SUPPLEMENTAL TEXT

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Validation of chromatin interactions in MCF7 and K562 cells by DNA-FISH and 3C-qPCR, related to **Figure 1**

(A) Quantitative DNA FISH data for positive (interaction) and negative (no interaction) hits randomly selected from MCF7 inter-chromosomal ChIA-PET data.

(B) An example of a chromatin interaction between chr11 and chr17 in MCF7 cells. This exemplifies that multi-gene complexes from different chromosomes could further converge to a common active nuclear compartment.

(C-G) Detailed 3C-qPCR validations for several long-range (up to ~17Mb) intra-chromosomal interactions and an inter-chromosomal interaction (D). Most of the intra-chromosomal interactions are tested in both MCF7 and K562 cell-lines. P-values are calculated using binomial test. Panel D, F and G represent local interactions at distant genomic loci converging to each other *via* long range *cis* or *trans* interactions. BACs used in the study are given in the Extended Experimental Procedure.

Figure S2: Detailed genomic features of distinct chromatin models, related to **Figure 2**

(A) Detailed examples from 6 different chromosomes illustrating the association of distinct chromatin architectures with genomic descriptors. Density of each descriptor (except %GC, which is measured in isochores) and the interacting anchors in each of our chromatin architectures is measured in each 1Mb domain across chromosomes and running mean over 5 values are plotted. Certain gene rich domains enriched in multi-gene (MG) models and depleted in single-gene (SG) models are highlighted in red, while relatively gene-poor domains enriched in SG and depleted in MG are marked in blue.

(B, C) Genomic features of BP, SG, MG models in MCF7 (B) and K562 (C) saturated libraries. The plots validate our observations on the combined pilot data presented in **Figure 2** of the main text.

Figure S3. Promoter properties and functional output (transcription) of different categories of chromatin models, related to **Figure 3**

(A, B) Tissue specificity measured by descriptor-1 (A) and descriptor-2 (B). Equations for tissue specificity descriptors are given in the Extended Experimental Procedures.

(C, D) Normalized CpG content (C) of promoters (+/-1500bp to TSS) and strand bias (CG-skew) (D) at promoters of genes in different models. Difference in the representation of High CpG (HCG) promoters (associated with housekeeping genes) and Low CpG (LCG) promoters (associated with tissue specific genes) is found to be significant between SG and MG complexes, while BP model has relatively negligible representation of LCG promoters suggesting their association primarily with housekeeping function. Similarly, CG-skew in (D) shows greater bias (associated with high and housekeeping expression) at promoter sites for BP and MG models, while lower bias (associated with lower and tissue specific expression) for SG model. These predictive measures support our observation in **Figure 3D** in the main text.

(E) Co-expression of interacting genes in K562 cells**.**

(F-K) Density plots for Pearson's Correlation Coefficient (PCC) values of gene pairs in MG complexes (red), rewired pairs and random gene pairs selected from a control dataset of the same distribution of genomic spans and gene density as MG pairs with an upper limit of 1Mb. The gene expression datasets analyzed are: **(F)** E2 induced time course microarray at 6 time points (Fullwood et al., 2009); **(G)** microarray dataset of 4,787 human samples covering a wide range of diversity in gene expression, like distinct tissues, gender, developmental and differentiation stages etc (Sahoo et al., 2008). Different controls are selected over genomic spans of BP, SG and MG genes; **(H-I)** ENCODE RNA-Seq datasets for 5 different cell-lines (K562, MCF7, Hela, HCT116 and GM12878) for MCF7 and K562 interactions; **(J)** PCC distribution for MG gene pairs belonging to the same and different functions (GO process); **(K)** PCC distribution for housekeeping (HK) and tissue specific (TS) gene pairs in MG units.

(L) Representation of gene families in random (#3383) and MG complexes (#1487) datasets with respect to expected probability 'p' of finding 2 proximal genes (within 1Mb proximity) from the same gene family. The method to compile random control and to calculate the probability of finding two proximal genes from the same gene family is given in the Extended Experimental Procedure. The plot suggests greater enrichment of gene families in multi-gene complexes.

Figure S4. Chromatin interactions and gene expression following siERα transfection in MCF7 cells, related to **Figure 4**

(A) Overlap of RNAPII loops with ERα loops at the *GREB1* locus. P1, P2, and P3 are RNAPII interacting sites; E1, E2, and E3 are ERα interacting sites.

(B) ERα knockdown by siERα as tested by Western blot and RT-qPCR.

(C) *GREB1* expression following 0, 3 and 6 hours of ethanol (ET) and estrogen (E2) treatment after siControl and siERα transfections.

(D) 3C-qPCR data for chromatin interactions at *GREB1* locus following ET and E2 treatment after siControl and siERα transfections.

(E) Estrogen induction and siERα knockdown led to correlated changes in the expression of interacting genes (*CCDC88C* and *GPR68*). ChIA-PET tracks clearly show that interaction between promoters of *GPR68* and *CCDC88C* is associated with RNAPII, while ERα binds only at promoter and gene-body of *CCDC88C*. Color codes of the bars are shown in Figure S4D.

Figure S5. Enrichment profiles of transcription factors and histone modification marks centered at the interaction anchor regions of RNAPII-bound chromatin interaction structures in K562 cells, and reporter gene assays in MCF7 cells, related to **Figure 5**.

(A) Aggregation plots of TFs enrichments centered at the RNAPII interaction sites, proximal to TSS (TSS) or distal to TSS (non-TSS). RNAPIII, as a negative control, shows negligible enrichment at the RNAPII interacting sites. *y*-axis: sliding median for ChIP-Seq enrichment in the region. *x*-axis: distance (bp) from RNAPII sites.

(B) TFs enriched at non-TSS (potential enhancer sites).

(C) Enrichment profile of chromatin remodeling and chromatin architectural factors.

(D) Enrichment profile of open chromatin and histone marks around RNAPII interacting sites. Clearly, the open chromatin mark DHS and active histone marks are substantially enriched at the RNAPII interacting sites, while the repressive histone marks show little enrichment.

(E) Map of pGL4.10 vector and cloning sites of promoters and enhancers for luciferase assays.

(F) Standard promoter and enhancer reporter assay for elements around the *CALM1* locus. The enhancer upstream of *CALM1* significantly enhanced the luciferase activity of the *CALM1* promoter, which was involved in the typical enhancer-promoter interaction.

(G) The *DDHD1* promoter, which is located on the same chromosome as *C14orf102-CALM1* locus and had no interaction with *CALM1*, showed no significant enhancement to the *CALM1* promoter activity in luciferase assays.

(H) Deletion promoter reporter assay around *CALM1* locus. The TATA box of the *CALM1* promoter was deleted (from – 133bp to +100bp, black arrow). The reporter construct containing this deletion promoter did not show any promoter activity in luciferase assays by itself or in any combinations with *C14orf102*.

(I) Non-MG promoters do not possess enhancer functions. The *ELFN1* and *DLD* promoters, which are located on the same chromosome as for the *INTS1-MAFK* locus and had no interaction with *MAFK*, did not enhance the promoter activity of *MAFK* in luciferase assays.

(J) Box plots of RNA-Seq data in log2 RPKM for the genes with low and high log ratio of H3K4me3/H3K4me1 in the pairs of interaction sites. The genes with higher log ratio in a pairing relationship have higher RNA-Seq counts on average than the interacting partner with lower log ratio.

(K) Box plots of normalized luciferase activities when the promoters with low log ratio of H3K4me3/me1 at the enhancer position of the luciferase constructs, or when the promoters with high log ratio of H3K4me3/me1 at the enhancer position of the luciferase constructs. The promoters with low log ratio of H3K4me3/me1 at the enhancer position of the luciferase constructs have higher enhancing effects in general.

(L) Swap of *INTS1* and *MAFK* promoters in positions in reporter gene construct for luciferase assays. The promoter sequence from *INTS1* (with lower log ratio of H3K4me3/me1 signals) enhanced the luciferase activity of *MAFK* (with higher log ratio of H3K4me3/me1 signals). On the reverse, the *MAFK* promoter showed no enhancer function.

Figure S6. Cell-specific interaction analysis, related to **Figure 6**

(A) All the Gene Ontology (GO) terms over-represented in the gene sets engaging cell-specific expression and interactions.

(B) Overrepresented GO terms in the gene set engaged in chromatin interactions common to both MCF7 and K562 cells. The abundance of housekeeping terms is apparent.

(C) K562-specific interactions around β-globin gene locus on chromosome 11. ChIA-PET loop tracks clearly show that there are chromatin interactions from β-globin genes to the locus control region (LCR) in K562 cells, but not in MCF7 cells. Correspondingly, the RNAPII and RNA-Seq show higher expression of β-globin genes in K562 cells, but not in MCF7 cells.

(D) MCF7-specific interactions around *GATA3* gene locus on chromosome 10. Contrast to the β-globin gene locus, the ChIA-PET loop tracks clearly show that there are chromatin interactions from *GATA3* gene locus to multiple enhancer sites in MCF7, but not in K562 cells. Correspondingly, the RNAPII binding and RNA-Seq showed high activity in MCF7 and low activity in K562 cells. Especially, one super-long-distance enhancer is about 1.2Mb away from *GATA3* promoter.

Figure S7. Examples of cell-type specific long range enhancer-promoter interactions, related to **Figure 6 and 7**.

(A-E) specific to MCF7 cells, and **(F-K)** specific to K562 cells

Most of regulatory sites in these long distance interaction examples are bypassing the nearest promoters and linking to other gene promoters.

(L) Several distant enhancers converging to *MYC* gene promoter. Cell-specific alternative usage of certain enhancers can be seen from the interaction loop views from MCF7 and K562 cells.

Figure S8. Assessment of technical noise, library reproducibility and saturation analysis, related to **Experimental Procedures**

(A-B) Heatmaps of PET sequence counts vs. genomic span for interactions identified from homo-dimer and hetero-dimer PETs from the combined pilot dataset.

(C) Densities of genomic spans of interactions from homo-dimer PETs and hetero-dimer PETs of the combined pilot data.

(D) Densities of genomic spans of interactions from re-wired PETs.

(E-G) Scatter plot of sequence reads per 10Kb from RNAPII ChIA-PET replicates: (E) MCF7 pilot datasets, (F) MCF7 saturated and (G) K562 saturated datasets.

(H) Scatter plot of sequence reads per 10Kb from K562 saturated and MCF7 saturated RNAPII ChIA-PET datasets.

(I-L) RNAPII binding site reproducibility of K562 saturated replicates. **(I)** Histogram of genomic distances between RNAPII peaks from replicates. **(J)** Venn diagram of RNAPII peak overlap between replicates. **(K)** Scatter plot of RNAPII peak intensities of replicates. **(L)** Box plot of peak intensities of RNAPII peaks common and unique in replicate 1.

(M-O) RNAPII interaction reproducibility of K562 saturated replicates. **(M)** Scatter plot of interaction PET counts between replicates. **(N)** Venn diagram of interaction overlaps between replicates. **(O)** Violin plot of interaction PET counts from common and unique interactions from replicate 1.

(P) Saturation assessment of chromatin interactions from K562 saturated RNAPII ChIA-PET replicates. The overlap ratio between replicates against the proportion of PETs sampled from K562 saturated replicate 1 (more details in the Extended Experimental Procedure; under saturation analysis).

(Q-R) RNAPII interaction region reproducibility of K562 saturated replicates. **(Q)** Scatter plot of interaction region PET counts between replicates. **(R)** Venn diagram of interaction region overlaps from replicates. *I*, *II*, and *III* for the top 25%, 50% and 75% interaction regions from K562 saturated replicate 1.

(S-T) Scatter plots of RNA-Seq reads per 10Kb in replicates from MCF7 **(S)** and K562 **(T)**.

SUPPLEMENTAL TABLES

Table S1. Statistical summary of different RNAPII ChIA-PET libraries

Pilot libraries	**Unique PETs**	**RNAPII Peaks**	**Interaction clusters**
HCT116_pilot	8,749,703	17,394	19,264*
K562_pilot	23,188,484	13,112	17,686
HeLa_pilot	19,079,666	15,333	200,952
NB4_pilot	14,023,893	12,707	54,232
MCF7_pilot_rep1	38,356,322	10,370	12,626
MCF7_pilot_rep2	22,967,674	11,234	8,111
Pilot_combined	118,523,881	14,604	892,991
Saturated libraries			
K562_saturated_rep1	14,177,547	26,922	64,565**
K562_saturated_rep2	14,365,592	27,046	65,489
MCF7_saturated_rep1	15,283,270	27,198	23,440
MCF7_saturated_rep2	15,622,720	27,683	24,126

Note:

* Pilot libraries: PET clusters with ≥ 2 PET counts are reported

** Saturated libraries: PET clusters with ≥ 3 PET counts are reported

Table S2. Percentage of intra-chromosomal PET clusters from MCF7 pilot and saturated libraries

	Total clusters*	intra-chrom clusters	inter-chrom clusters	intra-chrom %
Pilot library	4,379	3,498	881	79.88
Saturated library	51,134	50,387	747	98.54

Note: * PET clusters with ≥ 3 PET counts are reported here

Table S3. RNAPII peaks and high-confidence RNAPII interaction PET clusters from ChIA-PET libraries (in separate sheets of one Excel file)

Table S4. Gene family representation in multi-gene complexes (in one Excel file)

Table S5. Cell line specific enhancer-promoter interactions from MCF7 and K562 cells (in separate sheets of one Excel file)

Table S6. Summary of different interaction categories, and the number of genes in different chromatin models

Interaction categories	MCF7 pilot library	Combined pilot library	K562 saturated library	MCF7 saturated library
promoter-terminator interactions	524	938	16593	8293
enhancer-enhancer interactions	5760	4106	53610	10865
enhancer-promoter interactions	4355	6530	34228	18244
promoter-promoter interactions	2642	8282	22986	12985
Transcription models				
Basal promoter model (gene)*	5603 (7647)	3929 (4938)	294 (334)	2515 (2865)
Single-gene model (gene)	1490 (1760)	966 (1119)	2448 (2324)	2015 (2213)
Multi-gene model (gene)	1487 (6260)	1328 (11723)	1918 (19813)	2459 (14310)

Note: * Some interactions are linked to the bi-directional promoters or the alternative promoters of the same genes, which makes the gene numbers different from the number of chromatin models.

EXTENDED EXPERIMENTAL PROCEDURE

Cell culture and ChIP preparation

Five cell lines were used in this study for RNAPII ChIA-PET analysis. These are MCF7 (ATCC# HTB-22), K562 (ATCC# CCL-243), HCT116 (ATCC# CCL-247), HeLa (ATCC# CCL-2.2), and NB4 (Roussel and Lanotte, 2001) (provided by Dr. Sherman Weissman, Yale University). The cells were grown under standard culture conditions and harvested at log phase. The cells were treated with 1% formaldehyde at room temperature for 10 minutes on a plate rotator, followed by neutralization using 0.2M glycine. The cross-linked chromatin was obtained from the fixed cells by cell lysis and nuclear lysis. The chromatin was then subjected to fragmentation using Branson digital sonifier S450D to an average size of 300bp. The sonicated chromatin was pre-cleared overnight using Protein G magnetic beads to remove non-specific DNA. Simultaneously, RNAPII Monoclonal antibody 8WG16 (Covance, MMS-126R) was incubated with Protein G magnetic beads overnight to allow antibody coating to the beads. The pre-cleared chromatin was then immunoprecipitated overnight with the antibody-coated beads to capture chromatin of interest. The beads that were coated with DNA of interest were washed several times to remove non-specific binding. A portion of ChIP DNA was eluted off the beads for concentration quantification using Picogreen fluorimetry and enrichment checking using quantitative PCR.

ChIA-PET library construction

Immuno-precipitated chromatin fragments were subjected to ChIA-PET library construction following the protocol as previously described (Fullwood et al., 2010; Fullwood et al., 2009) with some modifications. Briefly, the chromatin DNA fragments bound to antibody beads were divided into two aliquots for DNA linker ligation by Linker A and Linker B, respectively. The two linkers have the same nucleotide sequences, except four nucleotides in the middle are different (Linker A with TAAG; Linker B with ATGT) as the nucleotide barcode. The linkers were in excess so as to saturate the ends of DNA fragments. After the linker ligation and removal of excess linkers by washing the beads, the two aliquots were combined together for proximity ligation in diluted conditions, in which DNA fragments in individual chromatin complexes would have the same specific linkers (either A or B). During proximity ligation, DNA fragments within the same chromatin complex with the same linker would ligate together, so as to generate ligation products with homo-dimer linker composition (AA or BB). However, if the ligation reactions took place between DNA fragments of different chromatin complexes, such non-specific ligation products would have a 50% chance of resulting with hetero-dimer linker composition (AB or BA). Hence, the hetero-dimer linker composition is an indicator of non-specific ligation, which can be used to assess the non-specific ligation rate of each ChIA-PET library and such non-specific ligation data can be removed from further analysis. After proximity ligation, Paired-End-Tag (PET) constructs were extracted from the ligation products, and the PET templates were subjected to Illumina GAIIx sequencing.

ChIA-PET library data processing

The ChIA-PET sequence reads were processed by ChIA-PET Tool (Li et al., 2010), a software package designed for ChIA-PET data analysis, with some modifications. Briefly, non-redundant PET sequence reads were first analyzed for linker barcode composition and identified as sequences with hetero-dimer AB linker (barcode TAAG / ATGT) derived from non-specific ligation products, or sequences with homo-dimer AA or BB linker (barcodes TAAG / TAAG or ATGT / ATGT) derived from specific ligation products. The linker composition information was used later for noise analysis. Then, the linker sequences were trimmed, and the PET tag sequences were mapped to the human reference genome (hg19). To further remove possible redundant PET sequences after genome mapping, the PETs with genomic locations from both head and tail tags within 2bp were merged to further reduce the library sequence redundancy arising from clonal PCR amplification. This step takes into account any Single Nucleotide Polymorphisms (SNPs) between the reference and the test genome and sequencing errors that may have occurred and resulted in a 1- or 2-bp difference in the tag sequences.

PET classification The mapping of PET sequences to the reference genome would reveal the nature of proximity ligation products, i.e. whether they are self-ligation products between the two ends of the same DNA fragment or are inter-ligation products between two DNA fragments that were captured in the same chromatin complex by protein interactions. Because the chromatin fragment sizes were sonicated within a narrower range from 100bp up to a few Kb, the mapping orientation and distance between the two tags of a PET sequence would indicate if the PET was derived from self-ligation or inter-ligation. For inter-ligation PETs, there are two categories: intra-chromosomal if the two tags of a PET were mapped in the same chromosome, and inter-chromosomal if the two tags were mapped in different chromosomes. For convenience, we call PETs derived from self-ligation “Self-Ligation PETs”, inter-ligation PETs mapped in the same chromosome “Intra-chromosomal PETs” and inter-ligation PETs mapped in different chromosomes “Inter-chromosomal PETs”.

Peak calling of RNAPII binding The coverage of all self-ligation PET sequences across the genome reflects the enrichment by RNAPII ChIP on specific locations, similar to ChIP-Seq mapping for protein binding sites. Using a similar method as that of the ChIP-Seq peak calling program MACS (Zhang et al., 2008), we performed peak calling on the ChIA-PET data. The local summits of the sequence coverage were called as potential peaks. The significances of the potential peaks were estimated with p-values from a Poisson distribution. The background parameters in the Poisson distribution were estimated from the maximum of the global tag density, tag density in a 10Kb window around the peak, and the tag density in a 20Kb window around the peak. The p-value was corrected as false discovery rate (FDR) with the Benjamini-Hochberg (B-H) method (Benjamini and Hochberg, 1995) for multiple hypothesis testing. The criteria for our final peaks were that 1) the sequence coverage is at least 5 and 2) the FDR is smaller than 0.05.

Interaction PET clusters Inter-ligation PETs potentially reflect long range chromatin interactions. However, inevitably, there are technical noises from various sources. To further distinguish true interaction signals from non-specific interaction noise, we reasoned that for true

interactions, multiple interaction PETs would be generated from the same interacting regions. To identify such chromatin interactions, mapping locations of the inter-ligation PETs were extended 1.5kb downstream, and the PETs that overlapped at both ends formed interaction PET clusters. Overlapping PET clusters are used to distinguish detectable interaction signals over background noise represented by singleton PETs, which could also include weak interaction events that are not distinct from background noise. The PET count of a PET cluster is the frequency of the interaction between the two locations involved. The statistical significance of such interactions was evaluated with p-values from a hyper-geometric distribution. The hyper-geometric model takes into consideration the tag counts from both anchor regions and the sequencing depth for p-value calculation, thus normalizing the effects of random ligations between two highly-enriched regions that would give rise to potentially noisy inter-ligation PETs. The p-values were corrected as false discovery rate (FDR) with the B-H method (Benjamini and Hochberg, 1995) for multiple hypothesis testing and the FDR cutoff is 0.05.

Transcription models from RNAPII peaks and chromatin interactions Each of these interactions identified by PET clusters is also termed as a duplex interaction because each of them involves a pair of interacting anchors. The duplex interactions are further collapsed based on the connectivity of overlapping anchors with other duplex interactions to form complex interactions. With the high-confidence RNAPII peaks and interaction PET clusters, we defined three transcription models based on how the genes were involved in the interaction regions: basal promoter (BP) models (gene promoters that overlapped with standalone RNAPII peaks, but did not overlap with interaction anchors), single gene (SG) interaction models and multi-gene (MG) interaction models.

ChIA-PET library statistics summary

For our RNAPII ChIA-PET analysis, we generated the data in two stages. In the pilot stage, we generated 6 individual datasets, including two biological replicates from MCF7 cell-line (MCF7_pilot_rep1 and MCF7_pilot_rep2). All these 6 individual datasets were combined into one human combined pilot RNAPII ChIA-PET library. In the saturated stage, we sequenced much deeper for MCF7 and K562 RNAPII ChIA-PET libraries, and used the PETs with homo-dimer linkers for further analysis.

The numbers of unique PETs, peaks and interactions from the individual libraries and the combined libraries are summarized in **Table S1**. RNAPII binding peaks and high-confidence interaction PET clusters from the combined MCF7 pilot library, human combined pilot library, and two saturated libraries are in **Table S3 (**in separate sheets of the Excel file). Interaction categories and transcription models are summarized in **Table S6**.

Noise analysis

Noise level in ChIA-PET data was measured with two different methods: one method based on barcode linkers and the other method based on randomly re-wired PETs. In the first method, two different barcode linkers have been designed in the ChIA-PET protocol. Using the linker nucleotide barcode embedded in the PET constructs (Li et al., 2010), we estimated that up to

94% PETs with homo-dimer linker barcodes in a ChIA-PET library are specific intra-chromatin fragment ligation products.

Second, we used the hetero-dimer PET (non-specific ligation) dataset to estimate the distribution of noisy data in terms of PET counts and the genomic spans covered by PET mapping. The heatmaps in Figure S8A-B show the frequencies of the interaction PET clusters with a specific genomic span and PET counts from the combined pilot dataset. The PET clusters from the homo-dimer PETs (specific ligation) appeared to be mostly in the area with high PET counts and small genomic span (less than 1Mb), while the PET clusters from hetero-dimer PETs (non-specific ligation) are mostly with low PET counts and long genomic span (more than 1Mb), as would be expected because the hetero-dimer PETs are formed by random ligations with very low chance.

When such data are shown as densities of the spans (Figure S8C), the non-specific PETs (blue curve) showed a single peak in the range of 1Mb-100Mb, whereas the specific PETs (red curve) showed two peaks. One peak in the span density of the homo-dimer PETs overlapped with the peak from the non-specific PETs, suggesting that this part of the homo-dimer PETs is probably non-specific. The other peak in the span density of the homo-dimer PETs is from 10Kb to 1Mb. Therefore, we chose a 1Mb genomic span as a cutoff to define high-confidence interacting PET data. We used this cutoff to reduce the number of false positives, although this cutoff may increase the number of false negatives by excluding some interactions that are beyond the 1Mb limit.

To further test the PET interaction cluster data, we randomly re-wired the tags from specific homo-dimer ligation PETs to break up the pairing relationship of each PET. The pairing information in the real inter-ligation PETs was removed first and the tags were randomly paired again to generate random re-wired PETs. As shown, the span distribution from the re-wired PETs (Figure S8D) is similar to the span distribution from the hetero-dimer linker PETs, but much different from the homo-dimer linker PETs. This means that our homo-dimer linker PETs are not random, but specific.

For the span distribution of all the intra-chromosomal PETs, we showed that PET mapping span <8Kb are likely to be derived from self-ligation of the same DNA fragments. Hence, we defined the following parameters for high-confidence intra-chromosomal inter-ligation PET clusters:

a) PET count ≥2 for each PET cluster from the pilot libraries, and PET count ≥3 for each PET cluster from the saturated libraries;

b) FDR < 0.05; and

c) Genomic span from 8Kb to 1Mb.

Such interaction clusters were mainly used for single-gene and multiple-gene complex study. Super long range intra-chromosomal interactions (beyond 1Mb span) were analyzed separately.

Reproducibility analysis of ChIA-PET libraries

To evaluate the robustness of the ChIA-PET method, we analyzed biological replicates of RNAPII ChIA-PET libraries from MCF7 and K562 cells at several different resolutions. There are two replicates from MCF7 in the pilot stage (MCF7_pilot_rep1 & MCF7_pilot_rep2), and two replicates from MCF7 (MCF7_saturated _rep1 & MCF7_saturated_rep2) and two replicates from K562 (K562_saturated _rep1 & K562_saturated_rep2) in the saturated stage. We assessed the reproducibility of the replicates at four different levels: reproducibility of library sequence reads, reproducibility of the RNAPII peaks, reproducibility of the individual interactions and reproducibility of RNAPII interaction regions. Some of the analysis results are presented in Figure S8.

Reproducibility of library reads To evaluate the reproducibility of library sequence reads, we divided the genome into bins with bin size 10Kb and counted the numbers of tags in the individual bins for each library. For two replicates from the same cell-line, we used the tag counts from the same bins as pairs of data values to generate scatter plots. The result (Figure S8E-H) shows that the correlation of library reads from different replicates of the same cell line is high (correlation coefficients > 0.97). In contrast, the library reads from MCF7 and K562 RNAPII ChIA-PET libraries showed lower correlation (correlation coefficient 0.52). This data suggests that at the library sequence level, ChIA-PET sequencing is highly reproducible, and can reflect cell specificity.

Reproducibility of RNAPII peaks Next, we evaluated the reproducibility of RNAPII peaks identified by ChIA-PET libraries. First, we examined the distribution of the distances from the RNAPII peaks in one replicate to the nearest peaks in another replicate in order to decide the distance cutoff to be used for the overlap of RNAPII peaks from different biological replicates. The maximum distance allowed was 10Kb. The results in Figure S8I show the distribution of distances between RNAPII peaks from K562 saturated biological replicates, with the peaks from the first biological replicate at the center. We saw that the peaks from the second biological replicate are symmetrically distributed around the peaks from the first biological replicate and noticed that most peaks from one biological replicate have a peak from another biological replicate within 200bp. Accordingly, 200bp was used as the distance cutoff to overlap peaks from different biological replicates.

Using 200bp as the cutoff for RNAPII peak overlap, we generated a Venn diagram (Figure S8J) and a scatter plot (Figure S8K) of the overlapped RNAPII peaks from biological replicates. The Venn diagram shows that the majority of the RNAPII peaks from two biological replicates are overlapped, and the scatter plot shows that the peak intensities of the overlapped peaks from two biological replicates are highly correlated.

In addition, we examined the peak intensities of the overlapped and non-overlapped RNAPII peaks from different biological replicates. The box plots in Figure S8L show the peak intensities from overlapped (or common) peaks and the non-overlapped (or unique) peaks for one of the biological replicates. The common peaks have higher peak intensities and the non-overlapped peaks have lower peak intensities. This suggests that most peaks with significant binding

signals are reproducible and reliable. The similar results of peak reproducibility were observed from MCF7 pilot replicates and saturated replicates. In summary, the RNAPII peak overlap analysis shows that the RNAPII peaks from different ChIA-PET biological replicates are highly reproducible.

Reproducibility of chromatin interaction PET clusters Third, we evaluated the reproducibility of interaction PET clusters from the replicates. Two interaction PET clusters are considered as overlapped if both anchors from one cluster have at least one base pair overlap with the anchors from another interaction PET cluster. Analysis result shows the scatter plot (Figure S8M) of the overlapped clusters from different biological replicates and the x-axis and y-axis show the PET counts from the overlapped clusters. If an interaction cluster from one replicate does not overlap with any clusters from another replicate, a dummy PET count 1 is assigned as the PET count from the counterpart replicate. The scatter plot shows that the interaction clusters from different replicates have a high overlap. The Venn diagram (Figure S8N) shows that the overlap ratios of the interactions from different replicates are up to 53%. The violin plot (Figure S8O) shows that the common interactions have higher PET counts and are more reliable, and the replicate-unique interactions have lower PET counts and are more dynamic. Similar results of interaction reproducibility were observed from MCF7 pilot replicates and saturated replicates.

Reproducibility of chromatin interaction regions Lastly, we analyzed the reproducibility of interaction regions from the biological replicates. Interaction regions are genomic regions covered by chromatin interactions, which include regions covered by standalone duplex interactions or complex interaction regions that are connected by a chain of multiple individual interactions. High-confidence intra-chromosomal interaction regions were used for this analysis. If two regions have at least one base overlap, the two regions are deemed to be overlapped. The overlap ratios of the different replicates are 46.5% (MCF7 pilot), 75% (MCF7 saturated), and 66% (K562 saturated, Figure S8Q). If we sort the interaction regions by PET counts, we found that most of the interaction regions with high PET count are overlapped in the two replicates, as shown the Venn diagram (Figure S8R), where *I*, *II*, and *III* are used to indicate the top 25%, 50% and 75% of the interaction regions from one replicate. This data suggests that most strong interaction regions identified in this study are reliable. The interactions with low PET counts are weak interactions and may vary between different replicates.

Saturation estimates of RNAPII ChIA-PET libraries

We assessed the saturation level of RNAPII interaction clusters identified by ChIA-PET based on the current sequencing depth and the overlap of the interaction clusters from the replicates. We assume that the total RNAPII interaction clusters in each cell line were sampled randomly and independently by the replicates. Given that most of the RNAPII interactions from the pairs of replicates are common to each other, the proportion of non-observed interaction clusters from the replicates will be the product of the proportions of non-observed interaction clusters from each replicate. Based on the assumption, we have the following formula:

$$\frac{N - \bigcup_{i=1,2} N_i}{N} = \frac{N - N_1}{N} \times \frac{N - N_2}{N}$$

Where N is the total number of RNAPII interaction clusters in the RNAPII interactome of the specific cell line, N_i is the number of RNAPII clusters found from replicate i, and $\bigcup_{i=1,2} N_i$ is the union of the interaction clusters from two replicates. This estimation is similar to the method called "capture-recapture" for population size estimation (Baillargeon and Rivest, 2007). The estimated numbers of interaction clusters at the same significance level (FDR < 0.05) and the similar sequence depth of the replicates are 4196 (MCF7 pilot), 48751 (MCF7_saturated), and 122244 (K562_saturated). Relative to these estimated numbers of the interaction clusters, the non-observed interactions are about 44% (MCF7 pilot), 26% (MCF7_saturated) and 22% (K562_saturated). Such estimates suggest that the number of clusters from the union of two different biological replicates is nearly saturated for MCF7_saturated and K562_saturated libraries, when compared to MCF7 (unsaturated) library, at a given significance level.

Another way to study the saturation is to generate the interaction clusters from the randomly-sampled PETs from one replicate and overlap them with the interaction clusters from another replicate. We used two saturated K562 replicates for testing. Different proportions (10%, 20% …, and 90%) of the PETs were randomly sampled from K562 saturated replicate 1, and interaction clusters were generated with the same pipeline. The overlap of the interaction clusters from different proportions of PETs with K562 saturated replicate 2 is near saturated to the overlap of the two replicates as in Figure S8P. This means that deeper sequencing from one replicate can't improve the overlap ratio between different replicates substantially. To increase the total coverage of the chromatin interactions, it is better to generate more biological replicates.

RNAPII immuno-fluorescence staining combined with DNA FISH

MCF7 cells were grown to 70% confluency in hybridization chambers and fixed with 4% paraformaldehyde (PFA) (Sigma-Aldrich) for 15 min at room temperature (RT) followed by permeabilization with 0.04% Triton X (Promega) for 30 min at RT. Prior to staining, cells were blocked with 10% normal donkey serum (Millipore) for 1 hr at RT and incubated with mouse RNA polymerase II 8WG16 monoclonal antibody (Covance, 1:1000) overnight at 4^{o}C. Cells were subsequently incubated with Cy3-conjugated donkey anti-mouse IgG polyclonal antibody (Millipore, 1:1000) for 1 hr at RT, after which slides were mounted with ProLong Gold antifade reagent containing DAPI (Invitrogen).

For combined RNAPII staining-DNA FISH, RNAPII-stained cells were post-fixed in 4% PFA for 10 min at RT after secondary antibody incubation and permeabilized in 0.5% Triton X for 10 min at RT. Cells are subsequently dehydrated through a 70-80-100% ethanol series, rehydrated with 2X SSC and denatured in 2X SSC/50% formamide at 80^{o}C for 40 min. Prior to FISH probe hybridization, cells were incubated in 2X SSC for 5 min at 4^{o}C and permeabilized in 2X SSC/0.5% Triton X for 5 min at 4^{o}C. Probe preparation and hybridization for FISH were

performed as described for DNA-FISH. Carl Zeiss Meta LSM confocal microscope was used to analyze the association of RNAPII foci with promoter-promoter interaction loci. 63x optical lens and inter-slice distance (Z-axis) of 0.4 µm was used. Data was analyzed using LSM image browser and % overlap was determined manually by counting 476 interphase nuclei.

DNA Fluorescence in-situ hybridization (DNA-FISH)

MCF7 cells were harvested by trypsinization and treated with 0.75M KCl for 15 minutes at 37°C. Subsequently cells were fixed with Methanol/Acetic acid (3/1) and dropped onto slides for FISH. Following overnight culture in LB media, BAC DNA was extracted with a Nucleobond PC500 Plasmid Purification Kit (Macherey-Nagel) and labeled by nick translation in the presence of biotin-16-dUTP or digoxigenin-11-dUTP using a Nick Translation Labeling Kit (ENZO Life Sciences). In the presence of 1µg/µl of Human Cot1 and Salmon sperm DNA (Invitrogen), labeled BAC clones were resuspended to 5 ng/µl in hybridization buffer (2xSSC, 10% dextran sulfate, 1X PBS, 50% formamide). Prior to hybridization, MCF-7 nuclei on slides were digested with 0.005% pepsin (Sigma)/0.01M HCl at 37°C for 3min followed by fixation with 1% formaldehyde (Merck - Calbiochem) and dehydrated through a 70-80-100% ethanol series. Labeled probes were denatured at 75°C for 5min and hybridized to pretreated slides at 37°C overnight. Post-hybridization washes were performed twice at 45°C in 2xSSC/50% formamide for 7min each followed by 2 washes in 2xSSC at 45°C for 7min each. Slides were revealed with avidin-conjugated fluorescein isothiocyanate (FITC) (Vector Laboratories, CA) for biotinylated probes and anti-digoxigenin- Rhodamine for digoxigenin-labeled probes (Roche). After washing, slides were mounted with vectashield (Vector Laboratories, CA) and observed under an epifluorescence microscope (ZEISS, Image.Z2) and under 63X lens magnification. Between 300-800 interphase nuclei were analyzed for each probe mix. Images were analyzed using metafer4. Center-to-center distance of 1µm was taken as cut-off for co-localization analysis. The BACs used in the experiments are listed below:

BACs used in the in-situ experiments

DNA FISH (inter-chr)

Test mix	**Control mix**
727F15 + 143M10	727F15 + 563H6
626F12 + 556I13	626F12 + 563H6
727F15 + 419E4	727F15 + 563H6
286L5 + 107L14	286L5 + 563H6
286L5 + 419E4	286L5 + 563H6
727F15 + 563H6	727F15 + 242D8

DNA FISH (intra-chrom)

Test mix	**Control mix**
766A9 + 463D24	80F13 + 979M22
80F13 + 795I20	80F13 + 482K16
191N14 + 915D14	191N14 + 92F20

DNA FISH-IF

Test MG locus	**Control locus**
973N23	699B7
399J13	699B7
34B20	699B7
143M10	699B7

Quantitative analysis of chromosome conformation capture assays (3C-qPCR)

The 3C-qPCR protocol used for this experiment was developed and modified from the previous publications (Fullwood et al., 2009; Hagege et al., 2007; Pan et al., 2008). Briefly, $1X10^7$ cells were fixed with 1% formaldehyde for 10 min at room temperature and lysed, and then nuclei were digested with EcoRI or HindIII (New England Biolabs) before ligation. All primers were designed within 150bp from EcoRI or HindIII digestion site in unidirectional side. The specificity and amplification efficiency of each primer were tested by performing quantitative PCR on serial dilution of the BACs which were mixed in equal molar before digestion and ligation, after which the standard curve was obtained. The linear range of the 3C template was determined by a serial dilution of 3C sample with looping primers and non-looping primers, and the optimal concentration of 3C sample for qPCR experiments was determined. Digestion efficiency, ligation efficiency, and sample purity were all checked as per established protocols (Hagege et al., 2007). To obtain the “cross-linking frequency”, Ct values of the experimental primer with the 3C template were first normalized with Ct values of the *GAPDH* control primer with the 3C template to ensure data between different cell types such as K562 and MCF7 were comparable. Then, we calculated each primer value of (Ct-b)/a according to a (slope) and b (intercept) based on the standard curve of the BACs sample. Finally, we transformed counted the values as 10^ (Ct-b)/a for each primer, and then divided the primer values (10^ (Ct-b)/a) by either the loading control (“Inner Ctrl”, meaning this control was designed against a genomic region shared with the BAC clone, where there were no restriction enzyme digestion sites) or an internal interaction control (inside the *GAPDH* region of the genome). The 3C quantitative results are presented as

the mean ± s.e.m from two to four independent preparations of 3C sample with duplicate qPCR data.

Histone modification and transcription factor ChIP-Seq data

To characterize the peaks and interaction anchors from RNAPII ChIA-PET libraries, we used histone modification and transcription factor ChIP-Seq data to measure their profile around the peaks and interaction anchors. Histones can undergo a variety of post-translational modifications that can be identified by chromatin immunoprecipitating DNA with an antibody specific to a particular modification of interest. Several histone marks have been shown to mark sites associated with transcription such as promoters of active and inactive genes, exons of transcribed genes, etc. For K562, we used publicly available histone modification ChIP-Seq libraries from the Broad Institute and the ENCODE Project targeting the following histone modifications: H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me1, H3K20me1, H3K27ac, H3K27me3, and H3K36me3 (Broad Institute and UCSC Genome Browser), and transcription factor ChIP-Seq libraries from Raha *et. al.* (Raha et al., 2010). Histone modification ChIP-Seq libraries in MCF7 were from Joseph *et. al.* (Joseph et al., 2010).

To determine whether the unique locations in RNAPII ChIA-PET libraries arose from promoters and enhancers, we used the log ratio of H3K4me3 and H3K4me1 tag counts from the unique locations. H3K4me3 is known to be a general mark for promoter regions, and H3K4me1 is known to be a mark for promoters and enhancers (Barski et al., 2007). For each location (RNAPII peaks or interaction anchors), the number of tags within +/- 2Kb from the center of the locations were counted. The log ratio of the tag counts from H3K4me3 and H3K4me1 was normalized with the sequencing depths of both libraries. The locations with positive log ratios were considered to be potential promoters and the locations with negative log ratios were considered to be potential enhancers. In the pairs of interactions sites from promoter-promoter interactions, the interaction sites with lower H3K4me3/me1 log ratio in the individual pairs are treated as enhancer-like sites, and the interaction sites with higher H3K4me3/me1 log ratio in the individual pairs are treated as promoter-like sites. Generally, enhancer-like RNAPII interaction sites showed lower transcriptional activity, and the promoter-like RNAPII sites showed higher transcription activity (Figure S5J).

From multi-gene models in MCF7 pilot library, we generated the heatmap with H3K4me3 and H3K4me1 profiles from the peaks and interaction anchor centers. For each RNAPII peak or interaction anchor (if no RNAPII peaks), 10 bins with bin size 500bp at each side of the centers were generated to count the histone modification ChIP-Seq reads. The profiles were normalized to the number of reads per bin per million reads. The signals are log-transformed. The rows in the heatmap correspond to the interaction anchors, and the columns correspond to the bins. The rows were clustered with kmeans method (k=7) from R package. From the heatmap, we can clearly see that H3K4me3 is enriched at more than half of the centers of the interaction anchors, while H3K4me1 is enriched in smaller number of interaction anchors. For some interaction anchors, both H3K4me3 and H3K4me1 are enriched at heterogeneity regions.

Analysis of histone modification and transcription factor binding profiles in chromatin interacting sites

To further characterize the transcription models from RNAPII ChIA-PET libraries, we used histone modification and transcription factor ChIP-Seq data to mark the locations from different transcription models. Signal tracks for all K562 ChIP-Seq experiments involving histone marks and transcription factors were downloaded from the ENCODE Project hosted by the UCSC Genome Browser (Kuhn et al., 2009). The signal levels for all tracks are given in terms of fold-change with regard to an equivalent distribution of reads and were generated with respect to human genome build hg19. All signal tracks were also corrected for local mappability and sequence depth.

Using these signal tracks and our region lists for TSS-associated and non-TSS-associated ChIA-PET peaks within each anchor, we used the ACT tool (Jee et al., 2011) to perform the signal aggregation over each list. The ACT tool operates on three distinct regions for each feature. Regions immediately upstream and downstream of each feature are subdivided into bins and the average signal level of a histone mark or transcription factor is calculated for each bin. For the feature itself, region widths are scaled to uniform lengths, subdivided into bins, and the median signal level for each bin calculated.

For our signal aggregations, peaks from ChIA-PET interactions were extended to a length of 500bp from the midpoint. We chose upstream and downstream regions of 1,000 bp, which were subdivided into 20 equal-sized bins of 50 bp. The peak regions were divided into 10 equal-sized bins of 50 bp. Values for each bin were plotted and a spline-smoothed line used to approximate signal levels for each histone mark or transcription factor in the neighborhood of a ChIA-PET anchor. Aggregations were performed across the 6 categories of locations.

For all transcription factors analyzed, ChIP-Seq peak locations in K562 were downloaded from the ENCODE Project Page within the UCSC Genome Browser. All peak lists were downloaded for human genome build hg19 and were unaltered, as the submitting labs are most familiar with the characteristics of each data set and presumably provided the highest quality peak calls possible. Each list of transcription factor regions was compared against ChIA-PET anchors in promoter, inter-genic, and intra-genic regions for our chromosome activity models. We required an at least 1-bp overlap between a ChIP-Seq peak and a ChIA-PET anchor region to associate a transcription factor binding site with the anchor.

RNA-Seq data

RNA-Seq library preparation and data generation Total RNA was extracted from MCF7 cells and then mRNA was isolated following the protocol previously described in Ruan *et. al.* (Ruan et al., 2007). Approximately one microgram of mRNA was used for fragmentation with RNase III, followed by gel-selection to obtain desired fragment range at 100-150nt. The randomly fragmented mRNA was then hybridized and ligated to a mixture of linkers and adaptors obtained from SOLiD for reverse transcription (RT) to generate cDNA. The resulting cDNA template was amplified by PCR and analyzed by RNA-Seq using ABI SOLiD platform

with single direction reads of 50 bp in length. Analysis was performed following the protocol and reagent kit provided from Life Technologies Inc., which generates strand-specific RNA reads.

RNA-Seq data processing and mapping After individual RNA reads were generated, initial filtering was undertaken to remove any noise sequences such as rRNA, tRNA, mitochondria RNAs and repeat sequences etc. The strand-specific reads were mapped on human reference genome hg19 with SOLiD whole transcriptome alignment pipeline and analyzed by ABI SOLiD Bioscope (version 1.0.1) analysis pipeline. Two mismatches were allowed in each 25 bp seed with progressive alignment to find full mapping location. Mapping score was computed for each read and any location that was scored <26 was filtered out and not processed further. A split-read mapping approach was also applied to map reads which spanned two exons. K562 RNA-Seq data was generated using the Illumina platform. Uniquely-mapped reads were mapped to human genome build hg19 and expression values obtained for all UCSC and RefSeq genes using RSeqTools (Habegger et al., 2010). Expression values were determined in terms of reads per kilobase per million mapped reads (RPKM) for coding sequences.

Reproducibility analysis of RNA-Seq data The reproducibility of RNA-Seq data was evaluated with scatter plots and correlation of the reads. The genome was divided into bins with bin size 10Kb and the number of reads in each bin was counted. The scatter plot of sequence reads per 10Kb in biological replicates (Figure S8S) shows good reproducibility of MCF7 RNA-Seq data (Pearson's correlation coefficient = 0.99). For K562 RNA-Seq data, we compared our in-house RNA-Seq library (as replicate 1) with that from Stanford (ENCODE data, as replicate 2). The scatter plot (Figure S8T) show reasonable correlation between the two K562 RNA-Seq replicates with Pearson's correlation coefficient = 0.72.

Computational characterization of chromatin interactions

Genomic span analysis of chromatin interactions To analyze the genomic spans of chromatin interactions identified by ChIA-PET, we calculated the distance between interacting loci at different PET count cut-offs. As a control, we randomly rewired the connections of the tags in the mapped PETs. Our results show the distribution of genomic spans of intra chromosomal interactions identified in MCF7 and K562 cell-lines. Difference between real and randomly rewired chromatin interactions can be seen. In particular, real chromatin interaction data shows a peculiar hierarchy of chromatin interactions. Short range interactions (~10Kb) , long range interactions (~100Kb) and super long range interactions (>1Mb) suggesting short range enhancer loops at local level, long range enhancer-promoter and promoter-promoter interactions at the middle level and super long range enhancer-promoter and promoter-promoter interactions at the top level respectively.

Analysis of genomic descriptors GC isochores were taken from Costantini *et. al.*, 2006 (Costantini et al., 2006) and Hg17 coordinates were converted to Hg19 using the *liftover* utility of the UCSC genome browser. Gene, SINE and LINE densities (per 200Kb) are mapped around unique RefSeq TSSs engaged in each model. Gene length was calculated by

subtracting TSS from TES coordinates of RefSeq genes present in the three models. To collect the genes present in the three models, an arbitrary distance of +/- 1500 bp from anchor boundary to TSS site is considered. Total gene-span covered by introns is divided by total span covered by exons to calculated intron/exon ratio. Chromosome-wide distribution of the genomic descriptors examined in this study is exemplified in six different chromosomes (**Figure S2A**). This analysis was done on the combined pilot cell-line data and validated at individual cell-line level (**Figure S2B and S2C**)

Promoter CpG analysis For HCG/LCG promoter analysis, the criterion adapted by Saxonov *et. al.* (Saxonov et al., 2006) was used. Briefly, all the promoters (+/- 1500 bp to the non-redundant TSSs) were pulled down from UCSC and the following equation was used to normalize the CpG content:

$$CpG(o/e) = \frac{P_{CG}}{\left[\left(P_C + P_G\right)/2\right]^2}$$

Where, P_{CG}, P_C and P_G are the frequencies of *CG, C* and *G* residues in the sequence respectively.

Promoter strand bias analysis Promoter GC skew is calculated in the sliding window of 100 bp across +/-1500bp around unique TSSs using following equation:

$$GC\ skew = \frac{P_C - P_G}{P_C + P_G}$$

Where, P_C= frequency of Cytosine and P_G= frequency of Guanine

Expression breadth analysis Gene expression data for 84 human tissues was downloaded from GNF Gene Atlas (http://biogps.gnf.org/) (Su et al., 2004). Genes showing at least 100 arbitrary units of probe difference value were considered as expressed. Probes not showing probe value >=100 in any of the tissues were not considered in the analysis. A more stringent cut-off of 150 does not alter our conclusions (data not shown). Expression breadth is defined as the number of tissues a gene is expressed in. Tissue specificity of genes in the enhancer model was further validated by following two descriptors (**Figures S3A and S3B**):

a. $$TS(1) = \frac{\max\left(x_i\right)}{median\left(x_i\right)}$$

b. $$TS(2) = \frac{\sum\left(1 - \frac{x_i}{\max\left(x_i\right)}\right)}{n-1}$$

Where, x_i is the expression of a gene in a tissue *i*, while *i* ranges from *1..n* when *n* is the total number of tissues (84 here).

PCC analysis for co-expression Gene expression data for 4,787 human microarray samples is downloaded from Boolean network dataset (http://gourd.stanford.edu/BooleanNet/), which covers wide range of expression differences like gender, tissues, development and differentiation stages etc (Sahoo et al., 2008). Probes mapping to MG complexes are paired as per their specific interactions in the MG units and Pearson Correlation Coefficient (PCC) is calculated. Supporting analysis based on one probe per gene suggests that multiple probes mapping to single gene does not alter our observations (data not shown). The genes are then randomly rewired to compile a random control with the same gene background, but different pairing than the original MG pairs. Proximal gene pairs (up to 1Mb) from the genomic spans over BP, SG and MG models are also taken as additional controls (**Figure S3G**). The same strategy was implemented for analysis (**Figure S3F**) on Estrogen induced time course Affymatrix data (6 time points) generated from MCF7 cell-line (Fullwood et al., 2009). GRO-Seq data used in Figure 3F was downloaded from (Hah et al., 2011). Similarly RNA-Seq data used in Figure S3H-I was downloaded from ENCODE UCSC page (http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/). Controls taken in the GRO-Seq and RNA-Seq PCC analysis were: (1) randomly rewired MG gene pairs and (2) randomly picked proximal (upto 1Mb) gene pairs from a control dataset having similar distribution of genomic span and gene density as MG regions.

Gene Ontology analysis For Gene Ontology (GO) analysis, loci were mapped to Entrez Genes and GO semantic similarities were calculated by the method defined by Fröhlich *et. al.* (Frohlich et al., 2007). To evaluate whether the higher PCC of paired genes in MG units is contributed by the genes in the same GO class, the PCC data described before was further split into two parts, one for the paired genes from the same GO class (GO similarity=1) and the other for the pairs of genes belonging to different GO classes (GO similarity=0) (**Figure S3L**). Gene pairs from the same GO classes have higher PCC in average.

Gene family analysis Gene family information was downloaded from http://www.genenames.org/genefamily.html and curated manually. The gene family information from the MG models is listed in **Table S4**. To calculate the probability of finding two proximal genes (within 1Mb) from the same gene family, we performed >1 million random simulations. This results to a distribution plot shown in **Figure S3L**, which suggests that gene families are significantly over-represented in MG complexes. To further scrutinize our observations, we compiled a random control dataset with the same genomic span and gene density distribution as of MG units. The dataset shows 3.4 % of regions with at least two genes from the same gene family. Incidentally, the value overlaps precisely with the mean probability of finding 2 proximal genes form the same gene family.

An example of a large multi-gene complex The largest multi-gene structure found in the human combined pilot RNAPII ChIA-PET dataset is on chromosome 11, spanning a 7.8Mb segment (chr11:60927774-68723699) and covering 317 genes (247 genes with promoters proximal to multi-gene anchors and 70 genes in intervening loop regions) (Figure S1G). The

large number of genes involved in multi-gene structures suggests that the multi-gene complexes are a major mechanism to organize multiple genes into distinct foci for efficient and coordinated transcription. It should also be noticed that there are 5090 genes that are included in multi-gene loop regions, but far from any interaction anchors. As loop regions may be far from high local concentrations of transcription-related proteins (Fullwood et al., 2009), genes resided in loop regions may potentially be displaced into transcriptionally inactive zones (**Figure S1G**).

Cell-line specificity analysis of chromatin interactions and genes therein

To identify the cell-line specific genes, we studied the gene expression from MCF7 and K562 cell lines. Each gene had a pair of expression levels from two cell lines, which were used to generate a scatter plot. The gene expression levels from the RNA-Seq library were measured as Reads Per Kilobase Per million reads (RPKM) and normalized with the median expression level from the library and log-tranformed (the expression level of the genes without any reads were replaced with the minimum expression level from the library). The smoothed scatter plot in **Figure 6A** in the main text clearly shows that some genes only expressed in one cell line, but not the other.

We applied a stringent criterion to identify the cell-line specific genes: if a gene has reads from one cell line, but does not have reads from another cell line, this gene is specific to the first cell line. With this criterion, we identified 2025 MCF7 specific genes and 2486 K562 specific genes.

We then examined their chromatin interaction patterns. For each gene, we measured the PET counts from the interactions linked to the gene promoter region. **Figure 6B** in the main text shows the smoothed scatterplot of the PET counts at the same genes from two cell lines. Clearly, cell-line specific genes have more interactions in their corresponding cell line, suggesting that the expression of cell-line specific genes is associated with their cell-line specific chromatin interactions. The cell-line specific genes can be classified into different categories based on the chromatin interaction models they are involved in: genes with enhancer-promoter (EP) interactions, genes with promoter-promoter (PP) interactions, or none of above categories. **Figure 7A** shows the percentage difference in the representation of genes in EP and PP models in cell line specific vs. common interactions. It is apparent that genes involved in enhancer-promoter interactions are significantly enriched in cell-type specific interactions, while genes enagged in PP interactions are enriched in chromatin interactions common in both cell-lines. Cell-line specific enhancer-promoter interactions from MCF7 and K562 cells are summarized in **Table S5**. **Figure 6D-E** shows two examples of the well known chromatin interactions: one for K562-specific chromatin interactions around HBA locus and one for MCF7-specific interactions around *GREB1* gene.

Gene Ontology analysis of cell-line specific genes For all the cell-line specific genes engaged in distinct interaction categories, we examined the enrichment of GO terms (biologocal process) using Panther (Thomas et al., 2003). The highlighted functions are shown as **Figure 6C** in the main text. The complete list of significant GO terms is shown in **Figure S6A**.The genesets common in both cell-lines show enrichment of housekeeping functions as shown in **Figure S6B**. Two cell-line specific interactions are shown in **Figure S6C** (around β–globin gene locus, K562 specific) and **Figure S6D** (around *GATA3* gene locus, MCF7 specific).

Promoter reporter gene assay

To explore the potential regulatory relationship between interacting chromatin loci, we performed the dual luciferase assay for various combinations of promoter and enhancer regions taken from RNAPII ChIA-PET data of MCF7 cells. In these analyses, approximately 500bp fragments of promoter and enhancer regions identified by RNAPII ChIA-PET data are isolated from MCF7 genomic DNA by PCR. All the primers used for plasmid construction are as follows.

Primers used for plasmid construction

Chr	Gene symbol	F-primer (5'-3')	R-primer (5'-3')	Size of genomic DNA fragment
14	*C14ORF102*	GACCCAGTTAGCAAGGATGG	TACAGAAGTCCCAGCTCCAG	705 (promoter)
14	*CALM1*	GACCCGCCGTAGAGGAAA	CTCGGTGGCAGTTCGAGT	716 (enhancer)
14	*CALM1*	CAATTGGCTGTGTGTTCCAC	GCCGCCTGACTACGAGTAAC	989 (promoter)
14	*CALM1*	CAATTGGCTGTGTGTTCCAC	GTCGGGGAAGCGTTCTAGG	756 (TATA-less promoter)
17	*INTS1*	CTCGTTCGCTCGCTCATT	AGCACCACCTGTGAATGGA	567 (promoter)
17	*MAFK*	GGCCTGAAACTCCTCTTCCT	ACTCAGCGGCCAGGTAGC	1158 (promoter)
7	*ELFN1*	CCTGGCATTTGGCATACTCT	CGAGGAATAGGCTGGAGGAT	765 (promoter)
7	*DLD*	TCTGTAGCCCCATCCCTTAT	TTTTCCTTTCCGCCAATACTT	588 (promoter)
14	*DDHD1*	CCTCTCCACCCGAAGTTTCTA	CTCGTATCCACCCACCTATG	567 (promoter)

These fragments are cloned into pGL4.10-basic vector (Promega, Madison, WI). For the promoter activity test, the fragments are cloned flanking to the 5' of the Luciferase gene, whereas for enhancer test the fragments are cloned in the distal location approximately 2Kb to the 5' of the luciferase gene, in either orientations.

For luciferase assay of reporter constructs, a constitutive renilla luciferase expressing vector pRL-SV40 (Promega, Madison, WI) was co-transfected as a control for transfection efficiency. MCF-7 cells were transiently transfected using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) and Opti-MEM® I Reduced Serum Media (Invitrogen, Carlsbad, CA). The luciferase activity was measured following the protocol as previously described (Pan et al., 2008), with minimal three replicates. We calibrated this reporter system using known examples of enhancers and promoters as positive and negative controls and tested a number of promoter-promoter and enhancer-promoter-promoter examples identified in MCF7 cells. The main results are presented in **Figure 5** in the main paper. Additional results are in **Figure S5**.

RNAPII ChIA-PET analysis revealed that the *CALM1* locus has a classic promoter and enhancer structure. RNA-Seq data showed that this gene is actively transcribed, and the H3K4me3 and H3K4me1 marks also characterized the expected promoter and enhancer properties. The normalized luciferase activities from the constructs (Figure S5F) suggested that the construct with promoter alone had significant signal above the background (basic construct and enhancer alone construct). The combination of enhancer and promoter construct showed significant increase of luciferase activity, suggesting that this reporter system is adequate for promoter and enhancer activity analysis.

As a negative control, we deleted the TATA box in the *CALM1* promoter and cloned this TATA-less promoter fragment into the reporter construct. The luciferase assay results (Figure S5H) showed that in all constructs, the TATA-less *CALM1* promoter had no activity whereas the native *CALM1* promoter showed appropriate activities.

We have shown that certain promoters in multi-gene complexes have enhancer functions to other promoters multi-gene complexes (Figure 5E-G). We asked if non-multi-gene promoters have the same property. We cloned the *DDHD1* (an active promoter not in a multi-gene complex) fragment into the *CALM1* promoter construct (Figure S5G), the ELNF1 (an inactive promoter not in a multi-gene complex) and the DLD (an active promoter not in a multi-gene complex) into the *MAFK* promoter constructs (Figure S5I). In all cases, these non-MG promoters did not show enhancements in *CALM1* and *MAFK* promoter activities.

Perturbation experiments by siRNA knockdown and estrogen induction

We used MCF-7 cells for siRNA knockdown experiments. Following standard log phase growth in T75 cell culture flasks, on the first day of the siERα knockdown experiment, we passaged the cells into 6-well cell culture plates, and maintained the cells in phenol red-free DMEM/F12 containing 5% charcoal-dextran-stripped fetal bovine serum to starve the cells of hormones. The next day, we changed the medium to DMEM/F12 without fetal bovine serum, and transfected the cells with siERα (on-targetplus SMARTpool, Dharmacon) or control siRNA (Dharmacon) with DharmaFECT 1 Transfection Reagent (Dharmacon). Successful siRNA knockdown was confirmed using Western Blot against ERα protein as well as RT-qPCR of *ESR1* gene. The next day, we changed the medium to phenol red-free DMEM/F12 containing 5% charcoal-dextran-stripped fetal bovine serum, left the cells to grow for another day, and in the following day, we induced the cells with 100 nM estradiol (Sigma Aldrich) or control vehicle, ethanol.

The quantitative Chromosome Conformation Capture (3C-qPCR) protocol used for this experiment was performed as described in the section “Quantitative analysis of chromosome conformation capture assays (3C-qPCR)” of this Extended Experimental Procedure. The 3C quantitative results are presented as the mean ± s.e.m from two to four independent preparations of 3C sample with duplicate qPCR data.

The cells were induced for 0h, 3h, or 6h, following which the cells were harvested and RNA was extracted using an RNAeasy kit (Qiagen). The RNA was then reverse-transcribed to cDNA using a Superscript III cDNA synthesis kit (Invitrogen). RT-qPCR was then performed using primers designed with Roche Universal ProbeLibrary Assay Design Center against the genes of interest, using a Roche SYBR Green I master mix on a Roche Lightcycler 480 machine. At each time point, the RT-qPCR data were double normalized against GAPDH, and against the siControl, ethanol-treated sample as a baseline. The RT-qPCR data are presented as mean ± s.d. from two independent preparations of RNA sample with duplicate to triplicate qPCR data.

Primer sequences used for RT-qPCR

Gene	Forward primer	Reverse primer
FLJ33534	ggacttcacaggtgaaaggact	aatgtgtcctttggcaacaa
C2orf50	agatgcagcccatctaggag	cgtgtcctgtcctgtgtaactc
PQLC3	ccctgcagaagtggatcatag	cctgagtctctcgtcttccac
PDIA6	ggatccaacaaaaacagacca	ctcagcgcagcatctacaat
NOL10	gatgggcatctattacattccag	ttcggtcaagttgtctaagaagg
ATP6V1C2	ctatgatgagaaggaaattgaaagg	aagagcaacacaggaagtttgat
KCNF1	cgacatctaccccaagacca	gcgatgacaccacacaagaa
ROCK2	acagcttgccccaaacaa	tggaagaatacgatcaccttga
ODC1	aaaacatgggcgcttacact	tggaattgctgcatgagttg
HPCAL1	ggcaacggctacatcagc	tcatcacagacgacaccatct
TAF1B	cgtcagcaaagcatcacaat	gattcccttctcctttcgttg
GRHL1	tgagccagatcacagcaaa	acatttttcagtacttgcactctgtt
E2F6	ggcgaggaagttacccagt	cagatttgcctgaataaacaacag
ESR1	ttactgaccaacctggcaga	atcatggagggtcaaatcca
GREB1	tgtggagtgcctgaagtgac	ctcagcagagacgaagaaagg
XBP1	ggagttaagacagcgcttgg	cactggcctcacttcattcc
HSCB	acctgaccccactcgagac	tcaactctgaaggaacggttg
CHEK2	caggtttagcgccactctg	gactcccgagacatcacgac
EWSR1	actgcaacctatgggcagac	tggacaggctggctgtatg
RFPL1	aggatgcggaagttccaag	ctgtgtgatgcacccacttc
NEFH	ccgacattgcctcctacc	ggccatctcccacttggt
CCDC88C	ggaggaaatcaagaaggtgct	tcctctttcctctcacactgg
GPR68	ttgagggagttataggcagagg	gcggctctcccagttctt
C14orf159	atgaggcccattcccaag	gctctttgattcccaacagttc

GLOSSARY

Here is the list of important terms used in this study. Some of the terms are similar to those used in our previous paper (Fullwood et al., 2009), and are reproduced here for convenience of the readers.

1. **Tag**: A short fragment of DNA sequence, here about 20-21 bp, derived from a ChIP DNA fragment. The tag sequence is mapped to the human reference genome, thus giving the genomic location of the tag.
2. **PET**: Abbreviation of “Paired-End diTag”. Sequenced construct consisting of two covalently linked tags, such that the relationship between the two tags is known. The tags come from two DNA ends.
3. **Self-ligation PET**: A PET arising from a ligation between the two ends of the same ChIP DNA fragment.
4. **Inter-ligation PET**: A PET arising from a ligation between the ends of two different ChIP DNA fragments as represented by two Tag-defined ChIP fragments.

5. **Inter-ligation PET cluster**: A group of inter-ligation PETs whose genomic locations of tag-defined ChIP fragments are directly overlapping. They are considered as derived from the same interacting pair of genomic regions.
6. **Interaction**: Generally refers to an interconnection between two or multiple loci (interacting anchors). Conceptually speaking, this is the genomic region defined by two or more overlapping tag-defined ChIP fragments. In data analysis, we defined interactions in the following way: as each inter-ligation PET was derived from two ChIP DNA fragments, the majority of which were less than 1,500 bp in size, we extended the mapped 20 bp tags to 1,500 bp along the reference genome to represent the virtual DNA fragments in pairs, and then we defined PETs with both extended tags overlapped as an interaction cluster. We used the number of PETs that connect two anchors as the measurement of interacting frequency or strength of an interaction between two regions, and 2 or more inter-ligation PETs was used as a cut-off to define an interaction.
7. **Interaction anchor**: In the genome browser visualization, and conceptually speaking, this is the genomic region covered by two or more overlapping tag-defined ChIP fragments from inter-ligation PETs. In the data analysis, as each inter-ligation PET was derived from two ChIP DNA fragments, the majority of which were less than 1,500 bp in size, we extended the mapped 20 bp tags to 1,500 bp along the reference genome to represent the virtual DNA fragments in pairs, and then we defined peaks from the profile. These peaks are considered equivalent to “anchors”, and are called as such.
8. **Duplex interaction**: The basic unit of chromatin interaction, involving two anchors and one intermediate loop region.
9. **Stand-alone duplex interaction**: Duplex interactions with anchors that do not overlap with anchor regions of other interactions. A stand-alone duplex is made up of only 1 interaction, and must have at least two overlapping inter-ligation PETs. On average, stand-alone duplex interactions are weaker than complex interactions.
10. **Complex interaction**: A complex interaction that has two or more anchors, indicating the interactions were further clustered upon clustering of duplex interactions. As a complex interaction is made up of at least 2 interactions, each of which has at least two inter-ligation PETs, a complex interaction must have at least four overlapping inter-ligation PETs. The 5’ most and 3’ most anchors on either end constitute the boundaries of the complex interaction.
11. **Interaction region**: Refer to a distinct genomic region after further clustering of overlapped duplex interactions. Such regions can be from stand-alone duplex interactions and complex interactions.
12. **Loop**: This is the genomic region between two adjacent anchors within an interaction region.
13. **TSS**: Transcription Start Site.
14. **TES**: Transcription End Site
15. **Anchor gene**: A gene with a TSS of a transcriptional unit that is proximal (±5Kb) of any anchor in a complex interaction or stand-alone duplex interaction.
16. **Loop gene**: A gene with a TSS of a transcriptional unit that does not fall within ±5Kb of any anchor in a complex interaction or stand-alone duplex interaction, but is within the interaction boundaries (and hence falls within the “loop” regions).
17. **ChIA-PET**: Chromatin Interaction Analysis with Paired-End-Tag sequencing

18. **Transcription Factory**: Active gene transcription units clustered in a eukaryotic nucleus, in discrete foci. Such foci can be visualized by tagging nascent RNA using immuno-labeled precursors (Br-UTP or Br-U) or by immuno-fluorescence in situ hybridization of RNA polymerase antibodies.
19. **Basal Promoter (BP) model**: the promoter regions of target genes are bound with standalone RNAPII peaks, but do not involve any chromatin interactions. The promoters likely involve only basal promoter function recognized by RNAPII binding; therefore, this model is termed the "basal promoter model".
20. **Single-Gene (SG) interaction complex**: RNAPII binds at promoter region of target gene and interacts with distal regulatory elements such as enhancers. Usually this model involves a single gene, and with one or multiple enhancer (multipartite enhancer) sites surrounding the target gene. Therefore, this model may include enhancer-to-enhancer and enhancer-to-promoter complex interactions.
21. **Multi-Gene (MG) interaction complex**: This model involves complex RNAPII interactions of multiple genes, in which distal regulatory elements and promoters of proximal genes interact with one another. Multiple genes are involved in the same interaction complexes.
22. **Super-MG complex**: Multiple MG units cross over large genic regions or even in trans and interconnect to each other to form higher order complex structures, termed "super-MG complex" or "higher order MG complex".

SUPPLEMENTAL REFERENCES

Baillargeon, S., and Rivest, L.P. (2007). Rcapture: Loglinear models for capture-recapture in R. Journal of Statistical Software *19*, http://www.jstatsoft.org/v19/i05.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B *57*, 289–300.

Costantini, M., Clay, O., Auletta, F., and Bernardi, G. (2006). An isochore map of human chromosomes. Genome Res *16*, 536-541.

Frohlich, H., Speer, N., Poustka, A., and Beissbarth, T. (2007). GOSim--an R-package for computation of information theoretic GO similarities between terms and gene products. BMC Bioinformatics *8*, 166.

Fullwood, M.J., Han, Y., Wei, C.L., Ruan, X., and Ruan, Y. (2010). Chromatin interaction analysis using paired-end tag sequencing. Curr Protoc Mol Biol *Chapter 21*, Unit 21 15 21-25.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., *et al.* (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58-64.

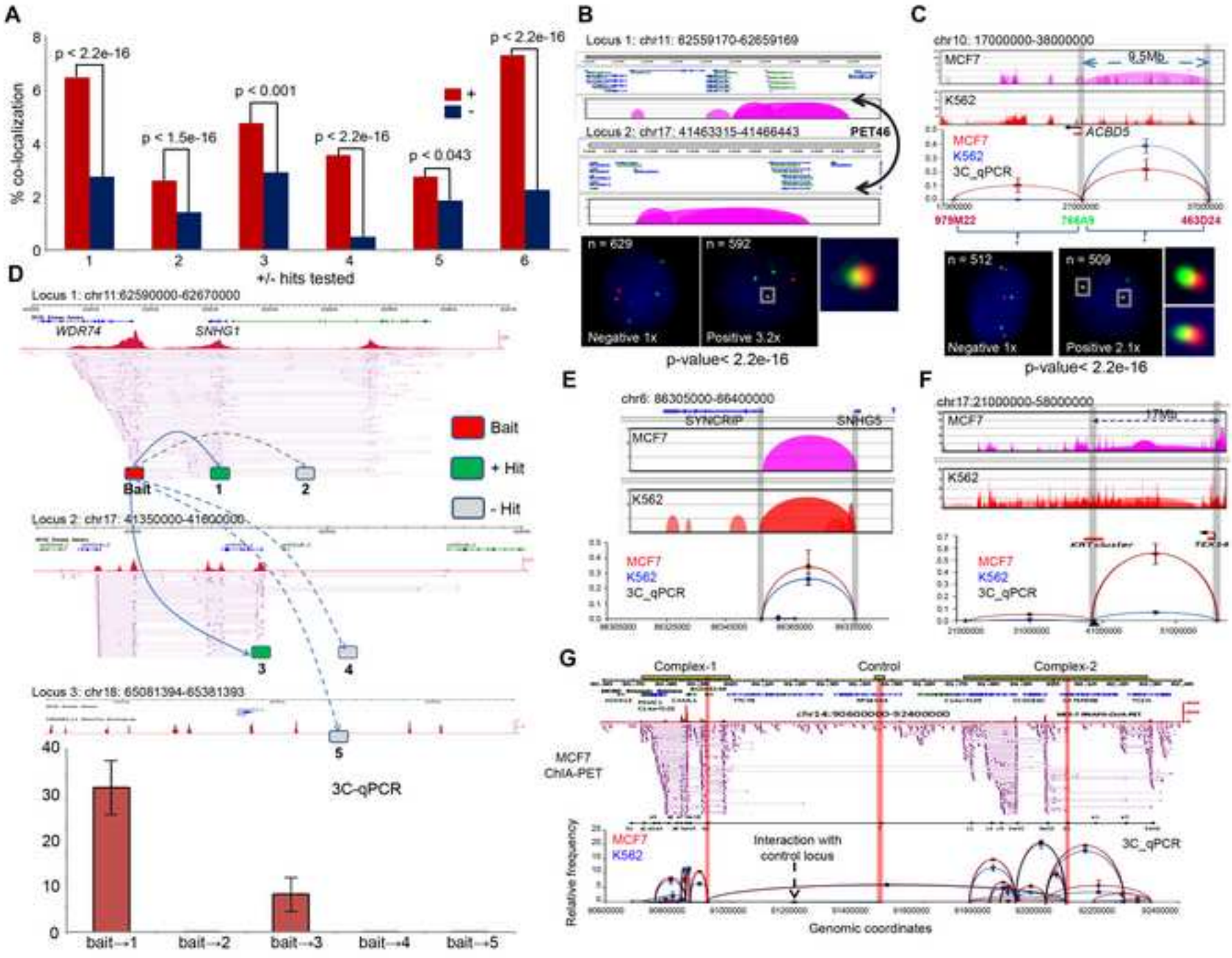
Habegger, L., Sboner, A., Gianoulis, T.A., Rozowsky, J., Agarwal, A., Snyder, M., and Gerstein, M. (2010). RSEQtools: A modular framework to analyze RNA-Seq data using compact, anonymized data summaries. Bioinformatics.

Hagege, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forne, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc *2*, 1722-1733.

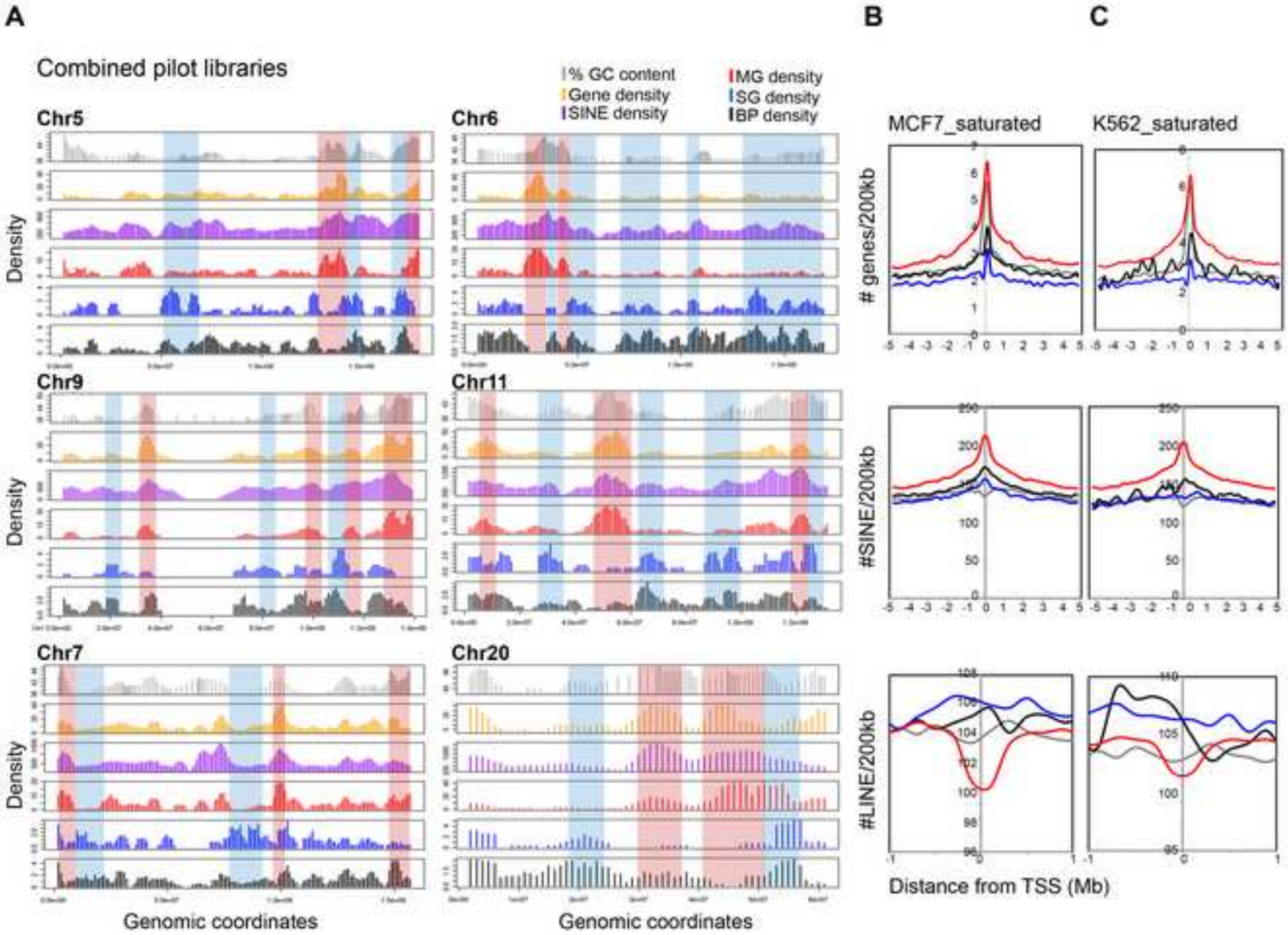
Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, W.L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell *145*, 622-634.

Jee, J., Rozowsky, J., Yip, K.Y., Lochovsky, L., Bjornson, R., Zhong, G., Zhang, Z., Fu, Y., Wang, J., Weng, Z.*, et al.* (2011). ACT: Aggregation and Correlation Toolbox for Analyses of Genome Tracks (in press). Bioinformatics.
Joseph, R., Orlov, Y.L., Huss, M., Sun, W., Kong, S.L., Ukil, L., Pan, Y.F., Li, G., Lim, M., Thomsen, J.S.*, et al.* (2010). Integrative model of genomic factors for determining binding site selection by estrogen receptor α. Molecular Systems Biology *6*, 456.
Kuhn, R.M., Karolchik, D., Zweig, A.S., Wang, T., Smith, K.E., Rosenbloom, K.R., Rhead, B., Raney, B.J., Pohl, A., Pheasant, M.*, et al.* (2009). The UCSC Genome Browser Database: update 2009. Nucleic Acids Res *37*, D755-761.
Li, G., Fullwood, M.J., Xu, H., Mulawadi, F.H., Velkov, S., Vega, V., Ariyaratne, P.N., Mohamed, Y.B., Ooi, H.S., Tennakoon, C.*, et al.* (2010). ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biol *11*, R22.
Pan, Y.F., Wansa, K.D., Liu, M.H., Zhao, B., Hong, S.Z., Tan, P.Y., Lim, K.S., Bourque, G., Liu, E.T., and Cheung, E. (2008). Regulation of estrogen receptor-mediated long range transcription via evolutionarily conserved distal response elements. J Biol Chem *283*, 32977-32988.
Raha, D., Hong, M., and Snyder, M. (2010). ChIP-Seq: a method for global identification of regulatory elements in the genome. Curr Protoc Mol Biol *Chapter 21*, Unit 21 19 21-14.
Roussel, M.J., and Lanotte, M. (2001). Maturation sensitive and resistant t(15;17) NB4 cell lines as tools for APL physiopathology: nomenclature of cells and repertory of their known genetic alterations and phenotypes. Oncogene *20*, 7287-7291.
Ruan, Y., Ooi, H.S., Choo, S.W., Chiu, K.P., Zhao, X.D., Srinivasan, K.G., Yao, F., Choo, C.Y., Liu, J., Ariyaratne, P.*, et al.* (2007). Fusion transcripts and transcribed retrotransposed loci discovered through comprehensive transcriptome analysis using Paired-End diTags (PETs). Genome Res *17*, 828-838.
Sahoo, D., Dill, D.L., Gentles, A.J., Tibshirani, R., and Plevritis, S.K. (2008). Boolean implication networks derived from large scale, whole genome microarray datasets. Genome Biol *9*, R157.
Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A *103*, 1412-1417.
Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G.*, et al.* (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A *101*, 6062-6067.
Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome Res *13*, 2129-2141.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

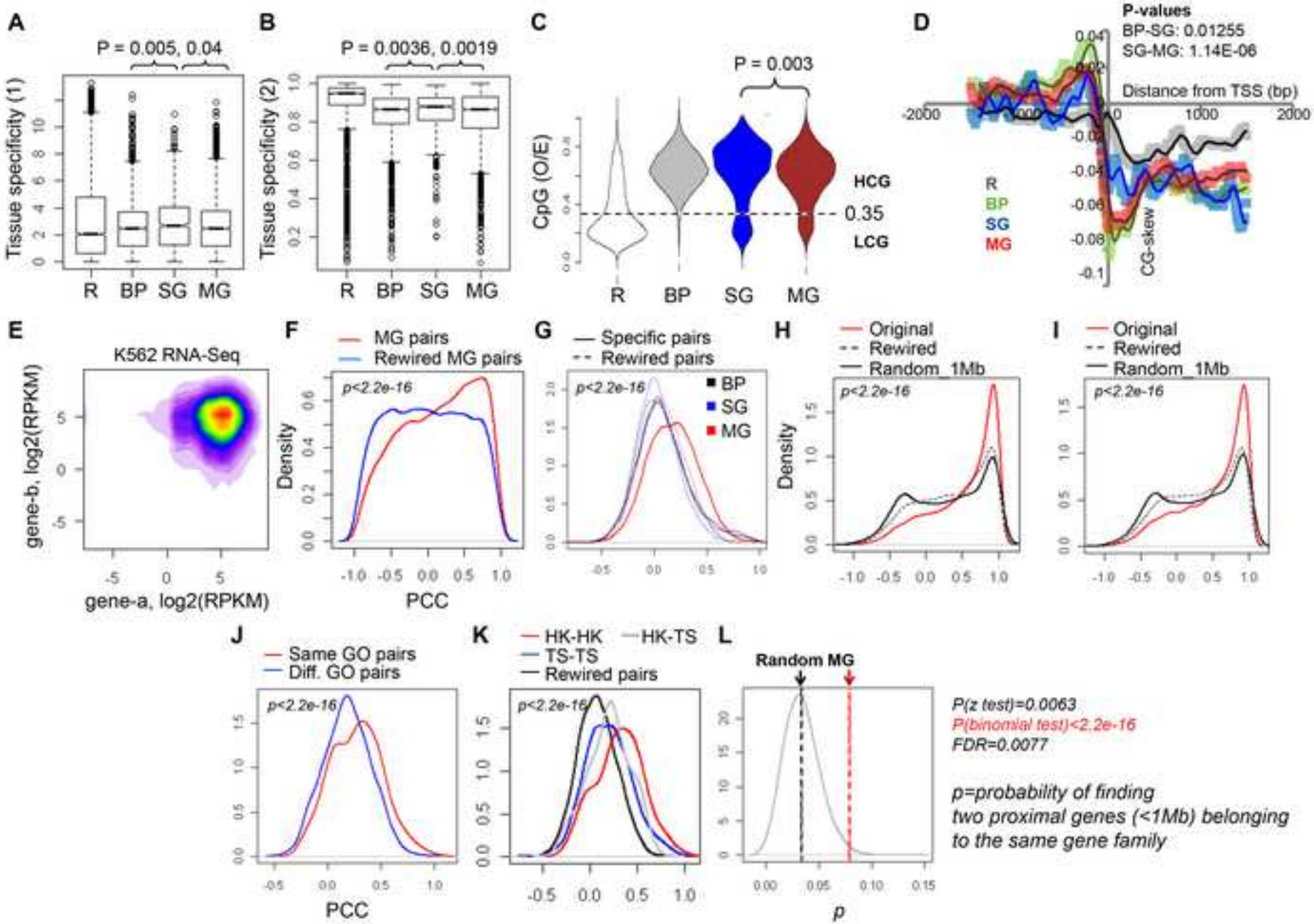
Supplemental Figure



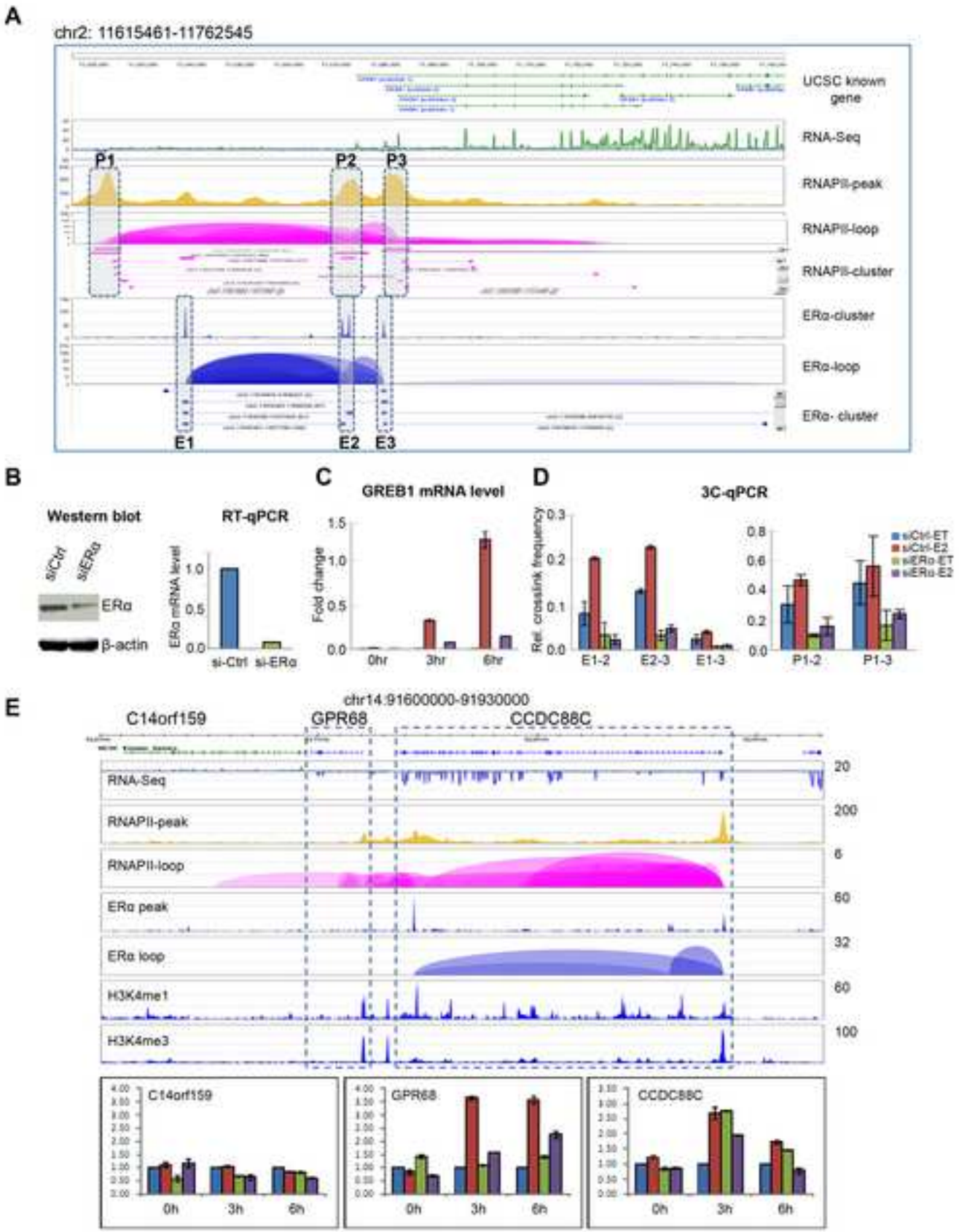
Supplemental Figure



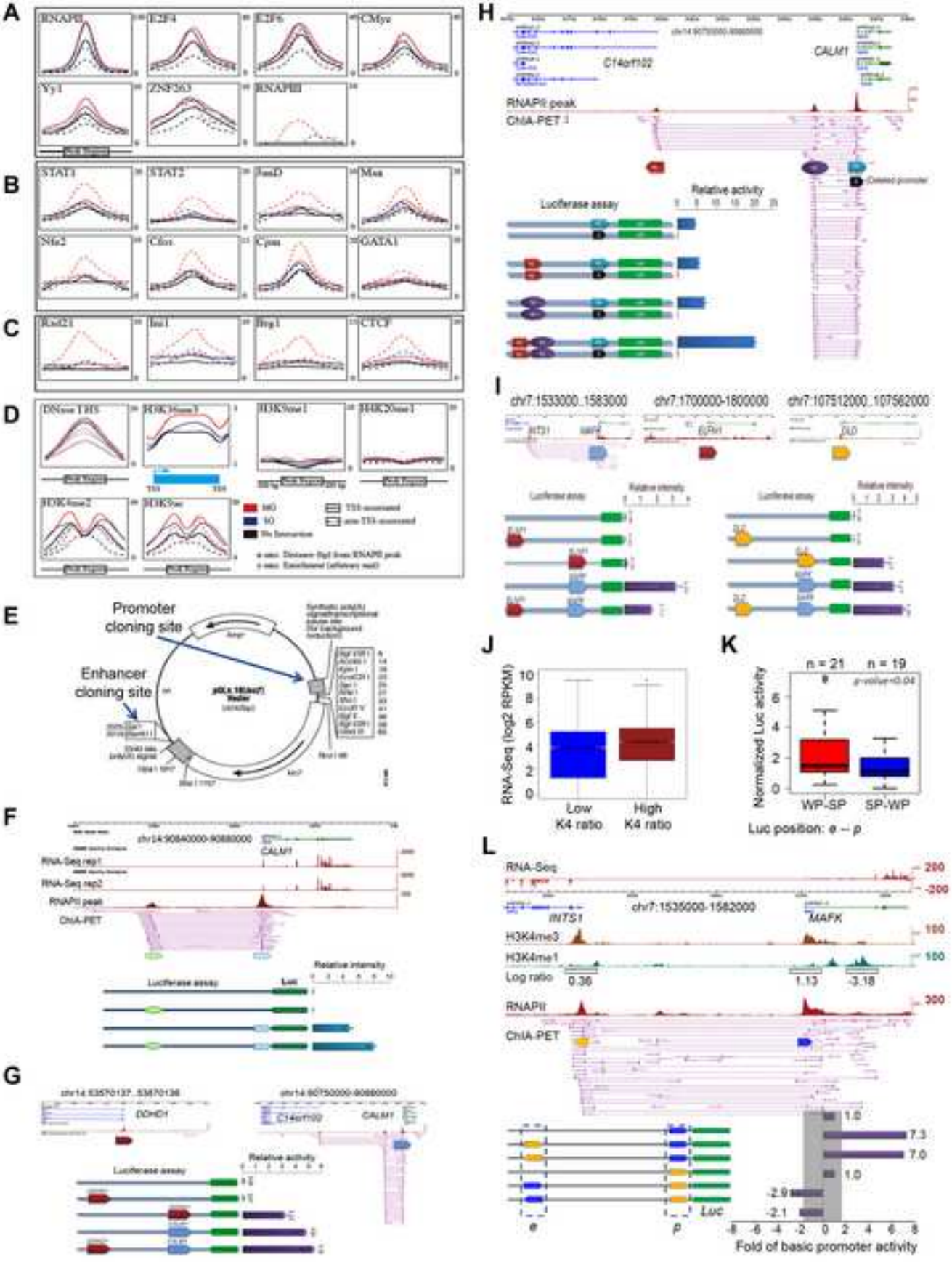
Supplemental Figure



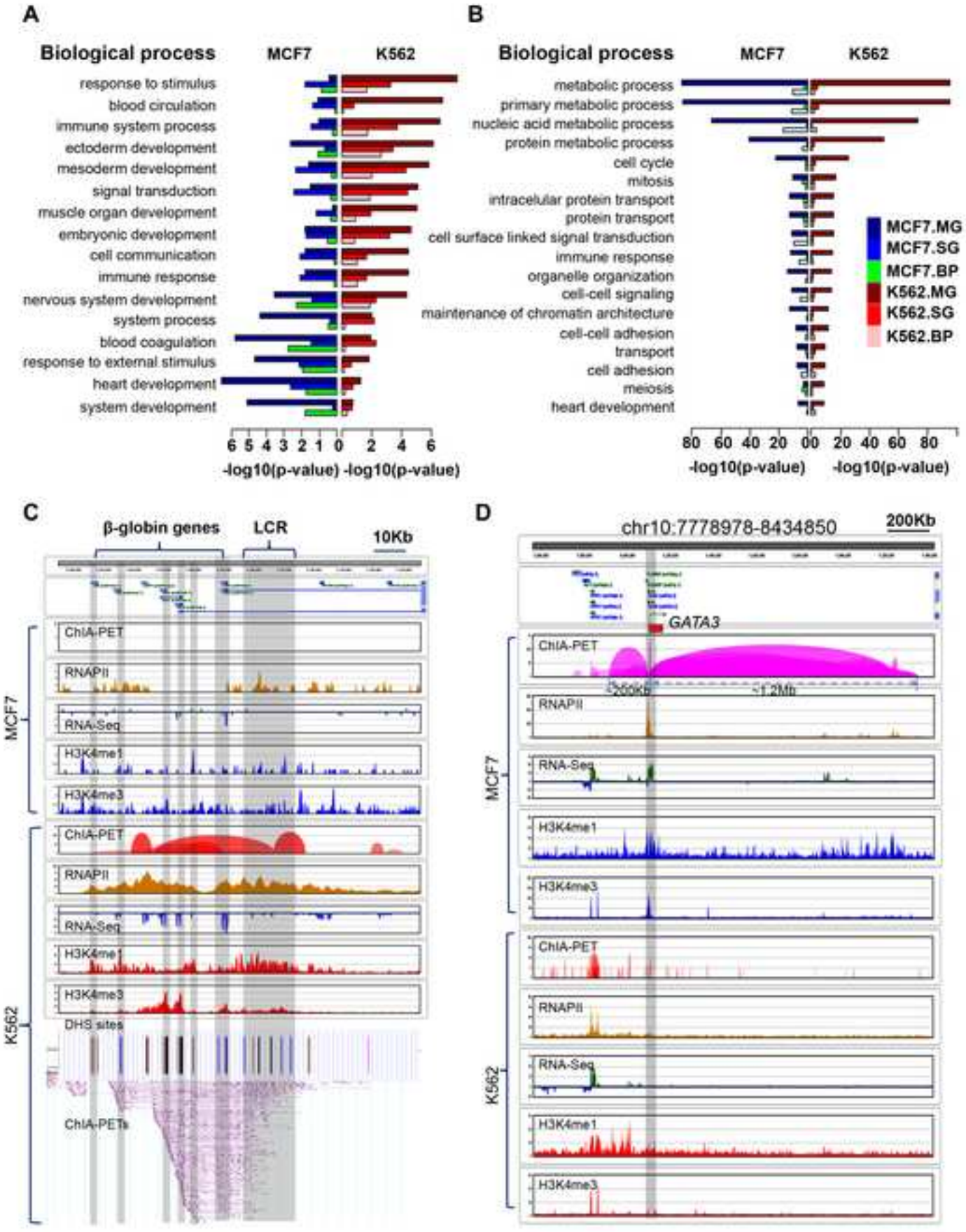
Supplemental Figure



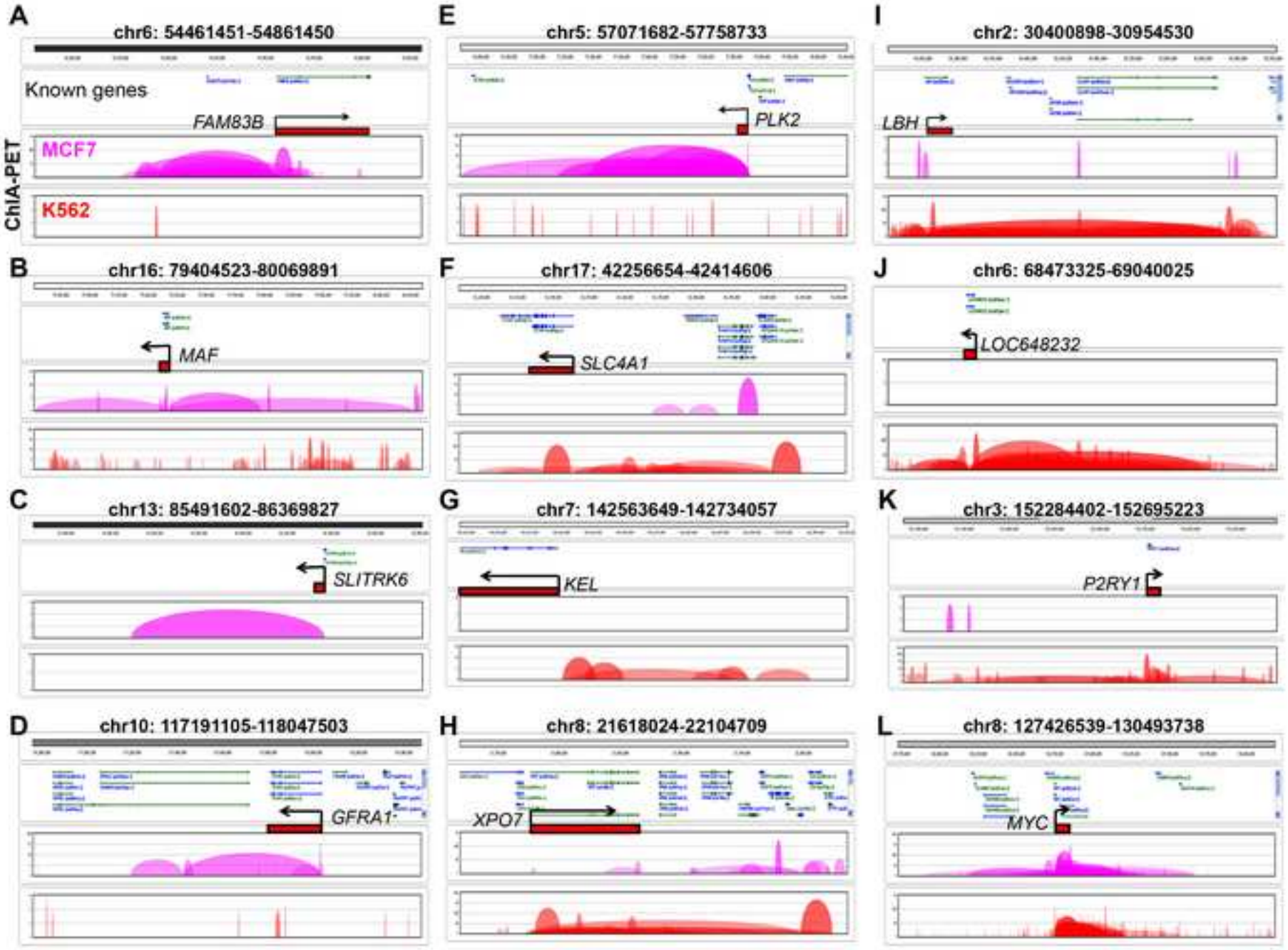
Supplemental Figure



Supplemental Figure



Supplemental Figure



Supplemental Figure