Supporting Information

Morikawa et al. 10.1073/pnas.1312717110

SI Materials and Methods

Processing of Cab Analysis of Gene Expression Data. Regions with differential levels of transcription initiation were defined based on a method described by Audic and Claverie (1). In brief, the probability of observing y mapped cab analysis of gene expression (CAGE) reads in a certain TSS region in sample 1; given x, the number of mapped reads in sample 2 is given by

$$p(y|x) = \left(\frac{N_2}{N_1}\right)^y \frac{(x+y)!}{x!y!\left(1+\frac{N_2}{N_1}\right)^{(x+y+1)}}, \quad \text{[S1]}$$

where N_1 and N_2 represent the total numbers of mapped CAGE tags in samples 1 and 2, respectively. A P value for differential transcription initiation in the TSS region was calculated from the cumulative distribution of these probabilities:

$$p(y|x) = \sum_{i=y}^{\infty} p(i\,|\,x) = 1 - \sum_{i=0}^{y-1} p(i\,|\,x). \quad \text{[S2]}$$

Finally, a correction for multiple testing was done by multiplying $p(y\,|\,x)$ by the number of tests (61,949; e.g., the number of TSS clusters containing at least one mapped read in the samples of interest). The threshold P value for significant differential transcription initiation was set to 0.01 (after correction for multiple testing).

Regulatory Motif Overrepresentation Analysis. A set of 543 position weight matrices (PWMs) was prepared from the Jaspar database (2), including binding preferences for human, mouse, and rat transcription factors as well as core promoter motifs. For each PWM, a threshold score was set in a way that resulted in about one predicted site per PWM per 5 kb.

For additional analysis, we made a distinction between the enrichment of regulatory motifs in (*i*) a set of promoters relative to the genomic set of promoters, (*ii*) a set of promoters relative to another set of promoters, and (*iii*) a set of regions relative to the genome-wide tendencies.

1. Overrepresentation index analysis of a set of promoters relative to the genomic set of promoters. Transcription start sites (TSSs) of CAGE data were decided in FANTOM5 (Functional Annotation of the Mammalian Genome 5) as robust peaks (version mm9 of the mouse genome was used), and promoter regions were defined as 500 bp upstream of summits of the TSS clusters. For a given set of promoter sequences S for each PWM motif, the overrepresentation index (ORI) was calculated as follows (3):

$$\mathrm{ORI}_s = \frac{Density_S}{Density_G} \times \frac{Proportion_S}{Proportion_G}, \quad \text{[S3]}$$

where $Density_S$ and $Density_G$ represent the average number of predicted transcription factor binding sites (TFBSs) per sequence for the PWM of interest in the set of promoters S and the genome-wide set of promoters G, respectively, and $Proportion_S$ and $Proportion_G$ represent the proportion of sequences containing at least one predicted TFBS for the PWM of interest in the set of promoters S and the genome-wide set of promoters G, respectively.

The significance of each ORI_S value was evaluated using a random sampling approach, where from the genome-wide set of promoters, a set of promoters was randomly sampled and an $\mathrm{ORI}_{Sampled}$ was calculated. To limit the influence of guanine cytosine content biases, this sampling was done in a way that the sampled set contains the same number of sequences that are annotated to be associated with cytosine-phosphodiester-guanine (CpG) islands as the original set S. Sampling was done 1 million times, and the P values of ORI_S were defined as the ratio of $\mathrm{ORI}_{Sampled}$ values being equal to or higher than ORI_S.

2. ORI analysis of a set of promoters relative to another set of promoters. The approach is basically identical to the one using the genome-wide set of promoters as reference, except that now ORI_S is defined as

$$\mathrm{ORI}_S = \frac{Density_S}{Density_{S'}} \times \frac{Proportion_S}{Proportion_{S'}}, \quad \text{[S4]}$$

where S' is a second set of promoters, different from the genome-wide set, used as reference. Here, a random sampling approach is also used for estimating the significance of ORI_S. Two sets of sequences are sampled: one corresponding to S and one corresponding to S'. From the predicted TFBSs in both sets, $\mathrm{ORI}_{Sampled}$ is calculated. Importantly, the randomly sampled sets contain the same number of sequences associated with CpG islands as in S and S', respectively. Sampling was done 10,000 times, and the P value of ORI_S was defined as described above.

3. ORI analysis a set of regions relative to the genome-wide tendencies. In this case, predicted TFBSs in the regions from position −500 to +500 relative to the summit of the mapped read peak representing differentially demethylated regions or forkhead box P3 (Foxp3) -bound regions were used. We observed that, in practice, these regions had specific GC content profiles that were hard to account for using random sampling approaches. Instead, we used the genome-wide tendencies of each motif to construct a large reference set of artificial sequences with the same GC content profile as the input set of sequences. The ORI_S was defined the same as in Eq. **S3**, where G' represents the large set of artificial sequences. The significance of ORI_S was estimated as described above by randomly sampling from G'.

In 20 sets of promoter sequences of genes with differential expression between regulatory T (Treg) and conventional T (Tconv) cells and between nonstimulated and stimulated Tconv and Treg cells, ORI_S values and corresponding P values were calculated for all 543 PWMs. A subset of 117 PWMs that had $\mathrm{ORI}_S > 2$ and P value <0.001 in at least 1 of 20 sets of promoters was picked up for the construction of a heat map (Fig. S4*A*). The heat map was constructed by hierarchical clustering using the pairwise Pearson correlation coefficient of log(ORI) values as a measure of similarity between pairs of PWMs (rows); 20 sets of promoters were clustered in the same way (columns). Color intensities in the heat map reflect ORI values.

1. Audic S, Claverie JM (1997) The significance of digital gene expression profiles. *Genome Res* 7(10):986–995.
2. Bryne JC, et al. (2008) JASPAR, the open access database of transcription factor-binding profiles: New content and tools in the 2008 update. *Nucleic Acids Res* 36(Database issue):D102–D106.
3. Bajic VB, Choudhary V, Hock CK (2004) Content analysis of the core promoter region of human genes. *In Silico Biol* 4(2):109–125.

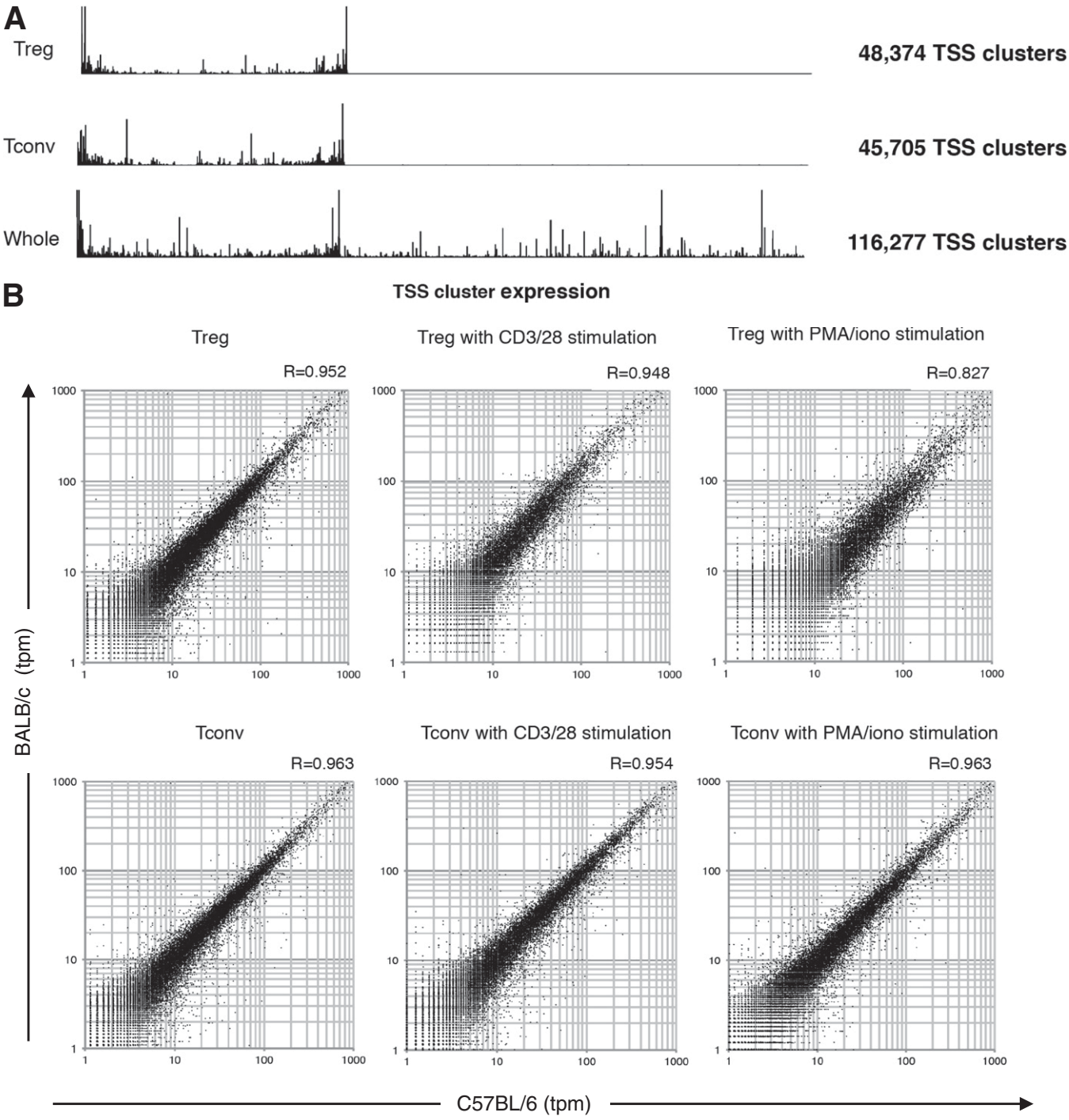
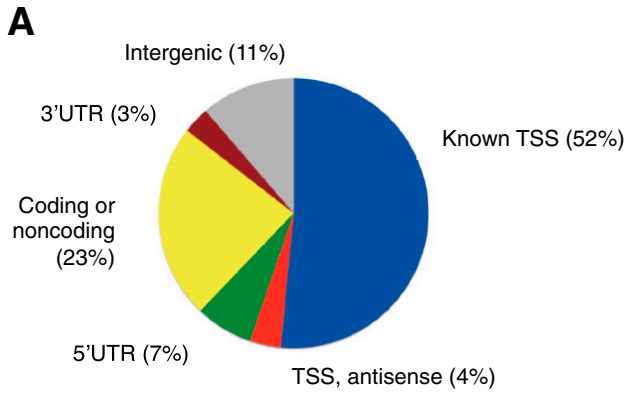
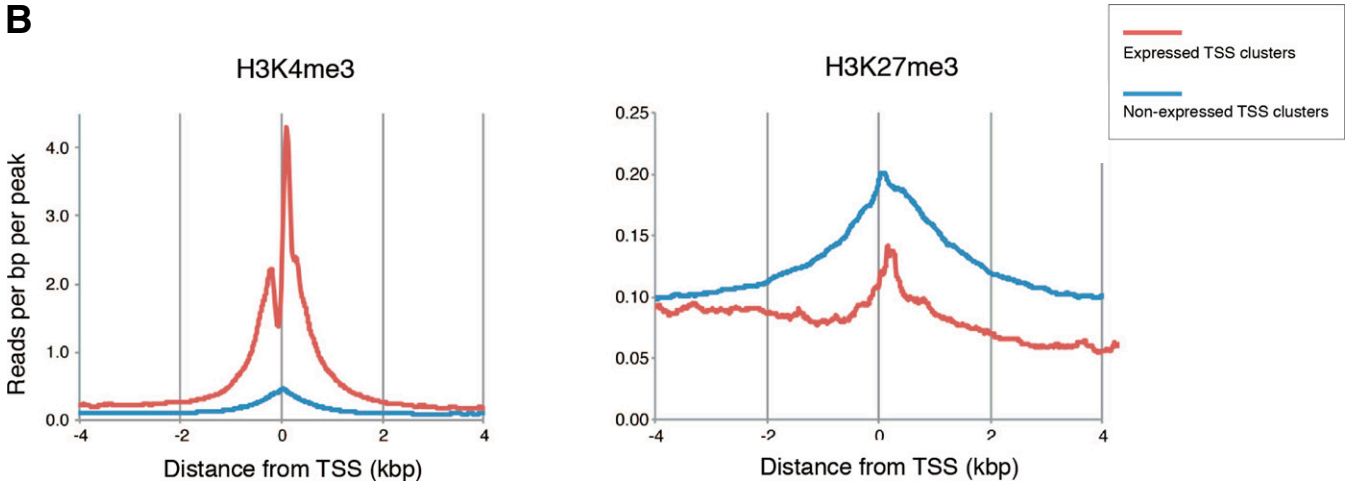


Fig. S1. Overview of identified TSS clusters. (*A*) Distribution of TSS clusters identified in Treg and Tconv cells by CAGE; *x* and *y* axes indicate TSS clusters and normalized expression levels, respectively. (*B*) TSS cluster expression profiles are mostly conserved between the C57BL/6 and BALB/c mouse strains. Comparison of TSS cluster expression profiles between C57BL/6 (*x* axis) and BALB/c (*y* axis) mouse strains. Expression profiles of Treg and Tconv cells were examined under nonstimulated, CD3/CD28-stimulated (CD3/28), and phorbol 12-myristate 13-acetate/ionomycin-stimulated (PMA/iono) conditions.



C

Motif	Name (Family)	ORI	p-value
	REL (Rel)	2.888	$<1.00\times10^{-6}$
	TATA-Box	2.868	$<1.00\times10^{-6}$
	TATA-binding (Tbp)	2.868	$<1.00\times10^{-6}$
	NF-kappaB (Rel)	2.859	$<1.00\times10^{-6}$
	RELA (Rel)	2.833	$<1.00\times10^{-6}$
	ISGF3G (Irf)	2.571	2.20×10^{-5}
	IRF4 (Irf)	2.305	1.60×10^{-5}
	NFKB1 (Rel)	2.175	$<1.00\times10^{-6}$



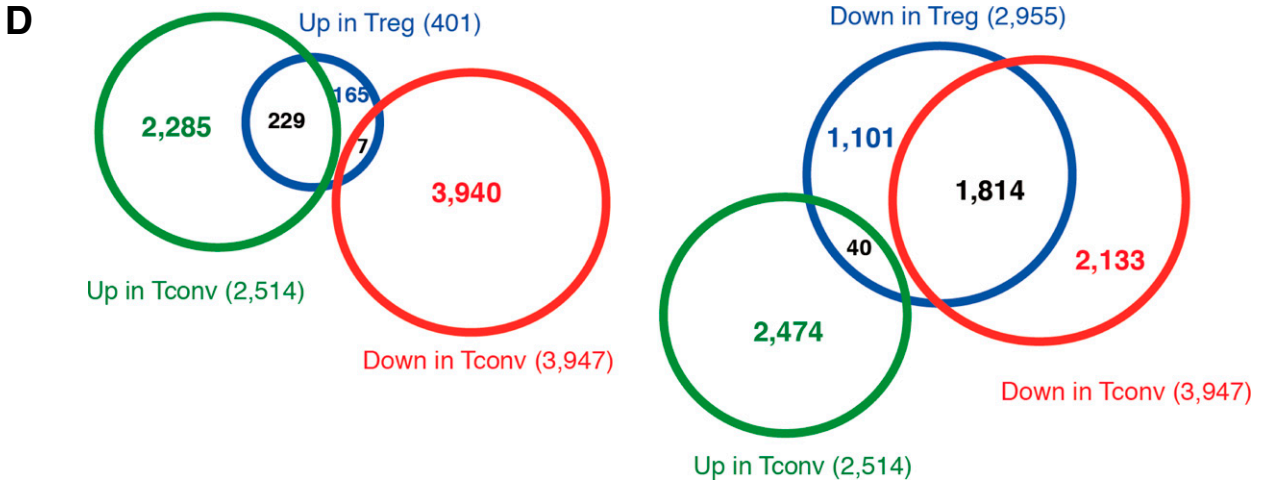


Fig. S2. Characteristics of TSS cluster in Treg and Tconv cells. (*A*) Annotation of TSS clusters identified in Tconv cells. (*B*) The positional relationships between histone H3 lysine 4 trimethylation (H3K4me3) and histone H3 lysine 27 trimethyaltion (H3K27me3) modification and TSS clusters. Density distribution of H3K4me3 or H3K27me3 ChIP-Seq tags (1) was calculated in the function of the distance from TSS clusters. (*C*) Shown are enriched transcription factor binding motifs in the promoter regions of TSS clusters that are up-regulated in Treg cells. Enriched transcription factor binding motifs, family names to which they belong, ORI scores, and *P* values are shown. (*D*) Venn diagrams illustrating the overlap in TSS clusters. Up- or down-regulated TSS clusters after T-cell receptor (TCR) stimulation in Treg cells were compared with those clusters in Tconv cells.

1. Wei G, et al. (2009) Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. *Immunity* 30(1):155–167.

A. Motifs enriched in up-regulated TSS clusters in Treg

Motif Name	Family	ORI	ORI_p
Hif1	Helix-Loop-Helix	4.6342	$<1.00\times10^{-6}$
Sp100	Sand	4.4966	$<1.00\times10^{-6}$
Myc	Helix-Loop-Helix	4.0624	$<1.00\times10^{-6}$
Mycn	Helix-Loop-Helix	3.7038	5.00×10^{-6}
Max	Helix-Loop-Helix	3.5703	1.30×10^{-5}
E2f3	E2F	3.4867	3.70×10^{-5}
Usf1	Helix-Loop-Helix	3.4627	2.31×10^{-4}
Nfya	NFY CCAAT-binding	3.3123	2.20×10^{-5}
Myc	Helix-Loop-Helix	3.2370	6.40×10^{-5}
Jundm2	Leucine Zipper	3.0566	1.11×10^{-4}
TATA-Box		2.8941	1.00×10^{-6}
Tbp	TATA-binding	2.8941	2.00×10^{-6}
Creb1	Leucine Zipper	2.7743	4.31×10^{-4}
Hdx	Homeo	2.6729	3.00×10^{-6}
Homez	Homeo	2.6189	9.70×10^{-5}
Gfi	BetaBetaAlpha-zinc finger	2.5683	1.61×10^{-4}
Rel	Rel	2.4711	1.70×10^{-4}
Atf1	Leucine Zipper	2.2991	7.17×10^{-4}

B. Motifs enriched in down-regulated TSS clusters in Treg

Motif Name	Family	ORI	ORI_p
GABPA	Ets	3.0733	$<1.00\times10^{-6}$
SFPI1	Ets	2.8290	$<1.00\times10^{-6}$
FEV	Ets	2.8101	$<1.00\times10^{-6}$
ELK4	Ets	2.5927	$<1.00\times10^{-6}$
EHF	Ets	2.5850	$<1.00\times10^{-6}$
EF5	Ets	2.5232	$<1.00\times10^{-6}$
SPI1	Ets	2.4644	$<1.00\times10^{-6}$
ELF3	Ets	2.4053	$<1.00\times10^{-6}$
ETS1	Ets	2.2357	$<1.00\times10^{-6}$
IRF1	Irf	2.1957	$<1.00\times10^{-6}$
ISGF3G	Irf	2.1738	$<1.00\times10^{-6}$
IRF4	Irf	2.1488	$<1.00\times10^{-6}$
SPDEF	Ets	2.1326	$<1.00\times10^{-6}$
NFYA	NFY CCAAT-binding	2.1108	$<1.00\times10^{-6}$
ELK1	Ets	2.0026	$<1.00\times10^{-6}$

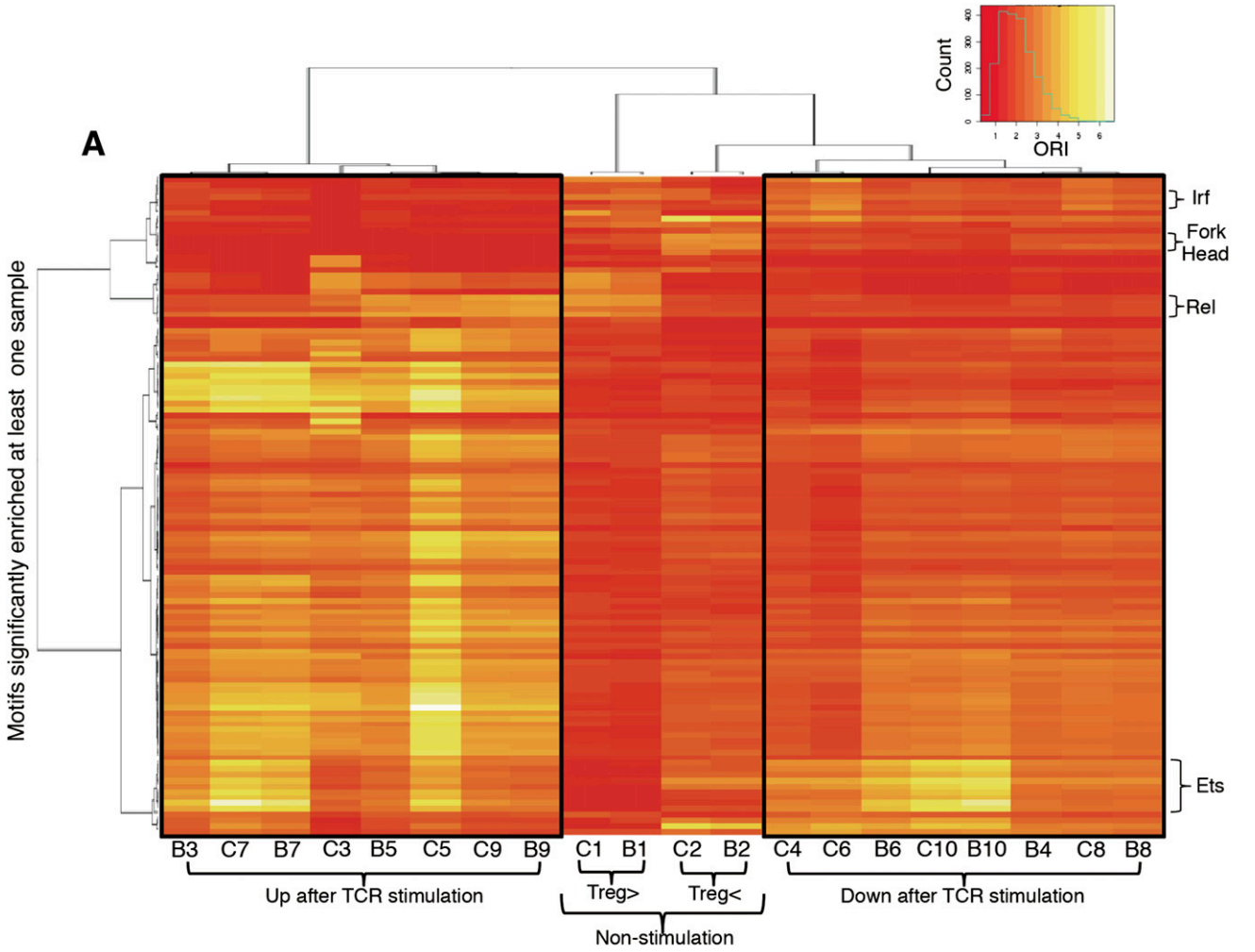
C. Motifs enriched in up-regulated TSS clusters in Tconv

Motif Name	Family	ORI	ORI_p
SPI1	Ets	5.4988	$<1.00\times10^{-6}$
MYC	Helix-Loop-Helix	4.8777	$<1.00\times10^{-6}$
USF1	Helix-Loop-Helix	4.8764	$<1.00\times10^{-6}$
MAX	Helix-Loop-Helix	4.5491	$<1.00\times10^{-6}$
MYCN	Helix-Loop-Helix	4.5263	$<1.00\times10^{-6}$
ELK4	Ets	4.4094	$<1.00\times10^{-6}$
HIF1A	Helix-Loop-Helix	4.3317	$<1.00\times10^{-6}$
BHLHB2	Helix-Loop-Helix	4.2156	$<1.00\times10^{-6}$
ETS1	Ets	4.2125	$<1.00\times10^{-6}$
GABPA	Ets	4.0577	$<1.00\times10^{-6}$
Arnt	Helix-Loop-Helix	3.8073	$<1.00\times10^{-6}$
ELK1	Ets	3.8061	$<1.00\times10^{-6}$
EHF	Ets	3.6765	$<1.00\times10^{-6}$
SP100	Sand	3.5380	$<9.40\times10^{-5}$
MIZF	BetaBetaAlpha-zinc finger	3.4071	$<1.00\times10^{-6}$
SP4	BetaBetaAlpha-zinc finger	3.1651	$<1.00\times10^{-6}$
SPDEF	Ets	3.1161	$<1.00\times10^{-6}$
CREB1	Leucine Zipper	2.8884	1.00×10^{-6}
NFYA	NFY CCAAT-binding	2.7775	$<1.00\times10^{-6}$
FEV	Ets	2.7112	$<1.00\times10^{-6}$
PAX2	Homeo	2.5710	1.00×10^{-5}
TBP	TATA-binding	2.4549	$<1.00\times10^{-6}$
JUNDM2	Leucine Zipper	2.4257	$<1.00\times10^{-6}$
CREB1	Leucine Zipper	2.4148	$<1.00\times10^{-6}$
ATF1	Leucine Zipper	2.4075	$<1.00\times10^{-6}$
ELF3	Ets	2.2986	$<1.00\times10^{-6}$
REL	Rel	2.2442	$<1.00\times10^{-6}$
SFPI1	Ets	2.1354	$<1.00\times10^{-6}$
LHX8	Homeo	2.0894	1.61×10^{-4}

D. Motifs enriched in down-regulated TSS clusters in Tconv

Motif Name	Family	ORI	ORI_p
ASCL2	Helix-Loop-Helix	2.2853	$<1.00\times10^{-6}$
BCL6B	BetaBetaAlpha-zinc finger	2.2315	$<1.00\times10^{-6}$
BREU		2.2047	1.00×10^{-6}
E2F2	e2f	2.2406	3.00×10^{-6}
EGR1	BetaBetaAlpha-zinc finger	2.1327	$<1.00\times10^{-6}$
EHF	Ets	2.2845	$<1.00\times10^{-6}$
ELF5	Ets	2.2192	$<1.00\times10^{-6}$
ELK4	Ets	2.0542	$<1.00\times10^{-6}$
FEV	Ets	2.4013	$<1.00\times10^{-6}$
GABPA	Ets	2.6024	$<1.00\times10^{-6}$
GATA6	Gata	2.1742	8.80×10^{-5}
GC-BOX		2.4041	$<1.00\times10^{-6}$
GMEB1	Sand	2.0807	1.53×10^{-4}
IRF1	Irf	2.3999	$<1.00\times10^{-6}$
IRF4	Irf	2.2019	$<1.00\times10^{-6}$
ISGF3G	Irf	2.1019	$<1.00\times10^{-6}$
KLF4	BetaBetaAlpha-zinc finger	2.2461	$<1.00\times10^{-6}$
KLF7	BetaBetaAlpha-zinc finger	2.3037	$<1.00\times10^{-6}$
MED-1		2.1549	4.40×10^{-4}
MTE		2.0875	4.00×10^{-6}
NFYA	NFY CCAAT-binding	2.2319	$<1.00\times10^{-6}$
SFPI1	Ets	2.1785	$<1.00\times10^{-6}$
SMAD3	Mh1	2.2806	$<1.00\times10^{-6}$
SP1	BetaBetaAlpha-zinc finger	2.3166	$<1.00\times10^{-6}$
SP100	Sand	2.1324	1.00×10^{-6}
SP4	BetaBetaAlpha-zinc finger	2.2822	$<1.00\times10^{-6}$
STAT1	Stat	2.1438	$<1.00\times10^{-6}$
XCPE1		2.1040	$<1.00\times10^{-6}$
ZFP161	BetaBetaAlpha-zinc finger	2.0797	8.19×10^{-4}
ZFP410	BetaBetaAlpha-zinc finger	2.3446	$<1.00\times10^{-6}$

Fig. S3. Overrepresented regulatory motifs in promoters of TSS clusters regulated by TCR stimulation. Shown are significantly overrepresented regulatory motifs in promoter regions of TSS clusters that were regulated by TCR stimulation. Enriched transcription factor binding motifs, their ORI scores, and *P* values are shown. Motifs found in promoters of TSS clusters (*A*) up- or (*B*) down-regulated after TCR stimulation in Treg cells. Motifs found in promoters of TSS clusters (*C*) up- or (*D*) down-regulated after TCR stimulation in Tconv cells.



	Treg vs. Tconv	
	Up	Down
C57BL/6	C1	C2
BALB/c	B1	B2

	Treg cells			
	CD3/28 up	CD3/28 down	PMA/iono up	PMA/iono down
C57BL/6	C3	C4	C5	C6
BALB/c	B3	B4	B5	B6

	Tconv cells			
	CD3/28 up	CD3/28 down	PMA/iono up	PMA/iono down
C57BL/6	C7	C8	C9	C10
BALB/c	B7	B8	B9	B10

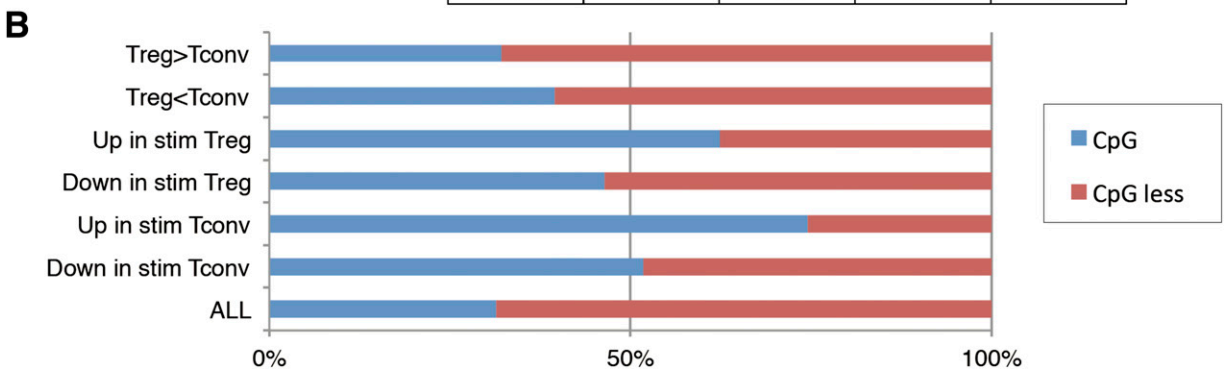
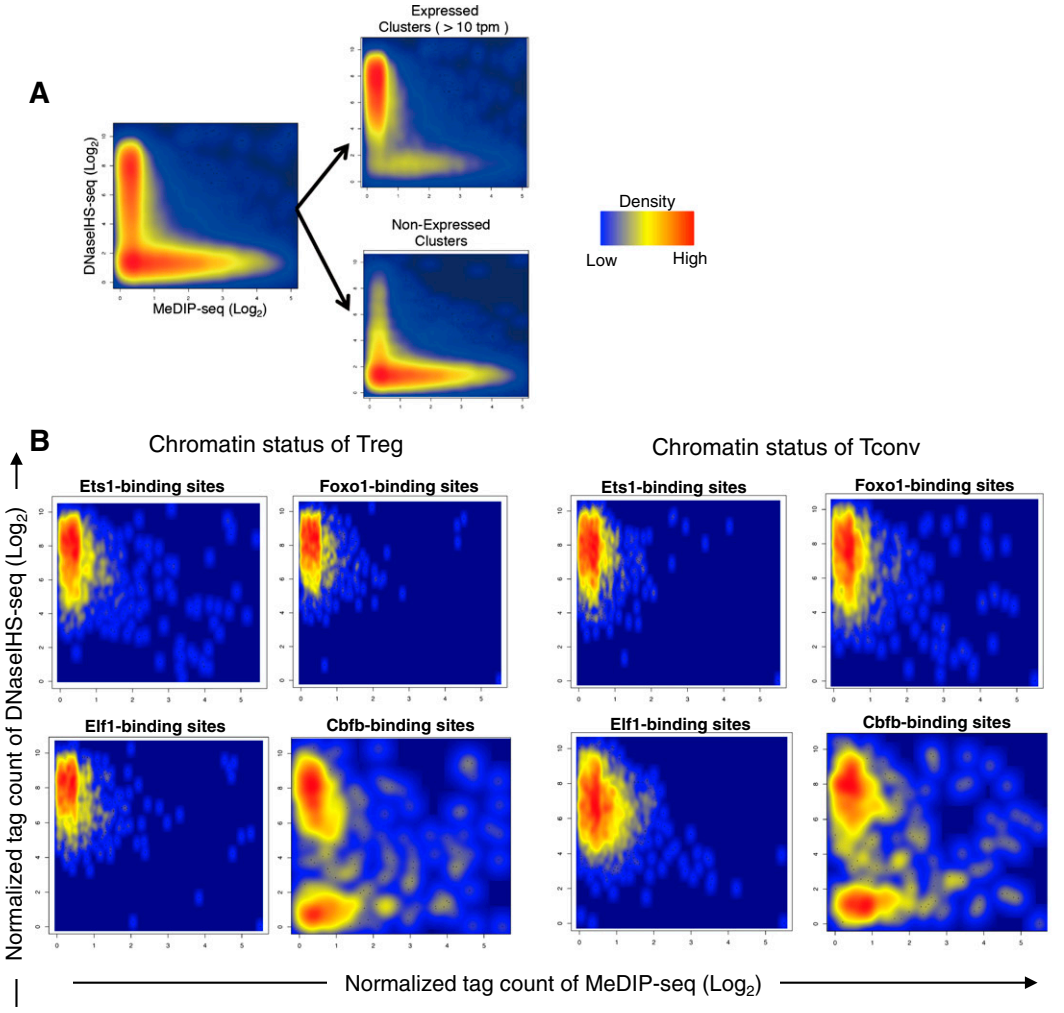


Fig. S4. Transcription factors associated with TCR stimulation-dependent regulation are mostly shared between Treg and Tconv cells. (*A*) Heat map showing ORI values of transcription factor binding motifs enriched in the promoter regions of TSS clusters significantly up- or down-regulated by TCR stimulation; *x* axis indicates sets of up- or down-regulated TSS clusters after TCR stimulation under different conditions. Samples are summarized in the table below the heat map. *y* Axis indicates motifs significantly enriched in at least one set of promoter regions. CD3/28, anti-CD3 and anti-CD28 antibody stimulation; down, down-regulated TSS clusters after TCR stimulation; Treg>, up-regulated TSS clusters in nonstimulated Treg cells compared with Tconv cells; Treg<, down-regulated TSS clusters in nonstimulated Treg cells compared with Tconv cells; up, up-regulated TSS clusters after TCR stimulation. (*B*) The ratio of TSS clusters possessing CpG islands in their promoters. Treg > Tconv and Treg < Tconv represent TSS clusters showing up- and down-regulation in Treg cells compared with Tconv cells, respectively. Up in stim and down in stim represent TSS clusters up- and down-regulated after TCR stimulation, respectively.



C

Motif	Name (Family)	ORI	p-value
	MYB (Myb)	4.0921	<1.00x10^{-6}
	ARNT (Helix-Loop-Helix)	3.5149	<1.00x10^{-6}
	HIF1A (Helix-Loop-Helix)	3.2057	<1.00x10^{-6}
	MTF (BetaBetaAlpha-zinc finger)	2.4081	<1.00x10^{-6}
	CREB1 (Leucine Zipper)	2.2204	<1.00x10^{-6}
	IRF5 (Irf)	2.3252	1.00x10^{-4}
	ATF1 (Leucine Zipper)	2.4824	3.00x10^{-4}
	SP100 (Sand)	2.0818	4.00x10^{-4}
	ETS1 (Ets)	2.0037	5.00x10^{-4}

D

Motif	Name (Family)	ORI	p-value
	GABPA (Ets)	9.4028	<1.00x10^{-6}
	FEV (Ets)	6.9282	<1.00x10^{-6}
	SRF (MADS)	4.7082	<1.00x10^{-6}
	ELF5 (Ets)	4.3643	<1.00x10^{-6}
	SPI1 (Ets)	4.2348	<1.00x10^{-6}
	ELF3 (Ets)	4.2142	<1.00x10^{-6}
	EHF (Ets)	4.1805	<1.00x10^{-6}
	IRF1 (Irf)	3.8712	<1.00x10^{-6}
	RUNX1 (Runt)	3.6989	<1.00x10^{-6}
	FOXD1 (Forkhead)	3.6649	<1.00x10^{-6}
	ETS1 (Ets)	3.6485	<1.00x10^{-6}
	SPDEF (Ets)	3.5522	<1.00x10^{-6}
	SFPI1 (Ets)	3.4774	<1.00x10^{-6}
	FOXO3 (Forkhead)	3.4041	<1.00x10^{-6}
	ISGF3G (Irf)	3.2942	<1.00x10^{-6}

Fig. S5. Genome status of several regulatory regions. (*A*) Relationships between DNA methylation status and DNaseI hypersensitivity. Heat maps show normalized tag counts of methylated DNA immunoprecipitating sequencing (MeDIP-seq; *x* axis) and DNaseI-HS-seq (*y* axis) within 500 bp from all TSS clusters. *Upper Right* and *Lower Right* show the distribution of TSS clusters with more than 10 tags per million tags (tpm) expression and those clusters without significant expression in Treg cells, respectively. (*B*) DNA methylation status and DNaseI hypersensitivity of several transcription factor binding sites. Heat maps show normalized tag counts of MeDIP-seq (*x* axis) and DNaseI-HS-seq (*y* axis) of the indicated transcription factor binding sites (500-bp region) in (*Left*) Treg and (*Right*) Tconv cells. (*C*) Transcription factor binding motifs significantly enriched within 500 bp from Treg-specific DNA demethylated regions (TSDRs). (*D*) Transcription factor binding motifs significantly enriched within 500 bp from Foxp3 binding sites.

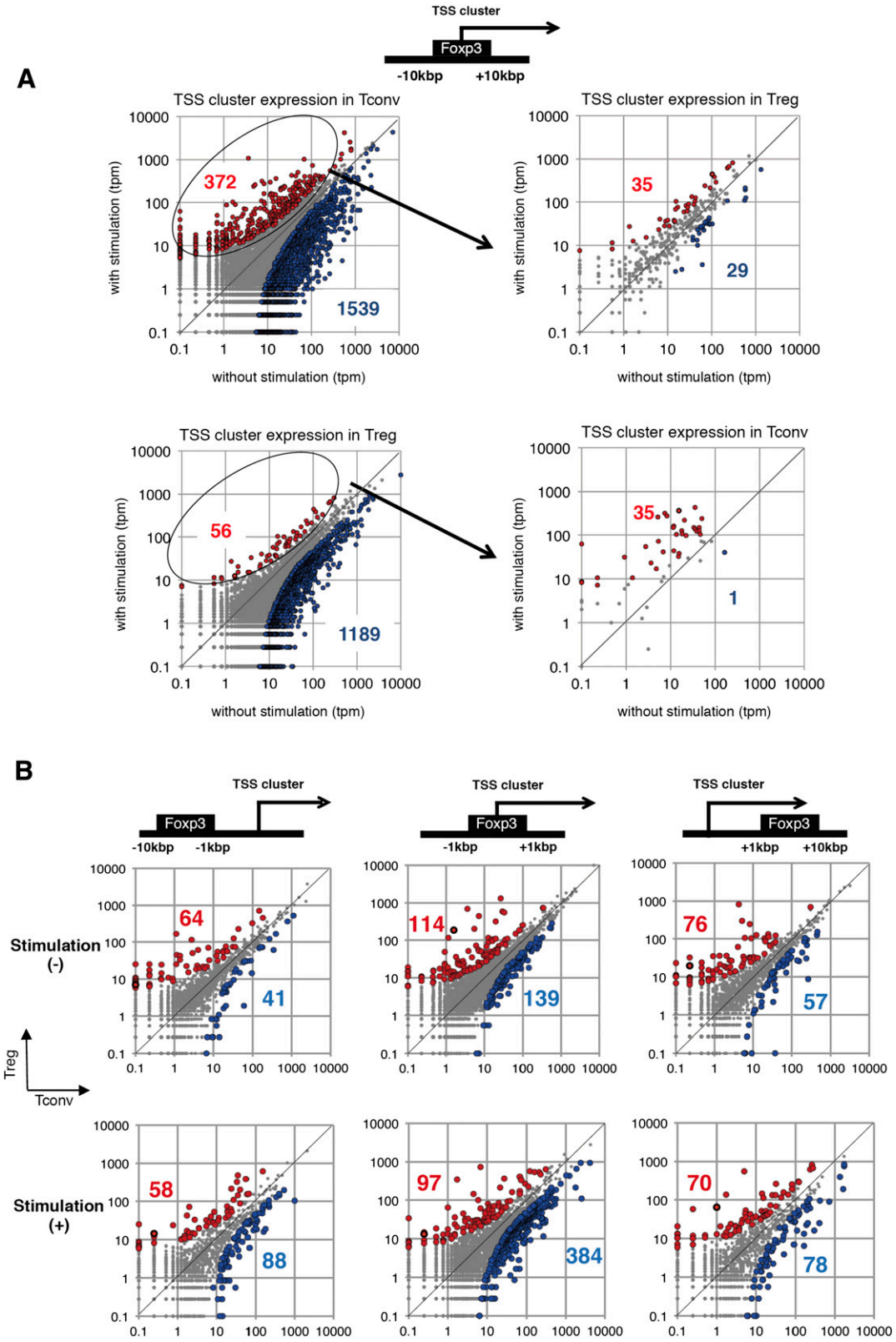


Fig. S6. Expression profiles of TSS cluster in relation to Foxp3 binding sites. (*A*, *Upper*) TSS clusters significantly up-regulated in Tconv cells after TCR stimulation were plotted onto the Treg expression profiles compared between nonstimulated and stimulated conditions. (*A*, *Lower*) TSS clusters up-regulated in Treg cells after TCR stimulation were also plotted onto the Tconv expression profiles under nonstimulated and stimulated conditions. (*B*) Expression profiles of TSS clusters sorted by the distance from Foxp3 binding sites to TSS clusters. *Upper* and *Lower* show TSS cluster expression without and with TCR stimulation, respectively. Red and blue dots indicate significantly up- or down-regulated TSS clusters in Treg cells compared with Tconv cells, respectively.

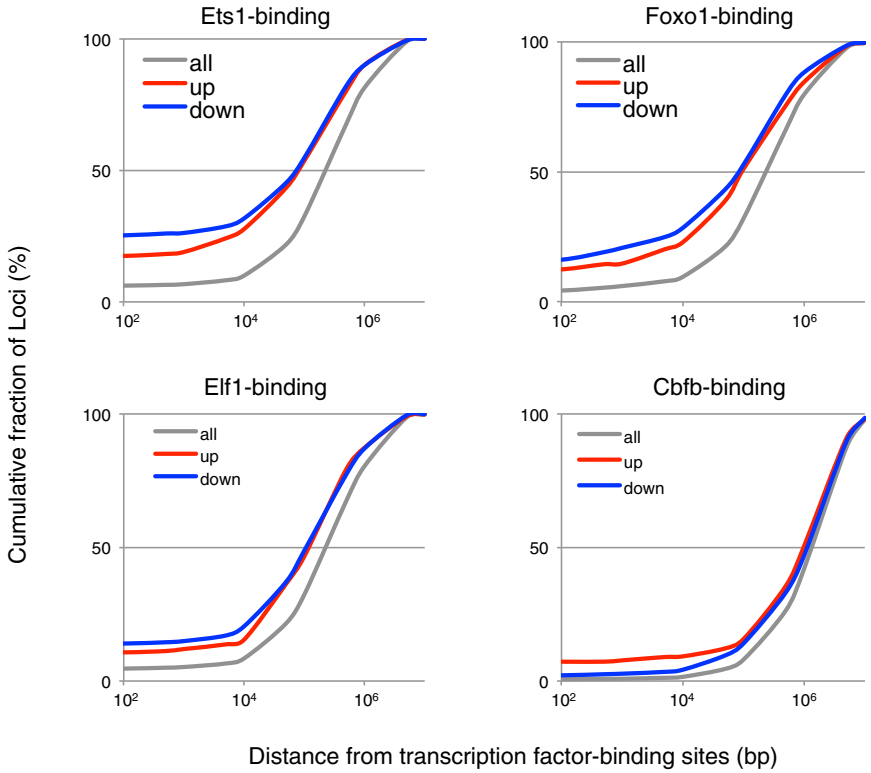


Fig. S7. Relationship between transcription factor binding sites and their transcriptional regulation. Cumulative distribution of TSS clusters within 10-Mbp regions from the indicated transcription factor binding sites. Red and blue lines indicate TSS clusters significantly up- and down-regulating TSS clusters in Treg cells after TCR stimulation, respectively. Cumulative distributions of all TSS clusters are also shown as a negative control (gray line).

A. Cage samples

Samples	Background	Total uniquely mapped tag count
$CD4^+CD25^+$ Treg	C57BL/6	3,644,175
$CD4^+CD25^+$ Treg with αCD3/28 stimulation	C57BL/6	3,565,379
$CD4^+CD25^+$ Treg with PMA/Ionomycin stimulation	C57BL/6	1,492,441
$CD4^+CD25^-$CD44low Tconv	C57BL/6	4,434,387
$CD4^+CD25^-$CD44low Tconv with αCD3/28 stimulation	C57BL/6	3,985,267
$CD4^+CD25^-$CD44low Tconv with PMA/Ionomycin stimulation	C57BL/6	7,852,126
$CD4^+CD25^+$ Treg	BALB/c	6,217,721
$CD4^+CD25^+$ Treg with αCD3/28 stimulation	BALB/c	3,041,346
$CD4^+CD25^+$ Treg with PMA/Ionomycin stimulation	BALB/c	3,660,522
$CD4^+CD25^-$CD44low Tconv	BALB/c	8,210,014
$CD4^+CD25^-$CD44low Tconv with αCD3/28 stimulation	BALB/c	7,893,290
$CD4^+CD25^-$CD44low Tconv with PMA/Ionomycin stimulation	BALB/c	4,927,526

B. Data Production

Sample	Insert Size (bp)	Read Size (bp)	Total Reads	Raw Bases (Gb)
Tconv	426	49	97,959,184	4.8
Treg	418	49	97,959,184	4.8

C. Reads Alignment

Sample	Total Reads	Mapped Reads	Mapped Rate (%)	Mapped Base (Gb)	Unique Mapped Reads	Unique Mapped Rate (%)
Tconv	97,959,184	86,085,258	87.88	4.2	37,191,993	37.97
Treg	97,959,184	86,304,240	88.10	4.8	38,534,628	39.34

D. Peak Analysis with MACS1.4

Sample	Total Peaks	Peak Mean Length (bp)	Peak Median Length (bp)	Peak Total Length (bp)	Peak Covered Rate In Genome (%)
Tconv	156743	1,077.2	987	168,842,581	9.03
Treg	176599	1,069.5	965	188,871,383	10.10
TSDR in Tconv vs Treg	301	799.6	775	240,666	0.01

Fig. S8. High-throughput data production by next generation sequencer. (*A*) Overview of CAGE datasets used in this study. Cell type, background, and uniquely mapped CAGE tag counts are shown. Data production and analysis of MeDIP-seq data are also shown. (*B*) Data production of MeDIP-seq data derived from Treg and Tconv cells using the Hiseq-2000 sequencer. (*C*) Read alignment of MeDIP-seq data by bowtie. (*D*) Peak analysis of all MeDIP-seq data by MACS.

Table S1. Significantly expressed TSS clusters in Treg vs. Tconv TOP100

Cluster location									
Chromosome	Start	End	Strand	Treg (tpm)	Tconv (tpm)	Log(1/*P* value)	Location	Distance	Related gene
chrX	7156792	7156811	+	826.5	4.3	744.5	TSS,S	0	FOXP3
chr2	11564469	11564490	+	691.0	3.6	640.8	TSSregion500,S	51	IL2RA
chr1	60965837	60965877	+	438.8	7.2	345.4	TSS,S	0	CTLA4
chr1	146096179	146096197	–	591.3	30.0	297.3	TSSregion500,S	36	RGS1
chr1	60965929	60965946	+	545.8	33.2	270.6	TSSregion500,S	61	CTLA4
chr4	155387944	155387958	+	298.8	5.2	252.4	TSSregion500,S	141	TNFRSF4
chr2	11564426	11564439	+	187.7	1.6	195.8	TSSregion500,S	8	IL2RA
chr7	105638493	105638511	+	164.9	0.5	193.7	Intergenic	0	NA
chr1	21315817	21315821	–	169.0	1.1	183.7	Intergenic	0	NA
chr4	155387967	155387981	+	248.1	8.4	183.1	TSSregion500,S	164	TNFRSF4
chr9	107199174	107199231	+	313.6	19.9	175.3	TSSregion500,S	155	CISH
chr15	78325467	78325478	–	202.5	5.4	167.5	TSSregion500,S	17	IL2RB
chr9	14664676	14664690	+	240.1	11.3	161.7	TSSregion500,S	−7	GPR83
chr9	14706968	14707003	–	702.5	146.6	155.5	5UTR_first_intron,S	1,391	FOLR4
chr4	150302128	150302144	+	256.8	20.1	139.4	TSSregion500,S	21	TNFRSF9
chr2	11564413	11564425	+	117.7	1.2	127.0	TSS,S	0	IL2RA
chr3	95543477	95543499	–	340.5	55.9	120.3	TSS,S	0	ECM1
chr13	43880537	43880572	+	116.0	2.7	109.1	TSSregion500,S	62	CD83
chr1	155587474	155587490	+	106.2	2.0	104.8	TSS,S	0	RGS16
chr12	104681863	104681883	–	730.8	331.7	99.2	TSSregion500,S	6	IFI27L2A
chr2	61416606	61416674	+	184.4	21.3	93.0	TSS,S	0	TANK
chr12	56593574	56593592	–	725.0	342.1	90.8	TSSregion500,S	41	NFKBIA
chr10	75398623	75398653	+	263.7	51.9	88.2	TSSregion500,S	306	CHCHD10
chr3	101101404	101101429	+	77.4	0.9	84.6	Intergenic	0	NA
chr6	72499424	72499472	+	107.0	6.1	81.4	TSS,S	0	CAPG
chr8	86568881	86568894	–	69.2	0.5	80.3	TSSregion500,S	−4	RLN3
chr17	35561787	35561801	+	341.9	94.0	80.3	TSSregion500,S	−20	H2-Q6
chr11	72904056	72904105	+	70.2	0.7	79.0	TSS,S	0	ITGAE
chr4	155400413	155400480	+	225.0	43.6	78.8	TSS,S	0	TNFRSF18
chr9	88222470	88222501	+	127.9	12.0	77.8	TSSregion500,S	24	NT5E
chr3	93359077	93359093	+	458.0	181.5	75.5	TSSregion500,S	39	S100A10
chr2	164268677	164268694	–	54.9	0.0	69.1	TSS,S	0	SDC4
chr8	130882965	130882985	+	61.7	0.9	67.0	TSS,S	0	NRP1
chr16	31039976	31040004	–	72.5	3.0	64.7	CDS_or_noncoding_inner_intron,S	41,513	AI480653
chr4	135699731	135699747	+	367.4	143.9	63.3	TSS,S	0	ID3
chr5	77543897	77543911	–	52.1	0.2	62.6	TSSregion500,S	236	HOPX
chr11	4397559	4397588	–	62.6	1.8	61.1	CDS_or_noncoding_inner_intron,S	97,229	MTMR3
chr4	4065921	4065945	–	54.6	0.9	58.9	TSSregion500,S	−330	PENK
chr18	60963492	60963510	+	127.0	21.0	58.4	TSS,S	0	CD74
chr19	9063767	9063785	+	134.7	24.8	57.3	TSS,S	0	AHNAK
chr1	153418571	153418601	+	65.0	3.2	56.5	TSSregion500,S	69	FAM129A
chr10	94879414	94879468	–	43.9	0.0	55.3	TSSregion500,S	22	SOCS2
chr7	50692502	50692544	+	52.4	1.4	53.3	TSS,S	0	NKG7
chr9	14708196	14708216	–	108.4	17.6	52.5	TSSregion500,S	178	FOLR4
chr6	29803983	29803998	+	48.6	0.9	52.1	CDS_or_noncoding_first_intron,S	85,541	AHCYL2
chr9	14708310	14708326	–	72.4	6.1	52.1	TSSregion500,S	68	FOLR4
chr4	46614743	46614783	–	57.3	2.5	52.1	TSSregion500,S	17	CORO2A
chr7	117365256	117365280	+	60.6	3.4	51.2	TSSregion500,S	40	SWAP70
chr12	114665543	114665607	–	111.3	19.9	50.7	Intergenic	0	NA
chr6	129189762	129189789	–	61.8	4.1	49.5	Intergenic	0	NA
chr9	78039813	78039830	+	56.2	2.9	48.9	TSSregion500,S	41	GSTA4
chr13	105412781	105412798	+	60.9	4.1	48.8	Intergenic	0	NA
chr3	142223536	142223566	+	63.7	5.0	48.0	5UTR_first_intron,S	521	GBP3
chr16	75909475	75909492	–	96.3	16.2	46.8	TSSregion500,S	18	SAMSN1
chr1	60967956	60967966	+	45.0	1.1	46.6	CDS_or_noncoding_first_intron,S	2,088	CTLA4
chr6	94650148	94650170	–	104.5	20.1	46.1	TSSregion500,S	−10	LRIG1
chr6	124867957	124867973	–	134.5	33.6	45.8	TSS,S	0	PTMS
chr6	108610610	108610676	+	49.9	2.3	45.7	TSS,S	0	BHLHE40
chr9	14185417	14185477	–	61.1	5.2	45.1	TSSregion500,S	208	ENDOD1
chr10	57514377	57514470	+	299.9	134.9	44.3	TSSregion500,S	28	SMPDL3A
chr7	108699289	108699321	–	51.0	3.0	43.9	Intergenic	0	NA

Table S1. Cont.

Cluster location									
Chromosome	Start	End	Strand	Treg (tpm)	Tconv (tpm)	Log(1/*P* value)	Location	Distance	Related gene
chr10	18731754	18731779	−	43.4	1.4	43.4	TSS,S	0	TNFAIP3
chr9	59788828	59788852	−	125.7	33.2	43.2	Intergenic	0	NA
chr13	23663251	23663261	+	236.0	97.4	42.7	TSSregion500,S	−7	HIST1H2BG
chr1	182947620	182947679	−	204.7	79.0	42.4	TSS,S	0	EPHX1
chr10	30375579	30375616	−	87.8	16.3	41.3	TSSregion500,S	56	NCOA7
chr1	146096229	146096243	−	51.6	3.8	41.0	TSS,S	0	RGS1
chr11	53705140	53705158	−	38.7	0.9	41.0	TSSregion500,S	46	SLC22A5
chr8	83866600	83866617	+	40.3	1.4	40.2	Intergenic	0	NA
chr1	162722930	162722963	−	45.0	2.5	39.7	TSSregion500,S	105	RABGAP1L
chr2	25562346	25562433	+	93.2	20.4	39.2	TSS,S	0	BMYC
chr15	101097270	101097290	+	181.7	67.4	39.0	TSS,S	0	NR4A1
chr10	128079984	128079997	−	32.4	0.2	38.6	CDS_or_noncoding_inner_intron,S	3,051	IKZF4
chr11	54776405	54776455	−	124.6	36.4	38.3	TSS,S	0	TNIP1
chr4	155388007	155388018	+	35.7	0.9	37.6	CDS_or_noncoding_first_exon,S	204	TNFRSF4
chr9	120001224	120001239	+	55.1	6.3	36.8	TSSregion500,S	−11	CCR8
chr19	11566512	11566543	+	68.6	11.3	36.8	TSS,S	0	GM8369
chr14	101876472	101876534	−	56.8	7.1	36.6	CDS_or_noncoding_inner_intron,S	131,873	TBC1D4
chr3	101101446	101101459	+	31.8	0.5	36.1	Intergenic	0	NA
chr6	124861654	124861694	−	43.1	3.0	36.1	TSSregion500,S	28	LAG3
chr13	105413416	105413432	−	36.2	1.4	35.7	Intergenic	0	NA
chr7	150238681	150238702	+	87.8	21.0	34.7	TSSregion500,S	27	CD81
chr2	10075190	10075254	+	54.3	7.0	34.6	TSSregion500,S	21	ITIH5
chr18	66618241	66618258	+	41.2	3.2	33.4	TSS,S	0	PMAIP1
chr6	29803746	29803790	+	35.9	1.8	33.3	CDS_or_noncoding_first_intron,S	85,304	AHCYL2
chr17	56092051	56092122	+	26.3	0.0	33.2	TSSregion500,S	6	EBI3
chr15	79722866	79722912	+	168.7	69.2	33.1	TSSregion500,S	29	APOBEC3
chr1	69729585	69729606	−	33.7	1.4	33.0	CDS_or_noncoding_inner_intron,S	2,927	IKZF2
chr15	78324299	78324312	+	36.0	2.0	32.4	5UTR_first_intron,AS	−1,183	IL2RB
chr3	60276553	60276588	+	52.1	7.3	32.1	TSSregion500,S	−163	MBNL1
chr8	84082901	84082910	+	46.7	5.4	31.9	Intergenic	0	NA
chr13	30841099	30841116	+	95.5	27.5	31.8	TSSregion500,S	−10	IRF4
chrX	7170351	7170362	+	25.0	0.0	31.5	3UTR_last_exon,S	13,550	FOXP3
chr3	68808374	68808389	−	80.4	20.1	31.3	TSSregion500,S	102	IFT80
chr5	115553549	115553569	−	138.0	52.8	31.3	Intergenic	0	NA
chr8	130883142	130883191	+	27.7	0.5	31.2	TSSregion500,S	170	NRP1
chr19	40734275	40734292	+	26.3	0.2	31.2	TSS,S	0	ENTPD1
chr1	69729731	69729752	−	25.5	0.2	30.2	CDS_or_noncoding_inner_intron,S	2,781	IKZF2
chrX	48499591	48499593	−	23.9	0.0	30.1	CDS_or_noncoding_inner_intron,S	59,415	MBNL3
chr17	26698460	26698537	+	53.4	8.9	30.0	TSSregion500,S	4	ERGIC1

CDS, Coding DNA sequence.