iScience, Volume *26*

Supplemental information

Low-affinity CTCF binding drives transcriptional regulation whereas high-affinity binding encompasses architectural functions

Ester Marina-Zárate, Ana Rodríguez-Ronchel, Manuel J. Gómez, Fátima Sánchez-Cabo, and Almudena R. Ramiro

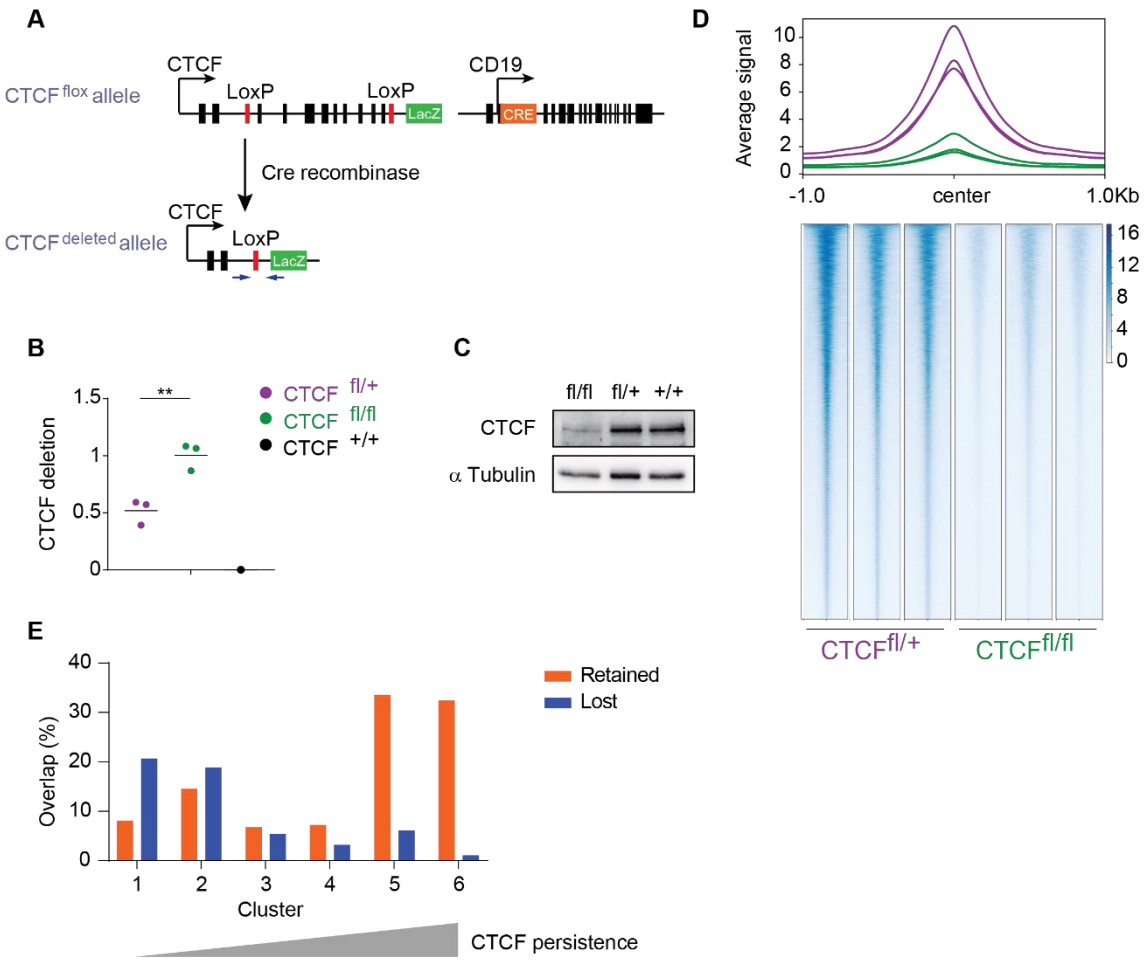

Figure S1. CTCF depletion model, Related to Figure 1. A) Representation of the CTCF^{flox} allele before and after Cre mediated deletion. Blue arrows represent the position of qPCR primers for quantification of CTCFflox deletion. B) qPCR analysis of CTCF deletion in B cells from CTCF$^{fl/+}$, CTCF$^{fl/fl}$ and CTCF$^{+/+}$ mice. p(deletion)=0.007. Statistical analysis was done with the two-tailed unpaired Student's t-test. Each dot represents an individual mouse. C) Western blot analysis of CTCF in B cells from CTCF$^{fl/fl}$, CTCF$^{fl/+}$ and CTCF$^{+/+}$ mice. α-Tubulin is shown as a loading control. D) Profile plot (top) and heatmap (bottom) showing the average CTCF distribution at CBSs. Regions 1kb upstream and downstream of the center of the CBS are shown. E) Percentage of lost or retained CBSs in our study overlapping with the CBSs clusters defined by Luan *et al*.[1]. CBSs coordinates were obtained from GSE150415.

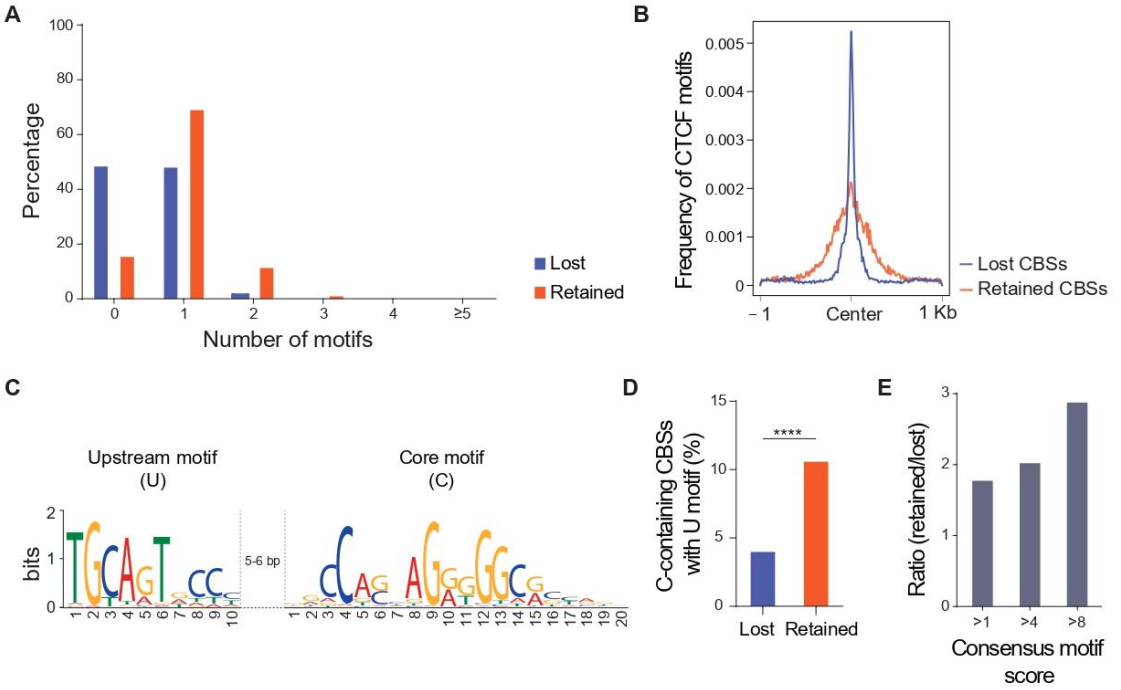

Figure S2. Lost and retained CBS molecular features, Related to Figure 2. A) Frequency of the number of CTCF motifs at lost CBSs (blue) and retained CBSs (orange). B) Density of CTCF motifs in the +/-1 kb regions flanking lost (blue) or retained (orange) CBSs. C) Sequence logo of the consensus Upstream (U) and Core (C) CTCF motifs separated by a spacer of 5-6 bp, as described by Nakahashi *et al.*[2]. D) Proportion of core-containing CBSs that harbor an upstream U motif at 5-6bp distance, as shown in panel C. Only motifs with a consensus motif score higher than 7 were considered. Statistical analysis was done with Fisher's exact test (p-value < 0.0001). E) Retained CBSs are enriched in high U+C score motif. Graph shows the ratio between U+C containing retained and lost CBSs for the indicated consensus U+C motif scores as calculated with Spamo.

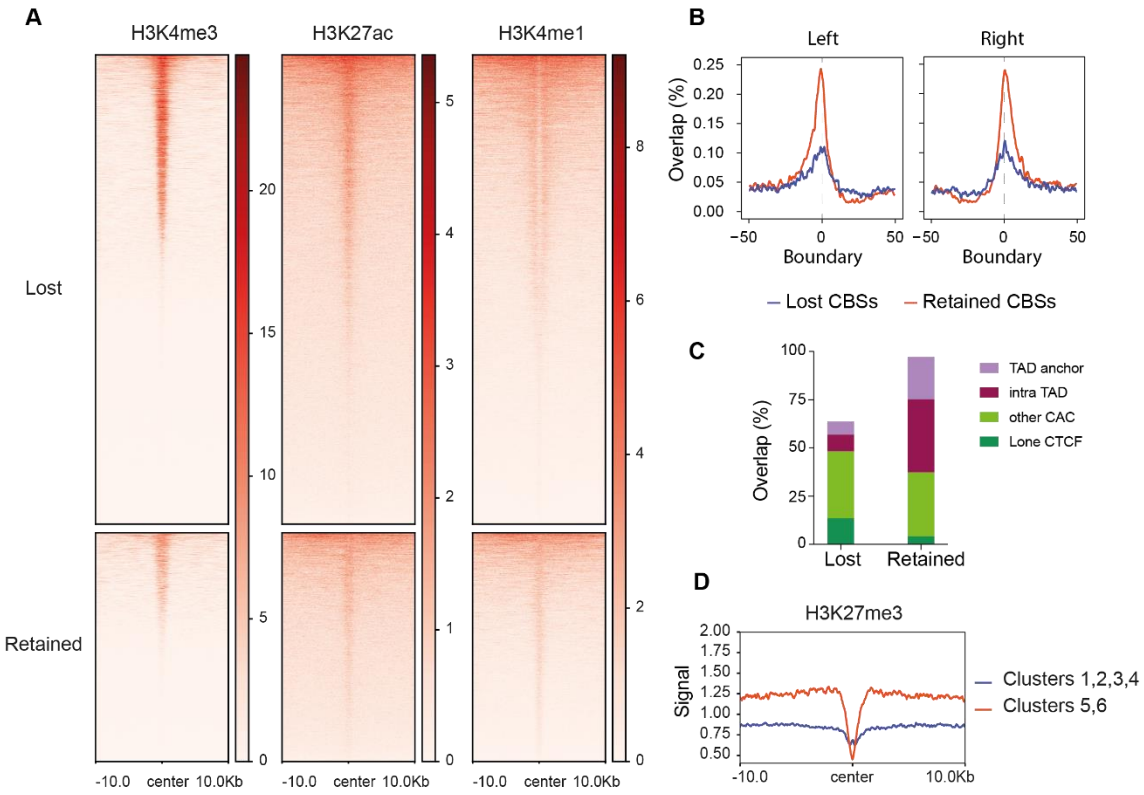

Figure S3. CBS histone marks and TAD and intra-TAD loop boundaries, Related to Figure 3 and Figure 4. A) Heatmaps of H3K4me3, H3K27ac and H3K4me1 at lost or retained CBSs (±10 kb from the center of the binding site). Histone datasets were downloaded from GSE82144[3]. B) Positional enrichment of lost or retained CBSs at loop boundaries obtained from Hi-C data in naïve B cells.Datasets were downloaded from GSE82144[3]**.** C) Percentage of lost and retained CBSs at Lone CTCF (CTCF sites lacking cohesin bound), other CAC sites (non anchor cohesion-and-CTCF sites), TAD anchors, and intra-TAD loop anchors. Datasets were downloaded from GSE102999[4]. D) H3K27me3 average signal at CBSs groups defined by Luan *et al*.[1]. CBSs coordinates were obtained from GSE150415. Histone dataset was downloaded from ENCODE (ENCFF978WJA).

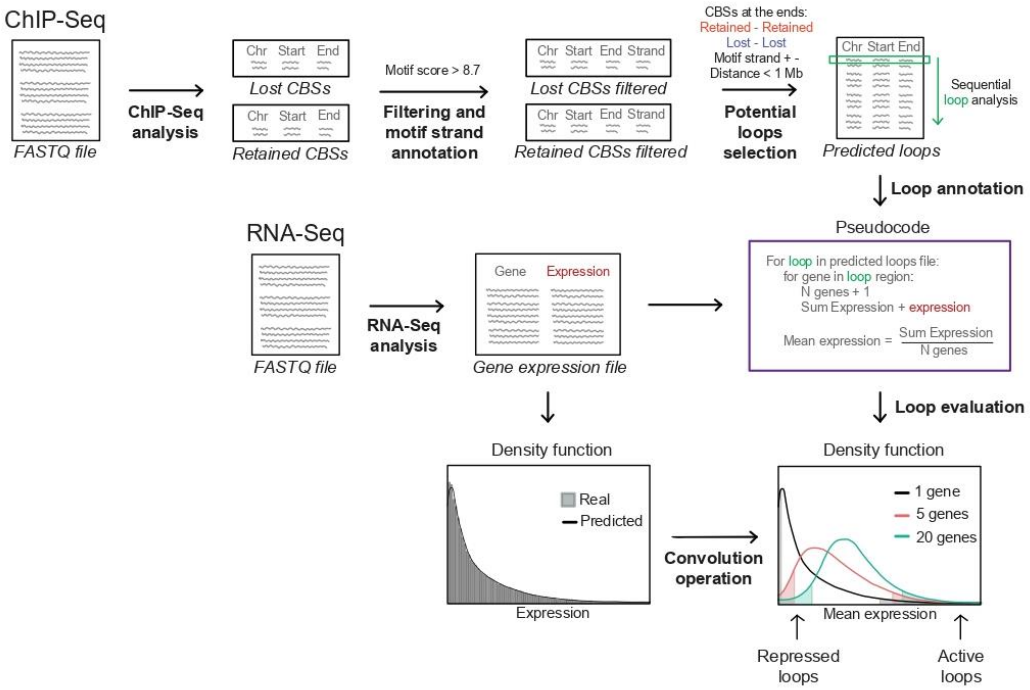

Figure S4. Bioinformatic approach for identification of potential repressed and active loops, Related to Figure 6. Within lost and retained CBSs only those containing a CTCF-motif were selected. CTCF motif-containing strand was annotated and then, using a custom python script, pairs of CBSs were selected to form the predicted loop group according to the following conditions: 1) both CBSs must be either lost or retained, 2) CTCF motif must be in forward orientation for the CBS located at 5' and in reverse orientation for the CBS located at 3' and 3) distance between both CBS must be shorter than 1 Mb. Then, a custom script was used to associate each of these loops with the genes contained in them and to calculate their mean expression using the expression data file obtained in the RNA-seq analysis. In addition, these expression data were used to estimate the probability density function (PDF) (see methods) of gene expression. To determine the PDF not only of a single gene but of a group of N genes, we convolved this function N times. Thus, we obtained the expression PDF of randomly selected gene clusters of various sizes. We then selected the potential loop regions with mean expression at the bottom 10% (repressed) or the top 10% (active) of their assigned distribution according the number of genes they contain.

REFERENCES

1. Luan, J., Xiang, G., Gómez-García, P.A., Tome, J.M., Zhang, Z., Vermunt, M.W., Zhang, H., Huang, A., Keller, C.A., Giardine, B.M., et al. (2021). Distinct properties and functions of CTCF revealed by a rapidly inducible degron system. Cell Rep *34*, 108783. 10.1016/j.celrep.2021.108783.
2. Nakahashi, H., Kieffer Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep *3*, 1678-1689. 10.1016/j.celrep.2013.04.024.
3. Kieffer-Kwon, K.R., Nimura, K., Rao, S.S.P., Xu, J., Jung, S., Pekowska, A., Dose, M., Stevens, E., Mathe, E., Dong, P., et al. (2017). Myc Regulates Chromatin Decompaction and Nuclear Architecture during B Cell Activation. Mol Cell *67*, 566-578.e510. 10.1016/j.molcel.2017.07.013.
4. Matthews, B.J., and Waxman, D.J. (2018). Computational prediction of CTCF/cohesin-based intra-TAD loops that insulate chromatin contacts and gene expression in mouse liver. Elife *7*. 10.7554/eLife.34077.