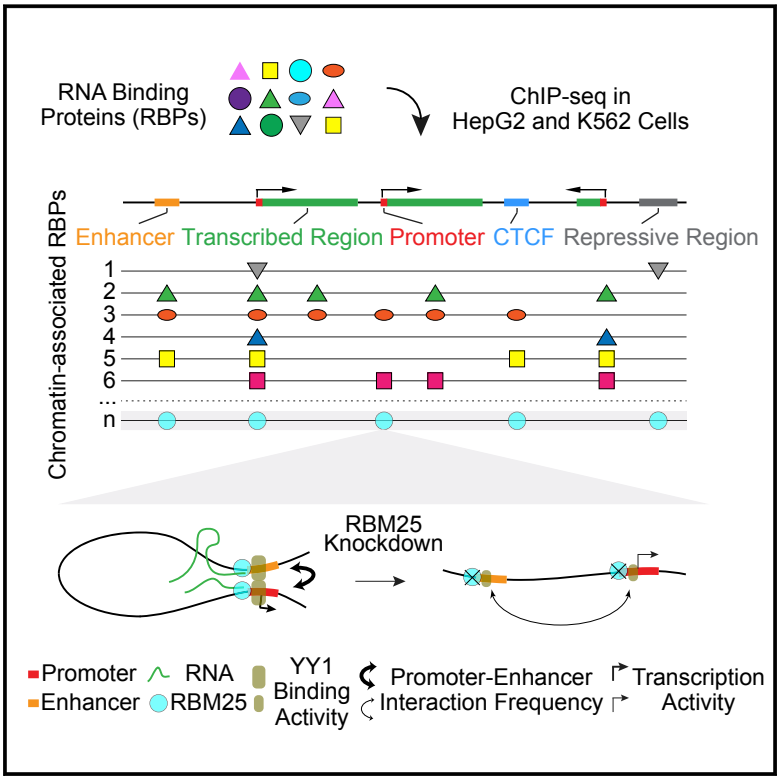
Article

Cell

Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription

Graphical Abstract



Authors

Rui Xiao, Jia-Yu Chen, Zhengyu Liang, ..., Michael Q. Zhang, Yu Zhou, Xiang-Dong Fu

Correspondence

xiaorui9@whu.edu.cn (R.X.), xdfu@ucsd.edu (X.-D.F.)

In Brief

Nuclear RNA-binding proteins are pervasive at gene promoters, with many directly participating in transcription through functional interaction with specific transcription factors.

Highlights

- RNA-binding proteins are prevalently localized on active chromatin regions
- Active gene promoters are major hotspots for interaction with RNA-binding proteins
- Multiple RNA-binding proteins are directly involved in transcription control
- RBM25 mediates YY1 function in chromatin binding, DNA looping, and transcription

Xiao et al., 2019, Cell *178*, 107–121
June 27, 2019
https://doi.org/10.1016/j.cell.2019.06.001

CellPress

Cell

Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription

Rui Xiao,[1,2,11,*] Jia-Yu Chen,[1,11] Zhengyu Liang,[1,3,11] Daji Luo,[1,4] Geng Chen,[5] Zhi John Lu,[3] Yang Chen,[3] Bing Zhou,[1] Hairi Li,[1] Xian Du,[5] Yang Yang,[5] Mingkui San,[2] Xintao Wei,[6] Wen Liu,[7] Eric Lécuyer,[8] Brenton R. Graveley,[6] Gene W. Yeo,[1] Christopher B. Burge,[9] Michael Q. Zhang,[3,10] Yu Zhou,[5] and Xiang-Dong Fu[1,12,*]

[1]Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, University of California, San Diego, La Jolla, CA 92093, USA
[2]Medical Research Institute, Wuhan University, Wuhan, Hubei 430071, China
[3]MOE Key Laboratory of Bioinformatics, Tsinghua University, Beijing 100084, China
[4]School of Basic Medical Sciences, Wuhan University, Wuhan, Hubei 430071, China
[5]College of Life Sciences and Institute for Advanced Studies, Wuhan University, Wuhan, Hubei 430072, China
[6]Department of Genetics and Genome Sciences, Institute for Systems Genomics, UConn Health Science Center, Farmington, CT 06030, USA
[7]School of Pharmaceutical Sciences, Xiamen University, Xiamen, Fujian 361102, China
[8]Institut de Recherches Cliniques de Montréal, Département de Biochimie and Médecine Moléculaire, Université de Montréal, Montréal, QC H2W 1R7, Canada
[9]Program in Computational and Systems Biology, Department of Biology, MIT, Cambridge, MA 02139, USA
[10]Department of Biological Sciences, Center for Systems Biology, University of Texas, Dallas, TX 75080, USA
[11]These authors contributed equally
[12]Lead contact
*Correspondence: xiaorui9@whu.edu.cn (R.X.), xdfu@ucsd.edu (X.-D.F.)
https://doi.org/10.1016/j.cell.2019.06.001

SUMMARY

Increasing evidence suggests that transcriptional control and chromatin activities at large involve regulatory RNAs, which likely enlist specific RNA-binding proteins (RBPs). Although multiple RBPs have been implicated in transcription control, it has remained unclear how extensively RBPs directly act on chromatin. We embarked on a large-scale RBP ChIP-seq analysis, revealing widespread RBP presence in active chromatin regions in the human genome. Like transcription factors (TFs), RBPs also show strong preference for hotspots in the genome, particularly gene promoters, where their association is frequently linked to transcriptional output. Unsupervised clustering reveals extensive co-association between TFs and RBPs, as exemplified by YY1, a known RNA-dependent TF, and RBM25, an RBP involved in splicing regulation. Remarkably, RBM25 depletion attenuates all YY1-dependent activities, including chromatin binding, DNA looping, and transcription. We propose that various RBPs may enhance network interaction through harnessing regulatory RNAs to control transcription.

INTRODUCTION

RNA-binding proteins (RBPs) have been studied on an individual basis for their functions in RNA metabolism, but recent global surveys of proteins that are UV crosslinkable to RNA reveal a large number of both canonical and non-canonical RBPs (Baltz et al., 2012; Bao et al., 2018; Castello et al., 2012; Kwon et al., 2013). Various typical DNA-binding proteins are also long known to bind both DNA and RNA (Cassiday and Maher, 2002), which has been extended to many transcription factors (TFs), such as CTCF (Kung et al., 2015; Saldaña-Meyer et al., 2014); enzymes involved in DNA repair, like Ku80/XRCC5 (Baltz et al., 2012; Ting et al., 2005); and transcription complexes, exemplified by polycomb complex 2 (PRC2) (Davidovich et al., 2015). Current estimates suggest that as many as 1,500 proteins have the capacity to bind RNA in the human genome (Gerstberger et al., 2014), and given such a large unexpected repertoire of RBPs in mammalian cells, we now need to study their functions beyond the traditional framework.

RBPs are involved in all aspects of RNA metabolism. Now, a well-accepted theme is that many RNA-processing events are tightly coupled with transcription (Bentley, 2014). Co-transcriptional RNA processing enables not only efficient and sequential recognition of emerging *cis*-acting regulatory elements in nascent RNA but may also affect downstream RNA fate, as documented for the role of gene promoters in specifying alternative splicing (Cramer et al., 1997; Moldón et al., 2008), RNA stability (Bregman et al., 2011; Trcek et al., 2011), alternative polyadenylation (Oktaba et al., 2015), and even translational control in the cytoplasm (Zid and O'Shea, 2014). These findings highlight functional integration of transcriptional and post-transcriptional machineries. As such, efficient coupling would require intimate interactions of key components of different machineries, suggesting that various RBPs may be directly involved in such integration processes through their actions on or in the proximity of chromatin.

It has also become clear that mammalian genomes are more actively transcribed than previously anticipated (Djebali et al.,

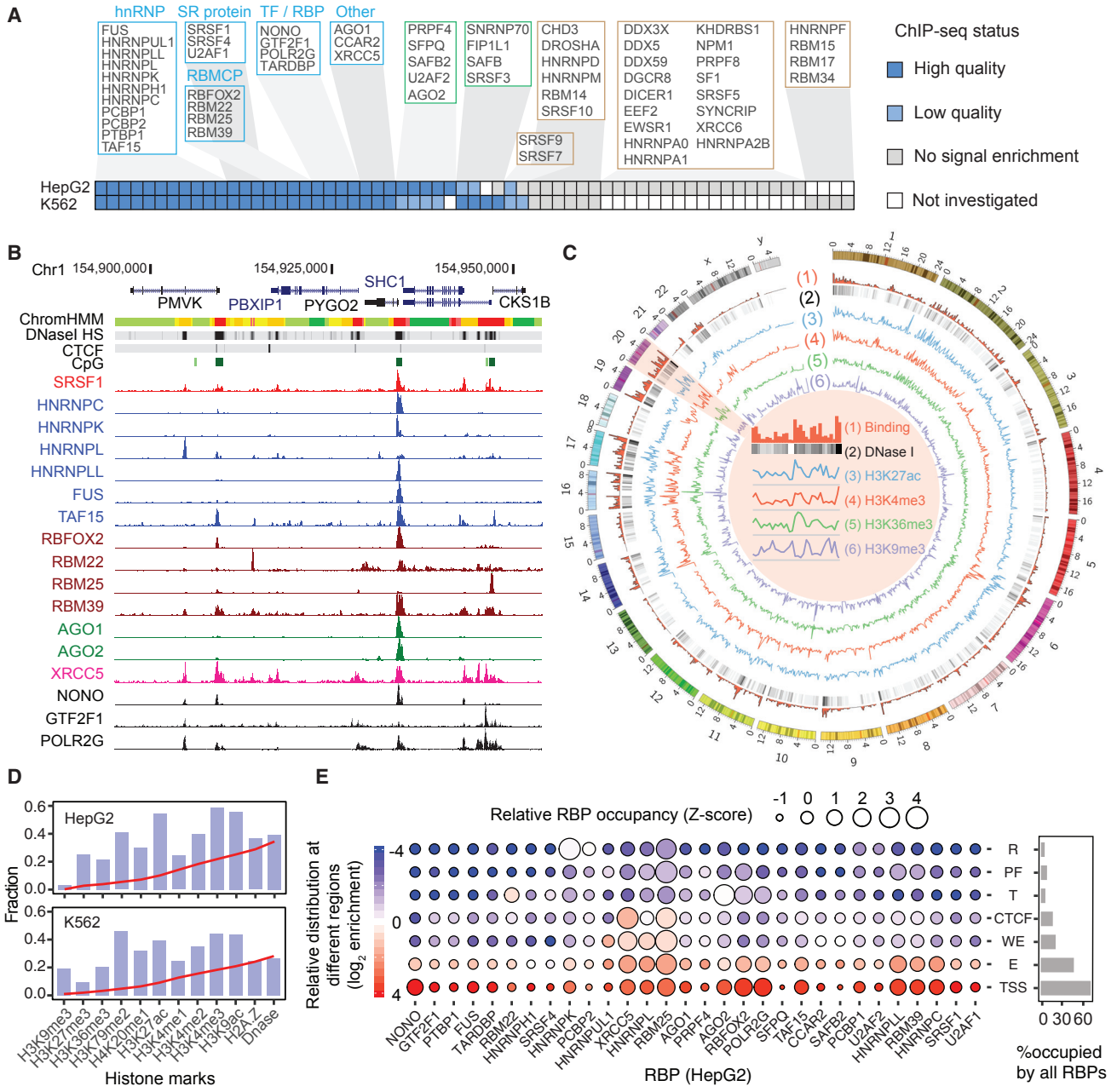


Figure 1. General Features of Chromatin-Associated RBPs

(A) Summary of RBPs surveyed by ChIP-seq in HepG2 and K562 cells. The 25 RBPs that produced high-quality ChIP-seq data are grouped into five classes: (1) hnRNP proteins, (2) SR proteins, (3) TFs that bind RNA, (4) proteins containing RNA-binding motif (RBM), and (5) others. Dark blue, high-quality data that met the ENCODE standards and showed a minimal number (≥200) of specific binding peaks; light blue, ChIP-seq data that met all other ENCODE standards except for sequencing depth (<10 million non-redundant reads) or the number of specific binding peaks (<200); gray, no signal enrichment after IP despite efficient IP detected by western blot; white, not investigated.

(B) A typical genomic region displaying annotated gene structures and four key chromatin features determined by ENCODE on HepG2 cells. The ChromHMM segments highlighted by red and orange correspond to promoters and enhancers, respectively.

(C) Circos plot showing the relationship between collective RBP-chromatin interactions, open chromatin regions detected by DNase I hypersensitivity, and key histone modification events in HepG2 cells. Chromosome 20 is magnified to illustrate positive and negative correlations with key histone modification events.

(D) Coverage of individual histone modification events by chromatin-associated RBPs in HepG2 and K562 cells and the accumulative coverage of all chromatin regions associated with at least one biochemical activity (red line).

(E) RBP occupancy on specific states of ENCODE-annotated genome segmentation in HepG2 cells. The seven states of segmentation are: R, repressive regions; PF, promoter flanking regions; T, transcribed regions; CTCF, CTCF-binding sites; WE, weak enhancers; E, enhancers; TSSs, transcription start sites/promoters.

(legend continued on next page)

2012). Besides the production of typical protein-coding mRNAs, mammalian genomes also generate numerous non-coding RNAs, including long non-coding RNAs (lncRNAs), many of which are directly involved in transcription control (Fu, 2014; Rinn and Chang, 2012; Vance and Ponting, 2014). Furthermore, transcriptional enhancers also produce enhancer-associated RNAs (eRNAs), which may mediate enhancer-promoter communications to enhance gene expression (Kim et al., 2010; Lam et al., 2014; Wang et al., 2011). In principle, various regulatory RNAs likely enlist specific RBPs to execute their functions. Indeed, increasing evidence suggests that many RBPs have direct roles in transcription, as exemplified by the elucidated function of typical splicing regulators in transcription, including SRSF2 (Ji et al., 2013), RBFox2 (Wei et al., 2016), NONO (Shav-Tal and Zipori, 2002), HNRNPL (Kuninger et al., 2002), and HNRNPK (Michelotti et al., 1996). In fact, PGC1α, an extensively characterized master regulator of metabolic programs in mammals, is structurally a typical RBP (Puigserver and Spiegelman, 2003). More recently, even the typical TF YY1 appears to bind enhancers in an RNA-dependent manner (Sigova et al., 2015), which may underlie its newly elucidated role in mediating enhancer-promoter looping (Weintraub et al., 2017). These findings raise the possibility that many RBPs may actually function as bona fide TFs.

This emerging picture for specific RBPs to participate in transcription and co-transcriptional RNA processing raises the question of how prevalently RBPs exert their functions at the level of chromatin. To address this global question, we participated in an ENCODE project to survey RBPs on chromatin by chromatin immunoprecipitation sequencing (ChIP-seq), initially focusing on RBPs that have specific antibodies available and are predominately or partially localized in the nucleus. Among 58 and 45 RBPs respectively analyzed on HepG2 and K562 cells, ~60% showed strong association with chromatin. We further took advantage of this rich resource to intersect RBP-chromatin interactions with ENCODE ChIP-seq profiles for TFs in the same cell lines to reveal numerous co-binding events, thus providing evidence for coordinated actions of TFs and RBPs.

Focusing on a recently elucidated regulatory paradigm in YY1-mediated gene expression, where YY1 appears to bind chromatin in an RNA-dependent fashion (Sigova et al., 2015) and to play a larger role than CTCF in mediating enhancer-promoter interactions in the human genome (Weintraub et al., 2017), we pursued the transcription function of a key YY1 co-binder RBM25, revealing that this RBP is required for strengthening nearly all YY1-dependent transcription activities in the human genome. These findings highlight the possibility that transcriptional and post-transcriptional RNA processing may be more functionally intertwined than just acting at similar times and places, with many traditional components of the prospective machineries playing direct and diverse functions in regulated gene expression.

RESULTS

Selection of RBPs for Large-Scale Survey by ChIP-Seq

To broadly investigate the potential function of RBPs at chromatin levels, we conducted a systematic ChIP-seq survey of 58 and 45 RBPs in HepG2 and K562 cells, respectively (Figure 1A; Table S1). These RBPs were selected for this discovery-driven study based on the following criteria: (1) partial or exclusive localization in the nucleus; (2) availability of antibodies capable of specific and efficient immunoprecipitation for detecting possible RBP-chromatin interactions based on a previous survey (Sundararaman et al., 2016) and additional screening efforts (Table S1); (3) representation of diverse RNA-binding domain structures (e.g., SR proteins, RNA-binding motif-containing proteins, KH domain-containing proteins, etc.) and functional classes (e.g., spliceosome components, RNA helicases, etc.); (4) prior information on certain RBPs either as components of specific TF-containing complexes or with well-documented effects on transcription (e.g., NONO, GTF2F1, POLR2G, SFPQ, and TARDBP). Many of these RBPs (~44) are expressed in both cell lines based on the existing ENCODE RNA sequencing (RNA-seq) data, thus enabling analysis of both common and cell-type-dependent functions.

To ensure the data quality, all ChIP-seq experiments were performed in replicate and following the ENCODE standards established for TFs (https://www.encodeproject.org/chip-seq/transcription_factor/). Because RBPs may not associate with chromatin as tightly as typical TFs, we made some modifications to enhance the ChIP efficiency (see Method Details). On average, we obtained ~12 million usable reads for each library after excluding low-quality, multi-mapped reads and PCR duplicates (Table S1). We identified confident peaks by using the SPP (sequencing processing pipeline) peak calling algorithm (Kharchenko et al., 2008), with the threshold for IDR (irreproducible discovery rate) set at 0.02 (Li et al., 2011), both according to the ENCODE Uniform ChIP-seq Processing pipeline (see Method Details). Our data for POLR2G (aka RBP7), an RNAPII subunit with the documented ability to bind RNA, were highly consistent with the previously produced TF ChIP-seq for POLR2A, the largest subunit of RNAPII (Figure S1A), indicating the robust data generated under our standardized conditions.

Global Features of RBP-Chromatin Interactions

Using the high-quality dataset, we first asked how many RBPs are associated with chromatin, finding that 51.7% (30 of 58) RBPs in HepG2 and 64.4% (29 of 45) in K562 showed extensive and specific interactions with chromatin (Figure 1A). These RBPs typically exhibited several hundred to more than 10,000 peaks on chromatin (Figure S1B), as exemplified on a multi-gene locus (Figure 1B). In most cases, RBPs showed strong binding in both HepG2 and K562 cells, with three RBPs in HepG2 and six RBPs in K562 cells exhibiting marginal association in one or both cell types (highlighted in light blue in Figure 1A; listed in Table S1).

For each RBP, the relative distribution of its occupied sites on individual segments (vertical comparison) is color coded (key on the left). For each class of segment annotation, the relative distribution of individual RBPs is represented by bubble size (horizontal comparison), as indicated by each RBP's *Z* score. Right: summed percentage of individual segment annotations covered by the surveyed RBPs.
See also Figure S1 and Tables S1 and S2.

These data reveal that a large portion of nuclear RBPs act at the chromatin level.

We next characterized global features of RBP-chromatin interactions. Focusing first on the data from HepG2 cells, it became immediately evident that RBPs generally prefer open chromatin regions according to ENCODE-annotated chromatin states (compbio.mit.edu/ChromHMM) and DNase I hypersensitive sites, which are often associated with CTCF-binding sites and CpG islands, as seen on a representative genomic segment (Figure 1B). This reflects a global trend, as RBP-chromatin interactions tended to positively correlate with active histone modifications (e.g., H3K27ac, H3K4me3, and H3K36me3, markers for activated enhancers, active promoters, and transcribed regions, respectively) but negatively correlate with repressive H3K9me3 marker (Figures 1C, further highlighted on chromosome 20 in the insert, and S1C; Table S2). This general pattern was quite consistent when extending the analysis to additional histone modifications and to both cell lines. Collectively, among RBPs showing detectable chromatin association, we found that ~30%–40% of biochemically active chromatin regions have evidence for association with at least one RBP (Figure 1D).

Promoters as Hotspots for RBP Binding

Strikingly, all chromatin-associated RBPs showed a general preference for gene promoters. However, it was equally clear that different RBPs exhibit preference for different sets of promoters while some specific promoters are bound by multiple RBPs (see examples in Figure 1B), reminiscent of similar but distinct chromatin-binding profiles of various TFs (Consortium, 2012; Gerstein et al., 2012; Moorman et al., 2006). To globally characterize the binding preference of individual RBPs surveyed, we assigned ChIP-seq peaks of each RBP to seven ENCODE-annotated genome segmentations and assessed the relative distribution of peaks for each RBP among these segmentations (Figures 1E and S1D). These data clearly suggest that, like TFs, active promoters are also hotspots for RBPs.

Collectively, the chromatin-associated RBPs we surveyed covered ~70% of promoters in HepG2 and ~80% in K562 cells (Figures 1E, right, and S1D). Power analysis indicates that all or almost all promoters are likely to be bound by one or more RBPs, even after we excluded the RBPs (i.e., NONO, GTF2F1, POLR2G, SFPQ, and TARDBP) that have previously been implicated in transcription control (Figure S1E). Importantly, individual RBPs appear to have distinct preferences for different promoters rather than binding indiscriminately to open chromatin (see Figure 1B). We illustrated this by randomly distributing the RBP-binding sites to open chromatin regions based on mapped DNase I hypersensitive sites and then counting co-localized RBP-binding sites, assuming that all RBP chromatin association would be mediated by their general affinity for open chromatin. The distribution of real data is clearly distinct with the simulated one, with the former showing a trend toward decreased co-binding observed in both cell types (Figure S1F).

We also determined the relative contribution of individual RBPs to each genomic segmentation type, as indicated by relative bubble sizes (see Figures 1E and S1D). This reveals that, relative to other RBPs, HNRNPK is a major RBP on repressive regions, consistent with its previously elucidated function in transcription repression (Pintacuda et al., 2017). In contrast, AGO2 is enriched in transcribed regions in HepG2 cells. The specific chromatin-binding activities of both AGO1 and AGO2 reinforce the more recent realization of the functionality of small RNA machineries within the nucleus in both plants and animals (Huang et al., 2013; Liu et al., 2018; Skourti-Stathaki et al., 2014; Taliaferro et al., 2013). Interestingly, the three RBPs XRCC5, HNRNPL, and RBM25 appear to be more generally linked to promoters and enhancers than other RBPs (see below).

Specificity and Conservation of RBP Promoter-Binding Activities

The fact that promoters are the primary interface for RBP-chromatin interactions prompted us to look further into the promoter-binding profiles of individual RBPs. By classifying promoters into subgroups based on either epigenetic marks or specific sequence features (see Method Details), we found that RBPs collectively show a general preference for bivalent promoters marked with both H3K4me3 and H3K27me3, for active promoters modified by H3K4me3 alone, and for CpG island promoters in both cell lines (Figures 2A and S2A). The enrichment of RBPs at bivalent and CpG island promoters might reflect the involvement of various RBPs in the dynamic regulation of gene expression via nascent RNAs (Wei et al., 2016) and/or the formation of R-loops (Chen et al., 2017). However, the regulation and maintenance of transcriptome is clearly distinct between HepG2 and K562 cells, as bivalent genes are more prevalent than H3K4me3-only genes in HepG2 cells, but the opposite is true in K562 cells (Figures 2A and S2B), suggesting that chromatin-associated RBPs may actively participate in cell-type-specific gene expression programs.

With respect to promoters of genes that function in different categories, we noted that four RBPs (i.e., RBM22, PRPF4, HNRNPUL1, and SNRNP70) exhibited additional enrichment on promoters of small RNA genes relative to promoters of protein coding and lncRNA genes (Figure 2B). Furthermore, PRPF4 was highly enriched on tRNA gene promoters and SNRNP70 on small nucleolar RNA (snoRNA) gene promoters (Figures 2C and S2C). The prevalent association of PRPF4 with almost all expressed tRNA gene promoters (Figure S2D) and its similar expression, localization, and protein domain with multiple RNAPIII subunits (Figure S2E) suggest that this RBP may be actively involved in tRNA transcription and/or co-transcriptional tRNA processing.

We next considered the position of RBP ChIP-seq peaks relative to transcription start sites (TSSs). We cataloged individual RBP ChIP-seq signals around annotated TSSs (Figure 2D; see Method Details). Although all RBPs showed binding events on both sides of TSSs, we noted three apparent classes of RBP-chromatin interaction profiles based on their binding profiles around TSS: (1) upstream TSS (i.e., RBM25), (2) centered on TSS (i.e., GTF2F1), and (3) downstream TSS (i.e., RBFOX2), the third class being representative of the majority of RBPs in both cell types (Figure 2D). The TSS-centered binding pattern of GTF2F1 likely reflects its function as part of the core transcription machinery (Aso et al., 1992), hinting that other RBPs with a similar association pattern might have a related function. The association of most RBPs with sequences downstream of TSS suggests that many RBPs may interact with chromatin in a

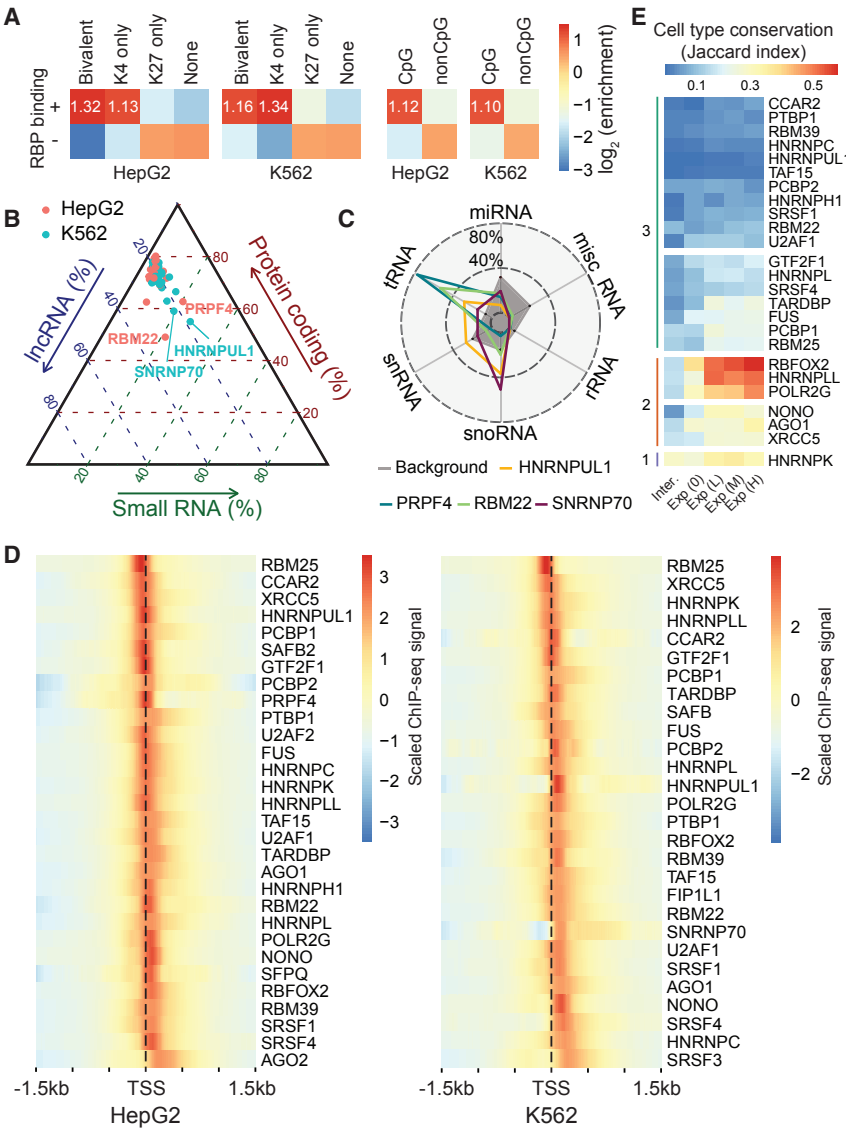


Figure 2. Distinct RBP-Chromatin Interaction Patterns on Different Promoter Classes

(A) Collective RBP preference for promoter subgroups segregated by sequence context (with or without CpG islands) or histone modification features, such as bivalent promoters marked by both H3K4me3 and H3K27me3 signals, promoters containing only H3K4me3 or H3K27me3 signals. Right: key for relative enrichment.

(B) Relative occupation frequencies of individual RBPs on different classes of gene promoters.

(C) The distribution of RBP ChIP-seq peaks among the six classes of small RNA gene promoters relative to background distribution.

(D) Composite RBP-binding signals around TSSs. RBPs are ordered based on their relative positions of signal maxima to TSS.

(E) Jaccard index for cell-type conservation of RBP-chromatin interactions between HepG2 and K562 cells (see Method Details). RBP occupation in genic regions are segregated according to expression levels: Exp (0), non-expressed; Exp (L), lowly expressed (bottom third of expressed genes); Exp (M), expressed at intermediate levels (middle third of expressed genes); Exp (H), highly expressed (top third of expressed genes).

See also Figure S2 and Table S2.

nascent RNA-dependent manner, as we recently documented with RBFox2 (Wei et al., 2016) and validated in this study with RBM22 and HNRNPL (see below).

Given the potential influence of nascent RNAs on RBP-chromatin interactions, we classified genes into subgroups with comparable expression levels between HepG2 and K562 cells and calculated the Jaccard Index (which measures the overlapped chromatin interactions for each RBP between the two cell types) to assess conservation of binding across cell types. This analysis uncovered at least three groups of RBPs based on hierarchical clustering of the Jaccard Index (Figure 2E). Specifically, HNRNPK, the only member in the first group, showed relatively constant conservation across the whole genome regardless of gene expression levels. The second group consists of a few RBPs that interacted with genes proportional to their levels of expression in both HepG2 and K562 cells. The third group, which includes most of the RBPs we surveyed, showed cell-type-specific association patterns across all subgroups, suggesting their distinct functions in different cell types.

Roles of Promoter-Associated RBPs in Different Levels of Gene Expression

Given the tight association of RBPs with gene promoters, we next asked whether RBP-promoter interactions might reflect their roles in transcription by examining the relationship between promoter interaction and transcription output for each RBP. Focusing on HepG2 cells, we compared changes in gene expression at steady state before and after knockdown of each RBP. Before knockdown, the interaction of most RBPs with promoters correlated to target gene transcription activities (Figure 3A, left), and upon knockdown, nearly all RBPs affected gene expression at steady state (Figure 3A, middle; Table S2). Note that knockdown of HNRNPUL1 or RBFOX2 induced little change in gene expression likely because of modest reduction at protein levels as noted earlier (Van Nostrand et al., 2018). Because changes in RNA-seq signals likely result from gene expression regulation at transcriptional and/or post-transcriptional levels, we selected a large subset of RBPs (n = 14) to directly measure their influence on gene expression at the transcriptional level by performing global run-on sequencing (GRO-seq) before and after knockdown (Tables S3 and S4; Figure S3A). We found that at least six of these RBPs (i.e., RBM22, XRCC5, RBM25, HNRNPK, HNRNPLL, and U2AF1) had considerable direct impacts on transcription, each

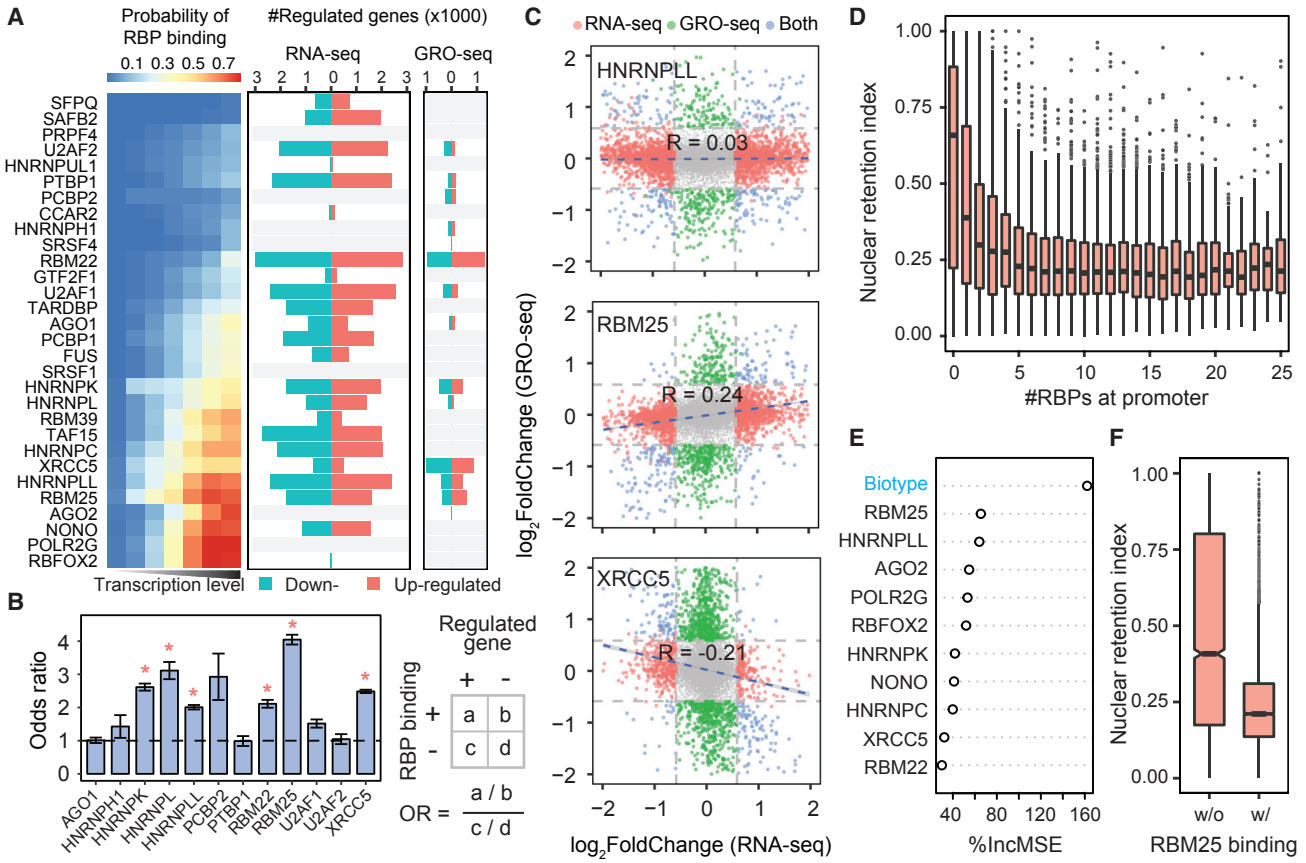


Figure 3. Correlation between RBP-Promoter Interaction and Gene Expression

(A) Correlation between the probability of RBP association with promoters and target gene transcription activities profiled by GRO-seq in HepG2 cells (left); response to knockdown of individual RBPs profiled by RNA-seq (middle) or GRO-seq (right). Significantly up- or down-regulated genes are determined by adjusted p value of ≤ 0.05 and fold-change of $\leq 2/3$ or $\geq 3/2$.

(B) Odds ratio of transcriptional response determined by GRO-seq on RBP-occupied promoters compared to non-occupied promoters. Right: definition of odds ratio. *p < 0.05 (Fisher's exact test).

(C) Comparison between changes in gene expression profiled by RNA-seq and GRO-seq upon knockdown of the three representative RBPs.

(D) The distribution of RNA nuclear retention index (nuclear/(nuclear+cytoplasmic)) for each group of genes whose promoters were occupied by different RBPs.

(E) Variable importance determined by Random Forest to evaluate the prediction power of each variable (see Method Details). Top ten RBPs are shown.

(F) The distribution of RNA nuclear retention for genes with or without evidence for binding of RBM25 on their promoters.

See also Figure S3 and Tables S2–S4.

inducing differential expression of >500 genes upon knockdown (Figure 3A, right). To determine whether the induced gene expression was linked to their promoter association, we further calculated the odds ratio for each RBP, asking whether RBP-associated promoters were more linked than unassociated promoters to induced gene expression measured by GRO-seq (Fisher's exact test). All six of these RBPs showed a significant odds ratio (>1, $p < 0.05$), based either on total regulated genes or separately on up- or down-regulated genes (Figures 3B and S3B). These data strongly suggest that these RBPs directly participate in transcriptional control.

Compared to GRO-seq, RNA-seq measures a combined consequence of regulated gene expression at both transcriptional and post-transcriptional levels. We thus compared between GRO-seq and RNA-seq signals and observed three distinct patterns with respect to global changes in transcription versus RNA levels at steady state (Figure 3C). For HNRNPLL, we saw little concordance between RNA-seq and GRO-seq upon knockdown (Figure 3C, top), indicating that this RBP may independently regulate transcriptional and post-transcriptional events. We observed a modest positive correlation with RBM25, suggesting that RBM25-mediated transcription may directly contribute to RNA levels at the steady state (Figure 3C, middle). In contrast, we detected a negative correlation with XRCC5 (aka Ku80) (Figure 3C, bottom), a helicase that has been mainly characterized for its role in DNA repair (Taccioli et al., 1994) and recently found to also bind mRNA (Baltz et al., 2012). Such a negative correlation might result from selective roles of XRCC5 in inducing less stable RNAs and/or repressing more stable RNAs, possibilities that would be interesting to follow up.

Using the resultant chromatin association and gene expression data, we were particularly interested in testing the so-called

promoter loading model, in which promoter association events are thought to instruct downstream RNA processing events, such as RNA stability, export, or translation, as reported on a few specific cases (Bregman et al., 2011; Moldón et al., 2008; Oktaba et al., 2015; Trcek et al., 2011; Zid and O'Shea, 2014). For this purpose, we calculated the ratio of nuclear retention based on the cell fractionation sequencing (CeFra-seq) data in HepG2 cells (Benoit Bouvrette et al., 2018) and RBP knockdown-induced splicing changes (assessed by "percent spliced in," or PSI) and determined the functional consequences in relationship with RBP chromatin-binding activities. Indeed, we noted a trend in which nuclear retention of mRNAs was inversely correlated with the number of RBPs detected at gene promoters (Figure 3D), which is the case for both protein-coding and non-coding genes (Figure S3C). We also found that different RBPs differentially contributed to this effect (Figure 3E), as determined by Random Forest (see Method Details). In particular, the promoter-association activity of RBM25 was most predictive of nuclear retention (Figure 3F). When separately analyzing coding and non-coding genes with or without evidence for RBM25 binding, it became more clear that the absence of RBM25 at promoters was pronouncedly linked to nuclear retention of lncRNAs (Figure S3D). These data suggest that many RBPs may co-transcriptionally facilitate RNA export, as previously reported (Köhler and Hurt, 2007). In contrast, we failed to detect any strong evidence for global coupling between promoter binding and RBP-dependent splicing (data not shown). The lack of such global coupling does not rule out the possibility for specific coupling events, which requires further dissection on individual cases, as global coupling between promoter interaction and regulated splicing might be largely masked by other RBP-mediated regulatory events at gene promoters or RBP-regulated splicing after transcription.

Network Interaction of RBPs and TFs in the Human Genome

To understand how RPBs might affect transcription, we next considered the possibility that, like TFs, RBPs may partner with other proteins, as we previously demonstrated with the splicing factors SRSF2 (Ji et al., 2013) and RBFox2 (Wei et al., 2016). We thus asked how RBP-chromatin interactions might be coordinated with one another and with specific TFs. To this end, we integrated our RBP ChIP-seq data with available ChIP-seq data for TFs in HepG2 cells (Table S2) and analyzed their co-association events by using a newly developed non-negative matrix factorization (NMF) approach (Li et al., 2017b). This method employs a "soft-clustering" strategy, allowing one RBP to be assigned to more than one group, as applied to genomic context-dependent protein-protein associations (Gerstein et al., 2012; Xiong et al., 2015). Using this approach (see Method Details), we first defined a unifying set of *cis*-regulatory elements (CREs) associated with individual RBPs and TFs. An occupancy profile was then constructed based on whether individual CREs were associated by each RBP or TF, followed by decomposing the occupancy profile into the mixture coefficient matrix and the basic matrix, with the former giving the resulting factor groups and coefficient values of individual RBPs and TFs in each group and the latter providing the probability of each CRE recognized by each factor group. In this analysis, it is also necessary to pre-determine a reasonable number of factor groups based on local maxima. In our case, we found 17 as an optimal group number (Figure 4A), with which the clustering became stabilized as revealed by cophenetic and dispersion correlation coefficients (Figure S4A) and the consensus matrix (Figure S4B). Each of these factor groups showed specific preference for different genomic regions according to ENCODE annotation (Figure 4B). It is important to emphasize that NMF analysis takes the binding patterns of all RBPs and TFs (n = 84) into consideration, and we grouped individual CREs based on dominant co-binding patterns. Therefore, any specific co-binding group does not necessarily exclude other relatively less dominant RBPs or TFs in binding to their CREs, and as a result, any specific pairwise co-binding events may be segregated into multiple CREs assigned to different NMF groups that were divided based on the most dominant co-binding events.

Consistent with the literature, group 3 contains three proteins, CTCF, RAD21 and SMC3, all of which are known to collaborate with one another in mediating long-distance genomic interactions (Merkenschlager and Odom, 2013; Ong and Corces, 2014), and consistently, we found that these proteins were all predominantly localized in CTCF-enriched genomic segments in HepG2 cells (Figure 4B). Overall, the RBPs currently surveyed could be segregated into 5 out of a total of 17 groups, one of which (group 10) was only composed of RBPs (Figure 4C). Importantly, multiple factors in each group have evidence for physical interactions at protein levels (Warde-Farley et al., 2010) (blue edges in Figure 4C). Previous studies revealed numerous high-occupancy target (HOT) regions in mammalian genomes that are targeted by an unusually high number of TFs (Moorman et al., 2006; Xie et al., 2013). By ranking RBPs based on their fraction of peaks in HOT regions in the human genome and segregating them into four ascending quartiles, we also found that more than half of the RBPs in group 4 ranked together with various TFs at the top quartile, including PRPF4, SRSF4, and others (Figure 4D). Thus, like TFs, RBPs also show great preference for HOT regions in the human genome.

Evidence for Cooperative Binding of RBM25 and YY1 on Chromatin

We noted that the TF YY1 was clustered into two separate co-association groups: group 9, which contains multiple TFs as well as the RBP NONO, which is long shown to be a multi-functional nuclear protein involved in RNA processing, DNA repair and recombination, and transcription (Shav-Tal and Zipori, 2002); and group 13, which comprises YY1, XRCC5, and RBM25 (Figure 4C), with the former two known to physically interact with one another (Sucharov et al., 2004). Given the recent finding that YY1 appears to interact with DNA in an RNA-dependent manner (Sigova et al., 2015), it is striking to note that 68% (25% for shared occupancy between YY1 and RBM25 and 43% for all three factors) of YY1 genomic-binding sites also exhibited RBM25 binding in HepG2 cells (Figure 4E, top), where 70% of their co-binding events occurred around TSS regions, although such co-occupancy was much less pronounced (14%, with 10% for co-occupancy between YY1 and RBM25 and 4% for all three factors) in K562 cells (Figure S4C). These

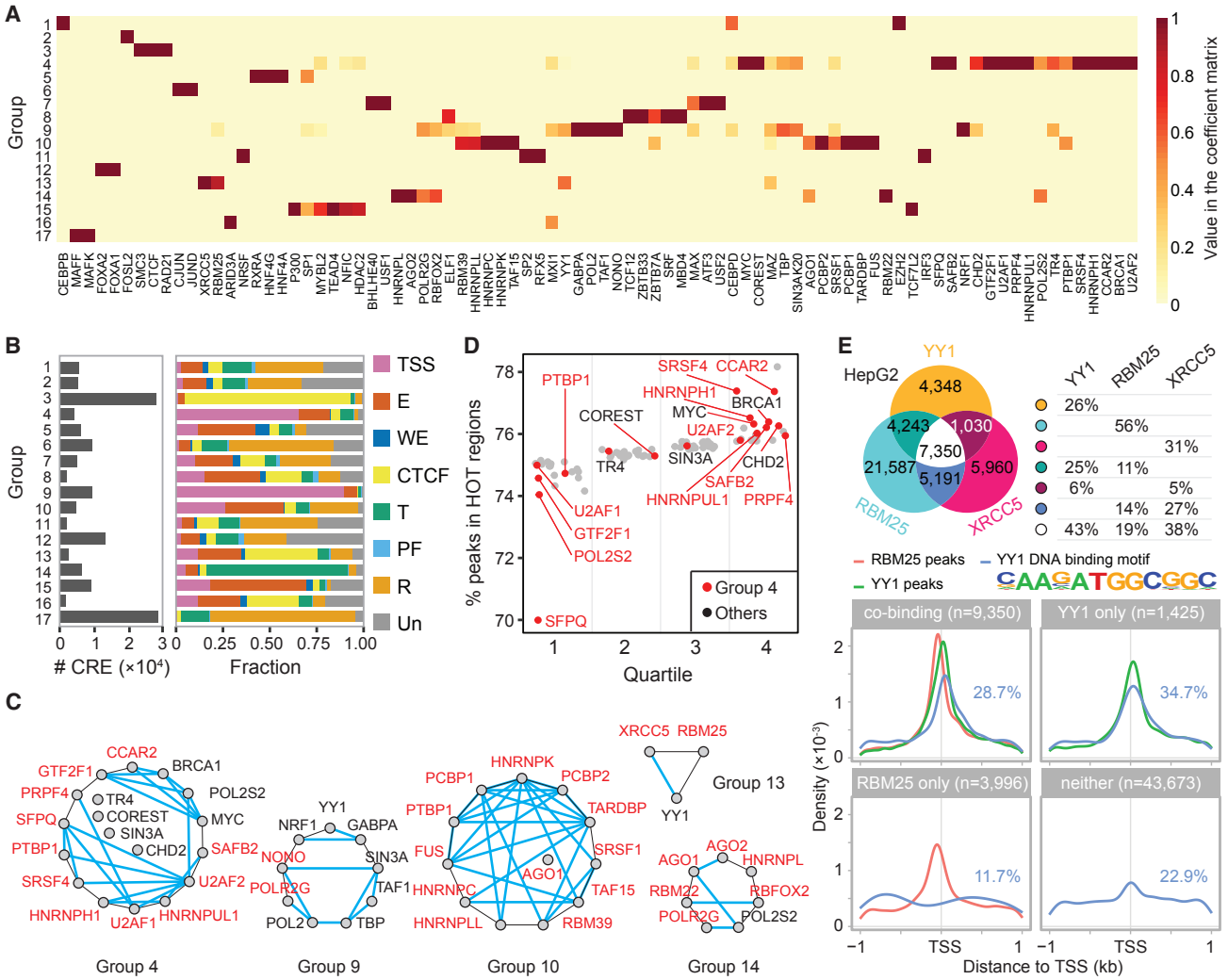


Figure 4. Integrated Analysis of Chromatin-Associated RBPs and TFs in HepG2 Cells

(A) Segregation of chromatin-associated RBPs and TFs into 17 groups by NMF-inferred coefficient matrixes (see Method Details).

(B) Coverage and annotation of total *cis*-regulatory elements (CREs, defined in Figure 1E) by individual NMF groups (left). Un, unannotated regions. Fractions of different CREs occupied by each group are shown on the right.

(C) Representative NMF-segregated groups. Blue line, known physical interactions between members in each group annotated by GeneMANIA.

(D) Preferential association of RBPs with HOT regions. RBPs are segregated into four quartiles (gray lines) based on their relative binding in HOT promoter regions. y axis: the percentage of total peaks that fall into HOT regions.

(E) Top: co-localization of RBM25, XRCC5, and YY1 in HepG2 cells. Each fraction of the Venn diagram was further quantified as the percentage of peaks for each RBP, based on which individual pairwise co-localization was calculated. Bottom: the distribution of the core YY1-binding motif relative to YY1, RBM25, and YY1-binding peaks. The number of TSSs and their percentage associated with the core YY1-binding motif are indicated in each case.

See also Figure S4 and Table S2.

observations suggest that RBM25 may play a broad role in modulating the transcriptional function of YY1 in a cell-type-specific manner.

To begin to explore this interesting possibility, we sought to showcase the utility of the current RBP ChIP-seq data to reveal previously unrecognized regulatory mechanisms for gene expression. We separately analyzed RBM25 and YY1-binding profiles in relationship to the distribution of YY1-binding motif around TSSs (Figure 4E, bottom) and observed that the core YY1-binding motif is enriched at the promoter regions that showed YY1 and RBM25 co-binding or YY1 binding alone, but not RBM25 binding alone, indicating that the core YY1-binding motif mediates YY1 binding. Interestingly, despite a similar distribution of the YY1-binding motif around TSSs, YY1 appears to bind more strongly to co-bound promoters than those bound by YY1 alone, suggesting that RBM25 might enhance YY1 binding. We further noted that the YY1-binding events were largely aligned with the YY1-binding motif on promoters bound by YY1 alone, but we detected a slight shift of YY1-binding events toward the composite RBM25-binding summit on co-bound

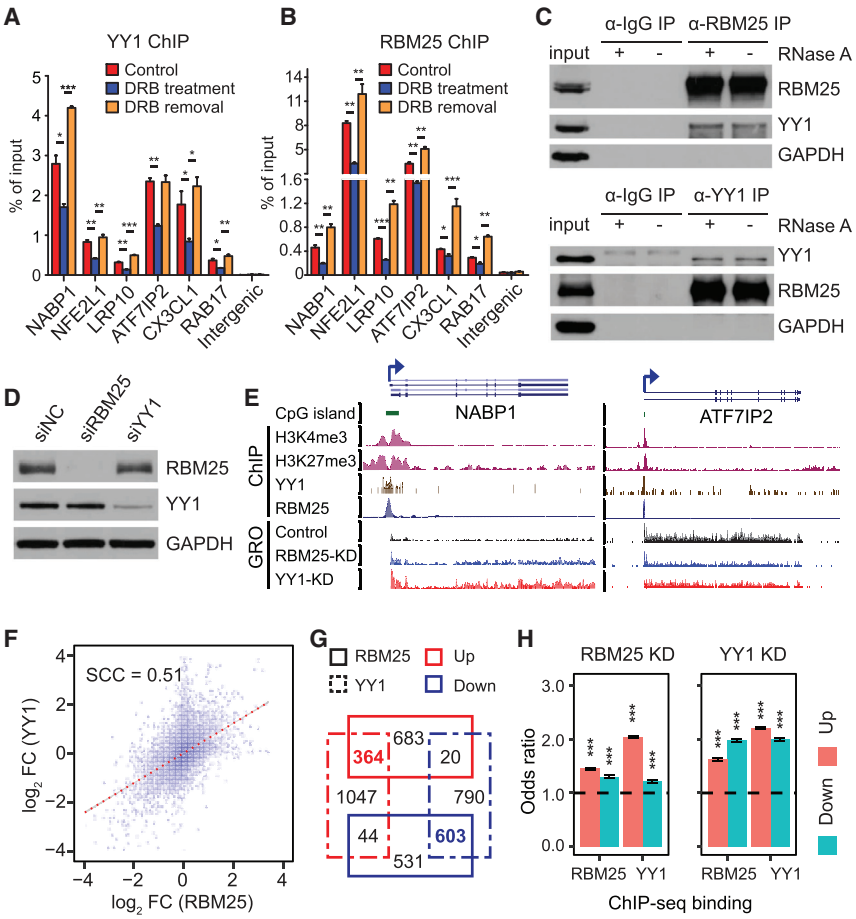


Figure 5. Co-regulation of Gene Expression by YY1 and RBM25 in HepG2 Cells

(A and B) ChIP-qPCR analysis of YY1 (A) and RBM25 (B) binding on representative target gene promoters upon DRB treatment and after DRB washout. Data are presented as mean ± SD (n = 3). *p < 0.05, **p < 0.01, ***p < 0.001 (unpaired Student t test).

(C) Reciprocal coIP of RBM25 and YY1 in an RNA-independent manner.

(D) Efficient knockdown of RBM25 or YY1 without affecting the other protein.

(E) RBM25 and YY1 ChIP-seq profiles on the two representative genic loci, one for up-regulated (left) and one for down-regulated (right) genes, as determined by GRO-seq.

(F) Global comparison of transcriptional responses to knockdown of RBM25 and YY1. Spearman's correlation coefficient (SCC) of fold-changes (FCs) is indicated.

(G) Overlap of differentially expressed genes in HepG2 cells depleted of RBM25 or YY1. The colored boxes and line types separately denote gene expression events up (red box)- or down (blue box)-regulated by RBM25 (solid line) or YY1 (dashed line).

(H) Positive association of RBM25 and YY1 binding at target gene promoters with induced gene expression determined by odds ratio. ***p < 0.001 (Fisher's exact test).

See also Figure S5 and Tables S5 and S6.

promoters. This suggests that presence of RBM25 may shift the position of YY1 more upstream.

Functional Corporation of YY1 and RBM25 in Transcription

To pursue the potential functional requirement of RBM25 in mediating YY1-dependent transcription, we first verified RNA-dependent YY1 binding on chromatin, as reported on mouse embryonic stem cells (mESCs) (Sigova et al., 2015). We selected multiple gene promoters that showed YY1-binding peaks in HepG2 cells from the ENCODE data (Table S2) and performed ChIP-qPCR for both YY1 and RBM25 in response to treatment with 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) to block transcription (Table S5). We found, as previously reported, that YY1 binding was attenuated on all target loci upon DRB treatment, which was restored upon DRB washout (Figure 5A). We obtained the same quantitative trend with RBM25 ChIP-qPCR in response to DRB treatment and after DRB removal (Figure 5B), although different promoter regions appeared to differ in relative occupancy by YY1 versus RBM25. Given the observation that most RBPs appear to bind chromatin downstream of TSSs (see Figure 2D), we extended the analysis to five more RBPs that showed impact on transcription determined by GRO-seq (see Figure 3A). Interestingly, we found that three of these RBPs (RBM22, XRCC5, and HNRNPL) showed nascent RNA-dependent binding, while the other two (HNRNPK and HNRNPLL) did not (Figure S5A). These data suggest that some RBPs may use nascent RNA to gain access to gene promoters, while others may directly bind DNA sequences or participate in complexes assembled on gene promoters via interactions with other promoter-bound proteins.

Focusing on the mechanism underlying their potential cooperative binding, we next asked whether YY1 and RBM25 formed a complex and found that they indeed interact with one another based on reciprocal co-immunoprecipitation (coIP), which was insensitive to RNase A treatment, indicating that they likely interact either directly or via an intermediary, but clearly not RNA (Figure 5C). Note that, because the RBM25 antibody is much more efficient than anti-YY1 in western blotting, we detected stronger RBM25 signals in both of the reciprocal coIP experiments, indicating a tight association between YY1 and RBM25 in HepG2 cells.

We then performed small interfering RNA (siRNA)-mediated depletion of YY1 or RBM25 in HepG2 cells, which showed specific knockdown effects on each protein, but without detectable effect on the non-targeted protein (Figure 5D). Based on GRO-seq analysis of these cells, which showed highly reproducible patterns (Figure S5B), we identified a large number of up- or down-regulated genes in each case (Figures S5C and S5D),

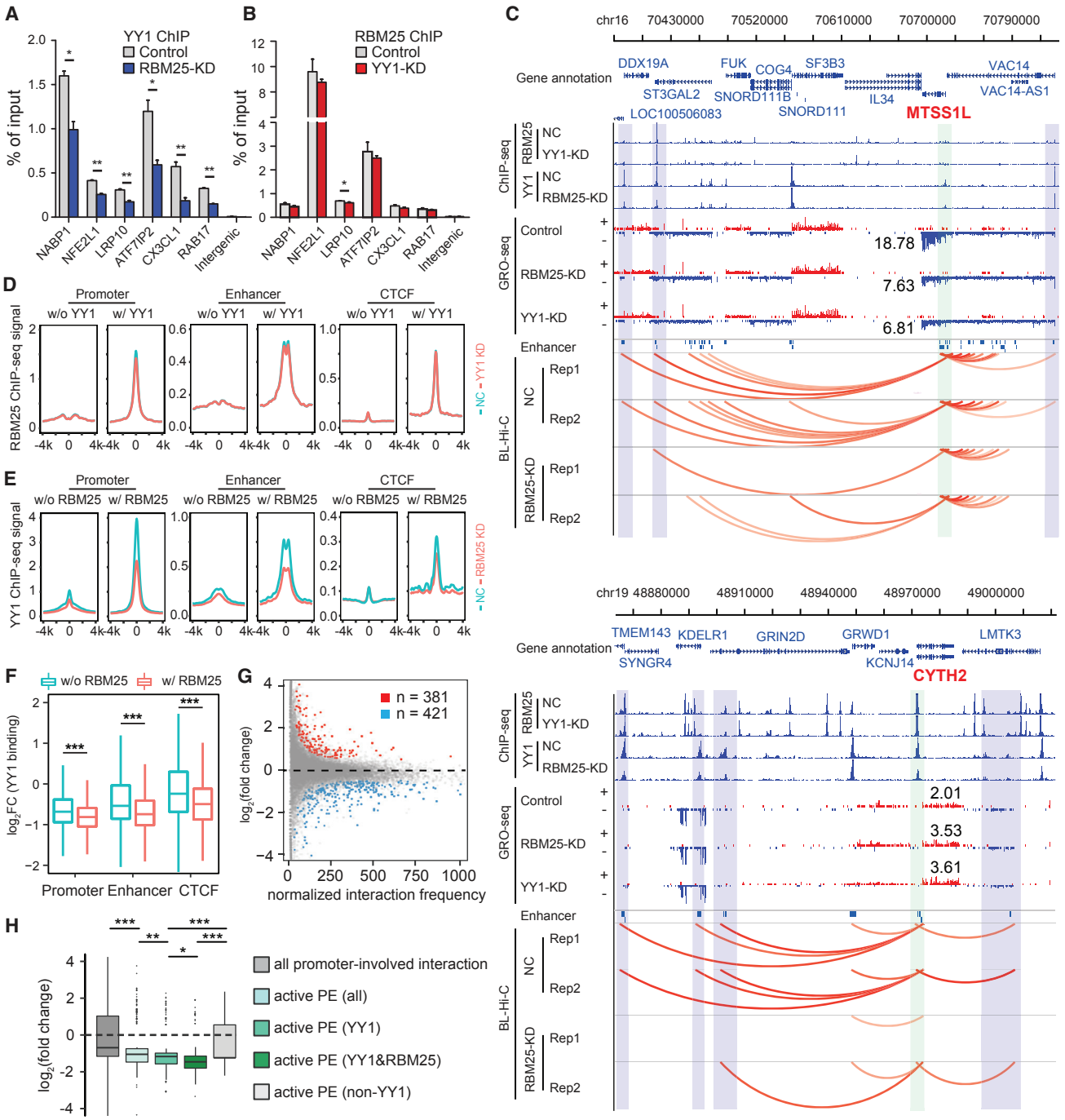


Figure 6. RBM25-Dependent YY1 Binding in the Human Genome

(A and B) ChIP-qPCR analysis of YY1 (A) and RBM25 (B) binding on representative target gene promoters upon knockdown of the other factor.

(C) Representative genomic loci showing down-regulated YY1 binding (ChIP-seq) and YY1-mediated chromatin looping (BL-Hi-C), as well as down (upper)-/up (lower)-regulated gene expression (insert numbers: expression levels measured by GRO-seq under individual conditions) upon knockdown of RBM25. Shaded regions highlight decreased interactions between the target gene promoter (green) and their enhancers (blue).

(D and E) Metagene analysis of RBM25 (D) or YY1 (E) ChIP-seq signals on promoters (left), enhancers (middle), and CTCF-binding sites (right) with or without evidence for YY1 (D) or RBM25 (E) binding and before (blue) or after (red) knockdown of YY1 (D) or RBM25 (E).

(F) Statistical analysis of the impact of RBM25 depletion on YY1 binding on all genomic loci without or with evidence for RBM25 occupancy.

(G) Using all annotated promoters as anchors, the $\log_2$-fold change of normalized PETs frequency upon knockdown of RBM25 is plotted against normalized PETs frequency in control sample. Highlighted are YY1-bound promoters with significantly increased (red) or decreased (blue) interactions.

(legend continued on next page)

as illustrated on typical up- and down-regulated genes (Figure 5E). Comparison between fold-change (FC) in GRO-seq signals revealed a global concordance (Spearman correlation coefficient = 0.51) between YY1 and RBM25 knockdown-induced gene expression (Figure 5F). As knockdown of each factor caused both down- and up-regulation of a large number of genes, we found that, among commonly affected genes, the vast majority of up- and down-regulated genes were affected in the same directions (Figure 5G). Furthermore, the up- and down-regulated genes in response to knockdown of RBM25 or YY1 were both linked to their promoter association, as evidenced by the highly significant odds ratios in all pairwise comparisons (Figure 5H). As a control, we performed similar analysis after knocking down RBM22, an RBP also implicated in transcription but with distinct binding pattern from RBM25, and found little co-binding or minimal numbers of co-regulated genes between YY1 and RBM22, and among commonly affected genes, there were few coordinated changes in the same directions (Figures S5E and S5F). These data provide strong evidence that YY1 and RBM25 function as a complex in regulated gene expression, although it is also clear that both have independent targets, likely due to their additional partnership with other TFs and/or RBPs.

RBM25-Dependent YY1-Chromatin Interactions

To understand the mechanism for coordinated binding and regulation of transcription by YY1 and RBM25, we next asked whether RBM25 binding depends on YY1 or vice versa. We first performed ChIP-qPCR on the same panel of YY1- and RBM25-targeted promoters (see Figure 5A) and found that RBM25 knockdown significantly reduced YY1 binding on all six target genes (Figure 6A), but not the other way around (Figure 6B). These observations indicate that RBM25 acts first on those target promoters, which is required for subsequent YY1 binding. To obtain global evidence for this, we performed ChIP-seq for YY1 before and after depleting RBM25 and ChIP-seq for RBM25 before and after depleting YY1 (Table S6). As exemplified by two genomic loci shown (Figure 6C), it is clear that most, if not all, of YY1-binding events decreased upon RBM25 depletion, but RBM25 ChIP-seq signals remained largely unaltered upon YY1 depletion. Metagene analysis of RBM25 ChIP-seq signals on promoters, enhancers, and CTCF-binding sites demonstrated that RBM25 ChIP-seq signals were essentially insensitive to YY1 depletion (Figure 6D). As expected, the YY1 ChIP-seq signals were more prevalent at RBM25-associated genomic loci, and RBM25 depletion significantly attenuated YY1 binding on all three classes of DNA elements (Figure 6E). Statistical analysis showed that FCs in YY1 binding were more affected on genomic loci occupied by RBM25 than those without evidence for RBM25 association (Figure 6F). These data demonstrated that RBM25 is required for efficient YY1 targeting genome-wide in HepG2 cells.

RBM25 Modulates YY1-Mediated Long-Distance Genomic Interactions

YY1 has recently been found to be a structural regulator of enhancer-promoter loops, which appears to play a larger role than CTCF in such genomic interactions (Weintraub et al., 2017). We thus wondered how RBM25 might modulate such YY1-mediated genomic interactions. To this end, we chose a newly developed Bridge-Linker-Hi-C (BL-Hi-C) technology, an improved version of Hi-C that utilizes a biotinylated linker to bridge HaeIII-cleaved GG/CC sites followed by biotin selection, library construction, and paired-end sequencing (Liang et al., 2017). As GGCC motifs are more frequently associated with promoters and enhancers in mammalian genomes, BL-Hi-C achieves a ~5-fold increase in sensitivity over Hi-C in detecting genomic interactions (Liang et al., 2017). We performed BL-Hi-C on HepG2 cells before and after knockdown of RBM25 and generated duplicated libraries under each condition, obtaining ~180 million raw reads from each library, and used the recommended chromatic interaction analysis by paired-end tag sequencing (ChIA-PET) software (Li et al., 2017a) to identify paired-end tags (PETs) from raw FASTQ data (see Method Details). We found that each library had ~80% valid PETs (Figure S6A), ~80% of which were unique after removing PCR duplicates (Figure S6B). The ratio of intra- versus inter-chromosomal interactions is ~8, indicating a rather robust signal-to-noise ratio. Using these data, we found that the overall chromatin structure was nearly identical, not only between the replicates, which demonstrates the reproducibility of our libraries, but also between mock-treated and siRBM25-treated cells, as indicated by similar A and B compartments (Figure S6C) and overall chromatin interactions matrix (Figure S6D). These data suggest that the three-dimensional (3D) genome remains largely unaltered in RBM25-depleted HepG2 cells.

We next focused on promoter-anchored interactions in the genome. On each annotated promoter, we counted PETs that connect the promoter to all other genomic loci and then calculated the ratio of such promoter-anchored PETs before and after RBM25 knockdown. We found that RBM25-knockdown-induced global changes in promoter-anchored PETs were slightly decreased (Figure 6G), likely reflecting both direct and indirect impacts of RBM25 depletion on transcription. We next used a q-value cutoff ($q < 0.05$) similar to the stringency in processing the HiChIP data from YY1-depleted mESCs (Weintraub et al., 2017) and focused on changes in significantly differential promoter-enhancer interactions. We found that promoter/enhancer (PE)-linked PETs showed similar changes in both directions among all annotated promoters in the human genome (Figure 6H). We then divided these PE-linked PETs into four subclasses: (1) all active promoters, (2) active promoters and enhancers associated with YY1 binding alone, (3) active promoters and enhancers associated with YY1 and RBM25 co-binding, and (4) non-YY1-bound active promoters (see Method Details). We observed that, compared to all promoters and non-YY1 active

(H) Gene promoters are subdivided into four different classes and separately analyzed for their interactions with active enhancers before and after RBM25 knockdown. The distribution of the $\log_2$-fold change of normalized PETs frequency is shown as boxplot (see Method Details). P, promoter; E, enhancer. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (unpaired Student t test).
See also Figure S6 and Tables S5 and S6.

promoters, YY1 and RBM25 co-bound promoters and enhancers showed more reduced than increased PETs in response to RBM25 depletion. Promoters and enhancers associated with YY1 binding alone showed weaker changes compared to co-bound ones (Figure 6H). The reduced PE interactions were further illustrated on the two representative YY1- and RBM25-regulated genes (see bottom BL-Hi-C tracks in Figure 6C). The reduced PE-linked PETs occurred on both up- and down-regulated genes, similar to the observation made in YY1-depleted mESCs (Weintraub et al., 2017), likely because YY1 functions as a transcriptional activator or repressor on different genes. Collectively, these data demonstrated the role of RBM25 in YY1-dependent transcription by modulating YY1 recruitment and YY1-mediated genomic interactions.

DISCUSSION

RBPs are widely known to participate in various co-transcriptional RNA-processing events; however, increasing evidence suggests that specific RBPs may also have roles in transcription. In this report, we embarked on a discovery-driven project to map the chromatin association profiles of a sizable set of nuclear RBPs by ChIP-seq. Despite the fact that we have only surveyed <5% of ~1,500 RBPs in two ENCODE cell lines, we found that these RBPs collectively occupy ~40% of chromatin regions that show at least one biochemical activity (as defined by DNase I hypersensitivity, histone modifications, and transcription) and ~80% of active gene promoters (as defined by H3K4me3 and H3K27ac). By extrapolation, it is possible that RBPs as whole may be involved in nearly all chromatin activities in the human genome. This suggests a general concept that transcription and co-transcriptional RNA processing may not simply be co-incident events in timing but may be more mechanistically integrated, with RBPs playing central roles in such integration.

It is important to emphasize that the RBP-chromatin interaction surveyed by ChIP-seq does not distinguish between direct and indirect binding to DNA. In fact, this also applies to many TFs, as TF-bound gene promoters often do not have sufficient underlying sequence motifs to support direct contacts, and to enhancers via the formation of various "mega-trans" complexes (Liu et al., 2014). RBPs may be part of these mega-trans complexes to connect with regulatory RNAs, as demonstrated with various lncRNAs (Fu, 2014; Rinn and Chang, 2012; Vance and Ponting, 2014). This may explain the increasing number of typical DNA-binding TFs that can also be UV-crosslinked to RNA, as exemplified by CTCF (Kung et al., 2015; Saldaña-Meyer et al., 2014), YY1 (Sigova et al., 2015), and PRC2 (Davidovich et al., 2013; Kaneko et al., 2013).

Comparison between RBP and TF ChIP-seq signals in HepG2 cells revealed a large number of TF and RBP co-occupancies on diverse DNA elements, particularly promoters and enhancers, suggesting their concerted functions at the chromatin levels. To explore the functional interplay between RBPs and TFs, we chose to focus on YY1 to perform detailed functional and mechanistic studies, given the recent unexpected finding that this typical TF appears to actually bind DNA in an RNA-dependent manner (Sigova et al., 2015). Given extensive co-binding between YY1 and RBM25, we detected reciprocal coIP between the two proteins and found that the two proteins regulate a large common set of genes at the level of transcription, where RBM25 appears to interact with target gene promoters first and then recruits YY1 to those promoters, both dependent on ongoing transcription. This functional interplay has been further linked to YY1-mediated genomic interactions. These findings illustrate a general framework to experimentally approach other RBP-TF co-occupancy events we have now uncovered.

A number of questions regarding YY1- and RBM25-regulated gene expression require further investigation. First, it is interesting that the DNA helicase XRCC5 may also participate in YY1- and RBM25-mediated regulation. Second, we currently do not know which type(s) of RNA is involved in the regulation. YY1 has been shown to bind both eRNAs and nascent RNAs transcribed from its target genes (Sigova et al., 2015). It is curious to note that the binding summit of RBM25 is slightly upstream of TSSs, which might result from looped enhancers. However, according to our preliminary crosslinking immunoprecipitation (CLIP) analysis, RBM25 appears to predominantly interact with pre-mRNAs and spliced mRNAs, rather than eRNAs, and as with other RBPs, its RNA-binding profile does not correlate well to its occupancy on DNA. Thus, our current data do not permit us to deduce RNA involved in mediating RBM25-dependent interactions of YY1 on specific target promoters and/or enhancers. Third, we would like to interpret our BL-Hi-C data with caution with respect to the detected DNA-DNA interactions as a cause or consequence of regulated transcription. We clearly detected specific changes in YY1-associated genomic interactions in response to RBM25 depletion, but most changes are rather modest, although comparable to those detected in YY1-depleted cells (Weintraub et al., 2017). In this regard, we note that the deterministic role of CTCF in genomic interactions is also under active debate (Barutcu et al., 2018; Kubo et al., 2017; Zuin et al., 2014). Thus, TFs and RBPs may only be able to modulate specific genomic interactions within the general framework of 3D genome.

The active participation of RBPs in regulated gene expression may be pertinent to an emerging concept of phase separation during gene expression (Hnisz et al., 2017). Compared to TFs, RBPs contain more low-complexity domains known to be instrumental in liquid-liquid phase separation (Lin et al., 2015; Molliex et al., 2015). Although not all RBPs interact with chromatin in an RNA-dependent manner, many do, as we show in the current study. Collectively, we imagine that functional interactions among TFs, RBPs, RNAs, and target DNA segments may be coordinated to form specific zones to establish separate phases for gene activation or repression in the nucleus, which may underlie the formation of specific gene networks and nuclear subdomains that are visible under a microscope.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2019.06.001.

ACKNOWLEDGMENTS

This project was supported by grants from NIH (HG007005, HG04659, and GM131796) to X.-D.F. Additional support was also provided by the National Natural Science Foundation of China (31870820) and the start fund from Wuhan University (413100022) to R.X. J.-Y.C. is a recipient of the NIH K99 award (DK120952).

AUTHOR CONTRIBUTIONS

R.X., J.-Y.C., and X.-D.F. designed the experiments. R.X. performed the majority of the experiments; J.-Y.C. processed and analyzed most genomic data; Z.L., Y.C., and M.Q.Z. performed BL-Hi-C experiments; D.L. contributed to GRO-seq experiments; G.C., Z.J.L., B.Z., X.D., Y.Y., M.S., W.L. and Y.Z. contributed to data analysis; H.L. sequenced all ChIP-seq libraries; X.W. coordinated with data submission to ENCODE; E.L., G.W.Y., C.B.B. and B.R.G. contributed to data interpretation. R.X., J.-Y.C., and X.-D.F. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: August 1, 2018
Revised: March 21, 2019
Accepted: May 31, 2019
Published: June 27, 2019

REFERENCES

Akdemir, K.C., and Chin, L. (2015). HiCPlotter integrates genomic data with interaction matrices. Genome Biol. *16*, 198.

Aso, T., Vasavada, H.A., Kawaguchi, T., Germino, F.J., Ganguly, S., Kitajima, S., Weissman, S.M., and Yasukochi, Y. (1992). Characterization of cDNA for the large subunit of the transcription initiation factor TFIIF. Nature *355*, 461–464.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell *46*, 674–690.

Bao, X., Guo, X., Yin, M., Tariq, M., Lai, Y., Kanwal, S., Zhou, J., Li, N., Lv, Y., Pulido-Quetglas, C., et al. (2018). Capturing the interactome of newly transcribed RNA. Nat. Methods *15*, 213–220.

Barutcu, A.R., Maass, P.G., Lewandowski, J.P., Weiner, C.L., and Rinn, J.L. (2018). A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. Nat. Commun. *9*, 1444.

Benoit Bouvrette, L.P., Cody, N.A.L., Bergalet, J., Lefebvre, F.A., Diot, C., Wang, X., Blanchette, M., and Lécuyer, E. (2018). CeFra-seq reveals broad asymmetric mRNA and noncoding RNA distribution profiles in *Drosophila* and human cells. RNA *24*, 98–113.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. *15*, 163–175.

Bregman, A., Avraham-Kelbert, M., Barkai, O., Duek, L., Guterman, A., and Choder, M. (2011). Promoter elements regulate cytoplasmic mRNA decay. Cell *147*, 1473–1483.

Cassiday, L.A., and Maher, L.J., 3rd. (2002). Having it both ways: transcription factors that bind DNA and RNA. Nucleic Acids Res. *30*, 4118–4126.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393–1406.

Chen, L., Chen, J.Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., et al. (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. Mol. Cell *68*, 745–757.e5.

Consortium, E.P.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, A.R. (1997). Functional association between promoter structure and transcript alternative splicing. Proc Natl Acad Sci USA *94*, 11456–11460.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. Nat. Struct. Mol. Biol. *20*, 1250–1257.

Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K.J., Gooding, A.R., Lee, J.T., and Cech, T.R. (2015). Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. Mol. Cell *57*, 552–558.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Fu, X.D. (2014). Non-coding RNA: a new frontier in regulatory biology. Natl. Sci. Rev. *1*, 190–204.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nat. Rev. Genet. *15*, 829–845.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489*, 91–100.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13–23.

Huang, V., Zheng, J., Qi, Z., Wang, J., Place, R.F., Yu, J., Li, H., and Li, L.C. (2013). Ago1 Interacts with RNA polymerase II and binds to the promoters of actively transcribed genes in human cancer cells. PLoS Genet. *9*, e1003821.

Ji, X., Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C.Y., Xiao, R., Burge, C.B., and Fu, X.D. (2013). SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. Cell *153*, 855–868.

Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1258–1264.

Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat. Biotechnol. *26*, 1351–1359.

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.

Köhler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nat. Rev. Mol. Cell Biol. *8*, 761–773.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

Kubo, N., Ishii, H., Gorkin, D., Meitinger, F., Xiong, X., Fang, R., Liu, T., Ye, Z., Li, B., Dixon, J., et al. (2017). Preservation of Chromatin Organization after Acute Loss of CTCF in Mouse Embryonic Stem Cells. bioRxiv. https://doi.org/10.1101/118737.

Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., del Rosario, B.C., Pinter, S.F., et al. (2015). Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. Mol. Cell *57*, 361–375.

Kuninger, D.T., Izumi, T., Papaconstantinou, J., and Mitra, S. (2002). Human AP-endonuclease 1 and hnRNP-L interact with a nCaRE-like repressor element in the AP-endonuclease 1 promoter. Nucleic Acids Res. *30*, 823–829.

Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1122–1130.

Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. Trends Biochem. Sci. *39*, 170–182.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. *5*, 1752–1779.

Li, G., Chen, Y., Snyder, M.P., and Zhang, M.Q. (2017a). ChIA-PET2: a versatile and flexible pipeline for ChIA-PET data analysis. Nucleic Acids Res. *45*, e4.

Li, Y.E., Xiao, M., Shi, B., Yang, Y.T., Wang, D., Wang, F., Marcia, M., and Lu, Z.J. (2017b). Identification of high-confidence RNA regulatory elements by combinatorial classification of RNA-protein binding sites. Genome Biol. *18*, 169.

Liang, Z., Li, G., Wang, Z., Djekidel, M.N., Li, Y., Qian, M.P., Zhang, M.Q., and Chen, Y. (2017). BL-Hi-C is an efficient and sensitive approach for capturing structural and regulatory chromatin interactions. Nat. Commun. *8*, 1622.

Lin, Y., Protter, D.S., Rosen, M.K., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Mol. Cell *60*, 208–219.

Liu, Z., Merkurjev, D., Yang, F., Li, W., Oh, S., Friedman, M.J., Song, X., Zhang, F., Ma, Q., Ohgi, K.A., et al. (2014). Enhancer activation requires trans-recruitment of a mega transcription factor complex. Cell *159*, 358–373.

Liu, C., Xin, Y., Xu, L., Cai, Z., Xue, Y., Liu, Y., Xie, D., Liu, Y., and Qi, Y. (2018). Arabidopsis ARGONAUTE 1 Binds Chromatin to Promote Gene Transcription in Response to Hormones and Stresses. Dev Cell *44*, 348–361.e7.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Merkenschlager, M., and Odom, D.T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. Cell *152*, 1285–1297.

Michelotti, E.F., Michelotti, G.A., Aronsohn, A.I., and Levens, D. (1996). Heterogeneous nuclear ribonucleoprotein K is a transcription factor. Mol. Cell. Biol. *16*, 2350–2360.

Moldón, A., Malapeira, J., Gabrielli, N., Gogol, M., Gómez-Escoda, B., Ivanova, T., Seidel, C., and Ayté, J. (2008). Promoter-driven splicing regulation in fission yeast. Nature *455*, 997–1000.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. Cell *163*, 123–133.

Moorman, C., Sun, L.V., Wang, J., de Wit, E., Talhout, W., Ward, L.D., Greil, F., Lu, X.J., White, K.P., Bussemaker, H.J., and van Steensel, B. (2006). Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. Proc. Natl. Acad. Sci. USA *103*, 12027–12032.

Oktaba, K., Zhang, W., Lotz, T.S., Jun, D.J., Lemke, S.B., Ng, S.P., Esposito, E., Levine, M., and Hilgers, V. (2015). ELAV links paused Pol II to alternative polyadenylation in the Drosophila nervous system. Mol. Cell *57*, 341–348.

Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. *15*, 234–246.

Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B.A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T.B., et al. (2017). hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. Mol Cell *68*, 955–969.e10.

Puigserver, P., and Spiegelman, B.M. (2003). Peroxisome proliferator-activated receptor-gamma coactivator 1 alpha (PGC-1 alpha): transcriptional coactivator and metabolic regulator. Endocr. Rev. *24*, 78–90.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes Dev. *28*, 723–734.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259.

Shav-Tal, Y., and Zipori, D. (2002). PSF and p54(nrb)/NonO–multi-functional nuclear proteins. FEBS Lett. *531*, 109–114.

Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science *350*, 978–981.

Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature *516*, 436–439.

Sucharov, C.C., Helmke, S.M., Langer, S.J., Perryman, M.B., Bristow, M., and Leinwand, L. (2004). The Ku protein complex interacts with YY1, is up-regulated in human heart failure, and represses alpha myosin heavy-chain gene expression. Mol. Cell. Biol. *24*, 8705–8715.

Sundararaman, B., Zhan, L., Blue, S.M., Stanton, R., Elkins, K., Olson, S., Wei, X., Van Nostrand, E.L., Pratt, G.A., Huelga, S.C., et al. (2016). Resources for the Comprehensive Discovery of Functional RNA Elements. Mol. Cell *61*, 903–913.

Taccioli, G.E., Gottlieb, T.M., Blunt, T., Priestley, A., Demengeot, J., Mizuta, R., Lehmann, A.R., Alt, F.W., Jackson, S.P., and Jeggo, P.A. (1994). Ku80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination. Science *265*, 1442–1445.

Taliaferro, J.M., Aspden, J.L., Bradley, T., Marwha, D., Blanchette, M., and Rio, D.C. (2013). Two new and distinct roles for Drosophila Argonaute-2 in the nucleus: alternative pre-mRNA splicing and transcriptional repression. Genes Dev. *27*, 378–389.

Ting, N.S., Yu, Y., Pohorelic, B., Lees-Miller, S.P., and Beattie, T.L. (2005). Human Ku70/80 interacts directly with hTR, the RNA component of human telomerase. Nucleic Acids Res. *33*, 2090–2098.

Trcek, T., Larson, D.R., Moldón, A., Query, C.C., and Singer, R.H. (2011). Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. Cell *147*, 1484–1497.

Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.-Y., Cody, N.A.L., Dominguez, D., et al. (2018). A Large-Scale Binding and Functional Map of Human RNA Binding Proteins. bioRxiv. https://doi.org/10.1101/179648.

Vance, K.W., and Ponting, C.P. (2014). Transcriptional regulatory functions of nuclear long noncoding RNAs. Trends Genet. *30*, 348–355.

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390–394.

Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T., et al. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. *38*, W214–W220.

Wei, C., Xiao, R., Chen, L., Cui, H., Zhou, Y., Xue, Y., Hu, J., Zhou, B., Tsutsui, T., Qiu, J., et al. (2016). RBFox2 Binds Nascent RNA to Globally Regulate Polycomb Complex 2 Targeting in Mammalian Genomes. Mol. Cell *62*, 875–889.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573–1588.e28.

Xie, D., Boyle, A.P., Wu, L., Zhai, J., Kawli, T., and Snyder, M. (2013). Dynamic trans-acting factor colocalization in human cells. Cell *155*, 713–724.

Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K., Hua, Y., Gueroussov, S., Najafabadi, H.S., Hughes, T.R., et al. (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science *347*, 1254806.

Zid, B.M., and O'Shea, E.K. (2014). Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. Nature *514*, 117–121.

Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van IJcken, W.F., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. USA *111*, 996–1001.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-SRSF1	Bethyl	A302-052A; RRID: AB_1604258
Rabbit polyclonal anti-SRSF3	MBL	RN080PW; RRID: AB_11160964
Rabbit polyclonal anti-SRSF4	Bethyl	A303-670A; RRID: AB_11204752
Rabbit polyclonal anti-SRSF7	MBL	RN079PW; RRID: AB_11161213
Rabbit polyclonal anti-SRSF9	MBL	RN081PW; RRID: AB_11160952
Rabbit polyclonal anti-U2AF1	Bethyl	A302-079A; RRID: AB_1604295
Rabbit polyclonal anti-U2AF2	Bethyl	A303-665A; RRID: AB_11204941
Goat polyclonal anti-HNRNPC	Santa Cruz Biotechnology	sc-10037; RRID: AB_2117316
Rabbit polyclonal anti-HNRNPH1	Aviva	ARP58479; RRID: AB_2615098
Rabbit polyclonal anti-HNRNPK	MBL	RN019P; RRID: AB_1953048
Rabbit polyclonal anti-HNRNPL	Aviva	ARP40368_P050; RRID: AB_2615153
Rabbit polyclonal anti-HNRNPLL	Cell Signaling Technology	4783S; RRID: AB_10547879
Rabbit monoclonal anti-HNRNPUL1	Abcam	ab68480; RRID: AB_2120657
Rabbit polyclonal anti-PTBP1	MBL	RN011P; RRID: AB_1570645
Rabbit polyclonal anti-PCBP1	MBL	RN024P; RRID: AB_1953051
Rabbit polyclonal anti-PCBP2	MBL	RN025P; RRID: AB_1953052
Rabbit polyclonal anti-FUS	Bethyl	A300-294A; RRID: AB_263410
Rabbit monoclonal anti-TAF15	Abcam	ab134916; RRID: AB_2614922
Goat polyclonal anti-SNRNP70	Santa Cruz Biotechnology	sc-9571; RRID: AB_2193707
Rabbit polyclonal anti-PRPF4	MBL	RN093PW; RRID: AB_11161200
Rabbit polyclonal anti-FIP1L1	Abclonal	A5016; RRID: AB_2614927
Rabbit polyclonal anti-RBFOX2	Bethyl	A300-864A; RRID: AB_609476
Rabbit polyclonal anti-RBM22	Bethyl	A303-923A; RRID: AB_2620272
Rabbit polyclonal anti-RBM25	Bethyl	A301-068A; RRID: AB_2175937
Rabbit polyclonal anti-RBM39	Bethyl	A300-291A; RRID: AB_263411
Rabbit monoclonal anti-AGO1	Cell Signaling Technology	5053; RRID: AB_2616013
Mouse monoclonal anti-AGO2	Abnova	H00027161-M01; RRID: AB_565459
Mouse monoclonal anti-SAFB	Zen BioScience	200580; RRID: AB_2616551
Rabbit polyclonal anti-SAFB2	Bethyl	A301-112A; RRID: AB_873125
Mouse monoclonal anti-NONO	Santa Cruz Biotechnology	sc-166702; RRID: AB_2152178
Rabbit polyclonal anti-GTF2F1	GeneTex	GTX114455; RRID: AB_11167162
Rabbit polyclonal anti-POLR2G	GeneTex	GTX108874; RRID: AB_1951315
Rabbit polyclonal anti-SFPQ	Bethyl	A301-321A; RRID: AB_937993
Rabbit polyclonal anti-TARDBP	Bethyl	A303-223A; RRID: AB_10973681
Mouse monoclonal anti-XRCC5	Zen BioScience	201004; RRID: AB_2616549
Mouse monoclonal anti-CCAR2	Zen BioScience	200093; RRID: AB_2636832
Mouse monoclonal anti-CHD3	Zen BioScience	200314-7F6; RRID: AB_2636830
Rabbit monoclonal anti-DROSHA	Cell Signaling Technology	3364; RRID: AB_2238644
Rabbit polyclonal anti-HNRNPD	Cell Signaling Technology	12382; RRID: AB_2616009
Mouse monoclonal anti-HNRNPM	Santa Cruz Biotechnology	sc-20001; RRID: AB_627740
Rabbit polyclonal anti-RBM14	MBL	RN069PW; RRID: AB_11124962
Rabbit polyclonal anti-SRSF10	MBL	RN064PW; RRID: AB_11124967
Mouse monoclonal anti-DDX3X	Zen BioScience	201041; RRID: AB_2616548

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Rabbit monoclonal anti-DDX5	Cell Signaling Technology	9877S; RRID: AB_10891054
Rabbit polyclonal anti-DDX59	Bethyl	A303-028A; RRID: AB_10755233
Rabbit polyclonal anti-DGCR8	Bethyl	A302-468A; RRID: AB_1944223
Rabbit polyclonal anti-DICER1	MBL	RN030PW; RRID: AB_10598025
Mouse monoclonal anti-EEF2	Zen BioScience	200559; RRID: AB_2616545
Rabbit polyclonal anti-EWSR1	Cell Signaling Technology	11910; RRID: AB_2616007
Rabbit polyclonal anti-HNRNPA0	MBL	RN061PW; RRID: AB_10794610
Mouse monoclonal anti-HNRNPA1	Santa Cruz Biotechnology	sc-32301; RRID: AB_627729
Mouse monoclonal anti-HNRNPA2B1	Santa Cruz Biotechnology	sc-32316; RRID: AB_2279639
Rabbit polyclonal anti-KHDRBS1	MBL	RN021P; RRID: AB_1953049
Mouse monoclonal anti-NPM1	Zen BioScience	200143; RRID: AB_2616543
Rabbit polyclonal anti-PRPF8	MBL	RN094PW; RRID: AB_11160957
Rabbit polyclonal anti-SF1	GeneTex	GTX104540; RRID: AB_1951880
Rabbit polyclonal anti-SRSF5	MBL	RN082PW; RRID: AB_11160960
Rabbit polyclonal anti-SYNCRIP	MBL	RN046PW; RRID: AB_10597869
Mouse monoclonal anti-XRCC6	Zen BioScience	200995; RRID: AB_2616544
Rabbit polyclonal anti-HNRNPF	GeneTex	GTX114476; RRID: AB_2037186
Rabbit polyclonal anti-RBM15	Bethyl	A300-821A; RRID: AB_2253435
Rabbit polyclonal anti-RBM17	Bethyl	A302-498A; RRID: AB_1966059
Rabbit polyclonal anti-RBM34	Bethyl	A302-293A; RRID: AB_1850203
Mouse monoclonal anti-YY1	Proteintech	66281-1-Ig; RRID: AB_2737053
Mouse monoclonal anti-GAPDH	Abmart	M20006; RRID: AB_2737054
Mouse monoclonal anti-β-actin	Sigma	A5441; RRID: AB_476744
Normal mouse IgG	Santa Cruz Biotechnology	sc-2025; RRID: AB_737182
Normal rabbit IgG	Cell Signaling Technology	2729; RRID: AB_2617119
Chemicals, Peptides, and Recombinant Proteins		
Protease Inhibitor Cocktail	Roche	11873580001
Protein A/G magnetic beads	Thermo	88802
Glycogen	Thermo	R0561
BSA	NEB	B9000S
tRNA	Thermo	15401011
Formaldehyde	Sigma-Aldrich	F8775-25ml
Proteinase K	Thermo	EO0492
RNase A	Thermo	EN0531
T4 DNA Polymerase	NEB	M0203S
Klenow DNA Polymerase	NEB	M0210S
T4 PNK	NEB	M0201S
Klenow Fragment (3′to 5′ exo-)	NEB	M0212S
T4 DNA Ligase	NEB	M0202M
Phusion DNA Polymerase	Thermo	F530L
dATP Solution (100 mM)	Thermo	R0141
DTT	Thermo	15508013
Lipofectamine RNAiMAX	Thermo	13778150
5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside (DRB)	Sigma-Aldrich	D1916-50MG
RNaseOUT	Thermo	10777019
Br-UTP	Sigma-Aldrich	B7166-5MG
ATP	Thermo	R0441

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GTP	Thermo	R0461
CTP	Thermo	R0451
Trizol LS	Thermo	10296010
30% sarkosyl	sigma	61747
Acid-Phenol:Chloroform, pH 4.5	Thermo	AM9722
RQ1 DNase	Promega	M6101
Antarctic Phosphatase	NEB	M0289S
BrdU antibody conjugated agarose beads	Santa Cruz	sc-32323 AC
E. coli Poly(A) Polymerase	NEB	M0276S
SuperScript III First-Strand Synthesis System	Thermo Fisher	18080051
Exonuclease I	NEB	M0293S
Circligase II ssDNA ligase	Epicenter	CL9021K
APE 1	NEB	M0282S
SYBR Gold Nucleic Acid Gel Stain	Thermo	S-11494
HaeIII	NEB	R0108L
Dynabeads M-280 Streptavidin	Thermo	11205D
I-Block Protein-Based Blocking Reagent	Thermo	T2015
Salmom Sperm DNA solution	Thermo	15632011
Quick Ligation Kit	NEB	M2200L
Lambda Exonuclease	NEB	M0262L
Q5 Hot Start DNA Polymerase	NEB	M0493L
AMPure XP	Beckman	A63881
Critical Commercial Assays		
MinElute PCR Purification Kit	QIAGEN	28004
MinElute Gel Extraction Kit	QIAGEN	28604
SuperSignal West Pico PLUS Chemiluminescent Substrate	Thermo	34577
Deposited Data		
Original gel imaging data	This study	https://doi.org/10.17632/svg4vyf2ry.1
Sequencing data of ChIP-seq experiments	This study	GEO: GSE120104
Sequencing data of GRO-seq experiments	This study	GEO: GSE120105
Sequencing data of BL-Hi-C experiments	This study	GEO: GSE120023
Experimental Models: Cell Lines		
Human: HepG2 cells	ATCC	HB-8065
Human: K562 Cells	ATCC	CCL-243
Oligonucleotides		
siRNAs for Knocking down experiments	This study	See Table S3
Primers for ChIP-qPCR experiments	This study	See Table S5
Software and Algorithms		
BWA	Li and Durbin, 2009	https://www.encodeproject.org/software/bwa/
AQUAS	Anshul Kundaje lab	https://github.com/kundajelab/chipseq_pipeline
SPP	Kharchenko et al., 2008	https://www.encodeproject.org/software/spp/
IDR	Li et al., 2011	https://www.encodeproject.org/software/idr/
HOMER	Heinz et al., 2010	http://homer.ucsd.edu/homer/ngs/ucsc.html
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools	Li et al., 2009	http://samtools.sourceforge.net/
Bedtools	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R	N/A	https://www.r-project.org/
Circos	Krzywinski et al., 2009	http://circos.ca
DEseq2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
RBPgroup	Li et al., 2017b	https://github.com/lulab/RBPgroup
ChIA-PET2 (version 0.9.2)	Li et al., 2017a	https://github.com/GuipengLi/ChIA-PET2
HiCPlotter	Akdemir and Chin, 2015	https://github.com/kcakdemir/HiCPlotter
HiC-Pro	Servant et al., 2015	https://github.com/nservant/HiC-Pro.

LEAD CONTACT AND MATERIALS AVAILABILITY

Please direct any requests for further information and resources to the Lead Contact, Xiang-Dong Fu (xdfu@ucsd.edu), Department of Cellular and Molecular Medicine, University of California, San Diego.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines and Culture Conditions

HepG2 (HB-805) cells derived from a 15-year-old male and K562 (CCL-243) cells from a 53-year-old female were both from ATCC and grown in accordance with ENCODE cell culture protocols. Briefly, HepG2 cells were cultured in MEM with 2mM L-glutamine (Cellgro, 10-010-CM) supplemented with 10% fetal bovine serum (Omega Scientific, FB-01), 1X non-essential amino acids (Cellgro, 25-025-CI), 100 U/mL penicillin, 100 μg/mL streptomycin (Cellgro, 30-002-CI), 1 mM sodium pyruvate (Cellgro, 25-000-CI), and 1.5g/L sodium bicarbonate (Cellgro, 25-035-CI). K562 cells were grown in RPMI 1640 with 2mM L-glutamine (Cellgro, 10-040-CM) supplemented with 10% fetal bovine serum, 100 U/mL penicillin, 100 μg/mL streptomycin and 1 mM sodium pyruvate.

Antibodies

Antibodies against RBPs have been extensively characterized with for their high efficiency in immunoprecipitation and specificity in response to specific RBP knockdown according to the ENCODE RBP antibody characterization standard (Key Resources Table**,** see also Table S1 for antibody ID number, which can be used to search the ENCODE website for the information on validating these antibodies.). Anti-YY1 from Proteintech (66281-1-Ig) was used for western blot, co-immunoprecipitation and chromatin immunoprecipitation. Anti-β-actin (Sigma, A5441) and anti-GAPDH (Proteintech, 60004-1-Ig) were used as loading control.

METHOD DETAILS

ChIP-seq Library Construction

RBP ChIP-seq library construction and data analysis were performed as previously described (Van Nostrand et al., 2018). Briefly, 1 to 2 × 10^7 HepG2 and K562 cells were crosslinked in 1% formaldehyde diluted in PBS for 20 min and then quenched with glycine. Cell nuclei were isolated, re-suspended in nuclear lysis buffer, and sonicated with Branson Sonifier cell disruptor for ~7 cycles, each with 10 s sonication at 40% output power followed by resting on ice for 1 min. Sheared chromatin fractions were examined on agarose gel to ensure the size range from 100–600 bp and 95% of nuclear lysate diluted in the final concentration of 1% Triton X-100, 0.1% sodium deoxycholate and 1X proteinase inhibitor cocktail was subjected to immunoprecipitation with antibody-coupled beads and 5% of nuclear lysate was saved for constructing input control library. After decrosslinking, RNase A digestion and proteinase K treatment, recovered DNA was used for library construction according to the instruction of the Illumina Preparing Samples for ChIP Sequencing. Each library was barcoded for pooled sequencing. DNA Libraries between 200-400 bp were gel purified, quantified with Qubit and subjected to Illumina HiSeq-2000/2500 for sequencing.

Identification and Annotations of RBP ChIP-seq Peaks

RBP ChIP-seq data were processed in accordance with ENCODE uniform transcription factor ChIP-seq pipeline (https://www.encodeproject.org/chip-seq/transcription_factor/) and using GRCh37 as the reference human genome. All datasets containing > 10 million usable reads from each replicate that passed IDR cutoff and generated > 200 peaks were used for final analysis. Annotated genome regions were according to GENCODE v19 gene annotation. DNaseI HS data, histone modification profiles generated by ENCODE/Broad Institute, and combined chromatin state segmentation by ChromHMM and Segway are available on the UCSC genome browser. The overlap between specific chromatin features and RBP ChIP-seq peaks was calculated using Bedtools (Quinlan and Hall, 2010). Quantitative analysis (Figure S1C) of the association between histone modifications and RBP ChIP-seq signals and

Circos visualization of the date (Krzywinski et al., 2009) (Figure 1C) were based on the sum of RBP ChIP-seq signals in each 2Mb interval in the human genome. Composite RBP ChIP-seq signals were determined and scaled around all TSSs for heatmap visualization (Figure 2D). HNRNPH1 is not included for data visualization on K562 cells due to scattered signals. A union set of ChIP-seq peaks was first determined as a reference for any pairwise comparison by merging all overlapped peaks (Figures 4E, S4C and S5E). The Homer software (Heinz et al., 2010) was used to search for YY1 DNA binding motifs within ± 1 kb promoter regions (Figure 4E).

Promoter Classification

Classification of bivalent, K4-only, K27-only, and none-K4-K27 promoters was as previously described (Wei et al., 2016). H3K27me3 and H3K4me3 ChIP-seq peaks were merged into clusters separately (Table S2), if peaks were within the window of 2 Kb. Bivalent promoters were defined as those at TSS ± 2Kb with at least 500 bp overlap with both H3K37me3 and H3K4me3 peaks. K4-only or K27-only promoters were those without substantial overlap between H3K4me3 and H3K27me3 peaks. Other promoters were defined as none-K4-K27 promoters if they lacked either of the two histone modification marks. Similarly, classification of promoters into CpG promoters or non-CpG promoters was based on the overlap or lack of overlap with CpG islands defined on the UCSC genome browser. Any promoter with 200 bp overlap with CpG islands was considered as a CpG promoter, and otherwise, non-CpG promoter. In terms of functional classification, promoters of protein-coding genes, small RNA genes and all other non-coding RNA genes were defined based on the biotype determined by GENCODE gene annotation (v19).

Cell Type Conservation of RBP-Chromatin Interactions

Genome regions were partitioned into genic and intergenic regions based on the GENCODE annotation (v19). For genic regions, the expression level of each gene was determined based on the RNA-seq data generated on HepG2 and K562 cells, as described (Van Nostrand et al., 2018). Genes were considered being expressed with matched levels between these two cell lines only when the percentile ranks of expression in two cell lines were comparable (rank difference $\leq$ 0.05). Genes with matched expression levels were grouped into non-expressed genes (RPKM = 0 in both cell lines) or expressed genes (RPKM > 0), with the latter further divided into three groups consisting of equal number of genes, i.e., lowly-expressed, intermediately expressed, and highly expressed genes. Jaccard Index (which is measured as ratio of intersection over union of occupancy of individual RBPs between two cell types) was then determined as a proxy of cross-cell-type conservation of RBP binding group by group.

GRO-seq and Data Analysis

GRO-seq was performed as described previously (Wei et al., 2016). Briefly, ~10^7 cells were harvested on ice. The nuclei were isolated with hypotonic buffer containing 0.5% NP-40 and used for the *in vitro* nuclear run-on reaction in the present of Br-UTP (Sigma). RNA was extracted with Trizol LS and nascent RNAs labeled with Br-UTP were purified on anti-BrdU agarose beads (Santa Cruz Biotechnology, sc-32323 AC). Purified nascent RNAs were treated with T4 polynucleotide kinase to remove phosphoryl group at 3′ end and add phoshoryl group at the 5′ end. After polyA-tailling, RNA was reverse transcribed; cDNA from a size range of 150-400 nt was PAGE-purified, and circularized by CircLigase II (epicenter, CL9021K). Circular cDNA was re-linearized with APE 1 (NEB) and amplified by PCR. The PCR products from 175-250 bp in size were PAGE-purified and sequenced on Illumina HiSeq-2500.

GRO-seq reads were mapped on the human reference genome using the Bowtie2 local model (Langmead and Salzberg, 2012), and non-redundant reads were determined and kept for downstream analysis by using Samtools (Li et al., 2009). The longest transcript was selected as the representative of each gene for expression quantification. The expression level of each gene was calculated as RPKM in the region from +300 bp of TSS to gene end. DEseq2 package (Love et al., 2014) was used to search for differentially expressed genes (DEGs), which was defined with the following cutoff: adjusted *p*-value of $\leq$ 0.05 and fold change of $\leq$ 2/3 or $\geq$ 3/2. To determine whether DEGs after knockdown of a specific RBP are associated with such RBP association with the corresponding promoters, equal number of genes with matched expression level either with or without RBP ChIP-seq signals was randomly sampled. Fisher's exact tests were then performed to determine the significant dependence between RBP-promoter interaction and differential expression. A significant dependence was defined when the null hypothesis was rejected at the level of 0.05 in at least 95 out of 100 times during random sampling.

Correlation Analysis of Nuclear Retention and RBP-Promoter Interaction

Taking advantage of the published fractional RNA-seq data (Benoit Bouvrette et al., 2018), we simply defined the nuclear retention index as the ratio of RPKM in the nuclear fraction over the sum of RPKM in all fractions for each gene. In search for RBP whose promoter occupation is tightly associated with nuclear retention, we built a regression model by using the random forest algorithm. Nuclear retention index was taken as response variable, while the biotype of the gene and whether the gene was associated with RBP in promoter regions as independent variables. The model accuracy was assessed by leave-one-out cross validation in terms of Pearson correlation coefficient between the actual response variable and the predicted value.

NMF Analysis

A unifying set of *cis*-acting regulatory elements (CREs) was defined based on ChIP-seq peaks for 54 TFs (Table S2) and 30 RBPs in HepG2 cells. ChIP-seq peaks with the distance from their summits $\leq$ 1 Kb were merged into one CRE by Bedtools. A matrix was then built, with 1/0 representing whether a CRE (row) was occupied or not by each TF/RBP (column). CREs occupied by multiple factors

were then subject to NMF analysis as described (Li et al., 2017b). First, the optimal rank was estimated in a given range from 2 to 35. The rank of 17 was selected due to the higher score for both cophenetic correlation coefficient (CPCC) and dispersion coefficient (see Figure S4A). Then, NMF was run with the selected rank (-r 17 –n 100 –m KL –p 50). A consensus merge of the segmentations produced by the ChromHMM and Segway from the UCSC genome browser, and protein-protein physical interaction information from GeneMANIA (http://genemania.org) were used to annotate the CRE elements, and the resulting TF/RBP groups by NMF. Each CRE was classified into one of the 17 NMF groups, each exhibiting the highest probability in recognizing the CRE than any other factor group (Figure 4B). CREs with occupation by more than half TF/RBP ($\geq$42) were defined as HOT regions.

Co-immunoprecipitation and western blot

Approximately 10^7 HepG2 cells were lysed in cold co-IP buffer (20 mM Tris-HCl pH8.0, 2mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.1% SDS and 1X proteinase inhibitor cocktail (Roche)) for 30 min at 4°C with rotation. After coupling with antibodies according to the ChIP protocol, beads were incubated with cleared-up cell lysates overnight at 4°C with rotation, washed 3 times with co-IP buffer and eluted with 10 mM DTT in TE buffer for 30 min at 37°C with shaking in ThermoMixer C (Eppendorf). Co-IP with RNase A treatment was performed as previously described (Ji et al., 2013). Total protein or co-IPed samples were resolved by 10% SDS-PAGE and transferred to PVDF membrane (Bio-Rad). The membrane was blocked with 5% nonfat milk for 1 h at room temperature, incubated with primary antibody (1:5000 dilutions of anti-RBM25 and anti-YY1, 1:10000 dilutions of anti-β-actin and anti-GAPDH) for 1 h at room temperature, and after wash several times, the blot was developed with HRP-conjugated secondary antibodies (Cell Signaling Technology) for 1 h at room temperature. Immunoblot signals were detected by autoradiography after application of SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo).

RBP Knockdown with siRNA

The siRNA duplex sequences were listed in Table S3. siRNAs against PCBP2, SRSF4, HNRNPH1, U2AF1, U2AF2, PTBP1, RBM22, AGO1, AGO2, XRCC5, HNRNPL, HNRNPK, HNRNPLL and RBM25 were purchased from Bioneer and siRNA against YY1 was synthesized by GenePharma. siRNAs were transfected into cells with Lipofectamine RNAiMAX (Life Technology) using the reverse transfection protocol. Western blot and functional studies were performed 72 h after transfection.

DRB treatment and ChIP-qPCR

DRB treatment was performed as described previously (Sigova et al., 2015). Briefly, ChIP-qPCR was performed after treatment of cells with 100 μM DRB for 30 min (DRB treatment) and for 3 h followed by two quick rinsing and replacing fresh media without DRB for 30 min (DRB removal). The primers used for ChIP-qPCR are listed in Table S5.

BL-Hi-C Library Construction and Data Analysis

BL-Hi-C library construction and data analysis were performed as described previously (Liang et al., 2017). Approximately, 5×10^6 HepG2 cells were chemically crosslinked by addition of 1/36 volume of fresh 37% formaldehyde solution to the medium and incubation for 10 min at room temperature with gentle shaking. Crosslinking was stopped by adding 2.5 M glycine to a final concentration of 0.2 M and incubating for 10 min at room temperature. After rinsing twice with PBS, cells were harvested in a 1.5 mL tube by scraping and centrifugation, and stored at −80°C until use. Cell pellets were resuspended in 1 mL BL-Hi-C Lysis Buffer 1 (50 mM HEPES-KOH pH7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 1X protease inhibitor cocktail) and incubated for 15 min on ice. After centrifugation at 1500 g for 5 min at 4°C, supernatant was removed, and after washing once with 1 mL BL-Hi-C Lysis Buffer 1, cell pellet was resuspended in 1 mL BL-Hi-C Lysis Buffer 2 (50 mM HEPES-KOH pH7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 1% SDS, 1X protease inhibitor cocktail) and rotated for 15 min at 4°C. Supernatant was removed after centrifugation. Cell pellet was washed once with 1 mL BL-Hi-C Lysis Buffer 1, resuspended in 50 μL of 0.5% SDS and incubated for 5 min at 62°C. At the end of incubation, SDS was quenched by adding 145 μL of ddH_2O and 25 μL of 10% Triton X-100 and incubation for 10 min at 37°C. Chromatin was cleaved by adding 25 μL of 10x NEBuffer 2 and 100 U HaeIII (NEB) for 2 h at 37°C followed by additional cleavage with another 30 U HaeIII for 3 h. Cleaved chromatin was A-tailed by adding 2.5 μL of 10 mM dATP solution (Thermo) and 2.5 μL Klenow Fragment (3′→5′ exo-) (NEB) with rotation at 900 rpm for 40 min at 37°C in ThermoMixer C. Proximity ligation was performed by adding 750 μL ddH_2O, 120 μL of 10X T4 DNA ligase buffer, 100 μL of 10% Triton X-100, 5 μL T4 DNA ligase (NEB) and 4 μL of 200 ng/μl biotinylated Bridge Linker S2 (annealed by /5Phos/CGCGATATC/iBIOdT/TATCTGACT and /5Phos/GTCAGATAAGATATCGCGT) with rotation for 4 h at 16°C. After centrifugation at 3500 g for 5 min at 4°C and removal of supernatant, the pellet was resuspended in 309 μL ddH_2O, 35 μL of Lambda Exonuclease buffer, 3 μL Lambda Exonuclease, 3 μL Exonuclease I and rocked at 900 rpm for 1 h at 37°C in the ThermoMixer C. After the addition of 45 μL of 10% SDS and 55 μL of 20 mg/mL Proteinase K, sample was incubated overnight at 55°C, and added in the next day with 65 μL of 5 M NaCl and incubated for 2 h at 68°C. DNA was recovered by extraction with 500 μL of phenol:chloroform:isoamyl alcohol (25:24:1) and ethanol precipitation with 1 μL glycoblue. After centrifugation and washing with 75% ethanol, DNA pellet was resolved in 130 μL Buffer EB (10 mM Tris-HCl pH8.0). DNA was sheared with Covaris S220 with the setting for DNA size of 300 bp. After washing twice with 2X B&W buffer (10mM Tris-HCl pH7.5, 1mM EDTA, 2 M NaCl), 30 μL Dynabeads M-280 Streptavidin (Thermo) were blocked with 100 μL of 1X I-Block buffer (2% I-Block Protein-Based Blocking Reagent (Thermo), 0.5% SDS) for 45 min at room temperature in a rotating wheel. Beads were washed twice with 100 μL of 1X B&W buffer (5 mM Tris-HCl pH7.5,

0.5 mM EDTA, 1 M NaCl) followed by resuspension in 1X B&W buffer containing 1 μg of pre-heated Salmon Sperm DNA solution and rotation for 30 min at room temperature. After washing twice with 100 μL of 1X B&W buffer, beads were resuspended with 130 μL of 2X B&W buffer, combined with sonicated DNA and rotated for another 45 min at room temperature. Beads were washed five times with 500 μL of 2X SSC containing 0.5% SDS, twice with 500 μL of 1X B&W buffer and once with 100 μL Buffer EB (QIAGEN). DNA on beads was end-repaired in the reaction containing 75 μL ddH_2O, 10 μL of 10X T4 DNA ligase buffer, 5 μL of 10 mM dNTP, 5 μL of T4 Polynucleotide Kinase, 4 μL of T4 DNA Polymerase and 1 μL of Large (Klenow) Fragment with shaking at 900 rpm for 30 min at 37°C in Thermomixer C. After washing twice with 500 μL of 1X TWB (5 mM Tris-HCl pH7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween 20) for 2 min at 55°C, DNA on beads was A-tailed in the reaction containing 80 μL of ddH_2O, 10 μL of NEBuffer 2, 5 μL of 10 mM dATP and 5 μL of Klenow Fragment (3′ → 5′ exo-) with shaking at 900 rpm for 30 min at 37°C in ThermoMixer C. Beads were washed twice with 500 μL 1X TWB for 2 min at 55°C and once with 50 μL of 1X Quick Ligase buffer. DNA on beads was ligated with adaptor in the reaction containing 6.6 μL ddH_2O, 10 μL of 2X Quick Ligase buffer, 2 μL of Quick ligase (NEB) and 0.4 μL of 20 μM Y-Adaptor (annealed by /5Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCAC and TACACTCTTTCCCTACACGACGCTCTTCCGATCT) for 45 min at room temperature. Beads were washed twice with 500 μL 1X TWB for 2 min at 55°C and once with 100 μL of Buffer EB. After resuspending in 60 μL of Buffer EB, bead suspension was aliquoted in 3X 20 μL for storage at −20°C. One aliquot of bead suspension was used as a template for PCR amplification with Q5 Hot Star DNA Polymerase, universal primer (AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACGAC) and index primer (CAAGCAGAAGACGGCATACGAGAT/index/GTGACTGGAGTTCAGACGTGT) for 10 cycles. PCR products between 300-700 bp were purified as BL-Hi-C library using Ampure XP beads and subjected to Illumina HiSeq X-10 (Annoroad Gene Technology at Beijing) for sequencing.

The ChIA-PET2 v0.9.2 software (Li et al., 2017a) was used for quality control and identification of chromatin interactions with the following parameter setting: -A ACGCGATATCTTATC -B AGTCAGATAAGATAT -s 1 -m 1 -t 4 -k 2 -e 1 -l 15 -S 500 -M ''-q 0.05.'' PCA analysis is then applied to 40-kb resolution interaction matrix generated by HiC-Pro (Servant et al., 2015), and regions of continuous positive or negative PC1 values were used for the identification of A or B compartments (Heinz et al., 2010). The interaction matrix was visualized by HiCPlotter (Akdemir and Chin, 2015). High confident interactions were defined as those with q-value < 0.05 for downstream analysis. To perform promoter-centered analysis, for each promoter (±2 kb TSS regions), PETs with one end overlapped with the promoter were counted in each condition, and normalized and tested by DESeq2 (Love et al., 2014). Fold change was then plotted against normalized PETs counts in control condition (Figure 6F). To specifically examine whether the YY1-mediated promoter-enhancer interaction is changed upon knockdown of RBM25, we then focused only on active promoters and enhancers mediated interactions, where active promoters and enhancers are defined as those overlapped with H3K27ac ChIP-seq peaks. Three subsets of interactions were further defined based on presence of YY1/RBM25 chromatin binding: i) YY1 at both ends of the interaction, ii) both RBM25 and YY1 at both ends, iii) YY1 at neither end. Promoters with significantly differential promoter-enhancer interaction before and after knockdown of RBM25 were then defined using DESeq2 (Love et al., 2014) at the cutoff of q-value < 0.05, with their fold changes of interaction frequency displayed as boxplots (Figure 6G).

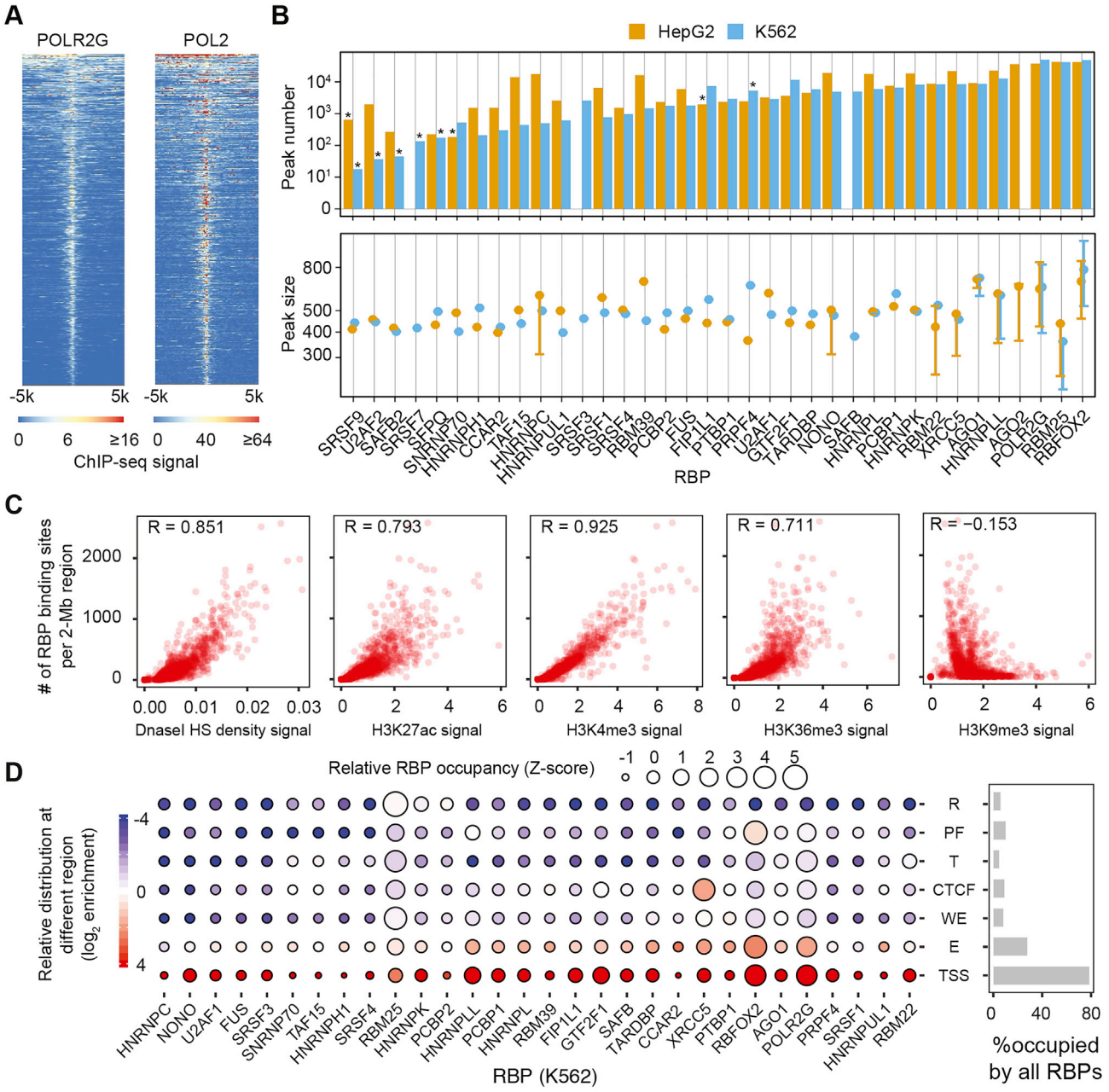
QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical parameters were reported either in individual figures or corresponding figure legends. Quantification data are in general presented as bar/line plots, with the error bar representing mean ± SEM or SD, or boxplot, showing the median (middle line), first and third quartiles (box boundaries), and furthest observation or 1.5 times of the interquartile (end of whisker). All statistical analyses were done in R. Whenever asterisks are used to indicate statistical significance, *stands for $p < 0.05$; **$p < 0.01$, and ***$p < 0.001$.

DATA AND CODE AVAILABILITY

The accession numbers for the raw data FASTQ files and processed BigWig files for all sequencing data deposited in NCBI GEO are GEO: GSE120104, GSE120105 and GSE120023. Original gel imaging data can be accessed from Mendeley (https://doi.org/10.17632/svg4vyf2ry.1).

Cell



(legend on next page)

Figure S1. Additional Characterization of RBP ChIP-Seq Data, Related to Figure 1

(A) Comparison between ChIP-seq profiles of the RNA binding RNAPII subunit POLR2G generated in this study and the existing ENCODE RNAPII. POLR2G ChIP-seq peaks are ordered based on signal intensity in the ± 5 kb window and aligned at peak center.

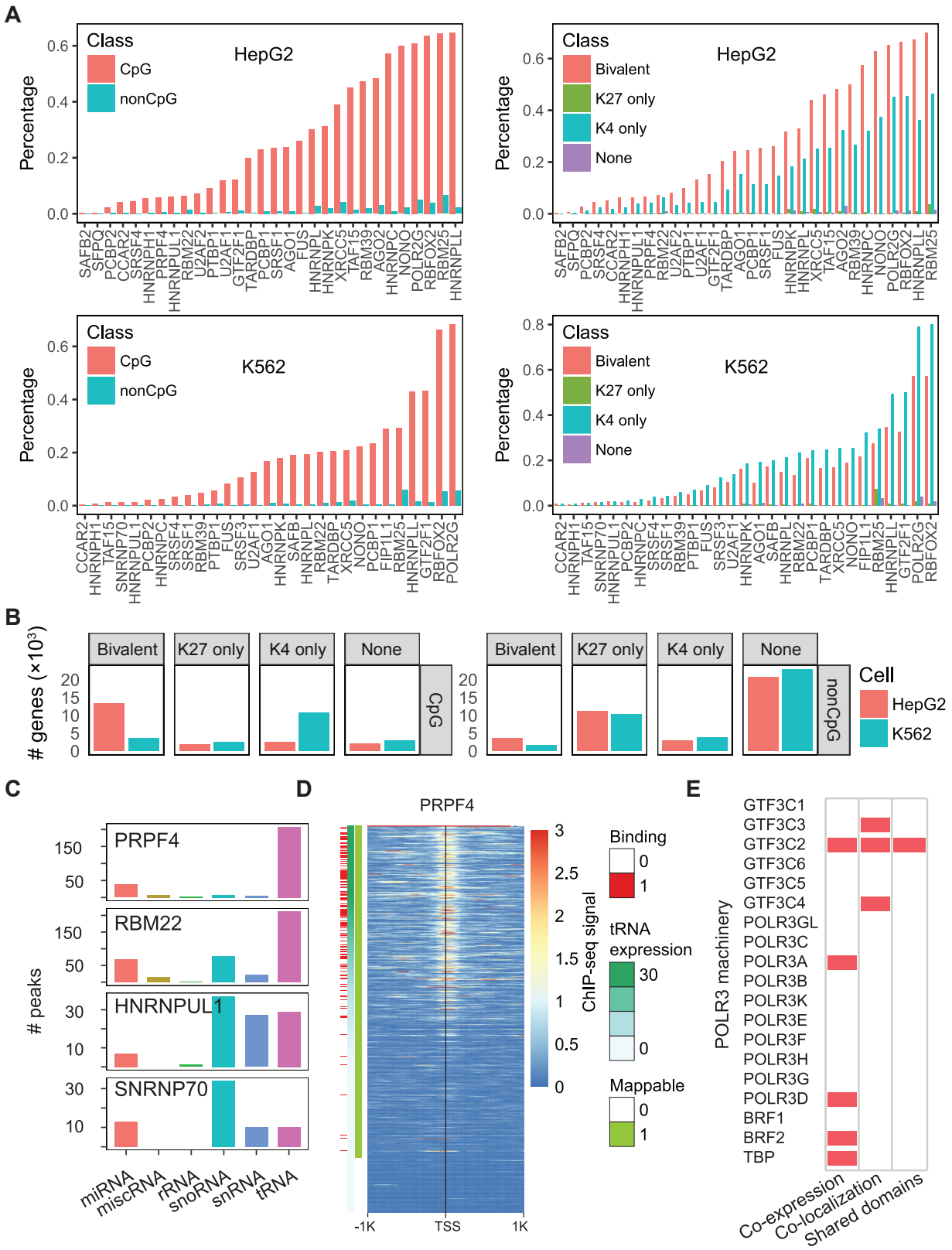
(B) RBP ChIP-seq peak number (top) and size range (bottom) in HepG2 (orange) and K562 (blue) cells. *indicates RBPs that fell short of the ENCODE cutoff, which correspond to the light blue group in Figure 1A.

(C) Correlation between RBP occupancy and specific histone modifications as indicated.

(D) Relative RBP ChIP-seq signals in different genome segmentations in K562 cells, as defined in Figure 1E.

(E) The accumulative coverage of gene promoters by surveyed RBPs in HepG2 (left) and K562 (right) cells. Pink line: The data with total RBPs surveyed that include previously annotated TFs (Class iii described in the legend for Figure 1A); Blue line: The data with total surveyed RBPs excluding the previously annotated TFs.

(F) Distribution of simulated versus observed RBP co-binding events. The simulated events are based the numbers of co-localized RBPs counted and the frequencies calculated, assuming that RBP binding sites are randomly distributed among DNase I hypersensitive sites. The real data are displayed in the same way for comparison.



(legend on next page)

Figure S2. RBP-Chromatin Interaction Preferences at Gene Promoters, Related to Figure 2

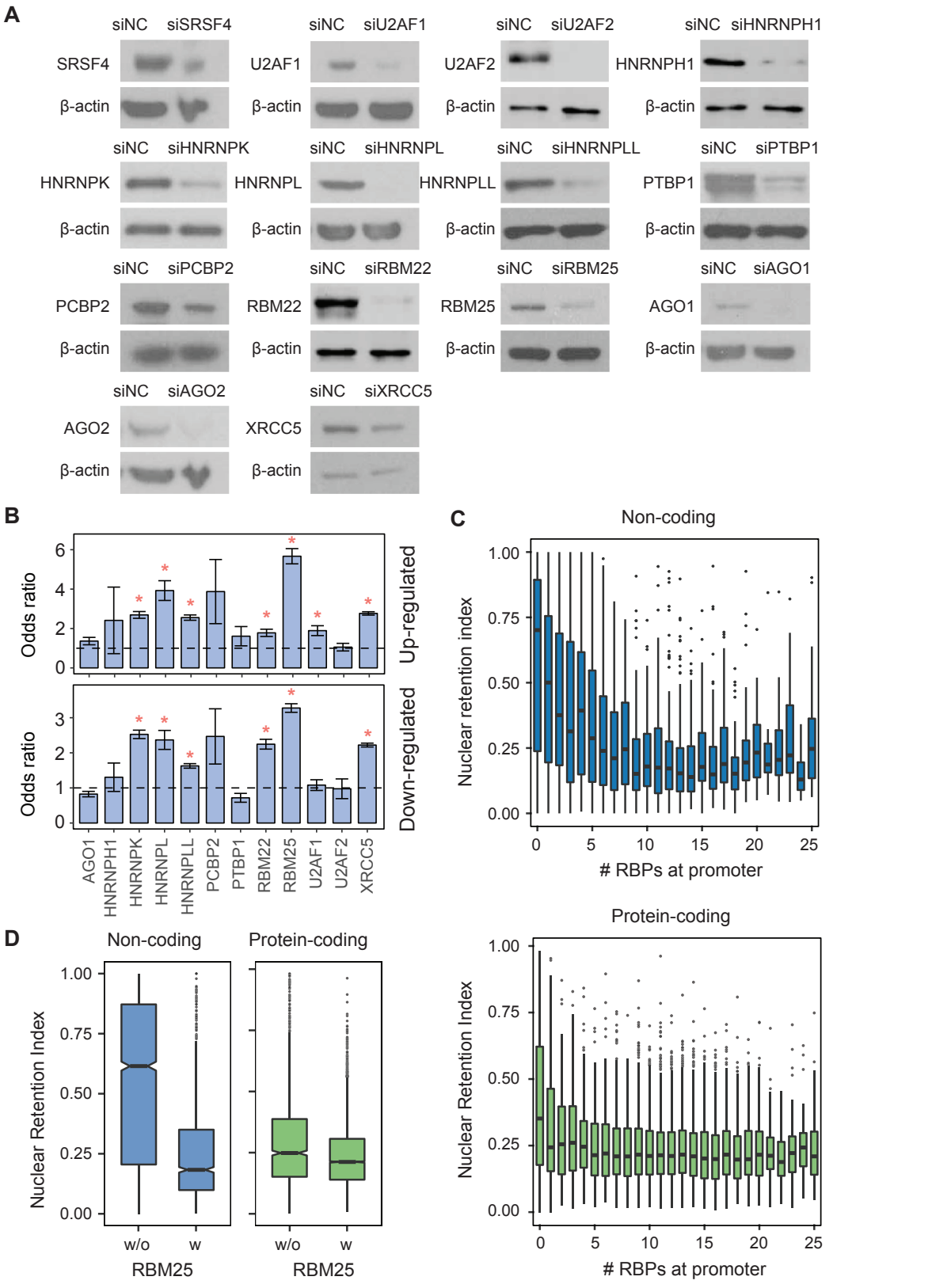
(A) Individual RBP's coverage of different promoter subgroups (as described in the legend for Figure 2A) in HepG2 and K562 cells.

(B) Number of genes in different subgroups collectively occupied by the surveyed RBPs in HepG2 and K562 cells.

(C) Peak number of different small RNA gene promoters by each of the four representative RBPs.

(D) The PRPF4 ChIP-seq profile at tRNA genes. The tRNA genes were first ordered by uniquely mapped reads within the −1 Kb to +1 Kb interval relative to their TSSs followed by their expression levels based on the GRO-seq data in HepG2 cells.

(E) Functional and structure features of PRPF4 in comparison with specific RNAPIII subunits.



(legend on next page)

Figure S3. Gene Expression Related to RBP Promoter Binding, Related to Figure 3

(A) Western blot analysis of 14 RBPs before and after knockdown, with β-actin as control.

(B) Odds ratio of RBP promoter occupancy in relationship to transcriptionally up- or down-regulated genes determined by GRO-seq in response to depletion of individual RBPs.

(C) The distribution of nuclear retention index for each group of non-coding genes (top) and coding genes (bottom) whose promoters were occupied with different numbers of RBPs.

(D) The distribution of RNA nuclear retention for coding and non-coding genes with or without evidence for the association of RBM25 with their promoters.

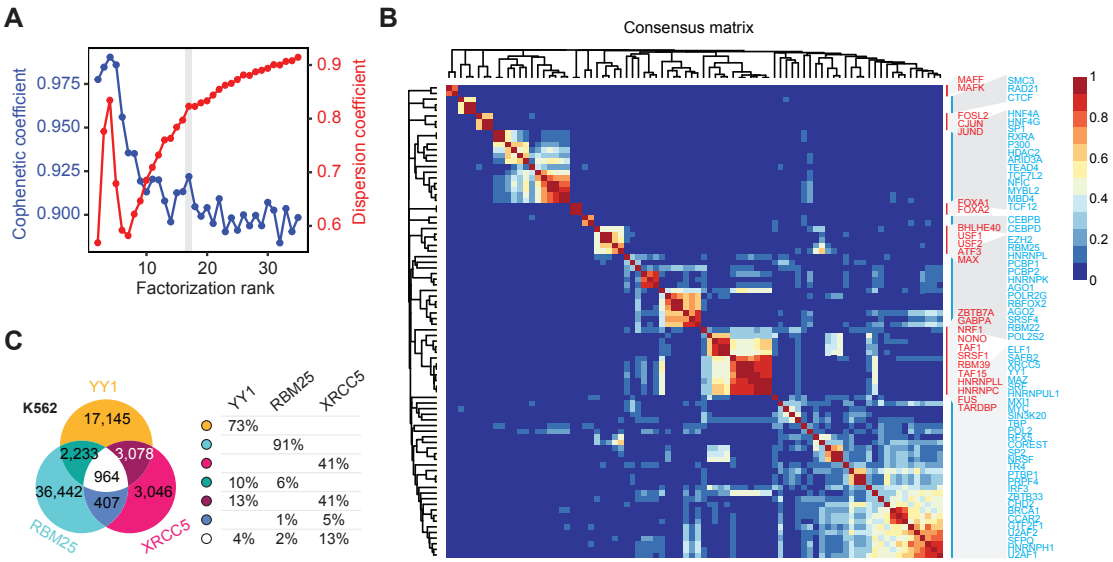
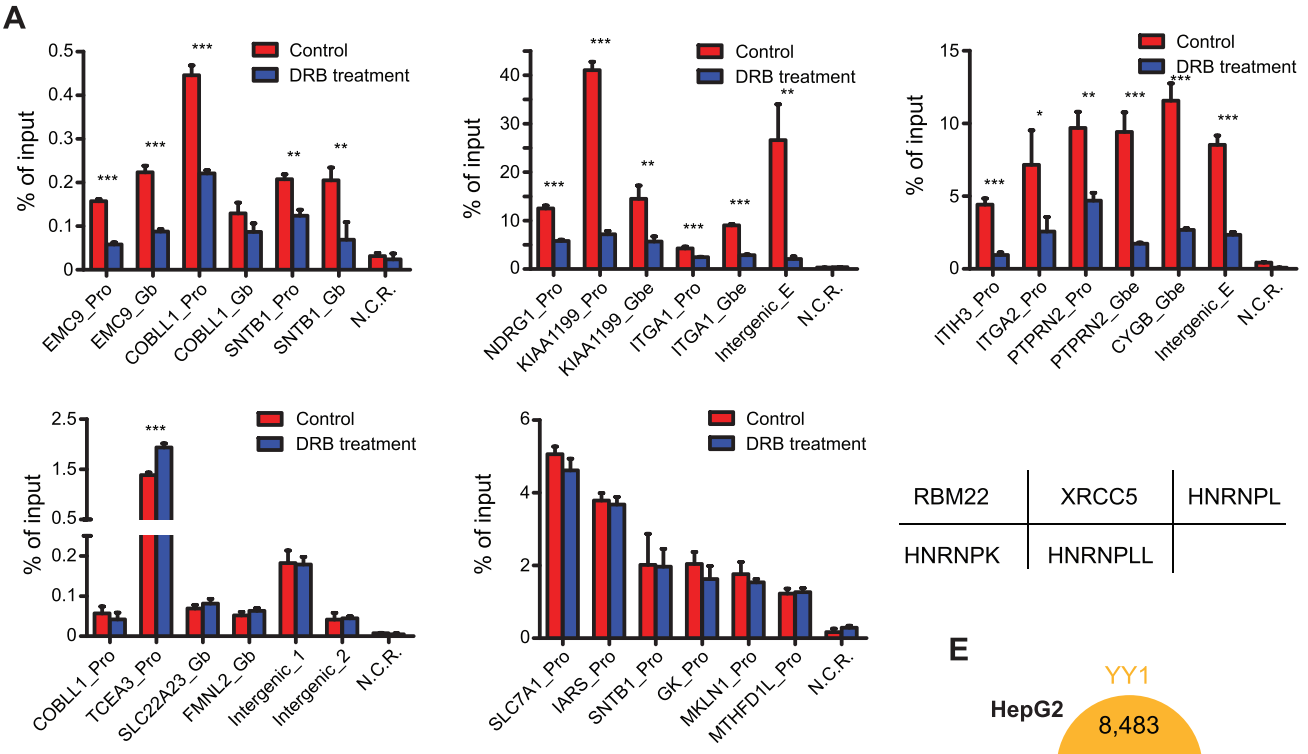


Figure S4. Evaluation of NMF Analysis, Related to Figure 4

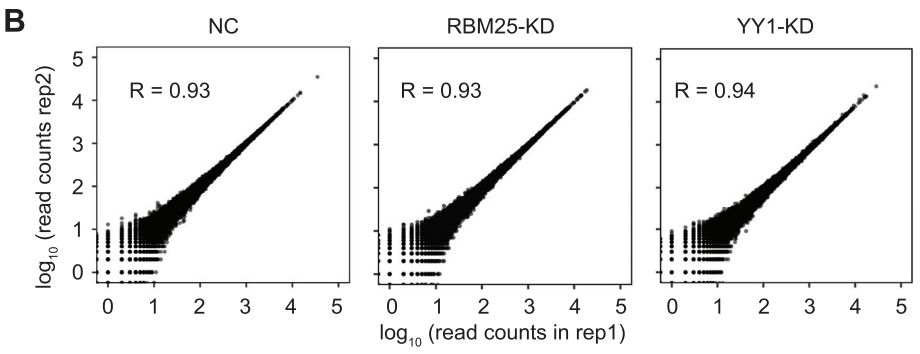
(A) Criteria for estimating a maximal stable factorization rank in NMF analysis. Grey line: the factorization rank of 17 was chosen for the representation of RBP and TF groups in Figure 4A.

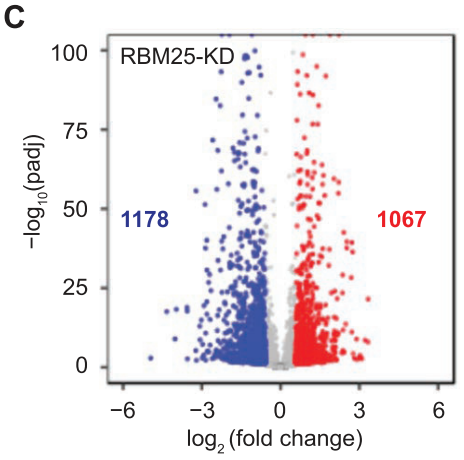
(B) The average connectivity matrix in NMF analysis, showing multiple groups that are distinct from each other.

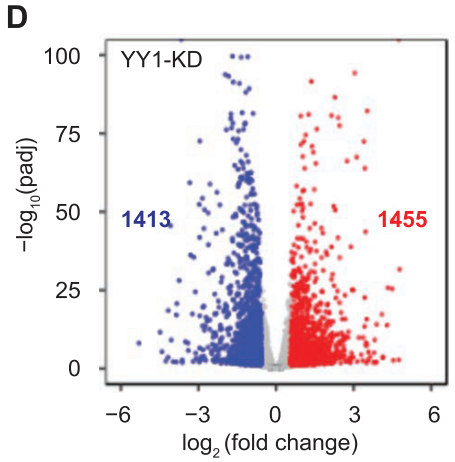
(C) Venn diagram of co-localization of RBM25, XRCC5, and YY1 in K562 cells, and as in Figure 4E, each fraction of the Venn diagram was quantified as the percentage of peaks for each RBP.

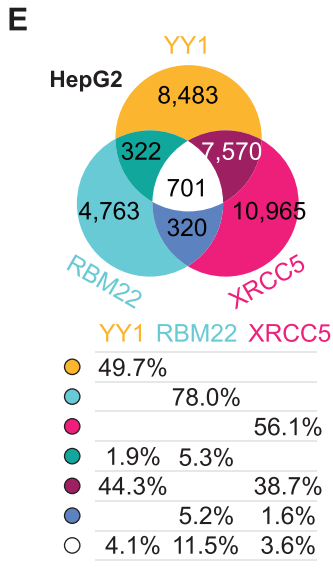


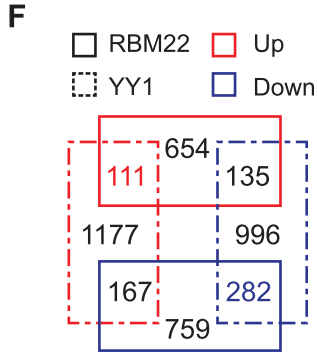
RBM22	XRCC5	HNRNPL
HNRNPK	HNRNPLL	











(legend on next page)

Figure S5. Transcriptional Response to Knockdown of RBM25 or YY1, Related to Figure 5

(A) ChIP-qPCR analysis of the interactions of indicated RBPs with their representative target sites at gene promoters (Pro), gene body (Gb), enhancers at gene body (Gbe), intergenic enhancers (intergenic_E), intergenic binding sites (intergenic_1 and 2) upon DRB treatment. Data are presented as mean ± SD (n = 3). *p < 0.05; **p < 0.01, ***p < 0.001, unpaired Student's t test.

(B) Reproducibility of GRO-seq library for negative control (left), knockdown of RBM25 (middle) and knockdown of YY1 (right). Spearman correlation coefficients are indicated.

(C and D) Volcano plots of gene expression change after knockdown of RBM25 (B) or YY1 (C) with numbers of significantly up- or down-regulated genes indicated.

(E) Venn diagram of co-localization among RBM22, YY1, and XRCC5 in HepG2 cells. Each fraction in the diagram was further quantified as the percentages of peaks for each RBP, based on which individual pairwise co-localizations were calculated, as shown on the right.

(F) Overlap of differentially expressed genes in response to depletion of RBM22 or YY1 in HepG2 cells. The colored boxes and line types separately denote gene expression events up- (red box) or down- (blue box) regulated by RBM22 (solid line) or YY1 (dashed line).

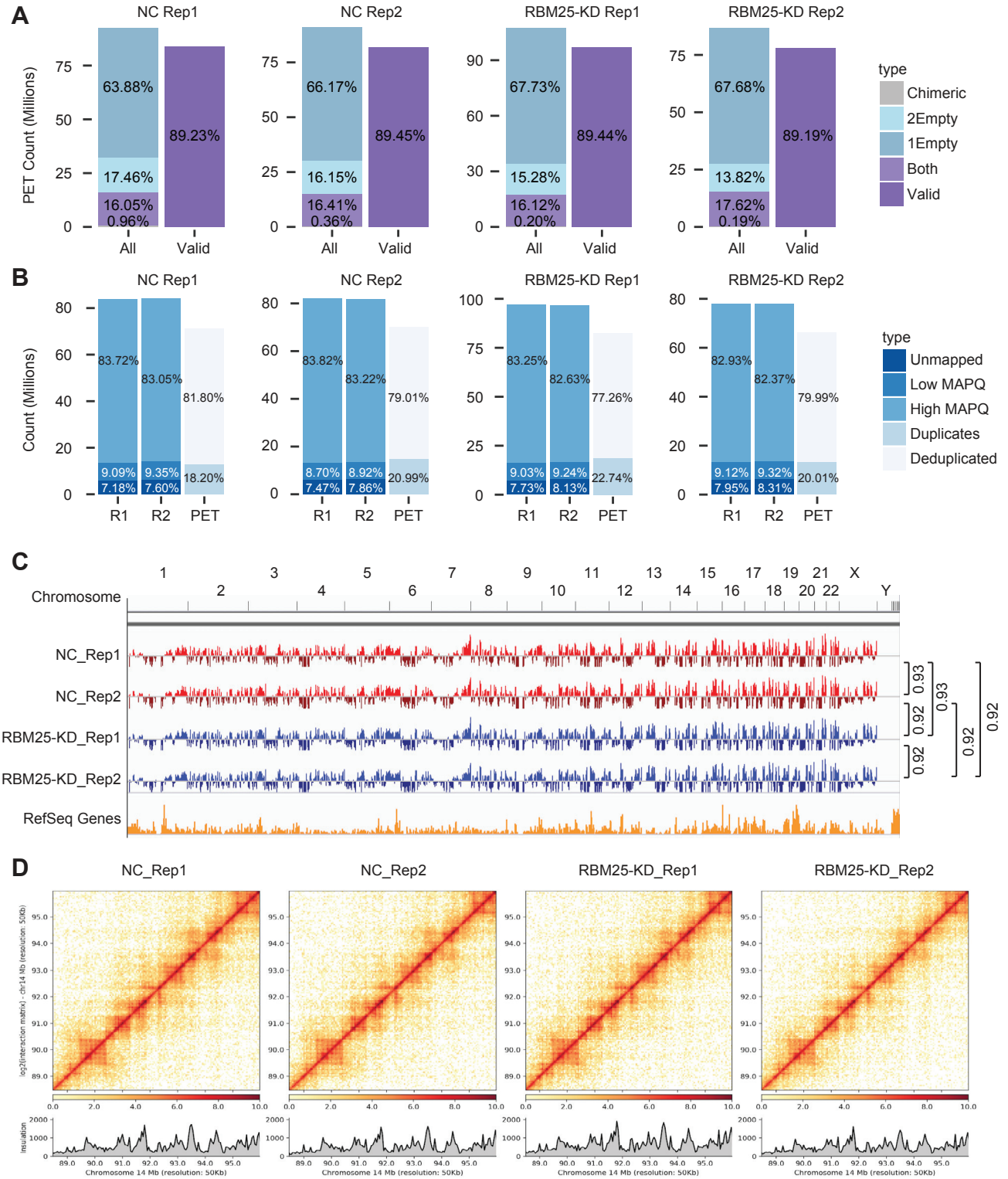


Figure S6. Overview of BL-Hi-C Data, Related to Figure 6

(A) Statistics metrics of linker trimming. Raw PETs are classified based on the presence of linker within each end (Both: both ends have linker; 1 Empty: no linker in one end; 2 Empty: neither of the two ends have linker; Chimeric: self-linker ligation) (left). The percentage of valid raw PETs for alignment is shown as a separate bar (right).

(legend continued on next page)

(B) Statistics metrics of read alignment are shown for each end individually. For read pairs with high mapping quality (MAPQ), proportions of PCR duplicates or unique PETs are indicated.
(C) The open A-type (upper) and closed B-type (lower) compartmentalization are shown for all human chromosome, with cross-sample correlations indicated on the right side.
(D) Chromatin interaction matrix for part of chromosome 14 before and after knockdown of RBM25.