SUPPLEMENTAL MATERIAL

Ungerbäck, Hosokawa, et al.

Contents

Reagent and Resource Tables

1. Antibodies
2. Chemicals, peptides, and recombinant proteins
3. Critical commercial assays
4. Deposited and Downloaded Data
5. Experimental models: Mouse strains and cell lines
6. Oligonucleotides
7. Recombinant DNA Software and algorithms

Other References for software and deposited datasets

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD44 eFluor-450, clone 1M7	eBioscience	Cat#: 48-0441-82
CD44 PE, clone 1M7	eBioscience	Cat#: 12-0441-83
CD117 (cKit) PE, clone 2B8	eBioscience	Cat#:12-1171-83
CD117 (cKit) Apce780, clone 2B8	eBioscience	Cat#: 47-1171-82
CD117 (cKit) Apc, clone 2B8	eBioscience	Cat#: 17-1171-83
CD25 Apc, clone PC61.5	eBioscience	Cat#: 17-0251-82
CD25 eFluor-450, clone PC61.5	eBioscience	Cat#: 48-0251-82
CD25 Apce780, clone PC61.5	eBioscience	Cat#: 47-0251-82
CD45 PECy7, clone 30-F11	eBioscience	Cat#: 25-0451-81
CD45 PE, clone 30-F11	eBioscience	Cat#: 12-0451-83
CD45 AF488, 30-F11	BioLegend	Cat#: 103122
CD11b APC-e780, clone M1/70	eBioscience	Cat#: 47-0112-82
CD11b PE, clone M1/70	eBioscience	Cat#: 12-0112-85
CD25 biotin, clone PC61.5	eBioscience	Cat#: 13-0251-86
NK1.1 biotin, clone PK136	BioLegend	Cat#: 108704
B220 biotin, clone RA3-6B2	BioLegend	Cat#: 103204
CD19 biotin, clone 1D3	BioLegend	Cat#: 115504
Ter119 biotin, clone TER-119	eBioscience	Cat#: 116204
Gr-1 biotin, clone RB6-8C5	eBioscience	Cat#: 13-5931-86
CD11b biotin, clone M1/70	eBioscience	Cat#: 13-0112-86
CD3e biotin, clone 145-2C11	BioLegend	Cat#: 100304
F4/80 biotin, clone BM8	eBioscience	Cat#: 13-4801-85
CD11c biotin, clone N418	eBioscience	Cat#: 13-0114-85
CD8a biotin, clone 53-6.7	eBioscience	Cat#: 13-0081-86
TCR$\gamma\delta$ biotin, clone eBioGL3	eBioscience	Cat#:13-5711-85
TCRβ biotin, clone H57-597	eBioscience	Cat#: 13-5961-85
Streptavidin PerCP-Cy5.5	eBioscience	Cat#: 45-4317-82
PU.1 AF647, clone 9G7	CellSignaling	Cat#: 2240S
HA-tag AF647, clone 6E2	CellSignaling	Cat#: 3444S
Rabbit anti-PU.1 polyclonal	Santa Cruz Biotechnology	Cat#: sc-352x
Rabbit HA-probe polyclonal IgG	Santa Cruz Biotechnology	Cat#: sc-805x
Rabbit anti-Ets1 polyclonal IgG	Santa Cruz Biotechnology	Cat#: sc-350x
Rabbit polyclonal anti-H3K27Ac	Abcam	Cat#: ab4729
Rabbit polyclonal anti-H3K4me2	Millipore	Cat#: 07-030
Rabbit polyclonal anti-H3K27me3	Millipore	Cat#: 07-449
Anti-Mouse IgG HRP	GE Healthcare	Cat#: #NA931V
Anti-Rabbit IgG HRP	GE Healthcare	Cat#: NA934V

Anti-hNGFR; PE-human CD271 (NGFR)	eBioscience	Cat#: 12-9400-42
Chemicals, Peptides, and Recombinant Proteins		
MEM Alpha	Gibco	Cat#: 12561-056
RPMI1640	Gibco	Cat#: 31800-022
DMEM	Gibco	Cat#: 12430-05
Fetal Bovine Serum	SigmaAldrich	Cat#: F7305
Human IL-7	PeproTech Inc	Cat#: 200-07
Human FLT-3-Ligand	PeproTech Inc	Cat#: 300-19
Stem Cell Factor	PeproTech Inc	Cat#: 250-03
2.4G2 cell supernatant.	This study	N/A
HBSS	Gibco	Cat#: 14175-095
HEPES	Gibco	Cat#: 15630-080
MEM NEAA	Gibco	Cat#: 11140-050
Sodium Puryvate	Gibco	Cat#: 11360-070
Pen Strep Glutamine	Gibco	Cat#: 10378-016
Retronectin	Takara	Cat#: T100B
MACS Streptavidin Microbeads	Miltenyi Biotec	Cat#: 130-048-101
NEBNext High-Fidelity 2XPCR MasterMix	NEB	Cat#: M0541L
Invitrogen SYBR Green I Dye	Invitrogen	Cat#: S7563
FuGENE 6 Transfection Reagent	Promega	Cat#: E2691
4-hydroxytamoxifen	SigmaAldrich	Cat#: H6278
Puromycin dihydrochloride	SigmaAldrich	Cat#: P8833
Complete, EDTA-free protease inhibitor cocktail	Roche	Cat#: 11873580001
Dynabeads M-280 Sheep anti-Mouse IgG	Invitrogen	Cat#: 11202D
Dynabeads M-280 Sheep anti-Rabbit IgG	Invitrogen	Cat#: 11204D
37% formaldehyde	ThermoFisher Scientific	Cat#: 28906
DSG (disuccinimidyl glutarate)	ThermoFisher Scientific	Cat#: 20593
7AAD	eBioscience	Cat#: 00-6993-50
ZombieAqua™ Viability Dye	BioLegend	Cat#: 423101
β-mercaptoethanol	SigmaAldrich	Cat#: M6250
Proteinase K Solution	ThermoFisher Scientific	Cat#: AM2548
Critical Commercial Assays		

Illumina Nextera DNA preparation Kit	Illumina	Cat#: FC-121-1030
Nextera Index Kit (96 indexes, 384 samples)	Illumina	Cat#: FC-121-1012
ChIP DNA Clean and Concentrator	Zymo Research	Cat#: D5205
Qiagen MiniElute PCR Purification Kit	Qiagen	Cat#: 28004
RNeasy Micro Kit	Qiagen	Cat#: 74004
NEBNext ChIP-Seq Library Preparation Kit	NEB	Cat#: E6240
NEBNext Ultra RNA Library Prep Kit	NEB	Cat#: E7530
High Sensitivity DNA Kit	Agilent Technologies	Cat#: 5067- 4626
Qubit dsDNA HS Kit	ThermoFisher Scientific	Cat#: Q32854
NE-PER Nuclear and Cytoplasmic Extraction Reagent	ThermoFisher Scientific	Cat#: 78833
SPRIselect reagent kit	Beckman Coulter	Cat#: B23318
Agencourt AMPure XP beads	Beckman Coulter	Cat#: A63880
ECL Reagents	GE Healthcare	Cat#: PRN2106
Deposited and Downloaded Data		
Constrained pioneering and partner factor redirection by PU.1 shape early T-cell gene regulation	GEO dataset generated in this study: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93755	GEO: GSE93755
PU.1 represses and activates gene expression in early T cells by redirection of transcription factor ensembles	GEO dataset generated in companion study (Hosokawa et al. 2018): https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse110020	GEO: GSE110020
Dynamic Transformations of Genome-wide Epigenetic Marking and Transcriptional Control Establish T Cell Identity	Downloaded GEO dataset: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31235 (Zhang et al. 2012),	GEO: GSE31235
PU.1 regulates T-lineage gene expression and progression via indirect repression during early T-cell development	Downloaded GEO datasets: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65344 (Champhekar et al. 2015)	GEO: GSE65344

CapStarr-seq: a high-throughput method for quantitative assessment of enhancer activity in mammals (ChIP-Seq)	Downloaded GEO datasets: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63732 (Vanhille et al. 2015)	GEO: GSE63732
Experimental Models: Cell Lines		
Scid.adh.2C2 cells	(Dionne et al. 2005).	N/A
Experimental Models: Organisms/Strains		
C57BL/6	Jackson laboratories	Stock nr 000664
Gt(ROSA)26Sortm1.1(CAG-cas9*,-EGFP)Fezh/J (CAS9)	Jackson laboratories	Stock nr 024858
Oligonucleotides		
PU1-ER(T2)_FOR ATG TGA ATT CAT GTT ACA GGC GTG CAA AAT GG	This paper: Scid.adh.PU.1-ERT2 experiment	N/A
PU1-ER(T2)_REV ATG TCT CGA GGT GGG GCG GGA GGC GC	This paper: Scid.adh.PU.1-ERT2 experiment	N/A
ert2_forward_pcr AAA AGA ATT CAT GTC TGC TGG AGA CAT GAG AGC	This paper: Scid.adh.PU.1-ERT2 experiment	N/A
ert2_reverse _pcr AAA AGC GGC CGC TCA AGC TGT GGC AGG GAA	This paper: Scid.adh.PU.1-ERT2 experiment	N/A
sgControl (Luciferase) ggcatttcgcagcctaccg	(Hosokawa et al. 2018)	
sgPU.1 #1 gccccagtactcacagggg	(Hosokawa et al. 2018)	N/A
sgPU.1 #2 catgactactactccttcg	(Hosokawa et al. 2018)	N/A
sgPU.1 #3 gcgatggagaaagccatag	(Hosokawa et al. 2018)	N/A
sgRunx1 #1 gctcgtgctggcatctacg	(Hosokawa et al. 2018)	N/A
sgRunx1 #2 agccccggcaagatgagcg	(Hosokawa et al. 2018)	N/A
sgRunx1 #3 agcggcgaccgcagcatgg	(Hosokawa et al. 2018)	N/A
Recombinant DNA		

Lzrs-EVGFP	www.addgene.org	Plasmid #21961
Lzrs-PU1WTHA	(Champhekar et al. 2015)	N/A
Lzrs-PU1ENGHA	(Champhekar et al. 2015)	N/A
Lzrs-PU1ETSHA	(Champhekar et al. 2015)	N/A
Lzrs-PU.1-ERT2	This study	N/A
Lzrs-ERT2	This study	N/A
pQCXIN-EF1a-mNeonGreen-P2A-Cas9 ("Cas9-GFP")	(Hosokawa et al. 2018)	N/A
E42 dTet-CFP	(Hosokawa et al. 2018)	N/A
pMxs-Myc-Flag-PU.1-IRES-hNGFR	(Hosokawa et al. 2018)	N/A
Software and Algorithms		
bedGraphToBigWig		http://hgdownload.soe.ucsc.edu/admin/exe/
Bedtools (v.2.17.0)	(Quinlan and Hall 2010)	http://bedtools.readthedocs.io/en/latest/
Bioconductor (v3.4)		http://bioconductor.org/
Bowtie (v1.1.1)	(Langmead et al. 2009)	http://bowtie-bio.sourceforge.net/index.shtml
CHOPCHOP	(Montague et al. 2014)	https://chopchop.rc.fas.harvard.edu/
Cluster3 (v1.52)	(de Hoon et al. 2004)	http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm
DESeq2 (v.1.14.1)	(Love et al. 2014)	https://www.bioconductor.org/packages/devel/bioc/html/DESeq2.html
EaSeq		http://easeq.net/
EdgeR (v.3.16.5)	(Robinson et al. 2010)	http://bioconductor.org/packages/release/bioc/html/edgeR.html
FlowJo (v10.0.8)		https://www.flowjo.com/
Ggplot2 (v.2.2.1)		http://ggplot2.org/

HOMER (v4.8)	(Heinz et al. 2010)	http://homer.ucsd.edu/homer/
HOMER-IDR		https://github.com/karmel/homer-idr
Limma (v.3.30.11)	(Ritchie et al. 2015)	http://bioconductor.org/packages/release/bioc/html/limma.htm
MatLab (R2016a)		https://www.mathworks.com/products/matlab.html
Pheatmaps (v1.0.8)		https://cran.r-project.org/web/packages/pheatmap/index.html
R (v3.3.2)		https://www.r-project.org/
RSEM (v1.2.25)	(Li and Dewey 2011)	http://deweylab.github.io/RSEM/
Rstudio (v1.0.136)		https://www.rstudio.com/
Samtools (v0.1.19-96b5f2294a)	(Li et al. 2009)	http://samtools.sourceforge.net/
STAR (v2.4.0)	(Dobin et al. 2013)	https://github.com/alexdobin/STAR/releases
TreeView (v1.1.6r4)	(Saldanha 2004)	http://jtreeview.sourceforge.net/
Trimmomatic (v.033)	(Bolger et al. 2014)	http://www.usadellab.org/cms/?page=trimmomatic
Other		
BD FACS Aria II Cell Sorter	BD Bioscience	
Illumina HiSeq 2500	Illumina	
Illumina NextSeq500	Illumina	
7900HT Fast Real-Time PCR System	Applied Biosystems	
BioRuptor	Diagenode	
iCyt Mission Technology Reflection Cell Sorter	Sony	
Miltenyi Biotech MACSQuant 10 Flow Cytometer	Miltenyi Biotec	

LITERATURE CITED FOR REAGENTS AND RESOURCES

Bolger AM, Lohse M, . Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Champhekar A, Damle SS, Freedman G, Carotta S, Nutt SL, . Rothenberg EV. 2015. Regulation of early T-lineage gene expression and developmental progression by the progenitor cell transcription factor PU.1. *Genes Dev* **29:** 832-848.

de Hoon MJ, Imoto S, Nolan J, Miyano S. 2004. Open source clustering software. *Bioinformatics* **20**: 1453-1454.

Dionne CJ, Tse KY, Weiss AH, Franco CB, Wiest DL, Anderson MK, . Rothenberg EV. 2005. Subversion of T lineage commitment by PU.1 in a clonal cell line system. *Dev Biol* **280**: 448-466.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.

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

Hosokawa H, Ungerbäck J, Wang X, Matsumoto M, Nakayama KI, Cohen SM, Tanaka T, Rothenberg EV. 2018. Transcription factor PU.1 represses and activates gene

expression in early T cells by redirecting partner transcription factor binding. *Immunity* **48**: 1119-1134.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. 2014. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucl Acids Res* **42**: W401-407.

Quinlan AR, Hall IM 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucl Acids Res* **43**: e47.

Robinson MD, McCarthy DJ, and Smyth GK 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Saldanha AJ 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**: 3246-3248.

Vanhille L, Griffon A, Maqbool MA, Zacarias-Cabeza J, Dao LT, Fernandez N, Ballester B, Andrau JC, Spicuglia S. 2015. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commu*n **6**: 6905.

Zhang JA, Mortazavi A, Williams BA, Wold BJ, and Rothenberg EV. 2012. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* **149**: 467-482.

COMPLETE EXPERIMENTAL PROCEDURES

Mice

C57BL/6, *B6.Cg-Tg(BCL2)25*^{Wehi}/J (Bcl2-tg) and B6.*Gt(ROSA)26Sor*$^{tm1.1(CAG-cas9*,-EGFP)Fezh}$/J (CAS9) mice were purchased from the Jackson Laboratory. Transgenics were maintained as homozygotes. For CRISPR experiments in primary cells, we used F_1 progeny of Bcl2-tg and CAS9 mice. All animals were bred and maintained in the California Institute of Technology Laboratory Animal Facility, under specific pathogen free conditions, and the protocol supporting animal breeding for this work was reviewed and approved by the Institute Animal Care and Use Committee of the California Institute of Technology.

Developmental cell-surface marker nomenclature conventions

A comment on marker nomenclature is needed for the cell population definition in the following. Antibodies used to characterize hematopoietic populations are normally referred to and listed by suppliers under the target molecules' common names, a codified nomenclature but one which does not always correspond to the names of the protein coding loci. This is important to note because some important cell type-specific determinants used to make developmental distinctions are actually post-translational modifications of proteins and not only the proteins themselves. For the experiments here, the most important markers are CD25, representing IL2RA; CD44; CD11b, representing ITGAM; c-Kit, representing KIT; CD45, representing PTPRC; CD4; and CD8, representing CD8A. Other important markers used for purification of the cells are CD11c, representing ITGAX; Gr1, representing LY6G; TER-119, representing LY76; NK1.1, representing KLRB1C; CD19; F4/80, representing ADGRE1; CD3ε, representing CD3E; B220,

representing a specific modified isoform of PTPRC called “CD45R”; TCRβ, representing any product of a productively rearranged *Tcrb* locus; and TCRγδ, representing a heterodimer of any products of productively rearranged *Tcrg* and *Tcrd* loci. The term “Lin” is used to refer to mature-cell markers detected by mixtures of antibodies that are used to remove unwanted cells from a progenitor population.

Cell sources and cell culture conditions

Fresh thymocytes used in the ATAC-seq analysis were taken from C57BL/6 animals bred in our mouse colony at Caltech, at less than 6 weeks of age. DN thymocytes were enriched by depleting CD4, CD8, TCR, and other lineage-marker expressing cells prior to staining and sorting to isolate DN subsets. In detail, 1.2×10^9 thymocytes from 3 B6 pups (3 wk old) were used for DN preparation. More mature cells were pre-depleted by staining with biotin-conjugated antibodies against CD8α, TCRγδ, TCRβ and Ter119 followed by streptavidin-magnetic bead depletion (MACS), yielding ~10 million DN thymocytes. After sorting DN1 and DN3 cells (see Flow Cytometry section, below), representative cell numbers for each population were DN1: 18×10^3 and DN3: 50×10^3. Two independent DN thymocyte preparations were performed for 2 biological replicates of each.

For other samples used in this study, primary cells were taken from in vitro differentiation of freshly isolated multipotent hematopoietic precursors rather than from fresh thymus, similarly to the strategy used in our previous report (Zhang et al. 2012), as the number of early-stage DN pro-T cells obtainable from a single fresh thymus is quite low. Except as noted, the ChIP-seq analyses and RNA-seq analyses after perturbation were done on samples of hematopoietic

progenitors that had been initiated into the T-cell developmental pathway and differentiated to DN1-DN3 stages in vitro, using OP9-DL1 stromal co-cultures as described in detail below. The precursors for these cultures were obtained from fetal liver for most of the experiments with primary cells. In some experiments to test the effect of acute CAS9-mediated deletion (Fig. 2D-G, Fig. 3B,C Table S5), precursors were obtained from adult bone marrow to ensure slower developmental progression *in vitro*. Details of different experimental protocols are presented in separate sections. Starting cells were prepared and sorted from the cultures as described below.

Fetal livers (FL) were dissected from E13.5 (day of plug, E0.5) C57BL/6 fetuses. Suspensions of FL cells were then prepared, stained for lineage markers using biotin-conjugated lineage antibodies (CD11c, Gr1, TER-119, NK1.1, CD19, F4/80), incubated with streptavidin-coated magnetic beads (Miltenyi Biotec), and Lin^+ cells were removed by passage through a magnetic column (Miltenyi Biotec). Lineage-depleted (Lin−) cells were eluted and stored in liquid nitrogen in freezing media [50% fetal bovine serum (FBS), 40% αMEM, 10% DMSO] for future use. For each experiment, frozen, lineage marker-depleted FL precursor cells from E13.5 (FLPs) were thawed and used to initiate OP9-DL1 cultures in OP9 medium (α-MEM, 20% FBS, 50 μM β-mercaptoethanol, Pen-Step-Glutamine) supplemented with IL-7 and FLT3L (5 ng/ml each), using slight modifications of methods previously described (Champhekar et al. 2015). On culture day 7, hematopoietic cells were harvested and enriched for $CD25^+$ cells (DN2-DN3 stages). Briefly, a biotin-conjugated antibody against CD25 (clone PC61.5, eBioscience) was used to stain the cells, and $CD25^+$ cells were then enriched using streptavidin bound magnetic beads (Miltenyi) per manufacturer's protocol.

For use in acute CAS9-mediated deletion experiments, bone marrow (BM) was removed from the femurs and tibiae of 2-3 month-old (Cas9 × Bcl2-tg)F_1 mice. Suspensions of BM cells were prepared and stained for lineage markers using biotin-conjugated lineage antibodies (CD11b, CD11c, Gr1, TER-119, NK1.1, CD19, CD3ε, B220), then depleted of Lin^+ cells on a magnetic column as above. Lin^- bone marrow cells were eluted and cultured on OP9-DL1 monolayers using OP9 medium supplemented with 10 ng/ml of IL-7 and 10 ng/ml of FLT3L.

To sort freshly-isolated DN ($CD4^-$ $CD8^-$) thymocytes, cells were stained with fluorescent antibodies against CD45 (PTPRC), CD44, c-Kit (KIT), and CD25 (IL2RA), and a biotin-conjugated lineage marker cocktail (CD11b, CD11c, Gr1, TER-119, NK1.1, CD19, CD3, CD8,TCRγδ,TCRβ), and were sorted on a BD FACSAria^{TM} to isolate DN1 progenitors (Lin^-CD45+c-$Kit^{hi}CD44^{hi}CD25^-$), DN2 progenitors (Lin^-CD45+c-$Kit^{hi}CD44^{hi}CD25^+$) or DN3 postcommitment cells (Lin^- $CD45^+$ c-Kit^{lo} $CD44^{lo}$ $CD25^+$). In principle, DN2 progenitors can be further subdivided according to their levels of c-Kit expression (Kit^{++} = DN2a, Kit^+ = DN2b), which distinguish pre- and post-commitment pro-T cells (Yui et al. 2010). Note that the cells referred to as “DN1” throughout this report, for simplicity, are in fact defined by the criteria for “ETP” (*bona fide* early T-cell precursors) (Allman et al. 2003).

Scid.adh.2C2 cells (Dionne et al. 2005) were cultured in RPMI1640 with 10% fetal bovine serum (Sigma-Aldrich), sodium pyruvate, non-essential amino acids, Pen-Strep-Glutamine and 50 μM β-mercaptoethanol. Cells were incubated at 5% CO_2 and 37°C.

Cell staining, flow cytometry and sorting methods

Antibodies used for these analyses were all standard, commercially available monoclonal reagents with widely established use to characterize immune cell populations in the mouse (see Reagent and Resource Tables, Antibodies). Prior to cell surface staining, Fcγ receptors were first blocked with 2.4G2 hybridoma cell supernatant.

Most of the bicistronic retroviral vectors used in this study (see below) also encoded GFP or mTurquoise (enhanced CFP). For sorting of transduced, CD25-enriched FLP derived cells, surface antibodies against c-Kit, CD25, CD44, CD45, and CD11b were used for staining of Lin^- FLP-derived $CD25^+$ cells. Fractions sorted were $7AAD^-$ $CD45^+$ GFP^+ $CD25^+$ ("$CD25^+$") for samples transduced with EVGFP, PU1WTHA, PU1ENGHA and PU1ETSHA or $7AAD^-$ $CD45^+$ GFP^+ $CD44^+$ $CD25^-$ ("$CD44^+$") (generated by PU1WTHA only). On the day of transduction, the purity of the CD25 enriched cells was confirmed with surface staining of the above described markers and analyzed using a bench top flow cytometer (MacsQuant 10, Miltenyi).

For ChIP experiments with Scid.adh.2C2 cells transduced with PU1WTHA, PU1ENGHA, PU1ETSHA, or PU.1-ERT2, the bicistronic vectors also encoded GFP, and transduced cells were sorted as 7AAD-GFP+. For measuring the effect of *Runx1* deletion on the ability of exogenous PU.1 to bind, the Myc-Flag-PU.1 construct in the pMXs vector with a tailless human NGFR (hNGFR) reporter was used instead, as described in detail (Hosokawa et al. 2018), and detected with PE-conjugated anti-hNGFR antibodies.

PU.1 and HA-tag intracellular staining using the BD cytofix/cytoperm kit (Becton Dickinson Immunocytometry Systems) was carried out with PU.1 (9G7) rabbit mAb-AlexaFluor 647 and

HA-tag (6E2) mouse mAb-AlexaFluor 647 and surface staining against c-Kit, CD25, CD44 and CD45 for FLP $CD25^+$ cells. For intracellular Scid.adh.2C2 staining experiments, surface staining was carried out against CD25 and CD11b. All intracellular staining experiments were analyzed using a bench top flow cytometer (MacsQuant 10, Miltenyi). For intracellular staining, ZombieAqua™ Viability Dye (BioLegend) was used for durable live-dead discrimination prior to fixation.

Cells were sorted on a BD FACSAria™ (Becton Dickinson), iCyt Mission Technology Reflection Cell Sorter (Sony Biotechnology Inc.), or Sony Synergy 3000 at the California Institute of Technology Flow Cytometry Facility. 7AAD was used as the viability marker except where otherwise noted.

Retroviral transduction

Retroviral transductions were carried out similarly to reported procedures (Del Real and Rothenberg 2013; Champhekar et al. 2015; Hosokawa et al. 2018). In general, retroviral vectors were packaged by transfecting Phoenix-Eco cells with retroviral constructs using Fugene 6 Transfection Reagent (Roche). For generation of Lzrs vector supernatants, stable cell lines were established under puromycin selection (1 μg/ml). After approximately 2 weeks, puromycin was removed and viral supernatant was collected at days 2 and 4. All other viral supernatants were made by transient transfection of Phoenix-Eco cells and viral supernatants collected on 2 and 3 d after transfection. For transductions, non-tissue culture plates were coated first with RetroNectin (40 μg/ml). After excess RetroNectin removal, wells were blocked with 1×PBS/2%BSA, washed once with 1×PBS, and viral supernatant was added to the wells. Plates were spun 2000×g, 32°C

for 2h followed by an additional 1×PBS wash. Then when cells were added for transduction, fresh viral supernatant was added to the plate, and plates with cells were spun at 300×g, 32°C for 20 min before being returned to incubator for the remainder of the transduction time.

To generate transduced cells for ChIP and RNA-collection, 10 million $CD25^+$ FLP-derived cells per well (6-well plate) were infected with 2 ml retroviral supernatant adsorbed to RetroNectin, in medium supplemented with IL-7 and FLT3L (5 ng/ml each), for 6 h. Cells were then returned to OP9-DL1 monolayers and cultured for an additional 40h prior to cell sorting. Because manipulations of PU.1 activity directly affect both CD44 and KIT (Champhekar et al. 2015), the two other markers that distinguish stages from DN2a to DN3, the transduced $CD25^+$ cells were not subdivided on the basis of these markers. However, both flow cytometric analysis and gene expression patterns shown in Supplementary Figs. S4B-D and Table S3 confirm that these cells were primarily late DN2b-DN3 cells at the time of transduction.

Scid.adh.2C2 cells were transduced with retroviral vectors essentially as previously described (Del Real and Rothenberg 2013; Champhekar et al. 2015; Hosokawa et al. 2018), except that the cells were typically incubated with virus for 6-18 h.

Cloning and constructs

A full description of the HA-tagged PU.1 LZRS constructs used in this study (PU1WTHA, PU1ENGHA, PU1ETSHA) is given in (Champhekar et al. 2015); structures are presented schematically in Supplementary Fig. S6A. PU.1 with Myc and Flag epitope tags in the pMXs vector with an hNGFR reporter was described in (Hosokawa et al. 2018).

For PU.1-ERT2, ERT2 was inserted into LZRS-IRES-GFP to make EV-ERT2, using a forward-primer containing a start codon and an EcoRI site, and a reverse primer containing a stop codon and a NotI site. Full length PU.1 was then PCR-amplified from LZRS-PU.1, using a forward-primer that again contains an EcoRI-site and a reverse primer that introduces an XhoI site and deletes the PU.1 stop codon, enabling PU.1 to be fused to ERT2 in-frame by cloning into EV-ERT2. For primer sequences, see Reagent and Resource Table (Oligonucleotides).

For sgRNA-expressing vectors (E42 dTet-CFP), we first made an empty sgRNA expression cassette with human U6 promoter and mTurquoise2 (brighter version of Cyan Fluorescent Protein with shorter EF-1α promoter) by modifying a pQCXIN backbone retroviral vector using Gibson cloning. 19-mer sgRNAs were designed using the CHOPCHOP web tool (https://chopchop.rc.fas.harvard.edu/) and inserted into the empty sgRNA-expression vector by PCR-based insertion. Three sgRNA-expression vectors were generated for *Spi1* (PU.1) or *Runx1*, and pooled retroviral plasmids were used to make retroviral supernatant. To make it possible to induce CAS9-mediated deletion of genes in cells without an intrinsic *Cas9* transgene, cDNA for *Cas9* was fused with sequences encoding a P2A cleavage site at the C-terminus of the green fluorescent protein, mNeonGreen, and this insert was cloned into a pQCXIN derivative with the short EF-1α promoter (pQCXIN-EF1a-mNeonGreen-P2A-Cas9, “Cas9-GFP” in the text).

CRISPR-CAS9 KO of *Spi1* in primary DN cells

These methods were described previously (Hosokawa et al. 2018). Briefly, bone marrow precursor cells from CAS9/Bcl2-transgenic mice were initiated into T-cell differentiation by

OP9-DL1 co-culture for 2d (ATAC-seq, ChIP-seq) or 4d (RNA-seq) using OP9 medium supplemented with 10 ng/ml of IL-7 and 10 ng/ml of FLT3L. Then, they were infected with a retroviral vector encoding sgRNA (Reagent and Resource Oligonucleotide table for protospacer sequences) against *Spi1*or control guide RNA. On day 2 (for ATAC-seq experiments) or day 4 (for RNA-seq experiments), cultured cells were disaggregated, filtered through 40-μm nylon mesh, transferred onto RetroNectin-coated, virus-bound plates, and cultured with OP9 medium supplemented with 10 ng/ml of IL-7, 10 ng/ml of FLT3L and 10 ng/ml of SCF for 6 h. Transduced cells were then returned to OP9-DL1 coculture and incubated 4 more days. Thus, transduced DN2 cells were sorted at d6 overall for ATAC-seq samples, or transduced $CD25^+$ (DN2-DN3) cells were sorted at d8 overall for RNA-seq samples. For collecting the samples, cells were stained with CD45, CD25, CD44, c-Kit and a biotin-conjugated lineage cocktail (CD8α, CD11b, CD11c, Gr1, TER-119, NK1.1, CD19, TCRβ, TCRγδ), and DN2 sgRNA-transduced cells were sorted (CFP^+ Lin^- $CD45^+$ CFP^+ c-Kit^{hi} $CD44^{hi}$ $CD25^+$) (ATAC-seq experiment) or $CD25^+$ (DN – DN3) infected cells (CFP^+ Lin^- $CD45^+$ $CD25^+$ CFP^+) (RNA-seq experiment). For a more extensive description, additional knockouts, and additional data analysis, see the related paper (Hosokawa et al. 2018).

Cas9-induced deletion of endogenous *Runx1* to determine effect on exogenous PU.1 binding in Scid.adh.2C2 cells

Methods for introducing *Cas9* as well as sgRNA into Scid.adh.2C2 cells were previously described (Hosokawa et al. 2018). Importantly, transduction efficiencies in these cells frequently exceed 90% in a single round; however, the large *Cas9* expression construct was less efficient. For CRISPR-CAS9-mediated deletion of *Runx1* in Scid.adh.2C2 cells, therefore, cells were first

transduced with the *Cas9-GFP* retroviral construct and sorted for GFP^+ cells two days later. These CAS9-expressing cells were then expanded for a week, and then sgRNA against *Runx1* was transduced. Two days after sgRNA introduction, the cells were transduced with Myc-Flag-PU.1-hNGFR, cultured two more days, and finally the triply infected cells (6×10^6) were harvested and subjected to ChIP-seq for PU.1 as described below.

Tamoxifen-dependent PU.1 mobilization time course

After transduction of Scid.adh.2C2 cells with Lzrs-PU.1-ERT2 (Scid.adh.2C2-PU.1-ERT2) or EV-ERT2 (Scid.adh.2C2-ERT2) control, cells were expanded for 48h and infected (GFP+) cells were sorted on a BD FACSAria™. The cells were further expanded *in vitro* before acute PU.1 mobilization was induced for 2, 8 or 24h with 0.1 μM 4-OHT (4-hydroxytamoxifen; Sigma-Aldrich).

Long-term culture (> 2 weeks) of uninduced Scid.adh.2C2-PU.1-ERT2 cells showed very limited effects on gene expression and only 93 genes were differentially expressed compared to long-term cultured control cells transduced with EV-ERT2 (EdgeR adj.p-value ≤ 0.05 and $|\log_2$ fold change$| \geq 1$). Of these 76 were up-regulated, suggesting that low levels of PU.1 are sufficient to activate a small subset of genes but insufficient to cause any major transcriptional or phenotypic changes. These effects, however, seem to be indirect since only 19 of the differentially expressed genes were associated with PU.1 binding at 0h.

mRNA-preparation and RNA-sequencing

Total RNA was isolated from 200,000-500,000 primary or Scid.adh.2C2 cells using an RNAeasy MicroKit (Qiagen) according to manufacturer's recommendations. Libraries were constructed using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB #E7530) from ~1 μg of total RNA following manufacturer's instructions. Libraries were sequenced on Illumina HiSeq2500 in single read mode with the read length of 50 nt following manufacturer's instructions. Base calls were performed with RTA 1.13.48.0 followed by conversion to FASTQ with bcl2fastq 1.8.4 and produced approximately 30 million reads per sample.

Chromatin Immunoprecipitation (ChIP) and ChIP-sequencing

Approximately 3-5 × 10^6 sorted primary $CD25^+$ cells or 10 × 10^6 Scid.adh.2C2 cells for ChIP were fixed for ChIP at RT by one of two protocols: (1) for 10 min of 1% FA only (for immune precipitations to detect PU.1, HA-tag, and histone H3-modifications), or (2) in 1 mg/ml DSG (Thermo Scientific) in PBS for 30 min followed by additional 10 min of incubation after addition of formaldehyde (FA) up to 1% (for ETS1 immune precipitations).. The reaction was quenched by addition of 1/10 volume of 1.25M glycine (c_f=0.125M) and the cells were washed in ice-cold 1xHBSS (Gibco). Cell pellets were snap frozen on dry ice and stored in -80°C. For cells crosslinked with FA only (protocol 1), cells were resuspended in 1% Lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl (pH 8). For DSG+FA crosslinked cells (protocol 2), nuclei were isolated by 10 min incubation in Nuclei Isolation buffer (50 mM Tris-pH 8.0, 60 mM KCl, 0.5% NP40) + protease inhibitor cocktail (PIC) (Roche) on ice. Pelleted nuclei were dissolved in 0.5% Lysis buffer [0.5% SDS, 10 mM EDTA, 0.5 mM EGTA, 50 mM Tris-HCl (pH 8)] + PIC. Lysates were then sonicated on a Bioruptor (Diagenode) for 18 cycles each of max power for 30s followed by 30s rest. Sonication was followed by pelleting of debris, the supernatant was

transferred to a new tube, and chromatin was diluted with 3 volumes of 1xHBSS+PIC, followed by diluting the resulting total with an equal volume of 2×RIPA buffer (2% Triton X-100, 2 mM EDTA, 200 mM NaCl, 20 mM Tris-HCl (pH 7.5), 1% sodium deoxycholate, 0.1% SDS + PIC). Approximately 10 μg per 10^7 cells of antibody was used for each reaction, as follows: Rabbit anti-PU.1 polyclonal IgG (sc-352x, Santa Cruz Biotechnology), Rabbit HA-probe polyclonal IgG (sc-805x, Santa Cruz Biotechnology), Rabbit anti-ETS1 polyclonal IgG (sc-350x, Santa Cruz Biotechnology), Rabbit polyclonal anti-H3K27Ac (ab4729, Abcam), Rabbit polyclonal anti-H3K27me3 (07-449, Millipore) or Rabbit polyclonal anti-H3K4me2 (07-030, Millipore).

Each antibody was first adsorbed to Dynabeads™ M-280 Sheep anti- Rabbit, Dynabeads™ M-280 Sheep anti-Mouse, or Dynabeads™ Protein A/G (Invitrogen), respectively in 1 ml 1×RIPA+PIC for 4h. Beads were then washed twice with 1×RIPA, resuspended in 100 μl 1×RIPA+PIC and added to the diluted chromatin. For polyclonal and monoclonal antibodies, 20 and 50 μl of Dynabeads™ were used per μg antibody, respectively. ChIP was performed by incubation overnight at 4°C, and the beads were subsequently washed, 1 time with 1 ml Low Salt Immune Complex Wash Buffer (0.1% SDS, 1 % Triton X-100, 2 mM EDTA, 50 mM Tris-HCl pH8, 150 mM NaCl), 1 time with 1 ml High Salt Immune Complex Wash Buffer (0.1% SDS, 1 % Triton X-100, 2 mM EDTA, 50 mM Tris-HCl pH8, 500 mM NaCl), 1 time with 1 ml LiCl Immune Complex Wash Buffer (0.25 M LiCl, 1% Igepal-CA630, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH8), and 2 times with 1 ml TE buffer (10 mM Tris–HCl, pH 8.0, 10 mM EDTA). Then, DNA was eluted from the beads for 6 h to O/N at 65°C in ChIP elution buffer (20 mM Tris-HCl, pH 7.5, 5 mM EDTA 50 mM NaCl, 1% SDS, and 50 μg proteinase K) treated and finally cleaned up using Zymo ChIP DNA Clean & Concentrator.

ChIP-seq libraries were constructed using NEBNext ChIP-Seq Library Preparation Kit (NEB #E6240) following manufacturer's instructions. Libraries were sequenced on Illumina HiSeq2500 in single read mode with the read length of 50 nt following manufacturer's instructions. Base calls were performed with RTA 1.13.48.0 followed by conversion to FASTQ with bcl2fastq 1.8.4 and produced approximately 30 million reads per sample.

Assay for Transposase Accessible Chromatin (ATAC-seq)

Eighty thousand Scid.adh.2C2, ~18 × 10^3 DN1 or 50 × 10^3 DN3 cells freshly isolated from the thymus or ~16-30 × 10^3 DN2 cells derived from *Cas9;Bcl2*-transgenic bone marrow precursor cells as described above were washed in ice cold PBS (Scid.adh.2C2) or HBSS-Hepes (primary cells) prior to Assay for Transposase Accessible Chromatin (ATAC-seq) library preparation as described in (Buenrostro et al. 2013). Briefly, rinsed cells were lysed using fresh ATAC lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.1% IGEPAL CA-630, 3 mM $MgCl_2$), the lysed nuclei were immediately processed by tagmentation reaction mix including Nextera Tn5 transposase (Illumina) and incubated for 30 min at 37°C, then cleaned up using a MinElute Kit (Qiagen). The library was constructed and barcoded using a Nextera library preparation kit (Illumina). For the primary cells, the final amplified libraries were purified and size-selected (at a bead-to-DNA solution ratio of 1.2 (v/v) for ~100-150bp cutoff DNA size) using SPRIselect-beads (Beckman-Coulter). Libraries were single-end sequenced on a HiSeq2500 (Illumina) (primary cells) or a NextSeq500 (Scid.adh.2C2) and produced approximately 30-50 million reads per sample. Scid.adh.2C2, DN1 and DN3 cells sample were processed and sequenced in

duplicates while the BM-derived DN2-cells with PU.1 CRISPR-CAS9 perturbation (or their corresponding controls) were processed and sequenced in triplicates.

Data processing

ChIP-seq

DN1 to DN3 PU.1 and H3K4me2 ChIP-seq data was downloaded from the GEO database (GSE31235) (Zhang et al. 2012) and mapped against the mm9 reference genome. DN1, DN2a and DN2b PU.1 peaks were derived with the HOMER platform (Heinz et al. 2010).
ChIP sequences were mapped to the mouse genome build NCBI37/mm9 using Bowtie (v1.1.1; http://bowtie-bio.sourceforge.net/index.shtml)(Langmead et al. 2009) with "-v 3 -k 11 -m 10 -t --best –strata" settings and HOMER tag directories were created with *makeTagDirectory* (Heinz et al. 2010) for further downstream analysis. In addition, read alignments from the Scid.adh.2C2-PU.1-ERT2 experiment were filtered against PCR duplicates using *Samtools* (Li et al. 2009) as well as genomic repeats and Broad blacklisted regions (Consortium 2012) using *Bedtools* (Quinlan and Hall 2010) (full script provided on request) prior to tag directory generation. Transcription factor peaks were identified with *findPeaks.pl* against a matched control sample using the settings "*-P .1 -LP .1 -poisson .1 -style factor*". Transcription factor peak reproducibility was determined by a HOMER adaptation of the IDR (Irreproducibility Discovery Rate) package (Li 2011; Karmel 2015) according to ENCODE guidelines (https://sites.google.com/site/anshulkundaje/projects/idr). Only reproducible high quality peaks, defined by normalized scores of at least 15 tags/10 million and an acceptable IDR score, were submitted to further analysis.

Motif enrichment analysis

Motif enrichment analysis was performed with the *findMotifsGenome.pl* command in the HOMER package using a 200bp window or a window size defined by the width of each identified peak (Scid.adh.2C2 PU.1-ERT2 4-OHT experiment). Overlapping peaks between samples were derived using *mergePeaks.pl* (default parameters). Annotation of peaks to genes and genomic regions *(e.g.* promoters, CG-rich regions, repeat regions) were performed with *annotatePeaks.pl* (*default settings).*

Site quality scoring

Position-weight matrix (PWM) log odds score analysis was performed with *annotatePeaks.pl* with the options *–m {motif file} -mscore* to derive the highest similarity score for each individual peak. Briefly, the top 12-mer motif was derived from a DN1-DN2b combined PU.1 peak list [reanalyzed from (Zhang et al. 2012), normalized peak score ≥ 15] using *findMotifsGenome.pl* with options *–size 200 –len 12*. The top identified ETS-family motif PWM (for PWM see Table S1) or alternatively a Macrophage derived PU.1 PWM (Pham et al. 2013) (Table S1) were provided as the motif files for the analysis. The highest reported log-odds similarity scores within each peak were imported into MATLAB (R2016a) and visualized as Violin distribution Plots (https://www.mathworks.com/matlabcentral/fileexchange/23661-violin-plots-for-plotting-multiple-distributions--distributionplot-m-/content/distributionPlot/distributionPlot.m) displaying 25,50 and 75 percentiles. Statistical significances of differences between groups were tested in GraphPad Prism 6 with a Kruskal-Wallis non-parametric test (Dunn's correction for multiple comparisons). Adjusted P-values ≤ 0.05 were considered as statistically significant. HOMER *scanMotifGenomeWide.pl* was used to predict motifs and derive log-odds score genome wide.

Previously published data and quality control comparisons

Publicly available data used in this study with accession numbers GSE31235 (Zhang et al. 2012), GSE65344 (Champhekar et al. 2015), and GSE63732 (Vanhille et al. 2015) were downloaded as raw sequence data (http://www.ncbi.nlm.nih.gov/geo) and remapped onto NCBI37/mm9 using the same settings as described above. For the samples described here and in our related work (Hosokawa et al. 2018), deposited under accession numbers GSE93755 and GSE110020, multiple independent biological repeats of each sample series were analyzed (n given for each experiment). Note that the sensitivity of detection from our current ChIP-seq conditions was higher than in our previous analysis of PU.1 binding in pro-T cells (Zhang et al. 2012), and we could therefore apply more stringent statistical criteria for peak calling. Peaks with normalized scores below 15 (tags/10M) were excluded, and only reproducible high quality peaks as defined by Irreproducibility Discovery Rate (IDR) analysis (Li 2011; Karmel 2015), were considered in further analysis, as already noted. In samples transduced with PU.1WTHA, the efficiency of αHA and αPU.1 could be compared on the same targets. Approximately 1.5× as many peaks were called with the αHA antibody as with the αPU.1, but this appeared to be related only to peak height above threshold. Peaks called with αPU.1 were 98.9% overlapping with peaks called with αHA.

ATAC-seq analysis

Sequences were mapped to mm9 as described in the Data processing-ChIP section. Additionally, Tag directories in which reads mapped to the mitochondrial chromosome were filtered out were created using the HOMER platform (Heinz et al. 2010) (*makeTagDirectory)*. ATAC-seq peaks

were identified using *findPeaks.pl* (Heinz et al. 2010) in HOMER (full parameter list available on request). Only regions consistently called in two independent experiments were defined as 'open' or 'accessible'. TF binding identified in these regions was scored as binding in 'open' chromatin whereas binding outside these regions was identified as binding in 'closed' chromatin. Abundance of ATAC-tags on ATAC or transcription factor ChIP-seq peaks was analyzed in HOMER using the *annotatePeaks.pl* command.

For analysis of dynamic chromatin changes around PU.1 bound sites, DN1 and DN3 ATAC read counts were derived from the tag directories in regions of 1000 bp surrounding the PU.1 occupancy sites defined by ChIP in DN1-DN2b, detecting endogenous PU.1 with αPU.1 (Supplementary Fig. S1A), or in transduced DN2b cells detecting exogenous PU1WTHA (Supplementary Fig. S5D). Regions were defined using HOMER (Heinz et al. 2010) (*annotatePeaks.pl –noadj –size 1000*) and were tested for differential ATAC peak size between the DN1 and DN3 stages, using the EdgeR (Robinson et al. 2010) and limma (v3.20.9; http://bioconductor.org/packages/release/bioc/html/limma.html) (Ritchie et al. 2015) Bioconductor packages via HOMER *getDifferentialExpression.pl.* Differential peaks with adjusted p-value ≤ 0.05 and $|\log_2$ fold change$| \geq 1$ were considered significantly different.

For ATAC-tag counting in 'open' and 'closed' regions (see definition above) a 2000 bp region was defined around each DN1-DN2b PU.1 peak and tags were counted in 50 bp windows. The row sum for each sample was then histogram plotted with the R package ggplot2 with bin size = 5 (Fig 1D). Thymocyte- and bone marrow derived DN2b ATAC overlapping peaks (Fig 2E) were derived using HOMER *mergePeaks.pl* (default parameters) (Heinz et al. 2010).

RNA-seq

RNA-sequenced reads were trimmed with Trimmomatic (v.033; http://www.usadellab.org/cms/?page=trimmomatic) (Bolger et al. 2014) for removal of adapter – and low quality sequences (settings: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). Resulting reads were then mapped onto the mouse genome build NCBI37/mm9 with STAR (v2.4.0) (Dobin et al. 2013) and post-processed with RSEM (v1.2.25; http://deweylab.github.io/RSEM/) (Li and Dewey 2011) according to the settings in the ENCODE long-rna-seq-pipeline (https://github.com/ENCODE-DCC/long-rna-seq-pipeline/blob/master/DAC/STAR_RSEM.sh) with the minor modifications that settings "*--output-genome-bam --sampling-for-bam*" was added to *rsem-calculate-expression*. STAR and RSEM reference libraries were created from genome build NCBI37/mm9 together with the Ensembl gene model file Mus_musculus.NCBIM37.66.gtf. The resulting RSEM-filtered bam-files were used to create HOMER (Heinz et al. 2010) tag directories (*makeTagDirectory* with –keepAll setting). For analysis of statistical significance among differentially expressed genes the raw gene counts were derived from each tag directory with *analyzeRepeats.pl* with the *–noadj -condenseGenes* options followed by the *getDiffExpression.pl* command using EdgeR (v3.6.8; http://bioconductor.org/packages/release/bioc/html/edgeR.html)(Robinson et al. 2010), or the gene count tables were imported into R and analyzed for differential expression using the DESeq2 package (v1.4.5; https://www.bioconductor.org/packages/devel/bioc/html/DESeq2.html) (Love et al. 2014). The same RSEM-filtered files were used to generate bigwig files for visualization on the UCSC genome browser as described below.

Statistical criteria for scoring differentially expressed genes (DEGs)

Differentially expressed gene (DEG) inclusion criteria were more or less inclusive depending on whether the purpose of the scoring was to identify definite regulation targets or simply to enrich for genes with some level of response for correlation with another property. For example, for analyses of in vitro-differentiated primary cells to compare genes subcategorized by TF binding status, less stringent DEG criteria were used to avoid excluding genes due to nondevelopmental differences among experiments in the co-culture system. The criteria selected for different experimental comparisons were as follows. In Figs. 2C, 4D,E, S1A, S4H, S5C, S6B,D, S6I : DESeq2, p.adj ≤ 0.1. In Fig. 3G and S3A: EdgeR (p.adj ≤ 0.05, $|\log2FC| \geq 1$). For the manually calculated q-value in Supplementary Fig. S4I, the Bioconductor qvalue package was applied on the DESeq2 derived p-values for transcriptional regulatory gene expression differences between PU1WTHA25 and EV control samples, and values of $q<0.05$ were taken as significant.
Note that the position of *Spi1 (Sfpi1)* in the gene lists derived from the PU1-construct overexpression experiments as an “upregulated gene” is a result of the overexpressed sequences from the exogenous PU1WTHA, PU.1ENGHA or PU.1ETSHA.

For gene expression data visualization in clustered heat maps, rpkm normalized reads were derived using the *analyzeRepeats.pl* command with the options *–count exons –condenseGenes –rpkm* followed by log transformation. Log transformed datasets were then subjected to geometric normalization against control (Supplementary Fig. S4J) or arithmetic mean row normalization (Supplementary Fig. S3B). The resulting normalized datasets were hierarchically clustered in MatLab (Supplementary Fig. S4J) or R (Supplementary Fig. S3B). In most heatmaps shown in

this study, clustering was done with “average” linkage. In Supplementary Fig. S6G, clustering used “complete” linkage. For Fig 2A, the simplicity of the pattern made it possible to use k-means clustering (k=2). Results were visualized with the MatLab (clustergram) or Pretty Heatmaps (pheatmaps) R package (https://cran.r-project.org/web/packages/pheatmap/index.html). Site Groups were aggregated manually from the unsupervised clustergrams as shown, except in Fig 2A, where Group 1 and Group 2 were defined from the two k-means clusters directly.

For the correlation plot in Supplementary Fig. S4D (R corrplot package), Pearson correlation was calculated based on 14,054 genes with RPKM ≥ 1 in at least one of the samples. Log-transformation and row mean normalizations were performed with Cluster3 (de Hoon et al. 2004) among samples within the same experiment first before combining with results from other independent replicate sets to prevent non-developmental differences in gene expression. To index the developmental state with principal component analysis (Supplementary Fig. S4C), the 14,054 genes with a PRKM ≥ 1 in any of the included samples (same genes as included in Supplementary Fig. S4D correlation analysis) were normalized as described above. The normalized signals from the samples were used to calculate the principal components using the prcomp function from the R stats package. The loadings for the first two principal components were plotted using the R package ggplot2.

Linking DNA-binding and gene expression analysis

Enriched peaks (TF and ATAC accessible regions) were annotated to closest gene/transcriptional start site with proximity based annotation using HOMER *annotatePeaks.pl* (Heinz et al. 2010)

(mm9 genome build). Entrez gene ID was then used to match gene expression tables (rpkm or DEG) with peak lists. Each gene category was assigned a unique value depending on the combination of features annotated to the genes. Empirical cumulative distribution (ECDF) plots were generated with the R package ggplot2 and statistical analysis was performed with Kolmogorov-Smirnov two-sided tests.

Tag distribution heatmap visualization

Tag density plots and heat maps were created with *annotatePeaks.pl* (*–hist* or *–hist* & *-ghist* respectively) in a 2000 bp region surrounding indicated TF peak center, normalizing data to 10 million mapped reads per experiment, and by hierarchical clustering the tag count profiles in Cluster3 (de Hoon et al. 2004) with average linkage (if not stated) followed by EaSeq (Lerdrup et al. 2016) visualization. Prior to EaSeq visualization, bed files were created from HOMER tag directories using *tagDir2bed.pl* with default settings and then imported into EaSeq. If not otherwise stated, each analysis was performed in duplicates.

UCSC Genome Browser bigwig visualization

BigWigs were generated from the aligned SAM or BED-file formats using *Samtools* (Li et al. 2009), *Bedtools* (Quinlan and Hall 2010) and the *UCSC genomeGoverageBed* and *bedGraphToBigWig* and normalized to 1 million reads. For visualization of RNA-seq tracks, *bamToBed* and *genomeCoverageBed* were used with the "-split" setting enabled. BiGWig files were up-loaded to the UCSC Genome Browser (http://genome.ucsc.edu) (Speir et al. 2016) for visualization.

LITERATURE CITED

Allman D, Sambandam A, Kim S, Miller JP, Pagan A, Well D, Meraz A, Bhandoola A. 2003. Thymopoiesis independent of common lymphoid progenitors. *Nat Immunol* **4**: 168-174.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213-1218.

Champhekar A, Damle SS, Freedman G, Carotta S, Nutt SL, Rothenberg EV. 2015. Regulation of early T-lineage gene expression and developmental progression by the progenitor cell transcription factor PU.1. *Genes Dev* **29**: 832-848.

Consortium Encode Project. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

de Hoon MJ, Imoto S, Nolan J, Miyano S. 2004. Open source clustering software. *Bioinformatics* **20**: 1453-1454.

Del Real MM, Rothenberg EV. 2013. Architecture of a lymphomyeloid developmental switch controlled by PU.1, Notch and Gata3. *Development* **140**: 1207-1219.

Dionne CJ, Tse KY, Weiss AH, Franco CB, Wiest DL, Anderson MK, Rothenberg EV. 2005. Subversion of T lineage commitment by PU.1 in a clonal cell line system. *Dev Biol* **280**: 448-466.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.

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

Hosokawa H, Ungerbäck J, Wang X, Matsumoto M, Nakayama KI, Cohen SM, Tanaka T, Rothenberg EV. 2018. Transcription factor PU.1 represses and activates gene expression in early T cells by redirecting partner transcription factor binding. *Immunity* **48**: 1119-1134.

Karmel A. 2015. homer-idr: Second pass updated [Data set].

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K. 2016. An interactive environment for agile analysis and visualization of ChIP-sequencing data. *Nat Struct Mol Biol* **23**: 349-357.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Li QB, James B.; Huang, Haiyan; Bickel, Peter J. 2011. Measuring reproducibility of high-throughput experiments. *Ann Appl Stat 5* **5**: 1752-1779.

Longabaugh WJR, Zeng W, Zhang JA, Hosokawa H, Jansen CS, Li L, Romero-Wolf M, Liu P, Kueh HY, Mortazavi A et al. 2017. Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. *Proc Natl Acad Sci U S A* **114**: 5800-5807.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. 2014. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucl Acids Res* **42**: W401-407.

Pham TH, Minderjahn J, Schmidl C, Hoffmeister H, Schmidhofer S, Chen W, Langst G, Benner C, Rehli M. 2013. Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1. *Nucl Acids Res* **41**: 6391-6402.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucl Acids Res* **43**: e47.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Saldanha AJ. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**: 3246-3248.

Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, Lee BT, Learned K, Karolchik D, Hinrichs AS et al. 2016. The UCSC Genome Browser database: 2016 update. *Nucl Acids Res* **44**: D717-725.

Vanhille L, Griffon A, Maqbool MA, Zacarias-Cabeza J, Dao LT, Fernandez N, Ballester B, Andrau JC, Spicuglia S. 2015. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun* **6**: 6905.

Yui MA, Feng N, Rothenberg EV. 2010. Fine-scale staging of T cell lineage commitment in adult mouse thymus. *J Immunol* **185**: 284-293.

Zhang JA, Mortazavi A, Williams BA, Wold BJ, Rothenberg EV. 2012. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* **149**: 467-482.

LIST OF SUPPLEMENTAL TABLES

Legends are also included as worksheets in Supplemental Tables S1-S6 themselves.

Table S1: PWM scoring of PU.1 occupancy sites

DN1-2b PWM: PWM and scores of individual motifs derived from occupancy sites in DN1, DN2a, and DN2b cells; Mac PWM: PWM and scores of individual motifs derived from occupancy sites in macrophages.

Table S2: Kinetic responses to PU.1-ERT2 mobilization in stably transduced Scid.adh.2C2 cells: Time course following 4OHT

RNA expression values at 0, 2h, 8h, and 24 h are given for genes defined as DEGs by 24 h post 4OHT treatment of PU.1-ERT2 expressing cells. Average RPKM (two independent experiments), fold change (FC), p values, and p.adj values are given, with statistical comparisons between the indicated pairs of samples. Columns K—M: PU.1-ERT2 0h vs. 2h 4OHT. Columns N—P: PU.1-ERT2 0h vs. 8h 4OHT. Columns Q—S: PU.1-ERT2 0h vs. 24h 4OHT.Columns T—V: PU.1-ERT2 0h vs. EV-ERT2 (“ERT2”) 0h. Columns W—Y: PU.1-ERT2 0h vs. EV-ERT2 24h. Columns Z—AE: RPKM values for indicated sample types (mean of two).

Table S3: Gene expression changes in primary pro-T cells in response to exogenous PU.1 constructs: analysis relative to endogenous expression and HA-PU.1 binding

RNA expression values (RPKM), fold changes from empty vector controls, p values, and p.adj values for genes defined as DEGs after transduction with PU1WTHA in cells remaining $CD25^+$ (PU1WTHA25) or in cells becoming $CD25^-$ $CD44^+$ (PU1WTHA44), or in $CD25^+$ cells after

transduction with PU1ENGHA or with PU1ETSHA. Each worksheet presents results from all samples for genes defined as differentially expressed in the named sample. In each worksheet: Columns D—O: Statistical features of effects; Columns P—T: RPKM (mean of four independent experiments); Column U: presence or absence of annotated PU.1 peak(s).

Table S4: Developmental index genes: regulation in normal pro-T development by PU.1

A curated list of 171 developmentally dynamic regulatory genes, defined as described in (Champhekar et al. 2015) from data in GSE31235 (Zhang et al. 2012), were monitored for changes in expression in PU1WTHA25, PU1WTHA44, PU1ENGHA, and PU1ETSHA samples. Columns D-N show RPKM values from GSE31235; columns O-AH show data from this study (GSE93755); and columns AI-AN show data from a previously reported shorter-term analysis of effects of PU1ENGHA and PU1ETSHA, from GSE65344 (Champhekar et al. 2015). RPKM, pvalues and p.adj values relative to EV-transduced controls in this study are shown, as well as q values for differential expression in response to PU1WTHA among genes within this list.

Table S5: Genes differentially expressed in primary DN2 cells in response to acute PU.1 knockout: full list with RNA-seq data

Full RNA expression data are shown for two independent experiments in which *Spi1* was disrupted by CAS9-mediated deletion. RPKM, fold changes, p values, and p.adj values are shown for the full transcriptome.

Table S6: High confidence PU.1-regulated genes in pro-T cells. Genes with concordant responses to loss and gain of PU.1 in primary pro-T cells. (A) List defined using padj<0.05 on

both, and indicating linked PU.1 binding Status; and (B) expanded list showing overlap of genes sensitive to PU.1 deletion with genes differentially expressed in PU1WTHA44

Table S7: Motifs enriched at functionally distinct classes of sites bound by exogenously introduced PU.1: effects of chromatin states and chromatin developmental dynamics

A. Motif analysis of Group 5,6 and 7 peaks defined in Fig. 4C. The top five motifs defined by Homer De Novo motif search are shown for each site category. **B.** Motif analysis of exogenous PU1WTHA-occupied non-promoter (distal) peaks in $CD25^{+}$ cells, as detected by αHA ChIP. Sites with quality scores above or below PWM score median are compared for co-enriched motifs, depending on whether they lie in normally inaccessible (PWM median score=9.74) or accessible (PWM median score=8.97) regions in DN3 cells. Note strongly similar pattern of motif preference for endogenous PU.1 at its natural sites (Fig. 1G, Supplementary Fig. S1D). The top five motifs defined by Homer De Novo motif search are shown for each site category. C. Top five motifs enriched at PU1WTHA αHA peaks in regions that change dynamically from DN1 to DN3.

SUPPLEMENTAL TABLE S7

TOP MOTIFS AT SITES OCCUPIED BY EXOGENOUS PU.1 IN PRO-T CELLS

A

Fig 4C group 5

Motif	Best match	% of Targets	% of Bg	P-value
	*SpiB(ETS)*	80.9	7.9	1e-20016
	*Cebpa*	2.4	0.01	1e-1476
	*Klf5*	2.6	0.06	1e-955
	*Runx2*	26.0	11.9	1e-942
	*Zic*	2.9	0.12	1e-844

Fig 4C group 6

Motif	Best match	% of Targets	% of Bg	P-value
	*SpiB(ETS)*	63.2	3.1	1e-8735
	*Nr2f2*	9.6	0.3	1e-1381
	*Dlx2*	8.7	0.2	1e-1365
	*SD0003.1 at AC acceptor*	9.2	0.3	1e-1348
	*Mecom*	10.3	0.4	1e-1342

Fig 4C group 7

Motif	Best match	% of Targets	% of Bg	P-value
	*PU.1(ETS)*	65.1	16.3	1e-6469
	*Runx1*	12.5	4.8	1e-498
	*Sp1*	15.2	9.0	1e-217
	*SPI1*	1.3	0.1	1e-213
	*BORIS*	1.1	0.3	1e-89

B

PU1WTHA closed above PWM median score

Motif	Best match	% of Targets	% of Bg	P-value
	*SpiB(ETS)*	86.5	1.7	1e-24217
	*Gcrf2*	9.6	0.2	1e-2204
	*Sox4*	8.2	0.2	1e-1772
	*Gsx6*	7.3	0.2	1e-1592
	*Oct*	6.3	0.1	1e-1548

PU1WTHA closed below PWM median score

Motif	Best match	% of Targets	% of Bg	P-value
	*Sfpi1*	84.0	15.3	1e-9022
	*Runx2*	36.5	20.8	1e-506
	*IRC900814*	13.3	5.5	1e-335
	*Cphx*	1.0	0.11	1e-111
	*Foxo1*	0.4	0.01	1e-91

PU1WTHA open above PWM median score

Motif	Best match	% of Targets	% of Bg	P-value
	*PU.1(ETS)*	42.6	3.0	1e-2123
	*Runx1*	17.7	6.8	1e-169
	*Ptf1a*	23.8	15.1	1e-67
	*Runx2*	22.8	15.1	1e-53
	*Ctcf*	1.8	0.3	1e-48

PU1WTHA open below PWM median score

Motif	Best match	% of Targets	% of Bg	P-value
	*Sfpi1*	52.7	22.7	1e-538
	*Runx1*	31.1	12.2	1e-315
	*Ctcf*	2.5	0.4	1e-69
	*E2A*	23.1	15.9	1e-45
	*Prdm1*	0.4	0.01	1e-39

C

PU1WTHA in closing regions

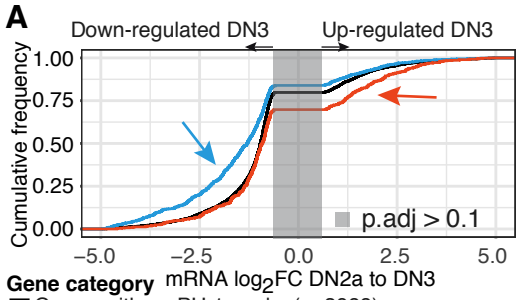
Motif	Best match	% of Targets	% of Bg	P-value
	*SpiB(ETS)*	64.6	2.8	1e-4969
	*Runx1*	26.5	8.9	1e-394
	*PU.1-IRF*	12.8	2.9	1e-291
	*Cebpa*	0.7	0.01	1e-67
	*PU.1(ETS)*	2.4	0.5	1e-59

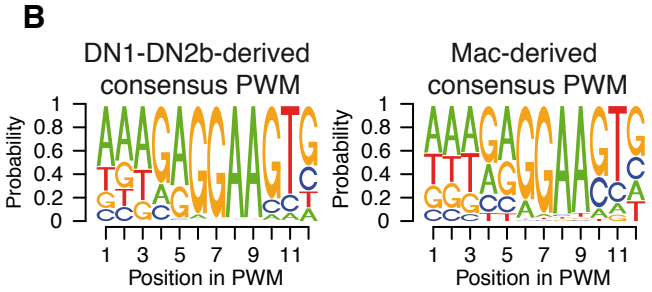
PU1WTHA in opening regions

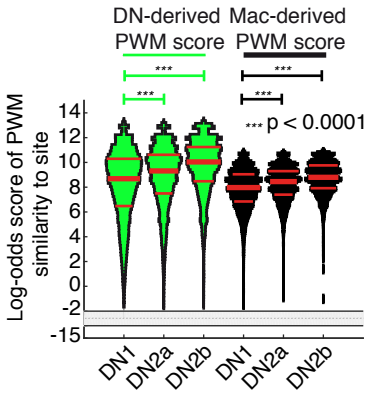
Motif	Best match	% of Targets	% of Bg	P-value
	PU.1(ETS)	64.3	5.9	1e-2208
	Runx2	20.7	9.2	1e-112
	Tcf3	30.1	17.0	1e-98
	*Tcf3(HMG)**	13.0	5.3	1e-81
	GATA	0.6	0.01	1e-32

**Tcf3(HMG) = Tcf7l1*

Figure S1







D

DN1-DN2a open above PWM median score

Motif	Best match	% of Targets	% of Bg	P-value
	SpiB(ETS)	55.9	3.7	1e-5534
	Runx2	32.8	16.4	1e-384
	PU.1:IRF8	9.0	4.5	1e-89
	Fra1	22.8	17.0	1e-54
	EWS:FLI1	4.3	1.9	1e-51
	Rel	10.5	6.7	1e-51
	Myf5	20.5	15.6	1e-41

DN1-DN2a closed above PWM median score

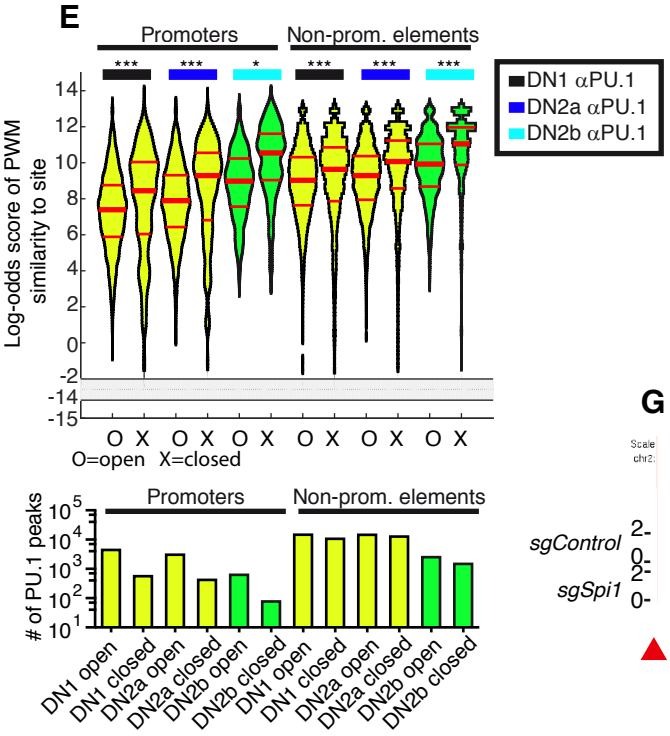
Motif	Best match	% of Targets	% of Bg	P-value
	SpiB(ETS)	86.8	2.4	1e-10356
	Foxh1	7.7	0.03	1e-1256
	Zic	8.5	0.08	1e-1142
	Hic1	7.9	0.11	1e-954
	NFIC	7.2	0.08	1e-938
	Tfap2c	6.1	0.03	1e-914
	Sox21	6.3	0.08	1e-772

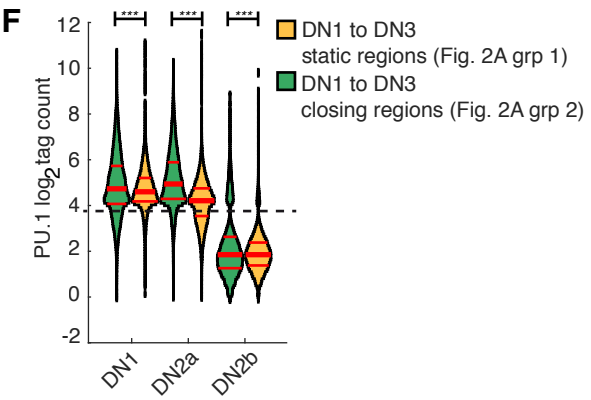
DN1-DN2a open below PWM median score

Motif	Best match	% of Targets	% of Bg	P-value
	Fli1(ETS)	59.0	28.4	1e-949
	Runx2	16.0	4.3	1e-475
	SPI1	4.6	0.9	1e-188
	Pol003 GC-box	21.1	12.4	1e-137
	BHLHE4	35.2	25.4	1e-112
	Trp(Myb)	29.2	21.0	1e-91
	CTCF	2.8	0.8	1e-74

DN1-DN2a closed below PWM median score

Motif	Best match	% of Targets	% of Bg	P-value
	Sfpi1	70.2	16.5	1e-2569
	Meis1	1.3	0.01	1e-166
	Nkx3	2.5	0.2	1e-166
	Mzf1	1.1	0.01	1e-141
	Jund	1.0	0.01	1e-129
	Sox11	0.9	0.01	1e-127
	Runx1	25.0	15.0	1e-124





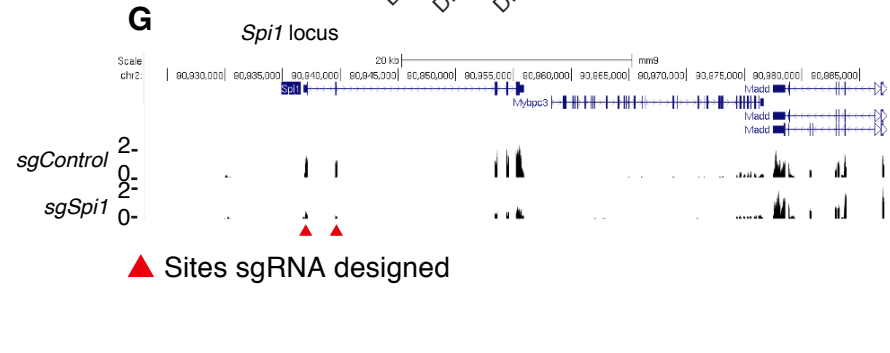


Figure S1: Quantitative aspects of endogenous PU.1 binding to open and closed genomic sites in early pro-T cells.

A) Correlation between developmental ATAC accessibility changes and developmental changes in local gene expression. Empirical cumulative distribution frequency plot showing association of PU.1 bound sites that change ATAC-accessibility between DN1 and DN3 stages of pro-T cell development (EdgeR; p.adj ≤ 0.05, $|\log_2$ fold change$| \geq 1$) with commitment-related changes in gene expression linked to those sites (shown, DN2a to DN3). Peaks were assigned to genes using proximity based annotation to nearest TSS using Homer *annotatePeaks.pl* (Heinz et al. 2010). *** Kolmogorov-Smirnov (KS) p-value ≤ 0.0001.

B) PU.1 binding motifs in pro-T cells are highly similar to those previously reported in macrophages. Sequence logo representations of the DN1-DN2b derived (left) or a Macrophage derived (right) PU.1 PWM in Table S1**.** The Macrophage derived PWM was derived from ref. (Pham et al. 2013) and the DN1-DN2b was derived as the top 12-mer motif from a DN1-DN2b combined PU.1 peak list [reanalyzed from (Zhang et al. 2012)].

C) Site quality distributions occupied by endogenous PU.1 as PU.1 concentrations decrease in development. Violin plots show the distribution of motif log-odds similarity score of DN1 to DN2b endogenous PU.1 peaks against a DN1-DN2b derived (green) or a Macrophage derived (Pham et al. 2013) (black) PU.1 PWM (Table S1). Median, 25% and 75% percentiles are shown. Dunn's corrected Kruskal-Wallis statistical test, *** $p \leq 0.0001$. Note that results are parallel but the more stringent DN1-DN2b PWM yields higher log-odds numerical scores.

D) Frequency of co-enriched transcription factor motifs at PU.1-occupied sites as functions of chromatin accessibility and PU.1 site quality. Motif analysis of DN1-DN2a PU.1 peaks above or below PWM score median in DN1 ATAC inaccessible (PWM median score=9.87, "DN1-DN2b"

PWM) or ATAC accessible (PWM median score=8.70, “DN1-DN2b” PWM)(cf. Fig. 1G). The top seven motifs defined by Homer De Novo motif search are shown for each site category.

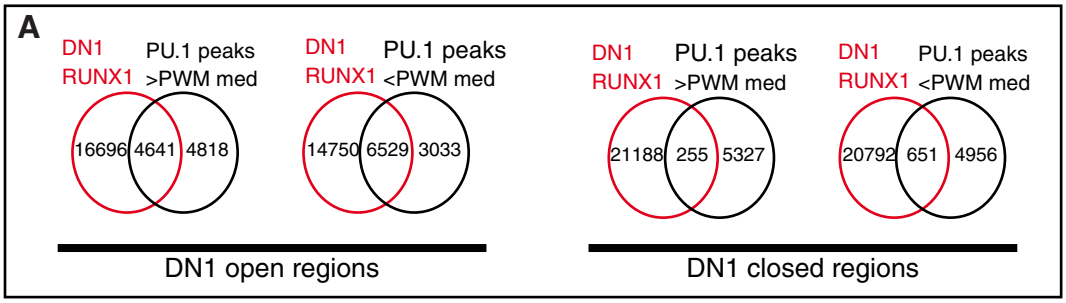
E) Site quality differences (top) and numbers of PU.1-occupied sites (bottom) at open and closed promoter and non-promoter PU.1 binding sites from DN1 to DN2b: note systematically lower site qualities for promoter sites even relative to other “open” sites at each stage. Violin plots show the distribution of motif log-odds similarity scores among DN1, DN2a and DN2b αPU.1 ChIP-seq peaks, separated based on ATAC-seq defined open (O) or closed (X) genomic regions as defined in DN1 (yellow, for reference with precommitment DN1 and DN2a cells) or DN3 (green, for reference with newly committed DN2b cells). Sites are also stratified for promoter or non-promoter (distal) elements. Scores are calculated in comparison to the PU.1 PWM-matrix derived from occupancy sites in DN1-DN2b cells (Table S1). Median, 25% and 75% quantiles are shown. Bar plots (below) show the numbers of occupancy peaks in each respective site category in each type of sample.

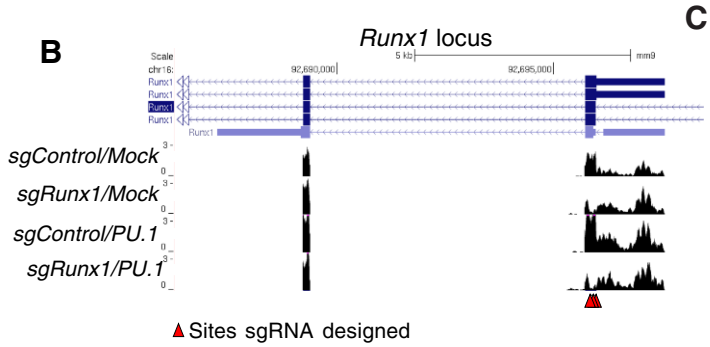
F) Parallel losses of PU.1 occupancy from conditionally and constitutively open sites as PU.1 levels decrease. Distribution of normalized PU.1 ChIP-seq tag counts in Group 1 (green) and Group 2 (yellow) in Figure 2A in DN1, DN2a and DN2b cells. Normalized tag counts for sites in Groups 1 and 2 (as defined in Fig 2A) were derived from each sample. Mann-Whitney.***, $p \leq 0.0001$.

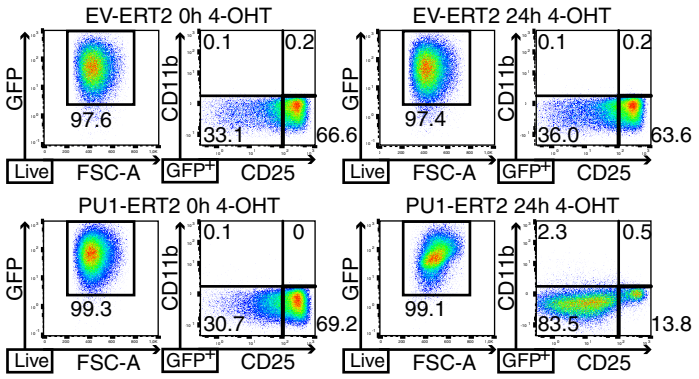
G) Evidence for efficient CAS9-mediated deletion of *Spi1* exons on both alleles, based on RNA-seq. Browser tracks show RNA transcribed from the *Spi1* and *Mybpc3* loci in CAS9-expressing cells that have been transduced with either guide RNA control (*sgControl*) or guide RNA against the first two coding exons of *Spi1* (sgPU.1) as diagrammed in Fig. 2D. Direction of transcription is from left to right. Red arrowheads show the sites targeted by the guide RNAs. Note severe loss

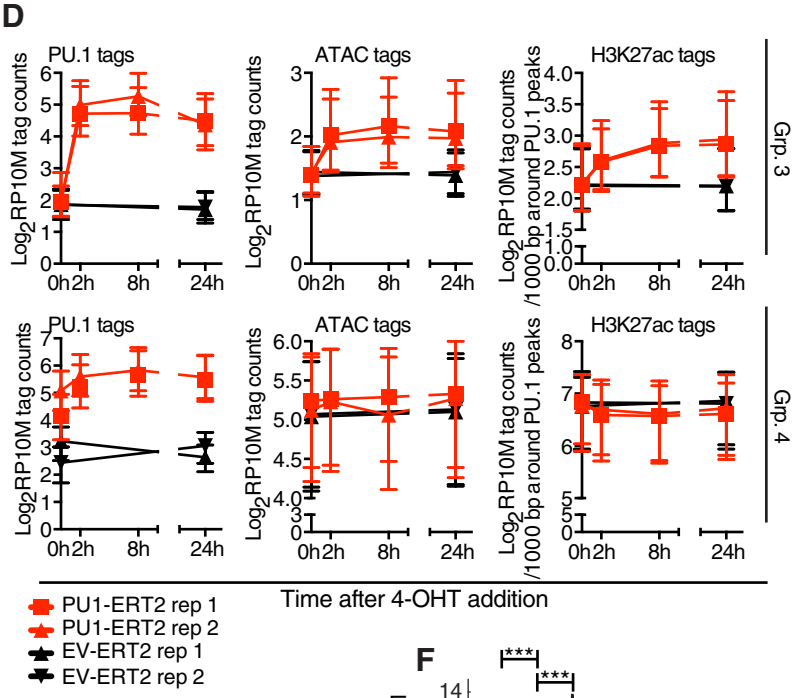
(>50% reduced) of targeted *Spi1* exon sequences from transcribed RNA, with no effect on transcripts from the neighboring *Mybpc3* gene.

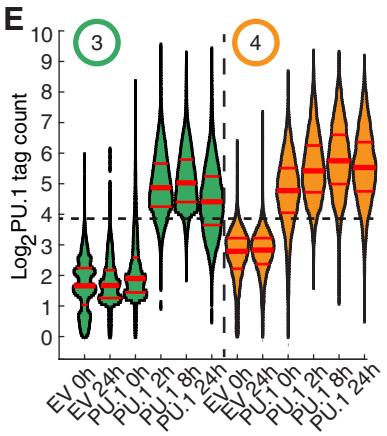
Figure S2

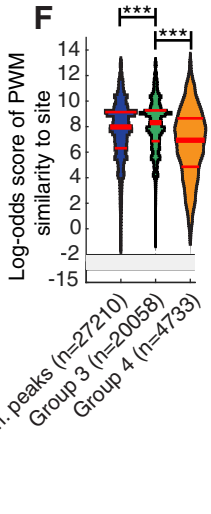


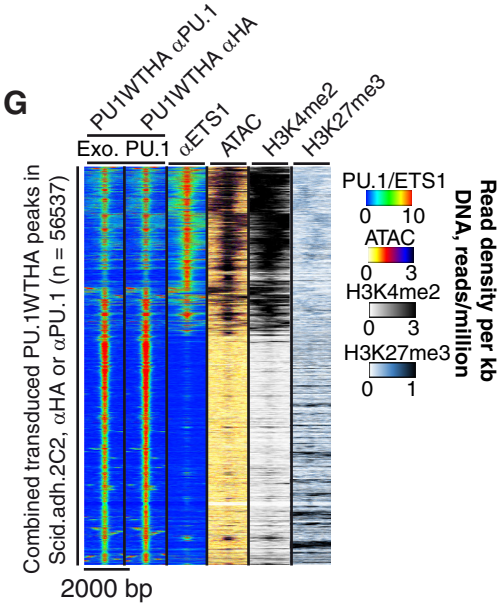












Top motifs after repeat masking	P-val.	% Targ.
3	1e-4338	85.5%
4	1e-1019	63.9%

Figure S2: Testing the role of RUNX1 in PU.1 binding and time course of responses to 4-hydroxytamoxifen (4-OHT) mobilized PU.1 in DN3-like Scid.adh.2C2 cells

A) RUNX1 specifically co-occupies open chromatin sites with endogenous PU.1 in primary cells. Venn diagrams of overlaps between RUNX1 binding in primary DN1 cells (Hosokawa et al. 2018) and PU.1 occupancy sites in primary DN1 cells, stratified by their ATAC status and PWM scores above or below median.

B) Evidence from RNA-seq analysis for efficient disruption of Runx first common exon coding sequence in Scid.adh.2C2 cells transduced with CAS9 and specific *sgRunx1* guide RNA. Control experiment for analysis in Fig. 3B,C shows sequences from the start of the coding region for the Runt domain in RUNX1 (transcription direction from right to left). Red arrowheads show sites targeted by three *Runx1*-specific guide RNAs directed just downstream of the initiation codon AUG in the Promoter 2 variant of *Runx1*; this sequence is in frame with sequences spliced in from RNAs initiating upstream from the Promoter 1 variant (not shown). RNA-seq tracks show that the targeted region experiences deletion by *sgRunx1* as compared to its expression in *sgControl* samples whether the cells are further transduced with PU.1 or not.

C) Rapid downregulation of CD25 in response to nuclear translocation of PU.1-ERT2. CD11b and CD25 surface staining in EV-ERT2- (empty vector control) or PU.1-ERT2-transduced Scid.adh.2C2 cells are shown 0h and 24h after 4-OHT stimulation. CD11b upregulation occurs later but only after 24h (not shown)(Del Real and Rothenberg 2013).

D) Quantitation of PU.1 binding, ATAC accessibility, and H3K27Ac in cells from Fig. 3D,E as a function of time after 4-OHT addition. Upper plots show status of Group 3 PU.1-binding sites,

lower plots show Group 4 sites. Data from each of the two independent replicate experiments are shown.

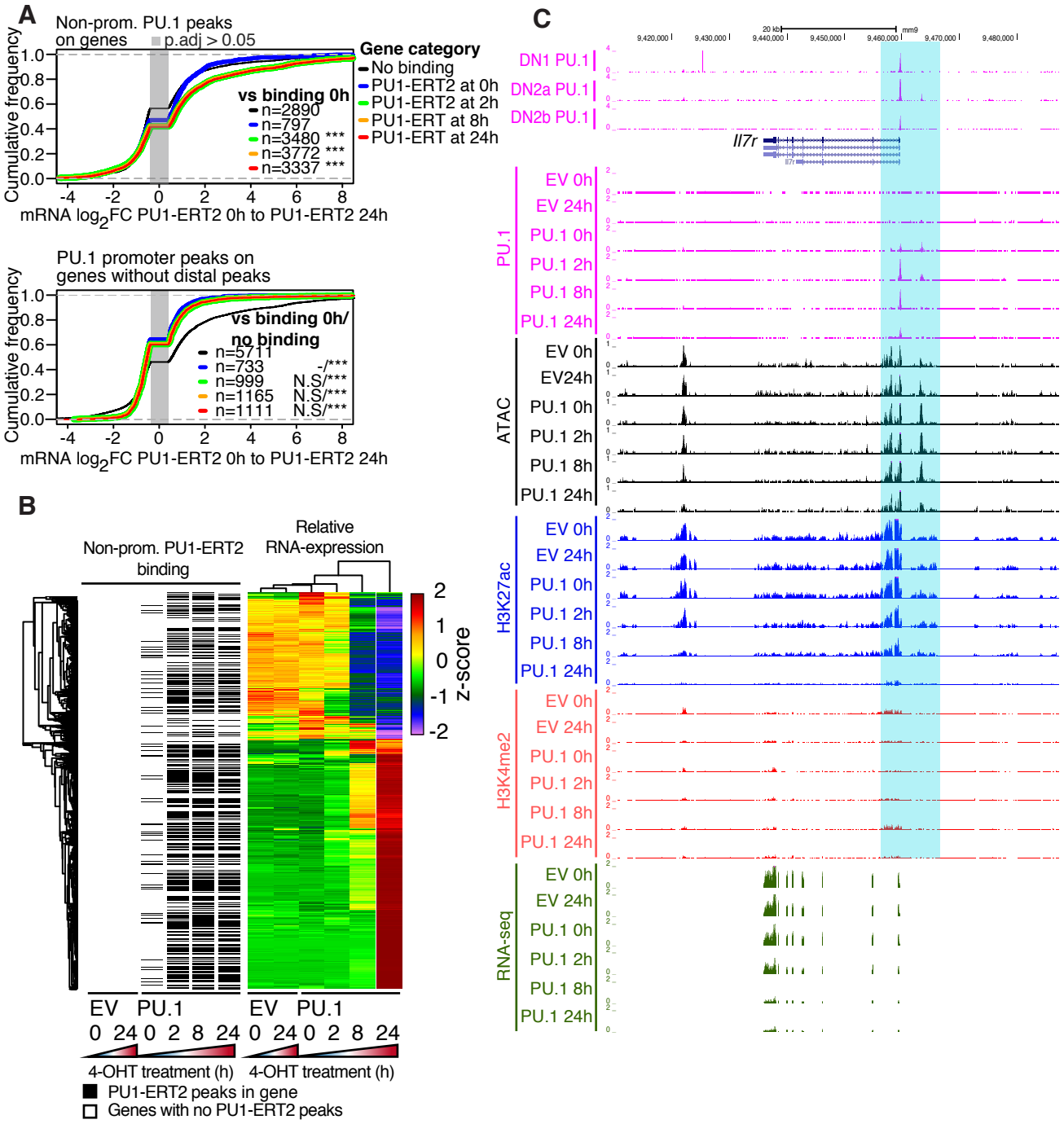
E) Time-dependent occupancy of closed and constitutively open sites (groups defined by Figure 3B clusters) by 4-OHT-mobilized PU.1-ERT2 in stably transduced Scid.adh.2C2 cells. Violin plots show increased PU.1-ERT2 occupancies at sites distinguished as Group 3 (green, initially closed) and Group 4 (yellow, constitutively open) in Fig. 3D, after 0-24hr of treatment with 4-OHT. Normalized ChIP-seq tag counts are shown. Signals from control cells transduced with EV-ERT2 (EV, empty vector) after 0 and 24h of 4-OHT are shown to establish background. The dashed line indicates the tag threshold for peaks considered significantly bound. Note that Group 3 sites go from baseline to full occupancy by 2h, whereas Group 4 sites show significant binding even before addition of 4-OHT.

F) Violin plots show the distribution of motif log-odds similarity scores of non-promoter peaks (promoter peaks filtered out) derived from Figure 3D Groups 3 (green; opening after 4-OHT mobilization) and 4 (orange; already open at 0h as well as after 4-OHT treatment) as well as list of combined Groups 3+4 PU.1 peaks (navy blue), scored using a DN1-DN2b derived PU.1 PWM-matrix (Table S1). Median, 25% and 75% percentiles are shown (red). Dunn's corrected Kruskal-Wallis statistical test, *** $p \leq 0.0001$. Sequence logos show HOMER identified top motifs (*findMotifsGenome.pl –size given*) in regions defined in Fig. 3B as Groups 3 and 4. Motif analysis in this case was carried out after masking of simple- sequence and UCSC-defined repeat sequences.

G) A separate exogenous PU.1 construct in transduced Scid.adh.2C2 cells, PU1WTHA, also preferentially binds to relatively closed sites, distinct from the open sites normally occupied by endogenous ETS1. The heat map shows PU1WTHA occupancy sites in the transduced cells

detected with either αPU.1 or αHA, compared with ETS1 binding detected with αETS1, ATAC-seq, H3K4me2 and H3K27me3 tag count distributions in non-transduced controls. ETS1 track shows merged tag directories from three independent ETS1 ChIP samples.

Figure S3


A
Non-prom. PU.1 peaks on genes
p.adj > 0.05
Cumulative frequency
mRNA log2FC PU1-ERT2 0h to PU1-ERT2 24h
Gene category
No binding
PU1-ERT2 at 0h
PU1-ERT2 at 2h
PU1-ERT at 8h
PU1-ERT at 24h
vs binding 0h
n=2890
n=797
n=3480 ***
n=3772 ***
n=3337 ***
PU.1 promoter peaks on genes without distal peaks
vs binding 0h/ no binding
n=5711
n=733 -/***
n=999 N.S/***
n=1165 N.S/***
n=1111 N.S/***
B
Non-prom. PU1-ERT2 binding
Relative RNA-expression
z-score
EV PU.1 EV PU.1
0 24 0 2 8 24 0 24 0 2 8 24
4-OHT treatment (h)
PU1-ERT2 peaks in gene
Genes with no PU1-ERT2 peaks
C
DN1 PU.1
DN2a PU.1
DN2b PU.1
Il7r
PU.1
ATAC
H3K27ac
H3K4me2
RNA-seq
EV 0h
EV 24h
EV24h
PU.1 0h
PU.1 2h
PU.1 8h
PU.1 24h

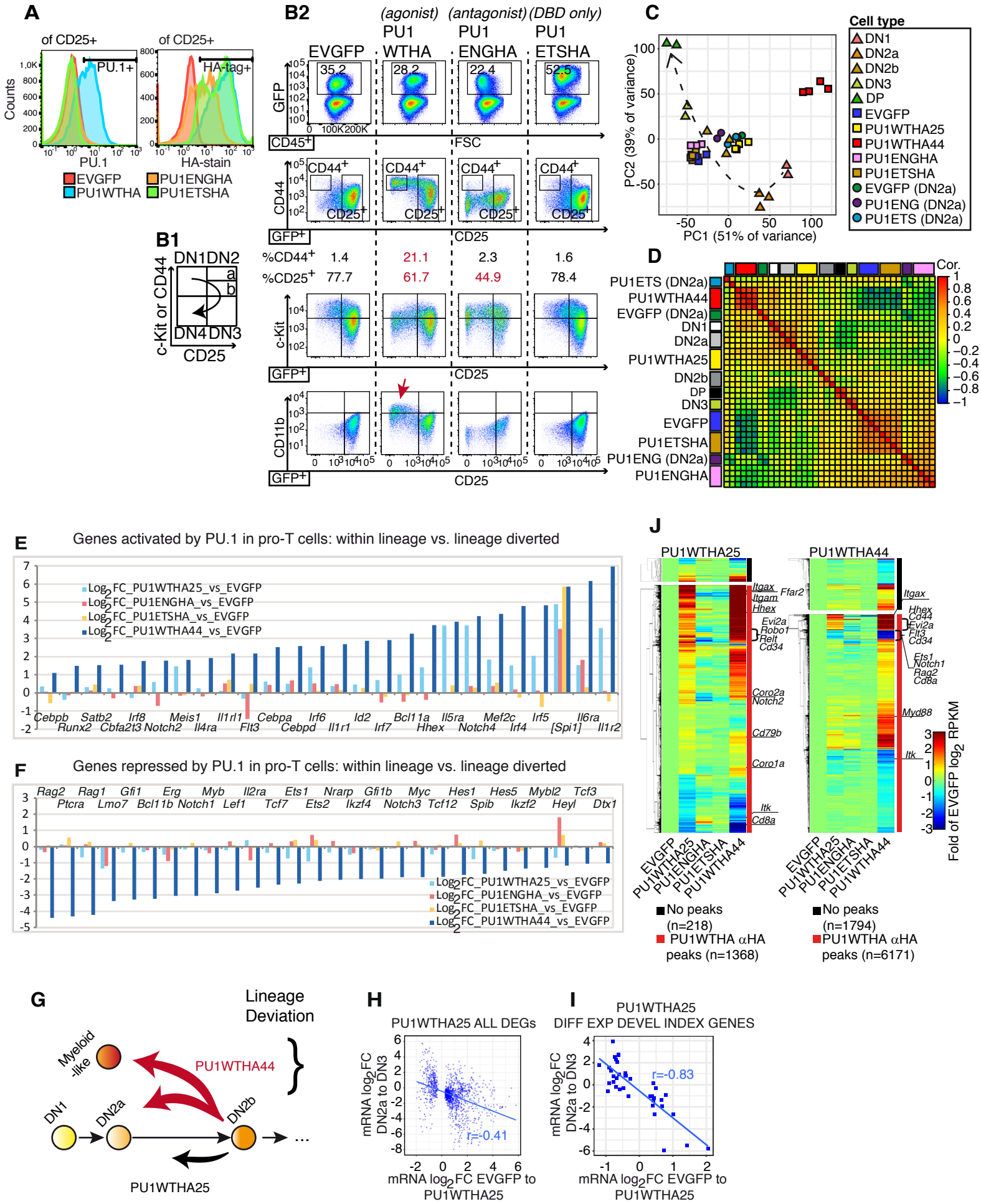
Supplementary Figure S3: Time course of gene regulation changes following chromatin opening triggered by PU.1-ERT2 mobilization in DN3-like Scid.adh.2C2 cells

A) Transcriptional changes induced upon 4-OHT stimulation are predictable by PU.1 occupancy of non-promoter elements as early as 2h. **Top:** Genes that change expression upon PU.1-ERT2 mobilization are compared in ECDF plots based on the timing when PU.1-ERT2 occupancy is detected at their linked non-promoter elements. Plots show cumulative fraction of genes in each PU.1 binding category (y axis) plotted according to their fold changes in expression as measured at 24h (x axis). The data show that the appearance of PU.1 at linked non-promoter sites at 2h predicts the response of the gene a day later as well as the occupancy of PU.1 at 24h. Samples with binding at 2h and 8h of 4-OHT treatment (green, orange curves) are plotted (thicker lines) but have gene expression responses superimposable with those showing binding at 24h (red curve; note overlap). Statistical significance of effects is determined vs. genes with background binding at 0h and indicated for genes with occupancy at each time point (n, number of genes in each category). *** Kolmogorov-Smirnov p-value ≤ 0.0001. Bottom: Same as Top, but only analyzing genes with PU.1 uniquely bound to promoters (no distal peaks). n, described number of genes per group. Note: because PU.1-ERT2 that enters the nucleus slowly without 4-OHT preferentially accumulates at promoters, as described in the text, significance of response is calculated both vs. genes with no promoter binding and also vs. genes with PU.1 binding to their promoters already at 0h. Peaks were assigned to genes using proximity based annotation to nearest TSS using Homer *annotatePeaks.pl* (Heinz et al. 2010).

B) Time course of RNA expression responses to PU.1-ERT2 nuclearization. Right, Heatmap displays hierarchically clustered RPKM values of differentially expressed genes in cells stably

transduced with EV-ERT2 (EV) or PU.1-ERT2 (PU.1) after 0h to 24h 4-OHT. Values for each gene were z-score row-normalized over all the samples. Left, presence of PU.1 occupancy (black line) linked to non-promoter sites of each of these genes at indicated time points. Note downregulation of some genes preceding upregulation of most genes in PU.1-ERT2 samples. **C)** Time course of PU.1 occupancy, ATAC accessibility change, histone modification and transcriptional activity at the *Il7r* locus in Scid.adh.2C2 cells when PU1-ERT2 is induced by 4-OHT. PU.1 dampens expression of this gene rather than activating it, in contrast to *Icam1* (Fig. 3D). Figure shows UCSC browser tracks of PU.1-ERT2 in stably transduced Scid.adh.2C2 cells at each timepoint of 4-OHT treatment as detected by αPU.1 ChIP (pink tracks), compared with the ATAC-seq (black tracks), H3K27ac (blue tracks), H3K4me2 (red tracks) and RNA-seq (green tracks) signals from the same samples. For comparison, top tracks (pink) show occupancy of endogenous PU.1 in primary DN1, DN2a, and DN2b cells. Note: at this locus, the endogenous PU.1 binding in primary cells is stronger than binding in the Scid.adh.2C2 transductants; note twofold difference in y axis scale for PU.1 binding in the primary cells. This system shows the timing of distinct changes in PU.1 binding and ATAC status (Cyan highlight) as *Il7r* expression begins to decline.

Figure S4


A
of CD25+
PU.1+
HA-tag+
Counts
PU.1
HA-stain
EVGFP
PU1WTHA
PU1ENGHA
PU1ETSHA
B1
DN1 DN2
a
b
DN4 DN3
c-Kit or CD44
CD25
B2
(agonist)
(antagonist)
(DBD only)
PU1 WTHA
PU1 ENGHA
PU1 ETSHA
EVGFP
35.2
28.2
22.4
52.5
GFP
CD45+
FSC
CD44+
CD25+
CD44
GFP+
CD25
%CD44+ 1.4 21.1 2.3 1.6
%CD25+ 77.7 61.7 44.9 78.4
c-Kit
CD11b
C
PC2 (39% of variance)
PC1 (51% of variance)
Cell type
DN1
DN2a
DN2b
DN3
DP
EVGFP
PU1WTHA25
PU1WTHA44
PU1ENGHA
PU1ETSHA
EVGFP (DN2a)
PU1ENG (DN2a)
PU1ETS (DN2a)
D
Cor.
PU1ETS (DN2a)
PU1WTHA44
EVGFP (DN2a)
DN1
DN2a
PU1WTHA25
DN2b
DP
DN3
EVGFP
PU1ETSHA
PU1ENG (DN2a)
PU1ENGHA
E
Genes activated by PU.1 in pro-T cells: within lineage vs. lineage diverted
Log2FC_PU1WTHA25_vs_EVGFP
Log2FC_PU1ENGHA_vs_EVGFP
Log2FC_PU1ETSHA_vs_EVGFP
Log2FC_PU1WTHA44_vs_EVGFP
Cebpb Satb2 Irf8 Meis1 Il1rl1 Cebpa Irf6 Id2 Bcl11a Il5ra Mef2c Irf5 Il6ra
Runx2 Cbfa2t3 Notch2 Il4ra Flt3 Cebpd Il1r1 Irf7 Hhex Notch4 Irf4 [Spi1] Il1r2
F
Genes repressed by PU.1 in pro-T cells: within lineage vs. lineage diverted
Rag2 Rag1 Gfi1 Erg Myb Il2ra Ets1 Nrarp Gfi1b Myc Hes1 Hes5 Mybl2 Tcf3
Ptcra Lmo7 Bcl11b Notch1 Lef1 Tcf7 Ets2 Ikzf4 Notch3 Tcf12 Spib Ikzf2 Heyl Dtx1
Log2FC_PU1WTHA25_vs_EVGFP
Log2FC_PU1ENGHA_vs_EVGFP
Log2FC_PU1ETSHA_vs_EVGFP
Log2FC_PU1WTHA44_vs_EVGFP
J
PU1WTHA25
PU1WTHA44
Itgax
Itgam
Hhex
Evi2a
Robo1
Relt
Cd34
Coro2a
Notch2
Cd79b
Coro1a
Itk
Cd8a
Ffar2
Cd44
Flt3
Ets1
Notch1
Rag2
Myd88
EVGFP
PU1WTHA25
PU1ENGHA
PU1ETSHA
PU1WTHA44
Fold of EVGFP log2 RPKM
No peaks (n=218)
PU1WTHA αHA peaks (n=1368)
No peaks (n=1794)
PU1WTHA αHA peaks (n=6171)
G
Lineage Deviation
Myeloid -like
PU1WTHA44
DN1
DN2a
DN2b
...
PU1WTHA25
H
PU1WTHA25 ALL DEGs
mRNA log2FC DN2a to DN3
r=-0.41
mRNA log2FC EVGFP to PU1WTHA25
I
PU1WTHA25
DIFF EXP DEVEL INDEX GENES
mRNA log2FC DN2a to DN3
r=-0.83
mRNA log2FC EVGFP to PU1WTHA25

Supplementary Figure S4: Definition and properties of the gene set specifically regulated by PU.1 within the T-cell program.

A) Expression of exogenous PU.1 and PU.1 variants transduced into primary DN2b-DN3 pro-T cells. Intracellular staining of PU.1 (left) and the HA-tag (right) ~40h after transduction of $CD25^+$ cells with EVGFP, PU1WTHA, PU1ENGHA and PU1ETSHA. Prior to transduction, $CD25^+$ cells were generated by *in vitro* differentiation from fetal liver hematopoietic progenitors as described in Methods. Data are presented as histograms with counts of $GFP^+/CD25^+$ cells at different levels of fluorescence. The data in the FACS plots are representative of three independent experiments.

B) Effects on developmental marker expression by reintroduction of PU.1, PU.1 antagonists, or empty vector (EV) into fetal liver-derived CD25+ **(**DN2b/DN3) pro-T cells. B1: Schematic showing normal developmental trajectory. B2: FACS analyses are shown 40h after transduction with the indicated constructs. Exogenous wildtype PU.1: PU1WTHA. For details of other constructs, see Supplementary Fig. S6 below. Note the increase in cells with elevated CD44 in cells transduced with PU1WTHA, including some that downregulate the DN2-DN3 marker CD25 and even upregulate the myeloid-cell marker CD11b. DBD: DNA binding domain. Data are representative of four independent experiments.

C) Principal component analysis of expressed genes (RPKM ≥ 1) in four independent samples of CD25+ DN2b/DN3 cells transduced with EVGFP, PU1WTHA, PU1ENGHA, or PU1ETSHA (square symbols), compared with normal reference cells (triangles) (Zhang et al. 2012). Samples are also compared with a separate series of DN2a PU.1 antagonist samples (circles) from a previous study (GSE65344)(Champhekar et al. 2015). Dashed arrow: approximate trajectory of normal development. Note that the PU1WTHA transduced cells remain close to the normal

trajectory if they retain CD25 (PU1WTHA25 cells), but sharply deviate from it if they switch to the CD25-negative, CD44-high, mostly $CD11b^+$ "PU1WTHA44") phenotype (red squares).
D) Correlation analysis of gene expression as described in panel C. This display with one dimension of separation shows the effect of wildtype PU.1 in pushing the cells toward a more "immature" state with normal DN1 and DN2a uncommitted cells. If anything, the trend of cells transduced with the PU.1 antagonist (PU1ENGHA) is toward the more "mature" pole.
E-F) Effects of PU.1 on gene expression in PU1WTHA44 cells indicate a lineage switch with many secondary regulatory changes.

E) Specific regulatory gene expression that is abnormally upregulated in PU1WTHA44 samples as compared to effects of PU.1 in PU1WTHA25. Effects of PU1 antagonists are also shown. Data taken from Table S3 show effects on non-T lineage genes or effector lymphocyte genes precociously expressed, among genes with p.adj<0.1 in PU1WTHA44 and |log2FC|>1. Graph shows $\log_2$ fold changes in the indicated samples relative to empty vector controls. Most show a stronger upregulation in PU1WTHA44 cells than in PU1WTHA25 cells, and some appear upregulated almost exclusively in the PU1WTHA44 cells. Note the selective stimulation of *Cebp* and *Irf* family genes in the PU1WTHA44 cells as compared to the other samples.

F) Specific regulatory gene expression: T-cell gene expression that is abnormally repressed in PU1WTHA44 as compared to PU.1 effects in PU1WTHA25. Data taken from Table S3 show effects on T-cell identity-associated regulatory genes repressed or silenced, among genes with p.adj<0.1 in PU1WTHA44 and $|\log_2 FC|>1$. Note repressive effects on Notch signaling associated genes *Ptcra, Dtx1, Heyl, Hes5, Hes1, Notch3,* and *Myc*, all downregulated specifically in the PU1WTHA44 samples. As for (E), graph shows $\log_2$ fold changes in the indicated samples relative to empty vector controls. Note the dramatic difference in the repression of these genes

between cells remaining within the T-cell pathway (PU1WTHA25) and cells being diverted from it (PU1WTHA44).

G) Schematic of PU.1 effects on pro-T cells when added back after normal downregulation: partial retrograde development within the T cell pathway vs. frank lineage diversion. Diagram is based on contrasting the effects of PU1WTHA on developmental gene expression (panels C-F) in cells remaining $CD25^{+}$ (PU1WTHA25) and cells becoming CD25-low $CD44^{+}$ (PU1WTHA44). The extensive secondary alterations in expression of other regulatory genes in PU1WTHA44 cells undergoing lineage diversion imply that the effects of PU.1 itself within this developmental program can be detected more cleanly in the PU1WTHA25 cells. This is the rationale for focusing site binding vs. function analyses on the PU1WTHA25 cells alone (Fig. 4, Supplementary Fig. S5, and below).

H) Transcriptional changes upon transduction of CD25+ cells with PU1WTHA as opposed to EV (blue) compared with normal transcriptional changes from DN2a to DN3 stages. Significantly differentially expressed genes in PU1WTHA25 cells as compared to empty vector controls (DESeq2, p.adj ≤ 0.1) are plotted with their fold changes in response to PU.1 against their natural fold changes in development, with Pearson's r shown.

I) Same as (**H**), for regulatory genes from a highly curated 171-gene developmental index list (Table S4)(Champhekar et al. 2015; Longabaugh et al. 2017). To select the genes shown in the plot, genes within the developmental index list with the most significant differential expression in response to PU.1 were scored manually. The Bioconductor qvalue package was applied on the DESeq2 derived p-values for regulatory gene expression differences between PU1WTHA25 and EV control samples, and values of $q<0.05$ within the list were taken as significant.

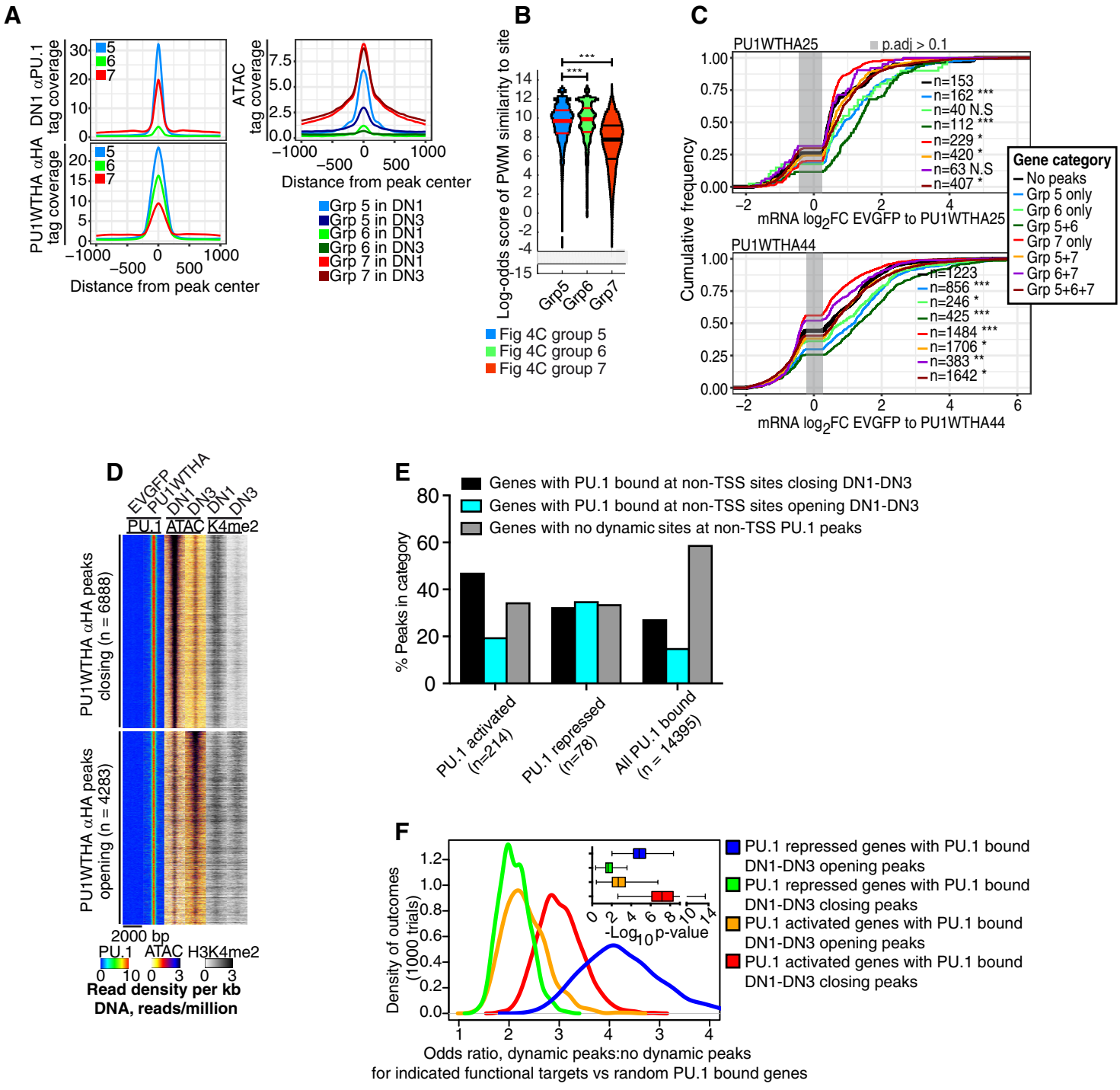

A
DN1 αPU.1 tag coverage
PU1WTHA αHA tag coverage
ATAC tag coverage
Distance from peak center
Grp 5 in DN1
Grp 5 in DN3
Grp 6 in DN1
Grp 6 in DN3
Grp 7 in DN1
Grp 7 in DN3
B
Log-odds score of PWM similarity to site
Grp5
Grp6
Grp7
Fig 4C group 5
Fig 4C group 6
Fig 4C group 7
C
PU1WTHA25
p.adj > 0.1
Cumulative frequency
mRNA log2FC EVGFP to PU1WTHA25
PU1WTHA44
mRNA log2FC EVGFP to PU1WTHA44
Gene category
No peaks
Grp 5 only
Grp 6 only
Grp 5+6
Grp 7 only
Grp 5+7
Grp 6+7
Grp 5+6+7
D
EVGFP
PU1WTHA
DN1
DN3
DN1
DN3
PU.1
ATAC
K4me2
PU1WTHA αHA peaks closing (n = 6888)
PU1WTHA αHA peaks opening (n = 4283)
2000 bp
PU.1
ATAC
H3K4me2
Read density per kb DNA, reads/million
E
Genes with PU.1 bound at non-TSS sites closing DN1-DN3
Genes with PU.1 bound at non-TSS sites opening DN1-DN3
Genes with no dynamic sites at non-TSS PU.1 peaks
% Peaks in category
PU.1 activated (n=214)
PU.1 repressed (n=78)
All PU.1 bound (n = 14395)
F
Density of outcomes (1000 trials)
-Log10 p-value
PU.1 repressed genes with PU.1 bound DN1-DN3 opening peaks
PU.1 repressed genes with PU.1 bound DN1-DN3 closing peaks
PU.1 activated genes with PU.1 bound DN1-DN3 opening peaks
PU.1 activated genes with PU.1 bound DN1-DN3 closing peaks
Odds ratio, dynamic peaks:no dynamic peaks
for indicated functional targets vs random PU.1 bound genes

Figure S5

Supplementary Figure S5: Exogenous PU.1-wt binding and gene regulation are modulated by chromatin status

A) DN1 αPU.1 (top left), exogenous PU1WTHA αHA (bottom left) and ATAC (top right) distribution plots within 1 kb of endogenous or exogenous PU.1 bound sites from Groups 5, 6, and 7, as defined in Fig. 4C.

B) Site quality for sequences occupied by exogenous PU.1 is higher at normally closed or closing sites than at constitutively open sites. Violin plots show the distribution of motif log-odds similarity scores for sites defined as closing (Group 5), closed (Group 6), or constitutively open (Group 7) in Fig. 4C. Scores were defined by the DN1-DN2b derived PU.1 PWM-matrix (Table S1). Median, 25% and 75% percentiles are shown. Dunn's corrected Kruskal-Wallis statistical test, *** $p \leq 0.0001$.

C) Functional impact of exogenous PU.1 binding at distinct site classes: qualitatively distinct contributions to local gene expression are additive in functional impact. ECDF plots shows changes in gene expression induced by exogenous PU1WTHA for genes linked to different classes of PU.1 sites. Genes are grouped based on their linkage to one or more PU.1 binding sites in Groups 5, 6, or 7 in Figure 4C. Fig. 4D shows a subset of these genes, namely those with sites of one class only. Here are also shown the more frequent genes that are linked to sites in combinations of classes. Note that additional Group 5 or Group 6 sites may offset the damping influence of Group 7 sites on genes linked to both kinds of sites as compared to genes with Group 7 sites alone (red), and the consistently higher upregulation of target genes that have Group 5 (blue) or Group 6 (light green) sites but not Group7 sites. Highest upregulation is seen for genes that have both Group 5 and Group 6 sites (dark green). Top shows genes affected in PU.1-transduced cells that remain in the pro-T cell pathway (PU1WTHA25), while bottom

shows the same genes in cells that include many diverting out of the T-cell pathway completely (PU1WTHA44). The results suggest that Group 7 site-associated binding may be associated with a higher likelihood of a gene's becoming repressed if and only if the cells are undergoing lineage diversion. Peaks were assigned to genes using proximity based annotation to nearest TSS using Homer *annotatePeaks.pl* (Heinz et al. 2010). *** Kolmogorov-Smirnov p-value ≤ 0.0001, ** p-value ≤ 0.001,* p-value ≤ 0.05.

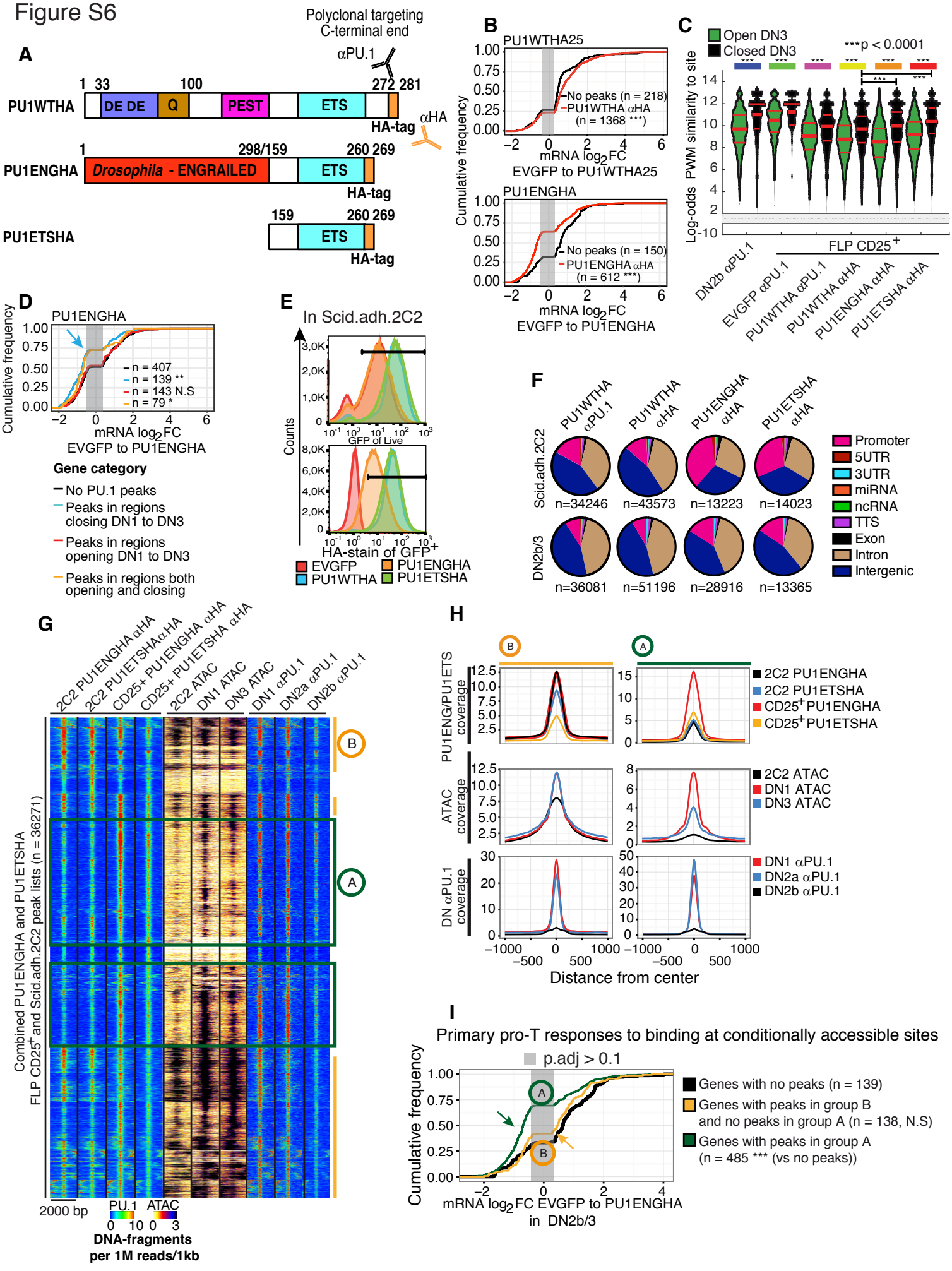
D) Definition of sites of exogenous PU1WTHA binding in transduced $CD25^{+}$ cells at sites that are naturally closing (top) or opening (bottom) *in vivo* from DN1 to DN3 stages. Heat maps show exogenous PU1WTHA (αHA) ChIP-seq, ATAC-seq, and H3K4me2 tag count distributions. Peaks are ordered by their PU1WTHA tag counts (high to low).

E) Test of enrichment of PU.1 sites with dynamic chromatin status at genes regulated by PU.1. Figure shows a tally of PU.1 binding non-promoter sites linked to highest-confidence PU.1 target genes (reciprocally responding to gain and loss of PU.1 function within the T-cell pathway; from Table S6), as compared to those at all genes with PU.1 binding to non-promoter sites (all PU.1 bound), in terms of the opening or closing ATAC status of their linked sites during the DN1 to DN3 developmental progression. Both PU.1-activated and PU.1-repressed targets in this high-confidence list were enriched for linkage to developmentally dynamic sites (black bars, closing DN1 to DN3; cyan bars, opening DN1 to DN3), as compared to the total.

F) Statistical significance of enrichment trends shown in (E). Main panel: distribution of odds ratios, dynamic sites vs. developmentally non-changing sites, in PU.1-repressed and PU.1-activated functional targets as compared to randomly chosen PU.1-bound control genes. Curves show odds ratios obtained in 1000 iterations of the indicated functional target gene groups vs. randomly chosen sets of 300 control genes with PU.1 binding in non-promoter regions. Inset

shows distribution of p values (Fisher's exact test). Significant enrichments were seen for developmentally closing PU.1 binding sites linked to PU.1-activated targets (red, most $p<10^{-6}$), and for developmentally opening PU.1 sites linked to PU.1-repressed targets (blue, most $p<10^{-4}$).

Figure S6


A
Polyclonal targeting C-terminal end
αPU.1
PU1WTHA
1 33 100 272 281
DE DE Q PEST ETS
HA-tag
αHA
PU1ENGHA
1 298/159 260 269
Drosophila - ENGRAILED ETS
HA-tag
PU1ETSHA
159 260 269
ETS
HA-tag
B
PU1WTHA25
No peaks (n = 218)
PU1WTHA αHA (n = 1368 ***)
mRNA log2FC EVGFP to PU1WTHA25
PU1ENGHA
No peaks (n = 150)
PU1ENGHA αHA (n = 612 ***)
mRNA log2FC EVGFP to PU1ENGHA
Cumulative frequency
C
Open DN3
Closed DN3
***p < 0.0001
Log-odds PWM similarity to site
FLP CD25+
DN2b αPU.1
EVGFP αPU.1
PU1WTHA αPU.1
PU1WTHA αHA
PU1ENGHA αHA
PU1ETSHA αHA
D
PU1ENGHA
n = 407
n = 139 **
n = 143 N.S
n = 79 *
mRNA log2FC EVGFP to PU1ENGHA
Gene category
No PU.1 peaks
Peaks in regions closing DN1 to DN3
Peaks in regions opening DN1 to DN3
Peaks in regions both opening and closing
E
In Scid.adh.2C2
Counts
GFP of Live
HA-stain of GFP+
EVGFP
PU1WTHA
PU1ENGHA
PU1ETSHA
F
PU1WTHA αPU.1
PU1WTHA αHA
PU1ENGHA αHA
PU1ETSHA αHA
Scid.adh.2C2
DN2b/3
n=34246
n=43573
n=13223
n=14023
n=36081
n=51196
n=28916
n=13365
Promoter
5UTR
3UTR
miRNA
ncRNA
TTS
Exon
Intron
Intergenic
G
2C2 PU1ENGHA αHA
2C2 PU1ETSHA αHA
CD25+ PU1ENGHA αHA
CD25+ PU1ETSHA αHA
2C2 ATAC
DN1 ATAC
DN3 ATAC
DN1 αPU.1
DN2a αPU.1
DN2b αPU.1
Combined PU1ENGHA and PU1ETSHA FLP CD25+ and Scid.adh.2C2 peak lists (n = 36271)
2000 bp
PU.1
ATAC
DNA-fragments per 1M reads/1kb
H
PU1ENG/PU1ETS coverage
ATAC coverage
DN αPU.1 coverage
2C2 PU1ENGHA
2C2 PU1ETSHA
CD25+ PU1ENGHA
CD25+ PU1ETSHA
2C2 ATAC
DN1 ATAC
DN3 ATAC
DN1 αPU.1
DN2a αPU.1
DN2b αPU.1
Distance from center
I
Primary pro-T responses to binding at conditionally accessible sites
p.adj > 0.1
Cumulative frequency
mRNA log2FC EVGFP to PU1ENGHA in DN2b/3
Genes with no peaks (n = 139)
Genes with peaks in group B and no peaks in group A (n = 138, N.S)
Genes with peaks in group A (n = 485 *** (vs no peaks))

Supplementary Figure S6: DNA binding specificity and chromatin-based site selectivity for PU.1 constructs lacking regions outside the DNA binding domain

A) Maps of PU1WTHA, PU1ENGHA, and PU1ETSHA, showing defined subdomains of wildtype PU.1. ETS: ETS-family DNA binding domain. The regions fused with the HA tag and recognized by the αPU.1 antibody used in these studies are also shown.

B) Physical linkage of sites for exogenous PU.1 construct binding, genome-wide, with genes changing expression significantly (p.adj<0.1) upon transduction of $CD25^+$ cells with PU1WTHA and PU1ENGHA. Peaks were assigned to genes using proximity based annotation to nearest TSS using Homer *annotatePeaks.pl* (Heinz et al. 2010). *** Kolmogorov-Smirnovp-value ≤ 0.0001.

C) Truncated PU.1 forms PU1ETSHA and PU1ENGHA (Champhekar et al. 2015) occupy sites in primary DN2b-DN3 cells with similar criteria to PU1WTHA. Violin plots show the site quality distribution of endogenous PU.1 in DN2b cells or $CD25^+$ DN2b-DN3 cells transduced with empty vector, compared with the site quality distributions occupied by exogenous PU1WTHA (detected both with αPU.1 and with αHA), PU1ENGHA detected with αHA, and PU1ETSHA detected with αHA. Motif log-odds similarity scores for each are compared in sites defined as open (green) -or closed (black) by ATAC-seq in primary DN3 cells. The pro-T cell derived PU.1 PWM-matrix was used (Table S1). Median, 25% and 75% quantiles are shown. Dunn's corrected Kruskal-Wallis statistical test***, $p \leq 0.0001$.

D) Truncated PU.1 antagonist construct is functional on a physiologically relevant set of sites where it binds: effects of PU1ENGHA on genes linked to sites in opening or closing regions. Arrow shows that obligate repressor effect is most strongly correlated with sites that are only open when endogenous PU.1 is expressed. ** Kolmogorov-Smirnov p-value ≤0.001.

E) Expression of exogenous PU.1 constructs PU1WTHA, PU1ENGHA, and PU1ETSHA in Scid.adh.2C2 cells. Intracellular staining of HA-tag (bottom) as well as analysis of GFP (top) expression ~40h after transduction of Scid.adh.2C2 cells with EVGFP, PU1WTHA, PU1ENGHA and PU1ETSHA. GFP^+ data are presented as counts of live cells, and HA-stain is presented as counts within the GFP^+ gate. The data in the FACS plots are representative of three independent experiments.

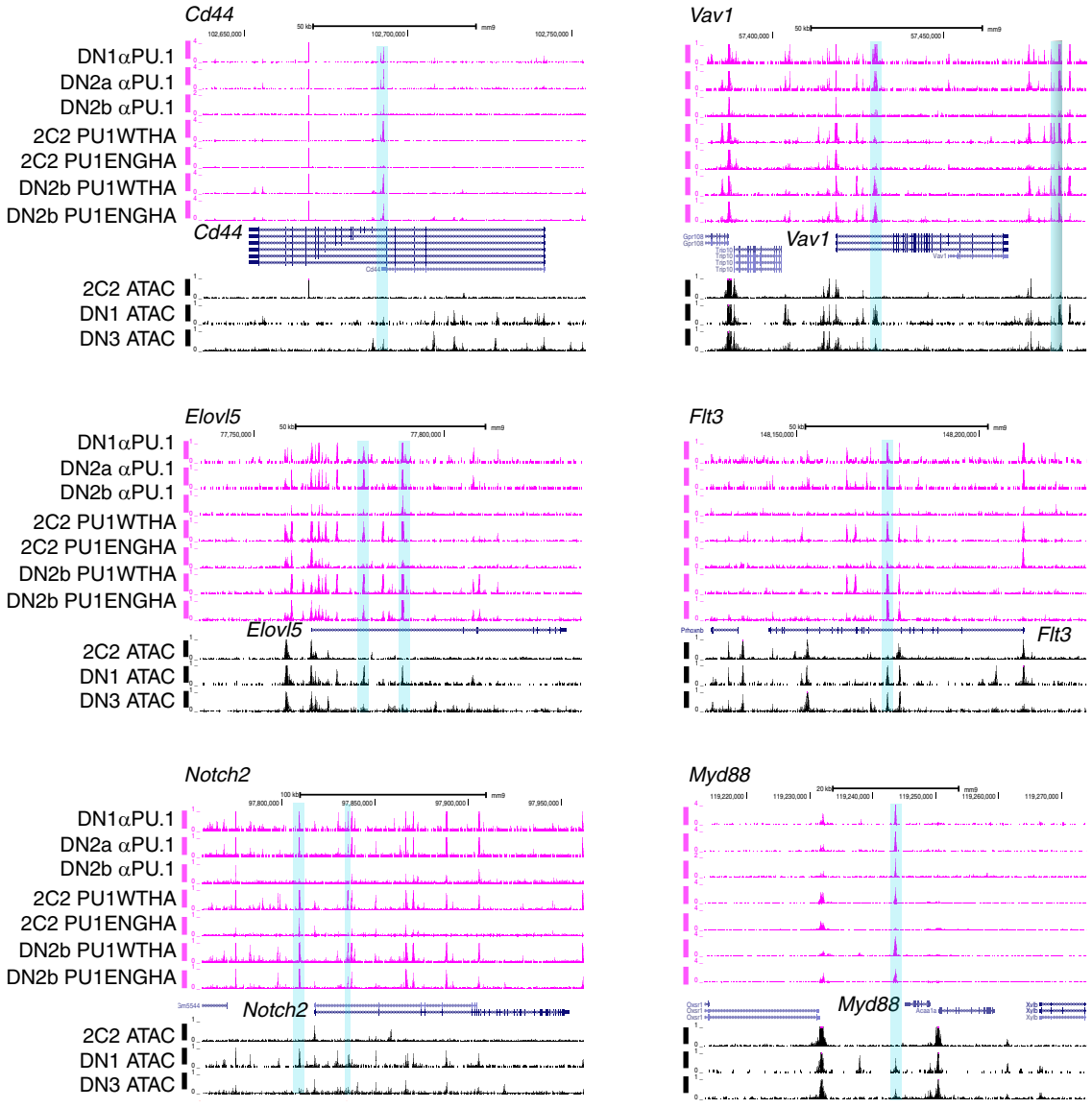
F) Preferential binding of PU1ENGHA and PU1ETSHA to promoters in both Scid.adh.2C2 cells and primary DN2b/3 cells.

G) Context-dependent differences in access of PU1ENGHA and PU1ETSHA to natural sites of endogenous PU.1 binding in Scid.adh.2C2 or fetal liver precursor-derived $CD25^+$ pro-T cells. Combined Scid.adh.2C2 and primary-cell PU1ENGHA & PU1ETSHA peak list was hierarchically clustered according to tag count profiles of PU1ENGHA or PU1ETSHA; DN1, DN3, and Scid.adh.2C2 ATAC-seq; and DN1, DN2a and DN2b endogenous PU.1. Shown are manually derived site groups that are open in both primary cells ($CD25^+$) and Scid.adh.2C2 cells (Group B) or in primary cells only (Group A).

H) Differential behavior of PU1ENGHA and PU1ETSHA in Scid.adh.2C2 cells (2C2) as compared to primary $CD25^+$ cells. Figure shows distribution plots of Scid.adh.2C2 –and $CD25^+$ PU.1ENGHA/PU1ETSHA binding tag counts are shown, as well as DN1, DN3, and Scid.adh.2C2 ATAC-seq signals and DN1, DN2a and DN2b endogenous PU.1 binding within 1 kb of PU.1ENGHA/PU1ETSHA bound sites, based on Groups A and B in panel A of this figure. Note the sharply reduced binding of these constructs to sites in Group A (closing in primary DN cells, completely closed in Scid.adh.2C2) in the Scid.adh.2C2 cells as compared to the primary cells.

I) Sites with context-dependent accessibility are uniquely potent for PU1ENGHA repressive function: Association of changes in linked gene expression in response to PU1ENGHA binding to sites accessible both in primary and Scid.adh.2C2 cells (panel F, Group B) or to sites accessible in primary cells but not Scid.adh.2C2 cells (panel F, Group A).

Figure S7


Cd44
DN1αPU.1
DN2a αPU.1
DN2b αPU.1
2C2 PU1WTHA
2C2 PU1ENGHA
DN2b PU1WTHA
DN2b PU1ENGHA
Cd44
2C2 ATAC
DN1 ATAC
DN3 ATAC
Vav1
Vav1
Elovl5
Elovl5
Flt3
Flt3
Notch2
Notch2
Myd88
Myd88

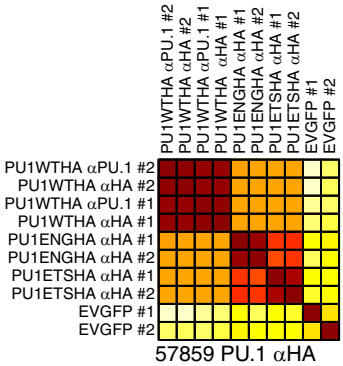
Supplementary Figure S7: Access to closed chromatin requires PU.1 domains beyond the ETS DNA binding domain: additional examples

UCSC Genome Browser tracks displaying binding of full-length PU.1 PU1WTHA and PU1ENGHA to open and closed regions in primary DN2b/3 and Scid.adh.2C2 cells. PU1ENGHA is limited to DN2b/3 open regions only and excluded from binding to ATAC inaccessible regions in Scid.adh.2C2 cells. Shown are regions encompassing the genes: *Cd44, Vav1, Elovl5, Flt3, Notch2,* and *Myd88*.

Figure S8

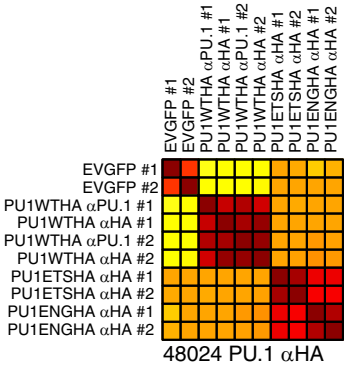
A) Correlation between ChIP-seq replicates in study

FLP DN2b/3 PU.1HA ChIP (n=2)



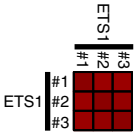
57859 PU.1 αHA and αPU.1 peaks

Scid.adh.2C2 PU.1HA ChIP (n=2)



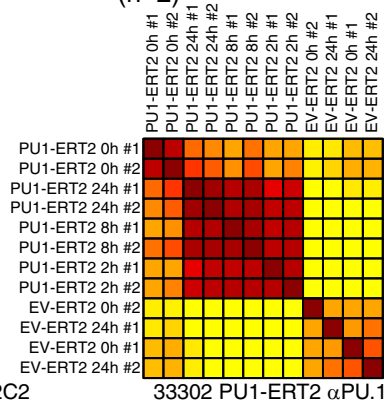
48024 PU.1 αHA and αPU.1 peaks

Scid.adh.2C2 ETS1 (n=3)



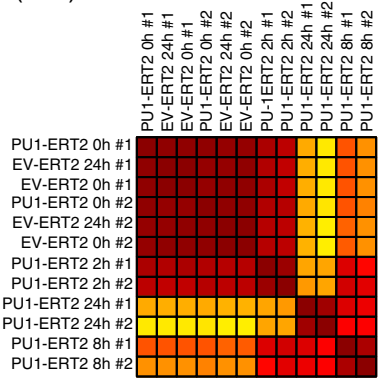
18188 Scid.adh.2C2 ETS1 peaks

PU1-ERT2 study PU.1 ChIP (n=2)



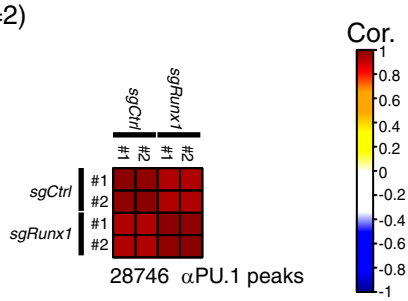
33302 PU1-ERT2 αPU.1 peaks

PU1-ERT2 H3K27ac ChIP (n=2)



Based on regions defined by 9617 PU1-ERT2 αPU.1 peaks with PU1-ERT2 H3K27ac 0 vs 24h dynamic status (read counts derived in a 1000bp region around PU.1 peaks)

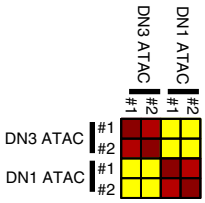
PU.1-pMxs, *Cas9* and *sgRunx1* or *sgCtrl* transduced Scid.adh.2C2 PU.1 ChIP. (n=2)



28746 αPU.1 peaks

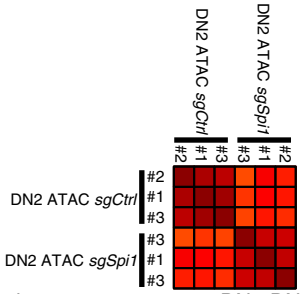
B) Correlation between ATAC-seq replicates in study

Primary ETP and DN3 ATAC (n=2)



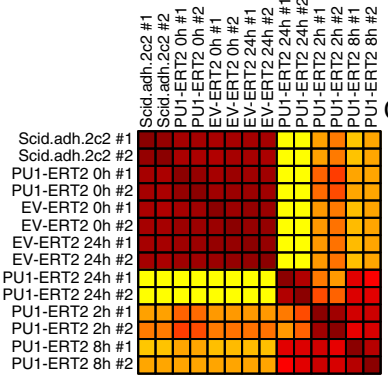
9386 DN1-DN2b PU.1-bound and DN1 to DN3 differentially ATAC-accessible peaks

BM culture with *sgSpi1* or *sgCtrl* (n=3)



1796 DN1-DN2b PU.1 bound and DN2 *sgCtrl* vs *sgSpi1* differential ATAC peaks

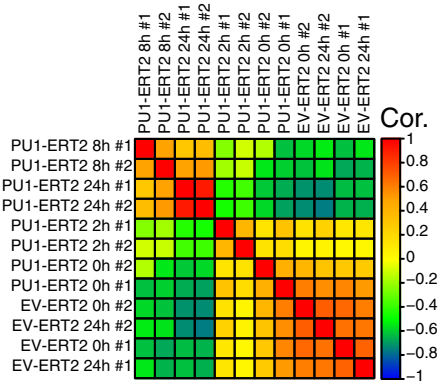
PU1-ERT2 ATAC (n=2)



Based on regions defined by 6901 PU1-ERT2 αPU1 peaks with PU1-ERT2 ATAC 0 vs 24h dynamic status (read counts derived in a 1000bp region around PU.1 peaks)

C) Correlation between RNA-seq replicates in PU1-ERT2 study

(n=2)



Based on 11812 expressed genes

Supplementary Figure S8: Correlations among results with independent replicate samples from ChIP-seq, ATAC-seq, and RNA-seq analyses

Number of replicates and correlation for ChIP, ATAC and RNA-sequencing experiments described in the study. Each comparison is divided by cell type or experiment type. Heat maps show Pearson correlation coefficients between the indicated pairs of samples analyzed for this study. **A)** Correlations among results of the ChIP-sequencing experiments in this study: PU.1, Ets1 and H3K27Ac ChIP-seq results. **B)** Correlations between results of the ATAC-sequencing experiments in the study. Only differential peaks were used for the correlation analysis, because most open ATAC peaks do not change among the sample types compared in this study. For A and B: Number of peaks and cell type for each comparison are described beneath each correlation plot. Comparisons are based on number of reads normalized to 10 million reads and read counts are summed over the entire peak width (1 kb of DNA) if not stated otherwise. Differential peaks underlying the ATAC-sequencing correlation calculations were calculated as explained in Supplemental Complete Experimental Procedures. **C)** Correlation of expressed genes between RNA-seq replicates (RPKM≥1) during the time course of RNA expression responses to PU1-ERT2 nuclearization based on all expressed genes in any of the samples (Supplementary Fig. S3B; see Fig. 3). Pearson correlation coefficients were calculated and hierarchically clustered with the R corrplot package. For correlations among primary cell RNA-seq sample data, see Supplementary Fig. S4D.