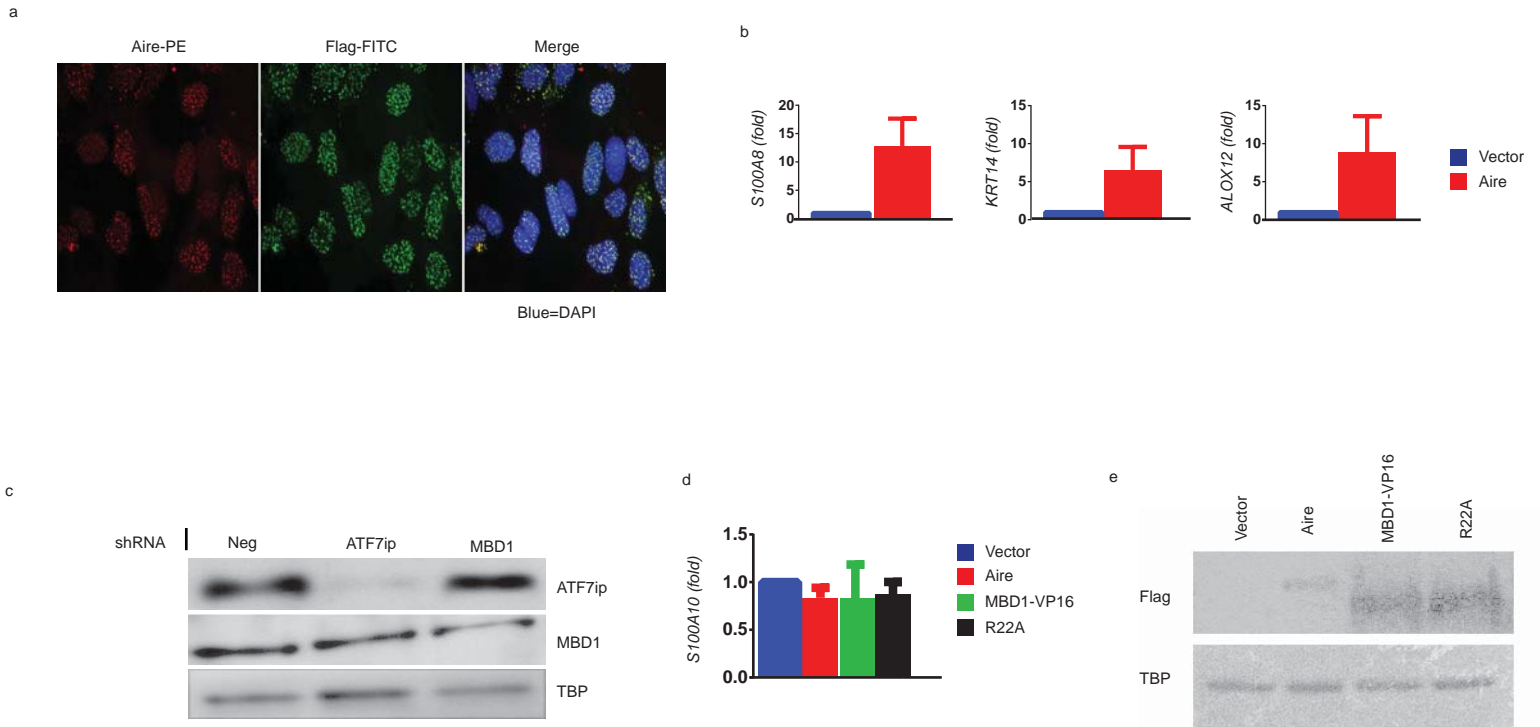
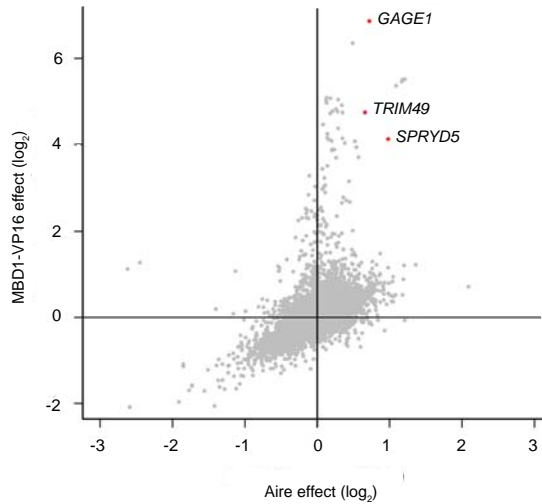


**Supplementary Figure 1** – The SAND domain of Aire interacts with ATF7ip and Aire binds to the CXXC domains of MBD1. **(a)** Outline of the strategy employed to generate the prey cDNA library from an Aire-GFP reporter transgenic mouse. Aire-expressing cells were purified by flow sorting MHC Class II Hi, GFP+ mTECs pre-gated on CD45-, EPCAM+ events from a collagenase treated thymus. **(b)** Structural domains of MBD1 including an N terminal MBD1 domain, a nuclear localization sequence (NLS), three CXXC domains, and a C terminal trans repression domain (TRD). Two truncation mutants are shown that are used for subsequent coimmunoprecipitation (Co-IP). **(c)** Representative example of yeast plates from pairwise mating using either HSR or SAND as bait and Aire or ATF7ip domain 2(D2) as prey. Each mating was performed in duplicate as indicated by arrows. **(d)** Plasmids expressing c-Myc-Aire and either Flag-MBD1 1-312 or Flag-MBD1 1-161 were transiently transfected into HEK 293 cells and Co-IP was performed with an irrelevant IgG or c-Myc antibody. TATA-binding protein (TBP) is included as a loading control. Data are representative of two experiments **(c,d)**.

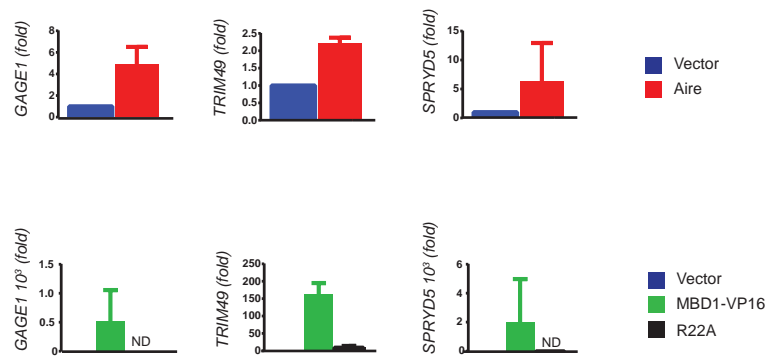


**Supplementary Figure 2** – A cell line stably expressing Aire recapitulates Aire-dependent gene expression. **(a)** Immunofluorescence of HEK 293 cells stably expressing Flag tagged Aire. The left panel shows Aire staining in red, the center panel shows Flag (green), the right panel shows the overlay (yellow). Blue indicates nuclear DAPI staining. **(b)** Relative expression of the Aire-dependent genes S100A8, KRT14, and ALOX12 in the Aire-stable cell line compared to a cell line expressing vector only. **(c)** Western blot showing the efficiency of knockdown by plasmids expressing a scrambled shRNA (shNeg), an ATF7ip shRNA (shATF7ip), or an MBD1 shRNA (shMBD1). TATA-binding protein (TBP) is included as a loading control. **(d)** Aire and MBD1-VP16 do not induce the non Aire-regulated gene S100A10. Relative expression of S100A10 after transient transfection with plasmids expressing Flag only (Vector), Aire-Flag (Aire), Flag-MBD1-VP16 (MBD1-VP16), or Flag-MBD1-VP16-R22A (R22A). **(e)** Western blot showing the expression of Flag only (Vector), Aire-Flag (Aire), Flag-MBD1-VP16 (MBD1-VP16), or Flag-MBD1-VP16-R22A (R22A) in HEK 293 cells. TATA-binding protein (TBP) is included as a loading control. Data are representative of three experiments (mean and s.e.m of pooled data from three biological replicates in **b,d**). Data are representative of two experiments (**a,c,e**).

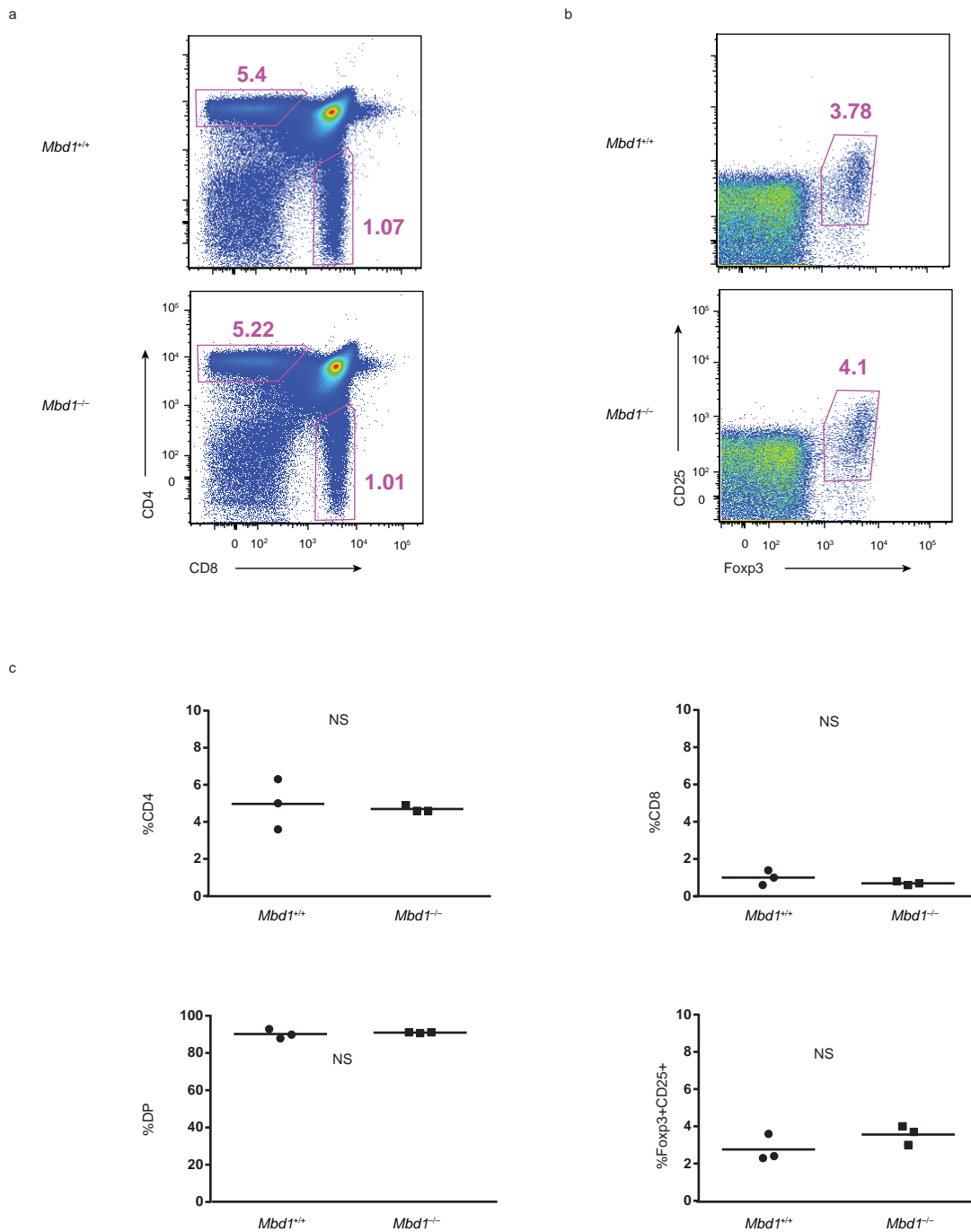
a



b

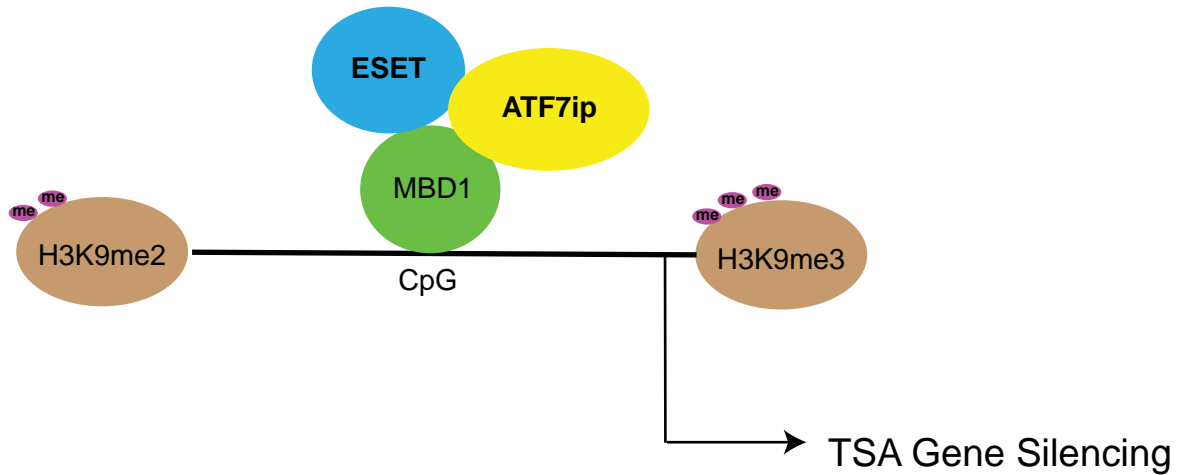


**Supplementary Figure 3** - Aire and MBD1-VP16 share many target genes. HEK 293 cells were transfected in triplicate with Flag only (Vector), Aire-Flag (Aire), or Flag-MBD1-VP16 (MBD1-VP16) and microarray analysis was performed on isolated RNA. **(a)** Comparison of effects of Aire and MBD1-VP16 on transcript levels (each dot represents a single gene). **(b)** qPCR confirmation of microarray analysis on three additional transcripts induced by both Aire and MBD1-VP16 as shown in **(a)**. Data are representative of three experiments **(a,b)**. Fold induction is calculated relative vector only.

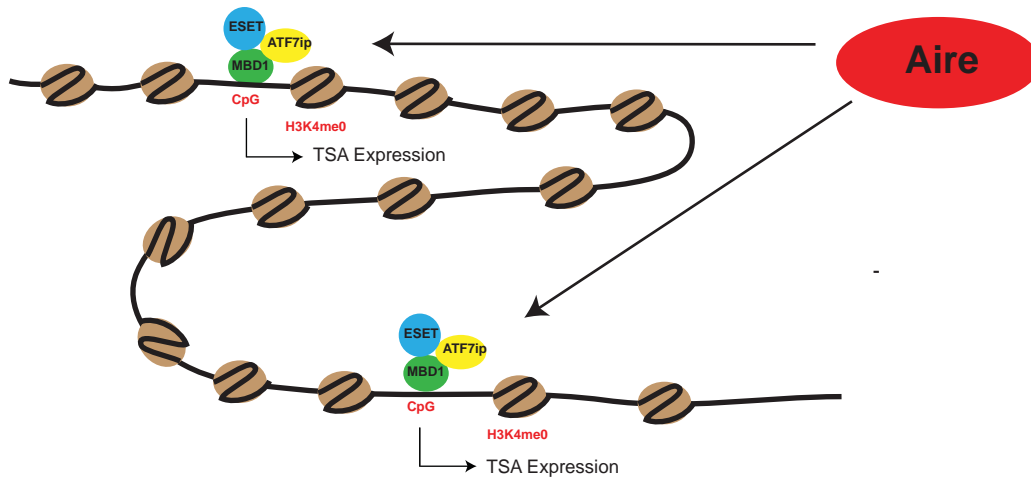


**Supplementary Figure 4** - *Mbd1*<sup>-/-</sup> mice have normal thymocyte populations. **(a)** Representative FACS plots of CD4 and CD8 thymocytes from 4 month old *Mbd1*<sup>+/+</sup> and *Mbd1*<sup>-/-</sup> mice. Numbers represent the percentage of cells in each gate. **(b)** Representative FACS plots of thymic Tregs (CD4<sup>+</sup>/CD25<sup>+</sup>/Foxp3<sup>+</sup>) from 4 month old *Mbd1*<sup>+/+</sup> and *Mbd1*<sup>-/-</sup> mice. Numbers represent the percentage of cells in each gate. **(c)** Graphs show the percentage of CD4, CD8, double positive (DP), and thymic Tregs from three mice of each genotype. NS = not statistically different. Each circle or square represents one mouse.

a



b



**Supplementary Figure 5** - Aire specifically targets repressed TSA loci by targeting the MBD1-ATF7ip-ESET protein complex. **(a)** Model shows the MBD1-ATF7ip-ESET complex interacting with methylated CpGs via MBD1 and creating the repressive histone mark H3K9me3. Data presented here indicates that this complex preferentially targets TSA genes for repression. **(b)** Aire binding to domain 2 (D2) of ATF7ip and to the CXXC domains of MBD1 targets Aire to the repressive MBD1-ATF7ip-ESET complex. Aire's binding to this complex specifically targets Aire to repressed TSA loci and promotes Aire's exquisite specificity for distinct TSA gene loci located throughout the genome. Aire's targeting of the MBD1-ATF7ip-ESET protein complex likely works in conjunction with Aire's interaction with H3K4me0 which may also be enriched at repressed TSA loci.

# MBD1 induced genes in mTECS

- TSA: 1-5 Tissues
- Housekeeping > 5 Tissues
- Not Accessible
- Below Threshold

Gene Symbol	GeneID	Mbd <sup>+/+</sup> vs Mbd1 <sup>-/-</sup> Fold Induction
Mbd1	17190	9.145298342
Mrps35	232536	6.533840784
Plcz1	114875	3.85141123
Synpr	72003	3.11568541
NAP043897-1		2.725867077
Hbb-b1	15129	2.561777232
Arml2	272322	2.545439487
Hbbp1	15199	2.347980628
Hbbd3	68695	2.334891062
NAP046281-1		2.324057824
BC064078*	408064*	2.294147562
Ces5	234673	2.289398244
Pap (Reg3b)	18489	2.216612695
Klhdc5	232539	2.109659755
Anxa13*	69787*	2.102752679
2010204N08Rik*	69983	2.063155969
Lrp1b	94217	2.044290519
Mro*	71263*	2.038691453
Muc15	269328	2.032227953
Apoc3	11814	2.009903405
Otor	57329	1.987474538
Tcfap2b	21419	1.939254094
P2y12	70839	1.91149809
AU020206*	101757*	1.897402732
Reg1	19692	1.865698137
Art2a*	11871*	1.863861439
Ube2e2	218793	1.850373564
Ttc18	76670	1.820800789
Hgf	15234	1.814313616
Me2	107029	1.804835314
F3	14066	1.793399261
Cci27	20301	1.790379958
BC052883 (Wes2)	381759	1.789235696
Gucy2b	14916	1.786500059
Afr2	14286	1.775534783
Mfsd4*	213006*	1.768320188
Ugt2b39	100559	1.765859507
Ddx25	30959	1.758791859
Nrl3	12355	1.757167787
Coq7	12850	1.756556616
C130053K05Rik (Sptc3)	228677	1.748246754
Hgf	15234	1.741859401
Ceacam18	72431	1.7364201
Hdh2	76987	1.736153867
Spt1	20770	1.730427333
Rhag	19743	1.723439531
Bicd1	12121	1.723420563
Lcn5	13863	1.723281567
Pcbp1	18795	1.719723374
Csn1s2a	12993	1.71785104
Nars	70223	1.717359168
Hoxb2	103889	1.709723535
Dkk2	56811	1.702940912
Cldn2	12738	1.70261461
LOC634026 (5033413D22Rik)	634026*	1.700403887
Orm1	18405	1.693426474
4932442A14Rik*	320136*	1.687848729
Rab27b	80718	1.686869095
AK036971		1.686705592
Cth (E0374403)	12628	1.685673088
1-Mir	72925	1.682859033
Cadps	27062	1.679966135
Apc*	11789*	1.679742857
Pdzd3	170761	1.676373672
LOC670762*	670762*	1.673954851
Ugt2b5	22238	1.671161674
Fgfr1op2	67529	1.67028279
Spt2	20771	1.667852657
Crisp3	11572	1.660773898
Clec2e	232409	1.641374395
Fkbp11	66120	1.630533373
1810023F06Rik	217845	1.629981363
Sloc1a1	28248	1.628671922
Ins2	16334	1.626888307
F10	14058	1.626189392
A1747448	99709	1.622312089
Cd244	18106	1.603785128
2610029K11Rik*	69250*	1.602757356
Kras	16653	1.601991098
AK162044*	211147*	1.601907049
Hsd17b6	27400	1.600421446
B930006L02Rik	319604	1.597839231
Cdh26	381409	1.59437262
AK077405		1.592829045
2210412D01Rik	70178	1.592458989
Ehrp2	18606	1.591270253
Smpc	17235	1.583822126
Gm94	225443	1.579071898
(Apoa1)		1.578198864
Phoc	18155	1.575937855
Cyp3a13*	13113*	1.57557296
D13000918Rik*	320249*	1.56947179
2810423A18Rik (Glpt1)	72178	1.564023014
Cts6	58518	1.56258383
Mgat4c	67569	1.560138111
D730047E02Rik*	77860*	1.558244743
Tm4sf20	66261	1.552889658
Oas11	243262	1.550602417
Tad2	56720	1.547800752
Asah2	54447	1.544407906
Jgjp1	60440	1.543385303
Lcn3	16820	1.540962352
Thns12	232078	1.540666837
Caena2d3	12294	1.532357076
Rerg	232441	1.53061055
NAP102794-1		1.529330904
Trim21	20821	1.528035321
Ngn	83485	1.527886955
Crisp1	11571	1.526720284
Hhr4	330723	1.526678556
Zfp97	22759	1.51981897
Iapp	15874	1.517442043
Amy1	11722	1.517030811
Oma1	17228	1.515283026
Dera	232449	1.510745714
Mki	74568	1.509253661
Sema3e	20349	1.505561794
Ugt2b37	112417	1.504301067
Sult1b1	56362	1.502498646
Ptptr	19281	1.502466541

**Supplementary Table 1** - Genes significantly upregulated ( $p < 0.01$ ) in *Mbd1*<sup>+/+</sup> relative to *Mbd1*<sup>-/-</sup> mTECs. Genes are labeled by gene symbol, gene ID, and ranked in descending order by fold change. Genes are color coded by whether they are tissue-specific antigens (TSAs, yellow), housekeeping (green), below the threshold in the database (blue), and not accessible in the database (white).