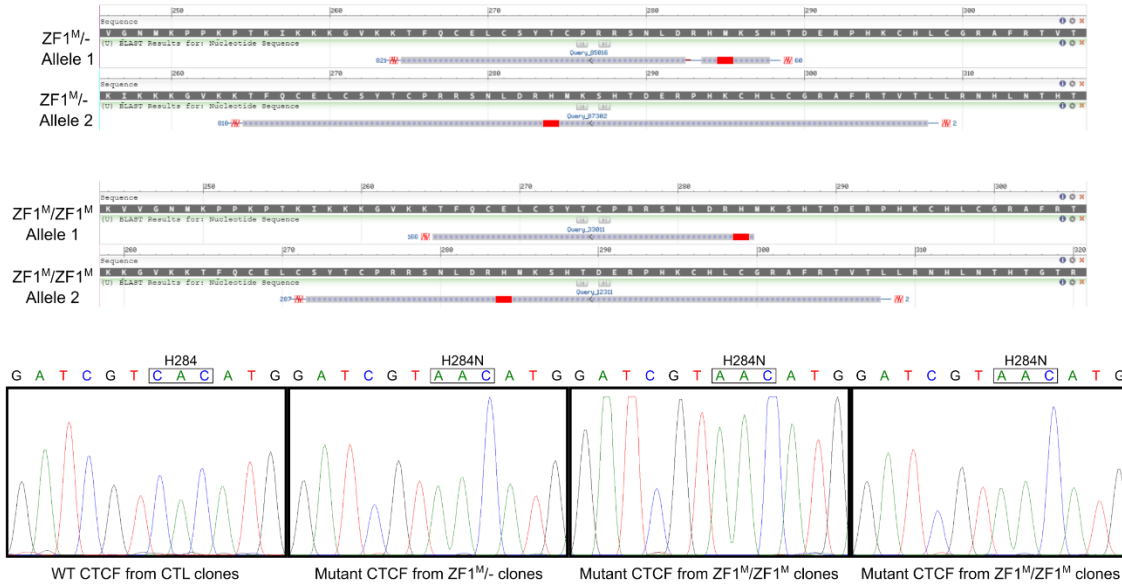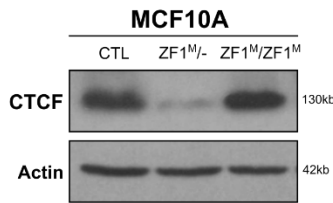# SUPPLEMENTARY DATA

A
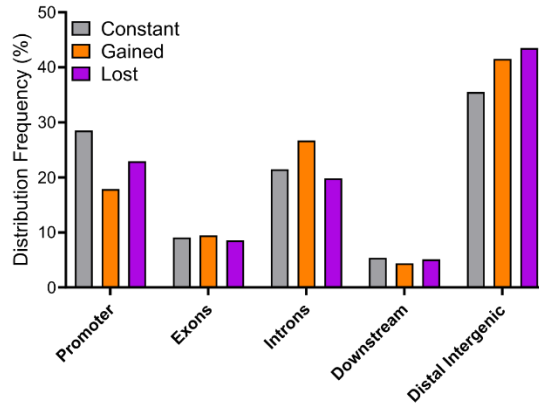
CTCF Amino Acid Sequence comparison with BlastX

| WT CTCF from CTL clones | Mutant CTCF from ZF1^M/- clones | Mutant CTCF from ZF1^M/ZF1^M clones | Mutant CTCF from ZF1^M/ZF1^M clones |

B

MCF10A

CTL   ZF1^M/-   ZF1^M/ZF1^M

CTCF — 130kb

Actin — 42kb

C

Legend: Constant, Gained, Lost

Y-axis: Distribution Frequency (%)

X-axis: Promoter, Exons, Introns, Downstream, Distal Intergenic

D

Compared with WT CTCF ChIP-Seq from Fritz et al. 2018

ZF1^M/ZF1^M: 92.22%, 87.76%

ZF1^M/-: 92.39%, 88.70%

Up or Down CBS with both CTCF WT datasets
Altered CBS only with our CTL CTCF ChIP-Seq
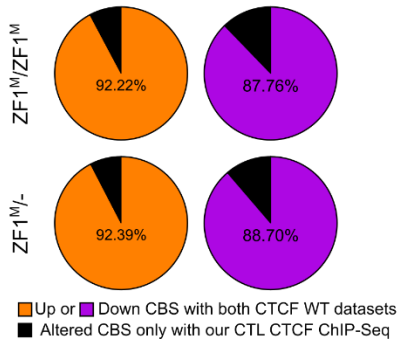
E

Compared with MACS2/DiffBind ChIP-Seq Analysis Pipeline

ZF1^M/ZF1^M: 81.97%, 79.01%

ZF1^M/-: 73.35%, 91.09%

Up or Down CBS called with both pipeline
Altered CBS only called by csaw
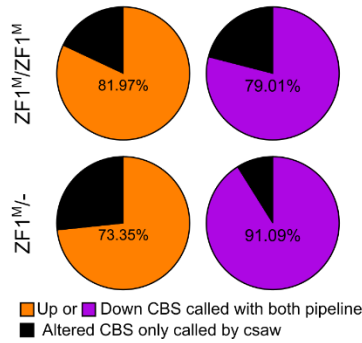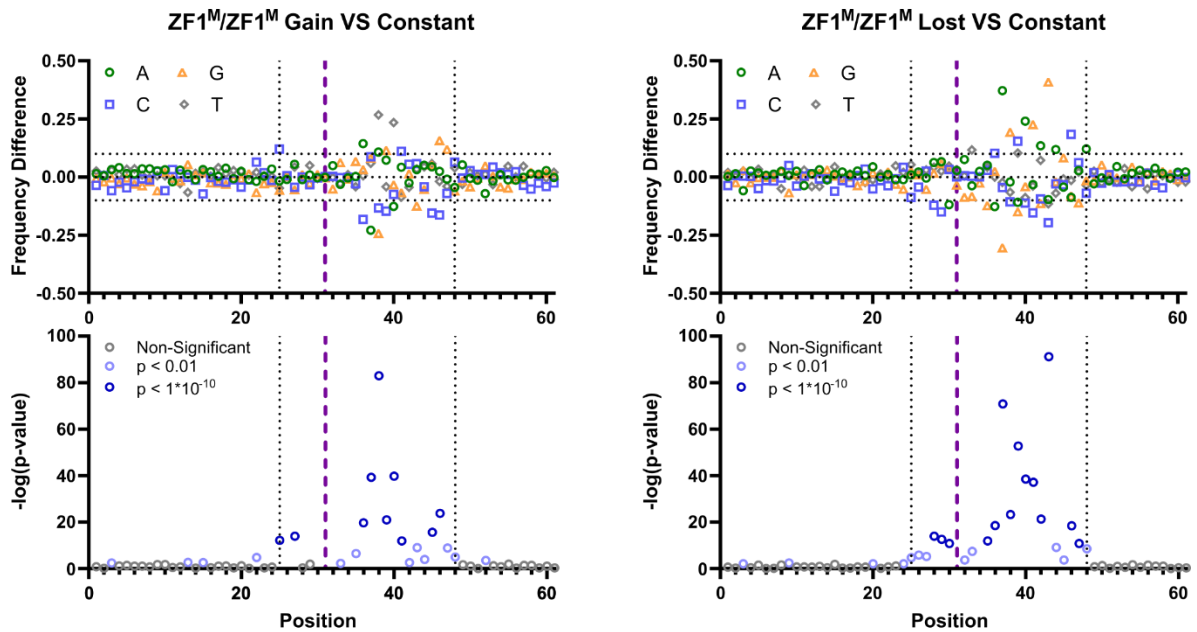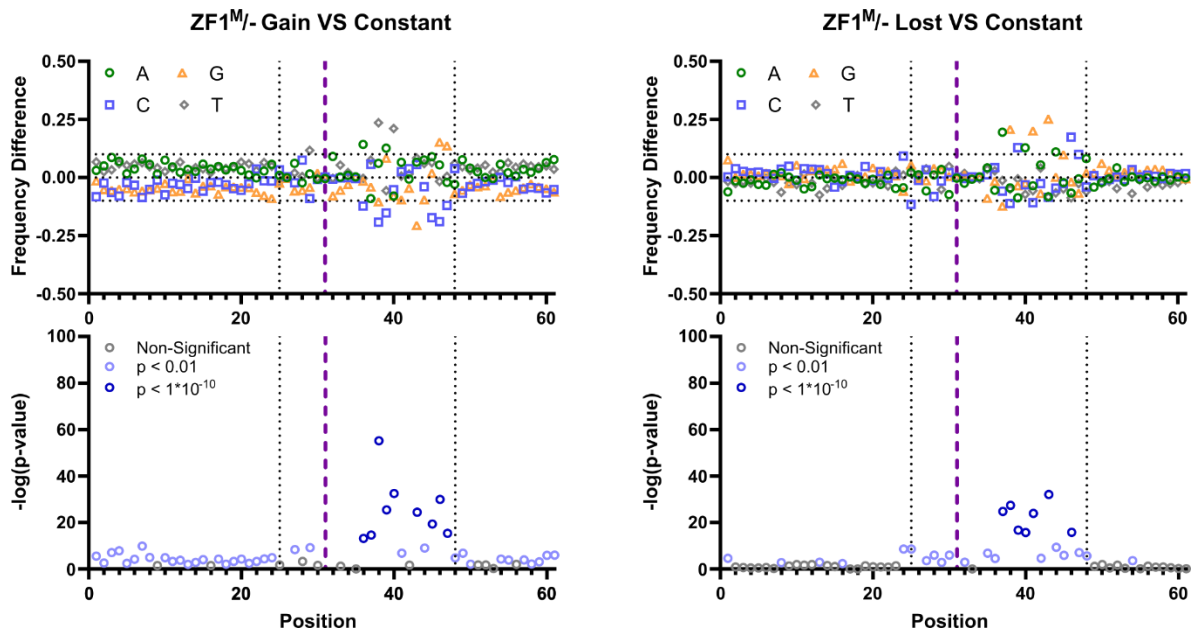
Supplementary Figure 1:

(**A**) BlastX Results from single allele Sanger sequencing showing the presence of the H284N mutation and frameshift deletion in the respective alleles of ZF1^M/ZF1^M and ZF1^M/- clones, respectively. Chromatograph results from one allele of the CTL clones and each mutated CTCF alleles of ZF1^M/- and ZF1^M/ZF1^M detailing the inserted mutation. (**B**) Western Blot of CTCF levels in the distinct CTCF ZF1M MCF10A clones. Actin used as loading controls. (**C**) Relative distribution of common CBS. Showing slight enrichment of altered CBS on distal intergenic elements and a slight enrichment of constant CBS on promoters. (**D**) Comparison between csaw called DB regions between ZF1^M/ZF1^M or ZF1^M/- and our CTL MCF10A or CTCF WT MCF10A ChIP-Seq from Fritz et al. 2018, showing that DB regions are intrinsic to the mutant clones. (**E**) Comparison between csaw and MACS2/DiffBind to identify differentially binding regions between ZF1^M/ZF1^M or ZF1^M/- and CTL MCF10A, showing that a majority of DB regions are called by both analysis methods.

A

Pre-processing	A sliding window approach to quantify binding intensity	Quality control
• Remove reads in blacklist • Filter out low-quality reads • Extend reads to the average fragment length	• Counting reads into windows of **10bp**	• Filtering windows by abundance • Normalizing for sample-specific trended biases

| Differentially binding analysis	Aggregating windows into regions	Results: (FDR<0.05)
• Empirical Bayes shrinkage to estimation **dispersion;** allows for sharing information between windows	• Adjacent windows less than 100bp apart are aggregated into regions	• Gained • Lost • Stable

Supplementary Figure 2:

(**A**) csaw flowcharts and specific settings used during this study.

Supplementary Figure 3:

(**A**) MoMotif analysis of base frequency difference and p-value of bases distribution difference around CTCF-Like motif in lost and gain CBS subsets compared to constant subset in CTCF ZF1^M/ZF1^M clones. (**B**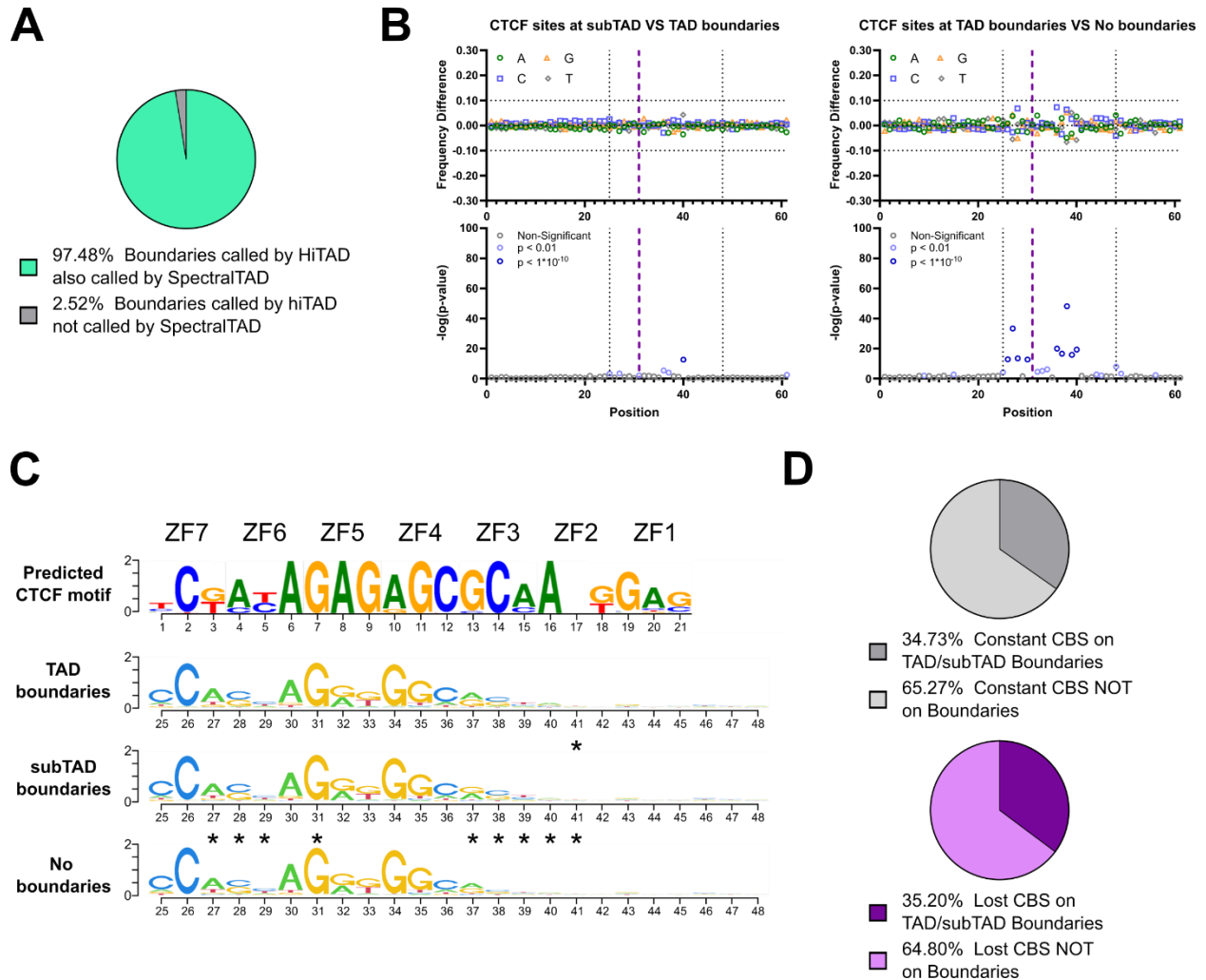) MoMotif analysis of base frequency difference and p-value of bases distribution difference around CTCF-Like motif in lost and gain CBS subsets compared to constant subset in CTCF ZF1^M/- clones. The purple line represents the middle of the CTCF Motif. The dotted line represented the selected region.
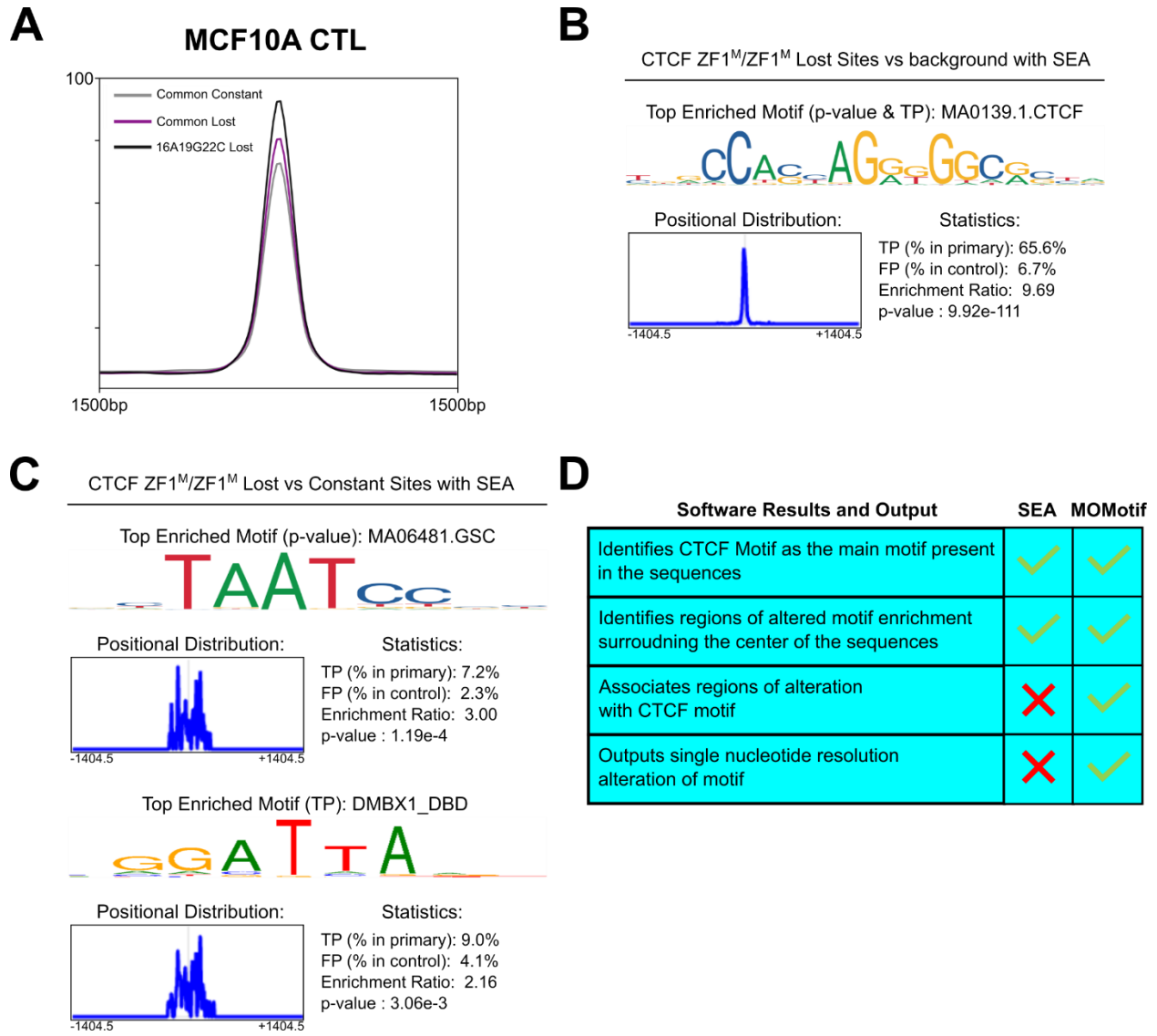
Supplementary Figure 4:

(**A**) Profile plot of CTCF ChIP-Seq read density in CTL MCF10A at commonly constant, lost and lost with the full extended motif. Showing that sites harboring the full extended motif have higher affinity for WT CTCF. (**B**) MEME-Suite SEA analysis and output of CTCF ZF1M/ZF1M lost sites compared to background showing CTCF motif as the top hit for p-value and True Positive (TP) and centrally located on the sequences. (**C**) MEME-Suite SEA analysis and output of CTCF ZF1M/ZF1M lost sites compared to constant sites. Top enriched motifs by p-value and TP shows a less than 10% TP and are located adjacent to the middle of the sequences. (**D**) Summary of the functional differences between SEA and MoMotif.

A

97.48% Boundaries called by HiTAD also called by SpectralTAD

2.52% Boundaries called by hiTAD not called by SpectralTAD

B

CTCF sites at subTAD VS TAD boundaries

CTCF sites at TAD boundaries VS No boundaries

○ A △ G
□ C ◇ T

○ Non-Significant
○ p < 0.01
○ p < 1*10^-10

C

ZF7 ZF6 ZF5 ZF4 ZF3 ZF2 ZF1

Predicted CTCF motif

TAD boundaries

subTAD boundaries

No boundaries

D

34.73% Constant CBS on TAD/subTAD Boundaries

65.27% Constant CBS NOT on Boundaries

35.20% Lost CBS on TAD/subTAD Boundaries

64.80% Lost CBS NOT on Boundaries

Supplementary Figure 5:

(**A**) Pie chart comparing the reproducibility of called TAD and subTAD boundaries from HiTAD (used for this study) and SpectralTAD, an alternative hierarchical TAD caller. (**B**) MoMotif analysis of base frequency difference and p-value of bases distribution difference around CTCF-Like motif in CBS at subTAD boundaries compared to an equal size subset of CBS at TAD boundaries and the subset of CBS at TAD boundaries compared to an equal size subset of CBS located within domains, therefore not on boundaries (+/- ½ bin/5kb) (n = 4915). The purple line represents the middle of the CTCF Motif. The dotted line represented the selected region, which was kept the same as in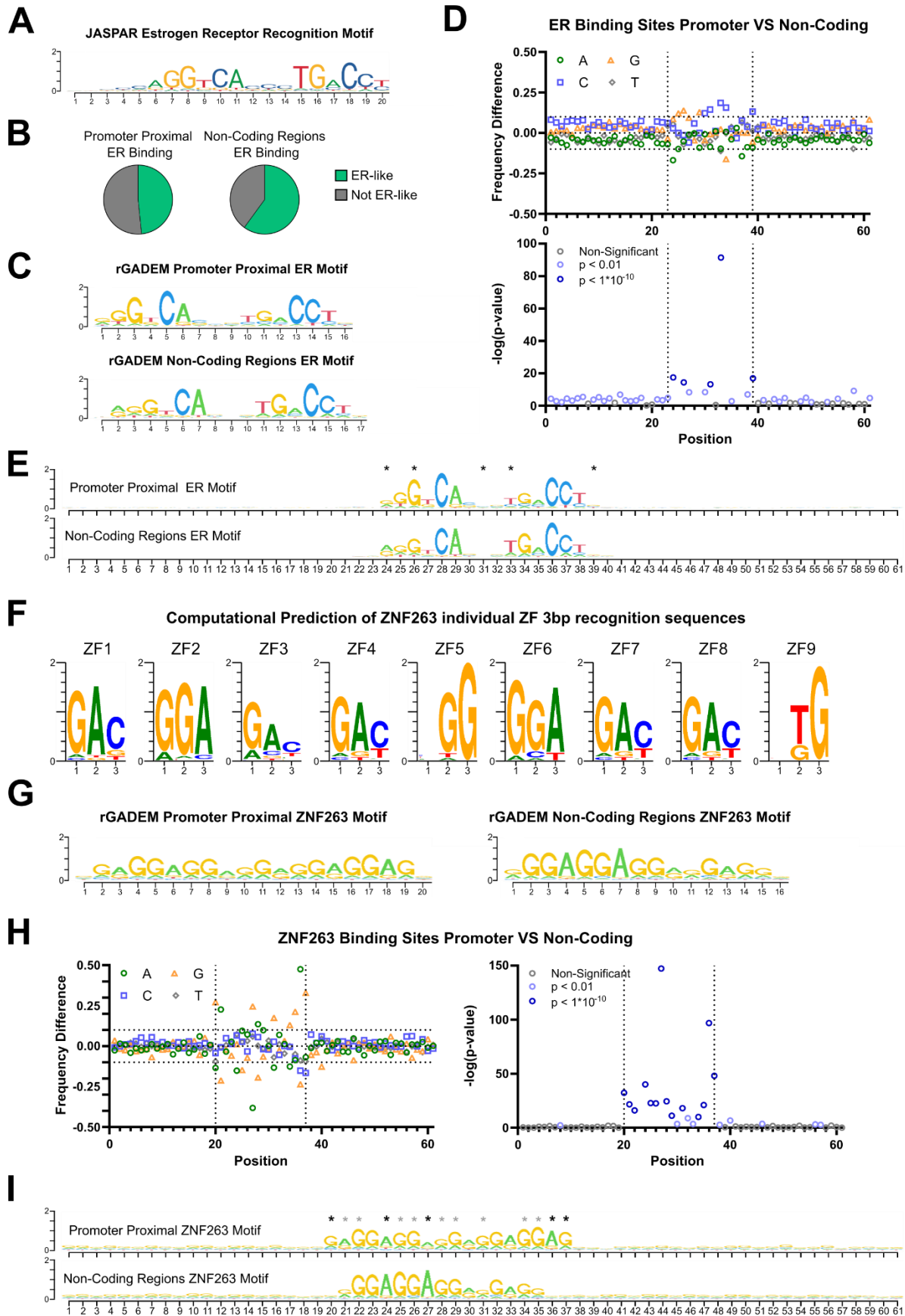 Figure 3 to ease comparison between figures and because no significant changes were observed outside of this region. (**C**) MoMotif results depiction as the height of each nucleotide representing the Shannon Entropy of its occurrence frequency at each position in each group. Asterisk marks individual position with significantly altered base frequencies compared to TAD, highlighting the decreased enrichment of the called bases at these position in CBS not colocalizing with boundaries. (**D**) Pie charts of the frequency of CBS found on TAD/subTAD boundaries or not on boundaries in all CBS constant or lost between CTL and CTCF ZF1M mutated MCF10A. Showing no enrichment of lost CBS in these specific topological contexts.

A

subTAD-B and Lost CBS

- ● **CBS on Boudary p-value = 0.008393**
- ▲ CBS anywhere within TAD p-value = 0.550154
- ■ No CBS p-value = 0.325697

subTAD-B and Gained CBS

- ● CBS on Boudary p-value = 0.407977
- ▲ CBS anywhere within TAD p-value = 0.330922
- ■ No CBS p-value = 0.751563
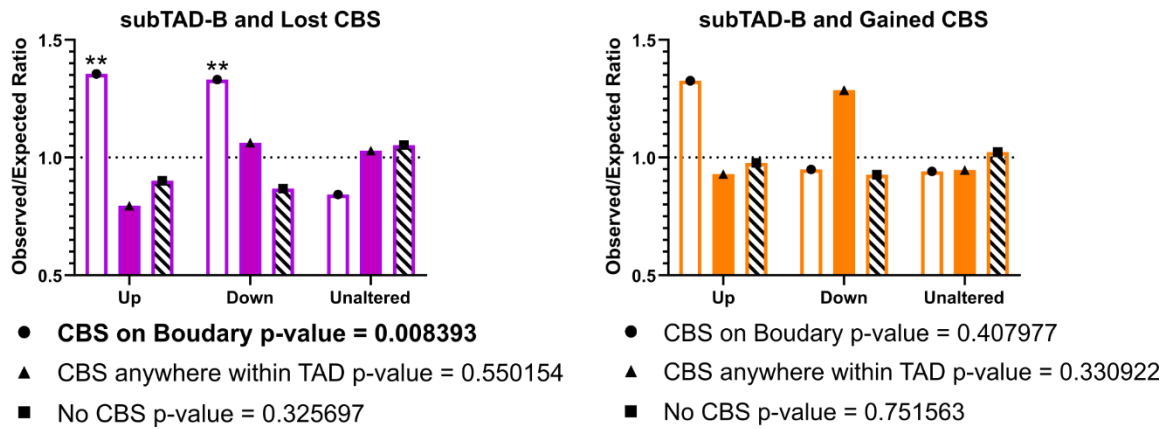
Supplementary Figure 6:

(**A**) Enrichment of strongly up and downregulated genes for different distribution of subTAD, TSS and CBS. Showing that lost of CTCF near a gene at the boundaries of subTAD is significantly predictive of its up or downregulation. (p-values were generated from Chi-Square test on distribution of altered genes)

A JASPAR Estrogen Receptor Recognition Motif

B Promoter Proximal ER Binding / Non-Coding Regions ER Binding
- ER-like
- Not ER-like

C rGADEM Promoter Proximal ER Motif
rGADEM Non-Coding Regions ER Motif

D ER Binding Sites Promoter VS Non-Coding
- A, G, C, T
- Non-Significant, p < 0.01, p < 1*10^-10

E Promoter Proximal ER Motif / Non-Coding Regions ER Motif

F Computational Prediction of ZNF263 individual ZF 3bp recognition sequences
ZF1 ZF2 ZF3 ZF4 ZF5 ZF6 ZF7 ZF8 ZF9

G rGADEM Promoter Proximal ZNF263 Motif / rGADEM Non-Coding Regions ZNF263 Motif

H ZNF263 Binding Sites Promoter VS Non-Coding

I Promoter Proximal ZNF263 Motif / Non-Coding Regions ZNF263 Motif

Supplementary Figure 7:

(**A**) Classical ER DNA binding (MA0112.2) motif from JASPAR. (**B**) Pie chart of the occurrence of the ER-like motif at promoter proximal or non-coding ligand-dependent ER binding sites. (**C**) ER-like motif identified by rGADEM and used as input for MoMotif analysis. (**D**) Single-nucleotide resolution base frequency difference and significance around the ER-like motif in promoter proximal versus non-coding sites. (**E**) MoMotif results depiction as the height of each nucleotide representing the Shannon Entropy of its occurrence frequency at each position for the regions analysed in **D**. Asterisks are placed above bases with significant difference between promoter proximal and non-coding ER binding sites. (**F**) Predicted 3bp sequences recognized by each ZF of ZNF263 by Persikov et al. 2014 and 2015. (**G**) ZNF263 motif identified by rGADEM and used as input for MoMotif analysis. (**H**) Single-nucleotide resolution base frequency difference and significance around the aligned ZNF263 GA rich motif in promoter proximal versus non-coding sites. (**I**) MoMotif results depiction as the height of each nucleotide representing the Shannon Entropy of its occurrence frequency at each position for the regions analysed in **H**. Dark Asterisks are placed above the five most significantly altered bases between promoter proximal and non-coding ER binding sites, showing an extension of the motif in promoter proximal binding sites and an enrichment intra-motif A in non-coding sites. Grey asterisks are placed above all significantly altered bases.

Name	Sequence
CTCF-H284N-Donor	ACATAGGTGTAAAGAAGACATTCCAGTGTGAGCTTTGCAGTTACACGTGTCCAC GGCGTTCAAATTTGGATCGTAACATGAAAAGCCACACTGATGAGAGACCACACA AGTGCCATCTCTGTGGCAGGGCATTCAGAACAGTCACCCTCC
CTCF-H284N-sgRNA F	CACCGCCACGGCGTTCAAATTTGGATCG
CTCF-H284N-sgRNA R	CGGTGCCGCAAGTTTAAACCTAGCCAAA

Supplementary Table 1: DNA donor to insert the H284N mutation coupled with the small guide RNAs targeting CTCF for the CRISPR-Cas9 experiment.