Dynamic recruitment of Ets1 to both nucleosome -occupied and -depleted enhancer regions mediates a transcriptional program switch during early T-cell differentiation

Pierre Cauchy[1-5,†], Muhammad A Maqbool[6,†], Joaquin Zacarias-Cabeza[1-3], Laurent Vanhille[4,5], Frederic Koch[1-3], Romain Fenouil[1-3], Marta Gut[7], Ivo Gut[7], Maria A Santana[1-3], Aurélien Griffon[4,5], Jean Imbert[4,5], Carolina Moraes-Cabé[8], Jean-Christophe Bories[8], Pierre Ferrier[1-3], Salvatore Spicuglia[4,5,*] and Jean-Christophe Andrau[6,*]

[1] CIML CNRS UMR7280, Case 906, Campus de Luminy, Marseille, F-13009, France
[2] CIML INSERM U1104, Case 906, Campus de Luminy, Marseille, F-13009, France
[3] Aix-Marseille University, 58 Bd Charles Livon, Marseille, F-13284, France
[4] Inserm U1090, Technological Advances for Genomics and Clinics (TAGC), Marseille, F-13009, France
[5] Aix-Marseille University UMR-S 1090, TAGC, Marseille, F-13009, France
[6] Institut de Génétique Moléculaire de Montpellier, CNRS UMR5535, 1919 Route de Mende, Montpellier, F-34293, France
[7] Centre Nacional D'Anàlisi Genòmica, Parc Científic de Barcelona, Baldiri i Reixac 4, Barcelona, ES-08028, Spain
[8] INSERM UMR 1126 Institut Universitaire d'Hématologie, Hôpital Saint-Louis, Paris, F-75475, France

* To whom correspondence should be addressed. Jean-Christophe Andrau; Tel: +33 4 34 35 96 52; Fax: +33 4 67 52 15; Email: jean-christophe.andrau@igmm.cnrs.fr. Correspondence may also be addressed to Salvatore Spicuglia; Tel +33 4 91 82 87 22; Fax: +33 4 91 82 87 01; Email: salvatore.spicuglia@inserm.fr.

† The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors

Present Address: Maria A Santana, Facultad de Ciencias, Universidad Autónoma del Estado de Morelos. Av. Universidad 1001, Chamilpa, Cuernavaca, Morelos, Mexico

TABLE OF CONTENTS

I. SUPPLEMENTARY TABLES AND FIGURES

Cells	Antibody (clone)	Origin	Reference	Cells / replicate	Antibody / beads	Washes (RIPA/TE)
Mouse Rag2 -/-	Ets1 (C-20)	rabbit polyclonal	Santa Cruz (sc-350x)	$1x10^8$	20µg / 200µl	8x/1x
	total (N-20)	rabbit polyclonal	Santa Cruz (sc-899x)	$2.5x10^7$	5µg / 50µl	8x/1x
	H3K4me1	rabbit polyclonal	Abcam (ab8895)	$5x10^6$	2µg / 20µl	8x/1x
	H3K4me3	rabbit polyclonal	Diagenode (pAb-003-050)	$5x10^6$	2µg / 20µl	8x/1x
	H3K27me3	mouse monoclonal	Abcam (ab6002)	$5x10^6$	2µg / 20µl	8x/1x
	Runx1	rabbit polyclonal	Abcam (ab23980)	$1x10^7$	10µg / 25µl	8x/1x
Mouse DP	E47	rabbit polyclonal	Santa Cruz (sc-349X)	$1x10^7$	10µg / 25µl	8x/1x
P5424 WT	Ets1 (C-20)	rabbit polyclonal	Santa Cruz (sc-350x)	$5x10^7$	20µg / 200µl	6x/1x
P5424 sh-scr	H3K4me1	rabbit polyclonal	Abcam (ab8895)	$5x10^6$	2µg / 20µl	6x/1x
	H3K4me3	rabbit polyclonal	Abcam (ab8580)	$5x10^6$	2µg / 20µl	6x/1x
P5424 sh-ets1	H3K4me1	rabbit polyclonal	Abcam (ab8895)	$5x10^6$	2µg / 20µl	6x/1x
	H3K4me3	rabbit polyclonal	Abcam (ab8580)	$5x10^6$	2µg / 20µl	6x/1x

Table S1. ChIP antibodies and conditions used in this study

Detailed experimental conditions used in this study.

Cells	Experiment	Platform	Aligned reads	Estimated Fragment size	Peaks
Mouse Rag2 $^{-/-}$	Ets1 ChIP-Seq biorep1	Illumina GA II	23144668	136bp	5893
	Ets1 ChIP-Seq biorep2	Illumina GA II	23363596	146bp	
	Pol II ChIP-Seq (N-20) biorep1	Illumina GA II	24416831	136bp	14385
	Pol II ChIP-Seq (N-20) biorep2	Illumina GA II	23536764	136bp	
	H3K4me1ChIP-Seq biorep1	Illumina GA II	22898035	146bp	98615
	H3K4me1ChIP-Seq biorep2	Illumina GA II	23892863	146bp	
	H3K4me3 ChIP-Seq biorep1	Illumina GA II	30644661	136bp	27429
	H3K4me3 ChIP-Seq biorep2	Illumina GA II	33685813	136bp	
	H3K27me3 ChIP-Seq biorep1 1	Illumina GA II	25742842	136bp	n.d.
	H3K27me3 ChIP-Seq biorep1 2	Illumina GA II	13182402	136bp	
	H3K27me3 ChIP-Seq biorep2 1	Illumina GA II	25721411	136bp	
	H3K27me3 ChIP-Seq biorep2 2	Illumina GA II	24774255	136bp	
	Short RNA	Illumina GA II	72126257	n.a.	n.a.
	Runx1 ChIP-Seq	SOLiD 4	12874449	125bp	7508
	MNase-Seq	SOLiD 4	86038441	146bp	n.a.
	Input	HiSeq 2000	99669331	208bp	n.a.
Mouse DP	E47 ChIP-Seq	SOLiD 4	21575830	94bp	n.d.
	MNase-Seq	SOLiD 4	71660704	146bp	n.a.
P5424 WT	Ets1 ChIP-Seq	HiSeq 2000	31423925	150bp	3565
P5424 sh-scr	H3K4me1 ChIP-Seq	HiSeq 2000	38493068	194bp	n.d.
	H3K4me3 ChIP-Seq	HiSeq 2000	46708023	197bp	n.d.
P5424 sh-ets1	H3K4me1 ChIP-Seq	HiSeq 2000	40453583	195bp	n.d.
	H3K4me3 ChIP-Seq	HiSeq 2000	37442004	170bp	n.d.

Table S2. Description and statistics of high-throughput sequencing runs used in this study.
Peak detection was performed for DN Ets1, Pol II, H3K4me1, H3K4me3, Runx1 and P5424 Ets1 only.

Table S3. Peak detection results for Ets1 ChIP-seq (as supplementary excel file)
Detailed CoCAS peak detection output for Ets1 DN and DP. Genomic coordinates in the mm9 reference genome

Table S4. Detailed list of distal Ets1 DN+DP peaks (as supplementary excel file)
Details of distal Ets1 DN+DP peaks, as well as name of nearest gene. Genomic coordinates in the mm9 reference genome

Table S5. Microarray expression values in DP *Ets1*$^{-/-}$ and WT cells (as supplementary excel file)
Normalized gene expression values in DP *Ets1*$^{-/-}$, DP *Ets1* WT and fold change.

Table S6. Details of the Ets^{can}-Ets1 composite motif (as supplementary excel file)
Mm9 genomic coordinates of Ets1 and Ets^{can} motifs in DN+DP peaks with Ets^{can} at a fixed distance of -4bases upstream of Ets1, as well as name of the closest gene and sequence of the composite motif.

Figure S1

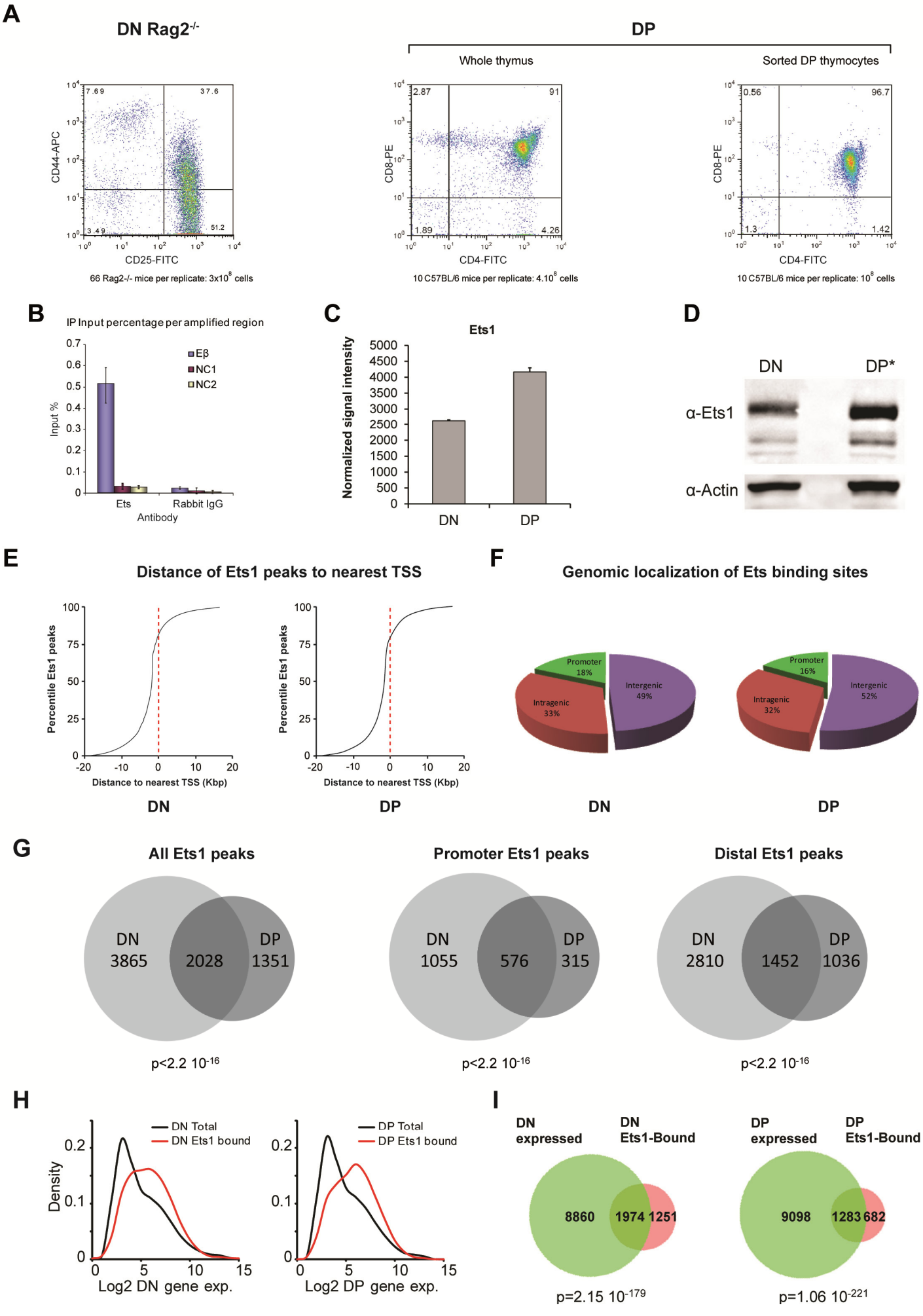

A
DN Rag2-/-
DP
Whole thymus
Sorted DP thymocytes
CD44-APC
CD25-FITC
CD8-PE
CD4-FITC
66 Rag2-/- mice per replicate: 3x10^8 cells
10 C57BL/6 mice per replicate: 4.10^8 cells
10 C57BL/6 mice per replicate: 10^8 cells
B
IP Input percentage per amplified region
Eβ
NC1
NC2
Input %
Ets
Rabbit IgG
Antibody
C
Ets1
Normalized signal intensity
DN
DP
D
DN
DP*
α-Ets1
α-Actin
E
Distance of Ets1 peaks to nearest TSS
Percentile Ets1 peaks
Distance to nearest TSS (Kbp)
DN
DP
F
Genomic localization of Ets binding sites
Promoter 18%
Intergenic 49%
Intragenic 33%
Promoter 16%
Intergenic 52%
Intragenic 32%
DN
DP
G
All Ets1 peaks
Promoter Ets1 peaks
Distal Ets1 peaks
DN 3865
2028
DP 1351
DN 1055
576
DP 315
DN 2810
1452
DP 1036
p<2.2 10-16
p<2.2 10-16
p<2.2 10-16
H
DN Total
DN Ets1 bound
DP Total
DP Ets1 bound
Density
Log2 DN gene exp.
Log2 DP gene exp.
I
DN expressed
DN Ets1-Bound
8860
1974
1251
DP expressed
DP Ets1-Bound
9098
1283
682
p=2.15 10-179
p=1.06 10-221

Figure S1. Experimental set-up, ChIP quality controls and peak statistics

(A) FACS analysis and purities of isolated DN and DP thymocytes. Right: profiles of CD25-FITC and CD44-APC staining in cells from murine Rag2$^{-/-}$ whole thymus. Right: profiles of CD4-FITC and CD8-PE staining in thymocytes from murine whole thymus and following AutoMACS purification using CD4 and CD8 antibodies. In both cases, similar results were obtained using non-crosslinked cells, e.g. for RNA-seq and MNase-seq. (B) ChIP-qPCR validation of Ets1 binding versus rabbit IgG in the enhancer β of the *Tcrb* locus using two negative, internal controls in intergenic regions. ChIPs were performed in duplicate. (C) Normalized gene expression microarray signal intensities for Ets1 at the DN and DP stage. (D) Western blot showing Ets1 protein levels at the DN and DP stages. DN denotes Rag2$^{-/-}$ thymocytes, and the asterisk denotes total wild type thymus. (E) Cumulative distributions of Ets1 peak distances to the nearest TSS in DN and DP revealing predominantly distal binding. (F) Proportions of Ets1 peaks in promoter, intragenic as well as intergenic regions in DN and DP. (G) Venn diagrams showing overlaps of Ets1 DN and DP peaks for all left), promoter (middle) and distal peaks (right). Hypergeometric significance p-values are indicated below each Venn diagram. (H,I) Ets1 is generally associated with positive regulation of transcription. Gene expression densities of all genes (black) and Ets1-bound genes (red) in DN (left) and DP stages (right) (H). (I) Venn diagrams showing overlaps of expressed genes (green) and Ets1-bound genes (red) in DN and DP (left, right).

Figure S2

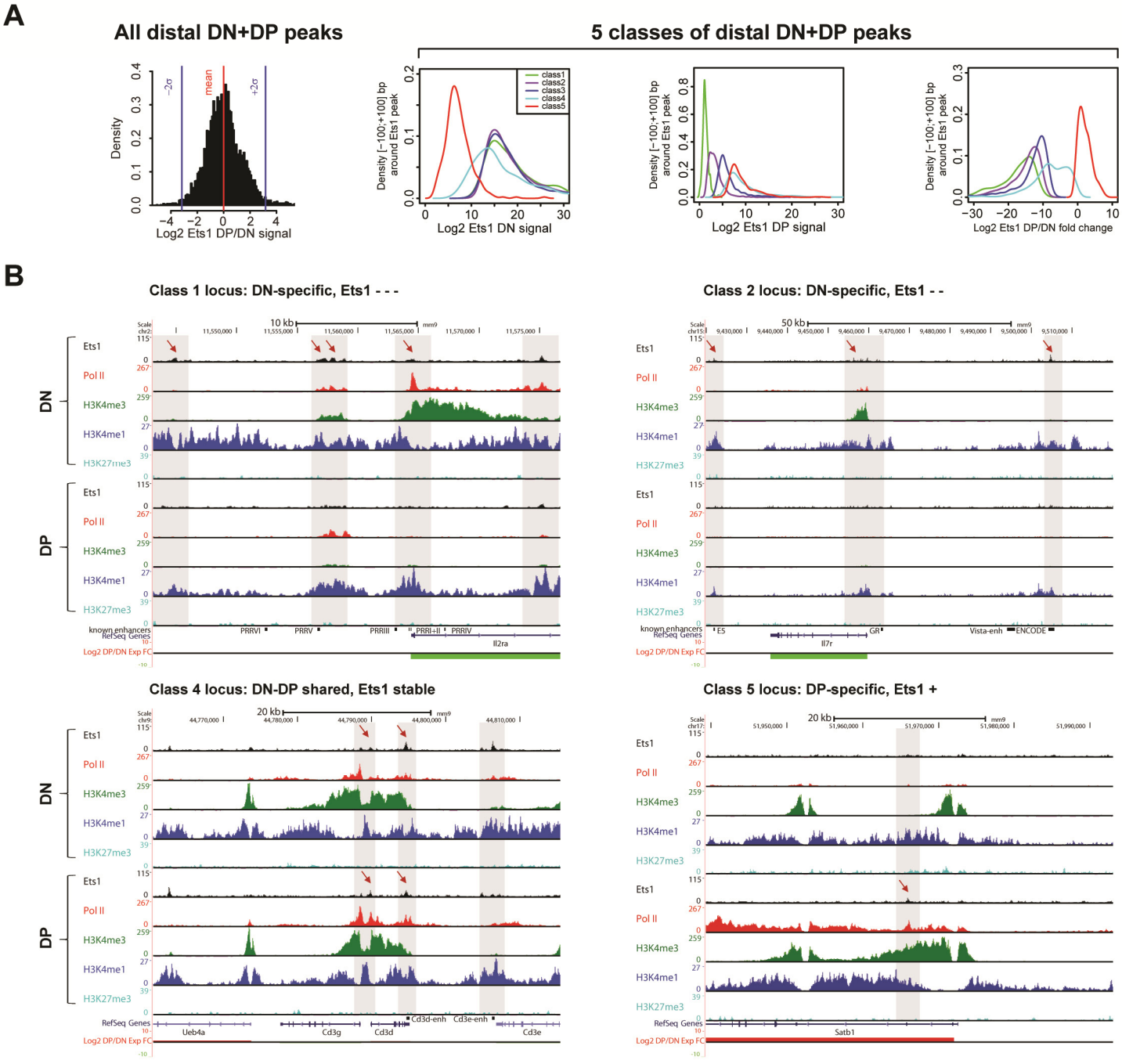

A
All distal DN+DP peaks
5 classes of distal DN+DP peaks
Density
−2σ
mean
+2σ
Log2 Ets1 DP/DN signal
Density [−100,+100] bp around Ets1 peak
class1
class2
class3
class4
class5
Log2 Ets1 DN signal
Log2 Ets1 DP signal
Log2 Ets1 DP/DN fold change
B
Class 1 locus: DN-specific, Ets1 - - -
Class 2 locus: DN-specific, Ets1 - -
Class 4 locus: DN-DP shared, Ets1 stable
Class 5 locus: DP-specific, Ets1 +
DN
DP
Ets1
Pol II
H3K4me3
H3K4me1
H3K27me3
known enhancers
RefSeq Genes
Log2 DP/DN Exp FC
Il2ra
Il7r
Satb1
Ueb4a
Cd3g
Cd3d
Cd3d-enh
Cd3e-enh
Cd3e
Vista-enh
ENCODE
C
R² = 0.7353
Log2 DP/DN exp. FC
Log2 DP/DN Ets1 FC
D
Av. log2 DP/DN exp. FC
Ets1 peak change/gene

Figure S2. Subdivision and functional role of Ets1 DN+DP peaks in 5 distinct classes

(A) Distributions, class subdivision of Ets1 DP/DN fold change as well as corresponding DN and DP signals. Leftmost: distribution of distal Ets1 DP/DN fold change showing an approximated normal distribution, allowing the use of statistical thresholds. Right: normalized DN, DP signals and fold change distribution across 5 classes showing distinct profiles in either DN or DP. (B) Dynamic enrichments for Ets1 plus core transcriptional hallmarks in example loci for class 1, 2, 4 and 5 sites. Genome browser screenshots of Ets1, Pol II, H3K4me1/3 and Pol II in DN and DP for the *Il2ra* (top left), *Il7r* (top right), *Cd3* (bottom; left) and *Satb1* loci (bottom right). DP/DN gene expression fold change is indicated at the bottom of each screenshot. (C) Positive, significant correlation between Ets1 and expression fold changes. Correlations of Ets1 and gene expression fold changes with fit. (D) Correlation between Ets1 peak number variation and expression fold change. Changes in peak numbers from DN to DP and corresponding gene expression fold change expressed as barplots.

Figure S3

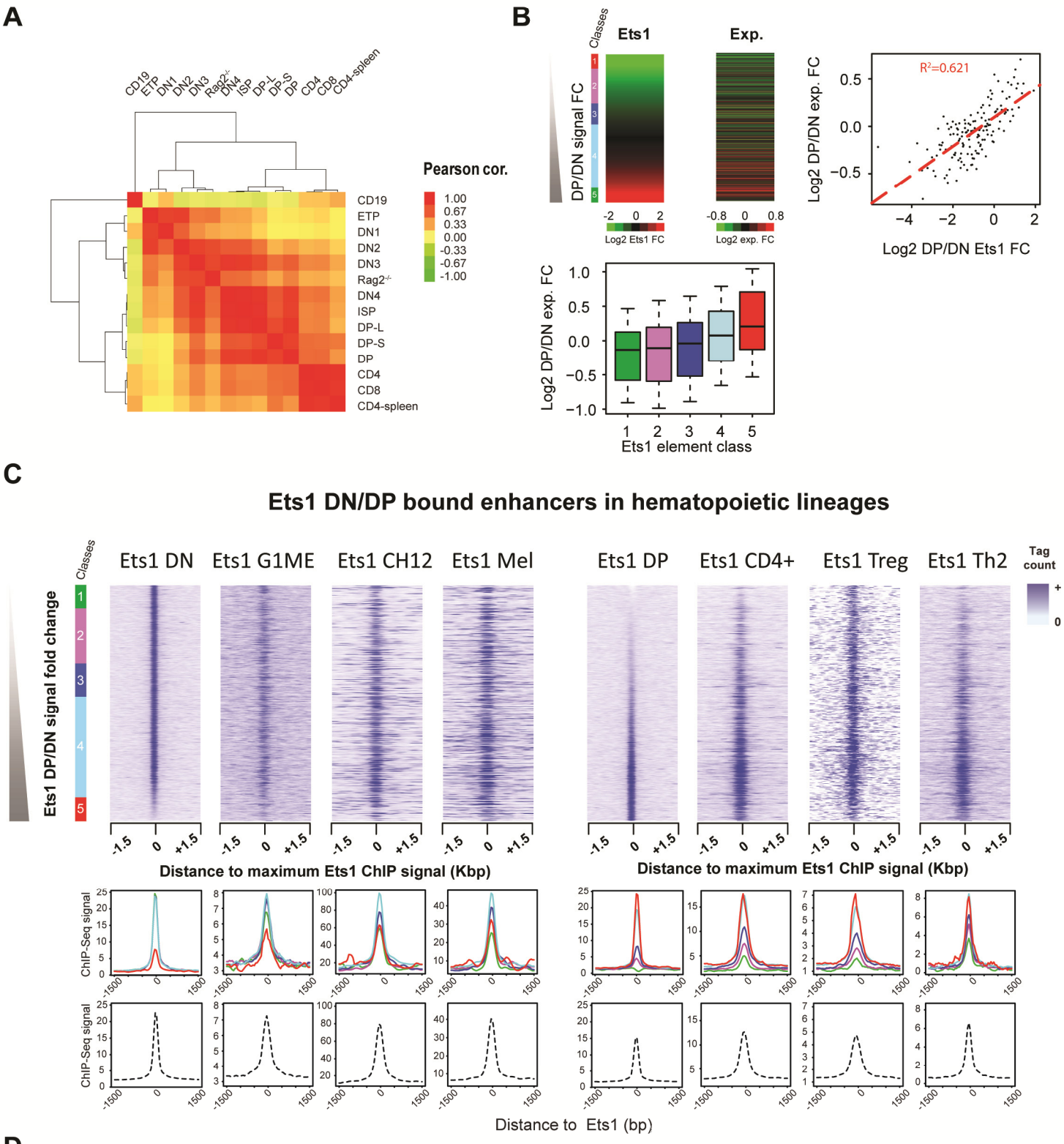
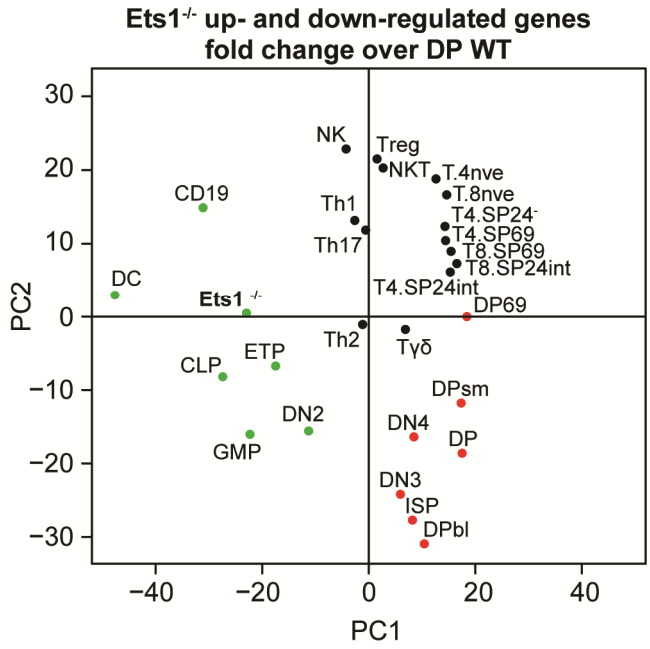

A
CD19
ETP
DN1
DN2
DN3
Rag2-/-
DN4
ISP
DP-L
DP-S
DP
CD4
CD8
CD4-spleen
Pearson cor.
1.00
0.67
0.33
0.00
-0.33
-0.67
-1.00
B
Classes
Ets1
Exp.
DP/DN signal FC
Log2 Ets1 FC
Log2 exp. FC
R²=0.621
Log2 DP/DN exp. FC
Log2 DP/DN Ets1 FC
Ets1 element class
C
Ets1 DN/DP bound enhancers in hematopoietic lineages
Ets1 DP/DN signal fold change
Ets1 DN
Ets1 G1ME
Ets1 CH12
Ets1 Mel
Ets1 DP
Ets1 CD4+
Ets1 Treg
Ets1 Th2
Tag count
Distance to maximum Ets1 ChIP signal (Kbp)
ChIP-Seq signal
Distance to Ets1 (bp)
D
1 kb
20 kb
DN + non T-cell hematopoietic
DP + T-cell
Ets1 DN
Ets1 G1ME
Ets1 CH12
Ets1 MEL
Ets1 DP
Ets1 CD4
Ets1_Treg
Ets1 Th2
UCSC Genes
Cbfa2t3
Tox

Figure S3. Lineage-specificity of DN and DP Ets1-bound sites

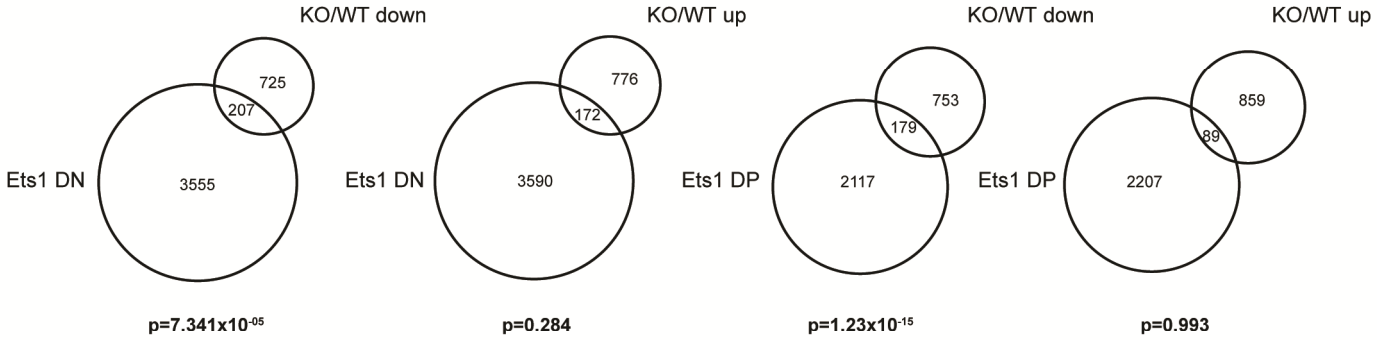
(A) Correlation clustering of expression values of differentially regulated genes (top 1% variance) from lymphocyte data from the Immunological Genome Project as well as Rag2-/- and DP thymocytes used in this study. (B) Left: heatmap of DP/DN3 gene expression fold change by increasing DP/DN Ets1 fold change (top). Bottom: boxplots of DP/DN3 gene expression fold change for each class of increasing DP/DN3 Ets1 elements. Right: correlation of DP/DN3 gene expression fold change and DP/DN Ets1 fold change. (C) DN Ets1 sites correspond to non T-cell hematopoietic lineages and DP Ets1 sites to T-cell ones. Top: heatmaps showing tag counts of Ets1 ChIP-seq signal from other hematopoietic lineages in distal Ets1 DN+DP sites. Bottom: average profiles for each class is shown with its corresponding color. Total average profiles are indicated with a dashed black line underneath. (D) Example loci showing differential Ets1 binding in DN + non T-cell hematopoietic and DP + T-cell lineages. Genome browser screenshots of Ets1 ChIP-seq in DN, G1ME, CH12, MEL lineages (DN + non T-cell hematopoietic, green) and in DP, CD4, Treg, Th2 lineages (DP + T-cell, red) for the *Cbfa2t3* and *Tox* loci (left, right).

Figure S4

A



B



p=7.341x10^{-05} **p=0.284** **p=1.23x10^{-15}** **p=0.993**

C

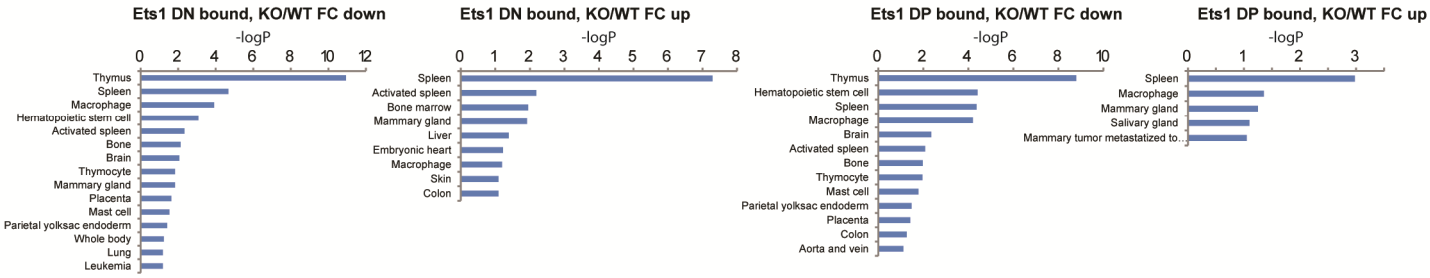
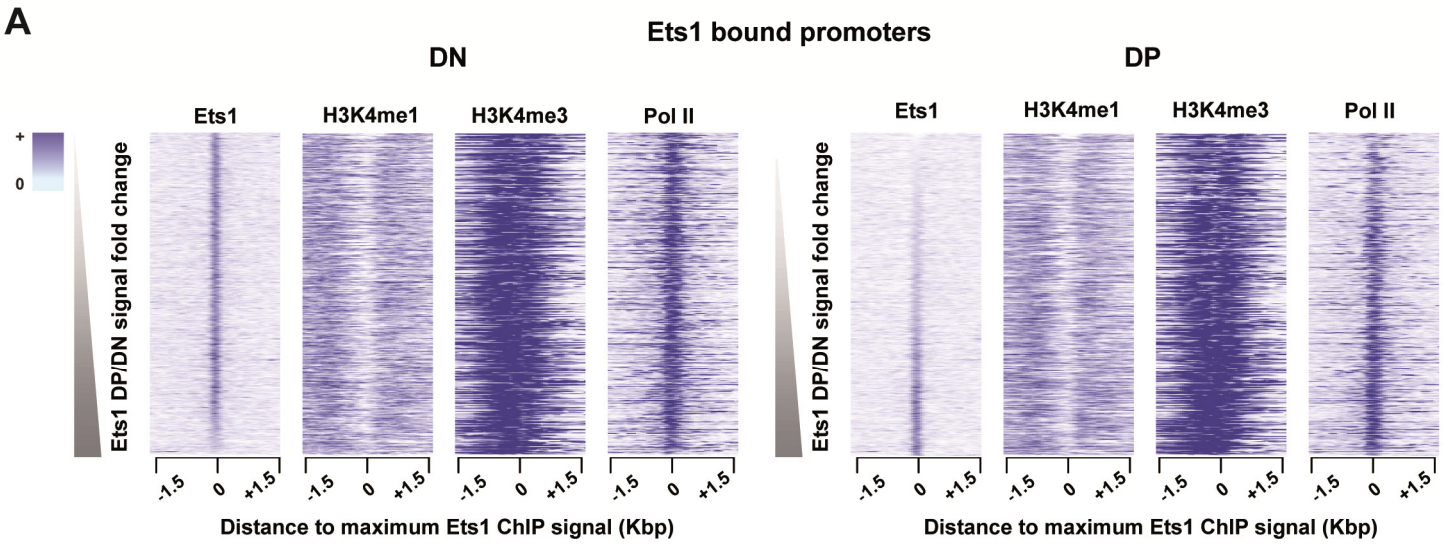
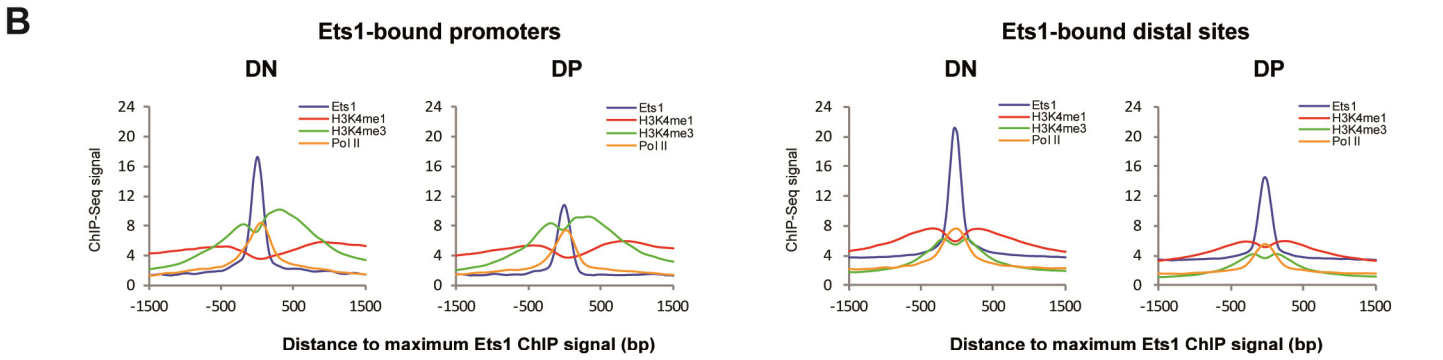
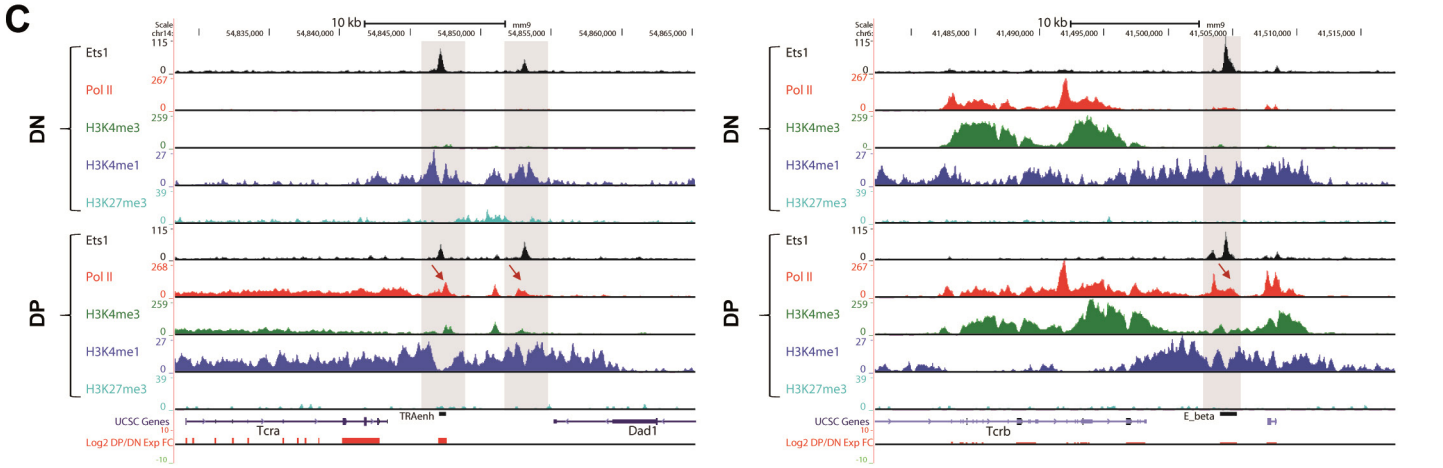


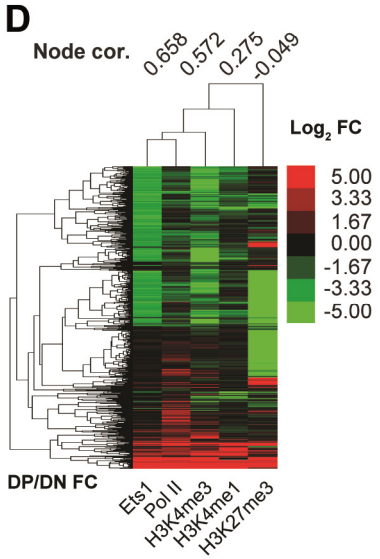
Figure S4. Loss of Ets1 impairs transcriptional switch towards T-specific expression program. (A) PCA of fold changes of microarray gene expression values of *Ets1*$^{-/-}$ against DP WT datasets, as well of relevant Immunological Genome datasets in T-cells and in other hemopoietic lineages with the addition of Th1, Th2, Th17 and Treg datasets from other studies against the DP WT dataset in both *Ets1*$^{-/-}$ up- and down -regulated genes. *Ets1*$^{-/-}$ / WT fold change is indicated in bold. (B) Venn diagram overlaps of Ets1 bound genes in DN and DP vs *Ets1*$^{-/-}$ up and down genes. (C) Tissue-specificity of intersecting genes in A).

Figure S5


A
Ets1 bound promoters
DN
DP
Ets1
H3K4me1
H3K4me3
Pol II
Ets1 DP/DN signal fold change
+
0
-1.5
0
+1.5
Distance to maximum Ets1 ChIP signal (Kbp)


B
Ets1-bound promoters
Ets1-bound distal sites
DN
DP
ChIP-Seq signal
Ets1
H3K4me1
H3K4me3
Pol II
Distance to maximum Ets1 ChIP signal (bp)


C
10 kb
DN
DP
Ets1
Pol II
H3K4me3
H3K4me1
H3K27me3
UCSC Genes
Tcra
TRAenh
Dad1
Tcrb
E_beta
Log2 DP/DN Exp FC


D
Node cor.
0.658
0.572
0.275
-0.049
Log2 FC
5.00
3.33
1.67
0.00
-1.67
-3.33
-5.00
DP/DN FC
Ets1
Pol II
H3K4me3
H3K4me1
H3K27me3

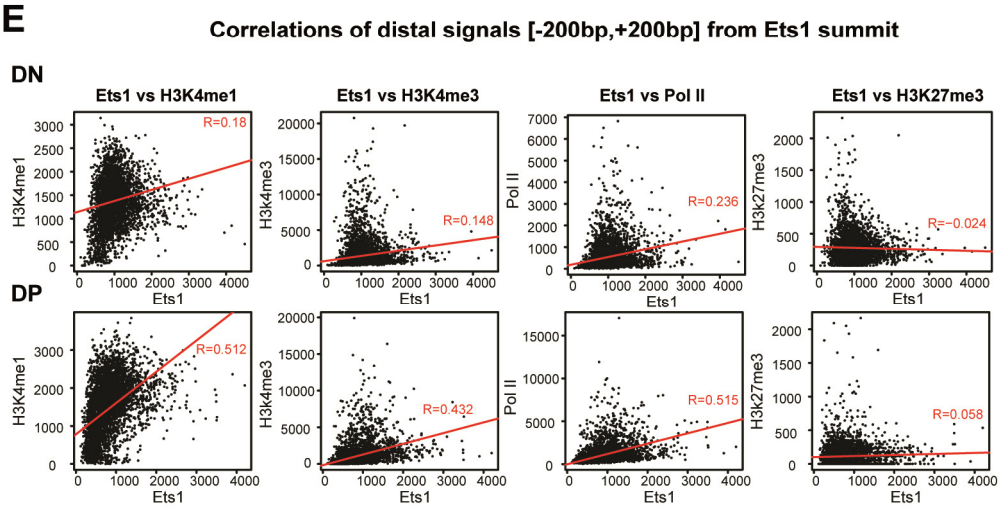

E
Correlations of distal signals [-200bp,+200bp] from Ets1 summit
DN
DP
Ets1 vs H3K4me1
Ets1 vs H3K4me3
Ets1 vs Pol II
Ets1 vs H3K27me3
R=0.18
R=0.148
R=0.236
R=-0.024
R=0.512
R=0.432
R=0.515
R=0.058
Ets1
H3K4me1
H3k4me3
Pol II
H3K27me3

Figure S5. Core transcriptional/epigenetic hallmarks at Ets1-bound promoters and recruitment of active marks at the DP stage

(A) Invariant enrichment of core transcriptional hallmarks at proximal Ets1 binding sites in DN and DP. Heatmaps of Ets1, H3K4me1, H3K4me3 and Pol II signals [-1500bp; +1500bp] around proximal Ets1-bound sites sorted by increasing Ets1 DP/DN fold change in DN and DP stages (left, right). (B) Core transcriptional hallmarks mirror Ets1 dynamics distally but not proximally. Total average profiles for Ets1, H3K4me1/3 and Pol II in promoter (left) and distal Ets1 DN+DP sites (right). (C) Recruitment of core transcriptional hallmarks by prior Ets1 poising. Genome browser screenshots showing ChIP-seq profiles for Ets1, Pol II, H3K4me1/3, and H3K27me3 at the *Tcra* and *Tcra* loci (left, right) in DN and DP stages. Gene expression fold change is indicated underneath. (D) Hierarchical clustering of Ets1, Pol II, H3K4me1, H3K4me3 and H3K27me3 DP/DN fold changes. Pearson correlation scores for each node are indicated on the left. (E) Ets1 signal correlates better with epigenetic and transcriptional marks in DP than in DN. Scatterplots of Ets1 distal signal versus H3K4me1, H3K4me3, Pol II and H3K27me3 (left to right) in DN and DP (top, bottom). Pearson correlation coefficients are indicated in red.

Figure S6

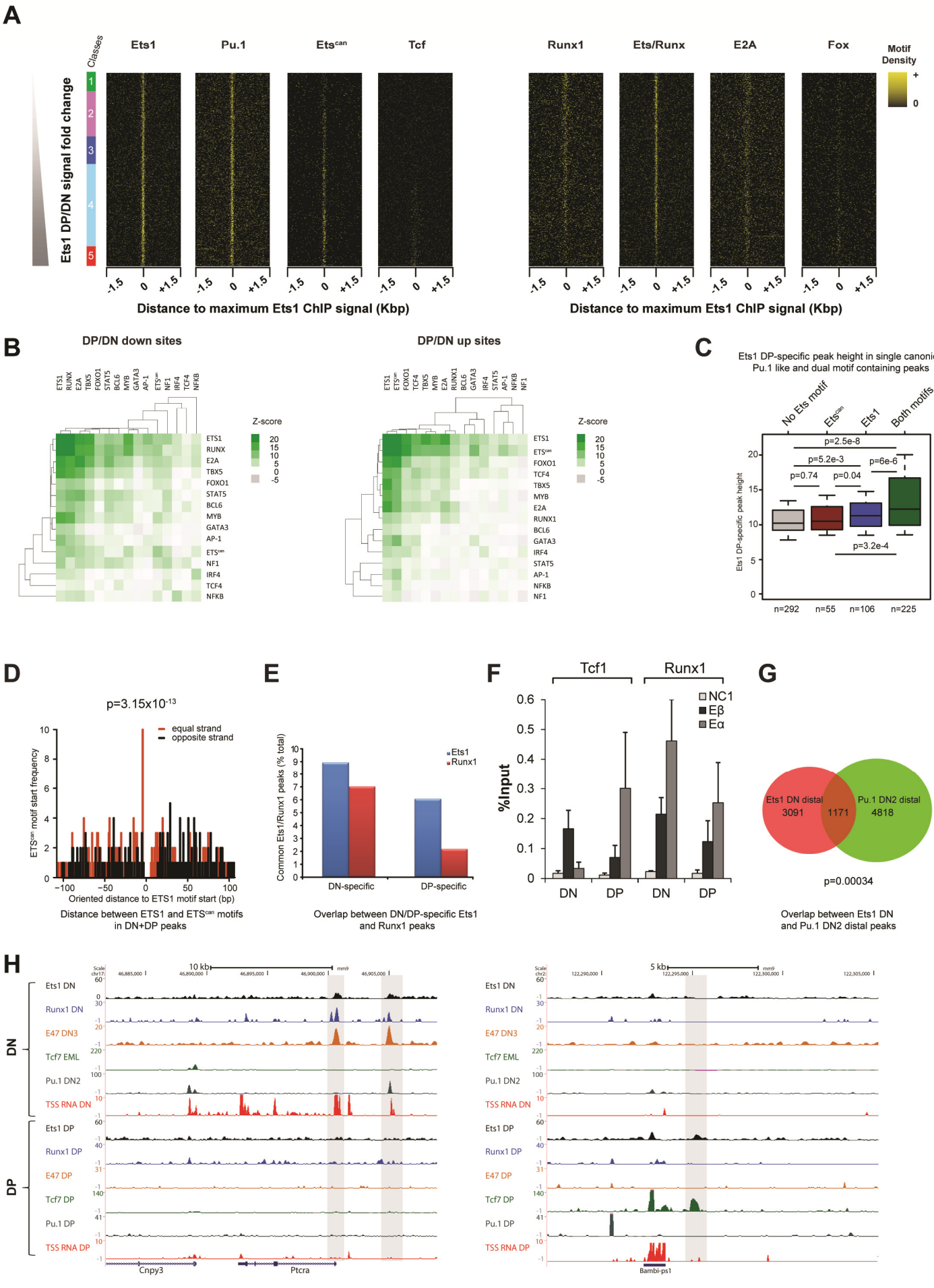
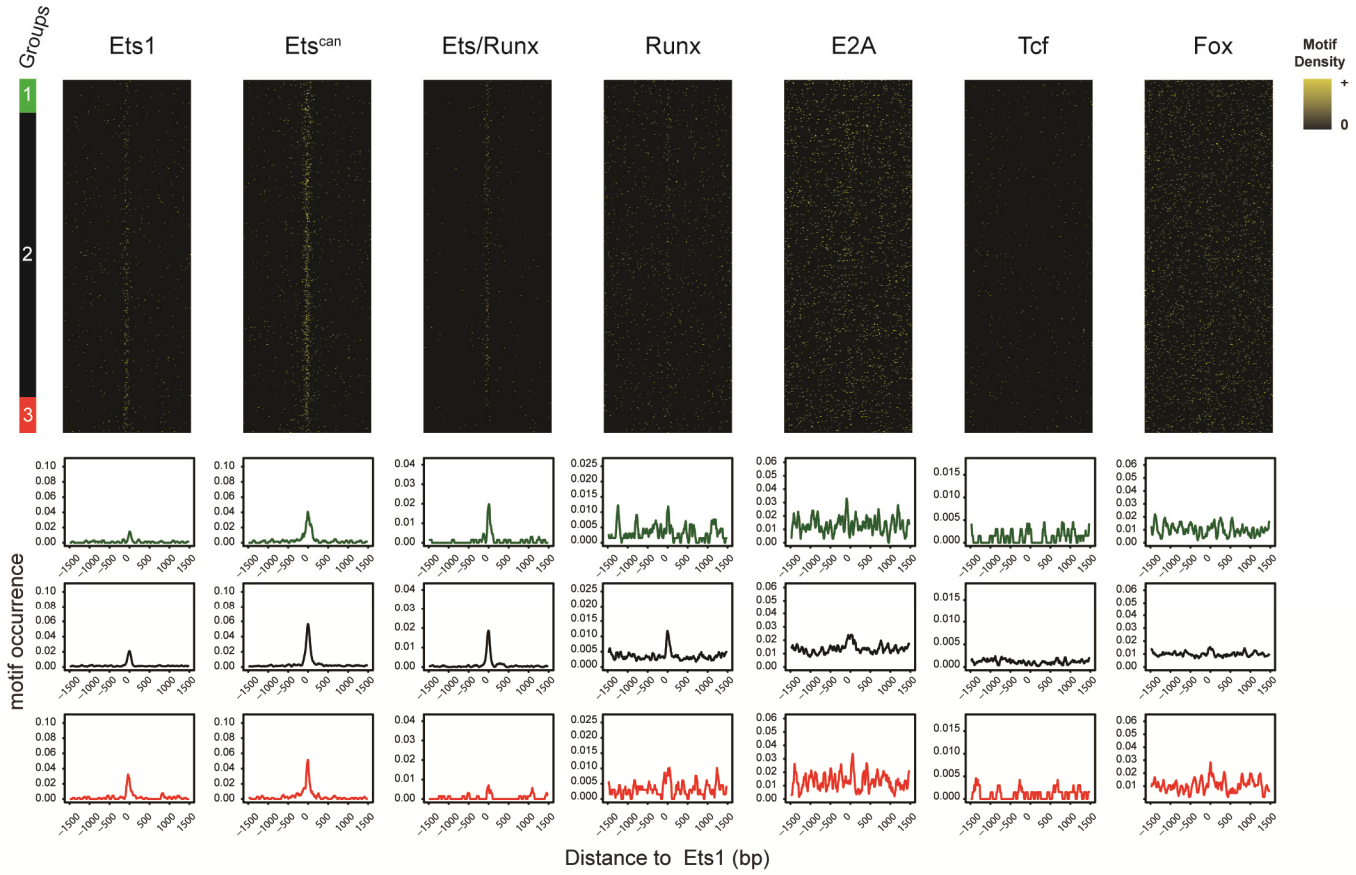

A
Ets1
Pu.1
Ets^can
Tcf
Runx1
Ets/Runx
E2A
Fox
Classes
Motif Density
Ets1 DP/DN signal fold change
Distance to maximum Ets1 ChIP signal (Kbp)
B
DP/DN down sites
DP/DN up sites
Z-score
C
Ets1 DP-specific peak height in single canonical, Pu.1 like and dual motif containing peaks
No Ets motif
Ets^can
Ets1
Both motifs
p=2.5e-8
p=5.2e-3
p=6e-6
p=0.74
p=0.04
p=3.2e-4
Ets1 DP-specific peak height
n=292
n=55
n=106
n=225
D
p=3.15x10^-13
equal strand
opposite strand
ETS^can motif start frequency
Oriented distance to ETS1 motif start (bp)
Distance between ETS1 and ETS^can motifs in DN+DP peaks
E
Common Ets1/Runx1 peaks (% total)
Ets1
Runx1
DN-specific
DP-specific
Overlap between DN/DP-specific Ets1 and Runx1 peaks
F
Tcf1
Runx1
NC1
Eβ
Eα
%Input
DN
DP
DN
DP
G
Ets1 DN distal
3091
1171
Pu.1 DN2 distal
4818
p=0.00034
Overlap between Ets1 DN and Pu.1 DN2 distal peaks
H
DN
DP
Ets1 DN
Runx1 DN
E47 DN3
Tcf7 EML
Pu.1 DN2
TSS RNA DN
Ets1 DP
Runx1 DP
E47 DP
Tcf7 DP
Pu.1 DP
TSS RNA DP
10 kb
5 kb
Cnpy3
Ptcra
Bambi-ps1

Figure S6. Differential co-association of motifs and their binding properties TFs in DN and DP

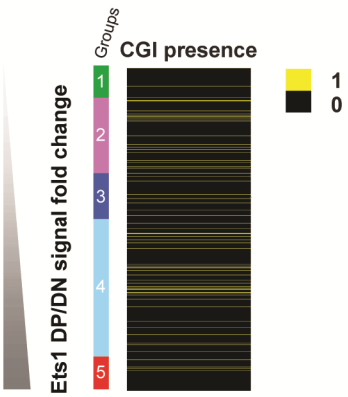
(A) Decrease of DN-specific and increase of DP-specific motif occurrences with increasing Ets1 DP/DN fold change. Motif heatmaps sorted by increasing Ets1 DP/DN fold change, as in **Figure 1B**. All Ets1 DN/DP classes are shown with a different color. (B) Stage-specific co-occurrences of motifs in DN and DP Ets1-bound sites. Hierarchical clustering of co-occurrence enrichments for motifs found in DN/DP down (left) and up (right) Ets1 bound sites. (C) Higher ChIP-seq signal for sites with both Ets motifs suggests Ets1 binds as a homodimer in DP de novo sites. Boxplot of ChIP-seq tag counts in class 5 elements subdivided by motif content. Ets1 enrichment is highest in sites bearing two distinct Ets motifs. (D) Ets1 and Runx1 co-localize more in DN-specific sites. Histogram of common Ets1/Runx1 peaks in DN- and DP-specific sites, as percentage of Ets1 (blue) or Runx1 peaks (red). (E) Ets^{can} motifs show increased frequency at -4 bases from the ETS1 motif on the same strand. Distribution of distances of ETS^{can} motif starts to ETS1 motif starts by ETS1 motif orientation. ETS^{can} motifs located on the same and opposite strand as the ETS motif are shown in red and black, respectively. (F) Dynamic binding of TCF1 and Runx1 between DN and DP stages. qPCR of TCF1 and Runx1 ChIP DNA at the DN and DP stages using primers for *NC1* control (gray), *Tcrb* (Eβ, black) and *Tcra* enhancers (Eα, dark gray). (G) A significant number of DN Ets1 peaks are bound by Pu.1 at the DN2 stage. Venn diagram showing overlap of distal DN Ets1 and DN2 Pu.1 peaks. (H) Example, stage-specific loci showing Ets1, Runx1 and E47 co-localizing in DN as well as Ets1 and TCF1 co-localizing in DP. Screenshots illustrating binding of Ets1, Runx1, E47, TCF1 and Pu.1 in DN and DP stages at the Ptcra and Bambi-ps1 loci. Short RNA-seq coverage is shown in both stages.

Figure S7

A



B



n=76

C

ETS-RUNX
p=10^{-14}
21.74% of targets
1.17% of background

Figure S7. Enrichment of Ets1 and co-associated motifs at promoters and intersection of distal Ets1 with CpG islands

(A) Heatmaps and average profiles of motif presences in Ets1 DN+DP promoters. (B) CpG island presence by increasing Ets1 DP/DN fold change. (C) Overrepresented motif in 76 Ets1 and CpG overlapping sites.

Figure S8

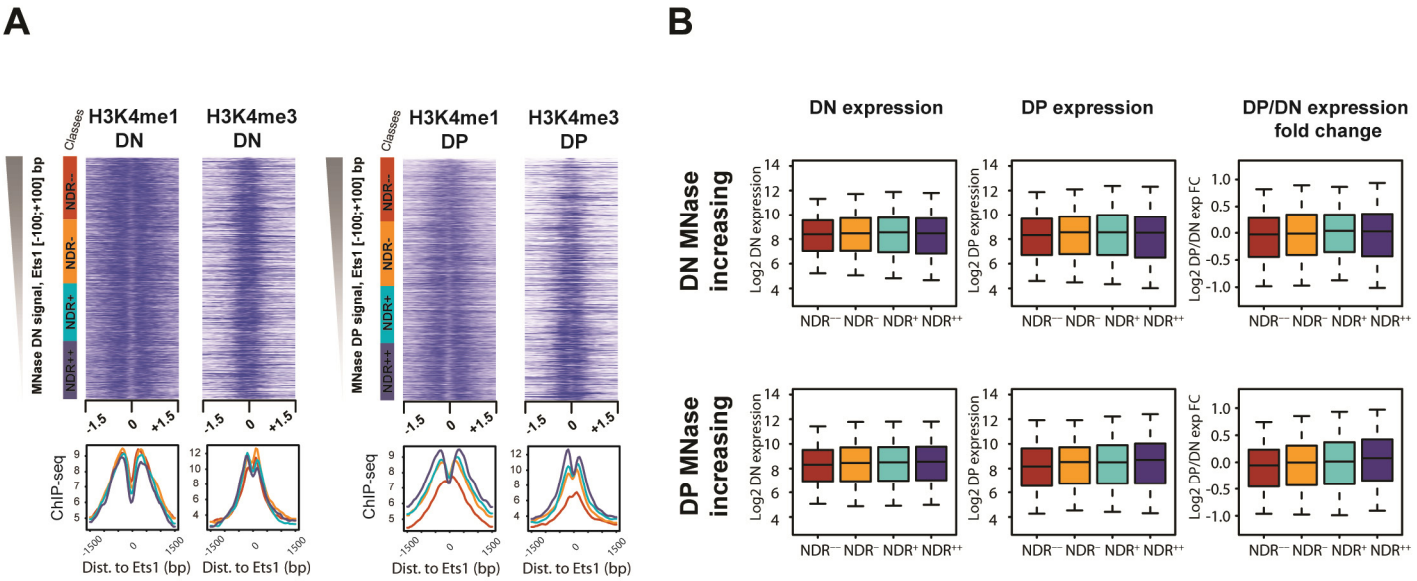

A
B
Classes
H3K4me1 DN
H3K4me3 DN
H3K4me1 DP
H3K4me3 DP
MNase DN signal, Ets1 [-100;+100] bp
MNase DP signal, Ets1 [-100;+100] bp
NDR--
NDR-
NDR+
NDR++
-1.5
0
+1.5
ChIP-seq
Dist. to Ets1 (bp)
DN expression
DP expression
DP/DN expression fold change
DN MNase increasing
DP MNase increasing
Log2 DN expression
Log2 DP expression
Log2 DP/DN exp FC
NDR⁻⁻ NDR⁻ NDR⁺ NDR⁺⁺

Figure S8. Higher transcriptional activity and sequence specificity of NDRs

(A) Stage-specific enrichments for activating histone modifications in NDRs and nucleosome-occupied regions in DN+DP Ets1 peak. Heatmaps of tag counts of DN- and DP-corresponding H3K4me1, H3K4me3 ChIP-seq sorted by increasing DN (left) and DP (right) MNase-seq, as in **Figure 5C**. Average profiles for all NDR classes are shown underneath. (B) DP NDRs show higher transactivation levels of the nearest gene. Boxplot representation of DN expression, DP expression and DP/DN expression fold change levels for all classes, sorted by increasing MNase DN (top) or DP signal (bottom).

Figure S9

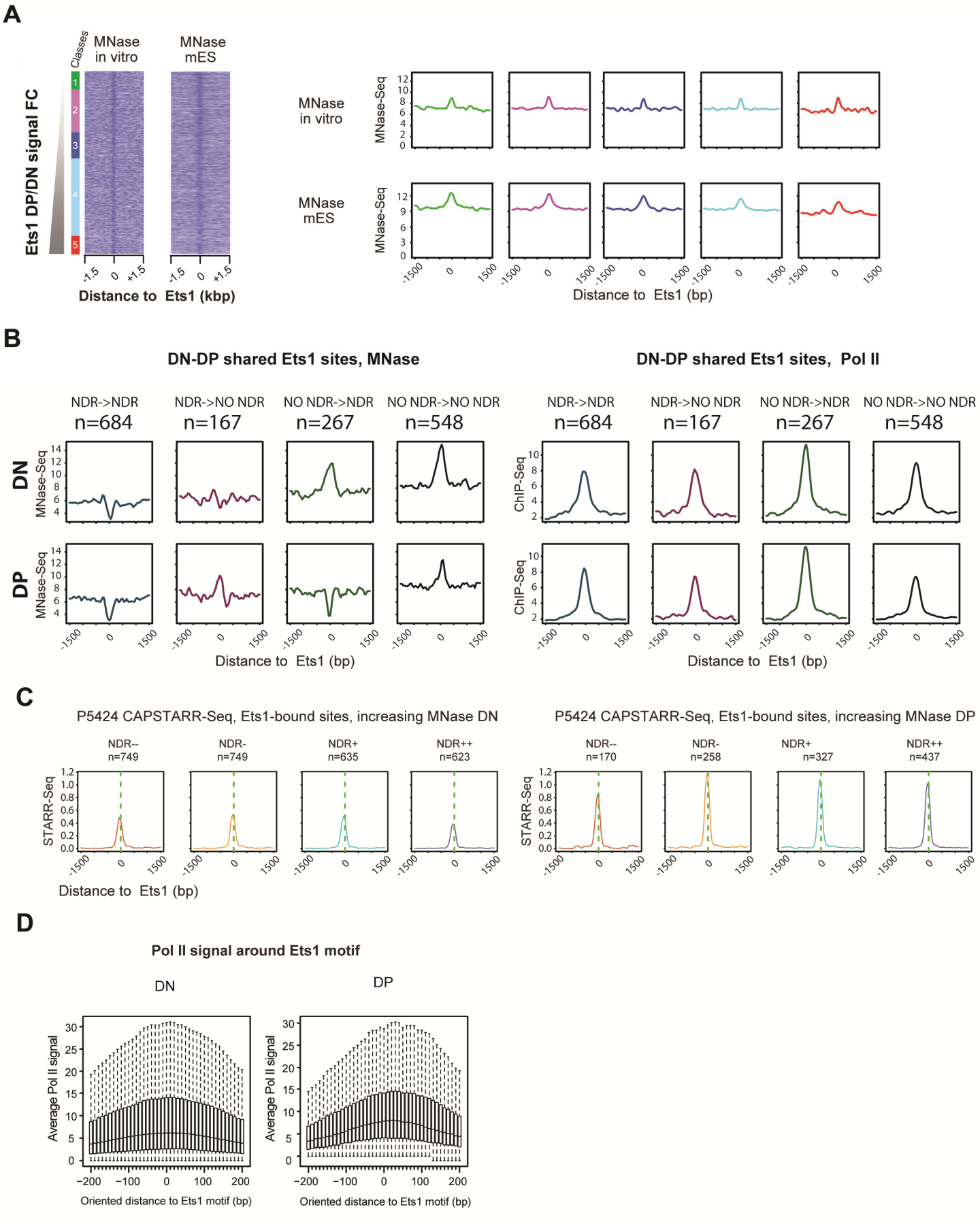

A
Classes
MNase in vitro
MNase mES
Ets1 DP/DN signal FC
Distance to Ets1 (kbp)
MNase in vitro
MNase mES
MNase-Seq
Distance to Ets1 (bp)
B
DN-DP shared Ets1 sites, MNase
DN-DP shared Ets1 sites, Pol II
NDR->NDR n=684
NDR->NO NDR n=167
NO NDR->NDR n=267
NO NDR->NO NDR n=548
DN
DP
MNase-Seq
ChIP-Seq
Distance to Ets1 (bp)
C
P5424 CAPSTARR-Seq, Ets1-bound sites, increasing MNase DN
P5424 CAPSTARR-Seq, Ets1-bound sites, increasing MNase DP
NDR-- n=749
NDR- n=749
NDR+ n=635
NDR++ n=623
NDR-- n=170
NDR- n=258
NDR+ n=327
NDR++ n=437
STARR-Seq
Distance to Ets1 (bp)
D
Pol II signal around Ets1 motif
DN
DP
Average Pol II signal
Oriented distance to Ets1 motif (bp)

Figure S9. Priming activity of Ets1

(A) Ets1-bound sites correspond to nucleosome-occupied regions in vitro and in mESC cells. In vitro reconstituted and mESC MNase-Seq signal sorted by increasing DP/DN fold change as heatmaps (left) and average profiles (right). (B) Increased Pol II recruitment at DN-DP transiting nucleosome-occupied and NDR Ets1 bound sites. Average profiles for MNase-Seq and ChIP-Seq in regions varying or stable in nucleosome occupancy between the DN (top) and DP stages (bottom). (C) P5424 CAPSTARR-Seq average profiles in Ets1 bound DN and DP NDR classes, as in **Figure 5D**. (D) Distances of Pol2 summit to ETS1 motif by ETS1 motif orientation in DN (left) and DP (right).

Figure S10

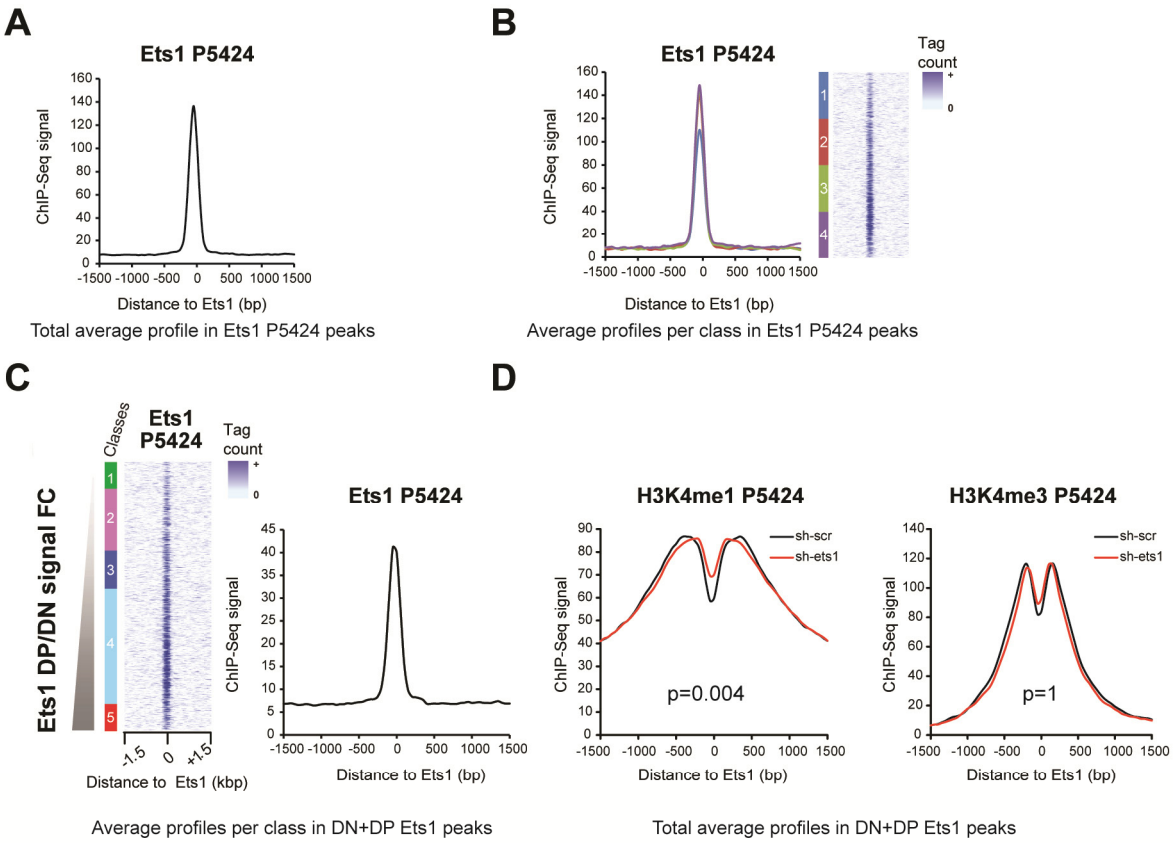

A
Ets1 P5424
ChIP-Seq signal
Distance to Ets1 (bp)
Total average profile in Ets1 P5424 peaks
B
Ets1 P5424
Tag count
ChIP-Seq signal
Distance to Ets1 (bp)
Average profiles per class in Ets1 P5424 peaks
C
Ets1 DP/DN signal FC
Classes
Ets1 P5424
Tag count
Distance to Ets1 (kbp)
Ets1 P5424
ChIP-Seq signal
Distance to Ets1 (bp)
Average profiles per class in DN+DP Ets1 peaks
D
H3K4me1 P5424
sh-scr
sh-ets1
p=0.004
ChIP-Seq signal
Distance to Ets1 (bp)
H3K4me3 P5424
sh-scr
sh-ets1
p=1
ChIP-Seq signal
Distance to Ets1 (bp)
Total average profiles in DN+DP Ets1 peaks

Figure S10. Loss of Ets1 induces chromatin remodeling via increased H3K4me1 signal genome-wide at DN+DP sites in the thymic P5424 cell line

(A) Total average profile for P5424 Ets1 at P5424 Ets1 sites, as in **Figure 6C**. (B) Heatmap and average profiles per class of H3K4me1 NDR for Ets1 in the thymic P5424 cell line at P5424 Ets1 sites, ranked by increasing P5424 sh-scr H3K4me1 signal [-500bp; +500bp] around the Ets1 site, as in **Figure 6E**. Classes of H3K4me1 NDR status are indicated left of heatmaps. (C) Heatmap showing Ets1 signal in the thymic P5424 cell line ranked by increasing DN+DP fold change, as in **Figure 1B** (left). Right: total average profile for P5424 Ets1 at DN+DP sites. (D) Total average profiles of H3K4me1 (left) and H3K4me3 (right) with and without Ets1 knockdown in the thymic P5424 cell line at Ets1 DN+DP sites, as in (C).

II. SUPPLEMENTARY MATERIALS AND METHODS

Experimental procedures

Cell preparation and sorting

For mouse thymocyte isolation, cohorts of litter mates were used as biological replicates. Mice were sacrificed humanely using a CO_2 chamber. For C57BL/6 $Rag2^{-/-}$ mice, a total of 122 mice were sacrificed and 10 for WT C57BL/6. *Ets1*$^{-/-}$ thymocytes were obtained at the Bories lab, Hôpital Saint Louis, Paris, France. Thymuses were stored in RPMI-1640 medium. To recover thymocytes, thymuses were subsequently passed twice through a 70µm Falcon cell strainer and stored in PBS. Cells were crosslinked 10 min in 10% formaldehyde followed by 5 min quenching in 250 mM glycine. DP cells were subsequently sorted via AutoMACS using, in turn, α-CD8-PE antibody conjugated to α-PE beads, and, after release, α-CD4 beads (Miltenyi). FACS profiles and cell purities were obtained for $Rag2^{-/-}$ and DP thymocytes via CD44-APC, CD25-FITC and CD8-PE, CD4-FITC staining, respectively **(Figure S1A)**.

Western blot for Ets1 protein levels in DN and DP

Thymocytes were rinsed in 1X PBS and lyzed in buffer A (1% NP-40, 50 mM Tris HCl, pH 8, 150 mM NaCl) containing 10 % Complete™ Protease Inhibitor Cocktail (Roche), 1% phosphatases inhibitor and PMSF 2 mM. One volume of Laemmli buffer (containing 5% β-mercaptoethanol) was added to 40 µg of protein and boiled for 5'. Boiled proteins were loaded on a 10% SDS-PAGE gels, transferred on nitrocellulose membranes and blotted with anti-Actin (Santa Cruz sc-1616) and anti-Ets-1 antibodies (Abcam ab109212).

Chromatin immunoprecipitation

Cells were aliquoted in crosslinked pellets of $5x10^7$ cells and lysed by rotation for 10 min at 4°C in 2x 1.25ml LB1 (50mM Hepes pH 7.5, 140mM NaCl, 1mM EDTA, pH 8, 10% glycerol, 0.75% NP-40, 0.25% Triton X-100). Following centrifugation at 1,350g for 5 minutes at 4°C, nuclei were subsequently treated with 2x1.25ml LB2 (200mM NaCl, 1mM EDTA pH 8, 0.5mM EGTA pH 8, 10mM Tris pH 8) on a rotating wheel for 10 minutes at 4°C. Chromatin was resuspended in 1.5ml LB3 (1mM EDTA pH8, 0.5mM EGTA pH 8, 10mM Tris pH 8, 100mM, NaCl, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine) in 15ml polypropylene falcon tubes and then sonicated for 10 cycles (30 seconds on, 30 seconds off) at 40W using a VCX130 sonicator (VibraCell) down to an average size of about 250bp. 50µl aliquots were collected as Input controls. All buffers contained final concentrations of 1X protease inhibitors (Roche, France) as well as 0.2mM PMSF and 1µg /ml pepstatin. ChIPs were conducted using either α-rabbit (Ets1) or protein-G (all other ChIPs) DynaBeads (Invitrogen). Per sample volumes of beads, amounts and nature of antibodies and cell numbers per experiment are detailed in **Table S1**. Antibody and bead mixes were washed three times with 1ml BSA blocking buffer (0.5% BSA in 1x DPBS), incubated and rotated overnight at 4°C, and subsequently washed another three times with 1ml BSA blocking buffer. Immunoprecipitation was carried out overnight in

750µl LB3 at 4°C on a rotating wheel. Samples were then washed 8 times in 1ml RIPA buffer (50mM Hepes pH 7.6, 500mM LiCl, 1mM EDTA pH 8, 1% NP-40, 0.7% Na-Deoxycholate) and once in 1ml TE buffer with 50mM NaCL. Two 15-min and 10-min subsequent elutions were carried out in a waterbath at 65°C, vortexing every other minute, for a final volume of 100µl. 100µl 1X elution buffer (50mM Tris pH 8, 10mM EDTA pH 8, 1% SDS) was added to eluates which were reverse-crosslinked overnight in a 65°C waterbath. Equal volumes of TE buffer (100µl) were added to samples as well as 0.2µg/ml RNAse A and placed at 37°C for 2 hours, followed by proteinase K treatment at 0.2µg/ml at 55°C. DNA purification was carried out via phenol-chloroform extraction followed by elution using QiaQuick PCR purification kits. (QiaGen) Samples were pooled per replicate then SpeedVac-ed down to 30µl. DNA quantification of eluates was obtained using Bioanalyzer High Sensitivity chips (Agilent). Library preparation was carried out as per manufacturer's instructions (Illumina) using 10ng material. Samples were subsequently size-selected via gel migration and extraction, with average cut sizes at about 230 bp for DN and 250 bp for DP, including adapters. Sequencing was performed using either Genome Analyzer II (Illumina) or SOLiD 5500xl sequencers (Applied Biosystems), whereby samples were sequenced at the CNAG, Barcelona, Spain and TAGC, Marseille, France, respectively. For SOLiD 5500xl sequencing, ChIP-seq libraries were prepared according to the Life Technologies protocol, SOLiD ChIP-seq kit guide (A12066), using 15 PCR cycles. The quality of the libraries was checked with the Bioanalyser DNA High sensitivity and by Q-PCR.

MNase-Seq

$2x10^7$ thymocytes were permeabilized in 50µl 150mM sucrose, 80mM KCl, 5mM K_2HPO_4, 5mM $MgCl_2$, 0.5mM $CaCl_2$, 35mM HEPES pH 7.4, 0.2% NP-40 for one minute at 37°C then treated with 10U MNase for 30 minutes at room temperature in 500µl 150mM sucrose, 50mM Tris pH8, 50mM NaCl, 2mM $CaCl_2$. Digestions were stopped by adding 1.45ml lysis buffer for 10 minutes at 4°C. 200µl aliquots were subsequently treated with equal amounts of TE buffer and 0.2µg/ml RNAse A, 0.2µg/ml proteinase K for 2 hours each at 37°C and 55°C, respectively. DNA was purified via phenol-chloroform extraction and subsequently eluted in 50µl water using QiaQuick PCR purification columns. Following gel size-selection of mononucleosomal fragment sizes, libraries were been prepared according to the Life Technologies protocol, SOLiD ChIP-seq kit guide (A12066), using 15 PCR cycles. The quality of the libraries was checked with the Bioanalyser DNA High sensitivity. Sequencing was carried out on a SOLiD 4 sequencer at TAGC, Marseille, France.

Short RNA-Seq

Total RNA was extracted using TRIzol (Invitrogen, USA) according to the manufacturer's instructions with some modifications to ensure higher recovery rates of small RNAs. $Rag2^{-/-}$ thymocytes were divided into $1x10^7$ aliquots and 1ml of TRIzol was added. Cells were lysed with vigorous vortexing and pipetting. Homogenized samples were incubated at room temperature for 5 minutes and 0.2ml of chloroform was added. Samples were vigorously shaken and incubated at room temperature for an additional 5 minutes. Phase separation was carried out by centrifugation at 4°C and 12,000g in a tabletop centrifuge for 15 minutes. The aqueous phase was transferred to fresh tubes and 1.5

volumes (approximately 1ml) of isopropanol together with 10µg of linear acrylamide (Ambion) were added. Samples were vortexed and incubated at room temperature for 15 minutes. The precipitated RNA was pelleted by centrifugation at 4°C and 12,000g in a tabletop centrifuge for 20 minutes. Pellets were washed with 80% ethanol, vortexed and centrifuged at 4°C and 7,500g in a tabletop centrifuge for 10 minutes. Pellets were allowed to air dry and resuspended in nuclease-free water (Ambion). DNA was digested using the rigorous Turbo DNase (Ambion) treatment as per manufacturer's instructions. RNA quantity was measured on a Nanodrop 1000 and the quality was verified using RNA Nano chips on a 2100 Bioanalyzer (Agilent).Small RNAs were purified from a 10% denaturing urea polyacrylamide gel, essentially as described in the Illumina Small RNA Rev. B protocol with some minor modifications. Approximately 10µg of total RNA was run at 200V for 1 hour, until the blue front reached the bottom of the gel. We used 21G needles to puncture holes into the bottom of two 0.5ml and placed them into 2ml eppendorf tubes. RNA corresponding to 15nt - 70nt was cut from the gel, diagonally cut in half and separately transferred into the 0.5ml tubes. The gel was crushed into the 2ml tubes by a two minute centrifugation at 14,000rpm. For gel elution by soaking, 0.4ml of 0.3M NaCl was added to each tube, before a 4 hour rotation at room temperature. After removal of gel particles using 0.22µm cellulose acetate filters, 10µg of linear acrylamide (Ambion) and 2.5 volumes (approximately 1ml) of ice-cold absolute ethanol were added. After 30-minute incubation at -80°C, the eluted RNA was precipitated by centrifugation at 4°C and maximum speed for 45 minutes. The pellet was washed with 1ml of room temperature 80% ethanol and resuspended in 5.7µl of water. This PAGE purification step was repeated after both the 5' and 3' adapter ligations, which were carried out according to the protocol. The resulting cDNA was purified with the QIAquick Gel Extraction kit (Qiagen) using the DNA cleanup protocol modification in order to retain DNA fragments as low as 70bp. DNA was quantified using a Nanodrop 1000 (Thermo Scientific) and verified using DNA High Sensitivity 2100 Bioanalyzer chips (Agilent). Libraries was clustered and sequenced using 76 cycles on a Genome Analyzer II (Illumina) at the CNAG, Barcelona, Spain, and on a HiSeq 2000 sequencer at the Gene Center Munich, Ludwig-Maximilian Universität München according to manufacturer's instructions.

Gene expression microarray assays and analysis

Total RNA from thymocytes was prepared using TRIzol reagent (Invitrogen, Paisley, UK) as recommended by the manufacturer. The samples were resuspended in Rnase-free water and treated with DNase I (Ambion) for 30 min at 37°C. RNA quality was then checked using a Bioanalyzer (Agilent Technologies). Only RNAs with RNA integrity number >9 were then used. Transcriptome experiments were performed in triplicate using the Affymetrix Mouse Gene 1.0 ST platform. RNA samples were amplified, labeled using Affymetrix's IVT labeling kit and hybridized according to the manufacturer's recommendations. DP assays were carried out at IPMC, Nice, France, whereas DP and DP *Ets1*$^{-/-}$ were carried out at the Hôpital Saint Louis, Paris, France. All assays were carried out and hybridized in biological triplicates. Expression data was obtained as Affymetrix CEL files, which were processed, background subtracted and normalized in R via the quantile normalization featured in the affy package for all microarray datasets, using the Mouse Gene 1.0 ST and 2.0 ST chip

definition files (CDF) for DP and DP *Ets1*$^{-/-}$ / WT microarrays, whereby probeset to gene annotation was carried out using the Affymetrix-supplied gene annotation. Assessment of individual gene expression levels was in agreement with published transcriptome data (1,2). For downloaded datasets, processing was done using the affy package and annotation packages were subsequently loaded in R and gene names retrieved using the mget() function of the AnnotationDBI R package. Unique gene expression data was subsequently derived as the average of all probe set values corresponding to one gene.

qPCR validation of ChIP

qPCR was performed using primer pairs for known Ets1 bindings site located in enhancer Eα and Eβ of the *Tcra* and *Tcrb* loci, and two negative controls (NC1 and NC2) located in a gene desert on chromosome 2. Primers were 5'-agccagaagtagaacaggaaatg-3' (Eα forward), 5'-gtgttaccaccaagacctgcaa-3' (Eα reverse), 5'-gggtggaagcatctcacccca-3' (Eβ forward), 5'-tggagctatggcaggtggcca-3' (Eβ reverse), 5'-aaggggcctctgcttaaaaa-3' (NC1 Forward), 5'-agagctccatggcaggtaga-3' (NC1 reverse), 5'-ccccttttctgaagcactctg-3' (NC2 forward), 5'-taaggcgtcatttcccaaag-3' (NC2 reverse). qPCRs were conducted using 2.5µl of 1:5 diluted ChIP sample (corresponding to 1% total ChIP), 10µl SYBR Green master mix (Applied Biosystems), 0.3µM of forward and reverse primer and H_2O in a final volume of 20µl, using 4 1:10 dilutions of Input as standard curves, and in triplicates for ChIPs. Cts were obtained using an ABI 7500 Real-Time PCR system (Applied Biosystems) and were expressed as % Input for each ChIP following exponential regression of each primer pair's standard curve (**Figure S1B**).

Bioinformatic analyses

High-throughput sequencing data processing and peak detection

For the Genome Analyzer II System, basecalls were performed using manufacturer's instructions using CASAVA 1.7. For the HiSeq 2000 system, basecalls were performed using the Sequencing Analysis Viewer Software v1.9.1 and converted to fastq files using bcl2fastq 1.8.4. For the SOLiD 4 System, high-quality reads were matched to individual samples based on the barcode tag and using default parameters in SOLiD Experimental Tracking Software (SETS) v4.0.1. For SOLiD 5500xl, basecalls were performed using 5500 Series Genetic Analyzers Instrument Control Software v1.2. Alignment of reads to the mm9 genome was conducted using CASAVA Eland and BFast 0.7.0 for Illumina and SOLiD sequencing runs, respectively, according to manufacturer's instructions. For HiSeq 2000 datasets, reads were aligned using bowtie using the --all --best --strata -v 2 -m 1 -S options to the mm9 genome. In all cases, only reads with a unique best scoring alignment were retained. Clonal artifacts were filtered out retaining the longer reads. Aligned reads were processed using a previously described in-house ChIP-seq analysis pipeline (3,4). Using a stepwise tag size extension in 10bp increments, fragment size was estimated in silico for DN and DP ChIP-Seq as well as MNase-Seq assays at tag extension sizes detailed in **Table S2**, for which strand overlap was maximal. Tags were then elongated by the estimated mean fragment size and converted to wig

format using a binning window of 50bp. Replicate tracks were subsequently merged by addition. Peak detection was carried out with CoCAS (5) using data averaged into 50bp windows and converted to general feature format (GFF). Thresholds were 7.5 x SD + mean for the main window and 5 x SD + mean for the extension window. Artifacts were detected and removed by performing peak detection in CoCAS (using 2 X SD + mean and SD + mean as main and extension window thresholds, respectively) in reads originating from the Input treatment. Peaks were annotated to the nearest TSS, except in the case of intragenic peaks, which were assigned to the surrounding gene. Peaks outside of [-5kb; +5kb] around known TSSs (refSeq, miRNA, rRNA, scRNA, snRNA, snoRNA and tRNA) were considered to be distal, while those falling within [-5kb; +5kb] around known TSSs were considered proximal. Distal peaks within a gene body were considered intragenic and those outside of gene bodies were considered intergenic. For short RNA-seq, we obtained the raw sequences and detected and removed the adapter sequences using the MIRO pipeline (http://seq.crg.es/main/bin/view/Home/MiroPipeline). The obtained reads were aligned against the mouse genome using the GEM aligner (http://gemlibrary.sourceforge.net). Only unambiguously aligned tags were used for further analysis. Coverage tracks were generated using non-extended tags and artifacts, showing >100 tags per bp, were removed.

DN and DP dataset tag count normalization

To account for differences in tag counts between DN and DP experiments, tag count normalization was performed by scaling DP datasets to DN tag counts. Visual inspection of invariant targets (e.g. enhancer Eβ of the *Tcrb* locus for Ets1, *β-actin* for Pol II, H3K4me1/2, MNase, the *Hoxa* locus for H3K27me3) allowed further fine tuning and scaling of DP tag-counts to reflect potential differences in signal to noise ratios between DN and DP experiments.

Ets1 fold change class definition, ranking and generation of corresponding heatmaps and average profiles

The union of all DN and DP peaks was obtained using the following methodology: first, we calculated the intersection of DN and DP peak maxima, which, to be common, had to be within 100bp of each other, since, when plotting the average summit distance between DN and DP peaks at a neighboring scale ([-400bp;+400bp] around the nearest DN peak), the distribution of distances followed a normal distribution with 2σ corresponding to 100bp. Peaks falling outside of this interval were considered specific. For all peaks, the summit tag count was assigned as the score for DN or DP peaks, respectively, using readily available summit tag counts in DN and DP for intersecting peaks, and retrieving DN or DP tag counts at the summit of specific peaks for peaks present in DN but not DP, and present in DP but not DN, respectively. The union of DN and DP peaks thus corresponded to the addition of intersecting peaks and specific peaks, or again DN∪DP = DN∩DP + DNU~DP + DPU~DN. Ets1 fold change was calculated as Log_2(DP/DN), using peak summit tag count as DN or DP score, and followed a normal distribution (**Figure S2A**). Ets1 DN+DP sites were subsequently ordered by increasing Log_2 Ets1 DP/DN fold change and split according to proximal or distal status, whereby the nearest gene had to be present on Affymetrix microarrays used to assess gene expression in DN and

DP (see below). Tag counts [-1500bp;+1500bp] every 50bp around the Ets1 summit were retrieved using a previously described methodology (4), whereby for overlapping intervals, the first interval was considered. Rows of tag counts were interpolated every three base pairs from original signal every 50bp. Artifactual signal corresponding to amplification, sequencing and alignment artifacts, characterized by the presence of > 20 tags at the same coordinate was removed. Resulting matrices featured 3900 distal sites and 1080 proximal sites, and were subsequently ranked by Log_2 DP/DN Ets1 fold change. Tag count and expresiso fold change heatmaps were generated using Java TreeView (6). Since while the distribution of Log_2 Ets1 fold change was normal, those of accordingly ranked Log_2 Ets1 DP/DN fold change did not follow normal distributions, we subdivided Log_2 DP/DN Ets1 fold change according to the following methodology: to capture the most varying populations, reflecting the normal distribution of Log_2 Ets1 DP/DN fold change, we chose the top and bottom 10% Log_2 Ets1 DP/DN enriched sites (DP/DN $Ets1^{--}$ and $Ets1^{+}$). Visual inspection of the Ets1 DN and DP signal heatmaps, ordered by increasing Log_2 Ets1 DP/DN fold change, showed that these classes corresponded to total disappearance and apparition from background of Ets1 between DN and DP. Three additional classes were also visible, which could be categorized by as corresponding to DP/DN $Ets1^{-}$, $Ets1^{-}$, showing diminished DP/DN signal loss and Ets1 stable, which corresponded to a large stable class. Class boundaries were optimized so that density plots of the isolated classes would show 5 Ets1 classes, showing unique, distinct profiles, either in DN or DP, and distinct profiles when plotting Log_2 Ets1 DP/DN fold change (**Figure S2A**). Tag counts for the coordinates of the union of distal Ets1 DN/DP sites for Ets1 DN, DP and tag counts for corresponding coordinates in other experiments, retrieved from 50bp wig tag coverage tracks, were retrieved for the entire union of coordinates and per class. Average profiles and corresponding matrices were obtained using R and heatmaps were plotted using Java TreeView.

Correlations between gene expression and Ets1 fold change

Gene names identified closest to Ets1 peaks but absent from the Mouse Gene 1.0 ST platform were discarded. Expression fold change boxplots were computed in R. The significance between gene expression levels of Ets1 fold change classes was assessed using pair wise T tests in R. For Ets1 fold change versus expression fold change correlation plots, total Ets1 tag counts and expression data were sorted by ascending Log_2 (DP over DN) Ets1 fold change and averaged into 100 points. Linear fits, Spearman correlation coefficients and graphs were computed in R. Correlation with WT DN3 and DP data from (7) was obtained using genes mapped with our Mouse Gene 1.0 ST data. Expression data for genes not present in the Mouse430_2 platform were annotated as NA.

Expression fold change and Ets1 ChIP-seq fold change clustering

To characterize groups of DN/DP varying, Ets1-bound sites correspond to significant changes in DN/DP expression, we selected sites exhibiting high normalized Log_2 DP/DN gene expression and Log_2 Ets1 DP/DN fold change, whereby significance thresholds were defined as mean ± 2σ. This selection resulted into a total of 203 sites for which Ets1 and expression fold changes were recorded as a table. Hierarchical clustering was performed using cluster 3.0 (8) using Euclidian distance

clustering, complete linkage, and array normalization. Subdivision into distal, genic, intergenic, and promoter genome locations yielded similar results (data not shown). Original, non cluster 3.0 normalized or scaled values were used to generate heatmaps. Heatmap images were generated using Java TreeView.

Gene ontology analysis

To maximize –Log p-values, the top and bottom 1,000 enriched Ets1 DP/DN sites were selected rather than classes 1 and 5 which only had 390 sites each. Gene ontology was performed using the gene ontology (–go) switch of the annotatePeaks function of the Homer package (9) on Ets1 peak coordinates, using the annotation of the nearest gene. MSigDB gene ontology annotation results were considered and were sorted by decreasing –Log p-value. P-values being available for all MSigDB classes, corresponding –Log p-values were retrieved for each analysis, e.g. for enriched classes in DN peaks, we recovered –Log p-values for those classes in DP peaks and plotted those side by side for each category in each analysis, rendering differential enrichment for gene ontology classes. For some classes, annotation had to be manually retrieved from MSigDB since not informative, notably all classes named “Module” followed by a number only.

Tissue specificity analysis of clusters of Ets1 and gene expression fold change

Gene names from the 203 top DP/DN Ets1 and expression fold change sites were uploaded to DAVID (10,11) as gene lists. Tissue expression analysis was performed using the Up_Tissue annotation tool. Results were recovered and annotated using four types of tissue specificities: Thymus (as from DAVID), Activated T-cell (as from DAVID), non T-hematopoietic (bone marrow, B-Cell, macrophage), and non hemapoietic (all remaining tissues). Counts were recovered for each Ets1/expression cluster type and expressed as percentages of gene counts in clusters. P-values of the highest enriched tissue were considered.

Public dataset retrieval and processing

We used our previously described DP ChIP-seq data, from series GSE29362: GSM726991 (Pol II N20), GSM726992 (Ets1), GSM726993 (H3K4me1), GSM726994 (H3K4me3), from series GSE38577: GSM945565 (H3K27me3) mouse DP ChIP-seq datasets as well as MNase-seq (GSM945575) and short RNA-seq (GSM945564) as well as mESC MNase-Seq (GSM945576) as they were processed (3,4), from series GSE45014: Runx1 DP (GSM1095815) (12) and from series GSE60029: CAPSTARR-Seq data for the P5424 cell line (GSM1463992 and GSM1463993) (13). Ets1 ChIP-seq datasets in other hemopoietic lineages were also downloaded from the Gene Expression Omnibus (GEO), from series GSE20898: GSM654875 (Ets1 Th2) (14), from series GSE31331: GSM777093 (Ets1 G1ME) (15,16), from series GSE40684: GSM999186 (Ets1 Treg) and GSM999187 (Ets1 $CD4^+$) (17), as well as mouse ENCODE datasets from series GSE36030: GSM1003774 (Ets1 CH12), GSM1003777 (Ets1 MEL) (18). Datasets for Tcf1 EML (series GSE31221, sample GSM773994) (19), Tcf1 Whole Thymus (WT) (series GSE46662, sample GSM1133644) (20,21), E47 DN (series GSE30518, sample GSM756892) (22), and series

GSE31233: Pu.1 DN2 (GSM774293) as well as Pu.1 DP (GSM774294) (23). Where mm9 aligned BED files were not available, SRA conversion to Fastq conversion was carried out using SRA Toolkit (24). Samples were processed as described above. We also downloaded the ENCODE WT DNaseI dataset (series GSE37074, sample GSM1014185) (18). For additional MNase-Seq assay comparison, we used previously published datasets following in vitro nucleosomal reconstitution (series GSE50762, sample GSM1228688) (25). Gene expression profiles for $CD4^+$ (GSM538374-6), $CD8^+$ (GSM538406-8) splenic $CD4^+$ T-cells (GSM605769), $CD19^+$ B-cells (GSM403988-90), common lymphoid progenitors (CLP, GSM538343-6), dendritic cells (DC, GSM854285-7), DN2 (GSM791136-8), DN3 (GSM791146-8), DN4 (GSM791154-6), DP (GSM399391-3), DP-L (GSM399397-9), DP-S (GSM399400-2), DP69 (GSM399394-6), earliest thymic progenitors (ETP, GSM854335-7), granulocyte macrophage progenitors (GMP, GSM791119-21), immature single positive cells (ISP, GSM399403-5), NK cells (GSM538315-7), NKT cells (GSM538322-4) and γδ (GSM920616-8), T4.SP69 (GSM399376-8), T4.SP24int (GSM399373-5), $T4.SP24^-$(GSM399370-2), T8.SP69 (GSM399385-7), T8.SP24int T-cells (GSM399382-4) were retrieved from the Immunological Genome Project (26) under GEO accession number GSE15907. Other Affymetrix gene expression data from WT DN3 (GSM1123162-3) and DP (GSM1123166-7) (7) mouse thymocytes were recovered from GEO accession number GSE46090. Th1 (GSM1182981-2) and Th17 (GSM1182977-8) expression microarray datasets (27) were retrieved from GEO accession number GSE48657. Th2 (GSM1194889) and Treg (GSM1619923-5) expression microarray datasets (28,29) were retrieved from GEO accession numbers GSE49166 and GSE66332, respectively.

Venn diagram overlaps

For overlaps of expressed genes and Ets1-bound genes, we assigned expressed state if gene expression was greater than the mean of all expression values, which we estimated via Gaussian fit was greater than the mean + 2σ of the Gaussian corresponding to the non expressed population. Gaussian fit was performed using the nls() function in R. Venn diagrams were generated using the venn_gchart function of the pybedtools package (30). To estimate significance of feature overlaps using intersects and Venn diagrams, we used two different approaches, based on the nature of features to overlap. For gene names, we used a hypergeometric test to determine the p-value of intersects, using the hypergeometric function provided in R, with the formula phyper($N_{intersect}$, N_A, N_T – N_A, N_B), where $N_{intersect}$ is the number of intersecting features, N_A the number of features from A, N_T the total number features (e.g. total number of genes in the genome), and N_B the number of features in B. For intersections of genomic intervals, we used the ChIPpeakAnno R package (31) which specifically implements hypergeometric tests for overlaps of genomic intervals.

ChIP-seq signal by occurrence of Ets motifs

To estimate Ets1 binding strength in de novo DP peaks with regards to the presence of Ets motif(s), thus inferring affinity of Ets1 for given motifs as well as the number of Ets1 proteins binding to motifs (17), we distinguished Ets1 peaks depleted of Ets motifs, peaks featuring the canonical T-cell Ets1 motif, Ets^{can} (CGGAAG) (32), our isolated Ets1 motif (AGGAAG) as well as peaks featuring both these

motifs. Total Ets1 DP tag counts [-100bp; +100bp] around peak summits were retrieved and plotted as boxplots in R by category. Unpaired t-tests were carried out in R using a two tailed distribution, and heteroscedastic variance.

Motif co-occurrence clustering

Motif discovery was carried out on Ets1 DN/DP varying populations using Homer, using 1σ cutoff to minimize motif p-values. Resulting outputs were converted to the BED format, and intersected with 1σ varying Ets1 DN and DP peaks using bedtools intersect (33), resulting in tables depicting motif presence for each location. To obtain motif co-occurrence, motif containing DN and DP peaks, respectively, were all intersected, using the intersection_matrix function of the pybedtools package (30). We restricted motif selections to significant, non-redundant motifs. For Ets motifs, we chose the Pu1-like motif (Ets1), AGGAAG, as well as the DP-only motif, CGGAAG (Ets^{can}). Since certain motifs will show inherently different representations than others due to differences in expected values, we sought to estimate over-representation of occurrence as compared to background co-occurrence in transcriptionally active sites. We thus chose the union of H3K4me1 DN/DP peaks as background since this mark generally represents transcriptional activity, and encompasses transcriptional poising as well as activation. Background co-occurrence was estimated using bootstrapping (1000 repetitions) of motif mapping and co-association counts (within 50bp) in randomly selected regions within the background, using equally sized populations as the original number of peaks (568 for Ets1 DN/DP down and 512 for Ets1 DN/DP up sites, corresponding to sites with fold changes greater than mean + 1σ). Motif mapping was carried out via the annotatePeaks function of Homer, using the top 15 non-redundant, non-composite matrices from the original motif discoveries. Co-localization frequency was obtained using the intersection_matrix function of pyBedTools. Enrichment was assessed by calculating a Z-score, which was represented by $Z=(X-\mu)/\sigma$, whereby μ and σ would represent background average co-occurrences and standard deviation of average, background co-occurrences. Z-scores were clustered via cluster 3.0 using hierarchical clustering, Pearson correlation for array and row clustering, using complete linkage. Heatmaps were plotted using Java TreeView.

ChIP-seq experiment correlation clustering

Tag count sums for [-1500;+1500] intervals around each coordinate of the distal Ets1 DN/DP union were retrieved and collated under the form of three matrices: 1) Ets1 DN, Ets1 DP, all other Ets1 datasets, 2) Ets1 DN, Runx1 DN, TCF1 EML, E47 DN3, and 3) Ets1 DP, Runx1 DP, TCF1 WT, E47 DP, respectively. Spearman correlation coefficients were computed for each matrix in R. Hierarchical clustering by spearman correlation was conducted using Eigen cluster 3.0, using spearman array and row correlation, single linkage, and corresponding heatmaps were subsequently generated using Java TreeView.

Motif discovery and heatmap/average profile generation

We used the findMotifsGenome function in Homer (9) for primary motif detection in each population, where de novo discovery was performed in regions ranging [-200bp; +200bp] around the maximum

Ets1 signal. Identified motif matrices corresponded to the closest Homer built-in matrix match (collated from public databases). Motif reinjection was carried out [-1500bp; +1500bp] around Ets1 summits using RSA Tools Matrix-Scan (34). Resulting outputs were converted as distances to each region center, and subsequently expressed as tables ranging [-1500bp; +1500bp] around the Ets1 maximum using custom Perl scripts, whereby motif frequencies were computed every 10 bp for all regions, ordered according to fold changes, using a previously described pipeline (4). Heatmaps were generated via Java Treeview. Motif densities were computed relative to each Ets1 peak maximum, where distances used were that between the start of each motif (regardless of the strand) and the Ets1 peak maximum.

Nucleosome depleted region determination and ranking

Nucleosome free regions (NDR) were determined by carrying out peak detection, using inverse MNase signal, whereby zero values were replaced by 0.01. Peak detection was carried out via CoCAS, with 100 as the main and extension threshold. Coordinates of the center of distal Ets1 DN+DP union were intersected with DN and DP NDRs. Classes of nucleosome free, and nucleosome-occupied regions were defined for DN and DP and subsequently ordered by increasing MNase signal [-100;+100] around the Ets1 site. Visual inspection of heatmaps allowed fine tuning of DN, DP NDR and nucleosome-occupied classes of Ets1 DN/DP union coordinates, resulting in 2044 nucleosome-occupied, 1856 NDR Ets1 sites in DN, and 2072 nucleosome-occupied, 1828 NDR Ets1 sites in DP. To better estimate nucleosome signal as well as those of TFs and histone modification marks around corresponding coordinates, nucleosome-occupied and NDR classes were further split into two equal classes for DN and DP, respectively. Following retrieval of corresponding tag counts, average profiles were plotted using R, and heatmaps generated using Java TreeView.

Intersection of CGI and Ets1 DN+DP distal peaks

The UCSC CpG island track (mm9_cpgIslandExt) was obtained as a bed file and intersected with Ets1 DN+DP distal peaks.

Promoter unions

Unions of promoter Ets1 DN+DP peaks were performed for peaks falling within [-5Kb; +5Kb] as described above for distal peaks. Classes were defined as < -1, ≥ -1 and ≤ 1, >1 $\log_2$ Ets1 DP/DN fold change. Average profiles and heatmaps were obtained as for the distal DN+DP unions.

CAPSTARR-Seq analyses

CAPSTARR-Seq data was retrieved and processed as other high-throughput datasets, and ultimately output as a 10bp wig file covering the mouse genome. Average profiles were plotted as for other high-throughput datasets.

Pol II directionality with regards to Ets1 motif

Average Pol II values were recovered in DN and DP around Ets1 motifs in Ets1 peaks, taking into account the orientation of the Ets1 motif. Values were plotted as boxplots for each 10bp window ranging [-200bp; +200bp] around Ets1 motifs.

Correlation clustering of T-cell gene expression datasets

A data matrix combining the Immunological Genome Project data as well as our $Rag2^{-/-}$ and DP data was created and normalized with limma using R. Differentially regulated genes were identified as the ones showing the top 1% variance among normalized $\log_2$ signal intensities. A Pearson correlation coefficient matrix was calculated for the data table, which was further clustered using cluster 3.0. Heatmap images were generated using Java Treeview.

Distances of ETScan motifs to ETS motifs

Positions of motifs relative to the reference motifs were first retrieved by mapping motifs [-200bp,+200bp] of DN+DP summits as bed files using Homer annotatePeaks, with very strict ETS^{can} and ETS1 position weight matrices (strictly CGGAAG and AGGAAG, respectively), using -size 400 -m -mbed as parameters. ETS^{can} motifs were annotated as the closest feature to ETS1 motifs via bedtools closest. Distances were derived by computing the difference between motif starts, taking into account the orientation of the reference motif (i.e. negating the distance if the reference motif was on the negative strand). Motifs on the same strand and opposite strand were separated. Oriented distances were plotted using R via the hist function. P-values for occurrences at a given position within 200bp were calculated using χ^2 tests. Strand occurrences p-values were computed using a binomial test using 0.5 as the probability of each strand event. P-values were obtained by computing the probability to obtain 10 motif pairs at a fixed spacing out of 346 at random (χ^2 p-value=3.22×10^{-10}), and that 10 out of these 10 pairs were located on the same strand at random (binomial p-value=9.77×10^{-4}). The global p-value for the occurrence of these two events corresponded to their product.

Ets1-/- DP thymocyte gene expression analyses

$Ets1^{-/-}$ DP thymocytes were generated as previously described (35). Gene expression microarray assays were performed as described for other microarray assays in this work and processed via limma in R. Ets KO down- and up- regulated genes were defined as those displaying a KO/WT fold change ≤-1 or ≥1, respectively. Heatmaps and boxplots were obtained as for the DP and DN datasets. To measure the Ets1 DP/DN summit tag count fold change in Ets1 KO down- and up- regulated genes, the Ets1 ChIP-Seq tag count at the DN+DP summits corresponding to one of those genes was recovered and its $\log^2$ DP/DN value plotted as boxplots using R.

Analysis of Ets1-/-, Ets1 WT, T-cell and other hemopoietic gene expression datasets

Gene expression profiles were retrieved from the Immunological Project (26) for $CD19^{+}$, CLP, DC, DN2, DN3, DN4, DP, DP69, DPbl, DPsm, ETP, GMP, ISP, NK, NKT, T.4nve, T.8nve, $T4.SP24^{-}$, T4.SP24int, T4.SP69, T8.SP24int, T8.SP69, and Tγδ cells. Th1, Th2, Th17 and Treg datasets, which are not present at the Immunological Genome project, were obtained from separate studies (27-29) . Data were processed as described above for the MoGene 1.0 ST platform and separately from the DP $Ets1^{-/-}$ and WT datasets which were carried out on the MoGene 2.1 ST platform. To combine these data, results from each platform were first aggregated by gene name in R using the aggregate() function, using the mean value of probesets, so as to obtain unique values per gene. Merging was carried out using the merge function in R. The resulting data frame was then normalized using the normalizeQuantiles() function of the limma R package. PCAs were carried out using the prcomp() function in R on either actual values or fold change over DP WT of the combined data frame for a list of genes corresponding to the union of both $Ets1^{-/-}$ up- and down- regulated genes. Expression boxplots were plotted for both $Ets1^{-/-}$ up- and down- regulated genes, whereby datasets were sorted by decreasing median expression value.

ChIP-Seq fold change clustering

To estimate which transcriptional marks showed highest correlations of fold changes between DN and DP tag count sums for [-1500;+1500] intervals around each coordinate of the distal Ets1 DN/DP union were retrieved and collated for Ets1 DN/DP, Pol II DN/DP, H3K4me1 DN/DP, H3K4me3 DN/DP, H3K27me3 DN/DP. Fold changes were subsequently computed for each pair. Hierarchichal clustering by Pearson correlation was conducted using cluster 3.0, using array and row normalization, Pearson array and row correlation, single linkage, and corresponding heatmaps were subsequently generated using Java TreeView. Node Pearson correlation coefficients represent the clustering distances of each branch. Original, non cluster 3.0 normalized or scaled values were used to generate heatmaps. P-values were obtained by performing t-tests on H3K4me1 and H3K4me3 signals comparing sh-scr and sh-ets1 signals, using the Bonferroni correction for multiple testing in R.

Average profiles and heatmaps of H3K4me1, H3K4me3 and H3K27ac ChIPs in Ets1 and control knockdown P5424 cells

Peak detection was carried out as in P5424 Ets1 using the same methodology and thresholds as for DN and DP Ets1 ChIP-Seq datasets, with proximal and distal peaks subsequently separated. Total average profiles were obtained as for other high-throughput datasets. For heatmaps and average profiles per class, total P5424 sh-scr H3K4me1 signal was retrieved [-500bp,+500bp] around distal P5424 Ets1 summits and sorted by increasing signal. Average profiles per class and heatmaps based on this ranking were derived using the same methodology as for other high-throughput datasets. To measure that the increases in H3K4me1 signals following knockdown were specific to the loss of Ets1, we used control sites comprising all distal ENCODE thymus DHSs depleted for Ets1 binding, i.e. overlapping with Ets1 in neither DN, DP nor P5424.

CAPSTARR-Seq signal by occurrence of Ets motifs

To estimate enhancer activity in de novo DP peaks with regards to the presence of Ets motif(s), we distinguished Ets1 peaks depleted of Ets motifs, peaks featuring the canonical T-cell Ets1 motif (CGGAAG) (32), our isolated Ets1 motif (AGGAAG) as well as peaks featuring both these motifs. Average P5424 CAPSTARR-Seq tag counts [-100bp; +100bp] around peak summits were retrieved and plotted as boxplots in R by category. Unpaired t-tests were carried out in R using a two tailed distribution, and heteroscedastic variance.

III. SUPPLEMENTARY REFERENCES

1. Hoffmann, R., Bruno, L., Seidl, T., Rolink, A. and Melchers, F. (2003) Rules for gene usage inferred from a comparison of large-scale gene expression profiles of T and B lymphocyte development. *J Immunol*, 170, 1339-1353.

2. Puthier, D., Joly, F., Irla, M., Saade, M., Victorero, G., Loriod, B. and Nguyen, C. (2004) A general survey of thymocyte differentiation by transcriptional analysis of knockout mouse models. *J Immunol*, 173, 6109-6118.

3. Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C. *et al.* (2011) Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol*, 18, 956-963.

4. Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J.Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I. *et al.* (2012) CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res*, 22, 2399-2408.

5. Benoukraf, T., Cauchy, P., Fenouil, R., Jeanniard, A., Koch, F., Jaeger, S., Thieffry, D., Imbert, J., Andrau, J.C., Spicuglia, S. *et al.* (2009) CoCAS: a ChIP-on-chip analysis suite. *Bioinformatics*, 25, 954-955.

6. Saldanha, A.J. (2004) Java Treeview--extensible visualization of microarray data. *Bioinformatics*, 20, 3246-3248.

7. Geimer Le Lay, A.S., Oravecz, A., Mastio, J., Jung, C., Marchal, P., Ebel, C., Dembele, D., Jost, B., Le Gras, S., Thibault, C. *et al.* (2014) The tumor suppressor Ikaros shapes the repertoire of notch target genes in T cells. *Sci Signal*, 7, ra28.

8. de Hoon, M.J., Imoto, S., Nolan, J. and Miyano, S. (2004) Open source clustering software. *Bioinformatics*, 20, 1453-1454.

9. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, 38, 576-589.

10. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4, 44-57.

11. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37, 1-13.

12. Lepoivre, C., Belhocine, M., Bergon, A., Griffon, A., Yammine, M., Vanhille, L., Zacarias-Cabeza, J., Garibal, M.A., Koch, F., Maqbool, M.A. *et al.* (2013) Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics*, 14, 914.

13. Vanhille, L., Griffon, A., Maqbool, M.A., Zacarias-Cabeza, J., Dao, L.T., Fernandez, N., Ballester, B., Andrau, J.C. and Spicuglia, S. (2015) High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun*, 6, 6905.

14. Wei, G., Abraham, B.J., Yagi, R., Jothi, R., Cui, K., Sharma, S., Narlikar, L., Northrup, D.L., Tang, Q., Paul, W.E. *et al.* (2011) Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. *Immunity*, 35, 299-311.

15. Dore, L.C., Chlon, T.M., Brown, C.D., White, K.P. and Crispino, J.D. (2012) Chromatin occupancy analysis reveals genome-wide GATA factor switching during hematopoiesis. *Blood*, 119, 3724-3733.

16. Chlon, T.M., Dore, L.C. and Crispino, J.D. (2012) Cofactor-mediated restriction of GATA-1 chromatin occupancy coordinates lineage-specific gene expression. *Mol Cell*, 47, 608-621.

17. Samstein, R.M., Arvey, A., Josefowicz, S.Z., Peng, X., Reynolds, A., Sandstrom, R., Neph, S., Sabo, P., Kim, J.M., Liao, W. *et al.* (2012) Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell*, 151, 153-166.

18. Raney, B.J., Cline, M.S., Rosenbloom, K.R., Dreszer, T.R., Learned, K., Barber, G.P., Meyer, L.R., Sloan, C.A., Malladi, V.S., Roskin, K.M. *et al.* (2011) ENCODE whole-genome data in the UCSC genome browser (2011 update). *Nucleic Acids Res*, 39, D871-875.

19. Wu, J.Q., Seay, M., Schulz, V.P., Hariharan, M., Tuck, D., Lian, J., Du, J., Shi, M., Ye, Z., Gerstein, M. *et al.* (2012) Tcf7 is an important regulator of the switch of self-renewal and differentiation in a multipotential hematopoietic cell line. *PLoS Genet*, 8, e1002565.

20. Li, L., Zhang, J.A., Dose, M., Kueh, H.Y., Mosadeghi, R., Gounari, F. and Rothenberg, E.V. (2013) A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. *Blood*, 122, 902-911.

21. Dose, M., Emmanuel, A.O., Chaumeil, J., Zhang, J., Sun, T., Germar, K., Aghajani, K., Davis, E.M., Keerthivasan, S., Bredemeyer, A.L. *et al.* (2014) beta-Catenin induces T-cell transformation by promoting genomic instability. *Proc Natl Acad Sci U S A*, 111, 391-396.

22. Miyazaki, M., Rivera, R.R., Miyazaki, K., Lin, Y.C., Agata, Y. and Murre, C. (2011) The opposing roles of the transcription factor E2A and its antagonist Id3 that orchestrate and enforce the naive fate of T cells. *Nat Immunol*, 12, 992-1001.

23. Zhang, J.A., Mortazavi, A., Williams, B.A., Wold, B.J. and Rothenberg, E.V. (2012) Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell*, 149, 467-482.

24. Leinonen, R., Sugawara, H. and Shumway, M. (2011) The sequence read archive. *Nucleic Acids Res*, 39, D19-21.

25. Barozzi, I., Simonatto, M., Bonifacio, S., Yang, L., Rohs, R., Ghisletti, S. and Natoli, G. (2014) Coregulation of Transcription Factor Binding and Nucleosome Occupancy through DNA Features of Mammalian Enhancers. *Mol Cell*, 54, 844-857.

26. Shay, T. and Kang, J. (2013) Immunological Genome Project and systems immunology. *Trends Immunol*, 34, 602-609.

27. Ichiyama, K., Chen, T., Wang, X., Yan, X., Kim, B.S., Tanaka, S., Ndiaye-Lobry, D., Deng, Y., Zou, Y., Zheng, P. *et al.* (2015) The methylcytosine dioxygenase Tet2 promotes DNA demethylation and activation of cytokine gene expression in T cells. *Immunity*, 42, 613-626.

28. Lin, C.C., Bradstreet, T.R., Schwarzkopf, E.A., Sim, J., Carrero, J.A., Chou, C., Cook, L.E., Egawa, T., Taneja, R., Murphy, T.L. *et al.* (2014) Bhlhe40 controls cytokine production by T cells and is essential for pathogenicity in autoimmune neuroinflammation. *Nat Commun*, 5, 3551.

29. Yang, S., Fujikado, N., Kolodin, D., Benoist, C. and Mathis, D. (2015) Immune tolerance. Regulatory T cells generated early in life play a distinct role in maintaining self-tolerance. *Science*, 348, 589-594.

30. Dale, R.K., Pedersen, B.S. and Quinlan, A.R. (2011) Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics*, 27, 3423-3424.

31. Zhu, L.J., Gazin, C., Lawson, N.D., Pages, H., Lin, S.M., Lapointe, D.S. and Green, M.R. (2010) ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*, 11, 237.

32. Hollenhorst, P.C., Chandler, K.J., Poulsen, R.L., Johnson, W.E., Speck, N.A. and Graves, B.J. (2009) DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genet*, 5, e1000778.

33. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841-842.

34. Turatsinze, J.V., Thomas-Chollier, M., Defrance, M. and van Helden, J. (2008) Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. *Nat Protoc*, 3, 1578-1588.

35. Eyquem, S., Chemin, K., Fasseu, M. and Bories, J.C. (2004) The Ets-1 transcription factor is required for complete pre-T cell receptor function and allelic exclusion at the T cell receptor beta locus. *Proc Natl Acad Sci U S A*, 101, 15712-15717.