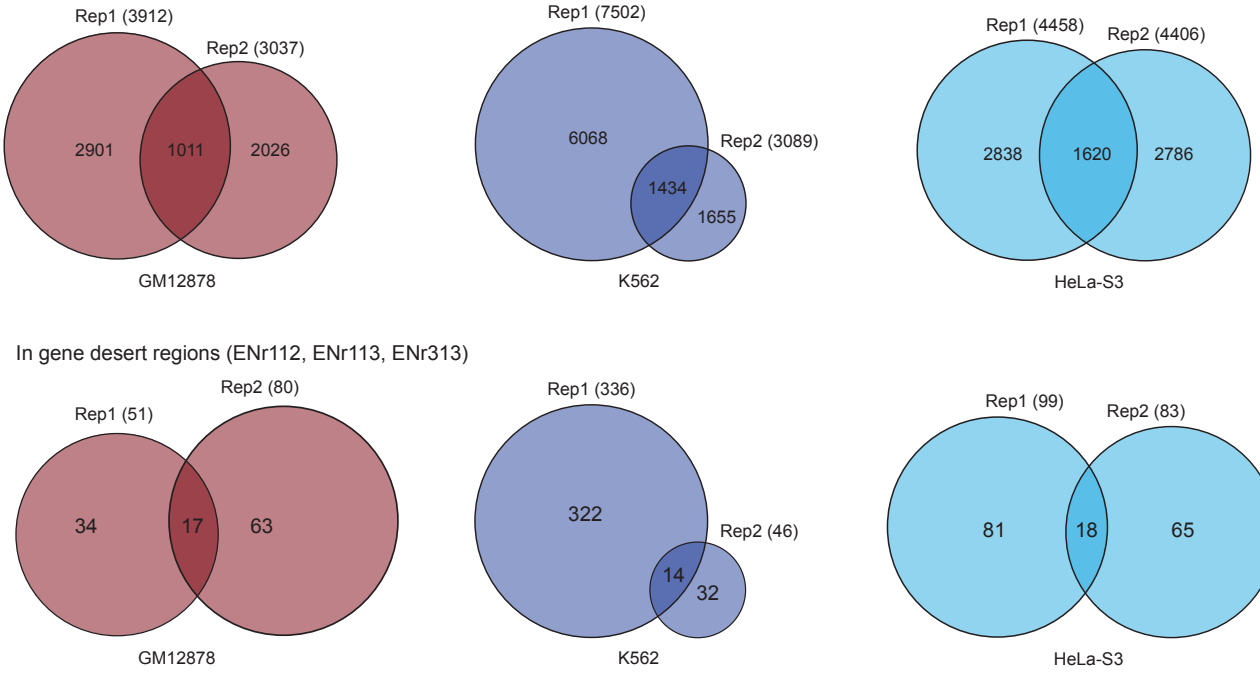
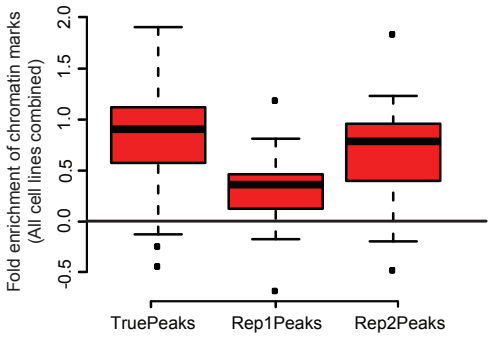
doi:10.1038/nature11279

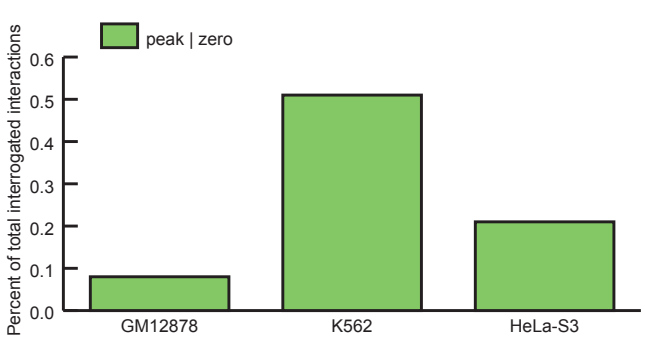
a Supplementary Figure 1



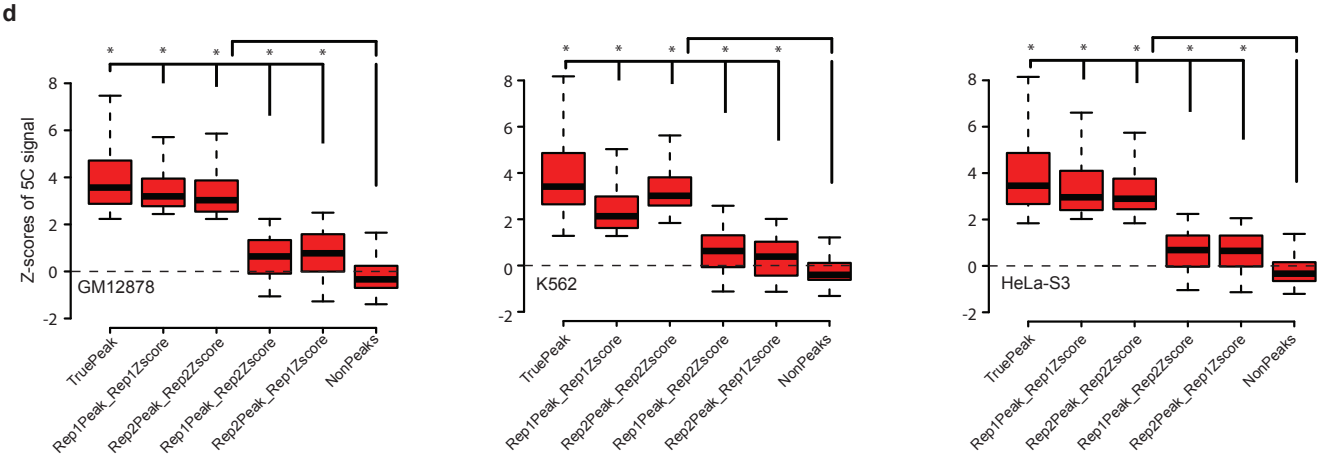
b

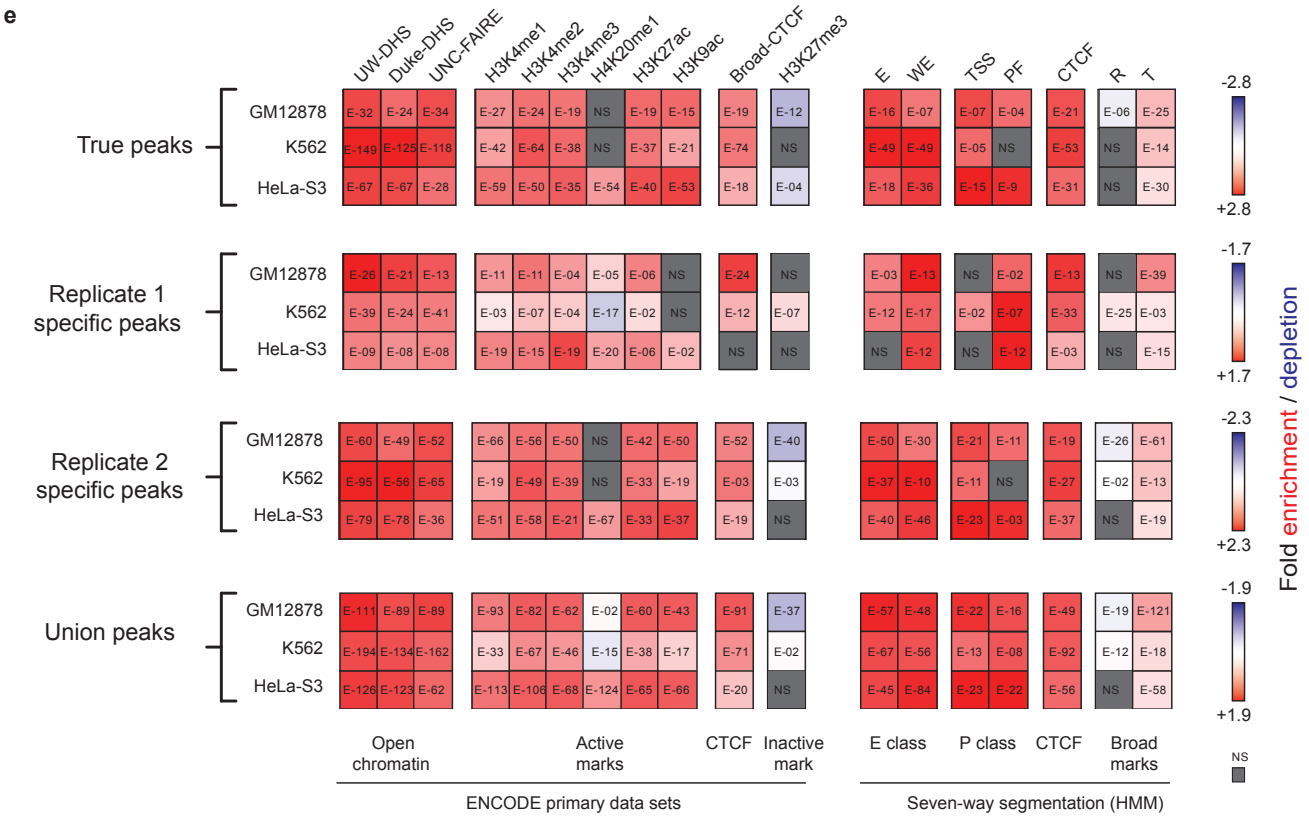


c

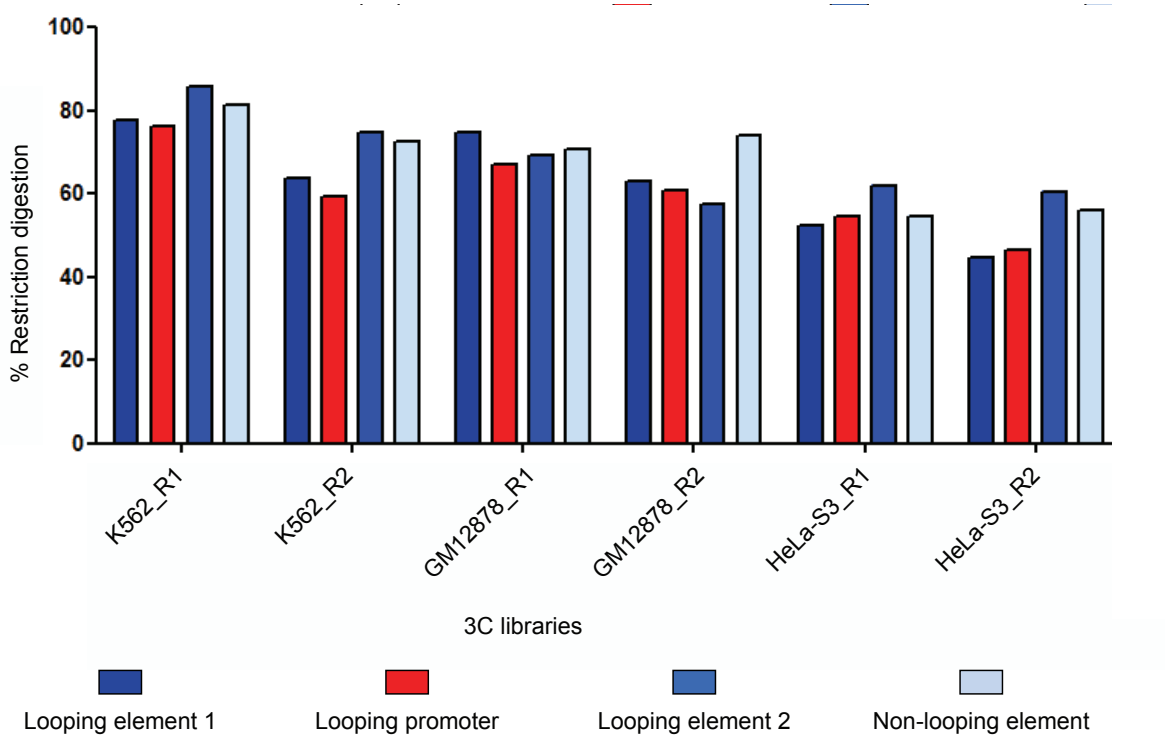


d





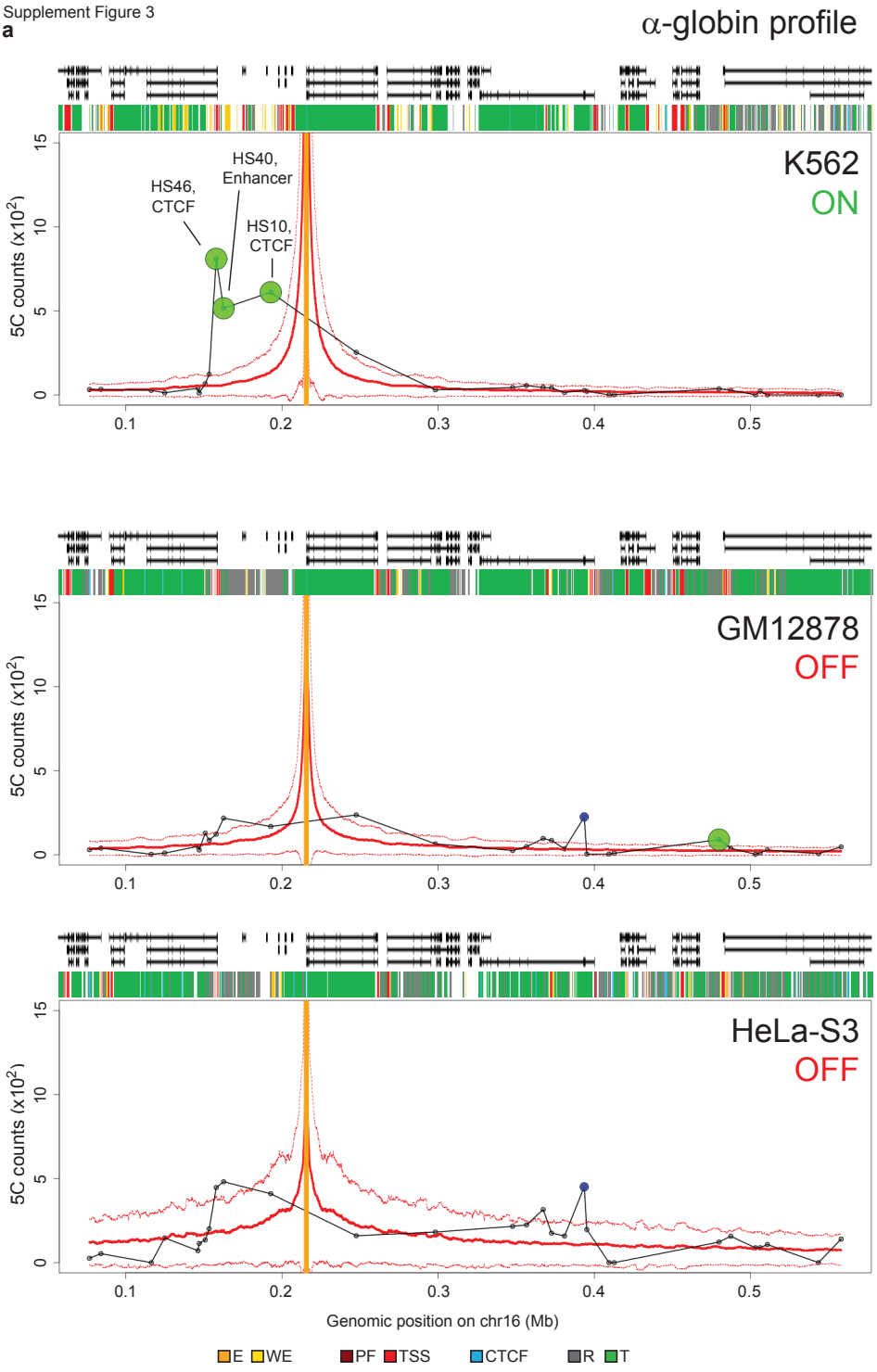
Supplementary Figure 1| Analysis of 5C reproducibility. **a,** Venn diagrams (top row) showing the numbers of the peak called looping interactions and their overlap in each of the biological replicates per cell line. Bottom row: Numbers and overlap of the significant looping interactions in each of the biological replicates for the three gene desert regions (ENr112, ENr113, ENr313) used to estimate false positive detection rates. **b,** Box plot showing the distribution of enrichment scores for looping interactions found in both biological replicates (TruePeaks), or looping interactions found exclusively in either replicate 1 (Rep1Peaks) or replicate 2 (Rep2Peaks). Data from all 3 cell lines are combined. Enrichment scores are from Supplementary Figure 1e. Loops found in all three cases are significant enriched for chromatin marks, but loops found in both biological replicates show a higher mean enrichment score. **c,** Bar graph showing the percentages of all interactions that are called a peak in one biological replicate and yield zero sequence reads in the other biological replicate of the same cell line. These interactions are caused by un-reliable 5C primers and represent a very small fraction of false positives in one biological replicate. Because these are significant in only 1 replicate, these interactions are excluded from the TruePeak set used in all other analyses. **d,** Box plot of z-scores distribution for TruePeak (peaks called in both replicates), Rep1Peak_Rep1Zscore (peak in rep1; z-score in rep1 plotted), Rep2Peak_Rep2Zscore (peak in rep2; z-score in rep2 plotted), Rep1Peak_Rep2Zscore (peak in rep1; z-score in rep2 plotted), Rep2Peak_Rep1Zscore (peak in rep2; z-score in rep1 plotted) and NonPeaks (not a peak in rep1 or rep2). Asterisk (*) denotes a significant difference ($P_{Wilcoxon} < 0.05$)) between the various z-score distributions compared to the NonPeak z-score distribution as determined by the Wilcoxon signed-rank test. This analysis shows that TruePeaks have a higher mean z-score and interactions that are called a peak in only one replicate still show a significantly higher mean z-score in the other replicate as compared to the non-peaks z-score distribution. **e,** Heatmap showing the enrichment/depletion of chromatin features in looping fragments compared to all interrogated fragments in True peaks, Replicate 1 and 2 specific peaks and union of peaks in Replicate 1 and 2 based on genome-wide datasets from ENCODE consortium (Supplemental Table 7) as described in Figure 2. Grey boxes represent non-significant (NS) enrichments/depletions.

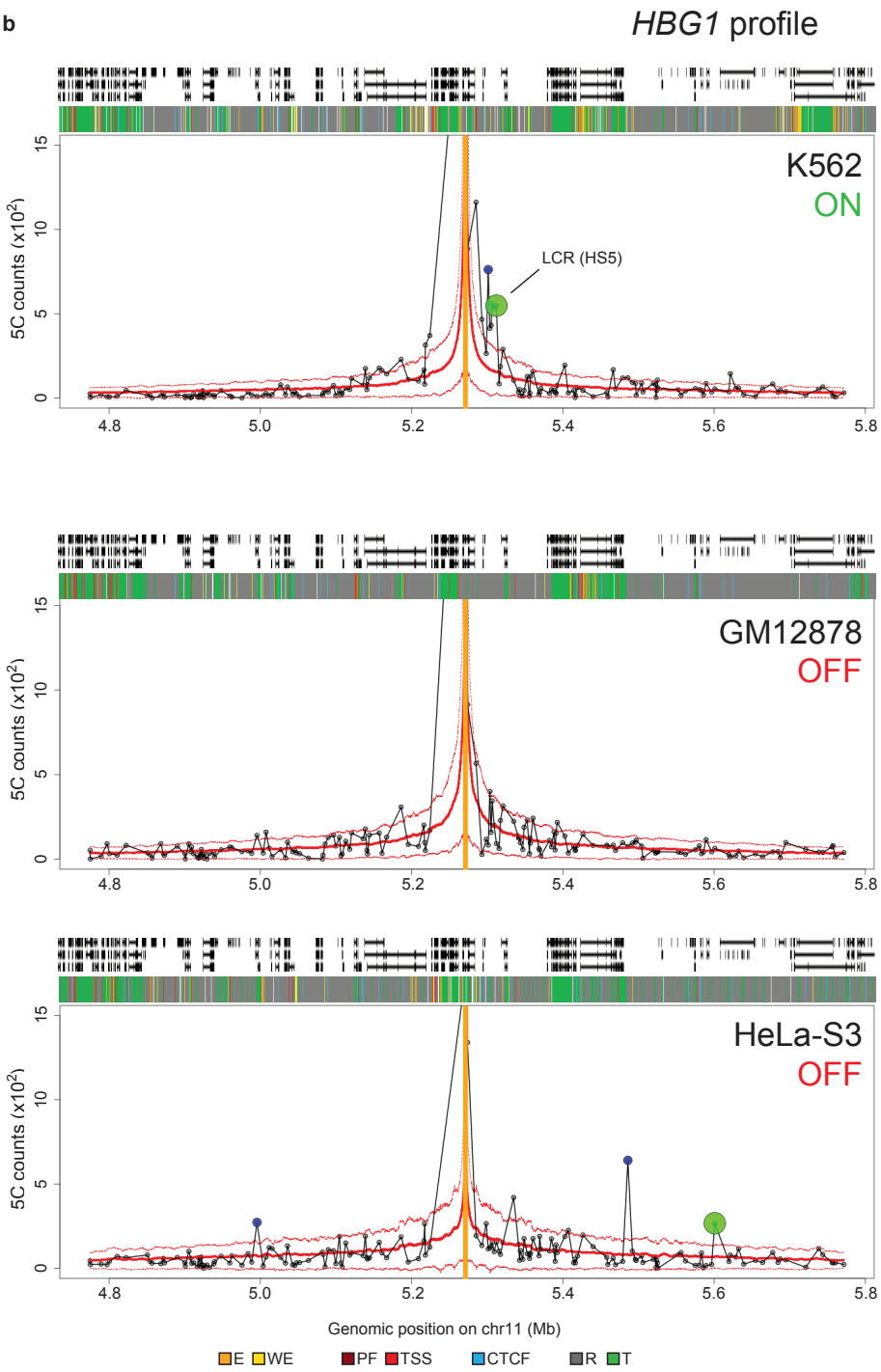


Supplementary Figure 2| Restriction digestion efficiencies of the 3C libraries that are used to make 3C carbon copies (5C libraries). Histogram showing the percentage of restriction digestion of 4 individual restriction sites based on qPCR data of two biological replicates of K562, GM12878 and HeLa-S3 3C libraries. Digestion efficiency was determined by qPCR using primers that were designed spanning specific HindIII restriction sites to estimate the fraction of DNA that was not digested. The control primers are designed within the restriction fragments. The primers used are:
Looping element 1 FOR: 5'-GAACCAAAAAGTGCAGAGTGC-3' REV: 5'-AAGCCCCCTCTAAAAACTGC-3'.
Looping promoter FOR: 5'- CCGAGTTGCAGTACCATGTG-3' REV: 5'-TGCTGGCCTAACTCCTTTGT-3';
Looping element 2 FOR: 5'- CAGTTGGATTCTAAAACCTGATACAA-3'
REV: 5'- GAAGTGTGGGTAATTCTAGGAAGC-3';
Non-looping element FOR: 5'- GGCTGGTACCTGCAACCTAA-3' REV: 5'- CCAGTTCACCTGGAATGAGG-3'.
The digestion efficiencies of restriction sites containing Looping element 1 and Looping promoter are calculated using Control primer 1 FOR: 5'- GAACACTGCTCCCCCAAATA-3' REV: 5'-TCACGTGGCATTCTTCTCTG-3' while for Looping element 2 and Non-looping element, digestion efficiencies are calculated using Control primer 2 FOR: 5'-CAGGGTGACGATCCTCAAGT-3'
REV: 5'-ACACCCTCGTCAACTTCGTC-3'. It should be noted that digestion efficiencies calculated here are under estimated by a few percent since the chromatin has already undergone the 3C protocol which includes digestion with a restriction enzyme followed by intra-molecular ligation of physically interacting cross-linked genomic fragments.

Supplement Figure 3

a





Supplementary Figure 3| Interaction profile of *HBG1* and α-globin genes in expressing (K562) and non-expressing cells (GM12878 and HeLa-S3). 5C interaction profile of reverse fragment (vertical orange bar) containing TSS of α- globin genes (*HBA1, HBA2, HBM*) or *HBG1* versus interrogated distal fragments in ENm008 (hg19; chr16:60002-559999) or ENm009 (hg19; chr11:4774421-5776011) regions respectively. The solid red line shows the expected interaction profile (LOWESS line) along the regions genomic coordinates and dashed red lines above and below indicates LOWESS ± 1 standard deviation. The 5C signals that are significantly higher than expected in two biological replicates (green circles) are considered as long range looping interactions between that TSS and the corresponding distal fragments. The blue circles denote interactions higher than expected in only one replicate (not considered as looping interactions). Gencode V7 genes and ENCODE seven-way segmentation data are displayed as tracks above the 5C interaction plots. Light/Dark Red – Promoter (TSS) and Promoter Flanking (PF);
Yellow/Orange – Weak Enhancer (WE) /Enhancer (E); Blue – Insulator (CTCF); Green – Transcribed (T);
White/Gray – Repressed (R)/Heterochromatin. **a,** In α-globin expressing K562 cells (ON), the 5C peak calling method accurately detects the known long-range interactions between the α-globin and its enhancer HS40 and the CTCF-containing HS46 and HS10 hypersensitive sites (indicated in top panel). These interactions are absent in cells (GM12878 and HeLa-S3) where α-globin is not expressed (OFF). **b,** In HBG1 expressing K562 cells (ON), the 5C peak calling method accurately detects the known long-range interactions between *HBG1* and the LCR element (HS5). These interactions are absent in cells GM12878 and HeLa-S3 cells where HBG1 is not expressed (OFF).

a Supplement Figure 4 GM12878 E class (ENm002_REV_19 : *IL13, AC004041*) GM12878 P class (ENm007_REV_135 : *LILRB4*) GM12878 CTCF (ENr323_REV_57 : *OSTM1, NR2E1*)

GM12878 Rep1

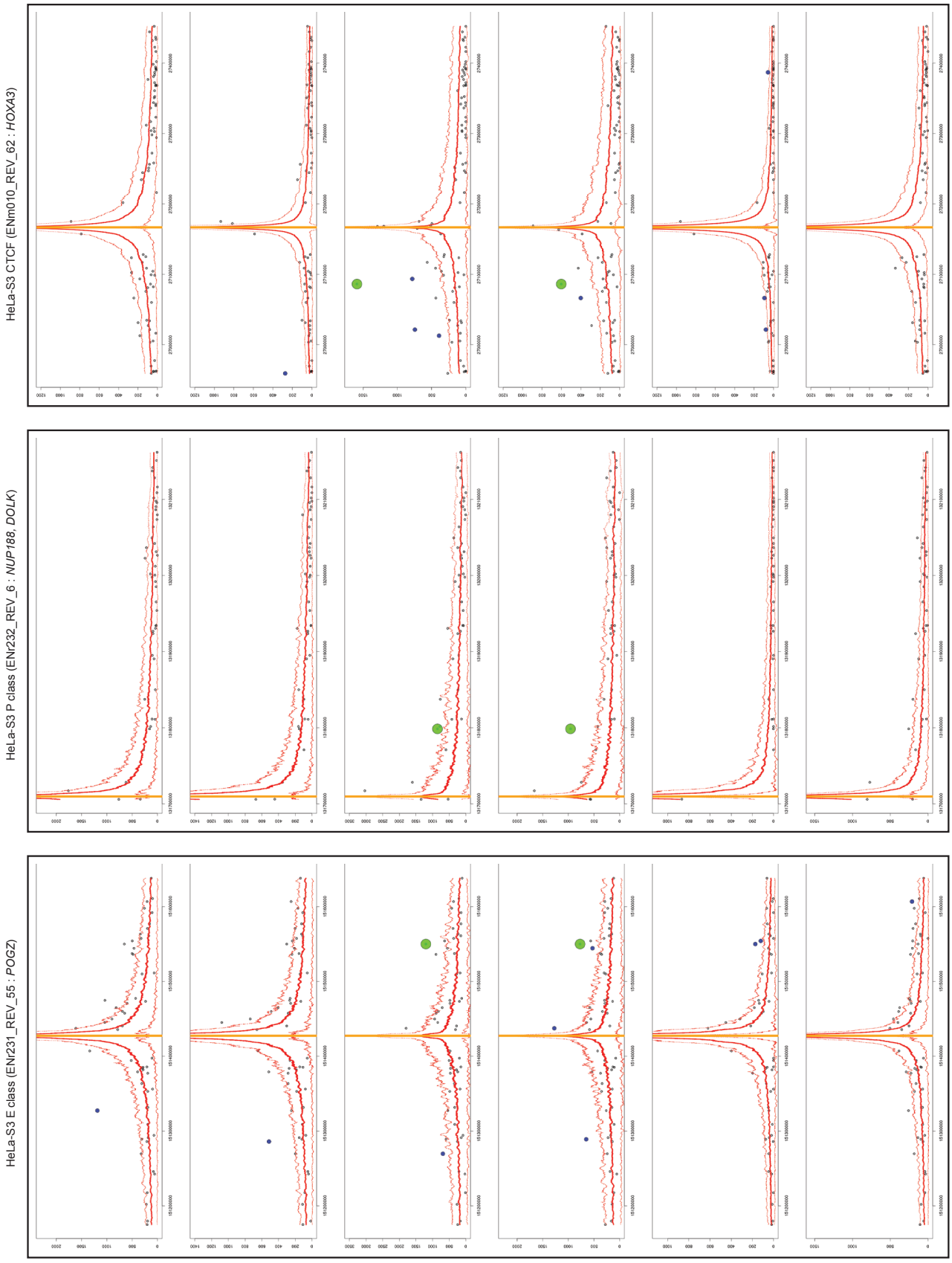
GM12878 Rep2

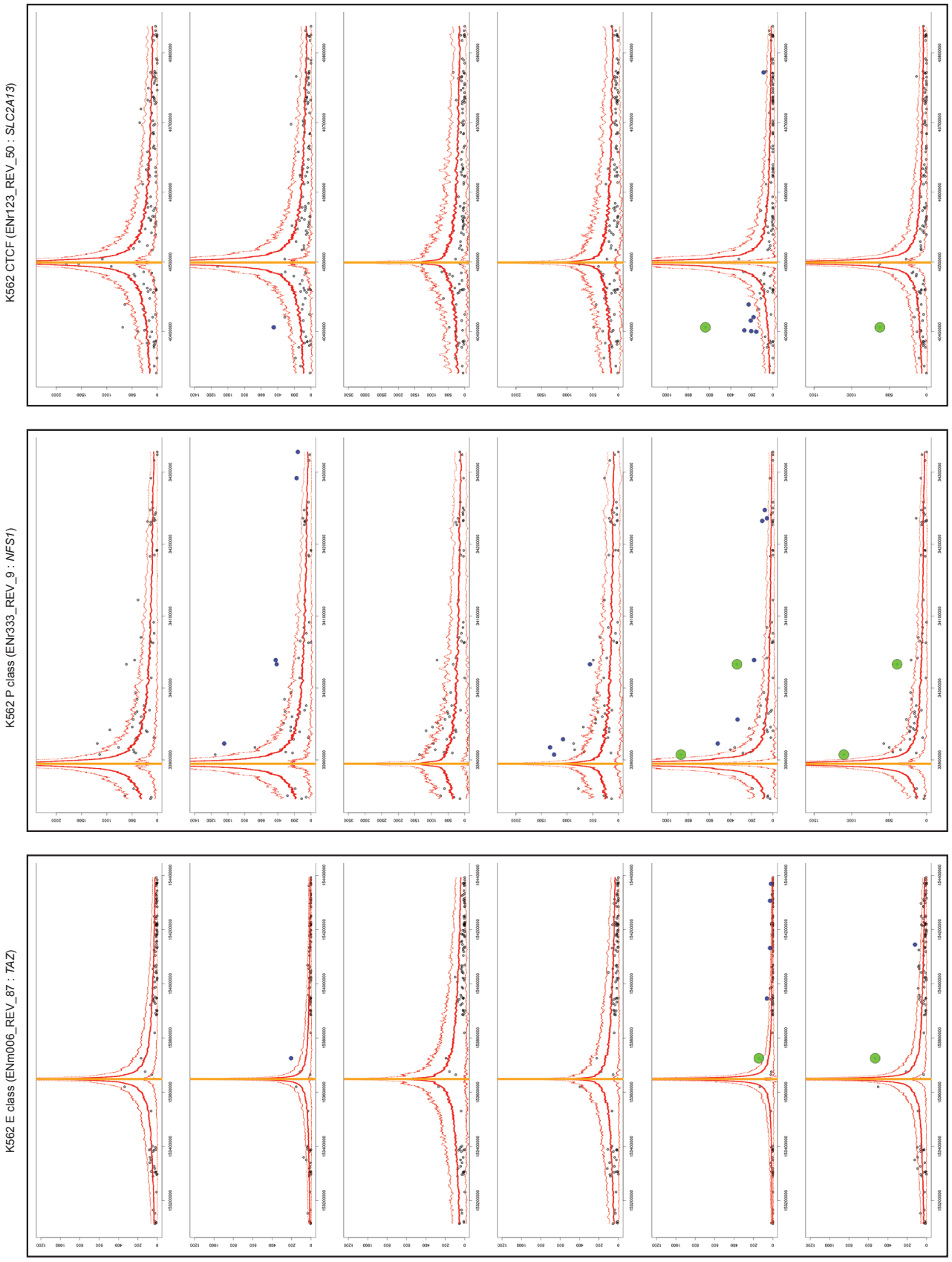
HeLa-S3 Rep1

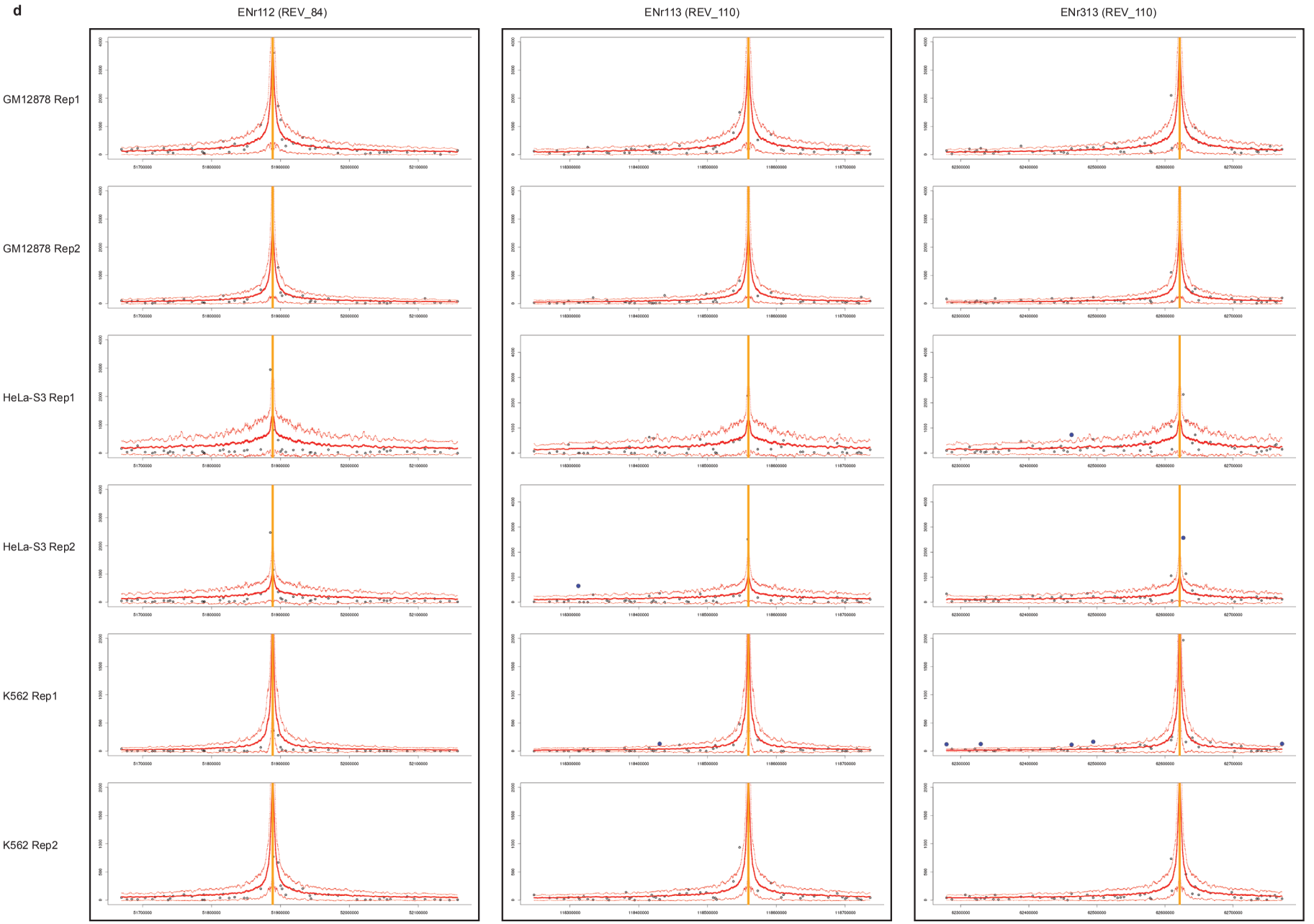
HeLa-S3 Rep2

K562 Rep1

K562 Rep2

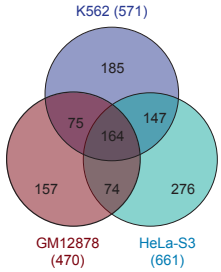

b
HeLa-S3 E class (ENr231_REV_55 : POGZ)
HeLa-S3 P class (ENr232_REV_6 : NUP188, DOLK)
HeLa-S3 CTCF (ENm010_REV_62 : HOXA3)
GM12878 Rep1
GM12878 Rep2
HeLa-S3 Rep1
HeLa-S3 Rep2
K562 Rep1
K562 Rep2


c
K562 E class (ENm006_REV_87 : TAZ)
K562 P class (ENr333_REV_9 : NFS1)
K562 CTCF (ENr123_REV_50 : SLC2A13)
GM12878 Rep1
GM12878 Rep2
HeLa-S3 Rep1
HeLa-S3 Rep2
K562 Rep1
K562 Rep2

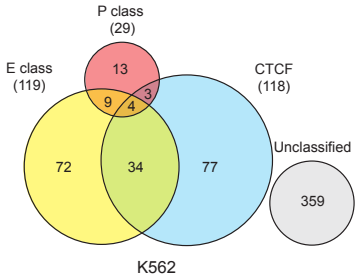


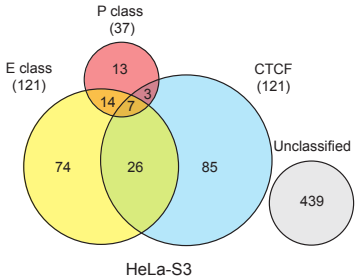
Supplementary Figure 4| Examples of 5C interaction profiles. 5C Interaction plots as described in Supplementary Figure 4. 5C interaction profiles for 2 biological replicates from each of 3 cell lines are shown separately. **a,** Examples of 5C interaction profiles for TSS-E, TSS-P, TSS-CTCF interactions specific to GM12878. **b,** Examples of 5C interaction profiles for TSS-E, TSS-P, TSS-CTCF interactions specific to HeLa-S3. **c,** Examples of 5C interaction profiles for TSS-E, TSS-P, TSS-CTCF interactions specific to K562. **d,** Examples of 5C interaction profiles observed in the 3 gene desert regions (ENr112, ENr113 and ENr313).

a Number of looping distal fragments

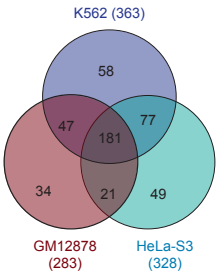


b Number of looping distal fragments per functional group

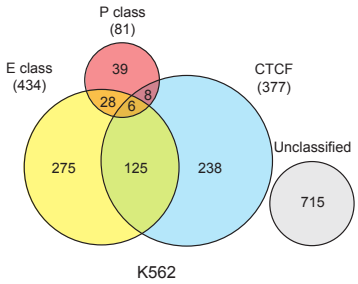


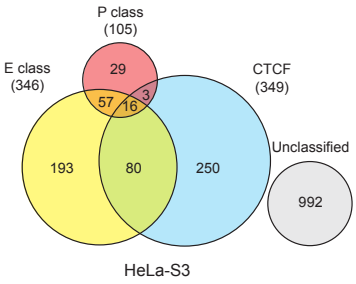


Number of looping TSS

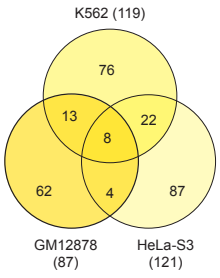


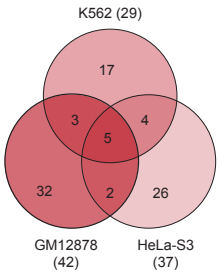
Number of looping interactions per functional group

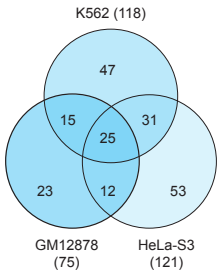


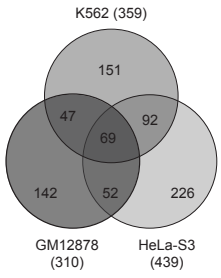


c Cell type specific distribution of looping distal fragments in different functional groups

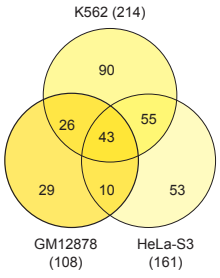


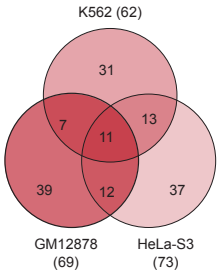


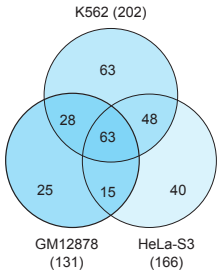


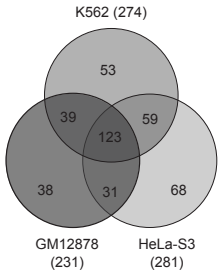


Cell type specific distribution of TSS looping to different functional groups

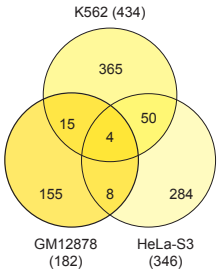


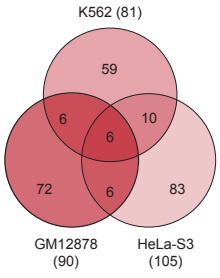


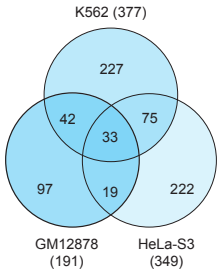


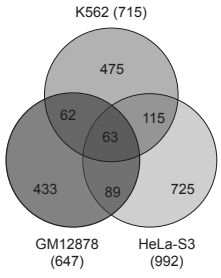


Cell type specific distribution of looping interactions in different functional groups

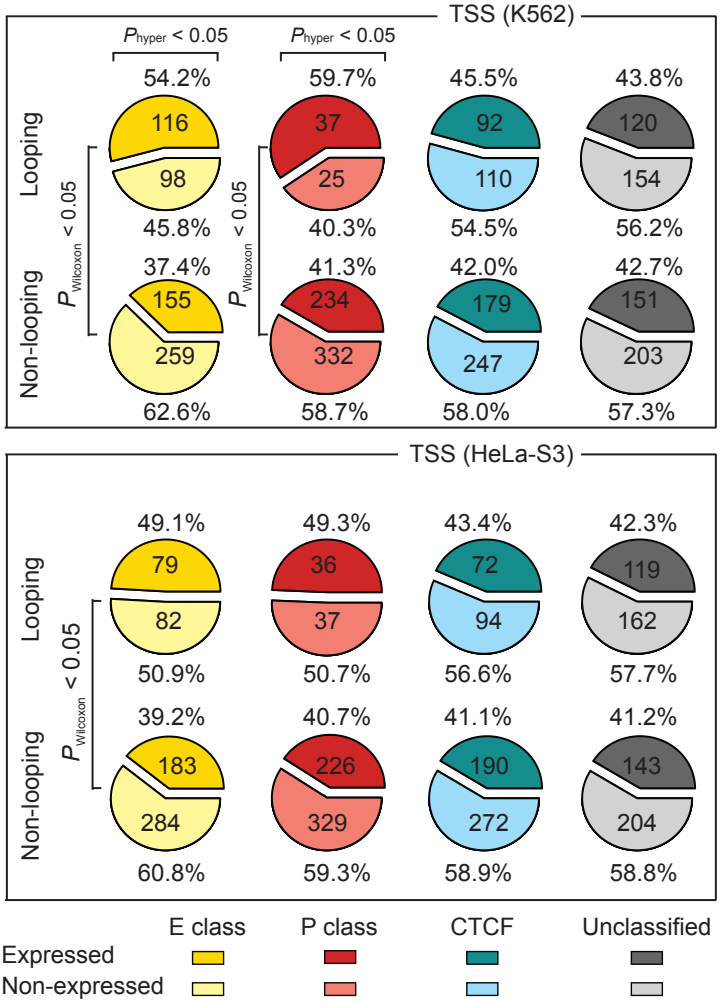








Supplementary Figure 5| Distribution of looping interactions across cell types and functional groups. **a,** Venn diagrams showing the unique and overlapping looping distal fragments (top) and looping TSSs (bottom) across 3 cell types (GM12878, K562, HeLa-S3). **b,** As described in figure 2c, looping interactions are classified into E class (yellow), P class (light magenta), CTCF (cyan) and Unclassified (grey) groups. Venn diagrams showing the distribution of looping distal fragments (above) and looping interactions (below) among the four groups in K562 and HeLa-S3. **c,** Venn diagrams showing the distributions of looping distal fragments (top), TSSs (middle) and looping interactions (bottom) across different cell types in each of the E class, P class, CTCF and Unclassified groups.



Supplementary Figure 6| Correlation between looping interactions of a particular group and gene expression in different cell types. As in figure 2d, CAGE expression data are used to assign expressions for each TSS in K562 and HeLa-S3. TSS with CAGE value >0 is considered as expressed. Different groups are represented as: E class (yellow), P class (magenta), CTCF (cyan) and Unclassified (grey).The top row in each panel of pie charts indicates percentages and numbers of expressed/non-expressed TSSs looping to a particular group (E class, P class, CTCF or Unclassified) of distal fragments. The bottom row in each panel of pie charts indicates percentages and numbers of expressed/non-expressed TSSs that are not involved in looping interactions. Significant enrichment for expressed TSSs in the looping or non-looping categories is indicated on top (hypergeometric test; $P_{hyper} < 0.05$). Significant differences in expression levels between TSS in the looping versus the non-looping category is indicated on the left (Wilcoxon signed-rank test; $P_{Wilcoxon} < 0.05$).

a

Interactions (looping/non-looping) with overlap

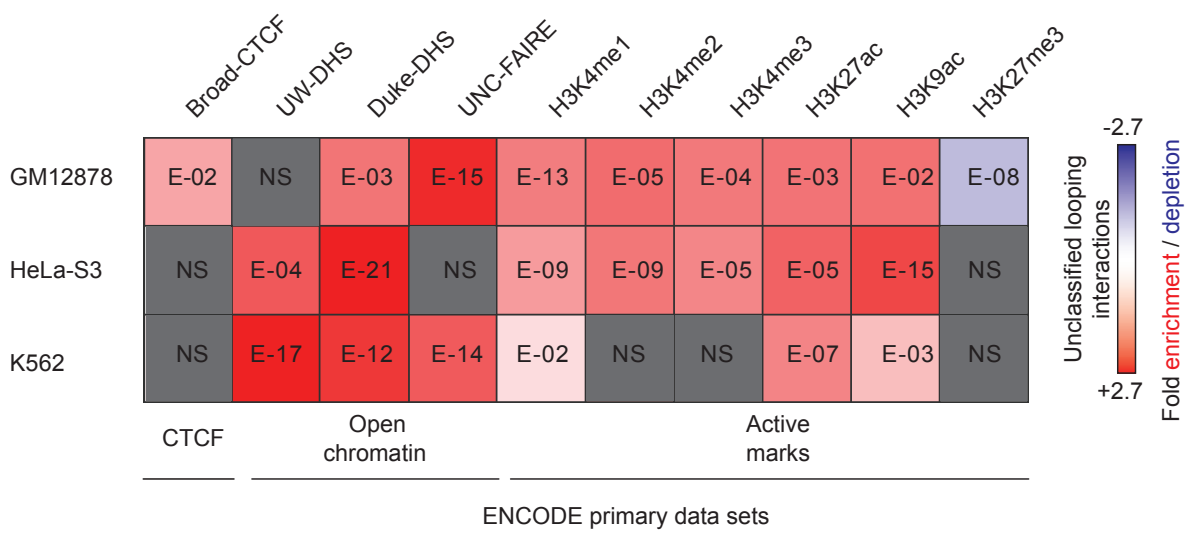
	CTCF	UW DHS	Duke DHS	UNC FAIRE	H3K4me1	H3K4me2	H3K4me3	H3K27ac	H3K9ac	H3K27me3	Others
GM12878 (647/65740)	72	13	49	100	183	63	59	50	34	213	190
	4209	1389	2369	3047	8905	2783	2783	2316	1519	26730	20399
K562 (715/63266)	56	99	94	146	218	71	56	122	151	262	203
	3602	2133	2665	4779	13318	3743	3519	4907	7952	18759	21118
HeLa-S3 (992/64757)	152	41	117	64	208	141	89	67	132	314	339
	7958	983	2065	3183	6782	3724	2604	1620	2884	17877	24642

b

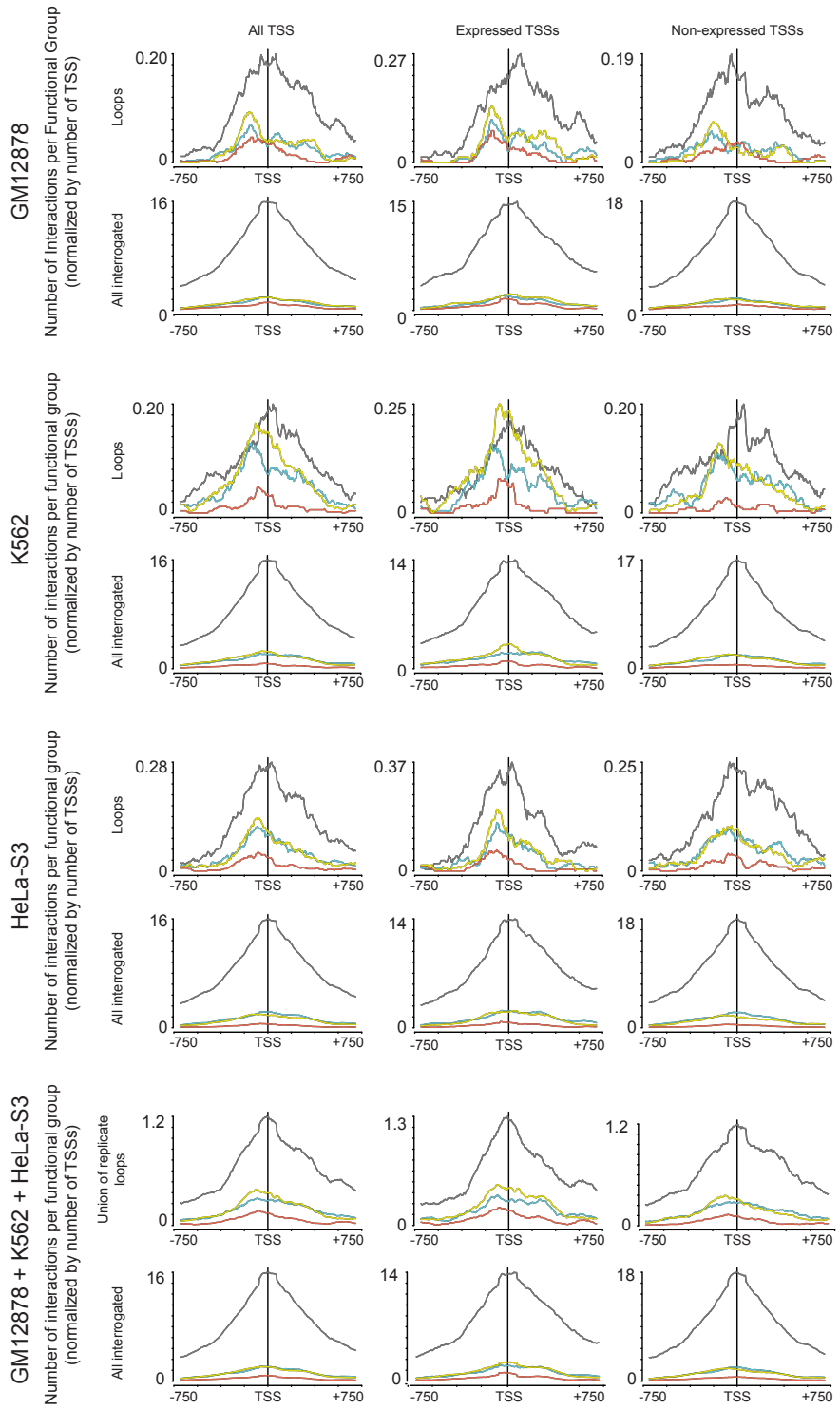
Fragments (looping/non-looping) with overlap

	CTCF	UW DHS	Duke DHS	UNC FAIRE	H3K4me1	H3K4me2	H3K4me3	H3K27ac	H3K9ac	H3K27me3	Others
GM12878 (310/4225)	29	8	19	41	68	26	27	27	15	114	103
	209	68	106	165	395	130	114	101	64	1556	1239
K562 (359/4176)	28	37	40	59	97	24	24	42	58	135	118
	150	82	127	241	475	140	136	153	244	975	1773
HeLa-S3 (439/4096)	53	19	34	33	80	56	37	23	36	125	185
	330	41	96	154	331	248	134	79	185	936	1519

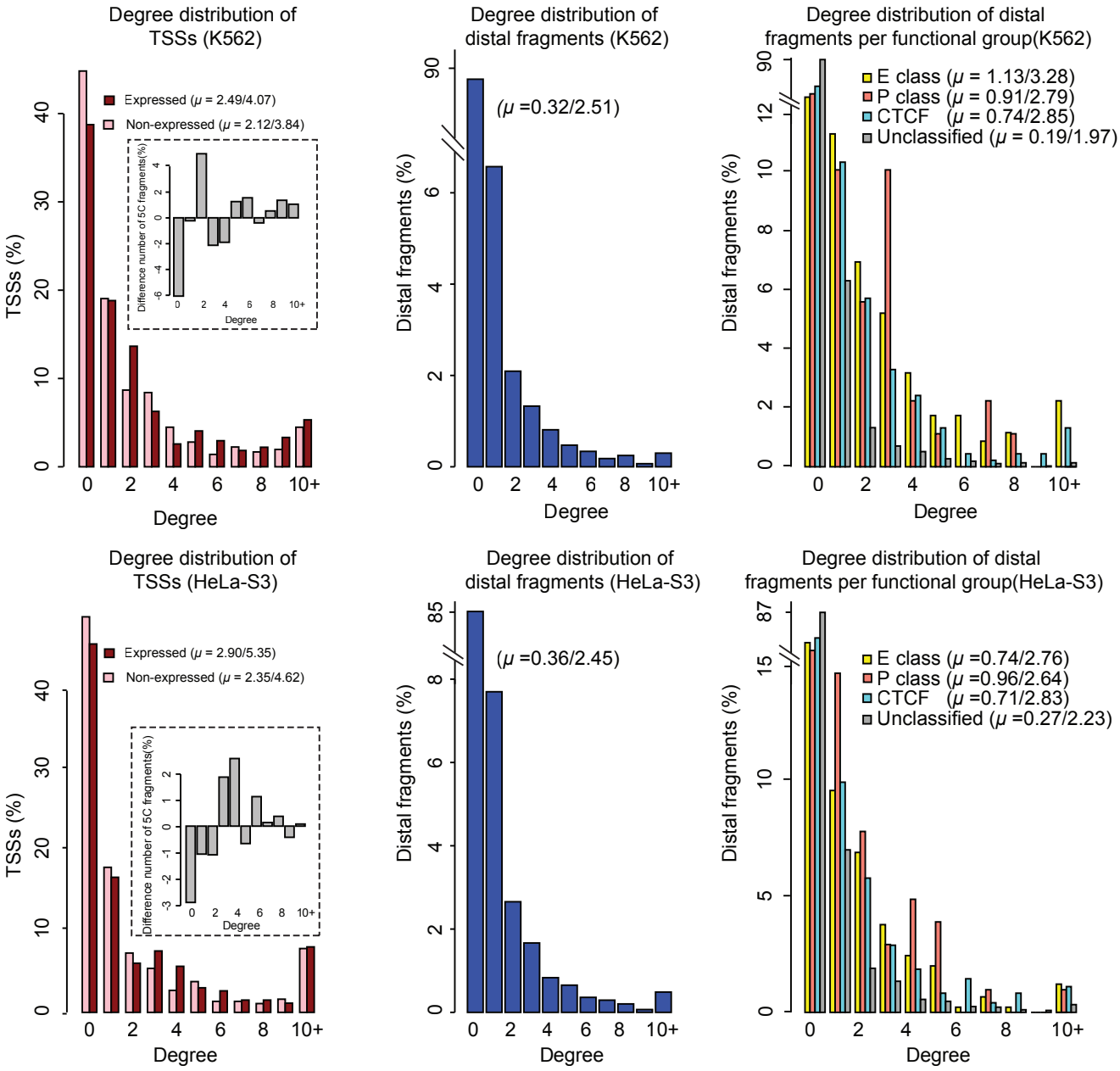
c



Supplementary Figure 7| Analysis of chromatin features of the "unclassified" category. **a,** Table of Unclassified Interactions with various chromatin marks. Table shows the number of looping / non-looping interactions belong to unclassified group that contain chromatin marks (by column). b, Table of Unclassified Fragments with chromatin marks. Table shows the number of looping / non-looping unclassified distal fragments that contain chromatin marks (by column). In a and **b,** numbers of looping interactions/fragments overlapping with additional chromatin marks that are not listed in the table are shown in "others" category. **c,** Heatmap representing the fold enrichment/depletion of unclassified looping interactions with chromatin marks as in Figure 2b.



Supplementary Figure 8| Average TSS-distal fragment looping landscape in different cell lines. Composite profiles of average number of group-specific looping interactions upstream and downstream of TSSs for each of the 3 cell lines. Each group is represented by different colors: E class – yellow, P class – magenta, CTCF – cyan and unclassified – grey. In each panel the top row shows the average looping profiles of all TSSs (left), of expressed TSSs (CAGE value of >0, middle) and of non-expressed TSSs (CAGE value = 0; right) with each of the four groups of distal elements. The bottom row of each panel shows plots with the corresponding profiles of all interrogated TSS-distal element interactions (left), of expressed TSSs (middle) and of non-expressed TSSs (right). All the interaction data for a particular group is binned with a sliding window of 150 Kb with step size of 5 Kb and the interactions values normalized by the number of TSSs. The bottom panel is as in Figure 3a, but now using the union of all significant interactions observed in each biological replicate instead of the intersection. The plots show the data for the 3 cell lines combined and again resulted in an asymmetric landscape. This tendency is weaker than when the intersection of the significant interaction is analyzed (using only those looping interactions that are significant in both biological replicates). This is probably the result of the presence of a higher percentage of false-positive interactions in the union set as compared to the intersection set (see Supplementary Materials).

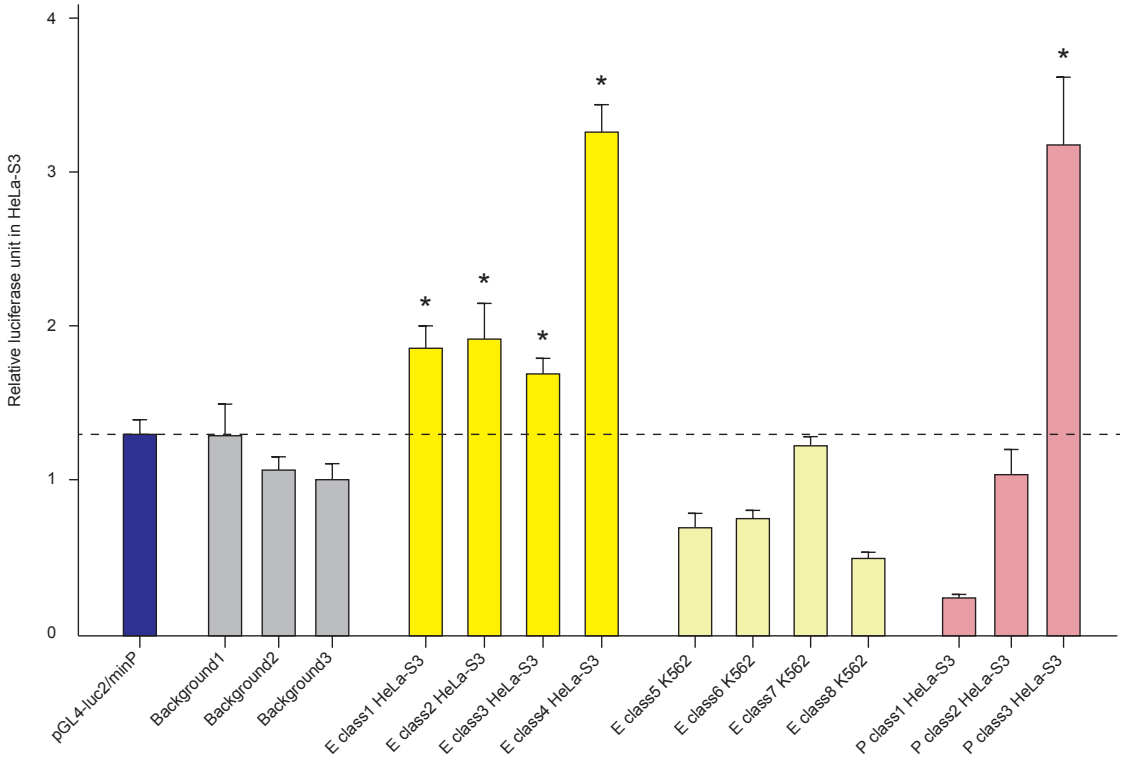


Supplementary Figure 9| Degree distribution of looping interactions of TSS and distal fragments in K562 and HeLa-S3. Histogram showing the number of TSSs (left, red) or distal fragments (middle, blue) in percentages that are involved in 0, 1, 2,..., 10 (and above) number of looping interactions (degree, x-axis) with distal fragments and TSSs respectively in K562 (top panel) and HeLa-S3 (bottom panel). All the values in degrees that are >9 are grouped and included in the category with degree 10+. The red bars represent the percentages of looping TSSs that are expressed (CAGE expression value >0) while light red bars represent the percentages of looping TSSs that are not expressed in the corresponding cell line. The difference of percentages between looping TSSs that are expressed and not expressed (red bar minus light red bar) for each degree is shown (inset). The right panel shows the degree distribution for each group of distal fragments. The average (mean, μ) degree for TSSs and distal fragments are indicated. The first value is the mean degree considering all the TSS/distal fragments (looping + non-looping) while the second value is the mean degree of looping TSS/distal fragments (degree >0).

a

	Looping fragments in E class with eRNA	Non-looping fragments in E class with eRNA	Looping fragments in E class	Non-looping fragments in E class	*P* value
GM12878	67	210	87	313	0.02
K562	94	156	119	225	0.01
HeLa-S3	98	234	121	328	0.01
All	259	600	327	866	0.00005

b



Supplementary Figure 10| Analysis of E class looping fragments. **a,** Correlation of E-class looping fragments with expression of enhancer RNA (eRNA; data from ENCODE consortium). The table shows the number of looping and non-looping fragments in the E class that express eRNA and the total number of looping and non-looping fragments in E class in each of the 3 cell types (GM12878, K562 and HeLa-S3). P values are determined by the hypergeometric test. P values that are >0.05 are considered as significant. **b,** Reporter assay of looping fragments belonging to the E class and P class as identified by 5C analysis in HeLa-S3 cells. The bar graph shows the luciferase activity (relative luciferase unit) ± standard deviation of different classes of looping fragments in HeLa-S3 cells. E class 1-4 are fragments that display significant looping interactions in HeLa-S3 cells while E class 5-8 are fragments that display significant looping interactions in K562 cells. P class 1-3 are the looping fragments in HeLa-S3 cells that are associated with P class while Background 1-3 are DNA sequences from repressed/heterochromatic regions.
All the fragments were PCR amplified from GM12878 genomic DNA and cloned in a Gateway modified pGL4 luc2/minP vector upstream of minimal promoter and the firefly luciferase reporter gene. HeLa-S3 cells are plated at a density of 2 X 10^4 cells per well in a 96 well plate one day prior transfection. Transient transfections were carried out in six replicates with Attractene (Qiagen) transfection reagent using 200 ng of indicated plasmid DNA per well along with 50 ng of Renilla luciferase construct (pGL4-hRluc/TK) as a normalizing control. The luminescence was measured after 24 hours using Dual-Glo Luciferase assay system (Promega) in a Victor3 1420 multilabel counter (Perkin Elmer). Luminescence from pGL4-luc2/minP vector transfected HeLa-S3 cells is used as the control for calculating the significant luciferase activity in different constructs. The asterisk (*) above the bar shows significant (P-value <0.05) upregulation in luciferase activity compared to the pGL4-luc2/minP as calculated by one tailed unpaired t-test. The coordinates of the looping elements tested are given below. Primers are designed upstream and downstream of the restriction fragments with 5' attB4 site for forward and 5' attB2.1 site for reverse primers for cloning in attR4-attR2 sites in pGL4-luc2/minP Gateway modified vector. Primer information will be made available upon request.

E-class 1-HeLa-S3 5C_301_ENm004_FOR_260|hg19|chr22:33096044-33097202
E-class 2-HeLa-S3 5C_298_ENm001_FOR_492|hg19|chr7:117305160-117306431
E-class 3-HeLa-S3 5C_1720_ENr212_FOR_102|hg19|chr5:142250469-142252650
E-class 4-HeLa-S3 5C_1725_ENr231_FOR_94|hg19|chr1:151591512-151592347

E-class 5-K562 5C_298_ENm001_FOR_275|hg19|chr7:116657331-116659879
E-class 6-K562 5C_306_ENm009_FOR_138|hg19|chr11:5217159-5218976
E-class 7-K562 5C_1724_ENr223_FOR_89|hg19|chr6:74126855-74128780
E-class 8-K562 5C_305_ENm008_FOR_29|hg19|chr16:191222-194336

P-class 1-HeLa-S3 5C_1733_ENr324_FOR_47|hg19|chrX:122995786-122997722
P-class 2-HeLa-S3 5C_1733_ENr324_FOR_26|hg19|chrX:122863474-122866031
P-class 3-HeLa-S3 5C_302_ENm005_FOR_404|hg19|chr21:35285017-35286762

Background1 ENr334_1100bp_Background1|hg19|chr6:41585743-41586842
Background2 ENr333_999bp_Background2|hg19|chr20:34337226-34338224
Background3 ENr232_1752bp_Background3|hg19|chr9:131954827-131956578