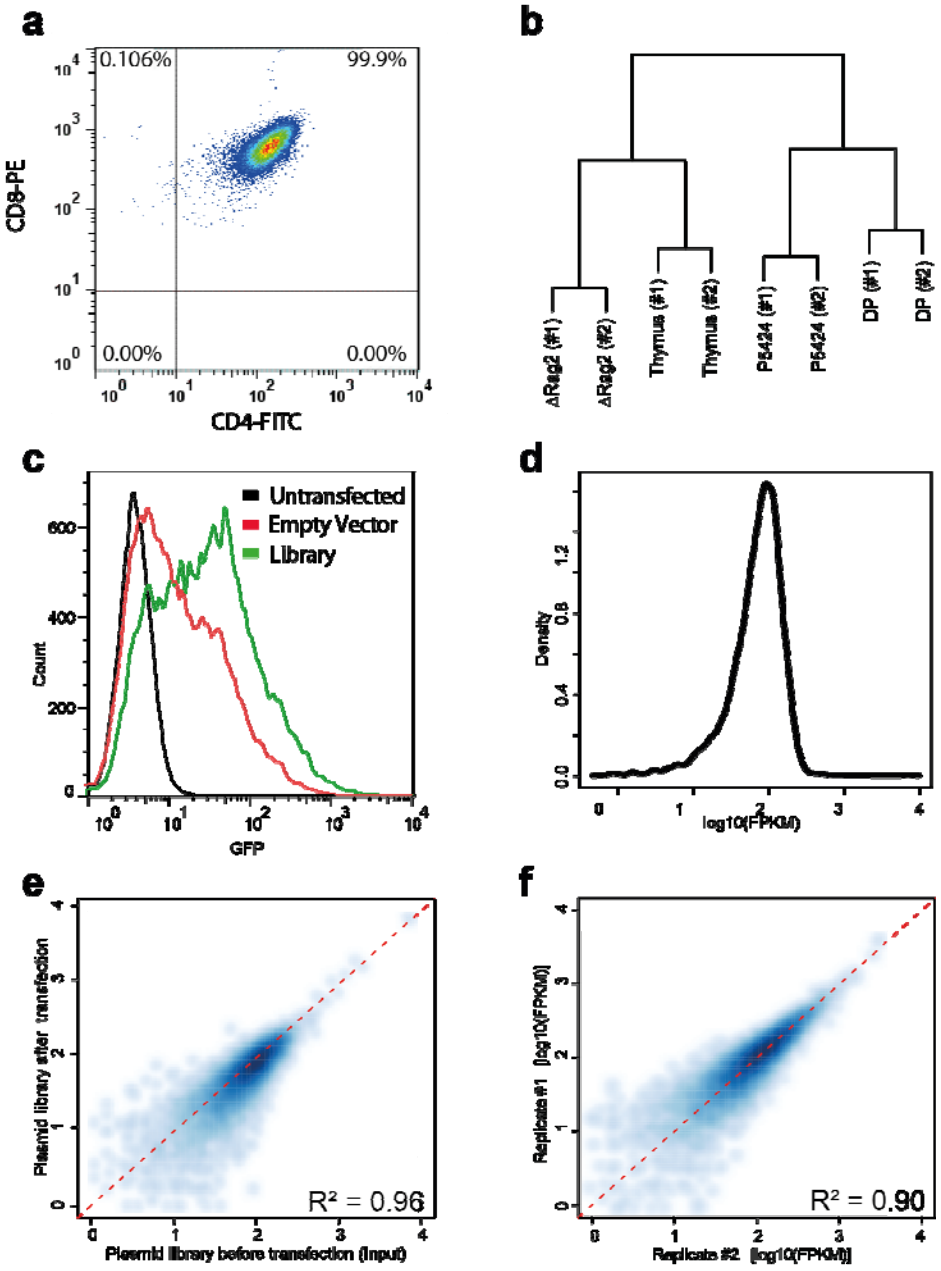
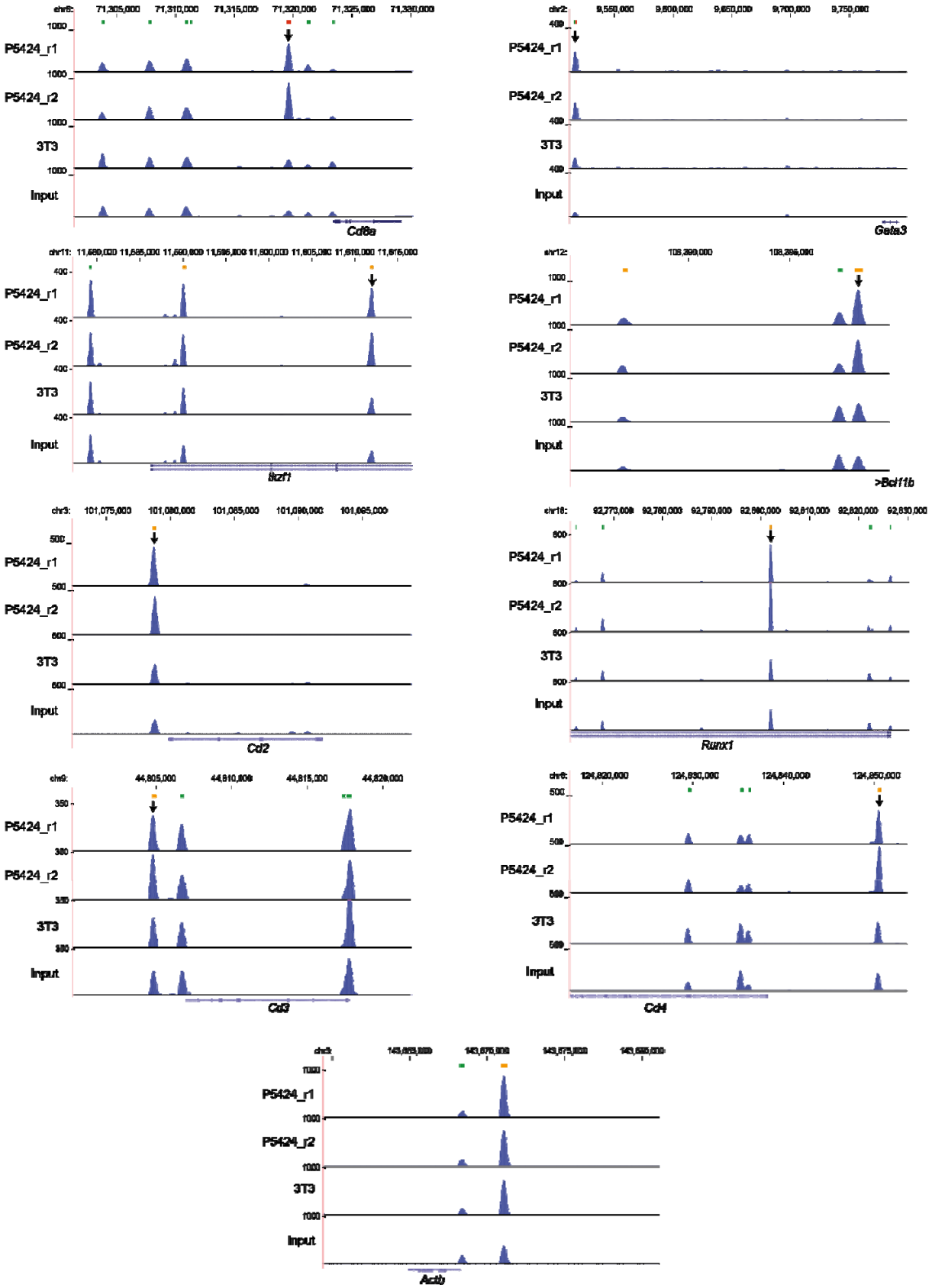
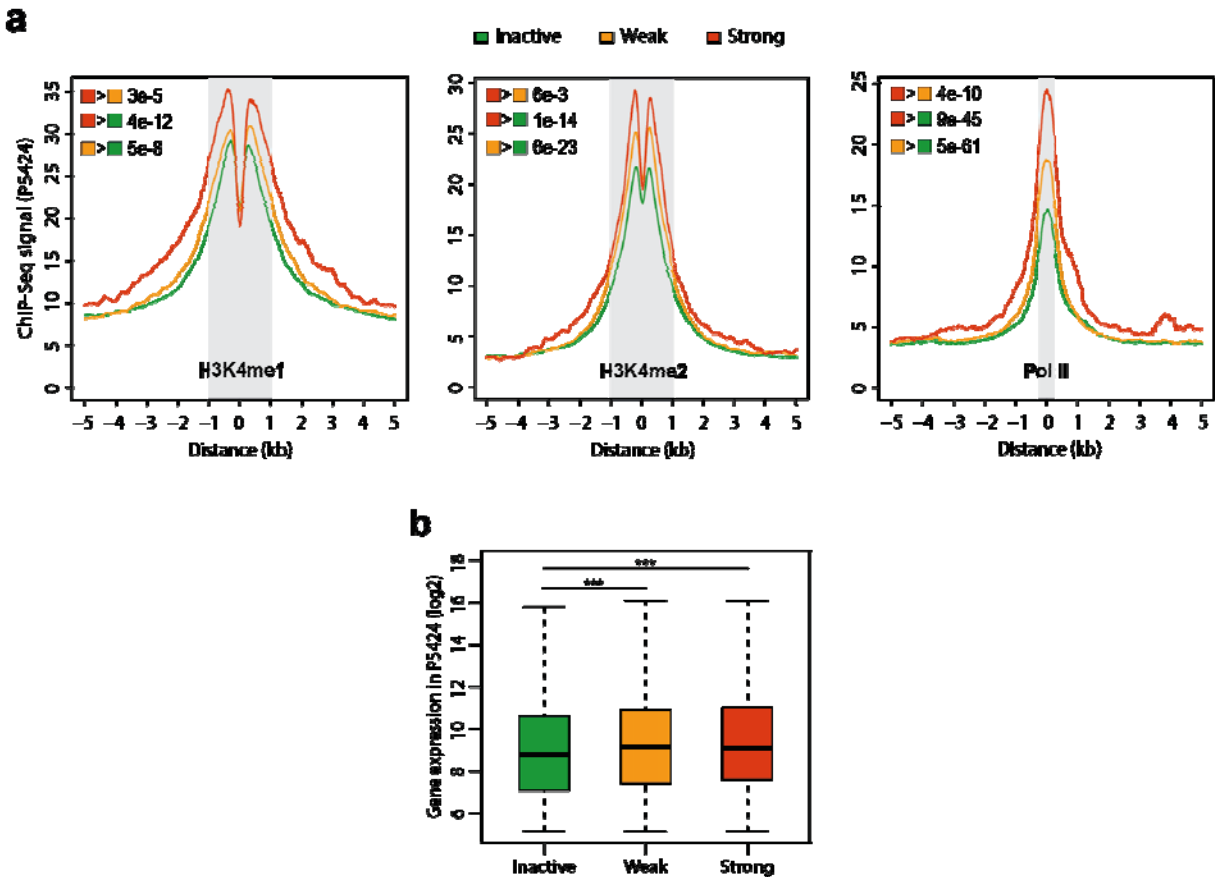


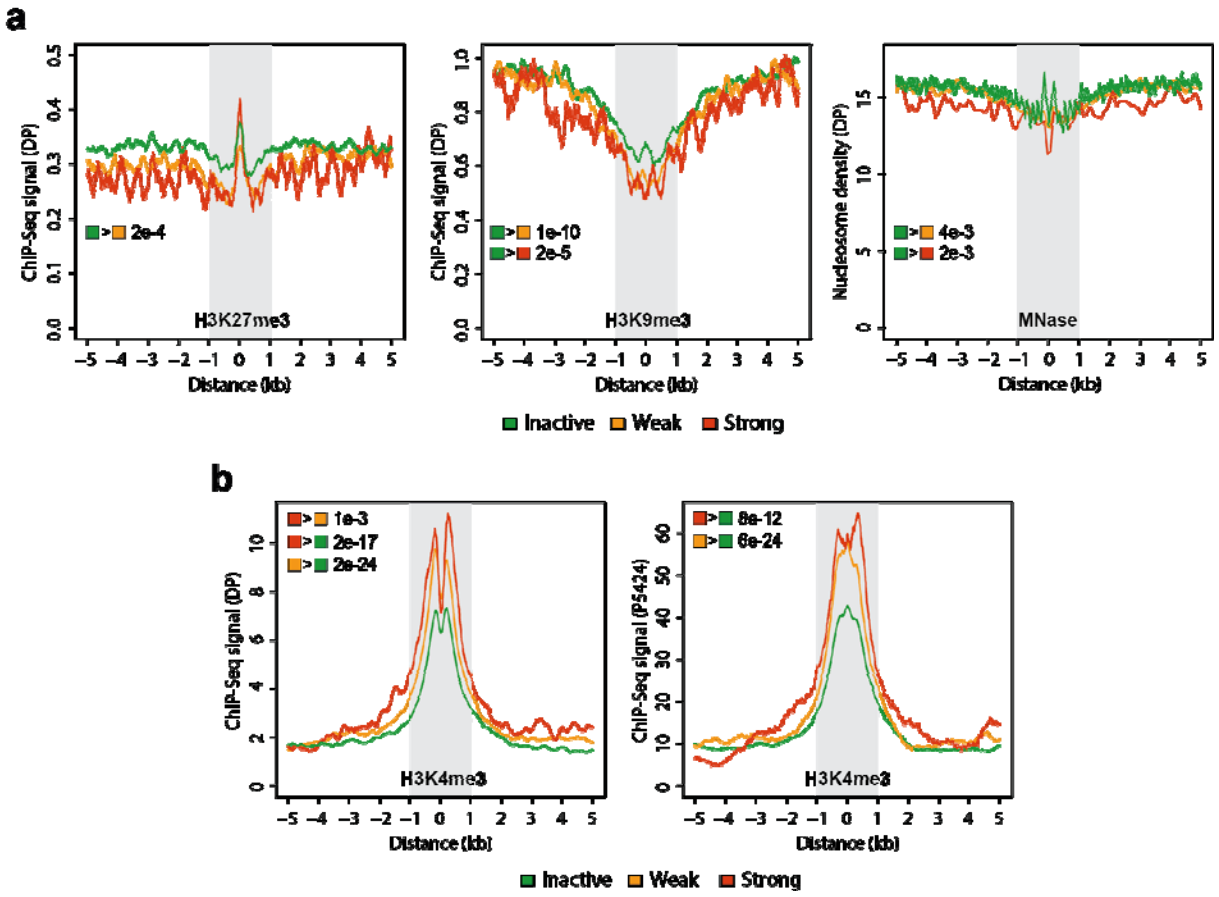
Supplementary Figure 1: Assessment of CapStarr-seq efficiency and reproducibility. (**a**) FACS analysis of P5424 cell line for CD4 and CD8 surface markers. (**b**) Hierarchical clustering of transcriptomic data from ΔRag2 thymocytes (CD4- and CD8- thymocytes), DP thymocytes, whole thymus and the P5424 cell line. Each sample was analyzed in duplicate (#1 and #2). The P5424 cell line correlates better with DP thymocytes than with ΔRag2 thymocytes or the whole thymus. (**c**) P5424 cells were transfected with the empty Starr-Seq screening vector (red line) or the screening vector containing the enriched library (green line). GFP was analyzed by FACS 24 hours after transfection. (**d**) Density plot of Fragments *Per* Kilobase *per* Million mapped reads (FPKM) obtained for each CRM after sequencing of the enriched plasmid library (Input). (**e**) Correlation plot between the input library and the plasmid library after transfection in P5424 cells. The high correlation between the two samples ($R^2 = 0.96$) reflects absence of bias due to transfection. (**f**) Correlation plot between the two CapStarr-Seq replicates performed in P5424 cells ($R^2 = 0.90$).



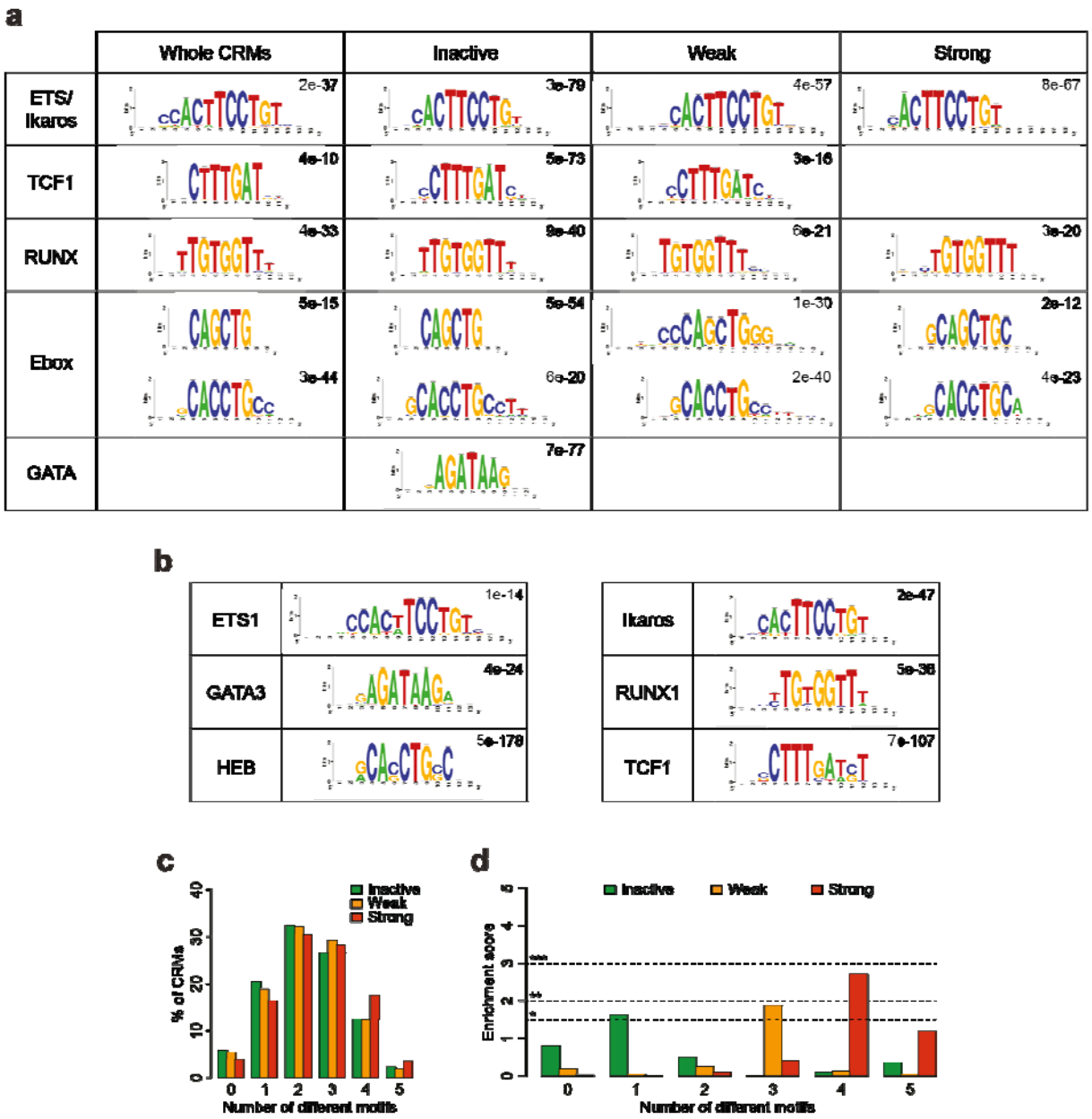
Supplementary Figure 2: CapStarr-seq profiles of known T-cell enhancers. Profiles are represented as in Figure 1c. Known T-cell enhancers shown in Figure 1b are indicated by an arrow. All the indicated enhancers were found to be inactive in the 3T3 cell line. The CapStarr-seq profile at the *Actb* locus is shown at the bottom panel. A distal CRM upstream *Actb* displayed weak enhancer activity in both 3T3 and P5424 cell lines.



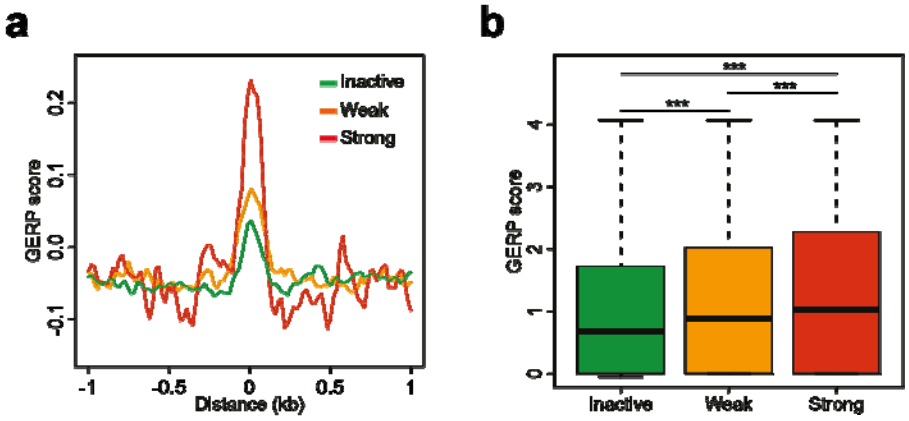
Supplementary Figure 3: Epigenetic features of CRMs in the P5424 cell line. (**a**) Average profiles of H3K4me1, H3K4me2 and Pol II ChIP-seq in the P5424 cell line for inactive, weak and strong CRMs identified by CapStarr-Seq in P5424 cells. Statistically significant differences ($p < 0.05$) calculated on regions highlighted in grey are indicated (Mann-Whitney U-test). (**b**) Boxplot showing the expression distribution in the P5424 cell line of genes associated with inactive, weak and strong CRMs defined in P5424 cells. Thick line represents the median expression, and the bottom and top whiskers indicate respectively the minimum and the maximum expression level. Gene expression levels were significantly higher for genes associated with active CRMs with respect to those associated with inactive CRMs (***: $p < 0.001$; Student *t*-test).



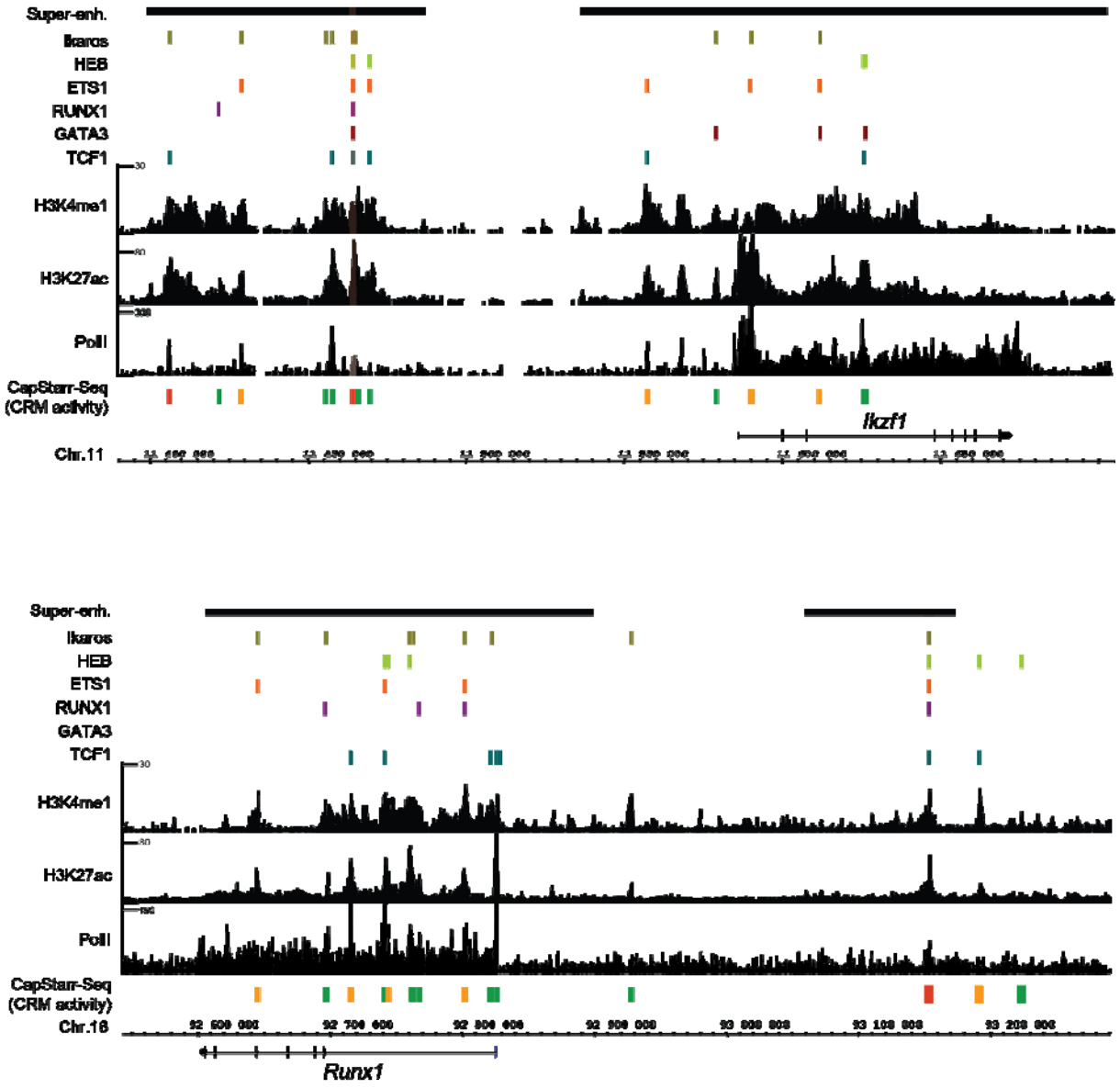
Supplementary Figure 4: Additional epigenetic modifications and nucleosome occupancy of CRMs. (**a**) Average profiles of the ChIP-seq for H3K27me3 and H3K9me3 as well as for the nucleosome occupancy (MNase-seq) in DP thymocytes for inactive, weak and strong CRMs identified in P5424 cells. (**b**) Average profiles of H3K4me3 ChIP-seq in DP thymocytes and P5424 cell line for inactive, weak and strong distal CRMs identified in P5424 cells. In panels a and b, statistically significant differences ($p < 0.05$) calculated on regions highlighted in grey are indicated (Mann-Whitney U-test).



Supplementary Figure 5: Analysis of TF binding motifs in the CRMs. (**a**) TF binding motifs found by a *de novo* motif analysis using the RSAT tool. Significant motifs found within inactive, weak and strong CRMs as well as with whole CRMs defined in P5424 cells are shown. (**b**) TF binding motifs identified by a *de novo* motif analysis within the peaks of indicated ChIP-seq datasets. (**c**) Distribution of CRMs in function of the number of different TF binding motifs found. Inactive, weak and strong CRMs were scanned with the ETS1, GATA3, HEB, RUNX1 and TCF1 matrices discovered in (b), and the percentage of CRMs with the indicated number of different sites was plotted. The number of different motifs per CRM is significantly higher in strong CRMs with respect to weak and inactive CRMs (p = 4e-3 and p = 3e-4, respectively) (Mann-Whitney U-test). (**d**) Enrichment of inactive, weak and strong CRM sets in function of the number of different TF binding motifs found. The enrichment scores were calculated as the inverted Log10 of the p-values obtained with a hypergeometric test. Statistically significant enrichments are indicated (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).



Supplementary Figure 6: DNA sequence conservation of the CRMs. (**a**) Average profile of the DNA sequence constraint (GERP score) around inactive, weak and strong CRMs defined in P5424 cells. Statistically significant differences ($p < 0.05$) calculated on regions highlighted in grey are indicated (Mann-Whitney U-test). (**b**) Boxplot showing the average GERP scores distribution within inactive, weak and strong CRMs defined in P5424 cells. Thick line represents the median average GERP scores within CRMs, and the bottom and top whiskers indicate respectively the minimum and the maximum average GERP scores. Statistical significance was calculated by a Mann-Whitney U-test (***: $p < 0.001$).



Supplementary Figure 7: Epigenomic profiles of the *Ikzf1* and *Runx1* loci. Legends are as in Figure 4a.

Supplementary Table 1: Information about CapStarr-seq, ChIP-seq and DNaseI-seq data generated in this study and submitted to the NCBI Gene Expression Omnibus.

Sample	# total reads	# mapped reads	Estimated fragment size (bp)	GEO ID
CapStarr-seq samples				
P5424_r1	5,199,314	5,158,023	318	GSE60029
P5424_r2	4,820,360	4,743,280	325	
3T3	4,952,162	4,869,506	299	
Input (library before transfection)	5,898,660	5,864,041	302	
Plasmid library after transfection	5,424,196	5,385,016	208	
ChIP-seq and DNaseI-seq in Double Positive thymocytes				
HEB	52,066,815	36,919,248	175	GSE63732
TCF1	34,084,990	26,469,671	165	
DNaseI	24,288,922	17,912,379	182	
H3K9me3	39,703,738	16,592,692	196	
ChIP-seq in P5424 cell line				
H3K4me1	48,787,435	36,198,602	186	GSE63416
H3K4me2	49,576,980	36,177,097	186	
H3K4me3	37,106,339	29,109,407	155	
Polymerase II	40,207,892	24,534,212	157	

Supplementary Table 2: Information about published ChIP-seq and MNase-seq datasets used in this study and downloaded from the NCBI Gene Expression Omnibus.

Sample	GEO ID	Reference
GATA3	GSE20898	Wei, et al. (2011). Immunity 35(2): 299-311
Ikaros	GSE32311	Zhang, et al. (2012). Nat Immunol 13(1): 86-94
CTCF	GSE27214	Ebert, et al. (2011). Immunity 34(2): 175-187
RUNX1	GSE44578	Lepoivre et al. (2013). BMC genomics 14: 914
Input		
ETS1	GSE29362	Koch, et al. (2011). N. S. M. B. 18(8): 956-963
H3K4me1		
H3K4me3		
Polymerase II		
H3K27me3	GSE38577	Fenouil, et al. (2012). Genome Res 22(12): 2399-2408
MNase		
H3K27ac	GSE63732	Lepoivre et al. (2013). BMC genomics 14: 914