APPENDIX

Supplementary figures and methods

Inducible chromatin priming is associated with the establishment of immunological memory in T cells.

Sarah L. Bevington, Pierre Cauchy, Jason Piper, Elisabeth Bertrand, Naveen Lalli, Rebecca C. Jarvis, L. Niall Gilding, Sascha Ott, Constanze Bonifer, and Peter N. Cockerill

Data sets accompanying this manuscript:

Dataset EV1 – 2882 primed DHSs, accompanying Figure 3A.
Dataset EV2 – 1895 inducible genes, accompanying Figure 3F.
Dataset EV3 – 683 inducible genes, accompanying Figure 3G.
Dataset EV4 – 1217 inducible DHSs, accompanying Figure 4A.
Dataset EV5 – 10 motifs identified in pDHSs or iDHSs, accompanying Figures 9A and 9B.

Supplementary Figure Legends

Appendix Figure S1

Properties of CD4+ naïve, blast and memory-phenotype T cells.

(A) Human CSF2, mouse Ccl1 and mouse Fos mRNA expression in CD4+ T N stimulated with Concanavalin A for the times indicated. The level of mRNA is normalized to Beta-2-microglobulin mRNA (B2M).

(B) Representative FACS plots of CD4+ T N, CD4+ T M and CD4+ T B cells stained with antibodies to distinguish the different cell types. Cells were labeled with CD4, B220, CD44, CD62L, CD3 and CD25.

Appendix Figure S2

UCSC browser screen shot of all of the mouse genome-wide data sets defined in this study.

This window spans the Th2 cytokine gene locus which includes the Il4, Il13 and Il5 genes, and includes many other more widely expressed genes. This view shows the scales that are used in each of the other screen shots depicted in this study. The CD4+ T N H3K27ac and the three H3K27me3 tracks are from published data sets. Shown underneath are Venn diagrams showing the overlaps between the major DHS peaks identified for the replicates of the DNAse-Seq analyses of CD4+ T N, CD4+ T N+, and CD4+ T B.
Appendix Figure S3

**Influence of primed DHSs on basal and inducible levels of gene expression.**

**A and B** pDHSs do not activate basal level transcription. These plots depict the average log2 mRNA array values for CD4+ Memory T cells compared to CD4+ naïve T cells in the absence of stimulation. Data are shown for the genes closest to the 2882 genes defined as primed DHSs (panel A) and for the 1893 genes defined as being at least 2-fold induced specifically in CD4+ memory T cells (panel B).

**C** Memory T cell-specific inducible genes have pDHSs in close proximity to iDHSs. Shown here are the log2 fold increases in mRNA expression detected in response to stimulation for loci that have iDHSs located within 150 kb of a promoter of an inducible TM-specific gene. The mRNA data is plotted versus the distances between the iDHS and the nearest pDHS.

Appendix Figure S4

**Overlaps for TF ChIP peaks with pDHSs and iDHSs.**

Venn diagram overlaps of ETS-1, RUNX1 and JUNB ChIP-Seq peaks with the defined subsets of 2882 pDHSs and 1217 iDHSs. ETS-1 peaks from TN and TB cells (top); RUNX1 peaks from TN, TB and TB+ cells (bottom left); JUNB peaks from TB and TB+ cells (bottom right).

Appendix Figure S5

**Suppression of inducible gene expression in memory-phenotype cells by an inhibitor of RUNX1 function.**

**A** Flow cytometric analyses of TB prepared from CD4+ T cells cultured with 2 μg/ml ConA in the presence or absence of 20 μM Ro5-3335 for 40 hours followed by 24 hours in 50 U/ml IL-2 in the presence or absence of 20 μM Ro5-3335. Cells were labeled with FITC- Annexin V and PI. A representative FACS plot from 3 experiments is shown.

**B** mRNA expression analyses of inducible gene expression in TB prepared in the presence and absence of the inhibitor Ro5-3335 at 20 μM. Cells prepared as in (A) were left unstimulated or stimulated with PMA/I for 2 hours. mRNA expression values were normalized to the levels of B2m.

Appendix Figure S6

**Composition of pDHSs and iDHSs**

**A** TM-specific regulation of Syne3 Left: UCSC genome browser shot of the Syne3 locus showing a pDHS (black arrow) and an iDHS (red arrow). Numbers represent the distance in kilobases from the Syne3 TSS. Lower: Log2 mRNA expression levels of Syne3 in untreated and PMA/I treated TN and TM taken from the microarray analyses. Right: DNA sequences of the +9 kb pDHS and the +6.5 kb iDHS. TF binding sites are underlined. The pDHS has a 3:1 ratio of constitutive to inducible TF binding sites and the iDHS a 2:1 ratio of inducible to constitutive TF sites.

**B, C** DNA sequences of the Rad50 (B) and Cxcr3 (C) pDHSs and iDHSs. TF binding sites are underlined.
Appendix Figure S1

Properties of CD4+ naïve, blast and memory-phenotype T cells.
Appendix Figure S2

UCSC browser screen shot of all of the mouse genome-wide data sets defined in this study.
Primed DHSs reside close inducible DHSs which are linked to inducible genes.

Appendix Figure S3

Influence of primed DHSs on basal and inducible levels of gene expression.
Overlaps for TF ChIP peaks with pDHSs and iDHSs

**Appendix Figure S4**

Overlaps for TF ChIP peaks with pDHSs and iDHSs.
Appendix Figure S5

Suppression of inducible gene expression in memory-phenotype cells by an inhibitor of RUNX1 function.
Appendix Figure S6
Composition of pDHSs and iDHSs
Supplementary Methods

Mice

C42 transgenic mice containing a 130 kb AgeI genomic DNA fragment of the human IL3/CSF2 locus were described previously (Mirabella et al., 2010). These mice were backcrossed onto C57 Black 6 mice (C57BL/6J) for greater than ten generations, and the genotype confirmed as essentially all C57BL/6J, before breeding the transgene back to homozygosity.

Cell culture, purification and stimulation

CD4 or CD8 single positive T cells were isolated using MACS CD4 (L3T4) or CD8a (Ly2) microbeads. CD4+ and CD8+ single positive naïve T cells (T_N) were isolated from the spleens of C42 transgenic mice using MACS CD4+ and CD8+ T cell Isolation kits followed by enrichment for naïve cells using MACS CD62L+ microbeads according to the manufacturer’s instructions (Miltenyi Biotec, Auburn, CA). CD4 positive memory T cells (T_M) were purified from splenocytes using the Easy Sep mouse memory CD4+ T cell Isolation Kit according to the manufacturer’s instructions (Stem Cell Technologies). Actively proliferating T blast cells (T_B) were generated from purified quiescent CD4 and CD8 positive splenic cells by culturing at 5x10^6 cells/ml in IMDM with 2μg/ml Concanavalin A for 40 hours to activate TCR signaling pathways. Proliferating cells were then maintained at 5x10^5 cells/ml in 50 U/ml recombinant mouse IL-2 (Peprotech) for an additional 2-3 days before harvesting. Cells were stimulated with 20 ng/ml phorbol myristate acetate (PMA) and 2 μM Calcium Ionophore A23187 (CaI) in IMDM supplemented with 50 U/ml IL-2 at a concentration of 1x10^6 cells/ml for up to 4 hours. Jurkat T cells were cultured in RPMI supplemented with 10% fetal bovine serum.

ConA Time course

CD4+ T_N cells were cultured in IMDM at 5x10^5 cells/ml. Cells were treated with Concanavalin A for 0, 3, 8 and 16 hours before harvesting for RNA extraction, reverse transcription and qPCR analysis of gene expression.

Flow cytometry (FACS)

CD4+ T_M, CD4+ T_N and CD4+ T_B cells were labeled with CD4 PerCP-Cy5.5 (BD Biosciences 561115), B220 eFluor 450 (e-biosciences 48-0452-82), CD44 PE-Cy7 (BD Biosciences 560569), CD62L APC (e-bioscience 17-0621-81), CD3 FITC (e-biosciences 11-0251-83) and CD25 PE (e-biosciences 12-0251-83) before analysis on the CyAn ADP analyser (Beckman Coulter).

DNase I hypersensitive site analysis

DNase I digestions were performed as previously described, by adding DNase I in the presence of NP40 to intact cells and digesting for just 3 min (Bert et al., 2007). A range of DNase I concentrations were used to achieve the optimal digestion, with the amount varying depending on the cell type and the concentration of cells in the reaction. Briefly, permeabilized cells were digested in DNase I buffer (60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 100 mM Tris-HCl pH 7.4, 1 mM EGTA pH 7.4, 0.3 M sucrose, 0.2% NP40, 1 mM CaCl2 and DNase I) at 1.5x10^6 cells/ml for T_M and 1.5x10^7 cells/ml for T_N and T_B. Digestions were carried out at 22°C for 3 minutes and SDS was then added to a final concentration of 0.5% to terminate the reaction. Cell lysates were treated with 0.5 mg/ml Proteinase K at 37°C overnight followed by 0.2 mg/ml RNase A at 37°C for 1 hour. DNA was purified by phenol/chloroform extraction.
For DNase-Seq, samples were first analyzed by qPCR to identify the optimally digested samples. The ratio of PCR signal was determined between active regions which were expected to be DNase I hypersensitive and inactive regions which should not be cleaved. Optimally digested samples gave a ratio of ~0.4 to 0.7 for active/inactive. DNA fragments (~80-300 bp) were isolated from favorable samples by agarose gel extraction and further validated by qPCR analysis before library preparation. Primer pairs used to validate samples were, for the active region, TBP promoter 5' – TGCAGTCAAGAGCGCAACTG; and 5' – CACCGCTACCAGGACTCGAT; and for the inactive region, a gene desert on chr 1 5'-TGCTCCACAGTGCCATGTACA; and 5'-AGCAATTTCATGGGTAGAGAA.

Southern blot DNA hybridization analysis of human T cells was performed as previously described (Baxter et al., 2012; Mirabella et al., 2010), using either (i) BamHI and the previously used probe for the human IL3 -34 kb DHS (Baxter et al., 2012), or (ii) using EcoRI and a PCR-amplified segment of the transgene prepared using the primers AGGTGCGGTATCTGTGGATG and TGGTCCAGCTATCATGAG for the human CSF2 +30 kb DHS region.

**Chromatin Immunoprecipitation**

For the antibodies H3K27ac (ab47299) and H3K4me2 (Millipore 07-030), cells were crosslinked in PBS at 1x10⁶ cells/ml with 1% formaldehyde (~0.34 M) for 10 minutes at room temperature with rotation. For the antibodies, JUNB (sc-46), RUNX1 (ab23980), ETS1 (sc-350), and BRD4 (Bethyl laboratories), cells were re-suspended in PBS at 3.3x10⁶ cells/ml and cross-linked with 0.83 mg/ml D(N-succinimidyl) glutarate (DSG - Sigma) for 45 minutes at room temperature with rotation. Cells were washed four times with PBS and further crosslinked in PBS at 2x10⁶ cells/ml with 1% formaldehyde (~0.34 M) for 10 minutes at room temperature. All crosslinking reactions were quenched by adding 4 volumes of PBS and 0.125 M glycine, the cells were washed twice in PBS and incubated with Buffer A (10 mM HEPES pH 8.0, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100) at 4°C for 10 minutes followed by Buffer B (10 mM HEPES pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.01% Triton X-100, 0.25% SDS) at 4°C for 10 minutes. Crosslinked chromatin was sonicated using a Biorupter™ (Diagenode) in IP buffer I (25 mM Tris 1 M, pH 8.0, 150 mM NaCl, 2 mM EDTA, pH 8.0, 1% TritonX-100 and 0.25 % SDS) at a concentration of 2x10⁷ cells/ml to give fragments of an average length of ~500 bp. The supernatant was diluted with 2 volumes of IP buffer II (25 mM Tris pH 8.0, 150 mM NaCl, 2 mM EDTA, pH 8.0, 1% TritonX-100, 7.5 % glycerol) and immunoprecipitation was performed for 2 hours at 4°C using 1-2 µg of antibody coupled to 15 µl protein G Dynabeads (Dynal) per 2x10⁶ cells. Beads were washed with Buffer 1 (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, pH 8.0, 1 % TritonX-100, 0.1 % SDS), twice with Buffer 2 (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, pH 8.0, 1 % TritonX100, 0.1 % SDS), LiCl buffer (10 mM Tris-HCl, pH 8.0, 250 mM LiCl, 1 mM EDTA, pH 8.0, 0.5 % NP40, 0.5 % Na-deoxycholate) and twice with TE/NaCl buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM EDTA). DNA and protein complexes were eluted in 100 µl elution buffer (100 mM NaHCO3, 1% SDS) and the crosslinks were reversed by incubating with proteinase K at 65 °C overnight. DNA was extracted using Agencourt AMPure (Beckman Coulter) beads according to the manufacturer’s instructions and analyzed by qPCR using SYBR Green reagents on the StepOnePlus (Applied Biosystems).

**Luciferase assays**

Assays of the GM-CSF promoter plasmids were carried out using the previously described luciferase reporter gene plasmids pGM and pGM-GME which contain the vector pXPG plus the -624 to +28 bp region of the human CSF2 (GM-CSF) promoter alone (pGM) or with the -3 kb CSF2 enhancer upstream of the promoter (pGM-GME) (Bert et al., 2000; Cockerill et al., 1999). Additional plasmids were constructed from pGM plus the indicated DNA fragments upstream of the CSF2 promoter. The +28 and +30 kb DHSs downstream of CSF2
were first sub-cloned by amplifying the entire +28 to +30 kb region using PCR primers to introduce a BgIII site upstream and a KpnI site downstream of the DHSs. The upstream BgIII primer sequence TTGTGAGATCTTACAGGATTTTT contains a natural BgIII site, and the downstream primer sequence TGGCTTGTGGACCACAAAAACCAA was used to introduce a KpnI site into the sequence TGGGTGGTGTGGGGACCAGAGCCAA. This +28/30 BgIII-KpnI segment was cloned into the BgIII and KpnI sites of pGM. Additional sub-clones were prepared by either (i) deleting a BamHI segment containing the +28 kb DHS to create pGM+30, or (ii) deleting a Sall segment containing the +30 kb DHS to create pGM+28. The minor +22 kb DHS downstream of CSF2 was sub-cloned as an Xhol-digested DNA fragment into the Xhol and Smal sites of pGM after using the PCR primers TACCAATCTCGAGGAAGCAGTT and GGCGAGCAATTGTCCCGCATGG to amplify a 1.6 kb region associated with a peak of H3K4me2 in blast cells which encompassed 2 clusters of highly conserved DNA sequences.

The co-transfected control Renilla luciferase plasmid pGME-RL was derived from pGM-GME by replacing the firefly luciferase gene with the Renilla luciferase gene from the pTK-TK vector (Promega). For each assay 4.5x10⁶ Jurkat T cells were transfected by electroporation using 5 µg of the test plasmid and 1µg of pGME-RL. Cells were cultured for 24 hours before stimulation with 2 µg/ml PMA and 2 µM calcium ionophore for 8 hours before harvesting for analysis. Assays were performed using the Promega Dual-Reporter Luciferase Assay System kit according to the manufacturer’s instructions.

Assays of the IL3 promoter plasmids were carried out as above but using the previously described luciferase reporter gene plasmid pIL3H (Hawwari et al., 2002), which contains the vector pXPG plus the -559 to +50 bp region of the human IL3 promoter, co-transfected with the control plasmid pRL-TK (Promega). These assays included previously described plasmids containing the human IL3 -34, -37 or -41 kb DHSs cloned upstream of the IL-3 promoter in pIL3H (Baxter et al., 2012). The plasmid pIL3H-4.1/-1.5 includes both the -4.1 and -1.5 kb IL3 DHSs, and was made inserting the entire -4.3 to -0.17 kb region upstream of the human IL3 promoter as a BamHI-Scal fragment in between the upstream HindIII site and the -0.17 kb Scal site within the IL3 promoter of pIL3H, by performing a partial Scal/HindIII digestion of pIL3H. Assays of Ccl1 gene elements were performed as above in the context of the IL3 promoter. Mouse Ccl1 locus DNA segments were amplified from Black 6 mouse DNA by PCR, followed by restriction enzyme digestion and cloning into appropriate sites upstream of the IL3 promoter in pIL3H. PCR amplification used the primers CTGGATCCACCCCTAAGCTGG and GGAACCAAGCCTCTAGTTGCT for cloning the Ccl1 -1.3 kb region into the BgIII and Scal sites, and TACACATGAAAGATCTTGTCTTCTTCTG and GATCTTCAAGAAGGCTATGTTGACCC for cloning the Ccl1 -35 kb region into the BamHI site. One of the -35 kb region primers has a mismatch (underlined) to create a BgIII site. The -35 kb PCR product was digested with BgIII and BamHI to give compatible ends for cloning into the pIL3H BamHI site.

**Gene expression analyses by RT-qPCR**

Total RNA was isolated using TRIzol™ (Invitrogen) according to the manufacturer’s instructions. Contaminating DNA was removed using the Turbo DNA-free™ kit (Ambion) and first strand cDNA synthesis was carried out using M-MLV Reverse Transcriptase (Invitrogen) according to the manufacturer’s recommendations. Gene expression was analyzed by qPCR using SYBR Green reagents and run on the StepOnePlus (Applied Biosystems). The mouse genome PCR primers used were as follows:

**B2m:** 5′-TTCTGTGCTTGTCTCACTG 5′-CAGATGTTTCCGCTTCCCATTC
**Jun:** 5′-GGCGGAAAAGGAAGCTGAGGCGG 5′-CTGTCTTCCGTAGCATGTGGCCG
**Fos:** 5′-TCCAGGCAGACAGATCACA 5′-TTTTTCTTCTTCTTCTAGATGTG
**Junb:** 5′-TCACGACGCATCTTACGCAG 5′-CCTTGAGACCCGCATAGGGA
**Jun:** 5′-GAGGGCTGATCACCAGCTTCCCA 5′-CCTCGCTAGCTTGCACTT
Human genome primers:

IL-3: 5'-GGACTTCAAACCACTTCACTGGG 5'-TTGAATGCTCCAGGTTTGG
GM-CSF: 5'-CAGTGCTGCTGAGATGAATGAAA 5'-GTCTGTAGGCAGGTCGGCTC
GAPDH: 5'-CCTGGCCAAGGTCATCCAT 5'-AGGGGCCATCCACAGTCTT

mRNA microarray analysis and clustering

Total RNA was isolated using Trizol™ (Invitrogen) according to the manufacturer's instructions. RNA quality was determined using an RNA 6000 Pico Chip run on the Bioanalyzer 2100 system (Agilent). 100 ng of RNA was labeled with Cyanine 3-CTP using the One-Color Microarray-Based Gene Expression Analysis protocol (Low Input Quick Amp Labeling) according to manufacturer's instructions (Agilent). Labeled RNA samples were purified using RNeasy mini spin columns (Qiagen) and the concentration of cRNA was determined using a Nanodrop spectrophotometer. 600 ng of each cRNA was hybridized to the SurePrint G3 Mouse GE 8x60K Microarray kit v1 design ID 028005 (Agilent) following the Agilent hybridization protocol. After washing, microarrays were scanned on an Agilent G2565C Microarray Scanner using the Profile AgilentG3_GX_1Color for 8x60K microarrays (Dye channel: Green; Scan region: Scan Area (61 x 21.6 mm); Scan resolution (μm): 3; Tiff: 20 bit). Probe signals were extracted via the Agilent Feature Extraction software (version 10.7.1.1), protocol GE1_107_Sep09, using grid number 028005_D_GeneList_20131202 (SurePrint G3 Mouse GE 8x60K). Arrays were normalized via quantile normalization in R using the limma package (Smyth et al., 2005). Transcript annotations were aggregated into unique gene names. The mean value was computed for genes with more than one transcript. To identify common trends in differentially regulated genes, the top 1% of genes showing the highest variance of expression among datasets was selected and a Pearson correlation matrix was computed for these genes in R. Clustering of correlations was carried out using cluster 3.0 (de Hoon et al., 2004), with -e 2 -g 2 -m mas parameters. T_M-specific genes were defined as those with a CD4+ T_M +PMA/I over CD4+ T_N expression fold change > Log2 1 and a T_M +PMA/I over T_M expression fold change > Log2 1 (1895 genes). Boxplots of gene expression fold changes were plotted using R.

CRISPR deletion of the IL3 -34 kb DHS, and analyses of mutated cells

Cloning of single cells after transfection:

Plasmids were constructed using the pSpCas9(BB)-2A-GFP vector (PX458-Addgene). Single guide RNAs (sgRNA) were designed using the online CRISPR design tool (http://tools.genome-engineering.org). Guides were chosen to have minimum off-target activity as computationally predicted by the design tool. Guides were designed to the 5’ and 3’ end of the DNase I hypersensitive site located 34 kb upstream of the IL-3 gene. 34 kb 5’ CACCGTGAATTCCAGCAATGGTTCG 5’ Reverse complement: AAACCGAACATTGCTGGAATTCAC 34 kb 3’: CACCGATGAGCACCCTTCGGGCTC 3’ Reverse complement: AAACGAGCCGAAGGGTGTCTCATC.
100 μM of the sense and antisense oligonucleotides were annealed and phosphorylated with 5U T4 Polynucleotide kinase at 37°C for 30 min followed by 95°C for 5 min before cooling slowly to room temperature. The annealed oligonucleotides were ligated into the PX458 vector using the Golden Gate assembly cloning strategy where digestion and ligation occur in the same reaction. PX458 (100 ng) was mixed with 1 μM of the annealed oligonucleotides, 20 U BbsI, 10 mM ATP, 5 μg BSA and 750 U T4 DNA ligase. Reactions were incubated at 37°C for 5 minutes and 20°C for 5 minutes for 20 cycles followed by 1 cycle at 80°C for 20 minutes. 1 μl of the ligation mix was transformed into DH5α E.coli. Plasmids were prepared from positive clones and the DNA sequenced to check for the insertion.

Transfection of CRISPR reagents:

2x10⁶ Jurkat T cells were resuspended in 300 μl antibiotic free RPMI supplemented with 10% fetal bovine serum and transfected with either 5 μg of PX458 (WT) or 5 μg of the plasmids PX458 containing the 5’ and 3’ sgRNAs (-34/-). Cells were electroporated at 240 V, 950 μF and transferred to a 6 well plate in a final volume of 5 ml antibiotic free RPMI supplemented with 10% fetal bovine serum. 48 hrs post transfection, GFP positive cells were sorted into 96 well plates after which clones were expanded for approximately 14 days before screening for deletions by PCR using the primers:

-34 kb F: CCCTCCCTGGAATCAGGATTTTG
-34 kb R: ATCACAGTCCAGGGATGTTGGTCC
followed by agarose gel electrophoresis and ethidium bromide staining. Single cell cloning was performed by culturing in 40% Jurkat T cell-conditioned media.

DNase I accessibility assay:

The -34/- clones, WT clones and untransfected Jurkat T cells were stimulated for 3 hours with PMA/I before harvesting with a range of DNase I concentrations as described previously. DNA was recovered and analyzed for the level of digestion by qPCR analysis using the following primers:

TBP Pr: 5' -CTGGCCGGAATGACATTATCAA
5' -GCCAGCGGAAGCGAAGTTA
Chr18: 5' -AGGTCCCAGGACATA
IL3 -37 DHS: 5' -CAGGTCTTCTCAACACCTGCC
5' -CATGCTTGGGATGAGATGACAC

Gene expression:

The -34/- clones, WT clones and untransfected Jurkat T cells were stimulated with PMA/I for the times indicated. RNA was extracted and reverse transcribed to cDNA as previously described. Analyses of human IL3 and JUN expression used the following primers:

IL-3: 5' -GGACTTCAACAACCTCAATGGG
5' -TTGAATGCCTCCAGGTTTGG
JUN 5' -GTTTGCAACTGCTGCGTTAG
5' -CAGGTGGCACAGCTTAAACA
GAPDH 5' -CCCACTCCTCCACCTTTGAC
5' -ACCCTGTTGCTGTAGGCAAAT

RUNX1 inhibitor studies

CD4+ T cells were purified using MACs CD4 (L3T3) microbeads according to the manufacturer’s instructions (Miltenyi Biotec, Auburn, CA). Cells were cultured at 5x10⁵ cells/ml in IMDM with 2 μg/ml Concanavalin A and either 20 μM CBFβ-Runx1 Inhibitor II, Ro5-3335 (CAS 30195-30-3 – Calbiochem) or an equivalent volume of DMSO. This inhibitor is reported to inhibit RUNX1 function, but has not been shown to interact directly with RUNX1 (Cunningham et al., 2012). After 40 hours the Concanavalin A was removed and the cells were maintained in IMDM at 5x10⁵ cells/ml with 20 μM Ro5-3335 or DMSO and 50 U/ml recombinant mouse IL-2 (Peprotech) for 24 hours.

The proportion of apoptotic and dead cells was determined using the PE Annexin V
Apoptosis detection kit I (559763) according to the manufacturer's instructions. Samples were analyzed on the CyAn ADP 8 analyser (Beckman Coulter). Dead cells were removed using the MACs dead cell removal kit (Miltenyi 130-090-101) according to the manufacturer's instructions prior to stimulation for 2 hours with PMA/I and subsequent RNA extraction and gene expression analyses.

Data analysis

Alignment, peak detection and generation of coverage tracks of DNase-Seq and ChIP-Seq datasets

The high throughput DNA sequence data presented in this paper is a compilation of the following data sets generated using either Illumina or SOLiD DNA sequencing technology:

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Experiment</th>
<th>Total reads</th>
<th>Aligned reads</th>
<th>Peaks</th>
<th>Sequencing method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve CD4</td>
<td>DHS biorep1</td>
<td>38309802</td>
<td>10203874</td>
<td>23708</td>
<td>SOLID</td>
</tr>
<tr>
<td>Naïve CD4</td>
<td>DHS biorep2</td>
<td>58330158</td>
<td>46938886</td>
<td>19231</td>
<td>Illumina</td>
</tr>
<tr>
<td>Naïve CD4+PMA</td>
<td>DHS</td>
<td>40377875</td>
<td>32713929</td>
<td>13896</td>
<td>Illumina</td>
</tr>
<tr>
<td>Naïve CD4</td>
<td>H3K4me2</td>
<td>91115520</td>
<td>74261046</td>
<td>53728</td>
<td>Illumina</td>
</tr>
<tr>
<td>Naïve CD4</td>
<td>Ets1</td>
<td>134062111</td>
<td>62938297</td>
<td>26048</td>
<td>Illumina</td>
</tr>
<tr>
<td>Naïve CD4</td>
<td>Runx1</td>
<td>63099439</td>
<td>54391910</td>
<td>9520</td>
<td>Illumina</td>
</tr>
<tr>
<td>Naïve CD8</td>
<td>DHS</td>
<td>39631761</td>
<td>10378873</td>
<td>32599</td>
<td>SOLID</td>
</tr>
<tr>
<td>T-Blast CD4</td>
<td>DHS biorep1</td>
<td>364257736</td>
<td>190745253</td>
<td>62430</td>
<td>SOLID/Illumina</td>
</tr>
<tr>
<td>T-Blast CD4</td>
<td>DHS biorep2</td>
<td>52712326</td>
<td>40420460</td>
<td>36798</td>
<td>Illumina</td>
</tr>
<tr>
<td>T-Blast CD4</td>
<td>H3K4me2</td>
<td>257985861</td>
<td>85122187</td>
<td>49567</td>
<td>SOLID/Illumina</td>
</tr>
<tr>
<td>T-Blast CD4</td>
<td>Ets1</td>
<td>148169018</td>
<td>103284815</td>
<td>44781</td>
<td>Illumina</td>
</tr>
<tr>
<td>T-Blast CD4</td>
<td>H3K27Ac</td>
<td>66001536</td>
<td>57001599</td>
<td>24357</td>
<td>Illumina</td>
</tr>
<tr>
<td>T-Blast CD4</td>
<td>Runx1</td>
<td>79060600</td>
<td>59829702</td>
<td>26990</td>
<td>Illumina</td>
</tr>
<tr>
<td>T-Blast CD4</td>
<td>JunB</td>
<td>112830344</td>
<td>53757476</td>
<td>13945</td>
<td>Illumina</td>
</tr>
<tr>
<td>T-Blast CD4</td>
<td>BRD4</td>
<td>38054961</td>
<td>30823748</td>
<td>28063</td>
<td>Illumina</td>
</tr>
<tr>
<td>T-Blast CD8</td>
<td>DHS</td>
<td>196310782</td>
<td>44891153</td>
<td>75724</td>
<td>SOLID</td>
</tr>
<tr>
<td>T-Blast+PMA CD4</td>
<td>DHS</td>
<td>176497317</td>
<td>145118657</td>
<td>62018</td>
<td>Illumina</td>
</tr>
<tr>
<td>T-Blast+PMA CD4</td>
<td>Runx1</td>
<td>39010278</td>
<td>31240495</td>
<td>28143</td>
<td>Illumina</td>
</tr>
<tr>
<td>T-Blast+PMA CD4</td>
<td>JunB</td>
<td>48380238</td>
<td>37298812</td>
<td>45301</td>
<td>Illumina</td>
</tr>
<tr>
<td>T-Blast+PMA CD4</td>
<td>H3K4me2</td>
<td>67439153</td>
<td>59018996</td>
<td>24640</td>
<td>Illumina</td>
</tr>
<tr>
<td>T-Blast+PMA CD4</td>
<td>H3K27Ac</td>
<td>51931052</td>
<td>42625817</td>
<td>31357</td>
<td>Illumina</td>
</tr>
<tr>
<td>T-Blast+PMA CD4</td>
<td>BRD4</td>
<td>45643715</td>
<td>36928604</td>
<td>35839</td>
<td>Illumina</td>
</tr>
<tr>
<td>Memory CD4</td>
<td>DHS</td>
<td>257065491</td>
<td>190457076</td>
<td>60794</td>
<td>Illumina</td>
</tr>
<tr>
<td>Memory+PMA CD4</td>
<td>DHS</td>
<td>217398014</td>
<td>149031692</td>
<td>63041</td>
<td>Illumina</td>
</tr>
</tbody>
</table>

DNA sequence reads were aligned to the July 2007 mouse genome data from the Build 37 assembly by NCBI and the Mouse Genome Sequencing Consortium (mm9) using the bowtie aligner (Langmead et al., 2009) with --all --best --strata --v 2 -m 1 as parameters. For the SOLID reads, the colorspace version of the mm9 genome was used, and reads were aligned as color fasta using the same parameters. For datasets split into several runs, and all datasets were aligned together using bowtie. In all cases, outputs were in the SAM format as instructed to bowtie using the -S switch, converted to BAM, sorted then indexed using the view, sort and index functions of the samtools package (Li et al., 2009). For the human
IL-3/GM-CSF transgene reads were aligned to a custom-built index of the human genome covering chromosome 5 of the March 2006 human genome data from the Build 36 assembly by NCBI and the Human Genome Sequencing Consortium (hg18), using the same parameters and methodology. This index was built using the bowtie-build function of the bowtie package, using the FASTA sequence of human chromosome 5 from the hg18 genome. For datasets requiring merging of Illumina and SOLiD outputs, sorted BAM files were merged using the merge function of the samtools package. Peak detection, tag extension and generation of coverage tracks was subsequently performed using the macs14 package (Zhang et al., 2008). In all cases, peak detection was performed with bowtie’s built-in calculation of the maximum number of tags at one location based on a binomial distribution, using the --keep-dup=auto switch. For datasets spanning less than one lane, this switch was used to compute coverages using the -w and -S switches. In the case of datasets spanning several lanes, the --keep-dup=all switch was used to compute coverages. Genome sizes were adjusted depending on whether the reference genome was mouse or human, using the -g mm or -g hs switches, respectively.

**Unions of data from CD4\(^+\) T\(_N\), T\(_B\), T\(_B^+\) and T\(_M\) T cells**

To compare sequence tag counts at two stages for common and specific regions, we generated the unions of the peaks. Due to the intrinsic imprecision of DNaseI fragment lengths, this can result in non-overlapping, redundant summit coordinates. To obtain non-redundant summit populations, pairwise unions were generated by concatenating, sorting then merging peak summits within 400 bp using sort and merge functions of the bedtools package (Quinlan and Hall, 2010). For T\(_N\) and T\(_M\) summits, a tag cut-off of 50 and 20 was used to discard ambiguous peaks. Tag counts were retrieved ±1 Kb around the union of summits from coverage files using the annotatePeaks function of the Homer package (Heinz et al., 2010) using -hist 10 -ghist -wig -size 2000 as parameters, summed up ±200 bp around the summit to retrieve tags corresponding to the length of a DHS, then subsequently ordered by Log\(_2\) fold change, whereby totals of value 0 were replaced by 1 in order to eliminate infinite values. A BED file representing summits sorted by Log\(_2\) fold change was created and corresponding DHS and ChIP tracks (e.g. T\(_M\) alongside the CD4\(^+\) T\(_B^+\)/T\(_N\) fold change axis) were retrieved using annotatePeaks using the same parameters, however sorted according to the Log\(_2\) fold change-sorted input BED file. Heatmap images were generated using Java Treeview (Saldanha, 2004).

**Definition of pDHSs, iDHSs and average profile generation**

To define the pDHSs, DHSs enriched more than 3-fold in the CD4\(^+\) T\(_M\) over T\(_N\) union were selected (3054 DHSs). From the 3054 T\(_M\)-specific DHSs, regions which were also called as peaks in the CD4\(^+\) T\(_B\), were defined as the pDHSs (2882 DHSs). For the inducible sites, 6823 iDHSs were enriched more than 5.5-fold in the T\(_B^+\) over T\(_B\) union. The 1217 strongest iDHSs were enriched more than 11-fold. To compute average DHS and ChIP profiles ±2 Kb around the pDHSs or iDHSs, the annotatePeaks function of the Homer package was used with -size 4000 -hist 10 -wig as parameters.

**Normalization of ChIP-Seq and DNase-Seq experiments**

For genome coverage tracks, scales were adjusted to control genes by visual inspection (TBP, ACTB, GAPDH, enhancers of the TCR \(\alpha\) and \(\beta\) for DHS and histone modification tracks, enhancers of the TCR \(\alpha\) and \(\beta\) for Ets1, Runx1 and Fli1-Ets1, Runx1 enhancers, respectively, and Fos, Jun promoters for JunB) to reflect signal to noise ratios rather than tag count normalization. For average profiles, scaling was performed by retrieving coverages in 10,000 randomly selected, 4000 bp wide regions in order to compute background levels, so that values of average profiles were scaled proportionally to the ratio of average background levels of all datasets with respect to CD4\(^+\) T\(_B\). For heatmaps, contrasts proportional to the numbers of reads divided by the number of lines in heatmaps were used.
Distance and gene expression analyses

pDHSSs or iDHSSs closest to the 1895 TM-specific genes were retrieved using the bedtools slop function. A RefSeq annotation of TSSs from the mm9 genome was used as a BED file. The distance to the TSS was obtained using the -D switch, which calculates the distance of features relative to their orientation. Genes were grouped according to distance from the pDHSS or iDHSS to the TSS of the TM specific genes in 25 kb intervals. The median fold change of gene expression in TM+PMA/I compared to TM was plotted relative to the distance. Genes were also grouped according to fold induction of expression in TM. The median distance to the closest iDHSS and pDHSS was calculated for each group. For the distance between iDHSSs and pDHSSs, the closest pDHSS was determined as outlined above and the distances were calculated using the coordinates of the nearest iDHSS. p-values were calculated via a χ² test using the random expectation value of each comparison as the expected value, vs the observed value. The empirical calculation using base-pair probabilities was performed as in the following example:

To calculate the random expectation value of 312 pDHSSs out of 683 falling within [-25 Kb, +25 Kb] of 6,823 iDHSS, the size of the base-space of iDHSSs +/- 25 Kb was first computed: there were a total of 6,823 iDHSSs thus corresponding to 6,823x2x25,000=341,150,000 bp. The size of the mm9 genome being 2,725,765,481 bp, the probability of encountering an element at random in this base-space is thus 341,150,000/2,725,765,481=0.125157502. The random expectation value of finding 312 elements out of 683 in this base-space is thus the number of these elements times the probability of encountering an element at random in this base-space or 683x0.125157502=85.48257421 pDHSSs. Calculations were then performed using the reciprocal population as the empirical reference (e.g. using the base-space of 683 pDHSSs as the reference and the probability of 6,823 iDHSSs falling within [-25 Kb, +25 Kb] of those).

Since these calculations yield one point with no standard deviation (the expected intersections), a χ² test was used to calculate the p-value with k=1 degree of freedom. χ² was calculated as (Observed – Expected)²/Expected. P-values were computed by obtaining the corresponding value of the computed χ² value in the χ² probability density distribution (function CHISQ.DIST.RT in Excel). The following example depicts how the p-value of finding 312 pDHSSs falling within [-25 Kb, +25 Kb] of 6,823 iDHSS while the random expectation value was of 85.48257421:

χ²=(OBSERVED – EXPECTED)²/EXPECTED
χ²=(312-85.48257421)²/312
χ²= 600.2409809
p(χ²)= CHISQ.DIST.RT(χ², k)= \frac{1}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}

where k=1 degree of freedom and \( \Gamma(n) = (n - 1)! \).

p(χ²)= 1.4838E-132

Random-expectation values were checked by simulating the background model using bootstrapping, using 10,000 iterations of intersections of random coordinates (generated via bedtools random using the system command in R) of size equal to one group vs the other actual group and subsequently swapped. The generated values were comparable to the calculated random-expectation values (e.g. 85.6 with σ=11.8 for the above example, yielding a Z-score Z=(X-µ)/σ of 19.119, corresponding to a p-value of 1.67 E-80 using the NORM.S.DIST function in Excel).
For comparisons that involved classes of gene expression fold change and distance of iDHSs to the TSS of an inducible gene or distance between pDHSs to the nearest iDHS around inducible genes (Fig 4F, H), p-values were calculated via t-tests using similarly-sized populations of random DHSs. For classes of expression fold change versus distance of iDHSs to the TSS of an inducible gene (Fig 4F), we selected 6823 DHSs at random while using the same inducible genes. For classes of expression fold change versus distance between pDHSs to the nearest iDHS around inducible genes (Fig 4H), we selected 6823 DHSs overlapping with no pDHSs. T-tests were carried out on all distance values for each class of fold change, between values for pDHSs and values for random DHSs, using a two-tail distribution.

Motif discovery, heatmaps and tables

De novo motif discovery was performed using the findMotifsGenome function of the Homer package. For motif heatmaps, top-enriched position weight matrices (PWM) identified via de novo motif discovery were matched against regions spanning ±1 Kb around summits of previously defined and sorted unions, using the annotatePeaks function of the Homer package, with -size 2000 -m as parameters. For consistency, the same, non-redundant matrices were used in all comparisons, with the ones showing the lowest p-value across datasets retained where applicable. Heatmap images were generated via Java Treeview. For the motif tables, the non-redundant matrices were matched to regions spanning ±200 bp around summit unions and the total number of occurrences for each motif was calculated. Motifs were then split into two classes: constitutive (RUNX, ETS, KLF, GATA, E-box) and inducible (AP-1, NFAT, EGR, NF-κB, CREB/ATF)

Motif co-occurrence clustering

In order to estimate whether pairs of motifs co-localized significantly in a given population of sequences, an enrichment Z-score was computed between actual motif intersection counts against 1,000 randomly selected sequences from the merged transcriptionally active space of T-cells corresponding to the scope of study as a background. For the enrichment of co-localization of motifs in pDHSs and iDHSs, this corresponded to the unions of DHSs from naïve, blast, memory T-cells and blast, stimulated blast CD4+ T-cells, respectively. A custom script performing motif annotation, intersection in actual populations as well as in randomly selected regions of the active space and enrichment calculation was written in the R statistical language. Background co-occurrence was thus estimated using bootstrapping (1,000 repetitions) of motif mapping and co-association counts (within 50bp) in randomly selected regions within the background, using equally sized populations as the original number of peaks (2882 for pDHSs and 1217 for iDHSs). Annotation of PWMs to sequences of pDHSs and iDHSs was performed using Homer annotatePeaks, using -size 400 -m -mbed as parameters. For the sake of consistency across datasets, built-in PWMs from Homer most closely related to PWMs enriched in de novo motif discovery were used. Resulting BED outputs were extended by 50bp using the slop function of the bedtools package. To obtain motif co-occurrence, extended motif BED files were all intersected using the intersection_matrix function of the pybedtools package (Dale et al., 2011). Enrichment was assessed by calculating a Z-score, which was represented by Z=(X-µ)/σ, whereby µ and σ would represent background average co-occurrences and standard deviation of average, background co-occurrences. Z-scores were clustered via cluster 3.0 using -e 2 -g 2 -m m as parameters. Heatmaps images were generated using Java TreeView.

Digital genomic footprinting

The CD4+ T8, T8+ and T8 datasets had >140,000,000 mouse-aligned read counts and signal to noise ratios high enough to undergo digital genomic footprinting. For this purpose, the wellington_footprints function of the Wellington suite (Piper et al., 2013) was used with standard parameters (p-value score, threshold 10^-20) on BED3-converted peaks. Briefly,
Wellington computes scores for each possible position of a footprint and for a number of different footprint sizes. The scores reflect the enrichment of positive strand reads starting near the 5’ end of a footprint and the enrichment of negative strand reads starting near the 3’ end of a footprint. To retrieve DNaseI cuts in the DHSs, the `dnase_wig_tracks` function of the `Wellington` suite was used with standard parameters on BED3-converted peaks. To compute average cut profiles and matrices on foot-printed motifs, PWMs were first matched to foot-printed regions using the `annotatePeaks` function of the Homer package, using -size given -m -mbed and thus output as BED files. Average cut profiles and strand imbalance matrices were generated using the `dnase_average_profiles` and `dnase_to_javatreeview` functions of `Wellington` using standard parameters (sort by footprint occupancy score for matrices) on BED3-converted footprinted motif files. For heatmaps of matrices corresponding to strand imbalance, `Java TreeView` was used for the generation of images. A Perl script was written so as to perform these steps automatically for all motifs, using `Java Treeview` from the command line using -f -x --o -s as parameters, whereby a scaling factor was computed so that all heatmaps were of equal height (600 pixels), with width being constant. Footprint Occupancy Scores (FOSs) are computed as the ratio of cuts observed within a binding site to those immediately outside (Neph et al., 2012). To retrieve FOSs, a development Python script of the `Wellington` suite (`wellington_score_heatmap.py`) was used, which outputs FOSs per input interval in an exploitable text format. To compute significant differences in overall occupancy for AP-1 motifs, a paired t-test was performed using FOSs at AP-1 motifs in TB and TB+. The p-value was retrieved using the two-tailed T distribution.

For heatmaps of footprints in TM cells containing ETS and RUNX motifs, TM footprints were first annotated with ETS and RUNX motifs using `bedtools annotatePeaks`, using -m -mbed as parameters. The resulting files were intersected with all TM footprints to reflect the ones containing these motifs and converted to bedGraph files. The presence of these motif-containing footprints was assessed by using `bedtools annotatePeaks` using the TM/TN DHS union as input and the motif containing footprints as the bedGraph coverage files, with -hist 10 –ghist -bedGraph as parameters. Heatmap images were generated using `Java TreeView`.

**Annotation of genomic features and gene ontology analyses**

In order to determine the genomic location of the DHSs (intergenic, intragenic, promoter, exonic, etc.) the `annotatePeaks` function of the `Homer` package was used with default options, with outputs redirected to text files. Numbers of occurrences for each genomic feature type were derived using the `DCOUNTA` function in Excel and plotted as percentages of total DHSs as piecharts. For gene ontology analyses, `annotatePeaks` was used using the -go switch, returning the closest transcription start site (TSS) and performing gene ontology analysis based on these. For gene ontology analyses of expression microarray results, the TSS of genes was used as input.

**Supplementary References**


