

# Aire Controls Heterogeneity of Medullary Thymic Epithelial Cells for the Expression of Self-Antigens

Hitoshi Nishijima,<sup>\*,1</sup> Minoru Matsumoto,<sup>\*,†,‡</sup> Junko Morimoto,<sup>\*</sup> Kazuyoshi Hosomichi,<sup>§</sup> Nobuko Akiyama,<sup>¶</sup> Taishin Akiyama,<sup>||</sup> Takeshi Oya,<sup>‡</sup> Koichi Tsuneyama,<sup>†</sup> Hideyuki Yoshida,<sup>#</sup> and Mitsuru Matsumoto<sup>\*</sup>

The deficiency of Aire, a transcriptional regulator whose defect results in the development of autoimmunity, is associated with reduced expression of tissue-restricted self-Ags (TRAs) in medullary thymic epithelial cells (mTECs). Although the mechanisms underlying Aire-dependent expression of TRAs need to be explored, the physical identification of the target(s) of Aire has been hampered by the low and promiscuous expression of TRAs. We have tackled this issue by engineering mice with augmented Aire expression. Integration of the transcriptomic data from Aire-augmented and Aire-deficient mTECs revealed that a large proportion of so-called Aire-dependent genes, including those of TRAs, may not be direct transcriptional targets downstream of Aire. Rather, Aire induces TRA expression indirectly through controlling the heterogeneity of mTECs, as revealed by single-cell analyses. In contrast, *Ccl25* emerged as a canonical target of Aire, and we verified this both *in vitro* and *in vivo*. Our approach has illuminated the Aire's primary targets while distinguishing them from the secondary targets. *The Journal of Immunology*, 2022, 208: 303–320.

Autoimmunity is caused by the wrong immune responses against self-Ags due to the multiple mechanisms that are yet to be determined. Among various types of autoimmune diseases, mutation of the *AIRE* for the development of autoimmune polyendocrine syndrome type 1 merits attention because it is a monogenic disorder (1, 2). Aire is expressed in thymic epithelial cells in the medulla (medullary thymic epithelial cells [mTECs]), and its structural features suggest that it acts as a transcriptional regulator (3). Indeed, comparison of the transcriptomes from mature mTECs between wild-type (WT) mice and Aire-deficient (Aire-KO) mice revealed that Aire controls the expression of many thousands of immunologically relevant genes in mTECs (4–8). Among the large numbers of downregulated genes in the mTECs from Aire-KO, much attention has been paid to the genes of tissue-restricted self-Ags (TRAs) because the expression of TRAs is considered to be critical for the elimination of autoreactive T cells and/or the production of regulatory T (Treg) cells in an Ag-specific manner (2, 4,

5, 9, 10). However, exact target genes controlled by Aire remain incompletely defined for a couple of reasons. First, the low and promiscuous nature of the TRA expression in mTECs makes it difficult to dissect the molecular action of Aire on the transcriptional control of wide varieties of TRAs. Second, because Aire may control the expression of other transcriptional regulators as downstream targets, it is difficult to identify Aire's primary targets while distinguishing them from the secondary targets. Third, the biochemical approach to Aire protein has been hampered by its subcellular localization to the nuclear matrix (11). These situations are compounded by the lack of availability of established cell lines constitutively expressing Aire.

High-throughput transcriptomic analysis is a powerful strategy to illustrate the biological properties of a given gene. A comparison of the transcriptomic data between WT and gene-targeted mice has provided us with valuable information on the action and/or pathways in which the gene is involved. However, there are some key points during the preparation of the samples to extract the correct information: the

<sup>\*</sup>Division of Molecular Immunology, Institute for Enzyme Research, Tokushima University, Tokushima, Japan; <sup>†</sup>Department of Pathology and Laboratory Medicine, Tokushima University Graduate School of Biomedical Sciences, Tokushima, Japan; <sup>‡</sup>Department of Molecular Pathology, Tokushima University Graduate School of Biomedical Sciences, Tokushima, Japan; <sup>§</sup>Department of Bioinformatics and Genomics, Graduate School of Medical Sciences, Kanazawa University, Ishikawa, Japan; <sup>¶</sup>Laboratory for Immunogenetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; <sup>||</sup>Laboratory for Immune Homeostasis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; and <sup>#</sup>YCI Laboratory for Immunological Transcriptomics, RIKEN Center for Integrative Medical Science, Yokohama, Japan

<sup>1</sup>Current address: Department of Immunology, Tokyo Medical University, Tokyo, Japan.

ORCID: 0000-0003-2712-1825 (K.H.); 0000-0002-0670-9868 (K.T.); 0000-0003-3844-8110 (Mitsuru Matsumoto).

Received for publication July 14, 2021. Accepted for publication November 4, 2021.

This work was supported in part by Japan Society for the Promotion of Science KAKENHI Grants 16H06496, 16K21731, 18K19564, and 19H03699 (to Mitsuru Matsumoto) and 17K08885 (to H.N.).

The dataset presented in this article has been submitted to the National Center for Biotechnology Information's Gene Expression Omnibus under accession number GSE155331.

Author contributions: H.N. and Mitsuru Matsumoto designed the experiments. H.N., Minoru Matsumoto, J.M., H.Y., and Mitsuru Matsumoto conducted the experiments.

K.H. performed the RNA-seq analysis. H.N., Minoru Matsumoto, J.M., N.A., T.A., and H.Y. performed scRNA-seq analysis. H.Y., Minoru Matsumoto, and Mitsuru Matsumoto analyzed the data with bioinformatics. Minoru Matsumoto, T.O., and K.T. evaluated the pathological changes in mice. H.N., Minoru Matsumoto, H.Y., and Mitsuru Matsumoto wrote the paper.

Address correspondence and reprint requests to Prof. Mitsuru Matsumoto, Institute for Enzyme Research, Tokushima University, 3-18-15 Kuramoto, Tokushima 770-8503, Japan. E-mail address: mitsuru@tokushima-u.ac.jp

The online version of this article contains supplemental material.

Abbreviations used in this article: 3xAire-KI, 3xAire-knock-in; Aire-KO, Aire-deficient mice; ATAC-seq, assay for transposase-accessible chromatin followed by high-throughput sequencing; *Ccl25*-KO, *Ccl25*-deficient mice; ChIP-seq, chromatin immunoprecipitation assay followed by sequencing; cTEC, cortical thymic epithelial cell; EpCAM, epithelial cell adhesion molecule 1; FC, fold change; FDR, false discovery rate; GSEA, gene set enrichment analysis; *Ins2*, insulin 2; KI, knock-in mouse; MFI, mean fluorescence intensity; MHC-II, MHC class II; mTEC, medullary thymic epithelial cell; Pol-II, RNA polymerase II; Sap1, salivary protein 1; scRNA-seq, single-cell RNA sequencing; TRA, tissue-restricted self-Ag; Treg, regulatory T; TSS, transcriptional start site; UEA-1, *Ulex Europaeus* agglutinin I; WT, wild-type.

This article is distributed under The American Association of Immunologists, Inc., [Reuse Terms and Conditions for Author Choice articles](#).

Copyright © 2022 by The American Association of Immunologists, Inc. 0022-1767/22/\$37.50

purity and/or homogeneity of the cells of interest. If the samples are a mixture of cells with different subpopulations and/or statuses, interpretation of the data would become complex. As such, there have been some difficulties in drawing the whole picture of Aire in mTECs. Because mTECs are a heterogeneous population in terms of not only the spectrum of TRAs but also the maturation stages and cell types (12, 13), we have to be careful with the interpretation of the results when the pooled mTECs were used for the analysis. For example, if Aire has any role in the differentiation of mTECs (14), it is not easy to conclude whether the altered expression of TRAs observed in Aire-KO mTECs is due to the loss of Aire's transcriptional activity on the corresponding TRAs or to the altered composition of the mTECs wherein a set of unique TRAs are expressed from each subpopulation. The issue has become more complex by the fact that TRA expression from the mTECs does not follow a conventional transcriptional control in its authentic cell types as exemplified by the Aire-dependent TRA of *insulin 2* (*Ins2*) (15, 16). Furthermore, we also need to keep in mind that Aire may control other transcriptional regulators as its downstream targets. In this case, genes changed in Aire-KO mTECs may not be the direct downstream targets of Aire. Instead, they might have been the indirect targets controlled by Aire via secondary Aire-dependent transcriptional regulators. Thus, both heterogeneities of mTECs and the complexity of transcriptional networks need to be considered for the characterization of Aire-controlled genes when the data were obtained from pooled mTECs. In this regard, the recent advent of single-cell analysis is a powerful approach to resolving this problem (5, 17–22). Although low cell number has been a limitation of single-cell analysis in some studies, the power of the single-cell approach was underscored by the identification of a novel subpopulation of tuft-like cells that play a unique role in creating a thymic microenvironment (23, 24).

Given that Aire is a transcriptional regulator, transcriptomic changes caused not only by the loss-of-function but also by overexpression of Aire might be useful for better characterization of the targets controlled by Aire: all the transcriptomic information on Aire in mTECs so far has been obtained by comparing WT with Aire-KO, and the information on the overexpression of Aire on the *in vivo* transcriptomic landscape of mTECs is still lacking. Because the transcriptional impact of Aire varies with cell type (25), additive Aire expression needs to be achieved using the endogenous Aire promoter/enhancer element: although we have previously examined the effect of the additive human AIRE expression in mice (7), human AIRE was expressed under the control of the MHC class II (MHC-II) promoter. To gain a better understanding of the transcriptional regulation by Aire in mTECs, we have established a novel strain of mice in which Aire expression was augmented in authentic Aire-expressing mTECs. By integrating the transcriptomic data from Aire-augmented and Aire-KO mTECs, we first revisited the Aire-regulated genes on a global scale. We then investigated the molecular and cellular basis for the transcriptomic changes caused by Aire's gene dosage using high-performance single-cell analyses. These approaches have provided a coherent and new look at the transcriptional regulation by Aire in mTECs with the identification of *Ccl25* as a canonical Aire target involved in disease development.

## Materials and Methods

### Mice

Mice expressing augmented Aire from authentic Aire-expressing cells were generated by constructing a knock-in (KI; 3xAire-KI) allele harboring two sets of a self-cleaving 2A peptide sequence (26) followed by mouse Aire cDNA into the Aire locus as shown in Fig. 1A. 3xAire-KI mice were generated by homologous recombination in embryonic stem cells on a C57BL/6 (B6) background. 3xAire-KI mice backcrossed onto NOD for eight generations were used for assay for transposase-accessible chromatin followed by high-throughput sequencing (ATAC-seq). Aire-KO mice on a NOD background have been reported previously (27). Mice deficient for *Ccl25* (28)

were purchased from The Jackson Laboratory. C57BL/6 and NOD mice were purchased from CLEA Japan. The mice were maintained under pathogen-free conditions and handled in accordance with the Guidelines for Animal Experimentation of Tokushima University School of Medicine.

### Immunohistochemistry

Thymi from heterozygous 3xAire-KI mice were harvested, embedded in optimal cutting temperature compound (Sakura Finetek), and frozen at  $-80^{\circ}\text{C}$ . Frozen tissue sections were fixed in acetone. Immunohistochemical analysis of the samples using rat anti-mouse Aire mAb (clone MM-525; prepared in our laboratory) and DAPI (Sigma-Aldrich) was performed as described previously (7).

### TEC preparation and flow cytometric analysis

Preparation of TECs and flow cytometric analysis with a FACSCalibur (BD Biosciences) and a FACSARIA II (BD Biosciences) were performed as described previously (7). In brief, thymic lobes were isolated from mice and cut into small pieces. The fragments were pipetted to remove the majority of thymocytes in RPMI 1640 medium (Life Technologies) supplemented with 10% heat-inactivated FCS (Bovogen Biologicals), 20 mM HEPES, 50 U/ml penicillin, 50  $\mu\text{g}/\text{ml}$  streptomycin, and 50  $\mu\text{M}$  2-ME, hereafter referred to as "R10." The resulting thymic fragments were digested with 0.15% collagenase D (Roche) and 10 U/ml DNase I (Roche) in R10 at  $37^{\circ}\text{C}$  for 30 min. The supernatants, containing dissociated TECs, were saved, and the remaining thymic fragments were digested with 0.15% collagenase/dispase (Roche) and DNase I at  $37^{\circ}\text{C}$  for 30 min. The supernatants from this digest were combined with the supernatants from the collagenase digests, and the mixture was centrifuged for 5 min at  $400 \times g$ . The cells were suspended in PBS containing 5 mM EDTA and 1% BSA and kept on ice until the staining. The mAbs used were anti-CD45 (clone 30-F11), anti-Ly51 (clone 6C3), anti-CD80 (clone 16-10A1), anti-CD4 (clone GK1.5), anti-CD8a (clone 53-6.7), anti-TCR $\beta$  (clone H57-597), anti-CD25 (clone PC61), anti-CD44 (clone IM7), anti-CD62L (clone MEL-14), anti-CD69 (clone H1.2F3), B220 (clone RA3-6B2), epithelial cell adhesion molecule 1 (EpCAM; clone G8.8), and anti-V $\alpha$ 2 (clone B20.1) purchased from BioLegend. The mAbs against Foxp3 (clone FJK-16s) and I-A/I-E (clone M5/114.15.2) were purchased from eBioscience. *Ulex Europaeus* agglutinin 1 (UEA-1) was from Vector Laboratories. Anti-V $\beta$ 5 mAb was from BD Biosciences. A rat anti-mouse Aire mAb (clone MM-525) was also used. Dead cells were excluded from the analysis using 7-aminoactinomycin D (BD Biosciences).

### Real-time PCR

RNA was extracted from TECs using RNeasy Mini Kits (Qiagen) and converted to cDNA with SuperScript III or VILO RT kits (Invitrogen) in accordance with the manufacturer's instructions. Real-time PCR for quantification of the *Aire*, *Ins2*, *salivary protein 1* (*Sap1*), *C-reactive protein*, *Ccl25*, and *Hprt* genes was performed as described previously (7). The following primers and dual-labeled probes (FAM/TAMRA) were used: *Aire* primers, 5'-TCTGC TAGTCACGACCCTGTTC-3' and 5'-GGCGTGACATGCTCTGGATG-3'; probe, 5'-ACGACCTGGAGTCCCTCCTCAATGAGC-3'. *Ins2* primers, 5'-GACCCACAAGTGGCACA-3' and 5'-ATCTACAATGCCACGCTTCTG-3'; probe, 5'-GCCCGGAGCAGGTGACCTT-3'. *Sap1* primers, 5'-ACTC CTTGTGTTGCTTGGTGT-3' and 5'-TCGACTGAATCAGAGGAATCA ACT-3'; probe, 5'-TCCTAGTCTCTTCCAGGACCCGGAGA-3'. *CRP* primers, 5'-TACTCTGGTGCCTTCTGATCATGA-3' and 5'-GGCTTCTT GACTCTGCTTCCA-3'; probe, 5'-CAGCTTCTCGGACTTTTGGTCAT GA-3'. *Ccl25* primers, 5'-TTGTCCATGCCCAAGGTGC-3' and 5'-CTGG CGGAAGTAGAATCTCACA-3'; probe, 5'-CACGTAGGTTGCAGCTTC-CACTCACTCC-3'. *Hprt* primers, 5'-TTCCCTGGTTAAGCAGTACAGC-3' and 5'-CCAACAAAGTCTGGCTGTATC-3'; probe, 5'-TTGAGAGG TCCTTTCCACGCAAGCT-3'.

### RNA-sequencing (RNA-seq) analysis

Total RNAs from sorted mTEC<sup>high</sup> were isolated using an RNeasy Mini Kit (Qiagen) following the manufacturer's instructions. Eluted RNAs were incubated with RT-grade DNase (Nippon Gene) for 10 min at room temperature, followed by purification on an RNeasy column (Qiagen). RNAs were verified quantitatively and qualitatively using an Agilent RNA 6000 Pico Kit run on an Agilent 2100 Bioanalyzer (Agilent Technologies). RNA integrity numbers were  $>7$ . Three independent biological replicates from each genotype were used for RNA-seq analysis. Sequencing libraries were generated from 1.0 ng of total RNA with the Ovation SoLo RNA-Seq System (NuGEN). The libraries were size selected with AMPure XP beads (Agilent Technologies) and used for cluster generation on the flow cells. Single-end 150-bp sequencing was performed with the Illumina MiSeq (Illumina), which generated 11.4 million–14.9 million reads per sample.

### Chromatin accessibility (ATAC-seq)

Chromatin accessibility by ATAC-seq was performed as follows. DAPI<sup>-</sup>CD45<sup>-</sup>EpCAM<sup>+</sup>Ly51<sup>-</sup>CD80<sup>high</sup> mTECs were isolated with a FACSAria III cell sorter from a homozygous 3xAire-KI female mouse (a total of 13,008 cells) and two WT female mice (a total of 9,050 cells), both at 8 wk of age on a NOD background, and were collected in a 1.5-ml tube containing 200  $\mu$ l of BAMBANKER (GC LYMPHOTEC). Tubes were frozen in a cell-freezing container filled with isopropyl alcohol at  $-1^{\circ}\text{C}/\text{min}$  rate and stored at  $-80^{\circ}\text{C}$  until library construction. ATAC-seq libraries were prepared as previously reported (29) with some modifications. Specifically, frozen samples were thawed, split into two tubes, and washed with 1 ml of ice-cold PBS containing protease inhibitors (cOmplete ULTRA, EDTA-free; Merck). Cell pellets were resuspended in 35  $\mu$ l of Tn5 transposase mixture (10 mM tris(hydroxymethyl)methylaminopropanesulfonic acid NaOH, pH 8.5, 5 mM MgCl<sub>2</sub>, 10% dimethylformamide, 0.2 mg/ml digitonin, and in-house Tn5) on ice. Then, the cells were incubated at  $37^{\circ}\text{C}$  for 30 min with agitation followed by DNA isolation using the Monarch PCR & DNA Cleanup Kit (New England Biolabs) according to the manufacturer's instruction. To construct libraries, PCR was performed in a 50- $\mu$ l total volume using KAPA HiFi DNA Polymerase (Roche) with a set of primers (0.3  $\mu$ M each) designed by Buenrostro et al. (30) with the following thermal cycles: 5 min at  $72^{\circ}\text{C}$ , 30 s at  $95^{\circ}\text{C}$  followed by 17 cycles ( $98^{\circ}\text{C}$  for 20 s,  $63^{\circ}\text{C}$  for 10 s), and a final extension at  $72^{\circ}\text{C}$  for 3 min. The PCR products were purified using a 1.8 $\times$  volume of AMPure XP beads (Beckman Coulter) and eluted in 25  $\mu$ l of EB (Qiagen). The libraries were quantified using a High Sensitivity DNA Assay (Agilent Technologies) and the Colibri Library Quantification Kit (Thermo Fisher Scientific), which were pooled and sequenced for paired-end sequencing (50 + 50 bp) on an Illumina HiSeq 2500 instrument in a rapid mode. After trimming adapter sequences and low-quality reads by fastp (version 0.20.0) (31), short reads were mapped to the mm10 reference genome by bowtie2 with options  $-X$  1000  $-fr$  (32), whereas non-uniquely mapped reads were filtered out by samtools view  $-q$  30 (33). Then, chrM mapped reads (2.3–4.7%) and duplicated reads (60–80%) were removed using samtools. A pile-up trace for each sample was computed by piling up the fragments defined by paired reads using macs2 with the  $-SPMR$  option (34).

### Method for informatics

After trimming low-quality reads and adapter sequences by fastp (31), short reads were mapped to mm10 by star (version 2.7.3a) (35) employing an Ensembl GTF file for mm10 downloaded from <https://www.ensembl.org>. Unmapped reads and reads mapped with a low quality score (mapping quality <5) were removed using samtools (version 1.9) (33), and duplicated reads were removed by Picard MarkDuplicates (<http://broadinstitute.github.io/picard>). Mapped reads on each transcript were counted using htseq-count (version 0.11.2) (36) with  $-stranded = \text{no}$  option and an Ensembl GTF file for mm10. Raw read counts on transcripts were processed using R (version 3.4.1) with the edgeR package (version 3.20.9) (37) for normalization and to identify differentially expressed genes. Gene Ontology enrichment analysis of a gene set was performed using R with the clusterProfiler package (version 3.6.0) (38). The data set associated with this project has been submitted to the National Center for Biotechnology Information (Gene Expression Omnibus accession number GSE155331).

### Chromatin immunoprecipitation assay followed by sequencing (ChIP-seq)

We used the ChIP-seq data sets (39), Gene Expression Omnibus accession number GSE92654 for Aire, RNA polymerase II (Pol-II) and IgG (control) in mTECs, and corresponding bigwig files were downloaded from <https://www.ncbi.nlm.nih.gov/geo/>. ChIP-seq signals on genomic regions of interests were extracted from bigwig files using R with the rtracklayer package (version 1.38.3) (40) and the GenomicRanges package (version 1.30.3) (41).

### Single-cell RNA-sequencing (scRNA-seq) analysis

FACS-sorted mTEC<sup>high</sup> from one mouse at 8 wk of age were prepared for scRNA-seq analysis. In brief, cell suspensions were loaded onto a Chromium instrument (10x Genomics) to generate a single-cell emulsion. scRNA-seq libraries were prepared using Chromium Single Cell 3' Reagent Kits version 2 Chemistry and sequenced in multiplex on the Illumina HiSeq2000 platform (rapid mode). FASTQ files were processed using Fastp, and reads were demultiplexed and mapped to the mm10 reference genome via Cell Ranger (version 3.1.0). Data processing by the Cell Ranger software was performed using supercomputers HOKUSAI at RIKEN and the Research Organization of Information and Systems, National Institute of Genetics. Expression count matrices were prepared by counting unique molecular identifiers. Genes that were expressed in more than 5 cells and cells expressing at least 200 genes were selected for the analysis. Seurat 3.0 (42), velocyto.R (43), and R 4.0 (<https://www.R-project.org/>) were used to analyze the scRNA-seq data.

### Gene set enrichment analysis (GSEA)

Averaged gene expression was computed by merging the raw read counts for cells in the corresponding clusters and normalized using the R package edgeR (version 3.30.3) (44). GSEA was performed employing the R package GO.db (version 3.11.4) and clusterProfiler (version 3.6.0) (38). The enrichment  $p$  value calculation (i.e., number of genes in the list that hit a given biology class compared with pure random chance) was performed using Benjamini-Hochberg multiple testing correction.

### Ccl25 reporter gene assay

cDNA coding for mouse Aire was subcloned into a pEGFP vector (Clontech) by replacing the enhanced GFP (EGFP) fragment (pCMV-mAire). A PCR-amplified Ccl25 promoter fragment (using Ccl25 promoter sense: 5'-GAACTTCAGGATACACTCGAGGATC-3' and Ccl25 promoter antisense: 5'-CGCTTGTGACAGGTTAGATCTCTCTCCAGATAC-3') was subcloned into the XhoI site of pGL3-Basic vector (Promega) in a forward (Ccl25-F) or reverse (Ccl25-R) orientation as shown in Supplemental Fig. 3A. The mTEC-like cell line 1C6 (kindly provided by Dr. M. Kasai, Kochi University, Kochi, Japan) was cultured in 12-well plates using DMEM (ThermoFisher) supplemented with FCS, GlutaMAX (1 $\times$ ), 2-ME (50  $\mu$ M), hydrocortisone hemisuccinate (1.5  $\mu$ g/ml), sodium pyruvate (1 mM), and antibiotics. The cells were transfected with 900 ng of Ccl25 reporter vectors (Ccl25-F or Ccl25-R) together with 10 ng of pRL-TK vector (Promega) and 90 ng of pEGFP vector or pCMV-mAire vector (i.e., 1000 ng of DNA in total amount) using X-tremeGENE HP DNA Transfection Reagent (D-0901; Roche Applied Science). Cells were harvested 24 h after transfection, and firefly luciferase activities were measured by Lumat LB9507 luminometer (Berthold) and normalized to Renilla luciferase activities using the Dual Luciferase Reporter Assay System (E1910; Promega) according to the manufacturer's instructions. Luciferase activities from the cells transfected with the EGFP-expressing vector were determined as 1.

### Pathology

Formalin-fixed tissue sections were subjected to H&E staining, and two pathologists independently evaluated the histology without being informed of the detailed condition of the individual mouse. Histological changes were scored as none (score 0), mild (score 1), moderate (score 2), and severe (score 3).

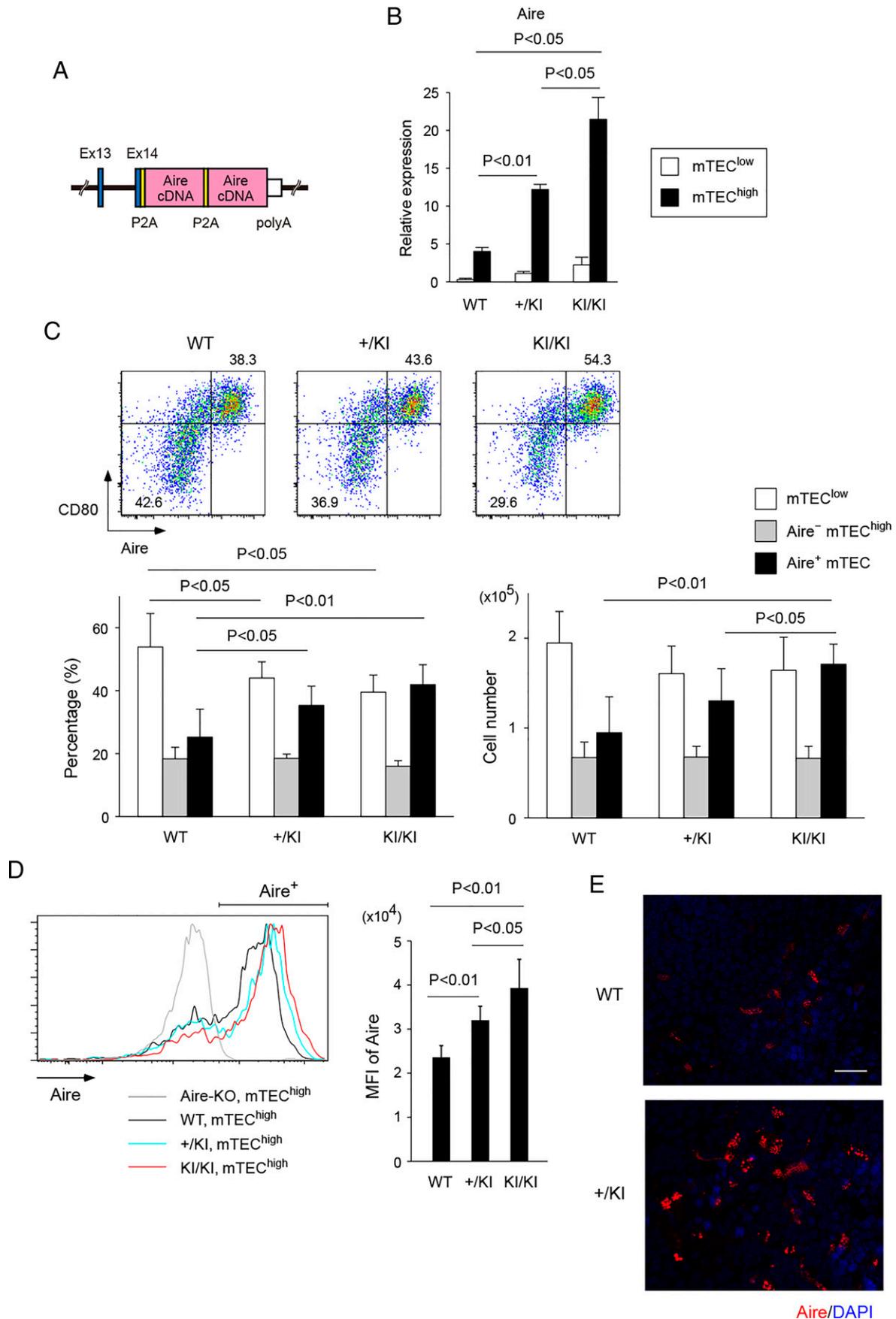
### Statistical analysis

All results are expressed as mean  $\pm$  SEM. Differences were considered significant at  $p < 0.05$ .

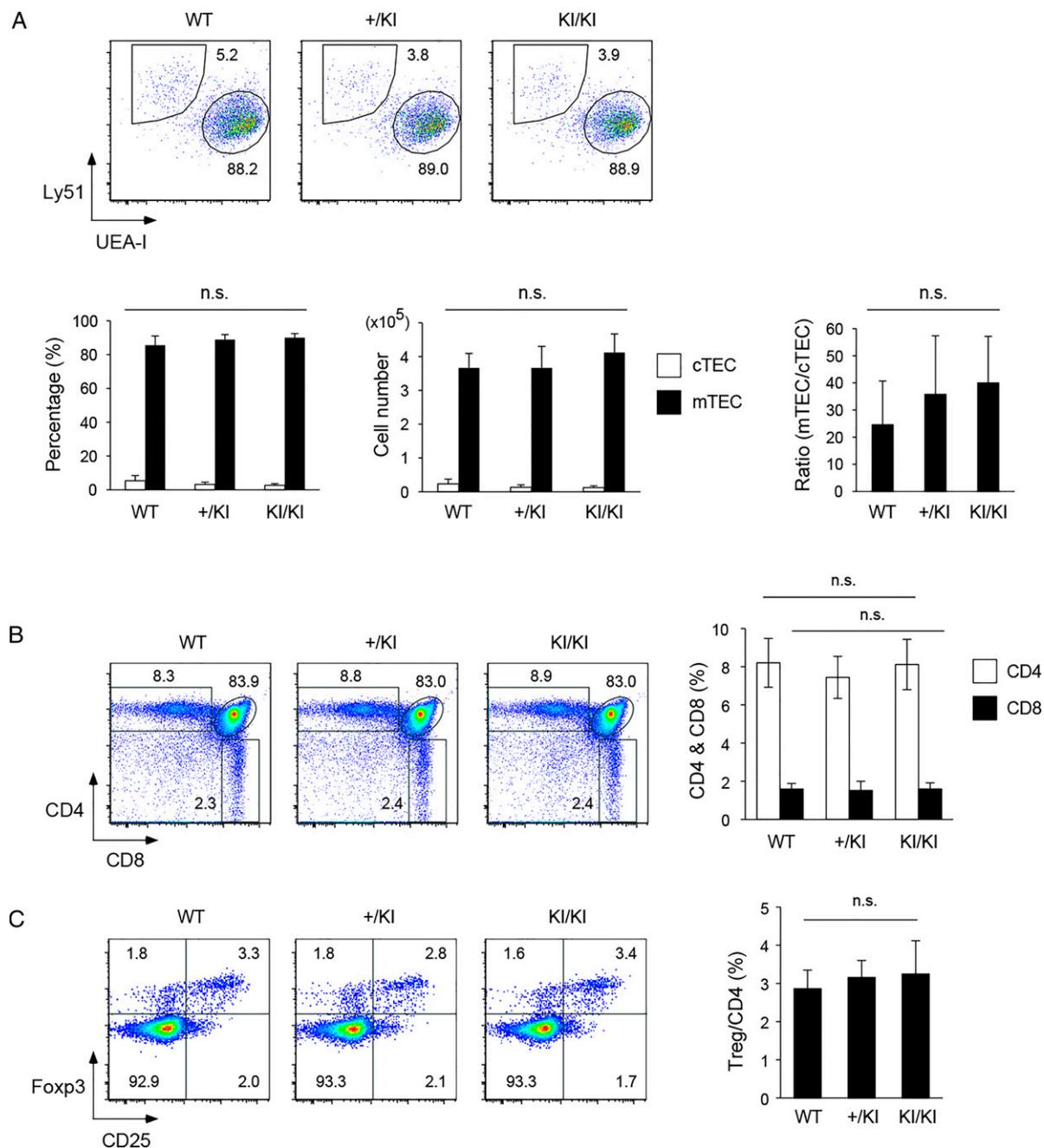
## Results

### Generation of a novel mouse strain expressing augmented Aire from authentic Aire-expressing cells

Augmented Aire expression from authentic Aire-expressing cells was achieved by constructing a KI allele harboring two sets of the self-cleaving 2A peptide sequence (P2A-linker) (26) followed by mouse Aire cDNA under the control of the endogenous Aire promoter/enhancer element (Fig. 1A). This makes it possible to transcribe three copies of mouse Aire mRNA from the KI allele (3xAire-KI). 3xAire-KI mice were generated by homologous recombination in embryonic stem cells on a C57BL/6 (B6) background. We first examined the levels of Aire mRNA expression from CD45<sup>-</sup>EpCAM<sup>+</sup>Ly51<sup>-</sup>UEA-1<sup>+</sup> mTECs after dividing them into immature CD80<sup>low</sup>MHC-II<sup>low</sup> (mTEC<sup>low</sup>) and mature CD80<sup>high</sup>MHC-II<sup>high</sup> (mTEC<sup>high</sup>) populations. We found that mTEC<sup>high</sup> from heterozygous and homozygous 3xAire-KI expressed almost double and triple the amount of Aire mRNA, respectively, compared with those from WT (Fig. 1B). We then evaluated the expression of Aire protein using flow cytometry and found that 3xAire-KI had more Aire<sup>+</sup> mTECs than WT, accompanied by a reduction in the percentage of mTEC<sup>low</sup> (Fig. 1C). Of note, the mean fluorescence intensities (MFIs) of Aire protein expression in mTEC<sup>high</sup> were also increased dose dependently by the augmented Aire expression (Fig. 1D). Immunohistochemical analysis showed that augmented Aire expression in heterozygous 3xAire-KI made each Aire nuclear dot slightly larger (Fig. 1E). In contrast to the mTECs, augmented Aire expression did not affect the cortical thymic epithelial cell (cTEC) component, maintaining the mTEC/cTEC ratios comparable to those



**FIGURE 1.** Generation of a novel mouse strain expressing augmented Aire from authentic Aire-expressing cells. **(A)** Schematic representation of the KI allele for augmented Aire expression. **(B)** Levels of Aire mRNA from immature mTEC<sup>low</sup> (white) and mature mTEC<sup>high</sup> (black) populations assessed using real-time PCR. The data were obtained from three repeated experiments using female mice at 4 wk of age with duplicate or triplicate samples for each experiment. **(C)** Aire<sup>+</sup> mTECs were examined by flow cytometry after gating for CD45<sup>-</sup>EpCAM<sup>+</sup>Ly51<sup>-</sup>UEA1<sup>+</sup> mTEC fractions. (Figure legend continues)



**FIGURE 2.** Phenotypic analysis of mice with augmented Aire expression from mTECs. **(A)** cTECs ( $\text{Ly51}^+\text{UEA-1}^-$ ) and mTECs ( $\text{Ly51}^-\text{UEA-1}^+$ ) were examined by flow cytometry after gating for  $\text{CD45}^-\text{EpCAM}^+$  fractions. One representative experiment from a total of five repeats is shown with percentages of cTECs and mTECs (upper row). The percentages, the numbers of cTECs and mTECs, and the ratios of mTEC/cTEC from each genotype are shown in the lower row. Seven or eight mice at 4 wk of age were analyzed for each genotype group. **(B)** Thymocyte development in mice at the age of 3–6 wk was examined by flow cytometry. One representative experiment from a total of nine repeats is shown on the left, and a summary of the results is shown on the right. WT,  $n = 9$ ; +/KI,  $n = 5$ ; KI/KI,  $n = 9$ . **(C)** Development of thymic Treg cells from mice shown in B after gating for CD4 single-positive T-cells.

from WT (Fig. 2A). Thus, we successfully generated mice with augmented Aire expression from authentic Aire-expressing cells. Overall, 3xAire-KI showed normal development of CD4 and CD8 single-positive thymocytes (Fig. 2B) and  $\text{CD4}^+\text{CD25}^+\text{Foxp3}^+$  mature Treg cells (Fig. 2C).

#### *A coherent and new look at transcriptional regulation by Aire in mTECs*

We have previously compared the transcriptomes of FACS-sorted  $\text{mTEC}^{\text{high}}$  populations between WT and Aire-KO on a NOD background (Supplemental Fig. 1A) (7). To investigate whether the

One representative experiment from a total of five repeats is shown with percentages of  $\text{CD80}^{\text{high}}\text{Aire}^+$  and  $\text{CD80}^{\text{low}}\text{Aire}^-$  cells (upper). Seven or eight mice at 4 wk of age were analyzed for each genotype group. **(D)** MFIs of Aire expression in  $\text{Aire}^+$  mTECs from mice analyzed in C.  $\text{mTEC}^{\text{high}}$  from Aire-KO served as a negative control. The  $\text{Aire}^+$  area used for MFI calculation is shown by the upper line. One representative experiment from a total of five repeats is shown on the left, and a summary of the results is shown on the right. **(E)** Immunohistochemical analysis of thymi from WT and heterozygous 3xAire-KI mice stained with anti-Aire mAb and DAPI. Scale bar, 20  $\mu\text{m}$ .

transcriptomic data obtained from Aire-KO on NOD can be integrated with those from 3xAire-KI on B6, we compared our Aire-KO on NOD data with publicly available data for Aire-KO on B6 (6), both obtained from FACS-sorted mTEC<sup>high</sup>. Use of a fold change (FC) –FC plot revealed a high correlation between the two sets of data ( $r^2 = 0.558$ ;  $p < 2.2 \times 10^{-16}$ ; Supplemental Fig. 1B), allowing us to integrate our transcriptomic data from Aire-KO on NOD with those from 3xAire-KI on B6.

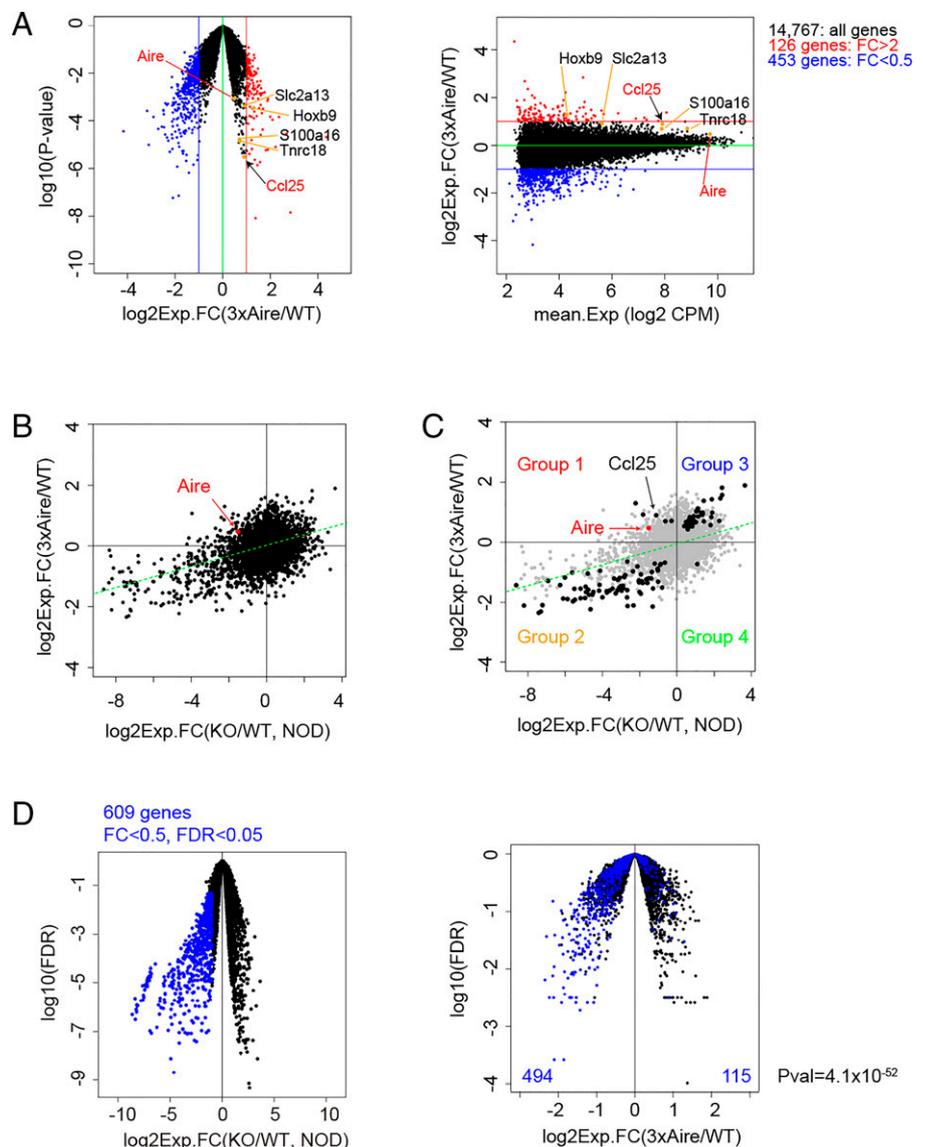
We obtained RNA-seq data using FACS-sorted mTEC<sup>high</sup> from WT and homozygous 3xAire-KI with triplicate samples for each (Fig. 3A). We then prepared an FC–FC plot between  $\log_2$  FC (Aire-KO/WT) and  $\log_2$  FC (3xAire-KI/WT) (Fig. 3B). When the false discovery rate (FDR) was set at  $<0.05$ , 125 genes were significantly changed, depending on the Aire status (Fig. 3C and Supplemental Table I). Interestingly, many genes showed similar alterations in the levels of expression upon deletion (in Aire-KO) or addition of Aire (in 3xAire-KI) (Fig. 3C). Indeed, when the genes downregulated in Aire-KO (FDR  $<0.05$ ; FC  $<0.5$ ) were selected (609 genes; Fig. 3D, left) and projected onto the volcano plot of  $\log_2$  FC (3xAire-KI/WT), more genes were downregulated than upregulated (494 versus 115 genes;  $p = 4.1e^{-52}$ ) (Fig. 3D, right). Here, we classified the genes into four groups depending on Aire status (Fig. 3C): group 1,

genes downregulated in Aire-KO and upregulated in 3xAire-KI (“Aire-induced” genes); group 2, genes downregulated in both Aire-KO and 3xAire-KI; group 3, genes upregulated in both Aire-KO and 3xAire-KI; and group 4, genes upregulated in Aire-KO and downregulated in 3xAire-KI (“Aire-suppressed” genes). *Aire* itself was in group 1, as expected (Fig. 3C).

*A large proportion of Aire-dependent genes, including those of TRAs, are noncanonical and indirect transcriptional targets downstream of Aire*

So far, many studies of autoimmune pathogenesis have focused on the genes downregulated by Aire deficiency, as exemplified by TRAs. In this context, it is noteworthy that our use of the transcriptomic data from 3xAire-KI revealed that genes downregulated in Aire-KO, so-called Aire-dependent genes, were further divisible into two groups, groups 1 and 2 (Fig. 3C), as we considered that the factors differing between these two groups would be of interest. To answer this issue, we first focused on the top 200 genes up- or downregulated in 3xAire-KI based on their FC (Supplemental Fig. 1C). We then examined the densities of Pol-II surrounding these genes because a previous study had suggested that the expression of Aire’s target genes would be marked by relatively high Pol-II densities (45); it was concluded that Aire activates target genes by releas-

**FIGURE 3.** A coherent and new look at transcriptional regulation by Aire in mTECs. **(A)** Comparison of the transcriptomes from FACS-sorted mTEC<sup>high</sup> between WT and homozygous 3xAire-KI. Genes upregulated (126 genes; FC  $>2$ ) and downregulated (453 genes; FC  $<0.5$ ) among 14,767 genes are marked in red and blue, respectively. *Ccl25* is marked by a black arrow, and several genes, including *Aire* (in red), are highlighted. **(B)**  $\log_2$  FC values for Aire-KO/WT ( $n = 3$ , x-axis) versus 3xAire-KI/WT ( $n = 3$ , y-axis). A total of 12,247 expressed genes were plotted, and *Aire* is highlighted in red. The green dotted line indicates the linear regression line,  $y = 0.175x - 0.0205$ ;  $r^2 = 0.159$ ;  $p < 2.2e^{-16}$ . **(C)** FDR-selected genes (FDR  $<0.05$  in both comparisons) shown in B are highlighted in black. Four groups of genes emerged, depending on the effect of Aire. *Ccl25* is marked by a black arrow, and *Aire* is highlighted in red, both in group 1. **(D)** Projection of the downregulated genes (609 genes; FC  $<0.5$  and FDR  $<0.05$ ) in mTEC<sup>high</sup> from Aire-KO shown in Supplemental Fig. 1A (left) onto the volcano plot for mTEC<sup>high</sup> between WT and homozygous 3xAire-KI shown in A, left (right).

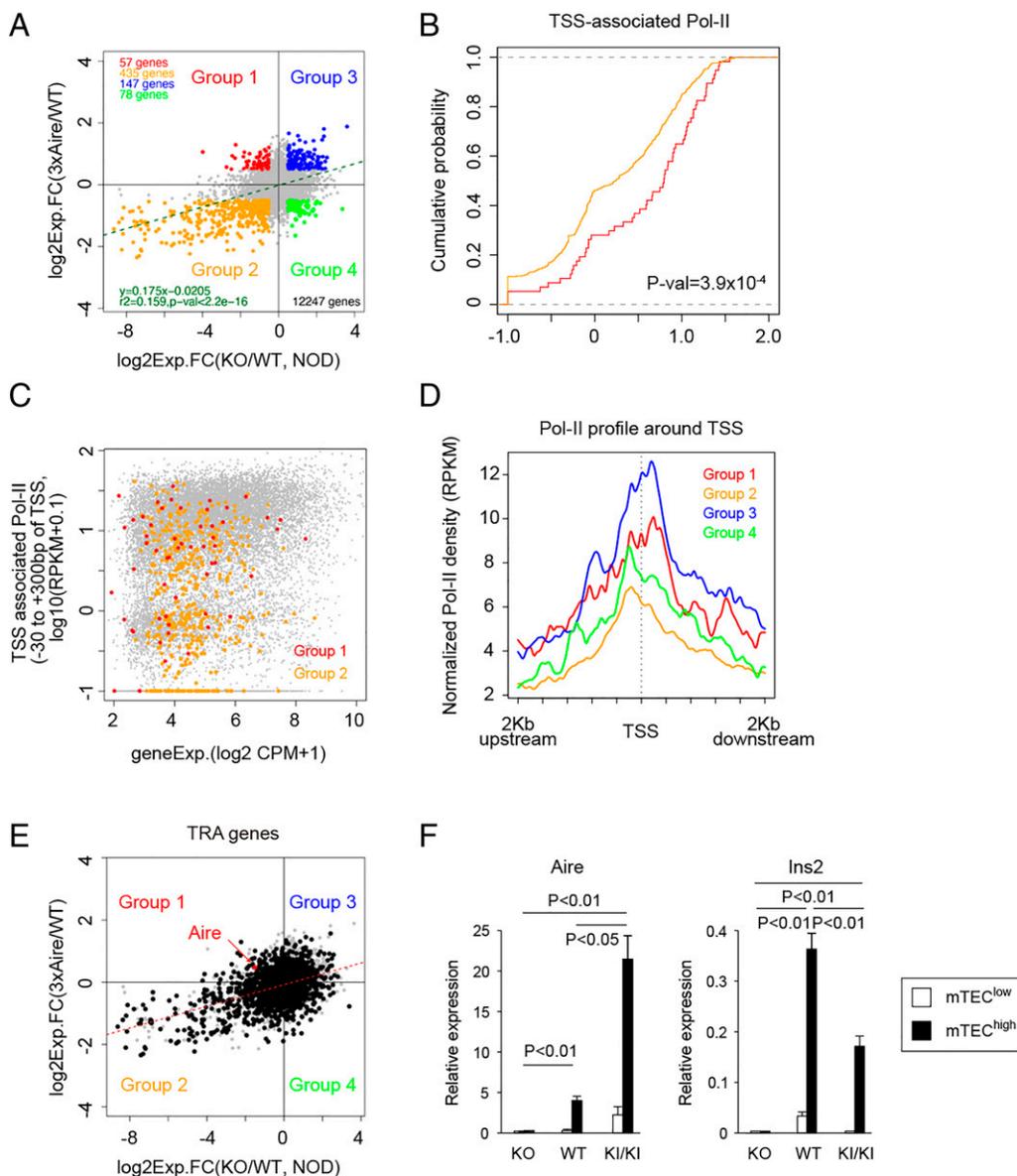


ing the stalled Pol-II. When we analyzed the Pol-II densities of the top 200 up- and downregulated genes in 3xAire-KI using ChIP-seq data obtained from WT mTECs (39), we found that the upregulated genes (induced genes) in 3xAire-KI had higher Pol-II densities than those that were downregulated (suppressed) in the vicinity of transcriptional start sites (TSSs) (Supplemental Fig. 1D, left and center) under conditions where the expression levels of these annotated genes were similar (Supplemental Fig. 1D, right). These results suggested that the genes upregulated in Aire-augmented mTECs from 3xAire-KI were more likely to be canonical Aire targets than the genes that were downregulated. With this in mind, we investigated the Pol-II densities of genes belonging to the four groups described above. For this analysis, we selected 57 genes from group 1, 435 genes from group 2, 147 genes from group 3, and 78 genes from group 4 based on their FC ( $\log_2$  FC >0.5 or  $\log_2$  FC less than -0.5) (Fig. 4A and Supplemental Table I). When we examined the Pol-II densities by focusing on the genes in groups 1 and 2, we found that genes in group 1 showed higher Pol-II densities than those in group 2 in the vicinity of TSSs (Fig. 4B) where the expression levels of these annotated genes were similar (Fig. 4C), suggesting that genes in group 1 might be better Aire target candidates. In

fact, genes in group 2 showed the lowest Pol-II densities among the four groups (Fig. 4D). Taken together, our results suggested that the genes in group 1 behaved as more likely canonical targets downstream of Aire than those in group 2.

We then investigated the groups to which the TRA genes belonged. By defining TRAs as genes that are expressed in <3 organs, 3,533 out of a total of 12,247 genes fitted this criterion (Fig. 4E). When  $\log_2$  FC >0.5 or  $\log_2$  FC less than -0.5 were set, a total of 361 TRA genes were differentially expressed: 24 genes in group 1, 237 genes in group 2, 60 genes in group 3, and 40 genes in group 4 (Supplemental Table I). The frequency of differentially expressed TRA genes among all the differentially expressed genes in each group (shown in Fig. 4A) was not so different: group 1, 24 out of 57 genes (42.1%); group 2, 237 out of 435 genes (54.5%); group 3, 60 out of 147 genes (40.8%); group 4, 40 out of 78 genes (51.3%). However, group 2 held the highest numbers of TRA genes (237 out of 361 total differentially expressed TRA genes [65.7%]). In contrast, group 1 contained only 24 out of 361 TRA genes (6.6%); group 3 and group 4 had 60 (16.7%) and 40 TRA genes (11.1%), respectively. Given that genes in group 1 exhibited better characteristics as Aire targets (Fig. 4B and 4D), a large number of TRAs

**FIGURE 4.** Expression of TRA genes as noncanonical and indirect transcriptional targets downstream of Aire. **(A)** Four groups of genes selected based on FCs ( $\log_2$  FC >0.5 and  $\log_2$  FC less than -0.5) shown in Fig. 3B: group 1, 57 genes (red); group 2, 435 genes (orange); group 3, 147 genes (blue); group 4, 78 genes (green). **(B)** A cumulative distribution function plot for TSS-associated Pol-II density. The red and orange lines represent genes for group 1 and group 2, respectively. **(C)** Similar expression levels of 57 genes in group 1 (red) and 435 genes in group 2 (orange) were used for deriving the cumulative distribution function plot shown in B. **(D)** Line plots for the normalized densities of Pol-II surrounding the genes in the four groups shown in A based on the ChIP-seq data from WT mTECs (39). **(E)** TRA genes (3,533 genes) among the genes shown in Fig. 3B (12,247 genes) are highlighted in black. The red dotted line indicates the linear regression line,  $y = 0.176x - 0.0614$ ;  $r^2 = 0.216$ ;  $p < 8.3e^{-189}$ . **(F)** The levels of expression of *Aire* and *Ins2* from mTEC<sup>low</sup> and mTEC<sup>high</sup> in Aire-KO, WT, and homozygous 3xAire-KI examined by real-time PCR. The data were obtained from three repeated experiments using mice at 4 wk of age.



present in group 2 may not have been direct targets of Aire. Rather, transcription of these genes might have been controlled by Aire indirectly through some other mechanisms beyond the canonical relationship between the transcriptional regulator and its target(s) (see below). Real-time PCR analysis showed that a prototypical Aire-dependent TRA of *Ins2* also behaved as a gene for group 2 (Fig. 4F).

#### *Altered heterogeneity of mTECs accounts for the paradoxical reduction in the expression of TRAs in Aire-augmented mice*

Given that a large proportion of Aire-dependent genes, including those of TRAs, may not be direct targets of Aire, we investigated the mechanisms underlying the paradoxical reduction in the expression of many TRAs in 3xAire-KI. We performed scRNA-seq analyses using FACS-sorted CD45<sup>-</sup>EpCAM<sup>+</sup> TECs from WT and homozygous 3xAire-KI. With stringent quality checking (i.e., filtering the cells that have unique feature counts >5000 or <1500, the cells that have >4% mitochondrial counts, and the cells that have unique RNA counts <5000) for 4 data sets (i.e., WT\_1: 1504 cells; WT\_2: 4585 cells; 3xAire-KI\_1: 4286 cells; 3xAire-KI\_2: 5238 cells), 12 clusters emerged (Fig. 5A). Three of them (clusters 1, 4, and 5) corresponded to mTEC<sup>high</sup> (Supplemental Fig. 2A), whereas seven (clusters 2, 3, 6, 7, 9, 11, and 12) corresponded to mTEC<sup>low</sup> (Supplemental Fig. 2B), based on their gene signatures defined by transcriptomic comparison between mTEC<sup>high</sup> and mTEC<sup>low</sup> (6). Aire was expressed from clusters 1 and 4 among mTEC<sup>high</sup> (Fig. 5B). Aire was also expressed from clusters 8 and 10 outside the conventional mTEC<sup>high</sup>, which showed active cell cycling (Fig. 5C). Remarkably, we found that the size of cluster 4 was significantly larger in homozygous 3xAire-KI than in WT, accompanied by a concomitant tendency for smaller sizes of clusters 1 and 5 (Fig. 5D), indicating an altered mTEC<sup>high</sup> heterogeneity by the augmented Aire expression.

We then examined the expression levels of Aire-dependent TRAs (FDR <0.05, FC >1.5, or FC <2/3 between WT and Aire-KO for which the scatterplot is shown in Supplemental Fig. 1A) in these Aire-expressing mTEC<sup>high</sup> clusters 1 and 4 (Supplemental Fig. 2C, combined data from WT and 3xAire-KI; Supplemental Fig. 2D, data from WT alone). We found that expression of Aire-induced TRAs was underexpressed in cluster 4 compared with cluster 1, either when the scRNA-seq data from WT and 3xAire-KI were combined (Fig. 5E, left; and Fig. 5F) or when the scRNA-seq data from WT alone were analyzed (Fig. 5G, left). Accordingly, the increased proportion of cluster 4 in 3xAire-KI (Fig. 5D) resulted in reduced TRAs when the TRA expression was evaluated using pooled mTEC<sup>high</sup> (Fig. 3A). Here, it is important to mention that averaged gene expression (not confined to TRAs) was comparable between WT and 3xAire-KI for the same clusters (i.e., cluster 1 from WT versus cluster 1 from 3xAire-KI; cluster 4 from WT versus cluster 4 from 3xAire-KI), although it was underexpressed in cluster 4 compared with cluster 1 in both WT and 3xAire-KI (Supplemental Fig. 2E). This result further supports our notion that reduced expression of TRAs in pooled mTEC<sup>high</sup> from 3xAire-KI was due to the altered proportion of clusters 1 and 4, not due to the reduced TRAs expressed from each cluster.

In contrast to Aire-dependent TRAs, expression of Aire-independent TRAs in clusters 1 and 4 was indistinguishable (Fig. 5E, right, combined data from WT and 3xAire-KI; Fig. 5G, right, data from WT alone). Accordingly, altered heterogeneity of Aire-expressing mTEC<sup>high</sup> in 3xAire-KI did not affect the expression levels of Aire-independent genes when analyzed using pooled mTEC<sup>high</sup>.

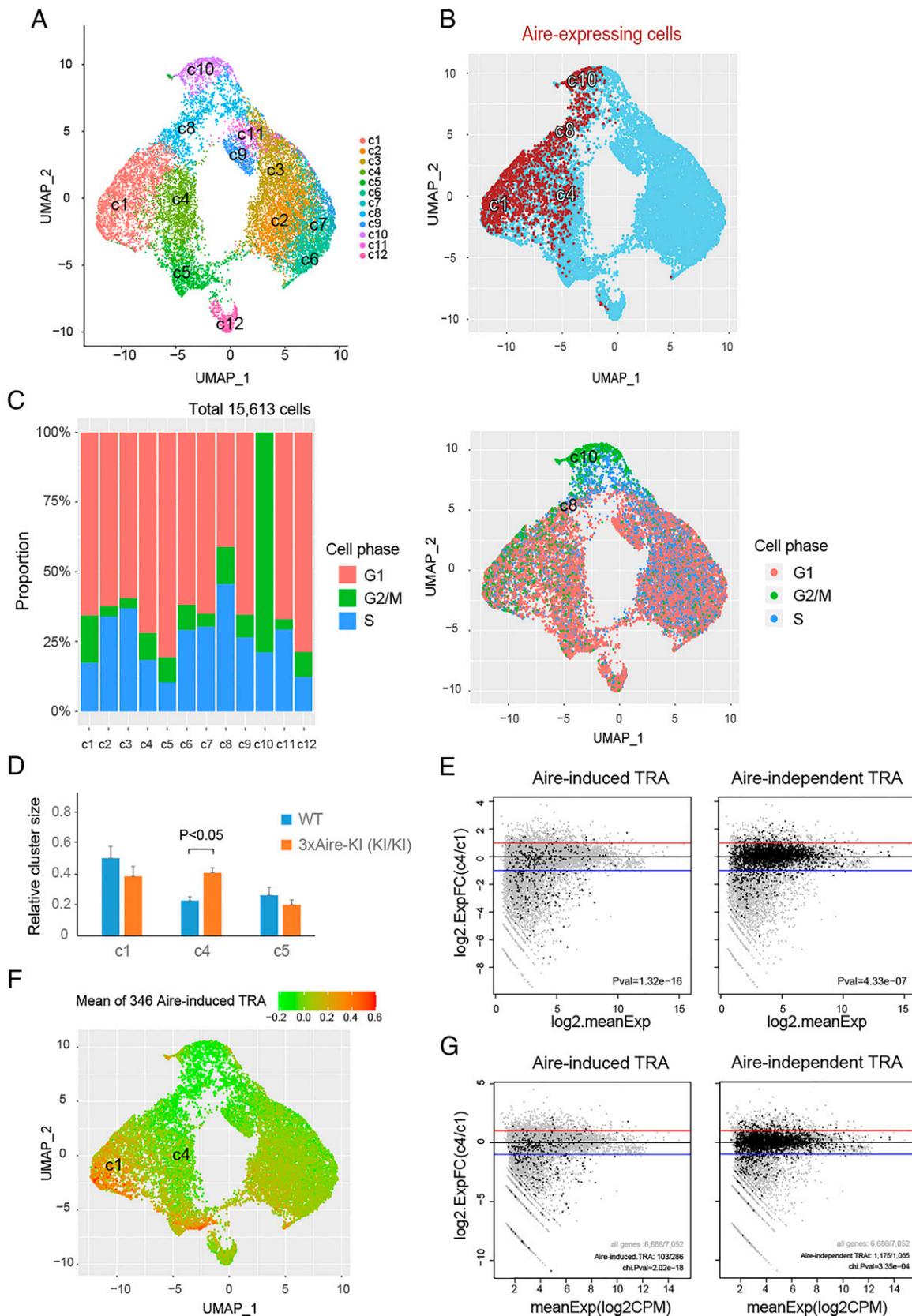
#### *Aire controls the differentiation program of mTECs*

To further characterize Aire-expressing mTEC<sup>high</sup> clusters 1 and 4, we performed GSEA. A signature of epidermal cell differentiation was

significantly enriched in cluster 1 (Fig. 6A and 6B), suggesting that clusters 1 and 4 were developmentally distinct. We then performed RNA velocity analysis to explore the developmental relationship between the clusters (43). Although the transition from mTEC<sup>low</sup> to mTEC<sup>high</sup> and the reverse transition were observed, no transition was evident between clusters 1 and 4 (Fig. 6C). We speculate that these two clusters are formed from a common precursor(s) wherein Aire may control the bifurcation process, thereby controlling the proportion of clusters 1 and 4 in mTEC<sup>high</sup>. Augmented Aire expression resulted in the enhanced differentiation toward cluster 4.

#### *Aire controls heterogeneity of mTECs for the expression of TRAs*

We also investigated the mechanisms underlying the reduced expression of many TRAs in Aire-KO with the same strategy using FACS-sorted CD45<sup>-</sup>EpCAM<sup>+</sup> TECs from both WT and Aire-KO with stringent quality checking (i.e., filtering the cells that have unique feature counts >6000 or <1500, the cells that have >4% mitochondrial counts, and the cells that have unique RNA counts <5000) for four data sets (i.e., WT\_1: 3608 cells; WT\_2: 3125 cells; Aire-KO\_1: 5521 cells; Aire-KO\_2: 5102 cells). There was very little variation within the same group of mice for WT and Aire-KO: two for each (Minoru Matsumoto, H. Yoshida, and Mitsuru Matsumoto, unpublished observation). Because of the version up of the reagent for scRNA-seq analysis, we analyzed the data independently from those obtained from 3xAire-KI. Fifteen clusters emerged (Fig. 7A), including Aire-expressing cluster 5 among CD80<sup>high</sup> (mTEC<sup>high</sup>) clusters (i.e., clusters 1–3, 5–7, and 14) (Fig. 7B), and Aire-expressing cluster 10 outside the conventional mTEC<sup>high</sup> with active cell cycling (Fig. 7C), as described in Fig. 5C. cTECs expressing *Psmb11*, post-Aire cells expressing *Krt10*, and tuft-like cells expressing *Pou2f3* formed clusters 15, 8, and 11, respectively (Fig. 7D). Similar to the case for 3xAire-KI (Fig. 5D), the proportion of mTEC<sup>high</sup> clusters was changed by the loss of Aire (Fig. 8A): the proportion of clusters 1, 2, and 7 was increased, whereas that of clusters 4 and 5 was decreased in Aire-KO. When the expression levels of Aire-induced TRAs were compared between the two cluster groups (i.e., clusters 1, 2, and 7 [overrepresented in Aire-KO] and cluster 5 [evident only in WT]), the former showed much lower levels of Aire-induced TRAs than the latter (Fig. 8B, left), accounting for the reduced expression of TRAs in pooled mTEC<sup>high</sup> from Aire-KO. In contrast, the expression levels of Aire-independent TRAs were comparable between the two groups (Fig. 8B, right). Interestingly, when averaged gene expression (not confined to TRAs) was compared between these two cluster groups, a signature of “keratinocyte differentiation” was identified by GSEA (Fig. 8C and 8D). These results suggested that deficiency of Aire resulted in the reduced expression of TRAs due to the altered heterogeneity of mTECs as we observed for 3xAire-KI. RNA velocity analysis further suggested the transition from cluster 10 with active cell cycle toward cluster 1 (evident only in Aire-KO) and cluster 5 (evident only in WT) (Fig. 8E), wherein the size of cluster 10 was not affected by the loss of Aire (Fig. 8A). This result may further support our idea that Aire controls the bifurcation process of mTEC development. Supporting the Aire-dependent generation of mTEC heterogeneity in both 3xAire-KI and Aire-KO, Gene Ontology analysis of the top 57 Aire-induced genes in group 1 (Fig. 4A) showed an association with “keratinization,” epidermal cell differentiation, and keratinocyte differentiation (Fig. 8F). Thus, we propose that Aire controls the expression of many TRAs indirectly through regulation of mTEC heterogeneity wherein expression of a set of TRAs is inherent in each mTEC subpopulation rather than through Aire’s direct action on the transcription machinery for TRAs. This idea was also supported by the fact that clusters expressing Aire and those expressing Aire-induced genes were not necessarily overlapped



**FIGURE 5.** Altered heterogeneity of mTECs accounting for the paradoxical reduction in the expression of TRAs in Aire-augmented mice. **(A)** UMAP plot for FACS-sorted TECs from WT and homozygous 3xAire-KI. Twelve clusters emerged from the two data sets from WT and homozygous 3xAire-KI for each. **(B)** Aire-expressing mTEC clusters among mTEC<sup>high</sup> (clusters 1 and 4) and outside conventional mTEC<sup>high</sup> (clusters 8 and 10). **(C)** Proportions of cells in the G<sub>1</sub>, G<sub>2</sub>/M, and S phases of the cell cycle in each cluster (left). Estimated cell cycle status was projected onto the UMAP plot shown in A with highlighting of clusters 8 and 10 (right). **(D)** The relative size of each cluster among the three mTEC<sup>high</sup> clusters from WT and homozygous 3xAire-KI. **(E)** Underexpression of Aire-induced TRAs (left), but not Aire-independent TRAs (right), in cluster 4 compared with those in cluster 1. The combined data (*Figure legend continues*)

(compare Aire in Fig. 7B, left, with Aire-induced genes in Fig. 8G). Furthermore, no significant correlation was observed between the Aire and Aire-induced TRAs in terms of the expression levels (Fig. 8H, upper left) and expression frequencies (Fig. 8H, lower left) (Pearson's  $r < 0.1$ ), similarly to the case for Aire-neutral TRA genes (Fig. 8H, right panels).

To verify the quality of our single-cell analyses demonstrated above, our data have been adapted for the subset reported by Bornstein and colleagues (23). Both single-cell data sets (WT versus homozygous 3xAire-KI shown in Fig. 5A and WT versus Aire-KO shown in Fig. 7A) contained cTECs and mTEC-I to mTEC-IV subsets, including the tuft-like cells (Fig. 9A). The proportion of each subset significantly varied in mTEC-I, mTEC-II, and mTEC-III in Aire-KO (Fig. 9A, lower), which is consistent with the results shown in Fig. 8A. Furthermore, we focused on the proportion of post-Aire mTECs using *Krt10* and *Involucrin* as markers, and there was a trend for the reduction of this subset in Aire-KO (Fig. 9B). Thus, altered heterogeneity of mTECs especially caused by the lack of Aire was evident when various kinds of mTEC classification were applied.

#### *Ccl25 is a canonical Aire target possibly contributing to Aire-dependent autoimmune disease*

Finally, we addressed whether the Aire-induced genes found in group 1 are inducible by Aire per se. For this, we chose *Ccl25* from this group because our transcriptomic analyses using both 3xAire-KI and Aire-KO highlighted this chemokine gene as a candidate of the Aire target with high confidence (FDR  $< 0.05$ ). Furthermore, *Ccl25* is one of the four minimal factors required for creating the thymic epithelial niche for hematopoietic progenitor cells downstream of *Foxn1* (46, 47). Expression of *Ccl25* was much higher in mTEC<sup>high</sup> than in mTEC<sup>low</sup> (Fig. 10A), and its Aire-dependent expression was confirmed by real-time PCR (Fig. 10B). We first characterized the promoter region of *Ccl25* by ATAC-seq. We found that the ATAC signal was substantially higher in a region ~3 kb upstream of the TSS (48) in Aire-augmented mTEC<sup>high</sup> from homozygous 3xAire-KI compared with those from WT (Fig. 10C, upper part). We then asked whether this region is a regulatory element for *Ccl25* expression by Aire using a reporter gene assay. For this, we subcloned the corresponding region (−2978/+229 relative to TSS [48]) upstream of the firefly luciferase gene in the pGL3 vector (Supplemental Fig. 3A) and transfected it into a mouse mTEC-like cell line, 1C6 (49), together with an Aire expression vector. As a negative control, we subcloned the same *Ccl25* promoter region into the same vector in a reverse orientation. We found that luciferase activity was induced in a dose-dependent manner with Aire when the *Ccl25* promoter region was in the forward, but not the reverse, orientation (Fig. 10D), indicating that *Ccl25* is a bona fide Aire-induced gene. Consistent with these findings, Aire bound to the *Ccl25* promoter region in mTECs according to the published CHIP-seq data (39) (Fig. 10C, lower part).

Given that *Ccl25* functions as a downstream target of Aire in mTECs, we examined any possible contribution of the reduced expression of *Ccl25* to the disease phenotypes in Aire-KO using *Ccl25*-deficient mice (*Ccl25*-KO) (50). We found that the *Ccl25*-KO had a significant degree of inflammatory cell infiltrations in the salivary gland and kidney (Fig. 10E), but no production of

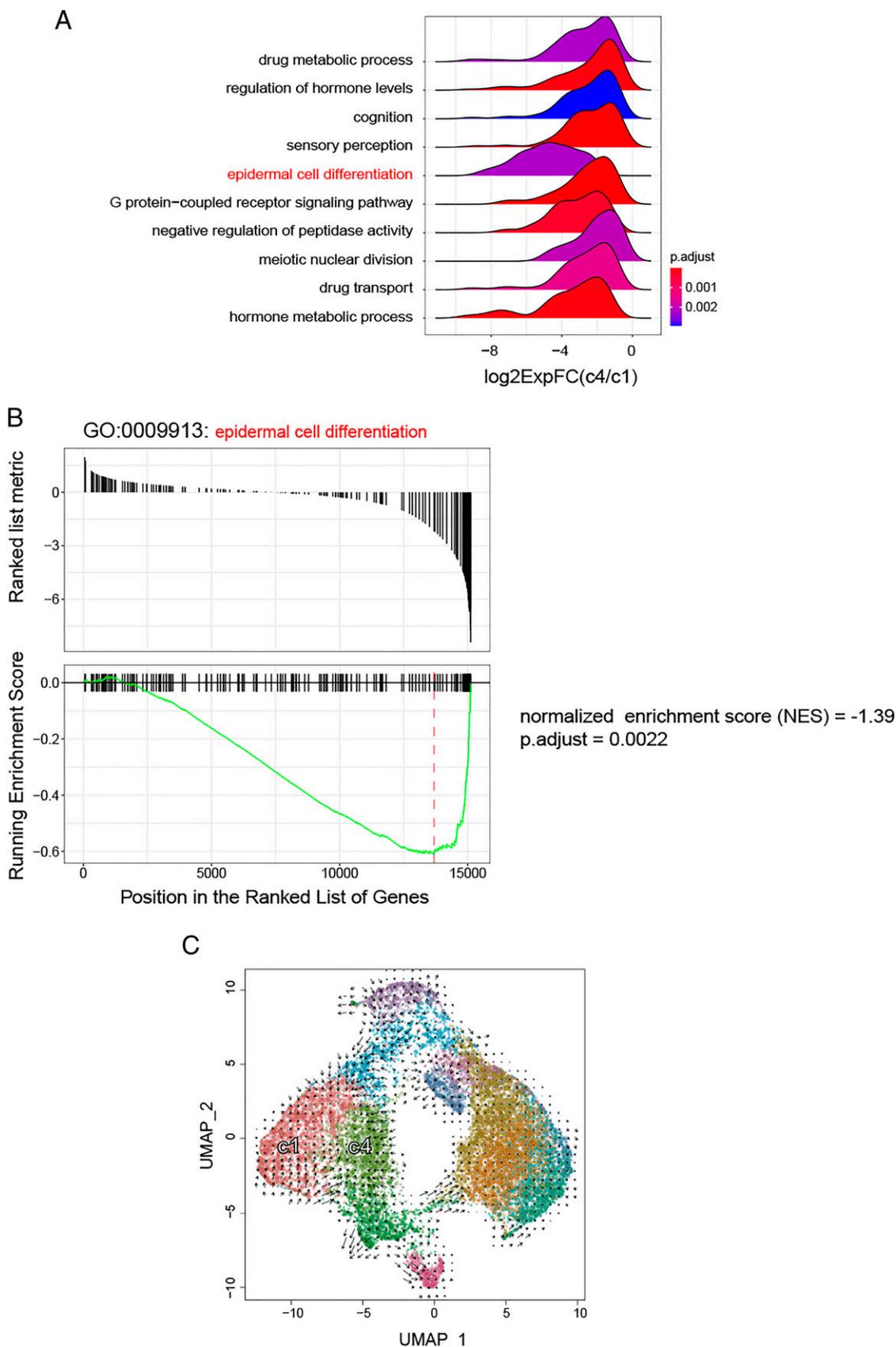
autoantibodies against these organs was evident (data not shown). In *Ccl25*-KO, expression of *Aire*, two prototypic Aire-dependent TRAs (i.e., *Ins2* and *Sap1*), and an Aire-independent TRA (i.e., *Crp*) from mTEC<sup>high</sup> remained unchanged (Supplemental Fig. 3B) with normal development of Aire<sup>+</sup> mTECs (Supplemental Fig. 3C). This suggested that the role of *Ccl25* for the disease phenotypes in Aire-KO is modifying (e.g., induction of inflammatory response) rather than initiating the autoimmune process.

## Discussion

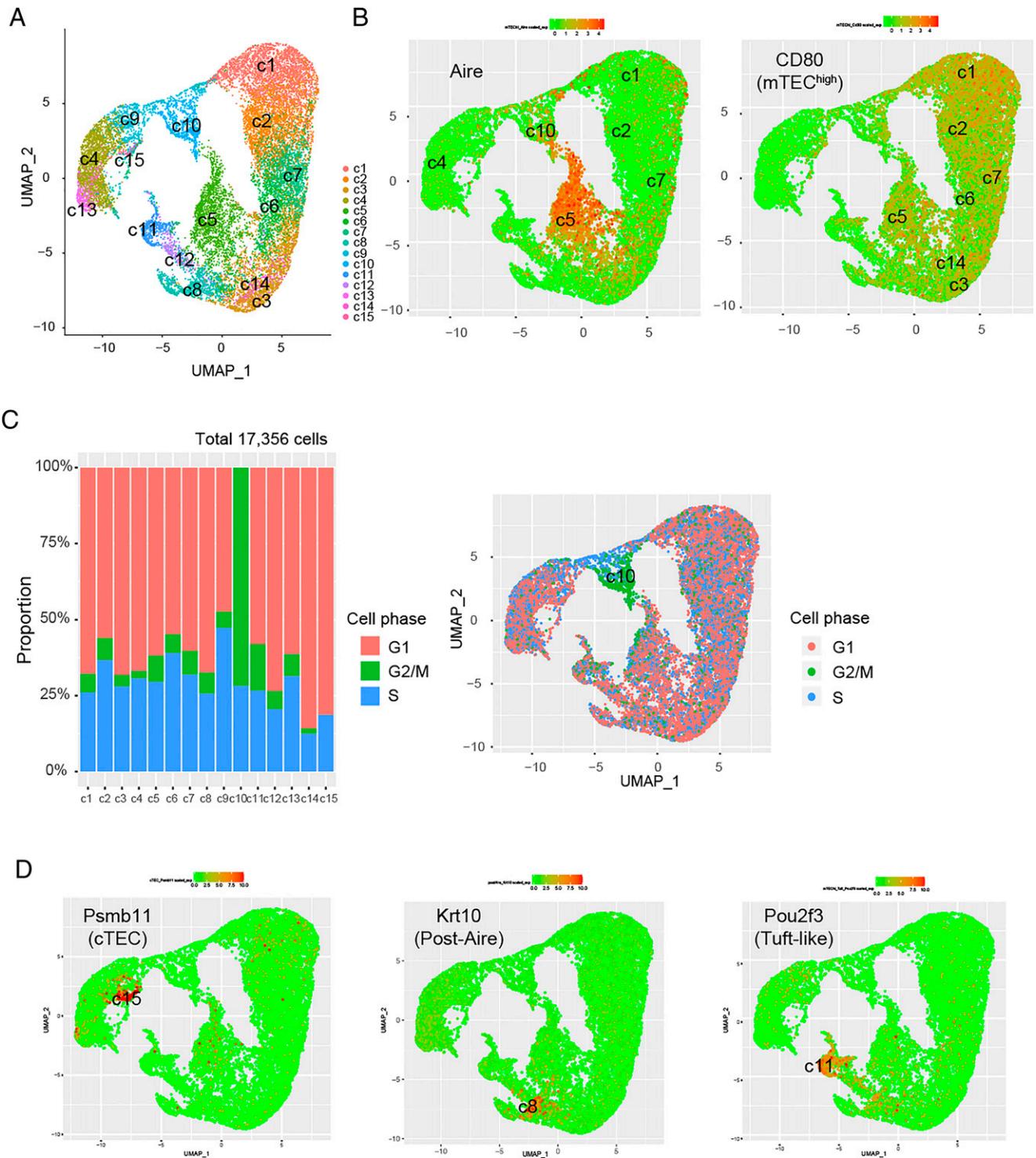
In the present study, we addressed a fundamental question regarding how mTECs acquire the capacity for promiscuous gene expression with the participation of Aire, with the hope that understanding the issue will help to unravel the molecular mechanisms responsible for the Aire-dependent thymic tolerance and its defect leading to the development of autoimmunity. Identification of the Aire target genes would be one of the most straightforward approaches to reach this goal. Because the previous transcriptomic analyses of mTECs to illustrate Aire's targets had been performed using mTECs lacking Aire alone, we additionally obtained the transcriptomic data from Aire-augmented mTECs to draw a coherent picture of the candidates. We were particularly careful about expressing the additive Aire from authentic Aire-expressing mTECs because the transcriptional impact of Aire varies with cell type (25). Unexpectedly, many TRAs downregulated in Aire-KO were also downregulated in Aire-augmented mTECs, and integration of the transcriptomic data from Aire-augmented and Aire-KO mTECs revealed that canonical Aire-induced genes found in group 1 contained rather small numbers of TRA genes. Interestingly, Aire-induced genes in this group were involved in the differentiation program of mTECs (i.e., keratinization, epidermal cell differentiation, and keratinocyte differentiation), which may be consistent with the previous notion that there were morphological changes in Aire-KO thymic stromal cells: globular mTECs without visible cellular projections were more prominent in the Aire-KO thymus (51). Altered distribution of mTECs committed to express Aire was also noted (52). Thus, our results underscored the role of Aire in the differentiation program of mTECs from a novel viewpoint (14).

Our single-cell analyses of mTECs from Aire-augmented and Aire-KO mice revealed altered heterogeneity of mTECs in both cases. Given that acquisition of the properties of TRA expression in mTECs depends on the maturation status of mTECs (52–54), perturbation of the maturation process, either by the augmented Aire expression or by the loss of Aire, might account for the impaired expression of TRAs. Although the changes in the mTEC<sup>high</sup> clusters from Aire-augmented and Aire-KO mice were not simply mirror imaged, and although the degree of the alteration seems to be more robust in Aire-KO mice than in Aire-augmented mice, these findings may be reasonable when considering that the presence of Aire is secured in the Aire-augmented mice, whereas Aire is completely absent in the Aire-KO mice. Thus, the expression of high levels of TRAs is ensured by the controlled mTEC differentiation program with the appropriate level of Aire expression. As for the paradoxical reduction in the expression of TRAs in Aire-augmented mice, we cannot completely exclude a possibility that feedback mechanisms evoked by the overexpression of Aire play some role in the transcriptional suppression of TRAs (55). Rate limiting of

from WT and homozygous 3xAire-KI shown in Supplemental Fig. 2C were used for the analysis. (F) Plot of the mean of 346 Aire-induced TRAs predominantly expressed from cluster 1 corresponding to E, left. (G) Underexpression of Aire-induced TRAs (left), but not Aire-independent TRAs (right), in cluster 4 compared with those in cluster 1. Data from WT alone shown in Supplemental Fig. 2D were used for the analysis.



**FIGURE 6.** mTEC differentiation program controlled by Aire. **(A)** Ridgeline plot for the expression distributions of GSEA. A signature of epidermal cell differentiation was significantly enriched in cluster 1. **(B)** GSEA enrichment plots for the epidermal cell differentiation gene set. Representative gene sets from GSEA for clusters 1 and 4 analyzed in Supplemental Fig. 2C are shown in A. GO, Gene Ontology. **(C)** RNA velocity analysis using the data from WT and 3xAire-KI. No obvious transition was observed between clusters 1 and 4.

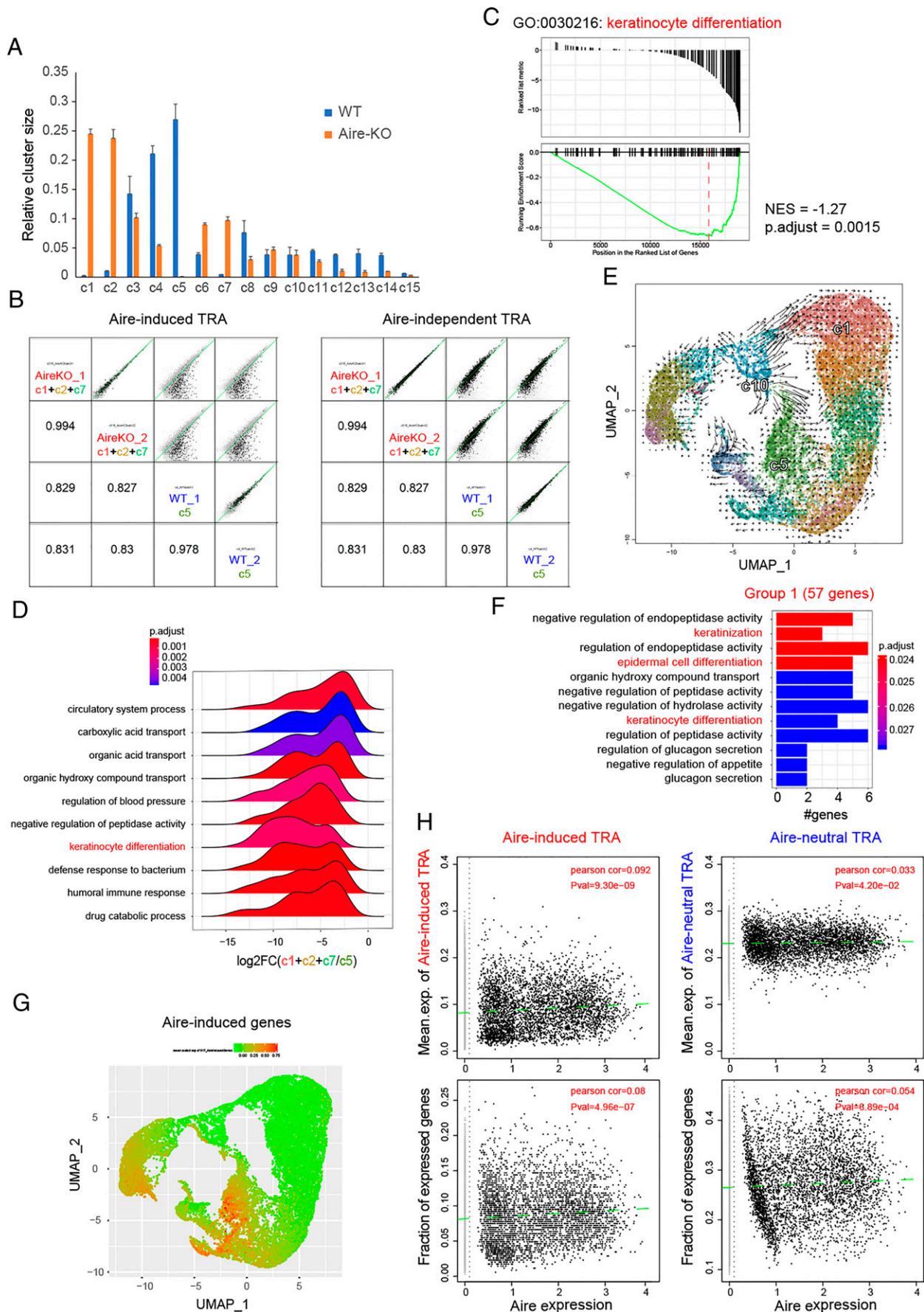


**FIGURE 7.** Single-cell analysis with Aire-KO mTECs. **(A)** UMAP plot for FACS-sorted TECs from WT and Aire-KO. Fifteen clusters emerged from the two data sets from WT and Aire-KO for each. **(B)** Expression of Aire (left) and CD80 (right) for TEC clusters. **(C)** Estimated cell cycle status of each TEC cluster (left). Proportions of cells in the G<sub>1</sub>, G<sub>2</sub>/M, and S phases of the cell cycle in each cluster. Cluster 10 contained a high proportion of cells in the active G<sub>2</sub>/M phase (right). **(D)** Expression of three additional representative genes for TEC clusters shown in B. The color scale above the plot illustrates the normalized expression values for each gene.

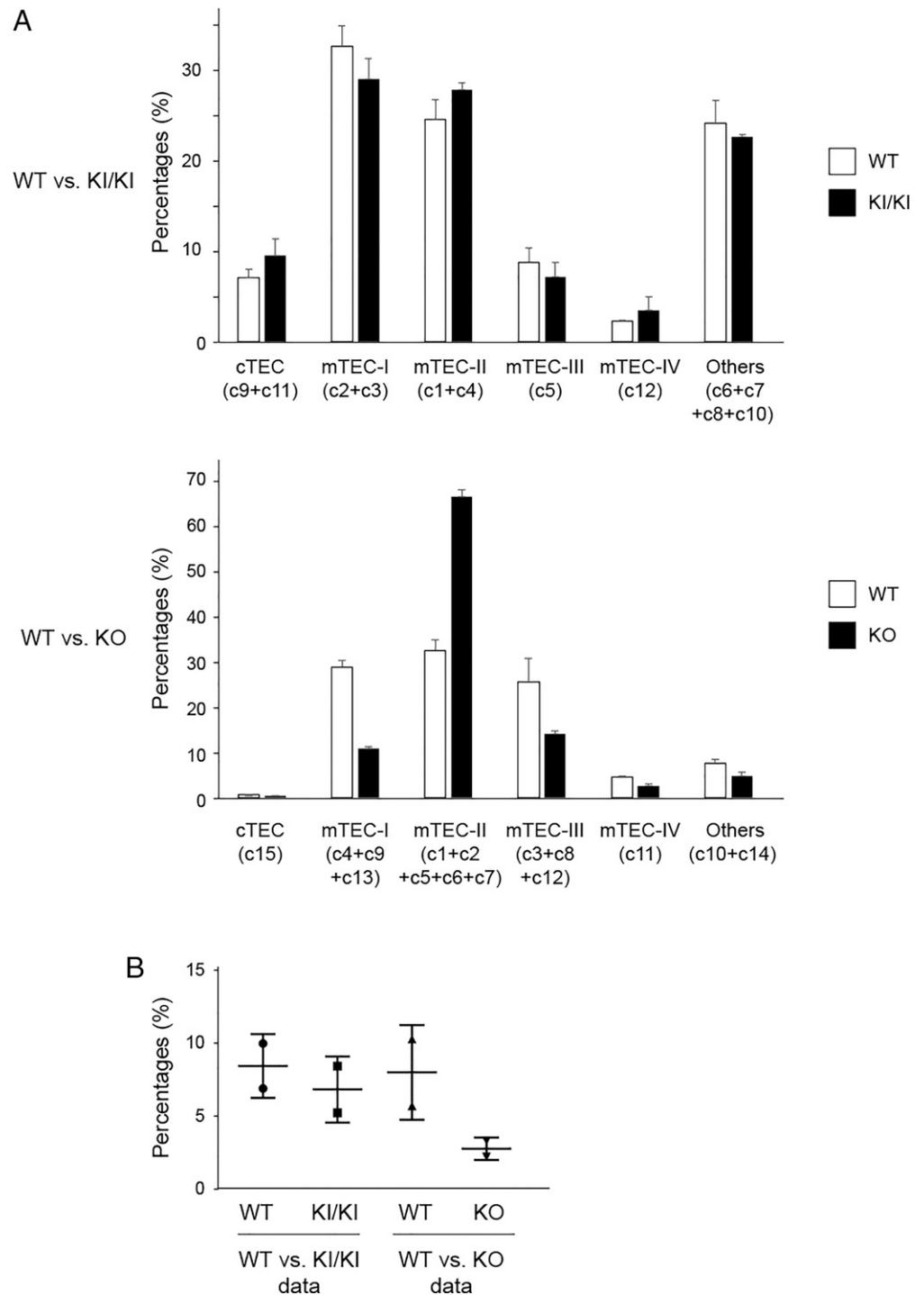
Aire's partner proteins might be another possibility for the downregulation of TRAs in Aire-augmented mice. Further studies on the molecular action of Aire are required to solve these issues.

So far, most studies using single-cell analyses of TECs have been conducted using normal mice, and the factors affecting the composition of the mTEC subpopulations remain largely unknown. In this regard, there is one study performing a single-cell analysis using the

TECs isolated from Aire-KO (22). The study was intended to determine whether the downregulation of transcripts in Aire-KO mTECs was due either to the decrease in the amount of transcript per cell or to the decrease in the proportion of cells expressing the transcript. Because of the underpowering of the sample size (i.e., 100 TECs from WT or Aire-KO for each), clustering of mTEC<sup>high</sup> into subpopulations was not possible. Accordingly, the altered heterogeneity of mTECs in Aire-KO was not revealed in this study. In



**FIGURE 8.** mTEC heterogeneity controlled by Aire for the expression of TRAs. **(A)** The relative size of each TEC cluster from WT and Aire-KO. **(B)** An averaged gene expression matrix for differentially represented clusters from WT and Aire-KO: clusters 1, 2, and 7 from Aire-KO versus cluster 5 from WT. Averaged gene expression not confined to TRAs is shown in gray, and that for Aire-induced TRAs (left) and Aire-independent (*Figure legend continues*)

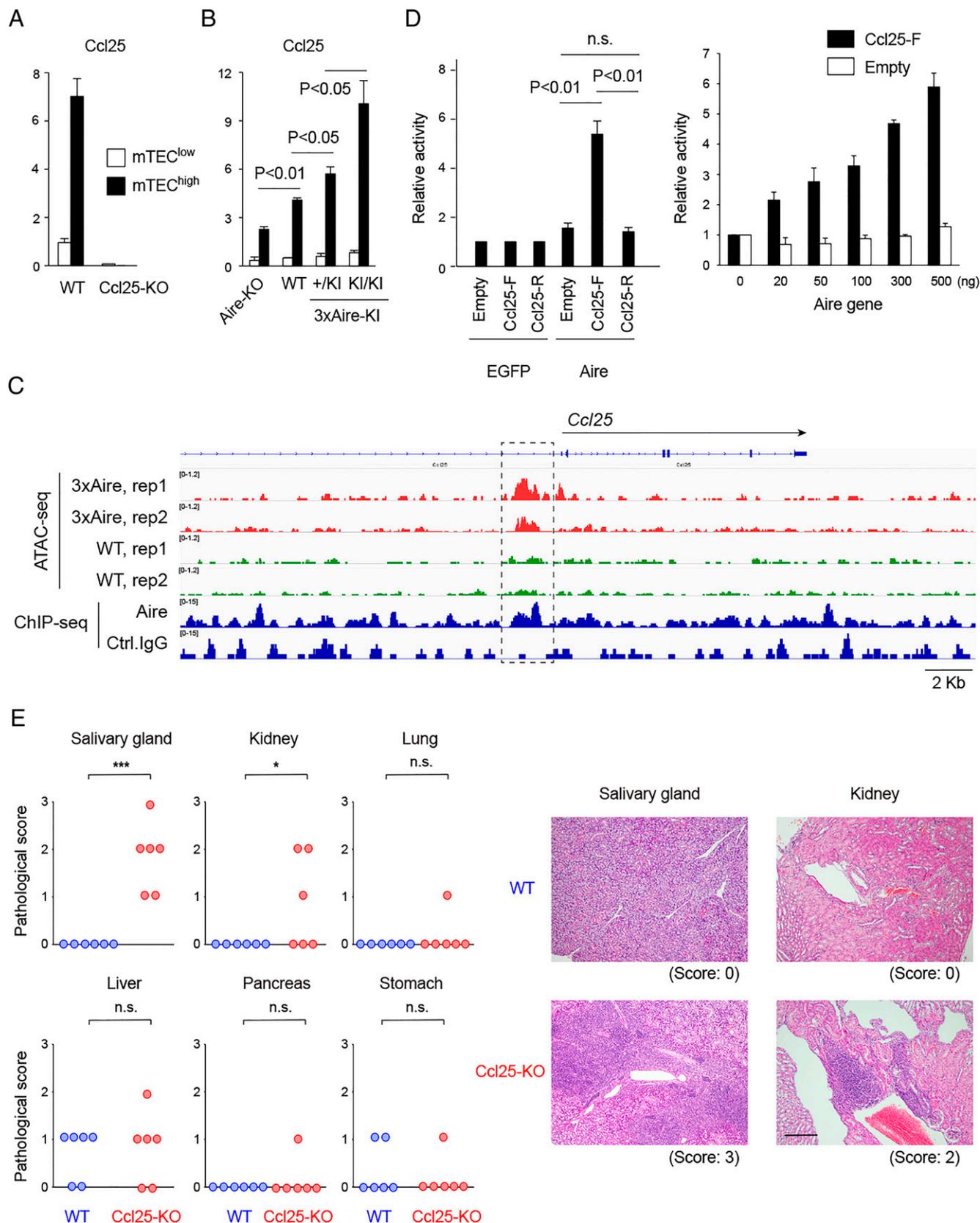


**FIGURE 9.** Aire controls heterogeneity of mTECs. **(A)** Both single-cell data sets (WT versus homozygous 3xAire-KI shown in Fig. 5A [upper] and WT versus Aire-KO shown in Fig. 7A [lower]) were adapted for the subset reported by Bornstein and colleagues (23). Marker genes for each subset are (in parentheses) as follows: mTEC-I (*Itga4* and *Itga6*), mTEC-II (*Aire* and *H2-Aa*), mTEC-III (*Ly6d* and *Pigr*), mTEC-IV representing tuft-like cells (*Lcam* and *Sox9*), and cTEC (*Ly75* and *Psm11*). **(B)** Percentages of post-Aire mTEC clusters expressing *Krt10* and *Involucrin* were calculated and plotted using two mice for each analysis described in A. One mark corresponds to one mouse analyzed.

contrast to this small-scale study, we have analyzed several thousands of TECs obtained from two sets of 3xAire-KI or Aire-KO together with two sets of WT for each experiment. It is also

important to mention that the batch effect among the individual mice was minimal in our study, thereby featuring the altered composition of mTEC<sup>high</sup> in both 3xAire-KI and Aire-KO mice. Thus,

TRAs (right) is highlighted in black. The correlation coefficient for each comparison is indicated. **(C)** GSEA enrichment plots of a representative gene set. GSEA for differentially represented cluster groups from WT and Aire-KO (clusters 1, 2, and 7 from Aire-KO versus cluster 5 from WT) identifying a signature of keratinocyte differentiation. **(D)** Ridgeline plot for the expression distributions of GSEA for two cluster groups analyzed in B (i.e., clusters 1, 2, and 7 from Aire-KO versus cluster 5 from WT). A signature of keratinocyte differentiation was identified. **(E)** RNA velocity analysis using the data from WT and Aire-KO. The transition from cluster 10 toward cluster 1 and cluster 5, which was evident only in Aire-KO and WT, respectively (A), was observed. **(F)** The plot of the significantly enriched Gene Ontology (GO) terms for the top 57 Aire-induced genes in group 1 (shown in Fig. 4A). **(G)** Mean expression of Aire-induced genes (not confined to TRAs) projected onto the UMAP plot shown in Fig. 7A. **(H)** The Pearson correlations between *Aire* and Aire-induced TRAs in terms of the expression levels (upper left) and expression frequencies (lower left) were computed and depicted on the top right in each plot (left panels). Plots for Aire-neutral TRA genes served as control (right panels).



**FIGURE 10.** Contribution of *Ccl25* to the development of Aire-dependent autoimmune disease. **(A)** Expression of *Ccl25* from mTEC<sup>low</sup> and mTEC<sup>high</sup> in WT. The data were obtained from three repeated experiments using mice at 4 wk of age. **(B)** Levels of *Ccl25* expression from isolated mTEC<sup>low</sup> and mTEC<sup>high</sup> in Aire-KO, WT, heterozygous, and homozygous 3x Aire-KI were examined by real-time PCR. The data were obtained from three repeated experiments using mice at 4 wk of age. **(C)** ATAC-seq from two experimental replicates using mTEC<sup>high</sup> from homozygous 3x Aire-KI and WT on a NOD background are shown in the upper rows. Pile-up traces of Aire and control (IgG) ChIP-seq signals for the *Ccl25* locus from the Gene Expression Omnibus data set (accession no. GSE92654) are shown in the lower rows. **(D)** *Ccl25* reporter gene assay. The mTEC-like cell line was transfected with the *Ccl25* reporter genes shown in Supplemental Fig. 3A. Left: Luciferase activity from the cells transfected with indicated reporter vectors and (Figure legend continues)

the size of the sample and its quality matter when pursuing single-cell biology.

Single-cell analyses in this study were performed using mice that were all on a B6 background, whereas transcriptomic comparison of sorted mTEC<sup>high</sup> between 3xAire-KI and Aire-KO mice was done using mice on different genetic backgrounds: 3xAire-KI (B6) and Aire-KO (NOD). Although we have confirmed a high correlation between the log<sub>2</sub> FC values from Aire-KO on NOD of our own study (7) and those from publicly available data for Aire-KO on B6 (6) ( $r^2 = 0.558$ ;  $p < 2.2 \times 10^{-16}$ ; Supplemental Fig. 1B), a direct comparison of the data set on the same genetic background will be necessary in future studies. The same applies to the ATAC-seq that was performed using 3xAire-KI on NOD because of the availability of mice (Fig. 10C). Here, we have selected the genes with FCs without employing an FDR <0.05 for Fig. 4A: four groups of genes were selected based on FCs (log<sub>2</sub> FC >0.5 and log<sub>2</sub> FC less than -0.5). Although it might be ideal to select the genes for each group by setting the FDR, we intended to maximize the number of genes for the later analyses to increase the reliability of statistics. We think that it may not be easy to decide what FCs should be applied to log<sub>2</sub> FC in this kind of analysis.

Altered heterogeneity in the disease was evidenced by the single-cell analysis of human oligodendrocytes in multiple sclerosis: some subclusters were underrepresented in the multiple sclerosis tissues, whereas others were more prevalent (56). In this case, an increase in the transcriptional programs responsible for myelination from mature oligodendrocytes has been implicated in the disease pathogenesis. However, the factors causing the altered heterogeneity of oligodendrocytes in multiple sclerosis remained unknown. Here, we have revealed that altered heterogeneity of mTECs was caused by the gene-dosage effect of Aire that was associated with the reduced expression of many TRAs leading to the development of autoimmunity. We speculate that a single-cell approach to studying other diseased tissues will unravel the pathogenesis of a broad spectrum of disorders before too long.

Because Aire is almost exclusively detected from mature mTECs with flow cytometric analysis, Aire expression from outside conventional mTEC<sup>high</sup> showing active cell cycling was unexpected. The significance of this population has recently been debated by other studies (17, 18, 57). The size of this cluster (cluster 10 defined by the single-cell analysis using WT and Aire-KO) was not affected without Aire. Based on the RNA velocity analysis, we speculate that cluster 1 (evident only in Aire-KO) and cluster 5 (evident only in WT) in mTEC<sup>high</sup> are formed from this cluster wherein Aire may control the bifurcation process, thereby controlling the composition of mTEC<sup>high</sup> clusters. This view may also apply to the altered composition of mTEC<sup>high</sup> in homozygous 3xAire-KI. Consistent with these ideas, Wells and colleagues suggested that this active cell-cycling population is a transit-amplifying population immediately preceding the Aire-expressing mTEC<sup>high</sup> (17). Identification of the specific cell markers for the cluster is required to better characterize this subpopulation by introducing a lineage-tracing experiment.

Our study does not address the question regarding how each mTEC subpopulation acquires the ability to express a set of unique TRAs. In other words, how each mTEC subpopulation selects the members of TRAs to be expressed (i.e., repertoire formation for TRAs) remains unsolved in the present study.

This is primarily due to the insufficient sensitivity for RNA detection from every single cell using the 10x Genomics Chromium System. However, the fact that the expression level of TRAs from each mTEC<sup>high</sup> cluster remained unchanged in 3xAire-KI and Aire-KO suggested that Aire has no major impact on the generation of the TRA repertoire within each cluster. Instead, Aire may be orchestrating the proportion of each cluster size through acting on the mTEC differentiation process, thereby avoiding making any “repertoire holes” of self-Ags in the negative selection niche of the medulla. A recent study suggesting that different maturational stages or classes of mTECs activate TRA expression differently might be consistent with this view (18).

Although single-cell analysis has provided us with a better understanding of the heterogeneity of mTECs as demonstrated in this study, there are some technical limitations. For example, genes expressed at low levels may be missed during the cell clustering (dropout phenomenon). Conversely, the expression level of Aire in each mTEC transcribed from the 3xAire-KI locus is not well reflected in the analysis, which inevitably occurs during the sequencing of the 3' ends of cDNAs. Because each TRA is generally expressed at low frequencies among mTECs, this problem, rather unique to mTECs, also needs to be overcome. In this regard, information gathered from scRNA-seq alone may not be sufficient to identify the precise mTEC subpopulations. Accordingly, the combination with other omics approaches (i.e., single-cell multiomics technologies [58] such as scRNA-seq combined with single-cell ATAC-seq [59]) will be required to gain more in-depth TEC biological information.

In contrast to our results, there is one study reporting that Aire deficiency resulted in the upregulation of Ccl25 from mTECs (60). We confirmed Aire-dependent Ccl25 expression using both Aire-KO and Aire-augmented mice. Furthermore, we have demonstrated that Aire controls the expression of Ccl25 as a canonical target downstream of Aire using ATAC-seq, CHIP-seq, and a reporter gene assay. The reason for the discrepancy between the two studies remains unknown.

Although Ccl25-deficient mice showed inflammatory changes in the salivary gland and kidney, Ccl25 deficiency alone showed no obvious signs of autoimmunity. Because Ccl25 does not affect the expression of TRAs and differentiation programs of mTECs, other autoimmune mechanisms beyond the Aire–Ccl25 axis are contributing to the full development of the organ-specific autoimmune disease in Aire deficiency. It is also possible that reduced Ccl25 expression is contributing to the target organ specificity after an Aire-dependent autoimmune trigger. The exact pathological role of reduced Ccl25 expression from Aire-KO mTECs in the development of organ-specific autoimmune disease awaits further study.

In conclusion, our study using the transcriptomic analyses of mTECs with conventional RNA-seq together with high-performance scRNA-seq by introducing a novel strain of Aire-augmented mice has revealed (to our knowledge) previously unrecognized transcriptional regulation by Aire in mTECs. Despite the importance of TRAs in Aire-mediated autoimmune pathogenesis, we suggest that Aire contributes to the expression of a large proportion of TRAs indirectly through controlling mTEC heterogeneity. Although the exact Aire targets that control the differentiation program of mTECs await further study, the role of Aire in organizing mTEC heterogeneity highlighted Aire's multitasking in immune tolerance (61). We also suggest that our present bioinformatics approach using mice not only deficient for

Aire-expressing vector. The value of the EGFP-expressing vector was determined as 1. Right: Reporter gene activities for Ccl25-F and the empty reporter upon increasing the amount of the Aire gene (500 ng of EGFP- and Aire-expressing vectors in total). Luciferase activities from the cells transfected with the EGFP-expressing vector alone were determined as 1. The data were obtained from three repeated experiments. (E) Organ-specific inflammatory lesions in WT and Ccl25-KO. Representative salivary gland and kidney pathologies with the corresponding scores are shown on the right. Scale bar, 200  $\mu$ m.

but also with augmented expression of a given gene product can be applied to searches for the primary targets of other transcription factors by distinguishing them from secondary targets.

## Acknowledgments

We thank F. Hirota for technical assistance.

## Disclosures

The authors declare that they have no conflict of interest.

## References

- Proekt, I., C. N. Miller, M. S. Lionakis, and M. S. Anderson. 2017. Insights into immune tolerance from AIRE deficiency. *Curr. Opin. Immunol.* 49: 71–78.
- Mathis, D., and C. Benoist. 2009. Aire. *Annu. Rev. Immunol.* 27: 287–312.
- Peterson, P., T. Org, and A. Rebane. 2008. Transcriptional regulation by AIRE: molecular mechanisms of central tolerance. *Nat. Rev. Immunol.* 8: 948–957.
- Anderson, M. S., E. S. Venanzi, L. Klein, Z. Chen, S. P. Berzins, S. J. Turley, H. von Boehmer, R. Bronson, A. Dierich, C. Benoist, and D. Mathis. 2002. Projection of an immunological self shadow within the thymus by the Aire protein. *Science* 298: 1395–1401.
- Sansom, S. N., N. Shikama-Dorn, S. Zhanybekova, G. Nusspaumer, I. C. Macaulay, M. E. Deadman, A. Heger, C. P. Ponting, and G. A. Holländer. 2014. Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. *Genome Res.* 24: 1918–1931.
- St-Pierre, C., A. Trofimov, S. Brochu, S. Lemieux, and C. Perreault. 2015. Differential features of AIRE-induced and AIRE-independent promiscuous gene expression in thymic epithelial cells. *J. Immunol.* 195: 498–506.
- Nishijima, H., T. Kajimoto, Y. Matsuoka, Y. Mouri, J. Morimoto, M. Matsumoto, H. Kawano, Y. Nishioka, H. Uehara, K. Izumi, et al. 2018. Paradoxical development of polyomysitis-like autoimmunity through augmented expression of autoimmune regulator (AIRE). *J. Autoimmun.* 86: 75–92.
- Derbinski, J., J. Gäbler, B. Brors, S. Tierling, S. Jonnakuty, M. Hergenhan, L. Peltonen, J. Walter, and B. Kyewski. 2005. Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. *J. Exp. Med.* 202: 33–45.
- Kyewski, B., and L. Klein. 2006. A central role for central tolerance. *Annu. Rev. Immunol.* 24: 571–606.
- Tomofuji, Y., H. Takaba, H. I. Suzuki, R. Benlaribi, C. D. P. Martinez, Y. Abe, Y. Morishita, T. Okamura, A. Taguchi, T. Kodama, and H. Takayanagi. 2020. Chd4 choreographs self-antigen expression for central immune tolerance. *Nat. Immunol.* 21: 892–901.
- Akiyoshi, H., S. Hatakeyama, J. Pitkänen, Y. Mouri, V. Doucas, J. Kudoh, K. Tsurugaya, D. Uchida, A. Matsushima, K. Oshikawa, et al. 2004. Subcellular expression of autoimmune regulator is organized in a spatiotemporal manner. *J. Biol. Chem.* 279: 33984–33991.
- Klein, L., B. Kyewski, P. M. Allen, and K. A. Hogquist. 2014. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat. Rev. Immunol.* 14: 377–391.
- Abramson, J., and G. Anderson. 2017. Thymic epithelial cells. *Annu. Rev. Immunol.* 35: 85–118.
- Matsumoto, M. 2011. Contrasting models for the roles of Aire in the differentiation program of epithelial cells in the thymic medulla. *Eur. J. Immunol.* 41: 12–17.
- Villasenor, J., W. Besse, C. Benoist, and D. Mathis. 2008. Ectopic expression of peripheral-tissue antigens in the thymic epithelium: probabilistic, monoallelic, misinitiated. *Proc. Natl. Acad. Sci. USA* 105: 15854–15859.
- Danso-Abeam, D., K. A. Staats, D. Franckaert, L. Van Den Bosch, A. Liston, D. H. Gray, and J. Dooley. 2013. Aire mediates thymic expression and tolerance of pancreatic antigens via an unconventional transcriptional mechanism. *Eur. J. Immunol.* 43: 75–84.
- Wells, K. L., C. N. Miller, A. R. Gschwind, W. Wei, J. D. Phipps, M. S. Anderson, and L. M. Steinmetz. 2020. Combined transient ablation and single-cell RNA-sequencing reveals the development of medullary thymic epithelial cells. *eLife* 9: e60188.
- Dhalla, F., J. Baran-Gale, S. Maio, L. Chappell, G. A. Holländer, and C. P. Ponting. 2020. Biologically indeterminate yet ordered promiscuous gene expression in single medullary thymic epithelial cells. *EMBO J.* 39: e101828.
- Kernfeld, E. M., R. M. J. Genga, K. Neherin, M. E. Magaletta, P. Xu, and R. Maehr. 2018. A single-cell transcriptomic atlas of thymus organogenesis resolves cell types and developmental maturation. *Immunity* 48: 1258–1270.e6.
- Miragaia, R. J., X. Zhang, T. Gomes, V. Svensson, T. Ilicic, J. Henriksson, G. Kar, and T. Lönnberg. 2018. Single-cell RNA-sequencing resolves self-antigen expression during mTEC development. *Sci. Rep.* 8: 685.
- Brennecke, P., A. Reyes, S. Pinto, K. Rattay, M. Nguyen, R. Küchler, W. Huber, B. Kyewski, and L. M. Steinmetz. 2015. Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. *Nat. Immunol.* 16: 933–941.
- Meredith, M., D. Zemmour, D. Mathis, and C. Benoist. 2015. Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nat. Immunol.* 16: 942–949.
- Bornstein, C., S. Nevo, A. Giladi, N. Kadouri, M. Pouzolles, F. Gerbe, E. David, A. Machado, A. Chuprin, B. Tóth, et al. 2018. Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. *Nature* 559: 622–626.
- Miller, C. N., I. Proekt, J. von Moltke, K. L. Wells, A. R. Rajpurkar, H. Wang, K. Rattay, I. S. Khan, T. C. Metzger, J. L. Pollack, et al. 2018. Thymic tuft cells promote an IL-4-enriched medulla and shape thymocyte development. *Nature* 559: 627–631.
- Guerau-de-Arellano, M., D. Mathis, and C. Benoist. 2008. Transcriptional impact of Aire varies with cell type. *Proc. Natl. Acad. Sci. USA* 105: 14011–14016.
- Kim, J. H., S. R. Lee, L. H. Li, H. J. Park, J. H. Park, K. Y. Lee, M. K. Kim, B. A. Shin, and S. Y. Choi. 2011. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One* 6: e18556.
- Niki, S., K. Oshikawa, Y. Mouri, F. Hirota, A. Matsushima, M. Yano, H. Han, Y. Bando, K. Izumi, M. Matsumoto, et al. 2006. Alteration of intra-pancreatic target-organ specificity by abrogation of Aire in NOD mice. *J. Clin. Invest.* 116: 1292–1301.
- Wurbel, M. A., M. Malissen, D. Guy-Grand, E. Meffre, M. C. Nussenzweig, M. Richelme, A. Carrier, and B. Malissen. 2001. Mice lacking the CCR9 CC-chemokine receptor show a mild impairment of early T- and B-cell development and a reduction in T-cell receptor  $\gamma\delta^+$  gut intraepithelial lymphocytes. *Blood* 98: 2626–2632.
- Corces, M. R., J. D. Buenrostro, B. Wu, P. G. Greenside, S. M. Chan, J. L. Koenig, M. P. Snyder, J. K. Pritchard, A. Kundaje, W. J. Greenleaf, et al. 2016. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* 48: 1193–1203.
- Buenrostro, J. D., B. Wu, U. M. Litzenburger, D. Ruff, M. L. Gonzalez, M. P. Snyder, H. Y. Chang, and W. J. Greenleaf. 2015. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523: 486–490.
- Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. Fastq: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34: i884–i890.
- Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357–359.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin; 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoutte, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9: R137.
- Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.
- Anders, S., P. T. Pyl, and W. Huber. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166–169.
- Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.
- Yu, G., L. G. Wang, Y. Han, and Q. Y. He. 2012. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* 16: 284–287.
- Bansal, K., H. Yoshida, C. Benoist, and D. Mathis. 2017. The transcriptional regulator Aire binds to and activates super-enhancers. *Nat. Immunol.* 18: 263–273.
- Lawrence, M., R. Gentleman, and V. Carey. 2009. Rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics* 25: 1841–1842.
- Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, and V. J. Carey. 2013. Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* 9: e1003118.
- Stuart, T., A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck 3rd, Y. Hao, M. Stoeckius, P. Smibert, and R. Satija. 2019. Comprehensive integration of single-cell data. *Cell* 177: 1888–1902.E21.
- La Manno, G., R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastrii, P. Lönnberg, A. Furlan, et al. 2018. RNA velocity of single cells. *Nature* 560: 494–498.
- McCarthy, D. J., Y. Chen, and G. K. Smyth. 2012. Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. *Nucleic Acids Res.* 40: 4288–4297.
- Giraud, M., H. Yoshida, J. Abramson, P. B. Rahl, R. A. Young, D. Mathis, and C. Benoist. 2012. Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. *Proc. Natl. Acad. Sci. USA* 109: 535–540.
- Zuklys, S., A. Handel, S. Zhanybekova, F. Govani, M. Keller, S. Maio, C. E. Mayer, H. Y. Teh, K. Hafen, G. Gallone, et al. 2016. Foxn1 regulates key target genes essential for T cell development in postnatal thymic epithelial cells. *Nat. Immunol.* 17: 1206–1215.
- Caldéron, L., and T. Boehm. 2012. Synergistic, context-dependent, and hierarchical functions of epithelial components in thymic microenvironments. *Cell* 149: 159–172.
- Ericsson, A., K. Kotarsky, M. Svensson, M. Sigvardsson, and W. Agace. 2006. Functional characterization of the CCL25 promoter in small intestinal epithelial cells suggests a regulatory role for caudal-related homeobox (Cdx) transcription factors. *J. Immunol.* 176: 3642–3651.
- Kasai, M., K. Hirokawa, K. Kajino, K. Ogasawara, M. Tatsumi, E. Hermel, J. J. Monaco, and T. Mizuuchi. 1996. Difference in antigen presentation pathways between cortical and medullary thymic epithelial cells. *Eur. J. Immunol.* 26: 2101–2107.
- Wurbel, M. A., M. Malissen, D. Guy-Grand, B. Malissen, and J. J. Campbell. 2007. Impaired accumulation of antigen-specific CD8 lymphocytes in chemokine CCL25-deficient intestinal epithelium and lamina propria. *J. Immunol.* 178: 7598–7606.

51. Gillard, G. O., J. Dooley, M. Erickson, L. Peltonen, and A. G. Farr. 2007. Aire-dependent alterations in medullary thymic epithelium indicate a role for Aire in thymic epithelial differentiation. *J. Immunol.* 178: 3007–3015.
52. Yano, M., N. Kuroda, H. Han, M. Meguro-Horike, Y. Nishikawa, H. Kiyonari, K. Maemura, Y. Yanagawa, K. Obata, S. Takahashi, et al. 2008. Aire controls the differentiation program of thymic epithelial cells in the medulla for the establishment of self-tolerance. *J. Exp. Med.* 205: 2827–2838.
53. Wang, X., M. Laan, R. Bichele, K. Kisand, H. S. Scott, and P. Peterson. 2012. Post-Aire maturation of thymic medullary epithelial cells involves selective expression of keratinocyte-specific autoantigens. *Front. Immunol.* 3: 19.
54. Wada, N., K. Nishifuji, T. Yamada, J. Kudoh, N. Shimizu, M. Matsumoto, L. Peltonen, S. Nagafuchi, and M. Amagai. 2011. Aire-dependent thymic expression of desmoglein 3, the autoantigen in pemphigus vulgaris, and its role in T-cell tolerance. *J. Invest. Dermatol.* 131: 410–417.
55. Koh, A. S., E. L. Miller, J. D. Buenrostro, D. M. Moskowitz, J. Wang, W. J. Greenleaf, H. Y. Chang, and G. R. Crabtree. 2018. Rapid chromatin repression by Aire provides precise control of immune tolerance. *Nat. Immunol.* 19: 162–172.
56. Jäkel, S., E. Agirre, A. Mendanha Falcão, D. van Bruggen, K. W. Lee, I. Knuesel, D. Malhotra, C. Ffrench-Constant, A. Williams, and G. Castelo-Branco. 2019. Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* 566: 543–547.
57. Gray, D., J. Abramson, C. Benoist, and D. Mathis. 2007. Proliferative arrest and rapid turnover of thymic epithelial cells expressing Aire. *J. Exp. Med.* 204: 2521–2528.
58. Hao, Y., S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, et al. 2021. Integrated analysis of multimodal single-cell data. *Cell* 184: 3573–3587.e29.
59. Delacher, M., M. Simon, L. Sanderink, A. Hotz-Wagenblatt, M. Wuttke, K. Schambeck, L. Schmidleithner, S. Bittner, A. Pant, U. Ritter, et al. 2021. Single-cell chromatin accessibility landscape identifies tissue repair program in human regulatory T cells. *Immunity* 54: 702–720.e17.
60. Laan, M., K. Kisand, V. Kont, K. Möll, L. Tserel, H. S. Scott, and P. Peterson. 2009. Autoimmune regulator deficiency results in decreased expression of CCR4 and CCR7 ligands and in delayed migration of CD4<sup>+</sup> thymocytes. *J. Immunol.* 183: 7682–7691.
61. Perniola, R. 2018. Twenty years of AIRE. *Front. Immunol.* 9: 98.