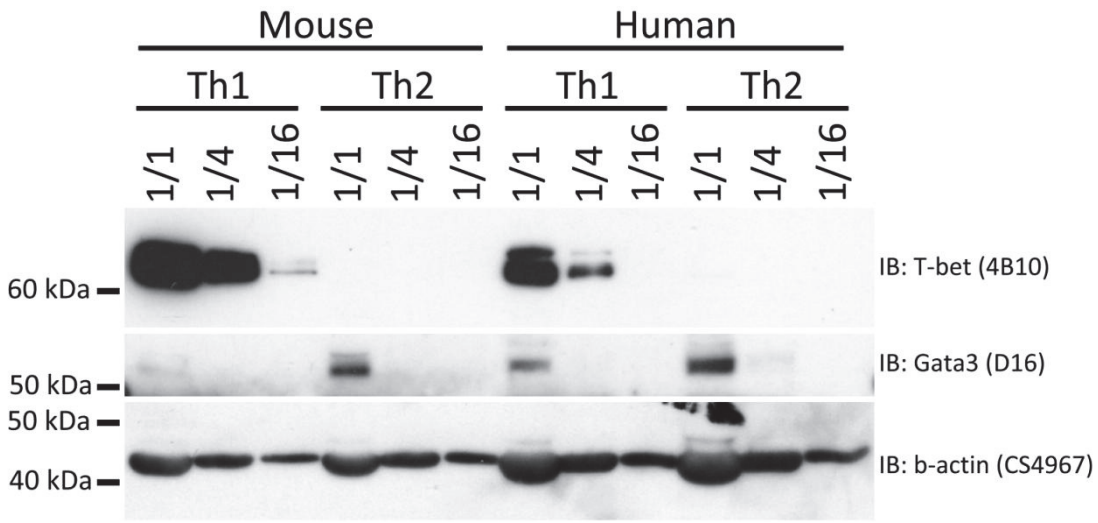
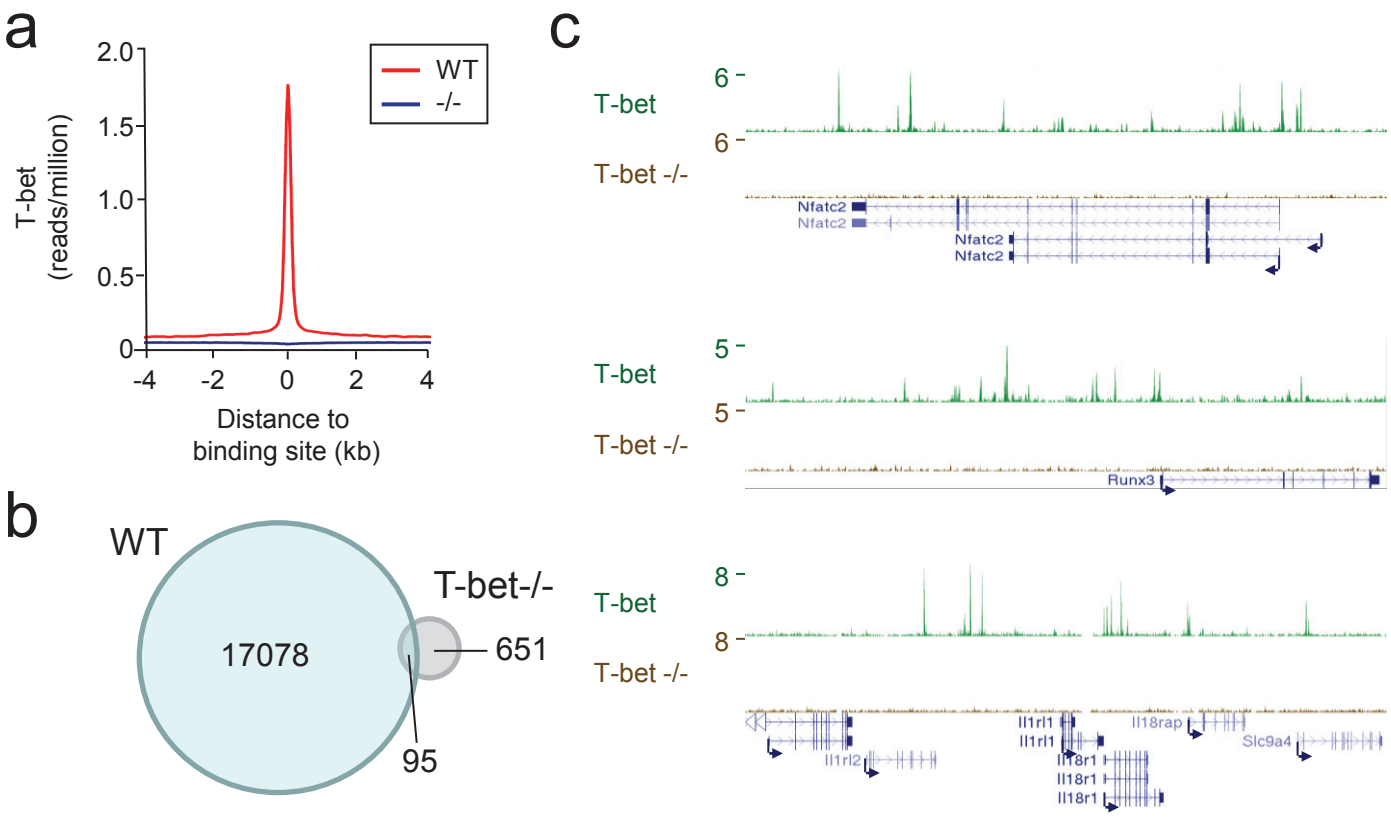


Supplementary Figure S1. T-bet and GATA3 expression in human and mouse Th1 and Th2 cells.

Immunoblotting for T-bet and Gata3 in serial dilutions of lysate from human and mouse Th1 and Th2 cells polarized in vitro from naïve precursors. The membrane was probed with anti-Gata3 antibody (D16, Santa Cruz), then T-bet (4B10, Santa Cruz) and then stripped and reprobed for β-actin (CS4967, Cell Signaling), which served as a loading control.



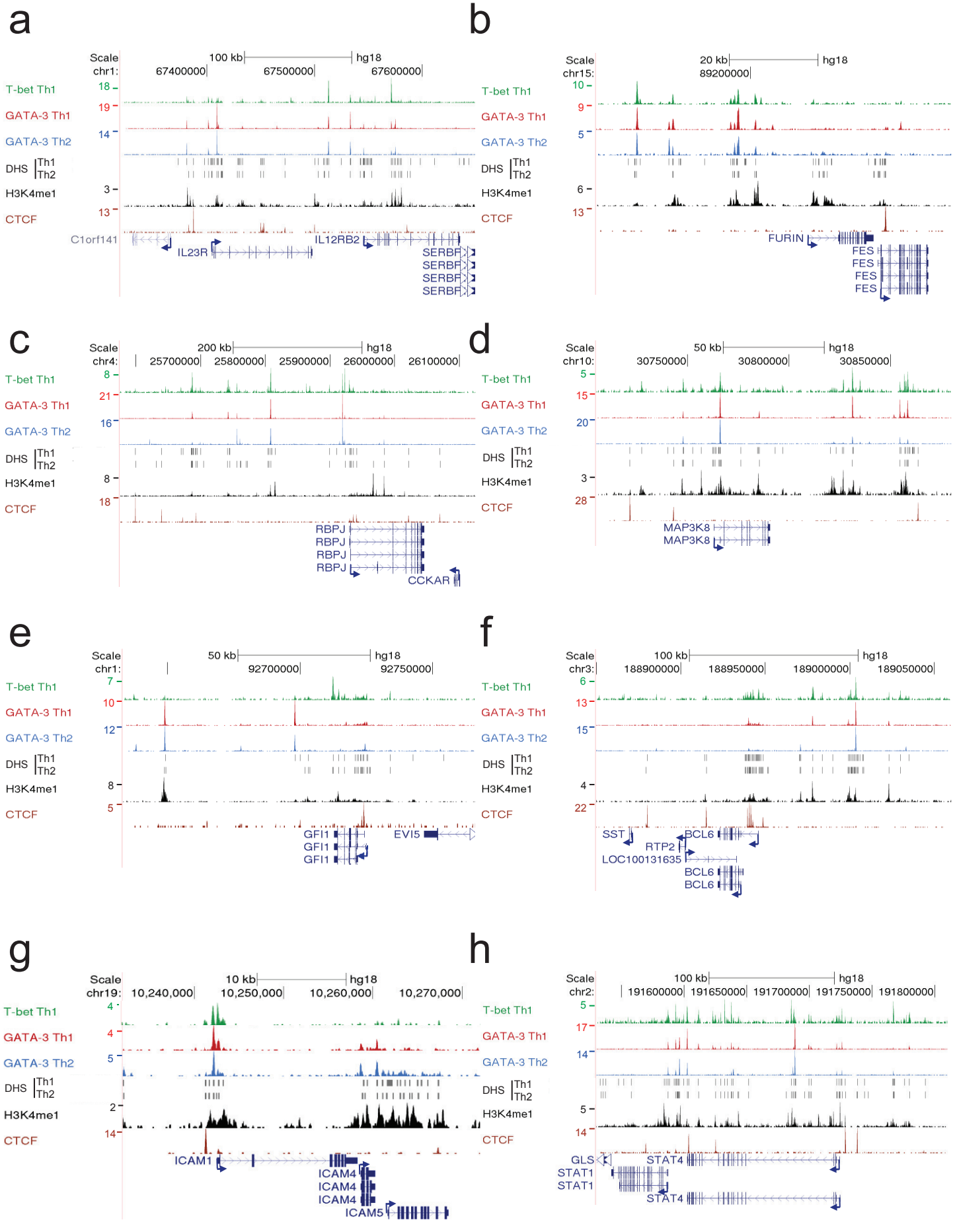
Supplementary Figure S2. ChIP-Seq for T-bet in Th1 cells cultured from wild-type and T-bet-null mice.

A. Average number of sequence reads from T-bet ChIP DNA (y-axis) from wild-type (red) and T-bet -/- (blue) mouse cells, plotted against the genomic distance from T-bet binding sites in wild-type cells.

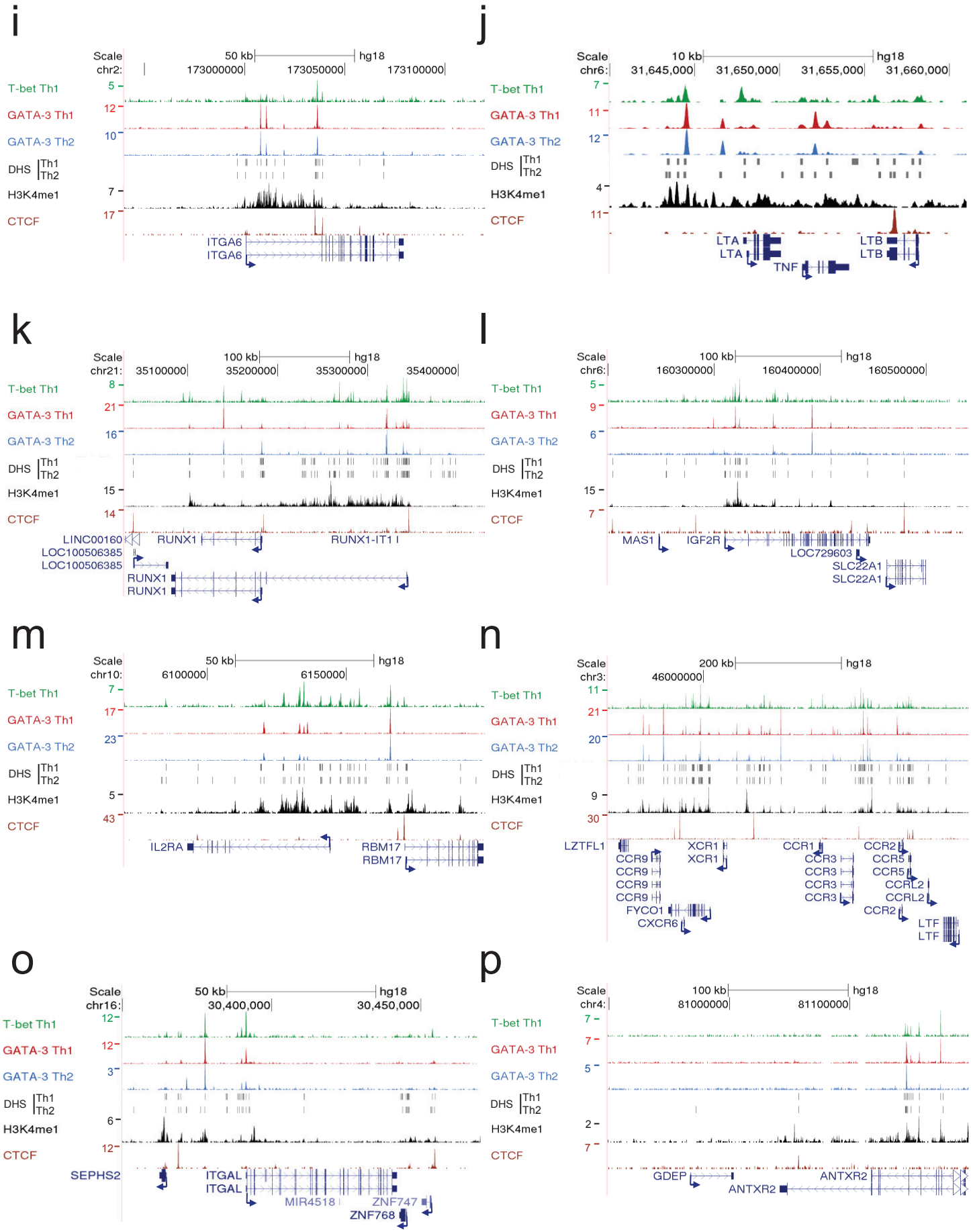
B. Venn diagram showing the overlap between the number of T-bet binding sites identified by MACS ($p<10^{-6}$) for T-bet in wild-type cells versus T-bet -/- cells. This shows that 0.6% of binding sites identified in wild-type cells are false-positives.

C. T-bet binding around the mouse genes Nfatc2 (region shown chr2:168,273,400-168,444,400), Runx3 (chr4:134,568,000-134,735) and the Il1r1, Il1rl2, Il1rl1, Il18r1, Il18rap cluster (chr1:40,309,000-40,690,000). The number of sequencing reads from T-bet ChIP-enriched DNA are plotted per million background-subtracted total reads and aligned with the mouse genome. Data for T-bet ChIP from T-bet-/- cells are shown for comparison. Gene structures are marked at the bottom of the figure and the start and direction of transcription by the blue arrow.

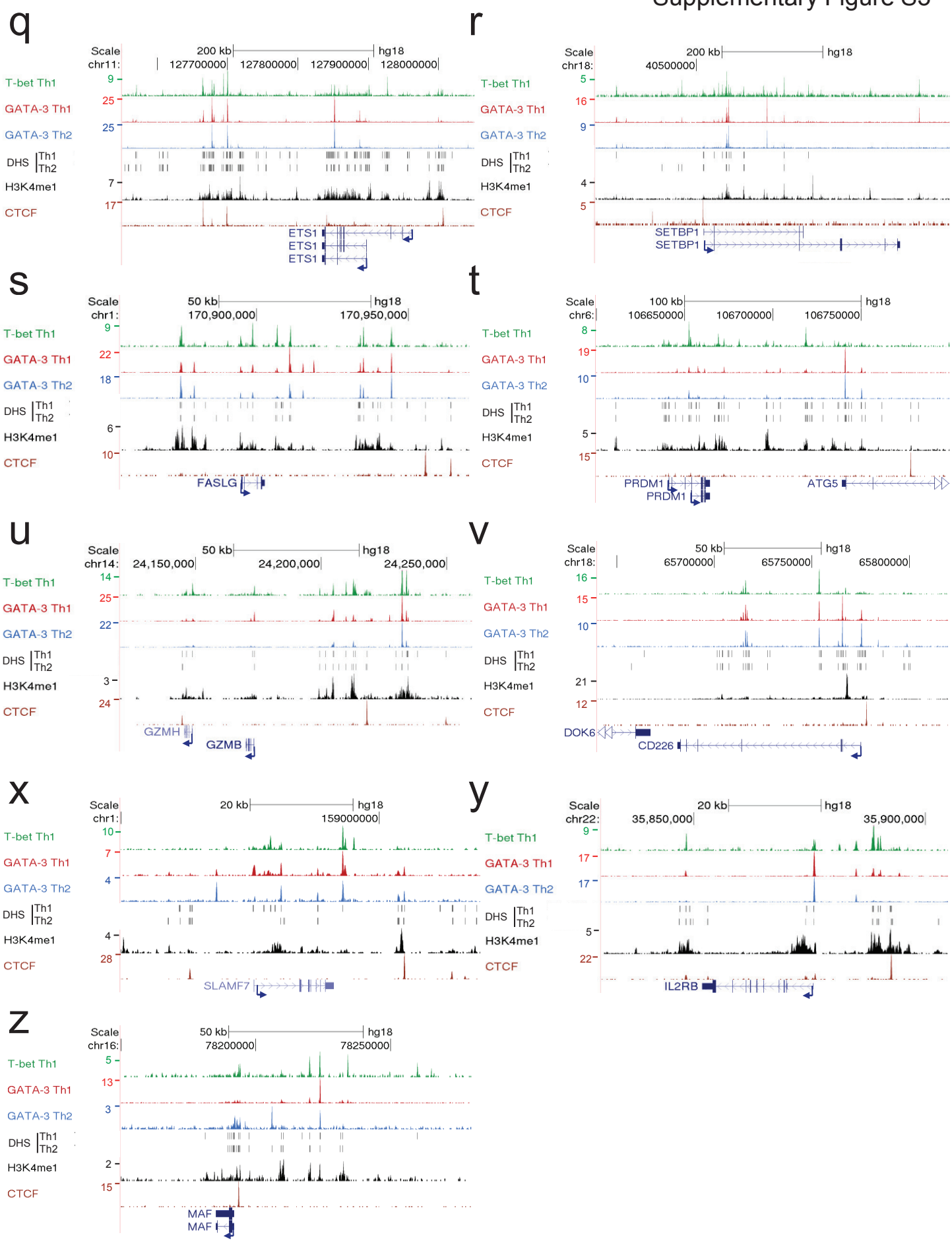
Supplementary Figure S3



Supplementary Figure S3

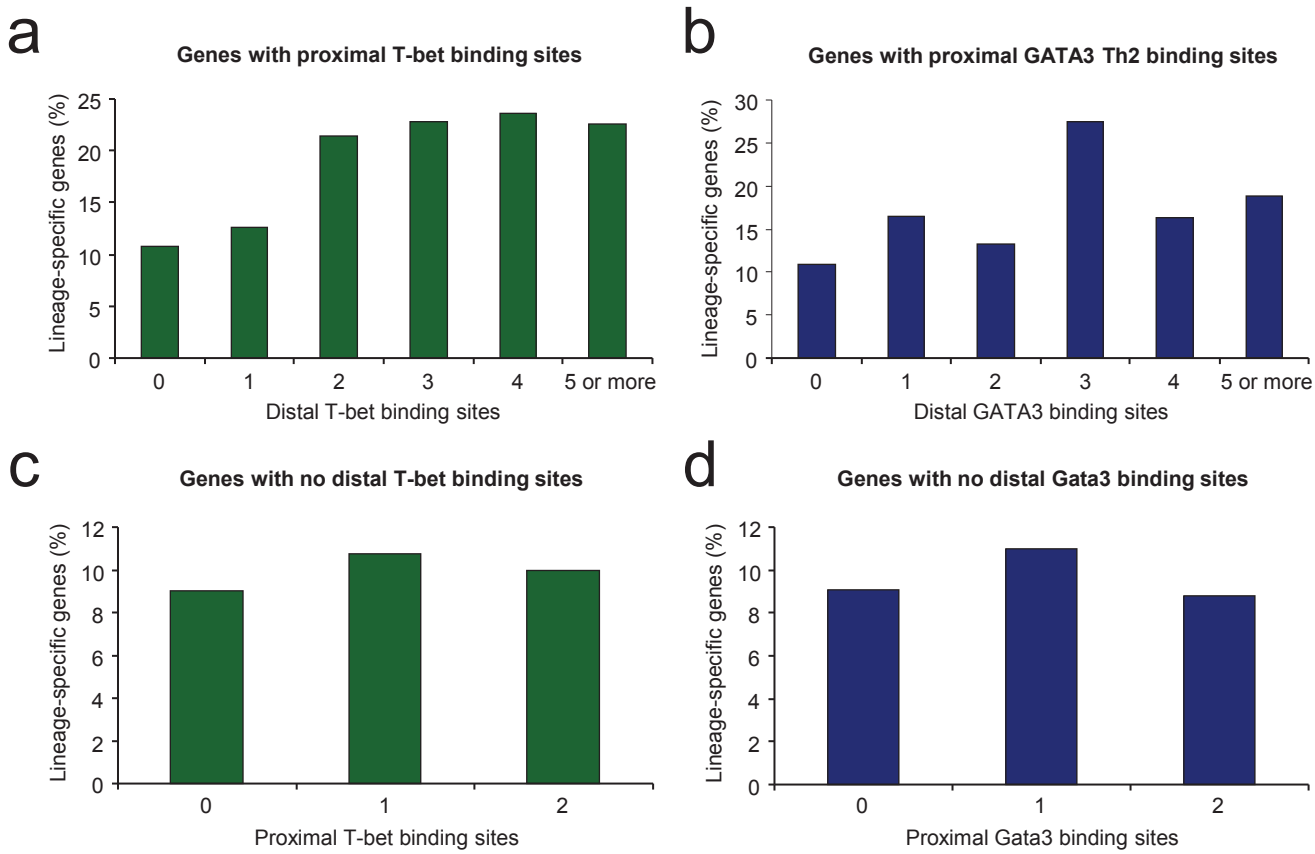


Supplementary Figure S3



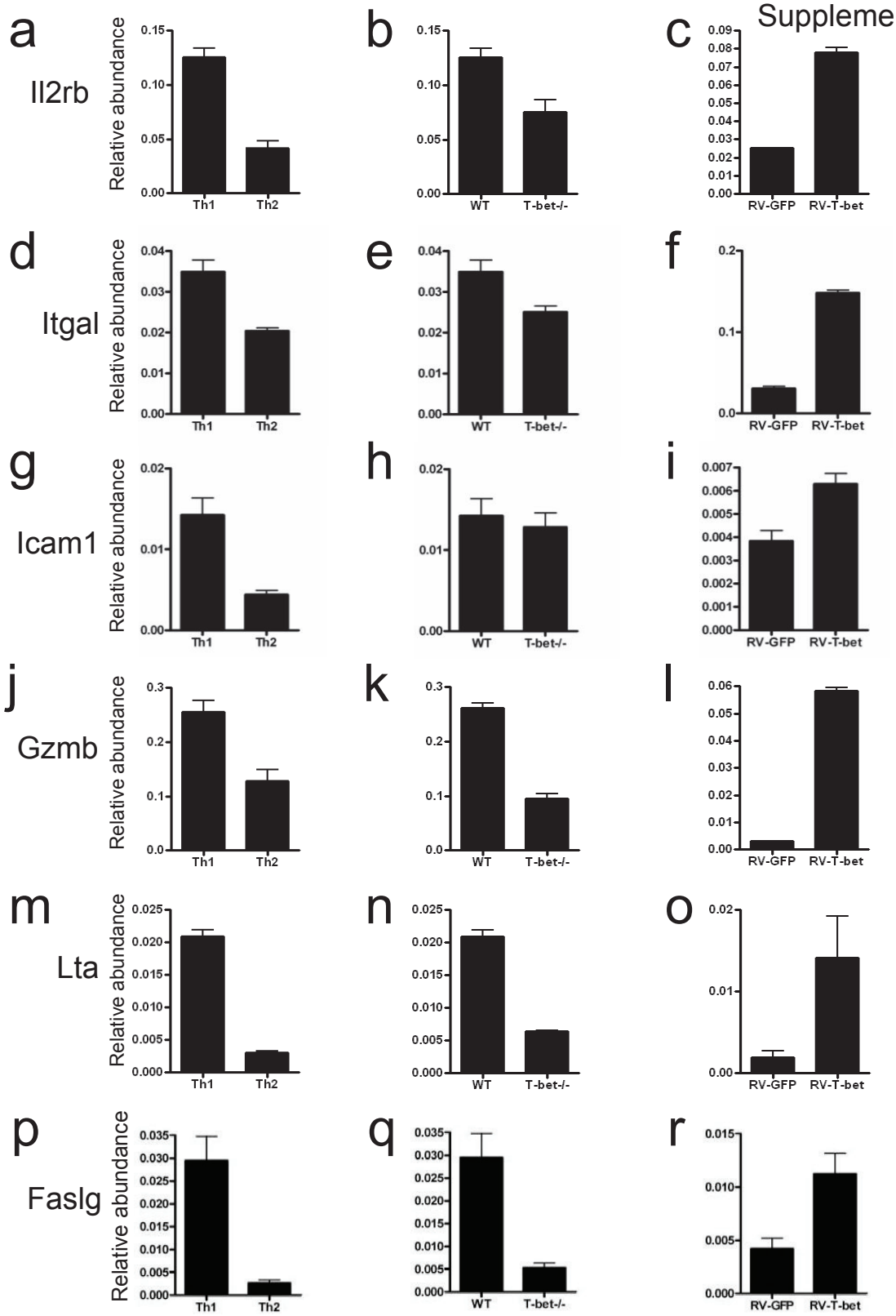
Supplementary Figure S3. Examples of genes with extensive distal T-bet and GATA3 binding.

A. T-bet and GATA3 binding around IL23R and IL12RB2 (region shown chr1:67,320,000-67,650,000). The number of sequencing reads from T-bet and GATA3 ChIP-enriched DNA are plotted per million background-subtracted total reads and aligned with the human genome. The position of DNaseI HS sites in Th1 and Th2 cells are marked by dashes and sites of H3K4me1 and CTCF occupancy by ChIP-Seq reads per million. Gene structures are marked at the bottom of the figure and the start and direction of transcription by the blue arrow.
B. As A., except for FURIN (chr15:89,165,000-89,245,000).
C. As A., except for RBPJ (chr4:25,575,000-26,125,000).
D. As A., except for MAP3K8 (chr10:30,705,000-30,880,000).
E. As A., except for GFI1 (chr1:92,633,700-92,766,600).
F. As A., except for BCL6 (chr3:188,850,000-189,060,000).
G. As A., except for ICAM1 (chr19: 10,232,000-10,272,000).
H. As A., except for STAT1 and STAT4 (chr2:191,530,000-191,810,000).
I. As A., except for ITAG6 (chr2:172,940,000-173,120,000).
J. As A., except for LTA (chr6: 31,640,000-31,661,000).
K. As A., except for RUNX1 (chr21:35,030,000-35,430,000).
L. As A., except for IGF2R (chr6:160,200,000-160,540,000).
M. As A., except for IL2RA (chr:10:6,070,000-6,200,000).
N. As A., except for CCR9, FYCO1, CXCR6, XCR1, CCR1, CCR3, CCR2, CCR5 and CCRL2 (chr3:45,820,000-46,500,000).
O. As A., except for ITGAL (chr16: 30,350,000-30,470,000).
P. As A., except for ANTXR2 (chr4:81,000,000-81,300,000).
Q. As A., except for ETS1 (chr11:127,550,000-128,065,000).
R. As A., except for SETBP1 (chr18:40,300,000-41,000,000).
S. As A, except for FASLG (chr1:170,855,000-170,970,000).
T. As A., except for PRDM1 (chr6:106,600,000-106,800,000).
U. As A., except for GZMB (chr14:24,120,000-24,260,000).
V. As A., except for CD226 (chr18:65,640,000-65,820,000).
X. As A., except for SLAMF7 (chr1:158,950,000-159,020,000).
Y. As A., except for IL2RB (chr22: 35,830,000-35,905,000).
Z. As A., except for MAF (chr16:78,150,000-78,280,000).

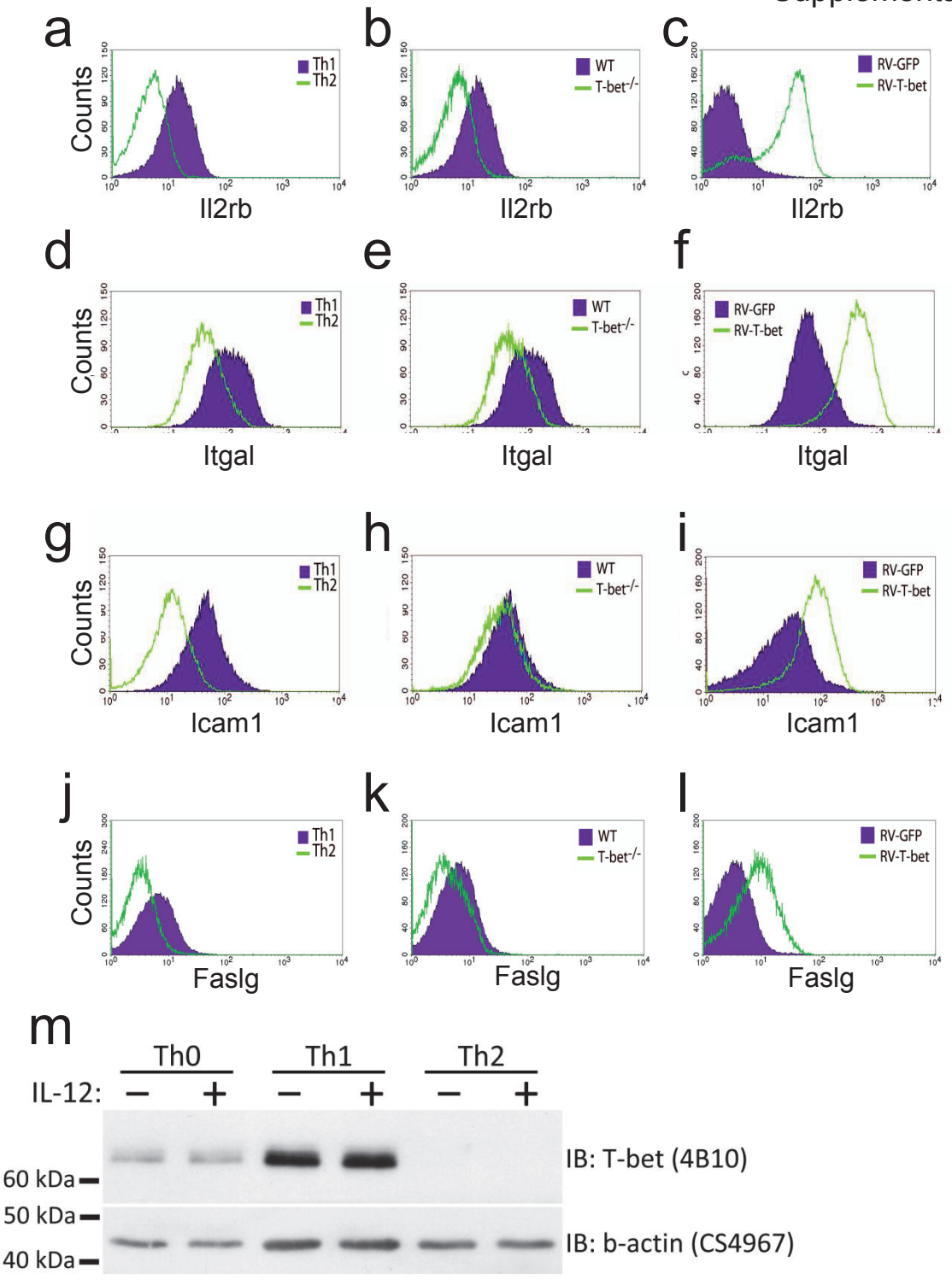


Supplementary Figure S4. Assessment of the relationship between the total number of T-bet and GATA3 binding sites and lineage-specific gene expression.
A. The proportion of human genes bound proximally by T-bet that show lineage-specific expression as a function of the total number of distal T-bet binding sites. There is not a proportional relationship between the number of binding sites and lineage-specificity.
B. As B., except for Th2 GATA3 binding sites.
C. The proportion of human genes that show lineage-specific expression as a function of the total number of proximal T-bet binding sites. Genes are filtered for those not bound distally by T-bet to observe the relationship between lineage-specific expression and the number of proximal binding sites alone.
D. As C., except for Th2 GATA3 binding sites.

Supplementary Figure S5



Supplementary Figure S5. Direct regulation of gene expression by T-bet.
Quantitative PCR measurements of Il2rb (**A-C**), Itgal (**D-F**), Icam1 (**G-I**), Gzmb (**J-L**), Lta (**M-O**) and Faslg (**P-R**) mRNA abundance relative to Actb mRNA abundance in in vitro differentiated mouse Th1 and Th2 cells (**A,D,G,J,M,P**), Th1 cells from wild-type mice and T-bet-/- mice (**B,E,H,K,N,Q**) and activated CD4+ T cells from T-bet-/- Ifnγ-/- mice transduced with retroviruses encoding GFP (RV-GFP) or T-bet (RV-T-bet) (**C,F,I,L,O,R**).



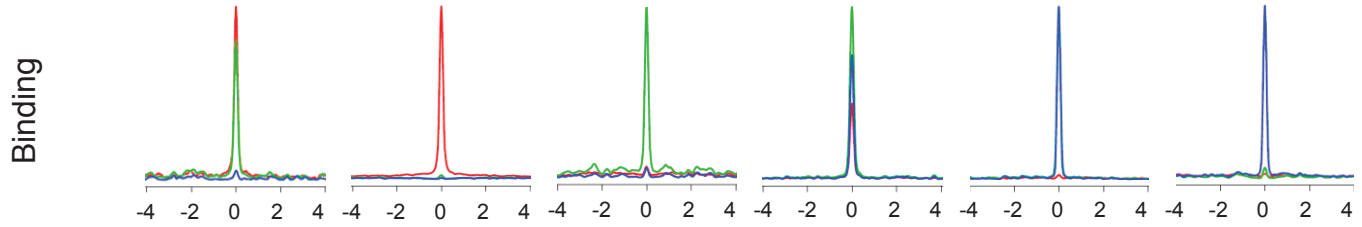
Supplementary Figure S6. Direct regulation of gene expression by T-bet leads to changes in the levels of protein products.

A-L. Il2rb (A-C), Itgal (D-F), Icam1 (G-I) and Faslg (J-K) protein expression measured by flow cytometry in *in vitro* differentiated mouse Th1 and Th2 cells (A,D,G, J), Th1 cells from wild-type mice and T-bet$^{-/-}$ mice (B,E,H, K) and activated CD4+ T cells from T-bet$^{-/-}$ Ifng$^{-/-}$ mice transduced with retroviruses encoding GFP (RV-GFP) or T-bet (RV-T-bet) (C,F,I,L).

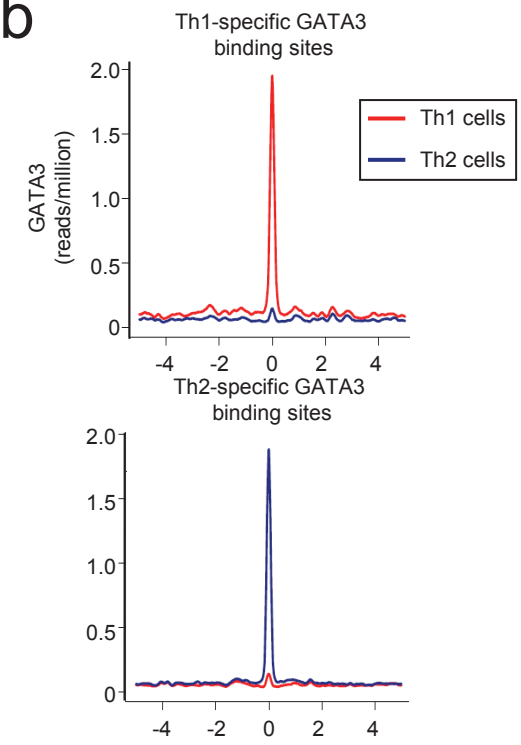
M. T-bet is expressed in Th0 cells. Immunoblotting for T-bet and b-actin in Th0, Th1 and Th2 cells polarized from naïve human CD4+ T cells for 6 days and then restimulated with anti-CD3 and anti-CD28 for 6 hours in the presence or absence of 20ng/ml IL12.

a

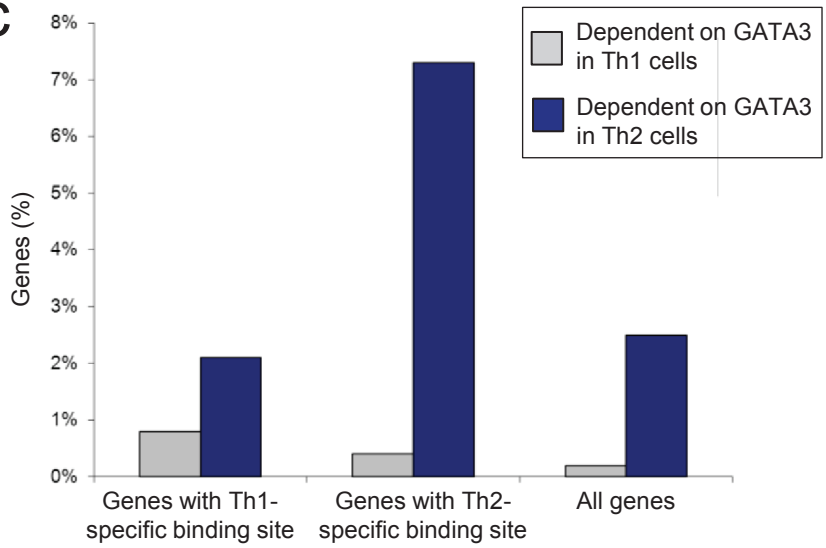
T-bet	+	+	-	+	-	-
GATA-3 Th1	+	-	+	+	+	-
GATA-3 Th2	-	-	-	+	+	+



b



c

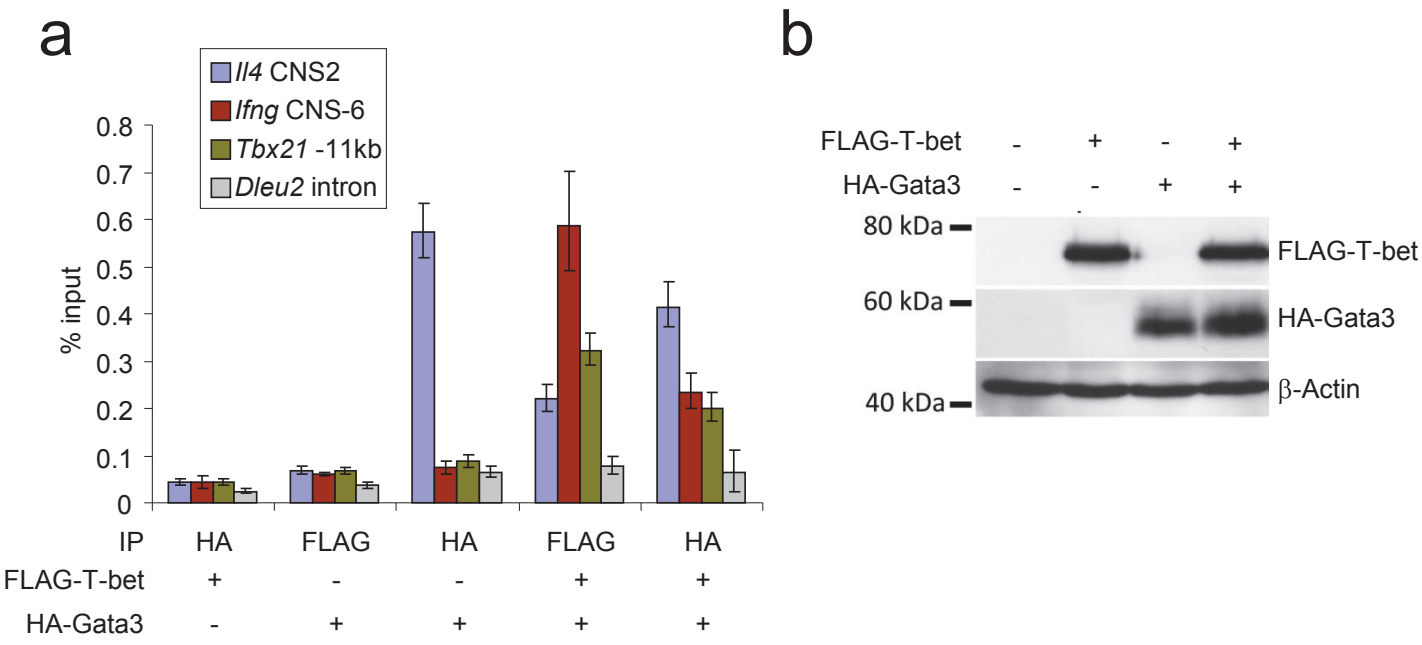


Supplementary Figure S7. Identification of specific combinations of T-bet and GATA3 binding.

A. Specific patterns of T-bet and GATA3 occupancy identified at distal binding sites; "+" indicates binding and "–" a lack of binding for T-bet in Th1 cells, GATA3 in Th1 cells and GATA3 in Th2 cells. The average binding profile for T-bet in Th1 cells (green), GATA3 in Th1 cells (red) and GATA3 in Th2 cells (blue) is plotted for each set of distal sites, as average ChIP-Seq reads per million plotted against the distance between the binding sites.

B. Average binding profile for GATA3 in Th1 cells (red) and Th2 cells (blue) at Th1-specific (left) or Th2-specific (right) sites. The Th1 and Th2-specific binding profiles of GATA3 are of similar intensities.

C. Percentage of genes that exhibit a difference in expression (at least 1.5-fold) in either *in vitro* polarised Th1 cells (light blue) or in *in vitro* polarised Th2 cells (dark blue) from Gata3$^{fl/fl}$ x Tnfrsf4-Cre mice compared to the same cells isolated and polarised from wild-type animals. Genes are divided into those that exhibit a Th1-specific GATA3 binding site, a Th2-specific GATA3 binding site or all genes.



Supplementary Figure S8. T-bet and Gata3 binding in a second batch of EL4 cell lines.
A. Enrichment of *Il4* CNS2, *Ifng* CNS-6, *Tbx21* -11kb and *Dleu2* intronic DNA by ChIP for FLAG-tagged T-bet or HA-tagged Gata3 from EL4 cells stably expressing the proteins indicated. DNA sequences are quantified relative to input DNA (mean and SD, n=3).
B. Western blotting of stable EL4 cell lines with anti-FLAG (M2, Sigma), anti-HA (16B12, Abcam) or anti-β-actin (CS4967, Cell Signaling).

Supplementary Methods

Mice

Wild type (WT) BALB/c mice were purchased from Harlan Laboratories (Indianapolis, IN, USA). T-bet$^{-/-}$ mice on a BALB/c background were purchased from Taconic (Ejby, Denmark) Ifnγ$^{-/-}$ mice were purchased from the Jackson Laboratory [39]. Mice were bred in the Biological Services Unit at KCL (UK Home Office project license PPL/70/6792) or at Charles River Laboratories International (Margate, UK).

Cell culture for ChIP-Seq and microarray analysis

Naïve human CD4+ T-cells (CD4+CD45RA+CD45RO-HLADR-CD62L+) were isolated from healthy donors by negative immunomagnetic selection (Miltenyi Biotec) and activated for 72 hours by plate bound anti-CD3 and anti-CD28 antibodies (2 μg/ml, BD Pharmingen). Cells were then cultured for 10 days with rhIL-2 (NCI, 100 U/ml) in the presence of rhIL-12 (10 ng/ml, R&D Systems) and anti-IL-4 (10 μg/ml, BD Pharmingen) for Th1 polarisation or rhIL-4 (10 ng/ml, R&D Systems) and anti-IFN-γ (10 μg/ml, BD Pharmingen) for Th2 polarisation.

Naïve murine CD4+ T-cells (CD4+CD25-CD62LhighCD44low) were purified by immunomagnetic selection (Miltenyi Biotec) from murine spleen and lymph nodes from WT and T-bet$^{-/-}$ mice (BALB/c background) or from WT and Gata3$^{fl/fl}$ x Tnfrsf4-Cre mice (C57BL/6 background). Cells were activated for 72 hours with plate-bound anti-CD3 and anti-CD28 monoclonal antibodies (both 2 μg/ml). For gene expression analysis, cells were then cultured for 7 days with 20ng/ml IL-2 (R&D Systems) in the presence of 20 ng/ml IL-12 (eBioscience) and

10 μg/ml anti-IL-4 (BioXCell) for Th1 polarisation or 20 ng/ml IL-4 (eBioscience) and 20 μg/ml anti-IFN-γ (BioXCell) for Th2 polarisation.

Chromatin immunoprecipitation and library preparation

We followed previously published protocols for ChIP [11]. Human Th1 and Th2 cells, pooled from multiple donors, were crosslinked by the addition of one-tenth volume of fresh 11% formaldehyde solution for 20 minutes at room temperature before the reaction was quenched by addition of glycine. Cells were rinsed twice with 1xPBS and flash frozen in liquid nitrogen. Cells were lysed with non-ionic detergent, nuclei washed and then lysed with ionic detegerent. Cells were sonicated on ice to solubilize and shear crosslinked DNA (24W for 10 x 30 second pulses using a Misonix Sonicator 3000). The resulting whole cell extract was cleared by centrifugation and then incubated overnight at 4°C with 100 μl of Dynal Protein G magnetic beads that had been preincubated with 10ul of purified serum (T-bet antibody 9856 [51]) or 10μg of anti-GATA3 (D-16, Santa Cruz). Beads were washed 6 times with RIPA buffer and 1 time with TE containing 50 mM NaCl. Bound complexes were eluted from the beads by heating at 65°C with occasional vortexing and crosslinks then reversed in IP and input DNA by overnight incubation at 65°C. IP and input DNA were then purified by treatment with RNAseA, proteinase K and phenol:chloroform extraction followed by ethanol precipitation. Libraries were constructed from IP and input DNA by standard Illumina protocols, except that DNA in the range 150-350bp was gel-purified after PCR-amplification. The library was quantified using an Agilent bioanalzer and subjected to 35bp single-end read sequencing with an Illumina Genome Analyzer II.

Gene expression microarrays

After polarisation for 7 days, murine T-cells were stimulated with PMA (50 ng/ml) and ionomycin (1 μg/ml) for 4 hours and then lysed in TRIzol (Invitrogen). Gene expression analysis was performed using Affymetrix GeneChip Mouse Gene 1.0 ST arrays (T-bet experiment) or Mouse Genome 430A 2.0 arrays (Gata3 experiment), according to the manufacturer's instructions. Arrays were scanned with an Affymetrix GeneChip 3000 7G scanner, images analysed using the Microarray Analysis Suite 5.0 algorithms with the default settings, and normalisation was performed using the robust multi-array average (RMA) method (T-bet data set) or linear global scaling (MAS5) (GATA3 dataset). For comparison to T-bet and GATA3 binding results, data for multiple transcripts were averaged for each gene. Gene expression data from human Th1 and Th2 cells was described previously [11]. Mouse array data are available at ArrayExpress (accession E-TABM-1187).

ChIP-Seq data analysis: Image processing and sequence alignment

Initial processing of sequencing images generated by the Illumina Genome Analyzer II was carried out using the CASAVA pipeline. Polony identification, base calling and QC statistics were performed using GOAT and Bustard modules. 36bp short reads were aligned using ELAND software to NCBI build 36 of human genome (UCSC hg18). Only reads with zero to two mismatches were aligned and only uniquely aligned reads passing quality threshold were retained. When multiple reads matched to same position, only a single read was considered for further analysis. Sequence reads were extended to 150bp. Wig files were generated by calculating tag density in 10bp windows and normalized to reads per million total reads using in-house R scripts. The data for ChIP sequencing runs were then background corrected using data from Th1

whole cell extract (WCE). Datasets for H3K4me1 [52], CTCF [52] and DNaseI HS (University of Washington ENCODE group [18]) were obtained in bed file format and transformed in the same way (except for background correction). Datasets were converted to bigwig format and visualized with the UCSC Genome browser (genome.ucsc.edu). Th1 and Th2 DNaseI data was visualized as peaks. All raw and processed ChIP-Seq data are available at GEO (accession GSE31320).

ChIP-Seq data analysis: Peak Calling

Regions of significant enrichment were identified using MACS version 1.3.7.1 [36] using Th1 WCE as background, with the following settings --tsize=36, --bw=150 and –mfold= 30. For general peak calling, T-bet and GATA-3 libraries were run with a p-value cut-off of 10^{-6}. 200 bp regions centred on peak positions were considered as binding sites for further analysis. Binding sites within 2 kb of any Refseq gene transcription start site (TSS) was considered as proximal binding sites and associated with that gene. Other binding sites were divided into those within a RefSeq gene (intragenic) and those outside of a gene (intergenic). For Gene Ontology and gene expression analysis, binding sites which were more than 4 kb away from any TSS were considered to be distal binding sites. Distal sites were then assigned to the gene with the closest TSS. Because the genes regulated by a distal binding site are not always possible to predict from its location, we only considered distal sites at genes that were also bound proximally by T-bet or GATA3. All binding sites assigned to genes bound proximally by T-bet and GATA3 are listed in Supplementary Table S1.

For comparison to gene expression data, sets of genes with only proximal (<2kb from TSS) T-bet or GATA3 binding sites or with both proximal and 3 or more distal sites (>4kb from TSS) were

identified. The significance of the association between transcription factor binding and change in gene expression was calculated using the hypergeometric distribution.

For functional analysis, sets of genes only associated with proximal sites or with both proximal sites and 4 or more distal sites were identified and then uploaded to DAVID (http://david.abcc.ncifcrf.gov). The complete set of RefSeq genes was used as a background.

ChIP-Seq data analysis: Specific binding combinations

To define specific combinations of T-bet and GATA3 binding, we developed a triple p-value cutoff approach to maximize true-positives and minimize false-negatives, such as incidences where one factor is judged not to be bound because the significance is just above the cutoff. We first identified T-bet, Th1 GATA3 and Th2 GATA3 binding sites at three p-value thresholds; 10^{-7}, 10^{-6} and 10^{-2}. We then identified distal sites with each of the 7 possible combinations of transcription factor binding using these different thresholds. For a site to be considered bound in one dataset only, the p-value for binding had to be below 10^{-7} and above 10^{-2} for the other two datasets. For a site to be considered bound in two datasets only, the p-value for binding had to be below 10^{-6} for each and above 10^{-2} for the other dataset. For a site to be considered bound in all three datasets, the p-value for binding has to be below 10^{-6} for each. For a binding site to be considered bound by two or more factors, the two 200bp binding regions (described above) also had to overlap by at least 100bp. To confirm that this strategy successfully identified specific binding combinations, we plotted the average binding profile for T-bet in Th1 cells, GATA3 in Th1 cells and GATA3 in Th2 cells (Supplementary Figure S7A). The number of sites in each binding category were as follows:

1. ++- bound by T-bet ($<10^{-6}$) and GATA3 in Th1 cells ($<10^{-6}$) but not Th2 cells ($>10^{-2}$). 795 sites.

2. +-- bound by T-bet (10^{-7}) and not bound by GATA3 in either Th1 or Th2 cells ($>10^{-2}$). 4037 sites.

3. -+- not bound by T-bet ($>10^{-2}$) and bound by GATA3 in Th1 cells ($<10^{-7}$) but not Th2 cells ($>10^{-2}$). 850 sites.

4. +++ bound by T-bet ($<10^{-6}$) and GATA-3 in both Th1 and Th2 cells ($<10^{-6}$). 1896 sites.

5. -++ not bound by T-bet ($>10^{-2}$) and bound by GATA3 in both Th1 and Th2 cells ($<10^{-6}$). 1550 sites.

6. --+ not bound by T-bet ($>10^{-2}$) or GATA3 in Th1 cells ($>10^{-2}$) but bound by GATA3 in Th2 cells ($<10^{-7}$). 1569 sites.

7. +-+ bound by T-bet ($<10^{-6}$) and GATA3 in Th2 cells ($<10^{-6}$) but not by GATA3 in Th1 cells ($>10^{-2}$). 89 sites.

The positions of distal binding sites exhibiting each of these combinations are given in Supplementary Table S2. Due to the small number of sites in category 7, this was not considered in the DNaseI HS, gene expression or motif analyses.

To calculate the number of Th1-specific GATA-3 binding sites we added together the number of sites in categories 1 and 3. To calculate the number of Th2-specific GATA3 binding sites, we added the number of sites in categories 6 and 7. For comparison to DNaseI HS data, we downloaded peak data from UCSC and identified those sites specific to Th1 or Th2 cells as those that did not overlap with any sites present in the other cell type. We then identified DNaseI HS

site that overlapped with T-bet or GATA3 binding sites. The frequency of DNaseI HS sites at a set of randomly selected sequences was also calculated as a control.

ChIP-Seq data analysis: Motif identification and analysis

MEME suit (version 4.3.0) was used for identification of significantly enriched motifs. These were identified from the 300 T-bet and GATA3 sites with the most significant p-value. For each set of binding sites, the most significant motifs between 6-30bp wide and occurring once per sequence were identified. From the motifs, we generated position-specific probability matrices (PSPM) for T-bet and GATA3 and used these to search for T-bet and GATA3 motifs at the six different binding combinations (described above). We also calculated the frequency of these motifs in a set of randomly selected sequences as a control. These searches were carried out at p-value $<10^{-4}$ using the FIMO program.

ChIP-Seq data analysis: Metagene analysis

For each dataset**,** a 10 kb region centered on each T-bet or GATA3 (Th2) binding site was extracted and the average reads/million for positions relative to the binding site calculated.

ChIP-Seq data analysis: Conservation analysis

Functionally important elements are more highly conserved than other parts of the genome. Distally bound sites were filtered for those that didn't overlap RefSeq exons. phastCons scores for multiple alignments of 28 vertebrate genomes to the human genome were obtained from the UCSC genome browser. Average conservation score was calculated using a 50 bp moving window and data extracted for each 10 kb region centred on a T-bet or GATA3 (Th2) binding

site. The average phastCons score for each position relative to the binding site was then calculated.

Enhancer reporter assays

-591 to +42bp relative to transcriptional start site of the murine *Ifng* promoter was amplified and restriction sites appended with primers;

F: GACAGATCTCGTTGACCCTGAGTGATTTGTAGTAG and

R: GACAAGCTTATCAGCTGATGTGTCTTCTCTAGGTC

and cloned into the firefly luciferase reporter vector pGL4.13 (Promega). Regions around distal T-bet and GATA3 binding sites were amplified from human genomic DNA and restriction sites appended using the following primers:

IFNG CNS-4 (TGCGAGCTCTCCTCAAAAATGGCCAGAAG and

CGTCTCGAGTAGACATCCCAGCCAACCTG),

IL12RB -2.7kb (AGTGAGCTCTTCAGGAGTTATGTTGCTGTCC and

ATGCTCGAGGGCCTTTCTATTTCCCCACT),

SETBP1 -42kb (GACGAGCTCTCATGTAAACCTTGGCAACC and

TACCTCGAGCCAAGGAATACTCTGGGAACC),

SETBP1 +78.6kb (CGTGAGCTCGGATTCCCACCTATGTTTTCC and

ACGCTCGAGTTGCCTTTCTCTCATTGTCC),

ANTXR2 -37.6kb (CAGGCTAGCGCATATTGTGTGACAGGCTTTT and

GACCTCGAGCCCTGGGAAGTTCAGATGC),

CD226 -46.4kb (CTAGAGCTCCTTTCCTTTCCAGGGTCTCC and

GATCTCGAGTGTGTGCTCTGAATGAGAGTTG),

IGF2R +13.8kb (GTCGAGCTCTGCTAGCTGTTTTTAAAAGGAAA and GTCCTCGAGAACAAGGGCAATTAATTTAACCA).

These sequences were then cloned upstream of the basal *Ifng* promoter and verified by DNA sequencing. Human CD4+ T-cells were activated with anti-CD3 and anti-CD28 antibodies for 72 hours and then cultured for an additional 72 hrs with IL2. $5x10^6$ cells were then transfected with 5μg of reporter plasmid, together with 1 μg of pRL-TK (Promega) using nucleofection (Nuclofector 4D, Lonza). After a 4 hour resting period, the cells were restimulated for 6 hours with anti-CD3/28 before lysis in passive lysis buffer. Firefly and Renilla luciferase activities were then determined using the Dual Luciferase Reporter Assay System (Promega) and a Lumat LB9507 luminometer (Berthold Technologies). Relative light units represent the ratio of firefly to renilla luciferase activity and normalized to the Ifnγ promoter alone. Estimates of statistical significance of the increase in reporter activity caused by the enhancer were performed using a one-sided T-test. We assumed biological and technical sources of error were normally distributed.

For assaying responsiveness to T-bet, we generated an EL4 cell line that stably expressed T-bet under Blasticidin resistance and also a control cell line expressing only the blasticidin resistance gene based on the pMY retroviral expression vector. Cells were transfected with 5 μg pGL4 luciferase reporter constructs and 0.3 μg of pRL-SV40 (Promega). Transfected cells were rested for 24 hrs before assaying. Estimates of statistical significance of the increase in reporter activity in the presence of T-bet were performed using a one-sided T-test.

Distal T-bet binding sites at *TBX21* were amplified and restriction sites appended by PCR with the following primers;

TBX21 -12kb (AGCGAGCTCCTGCCATTCTGAAAACTCTGC and

GTACTCGAGGAGGGTCAGTTGCTCGACAT),

TBX21 +8kb (AGCGAGCTCCTGCCATTCTGAAAACTCTGC and

GTACTCGAGGAGGGTCAGTTGCTCGACAT

and cloned upstream of the basal *Ifng* promoter. Distal elements in the murine and human *IFNG* loci were also cloned with the following primers:

Human CNS-16 (CGTCTCGAGTAGACATCCCAGCCAACCTG and

ACTCTCGAGTAGTGAGGAAGAAAAGCACATGA),

Mouse CNS-22 (GACGAGCTCTAAGGAAGCCAAAGAGAGTATCTT and

CAGCTCGAGTGCATTTTGTCATAGGAGCTTCT),

Human CNS-4 (TGCGAGCTCTCCTCAAAAATGGCCAGAAG and

CGTCTCGAGTAGACATCCCAGCCAACCTG),

Mouse CNS-6 (GACGAGCTCTGCAACCCTTGAAGCTGTGGGTAC and

TGACTCGAGAGATTGCCGTCTGGTCTTGGCGT).

To measure activity of *TBX21* and *IFNG* distal elements, naïve human and mouse CD4+ cells were activated for 72 hrs with anti-CD3/CD28 antibodies and cultured for an additional 7 days (human) or 72 hrs (mouse) with IL2 for Th0 polarisation and IL4 and anti-IFNγ for Th2 polarisation. $5x10^6$ cells were then transfected with 5μg of reporter plasmid, together with 1 μg of pRL-SV40 (Promega) using nucleofection (Nuclofector 4D, Lonza). After a 4 hr resting period, mouse cells were reactivated in the presence of 20 ng/ml IL-12 (eBioscience) for 6 hours before assaying luciferase activity.

Retroviral expression of T-bet in mouse T-cells

The cDNA encoding murine T-bet was amplified by PCR from Th1 cDNA and cloned in frame with a N-terminal double FLAG-tag into pcDNA3 (gift from M. Merkenschlager) and then ligated into pMIG bicistronic retroviral vector (gift from S. Reiner), upstream of the enhanced green fluorescent protein (EGFP) coding sequence under the control of an internal ribosome entry site (IRES). HEK293T cells were transfected with pMIG and pCL-Eco (Imgenex) using polyethyleneimine (Polysciences) and viral supernatants collected 48 hrs later. Naïve mouse T cells from Ifn$\gamma^{-/-}$ mice were sorted and maintained as before. After activation with anti-CD3 and anti-CD28 antibody for 24 hours, cells were transduced by spin-infection with viral supernatants containing 8 μg/ml polybrene. The cells were activated for an additional 48 hrs and then cultured for 3 days on uncoated plates. Total RNA was isolated with Trizol (Invitrogen), treated with Turbo-DNaseI (Ambion) and reverse transcribed with oligo-dT primers (Bioline). Quantitative PCR was performed in duplicate using Power SYBR Green PCR Master Mix (Applied Biosystems). Relative expression of *Tbx21* and *Hprt* (hypoxanthine guanine phosphoribosyl transferase) in cells infected with Tbx21 virus compared to GFP virus was calculated using the ΔΔC(t) method. Real-time PCR primer sequences are:

Tbx21-3'UTR forward: 5'-GTATAGTGATTGGTTGGAGAGGAAGC-3', Tbx21-3'UTR reverse: 5'-TGTAGTTCGGGCAGAGAAAGGG-3',

Hprt forward: 5'-TCAGTCAACGGGGGACATAAA-3', Hprt reverse: 5'-GGGGCTGTACTGCTTAACCAG-3'.

To assay the effect of T-bet on the expression of *Il2rb, Itgal, Icam1, Faslg,Grzb* and *Lta,* we transduced CD4+ T-cells from T-bet$^{-/-}$ Ifn$\gamma^{-/-}$ mice with T-bet and GFP-encoding retroviruses as

before. RNA was purified, reverse transcribed and real-time PCR performed as before. Primer sequences are available on request. To measure changes in protein levels corresponding to changes in mRNA abundance, cells were stained with antibodies from BD Biosciences, data acquired on a FACS Calibur and analyzed with CellQuest™ software. The purity of all cell populations used was determined by FACS and was routinely >98%.

T-bet and Gata3 co-expression and ChIP

The T-bet expression vector contained the N-terminally Flag-tagged murine Tbx21 coding region which is fused to the blasticidin resistance gene using a 2A sequence to obtain co-translational expression of both genes. The Tbx21 coding-region was amplified, digested with BglII/EcoRI and inserted into pcDNA-Flag. A 2A sequence from the *thosea asigna* virus was generated by annealing two DNA oligonucleotides with ends compatible with EcoRI (5') and NcoI (3'). The blasticidin deaminase gene (BSD) was amplified from pMY-IRES-BSD (gift from Adrian Hayday) and cut with NcoI/SalI. Flag-Tbx21 was released from the pcDNA-Flag vector with BamHI/EcoRI and ligated together with the annealed 2A sequence and the PCR-amplified BSD gene into BamHI/XhoI sites of pMY-IRES-GFP. To generate a control vector for the T-bet expression construct, a PCR-amplified BSD cassette was inserted into BamHI/XhoI sites of pMY-IRES-GFP.

The Gata3 expression vector contained a puromycin acetyl transferase gene (PAC) at the 5' end and a HA-tagged murine Gata3 cassette at the 3' end separated by a 2A sequence. An N-terminal HA-tag was added to Gata3 by inserting PCR-amplified Gata3 into BamHI/EcoRI sites of pcDNA-HA. The HA-Gata3 cassette was again amplified by PCR to introduce a SalI site at the 5'

and an EcoRI site at the 3' end of the coding region. The PAC gene was amplified by PCR using pMSCVpuro as a template and digested with BamHI and HindIII. Two DNA oligonucleotides were annealed to create the 2A sequence connecting PAC and Gata3. pMY-IRES-mPlum was linearised with BamHI/EcoRI and PAC, 2A sequence and HA-Gata3 were inserted. A corresponding empty control vector was created by inserting PAC into BamHI/XhoI sites of pMY-IRES-mPlum.

Viruses were made by transfection of 293T cells with the pMY-IRES constructs described above, along with the packaging plasmid pCL-Eco. EL4 cells were transduced with either both control viruses, FLAG-T-bet and mPlum viruses, HA-Gata3 and GFP viruses or FLAG-T-bet and HA-Gata3 viruses and selection was started 48 hrs later using both 4 μg/ml blasticidin and 6 μg/ml puromycin (Calbiochem). Expression of FLAG-T-bet and HA-Gata3 was confirmed by western blotting.

A second set of EL4 lines were also constructed using pMSCVpuro-HA-Gata3 expressed in combination with either pMY-BSD-GFP or pMY-Flag-T-bet-BSD-GFP. pMSCVpuro-HA-Gata3 was generated by releasing HA-Gata3 from pcDNA-HA by digestion with BglII/EcoRI and insertion into pMSCVpuro (Clontech). The data from these lines are shown in Supplementary Figure 8.

ChIP was performed as above, except chromatin was sonicated at 30W for 13 x 30 second pulses and chromatin was immunoprecipitated with anti-HA (3F10, Roche) and anti-FLAG (M2, Sigma) antibodies. Specific DNA sequences were quantified in triplicate in ChIP-enriched and input DNA by quantitative PCR using QuantiTect SYBR Green PCR master mix and enrichment

calculated relative to input and to a site within *Dleu2* that lacks T-bet or GATA3 binding according to the ChIP-Seq analysis. Standard curves were performed for all PCR runs to ensure linear amplification. Mouse genomic regions orthologous to human T-bet binding sites were identified using the Convert function of the UCSC genome browser and the following primers designed:

Il4 CNS2 forward: 5'- ATCACGTCGTCTTACCCAAACA-3', Il4 CNS2 reverse: 5'-TGTGGGAGAGCGTCTGATCTGT-3', Ifng CNS-6 forward: 5'-CACGGGCTTTGAAGGATACC-3', Ifng CNS-6 reverse: 5'-CCTCGCTAGTCTGGCCAATAGT-3', Tbx21 -11kb forward: 5'-TGCCCGAGAGTCAGTGCTT-3', Tbx21 -11kb reverse: 5'-TCCCAGCTTCGAGGAAACAC-3', Dleu2 intron: 5'- GGTGGACGAATAAGCTGTGC -3', Actin intergenic reverse: 5'-TCCAAGGAGTTAAGGGGTAGC-3'.