

# Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation

Guoliang Li,<sup>1,10</sup> Xiaolan Ruan,<sup>1,10</sup> Raymond K. Auerbach,<sup>2,10</sup> Kuljeet Singh Sandhu,<sup>1,10</sup> Meizhen Zheng,<sup>1</sup> Ping Wang,<sup>1</sup> Huay Mei Poh,<sup>1</sup> Yufen Goh,<sup>1</sup> Joanne Lim,<sup>1</sup> Jingyao Zhang,<sup>1</sup> Hui Shan Sim,<sup>1</sup> Su Qin Peh,<sup>1</sup> Fabianus Hendriyan Mulawadi,<sup>1</sup> Chin Thing Ong,<sup>1</sup> Yuriy L. Orlov,<sup>1</sup> Shuzhen Hong,<sup>1</sup> Zhizhuo Zhang,<sup>3</sup> Steve Landt,<sup>4</sup> Debasish Raha,<sup>4</sup> Ghia Euskirchen,<sup>4</sup> Chia-Lin Wei,<sup>1</sup> Weihong Ge,<sup>5</sup> Huaian Wang,<sup>6</sup> Carrie Davis,<sup>6</sup> Katherine I. Fisher-Aylor,<sup>7</sup> Ali Mortazavi,<sup>7</sup> Mark Gerstein,<sup>2</sup> Thomas Gingeras,<sup>6</sup> Barbara Wold,<sup>7</sup> Yi Sun,<sup>5</sup> Melissa J. Fullwood,<sup>1</sup> Edwin Cheung,<sup>1,8</sup> Edison Liu,<sup>1</sup> Wing-Kin Sung,<sup>1,3</sup> Michael Snyder,<sup>4,\*</sup> and Yijun Ruan<sup>1,9,\*</sup>

<sup>1</sup>Genome Institute of Singapore, Singapore 138672, Republic of Singapore

<sup>2</sup>Program in Computational Biology and Departments of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520, USA

<sup>3</sup>Department of Computer Science, School of Computing, National University of Singapore, Singapore 117417, Republic of Singapore

<sup>4</sup>Center for Genomics and Personalized Medicine, Department of Genetics, Stanford University, Stanford, CA 94305, USA

<sup>5</sup>Department of Molecular and Medical Pharmacology, UCLA, Los Angeles, CA 90095, USA

<sup>6</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11797, USA

<sup>7</sup>Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

<sup>8</sup>School of Biological Sciences, Nanyang Technological University, Singapore 637551, Republic of Singapore

<sup>9</sup>College of Life Sciences and Technology, Huazhong Agricultural University, Wuhan 430070, P.R. China

<sup>10</sup>These authors contributed equally to this work

\*Correspondence: [mpsnnyder@stanford.edu](mailto:mpsnnyder@stanford.edu) (M.S.), [ruanyj@gis.a-star.edu.sg](mailto:ruanyj@gis.a-star.edu.sg) (Y.R.)

DOI 10.1016/j.cell.2011.12.014

## SUMMARY

Higher-order chromosomal organization for transcription regulation is poorly understood in eukaryotes. Using genome-wide Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET), we mapped long-range chromatin interactions associated with RNA polymerase II in human cells and uncovered widespread promoter-centered intragenic, extragenic, and intergenic interactions. These interactions further aggregated into higher-order clusters, wherein proximal and distal genes were engaged through promoter-promoter interactions. Most genes with promoter-promoter interactions were active and transcribed cooperatively, and some interacting promoters could influence each other implying combinatorial complexity of transcriptional controls. Comparative analyses of different cell lines showed that cell-specific chromatin interactions could provide structural frameworks for cell-specific transcription, and suggested significant enrichment of enhancer-promoter interactions for cell-specific functions. Furthermore, genetically-identified disease-associated noncoding elements were found to be spatially engaged with corresponding genes through long-range interactions. Overall, our study provides insights into transcription regulation by three-dimensional chromatin

interactions for both housekeeping and cell-specific genes in human cells.

## INTRODUCTION

A fundamental question in biology is how genes and regulatory regions are organized and coordinated for transcription regulation. While operons, in which one promoter transcribes multiple genes in a single unit, are common in bacteria (Jacob et al., 1960), and bicistronic transcript structures have been described in worms and flies (Pauli et al., 1988; Zorio et al., 1994), eukaryotic genes are thought to be individually transcribed from their own promoters. However, evidence from in situ fluorescence studies in the last decade suggests that transcription is not evenly distributed and is instead concentrated within large discrete foci in mammalian nuclei, raising the possibility that genes are organized into “transcription factories” (Cook, 1999) containing RNA polymerase II (RNAPII) and other components for transcription. However, this theory lacks evidence with molecular and structural details. Thus, the question of how the regulation of genes is coordinated for transcription in mammalian cells remains largely open.

Mammalian genomes are known to be organized intensively into higher-order conformation inside the micron-sized nuclear space. Consequently, three-dimensional (3D) organization must have a role in the mechanisms for transcription regulation and coordination (Cremer and Cremer, 2001). Chromosome Conformation Capture (3C) and similar techniques (van Steensel and Dekker, 2010) along with traditional in situ techniques have demonstrated that chromatin interactions can regulate

transcriptional and epigenetic states (Cope et al., 2010). However, such analyses are either limited to certain specific domains or of low resolution and lack functional details. Therefore, a global and high-resolution map of functional chromatin interactions is likely to uncover underlying principles of the higher-order genomic architectures regulating transcription.

Recently, we developed Chromatin Interaction Analysis by Paired-End-Tag sequencing (ChIA-PET) for genome-wide investigation of chromatin interactions bound by specific protein factors (Fullwood et al., 2009). By immunoprecipitation of a factor of interest along with associated DNA fragments and followed by diluted proximity ligation of distant DNA fragments tethered together within individual chromatin complexes, we elucidated the association of regulatory information through nonlinear arrangements. We demonstrated that long-range chromatin interactions occur between the transcription factor Estrogen Receptor  $\alpha$  (ER $\alpha$ ) bound regions and their target promoters. To globally investigate how all active promoters dynamically interact with their corresponding regulatory regions in vivo, we used ChIA-PET to analyze genome-wide chromatin interactions associated with RNAPII. Our results provide insights into the 3D interplay of active promoters as well as regulatory regions and suggest an architectural model in which related genes in mega-base range are organized for efficient and potentially cooperative transcription.

## RESULTS

### Organizational Complexity of RNAPII-Associated Chromatin Interactions

We analyzed five different human cell lines (MCF7, K562, HeLa, HCT116, and NB4) using ChIA-PET with a RNAPII antibody (8WG16) that recognizes the initiation form of the protein. The cell lines originated from a wide range of lineages, and provided a broad representation of human cells. In our pilot analysis, about 20 million uniquely mapped paired-end reads were generated for each of the ChIA-PET experiments (Table S1A available online), which resulted in two genome-wide datasets: the ChIP-enriched RNAPII binding sites and the RNAPII-bound long-range chromatin interactions. Both intrachromosomal and interchromosomal interaction data were obtained, and the vast majority of chromatin interactions identified by ChIA-PET were intrachromosomal (Table S1B). Twenty-five intrachromosomal and seven interchromosomal interactions were validated either by 3C, DNA-FISH, or both (Figure S1 and inset of Figure 1C).

To present an inclusive view of the RNAPII-associated human chromatin interactome, we combined the ChIA-PET sequence reads from the six pilot experiments into one dataset for analysis (Table S1). Using embedded nucleotide barcode controls and statistical analyses, we assessed the data quality, filtered out the technical noise, and identified high-confidence binding sites and interacting PET clusters (Experimental Procedures). From the combined pilot dataset, we identified 14,604 high-confidence (FDR < 0.05) RNAPII binding sites as well as 19,856 high-confidence intrachromosomal interaction PET clusters (Table S3). The majority (83%) of RNAPII binding sites in the combined dataset were proximal to 5' Transcription Start Sites (TSS) of genes (Figure 1A). There were also distinct but relatively

weaker enrichments of peaks at the 3' Transcription End Sites (TES) of genes. Similar patterns were seen in all the individual experiments. Of the total RNAPII binding sites, 9,487 (65%) were involved in chromatin interactions and these sites showed higher RNAPII occupancy than those not involved in interactions (Figure 1B), indicating that most highly-enriched RNAPII binding sites are involved in looped chromatin conformations.

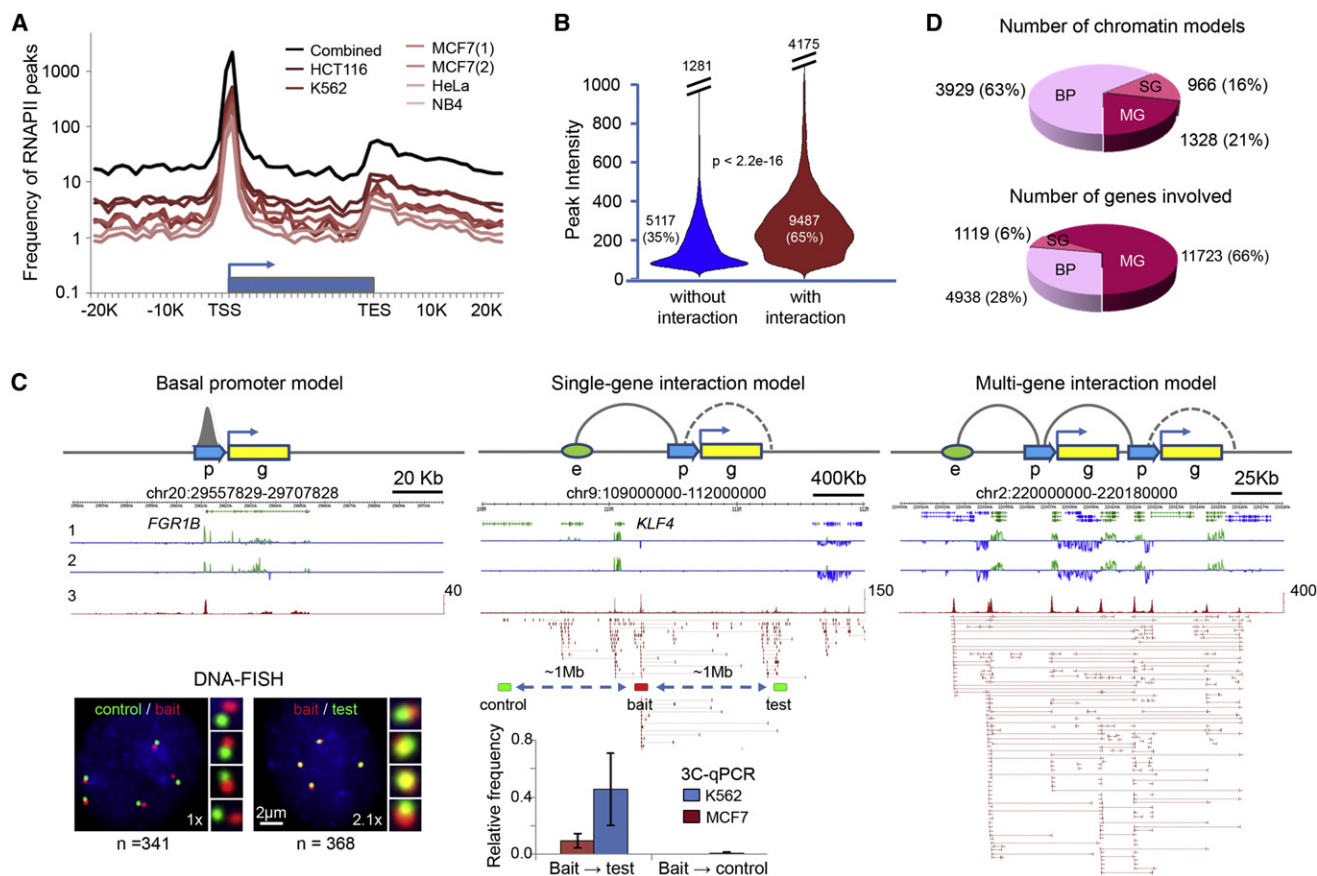
Three basic types of interactions were identified around gene promoters in the combined pilot dataset: intragenic (promoter to gene internal regions, 938, 5%), extragenic (promoter to distal regulatory elements such as enhancer, 6,530, 33%), and intergenic (promoter-promoter of different genes, 8,282, 42%). There was also a subcategory composed of intermediate enhancer-enhancer interactions (4,106, 20%). Some interactions (2,341, 12%) were standalone duplex interactions between two interacting anchor regions, whereas most (17,515, 88%) were further aggregated into 1,544 interaction complexes.

We speculated that the isolated RNAPII binding at promoter sites, which are not involved in interactions, may reflect the basal promoter function for gene transcription, and thus were termed "basal promoters." By contrast, RNAPII-associated interactions might constitute a structural basis for complex regulatory mechanisms. These basic interactions further aggregated into complex architectures which we classified as "single-gene" or "multigene" complexes depending on the number of genes involved (Figure 1C). The single-gene models consisted of single or multiple enhancer interactions with only one gene promoter, whereas the multigene models included intergenic promoter-promoter interactions and could also include intragenic and extragenic enhancer-promoter interactions. Moreover, several such complexes, distantly separated on a chromosome or on different chromosomes, further converged to form higher-order multigene interaction complexes (Figures S1B, S1D, S1F, and S1G). Many chromatin complexes had genomic spans of 150 kb–200 kb, and a few complexes spanned several megabases. Although there were only 1,328 multigene complexes in this combined pilot dataset, 11,723 genes were engaged in these complexes for an average of 8.8 genes per interaction complex (Figure 1D), indicating that promoter-promoter interactions were widespread and may play a significant role in transcription regulation.

To understand how these looping structures influence transcription, we characterized these RNAPII-associated chromatin models (basal promoters, single-gene and multigene complexes) for structural features (genomic property), functional output (transcription activity), and epigenomic marks (chromatin state).

### Distinct Genomic Properties of Single- and Multigene Interaction Models

To determine the genomic characteristics of RNAPII-associated chromatin structures, we mapped several genomic descriptors that were known to associate with the expressivity of the human genome (Versteeg et al., 2003), including GC content, gene density, SINE/LINE density, gene length, and the intron/exon ratio. In our analyses (Figure 2, Figure S2A), the multigene complexes were significantly enriched with higher GC content, higher gene and SINE density, and lower LINE density as



**Figure 1. Characterization of RNAPII Binding Peaks and Chromatin Interactions**

(A) RNAPII binding profile around gene body.

(B) Violin plots for intensities of RNAPII peaks involved (red, mean intensity = 281) and not involved in interactions (blue, mean intensity = 141).

(C) RNAPII-associated chromatin models: basal promoter (BP) with RNAPII binding but no chromatin interaction, single-gene (SG) complex with intra- and/or extragenic interactions and multigene (MG) complex with multiple genes in the interaction clusters. p, promoter; g, gene; and e, enhancer. The dotted curve for possible intragenic loop, and the solid curve for potential loop of enhancer-promoter and promoter-promoter interactions. Data tracks are: 1 and 2, strand specific RNA-Seq data of MCF7 and K562; 3, RNAPII binding peaks and ChIA-PET data. Inset (bottom): DNA-FISH and 3C-qPCR validations of the extragenic interaction at the *KLF4* locus, where the *KLF4* promoter and enhancer are ~1 Mb apart. Genomic locations used for 3C bait, test and control sites are indicated. The same locations were also used for DNA-FISH. The numbers (n) of nuclei counted and the fold change (x) in the number of instances showing close proximity ( $\leq 1 \mu\text{m}$ ) are indicated. 3C-qPCR mean values and standard error of means (SEM) from three independent experiments are shown.

(D) Distribution of chromatin models (BP, SG, MG) and the numbers of genes engaged in the models.

See also Figure S1, Table S1, Table S2, and Table S3.

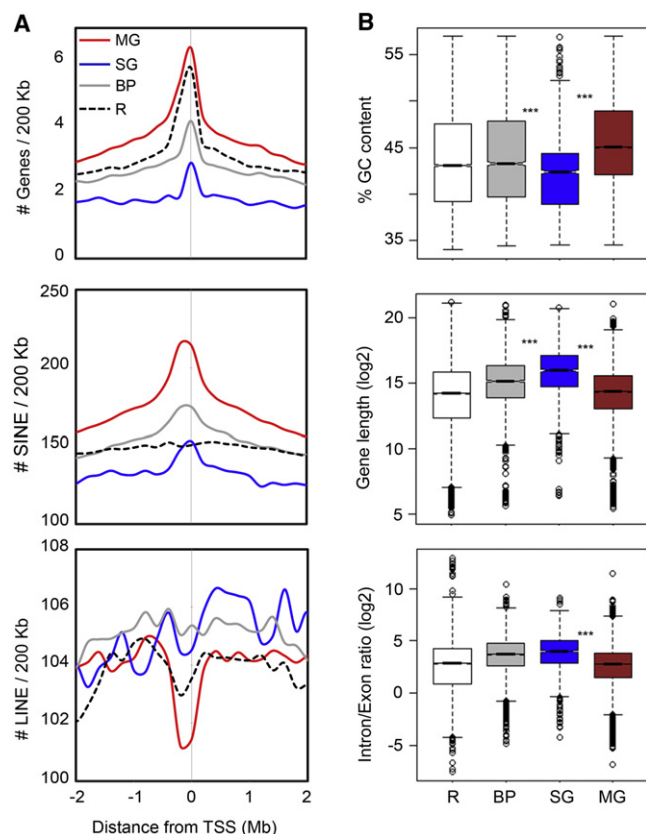
compared to the single-gene interaction complexes and the regions of basal promoters, suggesting that multigene complexes were located in open chromatin and highly transcribed regions. In addition, genes in the multigene complex regions were relatively shorter than other gene categories, which is yet another property of highly expressed genes (Eisenberg and Levanon, 2003). Conversely, genomic loci associated with the single-gene complexes lay in the regions with lower gene and SINE density. Moreover, the genes engaged in the single-gene complexes were significantly longer and had higher intron/exon ratios than the genes of other chromatin models (Figure 2B). These observations suggest that genes with enhancer-promoter interactions in single-gene complexes were more likely to be tissue-specific or developmentally regulated, in line with the previous findings that genes in gene-poor regions associated

with several distant regulatory elements, tended to be longer and had a higher noncoding to coding ratio than housekeeping genes (Eisenberg and Levanon, 2003; Taylor, 2005).

### Interacting Genes Show Correlated Expression

To investigate the functional output of genes involved in the different chromatin models, as defined by transcriptional activity, we focused our analyses on MCF7 cells, as it is a well-characterized human cancer cell model with complementary datasets including RNA-Seq (Experimental Procedures), time-course microarray gene expression (Fullwood et al., 2009), and GRO-Seq datasets (Hah et al., 2011).

Consistent with the combined pilot dataset, 90% binding sites in MCF7 cells were found proximal to known gene promoters and 97% genes with RNAPII present at their promoters had



**Figure 2. Genomic Properties of Promoter-Centered Chromatin Models**

(A) Aggregation plots showing enrichment of genes, SINE and LINE elements around the TSS of genes in different chromatin models. Unique RefSeq TSSs were used for analyses. Red curve stands for multigene (MG) model, blue for single-gene (SG) model, gray for basal promoter (BP) model, and black dotted line for the rest of the genes (R).

(B) Box-plots showing distribution of percentage GC content of GC isochore around different models, gene length, and intron/exon ratio of RefSeq genes involved in the models. Triple asterisks (\*\*\*) signifies  $p$  value <  $2.2 \times 10^{-16}$ . Red box stands for MG, blue for SG, and gray for BP. Open box is for R (rest of genic regions) as background.

See also Figure S2.

detectable transcriptional activity by RNA-Seq (Figure 3A). The interactive RNAPII binding sites that were distal to gene promoters included intra- and extragenic regulatory elements such as enhancers. Approximately 45% of the extragenic distal regulatory sites had detectable RNA signals that could represent possible noncoding RNA (ncRNA) transcripts.

For genes associated with the three chromatin models, we analyzed the transcription levels measured by RNA-Seq reads. As shown in Figure 3B, in general, RNAPII binding at promoter sites correlated well with the expression level of the corresponding genes. Interestingly, the genes involved in the single-gene and the multigene models showed higher correlation between RNAPII binding and RNA-Seq signal (Pearson's correlation coefficient: PCC: 0.46 and 0.45 respectively) as compared to basal promoter genes (PCC: 0.24). Moreover, we observed

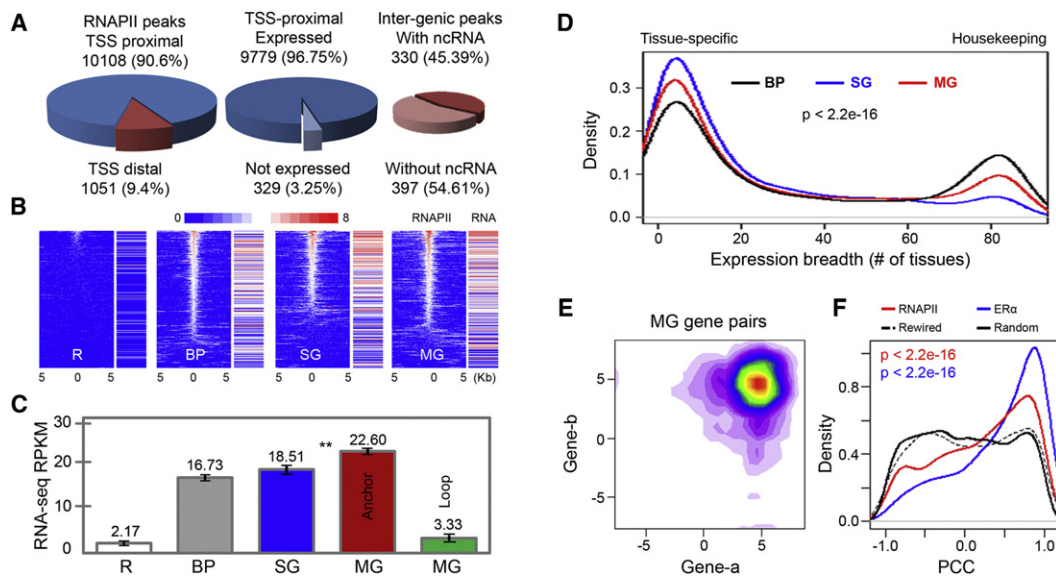
that genes linked by complex chromatin interactions, especially those in multigene complexes, had significantly higher expression levels than basal promoter genes (Figure 3C). This high expression appeared to be limited to genes interacting at the RNAPII anchor sites, as compared to genes located in the intervening chromatin loops. These data indicated that promoter-promoter interactions in multigene complexes were associated with higher transcriptional activity, which is consistent with our observations of their associated genomic features.

Next, we characterized the expression patterns of genes present in the interacting regions using microarray data derived from 84 human tissues (Su et al., 2002). We found distinct representation of tissue-specific and housekeeping genes in the three chromatin models (Figure 3D, Figures S3A and S3B). Most genes in single-gene complexes with enhancer-promoter connectivity were tissue-specific, consistent with growing evidence that the expression levels of developmental and tissue-specific genes are largely modulated through *cis*-remote regulatory elements and *trans*-protein factors (Hou et al., 2010; Schoenfelder et al., 2010), and consistent with their genomic features (less gene density, longer gene body and higher intron/exon ratio) as previously described. Conversely, genes involved in multigene complexes as well as the basal promoter genes were characterized as both tissue-specific and housekeeping categories. These observations were also supported by normalized CpG content and GC-skew at their promoter regions (Figures S3C and S3D).

As promoter-promoter interactions cluster multiple genes, they could provide an ideal topological framework for potential transcriptional coordination of both tissue-specific and housekeeping genes. This observation agrees with the evidence that "ridges," which are domains of highly transcribed genes, contain both housekeeping and tissue-specific genes (Versteeg et al., 2003). Since large numbers of genes are found in multigene complexes, we propose that promoter-promoter interactions could serve as a dominant mechanism for transcription regulation of both housekeeping and tissue-specific genes in mammalian genomes.

Next, we sought to determine whether genes with promoter-promoter interactions were more likely to be transcriptionally coordinated. RNA-Seq data showed that most of the paired genes with promoter-promoter interactions were expressed together at high levels (Figure 3E; Figure S3E). To further assess the coordinated transcription of paired genes across different conditions, we performed Pearson's correlation analysis using estrogen-induced time course of GRO-Seq data (Hah et al., 2011) that measured transcription initiation rates of estrogen responsive genes, and observed significant transcriptional correlation (Figure 3F;  $p$  value <  $2.2 \times 10^{-16}$ ). Interestingly, the correlation was even greater for ER $\alpha$ -mediated gene pairs derived from our earlier data (Fullwood et al., 2009), suggesting stronger correlation of transcription for genes involved in multigene complexes mediated by specific transcription factors. Similar correlation was also observed from other gene expression datasets (Figures S3F–S3I). As expected, housekeeping genes and genes belonging to the same GO classes showed even higher correlation than the rest (Figures S3J and S3K). Altogether, our analyses indicated that a significant proportion of gene pairs





**Figure 3. Transcriptional Activities in RNAPII-Associated Chromatin Models in MCF7 Cells**

(A) Pie charts of RNAPII binding peaks proximal (blue) or distal (red) to TSS of genes (left), RNA-Seq data for genes with RNAPII peaks near TSS (middle), and RNA-Seq enrichment around intergenic RNAPII peaks (right).

(B) Correlation of RNAPII binding in basal promoter (BP), single-gene (SG) and multigene (MG) models with gene transcription levels measured by RNA-Seq. The RNAPII enrichment heatmap shows binding intensity centered on TSS ( $\pm 5$  kb) along with corresponding gene transcription intensity.

(C) Bar plots of expression levels of genes in the three models (BP, SG, and MG). RNA-Seq mean values (RPKM) and standard error of means (SEM) from genes in the corresponding models are shown. MG complexes also contain “anchor genes” (TSS proximal to interacting anchors) and “loop genes” (distant from anchors, residing in loop regions). The remaining genes (R) not bound by RNAPII were included as a control. Double asterisks (\*\*) indicate significant differences between the mean expressions of genes from SG and MG models ( $p$  value  $< 4.02 \times 10^{-8}$ ).

(D) Expression breadth (number of tissues a gene is expressed in) of genes present in three different chromatin models.  $P$  value is calculated using the nonparametric test of Kruskal-Wallis.

(E) Contour plot of log-transformed RNA-Seq RPKM values for cotranscription of interacting genes involved in MG models in MCF7 cells.

(F) Distribution of PCC values for RNAPII- and ER $\alpha$ -bound interacting gene pairs, randomly rewired gene pairs, and randomly picked gene pairs from control regions with the same genomic span and gene density distribution as the multigene complex regions.

See also Figure S3.

involved in promoter-promoter interactions tended to be transcribed cooperatively.

### Multigene Complexes Provide Structural Framework for Cotranscription

Correlated expression of interacting genes suggests that the multigene interaction complex might provide a molecular basis for the postulated “transcription factory” (Cook, 1999). To elucidate the link between the multigene complexes revealed by ChIA-PET and transcription factories, we performed 3D DNA-FISH experiments using probes representing distinct multigene complexes in combination with RNAPII-IF staining in MCF7 nuclei (Experimental Procedures). All experiments on four genomic loci randomly chosen from multigene complexes revealed a significant association of the multigene complex loci with RNAPII foci (Figure 4A-B), adding further evidence to support our view that multigene complexes could provide a structural framework for cotranscription.

Furthermore, gene families were significantly over-represented ( $p$  value  $< 0.006$ ) in the multigene complexes (Figure S3L), such as *HIST*, *ZNF*, *KRT*, *HOXC*, etc. (Table S4). Taking the *HIST1H* family as an example, the 58 genes of this family located on chromosome 6 formed three multigene complexes, and

these three complexes converged into a higher-order super-complex, suggesting that all *HIST1H* genes were organized in a single chromatin architecture for coordinated transcription (Figure 4C). All *HIST1H* genes were actively transcribed in both MCF7 and K562 cells, and were highly coregulated across different tissues and cellular conditions (Figure 4D). Interestingly, *HFE*, a gene was not a part of the *HIST1H* family but was located in the middle of the first *HIST1H* multigene complex, was not anchored at the interaction sites and was not expressed. Similarly, the genes located in the intervening loop regions between the three *HIST1H* interacting complexes were relatively less active and much less coordinated for coregulation across different tissues and cellular conditions. This case exemplifies the model where multigene complexes organize genes with similar functions across genomic space for coordinated expression.

### Multigene Complexes Support Synergistic Transcription Regulation

To further investigate the likelihood that the multigene complex structure might provide a topological framework for transcriptional coregulation of interacting genes involved in such topology, we designed a set of perturbation experiments to

test this. After comparing the RNAPII and ER $\alpha$  ChIA-PET data from MCF7 cells, we found that the RNAPII-bound multigene complex at the *GREB1* locus partially overlaps with the ER $\alpha$ -bound chromatin loops, suggesting that this interaction complex, in part, is associated with ER $\alpha$ . Therefore, we performed siRNA experiments to knockdown the protein level of ER $\alpha$  in MCF7 cells, and monitored the alteration of chromatin interactions and gene transcription in the *GREB1* multigene complex. Several chromatin interaction loops at this locus were disrupted by siER $\alpha$  transfection as tested by 3C experiments (Figure 4E). In addition to *GREB1*, which had a strong response to estrogen induction and reduction by siER $\alpha$  knockdown (Figures S4A–S4D), we observed that the other genes in this complex such as *E2F6*, *KCNF1* and *ATP6VC12* also had various levels of response to induction by estrogen and reduction by siER $\alpha$  knockdown (Figure 4F). Interestingly, these genes did not directly interact with ER $\alpha$  at their promoter regions, but indirectly associated with ER $\alpha$  through RNAPII-bound chromatin loops. As a control, this effect was not seen in the nearby genes such as *NOL10* and *HPCAL1* that were in other RNAPII interaction complexes and also did not interact with ER $\alpha$  (Figure 4G). Similar results were observed at another interaction locus centered on the *GPR68* and *CCDC88C* genes (Figure S4E). Thus, these results indicate that a specific stimulus (estrogen) could lead to coactivation of genes organized primarily through RNAPII-bound multigene complexes, and perturbation at one gene locus (loss of ER $\alpha$  binding in this case) in a multigene complex could alter the transcriptional states of other interacting genes within the same complex. Although genes in close genomic distances with each other had been reported to be correlated in expression levels (Singer et al., 2005), our data suggests that the conjoint expression can be mediated through chromatin interactions. The functional significance of such coregulation needs further investigation.

### Epigenomic Marks Associated with Chromatin Interaction Sites

To study the association of transcription factors (TFs) with the RNAPII interactions, we examined the enrichment of 20 different TFs in K562 cells at the RNAPII interaction sites from the three chromatin models in our K562 ChIA-PET dataset (Figures 5A and 5B, Figures S5A–S5D). General TFs such as E2F4 and E2F6 (Figure 5A, Figure S5A) directly bound at TSS sites (Figure 5B for a specific example). By contrast, specific TFs such as JunD and Max preferentially bound to distal regulatory sites and marked potential enhancers (Figure S5B). Several chromatin remodeling factors and chromatin organization proteins such as INI1, BRG1, CTCF, and RAD21 associated primarily with non-TSS sites, suggesting that they may mediate long-range interactions with enhancer regions (Figure 5A, Figure S5C). This hypothesis is consistent with other observations that INI1 and BRG1, two subunits of the SWI/SNF complex, were involved in transcriptional looping (Euskirchen et al., 2011). A common observation among all the factors was that interaction sites in the multigene complexes consistently showed elevated levels of factor enrichment, suggesting that the cooperative binding of factors in gene-rich domains leads to higher transcriptional activity, or these transcriptionally active open chromatin

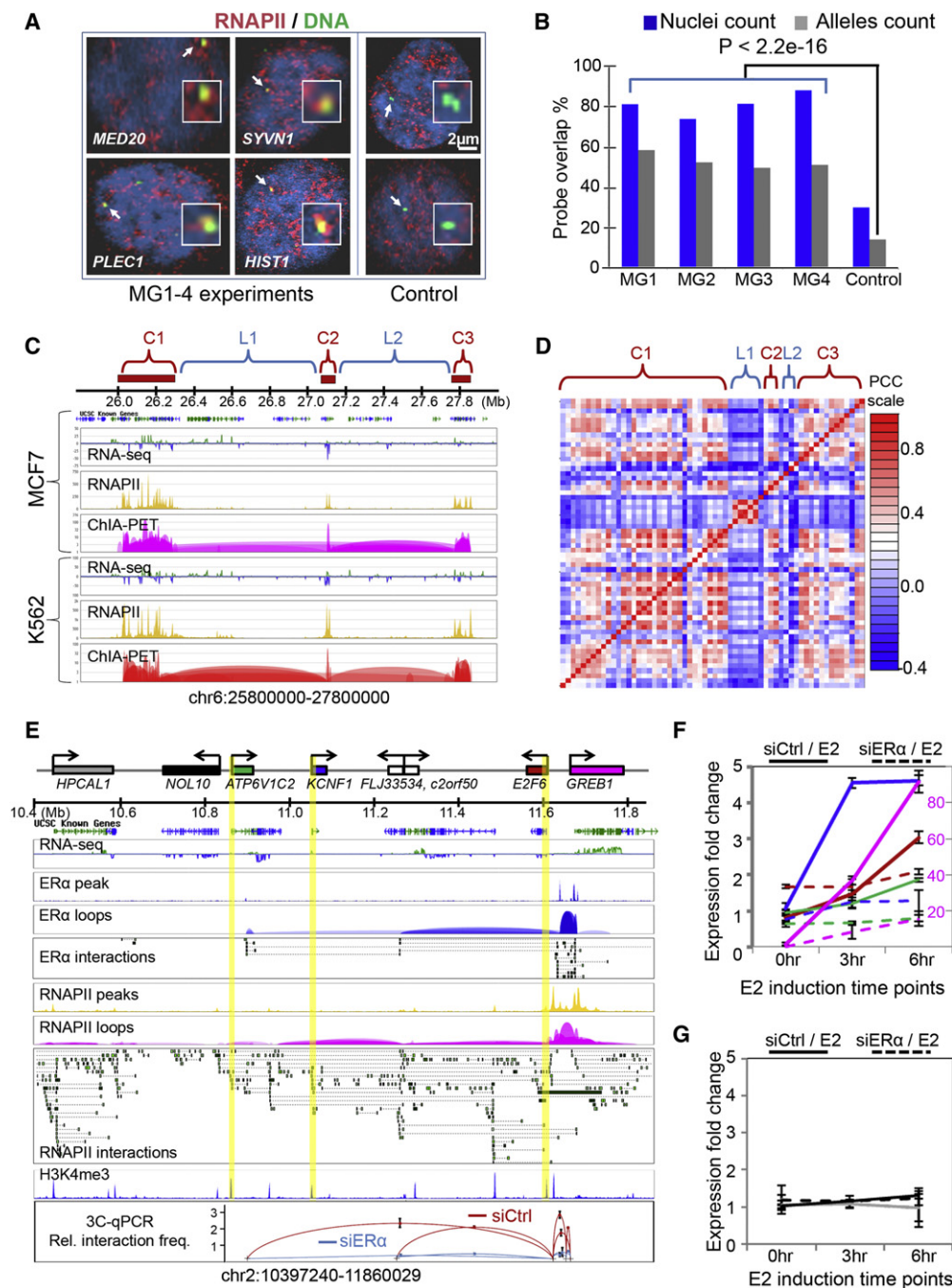
domains might converge to distinct specialized transcription factories, each enriched with general and specific TFs.

We further explored the histone modification data available from the ENCODE Consortium. Collectively, we found high enrichment of active histone modification marks coupled with a lack of repressive marks in RNAPII interaction sites, confirming that the RNAPII interaction sites mapped by our ChIA-PET data were located in promoter and distal regulatory regions engaged and/or poised for high transcription levels (Figure S5D). Interestingly, the enrichment of active marks was highest in the multigene complexes, indicating that these might constitute transcriptional hubs. Our observations matched previous findings that the enrichment of active histone modifications positively correlated with RNAPII occupancy (Barski et al., 2007).

We observed similar histone modification profiles in MCF7 cells (Figure 5C) using data that we generated previously (Joseph et al., 2010). In particular, we applied the log ratio of H3K4me3/H3K4me1 signal as a quantitative measurement of the likelihood that a genomic locus can act as a promoter or enhancer. Most noninteracting RNAPII sites proximal to TSS in basal promoter model showed high log ratios (Figure 5D, plot 1; median = 2.4; > 90% of the binding regions have log ratios > 0), whereas most of the RNAPII interaction sites distal to TSS in the single-gene complex model and the multigene complex model (conventional enhancer sites) showed low H3K4me3/me1 log ratios (Figure 5D, plot 4 and 6; median < –0.72), confirming that this log ratio could reflect relative capacities of promoters and enhancers. Surprisingly, examination of RNAPII interaction sites proximal to known TSSs in the multigene complexes (Figure 5D plot 5) revealed two peaks in the histogram of the log ratios, suggesting a mixture of enhancer and promoter elements in the promoter regions. Detailed profiles of H3K4me3 and H3K4me1 marks around the center ( $\pm 5$  kb) of those RNAPII interaction sites showed distinct characteristics of promoter-like, enhancer-like sub-groups (Figure 5D, heatmap). Moreover, enhancer-like RNAPII interaction sites, on average, showed lower transcriptional activity than the promoter-like RNAPII sites (Figure S5J). Thus, a large portion of interacting promoters may also have potential enhancer functions. We observed the same inverse correlation of H3K4me3/me1 log ratio at the TSS proximal and TSS distal RNAPII sites for K562 (Figure 5A), indicating that this observation is a general phenomenon applicable to all cell types.

### Interacting Promoters Possess Combinatorial Regulatory Functions

To examine potential enhancer activity of promoters, we performed luciferase reporter gene assays, a commonly used method for promoter and enhancer characterization (Pan et al., 2008). In these assays, approximately 500 bp fragments of the expected promoter regions were cloned upstream of a luciferase reporter gene construct either in a proximal position as the driving promoter or in a distal position as a presumed enhancer, and the constructs were transfected into MCF7 cells (Experimental Procedures, Figures S5E–S5I). As shown in Figure 5E, the two interacting loci *INTS1* and *MAFK* were 26 kb apart, and our RNA-Seq data suggested that both genes were active



**Figure 4. Transcriptional Coordination in Multigene Chromatin Complexes**

(A) Colocalization of multigene loci with RNAPII foci. Shown are the nuclear images of RNAPII IF-staining with four randomly-selected multigene loci (MG1-4) and 2 control loci. Representative gene loci are *MED20*, *SYVN1*, *HIST1*, and *PLEC1*.

(B) Quantitative analysis of nuclei ( $n = 476$ ) and alleles showing overlap of MG loci and RNAPII foci. Percentage overlaps from MG loci and those from control loci are significantly different.

(C) Super multigene complex of the histone gene family. Three distant clusters (C1, C2, C3) of *HIST1H* genes converge together in a super-MG complex. Shown are RNA-Seq, RNAPII and ChIA-PET tracks in MCF7 and K562 cells.

(D) Cotranscription of *HIST1H* genes in the super-MG complex in (C). Correlation matrix derived from publicly available microarray data of 4,787 samples (Supplemental Information). The rows and columns correspond to genes in each complex and the intervening regions.

(E) RNAPII-bound multigene complex at the *GREB1* locus. Shown are the ERα- and RNAPII-bound chromatin interactions. Highlighted promoters are anchored by RNAPII, but not by ERα. The bottom panel shows relative interaction frequency by 3C-qPCR data for the perturbation experiments using siERα knockdown and estrogen induction.

in MCF7 cells. However, the normalized log ratio of H3K4me3/me1 was 0.36 for the *INTS1* promoter and 1.13 for the *MAFK* promoter, suggesting that the *INTS1* promoter may have enhancer properties. To test this, we cloned the *INTS1* promoter fragment in both orientations upstream of the *MAFK* promoter flanking the luciferase gene. The luciferase reporter gene assay showed at least 7-fold enhancement of luciferase expression from the *MAFK* promoter activity by the *INTS1* promoter fragment, indicating that a bona fide promoter can act as an enhancer to augment the activities of other promoters.

In another example (Figure 5F), the promoter of *CALM1* interacts with an enhancer element 15 kb upstream and connects to the promoter of *C14orf102* further upstream in 65 kb. Both RNA-Seq data and the H3K4me3/me1 log ratio indicated that the *CALM1* promoter was strong, whereas the *C14orf102* promoter was weak and enhancer-like. The luciferase reporter gene assay showed marginal enhancement to the *CALM1* promoter reporter gene activity by the native *CALM1* enhancer and the *C14orf102* promoter individually. However, the combined *CALM1* enhancer and the *C14orf102* promoter together led to a significant ~3-fold enhancement of reporter expression from the *CALM1* promoter. This result further validates the enhancer function by interacting promoters and elucidates a possibility of combinatorial effect among interacting elements in multigene interaction complexes for transcription regulation.

Next, we asked whether promoters with enhancer activity act specifically on their target genes. We swapped the promoter elements in the two examples of *INTS1*-to-*MAFK* and *C14orf102*-to-*CALM1* for additional reporter genes assays (Figure 5G). Intriguingly, when placed upstream to the *CALM1* promoter, the *INTS1* promoter showed remarkable enhancement of *CALM1* promoter activity. Similarly, the combined construct of *C14orf102* promoter and *CALM1* enhancer also increased *MAFK* promoter activity significantly. Meanwhile, a TATA box deleted promoter and other control promoters (either active or inactive), taken from the nearby genes that are not involved in a promoter-promoter relationship, did not show cooperative enhancement to *MAFK* and *CALM1* promoter activities (Figures S5H and S5I). Thus, these results suggest a common property for promoters with enhancer capacity that could influence other promoters.

In addition, we also tested the combination of inserting the enhancer-like promoter fragment in the position proximal to luciferase gene and the strong promoter in the distal position in the reporter gene construct. Of the 20 such luciferase experiments, we observed that the weaker promoters conveyed significant enhancer function to their stronger interacting partners in luciferase activity rather than the reverse (Figure S5K). In the case of interacting pair *INTS1* (enhancer-like promoter) and *MAFK* (strong promoter), the strong promoter *MAFK* did not demonstrate significant enhancer activity (Figure S5L). Thus,

at promoter sites, there is an inverse relationship between enhancer and promoter functions.

### Cell-Line Specificity of Long-Range Chromatin Interactions

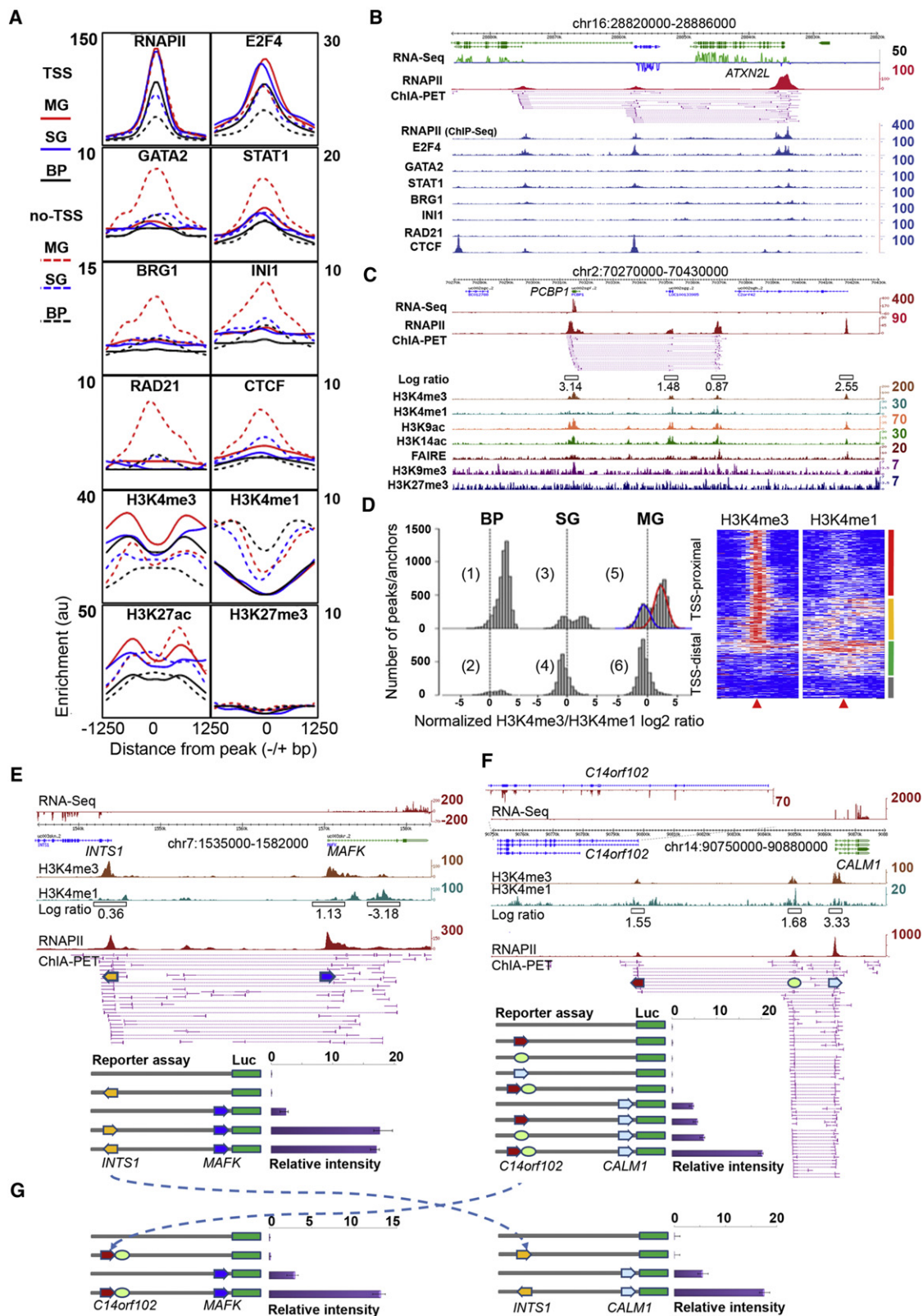
To elucidate the cell-line specificity of chromatin interactions, we saturated the coverage of chromatin interactions through deep sequencing of more MCF7 and K562 ChIA-PET replicates (Experimental Procedures). The saturated libraries are highly reproducible for interactions, and thus highly reliable for intercell line comparative analysis. These libraries exhibit the same pattern of genomic descriptors as the pilot libraries (Figures S2B and S2C). With comprehensive ChIA-PET and RNA-Seq datasets, we performed comparative analysis between the two cell lines and identified cell-line specific genes and chromatin interactions (Figure 6A). Most of the genes specifically expressed in their respective cells also showed cell-specific interactions (Figure 6B), implying that cell-specific chromatin interactions provide the structural basis for cell-specific transcription. Gene Ontology (GO) analysis revealed significant enrichment of erythroid related GO terms such as response to stimulus and blood circulation for genes with specific expression and chromatin interactions in K562 cells, whereas GO terms such as ectoderm development and related biological process were enriched in MCF7 cells (Figure 6C, Figure S6A). As expected, the genes common in both cell lines showed enrichment of house-keeping functions like metabolism, cell-cycle and signal transduction (Figure S6B).

Among the chromatin interactions specific to K562 cells, we captured many previously characterized interactions including the  $\alpha$ - and  $\beta$ -globin loci (Bau et al., 2011; Hou et al., 2010). Figure 6D shows extensive interactions identified by ChIA-PET data between the  $\alpha$ -globin gene locus and the DNase hypersensitive (DHS) sites present in the gene body of the *C16orf35* gene. Additionally, we found that the  $\alpha$ -globin locus in K562 extended its interactions to the neighboring domains, which were constitutively active in both K562 and MCF7 cells, whereas the interactions to  $\alpha$ -globin genes are K562-specific, suggesting a complex chromatin architecture for spatiotemporal regulation of both constitutive and cell-specific transcription. Similarly, the  $\beta$ -globin gene locus also displayed previously known K562-specific interactions with the nearby locus control region (Figure S6C).

*GREB1* is a well characterized MCF7-specific gene. As expected, we found abundant chromatin interactions associated with RNAPII at this locus in MCF7, but not in K562 cells (Figure 6E). In addition to recapitulating the previously identified ER $\alpha$ -associated interactions (Fullwood et al., 2009), RNAPII interaction data showed an additional interaction site on the far most upstream (left in Figure 6E) side of this complex. A strong H3K4me1 mark on this site suggested that this is potentially an enhancer site for a transcription factor other than ER $\alpha$ .

(F and G) Time course RT-qPCR following estrogen (E2) induction after siControl (solid) and siER $\alpha$  (dashed) transfections of MCF7 cells. Colors of the curves correspond to genes shown in (E). A secondary axis (red, right side) is used for *GREB1* expression to accommodate its high expression level. Expression data of genes involved in the *GREB1* multigene complex are in (F), and the data for genes outside of the complex are in (G). RT-qPCR mean values and standard deviations (SD) from two independent experiments are shown. See also Figure S4 and Table S2.





Intriguingly, a significant RNA-Seq peak was also identified at this site, indicating a possible enhancer RNA transcript, a new class of noncoding RNA species (Kim et al., 2010).

### Long-Range Enhancer-Promoter Interactions and Disease-Associated Noncoding Elements

Our data showed that the enhancer-promoter interactions were significantly enriched over other types of interactions for cell-specific genes (Figure 7A) when compared to genes commonly expressed in both cell lines. This finding supported the general view that distant-acting enhancers tend to be specifically involved in tissue-specific genes, and was consistent with our analysis in Figure 3D. Although potential enhancer sites can be identified using high throughput approaches (Heintzman et al., 2009), it is still challenging to connect enhancers to their target genes that are hundreds of kilobases away. Moreover, many remote enhancers could be embedded in intronic regions of other distantly located genes (Visel et al., 2009), making it notoriously difficult to relate enhancers to their specific target genes. In this study, we identified tens of thousands enhancer-promoter interactions (Table S1C) including approximately 1000 ultra-long-distance (500 kb to megabases) events. We observed that  $\geq 40\%$  of enhancers do not interact with their nearest promoters and instead jump over to their target promoters, bypassing several intervening genes (Figure 7B, Figure S7).

An interesting example is the *SHH* gene that was expressed in MCF7 but not in K562 cells (Figure 7C). *SHH* is important in development and related to certain cancers (Lettice et al., 2002). Transcription of *SHH* is controlled by its enhancer which is located 1 Mb away and embedded in the intronic region of *LMBR1*; point mutation in this enhancer site is known to cause *preaxial polydactyly*, a common congenital limb malformation in mammals (Lettice et al., 2002). We found abundant interaction data between the *SHH* promoter and the previously characterized *SHH* enhancer site in the *LMBR1* intronic region in MCF7 cells, but no interaction data in K562 cells (Figure 7C), which correlated well with their *SHH* transcription status. This is consistent with earlier observations (Amano et al., 2009).

In another interesting example, we identified two major interaction sites located  $\sim 600$  kb and  $\sim 1$  Mb downstream from the *IRS1* gene promoter. *IRS1* is known to participate in type-2 diabetes (T2D) mellitus, and is found specifically expressed in MCF7 cells (Figure 7D). A recent GWAS study uncovered a cluster of SNPs that is genetically associated with high risk to insulin resistance, T2D, and coronary artery heart disease (Kilpelainen et al., 2011). This high risk locus is found located in one of the *IRS1* enhancer sites (Figure 7D). Thus, our data provides experimental evidence to suggest that this disease-risk locus could be physically connected with the *IRS1* promoter, potentially serving as a critical long-range enhancer to regulate the expression of *IRS1*, in a similar manner as the *SHH* locus. Other examples of long-range and cell-specific enhancer-promoter interactions in MCF7 and K562 are shown in Figure S7. Taken together, these results suggest that ChIA-PET interaction data may better inform the association of a SNP with a gene involved in a disease process by providing evidence for direct physical interactions.

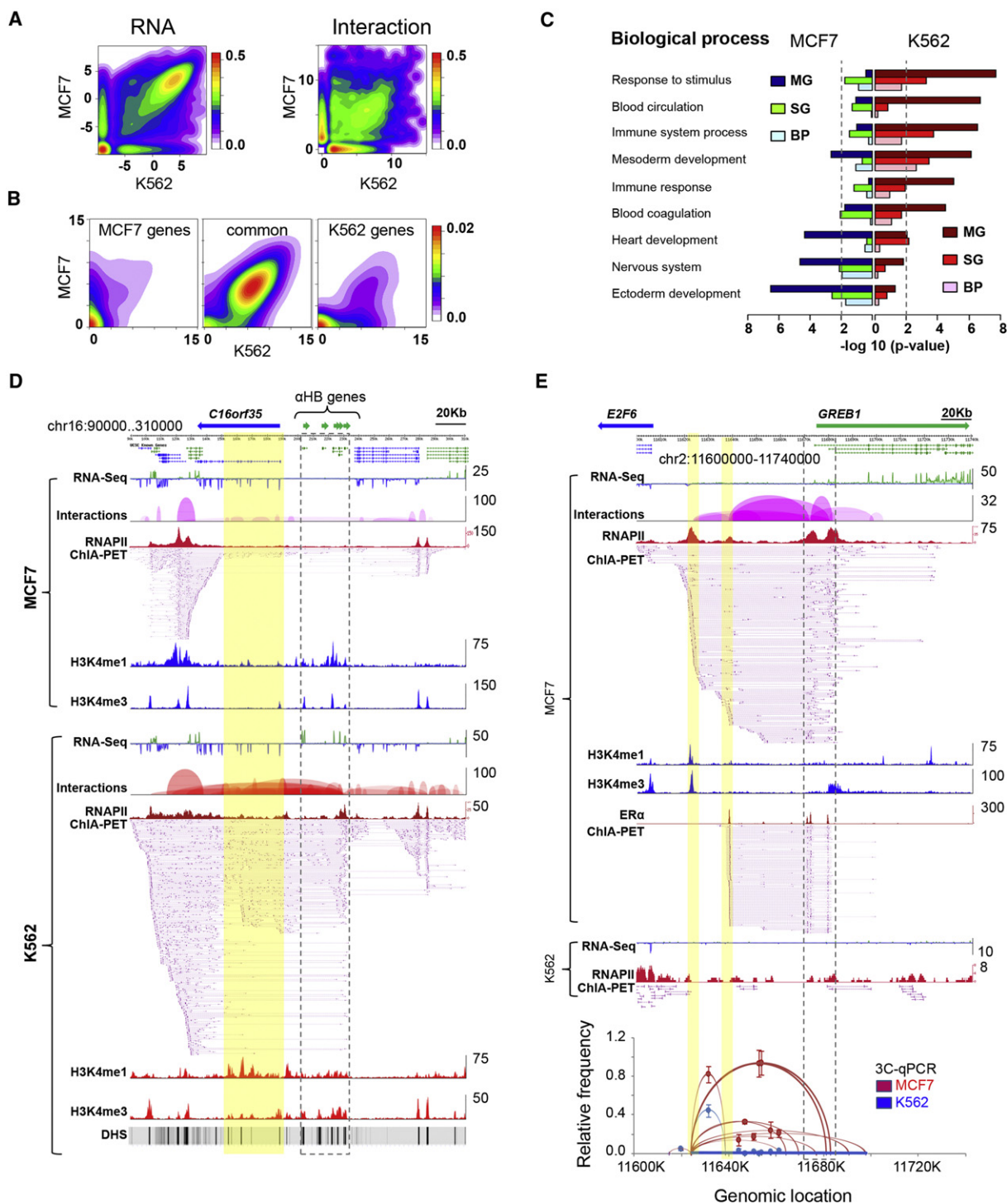
### DISCUSSION

Through genome-wide mapping, we comprehensively analyzed RNAPII-associated long-range chromatin interactions. Our most interesting finding was the extensive promoter-promoter interactions among proximal and distant genes from 5 human cell-lines, which indicated that this mechanism is common in cells. Our work with reporter gene and siRNA knockdown assays provided experimental evidence that many promoters in the multigene complexes can cooperatively regulate the activity of other promoters with which they interact. Our observations thus blurred the conventional definition of promoter and regulatory elements for transcription. With such promoter-promoter interactions, we speculate that genetic error at one particular promoter might also propagate to other promoters and hence could lead to pleiotropic consequences depending on the interaction network within a cell type. Intriguingly, the multigene complexes illustrated in this study are, in principle, akin to the

### Figure 5. Epigenomic Profiles of Chromatin Interactions and Combinatorial Regulation of Interacting Promoters

- (A) Enrichment profiles of TFs and histone modifications centered on RNAPII peaks ( $\pm 1250$  bp) of interacting loci of the three models in K562 cells. Solid lines represent "TSS" proximal regions and dotted lines depict "non-TSS" regions. y axis: sliding median for ChIP-Seq enrichment in the region.
- (B) Examples of TF enrichment at RNAPII interacting loci in K562 cells.
- (C) Histone modification marks and open chromatin mark (FAIRE) associated with chromatin interaction sites in MCF7 cells. The width of the open boxes in the log ratio track reflects the region where the H3K4me3 and H3K4me1 data were used for the log ratio calculation.
- (D) Histograms of normalized H3K4me3/me1 log ratio at RNAPII sites proximal to TSS (TSS) and distal to TSS (non-TSS) of genes in the three chromatin models in MCF7 cells. Two peaks are seen in plot #5 (blue curve for enhancer-like, and the red for promoter-like). The heatmap shows detailed H3K4me3 and H3K4me1 enrichments around RNAPII interaction sites ( $\pm 5$  kb) proximal to TSS. Four distinct clusters of RNAPII sites are promoter-like (red), enhancer-like (green), heterogeneous (yellow) and weak signals (gray).
- (E–G) Reporter gene assay of interacting promoters in MCF7 cells. RNA-Seq, H3K4me3, H3K4me1, H3K4me3/me1 ratio, and RNAPII ChIA-PET data tracks are shown. Numbers on the right side for each track indicate the highest peak intensity. The mean values and standard deviations (SD) of the luciferase activities from at least three independent experiments are shown.
- (E) Promoter-promoter interaction at the *INTS1-MAFK* locus. The arrow boxes indicate the aligned promoter regions which were cloned in reporter gene constructs for luciferase assay.
- (F) Promoter-enhancer-promoter interactions at the *C14orf102-CALM1* locus. RNA-Seq data showed that *CALM1* was highly expressed, whereas *C14orf102* only marginally transcribed (enlarged RNA-Seq track of the *C14orf102* locus).
- (G) Swap assay of DNA fragments from different multigene complexes. The dotted arrow lines show the swap of elements cloned in the distal positions in the reporter gene constructs for luciferase assay.

See also Figure S5 and Table S2.



**Figure 6. Cell-Specific Chromatin Interactions**

(A) Contour plots of RNA-Seq data (log RPKM, left) and chromatin interactions (log PET counts, right) in MCF7 and K562 cells, showing common and cell-specific gene expression and chromatin interactions.

(B) Contour plots of interaction data (log PET counts) for genes specifically and commonly expressed in MCF7 and K562 cells.

(C) Enrichment of cell-specific GO terms in genes and chromatin interactions specific in MCF7 and K562 cells. The p value of 0.01 is marked as dotted line.

(D) An example of K562-specific chromatin interactions.  $\alpha$ -globin genes (in dotted line box) interact with distantly located ( $\sim 20$  kb) DHS sites (highlighted in yellow) which are known to interact with  $\alpha$ -globin genes. In sharp contrast, the  $\alpha$ -globin genes in MCF7 cells are not expressed and have no interactions with the DHS sites.



bacterial operon as a mechanism for coordinated transcriptional regulation of related genes, suggesting the possibility of a chromatin-based operon mechanism (chro-operon or chroperon) for spatiotemporal regulation of gene transcription in eukaryotic nuclei. However, the “chroperon” expression is not dependent on the linear arrangement of the genes, but is highly dynamic and can adopt a multitude of cassette configurations because of the combinatorics permitted by the looping interactions. Alternatively, these interactions could reflect stochastic movement of proximal and distant active genes to localized transcription factories.

An important question is how these multigene complexes are organized. A likely model is that a suite of protein factors for modulating gene expression in a functional regulatory cassette may result in optimal stoichiometry when aggregated in 3D space. This clustering also draws the regulated genes into a common spatial domain, similar to how the nucleolus is organized. The interacting regions can be established and/or maintained by potential chromatin bridging proteins such as cohesins (Merkenschlager, 2010) and CTCF (Handoko et al., 2011), and this process might be facilitated by chromatin remodeling proteins (Euskirchen et al., 2011), all of which are enriched at the interacting sites defined by RNAPII ChIA-PET data.

Long-range chromatin interactions including enhancer-promoter interactions are increasingly being recognized as an important mechanism to regulate many important genes. However, methods to identify such long-range relationships have been technically challenging. High-throughput approaches such as ChIP-Seq and DNase-Seq are efficient in identifying potential regulatory sites, but lack the ability to interrogate the connectivity between the prospective enhancers and their target gene promoters. In this study using RNAPII as the protein target for ChIA-PET analysis, we identified a comprehensive repertoire of distant regulatory elements directly interacting with gene promoters. Many of them act through ultra-long-range chromatin interactions. Such distal enhancer-promoter relationships are particularly difficult to be identified by other approaches. As demonstrated in the cases of *SHH* and *IRS1*, long range interactions derived from ChIA-PET data could provide the connectivity of GWAS-identified high-risk loci to their target genes, and thus offer possible mechanistic explanations to the function of disease-associated noncoding elements. Further investigation of spatial architectures revealed in this study will enhance our understanding of transcription regulation in normal and diseased conditions of human cells.

## EXPERIMENTAL PROCEDURES

### Cell Culture

Five cell lines, namely MCF7 (ATCC# HTB-22), K562 (ATCC# CCL-243), HCT116 (ATCC# CCL-247), HeLa (ATCC# CCL-2.2), and NB4, were grown under standard culture conditions and harvested at log phase.

### ChIA-PET

Harvested cells were cross-linked using 1% formaldehyde followed by neutralization with 0.2M glycine. Chromatin was isolated and subjected to the ChIA-PET procedure (Fullwood et al., 2009). The ChIA-PET sequence reads were analyzed using ChIA-PET Tool (Li et al., 2010). The data are available from NCBI/GEO (ID GSE33664). Control and reproducibility analyses are described in Figure S8.

### RNA-Seq Data

MCF7 mRNA was isolated following the protocol described in Ruan et al. (Ruan et al., 2007) for strand-specific RNA-Seq analysis by SOLiD sequencing platform. The rest of the RNA-Seq datasets for other cell-lines were retrieved from the ENCODE data repository site (<http://genome.ucsc.edu/ENCODE/>).

### ChIP-Seq Data

The ChIP-Seq data were retrieved from (Joseph et al., 2010), (Raha et al., 2010) and the ENCODE data repository site (<http://genome.ucsc.edu/ENCODE/>).

### RNAPII IF Stain and DNA-FISH

MCF7 cells were fixed using 4% formaldehyde followed by permeabilization with 0.04% Triton-X. After blocking with donkey serum, cells were incubated with primary antibody (8WG16) overnight followed by Cy3 conjugated secondary antibody for 1 hr. IF-stained cells were post-fixed and subjected to dehydration by 70, 80, 100% ethanol series, rehydration with 2× SSC and denaturation in 2× SSC/50% formamide at 80°C for 40 min. Biotin-16-dUTP and digoxigenin-11-dUTP labeled DNA probes were hybridized to cells at 37°C overnight in a humid chamber. Slides were washed, stained with DAPI, mounted and visualized by a Carl Zeiss LSM confocal microscope.

### Quantitative Chromosome Conformation Capture Analysis

Targeted 3C products were analyzed by qPCR. The 3C-qPCR protocol was adapted and modified from the previous publication (Fullwood et al., 2009).

### Luciferase Reporter Gene Assay

Dual luciferase assays were performed as described (Pan et al., 2008). Testing fragments were cloned into pGL4.10-basic vector. Constructs were transfected into MCF7 cells, and luciferase activities were measured following standard protocols.

### Statistical Analysis

All the statistical tests were executed using the R statistical package (<http://www.r-project.org/>).

More details are available in Extended Experimental Procedures.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five tables, and eight figures and can be found with this article online at doi:10.1016/j.cell.2011.12.014.

## ACKNOWLEDGMENTS

We acknowledge the Genome Technology and Biology Group at the Genome Institute of Singapore for technical support. This work was supported by Singapore A\*STAR and an NIH grant (HG004456) to Y.R.

Received: July 18, 2011

Revised: October 21, 2011

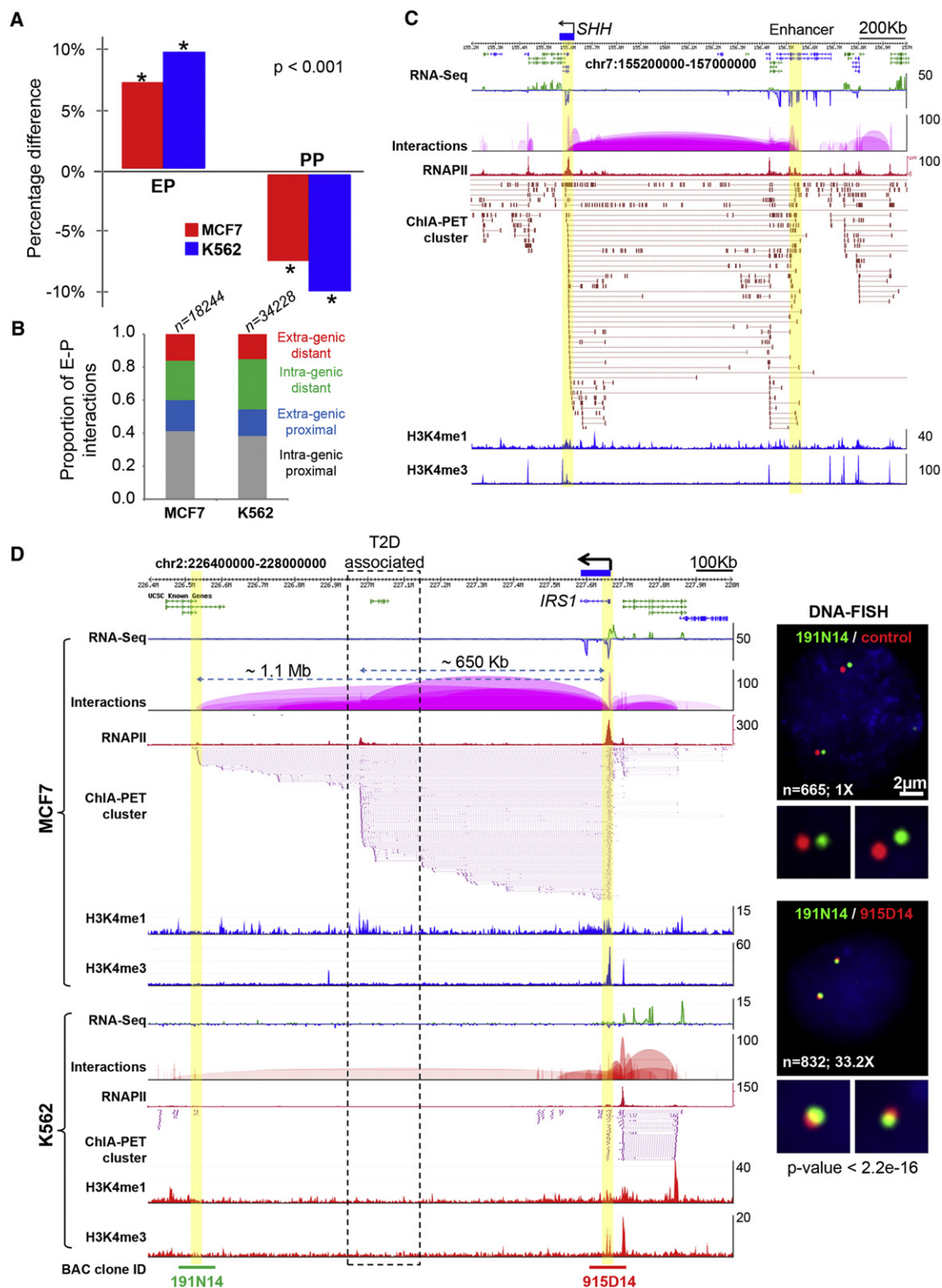
Accepted: December 12, 2011

Published: January 19, 2012

(E) An example of MCF7-specific chromatin interactions around the *GREB1* locus. The far left highlighted yellow is a RNAPII interaction site that is not overlapped by ER $\alpha$ -bound interactions in this region. It is also the bait site for independent 3C validation of interactions in this region. Tracks included in (D) and (E) are RNA-Seq data, interaction loop view, RNAPII ChIA-PET peaks and interaction PETs, ChIP-Seq density profile of H3K4me1 and H3K4me3, and the ER $\alpha$ -ChIA-PET in (E). The numbers on the right of each track are the highest density value. 3C-qPCR mean values and standard error of means (SEM) from three independent experiments are shown.

See also Figure S6 and Figure S7.





**Figure 7. Long-Range Enhancers and Disease-Associated Noncoding Elements**

(A) Percentage difference of enhancer-promoter (EP) and promoter-promoter (PP) interactions in cell-specific versus common genes from MCF7 and K562 cells. The representation of EP interactions is significantly increased in cell-specific interactions, while the representation of PP interactions is decreased, when compared to interactions that are common to both cell lines.

## REFERENCES

- Amano, T., Sagai, T., Tanabe, H., Mizushima, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal dynamics at the *Shh* locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* 16, 47–57.
- 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.
- Bau, D., Sanyal, A., Lajoie, B.R., Capriotti, E., Byron, M., Lawrence, J.B., Dekker, J., and Marti-Renom, M.A. (2011). The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.* 18, 107–114.
- Cook, P.R. (1999). The organization of replication and transcription. *Science* 284, 1790–1795.
- Cope, N.F., Fraser, P., and Eskiwi, C.H. (2010). The yin and yang of chromatin spatial organization. *Genome Biol.* 11, 204.
- Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* 2, 292–301.
- Eisenberg, E., and Levanon, E.Y. (2003). Human housekeeping genes are compact. *Trends Genet.* 19, 362–365.
- Euskirchen, G.M., Auerbach, R.K., Davidov, E., Gianoulis, T.A., Zhong, G., Rozowsky, J., Bhardwaj, N., Gerstein, M.B., and Snyder, M. (2011). Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches. *PLoS Genet.* 7, e1002008.
- 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.
- 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.
- Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W., Ye, C., Ping, J.L., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* 43, 630–638.
- Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112.
- Hou, C., Dale, R., and Dean, A. (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. USA* 107, 3651–3656.
- Jacob, F., Perrin, D., Sanchez, C., and Monod, J. (1960). C. R. Hebd. Seances Acad. Sci. 250, 1727–1729.
- 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  $\alpha$ . *Mol. Syst. Biol.* 6, 456.
- Kilpelainen, T.O., Zillikens, M.C., Stancakova, A., Finucane, F.M., Ried, J.S., Langenberg, C., Zhang, W., Beckmann, J.S., Luan, J., Vandenput, L., et al. (2011). Genetic variation near *IRS1* associates with reduced adiposity and an impaired metabolic profile. *Nat. Genet.* 43, 753–760.
- Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187.
- Lettice, L.A., Horikoshi, T., Heaney, S.J., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N., et al. (2002). Disruption of a long-range cis-acting regulator for *Shh* causes preaxial polydactyly. *Proc. Natl. Acad. Sci. USA* 99, 7548–7553.
- 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.
- Merkenschlager, M. (2010). Cohesin: a global player in chromosome biology with local ties to gene regulation. *Curr. Opin. Genet. Dev.* 20, 555–561.
- 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.
- Pauli, D., Tonka, C.H., and Ayme-Southgate, A. (1988). An unusual split *Drosophila* heat shock gene expressed during embryogenesis, pupation and in testis. *J. Mol. Biol.* 200, 47–53.
- 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.
- Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., et al. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* 42, 53–61.
- Singer, G.A., Lloyd, A.T., Huminiecki, L.B., and Wolfe, K.H. (2005). Clusters of co-expressed genes in Mamm. Genomes are conserved by natural selection. *Mol. Biol. Evol.* 22, 767–775.
- Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinosa, L.M., Moqrich, A., et al. (2002). Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. USA* 99, 4465–4470.

(B) Proportional distribution of 4 classes of enhancers observed in two cell lines based on locations in relation to gene coding regions. “Intragenic proximal” enhancers locate inside of gene-body (mostly introns) and interact with the nearby promoters. “Extragenic proximal” enhancers locate outside of gene body and interact with the nearby promoters. “Intragenic distal” enhancers locate inside of gene body (mostly introns), bypass nearby genes and interact with faraway gene promoters in long-distance. “Extragenic distal” enhancers locate outside of gene body, bypass nearby genes and interact with faraway gene promoters in long-distance.

(C) Long-range interactions between *SHH* (highlighted in yellow, left) and its enhancer located about 1 Mb away in an intron of *LMBR1* (highlighted yellow, right). The *SHH* expression is specifically seen in MCF7 cells.

(D) Long-range interactions between *IRS1* promoter and two enhancers as well as strong *IRS1* expression are seen in MCF7, but not in K562 cells. The dotted line box indicates the enhancer region that contains SNPs associated with insulin resistance, type-2 diabetes (T2D) and coronary artery heart disease identified by a GWAS study. The interactions of enhancer located 1.1 Mb away to *IRS1* promoter (highlighted in yellow) is validated by DNA-FISH (right). The BAC clones and genomic segments used for DNA-FISH are indicated at the bottom.

Tracks included in (C) and (D) are RNA-Seq density profile, interaction loop view, RNAIIP peaks, ChIA-PET interaction PETs, ChIP-Seq density profile of H3K4me1 and H3K4me3 marks.

See also Figure S7 and Table S5.

- Taylor, J. (2005). Clues to function in gene deserts. *Trends Biotechnol.* 23, 269–271.
- van Steensel, B., and Dekker, J. (2010). Genomics tools for unraveling chromosome architecture. *Nat. Biotechnol.* 28, 1089–1095.
- Versteeg, R., van Schaik, B.D., van Batenburg, M.F., Roos, M., Monajemi, R., Caron, H., Bussemaker, H.J., and van Kampen, A.H. (2003). The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res.* 13, 1998–2004.
- Visel, A., Rubin, E.M., and Pennacchio, L.A. (2009). Genomic views of distant-acting enhancers. *Nature* 461, 199–205.
- Zorio, D.A., Cheng, N.N., Blumenthal, T., and Spieth, J. (1994). Operons as a common form of chromosomal organization in *C. elegans*. *Nature* 372, 270–272.