SUPPLEMENTAL FIGURES

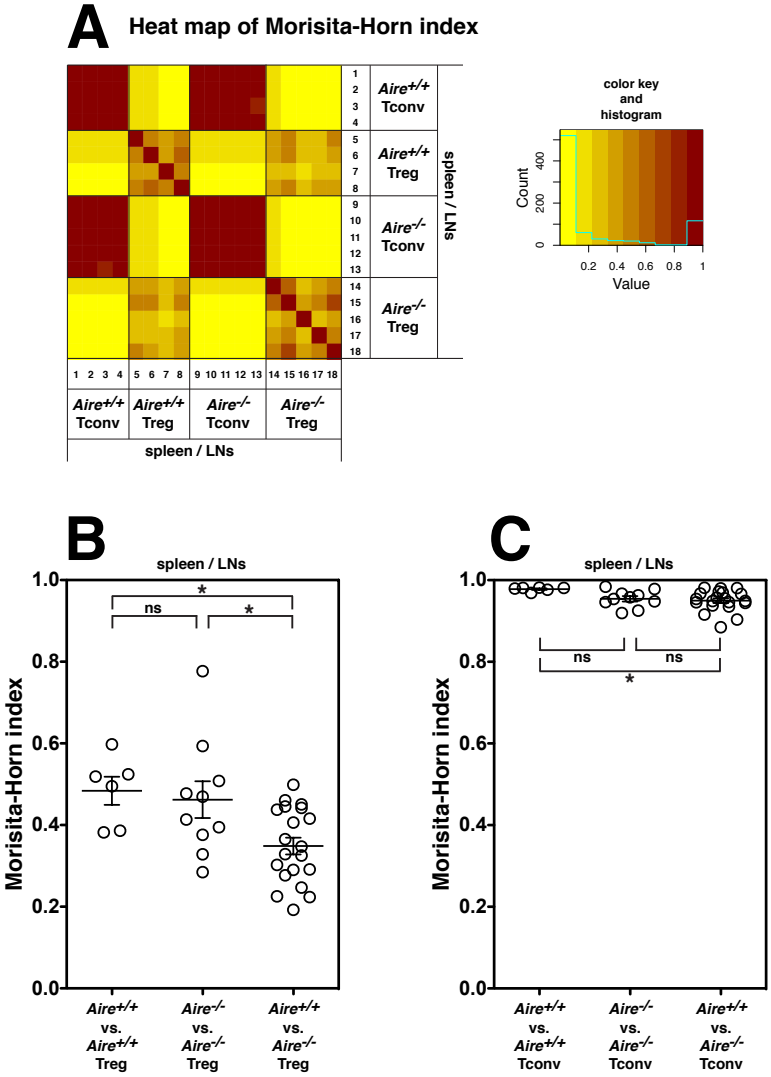

Figure S1 (Related to Figure 2): Analysis of TCR Repertoire Similarity Using the Morisita-Horn Similarity Index.

For the 18 complete TCRα sequencing samples from the pooled spleen and lymph nodes, repertoire overlap was assessed using the Morisita-Horn (MH) similarity index (see Supplemental Experimental Procedures). TCRα chains were assessed solely based on their predicted CDR3 segment, regardless of V-region usage.

(A). Heat map of the MH index for pairwise comparisons of the 18 samples, denoted 1-18 and organized according to the indicated T cell subset (Tconv cell or Treg cell) and *Aire* genotype (*Aire*$^{+/+}$ or *Aire*$^{-/-}$). For the MH index, a value of 1 indicates identity and a value of 0 denotes complete dissimilarity. A color key for the heat map and an overlaid histogram of MH comparisons (counts) vs. MH value are shown.

(B and C). Scatter plots of select MH pairwise comparisons from panel A. The impact of Aire on the Treg cell (B) or Tconv cell (C) TCR repertoire in the spleen/LNs was analyzed by plotting the indicated MH pairwise comparisons. For example, in the first column of panel B, the four *Aire*$^{+/+}$ Treg cell samples (5-8) are compared to each other. Likewise, in the third column of panel B, the four *Aire*$^{+/+}$ Treg cell samples (5-8) are compared to the five *Aire*$^{-/-}$ Treg cell samples (14-18). Asterisks indicate $P < 0.05$; ns, not significant. Statistical comparisons were made using ANOVA and Tukey's multiple comparison test. See also Table S1.

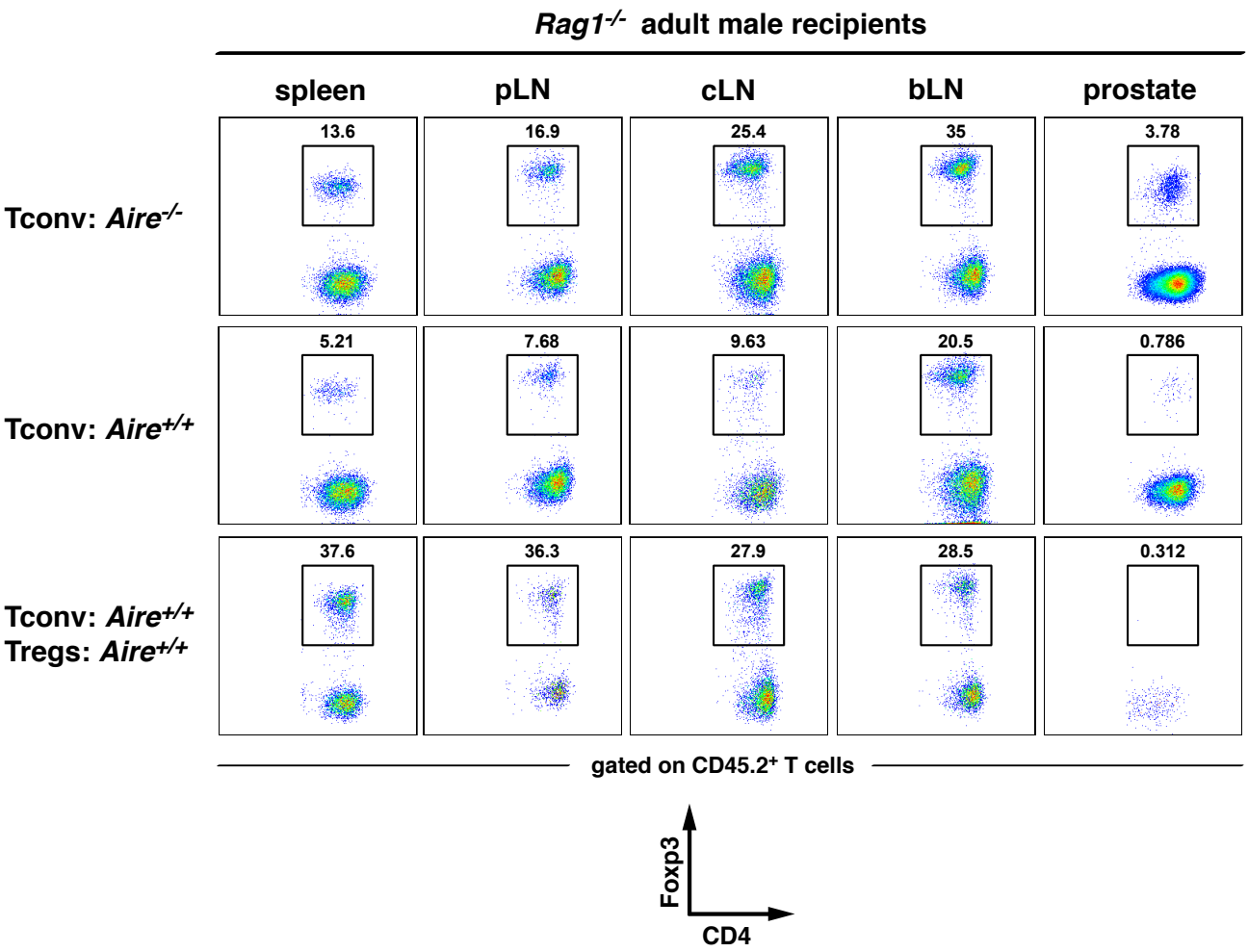

Figure S2 (Related to Figure 3): Transfer of Tconv Cells From *Aire*$^{+/+}$ Mice Into *Rag1*$^{-/-}$ Adult Males Induces Prostatic T Cell Infiltration.

7 x 10^6 CD4$^+$Foxp3^{neg} Tconv cells from the pooled spleen and lymph nodes of 4-5-week-old CD45.2$^+$ *Aire*$^{-/-}$ males (top row) or *Aire*$^{+/+}$ males (middle row) were transferred into 6-8-week-old male CD45.1$^+$ *Rag1*$^{-/-}$ recipients. In the bottom row, 7 x 10^6 Tconv cells from *Aire*$^{+/+}$ mice were co-transferred with 1 x 10^6 CD4$^+$Foxp3$^+$ Treg cells from *Aire*$^{+/+}$ mice. 3 weeks post-transfer, the fate of CD45.2$^+$ donor cells was assessed at different sites. Representative flow cytometric plots of Foxp3 vs. CD4 expression are shown for CD4$^+$CD45.2$^+$ donor T cells isolated from the indicated sites. The percentages of cells falling within the denoted gates are shown. pLN, periaortic lymph nodes; cLN, cervical lymph nodes; bLN, brachial lymph nodes. A representative experiment is shown.

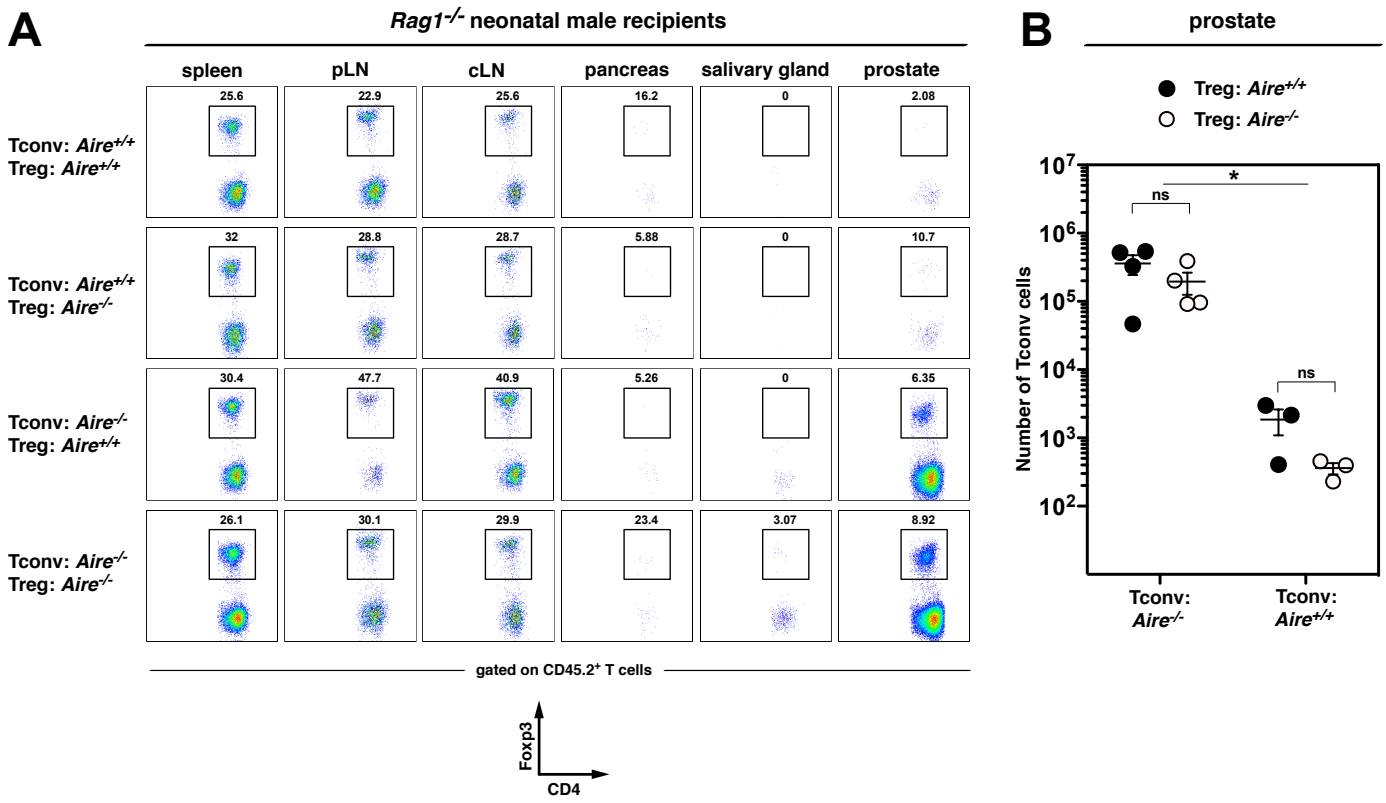

Figure S3 (Related to Figure 3): The Autoimmune Defect in *Aire*$^{-/-}$ Mice Maps to the T Conventional Cell Compartment.

A mixture of 7 x 10^6 CD4$^+$Foxp3^{neg} Tconv cells and 1 x 10^6 CD4$^+$Foxp3$^+$ Treg cells, isolated from the pooled spleen and lymph nodes of 4-5-week-old CD45.2$^+$ *Aire*$^{+/+}$ or *Aire*$^{-/-}$ males, as indicated, was transferred into 3-day-old male CD45.1$^+$ *Rag1*$^{-/-}$ recipients. 9 weeks post-transfer, the fate of CD45.2$^+$ donor cells was assessed at different sites.

(A) Representative flow cytometric plots of Foxp3 vs. CD4 expression are shown for CD4$^+$CD45.2$^+$ donor T cells isolated from the indicated organs. The percentages of cells falling within the denoted gates are shown. pLN, periaortic lymph nodes; cLN, cervical lymph nodes; bLN, brachial lymph nodes.

(B) Summary plot of the number of CD45.2$^+$ donor Tconv cells recovered from the prostate. The source (*Aire*$^{+/+}$ or *Aire*$^{-/-}$) of donor Treg cell and Tconv cells is indicated. Data are pooled from $N = 3$ experiments. The mean +/- SEM is denoted. Asterisks indicate $P < 0.05$; ns, not significant. Statistical comparisons of all groups were made using ANOVA and Tukey's multiple comparison test.

Table S1 (Related to Figure 2 and Figure 4). Complete TCRα Sequencing Frequency Table (Pooled Spleen and Lymph Nodes).
For samples subjected to complete TCRα sequencing (see Supplemental Experimental Procedures), the frequencies of each TCR CDR3α segment, as a percentage of all TCRα sequence reads within a given sample, are listed. For this study, TCRs were cataloged solely based on predicted TCR CDR3α sequence, regardless of V-region usage. The table lists the Sample ID (1-18) for the complete TCRα sequencing data set, the organ from which T cells were isolated (SP/LN, pooled spleen and lymph nodes), the *Aire* genotype, and the T cell subset (Tconv cell or Treg cell). For each biological sample, the total number of in-frame TCR sequence reads and the number of unique TCR sequence reads are also listed.

Table S2 (Related to Figure 5). Complete TCRα Sequencing Frequency Table (*Aire*$^{-/-}$ Prostates).
For samples subjected to complete TCRα sequencing (see Supplemental Experimental Procedures), the frequencies of each TCR CDR3α segment, as a percentage of all TCRα sequence reads within a given sample, are listed. For this study, TCRs were cataloged solely based on predicted TCR CDR3α sequence, regardless of V-region usage. The table lists the Sample ID (*Aire*$^{-/-}$ prostate 1 through *Aire*$^{-/-}$ prostate 9) for the complete TCRα sequencing data set, the organ from which T cells were isolated (PR, prostate), the *Aire* genotype (*Aire*$^{-/-}$), and the T cell subset (Tconv cell). For each biological sample, the total number of in-frame TCR sequence reads and the number of unique TCR sequence reads are also listed.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

TCRα sequence analysis

$CD4^+Foxp3^+$ Treg cells and $CD4^+Foxp3^{neg}$ Tconv cells were FACS-purified from different anatomical sites of 9-week-old TCRβTg *Foxp3*^{GFP} males on an *Aire*$^{+/+}$ and *Aire*$^{-/-}$ background, and cDNA was subjected to TCRα sequence analysis using either "TRAV14 TCRα sequencing" or "complete TCRα sequencing", described below.

TRAV14 TCRα sequencing

For "TRAV14 TCRα sequencing", cDNA from FACS-sorted T cell subsets was PCR-amplified using the following primers. The forward fusion primers follow the sequence PrimerA-Barcode-mVregion, and the reverse fusion primers follow the sequence PrimerB-Barcode-mCregion. Sequences are as follows:
PrimerA 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3'
PrimerB 5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3'
mVα2.P2 5'-GAGAAAAAGCTCTCCTTGCAC-3'
mCα.P2 5'-CCATGGTTTTCGGCACATTG-3'

The barcodes used were those recommended by the manufacturer for use with GS FLX Titanium emPCR (LIB-L) chemistry (Roche). PCR amplicons were gel-purified using a Gel Extraction Kit (Qiagen), and samples were processed for sequencing. Sequencing was performed at the Utah State University Center for Integrated Biosystems using the Titanium sequencing chemistry on a Roche GS FLX system. The use of barcoded samples permitted the multiplexing of up to 14 different samples.

For TRAV14 TCRα sequencing data analysis, each sequence read was processed using a custom written python script (available upon request) to demultiplex the sample barcodes and parse out and translate the CDR3 regions. Using the parsed CDR3 sequences we generated a count table which had for each CDR3 and each sample a count of how many times that CDR3 element appeared. Typically, $>2 \times 10^3$ TCR sequence reads were obtained per sample.

Complete TCRα sequencing

For "complete TCRα sequencing", RNA from FACS-sorted T cell subsets was subjected to TCRα sequencing using the Amp2Seq service from iRepertoire, a platform based on semi-quantitative multiplex PCR coupled with Illumina sequencing. This approach allows analysis of the complete TCRα repertoire, regardless of variable-region usage. Typically, $>7 \times 10^5$ TCR sequence reads were obtained per sample.

Differential TCR Representation Analysis.

For complete TCRα sequencing, we used edgeR from the R/Bioconductor package, a standard method from RNAseq analysis. TCRα chains were analyzed solely based on the predicted CDR3 sequence, regardless of V-region usage. We first filtered for TCRs with CDR3 segments between 7-17 amino acids in length and removed those TCRs with counts that were less than 10 in N_s-1 samples where N_s was the size of the smallest group. For example, when comparing two data sets with N = 4 and N = 5, N_s-1 = 3. In this case, TCRs with fewer than 10 counts in 3 or more samples were removed from the differential analysis. This approach allowed focusing the analysis on recurrent TCRs. The samples were then normalized using the standard edgeR normalization routine (calcNormFactors) and p-values were computed using the edgeR standard test. P-values were corrected using the false discovery rate (FDR) method and the standard threshold of 0.05 was set. To enable fold-change calculations, a pseudocount approach was used. Differential representation of recurrent TCRs was displayed in "volcano plots" depicting FDR vs. $\log_2$ fold-change in *Aire*$^{+/+}$ samples vs. *Aire*$^{-/-}$ samples.

Morisita-Horn similarity index

To analyze TCR repertoire similarity, we analyzed CDR3 elements using the Morisita-Horn Similarity Index (Magurran, 1988). The index computes the similarity between two populations of "species", originally taxonomic species, but here CDR3 species. If x_i is the number of time species "i" occurs in sample X and y_i the number in sample Y then the index is defined as:

$$C_H = \frac{2\sum_{i=1}^{S} x_i y_i}{\left(\frac{\sum_{i=1}^{S} x_i^2}{X^2} + \frac{\sum_{i=1}^{S} y_i^2}{Y^2}\right) XY}$$

Where X is the sum of x_i, Y is the sum of y_i and S is the number of species. The resulting matrix of pair-wise similarities between all samples was then clustered using standard hierarchical clustering using 1-C_H as the distance metric and using average linkage for the joining.