Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2020 September 29.

Published in final edited form as:

Cell Rep. 2020 September 22; 32(12): 108178. doi:10.1016/j.celrep.2020.108178.

Binding of the Treslin-MTBP Complex to Specific Regions of the Human Genome Promotes the Initiation of DNA Replication

Akiko Kumagai[1], **William G. Dunphy**[1,2,*]

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

[2]Lead Contact

SUMMARY

The processes that control where higher eukaryotic cells initiate DNA replication throughout the genome are not understood clearly. In metazoans, the Treslin-MTBP complex mediates critical final steps in formation of the activated replicative helicase prior to initiation of replication. Here, we map the genome-wide distribution of the MTBP subunit of this complex in human cells. Our results indicate that MTBP binds to at least 30,000 sites in the genome. A majority of these sites reside in regions of open chromatin that contain transcriptional-regulatory elements (e.g., promoters, enhancers, and super-enhancers), which are known to be preferred areas for initiation of replication. Furthermore, many binding sites encompass two genomic features: a nucleosome-free DNA sequence (e.g., G-quadruplex DNA or AP-1 motif) and a nucleosome bearing histone marks characteristic of open chromatin, such as H3K4me2. Taken together, these findings indicate that Treslin-MTBP associates coordinately with multiple genomic signals to promote initiation of replication.

Graphical Abstract

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

*Correspondence: dunphy@caltech.edu.

AUTHOR CONTRIBUTIONS

A.K. and W.G.D. conceived the study; A.K. carried out the experiments; and A.K. and W.G.D. analyzed the data and wrote the paper.

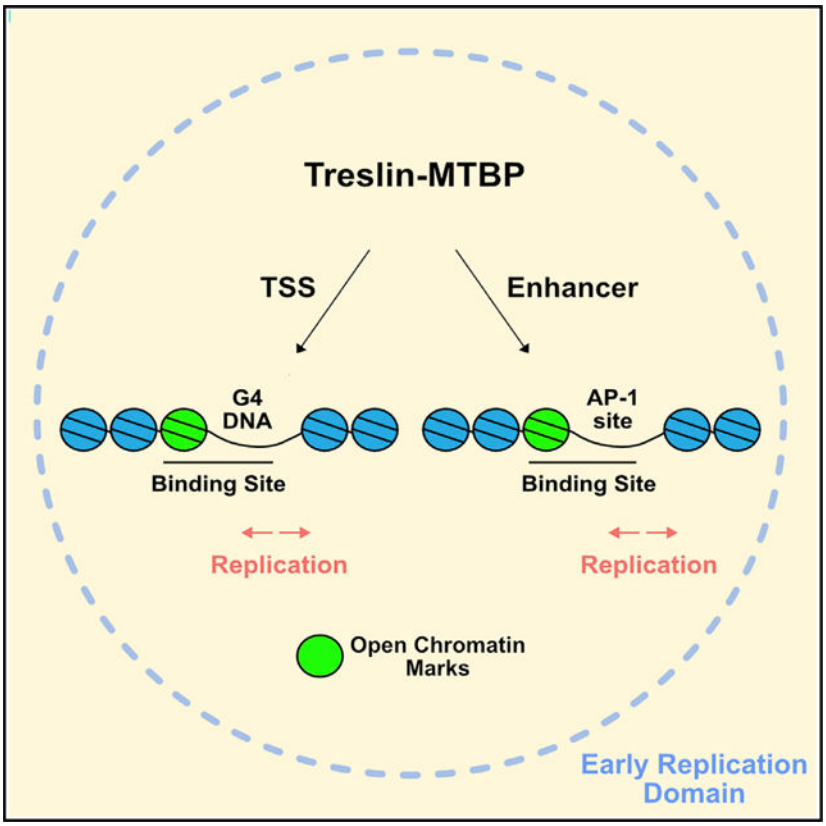
SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.108178.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



In Brief

Kumagai and Dunphy show that Treslin-MTBP, activator of the replicative helicase, binds to at least 30,000 sites in the human genome. Many sites contain a nucleosome with active chromatin marks and nucleosome-free DNA (G-quadruplex or AP-1 site). Thus, Treslin-MTBP associates with multiple genomic elements to promote initiation of DNA replication.

INTRODUCTION

Human cells initiate DNA replication at tens of thousands of origins throughout the genome (Ganier et al., 2019; Marchal et al., 2019; Prioleau and MacAlpine, 2016). A critical aspect of this process involves activation of the replicative helicase, which en-tails an intricate chain of events (Parker et al., 2017; Siddiqui et al., 2013; Tanaka and Araki, 2013). Initially, the origin recognition complex (ORC), Cdc6, and Cdt1 enable loading of the minichromosome maintenance (MCM) complex onto potential origins to establish the pre-replication complex (pre-RC). Thereafter, the Cdc45 and GINS proteins associate with the MCM complex to form the Cdc45-MCM-GINS (CMG) complex, which comprises the core of the activated helicase.

In metazoans, the integration of Cdc45 and GINS with the MCM proteins is mediated by the Treslin-MTBP complex (Boos et al., 2013; Guo et al., 2015; Kumagai and Dunphy, 2017; Kumagai et al., 2010; Sansam et al., 2010). Significantly, this function of Treslin-MTBP depends on phosphorylation of Treslin by the S-phase cyclin-dependent kinase (S-CDK)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Boos et al., 2011; Kumagai et al., 2011; Sansam et al., 2015). This property of Treslin-MTBP helps to explain how S-CDK promotes the occurrence of S-phase in higher eukaryotic cells. Budding yeast Sld3-Sld7, the counterpart of Treslin-MTBP, functions in a similar manner (reviewed in Siddiqui et al., 2013; Tanaka and Araki, 2013).

The exact nature of replication origins in higher eukaryotic cells has remained enigmatic. Unlike budding yeast, higher eukaryotes do not rely upon recognition of specific origin sequences by ORC (Ekundayo and Bleichert, 2019; Parker et al., 2017). Replication in eukaryotic cells exhibits a “replication timing” program in which various domains in the genome undergo DNA synthesis at characteristic times during S-phase (Fragkos et al., 2015; Fu et al., 2018; Marchal et al., 2019). Within given domains, especially early-replicating ones, clusters of origins appear to fire coordinately (Chagin et al., 2016). Moreover, various studies have indicated that initiation is stochastic, in that certain origins fire in some cells but not others during a given cell cycle (Rhind and Gilbert, 2013).

In this context, the term “stochastic” does not necessarily imply that initiation occurs entirely at random throughout the genome. Indeed, certain regions of the genome have a predisposition for initiation. For example, regions of open chromatin around transcription start sites (TSSs) and at enhancers have a higher initiation propensity (Blin et al., 2019; Chen et al., 2019; Dellino et al., 2013; Karnani et al., 2007, 2010; MacAlpine et al., 2004, 2010; Schübeler et al., 2002; Sequeira-Mendes et al., 2009; Sima et al., 2019; Woodfine et al., 2004). Likewise, some DNA sequence features, such as CpG islands and G-quadruplex (G4) DNA, correlate positively with initiation (Besnard et al., 2012; Cadoret et al., 2008; Cayrou et al., 2011, 2012; Delgado et al., 1998; Prorok et al., 2019; Valton et al., 2014). In addition, various epigenetic modifications on histone tails have also been implicated in the regulation of initiation (Cayrou et al., 2015; Parker et al., 2017; Picard et al., 2014; Smith et al., 2016).

Because Treslin-MTBP acts at one of the last steps prior to full activation of the replicative helicase, we have explored whether its localization on chromatin might serve as a potential indicator for sites of initiation and thereby shed light on the distribution and properties of replication origins. Another motivation for these studies is that MTBP contains a C-terminal DNA-binding domain that is important for initiation (Kumagai and Dunphy, 2017). This region, called the CTM domain, can bind to various DNA structures, including double-stranded and G4 DNA. Although a growing body of evidence has implicated G4 DNA in the initiation of replication, the molecular basis of this connection has been unclear (Prioleau, 2017).

In this study, we have mapped the binding sites in the human genome for the MTBP subunit of Treslin-MTBP by using the CUT&RUN method (Skene and Henikoff, 2017). Our results have indicated that MTBP associates reproducibly with several tens of thousands of sites. A large proportion of these sites reside in regions of open chromatin that contain transcriptional-regulatory elements. Many binding sites contain either G4 DNA or an AP-1 motif in a nucleosome-free region at TSSs or enhancers, respectively. The binding sites also encompass an adjacent nucleosome bearing certain histone marks characteristic of open chromatin, most notably H3K4me2. We have also found that many binding sites for MTBP

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

are concentrated in early-replicating zones with an increased tendency for DNA looping. By analogy with promoter-enhancer interactions, looping of DNA may facilitate interaction of Treslin-MTBP with distantly bound MCM complexes to promote more dispersed patterns of initiation. Taken together, our findings suggest that Treslin-MTBP associates with multiple genomic signals in a spatially regulated manner to promote the initiation of DNA replication.

RESULTS

Development of Cell Lines Expressing FLAG-Tagged Versions of Endogenous MTBP

To map the binding sites for Treslin-MTBP in the human genome, we used the CRISPR-Cas9 system to attach a 3X-FLAG tag or an auxin-inducible degron (mAID) onto the endogenous copies of MTBP in human DLD-1 colorectal adenocarcinoma cells (see Method Details). Overall, we generated three cell lines (Figures S1A and S1B). One line contained the mAID sequence on both copies of MTBP (MTBP-mAID/MTBP-mAID). Another line contained the mAID sequence or the 3X-FLAG tag on either copy of endogenous MTBP (MTBP-mAID/MTBP-WT [wild-type]-FLAG). Finally, we also prepared a cell line in which we added the mAID sequence to one copy of MTBP but deleted the CTM domain from the other copy of MTBP and also attached a 3X-FLAG tag (MTBP-mAID/MTBP-ΔC-FLAG). This line would allow us to examine the properties of the MTBP-ΔC mutant.

Characterization of Cell Lines Expressing Tagged Versions of Endogenous MTBP

Next, we examined the expression of MTBP proteins in these cell lines in the presence of auxin. The amount of MTBP-mAID protein was greatly diminished in auxin-treated MTBP-mAID/MTBP-mAID cells (Figure 1A). Treslin was also greatly reduced in these cells because Treslin and MTBP depend upon one another for stability (Kumagai and Dunphy, 2017). For the MTBP-mAID/MTBP-WT-FLAG and MTBP-mAID/MTBP-ΔC-FLAG cell lines, we observed good expression of the MTBP-WT-FLAG and MTBP-ΔC-FLAG proteins, but no detectable MTBP-mAID protein, in the presence of auxin (Figure 1A, lanes 3 and 4). Moreover, Treslin was also present in normal amounts.

We examined chromosomal DNA replication in these cell lines in the absence and presence of auxin (Figures 1B and S1C). We synchronized cells in late G1 with Palbociclib, an inhibitor of Cdk4/6 (Cdk4/6i). Next, we released cells from the arrest and incubated them with EdU to monitor DNA replication. Auxin-treated cells with two alleles of MTBP-mAID did not incorporate EdU above background levels at t = 6 h in comparison with cells incubated without auxin (Figures 1B and S1C). This observation underscores the importance of MTBP for entry into S-phase (Boos et al., 2013; Köhler et al., 2019; Kumagai and Dunphy, 2017). On the other hand, the MTBP-mAID/MTBP-WT-FLAG cells displayed similar incorporation of EdU at t = 6 h with and without auxin (Figures 1B and S1C).

As described previously, cells expressing the MTBP-ΔC protein incorporate EdU reasonably well (Figure S1C), but exhibit delayed S-phase progression due to diminished origin firing (Kumagai and Dunphy, 2017). To examine this issue for the auxin-regulated cell lines, we incubated auxin-treated cells expressing the MTBP-WT-FLAG or MTBP-ΔC-FLAG proteins

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

with EdU at various times after release from a G1 arrest and scored cells according to whether EdU-labeling patterns were characteristic of early, mid, or late S-phase (Figure 1C). We observed that most cells expressing MTBP-WT-FLAG had completed S-phase by 12.5 h. However, cells containing the MTBP-ΔC-FLAG protein progressed through S-phase more slowly. In particular, more than 50% of S-phase cells carrying MTBP-ΔC had mid-S staining patterns at 10 and 12.5 h. Overall, these experiments indicated that DLD-1 cells expressing only MTBP-WT-FLAG replicate their DNA efficiently. Furthermore, in accordance with previous studies, cells containing a ΔC mutant version of MTBP-FLAG display impaired S-phase progression.

MTBP Associates with Tens of Thousands of Sites Throughout the Human Genome

We used the CUT&RUN method to map the binding sites for the MTBP-FLAG proteins in the genome (Skene and Henikoff, 2017). For these experiments, we treated cell lines expressing MTBP-WT-FLAG or MTBP-ΔC-FLAG protein with auxin for 16 h prior to the CUT&RUN procedure. As the control, we used the same protocol on MTBP-mAID/MTBP-mAID cells incubated in the absence of auxin because these cells contain no FLAG-tagged MTBP. To compare binding accurately between samples, we utilized yeast spike-in DNA as a quantitative reference.

We carried out the CUT&RUN experiments three times and determined the locations of the sequencing reads in the genome. As shown in Figure 2A, we have depicted representative genome browser profiles for loci that encode lamin B2 (*LMNB2*) and the *HoxA* cluster. The *LMNB2* locus is a well-characterized area for origin activity, and the *HoxA* cluster contains numerous origins within about 150 kb (Abdurashidova et al., 2000; Besnard et al., 2012; Cadoret et al., 2008; Giacca et al., 1994; Langley et al., 2016). In the case of *LMNB2*, we detected a pronounced peak for binding of MTBP downstream of the *LMNB2* gene and immediately upstream of the *TIMM13* gene, which coincides with an area of established origin activity. For example, this region exhibited a strong peak of initiation activity in a published study that used the Ini-seq method for detection of early origins (Langley et al., 2016). For the *HoxA* cluster, we noticed a high density of MTBP peaks that correlated with increased origin activity.

To analyze the nature of the MTBP binding sites, we utilized the MACS2-peak calling algorithm to locate the sites in the genome. To increase confidence in the identifications, we selected peaks that appeared in at least two of three replicates. By this procedure, we identified 28,981 reproducible, high-confidence binding sites for MTBP-WT protein (Figures 2B, S2A, and S2B). For the MTBP-ΔC protein, we observed 14,600 reproducible, high-confidence sites (Figures 2B, S2A, and S2B). Most of the high-confidence sites for MTBP-ΔC overlapped with those for MTBP-WT (Figure 2B). Furthermore, heatmaps showed that sequencing reads from cells expressing WT or ΔC versions of MTBP mostly co-localized in the genome (Figure 2C). We noted that sites that scored as WT only typically contained reads for the ΔC protein that fell below the peak-calling threshold for MACS2.

In this regard, the vast majority of peaks for MTBP-WT had higher read scores than for MTBP-ΔC (Figures 2C, 2D, 2E, and S2C). We quantified this difference by determining total normalized read counts per peak across the genome and found that the mean values for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

MTBP-ΔC were 53% of those for MTBP-WT (Figure 2E). Paired Wilcoxon tests indicated that this difference is highly significant ($p < 2e\text{-}16$). Thus, the ΔC protein associates significantly less well with the genome than the WT protein. Finally, there was a shoulder on either side of the center of peak-summit plots for both proteins, which suggested some complexity in the binding mechanism (Figure 2D).

As another means to characterize the DNA-binding properties of Treslin-MTBP, we incubated biotinylated DNA of various lengths in nucleoplasmic extract (NPE) fractions from *Xenopus* eggs. Treslin-MTBP required at least 250 to 300 bp for binding, whereas ORC and TopBP1 could associate with shorter DNA (Figure S2D). Moreover, since NPE fractions do not support formation of the pre-RC, this binding occurs independently of the MCM complex (Walter et al., 1998). Taken together, these observations suggest that Treslin-MTBP does not associate with a single, short DNA motif.

Overall, these experiments illustrate that Treslin-MTBP associates with a reproducible set of binding sites in the human genome. Moreover, Treslin-MTBP appears to contain at least two chromatin-binding domains because a complex containing MTBP-ΔC can still associate with chromatin, albeit more weakly. We combined the peaks for both the MTBP-WT and MTBP-ΔC proteins (29,858 peaks in total) and used this dataset for subsequent analyses (Figure 2B). These peaks were enriched in 5′ UTRs, promoter regions, and CpG islands (Figure S2E). Furthermore, 38.7% of the MTBP peaks also co-localized well with early replication origins identified by Ini-seq (Langley et al., 2016; Figure S2F). This overlap is notable because Ini-seq identifies only early origins and this study utilized a different cell type (human bladder carcinoma).

A Majority of Binding Sites for MTBP Reside at Transcriptional-Regulatory Elements

Next, we sought to characterize the nature of the binding sites for MTBP. We used the discovery tool ChromHMM, which identifies various functional elements in the genome (e.g., promoters, enhancers, and so forth) according to chromatin-state signatures (Ernst and Kellis, 2017). Due to the relative paucity of genome-wide data from DLD-1 cells, we utilized data from HCT-116 cells, a very similar colon cancer cell line (ENCODE Project Consortium, 2012). We observed MTBP peaks at or near TSSs, enhancers and super-enhancers, and some CTCF-binding sites (Figure S3A). We proceeded to segregate the MTBP peaks into three classes according to their locations: promoter-TSS; enhancer and super-enhancer; and others (Figure 3A).

For these three classes, we examined heatmaps for the WT and ΔC MTBP proteins alongside maps for other relevant markers from public datasets (Figure 3B). For example, we examined the binding sites for Orc2, a component of ORC. We also assessed the localization G4 structures, which have been implicated in initiation. The data for Orc2 was derived from a chromatin immunoprecipitation sequencing (ChIP-seq) study in K562 cells (Miotto et al., 2016), which is consistent with previous studies (Dellino et al., 2013; MacAlpine et al., 2010). The localization of G4 structures was analyzed by ChIP-seq in HaCaT cells with a monoclonal antibody (BG4) that specifically recognizes G4 DNA (Hänsel-Hertsch et al., 2018). We found that both Orc2 and BG4 co-localized well with MTBP peaks at promoter-TSS regions. There was also co-localization, but to a somewhat

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

lower extent, at enhancers and super-enhancers, perhaps due to differences between cell lines.

We first analyzed localization of MTBP to promoter-TSS regions in greater detail. We selected MTBP peaks that resided near a TSS and aligned them so that the TSS would be located at the center of the plots (Figure S3B). Interestingly, we observed two closely overlapping peaks for both MTBP-WT and MTBP-ΔC immediately upstream and downstream of the TSS. Both Orc2 and BG4 displayed relatively broad peaks around the TSS, consistent with previous observations (Dellino et al., 2013; Hänsel-Hertsch et al., 2018; Miotto et al., 2016).

As discussed above, G4 structures have been implicated in the regulation of replication initiation (Prioleau, 2017; Prorok et al., 2019). Moreover, we have previously observed direct binding of the CTM domain from MTBP to multiple G4 structures. Accordingly, we assessed the relationship between binding of MTBP to TSS regions and the presence of G4 structures. We segregated TSSs into two groups that either contain or lack sequences that could potentially form G4 structures. For this procedure, we identified potential G4 structures with the G4Hunter program (Bedrat et al., 2016). Next, we examined peaks for MTBP (WT and ΔC), Orc2, and BG4 around these two classes of TSSs (Figure 3C).

First, we observed that TSSs with potential G4 structures contained many more BG4 peaks than TSSs without G4 DNA, which validates analysis with G4Hunter. We found that the MTBP-WT and ΔC proteins as well as Orc2 were enriched at TSSs with potential G4 structures relative to TSSs without these structures. Moreover, as was the case for the whole-genome analysis, the peaks for the MTBP-ΔC protein were substantially lower at the G4-containing TSSs (as well as at the lower proportion of G4-lacking TSSs) in comparison with the WT protein. These observations indicate that MTBP associates preferentially with TSSs containing potential G4 structures. However, the ΔC mutant of MTBP that lacks the CTM domain still associates with TSSs containing G4 structures, albeit to a substantially lower extent. Thus, the CTM domain does not by itself direct Treslin-MTBP to these locations (see Discussion).

Relationship Between Binding of MTBP and Transcription

Initiation of DNA replication has been shown to correlate with transcription (see Introduction). To examine the relationship between the binding of MTBP and transcription, we sorted TSSs according to total RNA sequencing (RNA-seq) read scores from DLD-1 cells (Rokavec et al., 2017) and examined peaks for MTBP-WT and MTBP-ΔC (Figure 3D). In parallel, we examined the distribution of Orc2, BG4, and H3K27ac, a marker for transcriptionally active chromatin. As expected, H3K27ac correlated well with transcription. Both MTBP proteins associated preferentially with transcriptionally active loci, but there did not appear to be a quantitative correlation between the level of transcription and the amount of bound MTBP. The results were similar for Orc2 and BG4. Furthermore, plotting of read scores for the MTBP-WT peaks against RNA-seq read scores did not reveal any particular correlation (Figure S3C). Overall, these observations indicate that a large proportion of binding sites for MTBP reside in transcriptionally active chromatin. However, the level of bound MTBP does not correlate in a quantitative manner with the degree of transcriptional

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

activity at individual loci (as estimated by RNA-seq, which does not account for RNA turnover).

Different Sequence Elements Correlate with the Localization of MTBP at TSSs versus Enhancers

To assess the overall contribution of G4 structures to the localization of MTBP, we compared the distribution around TSSs, as well as at other sites, of MTBP peaks that either do or do not contain potential G4 structures. We observed that MTBP peaks that contain G4 structures were enriched around TSSs in comparison with peaks that do not (Figure 4A, top left). Kolmogorov-Smirnov tests indicated that this enrichment is highly significant (Table S1). Because MTBP also localizes at enhancers and super-enhancers (see Figures 3A, 3B, and S3A), we examined the disposition of G4 structures in such areas. For this analysis, we focused on the distribution of MTBP peaks around the centers of individual enhancers. We found that the distribution was not very different depending on whether the MTBP peaks contained a potential G4 structure or not (Figure 4A, top right; Table S1).

This observation motivated us to search for other sequence motifs that might be enriched in the MTBP peaks, both in general and at subsets of binding sites. Using HOMER, we found that the TGAG/CTCA (AP-1) motif was highly enriched in MTBP peaks (Figure S4A; see Method Details). RUNX1 and TEAD motifs were also enriched, but to a lesser extent. Moreover, MTBP peaks containing the AP-1 motif were enriched at enhancers and super-enhancers, but not at TSSs (Figure 4A, compare bottom panels on left and right; super-enhancers, not shown). Kolmogorov-Smirnov tests showed that the difference in distribution of MTBP peaks with or without AP-1 motifs is highly significant (Table S1). AP-1 is a heterodimeric transcription factor containing proteins in the families of c-Fos, c-Jun, and others (Lee et al., 1987). Notably, the AP-1 motif has been implicated in both viral and cellular DNA replication previously (Cadoret et al., 2008; Guo and DePamphilis, 1992; Puzzi et al., 2015), but the functional basis has been unclear.

Finally, we tallied what proportion of MTBP peaks have various motifs (Figure 4B). Overall, at least 20.4% of MTBP peaks possess a G4 structure ($p < 2e\text{-}16$, Fisher's exact test). Most of these G4 structures reside at promoter-TSS regions. Among this class, 50.3% have a G4 structure. Moreover, 28.2% of total MTBP peaks have an AP-1 motif ($p < 2e\text{-}16$, Fisher's exact test). Almost half (45.9%) of MTBP peaks in enhancer and super-enhancer domains have an AP-1 site.

Binding Sites for MTBP Contain AP-1 and G4 Motifs at Characteristic Locations

To gain further insight into the role of the AP-1 motifs, we proceeded to analyze the position of these short motifs within the larger MTBP peaks. In genome-wide compilations, we observed two peaks of AP-1 motifs at around 145 bp (±10 bp) on either side of MTBP peak summits (Figure 4C). This pattern was most prominent at enhancer/super-enhancer regions. We did not observe any preference between TGAGTCA and TGACTCA versions of the AP-1 motif. Other less common motifs from the HOMER searches, namely, a variant AP-1 motif denoted "version 2" (TGAG/CTAA), RUNX1 motif (AACCACA), and TEAD motif

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(A/GCATTCC), all resided at very similar locations as the AP-1 motifs. However, the heights of the peaks were substantially lower (Figure S4B).

Because the whole-genome compilations showed two peaks of AP-1 motifs (one on either side of MTBP peak summits), we asked whether individual MTBP binding sites contain one or two AP-1 motifs. To examine this question, we classified individual MTBP binding sites as having an AP-1 motif on the left, right, or both sides of the peaks (Figure S4C). We found that most of the MTBP peaks (89.5%) have one AP-1 motif, whereas 10.5% of the peaks have AP-1 motifs on both sides. Upon segregation of the peaks into classes that have the AP-1 motif on the left or right side, we observed an asymmetric nature in the peak-summit plots for MTBP (Figure S4C, bottom two panels). In particular, the number of reads was higher on the side that contained the AP-1 motif. This observation suggests that the CUT&RUN method reveals a complexity in the binding of MTBP that may correspond to multiple types of interaction with chromatin.

We performed a similar analysis for the MTBP binding sites at TSS regions (Figure S4D). We observed that the summits of individual MTBP peaks resided on either the left or right side of each TSS in 89.4% of the cases (Figure S4D). Interestingly, however, there was no correlation with the direction of transcription. Moreover, 10.6% of peaks have a TSS on both sides. As was the case for AP-1 motifs, upon segregation of MTBP peaks into classes that have the TSS on either the left or right side, there was an asymmetry in the peak-summit profile for MTBP (Figure S4D, bottom two panels). In particular, the height of the profile was somewhat higher on the side containing the TSS.

MTBP Associates Coordinately with Adjacent Nucleosome-free and Nucleosome-Containing Regions

Next, we assessed the nature of chromatin around the binding sites for MTBP. For this purpose, we examined DNase sequencing (DNase-seq) and MNase sequencing (MNase-seq) datasets from HCT-116 cells (Figure 4D). DNase-seq detects regions that are hypersensitive to DNase I, which is an indicator for nucleosome-free regions. By contrast, MNase-seq identifies regions of the DNA that are resistant to digestion with micrococcal nuclease due to the presence of nucleosomes. ORC binds to nucleosome-free regions (Miotto et al., 2016; Parker et al., 2017). Strikingly, we observed that the MTBP peak summits co-localized with the MNase-seq peaks, which indicates the presence of a nucleosome at these locations. Moreover, the MTBP peak summits resided between two closely spaced DNase-seq peaks that are indicative of open chromatin. Taken together, these observations indicate that Treslin-MTBP associates coordinately with adjacent nucleosome-containing and nucleosome-free regions.

As stated above, we found that MTBP is enriched in some CTCF-containing regions (see Figure S3A). Therefore, we examined the localization of CTCF as well as the YY1 protein relative to MTBP (Figure 4D). Both CTCF and YY1 promote long-distance DNA interactions by looping mechanisms (Dixon et al., 2012; Weintraub et al., 2017). We observed a broad distribution of CTCF in areas containing MTBP peaks. In contrast, YY1 exhibited a strong co-localization with MTBP.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In view of the observation that MTBP associates with sites containing a nucleosome, we investigated whether there was a correlation between the binding sites for MTBP and epigenetic marks on histones. We surveyed various modified forms of histones H3 and H4 that have been implicated in DNA replication and transcription as well as other processes (Cayrou et al., 2015; Parker et al., 2017; Picard et al., 2014; Smith et al., 2016). We also examined the distribution of H2A.Z, which was recently implicated in the recruitment of ORC to chromatin (Long et al., 2020). Specifically, we analyzed ChIP-seq datasets from HCT-116 cells and found that the MTBP peaks co-localized with H3K27ac, H3K4me1, H3K4me2, H3K4me3, H3K9ac, and H2A.Z in a highly significant manner (Figures S5A and S5B; Table S2). More than half of the MTBP peaks overlapped with these markers at the genome-wide level (Figure S5B). The highest overlap occurred with H3K4me2. By contrast, there was not an apparent correlation with H3K27me3, H4K20me1, H4K20me3, H3K9me2, H3K9me3, and H3K36me3 (Figure S5A).

To pursue these observations, we performed CUT&RUN experiments to identify the distribution of H3K4me2 on chromatin in DLD-1 cells. In genome browser profiles, we observed a clear correspondence between H3K4me2 and MTBP (Figure 5A). More specifically, 99% of the MTBP peaks overlapped with H3K4me2 (Figure 5B). Both Fisher's exact tests ($p < 2e\text{-}16$) and permutation analyses (Table S2) indicated that this overlap is highly significant. Moreover, we separated the H3K4me2 peaks into two groups: one group with either AP-1 or G4 motifs within 200 bp of the H3K4me2 peak summits and another group without these motifs (Figure 5C). There was significant enrichment of MTBP binding sites in H3K4me2-containing regions that also include either AP-1 or G4 motifs ($p < 2e\text{-}16$, Fisher's exact test). Interestingly, heatmaps showed co-localization of MTBP and H3K4me2 not only at TSSs and enhancers/super-enhancers, but also in the "others" category of binding sites (Figure 5D). Taken together, these experiments indicate that MTBP associates with H3K4me2-containing nucleosomes in a highly significant manner.

Binding Sites for MTBP at TSSs and Enhancers Have a Similar Organization

As described above, we found that MTBP binding sites possess an asymmetric nature. Therefore, we attempted to relate these observations to the presence of both nucleosome-containing and nucleosome-free regions within the binding sites. First, we aligned MTBP sites containing AP-1 motifs so that these motifs would reside on the right-hand side of the MTBP peak summits. Next, we examined the distribution of the various histone marks. We observed that the MTBP peak summits resided at nucleosomes (identified by MNase-seq) containing H3K4me1, H3K4me2, H3K4me3, H3K27ac, and H3K9ac (Figures 5E, 5G, and S5C). Moreover, the AP-1 motifs to the right of the nucleosome resided in open chromatin identified by DNase-seq. Finally, YY1 also localized to this nucleosome-free region.

We repeated these analyses for TSS-associated MTBP binding sites. We aligned these MTBP sites so that the TSSs would be on the right side (irrespective of the direction of transcription). We found that binding sites for MTBP at TSSs have a similar arrangement as at enhancers (i.e., a nucleosome with an adjacent nucleosome-free region; Figures 5F and 5G and S5D). These MTBP peaks co-localized with nucleosomes modified with H3K4me2, H3K4me3, H3K27ac, and H3K9ac. The level of H3K4me1 appeared lower than was the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

case for enhancers. Moreover, G4 structures, identified with G4Hunter, resided on the right in the adjacent open chromatin. YY1 was also detected in these regions. Overall, these analyses indicate that binding sites for MTBP have a characteristic architecture, consisting of a nucleosome with marks of open chromatin and an adjacent nucleosome-free area containing G4 or AP-1 motifs at TSSs and enhancers, respectively. Furthermore, in the case of TSSs, the MTBP-binding nucleosome can reside predominantly either upstream or downstream of the TSS without regard to the direction of transcription (see Figures S3B, 6E, and S6D).

Potential G4 structures and AP-1 motifs are highly abundant in the human genome (approximately 342,000 and 246,000 copies, respectively), which makes it unlikely that these features alone could regulate initiation. However, if we impose the constraint that these features reside in DNase I-hypersensitive regions, the values drop considerably to 11,919 and 15,065 sites for G4 DNA and AP-1 sites, respectively (Table S2). Significantly, in these categories, 44% of the G4 structures ($p < 2e\text{-}16$, Fisher's exact test) and 36.7% of the AP-1 sites ($p < 2e\text{-}16$, Fisher's exact test) reside within MTBP binding sites (see Discussion).

MTBP Is Concentrated in Early-Replicating Zones

Finally, we sought to assess explicitly whether the distribution of MTBP peaks could account for replication patterns in the genome. In DLD-1 cells, we observed that MTBP peaks occur at a density of 1.02 peaks per 100 kb in the genome. We segmented replication-timing data from HCT-116 cells (Weddington et al., 2008) into early, mid, and late domains and counted the number of MTBP peaks in each type of domain (Figure 6A). We observed that MTBP peaks were not distributed evenly in the genome. Specifically, the mean peak density in early domains was 5.7-times higher than in late domains (Figure 6B). A majority of the MTBP peaks in early domains resided in promoter-TSS and enhancer/super-enhancer regions (Figure S6A), whereas most peaks in late domains occurred outside such regions.

To study the relationship between the binding sites for MTBP and initiation of DNA replication more closely, we carried out labeling studies with EdU in DLD-1 cells. We decided to focus on early replication zones (see Method Details). To label these zones, we synchronized auxin-treated DLD-1 cells expressing MTBP-WT-FLAG or MTBP-ΔC-FLAG in late G1 by treatment with Cdk4/6i for 20 h. Next, we removed Cdk4/6i and incubated for an additional 3 h in the presence of EdU, HU, and aphidicolin (Macheret and Halazonetis, 2018; Tubbs et al., 2018). During this time, only early origins can undergo firing and the resulting nascent strands do not elongate extensively due to the presence of aphidicolin.

We observed that peaks of EdU-labeled nascent DNA were highly reproducible between three replicates (Figure S6B). Nascent DNA from S-phase cells expressing MTBP-WT showed a strong enrichment over control Cdk4/6i-arrested cells (Figure 6A). By contrast, levels of nascent DNA from cells expressing MTBP-ΔC were noticeably lower. We identified 2,473 early replication initiation zones with a median size of 48 kb for the MTBP-WT protein (Figure S6C). These regions covered 4.9% of the genome and contained 6,860 MTBP peaks (23.0% of the total). We compared average EdU-seq reads for cells expressing MTBP-WT or MTBP-ΔC and found that incorporation of EdU into these early replication initiation zones was significantly compromised in the case of the MTBP-ΔC protein (Figure

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

6C). Thus, the reduced binding of the MTBP-ΔC protein to chromatin results in significantly impaired initiation.

To assess further the relationship between the MTBP peaks and initiation, we examined the distribution of EdU-seq reads 100 kb upstream and downstream around MTBP peaks in the three categories of binding sites: promoter-TSS; enhancer/super-enhancer; and others. We also examined EdU-seq reads around 10,000 randomly generated sites as a control. We observed broad EdU-seq peaks surrounding all three categories of MTBP peaks, but not around control sites (Figure 6D). Thus, there is clear correlation between initiation of replication and the locations of the MTBP peaks. We also noted that the MTBP peaks in the promoter-TSS category exhibited a spike of EdU-seq signal near the peaks in addition to the broad distribution of EdU reads around the peaks. We decided to analyze this observation further.

As shown above, the binding profile of MTBP at TSSs exhibits two closely spaced peaks due to binding on the 5′ or 3′ side of the TSS (Figure S3B). We segregated the MTBP peaks according to whether their summits resided at the 5′ or 3′ side of the TSS (Figure 6E). These two classes of peaks are 300 bp apart and co-localize with nucleosomes, as indicated by MNase-seq (Figures 6E, 6G, and S6D). Next, we plotted EdU-seq reads around each group of MTBP binding site as well as around control TSSs with little or no MTBP (Figure 6F). For both groups containing MTBP binding sites, we observed two types of EdU-seq peaks: one sharp peak that lies within 500 bp of the TSS, and a broad peak centered around 10 kb upstream. In contrast, control TSSs with little or no MTBP showed very low EdU-seq reads. These results indicate that some sites of initiation reside close to the TSS, whereas others occur in a broad zone upstream of the TSS, which is consistent with past studies (see Chen et al., 2019). The lower frequency of initiation downstream of the TSS is most likely due to transcription-dependent displacement of MCMs (Chen et al., 2019; Gros et al., 2015; Powell et al., 2015)

MTBP-Containing, Early-Replicating Zones Exhibit Elevated DNA Looping

Next, we considered how replication might be occurring at sites distal to the location of MTBP near the TSS. We noted that the density of MTBP peaks in early initiation zones (identified by EdU-seq) was very high, approximately 4.75 peaks per 100 kb (Figure 7A, left). As described above, MTBP co-localizes very well with YY1 at promoter-TSS regions (see Figure 4D). Accordingly, we examined the distribution of YY1 in early replication zones and observed that the density of YY1 peaks within these zones was likewise more than five times higher than in randomly selected regions (Figure 7A, right; Figures S7A and S7B). Moreover, in heatmaps, we observed a co-localization of EdU-seq reads, MTBP, and YY1 in the early initiation zones (Figure 7B). We also noted the presence of the TSS/enhancer-associated histone marks (e.g., H3K4me1, H3K4me2, H3K4me3, H3K9ac, and H3K27ac) in the initiation zones (Figure 7B). To evaluate these observations further, we sorted these TSSs according to the direction of transcription and mapped their locations around initiation zones (Figure S7A). We found that TSSs involved in leftward versus rightward transcription are preferentially associated with the left and right boundaries of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

initiation zones, respectively. Thus, initiation zones are demarcated by TSSs involved in transcription outward from these zones.

YY1 acts as a structural regulator of the three-dimensional organization of chromatin by binding to promoter and enhancer regions (Weintraub et al., 2017). Specifically, YY1 dimerizes and thereby creates loops that facilitate interaction of these DNA elements. We analyzed YY1 HiChIP data (Weintraub et al., 2017) and found that YY1-mediated loops are enriched in early initiation zones (Figures 7C and S7B). In particular, the PET (paired-end tag) scores from the YY1 HiChIP data were significantly higher in initiation than control zones. Taken together, these observations indicate that early initiation zones form distinct structures that contain a high concentration of MTBP as well as various markers of active chromatin. Furthermore, these zones exhibit elevated DNA looping. Because MCM complexes would be more broadly distributed throughout the genome than Treslin-MTBP, these observations raise the possibility that looping of DNA may allow Treslin-MTBP to activate MCM complexes in a dispersed manner in early replication zones (Figure 7D).

DISCUSSION

In this study, we have explored whether the distribution of Treslin-MTBP on chromatin would shed light on the properties of sites for initiation of DNA replication. Our findings have indicated that Treslin-MTBP possesses a set of at least 30,000 binding sites. A majority of sites reside in regions that have been implicated in transcriptional regulation, which fits with a substantial body of data that such zones are also preferred areas for initiation of replication. A variety of observations suggest that these binding sites for MTBP are functionally important for DNA replication. First, the number of binding sites generally fits with estimates for the number of replication initiation sites or zones in human cells. Although estimates vary depending on the methodology, the consensus seems to be that human cells would utilize approximately 10,000 to 50,000 origins (Chagin et al., 2016; Ganier et al., 2019). We have also observed a good correspondence (38.7% overlap) between the set of MTBP binding sites and replication initiation sites that were identified with the Ini-seq method (Langley et al., 2016). Finally, we have directly identified early replication initiation zones in DLD-1 cells by using an EdU-seq method and found an elevated concentration of MTBP specifically in these areas.

We previously reported that the CTM domain of MTBP can bind to various DNA structures, including double-stranded and G4 DNA. Consistent with this observation, the MTBP-ΔC mutant that lacks this domain exhibits substantially reduced binding to chromatin. Furthermore, cells expressing the MTBP-ΔC mutant displayed greatly reduced initiation in early replication zones, as monitored by EdU-seq. However, the fact that this mutant still does associate with chromatin to a significant extent suggests that there is at least one additional chromatin-binding domain in the Treslin-MTBP complex. In this regard, we have found that the binding sites for Treslin-MTBP encompass multiple genomic features. The implication is that Treslin-MTBP must associate with multiple genomic elements to promote initiation optimally.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We have observed a highly significant correlation between binding sites for MTBP and locations of G4 motifs, specifically those in the vicinity of TSSs. Nonetheless, although the ΔC mutant displays considerably lower binding to these regions, it still does localize to them. Thus, the CTM domain by itself appears not to direct the Treslin-MTBP complex to these areas. Other regulators of initiation, including Orc1 and RecQ4, have also been reported to interact with G4 DNA (Hoshina et al., 2013; Keller et al., 2014). RecQ4 also associates with TopBP1, a known Treslin-interacting protein, which raises the possibility that Treslin-MTBP and RecQ4 could act cooperatively in binding to G4 structures (Hoshina et al., 2013; Kumagai et al., 2010). It should also be noted that the step at which G4 DNA would act in the overall process of initiation is unknown. Nonetheless, our results have indicated that there is a striking co-localization of Treslin-MTBP and G4 DNA at TSSs.

In our analyses, we have also obtained evidence that additional sequence motifs correlate with the localization of MTBP to areas that lie outside promoter-TSS regions. In particular, we have observed an enrichment of AP-1 motifs in the binding sites for MTBP at enhancers and super-enhancers. This motif has been implicated previously in both viral and cellular DNA replication (Cadoret et al., 2008; Guo and DePamphilis, 1992; Puzzi et al., 2015). In the case of the *LMNB2* locus, Puzzi et al. (2015) reported that c-Fos and c-Jun, components of AP-1, had dissociated from the origin region by the time of the G1/S transition, which would allow access of other proteins.

Both potential G4 structures and AP-1 motifs are highly abundant in the human genome. Therefore, it would not seem plau-sible that these elements by themselves could mark areas for initiation. However, in our analyses, we have also observed that binding sites for MTBP possess a region of open chromatin sensitive to DNase I at the positions of the G4 and AP-1 sites. There are only approximately 12,000 sites for G4 DNA and 15,000 AP-1 sites that conform to this more stringent criterion. Notably, MTBP associates with a substantial proportion of these sites (37% to 44%).

There is also a correlation between the binding sites for MTBP and certain types of histones and histone marks, which could provide an additional level of specificity. The best correlation occurred with H3K4me2. In DLD-1 cells, 99% of the MTBP binding sites contain this modification. Significantly, a number of studies have implicated methylation of H3K4 in DNA replication (Liu et al., 2010; Rondinelli et al., 2015). MLL/KMT2A, a methyl-transferase for H3K4, regulates binding of Cdc45 to chromatin (Liu et al., 2010). Furthermore, JARID1C/KDM5C/SMCX, which demethylates H3K4me3 to di- and mono-methylated forms, is also involved in the loading of Cdc45 onto chromatin (Rondinelli et al., 2015).

Treslin-MTBP must presumably interact with MCM complexes to carry out activation of the replicative helicase. However, a number of observations argue that interaction of Treslin-MTBP with MCMs is not sufficient to explain its localization to chromatin. For example, we have observed both pre-RC-dependent and pre-RC-independent binding of Treslin-MTBP to chromatin in *Xenopus* egg extracts (Kumagai et al., 2010). The pre-RC-independent mode is especially prominent in egg extracts treated with the CDK-inhibitor p27 to block initiation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Furthermore, in *Xenopus* NPE fractions, where the pre-RC cannot form, we observe rapid and robust binding of Treslin-MTBP to chromatin.

Another consideration is that MCM complexes are more abundant and broadly distributed throughout the genome than Treslin-MTBP (Sugimoto et al., 2018). MCMs are loaded onto chromatin in a large surplus (McIntosh and Blow, 2012). Furthermore, inactive MCM complexes can also move on the DNA after loading due to transcription and other processes. For example, the transcriptional machinery can essentially push MCMs along the DNA (Gros et al., 2015; Powell et al., 2015). Overall, it seems that mechanisms besides or in addition to interaction with MCMs are responsible for the recruitment and/or maintenance of Treslin-MTBP on chromatin. As described here, we have identified multiple genomic features that correlate with the binding of Treslin-MTBP to chromatin.

An important feature of transcriptional regulation is that various proteins at promoters and enhancers/super-enhancers associate by looping of the DNA. Because Treslin-MTBP localizes to these regions, it would also likely have the opportunity to interact with distant sequences by such looping mechanisms. It is known that higher-order organization of chromatin is important for DNA replication (Guillou et al., 2010; Sima et al., 2019). We have observed enrichment of YY1 and increased occurrence of DNA loops in early replication zones with elevated concentrations of MTBP. Depletion of YY1 does not affect the global replication timing program (Sima et al., 2019), but the effect on local patterns of replication is unknown. Moreover, additional mechanisms could contribute to higher-order chromatin organization in early replication zones. A looping mechanism could allow Treslin-MTBP to promote a more dispersed pattern of initiation around its binding sites. In this scheme, the Treslin-MTBP complex could occupy relatively specific sites but nonetheless activate more broadly distributed MCM complexes. This process could involve direct, looping-mediated contact. Another possibility is that increased local concentrations of the two complexes relative to one another due to higher-order chromatin organization could favor diffusion-mediated interactions.

Promoters and enhancers/super-enhancers contain proteins such as Med1 and Brd4 that contain intrinsically disordered regions (IDRs) and form liquid-liquid phase condensates, which may promote gene expression (Chong et al., 2018; Sabari et al., 2018). Likewise, various pre-RC components possess IDRs, which may facilitate loading of MCMs (Parker et al., 2019). Interestingly, both Treslin and MTBP contain extensive potential IDRs in their C-terminal regulatory domains (Dosztányi et al., 2005) (A.K. and W.G.D., unpublished data). Brd4, a known IDR-containing protein, also interacts with Treslin (Sansam et al., 2018). Although speculative, incorporation of Treslin-MTBP into phase-separated condensates that form among clustered loops of chromatin could also serve to regulate initiation.

In closing, localization of Treslin-MTBP to transcriptional-regulatory elements underscores the importance of these regions for initiation of DNA replication. In the future, it will be important to understand how these sites switch between replicational and transcriptional modes in an orderly manner to maintain genomic integrity.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

STAR★METHODS

RESOURCE AVAILABILITY

Lead Contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, William G. Dunphy (dunphy@caltech.edu).

Materials Availability—Unique reagents and resources generated in this study are available upon request.

Data and Code Availability—Raw sequencing data and processed files were deposited at the NCBI GEO under accession number GSE143481. Further details on data processing are described in Supplemental Information (see Data S1) and in a GitHub repository (https://github.com/dunphylab/kumagai_dunphy_2020).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human tissue culture cells—Human DLD-1 cells (ATCC CCL-221) were cultured in DMEM supplemented with 10% fetal bovine serum, 50 U/ml penicillin, and 50 μg/ml streptomycin. Cells were tested routinely for mycoplasma by DAPI staining.

***Xenopus* egg extracts—**Nucleoplasmic extract (NPE) fractions from *Xenopus* egg extracts were prepared and processed as described previously (Kumagai and Dunphy, 2017). All procedures were approved by the Caltech Institutional Animal Care and Use Committee (IACUC).

METHOD DETAILS

Selection of cell line for CUT&RUN studies—In order to map the binding sites for Treslin-MTBP in the human genome, we set out to attach a 3X-FLAG tag onto the endogenous MTBP by using the CRISPR-Cas9 system. Our previous studies have indicated that Treslin and MTBP associate quantitatively with one another and also mutually depend upon each other for stability (Kumagai and Dunphy, 2017). Accordingly, the localization of MTBP would reflect the presence of the Treslin-MTBP complex. As a consideration for this approach, we also wished to verify that cells could replicate their DNA normally in the presence of only a FLAG-tagged version of MTBP. We also wanted to establish a system in which we could examine mutants of MTBP. We used a multi-step strategy to address these issues. First, we tagged the endogenous copies of MTBP in a human cell line with the mini auxin-inducible degron (mAID). We initially considered using HCT-116 colon cancer cells for these experiments due to the fact that use of the AID system has been well characterized in these cells (Natsume et al., 2016). However, HCT-116 cells have three copies of the gene encoding MTBP due to a chromosomal duplication, which complicated the use of CRISPR-Cas9 technology (Tym et al., 2016). Accordingly, we chose to use human DLD-1 cells, a similar colorectal adenocarcinoma cell line. Other advantages of DLD-1 cells are that they are pseudo-diploid and maintain relatively normal chromosomes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Preparation of DLD-1-OsTIR1 cells—To introduce OsTIR1 into the safe-harbor AAVS1 locus, we transfected DLD-1 cells (2 mL culture) with 1 μg AAVS1 T2 CRISPR in pX330 (Addgene #72833) and 1μg pMK232 (CMV-OsTIR1-PURO) (Addgene #72834) by using 8 μl Lipofectamine 2000. After 48 hr, cells were diluted and plated in DMEM containing 2 μg/ml puromycin. After 12 days, clones were picked and grown on 24-well plates. Genotyping was performed for each clone by PCR with LongAmp Taq DNA polymerase.

Plasmid construction—CRISPR/Cas9 vectors were constructed according to a standard protocol (Ran et al., 2013) using pX330-U6-Chimeric_BB-CBh-hSpCas9-hGem(1/110) (Gutschner et al., 2016) with guide oligonucleotides for Exons 19 and 22 of MTBP to yield pX330-Cas9-MTBP-Exon19 and pX330-Cas9-MTBP-Exon22, respectively. Donor template arms for Exons 19 and 22 of MTBP (1,600 bp each) were amplified from genomic DNA of DLD-1 cells using Q5 DNA polymerase and cloned into pBluescript to yield pBlue-script-MTBP-Exon19 and pBluescript-MTBP-Exon22, respectively. DNA fragments corresponding to mAID-Neo, mAID-Hygro, and 3X-FLAG-Hygro were derived from pMK286 (Addgene plasmid # 72824), pMK287 (Addgene plasmid # 72825), and pMK284 (Addgene plasmid # 72800), respectively. These fragments were cloned into pBluescript-MTBP-Exon22 to make an in-frame fusion of the MTBP C-terminal end and these tags (pBluescript-MTBP-mAID-Neo, pBluescript-MTBP-mAID-Hygro, and pBluescript-MTBP-WT-FLAG-Hygro). In addition, the sequence 3X-FLAG-Hygro from pMK284 was cloned into pBlue-script-MTBP-Exon19 to make an in-frame fusion after amino acid 817 in order to delete 87 amino acids from the C-terminal end of MTBP (pBluescript-MTBP-ΔC-FLAG-Hygro). The appropriate silent mutations were introduced into the gRNA target sites of all the donor templates.

Preparation of endogenously tagged cell lines—For preparation of MTBP-mAID/MTBP-mAID and MTBP-mAID/MTBP-WT-FLAG cell lines, DLD-1 cells harboring CMV-OsTIR1 were transfected with pX330-Cas9-MTBP-Exon 22 and combinations of either pBluescript-MTBP-mAID-Neo and pBluescript-MTBP-mAID-Hygro or pBluescript-MTBP-mAID-Neo and pBluescript-MTBP-WT-FLAG-Hygro. Cells were selected in the presence of both 300 μg/ml Hygromycin and 800 μg/ml G418. Drug-resistant clones were selected for genotyping.

For preparation of the MTBP-mAID/MTBP-ΔC-FLAG cell line, DLD-1 cells harboring CMV-OsTIR1 were transfected with pX330-Cas9-MTBP-Exon22 and pBluescript-MTBP-mAID-Neo. Cells were selected in the presence of 800 μg/ml G418. Drug-resistant clones were picked for genotyping to identify cell lines in which both MTBP alleles were tagged with mAID at the C-terminal end. These cells were transfected with pX330-Cas9-MTBP-Exon19 and pBluescript-MTBP-ΔC-FLAG-Hygro. Clones were selected in the presence of both 300 μg/ml Hygromycin and 800 μg/ml G418. Drug-resistant clones were picked for genotyping to identify cell lines with one allele of MTBP-mAID and one allele of MTBP-ΔC-FLAG.

DNA replication assays—Cells were arrested in G1 in the presence of 1 μM Cdk4/6i for 20 hr. Cells were labeled with 10 μM EdU for 20 min either at the end of this arrest or at

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

various times after release from arrest and incubation in the absence or presence of 500 μM auxin. After fixation in 2% formaldehyde for 10 min, cells were permeabilized in phosphate-buffered saline (PBS) containing 0.5% Triton X-100. Click reactions were performed in Tris-buffered saline (TBS) containing 4 μM $CuSO_4$, 4 mM Alexa Fluor 488 azide, and 10 mM sodium ascorbate. Nuclei were stained with 0.5 μg/ml DAPI for 10 min. Images were obtained with a Zeiss LSM 800 microscope with a 20 × Plan-Apochromat objective in wide-field mode. Incorporation of EdU and staining with DAPI in individual nuclei were quantified with CellProfiler. More than 1,500 nuclei were counted for each time point. For determination of replication-timing patterns, images were obtained with a Zeiss LSM 800 laser scanning confocal microscope using a 63 × Plan-Apochromat objective.

CUT&RUN analyses—For depletion of the MTBP-mAID protein, asynchronous cells were incubated in the presence of auxin for 16 hr prior to harvesting. CUT&RUN reactions were performed on 500,000 cells bound to BioMag Plus Concanavalin A beads (Bangs Laboratory) according to the original protocol (version 1) (Skene et al., 2018; Skene and Henikoff, 2017). Cells were incubated with mouse monoclonal anti-FLAG M2 antibody (Sigma) at 10 μg/ml in 200 μl digitonin-wash buffer containing 0.04% digitonin (Millipore #300410). Next, cells were incubated successively with rabbit anti-mouse IgG and protein A-micrococcal nuclease (pA-MNase). Thereafter, $CaCl_2$ was added to initiate digestion with pA-MNase, and DNA fragments were subsequently released by incubation at 37°C for 10 min in Stop buffer containing yeast spike-in DNA. The use of yeast spike-in DNA allowed a quantitative comparison between samples. After removal of magnetic beads containing undigested chromatin, released DNA fragments were purified using the Monarch PCR & DNA Cleanup Kit (NEB). Purified fragments were end-repaired, adaptor-ligated, and PCR-amplified with NEBNext Multiplex Oligos using the NEBNext Ultra II DNA Library Prep Kit. Adapters and PCR primers were removed with AMPure beads (Beckman Coulter). Libraries were sequenced in the paired-end mode (150 bp) at a commercial facility (Novogene) on an Illumina HiSeq platform. For CUT&RUN analysis of H3K4me2, we incubated auxin-treated MTBP-mAID/MTBP-WT-FLAG cells with anti-H3K4me2 rabbit antibodies or control rabbit antibodies and thereafter with pA-MNase.

CUT&RUN data processing and peak calling—CUT&RUN sequence data was processed using Trim Galore (version 0.5.0) to remove adaptor and low-quality reads ($q < 20$). Next, reads were mapped to the human hg38 genome assembly (GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.bowtie_index) as described (Skene et al., 2018) using Bowtie2 (version 2.3.5) with options:–local–very-sensitive-local–no-unal–no-mixed–no-discordant–phred33 -I 10. Reads that mapped to mitochondrial DNA and random contigs were removed. Also, reads were mapped to the yeast genome assembly in order to quantify yeast spike-in DNA. MACS2 (version 2.1.2) was used to call peaks on the reads for each of three replicates for the MTBP-WT and MTBP-ΔC CUT&RUN experiments using options:–keep-dup 1 -f BAMPE–nolambda -q 0.01. Peaks identified in at least two replicates were selected as high-confidence peaks. For H3K4me2 CUT&RUN, we performed the experiments twice and identified peaks that overlapped in the replicates. Peaks were called again on pooled replicated reads in MACS2 and reproducible peaks were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

identified. These reproducible peaks were used for analyses unless indicated otherwise. Summits were also called in MACS2 and the highest summit for each peak was identified.

Generation of genome browser track files—Genome browser coverage track (bigwig) files were generated using Deeptools (version 3.3.0) bamCoverage with parameters:–binsize 10–normalizeUsing RPGC (reads per genomic content)–ignoreDuplicates–blackListFileName hg38.blacklist.bed (obtained from http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg38-human/)–scaleFactor x. Spike-in-ratios (number of reads aligned to human genome)/(number of reads aligned to yeast genome) were calculated for each experiment. Scale factors were calculated by using spike-in-ratio (control or MTBP-ΔC)/spike-in-ratio (MTBP-WT) to normalize the reads to MTBP-WT. Scale factors ranged between 0.82 and 1.09. Three replicated genome browser track files were combined using bigWigMerge and the resultant bedgraph file was converted to a bigwig file with bedGraphToBigWig. Quantification of peaks was performed with multiBigwigSummary in Deeptools using the BED-file mode with MTBP peaks identified in MACS2.

Locations of TSSs, enhancers, and super-enhancers—Locations of TSSs were downloaded and extracted from GENCODE (gencode.v32.basic.annotation.gtf). Regions from 2 kb upstream to 2 kb downstream of each TSS were designated as promoter-TSS. Locations of enhancer and super-enhancers from HCT-116 cells were obtained from Hnisz et al. (2013).

Annotation of peaks and identification of motifs—Annotation of peaks and determination of the location of peaks relative to TSSs were performed with HOMER (annotatePeaks.pl) using the GENCODE (v32) basic gene annotation file. HOMER (findMotifsGenome.pl) was used to search for motifs enriched in the MTBP peaks. Locations of AP-1, RUNX, and TEAD motifs in the genome were identified using HOMER (scanMotifGenomeWide.pl). G4 motifs in human genome were mapped using G4Hunter with a strict threshold of 2 (Bedrat et al., 2016).

EdU-seq—5×10^6 DLD-1 cells (MTBP-mAID/MTBP-WT-FLAG and MTBP-mAID/MTBP-ΔC-FLAG) were arrested in G1 in the presence of 1 μM Cdk4/6i for 20 hr. Cells were washed and released from the arrest into fresh medium containing 500 μM auxin, 10 μM EdU, 2 mM HU, and 0.5 μg/ml aphidicolin for 3 hr (WT_EdU and ΔC_EdU). HU and aphidicolin cause arrest in early S-phase and aphidicolin further limits elongation at existing nascent strands. One batch of MTBP-WT cells was also incubated without release from G1 in the presence of 1 μM Cdk4/6i as well as 500 μM auxin, 10 μM EdU, 2 mM HU, and 0.5 μg/ml aphidicolin for 3 hr (Control_EdU). Following incubation, cells were washed and incubated in fresh medium lacking EdU, HU, and aphidicolin for another 30 min to allow for extension of replication forks. Cells were trypsinized, washed in PBS twice, and fixed with 70% ice-cold ethanol. After an overnight incubation, cells were washed in PBS and incubated in click-reaction buffer containing 100 μM biotin-dPEG$_{11}$-azide (Quanta Biodesign), 2 mM $CuSO_4$, and 10 mM sodium ascorbate in TBS for 15 min. After washing with TBS, cells were suspended in 0.3 mL 10 mM Tris-HCl (pH 8), 0.1 M NaCl, and 10

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mM EDTA and incubated overnight at 65°C in the presence of 1% SDS and 0.3 mg/ml proteinase K. Genomic DNA was isolated by phenol/chloroform/isoamyl alcohol extraction, chloroform extraction, and finally ethanol precipitation. The DNA was sonicated to 300–500 bp with a Branson 450 sonifier. Biotinylated DNA was collected for 2 hr on 20 μl of Dynabeads MyOne Streptavidin T1 magnetic beads in 5 mM Tris-HCl (pH 7.5) containing 1 M NaCl, 0.5 mM EDTA, 0.1% Tween 20, and 1% bovine serum albumin. After extensive washing, the collected DNA was end-repaired and the appropriate adaptor was ligated to the DNA on the beads. Libraries were amplified and sequenced as described above. These experiments were conducted three times.

Data processing for EdU-seq—EdU-seq data was processed and aligned to the human hg38 genome as described for the CUT&RUN analyses. Genome browser coverage track (bigwig) files for EdU-seq were generated using Deeptools (version 3.3.0) bamCoverage with parameters:–binsize 50–normalizeUsing RPGC (reads per genomic content)–ignoreDuplicates–blackListFileName hg38.blacklist.bed (obtained from mitra.stanford.edu/kundaje/release/blacklists/hg38-human). Thus, reads for each experiment were normalized to the whole genome (RPGC, reads per genomic content). Using Deeptools bigwigCompare, log2 ratios of WT_EdU signal over control_EdU signal and ΔC_EdU signal over control_EdU signal were calculated for each replicate. The three replicates were combined to produce EdU_seq_WT.bw and EdU_seq_ΔC.bw files. Peaks of EdU-seq for MTBP-WT were called using HOMER findPeaks (region -size 6000 -minDist 12000 -F 2). Peaks from the three replicates were merged using HOMER mergePeaks to generate the Initiation_zones.txt file. For quantification of average read scores over control, average read scores were computed with Deeptools multiBigwigSummary with three bigwig files each for WT_EdU, ΔC_EdU, and control_EdU in the BED-file mode using the initiation zones file. Control zones were generated from random genomic loci using regioneR.

Data processing for YY1 HiChIP—YY1 HiChIP data was downloaded from GSE99521 (GSM2774000) (Weintraub et al., 2017). Paired-end tags (PETs) with confidence scores > 0.9 were selected. PETs connecting adjacent bins were removed. PETs within initiation zones and randomly generated control zones of matching sizes were counted. Control zones from random genomic loci were made using regioneR.

QUANTIFICATION AND STATISTICAL ANALYSIS

For determination of p values, calculations were performed with Anova, Wilcoxon, or Fisher's exact tests, as indicated in the text, figures, and figure legends. Statistical analyses of densities for distributions of MTBP peaks were carried out with Kolmogorov-Smirnov tests (Table S1). Statistical analyses of associations with genomic regions were performed with permutation tests (Table S2). Numbers of biological replicates are indicated in the figure legends.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ACKNOWLEDGMENTS

We are grateful to Xun Wang, Hiroyuki Hosokawa, and Ellen Rothenberg as well as Fan Gao and Lior Pachter (Bioinformatics Resource Center in the Beckman Institute at Caltech) for valuable discussions. Alexander Varshavsky and Ke Lyu provided helpful comments on the manuscript. We also thank Diane Trout for advice about computation and Igor Antoshechkin for DNA sequencing. Imaging studies were performed in the Caltech Biological Imaging Center (supported by the Beckman Institute and the Arnold and Mabel Beckman Foundation). Kanomi Sasaki-Capela provided technical assistance. This work was supported by NIH grants GM043974 and GM070891 to W.G.D.

REFERENCES

Abdurashidova G, Deganuto M, Klima R, Riva S, Biamonti G, Giacca M, and Falaschi A (2000). Start sites of bidirectional DNA synthesis at the human lamin B2 origin. Science 287, 2023–2026. [PubMed: 10720330]

Bedrat A, Lacroix L, and Mergny JL (2016). Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Res. 44, 1746–1759. [PubMed: 26792894]

Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, Marin JM, and Lemaitre JM (2012). Unraveling cell type-specific and re-programmable human replication origin signatures associated with G-quadruplex consensus motifs. Nat. Struct. Mol. Biol 19, 837–844. [PubMed: 22751019]

Blin M, Le Tallec B, Nähse V, Schmidt M, Brossas C, Millot GA, Prioleau MN, and Debatisse M (2019). Transcription-dependent regulation of replication dynamics modulates genome stability. Nat. Struct. Mol. Biol 26, 58–66. [PubMed: 30598553]

Boos D, Sanchez-Pulido L, Rappas M, Pearl LH, Oliver AW, Ponting CP, and Diffley JF (2011). Regulation of DNA replication through Sld3-Dpb11 interaction is conserved from yeast to humans. Curr. Biol 21, 1152–1157. [PubMed: 21700459]

Boos D, Yekezare M, and Diffley JF (2013). Identification of a heteromeric complex that promotes DNA replication origin firing in human cells. Science 340, 981–984. [PubMed: 23704573]

Cadoret JC, Meisch F, Hassan-Zadeh V, Luyten I, Guillet C, Duret L, Quesneville H, and Prioleau MN (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc. Natl. Acad. Sci. USA 105, 15837–15842. [PubMed: 18838675]

Cayrou C, Coulombe P, Vigneron A, Stanojcic S, Ganier O, Peiffer I, Rivals E, Puy A, Laurent-Chabalier S, Desprat R, and Méchali M (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res. 21, 1438–1449. [PubMed: 21750104]

Cayrou C, Coulombe P, Puy A, Rialle S, Kaplan N, Segal E, and Méchali M (2012). New insights into replication origin characteristics in metazoans. Cell Cycle 11, 658–667. [PubMed: 22373526]

Cayrou C, Ballester B, Peiffer I, Fenouil R, Coulombe P, Andrau JC, van Helden J, and Méchali M (2015). The chromatin environment shapes DNA replication origin organization and defines origin classes. Genome Res. 25, 1873–1885. [PubMed: 26560631]

Chagin VO, Casas-Delucchi CS, Reinhart M, Schermelleh L, Markaki Y, Maiser A, Bolius JJ, Bensimon A, Fillies M, Domaing P, et al. (2016). 4D Visualization of replication foci in mammalian cells corresponding to individual replicons. Nat. Commun 7, 11231. [PubMed: 27052570]

Chen YH, Keegan S, Kahli M, Tonzi P, Fenyö D, Huang TT, and Smith DJ (2019). Transcription shapes DNA replication initiation and termination in human cells. Nat. Struct. Mol. Biol 26, 67–77. [PubMed: 30598550]

Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, Heckert A, Banala S, Lavis L, Darzacq X, and Tijian R (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science 361, eaar2555. [PubMed: 29930090]

Delgado S, Gómez M, Bird A, and Antequera F (1998). Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO J. 17, 2426–2435. [PubMed: 9545253]

Dellino GI, Cittaro D, Piccioni R, Luzi L, Banfi S, Segalla S, Cesaroni M, Mendoza-Maldonado R, Giacca M, and Pelicci PG (2013). Genome-wide mapping of human DNA-replication origins:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

levels of transcription at ORC1 sites regulate origin selection and replication timing. Genome Res. 23, 1–11. [PubMed: 23187890]

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, and Ren B (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380. [PubMed: 22495300]

Dosztányi Z, Csizmok V, Tompa P, and Simon I (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21, 3433–3434. [PubMed: 15955779]

Ekundayo B, and Bleichert F (2019). Origins of DNA replication. PLoS Genet. 15, e1008320. [PubMed: 31513569]

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. [PubMed: 22955616]

Ernst J, and Kellis M (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc 12, 2478–2492. [PubMed: 29120462]

Fragkos M, Ganier O, Coulombe P, and Méchali M (2015). DNA replication origin activation in space and time. Nat. Rev. Mol. Cell Biol 16, 360–374. [PubMed: 25999062]

Fu H, Baris A, and Aladjem MI (2018). Replication timing and nuclear structure. Curr. Opin. Cell Biol 52, 43–50. [PubMed: 29414592]

Ganier O, Prorok P, Akerman I, and Méchali M (2019). Metazoan DNA replication origins. Curr. Opin. Cell Biol 58, 134–141. [PubMed: 31200185]

Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, and Malinverni R (2016). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics 32, 289–291. [PubMed: 26424858]

Giacca M, Zentilin L, Norio P, Diviacco S, Dimitrova D, Contreas G, Biamonti G, Perini G, Weighardt F, Riva S, et al. (1994). Fine mapping of a replication origin of human DNA. Proc. Natl. Acad. Sci. USA 91, 7119–7123. [PubMed: 8041756]

Gros J, Kumar C, Lynch G, Yadav T, Whitehouse I, and Remus D (2015). Post-licensing specification of eukaryotic replication origins by facilitated Mcm2–7 sliding along DNA. Mol. Cell 60, 797–807. [PubMed: 26656162]

Guillou E, Ibarra A, Coulon V, Casado-Vela J, Rico D, Casal I, Schwob E, Losada A, and Méndez J (2010). Cohesin organizes chromatin loops at DNA replication factories. Genes Dev. 24, 2812–2822. [PubMed: 21159821]

Guo ZS, and DePamphilis ML (1992). Specific transcription factors stimulate simian virus 40 and polyomavirus origins of DNA replication. Mol. Cell. Biol 12, 2514–2524. [PubMed: 1317005]

Guo C, Kumagai A, Schlacher K, Shevchenko A, Shevchenko A, and Dunphy WG (2015). Interaction of Chk1 with Treslin negatively regulates the initiation of chromosomal DNA replication. Mol. Cell 57, 492–505. [PubMed: 25557548]

Gutschner T, Haemmerle M, Genovese G, Draetta GF, and Chin L (2016). Post-translational regulation of Cas9 during G1 enhances homology-directed repair. Cell Rep. 14, 1555–1566. [PubMed: 26854237]

Guzman C, and D'Orso I (2017). CIPHER: a flexible and extensive workflow platform for integrative next-generation sequencing data analysis and genomic regulatory element prediction. BMC Bioinformatics 18, 363. [PubMed: 28789639]

Hänsel-Hertsch R, Spiegel J, Marsico G, Tannahill D, and Balasubrama-nian S (2018). Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. Nat. Protoc 13, 551–564. [PubMed: 29470465]

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589. [PubMed: 20513432]

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, and Young RA (2013). Super-enhancers in the control of cell identity and disease. Cell 155, 934–947. [PubMed: 24119843]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Hoshina S, Yura K, Teranishi H, Kiyasu N, Tominaga A, Kadoma H, Na-katsuka A, Kunichika T, Obuse C, and Waga S (2013). Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. J. Biol. Chem 288, 30161–30171. [PubMed: 24003239]

Karnani N, Taylor C, Malhotra A, and Dutta A (2007). Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. Genome Res. 17, 865–876. [PubMed: 17568004]

Karnani N, Taylor CM, Malhotra A, and Dutta A (2010). Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. Mol. Biol. Cell 21, 393–404. [PubMed: 19955211]

Keller H, Kiosze K, Sachsenweger J, Haumann S, Ohlenschläger O, Nuutinen T, Syvaöja JE, Görlach M, Grosse F, and Pospiech H (2014). The intrinsically disordered amino-terminal region of human RecQL4: multiple DNA-binding domains confer annealing, strand exchange and G4 DNA binding. Nucleic Acids Res. 42, 12614–12627. [PubMed: 25336622]

Khan A, and Mathelier A (2017). Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics 18, 287. [PubMed: 28569135]

Köhler K, Sanchez-Pulido L, Höfer V, Marko A, Ponting CP, Snijders AP, Feederle R, Schepers A, and Boos D (2019). The Cdk8/19-cyclin C transcription regulator functions in genome replication through metazoan Sld7. PLoS Biol. 17, e2006767. [PubMed: 30695077]

Kuhn RM, Haussler D, and Kent WJ (2013). The UCSC genome browser and associated tools. Brief. Bioinform 14, 144–161. [PubMed: 22908213]

Kumagai A, and Dunphy WG (2017). MTBP, the partner of Treslin, contains a novel DNA-binding domain that is essential for proper initiation of DNA replication. Mol. Biol. Cell 28, 2998–3012. [PubMed: 28877985]

Kumagai A, Shevchenko A, Shevchenko A, and Dunphy WG (2010). Treslin collaborates with TopBP1 in triggering the initiation of DNA replication. Cell 140, 349–359. [PubMed: 20116089]

Kumagai A, Shevchenko A, Shevchenko A, and Dunphy WG (2011). Direct regulation of Treslin by cyclin-dependent kinase is essential for the onset of DNA replication. J. Cell Biol 193, 995–1007. [PubMed: 21646402]

Langley AR, Gräf S, Smith JC, and Krude T (2016). Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq). Nucleic Acids Res. 44, 10230–10247. [PubMed: 27587586]

Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. [PubMed: 22388286]

Lashgari A, Millau JF, Jacques PE, and Gaudreau L (2017). Global inhibition of transcription causes an increase in histone H2A.Z incorporation within gene bodies. Nucleic Acids Res. 45, 12715–12722. [PubMed: 29036442]

Lee W, Haslinger A, Karin M, and Tjian R (1987). Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. Nature 325, 368–372. [PubMed: 3027570]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R; 1000 Genome Project Data Processing Sub-group (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. [PubMed: 19505943]

Liu H, Takeda S, Kumar R, Westergard TD, Brown EJ, Pandita TK, Cheng EH, and Hsieh JJ (2010). Phosphorylation of MLL by ATR is required for execution of mammalian S-phase checkpoint. Nature 467, 343–346. [PubMed: 20818375]

Long H, Zhang L, Lv M, Wen Z, Zhang W, Chen X, Zhang P, Li T, Chang L, Jin C, et al. (2020). H2A.Z facilitates licensing and activation of early replication origins. Nature 577, 576–581. [PubMed: 31875854]

MacAlpine DM, Rodríguez HK, and Bell SP (2004). Coordination of replication and transcription along a Drosophila chromosome. Genes Dev. 18, 3094–3105. [PubMed: 15601823]

MacAlpine HK, Gordân R, Powell SK, Hartemink AJ, and MacAlpine DM (2010). Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Res. 20, 201–211. [PubMed: 19996087]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Macheret M, and Halazonetis TD (2018). Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. Nature 555, 112–116. [PubMed: 29466339]

Marchal C, Sima J, and Gilbert DM (2019). Control of DNA replication timing in the 3D genome. Nat. Rev. Mol. Cell Biol 20, 721–737. [PubMed: 31477886]

McIntosh D, and Blow JJ (2012). Dormant origins, the licensing checkpoint, and the response to replicative stresses. Cold Spring Harb. Perspect. Biol 4, a012955. [PubMed: 22904560]

McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Kar-hohs KW, Doan M, Ding L, Rafelski SM, Thirstrup D, et al. (2018). CellProfiler 3.0: next-generation image processing for biology. PLoS Biol. 16, e2005970. [PubMed: 29969450]

Miotto B, Ji Z, and Struhl K (2016). Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. Proc. Natl. Acad. Sci. USA 113, E4810–E4819. [PubMed: 27436900]

Natsume T, Kiyomitsu T, Saga Y, and Kanemaki MT (2016). Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. Cell Rep. 15, 210–218. [PubMed: 27052166]

Olshen AB, Bengtsson H, Neuvial P, Spellman PT, Olshen RA, and Se-shan VE (2011). Parent-specific copy number in paired tumor-normal studies using circular binary segmentation. Bioinformatics 27, 2038–2046. [PubMed: 21666266]

Parker MW, Botchan MR, and Berger JM (2017). Mechanisms and regulation of DNA replication initiation in eukaryotes. Crit. Rev. Biochem. Mol. Biol 52, 107–144. [PubMed: 28094588]

Parker MW, Bell M, Mir M, Kao JA, Darzacq X, Botchan MR, and Berger JM (2019). A new class of disordered elements controls DNA replication through initiator self-assembly. eLife 8, e48562. [PubMed: 31560342]

Picard F, Cadoret JC, Audit B, Arneodo A, Alberti A, Battail C, Duret L, and Prioleau MN (2014). The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. PLoS Genet. 10, e1004282. [PubMed: 24785686]

Powell SK, MacAlpine HK, Prinz JA, Li Y, Belsky JA, and MacAlpine DM (2015). Dynamic loading and redistribution of the Mcm2–7 helicase complex through the cell cycle. EMBO J. 34, 531–543. [PubMed: 25555795]

Prioleau MN (2017). G-Quadruplexes and DNA replication origins. Adv. Exp. Med. Biol 1042, 273–286. [PubMed: 29357063]

Prioleau MN, and MacAlpine DM (2016). DNA replication origins—where do we begin? Genes Dev. 30, 1683–1697. [PubMed: 27542827]

Prorok P, Artufel M, Aze A, Coulombe P, Peiffer I, Lacroix L, Guédin A, Mergny JL, Damaschke J, Schepers A, et al. (2019). Involvement of G-quadruplex regions in mammalian replication origin activity. Nat. Commun 10, 3274. [PubMed: 31332171]

Puzzi L, Marchetti L, Peverali FA, Biamonti G, and Giacca M (2015). DNA-protein interaction dynamics at the Lamin B2 replication origin. Cell Cycle 14, 64–73. [PubMed: 25483070]

Quinlan AR, and Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. [PubMed: 20110278]

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, and Manke T (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44 (W1), W160–W165. [PubMed: 27079975]

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, and Zhang F (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc 8, 2281–2308. [PubMed: 24157548]

Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, et al. (2017). Cohesin loss eliminates all loop domains. Cell 171, 305–320.e24. [PubMed: 28985562]

Rhind N, and Gilbert DM (2013). DNA replication timing. Cold Spring Harb. Perspect. Biol 5, a010132. [PubMed: 23838440]

Rokavec M, Horst D, and Hermeking H (2017). Cellular model of colon cancer progression reveals signatures of mRNAs, miRNA, lncRNAs, and epigenetic modifications associated with metastasis. Cancer Res. 77, 1854–1867. [PubMed: 28130225]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Rondinelli B, Schwerer H, Antonini E, Gaviraghi M, Lupi A, Frenquelli M, Cittaro D, Segalla S, Lemaitre JM, and Tonon G (2015). H3K4me3 demethylation by the histone demethylase KDM5C/JARID1C promotes DNA replication origin firing. Nucleic Acids Res. 43, 2560–2574. [PubMed: 25712104]

Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, Abraham BJ, Hannett NM, Zamudio AV, Manteiga JC, et al. (2018). Co-activator condensation at super-enhancers links phase separation and gene control. Science 361, eaar3958. [PubMed: 29930091]

Sansam CL, Cruz NM, Danielian PS, Amsterdam A, Lau ML, Hopkins N, and Lees JA (2010). A vertebrate gene, ticrr, is an essential checkpoint and replication regulator. Genes Dev. 24, 183–194. [PubMed: 20080954]

Sansam CG, Goins D, Siefert JC, Clowdus EA, and Sansam CL (2015). Cyclin-dependent kinase regulates the length of S phase through TICRR/TRESLIN phosphorylation. Genes Dev. 29, 555–566. [PubMed: 25737283]

Sansam CG, Pietrzak K, Majchrzycka B, Kerlin MA, Chen J, Rankin S, and Sansam CL (2018). A mechanism for epigenetic control of DNA replication. Genes Dev. 32, 224–229. [PubMed: 29483155]

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. [PubMed: 22743772]

Schübeler D, Scalzo D, Kooperberg C, van Steensel B, Delrow J, and Groudine M (2002). Genome-wide DNA replication profile for Drosophila mel-anogaster: a link between transcription and replication timing. Nat. Genet 32, 438–442. [PubMed: 12355067]

Sequeira-Mendes J, Díaz-Uriarte R, Apedaile A, Huntley D, Brockdorff N, and Gómez M (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS Genet. 5, e1000446. [PubMed: 19360092]

Siddiqui K, On KF, and Diffley JF (2013). Regulating DNA replication in eu-karya. Cold Spring Harb. Perspect. Biol 5, a012930. [PubMed: 23838438]

Sima J, Chakraborty A, Dileep V, Michalski M, Klein KN, Holcomb NP, Turner JL, Paulsen MT, Rivera-Mulia JC, Trevilla-Garcia C, et al. (2019). Identifying cis elements for spatiotemporal control of mammalian DNA replication. Cell 176, 816–830.e18. [PubMed: 30595451]

Skene PJ, and Henikoff S (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. eLife 6, e21856. [PubMed: 28079019]

Skene PJ, Henikoff JG, and Henikoff S (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat. Protoc 13, 1006–1019. [PubMed: 29651053]

Smith OK, Kim R, Fu H, Martin MM, Lin CM, Utani K, Zhang Y, Marks AB, Lalande M, Chamberlain S, et al. (2016). Distinct epigenetic features of differentiation-regulated replication origins. Epigenetics Chromatin 9, 18. [PubMed: 27168766]

Sugimoto N, Maehara K, Yoshida K, Ohkawa Y, and Fujita M (2018). Genome-wide analysis of the spatiotemporal regulation of firing and dormant replication origins in human cells. Nucleic Acids Res. 46, 6683–6696. [PubMed: 29893900]

Tanaka S, and Araki H (2013). Helicase activation and establishment of replication forks at chromosomal origins of replication. Cold Spring Harb. Perspect. Biol 5, a010371. [PubMed: 23881938]

Tubbs A, Sridharan S, van Wietmarschen N, Maman Y, Callen E, Stanlie A, Wu W, Wu X, Day A, Wong N, et al. (2018). Dual roles of poly(dA:dT) tracts in replication initiation and fork collapse. Cell 174, 1127–1142 e1119. [PubMed: 30078706]

Tym JE, Mitsopoulos C, Coker EA, Razaz P, Schierz AC, Antolin AA, and Al-Lazikani B (2016). canSAR: an updated cancer research and drug discovery knowledgebase. Nucleic Acids Res. 44 (D1), D938–D943. [PubMed: 26673713]

Valton AL, Hassan-Zadeh V, Lema I, Boggetto N, Alberti P, Saintomé C, Riou JF, and Prioleau MN (2014). G4 motifs affect origin positioning and efficiency in two vertebrate replicators. EMBO J. 33, 732–746. [PubMed: 24521668]

Walter J, Sun L, and Newport J (1998). Regulated chromosomal DNA replication in the absence of a nucleus. Mol. Cell 1, 519–529. [PubMed: 9660936]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Weddington N, Stuy A, Hiratani I, Ryba T, Yokochi T, and Gilbert DM (2008). ReplicationDomain: a visualization tool and comparative database for genome-wide replication timing data. BMC Bioinformatics 9, 530. [PubMed: 19077204]

Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, Abraham BJ, Cohen MA, Nabet B, Buckley DL, et al. (2017). YY1 is a structural regulator of enhancer-promoter loops. Cell 171, 1573–1588 e1528. [PubMed: 29224777]

Woodfine K, Fiegler H, Beare DM, Collins JE, McCann OT, Young BD, Debernardi S, Mott R, Dunham I, and Carter NP (2004). Replication timing of the human genome. Hum. Mol. Genet 13, 191–202. [PubMed: 14645202]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, and Liu XS (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137. [PubMed: 18798982]

Zhao H, Sun Z, Wang J, Huang H, Kocher JP, and Wang L (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics 30, 1006–1007. [PubMed: 24351709]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

- Replication initiator Treslin-MTBP binds to at least 30,000 sites in human genome
- Binding sites contain a nucleosome with active chromatin marks, such as H3K4me2
- Binding sites also contain nucleosome-free DNA (G-quadruplex DNA or AP-1 motif)
- Treslin-MTBP is concentrated in early replication zones with elevated DNA looping

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

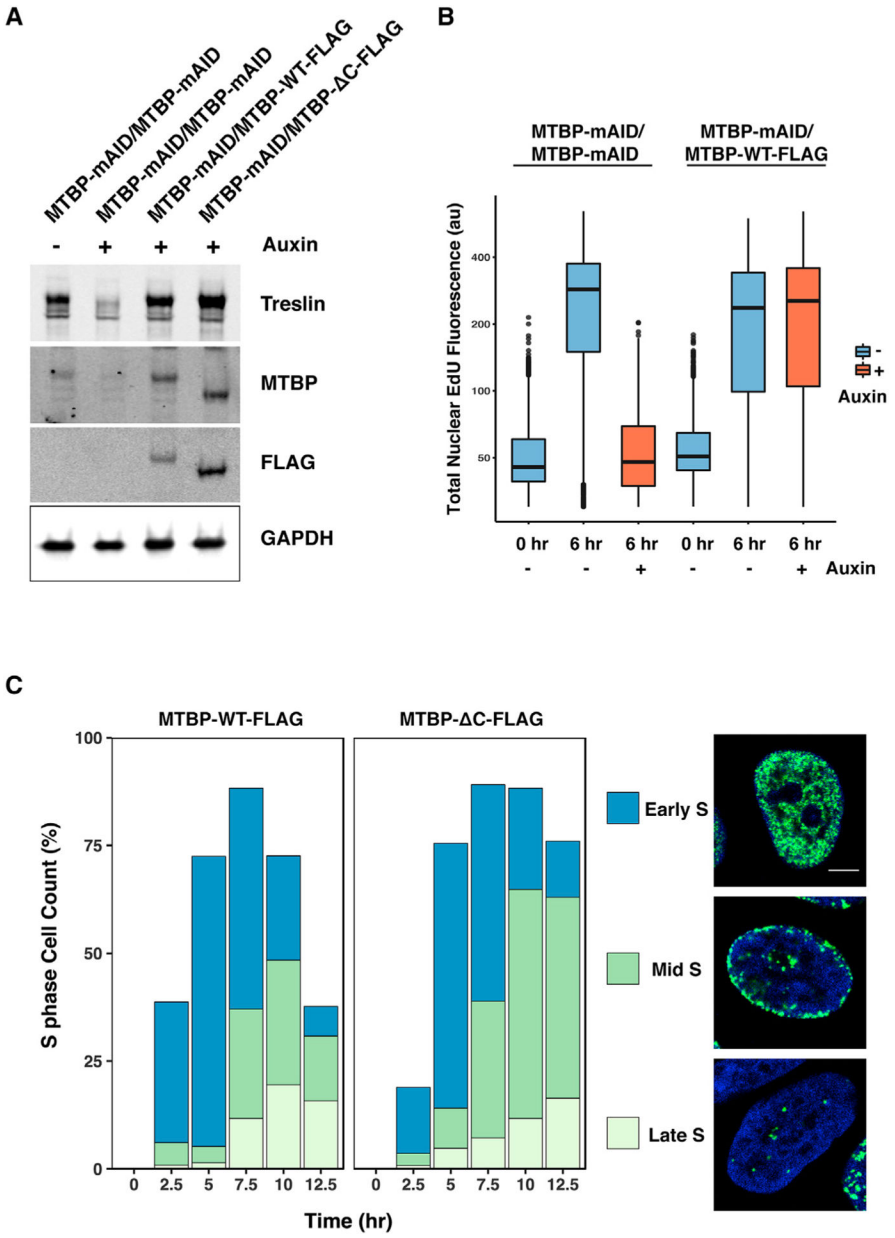


Figure 1. Characterization of Cell Lines Harboring Endogenously Tagged Versions of MTBP

(A) The indicated cell lines were incubated in the absence (lane 1) or presence of auxin (lanes 2 to 4) for 16 h. Cell lysates were immunoblotted with the indicated antibodies. Immunoblots were analyzed with the Odyssey system.

(B) The indicated cell lines were arrested with Cdk4/ 6i. After 20 h, cells were incubated with EdU for 20 min (t = 0) or released from arrest in the presence or absence of auxin for 6 h prior to incubation with EdU for 20 min (t = 6 h). Incorporated EdU was labeled with Alexa 488 azide and quantified (indicated in au [arbitrary units]). Boxplots depict second quartiles, medians, and third quartiles. Representative of two experiments.

(C) The indicated cell lines from (A) were arrested with Cdk4/6i for 20 h. Cells were released from arrest in the presence of auxin and incubated at the indicated times for 20 min

with EdU. Alexa 488-labeled cells were observed by fluorescence microscopy and classified as being in early, mid, or late S-phase (see patterns on right). At least 200 cells were examined per time point. Representative of two experiments. Scale bar, 5 μm.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

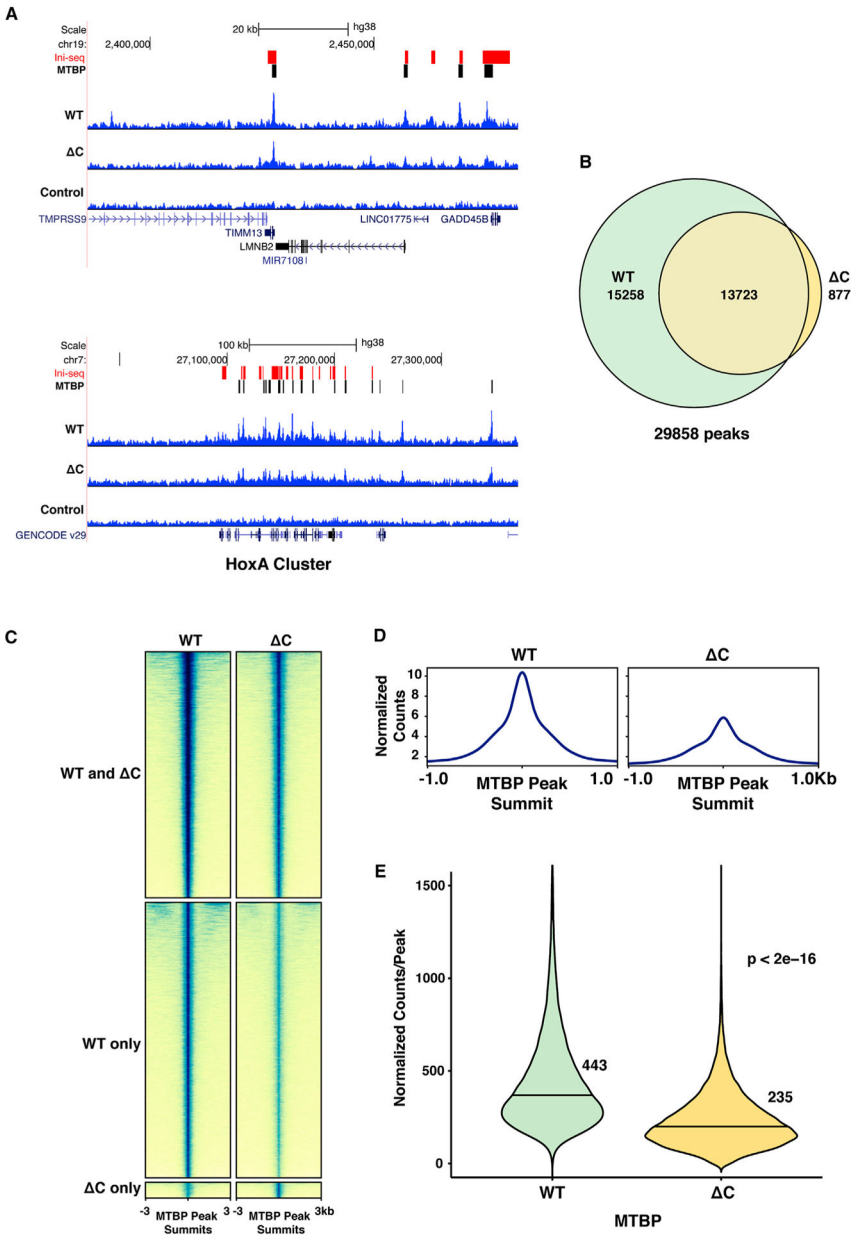


Figure 2. MTBP Associates with Tens of Thousands of Sites in the Human Genome

(A) Browser profiles of regions containing *LMNB2* (top) and the *HoxA* cluster (bottom). Normalized read counts were plotted for CUT&RUN analyses from MTBP-WT, MTBP-ΔC, and control samples. MTBP-WT peaks are indicated with black bars. Replication origins identified by Ini-seq are marked with red bars.

(B) Venn diagram of peaks for MTBP-WT and MTBP-ΔC.

(C) Heatmaps of read distributions for MTBP-WT and MTBP-ΔC centered at the MTBP peak summits. Peaks identified by MACS2 in only MTBP-WT and MTBP-ΔC datasets were denoted as WT only and ΔC only, respectively.

(D) Average normalized read counts for MTBP-WT and MTBP-ΔC were centered at the summits for MTBP peaks.

(E) Violin plots of normalized read counts per peak for MTBP-WT and MTBP-ΔC. Medians are shown as lines and mean values are indicated. Paired Wilcoxon test was performed.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

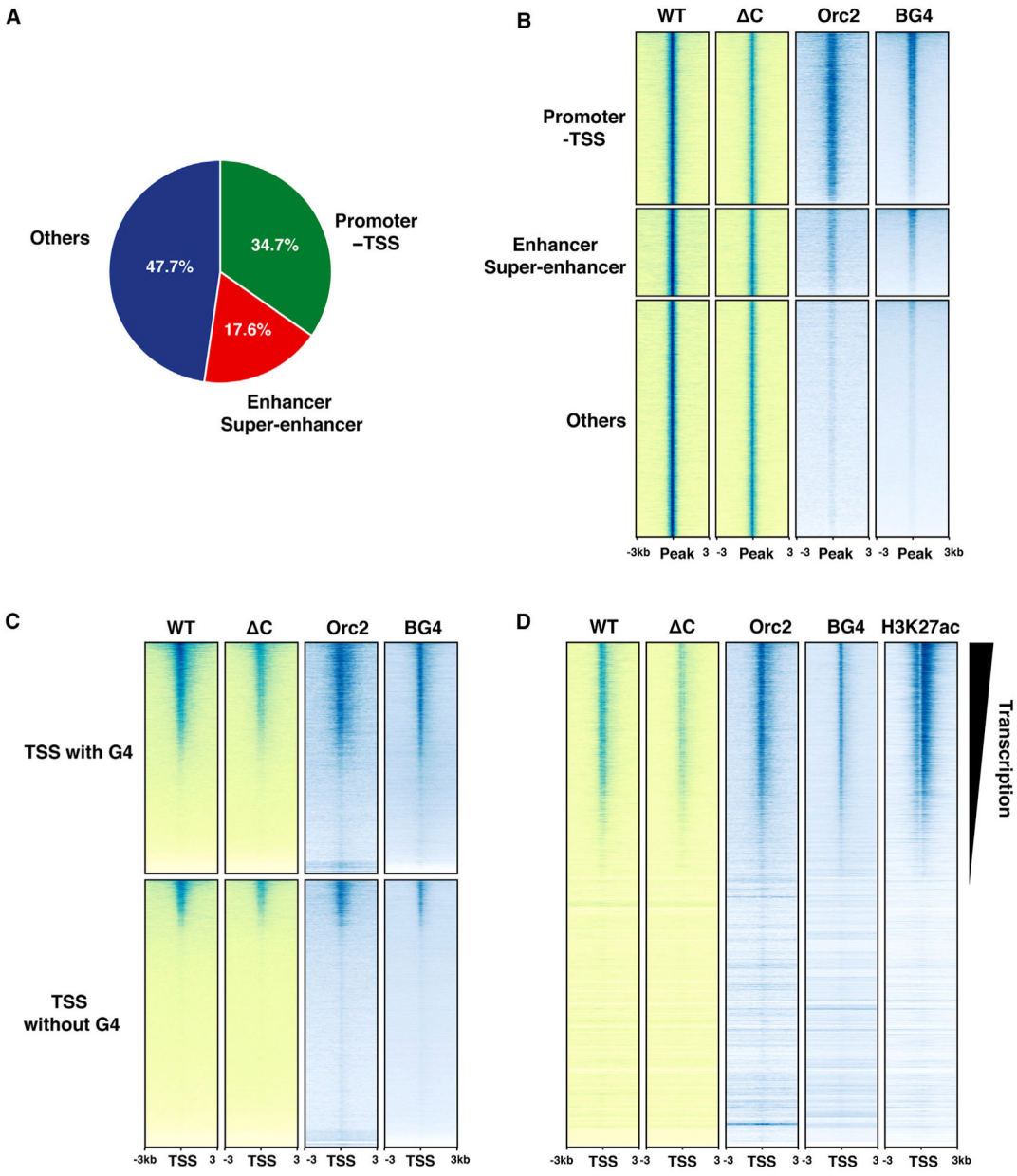


Figure 3. Nature of Binding Sites for MTBP in the Genome

(A) Classification of MTBP binding sites: promoter-TSS (10,365 peaks); enhancer/super-enhancer (5,257 peaks); and others (14,236 peaks).

(B) Heatmaps of MTBP-WT, MTBP-ΔC, Orc2, and BG4 (G4-ChIP) centered around MTBP peak summits (sorted from high to low on the basis of the BG4 data).

(C) TSSs were separated into classes that either contain or lack a G4 structure within 1 kb upstream and downstream. Potential G4 structures were identified with G4Hunter. Heatmaps were centered around TSSs, oriented in the direction of transcription (left to right), and sorted from high to low on the basis of the MTBP-WT data.

(D) TSSs from DLD-1 cells were ordered according to RNA-seq read scores (RPKM, reads per kilobase per million mapped reads). Heatmaps for MTBP-WT, MTBP-ΔC, Orc2, BG4, and H3K27ac were centered around the TSSs. Direction of transcription is left to right.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

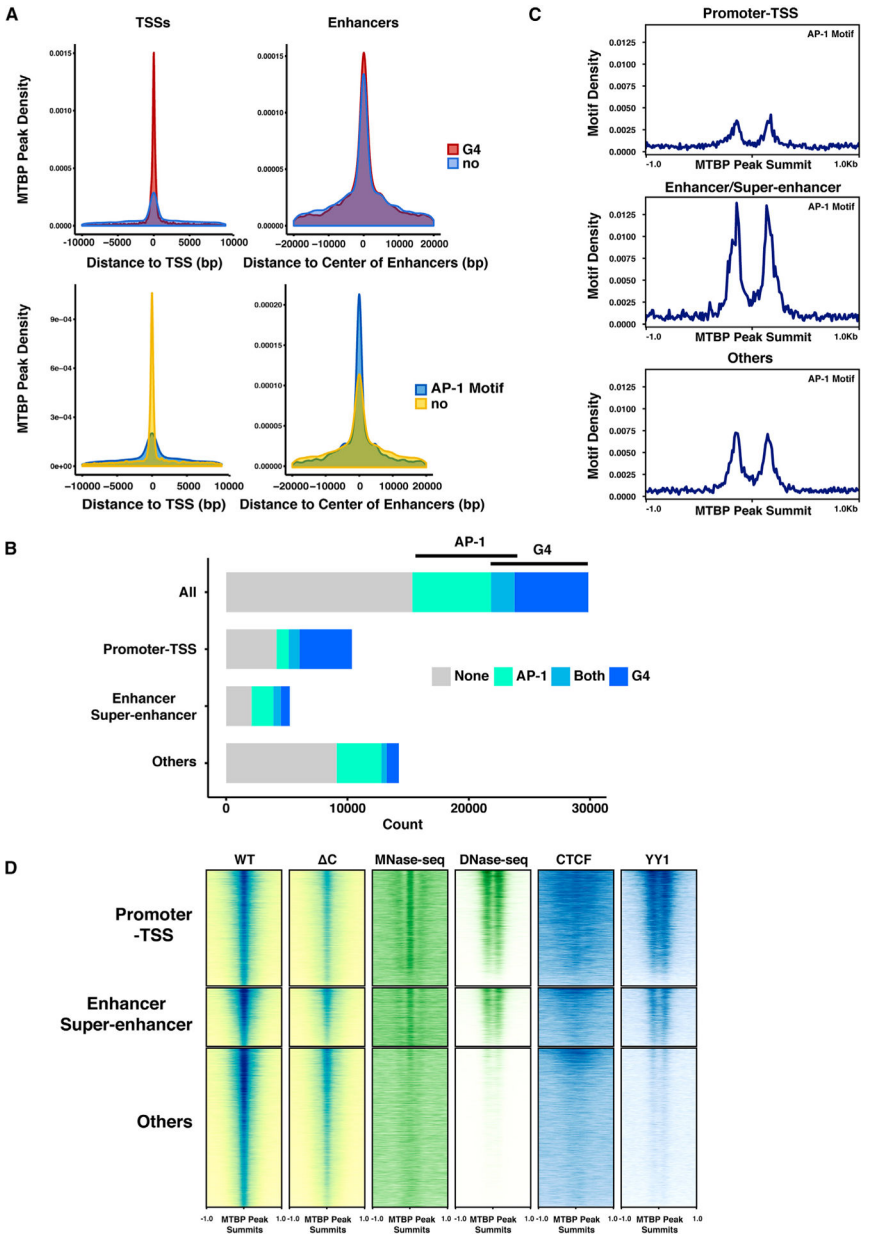


Figure 4. Different Sequence Elements Correlate with the Localization of MTBP at TSSs versus Enhancers

(A) Density plots of MTBP peaks around TSSs and in enhancers. Top two panels: densities of MTBP peaks that contain (red) or lack potential G4 structures (blue) were plotted around TSSs (left) and the centers of enhancers (right). Mean size of enhancers is 4,150 bp. Bottom two panels: densities of MTBP peaks that contain (blue) or lack an AP-1 motif (yellow) were plotted around TSSs (left) and the centers of enhancers (right).

(B) Bar graphs for numbers of MTBP peaks with predicted G4 structures and AP-1 motifs according to location.

(C) Locations of AP-1 motifs mapped around the summits of MTBP peaks in the indicated categories.

(D) Heatmaps of read distributions for MTBP-WT, MTBP-ΔC, MNase-seq, DNase-seq, and ChIP-seq for CTCF and YY1 around MTBP peak summits (scale is indicated in kilobases).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

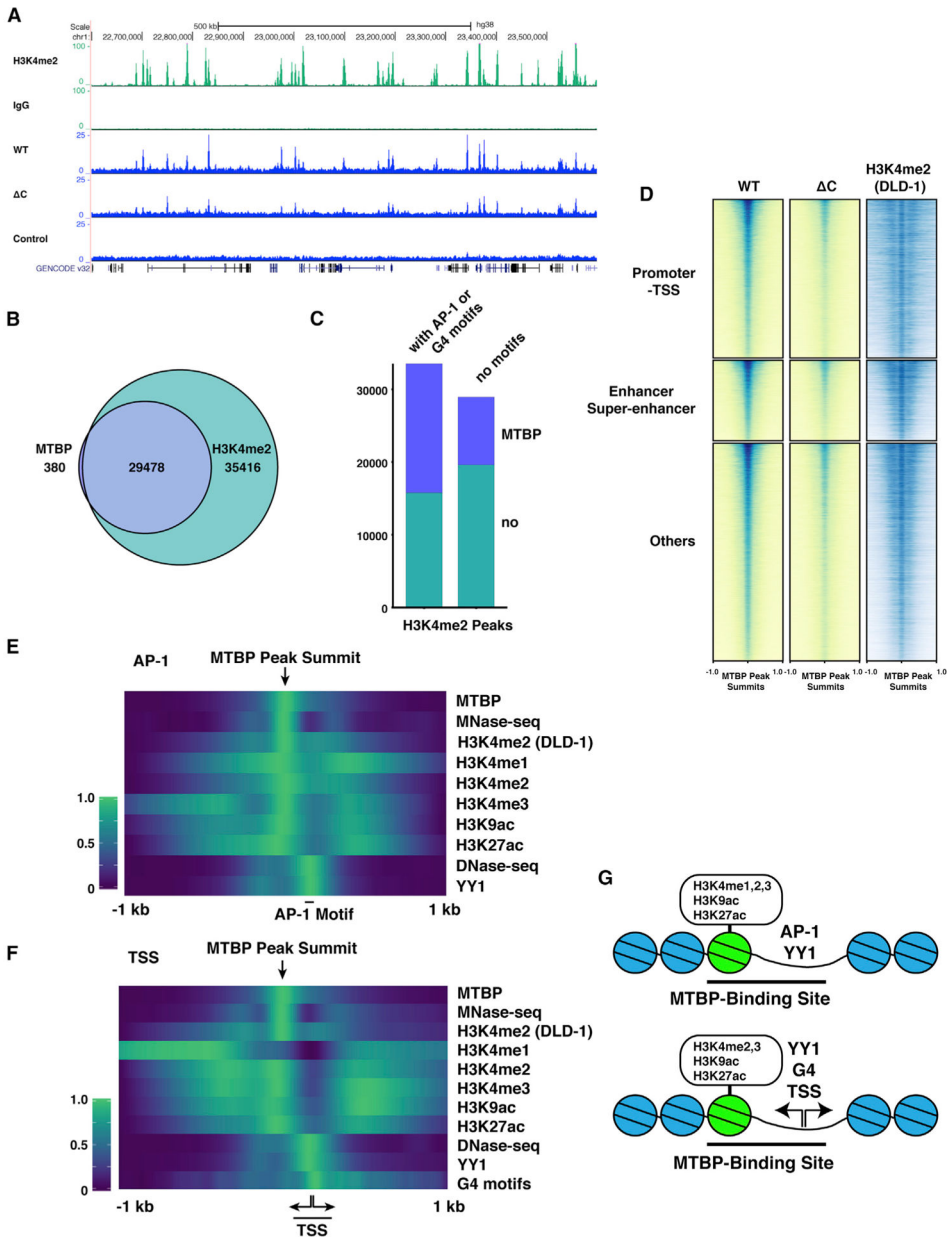


Figure 5. MTBP Associates Coordinately with Adjacent Nucleosome-Containing and Nucleosome-Free Regions

(A) Browser profiles of normalized read counts for MTBP-WT, MTBP-ΔC, and H3K4me2 from CUT&RUN analyses in DLD-1 cells.

(B) Venn diagram of peaks for MTBP-WT and H3K4me2 in DLD-1 cells.

(C) Bar graphs for the distribution of MTBP peaks among H3K4me2-containing regions that contain or lack AP-1 or G4 motifs.

(D) Heatmaps of read distributions for MTBP-WT, MTBP-ΔC, and H3K4me2 from CUT&RUN in DLD-1 cells.

(E) For MTBP peaks containing an AP-1 site, regions 1 kb upstream to 1 kb downstream of the MTBP peak summits were oriented so that the AP-1 site would be on the right side.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Values from the indicated datasets were scaled from 0 to 1 to show the distribution of each signal around the MTBP peak summits. For original data, see Figure S5C.

(F) For MTBP peaks containing a TSS, regions 1 kb upstream to 1 kb downstream of the peak summits were oriented so that the TSS would be on the right side. Values from the indicated datasets were scaled from 0 to 1 to show the distribution of each signal. For original data, see Figure S5D.

(G) Cartoons summarizing features of MTBP binding sites at TSSs and enhancers.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

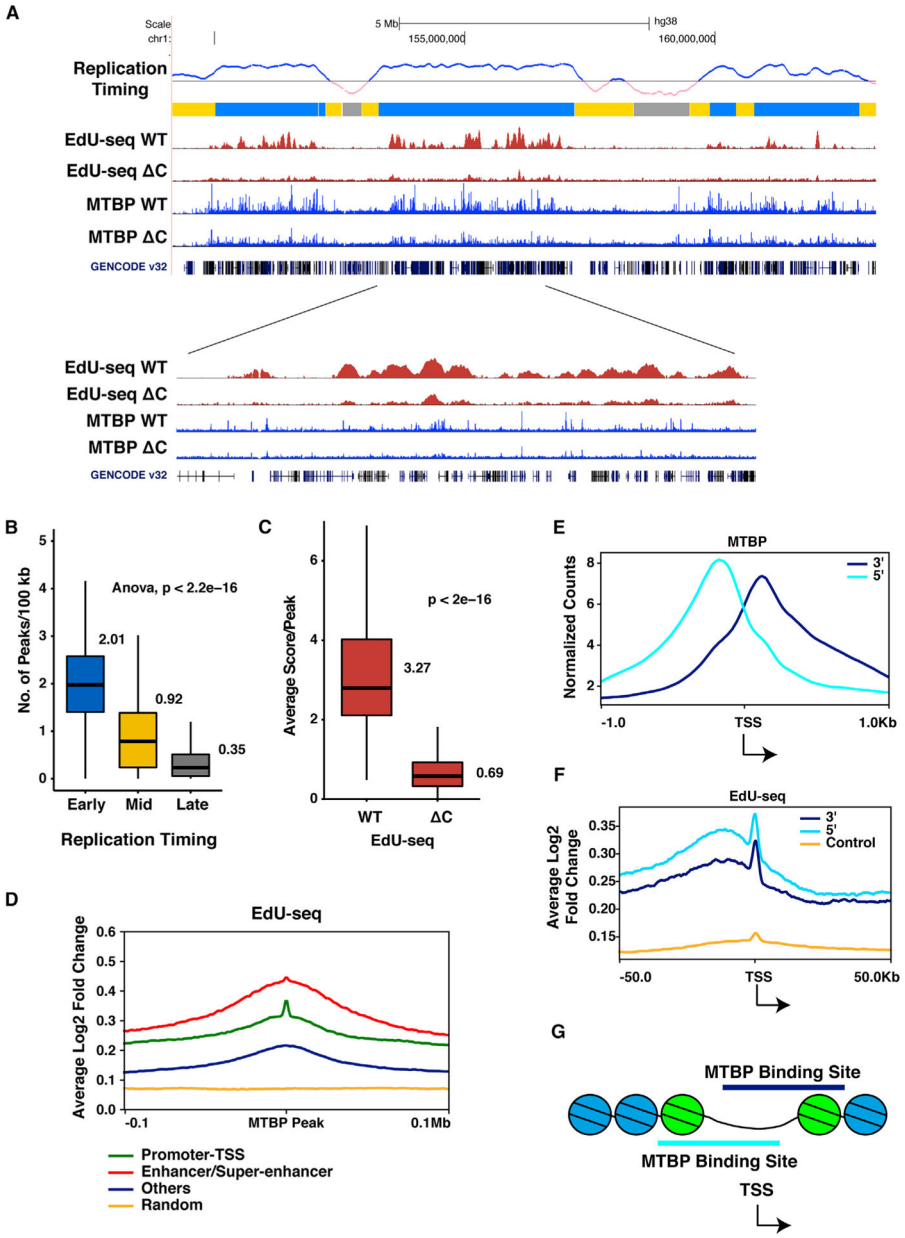


Figure 6. MTBP Is Concentrated in Early-Replicating Zones in DLD-1 Cells

(A) Replication timing tracks from HCT-116 cells, plotted as log2(Early/Late), were downloaded from a public database (see Method Details) and segmented as follows: early (blue); mid (yellow); and late (gray). EdU-seq tracks for MTBP-WT and MTBP-ΔC, plotted as log2(WT or ΔC/control), and CUT&RUN track read coverages for MTBP-WT and MTBP-ΔC are shown. Bottom, blowup of one chromosomal region.

(B) Densities of MTBP-WT peaks in early, mid, and late replication timing zones. Boxplots depict second quartiles, medians, and third quartiles. Mean values are also shown.

(C) Average reads over control of EdU-seq (WT and ΔC) for each initiation zone. Boxplots depict second quartiles, medians, and third quartiles. Mean values are also listed. Paired Wilcoxon test was performed.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(D) Average EdU-seq reads around MTBP-WT peaks in the following regions: promoter-TSS (green); enhancer/super-enhancer (red); and others (blue). Reads in 10,000 randomly chosen positions were also plotted (orange).
(E) MTBP-WT peaks associated with the 5′ (light blue) or 3′ side of TSSs (dark blue), respectively, were separated into two groups. Normalized counts for MTBP signals in each group were plotted.
(F) Average EdU-seq reads (log2 EdU/control) around TSSs with MTBP-WT located at the 5′ (light blue) or 3′ side (dark blue) and around control TSSs (orange).
(G) Cartoon illustrating that MTBP can associate with either side of a TSS.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

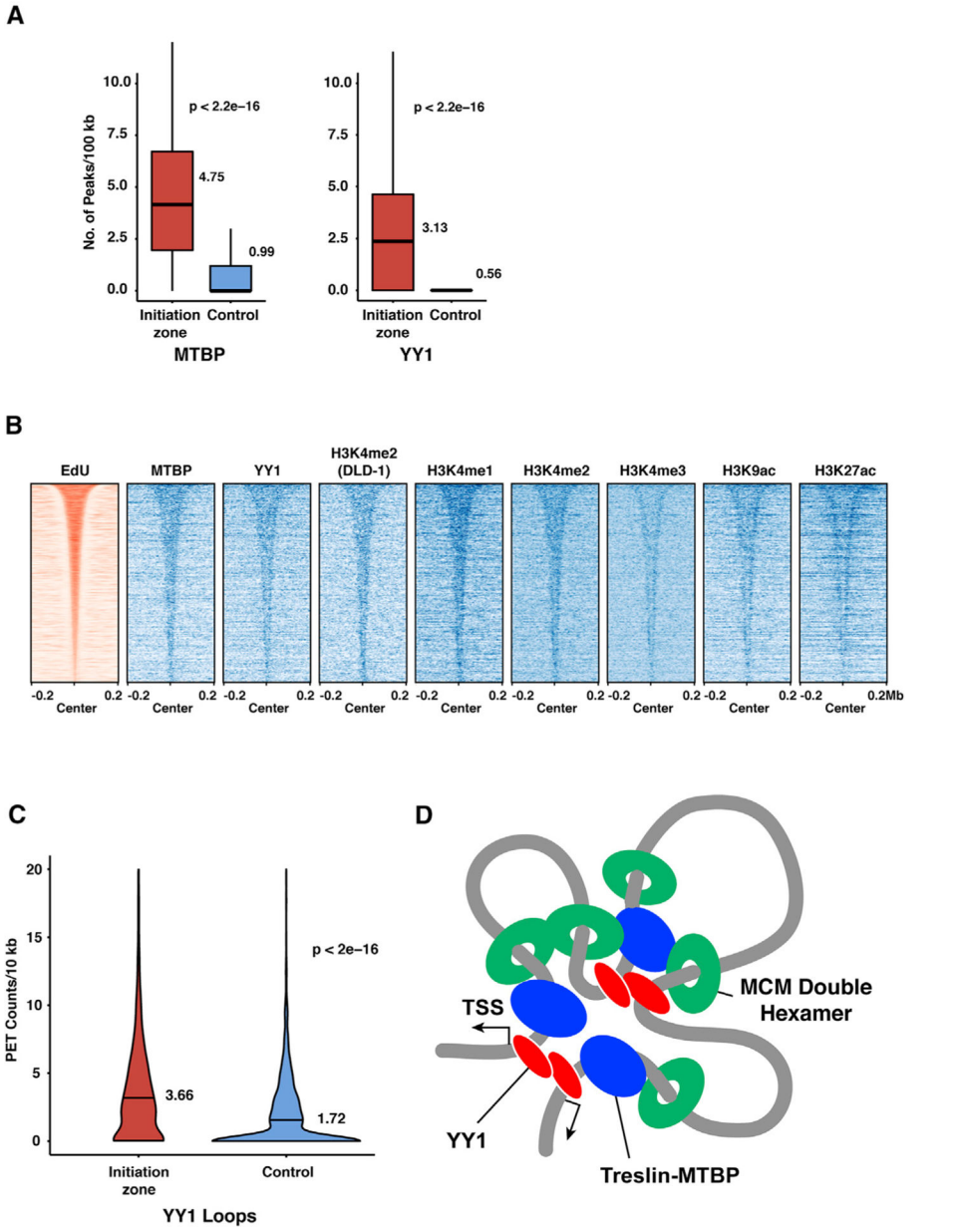


Figure 7. MTBP-Containing, Early-Replicating Zones Contain More DNA Loops

(A) Densities of MTBP-WT peaks (left) and YY1 peaks (right) in initiation and control zones. Boxplots depict second quartiles, medians, and third quartiles. Mean values are also shown. p values were determined by Wilcoxon test.

(B) Heatmap of EdU-seq centered around midpoints of initiation zones. CUT&RUN reads for MTBP-WT and H3K4me2 from DLD-1 cells as well as ChIP-seq data for YY1, H3K4me1, H3K4me2, H3K4me3, H3K9ac, and H3K27ac were plotted in parallel. Initiation zones were sorted according to the size.

(C) Violin plots of PET counts per 10 kb from YY1 HiChIP experiments in initiation and control zones. Medians are shown with lines and mean values are also indicated. p values were determined by Wilcoxon test.

(D) Model for structure of initiation zones. Treslin-MTBP and YY1 are enriched in initiation zones. YY1 could form dimers connecting distal DNA elements. Treslin-MTBP could potentially activate MCMs located at both nearby and distal locations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-FLAG antibody	Sigma-Aldrich	Cat#F1804; RRID: AB_262044
Rabbit anti-mouse IgG antibody	Jackson ImmunoResearch Labs	Cat#315-005-003; RRID: AB_2340033
Anti-Treslin antibodies (human and *Xenopus*)	Kumagai et al., 2010	N/A
Anti-MTBP antibody (B5)	Santa Cruz Biotechnology	Cat#sc-137201; RRID: AB_2147103
Anti-H3K4me2 antibody (C64G9)	Cell Signaling Technology	Cat#9725; RRID: AB_10205451
Anti-GAPDH antibody (14C10)	Cell Signaling Technology	Cat#2118; RRID: AB_561053
Anti-rabbit IgG (Dylight 800 4xPEG conjugate)	Cell Signaling Technology	Cat#5151; RRID: AB_10697505
Anti-mouse IgG (Dylight 800 4xPEG conjugate)	Cell Signaling Technology	Cat# 5257; RRID: AB_10693543
Chemicals, Peptides, and Recombinant Proteins		
Protein A-MNase	Gift from Steven Henikoff	Skene and Henikoff, 2017
Palbociclib (PD-0332991 isethionate; Cdk4/6i)	Sigma-Aldrich	Cat#PZ0199
Auxin (Indole-3-acetic acid)	Sigma-Aldrich	Cat#I5148
EdU (5′-ethynyl-2′-deoxyuridine)	Thermo Fisher Scientific	Cat#C10337
Alexa 488 azide	Thermo Fisher Scientific	Cat#A10266
Biotin-dPEG$_{11}$-azide	Quanta Biodesign	Cat#10784
Dynabeads MyOne Streptavidin T1 beads	Thermo Fisher Scientific	Cat#65601
DAPI (4′,6-Diamidino-2-phenylindole dihydrochloride)	Sigma-Aldrich	Cat#D9542
BioMag Plus Concanavalin A beads	Bangs Laboratory	Cat#BP531
Digitonin	Millipore EMD	Cat#300410
AMPure XP	Beckman Coulter	Cat#A63880
Q5 DNA polymerase	NEB	Cat#M0491S
LongAmp Taq DNA polymerase	NEB	Cat#M0323S
Critical Commercial Assays		
NEBNext Ultra II DNA Library Prep Kit for Illumina	NEB	Cat#E7645S
NEBNext Multiplex Oligos for Illumina (Set 1)	NEB	Cat#E7335S
NEBNext Multiplex Oligos for Illumina (Set 2)	NEB	Cat#E7500S
NEBuilder HiFi DNA Assembly Master Mix	NEB	Cat#E2621S
Monarch PCR & DNA Cleanup Kit	NEB	Cat#T1030S
Raw and Analyzed Data	This paper	GEO: GSE143481
Orc2	Miotto et al., 2016	GEO: GSE70165
CTCF	ENCODE	ENCSR240PRQ
YY1	ENCODE	ENCSR000BNX
DNase-seq	ENCODE	ENCSR000ENM

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MNase-seq	Guzman and D'Orso, 2017	GEO: GSM2391842
H3K4me1	ENCODE	ENCSR161MXP
H3K4me2	ENCODE	ENCSR794ULT
H3K4me3	ENCODE	ENCSR000DTQ
H3K9ac	ENCODE	ENCSR093SHE
H3K27ac	ENCODE	ENCSR000EUT
H3K27me3	ENCODE	ENCSR810BDB
H4K20me1	ENCODE	ENCSR474DOV
H4K20me3	Rao et al., 2017	GEO: GSM2809633
H3K9me2	ENCODE	ENCSR555LYM
H3K9me3	ENCODE	ENCSR179BUC
H3K36me3	ENCODE	ENCSR091QXP
H2A.Z	Lashgari et al., 2017	GEO: GSM2702721
Enhancers and Super-Enhancers	Hnisz et al., 2013	N/A
Ini-seq	Langley et al., 2016	N/A
BG4	Hänsel-Hertsch et al., 2018	GEO: GSM2635754
RNA-seq (DLD-1)	Rokavec et al., 2017	GEO: GSE85685
Replication Timing (HCT-116)	Weddington et al., 2008	https://www2.replicationdomain.com/
YY1 HiChIP	Weintraub et al., 2017	GEO: GSM2774000
Experimental Models: Cell Lines		
Human DLD-1 cells	ATCC	CCL-221
Experimental Models: Organisms/Strains		
Xenopus laevis	Nasco	Cat# LM00535
Oligonucleotides		
Guide Oligonucleotides	This paper	See Table S3.
Oligonucleotides for Cloning of Genomic MTBP Fragments for Donor Templates	This paper	See Table S4.
Oligonucleotides for Construction of Donor Templates	This paper	See Table S5.
Recombinant DNA		
AAVS1 T2 CRISPR in pX330	Natsume et al., 2016	RRID: Addgene_72833
pMK232 (CMV-OsTIR1-PURO)	Natsume et al., 2016	RRID: Addgene_72834
pMK286 (mAID-NeoR)	Natsume et al., 2016	RRID: Addgene_72824
pMK287 (mAID-Hygro)	Natsume et al., 2016	RRID: Addgene_72825
pMK284 (3X-FLAG-Hygro)	Natsume et al., 2016	RRID: Addgene_72800
pX330-U6-Chimeric_BB-CBh-hSpCas9hGem(1/110)	Gutschner et al., 2016	RRID: Addgene_71707
pBluescript-MTBP-mAID-Hygro	This paper	
pBluescript-MTBP-mAID-Neo	This paper	
pBluescript-MTBP-WT-FLAG-Hygro	This paper	
pBluescript-MTBP-DC-FLAG-Hygro	This paper	
Software and Algorithms		

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Fiji (version 2.0.0-rc-69/1.52p)	Schindelin et al., 2012	RRID: SCR_002285
CellProfiler (3.1.5)	McQuin et al., 2018	RRID: SCR_007358
TrimGalore (version 0.5.0)	http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/	RRID: SCR_016946
Bowtie2 (version 2.3.5)	Langmead and Salzberg, 2012	RRID: SCR_005476
Samtools (version 1.9)	Li et al., 2009	RRID: SCR_002105
MACS2 (version 2.1.2)	Zhang et al., 2008	RRID: SCR_013291
Deeptools (version 3.3.0)	Ramírez et al., 2016	RRID: SCR_016366
HOMER (version 4.9.1)	Heinz et al., 2010	RRID: SCR_010881
Bedtools (version 2.27.1)	Quinlan and Hall, 2010	RRID: SCR_006646
Intervene (0.6.4)	Khan and Mathelier, 2017	https://github.com/asntech/intervene
UCSC-genome-browser-tools	Kuhn et al., 2013	https://genome.ucsc.edu/
CrossMap (0.2.7)	Zhao et al., 2014	RRID: SCR_001173
G4Hunter	Bedrat et al., 2016	https://academic.oup.com/nar/article/44/4/1746/1854457
ChromHMM (version 1.18)	Ernst and Kellis, 2017	http://compbio.mit.edu/ChromHMM/
regioneR	Gel et al., 2016	http://bioconductor.org/packages/release/bioc/html/regioneR.html
RStudio (version 1.2.5001)	RStudio Team	RRID: SCR_000432
DNAcopy (version 1.56.0)	Olshen et al., 2011	RRID: SCR_012560
Bioconductor	http://www.bioconductor.org/	RRID: SCR_006442
R (version 3.5.1)	R Core Team	https://www.r-project.org
Benchling	https://benchling.com	RRID: SCR_013955