SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure S1. Micro-C and Hi-C maps of human pluripotent and differentiated cell types. Related to Figure 1.

(A) Examples of intermediate stages of the Micro-C protocol for ESCs (top panel) and HFFs (bottom panel). In both panels, lanes include Marker (M), Input (I), and Proximity Ligation (PL) samples. Key features shown here are the digestion level of Input material to ~90% mononucleosomes, and the shift following proximity ligation to dinucleosome-sized ligation products. Designations above samples (eg R2T1) indicate the biological replicate (R1-3) and technical replicate (T1 etc.) loaded.

(B) Hi-C interaction decay curves for differentiated cells in this study are comparable with prior in situ Hi-C datasets for differentiated cells. Interaction decay curves are shown for in situ Hi-C and Micro-C analysis of HFFs (this study), compared to two published in situ Hi-C datasets for GM12878 cells (Rao et al., 2014).

(C) Interaction decay curves for ESCs showing minor effects of downsampling Micro-C data to match the cis-interacting read depth from the Hi-C dataset (1.09 billion reads).

(D) Interaction decay curve with data renormalized following removal of all interactions between loci closer than 10,000 bp. It is important to note here that removal of interactions from short ranges will by necessity result in an upward shift of the Micro-C curve. This is due to the better coverage of short range interactions in Micro-c relative to Hi-C. The impact of this effect depends on the distance chosen for such an exercise. Micro-C will continue to exhibit lower long-distance (>1 Mb) interactions than Hi-C (as seen here), or ultimately after more aggressive removal of short-range data the ultra-long distance interactions will overlap. However, this would artifactually inflate interactions throughout intermediate distances (~10 kb to 1 Mb) for Micro-C.

In other words, selective read removal followed by normalization could in principle lead to a range spanning three interpretations: Micro-C exhibits equivalent long-distance interactions to Hi-C but much higher intermediate range interactions (short range reads removed), Micro-C exhibits similar mid-range interactions with higher short-range interactions and a lower noise floor (**Figure 1C**), or Micro-C captures more short-range interactions with reduced capture of intermediate range interactions (long range reads removed). In our view the second interpretation, which does not involve arbitrarily choosing interactions to remove and which is presented in **Figure 1C**, is the most natural normalization scheme, but as with any genome-wide dataset it is always valuable to consider effects of normalization on biological interpretations.

(E) Interaction decay curves for Micro-C maps of ESCs. Here, read pairs are separated according to their relative orientation. Note the excess of IN-IN interactions at short distance, attributable to undigested dinucleosomes. Inset shows zoom-in.

(F) As in panel (E) inset, but with reads shifted 73 bp to the nucleosome dyad, thereby aligning IN-IN/IN-OUT/OUT-OUT read pairs.

(G) Step-wise decrease in interactions between adjacent nucleosomes. Each bar shows the ratio of ligation product abundance across two adjacent nucleosomes – the first set of bars shows the ratio of N/N+1 over N/N+2. Bars are shown for IN/IN, IN-OUT, OUT-IN, and OUT-OUT orientations. Notably, there is a greater dropoff from N+1 to N+2, and from N+3 to N+4,

than there is between the nucleosome pairs N+2/3 and N+4/5, potentially consistent with a zig-zag fiber architecture. These data suggest that in humans, in contrast to budding yeast, compacted chromatin fiber may extend beyond tetranucleosomes to organize somewhat longer stretches of the genome. Nonetheless, there is a marked dropoff from N/N+2 to N/N+4 and again from N/N+4 to N/N+6 – whether this represents inefficient extension of chromatin fiber compaction beyond the tetranucleosome, or a technical inability to recover N/N+4 products in a zig-zag fiber thanks to the interposed N+2 nucleosome, remains to be determined, although several independent analyses of in vivo chromatin folding (Ou et al., 2017; Ricci et al., 2015) support the former view: that chromatin forms fairly short, heterogeneous “clumps” of zig-zag fiber in vivo.

(H) Plots show frequency of Hi-C or Micro-C interactions > 10 kb (y axis) for genomic 2 kb bins with varying numbers of DpnII target sequences (x axis), revealing strong bias for poor Hi-C coverage for genomic intervals depleted of DpnII target sites.

Supplementary Figure S2. Increased compartmental organization of HFFs. Related to Figure 1.

(A) Single-gene scale compartment. Left panel shows a broad zoom (chr11: 27,384,310-33,423,664) showing compartment signature (enriched interactions at long distance with the compartment checkerboard pattern) for this gene, while right panel shows the gene-scale zoom-in.

(B) Micro-C contact maps for ESCs and HFFs (above and below the diagonal, respectively) for the indicated chromosomes. The characteristic checkerboard compartment pattern is clearly stronger in HFFs compared to ESCs.

(C-D) Differences in expression between ESCs and HFFs are correlated with changes in compartment organization. Panel (C) shows a typical Volcano plot, in which genes are plotted according to the difference in expression between HFFs and ESCs (x axis) and the p value for the significance of the change in expression (y axis). Dots are colored according to the compartment change for each gene between the two cell types – red dots show genes preferentially located in the active compartment in ESCs relative to HFFs, and these genes are primarily those that are more highly-expressed in ESCs. This was highly statistically-significant: confining the analysis to highly induced/repressed genes (log2(fold change) > 5 and -log10(p-value) > 5), we find a Spearman Correlation of 0.469 between changes in mRNA abundance and compartment score, with a p-value of 1.18e-139. Panel (D) shows a compartment-focused analysis, with each gene’s relative ranking in compartment score (higher = enriched in active compartment) plotted for the two cell types. Genes are colored by the difference in expression between ESCs and HFFs. Again, genes with higher active compartment scores in HFFs (dots in the lower right corner) are enriched for genes that are activated in HFFs vs. ESCs (blue dots), and vice versa.

Supplementary Figure S3. Characterization of boundaries between contact domains. Related to Figures 2 and 3.

(A) Near-diagonal interactions at TSSs broken down by gene orientation and expression level. Top and bottom panels show all + and - strand genes, respectively. For each group, three panels are shown for low, intermediate, and high expression as defined by CAGE-Seq data (ENCODE accession# ENCFF038OTF). Notable here is the well-positioned nucleosome apparent at the +1 position, and the increased boundary-related “clearing” in the lower left/upper right quadrants at increased transcription rates. Also curious is the asymmetry in boundary-crossing interactions, with the +1 nucleosome interacting with upstream sequences but little to no interaction of -1/-2 nucleosomes with the downstream coding sequence.

(B) Boundary calls are not sensitive to parameter choice. Boundaries were identified using a sliding window approach in which, for any given locus, the number of crossing interactions was calculated within some distance, relative to interaction frequency in windows on either side of the locus. Boundaries are identified as local minima in this insulation score. This approach has two free parameters: data bin size, and width of the sliding window. Here, boundaries were calculated using one set of parameters (100 bp bins, 1000 bp window), and insulation scores are shown in heatmaps for these and two additional parameter choices.

(C) Insulation scores scale with promoter transcriptional activity. Here, we analyzed all annotated promoters in human, splitting promoters into quintiles based on the mRNA abundance of the downstream gene in either ESCs or HFFs (Quintile 1 = highest expression, Quintile 5 = lowest). For each quintile, the distribution of Micro-C insulation scores (negative = stronger insulation) is shown as a violin plot, revealing a strong overall relationship between promoter activity and insulation.

(D) Changes in transcription during differentiation are accompanied by changes in promoter insulation scores. Scatterplot shows insulation scores at all promoters exhibiting a 2-fold change in expression between ESCs and HFFs, comparing insulation scores in Micro-C datasets for HFFs (x axis) and ESCs (y axis). Dots are color-coded by changes in associated mRNA abundance, with genes upregulated in HFFs colored red and those downregulated in HFFs colored blue. Although insulation scores are generally well-correlated between cell types (dots along diagonal), two prominent groups of promoters fall off the diagonal, with HFF-upregulated promoters showing stronger insulation in HFFs, and HFF-downregulated promoters losing insulation activity.

(E) Overview of molecular features of weak boundaries. Heatmap from **Figure 3G**, with complete list of all feature enrichments at Cluster IV boundaries.

(F) Boundary-level view of molecular correlates of weak boundaries. Heatmaps show ChIP (or DNase) signal for Cluster IV boundaries from **Figures 3E-G**. Several features, notably nuclease hypersensitivity, are associated with weak boundary elements.

Supplementary Figure S4. Examples of enhancer-promoter and Polycomb-mediated looping interactions. Related to Figure 4.

(A-B) Contact maps are shown (rotated 45 degrees) for the indicated genomic intervals showing enhancer-promoter interactions (A) or Polycomb-associated interactions (B), along with public data for CTCF and select other marks related to chromosome architecture. A subset of looping

interactions are indicated with arrowheads, highlighting examples both of loops linking CTCF binding sites, as well as CTCF-depleted looping interactions.

(C) Loop averages for HFFs. Data for Hi-C only, Micro-C only, and shared loop anchors are shown as in **Figure 4C**, but for the HFF dataset.

Supplementary Figure S5. CTCF-associated looping interactions. Related to Figure 4.

(A) Distribution of CTCF ChIP-Seq enrichment for all ESC loop anchors, compared to CTCF ChIP enrichment for an equal number of loci randomly shifted by distances between 80-160 kb from loop anchors.

(B-C) Loop averages in ESCs for CTCF motif pairs in all four orientations for the Hi-C dataset. (B) shows data for 40 kb windows surrounding CTCF motifs, while (C) shows a 3 kb zoom-in.

(D-E) Loop averages for the Micro-C dataset, arranged as in panels (B-C). Notable here is the presence of signal for all four motif orientations, with asymmetric loop extrusion flares in the two tandem orientations consistent with CTCF motifs oriented “towards” a loop extruder exerting much stronger barrier activity than outwardly-oriented CTCF motifs. In addition, although “blurry” signal is apparent in Hi-C maps at lower resolution (compare (B) and (D)), the signal is much sharper in Micro-C maps at this resolution. At the higher resolution shown in (C) and (E), the blurring of the Hi-C signal almost completely obscures loop anchors, while these are clearly resolved in the Micro-C dataset.

Supplementary Figure S6. Molecular characteristics of loop anchors. Related to Figure 4.

(A) Heatmap from **Figure 4E** is reproduced here in greater detail

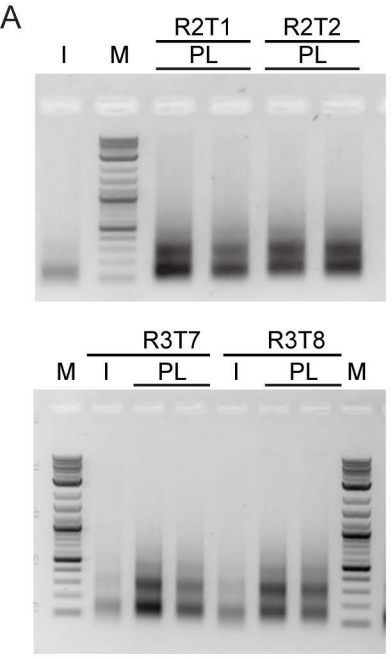
(B-C) Chromatin states for CTCF-enriched and -depleted loop anchors, using two different definitions for CTCF enrichment/depletion (see **Supplementary Figure S5A**) as indicated.

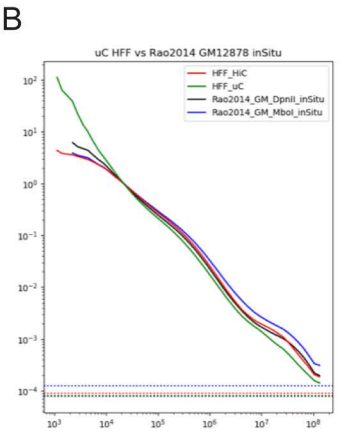
Supplementary Figure S7. Looping interactions at CTCF-depleted loci. Related to Figure 4.

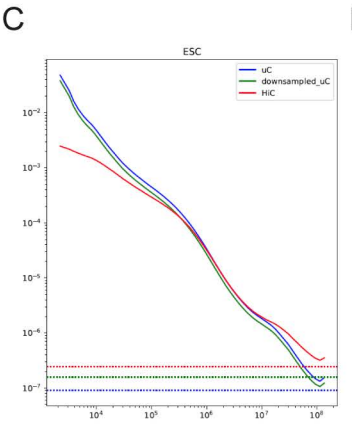
(A) Average contact maps for TSSs (center of x axis) with the nearest enhancer, in both cases excluding regulatory elements overlapping with a significant CTCF ChIP-Seq peak. TSSs are sorted into quintiles based on PolII ChIP-seq signal at promoters.

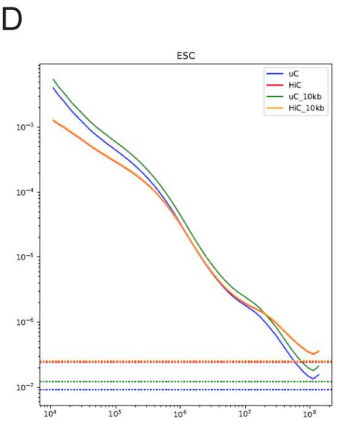
(B) Loops between paired sites for various structural proteins and histone marks. In all cases, we first excluded ChIP peaks for these factors if they fell within 10 kb of a CTCF peak. From this set of peaks, heatmaps show averaged signal for peak pairs falling farther than 5 kb from one another. Red labels indicate factors with <500 peaks in this analysis.

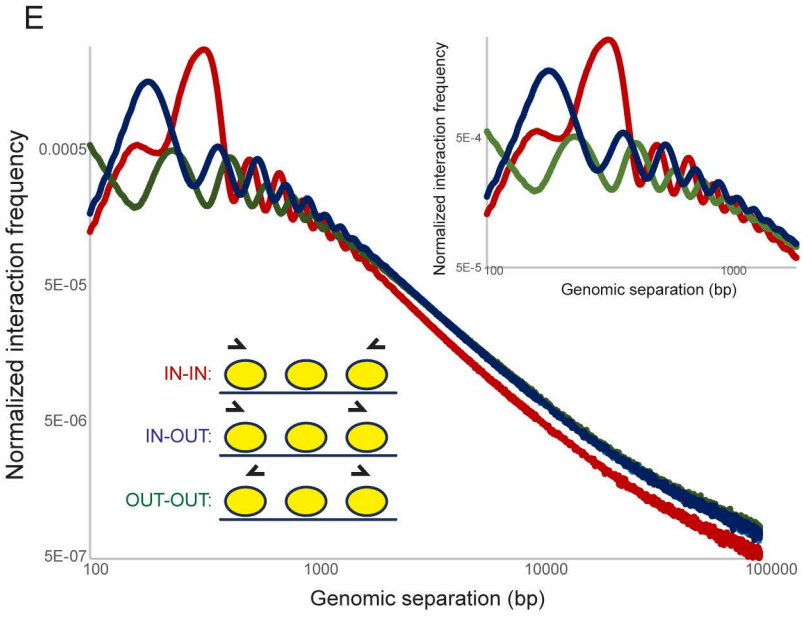
Figure S1

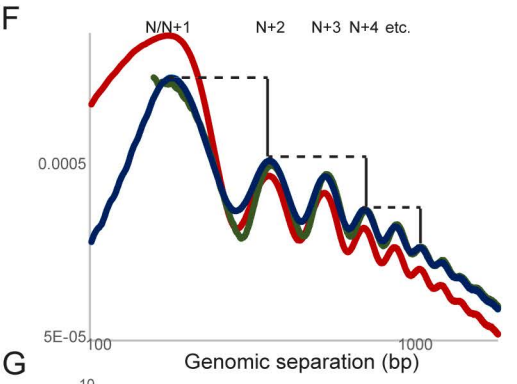

A
I
M
R2T1
PL
R2T2
PL
R3T7
R3T8
M
I
PL
I
PL
M

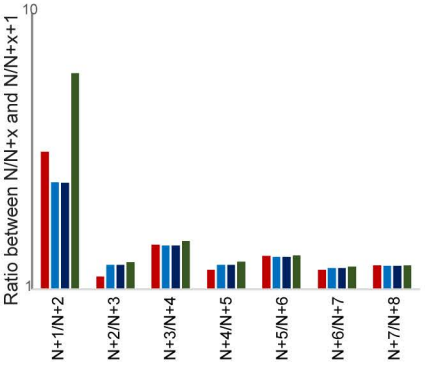

B
uC HFF vs Rao2014 GM12878 inSitu
HFF_HiC
HFF_uC
Rao2014_GM_DpnII_inSitu
Rao2014_GM_MboI_inSitu


C
ESC
uC
downsampled_uC
HiC


D
ESC
uC
HiC
uC_10kb
HiC_10kb


E
Normalized interaction frequency
0.0005
5E-05
5E-06
5E-07
100
1000
10000
100000
Genomic separation (bp)
Normalized interaction frequency
5E-4
5E-5
100
1000
Genomic separation (bp)
IN-IN:
IN-OUT:
OUT-OUT:


F
N/N+1
N+2
N+3
N+4
etc.
0.0005
5E-05
100
1000
Genomic separation (bp)


G
Ratio between N/N+x and N/N+x+1
10
1
N+1/N+2
N+2/N+3
N+3/N+4
N+4/N+5
N+5/N+6
N+6/N+7
N+7/N+8

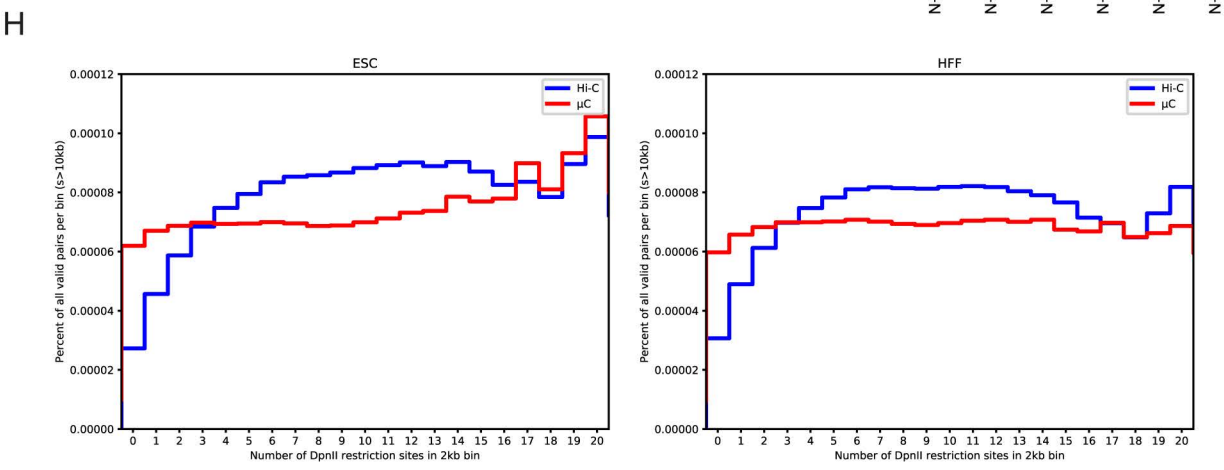

H
ESC
HFF
Hi-C
µC
Percent of all valid pairs per bin (s>10kb)
0.00012
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Number of DpnII restriction sites in 2kb bin

Figure S2

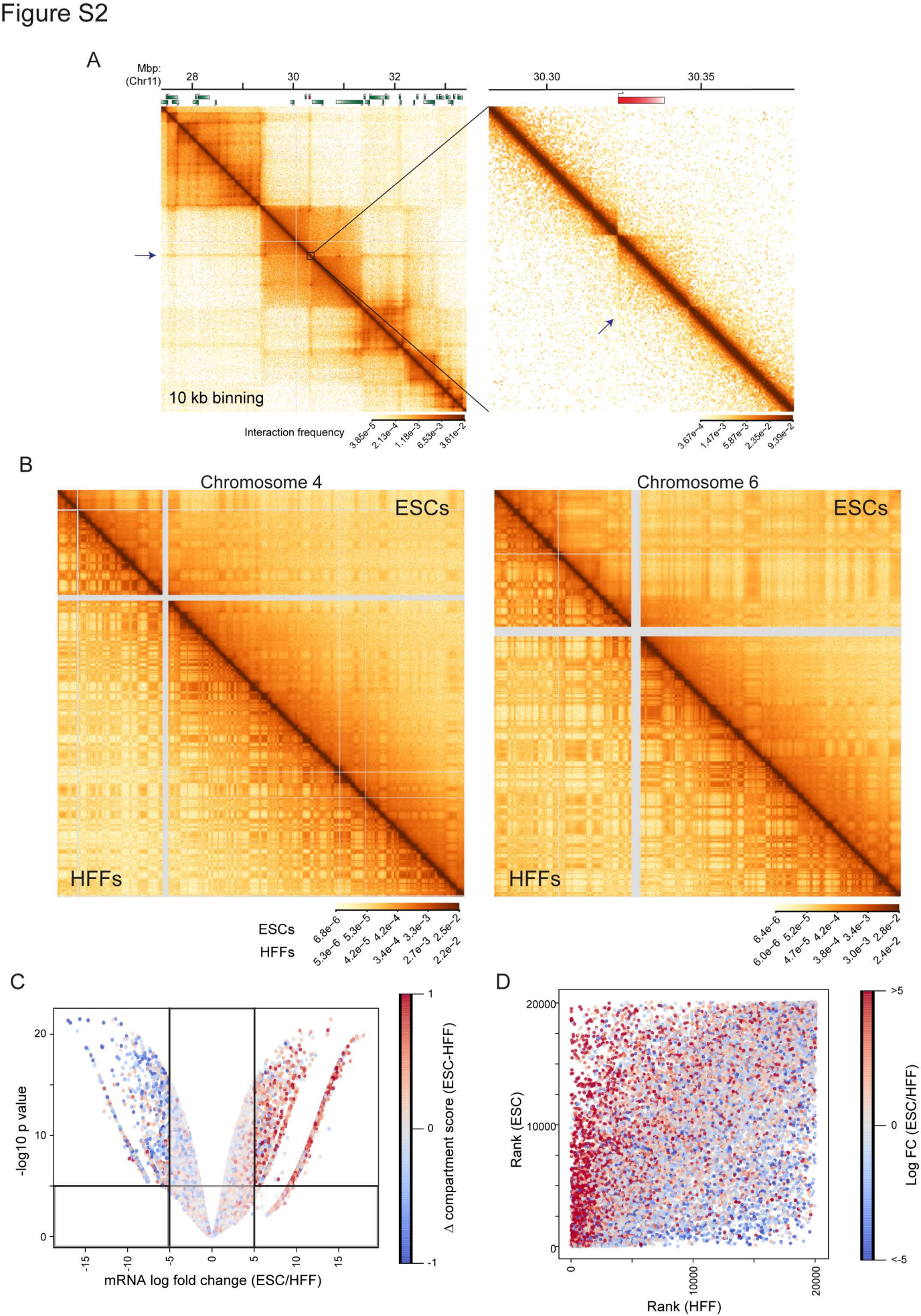

A
Mbp:
(Chr11)
28
30
32
30.30
30.35
10 kb binning
Interaction frequency
3.85e−5
2.13e−4
1.18e−3
6.53e−3
3.61e−2
3.67e−4
1.47e−3
5.87e−3
2.35e−2
9.39e−2
B
Chromosome 4
ESCs
HFFs
Chromosome 6
ESCs
HFFs
ESCs
HFFs
6.8e−6
5.3e−5
4.2e−4
3.3e−3
2.5e−2
5.3e−6
4.2e−5
3.4e−4
2.7e−3
2.2e−2
6.4e−6
5.2e−5
4.2e−4
3.4e−3
2.8e−2
6.0e−6
4.7e−5
3.8e−4
3.0e−3
2.4e−2
C
-log10 p value
mRNA log fold change (ESC/HFF)
Δ compartment score (ESC-HFF)
1
0
-1
D
Rank (ESC)
Rank (HFF)
Log FC (ESC/HFF)
>5
0
<-5

Figure S3

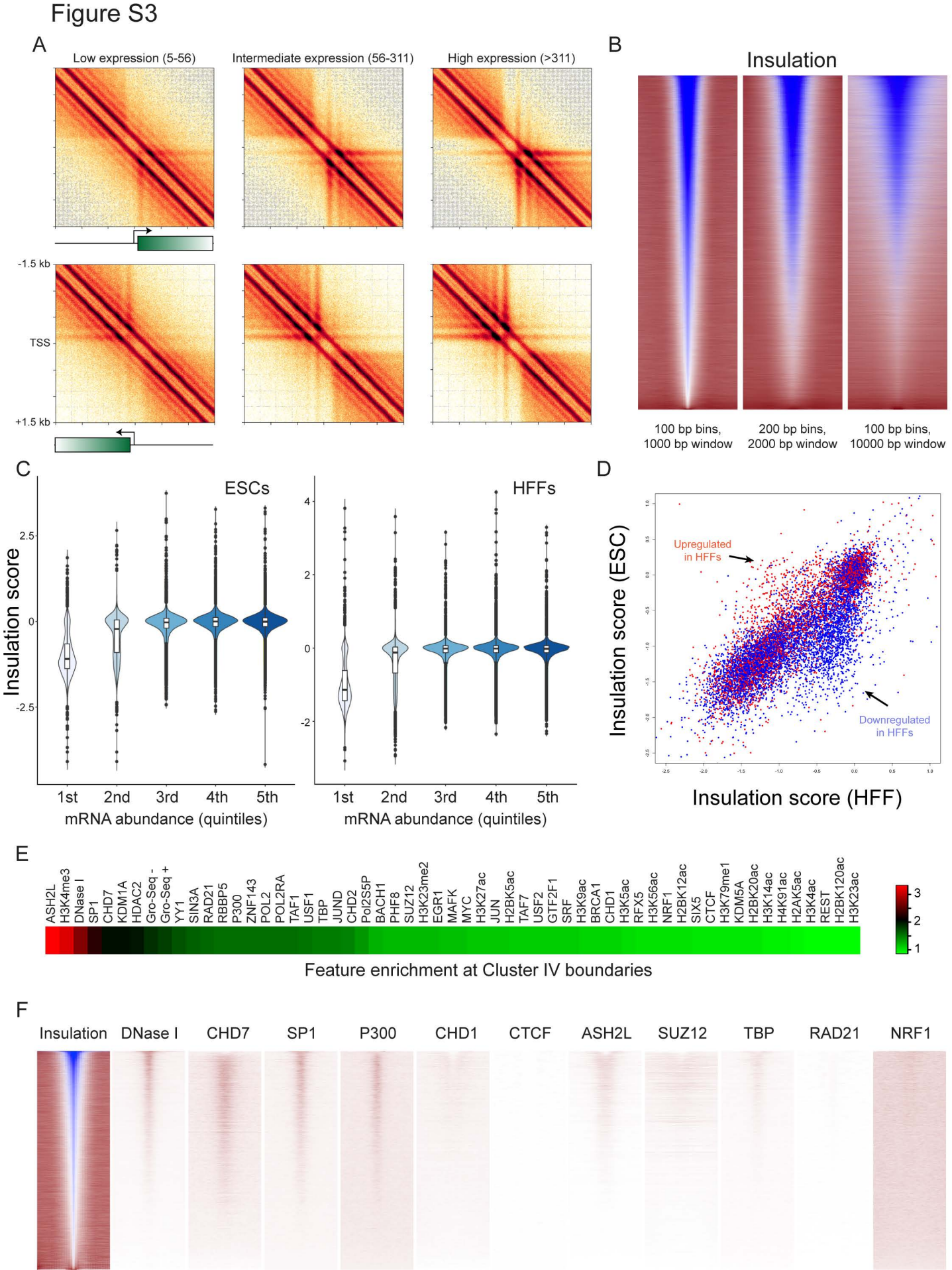
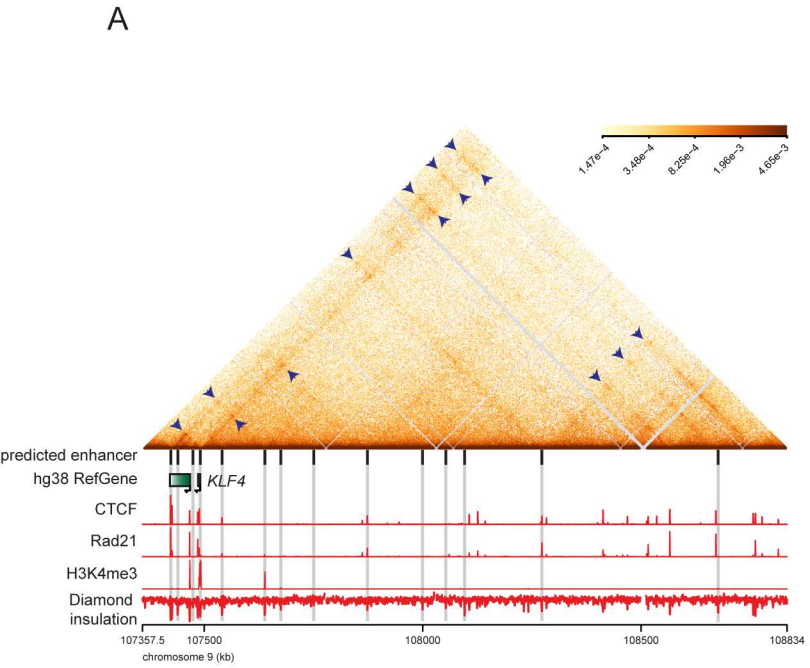
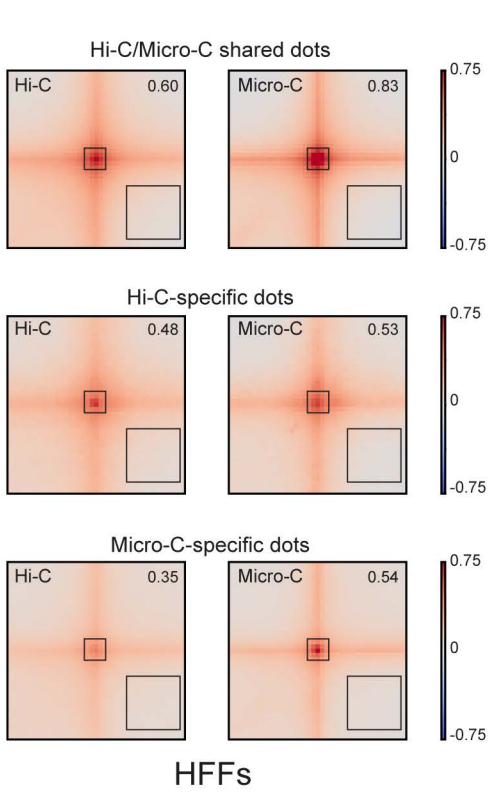

A
Low expression (5-56)
Intermediate expression (56-311)
High expression (>311)
-1.5 kb
TSS
+1.5 kb
B
Insulation
100 bp bins, 1000 bp window
200 bp bins, 2000 bp window
100 bp bins, 10000 bp window
C
ESCs
HFFs
Insulation score
mRNA abundance (quintiles)
1st
2nd
3rd
4th
5th
D
Insulation score (ESC)
Insulation score (HFF)
Upregulated in HFFs
Downregulated in HFFs
E
ASH2L
H3K4me3
DNase I
SP1
CHD7
KDM1A
HDAC2
Gro-Seq -
Gro-Seq +
YY1
SIN3A
RAD21
RBBP5
P300
ZNF143
POL2
POL2RA
TAF1
USF1
TBP
JUND
CHD2
Pol2S5P
BACH1
PHF8
SUZ12
H3K23me2
EGR1
MAFK
MYC
H3K27ac
JUN
H2BK5ac
TAF7
USF2
GTF2F1
SRF
H3K9ac
BRCA1
CHD1
H3K5ac
RFX5
H3K56ac
NRF1
H2BK12ac
SIX5
CTCF
H3K79me1
KDM5A
H2BK20ac
H3K14ac
H4K91ac
H2AK5ac
H3K4ac
REST
H2BK120ac
H3K23ac
Feature enrichment at Cluster IV boundaries
F
Insulation
DNase I
CHD7
SP1
P300
CHD1
CTCF
ASH2L
SUZ12
TBP
RAD21
NRF1

Figure S4


A
1.47e−4
3.48e−4
8.25e−4
1.96e−3
4.65e−3
predicted enhancer
hg38 RefGene
KLF4
CTCF
Rad21
H3K4me3
Diamond insulation
107357.5
107500
108000
108500
108834
chromosome 9 (kb)


C
Hi-C/Micro-C shared dots
Hi-C
0.60
Micro-C
0.83
0.75
0
-0.75
Hi-C-specific dots
Hi-C
0.48
Micro-C
0.53
0.75
0
-0.75
Micro-C-specific dots
Hi-C
0.35
Micro-C
0.54
0.75
0
-0.75
HFFs

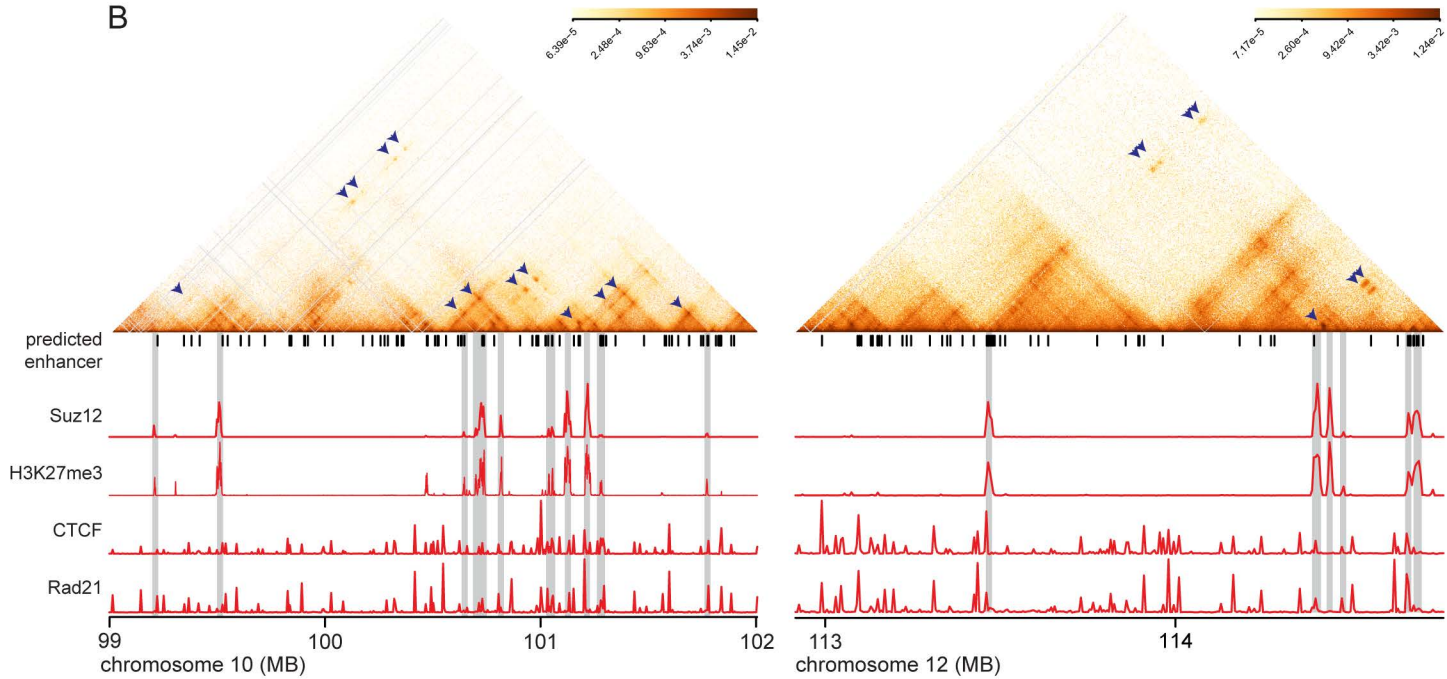
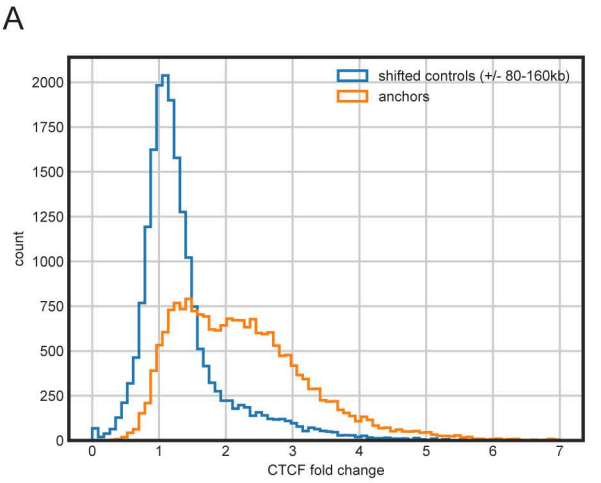
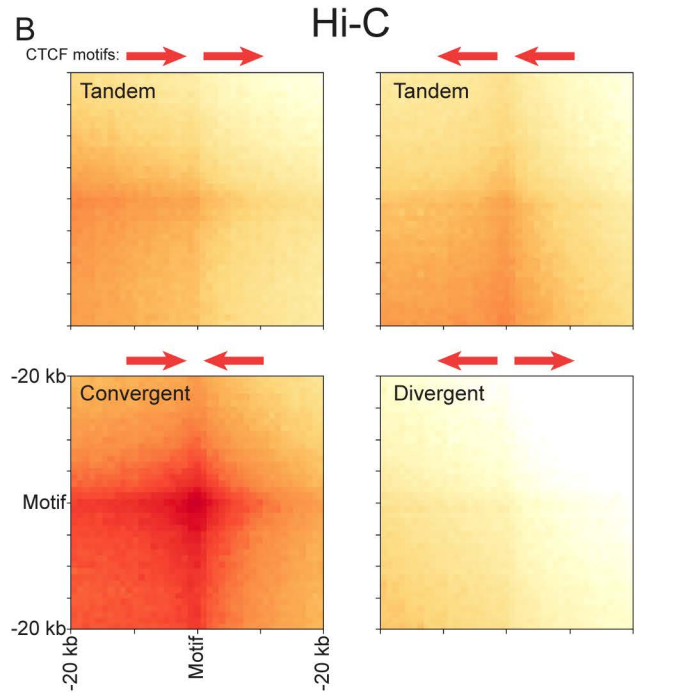
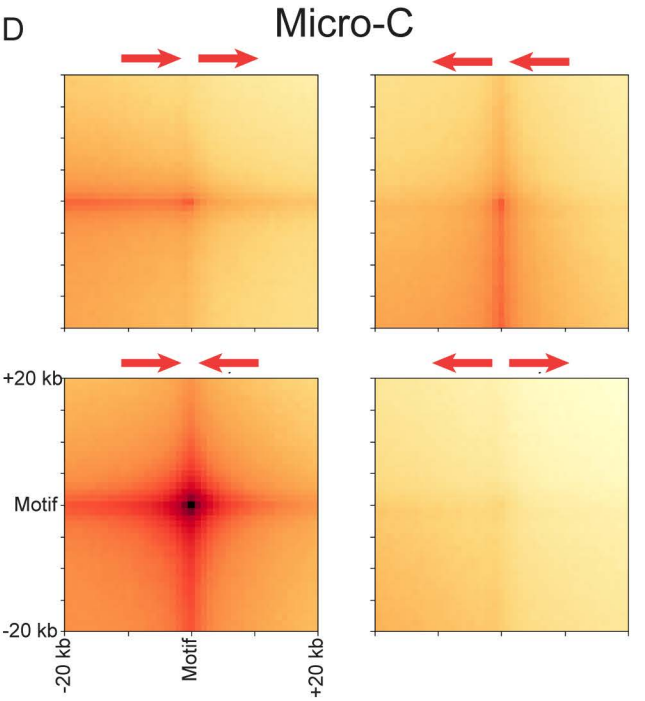
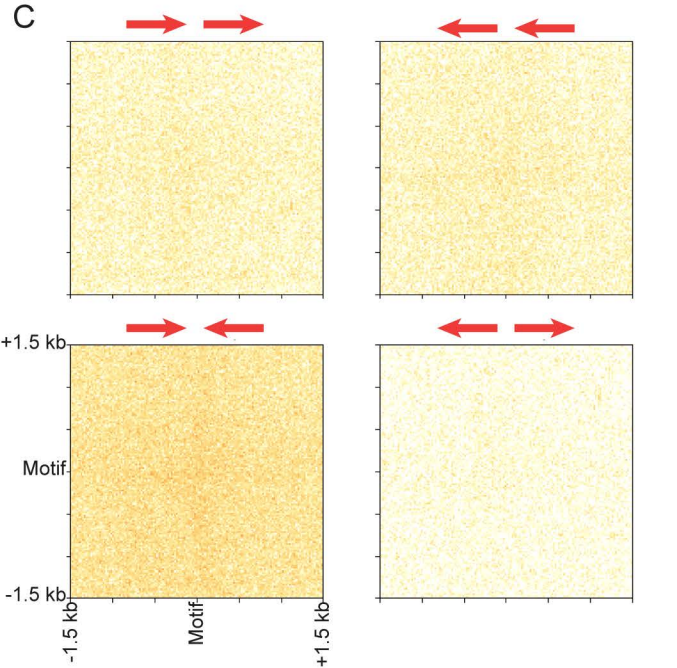

B
6.39e−5
2.48e−4
9.63e−4
3.74e−3
1.45e−2
7.17e−5
2.60e−4
9.42e−4
3.42e−3
1.24e−2
predicted enhancer
Suz12
H3K27me3
CTCF
Rad21
99
100
101
102
chromosome 10 (MB)
113
114
chromosome 12 (MB)

Figure S5


A
shifted controls (+/- 80-160kb)
anchors
count
CTCF fold change
0
250
500
750
1000
1250
1500
1750
2000
0
1
2
3
4
5
6
7


B
Hi-C
CTCF motifs:
Tandem
Tandem
Convergent
Divergent
-20 kb
Motif
-20 kb
-20 kb
Motif
-20 kb


D
Micro-C
+20 kb
Motif
-20 kb
-20 kb
Motif
+20 kb


C
+1.5 kb
Motif
-1.5 kb
-1.5 kb
Motif
+1.5 kb

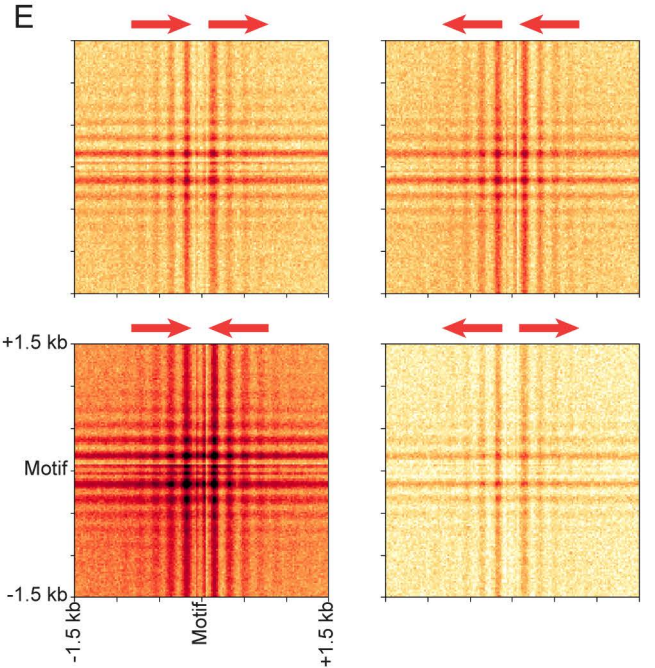
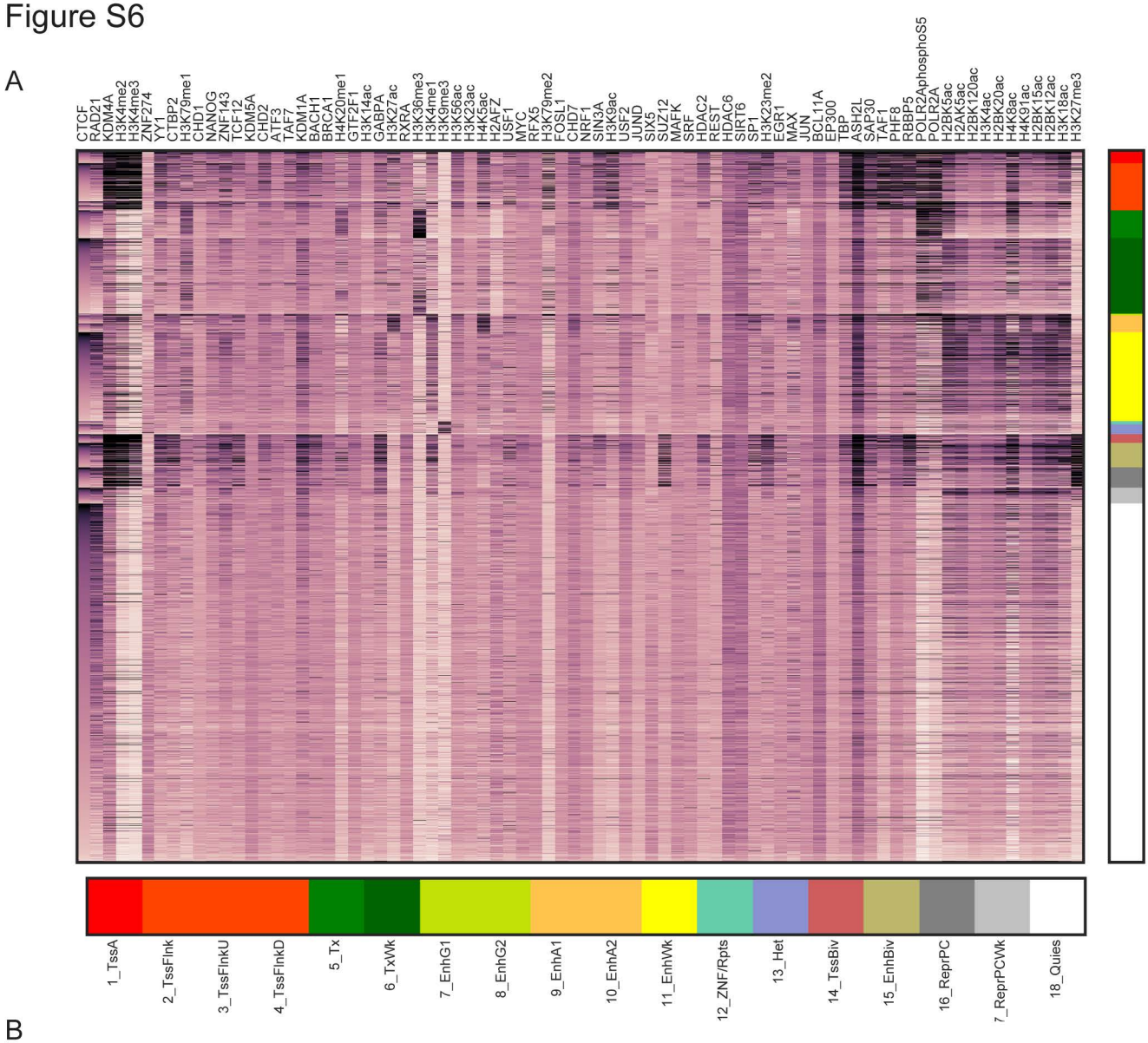

E
+1.5 kb
Motif
-1.5 kb
-1.5 kb
Motif
+1.5 kb

Figure S6


A
CTCF
RAD21
KDM4A
H3K4me2
H3K4me3
ZNF274
YY1
CTBP2
H3K79me1
CHD1
NANOG
ZNF143
TCF12
KDM5A
CHD2
ATF3
TAF7
KDM1A
BACH1
BRCA1
H4K20me1
GTF2F1
H3K14ac
GABPA
H3K27ac
RXRA
H3K36me3
H3K4me1
H3K9me3
H3K56ac
H3K23ac
H4K5ac
H2AFZ
USF1
MYC
RFX5
H3K79me2
FOSL1
CHD7
NRF1
SIN3A
H3K9ac
USF2
JUND
SIX5
SUZ12
MAFK
SRF
HDAC2
REST
HDAC6
SIRT6
SP1
H3K23me2
EGR1
MAX
JUN
BCL11A
EP300
TBP
ASH2L
SAP30
TAF1
PHF8
RBBP5
POLR2AphosphoS5
POLR2A
H2BK5ac
H2AK5ac
H2BK120ac
H3K4ac
H2BK20ac
H4K8ac
H4K91ac
H2BK15ac
H2BK12ac
H3K18ac
H3K27me3
1_TssA
2_TssFlnk
3_TssFlnkU
4_TssFlnkD
5_Tx
6_TxWk
7_EnhG1
8_EnhG2
9_EnhA1
10_EnhA2
11_EnhWk
12_ZNF/Rpts
13_Het
14_TssBiv
15_EnhBiv
16_ReprPC
7_ReprPCWk
18_Quies

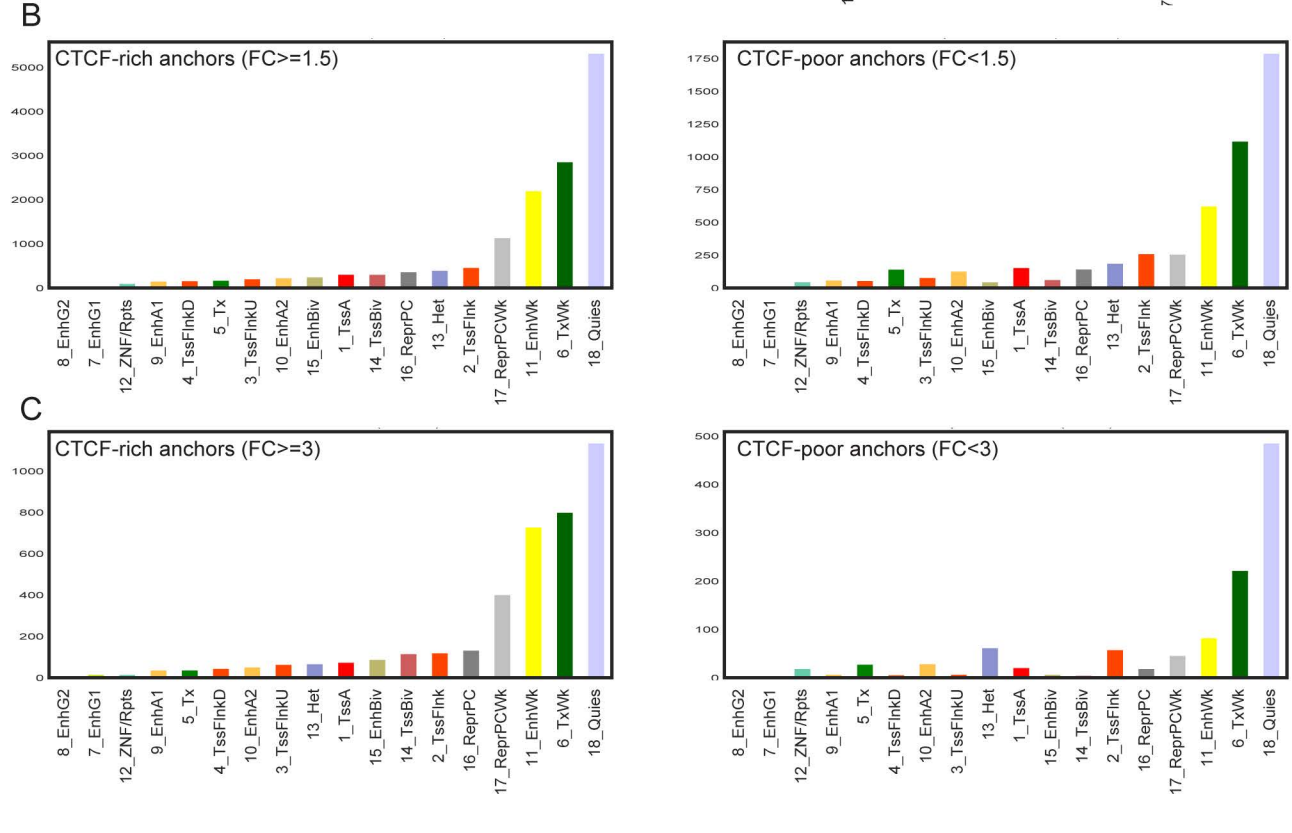
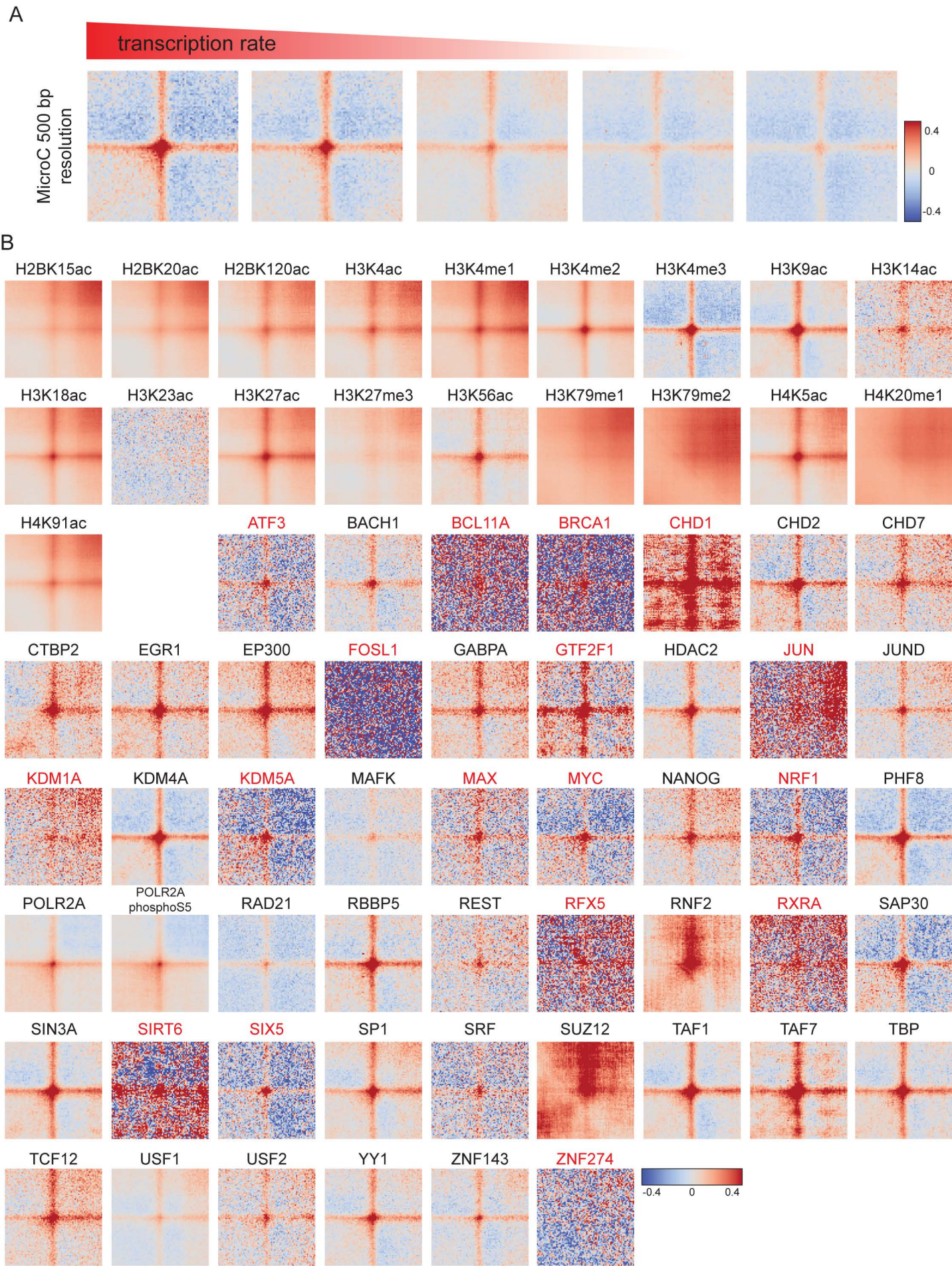

B
CTCF-rich anchors (FC>=1.5)
CTCF-poor anchors (FC<1.5)
C
CTCF-rich anchors (FC>=3)
CTCF-poor anchors (FC<3)

Figure S7


A
transcription rate
MicroC 500 bp resolution
0.4
0
-0.4
B
H2BK15ac
H2BK20ac
H2BK120ac
H3K4ac
H3K4me1
H3K4me2
H3K4me3
H3K9ac
H3K14ac
H3K18ac
H3K23ac
H3K27ac
H3K27me3
H3K56ac
H3K79me1
H3K79me2
H4K5ac
H4K20me1
H4K91ac
ATF3
BACH1
BCL11A
BRCA1
CHD1
CHD2
CHD7
CTBP2
EGR1
EP300
FOSL1
GABPA
GTF2F1
HDAC2
JUN
JUND
KDM1A
KDM4A
KDM5A
MAFK
MAX
MYC
NANOG
NRF1
PHF8
POLR2A
POLR2A phosphoS5
RAD21
RBBP5
REST
RFX5
RNF2
RXRA
SAP30
SIN3A
SIRT6
SIX5
SP1
SRF
SUZ12
TAF1
TAF7
TBP
TCF12
USF1
USF2
YY1
ZNF143
ZNF274
-0.4
0
0.4