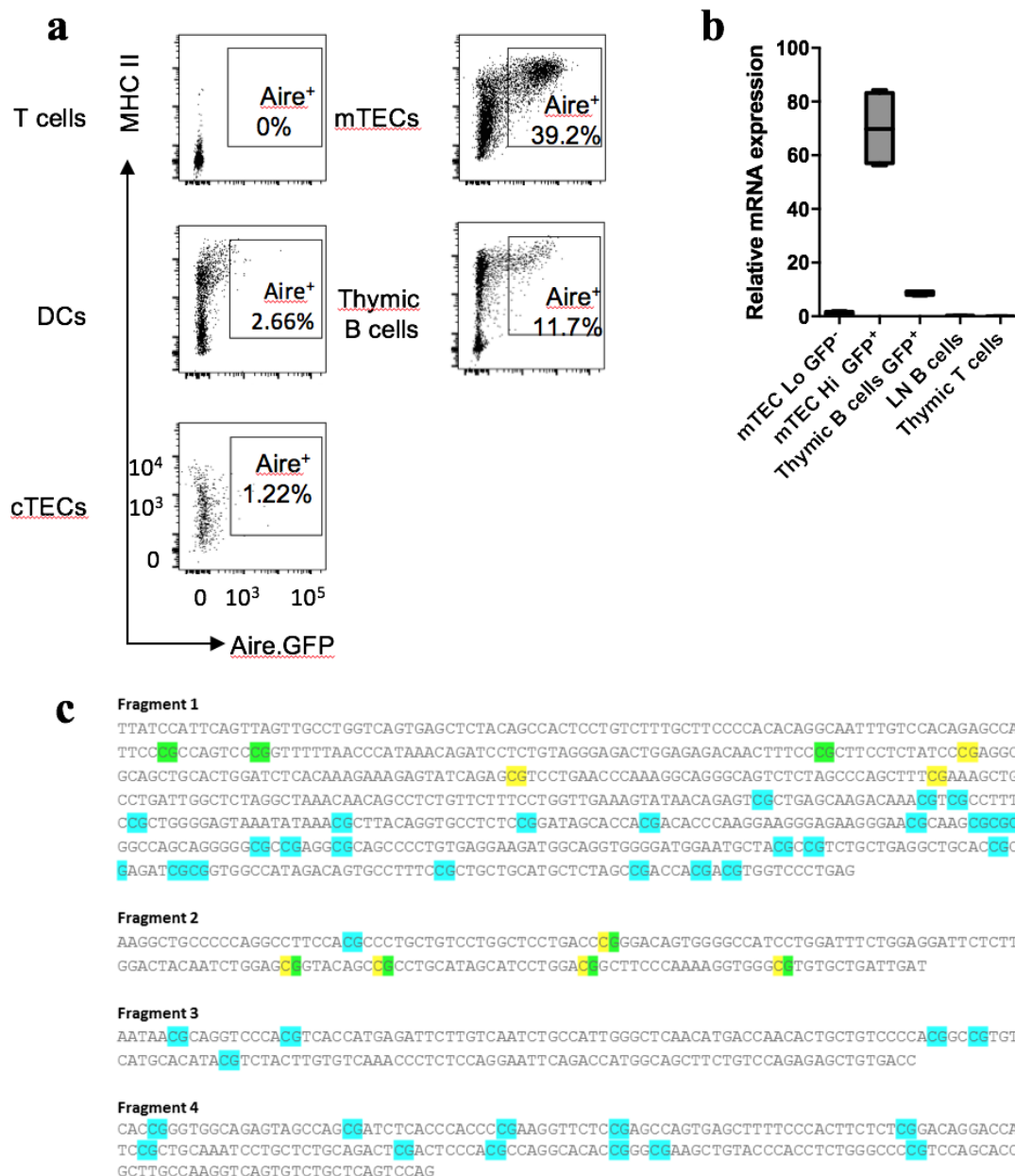
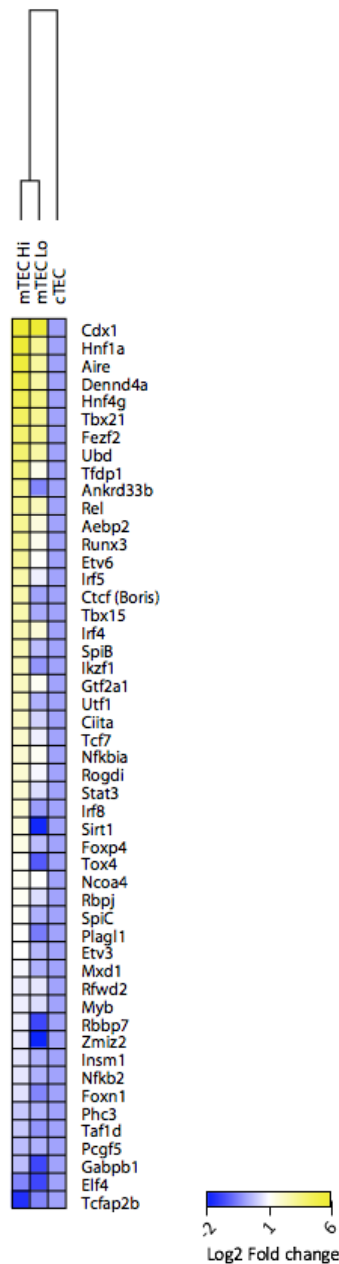




Supplementary Figure 1

Aire expression in various immune-cell populations.

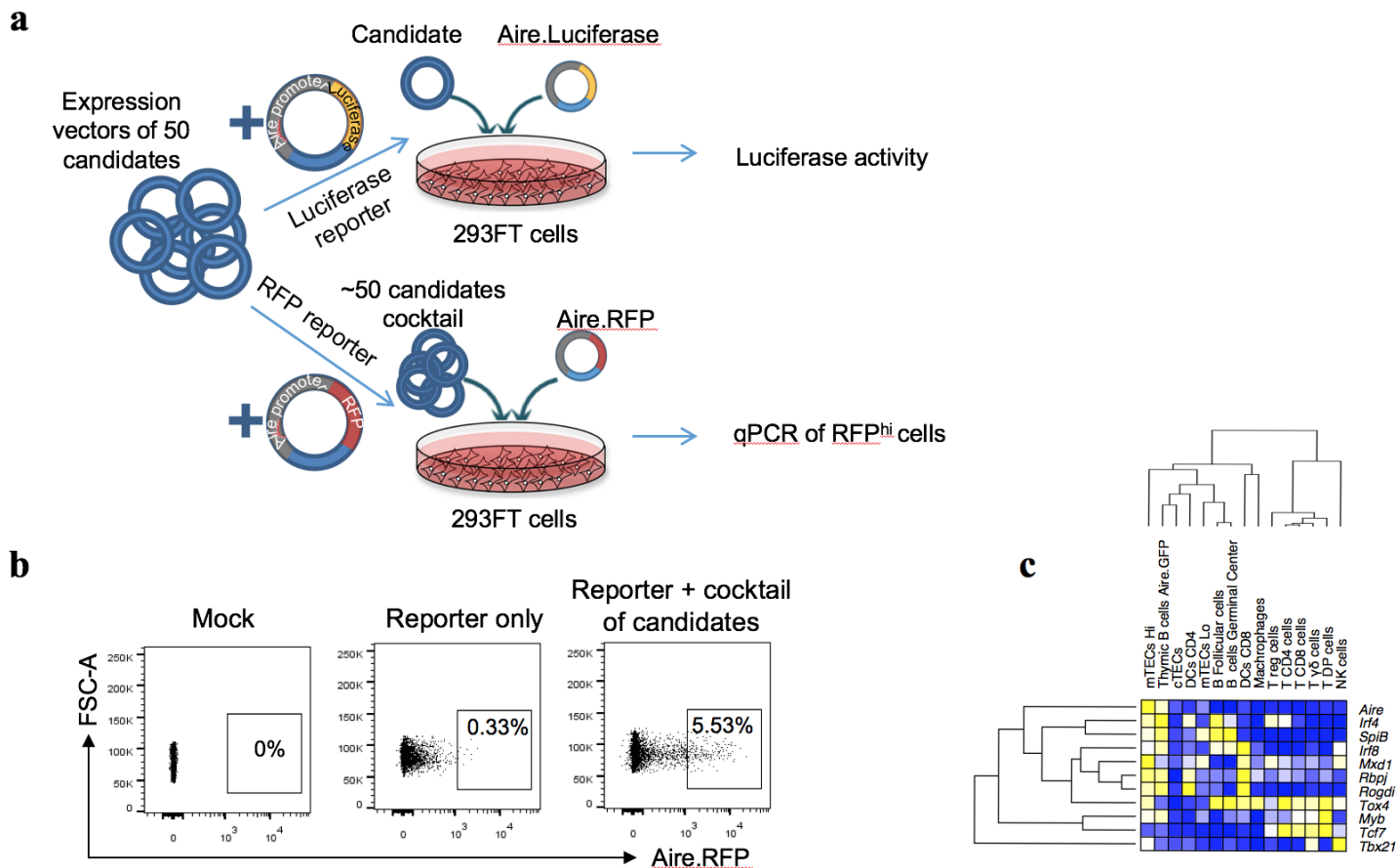
(a) Representative flow cytometry analyses of the expression of Aire.GFP (x-axis) and MHC class II (MHCII) (y-axis) in cell populations obtained from (n=3) 6-7week-old Aire.GFP mice, gated on thymic CD45⁺ EpCAM⁻ CD11c⁻ (T cells); CD45⁺ EpCAM⁻ CD11c⁺ (Dendritic Cells; DCs); CD45⁺ EpCAM⁺ Ly51⁺ (cortical thymic epithelial cells; cTECs); CD45⁺ EpCAM⁺ Ly51⁻ (medullary thymic epithelial cells; mTECs) or CD45⁺ EpCAM⁻ CD19⁺ (B cells) populations. Numbers in outlined areas indicate %GFP⁺ cells, gated on an age-matched WT mouse. Data are representative of two independent experiments with similar results. (b) Quantitative PCR analysis of mean Aire expression in various sorted cell types obtained from same mice as indicated in (a). Results are normalized to the expression of *Hprt* and are presented relative to the expression in mature mTECs. (c) Genomic sequence of the predicted CpG islands. Color coded are the CG pairs. In light green are mTEC-specific differentially demethylated and in yellow are immature (MHCII^{lo}) mTEC-specific differentially demethylated cytosine residues.



Supplementary Figure 2

Quantitative PCR analysis verifies the mTEC^{hi}-cell-specific transcriptional regulator signature.

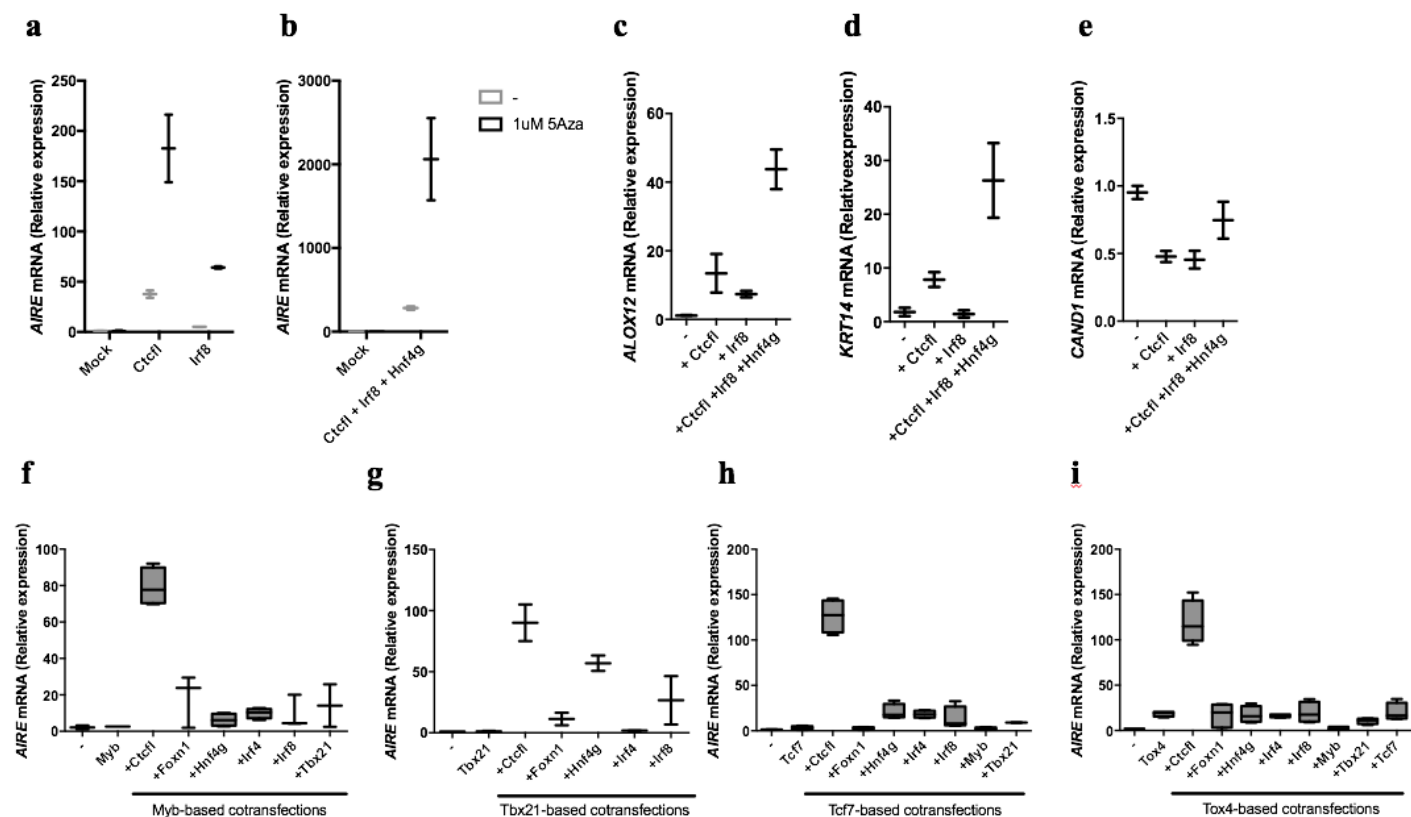
Clustered heat map depicting relative mRNA expression of indicated transcription regulators in sorted EpCAM⁺ Ly51^{neg-lo} mature (MHCII^{hi}) *vis-à-vis* immature (MHCII^{lo}) medullary or cortical (Ly51^{hi}) thymic epithelial cells obtained from 6-7 weeks old mice (n=3). Shown are relative signal values normalized to expression of *Hprt* and relative to cTECs. Colors represent high (yellow) or low (blue) expression levels. Data are representative of at least three independent experiments with similar results.



Supplementary Figure 3

Aire reporter assays highlight potential mTEC-specific *Aire* regulators.

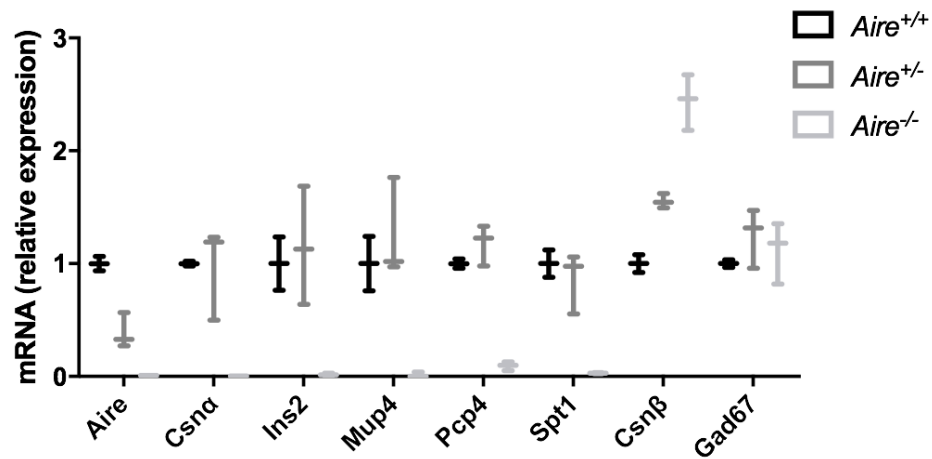
(a) Outline of the *Aire* promoter-based luciferase and RFP screening systems. (b) Representative flow cytometry analysis showing *Aire*.RFP reporter activity (x-axis) in HEK 293FT cells co-transfected with the *Aire*.RFP reporter vector alongside either an empty vector or expression vectors encoding all ~50 candidate mTEC^{hi}-specific transcription regulator genes for 48h. Numbers outlined present %RFP^{hi} cells. Data are representative of at least three independent experiments with similar results. (c) Clustered heat map depicting mRNA expression profiles of candidate *Aire*-regulators that are predominantly expressed in antigen presenting cells or T cells and are predicted to activate the *Aire* promoter by either the Luciferase or the RFP reporter assays (Fig. 3e). Shown are normalized relative expression values of indicated immune cell populations. Colors represent high (yellow) or low (blue) expression levels. DCs, Dendritic cells; T $\gamma\delta$, gamma delta T cells; T DP, Double positive (CD4⁺ CD8⁺) T cells; NK, Natural Killer cells.



Supplementary Figure 4

AIRE and *AIRE*-dependent genes are induced following expression of various transcriptional regulators.

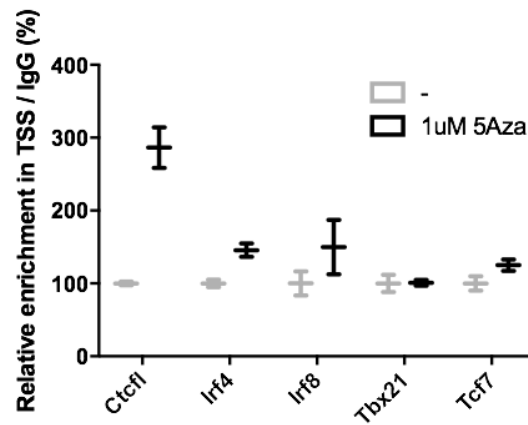
(a+b) *AIRE* mRNA expression in HEK 293FT cells, treated with or without 5-Aza, encoding highlighted transcriptional regulators (a) or their combinations (b), and tested 48h later. (c-e) qPCR analyses showing mRNA expression of the *AIRE*-dependent genes *ALOX12* (c) and *KRT14* (d) or *AIRE* independent gene *CAND1* (e) in 5-Aza treated cells transfected with the indicated factors and tested 48h later. (f-i) qPCR analyses showing *AIRE* mRNA levels in 5-Aza treated HEK 293FT cells transfected with expression vectors of Myb (f), Tbx21 (g), Tcf7 (h) or Tox4 (i) alongside short-listed candidates and measured 48h later; Results are normalized to the expression of *HPRT* and are presented relative to the expression in untreated cells transfected with an empty expression vector. Data in all experiments are representative of at least two independent experiments with similar results (mean and S.E.M of $n = 2$ biological replicates).



Supplementary Figure 5

mTECs with 50% Aire expression show normal TRA expression.

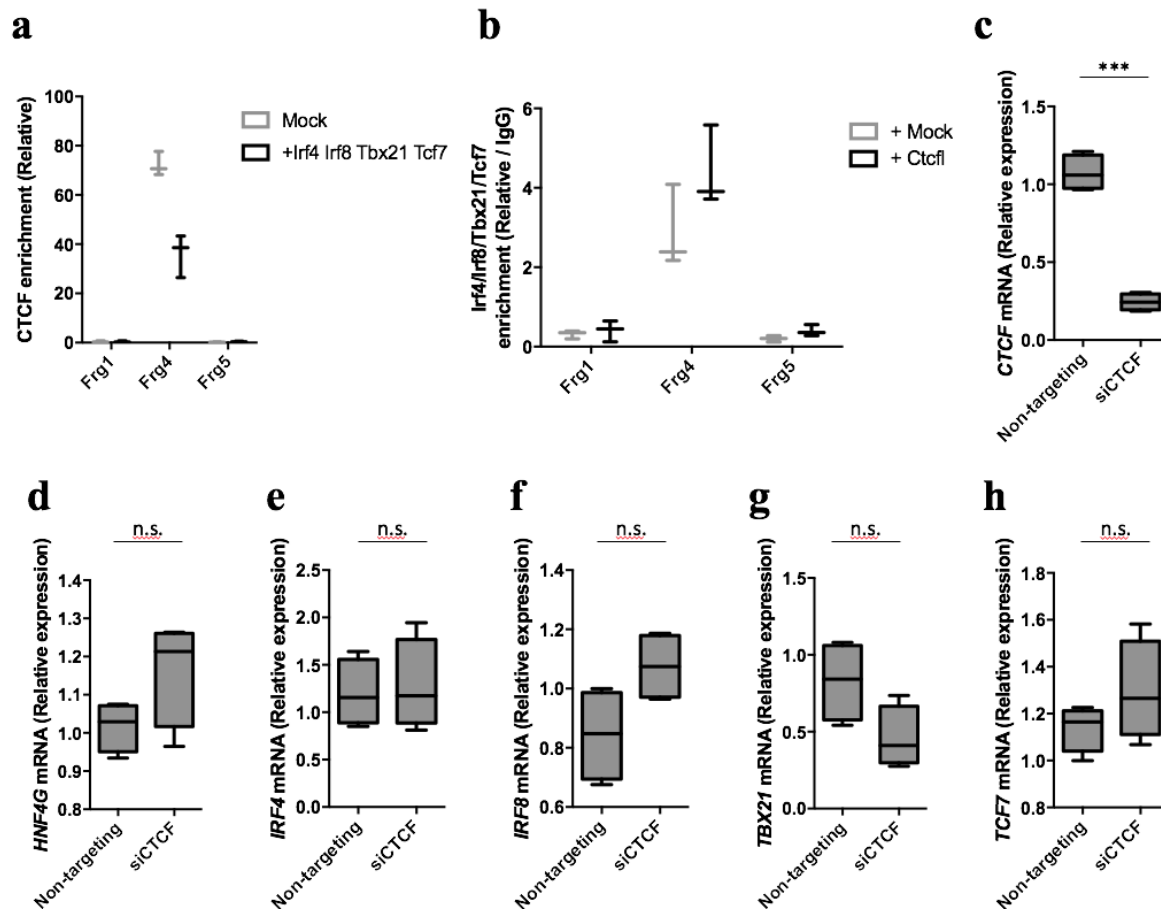
Quantitative PCR analysis of expression of *Aire*, several Aire-responsive TRAs (*Ins2*, *Csna*, *Mup4*, *Pcp4* and *Spt1*), and several Aire-neutral TRAs (*Csnb* and *Gad67*) in FACS-sorted mTEC^{hi} cells obtained from 6-8 week old *Aire*^{+/+}, *Aire*^{+/-} or *Aire*^{-/-} mice (n=3). Results are normalized to the expression of *Hprt* and are presented relative to the expression values in *Aire*^{+/+} mTECs. Data are representative of two independent experiments with similar results (mean and S.E.M).



Supplementary Figure 6

Most candidate regulators bind the *AIRE* promoter independently of DNA-methylation status.

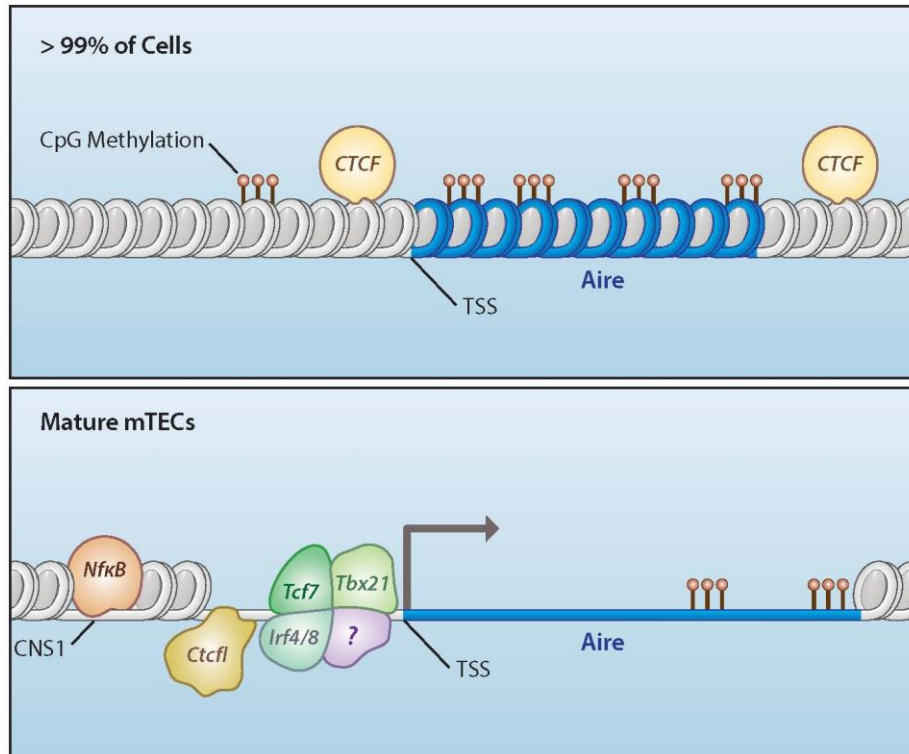
Chromatin immunoprecipitation followed by quantitative PCR assessing relative enrichment of indicated transfected candidates (vs. IgG control) along the *AIRE* TSS (corresponding to fragment 3 in Fig 6b), in HEK 293FT cells treated 48h or untreated with 5-Aza. Results are normalized to values obtained from 5-Aza untreated cells. Data are representative of at least two independent experiments with similar results (mean and S.E.M of $n = 2$ biological replicates).



Supplementary Figure 7

Binding dynamics and endogenous expression of candidate transcriptional regulators.

(a-b) Chromatin immunoprecipitation followed by quantitative PCR assessing relative enrichment of CTCF following expression of Irf4/Irf8/Tbx21/Tcf7 (a) or of HA-tagged Irf4/Irf8/Tbx21/Tcf7 following expression of Ctcf1 (b) in 5-Aza treated HEK 293FT cells transfected for 48h (n=3); (c-g) Quantitative PCR analyses showing endogenous expression of *CTCF* (c) *HNF4G* (d), *IRF4* (e), *IRF8* (f), *TBX21* (g) and *TCF7* (h) in 5-Aza treated HEK 293FT cells (pooled n=4) transfected with an siRNA set targeting expression of CTCF, and tested 48h later. Results are normalized to the expression of *HPRT* and are presented relative to the expression in the same cells transfected with a non-targeting siRNA set; Data are representative of at least two independent experiments with similar results. n.s. not significant, *P < 0.05, **P < 0.01 and ***P < 0.001 (Student's t-test, S.E.M).



Supplementary Figure 8

The molecular mechanisms that control the expression of *Aire*.

Aire is not expressed in the vast majority of cells in the body, except for a few cell types, primarily mature mTECs. To achieve such regulation, the *Aire* expression is regulated at multiple levels: First, the *Aire* locus is physically inaccessible and hypermethylated at specific CpG residues upstream and downstream to the TSS in cells and tissues that do not express it. Second, *Aire* locus is insulated by a global chromatin organizer - CTCF, which specifically binds at *Aire*'s TSS and downstream of the last *Aire* exon; Third, in mTECs, *Aire* expression is facilitated by concurrent eviction of CTCF, specific DNA demethylation and coordinated action of several transcription activators, including Irf4, Irf8, Tbx21, Tcf7 and Ctcf, which act on mTEC-specific accessible regions in the *Aire* locus .

Supplementary Table 1 – Transcription regulator genes upregulated in mature mTECs vis-à-vis immature mTECs and cTECs.

Name	Fold change mTEC Hi/Lo	Fold changemTEC Hi/cTEC
Aire	11.4100963	10.67106
Tcfap2b	8.98452746	8.364344
Dennd4a	8.71804246	14.66115
Sirt1	8.05024159	14.35294
Utf1	7.39779657	7.301556
Ankrd33b	5.60652144	11.86995
Phc3	5.55497389	6.594924
Irf5	5.04240688	4.928883
Irf4	4.74263033	6.895375
Ncoa4	4.62686207	5.773353
Plagl1	4.08769744	16.33666
Hnf4g	4.03769368	5.224291
Tbx15	3.91150184	3.835202
Tfdp1	3.62959876	4.640333
Foxn1	3.59373554	3.104075
Irf8	3.57619169	3.858744
Rel	3.52249268	5.241566
Tbx21	3.44948838	2.236881
Etv3	3.44264711	4.486559
Tcf7	3.39900219	2.238772
Aebp2	3.30079972	4.082422
Runx3	3.24348469	5.152152
Hnf1a	3.19610281	3.379897
Fezf2	3.17759271	12.29287
Rbpj	3.03087115	6.527417
Insm1	2.90392642	4.350036
Pcgf5	2.89077867	5.860701
Ciita	2.82972798	3.428721
Nfkbia	2.77274131	4.581448
Gtf2a1	2.7568068	4.44334
Ikzf1	2.71968333	3.870083
Tox4	2.62980656	4.474972
Rfwd2	2.62367036	4.246492
Elf4	2.56809129	8.318809
Rbbp7	2.29855073	4.20985
Foxp4	2.2741079	3.791974
Rogdi	2.24674421	2.345978
Nfkb2	2.22155578	4.085591
Mxd1	2.21177256	5.152426
Zmiz2	1.94620437	3.613077
Myb	1.88048958	4.752634
Ctcf1 (Boris)	1.68022255	1.079153
Stat3	1.65916194	3.448065
Cdx1	1.62163161	1.319616
Gabpb1	1.58578471	3.206148
Etv6	1.45836195	5.98898
SpiB	1.16874494	8.217068
Taf1d	1.05496154	1.329289

Supplementary Table 2 – Expression of mTEC^{hi}-specific transcriptional regulators in Aire.GFP⁺ thymic B cells vs. splenic B cells

Name	AireGFP ⁺ Thymic/Splenic B cells	Mean expression
Aire	9.069605372	222.4404
Tcf7	6.679928897	154.4789
Myb	3.73302823	331.6291
Tfdp1	3.575457068	246.369
Rbpj	2.742463946	300.5194
Etv6	2.60503916	231.3065
Runx3	2.576029358	445.0536
Ankrd33b	2.429169615	162.4542
Nfkb1a	2.347676318	855.8783
Mxd1	1.841652703	181.6231
Rogdi	1.800207103	385.6251
Aebp2	1.73822991	183.2502
Ciita	1.728731552	718.2777
Rbbp7	1.694879029	792.1886
Tbx21	1.628119619	99.88379
Foxp4	1.576560834	265.4678
Nfkb2	1.561903985	499.5257
Pcgf5	1.523377318	67.7937
Zmiz2	1.51265162	896.6524
Irf5	1.512611248	542.1213
Insm1	1.467949814	62.81354
Plagl1	1.463414426	33.67529
Irf4	1.442455029	398.4614
Dennd4a	1.322383507	1459.764
Etv3	1.237677057	214.3112
Cdx1	1.199206814	51.41575
Gtf2a1	1.176635329	431.9736
Gabpb1	1.174270921	206.8966
Utf1	1.129320116	23.63227
Rfwd2	1.117321546	148.3367
Fezf2	1.109492041	54.19175
Irf8	1.09713091	1370.767
Stat3	1.094667953	578.793
Phc3	1.090953703	274.3574
Ncoa4	1.081089868	466.0937
Taf1d	1.032198527	121.1544
Tbx15	0.919026968	13.54754
Ctcf1	0.904254381	27.74701
Hnf4g	0.859103402	14.83876
SpiB	0.851087595	1114.948
Ikzf1	0.850856701	780.5868
Tcfap2b	0.850409473	19.77215
Sirt1	0.829877556	225.2203
Tox4	0.803972365	276.7259
Foxn1	0.75894829	47.80004
Hnf1a	0.748182066	53.08571
Elf4	0.684834759	687.978
Rel	0.587924554	1507.865

Data obtained from Affymetrix gene microarray on sorted Thymic/Splenic B cells from Aire.GFP⁺ mice. The color code indicates candidate Aire regulators that have >100 mean expression levels and have either >1.5 fold change (dark red), between 1 and 1.5 fold change (pink) or <1 fold change (blue) compared to splenic B cells.

Supplementary Table 3 – Generation of a library of expression vectors of all the candidate Aire-regulators

Name	Expression Vector Generated	Proper expression in 293FT
Aebp2	+	+
Aire	+	+
Ankrd33b	+	+
Cdx1	+	+
Ciita	+	+
Ctcf1 (Boris)	+	+
Dennd4a	+	-
Elf4	-	-
Etv3	+	+
Etv6	+	+
Fezf2	+	+
Foxn1	+	+
Foxp4	+	+
Gabpb1	-	-
Gtf2a1	+	+
Hnf1a	+	+
Hnf4g	+	+
Ikzf1	+	+
Insm1	+	-
Irf4	+	+
Irf5	+	+
Irf8	+	+
Mxd1	+	+
Myb	+	+
Ncoa4	+	+
Nfkb2	+	+
Nfkbia	+	+
Pcgf5	+	+
Phc3	+	+
Plagl1	+	-
Rbbp7	+	+
Rbpj	+	+
Rel	+	+
Rfwd2	+	-
Rogdi	+	+
Runx3	+	+
Sirt1	+	+
SpiB	+	+
SpiC	+	+
Stat3	+	+
Taf1d	+	-
Tbx15	+	+
Tbx21	+	+
Tcf7	+	+
Tcfap2b	+	+
Tfdp1	+	+
Tox4	+	+
Utf1	+	+
Zmiz2	-	-