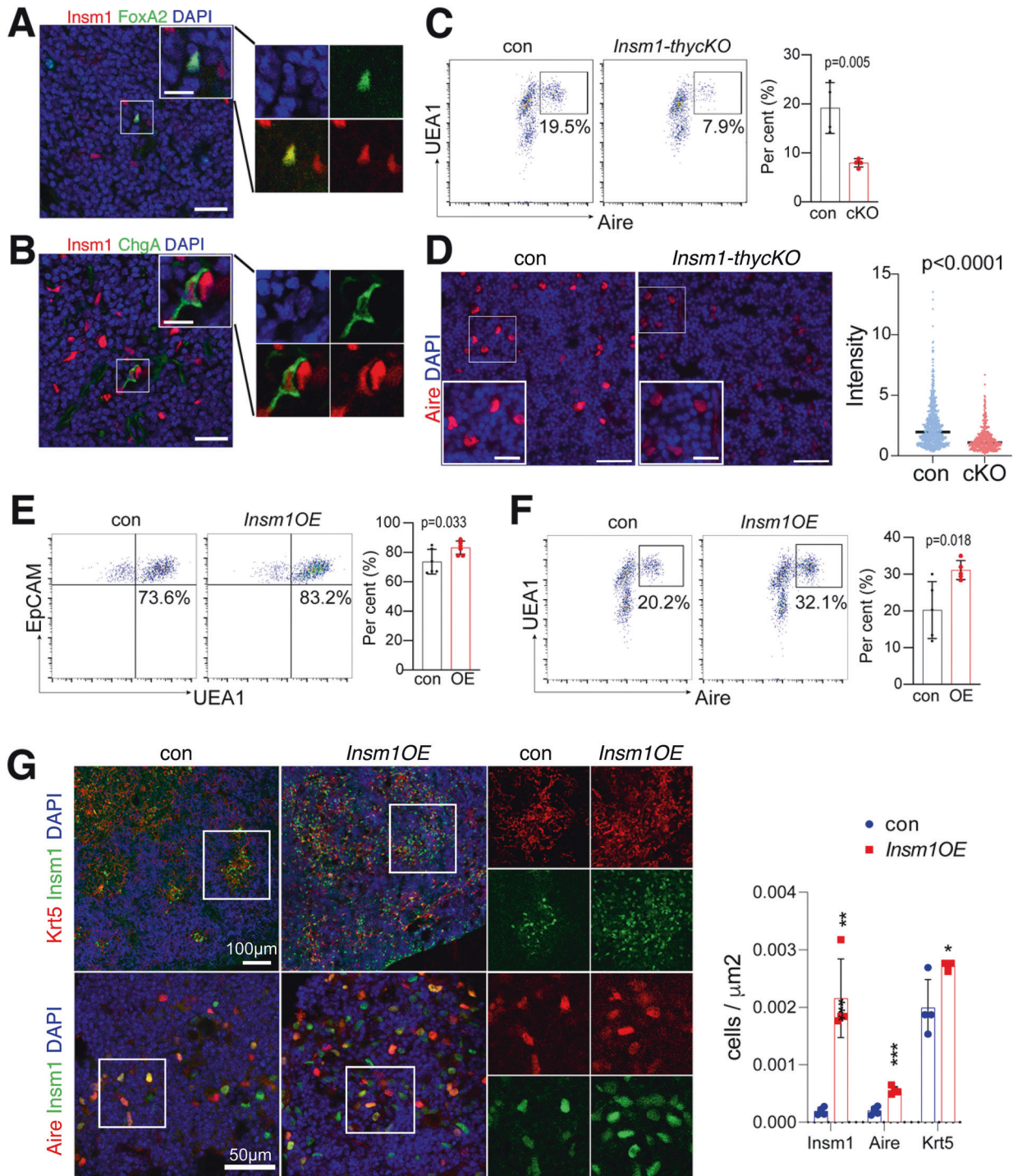
Fig. 2 Insm1 regulates Aire expression. Immunofluorescence analysis of thymuses of 6-week-old mice using antibodies against (**A**) Foxa2 (green) and Insm1 (red) and (**B**) ChgA (green) and Insm1 (red). Scale bars show 25 μm and 10 μm in the main and magnified panels, respectively. **C** Flow cytometry analysis of Aire+ cells in adult *Insm1-thycKO* and wild-type control animals (gate on CD45−/EpCAM+). Quantifications of the cell percentage are shown on the right (animal number $n = 4$ for each genotype). **D** Immunofluorescence analysis of the expression of Aire (red) in the thymuses of adult *Insm1-thycKO* and control mice. DAPI staining was used as a counterstain. Quantifications of the fluorescence intensity are shown on the right (animal number $n = 4$ for each genotype; 818 and 490 Aire+ cells were counted in control and mutant mice, respectively). Scale bar shows 50 μm and 20 μm in the main and magnified panels, respectively. Flow cytometry analysis of CD45−/EpCAM+/UEA1+ mTECs (**E**) and Aire+ cells (CD45−/EpCAM+/UEA1+/Aire+) (**F**) in thymuses of wild-type control and *Insm1OE* mice. Quantifications of the cell percentage are shown on the right (animal number $n = 6$ (**E**) and $n = 5$ (**F**) for each genotype). **G** Immunofluorescence analysis of the expression of Insm1 (green), Krt5 (red, upper panels) and Aire (red, lower panels) in *Insm1*-overexpressing (*Insm1OE*) thymuses of postnatal 2-day-old animals. DAPI was used as a counterstain. Quantifications of the cell number are shown on the right (animal number $n = 4$, 4 sections were used for each animal). Data are presented as scale dots with the median shown in (**D**) and as the means ± SDs in (**C**, **E**–**G**). Statistical significance was assessed by 2-tailed unpaired Student's *t* test, and *p* values are shown or defined as *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$
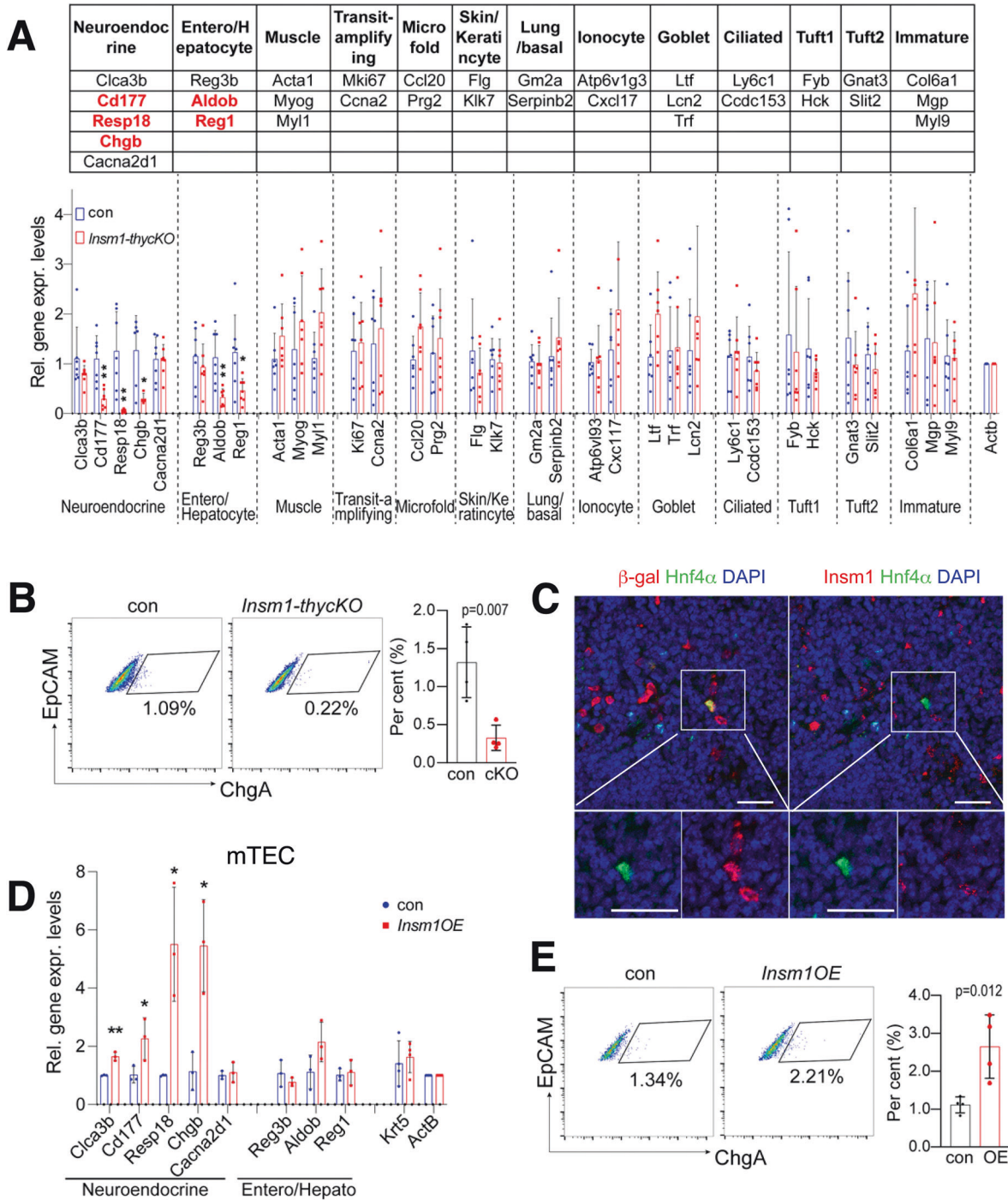
Fig. 3 Insm1 regulates neuroendocrine mimetic cell development. **A** Quantitative RT–PCR analysis of the expression of mimetic cell-specific genes in 7- to 8-week-old female and male control and *Insm1-thycKO* mice. The upper panel shows the genes specifically expressed in the mimetic cell types according to published data [7]. Red labels the downregulated genes in *Insm1* mutants. The lower panel shows the qRT–PCR results. Significantly downregulated genes are labeled red in the upper panel. Data in the figure panel are presented as the means ± SDs, and statistical significance was assessed by multiple unpaired *t* tests (animal number $n = 7$). Statistical significance was defined as *$p < 0.05$, **$p < 0.01$. **B** Flow cytometry analysis of ChgA+ neuroendocrine cells in 6- to 8-week-old *Insm1*$^{flox/flox}$ control and *Insm1-thycKO* mice. The cells were gated from CD45-/EpCAM+ cells. Quantification of the cell percentage is shown on the right, and statistical significance was identified by a 2-tailed unpaired Student's *t* test (animal number $n = 4$). **C** Immunofluorescence analysis of thymuses of 6-week-old *Insm1*$^{lacz/+}$ mice using antibodies against β-gal, Insm1 and Hnf4α. DAPI was used as a counterstain. The scale bar represents 25 μm in the main and magnified panels. Separated channels are shown for β-gal, Insm1 and Hnf4α in the same section. **D** The expression of neuroendocrine and enterocyte/hepatocyte marker genes in mTECs from postnatal 2-day *Insm1OE* and wild-type control mice ($n = 3$). Data are presented as the means ± SDs, statistical significance was determined by a 2-tailed unpaired Student's *t* test, and statistical significance was defined as *$p < 0.05$ and **$p < 0.01$. **E** Flow cytometry analysis of ChgA+ neuroendocrine cells in 6- to 8-week-old wild-type control and *Insm1OE* mice. The cells were gated from CD45-/EpCAM+ cells. Quantification of the cell percentage is shown on the right, and statistical significance was determined by a 2-tailed unpaired Student's *t* test (animal number $n = 4$)

However, the upregulated genes were involved in cell growth regulation, including ribosome processes and energy production (Tables S4, 5). These data indicated that Insm1 contributed to peripheral tissue type-specific gene expression in mTECs and the regulation of immune responses.

A test of differential gene expression by DESeq2 under cutoffs of FDR ≤ 0.1 and fold change ≥ 1.5 identified 63 and 97 dysregulated genes in fetal and adult mice, respectively, and *Aire* was among the downregulated genes in both datasets (Fig. 4A, B and Tables S6, S7). Considering that each TRA is typically expressed in only 1–5% of mTECs [30], we next used a lower cutoff ($p ≤ 0.05$ and fold change ≥ 1.5), which identified 332 and 716 dysregulated genes in fetal and adult *Insm1* mutants, respectively (Fig. 4A, B). Among these, 78 and 81 corresponded to downregulated TRA genes in fetal and adult Insm1 mutants, respectively (Fig. S4A, Fig. 4C and Tables S8, S9 and S10). Using qRT–PCR, we verified the altered expression of a subset of TRA genes (Fig. 4D and Fig. S4B). Among the 81 downregulated TRA genes identified in adult *Insm1-thycKO* mTECs, 26 were also affected by Aire mutations, and the remaining 55 were unique to *Insm1* mutants (Fig. 4C). Thus, the comparison with published *Aire* mutant datasets [31] showed that Aire-dependent and Aire-independent TRAs were downregulated (Fig. S4A and Fig. 4C). Fezf2 regulates a subset of Aire-independent TRAs [32], and we asked whether Insm1 and Fezf2 regulate a common set of TRAs by comparing Insm1-dependent and Fezf2-dependent TRAs. A relatively lower proportion of common TRAs was detected in the Insm1- and Fezf2-dependent TRAs (Fig. S4A and Fig. 4C).

Next, we assigned the downregulated TRAs to mimetic cell types using published scRNA-seq data [7]. This finding again demonstrated that the downregulated TRAs were not restricted to neuroendocrine mimetic cells. Instead, in both fetal and adult animals, downregulated TRAs were found in multiple mimetic cell types, e.g., neuroendocrine cells, enterocytes/hepatocytes, microfold cells, and basal (lung) cells, as well as in Aire-expressing mTECs (Fig. 4E and Table S11). These data indicate a broad regulatory role of Insm1 in mTECs and support the notion that Insm1 function is not restricted to neuroendocrine mimetic cells. We further assigned the downregulated TRA genes to tissues using public microarray data [33]. This finding indicated that Insm1-dependent TRAs were expressed in many peripheral tissues, for instance, in the salivary gland, pancreas, small intestine and large intestine (Fig. S4C).

We further analyzed the expression of TRA genes in *Insm1OE* mice. Approximately half of the tested TRA genes that were downregulated in *Insm1* mutant thymuses were significantly upregulated in mTECs or thymuses of *Insm1-OE* animals (Fig. 4F and Fig. S4D). Thus, loss of Insm1 reduces and overexpression promotes the expression level of TRA genes in the thymus.

To investigate whether the expression of TRAs in individual cells is altered, we analyzed the immunofluorescence intensity of Calca, Sst, Csn2 and ChgA in thymic sections. The immunofluorescence intensity of Calca in thymic cells was decreased in the *Insm1-thycKO* mice and increased in the *Insm1-OE* mice compared to the control mice (Fig. 4G, H). However, comparable levels of the immunofluorescence intensity of Sst, Csn2 and ChgA were detected in the *Insm1-thycKO* mice or in the *Insm1-OE* mice compared to the control mice (Fig. S5A, B). Thus, Insm1 regulates the expression of certain TRAs in thymic cells.

Insm1 binds to promoters and superenhancers in mTECs
We previously observed that Insm1 binds to the chromatin of endocrine cells [27]. To investigate the mechanism of Insm1 function in the thymus, we examined the chromatin binding of Insm1 using CUT&Tag analysis. We detected 5206 Insm1 binding sites (union of 3 repeats, $q < 1e-5$) in mTECs isolated from the fetal thymus (E18.5) and 1458 (union of 2 repeats, $q < 1e-5$) sites in mTECs isolated from adults (Fig. 5A, B). Furthermore, 1382 (94.7%)

of the sites identified in adults overlapped with those in fetuses. Thus, Insm1 binds to a largely overlapping set of genomic sites in adult and developing mTECs.

The majority of Insm1 binding sites were located in promoter regions in both fetal (50%) and adult (65.8%) datasets (Fig. 5A, B). However, a comparison of binding sites and dysregulated genes (cutoff $p ≤ 0.05$ and FC ≥ 1.5 in the RNA-seq data of *Insm1-thycKO*) showed that Insm1 bound to only a minority of the dysregulated genes at their promoter (Fig. 5C–E and Fig. S5C). Thus, Insm1 binding to promoters does not significantly impact gene expression.

Next, we compared the chromatin binding sites of Insm1 with previously identified binding sites of Aire [31, 34]. The majority of Insm1 binding sites (72–78%) were co-occupied by Aire (Fig. S5D). Aire was reported to bind to superenhancers [19], and in accordance, Insm1 binding sites were enriched on superenhancers (Fig. 5F–H). Furthermore, the number of dysregulated genes lying within 500 kb of Insm1-binding superenhancers was modestly overrepresented in the mTECs of fetal (Fig. 5I) but not adult animals (Fig. S5E). These results indicate that Insm1 and Aire cooperate in gene expression by binding to superenhancers in perinatal stages.

The promoter and distal regulatory sequences are essential for Aire expression [24, 35, 36]. In the CUT&Tag data, we identified that Insm1 bound to both sequences (Fig. 5J). We verified Insm1 binding by ChIP-PCR on the *Aire* promoter and distal regulatory sequences in wild-type mice, while binding was lost in *Insm1* mutants (Fig. 5K). We further examined the regulatory role of Insm1 binding to these sequences using a dual-luciferase assay. The presence of Insm1 significantly promoted the expression of the firefly luciferase gene, which was driven by the promoter, distal sequence or promoter-distal fusion sequences of *Aire* (Fig. 5L). Thus, Insm1 binds to the *Aire* promoter and the distal regulatory sequence and promotes *Aire* expression.

Autoimmune responses in nude mice transplanted with *Insm1* mutant thymuses
Next, we used thymus transplantation in nude mice to test whether the downregulation of TRA genes in the Insm1 mutant thymus results in autoimmune reactions. For this, T-cell-depleted thymuses that were isolated from fetal control (*Insm1$^{+/lacZ}$*) and *Insm1-null* mutant (*Insm1$^{lacZ/lacZ}$*) mice were transplanted into the renal capsule of six-week-old nude mice [37] (Fig. 6A, B). Eight weeks after transplantation, the size or structure of the transplanted thymuses was similar in the nude mice that received the thymuses from heterozygous (Het/nu) or mutant (Ko/nu) animals (Fig. 6C and Fig. S6), and the numbers of newly generated CD4+ and CD8+ cells were comparable in the thymuses detected by flow cytometry (Fig. 6D). However, spleens were larger and heavier with smaller spleen follicles in the *KO/nu* mice than in the *Het/nu* mice (Fig. 6E, F). The structure of the lymph nodes was unchanged (Fig. S6). We investigated lymphocyte infiltration of various organs using hematoxylin and eosin (H&E) staining, which detects the accumulation of lymphocyte nuclei. Increased infiltration was observed in pancreatic islets, lungs, kidneys and salivary glands of the *KO/nu* mice (Fig. 6G, H) but not in other investigated tissues (Fig. S6).

We further analyzed autoimmune reactions by the detection of autoimmune antibodies. Among the 13 investigated tissues, we observed autoimmune antibody reactions in pancreatic islets, kidney, and testis by staining these tissues with the serum obtained from the *KO/nu* mice (Fig. 6I and Fig. S7). In particular, the serum of the *KO/nu* mice showed subsets of pancreatic polypeptide (Pp)-positive cells, glucagon (Gcg)-positive alpha cells, and somatostatin (Sst)-positive delta cells, whereas insulin (Ins)-positive beta cells were rarely detected (Fig. 6J). Thus, autoimmune antibodies against PP-, alpha- and delta-cells were present in *KO/nu* mice. In summary, the transplantation of *Insm1*
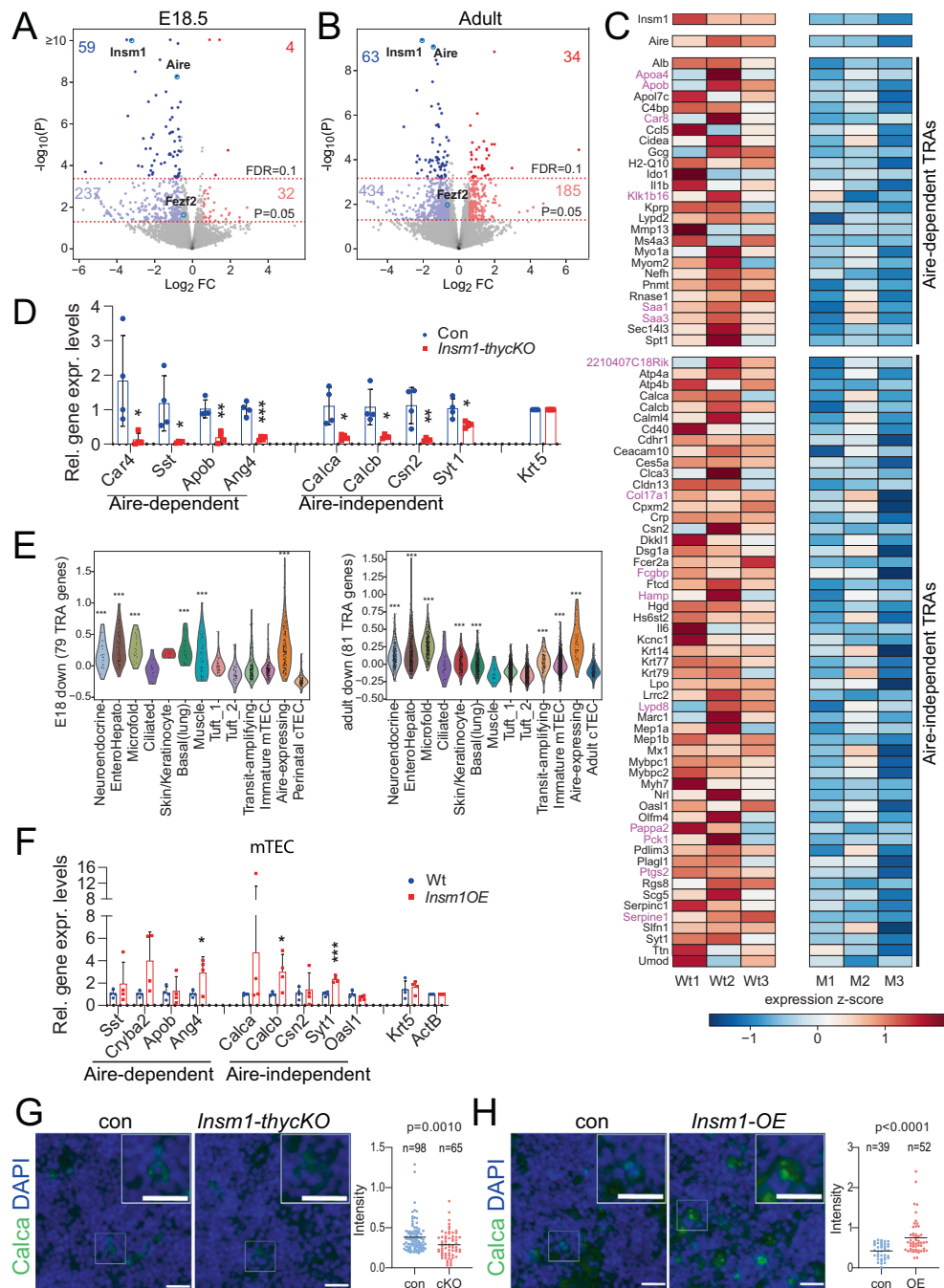
Fig. 4 Insm1 mutation results in decreased TRA levels. Volcano blots of RNA-seq data in *Insm1* mutants versus wild-type controls at the fetal (**A**) and adult stages (**B**). FDR ≤ 0.1 or *p* ≤ 0.05 combined with absolute fold change ≥ 1.5 was used for identifying the dysregulated genes. The numbers of dysregulated genes are shown in the figures accordingly. **C** Heatmap of dysregulated TRA genes detected in mTECs of *Insm1-thycKO* adult mice. Fezf2-dependent TRAs are labeled in magenta. **D** Quantitative RT–PCR analysis of the expression of TRA genes in mTECs of 6- to 8-week-old control and *Insm1-thycKO* mice (*n* = 4). Data in the figure panel are presented as the means ± SDs, and statistical significance was assessed by 2-tailed unpaired Student's *t* test. ns: *p* > 0.05; *p < 0.05, **p < 0.01, ***p < 0.001. **E** Expression distribution of the Insm1-dependent TRA genes (*p* ≤ 0.05, FC ≥ 1.5) in mimetic cells. Violin plots of the gene score for each cell of the post-Aire mTECs. Left panel, the dysregulated TRA genes in E18.5 *Insm1* mutant mice plotted on published perinatal scRNA-seq data; right panel, the dysregulated TRA genes in adult *Insm1* mutant mice plotted on published adult scRNA-seq data. The significance of whether a subtype has higher gene scores than all other cells (one-tailed Wilcoxon test) is shown above each violin: *p < 0.05, **p < 0.01, ***p < 0.001. **F** Analysis of TRAs expression in mTECs (animal number *n* = 4) of postnatal 2-day wild-type control and *Insm1OE* mice using qRT–PCR. Data are presented as the means ± SDs, and statistical significance was assessed by 2-tailed unpaired Student's *t* test. *p < 0.05, ***p < 0.001. Immunofluorescence analysis of Calca expression in thymic sections of 6-week-old control and *Insm1-thycKO* mice (**G**) or in 4-week-old control and *Insm1OE* mice (**H**). Statistical analysis of fluorescence intensity is shown on the right of the figure. Data are presented as scatter dot blots with the mean shown. Statistical significance was assessed by 2-tailed unpaired Student's *t* test, and *p* values are shown for each test. Three animals of each genotype were used, and at least 3 sections from each mouse were collected for immunostaining. The cell number counted for intensity analysis is shown above each dot blot. The fluorescence intensity was determined using ImageJ software
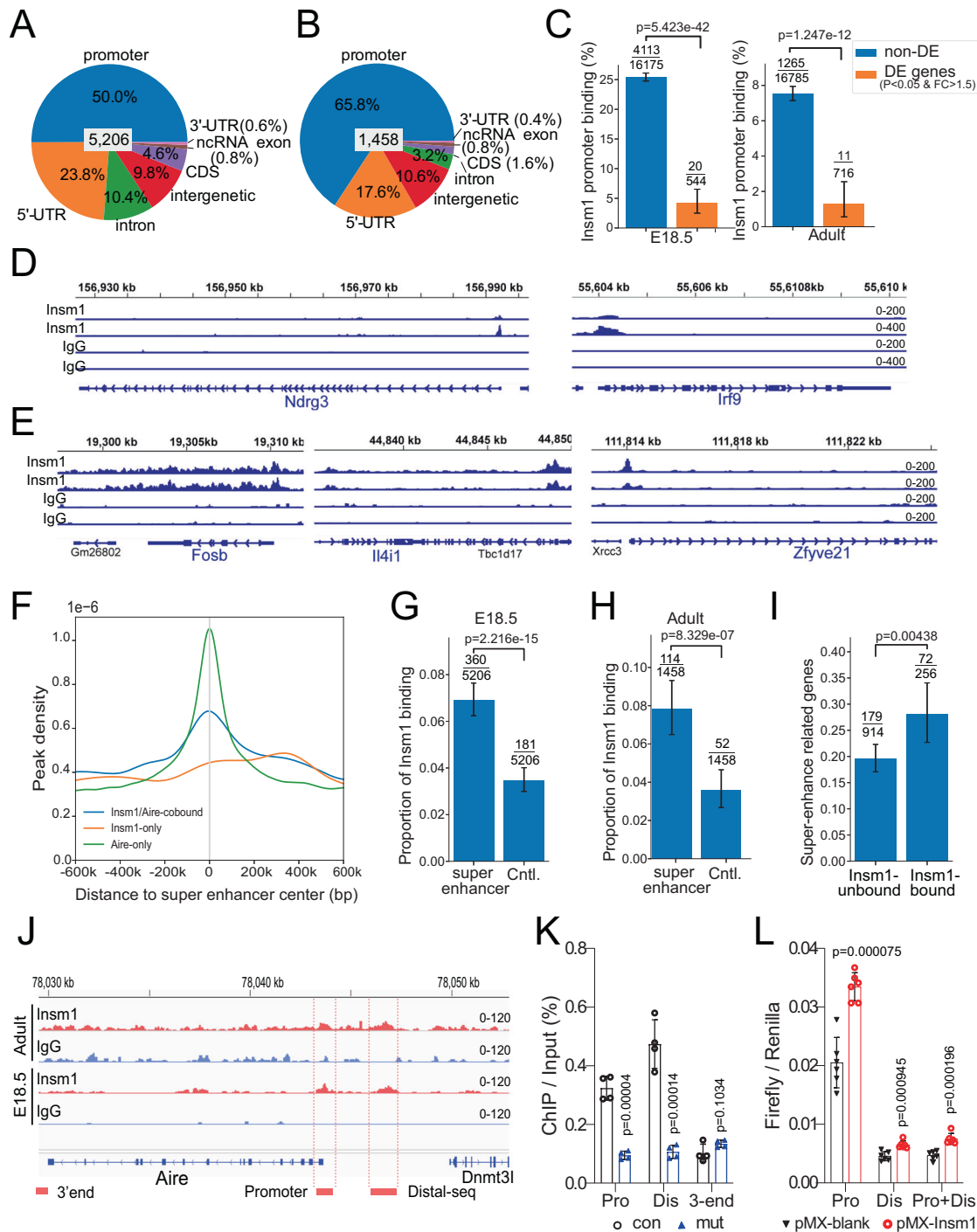
Fig. 5 Insm1 binds to superenhancer loci. Distribution of Insm1 binding sites on different genome regions in mTECs at fetal (**A**) and adult (**B**) stages. **C** Proportions of genes with Insm1 binding sites in the promoter regions (−2000 bp to +500 bp of TSS) at the fetal (left) and adult (right) stages. Insm1 binding sites are significantly depleted on the promoters of dysregulated genes. Insm1 binding traces on dysregulated genes at the fetal (**D**) and adult (**E**) stages. **F** Density curves of detected loci of Insm1 and Aire cobinding, Insm1-only binding (orange), and Aire-only binding (green), aligned to the center (0 position) of the superenhancers. Proportion of Insm1 binding peaks overlapping with superenhancer and control loci at the fetal (**G**) and adult (**H**) stages. An identical length sequence located 200 kb plus the length of the superenhancer away from each superenhancer was selected as the control sequence. **I** Proportions of superenhancers located within ±500 kb of genes that were dysregulated ($p \leq 0.05$, FC ≥ 1.5) in *Insm1* mutant mTECs at the fetal stage. Stratified by the superenhancers bound or unbound by Insm1. In (**C**), (**G**), (**H**) and (**I**), numerators and denominators of each proportion are given above the bars. The bars show 95% confidence intervals. The *p* values of Fisher's test are also given. **J** Insm1 binding traces at the Aire locus. Insm1 binding sites on promoter and distal enhancer sequences are indicated with blocks at the bottom of the panel. A non-Insm1 binding site at the 3′ end of Aire was also indicated with a block. **K** ChIP-PCR performed using mTECs isolated from wild-type and *Insm1-thycKO* adult mice. Animal number $n = 12$ for each genotype, mTECs from 3 animals were pooled for one ChIP. **L** Dual-luciferase reporter assays of the Aire promoter and Aire distal enhancer sequences. Experiment repeats $n = 6$. Data are presented as the means ± SDs in **K** and **L**, and statistical significance was assessed by a 2-tailed unpaired Student's *t* test. *P* values are shown in the figures
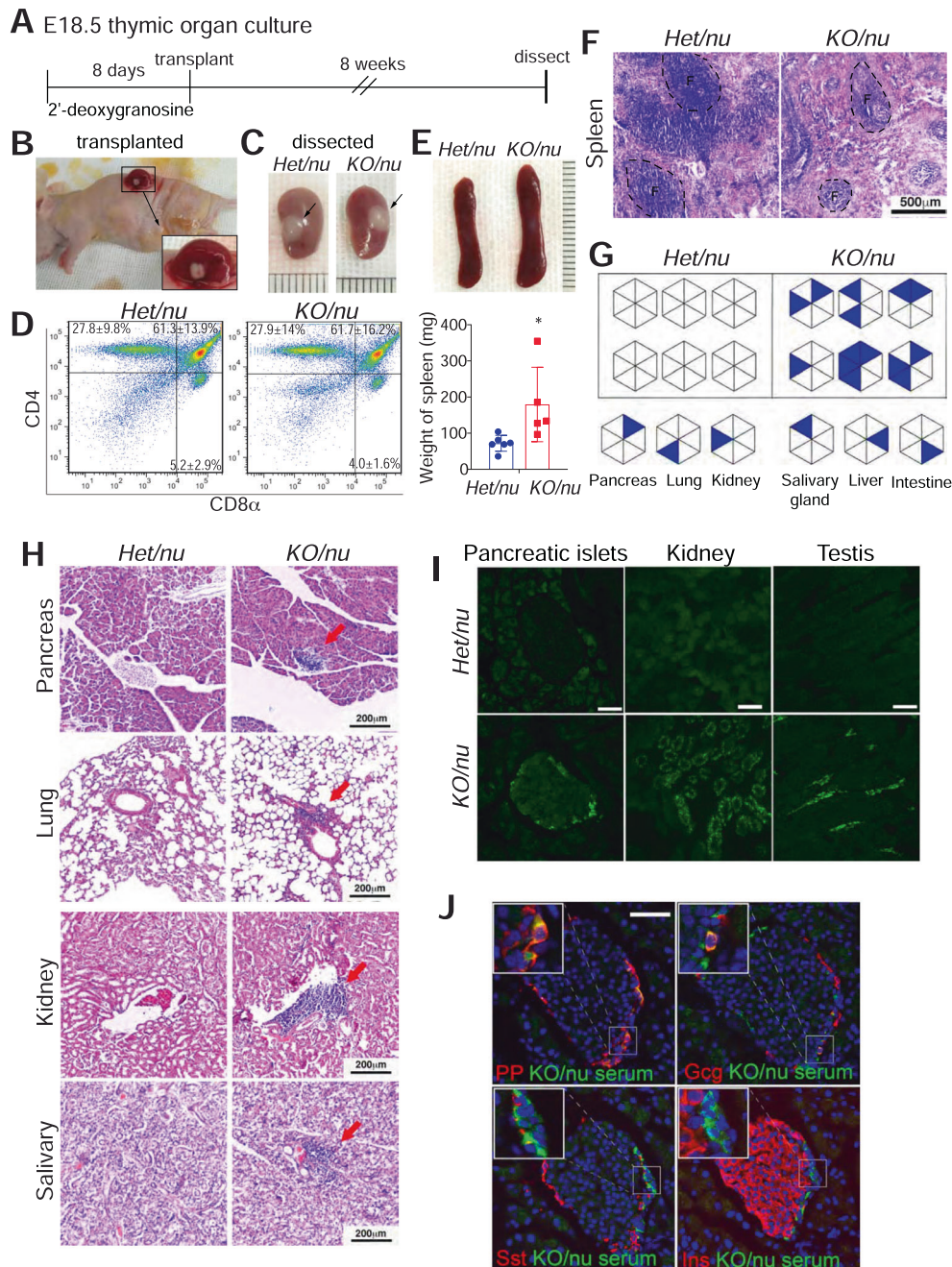
Fig. 6 Transplantation of *Insm1* mutant thymuses in nude mice. **A** Schematic outline for transplantation experiments. Thymuses were isolated, depleted of lymphocytes by 8-day culture in 2-deoxygranosine, and transplanted into nude mice, and the mice were analyzed 8 weeks after transplantation. **B** Transplantation of the thymus under the kidney capsule of the nude mouse. The enlarged picture shows the kidney with the transplanted thymus. **C** Isolated kidneys showing the thymuses under the kidney capsule 8 weeks after transplantation. *Het/nu*: the thymus was isolated from an *Insm1*$^{+/lacZ}$ mouse and transplanted into a nude mouse; *KO/nu*: the thymus was isolated from an *Insm1*$^{lacZ/lacZ}$ mouse and transplanted into a nude mouse. **D** CD4 and CD8α staining and flow cytometry analysis of thymocytes isolated from the transplanted thymuses of *Het/nu* and *KO/nu* mice. Data are presented as the means ± SDs, and significance was assessed by 2-tailed unpaired Student's *t* test, $p = 0.97$ for CD4 + CD8- cells, $p = 0.96$ for CD4 + CD8+ cells and $p = 0.34$ for CD8 + CD4− cells ($n = 3$). **E** Appearance (upper panel) and quantification of the weight (lower panel) of spleens isolated from *Het/nu* and *KO/nu* mice. Statistical data are presented as the means ± SDs, and significance was assessed by 2-tailed unpaired Student's *t* test. *$p < 0.05$ (animal number $n = 6$ for *Het/nu* and $n = 5$ for *KO/nu*). **F** H&E staining showing the structures of spleens isolated from *Het/nu* and *KO/nu* mice. **G** Summary of lymphocyte infiltration in multiple tissues of *Het/nu* and *KO/nu* mice. Upper panel, each hexagon represents a nude mouse transplanted with the indicated thymus, and the triangles indicate different organs. The blue triangle indicates the infiltrated organs (animal number $n = 6$). In the lower panel, each organ represented by the triangle in the hexagon is labeled with blue. **H** H&E staining of the pancreas, lung, kidney and salivary gland isolated from *Het/nu* and *KO/nu* mice. Three to four sections were used in H&E staining for each tissue of each animal. Red arrows indicate sites of lymphocyte infiltration. **I** Immunostaining using serum (green) isolated from *Het/nu* and *KO/nu* mice on sections that were prepared from *Rag1*$^{-/-}$ mice. Three to four sections collected from the nonserial sections of two *Rag1*$^{-/-}$ mice were used for the serum test for each tissue. Four of 6 *KO/nu* mice had autoimmune reaction antibodies. **J** Costaining of *KO/nu* serum (green) from *KO/nu* mice with pancreatic islet-specific hormones (red, as indicated in the panels), e.g., PP, Gcg, Sst and insulin
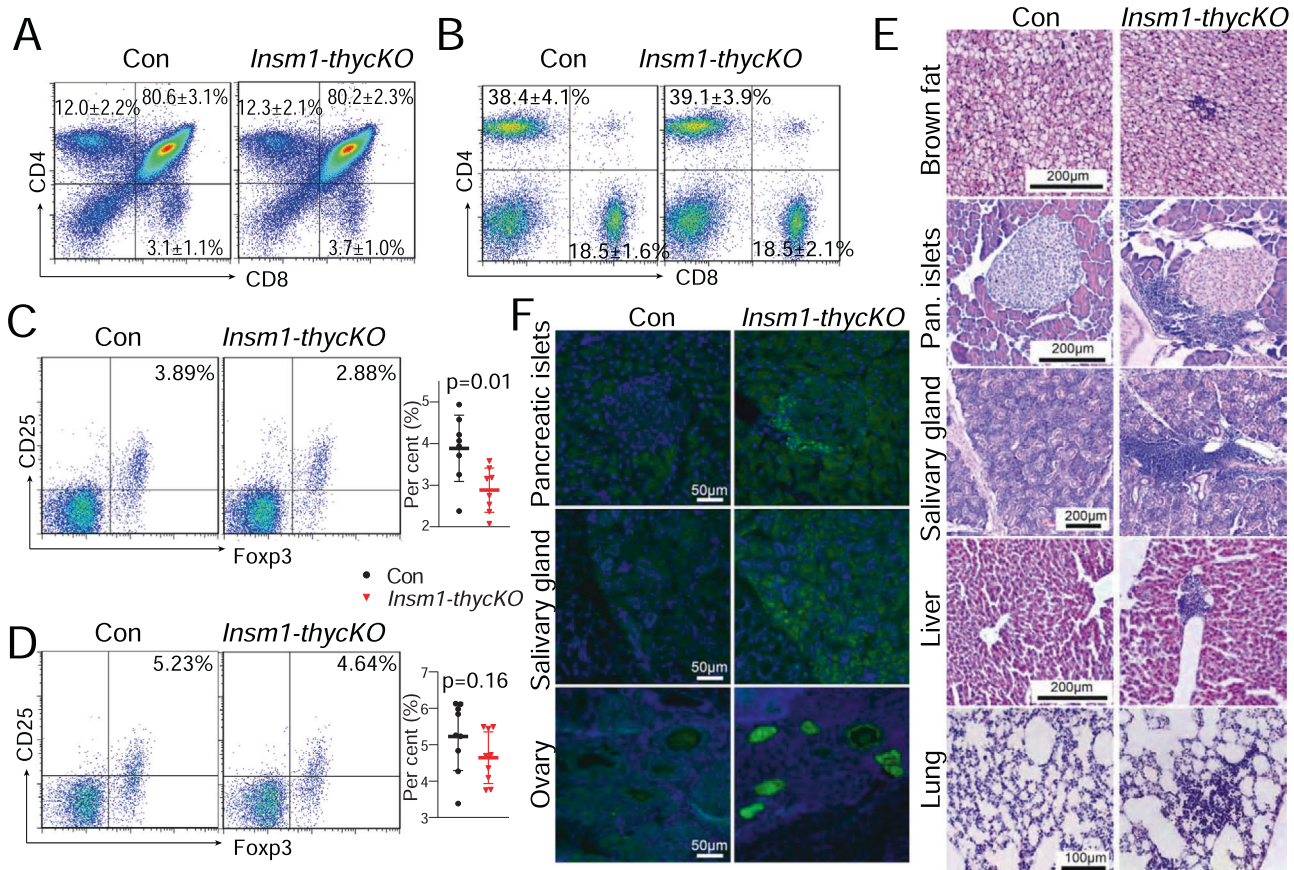
Fig. 7 Autoimmune phenotype in thymus-specific *Insm1* mutant mice. **A** CD4 and CD8α staining and flow cytometry analysis of thymocytes from thymuses of wild-type and *Insm1-thycKO* mice. The mean percentage and SD are shown in the figures, and significance was assessed by a 2-tailed unpaired Student's *t* test, $p > 0.05$ ($n = 10–12$). **B** CD4 and CD8α staining and flow cytometry analysis of thymocytes from the axillary lymph nodes of wild-type and *Insm1-thycKO* mice. The mean percentage and SD are shown in the figures, and significance was assessed by a 2-tailed unpaired Student's *t* test, $p > 0.05$ ($n = 9$). **C, D** Flow cytometry analysis of CD4$^+$/CD25$^+$/Foxp3$^+$ Treg cells from the thymus (**C**, $n = 8$) and the axillary lymph node (**D**, $n = 9$) of *Insm1-thycKO* and wild-type mice. Significance was assessed by 2-tailed unpaired Student's *t* test, and the *p* value is shown in the quantification chart. **E** H&E staining of brown fat, pancreas, salivary gland, liver and lung in *Insm1$^{flox/flox}$* control and *Insm1-thycKO* mice. Three to four sections were used in H&E staining for each tissue of each animal. The animal numbers are listed in Table 1. **F** Immunostaining using serum isolated from *Insm1$^{flox/flox}$* control and *Insm1-thycKO* mice on sections of pancreas, salivary gland and ovary prepared from *Rag1$^{-/-}$* mice. Three to four sections collected from the nonserial sections of two *Rag1$^{-/-}$* mice were used for the serum test for each tissue. The animals from which the serum was collected are listed in Table 2

mutant thymuses into nude mice results in autoimmune responses in multiple tissues.

Thymus-specific *Insm1* mutant mice show autoimmune responses

We next employed *Insm1-thycKO* and littermate *Insm1$^{flox/flox}$* control mice to further examine the autoimmune phenotype. *Insm1-thycKO* mice displayed no obvious abnormalities in adulthood, and comparable proportions of CD4$^+$ and CD8$^+$ T cells were generated in the thymus (Fig. 7A) and presented in lymph nodes (Fig. 7B). However, Treg cell (CD8$^-$CD4$^+$Foxp3$^+$CD25$^+$) numbers in the thymus were significantly decreased in *Insm1-thycKO* mice (Fig. 7C), and we observed a trend toward reduced numbers of Treg cells in lymph nodes that did not reach statistical significance (Fig. 7D).

Investigation of lymphocyte infiltration at various ages (6 weeks, 6 months, 12 months and 18 months) showed pronounced lymphocyte infiltration of multiple organs in a subset of the *Insm1-thycKO* animals, and the frequency of infiltration increased with age (Table 1). Infiltrated lymphocytes were observed in the pancreas, salivary gland, liver, lung and brown fat (Table 1 and Fig. 7E). Furthermore, we detected autoimmune antibodies against multiple tissues, including the pancreas, salivary gland,

and ovary, using the serum of *Insm1-thycKO* mice (Table 2 and Fig. 7F). Thus, in agreement with the expression of Insm1 and altered TRAs expression after *Insm1* ablation, loss of Insm1 function resulted in autoimmune responses in multiple peripheral organs in a subset of the mutant mice.

DISCUSSION

We show here that Insm1 functions as a pivotal factor for the acquisition of autoimmunity by controlling the levels of TRAs in the thymus. Insm1 is broadly expressed in transit-amplifying and Aire-expressing mTECs and thus in precursors of mimetic cells. As these mTECs mature, Insm1 expression is downregulated in most mimetic cell types but maintained in neuroendocrine mimetic cells. Nevertheless, loss of Insm1 function affects not only TRAs expressed by neuroendocrine mimetic cells but also other mimetic cell types. We believe this phenomenon is due to Insm1 in mTECs that express TRAs in the thymus.

Insm1 regulates the development of Aire-expressing mTECs and neuroendocrine mimetic cells

Both Aire-expressing mTECs and mimetic cells express TRAs and contribute to pools of self-antigen recognition for T-cell negative

Table 1. Lymphocyte infiltration in tissues

Age	17-month									6-month								6-week							
Genotypes	Insm1-thycKO				Ctl					Insm1-thycKO				Ctl				Insm1-thycKO				Ctl			
Mice	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Gender	F	F	F	M	M	F	F	M	M	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
Pancreas	+	-	-	+	++	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Salivary gland	++	+/−	-	+	++	+/−	+/−	+/−	+/−	+/−	+/−	+	+	nt	+/−	+/−	-	+	-	+	-	-	-	-	-
Liver	++	+/−	+/−	++	nt	+/−	+/−	nt	+/−	+/−	++	++	+/−	-	+/−	+/−	-	+	-	-	-	-	-	-	-
Kidney	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Spleen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Brain	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ovary	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Intestine	-	-	-	nt	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	nt	-	-	-
Colon	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Stomach	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Eye	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ganglia	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Heart	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	nt	-	nt	-	-	nt	-	-
Lung	++	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Skeletal muscle	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Brown fat	-	+	+/−	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Testis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Severe infiltration: ++; frequently observed infiltration: +; occasionally observed infiltration: +/−; negative: − *nt* not tested, *F* female, *M* male

Table 2. Autoimmune antibody reaction in tissues

Age	17-month									12-month						6-month										
Genotypes	Insm1-thycKO					Ctl				Insm1-thycKO			Ctl			Insm1-thycKO				Ctl				Insm1-thycKO		Ctl
Mice	1	2	3	4	5	6	7	8	9	26	27	28	29	30	31	10	11	12	13	14	15	16	17	32 33		34 35
Gender	F	F	F	M	M	F	F	M	M	F	F	M	F	F	M	F	F	F	F	F	F	F	F	M M		M M
Pancreas	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Salivary gland	-	-	-	++	+	-	-	-	-	+	-	-	-	-	-	+	-	-	+	-	-	-	-	- -		- -
Liver	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Kidney	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Brain	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Ovary	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	- -		- -
Intestine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Colon	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Stomach	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Eye	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Ganglia	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Nerve fiber	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Heart	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Lung	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Skeletal muscle	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Brown fat	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	- -		- -
Testis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	- -		- -

Severe positive: ++; positive: +; negative: − *F* female, *M* male

selection [4, 5, 7, 10]. Our study demonstrated that Insm1 is crucial for the proper generation of Aire-expressing mTECs and neuroendocrine mimetic cells, and a mutation in Insm1 resulted in impaired development of both cell types, leading to a decrease in the population of TRA-expressing cells within the thymus.

While most mimetic cell types originate from Aire-expressing cells, it is important to note that Aire expression is not strictly needed in all mimetic cells [9]. However, a mutation in *Aire* leads to reduced cell numbers and diminished TRA expression in certain mimetic cell types, such as FoxA+ and Sox+ cells [7, 11]. Recent studies have highlighted the importance of lineage-specific transcription factors, such as Spib and Sox8, in the development of particular mimetic cell subtypes, such as microfold cells [7]. Our research further revealed that Insm1 is essential for the development of neuroendocrine mimetic cells, implying that Insm1 serves as a neuroendocrine lineage factor. Notably, unlike other lineage-specific transcription factors that are restricted to specific mimetic cell types, Insm1 is broadly expressed in mTECs and plays a role in regulating the development of Aire-expressing mTECs. This finding suggests that Insm1 has multiple functions in both mTECs and neuroendocrine cells. Nevertheless, additional investigations are necessary to determine the lineage-specific role of Insm1 in regulating neuroendocrine mimetic cells.

Insm1 plays a regulatory role in the expression of Aire and certain TRAs
A broad function of Insm1 in the development of mTECs and mimetic cells is supported by the fact that *Insm1* regulates *Aire* expression. When *Insm1* is ablated, *Aire* expression is down-regulated, whereas overexpression of *Insm1* leads to broader *Aire* expression in mTECs. Aire-expressing cells give rise to various types of mimetic cells, and the absence of Aire affects the expression of TRAs in many types of peripheral tissues [5, 6, 13, 38, 39]. Interestingly, many of the DNA regions to which Insm1 binds are also occupied by Aire. This finding suggests that Insm1 not only regulates the expression of Aire but also collaborates with it to control gene expression in mTECs. Most of the DNA regions bound by Insm1 and Aire are enriched in superenhancers, which are, however, not associated with dysregulated TRAs, suggesting a role in cell fate regulation. Nevertheless, Insm1 can also independently regulate the expression of individual TRAs, such as Calca. Further research is needed to fully understand the role of Insm1 in cell fate determination and the regulation of TRA expression.

Mutation of *Insm1* results in downregulation of *Aire*. This downregulation might contribute to the observed changes in TRA expression in Insm1 mutants. However, the impact of Aire dosage on TRA regulation has been discussed controversially in the literature. Several studies have suggested that Aire dosage can control TRA expression [40, 41]. In particular, the expression of specific TRAs, such as Ins2, Tff3, Mup1, and Spt1, was strongly affected in $Aire^{+/-}$ mice [39]. In contrast, *Herzig* et al. proposed that a 50% decrease in Aire expression in $Aire^{+/-}$ mice does not affect Aire-dependent TRA expression [24]. Notably, Aire regulates its own expression, and Insm1 binds to both the promoter and distal regulatory sequences of the *Aire* gene, which are also occupied by Aire [34]. This phenomenon suggests potential cooperation between Insm1 and Aire at these sites to regulate *Aire* expression.

Insm1 is essential for the prevention of autoimmune reactions
Our study found that deleting Insm1 in the thymus resulted in autoimmune reactions in multiple tissues. We observed lymphocyte infiltration and the presence of autoimmune antibodies in various organs of both nude mice transplanted with *Insm1* mutant thymuses and mice with a specific *Insm1* mutation in the thymus. The affected tissues varied and included the pancreatic islets, salivary glands, lung, kidney, liver, and brown fat. Pancreatic islets

and the salivary gland were frequently affected in both transplanted nude mice and thymus-specific *Insm1* mutants. Only occasionally were there autoimmune reactions in other tissues. The various tissues that are attacked by autoimmune reactions were previously reported in Aire mutants [4, 5]. However, autoimmune reactions in pancreatic islets and brown adipose tissue have not been reported in *Aire* mutant mice [4, 5, 14, 15, 20, 42]. A broad set of tissues displaying autoimmune responses reflects our findings on changes in TRA gene expression in *Insm1* mutants. Thus, our data demonstrate that the establishment of central tolerance critically depends on Insm1 function in the thymus and that a loss of Insm1 causes autoimmune reactions against various cell types and tissues.

Additionally, we discovered that regulatory T cells (Treg cells) were decreased in the Insm1 mutant thymus, which was not observed in *Aire* mutants [4, 5] but has been observed in *Fezf2* mutants [32]. This result suggests that Insm1 plays an unknown role in regulating Treg cell development in the thymus, along with its role in regulating TRA-expressing cells. Insm1 mutation might eliminate certain TRAs that are crucial for Treg cell development. Alternatively, Insm1 might regulate the paracrine function in neuroendocrine mimetic cells and affect the development of Treg cells. Further research is needed to fully understand the mechanisms by which Insm1 regulates Treg cells.

MATERIALS AND METHODS
Animals
$Insm1^{lacZ/lacZ}$ mice were described previously [25]. E18.5 embryos were collected, genotyped, and analyzed. Wild-type littermates were used as controls in most experiments, while $Insm1^{lacZ/+}$ animals were used as controls in experiments that relied on monitoring β-gal expression. A thymus-specific mutation of *Insm1* was introduced by crossing $Insm1^{flox/flox}$ [27] and $Foxn1Cre$ mice (JAX No. 018448). Tissues collected from 8-week-old $Rag1^{-/-}$ mice (GemPharmatech No. T004753) were used for the autoimmune antibody test. The $Insm1KI$ mouse strain was generated by introducing Insm1 coding sequences into the $Rosa26$ locus. The Insm1 coding sequences are preceded in the 5′ region by a stop cassette that can be removed by Cre (Cyagen, Suzhou, China) (Fig. S3A). Both male and female mice were used in the work. All animal experiments were approved by the institutional animal care and use committee of Jinan University (IACUC-20211123-02).

Thymus transplantation
Thymuses were isolated from E18.5 $Insm1^{+/lacZ}$ and $Insm1^{lacZ/lacZ}$ mice (CD1 background), cultured for 8 days in RPMI-1640 containing 10% FBS and 1.25 mM 2′-deoxyguanosine (to deplete donor lymphocytes), and transplanted as described previously [5, 37]. In brief, one of the lymphocyte-depleted donor thymuses was transplanted into the renal capsule of a nude mouse (6 weeks old, CD1 background). Eight weeks after transplantation, the thymus and newly generated lymphocytes were collected, and various tissues were also collected and analyzed.

Thymic cell dissociation and mTEC isolation
The fetal (E18.5) and adult (6–8 weeks) thymic cells were dissociated using Liberase [43]. In brief, the thymuses were separated, and connective tissues and fat were removed. For E18.5 animals, 2–3 thymuses with the same genotype were pooled, and thymic lobes were cut into small pieces in RPMI to release lymphocytes. Lymphocytes were collected in those cases when the entire thymic cell population was analyzed. Otherwise, the thymuses were digested in 100 μl of RPMI 1640 containing 0.1 mg/ml Liberase (Roche, Germany) and 20 U/ml DNase I at 37 °C for 4 min. The digestion procedure was repeated twice, and cells were collected after each round of digestion.

mTEC isolation was performed using a magnetic bead-based method as previously described [44]. In brief, 1 μl of biotinylated UEA1 (Vector Laboratories, B-1065-2) was incubated with 25 μl of Biotin Binder Dynabeads (Thermo Fisher Scientific, 11047) for 30 min at room temperature. Unbound UEA1 was removed by washing in 1 ml of FACS buffer (PBS with 0.02% BSA and 5 mM EDTA). Thymuses from 2–3 E18.5 fetal mice or one adult mouse were incubated with 25 μl of UEA1-coated beads at 4 °C for 30 min. Unbound cells were removed by three washes of

FACS buffer and by additional incubation with CD45 S-pluriBeads (pluriSelect Life Science, SKU#70-50010-11) before mTECs were collected for analysis. The isolated mTECs were used for RNA-seq, Cut&Tag and qRT–PCR analysis.

Flow cytometry analysis

For analysis of lymphocytes in the transplanted thymuses by flow cytometric analysis, thymic lobes were gently cut, and released lymphocytes were collected and directly stained with CD4 and CD8 antibodies in FACS buffer. For analysis of thymic cells, stroma cells were dissociated using Liberase as described above. Staining with antibodies recognizing cell surface proteins was performed directly in FACS buffer, whereas staining with antibodies recognizing nuclear antigens was performed after fixation in Foxp3/Transcription Factor Staining Buffer (Invitrogen, 00-5523-00). Immunostaining procedures were performed as described [45]. The antibodies used are listed in Supplementary Table 12.

BD FACSCanto II or FACSAria II was used to analyze and collect cells (Beckton Dickenson, Franklin Lakes, USA). Cell quantifications were performed using FloJo 7.6.5 software.

Immunohistochemistry, western blot analysis and hematoxylin and eosin (H&E) staining

Immunohistochemistry and western blot analysis were performed as previously described [27]. Fluorescence was imaged on a Zeiss LSM 700 confocal or a Leica DMi8 microscope, and the images were processed using Adobe Photoshop software. The antibodies used are listed in Supplementary Table 12. For the autoimmune antibody tests, serum of the KO/nu or Insm1-thycKO mice was used at a concentration of 1:25 on tissue sections obtained from $Rag1^{-/-}$ mice.

The Lillie-Mayer method was used for H&E staining. Five percent aluminum ammonium sulfate and 0.5% hematoxylin were used to stain the nuclei, followed by incubation with 0.3% acid alcohol for the differentiation of nuclear staining. Acetified eosin was used for counterstaining to reveal cellular details.

RNA-seq analysis

For E18.5 mice, six to nine thymuses of each genotype (Insm1 mutant and wild type) were collected, and mTECs were isolated after pooling 2–3 thymuses of the same genotype. For adult mice (6–8 weeks old), mTECs were isolated from one thymus, and 3 animals were used for each genotype. Total RNA was isolated using TRIzol reagents. Three independent sequencing libraries for each genotype were generated using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, USA). An Illumina NovaSeq 6000 was used for sequencing, and 150-nt paired-end reads were generated. Sequencing data (6 Gb or more) with more than 94% of bases scoring above Q30 (accuracy rate 99.9%) were produced from each library sample. RNA-seq fastq files were aligned to the mouse genome (mm10) using STAR (v2.5.2a) [46] with the parameter --outFilterMismatchNmax 2. The proportions of uniquely mapped reads were 79.4–85.9%. Gene read numbers were counted using HTSeq (v0.11.0) [47] based on mouse gene annotation Ensembl v79, and the value of fragments per kilobase of exon per million reads mapped (FPKM) was calculated. Next, we used DESeq2 (v1.34.0) [48] to detect dysregulated genes under cutoffs of FDR ≤ 0.1 and fold change ≥ 1.5. For analysis of downregulated TRAs and dysregulating-trend genes for downstream analysis, a looser cutoff ($p ≤ 0.05$ and fold change ≥ 1.5) was applied. All RNA-seq data are accessible on the Gene Expression Omnibus (GSE193929).

For comparison of genes dysregulated after Aire mutation, published data were used [31]. Gene read counts were downloaded from GEO (ID: GSE144877). Differentially expressed genes were analyzed according to the method described previously [31].

GO enrichment analysis was performed by the R package topGO (v2.38.1) [49] with the "classic" algorithm and "Fisher exact test", followed by Benjamini–Hochberg multiple test correction using FDR = 0.05 as a cutoff.

qRT–PCR analysis

Cells were lysed, and total RNA was isolated using TRIzol reagent (Invitrogen). For qRT–PCR analysis, RNA was isolated from thymuses or mTECs of E18.5 fetal mice. cDNA was synthetized using HiScript II Q RT SuperMix for qPCR (+gDNA wiper) (R223-01, Vazyme, China) and analyzed using SYBR Green I-based real-time quantitative PCR on a CFX96 RT–PCR system (Bio-Rad). Expression levels were determined using the $2^{-\Delta\Delta Ct}$

method and ActB or Krt5 as internal standards and are displayed as a proportion of the control. The primers used for quantitative analysis are listed in Supplementary Table 13.

Insm1 CUT&Tag analysis

CUT&Tag (Cleavage Under Targets and Tagmentation) analysis was performed according to the manufacturer's instructions (Novoprotein, N259-YH01) and previously published protocols [50]. In brief, mTECs were isolated from wild-type mice using UEA1-coupled Dynabeads. Then, 0.5 µl of Insm1 antibody or 1 µg of IgG antibody was incubated overnight at 4 °C with 5×10^4 mTECs in a volume of 100 µl of primary antibody buffer. A donkey anti-guinea pig (Jackson ImmunoResearch, 706-005-148) secondary antibody was used and incubated with mTECs for 45 min in 100 µl of secondary antibody buffer at room temperature. After washes to remove unbound antibodies, the ChiTag transposon was incubated with mTECs for 1 h at room temperature. Fragmentation, DNA purification and library construction were performed according to the manufacturer's protocol. At least three biological replicates were obtained.

An Illumina NovaSeq 6000 was used for sequencing, and 150-nt paired-end reads were generated. At least twelve Gigabyte pass-filtering sequencing data were generated. Raw CUT&Tag sequencing data were trimmed to remove adaptor sequences using fastp (v0.12.0) [51] and aligned to the mouse genome (mm10) using Bowtie2 (v2.2.9) [52] with the parameters "--local --very-sensitive-local --no-unal --no-mixed --no-discordant --phred33 -X 700". Peaks were called using MACS2 (v2.2.6) [53] with the parameter "--keep-dup auto". IgG data were used as a control. Insm1-binding peaks identified in three replicates with $q < 10e-5$ were merged for subsequent analysis.

For calculation of the proportions of Insm1 binding on different genomic regions, peaks were assigned to different genomic regions based on their summit positions on the Ensembl transcriptome annotation according to the priority order of CDS, 5'-UTR, 3'-UTR, promoter, ncRNA, intron, and intergenic, where the promoters were defined as the regions of −2000 bp to +500 bp of transcription start site (TSS). The CUT&Tag sequencing data are accessible on Gene Expression Omnibus (GSE193929).

ChIP-PCR

For ChIP-PCR, we used mTECs isolated from adult mice. mTECs (1×10^6) were used for each ChIP experiment. Chromatin precipitation was performed based on protein A beads (Invitrogen) for Insm1/IgG pulldown. The procedure was performed as described [27]. The primers used are listed in Supplementary Table 13.

Dual-luciferase reporter assay

The promoter, distal regulatory sequence and fusion sequences of the promoter and distal regulatory sequence were cloned and inserted into the pGL4.23 minimal promoter vector. A dual-luciferase reporter assay was performed using 293 T cells in 96-well plates with cotransfection of the internal control pRL-TK-Rinella (10 ng/well) with the cloned Pgl4.23-firefly plasmid (100 ng/well). pMX-blank (90 ng/well) or pMX-Insm1 (90 ng/well) plasmid was cotransfected with luciferase plasmids. Lipofectamine 3000 was used for transfection. Data are presented as the ratio of firefly versus Renilla.

Single-cell RNA sequencing data analysis

A preprocessed transcript-by-cell matrix of scRNA-seq data of Pdpn⁻CD104⁻mTEC^{lo} from perinatal and adult mice was downloaded from GEO (ID: GSM5831744) and read by scanpy [54] for downstream analysis. First, the valid 8236 mTECs used in the study [7] were selected according to the barcode list provided. The gene expression matrix was normalized to counts per 10,000 counts, followed by log-transfer and scaling to unit variance and zero mean. Next, the top 2000 highly variable genes were identified and selected, and PCA was performed for dimensionality reduction. According to the elbow point of the PC contribution curve, the top 30 PCs were used for neighborhood graph construction. Cell clustering was performed using the Leiden method (resolution = 1.9). The scores of gene sets were calculated using the scanpy function tl.score_genes. For visualization consistency with the published paper [7], we used the original coordinates of UMAP from the authors.

TRA gene analysis

The list of TRA genes were identified in an earlier study [38] and provided by the authors (Prof. Perreault C, Université de Montréal, personal

communication). Aire-dependent TRAs were identified by the overlaps of downregulated genes from the public Aire mutant RNA-seq data [31]. The TRA expression profiles of various mouse tissues were generated in previous microarray studies [33] and downloaded from GEO (GSE10246). For visualization of TRA expression in different tissues (Fig. S4C), samples of each tissue were merged by mean expression values, followed by *z* score normalization on each gene.

Superenhancer analysis

Superenhancer regions were identified in the published study [19] and were provided to the author via email (Prof. Diane Mathis, Harvard Medical School, personal communication). The control regions used in Fig. 5G, H were the sequences with the same length as the super enhancer and with an offset of the distance of the super enhancer length plus 200 kb.

REFERENCES

1. Takahama Y. Journey through the thymus: stromal guides for T-cell development and selection. Nat Rev Immunol. 2006;6:127–35.
2. Alves NL, Takahama Y, Ohigashi I, Ribeiro AR, Baik S, Anderson G, et al. Serial progression of cortical and medullary thymic epithelial microenvironments. Eur J Immunol. 2014;44:16–22.
3. Kyewski B, Derbinski J. Self-representation in the thymus: an extended view. Nat Rev Immunol. 2004;4:688–98.
4. Ramsey C, Winqvist O, Puhakka L, Halonen M, Moro A, Kampe O, et al. Aire deficient mice develop multiple features of APECED phenotype and show altered immune response. Hum Mol Genet. 2002;11:397–409.
5. Anderson MS, Venanzi ES, Klein L, Chen Z, Berzins SP, Turley SJ, et al. Projection of an immunological self shadow within the thymus by the aire protein. Science. 2002;298:1395–401.
6. Wells KL, Miller CN, Gschwind AR, Wei W, Phipps JD, Anderson MS, et al. Combined transient ablation and single-cell RNA-sequencing reveals the development of medullary thymic epithelial cells. Elife. 2020;9:e60188.
7. Michelson DA, Hase K, Kaisho T, Benoist C, Mathis D. Thymic epithelial cells co-opt lineage-defining transcription factors to eliminate autoreactive T cells. Cell. 2022;185:2542–58 e18.
8. Dhalla F, Baran-Gale J, Maio S, Chappell L, Hollander GA, Ponting CP. Biologically indeterminate yet ordered promiscuous gene expression in single medullary thymic epithelial cells. EMBO J. 2020;39:e101828.
9. Miller CN, Proekt I, von Moltke J, Wells KL, Rajpurkar AR, Wang H, et al. Thymic tuft cells promote an IL-4- enriched medulla and shape thymocyte development. Nature. 2018;559:627–31.
10. Bornstein C, Nevo S, Giladi A, Kadouri N, Pouzolles M, Gerbe F, et al. Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. Nature. 2018;559:622–6.
11. Michelson DA, Mathis D. Thymic mimetic cells: tolerogenic masqueraders. Trends Immunol. 2022;43:782–91.
12. Chuprin A, Avin A, Goldfarb Y, Herzig Y, Levi B, Jacob A, et al. The deacetylase Sirt1 is an essential regulator of Aire-mediated induction of central immunological tolerance. Nat Immunol. 2015;16:737–45.
13. Matsumoto M, Nishikawa Y, Nishijima H, Morimoto J, Matsumoto M, Mouri Y. Which model better fits the role of aire in the establishment of self-tolerance: the transcription model or the maturation model? Front Immunol. 2013;4:210.
14. Giraud M, Yoshida H, Abramson J, Rahl PB, Young RA, Mathis D, et al. Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. Proc Natl Acad Sci USA. 2012;109:535–40.
15. Danso-Abeam D, Humblet-Baron S, Dooley J, Liston A. Models of aire-dependent gene regulation for thymic negative selection. Front Immunol. 2011;2:14.
16. Abramson J, Giraud M, Benoist C, Mathis D. Aire's partners in the molecular control of immunological tolerance. Cell. 2010;140:123–35.
17. Anderson AO, Shaw S. Conduit for privileged communications in the lymph node. Immunity. 2005;22:3–5.
18. Koh AS, Miller EL, Buenrostro JD, Moskowitz DM, Wang J, Greenleaf WJ, et al. Rapid chromatin repression by Aire provides precise control of immune tolerance. Nat Immunol. 2018;19:162–72.
19. Bansal K, Yoshida H, Benoist C, Mathis D. The transcriptional regulator Aire binds to and activates superenhancers. Nat Immunol. 2017;18:263–73.
20. Guerau-de-Arellano M, Martinic M, Benoist C, Mathis D. Neonatal tolerance revisited: a perinatal window for Aire control of autoimmunity. J Exp Med. 2009;206:1245–52.
21. Gabler J, Arnold J, Kyewski B. Promiscuous gene expression and the developmental dynamics of medullary thymic epithelial cells. Eur J Immunol. 2007;37:3363–72.
22. Yang S, Fujikado N, Kolodin D, Benoist C, Mathis D. Immune tolerance. Regulatory T cells generated early in life play a distinct role in maintaining self-tolerance. Science. 2015;348:589–94.
23. Baran-Gale J, Morgan MD, Maio S, Dhalla F, Calvo-Asensio I, Deadman ME, et al. Ageing compromises mouse thymus function and remodels epithelial cell differentiation. Elife. 2020;9:e56221.
24. Herzig Y, Nevo S, Bornstein C, Brezis MR, Ben-Hur S, Shkedy A, et al. Transcriptional programs that control expression of the autoimmune regulator gene Aire. Nat Immunol. 2017;18:161–72.
25. Gierl MS, Karoulias N, Wende H, Strehle M, Birchmeier C. The zinc-finger factor Insm1 (IA-1) is essential for the development of pancreatic beta cells and intestinal endocrine cells. Genes Dev. 2006;20:2465–78.
26. Jia S, Wildner H, Birchmeier C. Insm1 controls the differentiation of pulmonary neuroendocrine cells by repressing Hes1. Dev Biol. 2015;408:90–8.
27. Jia S, Ivanov A, Blasevic D, Muller T, Purfurst B, Sun W, et al. Insm1 cooperates with Neurod1 and Foxa2 to maintain mature pancreatic beta-cell function. EMBO J. 2015;34:1417–33.
28. Welcker JE, Hernandez-Miranda LR, Paul FE, Jia S, Ivanov A, Selbach M, et al. Insm1 controls development of pituitary endocrine cells and requires a SNAG domain for function and for recruitment of histone-modifying factors. Development. 2013;140:4947–58.
29. Anderson G, Baik S, Cowan JE, Holland AM, McCarthy NI, Nakamura K, et al. Mechanisms of thymus medulla development and function. Curr Top Microbiol Immunol. 2014;373:19–47.
30. Derbinski J, Schulte A, Kyewski B, Klein L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. Nat Immunol. 2001;2:1032–9.
31. Tomofuji Y, Takaba H, Suzuki HI, Benlaribi R, Martinez CDP, Abe Y, et al. Chd4 choreographs self-antigen expression for central immune tolerance. Nat Immunol. 2020;21:892–901.
32. Takaba H, Morishita Y, Tomofuji Y, Danks L, Nitta T, Komatsu N, et al. Fezf2 Orchestrates a Thymic Program of Self-Antigen Expression for Immune Tolerance. Cell. 2015;163:975–87.
33. Lattin JE, Schroder K, Su AI, Walker JR, Zhang J, Wiltshire T, et al. Expression analysis of G Protein-Coupled Receptors in mouse macrophages. Immunome Res. 2008;4:5.
34. Goldfarb Y, Givony T, Kadouri N, Dobes J, Peligero-Cruz C, Zalayat I, et al. Mechanistic dissection of dominant AIRE mutations in mouse models reveals AIRE autoregulation. J Exp Med. 2021;218:e20201076.
35. LaFlam TN, Seumois G, Miller CN, Lwin W, Fasano KJ, Waterfield M, et al. Identification of a novel cisregulatory element essential for immune tolerance. J Exp Med. 2015;212:1993–2002.
36. Murumagi A, Vahamurto P, Peterson P. Characterization of regulatory elements and methylation pattern of the autoimmune regulator (AIRE) promoter. J Biol Chem. 2003;278:19784–90.
37. Wang J, Chen G, Cui Q, Song E, Tao W, Chen W, et al. Renal Subcapsular Transplantation of 2'- Deoxyguanosine-Treated Murine Embryonic Thymus in Nude Mice. J Vis Exp. 2019:e59657, https://doi.org/10.3791/59657.
38. St-Pierre C, Trofimov A, Brochu S, Lemieux S, Perreault C. Differential Features of AIRE-Induced and AIREIndependent Promiscuous Gene Expression in Thymic Epithelial Cells. J Immunol. 2015;195:498–506.
39. Kont V, Laan M, Kisand K, Merits A, Scott HS, Peterson P. Modulation of Aire regulates the expression of tissue-restricted antigens. Mol Immunol. 2008;45:25–33.
40. Fornari TA, Donate PB, Macedo C, Marques MM, Magalhaes DA, Passos GA. Age-related deregulation of Aire and peripheral tissue antigen genes in the thymic stroma of non-obese diabetic (NOD) mice is associated with autoimmune type 1 diabetes mellitus (DM-1). Mol Cell Biochem. 2010;342:21–8.
41. Oliveira EH, Macedo C, Collares CV, Freitas AC, Donate PB, Sakamoto-Hojo ET, et al. Aire Downregulation Is Associated with Changes in the Posttranscriptional Control of Peripheral Tissue Antigens in Medullary Thymic Epithelial Cells. Front Immunol. 2016;7:526.
42. Oven I, Brdickova N, Kohoutek J, Vaupotic T, Narat M, Peterlin BM. AIRE recruits P-TEFb for transcriptional elongation of target genes in medullary thymic epithelial cells. Mol Cell Biol. 2007;27:8815–23.
43. Seach N, Wong K, Hammett M, Boyd RL, Chidgey AP. Purified enzymes improve isolation and characterization of the adult thymic epithelium. J Immunol Methods. 2012;385:23–34.
44. Dohr D, Engelmann R, Muller-Hilke B. A novel method to efficiently isolate medullary thymic epithelial cells from murine thymi based on UEA-1 MicroBeads. J Immunol Methods. 2019;467:12–8.
45. Griger J, Schneider R, Lahmann I, Schowel V, Keller C, Spuler S, et al. Loss of Ptpn11 (Shp2) drives satellite cells into quiescence. Elife. 2017;6:e21552.
46. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

47. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.

48. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15:550.

49. Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 2006;22:1600-7. https://doi.org/10.1093/bioinformatics/btl140.

50. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun. 2019;10:1930.

51. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–i90.

52. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

53. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137.

54. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19:15.

AUTHOR CONTRIBUTIONS
WT designed the study and performed the molecular experiments. ZY and YW contributed to the animal experiments, tissue analysis and flow cytometry analysis. ZY, YW, JW, WY, and GY contributed to the molecular, cellular and histological experiments. JX performed and managed the bioinformatic analysis and participated in manuscript preparation. SJ supervised the project, analyzed the data and wrote the manuscript. SJ is the guarantor of this work and takes responsibility for the integrity of the data and the accuracy of the data analysis.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41423-023-01102-0.

Correspondence and requests for materials should be addressed to Jieyi Xiong or Shiqi Jia.

Reprints and permission information is available at http://www.nature.com/reprints