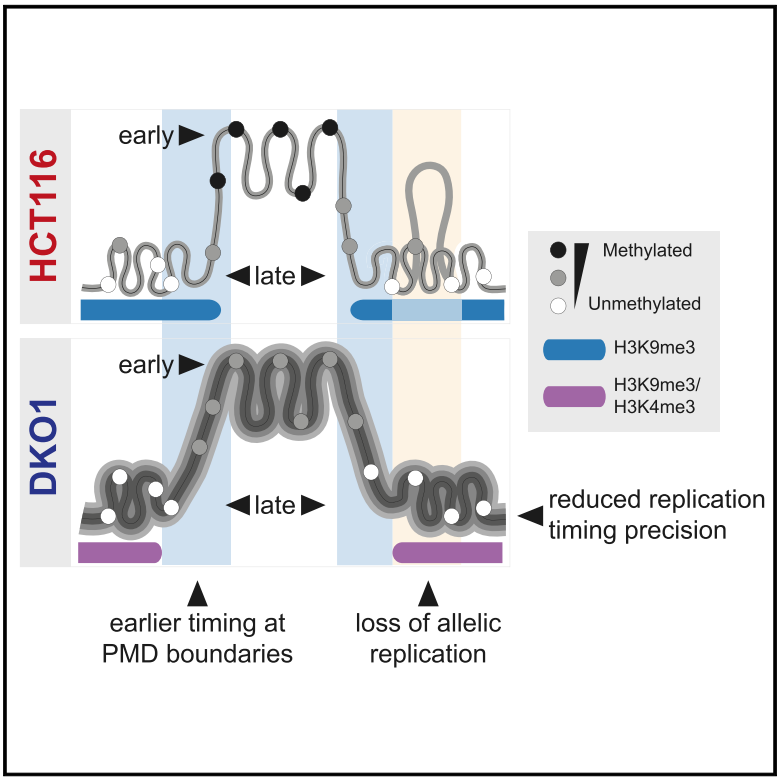
Cell Reports

Article

DNA methylation is required to maintain both DNA replication timing precision and 3D genome organization integrity

Graphical abstract



Highlights

- DNA hypomethylation is associated with single-cell replication timing heterogeneity
- Hypomethylation induces loss of allelic replication at cancer-related gene loci
- Hypomethylation alters higher-order genome architecture of PMD boundaries
- Non-canonical H3K4me3-H3K9me3 domains form to protect silent late replication

Authors

Qian Du, Grady C. Smith, Phuc Loi Luu, ..., Clare Stirzaker, Joanna Achinger-Kawecka, Susan J. Clark

Correspondence

s.clark@garvan.org.au

In brief

Du et al. demonstrate the importance of DNA methylation in the maintenance of 3D genome regulation. They show that hypomethylation, a hallmark of cancer, is associated with the disruption of 3D genome compartmentalization, gain in replication timing heterogeneity, and loss of allelic replication, potentially through the deregulation of DNA replication and genome organization pathways.

Du et al., 2021, Cell Reports *36*, 109722
September 21, 2021
https://doi.org/10.1016/j.celrep.2021.109722

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

DNA methylation is required to maintain both DNA replication timing precision and 3D genome organization integrity

Qian Du,[1,2] Grady C. Smith,[1] Phuc Loi Luu,[1,2] James M. Ferguson,[3] Nicola J. Armstrong,[4] C. Elizabeth Caldon,[1,2] Elyssa M. Campbell,[1] Shalima S. Nair,[1,2] Elena Zotenko,[1] Cathryn M. Gould,[1] Michael Buckley,[1] Kee-Ming Chia,[1] Neil Portman,[1] Elgene Lim,[1,2] Dominik Kaczorowski,[5] Chia-Ling Chan,[5] Kirston Barton,[3] Ira W. Deveson,[2,3] Martin A. Smith,[2,3] Joseph E. Powell,[5,6] Ksenia Skvortsova,[1,2] Clare Stirzaker,[1,2] Joanna Achinger-Kawecka,[1,2] and Susan J. Clark[1,2,7,*]

[1]Garvan Institute of Medical Research, Sydney, NSW 2010, Australia
[2]St Vincent's Clinical School, University of New South Wales, Sydney, NSW 2010, Australia
[3]The Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia
[4]Mathematics and Statistics, Murdoch University, Murdoch, WA 6150, Australia
[5]Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia
[6]UNSW Cellular Genomics Futures Institute, School of Medical Sciences, UNSW Sydney, NSW 2010, Australia
[7]Lead contact
*Correspondence: s.clark@garvan.org.au
https://doi.org/10.1016/j.celrep.2021.109722

SUMMARY

DNA replication timing and three-dimensional (3D) genome organization are associated with distinct epigenome patterns across large domains. However, whether alterations in the epigenome, in particular cancer-related DNA hypomethylation, affects higher-order levels of genome architecture is still unclear. Here, using Repli-Seq, single-cell Repli-Seq, and Hi-C, we show that genome-wide methylation loss is associated with both concordant loss of replication timing precision and deregulation of 3D genome organization. Notably, we find distinct disruption in 3D genome compartmentalization, striking gains in cell-to-cell replication timing heterogeneity and loss of allelic replication timing in cancer hypomethylation models, potentially through the gene deregulation of DNA replication and genome organization pathways. Finally, we identify ectopic H3K4me3-H3K9me3 domains from across large hypomethylated domains, where late replication is maintained, which we purport serves to protect against catastrophic genome reorganization and aberrant gene transcription. Our results highlight a potential role for the methylome in the maintenance of 3D genome regulation.

INTRODUCTION

The genome is arranged into many higher-order architectural layers, including DNA replication timing and three-dimensional (3D) chromatin conformation, which serve to functionally compartmentalize genomic regulation. DNA replication follows a highly organized replication timing program, whereby genomic domains are replicated in a specific temporal order during S phase, from early to late. The genome is also organized in nuclear space into 3D clusters formed by *cis*-chromatin interactions (Dixon et al., 2012; Lieberman-Aiden et al., 2009; Nora et al., 2012; Sexton et al., 2012). These clusters partition the genome into two large-scale compartments: transcriptionally active, open A-compartments and silenced, mostly closed B-compartments (Lieberman-Aiden et al., 2009). Integration of replication timing (Repli-Seq) and 3D chromatin conformation (Hi-C) sequencing data has revealed that early replication timing domains correspond to the active A-compartments and late replication timing to the repressive B-compartments (Dixon et al., 2012; Ryba et al., 2010). Alterations in DNA replication timing and 3D chromatin organization correspond to transcriptional and epigenomic changes during differentiation (Miura et al., 2019; Rivera-Mulia et al., 2015) and carcinogenesis (Achinger-Kawecka et al., 2020; Du et al., 2019; Taberlay et al., 2016). These results highlight that DNA replication timing and 3D genome organization together may play a coordinated role in the higher-order regulation of the genome.

The DNA methylation landscape, in particular, long-range domains of low DNA methylation, partially methylated domains (PMDs), are associated with late-replicating domains (Berman et al., 2011; Du et al., 2019) and B-compartments (Nothjunge et al., 2017; Xie et al., 2017). These associations prompted us to ask whether genome-wide loss of DNA methylation, one of the major epigenomic hallmarks of cancer (Shen and Laird, 2013), can also alter the shape of the DNA replication timing program and 3D genome organization. Previous loci-specific studies reported that DNA methylation loss is related to changes in replication timing at candidate regions. For example, a shift in

replication timing from late to early timing of the inactive chromosome X was associated with DNA demethylation in patients with immunodeficiency disease (ICF) (Hansen et al., 2000), and DNMT1 knockout (KO) mouse embryonic stem cells displayed earlier replication timing of pericentromeric major satellite repeat elements (Jørgensen et al., 2007). However, little is understood about the effect of DNA methylation loss on genome-wide higher-order genome organization.

Here, we address whether global DNA methylation loss that typically occurs in cancer is associated with alterations in DNA replication timing and 3D genome structure. We used Repli-Seq, single-cell Repli-Seq, and Hi-C in a colorectal cancer cell line (HCT116) and its isogenic cell line with double KO (DKO) of maintenance DNA methyltransferase *DNMT1* and *de novo* DNA methyltransferase *DNMT3B* (DKO1) (Rhee et al., 2002), as well as hypomethylated tumors and tumor patient-derived xenograft (PDX) samples treated with decitabine, to show that hypomethylation is associated with concordant loss of replication timing precision and deregulation of 3D genome organization. Our results highlight a previously underappreciated role for DNA methylation in the regulation of the 3D genome.

RESULTS

Coordinate alterations to DNA replication timing and 3D genome organization

To determine the impact of DNA methylation levels on higher-order genome architecture, we investigated changes in both DNA replication timing and 3D chromatin conformation after global DNA hypomethylation. We used a well-described model of DNA methylation loss; HCT116, a colorectal cancer cell line; and DKO1, the isogenic cell line with DKO of the maintenance DNA methyltransferase *DNMT1* and the *de novo* DNA methyltransferase *DNMT3B* (Rhee et al., 2002). We found that the DKO1 cells show DNA hypomethylation genome-wide (~56% methylation loss) compared to HCT116 cells (Figure 1A). To determine whether global DNA methylation loss results in changes in DNA replication timing and/or 3D genome structure, we performed Repli-Seq and *in situ* Hi-C in replicates in HCT116 and DKO1 cells (seeFigures S1 and S2A; Method details).

We examined genome-wide trends in DNA replication timing and Hi-C data between HCT116 and DKO1 cells and found that in contrast to global DNA hypomethylation (Figure 1A), replication timing weighted average (WA) values between HCT116 and DKO1 are highly correlated (Figure 1B, Spearman's = 0.9613). A representative example is shown in Figure 1C. We next used a WA difference of 15 ($|\Delta WA| > 15$) as a stringent approach to define regions with large replication timing alterations, and a WA difference of 5 ($|\Delta WA| < 5$) to measure minimal changes in timing (see Figures S1E and S1F; Method details). We found that 66.81% of the genome shows close conservation of replication timing ($|\Delta WA| < 5$) and that 29.87% of the genome displays a moderate shift in replication timing ($5 < |\Delta WA| < 15$). Notably, at the most stringent cutoff ($|\Delta WA| > 15$), a distinct fraction of the genome displayed a large shift in replication timing, that is, to either replicate earlier (1.96%) or replicate later (1.36%) in DKO1 compared to HCT116 (Figure 1D). These earlier and later regions are on average ~300 kb in size and are primarily in intergenic and intronic regions (Figures S1G and S1H).

Next, to investigate whether loss of DNA methylation also results in alterations to large-scale genome compartmentalization, we performed compartment analyses to define A/B-compartment switching from HCT116 to DKO1. We defined A/B compartment status with the first principal-component (PC1) values, which represent euchromatin/heterochromatin neighborhoods, respectively. Similar to replication timing, we observe that HCT116 and DKO1 Hi-C data show a high correlation between PC1 values (Figures 1C, 1E, and S2B) and comparable proportions of A- and B-compartments (Figure S2C). We found that 13.63% of the genome had switched A/B-compartments comprising switching from A to B (9.09%) and switching B to A (4.54%) (Figure S2D). We found that while compartment switches do cross the midline that defines A- versus B-compartments, a large proportion (~45%) of these switches are at regions with low compartment values, centered around PC1 values of <0.5 and > PC1 >−0.5 (Figures S2E and S2F). Therefore, to identify regions that show strong compartment shifts, we defined switching regions by their ΔPC1 score, using a ΔPC1 cutoff of $\geq |1|$ (Figures 1D and S2G; Method details). Using the more stringent criteria, we found up to 7.75% of the genome architecture had changed; specifically, 4.43% of the genome increased PC1 values in DKO1 and therefore is more A-type than HCT116, and 3.32% of the genome decreased PC1 values, where DKO1 is more B-type than HCT116. ΔPC1 regions are on average ~162 kb in size, and are similar in intergenic and intronic regions (Figures S2H and S2I).

Lastly, we sought to determine whether there is concordance between regions that show changes in DNA replication timing and regions that show a change in genome compartmentalization, following DNA hypomethylation. We found that where DNA replication timing has become earlier in DKO1 compared to HCT116, the region also moves more toward the "active" A-compartment (positive PC1 values) and where DNA replication timing has become later in DKO1, the region moves more toward the "inactive" B-compartment (Figures 1F and 1G). Exemplary regions are shown in Figure 1D. These data support that global DNA hypomethylation is associated with coordinate changes in higher-order genome architecture at distinct domains.

Changes in higher-order genome architecture occur at partially methylated domain boundaries

We next examined whether the genomic domains that displayed the larger changes in replication timing and genome organization are related to the degree of methylation loss. Early replicating loci are highly methylated and are therefore able to lose methylation, while late-replicating loci are lowly methylated and thus less amenable to methylation loss (Figure 1C). Not surprisingly, loci with the largest degree of methylation loss are highly methylated in HCT116, and conversely, loci with no methylation change or methylation gain are lowly methylated in HCT116 (Figure S3A). In particular, we find that loci with DNA methylation loss ($\Delta Meth \leq -0.4$: 45.12% of all measurable CpG sites) are significantly associated with loci that replicated later or showed a negative ΔPC1 difference from HCT116 to DKO1 (Figures S3B and S3C), and these regions were highly methylated in

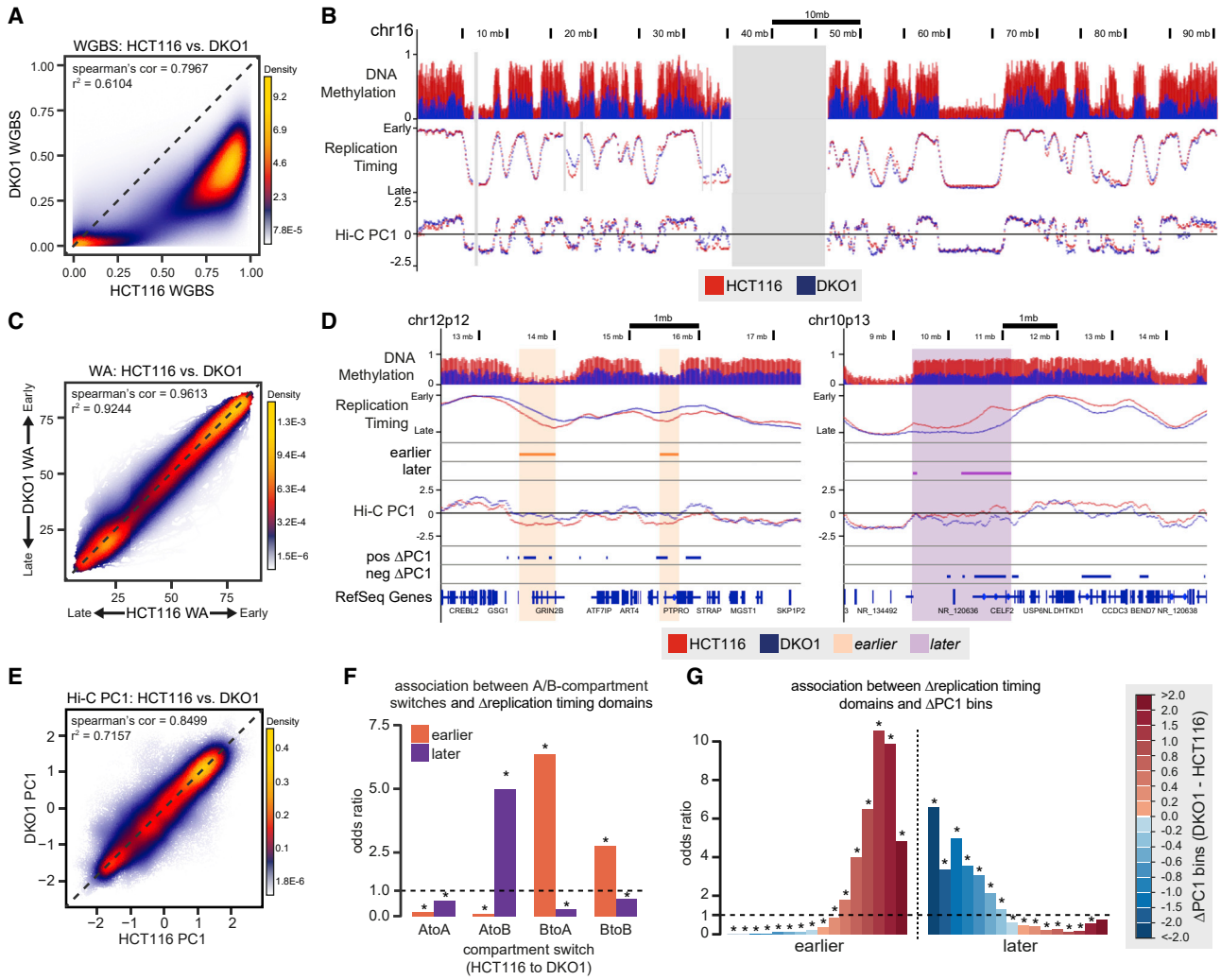


Figure 1. Coordinate change in genome organization and DNA replication timing following DNA methylation loss

(A) Replicate averaged DNA methylation levels (WGBS), HCT116 versus DKO1.

(B) Replicate averaged replication timing values (WA), HCT116 versus DKO1.

(C) Representative example of replication timing and Hi-C PC1 profiles of HCT116 (red) and DKO1 (blue). Gray bars indicate no data.

(D) Representative examples of concordant change in replication timing and Hi-C PC1.

(E) Replicate averaged Hi-C PC1 values of HCT116 versus DKO1.

(F) Fisher's exact association test between A-/B-compartment switches and earlier/later replicating loci.

(G) Fisher's exact association test between ΔPC1 intervals and earlier/later replicating loci.

For (F) and (G), asterisks indicate significant associations (false discovery rate [FDR] < 0.05).

HCT116 (Figure S3D). In contrast, loci that replicated earlier or showed a positive ΔPC1 difference from HCT116 to DKO1 are significantly associated with a slight gain in methylation (ΔMeth $\geq$ 0.1: 1.46% of CpG sites) (Figures S3B and S3C) and were lowly methylated in HCT116 (Figure S3D). Therefore, distinct higher-order genome architectural changes appear to occur coordinately and are associated with the degree of DNA methylation change.

We next observed that changes in replication timing and chromatin conformation frequently co-occur at troughs in the DNA methylation profile, also known as PMDs (Figures 2A, 2B, and S3E). Interestingly, we found that despite widespread hypomethylation in DKO1 cells, the majority of HCT116 PMD regions persist in DKO1 (Figure 2C). Both HCT116 and DKO1 PMDs are late replicating and form "troughs" in the DNA methylation profile, replication timing profile, and PC1 compartment values (Figure S3F). However, DKO1 PMDs have less well-defined DNA methylation boundaries due to the global hypomethylation (Figure S3F) and also show shallower troughs in both the replication timing and Hi-C PC1 profiles.

To explore the alterations in PMD structure, we divided PMDs into regions that are maintained, lost, or gained from HCT116 to DKO1 (see Method details). Regions that lose PMD definition associate with the same bins of methylation change as earlier

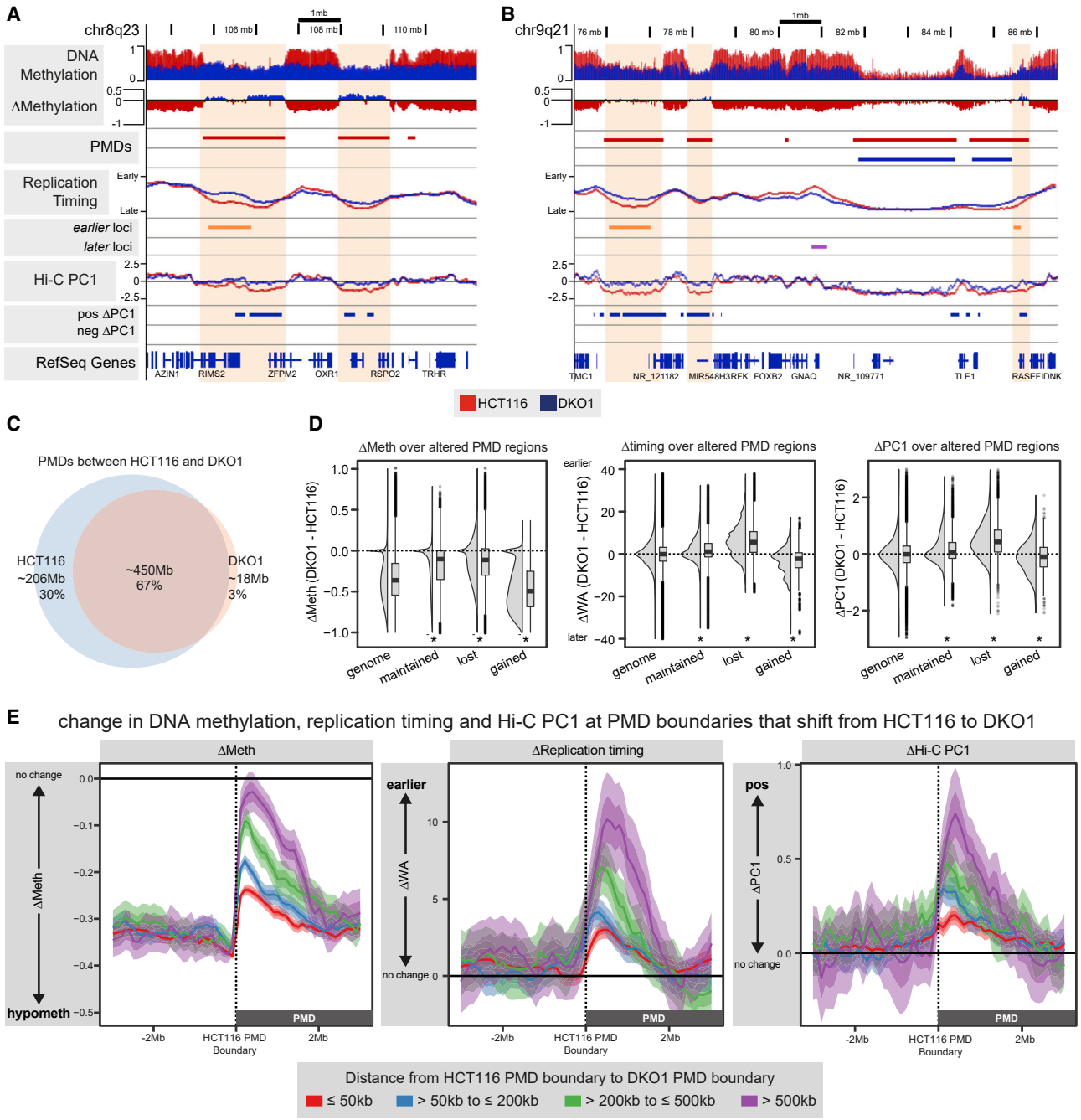


Figure 2. Changes in replication timing and genome organization occur at partially methylated domain boundaries

(A and B) Representative examples of regions containing lost PMDs.

(C) Overlap of PMDs between HCT116 and DKO1.

(D) Change in methylation, replication timing and Hi-C PC1 at maintained, lost, or gained PMDs compared to "genome"-wide. Asterisks indicate significance ($p < 0.05$) in 1-tailed Mann-Whitney-Wilcoxon test against "genome."

(E) HCT116 PMDs are grouped based on the distance of shift inward of the DKO1 PMD boundary at the 5′ and/or 3′ end. Profile plots of the average change in DNA methylation, replication timing, and Hi-C PC1 of the 4 groups are shown; width of shading indicating confidence intervals.

replicating and positive ΔPC1 compartment regions (Figure S3G). Furthermore, lost PMD regions show a clear shift toward early replication timing and positive PC1 values in DKO1 compared to HCT116 (Figure 2D), contributing to ~50% of all earlier loci and ~34% of all positive ΔPC1 regions (Figure S3H). Lost PMDs are also enriched for earlier and positive ΔPC1 regions (Figure S3I). We further observed that PMD loss appears to occur specifically at the boundaries of the HCT116 PMD

and shifts inward from HCT116 to DKO1 (Figures 2B and S3E). Figure 2E shows that larger shifts in PMD boundaries display increasingly earlier replication timing and corresponding increases in Hi-C PC1 values (i.e., B to A shift) and minimal methylation change between HCT116 and DKO1 (Figure 2E). Therefore, our results suggest that the shift in PMD boundaries, due to the loss of the stratification of the DNA methylation landscape, results in a corresponding shift in higher-order genome architecture.

DNA hypomethylation reduces precision of DNA replication timing

Interestingly, even though replication timing values from whole-population data between HCT116 and DKO1 are highly correlated (Figure 1B; Spearman's = 0.9613), we observed that the range of pre-normalized replication timing (WA) values in DKO1 cells is smaller compared to that of HCT116 (Figure S1D). In the Repli-Seq method, we calculate WA values from six-fraction values across S phase (see Method details). To investigate the difference in the spread of WA values between HCT116 and DKO1, we determined the variance of the six-fraction signal. A higher score indicates that the majority of the signal is coming from a small number of fractions (precise timing), whereas a lower score indicates that the signal is more evenly distributed between the six-fractions (varied timing) (Figure S4A). We found that DKO1 showed lower variance than HCT116 (Figure 3A), suggesting that replication in DKO1 is more spread over the six-fractions and therefore less precise. Representative examples of regions with decreased precision in DKO1 compared to HCT116 are shown in Figure S4B.

A standard variance calculation disregards the order of the six-fractions, whereas the order of the replication timing fractions across S phase is biologically important. Therefore, we re-examined the variation using a weighted variance calculation (see Method details). Here, a lower value indicates more precision in timing (Figure S4E). Using the weighted variance score, we again showed that the majority of loci show more variance and reduced precision of DNA replication timing in DKO1 compared to HCT116 (Figure 3B). We further showed that a similar loss of replication timing precision also occurs in hypomethylated prostate cancer LNCaP cells compared to normal prostate PrEC cells (Figure S4F). These data suggest that the decrease in the synchronization of DNA replication timing occurs as a result of DNA hypomethylation.

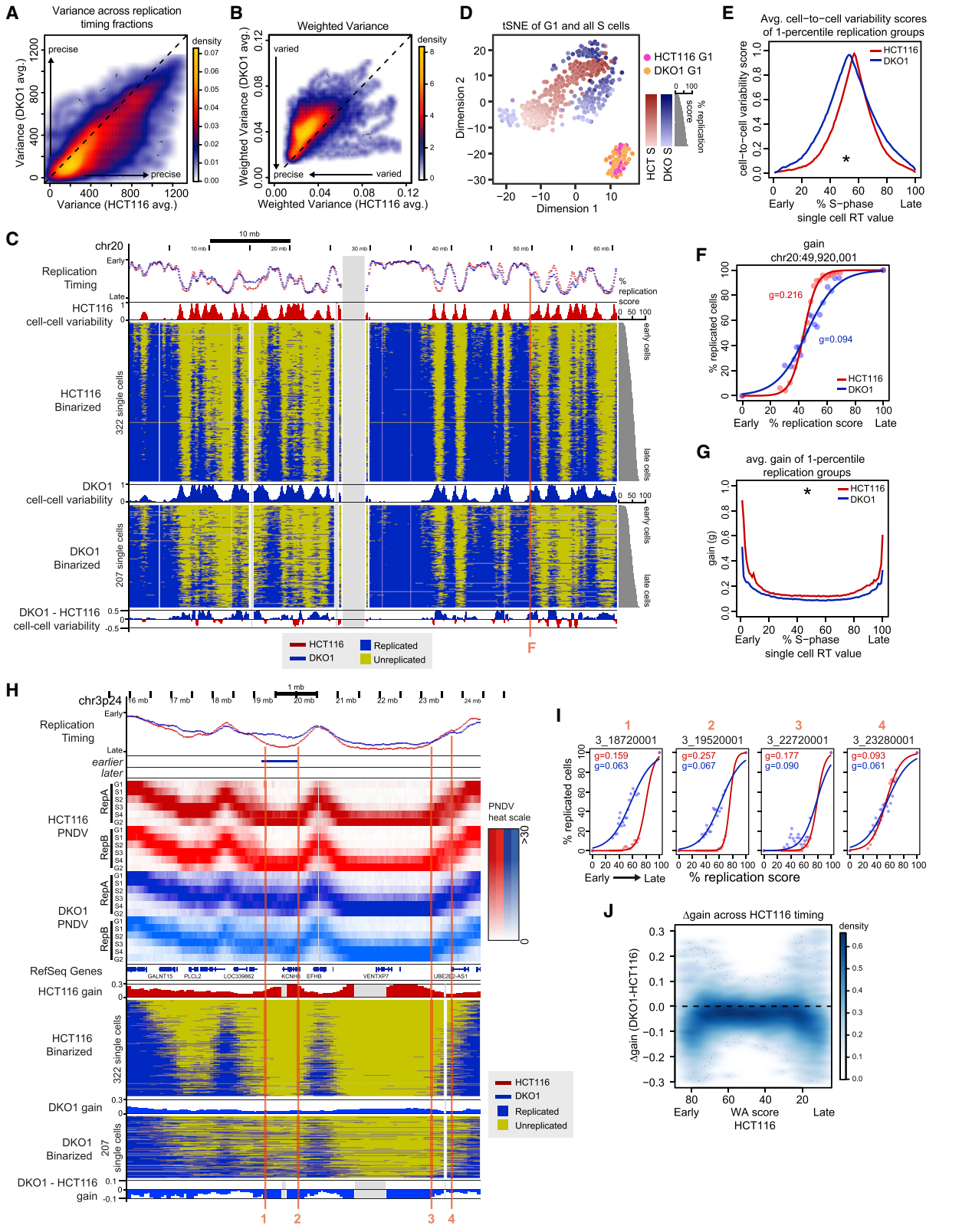
DNA hypomethylation increases cell-to-cell heterogeneity of DNA replication timing

To next determine whether the reduction in the precision of population-level Repli-Seq data is due to an increase in cell-to-cell variability within the cell population, we performed single-cell replication timing sequencing (scRepli-Seq, see Method details) (Dileep and Gilbert, 2018; Takahashi et al., 2019). Single-cell libraries were generated on the Single Cell CNV Solution platform from 10x Genomics (Figures 3C and 3D; Method details). Single G1 and S phase cells cluster within their respective cell-cycle states between HCT116 and DKO1 (Figures 3D and S5E). Early S phase cells are closer to the G1 population and late S phase cells are further from the G1 population along dimension 1 (Figure 3D). In line with the lack of replication timing precision, DKO1 cells are more disparately distributed within the S phase cell cluster relative to HCT116 cells. We next calculated cell-to-cell variability using mid-S phase cells (40%–70% replication) for 80-kb bins across the genome. Cell-to-cell variability scores are highest at mid-replication timing (Figure S5F) for mid-S phase cells as expected from prior scRepli-Seq studies (Dileep and Gilbert, 2018; Takahashi et al., 2019). However, average cell-to-cell variability scores of loci in one-percentile groups across replication timing were statistically higher in DKO1 compared to HCT116 (Figure 3E), confirming increased cell-to-cell variability in the hypomethylated DKO1 cells.

To address whether the degree of replication heterogeneity throughout S phase is also higher in DKO1 cells compared to HCT116 cells, we performed sigmoidal curve modeling of the single-cell data to obtain replication kinetics for 80-kb bins across the genome (see Method details). An exemplary region from Figure 3C with sigmoidal curve modeling is shown in Figure 3F. The gain (or slope) of each sigmoid curve indicates how heterogeneously a locus is replicated between cells. A steep curve (large gain value in HCT116, e.g., g = 0.216) indicates synchronous replication of the loci among cells, and a flatter curve (small gain value in DKO, e.g., g = 0.094) indicates heterogeneous replication of the loci among cells. The gain value is highest toward the earliest and latest extremes of replication timing, indicating that the start and end of replication has the least cell-to-cell heterogeneity (Figure S5G). DKO1 cells showed overall lower gain values genome-wide than HCT116 across replication timing, further indicating more heterogeneous replication within the DKO1 cell population (Figure 3G). Representative examples of the consistent lower gain values in DKO cells can be seen in Figure 3H (lower panel DKO1-HCT116) and Figure 3I. This trend in increased heterogeneity occurs genome-wide, including regions with (Figure 3I, loci 1,2) and without a change in replication timing (Figures 3I, loci 3,4, and S5H), and particularly at very early and very late regions (Figure 3J). The increase in heterogeneity at the beginning and end of S phase agrees with the reduced range of population-level Repli-Seq WA values in DKO1 (Figure S1D). Our data support that DNA hypomethylation is associated with the ''erosion'' of the precise regulation of replication timing genome-wide.

DNA hypomethylation reduces integrity of 3D genome compartmentalization

We next examined the effects of reduced replication timing precision on 3D genome organization. As Hi-C is a cell population-level assay, we hypothesize that similar to the loss of the precision of replication timing, we may observe a loss of strength of organization. In corroboration with replication timing cell-to-cell heterogeneity, DKO1 cells have reduced compartmentalization strength of both A- and B-compartments, indicated by the reduction of intra-compartment contact frequencies (A-A, B-B) and an increase of inter-compartmental (A-B) contact frequency (Figures 4A and 4B). The same finding can also be summarized in a single compartmentalization score (Figure 4C). In addition, we found a clear loss of compartmentalization between HCT116 and a second more hypomethylated colorectal cancer cell



(legend on next page)

LS174T (Figure S5I) and notable loss of compartmentalization between hypomethylated colorectal tumor samples compared to normal colorectal tissue (Figure 4D). Furthermore, we found that a loss of compartmentalization occurs in *DNMT* knockdown or KO systems in different cell-type hypomethylation models (Figures S5J and S5K). Finally, we demonstrated that the loss of compartmentalization occurs following DNA hypomethylation that was induced in breast cancer xenografts treated with a US Food and Drug Administration (FDA)-approved DNA methyltransferase inhibitor decitabine (5-aza-2′-deoxycytidine) (Figure 4E).

In agreement with the clear loss of compartmentalization, we show that DKO1 has a smaller percentage of the genome organized into topologically associated domains (TADs) (Figure 4F). Reduced compartment strength and TAD structure suggest a decrease in the number of cells within the population sharing similar chromatin conformation and thus an increase in chromatin conformation heterogeneity. Figure 4G shows representative examples of chromosomes with reduced integrity of compartmentalization. Our data show that the increase in cell-to-cell heterogeneity of DNA replication is also reflected in the increased blurring of 3D compartmentalization.

DNA hypomethylation is associated with loss of allelic replication

In exploring the decrease in replication timing precision using the WA value, we identified loci that showed a gain in replication timing precision specifically associated with later replication timing (Figure S6A). This was contrary to the loss of precision in the rest of the genome. Visual inspection of these loci revealed that they appear to be asynchronously or biphasically replicating loci in HCT116. We therefore asked whether DNA hypomethylation can also result in changes in biphasic replication. Biphasically replicating regions occur where the locus is replicated, both in early and late timing within the same cell type (Hansen et al., 2010). We called biphasic regions in our HCT116 and DKO1 Repli-Seq datasets using both our weighted variance score and the scoring method from Hansen et al. (2010) (see Method details). Both methods identified more biphasic regions in HCT116 than DKO1 (Figures 5A and S6B), with ~70%–85% of the biphasic regions lost in DKO1 (Figures 5B and S6C). Representative examples of lost biphasic regions are shown in Figures 5C and S6D.

Biphasic replicating regions in cell population-level data may be due either to allelic replication or the presence of two subpopulations with differential replication at the same locus. To determine whether biphasic regions show allele-specific replication, we used long-read Nanopore sequencing followed by variant calling and phasing to obtain phased haplotypes for both HCT116 and DKO1 (see Method details). Allelically separated replication timing WA scores show that biphasic regions comprise one early replicating allele and one late replicating allele in HCT116, and the alleles become more synchronously replicated in DKO1 (Figures 5D, S6E, and S6F). We confirmed that the biphasic replicating regions are allelic in our single-cell DNA-seq data (Figures S6G and S6H). Using allelic replication timing WA scores alone, we confirmed that ~50% of allele-specific replicating regions are lost after DNA methylation reduction in DKO1 cells (Figure 5E). In three regions, the loss of allelic replication timing of the early allele resulted in a significant shift to later replication timing in population Repli-Seq data (Figures 5C, 2nd and 3rd panels, and S6D, 2nd panel). Further representative examples of loss of allele-specific replication are shown in Figure S6I. The population-level biphasic data, together with the Nanopore data, show that the hypomethylation predominantly results in a loss of asynchronous allele-specific replication.

As imprinting loci are also known to show allelic replication (Kitsberg et al., 1993), we next asked whether allele-specific replication in HCT116 occurs at imprinted loci. We compared allelically replicated genes against a list of known or predicted imprinted genes (see Method details) (Luedi et al., 2007). Of a total of 111 genes located at allelically replicating regions in either HCT116 or DKO1, only 35 were protein coding (Table S1) and only 2 of these, *PRIM2* and *DGCR6*, were found to be in the imprinted gene database. This suggests that the majority of allelic replicating loci in HCT116 are not at developmentally imprinted regions. Interestingly, we found that 16 of the 35 allelically replicating protein-coding genes are also known or predicted to be mono-allelically expressed (Gimelbrant et al., 2007; Savova et al., 2016). Furthermore, many of the allele-specific replication genes are reported to be cancer related (Table S1). Notably, three genes that lose allelic replication in DKO1 cells are also reported to be associated with colon cancer: *NRG1* (Luraghi et al., 2017; Stahler et al., 2017), *PCDH7* (Li et al., 2020; van Roy, 2014), and *DLC1* (Durkin et al., 2007; Peng et al., 2013) (Figures 5C, 5D, S6D, and S6E). *NRG1* and *PCDH7* are also mono-allelically

Figure 3. Loss of global methylation increases cell-to-cell heterogeneity of DNA replication timing

(A) Replicate averaged variance scores, HCT116 versus DKO1.

(B) Replicate averaged weighted variance scores, HCT116 versus DKO1.

(C) Representative example of whole-population Repli-Seq with single-cell Repli-Seq (scRepli-Seq). Binarized single-cell data are ordered from early to late (low to high % replication score). Red vertical line refers to the 80-kb bin shown in (F).

(D) tSNE of G1 and S phase cells from HCT116 and DKO1. Early to late gradient of S phase cells is denoted by transition from dark to light shading.

(E) Average cell-to-cell variability score per 1-percentile of bins across single-cell RT values. Asterisk indicates significant difference ($p < 0.05$, permutation test) between HCT116 and DKO1.

(F) Sigmoid model curves and gain values of example locus from (C). Dots are real data points and lines are the fitted curves.

(G) Average gain scores per 1-percentile of bins across single cell RT values. Asterisk indicates significant difference ($p < 0.05$, permutation test) between HCT116 and DKO1.

(H) Representative region with increased heterogeneity (loss of gain value) in DKO1 compared to HCT116. All 6 Repli-Seq fractions (PNDV) are shown in order of G1, S1, S2, S3, S4 and G2. Red vertical lines refer to 80-kb bins shown in (I).

(I) Sigmoid model curves and gain values of loci from (H).

(J) Gain differences (DKO1–HCT116) across whole-population replication timing values (WA) of HCT116.

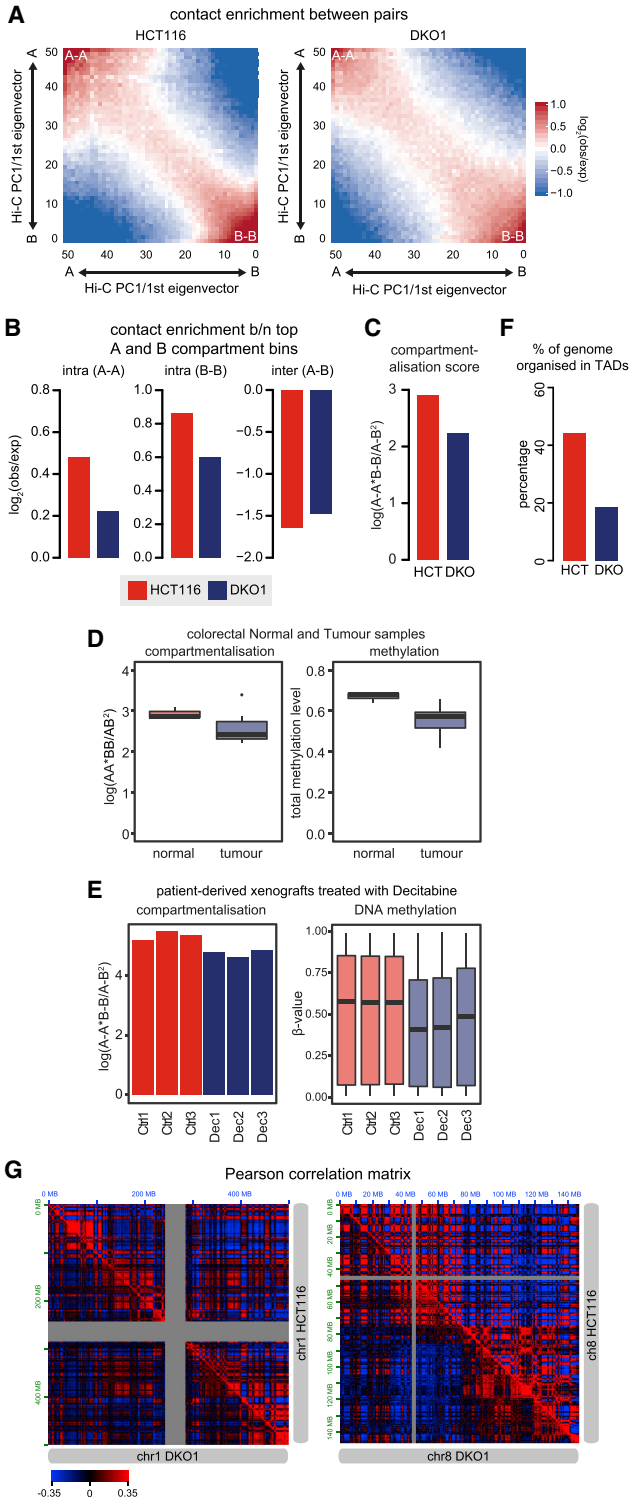


Figure 4. Loss of global methylation reduces integrity of 3D compartmentalization

(A) Saddle plot showing contact enrichments (log_2(Obs/Exp)) between pairs of 100-kb bins ordered in 50 PC1 quantile groups.

(B) Mean contact enrichments within and between A- and B-compartments (top and bottom 20% of PC1 quantiles).

(C) Compartmentalization score of HCT116 compared to DKO1.

(D) Boxplots of compartmentalization scores and corresponding total methylation levels of patient normal (n = 3) and tumor (n = 12) samples, re-analyzed from Johnstone et al. (2020).

(E) Compartmentalization scores of patient-derived xenografts, control ("Ctrl") and treated with DNA demethylation agent decitibine ("Dec"). Boxplots of the corresponding methylation β values are shown.

(F) Percentage of the genome called TADs.

(G) Representative examples (Pearson correlation matrix) of loss of compartmentalization in DKO1 compared to HCT116.

expressed genes, and *DLC1* was recently reported to be partially mono-allelically expressed (Gupta et al., 2020). All three genes are significantly downregulated in DKO1 compared to HCT116 (Figure S6J); interestingly, they show loss of the early replicating allele in DKO1 (Figures 5D and S6E). These data suggest that DNA hypomethylation promotes an alteration of allelic replication timing and reduced expression of some key cancer-related genes.

Chromatin modifications and gene expression changes following DNA hypomethylation are associated with altered 3D genome architecture

Chromatin changes in repressive heterochromatin histone marks (i.e., H3K27me3, H3K9me3) have been shown to occur following the loss of *DNMT* expression (Espada et al., 2004; Reddington et al., 2013; Saksouk et al., 2014). Furthermore, heterochromatin has been shown to be important in establishing and maintaining global nuclear organization and compartmentalization (Belaghzal et al., 2019; Falk et al., 2019; MacPherson et al., 2018). Therefore, we asked where coordinated histone modification changes occur in the genome after DNA hypomethylation in DKO1 cells and whether this relates to higher-order genome architecture. To do this, we called chromHMM states for both HCT116 and DKO1 and found that the majority of state changes between HCT116 and DKO1 were the result of a gain of H3K4me3, H3K4me1, and H3K27me3, and a loss of H3K36me3 and H3K9me3 (Figures S7A and S7B; Method details). These changes can also be observed in the histone marks themselves (Figure 6A). We observed an overall increase in H3K27me3 and H3K4me3, and a loss of H3K9me3 mostly at late-replicating loci. Surprisingly, the gain in H3K4me3 occurs at both early- and late-replicating loci. The gain in H3K27me3 and loss of H3K9me3 reflects our previous observations at hypomethylated late-replicating regions in prostate and breast cancer cells (Du et al., 2019).

Next, to explore whether gene deregulation, between HCT116 and DKO1, is related to replication timing and chromatin conformation changes, we examined differential gene expression (Figure S7C). First, we identified genes that replicated later and located in A-B shifted compartments (negative ΔPC1) are associated with downregulation of expression, and genes that replicated earlier and located in B-A shifted compartments (positive ΔPC1) are associated with upregulation (Figure S7D). Genes located within late replication timing and B-compartments (negative PC1 values) also show the most expression upregulation in DKO1 (Figure S7E), suggesting that the loss of DNA methylation leads to aberrant gene activation in late-replicating B-compartments.

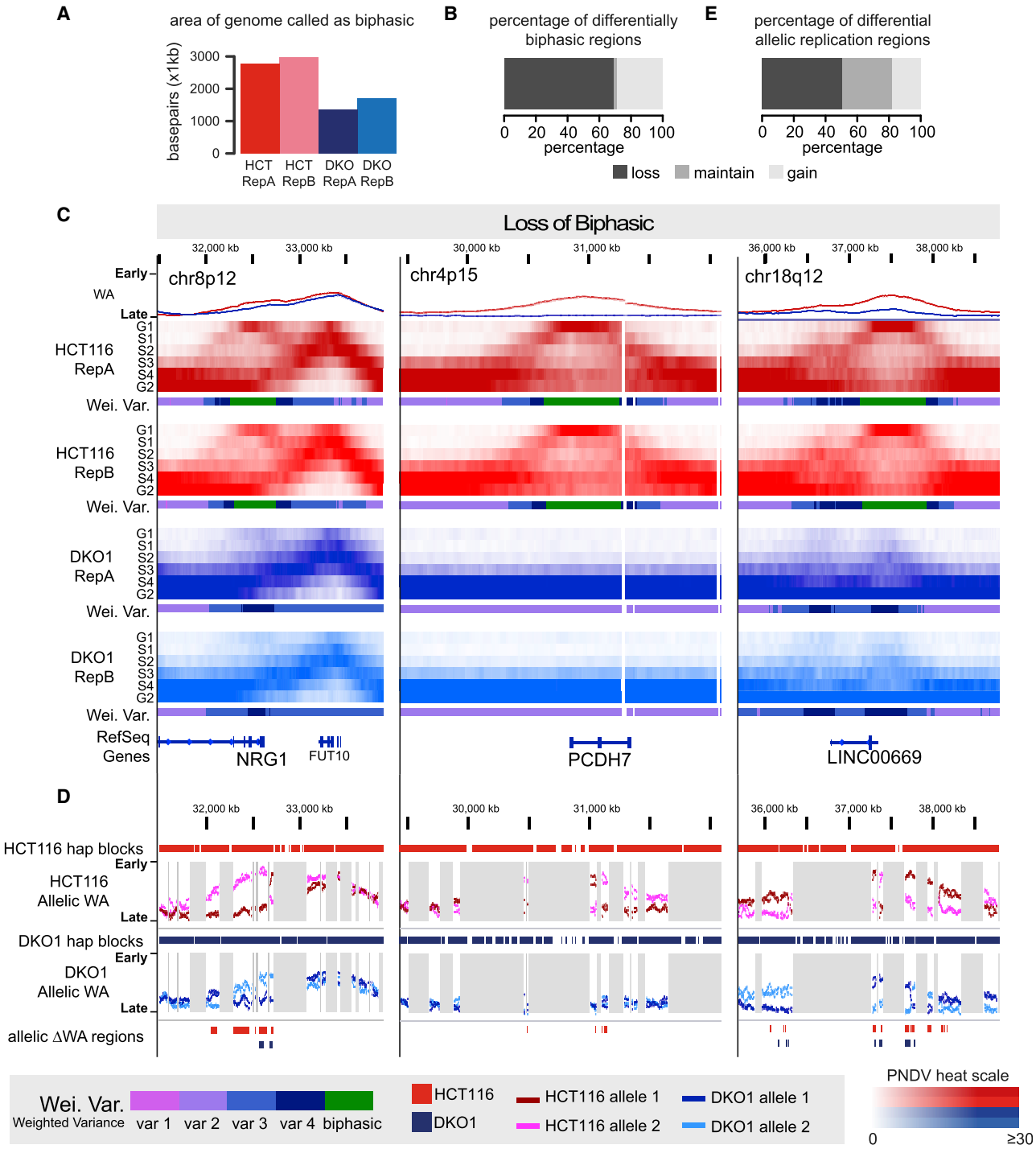


Figure 5. Loss of DNA methylation causes loss of allelic DNA replication

(A) Area of the genome (kb) called "biphasic" using the weighted variance score.

(B) Percentage of biphasic replication regions that are lost, maintained, or gained from HCT116 to DKO1.

(C) Examples of regions that changed or maintained biphasic status between HCT116 and DKO1. Colors indicating bins of the weighted variance score is shown below each set of 6-fractions. Green indicates biphasic regions.

(D) Allelically separated replication timing (WA) scores of regions in (C). Haplotype blocks are shown above WA scores. Gray shading indicates either no data or a break between haplotype blocks.

(E) Percentage of allelically replicating regions that are lost, maintained, or gained from HCT116 to DKO1.

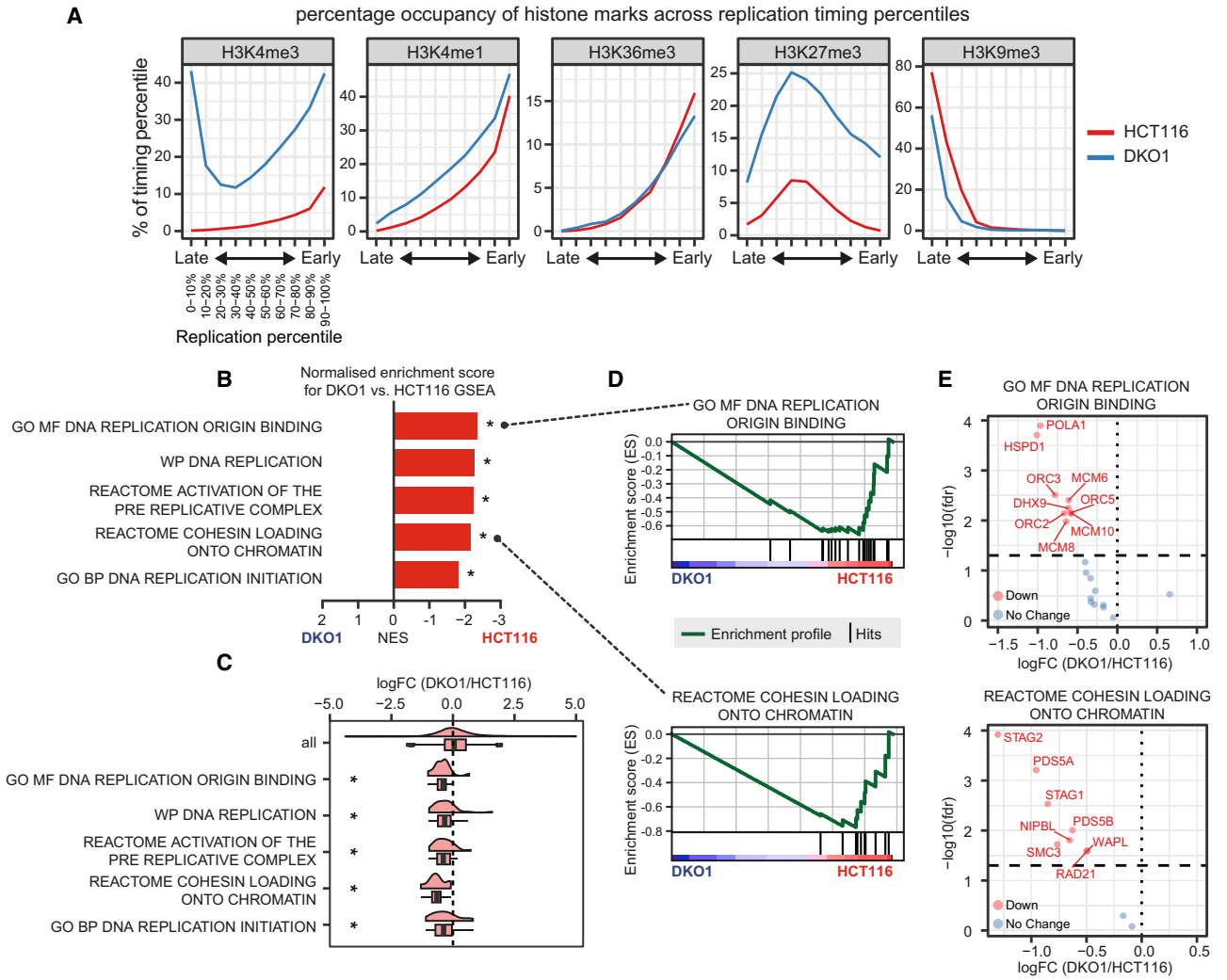


Figure 6. Chromatin modifications and gene expression changes are associated with altered 3D genome architecture

(A) Percentage occupancy of histone marks for 1-kb loci across replication timing bins in HCT116 and DKO1. Replicated peaks were used where replicates were available (H3K4me3, H3K4me1, H3K27me3, H3K9me3).

(B) Normalized enrichment scores (NESs) for gene sets significantly enriched (FDR < 0.05) in DKO1 versus HCT116 GSEA analysis.

(C) Boxplots and density plots of differential expression (log fold change [logFC]) of genes within each gene set from (B). Asterisks indicate significance in 1-tailed Mann-Whitney-Wilcoxon test of each gene set against genome-wide differential expression "all," where the alternative is "less."

(D) Enrichment plots for gene sets "DNA replication origin binding" and "cohesin loading onto chromatin."

(E) Differential expression volcano plots (DKO1/HCT116) for genes within gene sets from (D).

Second, to determine whether our observations of genome decompartmentalization and reduced replication precision are related to gene deregulation, we performed a gene set enrichment analysis (GSEA). We found that terms related to DNA replication and chromosome organization are significantly downregulated in DKO1 compared to HCT116 (Figures 6B and 6C)—in particular, terms relating to DNA replication origins and cohesin (Figure 6D). Significant terms include *ORC* and *MCM* genes, which are involved in DNA replication origins (Figure 6E) (Bell and Stillman, 1992; Remus et al., 2009), and *WAPL* and *RAD21*, which are involved in cohesin activity and chromosome organization (Figure 6E) (Haarhuis et al., 2017; Wutz et al., 2017). Therefore, the loss of DNA methylation may contribute to altered 3D genome architecture through the gene deregulation of key components of the DNA replication and genome organization machinery.

Gain of broad H3K4me3 in H3K9me3 domains after hypomethylation may protect against genome reorganization

One of the most significant changes in chromHMM states between HCT116 and DKO1 is the change from the H3K9me3 enrichment (heterochromatin, Het) to the H3K4me3 enrichment (active TSS, TssA) chromatin state. This appears to occur in large domains that coincide with late-replicating domains in both HCT116 and DKO1 (Figures 7A and S8A). These regions have maintained H3K9me3 between HCT116 and DKO1. However,

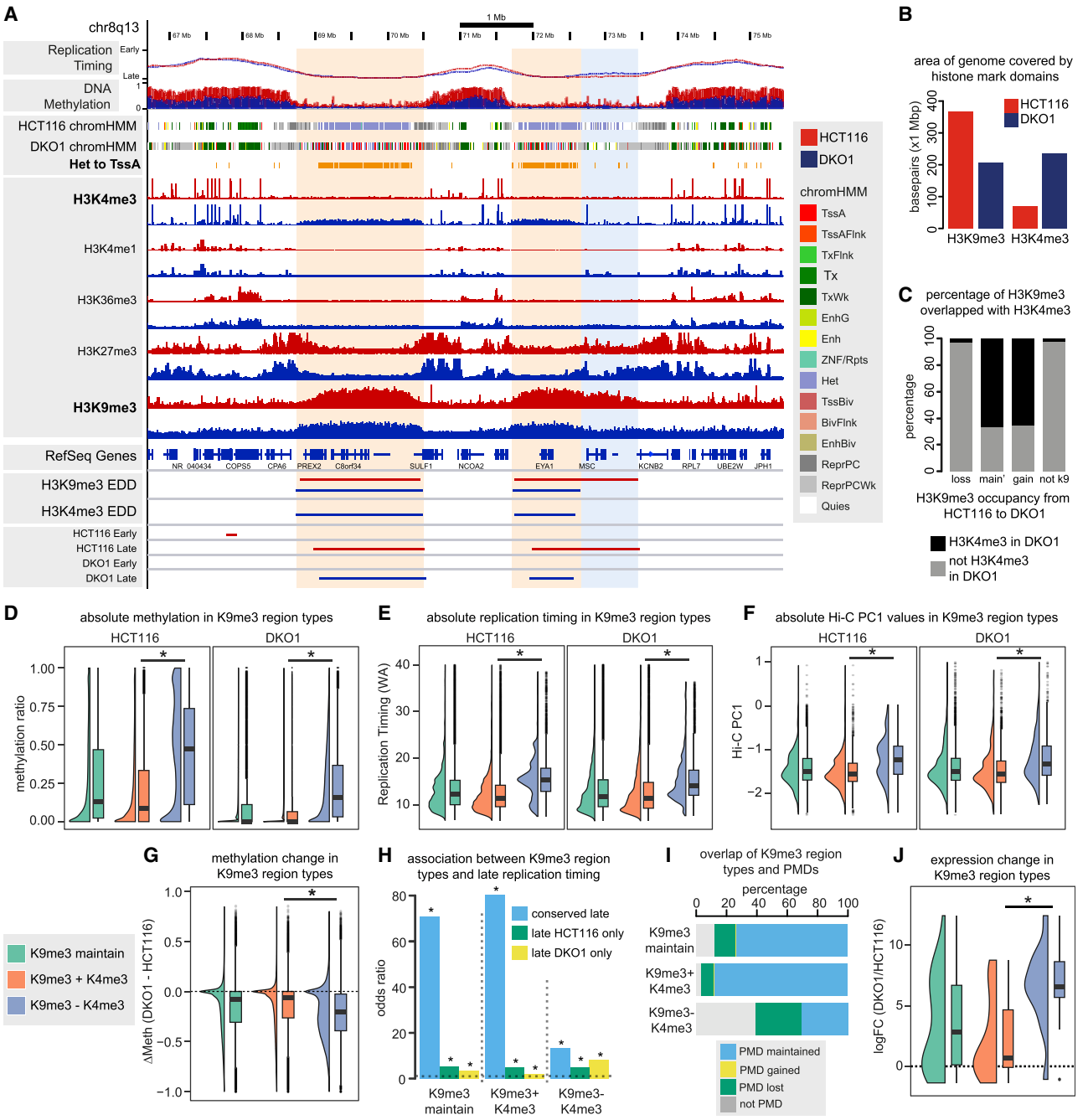


Figure 7. Non-canonical H3K9me3 and H3K4me3 bivalent domains occur at conserved late-replicating B-compartments and maintains expression silencing

(A) Representative example of broad H3K9me3 and H3K4me3 enrichment. "Het to TssA" track shows loci with corresponding chromatin state transitions (see Figure S7). Beige shading indicates maintained H3K9me3 domains, and blue shading indicates lost H3K9me3 domains from HCT116 to DKO1.

(B) Area of the genome (Mb) covered by H3K9me3 and H3K4me3 board domains.

(C) Percentage occupancy of H3K9me3 and H3K4me3 broad domains between HCT116 and DKO1.

(D–F) Absolute DNA methylation levels, absolute replication timing levels, and absolute Hi-C PC1 levels in H3K9me3/H3K4me3 region types. Asterisks indicate that the K9me3+K4me3 regions are less than the K9me3-K4me3 regions ($p < 0.05$, 1-tailed Mann-Whitney-Wilcoxon).

(G) DNA methylation change in H3K9me3/H3K4me3 region types. Asterisks indicate that the K9me3+K4me3 region is greater than the other category ($p < 0.05$, 1-tailed Mann-Whitney-Wilcoxon).

(legend continued on next page)

DKO1 also displays a low but broad gain in H3K4me3 enrichment across the same region (Figures 7A and S8A). The broad H3K4me3 domains do not resemble the typically observed punctate H3K4me3 peaks and interestingly, are only observed when H3K9me3 is maintained from HCT116 to DKO1 (Figures 7A and S8A, beige shading), but not when H3K9me3 is lost in DKO1 (Figures 7A and S8A, blue shading).

To examine the H3K9me3/H3K4me3 patterns genome-wide, we called broad domains using enriched domain detector (EDD) (Lund et al., 2014). We found that maintained H3K9me3 regions showed the highest increase in H3K4me3 enrichment in DKO1 (Figure S8B). There is no obvious increase in H3K4me1, H3K36me3, or H3K27me3 in the H3K9me3-maintained regions (Figure S8C), suggesting that this enrichment is H3K4me3 specific. We next tabulated the co-occupancy of H3K9me3 and H3K4me3 between HCT116 and DKO1 (see Method details) and observed that while DKO1 lost half of the broad H3K9me3 domains, more than double the area of the genome gained the broad H3K4me3 domains (Figure 7B). Approximately 66% of H3K9me3 domains in DKO1 overlap with newly acquired H3K4me3 domains in DKO1 (Figure 7C). In contrast, there were minimal novel H3K4me3 domains in regions that lacked H3K9me3 domains (Figure 7C). These results show that after DNA methylation loss, regions of the genome that maintain H3K9me3 become broadly marked by H3K4me3 to form non-canonical bivalent domains.

H3K9me3 is highly abundant in late replication timing loci in cancer cells (Du et al., 2019); therefore, regions with H3K9me3 are expected to have low DNA methylation and low Hi-C PC1 values. However, we observed that H3K9me3+H3K4me3 domains are consistently less methylated, display later timing, and have lower Hi-C PC1 values than H3K9me3-H3K4me3 domains (Figures 7D–7F). As non-canonical bivalent H3K9me3+H3K4me3 domains have very low methylation in HCT116 cells, only minimal loss of methylation was observed (Figure 7G), with loci becoming completely unmethylated in DKO1 (Figure 7D). The gain of H3K4me3 may therefore be a response to a complete methylation loss at these specific regions in DKO1 cells. Non-canonical bivalent H3K9me3+H3K4me3 domains also show more association with conserved late replication between HCT116 and DKO1, compared to H3K9me3-H3K4me3 (Figure 7H). Agreeing with the overlap with conserved late regions, H3K9me3+H3K4me3 domains also overlap more with maintained PMDs compared to H3K9me3-H3K4me3 domains (Figure 7I). Therefore, in response to global DNA methylation loss, late replication and genome organization appears to be maintained by the formation of these non-canonical broad bivalent H3K9me3+H3K4me3 domains.

We finally asked whether the increase in H3K4me3 facilitates a gain in transcriptional activity within H3K9me3-maintained regions. Interestingly, bivalent H3K9me3+H3K4me3 domains showed lower expression upregulation compared to H3K9me3-H3K4me3 domains (Figure 7J), indicating that the H3K9me3+H3K4me3 co-occupancy protects these regions against an overall gene activation that occurs upon global hypomethylation in DKO1 cells (see Figure S7C). These results suggest that the gain in broad H3K4me3 occupancy in H3K9me3-maintained domains provides protection against aberrant gene activation after DNA methylation loss.

DISCUSSION

DNA hypomethylation is a hallmark of cancer cells, but its impact on the integrity of the 3D cancer genome architecture is not well understood. Here, we reveal that DNA methylation facilitates the maintenance of replication timing and higher-order genome architecture. We show that genome-wide loss of DNA methylation is associated with a remarkable reduction in replication timing precision and corresponding loss of 3D genome compartmentalization. Our findings suggest that DNA methylation is not only associated with transcriptional regulation but also plays a critical role in the maintenance of higher-order genome architecture genome-wide.

To date, there are limited studies on the consequences of DNA hypomethylation on 3D genome organization. Imaging studies reported de-condensation at chromocenters after *Dnmt1* KO in mouse embryonic fibroblasts (Casas-Delucchi et al., 2012) and an increase in DNase-I sensitivity of the inactive X chromosome after 5-azacytidine treatment of gerbil lung fibroblasts (Jablonka et al., 1985). However, a prior study using the HCT116/DKO1 hypomethylation cell model did not find an increase in genome-wide open chromatin, suggesting no chromosomal de-condensation (Pandiyan et al., 2013). In contrast, we found using sequencing approaches that DKO1 cells, as well as other *DNMT* knockdown and KO cell line models (Nothjunge et al., 2017; Sati et al., 2020), hypomethylated patient tumor samples (Johnstone et al., 2020) and patient-derived xenografts treated with decitabine, that a common genome-wide loss of compartmentalization was associated with DNA hypomethylation, indicating that DNA methylation is indeed required to maintain the stability of 3D genome organization. Furthermore, GSEA revealed downregulation of cohesin and cohesin-interacting factors, suggesting that the loss of genome organization machinery may be involved in decompartmentalization.

Importantly, no studies thus far have examined the global consequence of hypomethylation on DNA replication timing. Using single cell sequencing, we show that DNA hypomethylation is associated with the erosion of precise replication timing regulation. As we observed the downregulation of replication origin factors, we speculate that this may result in fewer origins, leading to the loss of replication timing precision. In addition, DNA methylation loss may destabilize origin localization, leading to cell-to-cell replication timing heterogeneity. Previous studies reported that DNA methylation can stabilize interactions between H3K9me3 and origin recognition complex proteins (Bartke

(H) Association between H3K9me3/H3K4me3 region types and either conserved late domains, HCT116-only late domains, or DKO1-only late domains. Asterisks indicate significant associations (Fisher"s exact test, FDR < 0.05).

(I) Overlap of H3K9me3/H3K4me3 region types with maintained, lost, and gained PMDs between HCT116 and DKO1.

(J) Differential expression (logFC, DKO1/HCT116) of transcripts within H3K9me3/H3K4me3 region types based on promoter overlaps. Asterisks indicate that K9me3+K4me3 is less than K9me3-K4me3 ($p < 0.05$, 1-tailed Mann-Whitney-Wilcoxon).

et al., 2010). Importantly, regulated removal of H3K9me3 by Kdm4d at replication origins is essential for origin initiation and elongation (Wu et al., 2017), and loss of DNA methylation results in the aberrant loss of H3K9me3 in mouse and human cells (Espada et al., 2004; Saksouk et al., 2014). Thus, we propose that DNA methylation may play a role in stabilizing origin location by preventing ectopic origin activation, and hence loss of methylation leads to replication timing imprecision via disorganized origin activation.

Another potential mechanism driving increased cell-to-cell replication timing heterogeneity involves the loss of H3K9me3. Our findings that late-replicating B-compartment regions show significantly increased expression and associated loss of the heterochromatic mark H3K9me3, support that late-replicating heterochromatic regions are predominantly compromised after DNA methylation loss. Heterochromatin, particularly H3K9me3 and its reader HP1, are reported to be important in establishing and/or maintaining global chromatin conformation, phase separation and compartmentalization, and DNA replication timing (Falk et al., 2019; Klein et al., 2021; Larson et al., 2017; MacPherson et al., 2018). Loss of the H3K9me3 methyltransferases (*Suv39h1,2*) or *HP1* also leads to earlier replication timing of chromocenters and centromeric repeats (Schwaiger et al., 2010; Wu et al., 2006), similar to loci-specific studies of DNA methylation loss. Therefore, the loss of H3K9me3 in DKO1 may destabilize the organization of the genome, hence driving global de-compartmentalization and decreased replication timing stability.

Notably, we found the majority of 3D genome architectural changes involved switching to earlier DNA replication timing at the boundaries of PMDs, accompanied by a similar shift toward active A-compartments. We show that the loss of methylation at PMD boundaries is associated with blurring of the methylation boundaries and shrinking of PMDs. PMDs are typically heterochromatic, marked by H3K27me3 and/or H3K9me3 (Berman et al., 2011; Hon et al., 2012). Interestingly, loss of DNA methylation was previously reported to result in an earlier onset of DNA replication at a number of heterochromatic loci such as chromocenters (Casas-Delucchi et al., 2012), pericentromeric major satellite repeats (Jørgensen et al., 2007), and the inactive X chromosome (Hansen et al., 2000; Jablonka et al., 1985). Our results indicate that earlier replication at discrete heterochromatic PMD loci occurs genome-wide and is more common than previously described.

We next found that discrete loci showing allele-specific replication in HCT116 are lost after DNA hypomethylation in DKO1. The same loss of allelic replication at key cancer-related genes (*TP53* and *RB1*) was found in cancer patient lymphocytes treated with the demethylation agent 5-azacytidine (Dotan et al., 2004, 2008; Nagler et al., 2010). Early studies reported that DNA methylation-imprinted regions can show allele-specific replication (e.g., *IGF2*, *H19, SNRPN*, chromosome X) (Kitsberg et al., 1993; Takagi and Oshimura, 1973). However, similar to other studies (Hansen et al., 2010; Rivera-Mulia et al., 2018; Zhao et al., 2020), we found that few allelically replicating loci contain known or predicted imprinted genes. We therefore suggest that allelic replication may be a separate regulatory mechanism to classic gene imprinting. Interestingly, a substantial proportion of the allelically replicating regions contain genes previously identified as showing monoallelic expression. Monoallelic expression is reported to differ in frequency between grades of brain tumors (Walker et al., 2012) and in colorectal tumors compared to normal tissue (Liu et al., 2018), suggesting that cancer cells can acquire ectopic allele-specific expression during transformation. DNA methylation is also involved in non-imprinted monoallelic expression in both humans and mice (da Rocha and Gendrel, 2019; Gupta et al., 2020; Schalkwyk et al., 2010). We found that the majority of allele-specific replicating loci lost their allelic replication after DNA hypomethylation, suggesting that DNA methylation may also play a role in the regulation of allelically replicating genes that are also monoallelically expressed. Furthermore, we identified that loss of allelic replication was associated with downregulation at three colon cancer-related monoallelically expressed genes. Our results support that altered allelic replication, induced by global changes in DNA methylation, is a common mechanism of gene deregulation within cancer cells.

Finally, it is well established that bivalent H3K4me3-H3K27me3 states commonly occur in normal cells at silent CpG island promoters, and that in cancer, bivalency is lost when these promoters gain DNA methylation (Baylin and Jones, 2016). We were therefore intrigued to find that global loss of DNA methylation was associated with the formation of non-canonical bivalent domains of broad H3K4me3 enrichment in H3K9me3 domains genome-wide. Broad H3K4me3 "mesas" have previously been described in senescent cells and form over lamina-associated domains (LADs) (Shah et al., 2013), and examples of broad H3K4me3/H3K9me3 loci have previously been noted in DKO1 cells (Lay et al., 2015). Here, we further demonstrated that these H3K9me3+H3K4me3 domains are associated with late replication and PMD structure and show the least aberrant gene activation, despite the gain of the H3K4me3 active mark. More recently, broad H3K4me3 domains have been identified in mouse oocytes (Dahl et al., 2016; Liu et al., 2016; Zhang et al., 2016). These non-canonical H3K4me3 domains appear following the wave of DNA methylation erasure that occurs during the reprogramming of primordial germ cells and during a period of genomic silencing in late-stage oocytes (Dahl et al., 2016; Zhang et al., 2016). Similar to our findings, the non-canonical H3K4me3 domains also co-occur with PMDs (Zhang et al., 2016) and are anti-correlated with DNA methylation (Dahl et al., 2016). Active *de novo* methylation by DNMT3A and -3B is required to protect regions against acquiring this form of H3K4me3 (Hanna et al., 2018). Therefore, KO of *de novo* methyltransferase *DNMT3B* may be the main driver of H3K4me3 domain formation in DKO1 cells. The H3K4me3 domains further function in oocytes to maintain zygotic genome silencing (Zhang et al., 2016), suggesting that the ectopic H3K4me3 bivalent domains in DKO1 cells are protecting late-replicating H3K9me3-marked domains from spurious activation in response to extreme DNA methylation loss. We propose that KO of *DNMT1* and *DNMT3B* in DKO1 cells is creating a similar demethylation event that occurs in oocytes, subsequently causing the acquisition of non-canonical H3K4me3 domains, which protects against catastrophic genome reorganization and corresponding gene deregulation.

In summary, we show that DNA hypomethylation, a hallmark of cancer cells, can lead to the disruption of DNA replication timing precision, cell-to-cell heterogeneity, and higher-order genome reorganization. We hypothesize that the overall loss of distinction in the DNA methylation profile leads to the reduced organization of replication origin activation, blurring' of methylation boundaries at PMDs, and loss of heterochromatin domains. The resulting 3D genome heterogeneity may therefore create an unexpected source of variation for clonal selection in cancer evolution. It will be important in future studies to use more single-cell epigenomic approaches to understand the temporal relationship between changes in the epigenome and the 3D genome architecture during tumor progression.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Expression qRT-qPCR of DNMT genes
  - Western blotting for DNMT proteins
  - Nanopore sequencing and base calling
  - Calling structural variants
  - Whole genome bisulphite sequencing and processing
  - Repli-Seq data generation and processing
  - Defining partially methylated domain boundary shifts
  - Hi-C library preparation
  - Hi-C data processing
  - TADs
  - A/B compartments
  - Defining domains of genome reorganization
  - Hi-C compartment strength calculation
  - Weighted variance calculation for Repli-Seq fractions
  - Single cell replication timing
  - Identifying biphasically replicating loci
  - Allelic replication timing
  - Profile plots
  - ChIP-seq processing
  - ChromHMM analysis
  - RNA-seq data generation and processing
  - Genomic annotation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109722.

ACKNOWLEDGMENTS

We thank members of the Clark laboratory for helpful discussions and careful reading of the manuscript. We thank Prof. Stephen Baylin for kindly providing the HCT116 and DKO1 cells. We thank Phillippa C. Taberlay and Yi Cai for advice with tissue culture. S.J.C. is a National Health and Medical Research Council, Australia (NHMRC) Senior Principal Research Fellow #1063559. Q.D. is a NHMRC Investigator Grant recipient #1177792. This work was supported by NHMRC Project Grants APP1147974 and APP1163939 (to S.J.C.) and was partially enabled by a Cancer Institute New South Wales, Australia (CINSW) equipment grant REG181268 (to M.A.S.). The contents of the published material are solely the responsibility of the administering institution and individual authors and do not reflect the views of the NHMRC.

AUTHOR CONTRIBUTIONS

Conception and design, Q.D. and S.J.C.; methodology, Q.D., M.B., C.E.C., K.-M.C., N.P., D.K., and J.E.P.; investigation, Q.D., G.C.S., E.M.C., S.S.N., D.K., C.-L.C., K.B., and J.A.-K.; analysis, Q.D., P.L.L., J.M.F., N.J.A., C.M.G., E.Z., M.B., and J.A.-K.; data interpretation, Q.D., K.S., J.A.-K., and S.J.C.; writing – original draft, Q.D. and S.J.C.; writing – review & editing, Q.D., C.S., J.A.-K., and S.J.C.; resources, M.A.S., E.L., J.E.P., and I.W.D.; funding acquisition, S.J.C. and C.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 23, 2020
Revised: June 22, 2021
Accepted: August 25, 2021
Published: September 21, 2021

REFERENCES

Abbott, K.L., Nyre, E.T., Abrahante, J., Ho, Y.Y., Isaksson Vogel, R., and Starr, T.K. (2015). The Candidate Cancer Gene Database: a database of cancer driver genes from forward genetic screens in mice. Nucleic Acids Res. *43*, D844–D848.

Achinger-Kawecka, J., Valdes-Mora, F., Luu, P.L., Giles, K.A., Caldon, C.E., Qu, W., Nair, S., Soto, S., Locke, W.J., Yeo-Teh, N.S., et al. (2020). Epigenetic reprogramming at estrogen-receptor binding sites alters 3D chromatin landscape in endocrine-resistant breast cancer. Nat. Commun. *11*, 320.

Achinger-Kawecka, J., Stirzaker, C., Chia, K.-M., Portman, N., Campbell, E., Du, Q., Laven-Law, G., Nair, S.S., Yong, A., Wilkinson, A., et al. (2021). Epigenetic therapy suppresses endocrine-resistant breast tumour growth by re-wiring ER-mediated 3D chromatin interactions. bioRxiv. https://doi.org/10.1101/2021.06.21.449340.

Alavattam, K.G., Maezawa, S., Sakashita, A., Khoury, H., Barski, A., Kaplan, N., and Namekawa, S.H. (2019). Attenuated chromatin compartmentalization in meiosis and its maturation in sperm development. Nat. Struct. Mol. Biol. *26*, 175–184.

Bakker, B., Taudt, A., Belderbos, M.E., Porubsky, D., Spierings, D.C., de Jong, T.V., Halsema, N., Kazemier, H.G., Hoekstra-Wakker, K., Bradley, A., et al. (2016). Single-cell sequencing reveals karyotype heterogeneity in murine and human malignancies. Genome Biol. *17*, 115.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell *143*, 470–484.

Baylin, S.B., and Jones, P.A. (2016). Epigenetic Determinants of Cancer. Cold Spring Harb. Perspect. Biol. *8*, a019505.

Belaghzal, H., Borrman, T., Stephens, A.D., Lafontaine, D.L., Venev, S.V., Weng, Z., Marko, J.F., and Dekker, J. (2019). Compartment-dependent chromatin interaction dynamics revealed by liquid chromatin Hi-C. bioRxiv. https://doi.org/10.1101/704957.

Bell, S.P., and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128–134.

Berman, B.P., Weisenberger, D.J., Aman, J.F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C.P., van Dijk, C.M., Tollenaar, R.A., et al. (2011).

Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. Nat. Genet. *44*, 40–46.

Bert, S.A., Robinson, M.D., Strbenac, D., Statham, A.L., Song, J.Z., Hulf, T., Sutherland, R.L., Coolen, M.W., Stirzaker, C., and Clark, S.J. (2013). Regional activation of the cancer genome by long-range epigenetic remodeling. Cancer Cell *23*, 9–22.

Casas-Delucchi, C.S., van Bemmel, J.G., Haase, S., Herce, H.D., Nowak, D., Meilinger, D., Stear, J.H., Leonhardt, H., and Cardoso, M.C. (2012). Histone hypoacetylation is required to maintain late replication timing of constitutive heterochromatin. Nucleic Acids Res. *40*, 159–169.

da Rocha, S.T., and Gendrel, A.V. (2019). The influence of DNA methylation on monoallelic expression. Essays Biochem. *63*, 663–676.

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y., et al. (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature *537*, 548–552.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. Gigascience *10*, giab008.

Dileep, V., and Gilbert, D.M. (2018). Single-cell replication profiling to measure stochastic variation in mammalian replication timing. Nat. Commun. *9*, 427.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dotan, Z.A., Dotan, A., Ramon, J., and Avivi, L. (2004). Altered mode of allelic replication accompanied by aneuploidy in peripheral blood lymphocytes of prostate cancer patients. Int. J. Cancer *111*, 60–66.

Dotan, Z.A., Dotan, A., Ramon, J., and Avivi, L. (2008). Aberrant allele-specific replication, independent of parental origin, in blood cells of cancer patients. BMC Cancer *8*, 390.

Du, Q., Bert, S.A., Armstrong, N.J., Caldon, C.E., Song, J.Z., Nair, S.S., Gould, C.M., Luu, P.-L., Peters, T., Khoury, A., et al. (2019). Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. Nat. Commun. *10*, 416.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016a). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. *3*, 99–101.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016b). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. *3*, 95–98.

Durkin, M.E., Yuan, B.Z., Zhou, X., Zimonjic, D.B., Lowy, D.R., Thorgeirsson, S.S., and Popescu, N.C. (2007). DLC-1:a Rho GTPase-activating protein and tumour suppressor. J. Cell. Mol. Med. *11*, 1185–1207.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.

Espada, J., Ballestar, E., Fraga, M.F., Villar-Garea, A., Juarranz, A., Stockert, J.C., Robertson, K.D., Fuks, F., and Esteller, M. (2004). Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern. J. Biol. Chem. *279*, 37175–37184.

Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B.R., Leonhardt, H., Joffe, B., Dekker, J., Fudenberg, G., Solovei, I., and Mirny, L.A. (2019). Heterochromatin drives compartmentalization of inverted and conventional nuclei. Nature *570*, 395–399.

Feng, X., Grossman, R., and Stein, L. (2011). PeakRanger: a cloud-enabled peak caller for ChIP-seq data. BMC Bioinformatics *12*, 139.

Fiziev, P., Akdemir, K.C., Miller, J.P., Keung, E.Z., Samant, N.S., Sharma, S., Natale, C.A., Terranova, C.J., Maitituoheti, M., Amin, S.B., et al. (2017). Systematic Epigenomic Analysis Reveals Chromatin States Associated with Melanoma Progression. Cell Rep. *19*, 875–889.

Gimelbrant, A., Hutchinson, J.N., Thompson, B.R., and Chess, A. (2007). Widespread monoallelic expression on human autosomes. Science *318*, 1136–1140.

Gupta, S., Lafontaine, D.L., Vigneau, S., Vinogradova, S., Mendelevich, A., Igarashi, K.J., Bortvin, A., Alves-Pereira, C.F., Clement, K., Pinello, L., et al. (2020). DNA methylation is a key mechanism for maintaining monoallelic expression on autosomes. bioRxiv. https://doi.org/10.1101/2020.02.20.954834.

Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yáñez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B., et al. (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell *169*, 693–707.e14.

Hanna, C.W., Taudt, A., Huang, J., Gahurova, L., Kranz, A., Andrews, S., Dean, W., Stewart, A.F., Colomé-Tatché, M., and Kelsey, G. (2018). MLL2 conveys transcription-independent H3K4 trimethylation in oocytes. Nat. Struct. Mol. Biol. *25*, 73–82.

Hansen, R.S., Stöger, R., Wijmenga, C., Stanek, A.M., Canfield, T.K., Luo, P., Matarazzo, M.R., D'Esposito, M., Feil, R., Gimelli, G., et al. (2000). Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. Hum. Mol. Genet. *9*, 2575–2587.

Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O., Gartler, S.M., and Stamatoyannopoulos, J.A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc. Natl. Acad. Sci. USA *107*, 139–144.

Hardwick, S.A., Chen, W.Y., Wong, T., Deveson, I.W., Blackburn, J., Andersen, S.B., Nielsen, L.K., Mattick, J.S., and Mercer, T.R. (2016). Spliced synthetic genes as internal controls in RNA sequencing experiments. Nat. Methods *13*, 792–798.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

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

Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., et al. (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res. *22*, 246–258.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods *9*, 999–1003.

Jablonka, E., Goitein, R., Marcus, M., and Cedar, H. (1985). DNA hypomethylation causes an increase in DNase-I sensitivity and an advance in the time of replication of the entire inactive X chromosome. Chromosoma *93*, 152–156.

Johnstone, S.E., Reyes, A., Qi, Y., Adriaens, C., Hegazi, E., Pelka, K., Chen, J.H., Zou, L.S., Drier, Y., Hecht, V., et al. (2020). Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer. Cell *182*, 1474–1489.e23.

Jørgensen, H.F., Azuara, V., Amoils, S., Spivakov, M., Terry, A., Nesterova, T., Cobb, B.S., Ramsahoye, B., Merkenschlager, M., and Fisher, A.G. (2007). The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells. Genome Biol. *8*, R169.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D.J., Nicholls, R.D., and Cedar, H. (1993). Allele-specific replication timing of imprinted gene regions. Nature *364*, 459–463.

Klein, K.N., Zhao, P.A., Lyu, X., Sasaki, T., Bartlett, D.A., Singh, A.M., Tasan, I., Zhang, M., Watts, L.P., Hiraga, S.I., et al. (2021). Replication timing maintains the global epigenetic state in human cells. Science *372*, 371–378.

Kruse, K., Hug, C.B., Hernández-Rodríguez, B., and Vaquerizas, J.M. (2016). TADtool: visual parameter identification for TAD-calling algorithms. Bioinformatics *32*, 3190–3192.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol. *9*, e1003118.

Lay, F.D., Liu, Y., Kelly, T.K., Witt, H., Farnham, P.J., Jones, P.A., and Berman, B.P. (2015). The role of DNA methylation in directing the functional organization of the cancer epigenome. Genome Res. *25*, 467–477.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

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

Li, T., Li, Z., Wan, H., Tang, X., Wang, H., Chai, F., Zhang, M., and Wang, B. (2020). Recurrence-Associated Long Non-coding RNA LNAPPCC Facilitates Colon Cancer Progression via Forming a Positive Feedback Loop with PCDH7. Mol. Ther. Nucleic Acids *20*, 545–557.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H., et al. (2016). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature *537*, 558–562.

Liu, Z., Dong, X., and Li, Y. (2018). A Genome-Wide Study of Allele-Specific Expression in Colorectal Cancer. Front. Genet. *9*, 570.

Luedi, P.P., Dietrich, F.S., Weidman, J.R., Bosko, J.M., Jirtle, R.L., and Hartemink, A.J. (2007). Computational and experimental identification of novel human imprinted genes. Genome Res. *17*, 1723–1730.

Lund, E., Oldenburg, A.R., and Collas, P. (2014). Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. Nucleic Acids Res. *42*, E92.

Luraghi, P., Bigatto, V., Cipriano, E., Reato, G., Orzan, F., Sassi, F., De Bacco, F., Isella, C., Bellomo, S.E., Medico, E., et al. (2017). A molecularly annotated model of patient-derived colon cancer stem-like cells to assess genetic and nongenetic mechanisms of resistance to anti-EGFR therapy. Clin. Cancer Res. *24*, 807–820.

MacPherson, Q., Beltran, B., and Spakowitz, A.J. (2018). Bottom-up modeling of chromatin segregation due to epigenetic modifications. Proc. Natl. Acad. Sci. USA *115*, 12739–12744.

Miura, H., Takahashi, S., Poonperm, R., Tanigawa, A., Takebayashi, S.I., and Hiratani, I. (2019). Single-cell DNA replication profiling identifies spatiotemporal developmental dynamics of chromosome organization. Nat. Genet. *51*, 1356–1368.

Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273.

Nagler, A., Cytron, S., Mashevich, M., Korenstein-Ilan, A., and Avivi, L. (2010). The aberrant asynchronous replication - characterizing lymphocytes of cancer patients - is erased following stem cell transplantation. BMC Cancer *10*, 230.

Nair, S.S., Luu, P.L., Qu, W., Maddugoda, M., Huschtscha, L., Reddel, R., Chenevix-Trench, G., Toso, M., Kench, J.G., Horvath, L.G., et al. (2018). Guidelines for whole genome bisulphite sequencing of intact and FFPET DNA on the Illumina HiSeq X Ten. Epigenetics Chromatin *11*, 24.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nothjunge, S., Nührenberg, T.G., Grüning, B.A., Doppler, S.A., Preissl, S., Schwaderer, M., Rommel, C., Krane, M., Hein, L., and Gilsbach, R. (2017). DNA methylation signatures follow preformed chromatin compartments in cardiac myocytes. Nat. Commun. *8*, 1667.

Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics *32*, 292–294.

Pandiyan, K., You, J.S., Yang, X., Dai, C., Zhou, X.J., Baylin, S.B., Jones, P.A., and Liang, G. (2013). Functional DNA demethylation is accompanied by chromatin accessibility. Nucleic Acids Res. *41*, 3973–3985.

Pedersen, B.S., Eyring, K., De, S., Yang, I.V., and Schwartz, D.A. (2014). Fast and accurate alignment of long bisulfite-seq reads. arXiv, 1401.1129v2. https://arxiv.org/abs/1401.1129v2.

Peng, H., Long, F., Wu, Z., Chu, Y., Li, J., Kuai, R., Zhang, J., Kang, Z., Zhang, X., and Guan, M. (2013). Downregulation of DLC-1 gene by promoter methylation during primary colorectal cancer progression. BioMed Res. Int. *2013*, 181384.

Pidsley, R., Zotenko, E., Peters, T.J., Lawrence, M.G., Risbridger, G.P., Molloy, P., Van Djik, S., Muhlhausler, B., Stirzaker, C., and Clark, S.J. (2016). Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. *17*, 208.

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

R Development Core Team (2015). R: a language and environment for statistical computing (R Foundation for Statistical Computing).

Reddington, J.P., Perricone, S.M., Nestor, C.E., Reichmann, J., Youngson, N.A., Suzuki, M., Reinhardt, D., Dunican, D.S., Prendergast, J.G., Mjoseng, H., et al. (2013). Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. Genome Biol. *14*, R25.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell *139*, 719–730.

Rhee, I., Bachman, K.E., Park, B.H., Jair, K.W., Yen, R.W., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., et al. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature *416*, 552–556.

Rivera-Mulia, J.C., Buckley, Q., Sasaki, T., Zimmerman, J., Didier, R.A., Nazor, K., Loring, J.F., Lian, Z., Weissman, S., Robins, A.J., et al. (2015). Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. Genome Res. *25*, 1091–1103.

Rivera-Mulia, J.C., Dimond, A., Vera, D., Trevilla-Garcia, C., Sasaki, T., Zimmerman, J., Dupont, C., Gribnau, J., Fraser, P., and Gilbert, D.M. (2018). Allele-specific control of replication timing and genome organization during development. Genome Res. *28*, 800–811.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res. *20*, 761–770.

Sadedin, S.P., Pope, B., and Oshlack, A. (2012). Bpipe: a tool for running and managing bioinformatics pipelines. Bioinformatics *28*, 1525–1526.

Saksouk, N., Barth, T.K., Ziegler-Birling, C., Olova, N., Nowak, A., Rey, E., Mateos-Langerak, J., Urbach, S., Reik, W., Torres-Padilla, M.E., et al. (2014). Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. Mol. Cell *56*, 580–594.

Sati, S., Bonev, B., Szabo, Q., Jost, D., Bensadoun, P., Serra, F., Loubiere, V., Papadopoulos, G.L., Rivera-Mulia, J.C., Fritsch, L., et al. (2020). 4D Genome Rewiring during Oncogene-Induced and Replicative Senescence. Mol. Cell *78*, 522–538.e9.

Savova, V., Chun, S., Sohail, M., McCole, R.B., Witwicki, R., Gai, L., Lenz, T.L., Wu, C.T., Sunyaev, S.R., and Gimelbrant, A.A. (2016). Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. Nat. Genet. *48*, 231–237.

Schalkwyk, L.C., Meaburn, E.L., Smith, R., Dempster, E.L., Jeffries, A.R., Davies, M.N., Plomin, R., and Mill, J. (2010). Allelic skewing of DNA methylation is widespread across the genome. Am. J. Hum. Genet. *86*, 196–212.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Schwaiger, M., Kohler, H., Oakeley, E.J., Stadler, M.B., and Schübeler, D. (2010). Heterochromatin protein 1 (HP1) modulates replication timing of the Drosophila genome. Genome Res. *20*, 771–780.

Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., and Schatz, M.C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. Nat. Methods *15*, 461–468.

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

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Shah, P.P., Donahue, G., Otte, G.L., Capell, B.C., Nelson, D.M., Cao, K., Aggarwala, V., Cruickshanks, H.A., Rai, T.S., McBryan, T., et al. (2013). Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. Genes Dev. *27*, 1787–1799.

Sheffield, N.C., and Bock, C. (2016). LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics *32*, 587–589.

Shen, H., and Laird, P.W. (2013). Interplay between the cancer genome and epigenome. Cell *153*, 38–55.

Song, Q., Decato, B., Hong, E.E., Zhou, M., Fang, F., Qu, J., Garvin, T., Kessler, M., Zhou, J., and Smith, A.D. (2013). A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. PLoS ONE *8*, e81148.

Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., Gomez, A., Collombet, S., Berenguer, C., Cuartero, Y., et al. (2018). Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat. Genet. *50*, 238–249.

Stahler, A., Heinemann, V., Neumann, J., Crispin, A., Schalhorn, A., Stintzing, S., Giessen-Jung, C., Fischer von Weikersthal, L., Vehling-Kaiser, U., Stauch, M., et al. (2017). Prevalence and influence on outcome of HER2/neu, HER3 and NRG1 expression in patients with metastatic colorectal cancer. Anticancer Drugs *28*, 717–722.

Stempor, P., and Ahringer, J. (2016). SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics. Wellcome Open Res. *1*, 14.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Taberlay, P.C., Statham, A.L., Kelly, T.K., Clark, S.J., and Jones, P.A. (2014). Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. Genome Res. *24*, 1421–1432.

Taberlay, P.C., Achinger-Kawecka, J., Lun, A.T., Buske, F.A., Sabir, K., Gould, C.M., Zotenko, E., Bert, S.A., Giles, K.A., Bauer, D.C., et al. (2016). Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. Genome Res. *26*, 719–731.

Takagi, N., and Oshimura, M. (1973). Fluorescence and Giemsa banding studies of the allocyclic X chromosome in embryonic and adult mouse cells. Exp. Cell Res. *78*, 127–135.

Takahashi, S., Miura, H., Shibata, T., Nagao, K., Okumura, K., Ogata, M., Obuse, C., Takebayashi, S.I., and Hiratani, I. (2019). Genome-wide stability of the DNA replication program in single mammalian cells. Nat. Genet. *51*, 529–540.

van der Weide, R.H., van den Brand, T., Haarhuis, J.H.I., Teunissen, H., Rowland, B.D., and de Wit, E. (2021). Hi-C analyses with GENOVA: a case study with cohesin variants. NAR Genom Bioinformatics *3*, lqab040.

van Roy, F. (2014). Beyond E-cadherin: roles of other cadherin superfamily members in cancer. Nat. Rev. Cancer *14*, 121–134.

Walker, E.J., Zhang, C., Castelo-Branco, P., Hawkins, C., Wilson, W., Zhukova, N., Alon, N., Novokmet, A., Baskin, B., Ray, P., et al. (2012). Monoallelic expression determines oncogenic progression and outcome in benign and malignant brain tumors. Cancer Res. *72*, 636–644.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer).

Wu, R., Singh, P.B., and Gilbert, D.M. (2006). Uncoupling global and fine-tuning replication timing determinants for mouse pericentric heterochromatin. J. Cell Biol. *174*, 185–194.

Wu, R., Wang, Z., Zhang, H., Gan, H., and Zhang, Z. (2017). H3K9me3 demethylase Kdm4d facilitates the formation of pre-initiative complex and regulates DNA replication. Nucleic Acids Res. *45*, 169–180.

Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J., et al. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J. *36*, 3573–3599.

Xie, W.J., Meng, L., Liu, S., Zhang, L., Cai, X., and Gao, Y.Q. (2017). Structural Modeling of Chromatin Integrates Genome Features and Reveals Chromosome Folding Principle. Sci. Rep. *7*, 2818.

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., et al. (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature *537*, 553–557.

Zhao, P.A., Sasaki, T., and Gilbert, D.M. (2020). High-resolution Repli-Seq defines the temporal choreography of initiation, elongation and termination of replication in mammalian cells. Genome Biol. *21*, 76.

Zhou, A., Lin, T., and Xing, J. (2019). Evaluating nanopore sequencing data processing pipelines for structural variation identification. Genome Biol. *20*, 237.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-DNMT1 (N-term)	Sigma-Aldrich	Cat#D4692; RRID: AB_262096
Anti-DNMT1 [EPR3522] (C-term)	Abcam	Cat#ab92314; RRID: AB_10562817
Anti-GAPDH [6C5]	Thermo Fisher Scientific	Cat#AM4300; RRID: AB_2536381
Anti-BrdU	BD Biosciences	Cat#555627; RRID: AB_395993
Chemicals, peptides, and recombinant proteins		
5-Bromo-2′-deoxyuridine (BrdU)	Sigma	Cat#B5002; CAS: 59-14-3
Vybrant DyeCycle Violet Ready Flow	Invitrogen	Cat#R37172
Critical commercial assays		
EZ DNA Methylation-Gold Kit	Zymo Research	Cat#D5005
CEGX TrueMethyl® Whole Genome Kit (v3.1)	CEGX	Cat#CEGXTMWG
Epicenter EpiGnome™ Methyl-Seq Kit	Illumina	Cat#EGMK81312
Arima-HiC+ kit	Arima Genomics Inc.	Cat#A510008
Chromium Single Cell DNA Library & Gel Bead Kit	10X Genomics	Cat#PN-1000041
Chromium Single Cell DNA Cell Bead Kit	10X Genomics	Cat#PN-1000057
Chromium Chip C Single Cell DNA Kit	10X Genomics	Cat#PN-1000032
Chromium Chip D Single Cell DNA Kit	10X Genomics	Cat#PN-1000036
Ligation Sequencing Kit 1D	Oxford Nanopore Technologies	Cat#SQK-LSK109
TruSeq Stranded mRNA Library Prep kit	Illumina	Cat#RS-122-2102
Deposited data		
HCT116 and DKO1: Repli-Seq, Hi-C, WGBS, single cell Repli-Seq, Nanopore, RNA-seq	This paper	GEO: GSE158011
HCT116 and DKO1: ChIP-seq	(Lay et al., 2015)	GEO: GSE58638
PDX tumor mouse model: Hi-C and DNA methylation EPIC array	(Achinger-Kawecka et al., 2021)	GEO: GSE171074
PrEC and LNCaP prostate cell lines: Repli-Seq	(Du et al., 2019)	GEO: GSE98732
PrEC and LNCaP prostate cell lines: WGBS	(Pidsley et al., 2016)	GEO: GSE86833
Human colorectal normal and tumor samples	(Johnstone et al., 2020)	GEO: GSE133928
WI38 human fibroblast *DNMT1* siRNA: Hi-C and WGBS	(Sati et al., 2020)	GEO: GSE130306
Mouse cardiomyocyte WT and *Dnmt3a/b* DKO: Hi-C and WGBS	(Nothjunge et al., 2017)	NCBI SRA: PRJNA378914 & PRJNA229470
Experimental models: cell lines		
HCT116 and DKO1	Laboratory of Stephen Baylin	(Rhee et al., 2002)
Oligonucleotides		
Primer: *DNMT1* Forward: CATCCTGTACCGAGTTGGTG	This paper	N/A
Primer: *DNMT1* Reverse: TTTCACGGGACTGGACAGC	This paper	N/A
Primer: *DNMT3A* Forward: CCCATTCGATCTGGTGATTG	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Primer: *DNMT3A* Reverse: CAGGAGGCGGTAGAACTCAAAG	This paper	N/A
Primer: *DNMT3B* Forward: TCTCTCAAATGTGAATCCAGCC	This paper	N/A
Primer: *DNMT3B* Reverse: GGCGTGAGTAATTCAGCAGG	This paper	N/A
Software and algorithms		
Custom code used in this study	This paper	https://github.com/qianxidu/Replication_Timing_Du_et_al_2021; DOI: 10.5281/zenodo.5240900
ImageJ	(Schneider et al., 2012)	https://imagej.nih.gov/ij/
Meth10X (v1.2)	(Nair et al., 2018)	https://github.com/luuloi/Meth10X
MethPipe (v3.4.2)	(Song et al., 2013)	http://smithlabresearch.org/software/methpipe/
Repli-Seq processing	(Du et al., 2019)	https://github.com/clark-lab/Replication-Timing/
HiC-Pro (v2.11.4)	(Servant et al., 2015)	https://github.com/nservant/HiC-Pro
Juicer (v1.6)	(Durand et al., 2016b)	https://github.com/aidenlab/juicer
JuiceBox (v0.76)	(Durand et al., 2016a)	https://github.com/aidenlab/Juicebox
TADtool (v0.76)	(Kruse et al., 2016)	https://github.com/vaquerizaslab/tadtool
Homer (v4.6)	(Heinz et al., 2010)	http://homer.ucsd.edu/homer/
GENOVA (v1.0.0)	(van der Weide et al., 2021)	https://github.com/robinweide/GENOVA
10X Genomics Cell Ranger DNA (v1.1.0)	10X Genomics	https://support.10xgenomics.com/single-cell-dna/software/pipelines/latest/what-is-cell-ranger-dna
AneuFinder (v1.14.0)	(Bakker et al., 2016)	https://github.com/ataudt/aneufinder
Guppy (v3.3.0)	Oxford Nanopore Technologies	https://community.nanoporetech.com/sso/login?next_url= %2Fdownloads
minimap2 (v.2.17-r943-dirty)	(Li, 2018)	https://github.com/lh3/minimap2
medaka_variant (v0.11.4)	Oxford Nanopore Technologies	https://nanoporetech.github.io/medaka
NGMLR (v0.2.7)	(Sedlazeck et al., 2018)	https://github.com/philres/ngmlr
Sniffles (v1.0.9)	(Sedlazeck et al., 2018)	https://github.com/fritzsedlazeck/Sniffles
bcftools (v1.9)	(Danecek et al., 2021)	https://samtools.github.io/bcftools/
bowtie (v1.1.0)	(Langmead et al., 2009)	http://bowtie-bio.sourceforge.net/index.shtml
SAMtools (v1.10)	(Li et al., 2009)	http://www.htslib.org
BEDtools (v2.26.0)	(Quinlan and Hall, 2010)	https://github.com/arq5x/bedtools2
PeakRanger (v1.16)	(Feng et al., 2011)	http://ranger.sourceforge.net
Enriched Domain Detector (EDD)	(Lund et al., 2014)	https://github.com/CollasLab/edd
chromHMM (v1.10)	(Ernst and Kellis, 2012)	http://compbio.mit.edu/ChromHMM/
Trim Galore (v0.4.0)	Babraham Institute	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
STAR (v2.5.3a)	(Dobin et al., 2013)	https://github.com/alexdobin/STAR
rsem (v1.2.21)	(Li and Dewey, 2011)	http://deweylab.github.io/RSEM/
edgeR (v3.12.1)	(Robinson et al., 2010)	https://bioconductor.org/packages/release/bioc/html/edgeR.html
GSEA (v4.1.0)	(Subramanian et al., 2005) & (Mootha et al., 2003)	http://www.gsea-msigdb.org/gsea/
SeqPlots	(Stempor and Ahringer, 2016)	https://github.com/Przemol/seqplots

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Susan Clark (s.clark@garvan.org.au).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data generated in this study have been deposited at NCBI Gene Expression Omnibus (GEO) and are publically available as of the date of publication. Accession numbers are listed in the Key resources table. This paper analyses existing, publically available data. Accession numbers for public datasets are also listed in the Key resources table.
- All custom code has been deposited at https://github.com/qianxidu/Replication_Timing_Du_et_al_2021 (https://doi.org/10.5281/zenodo.5240900) and is publically available as of the date of publication.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

HCT116 and DKO1 cells were kindly provided by Prof. Stephen Baylin (The Johns Hopkins School of Medicine, The Sidney Kimmel Comprehensive Cancer Center). HCT116 human male colorectal cancer cells were cultured in McCoy's 5A modified medium (GIBCO, #16600-082) supplemented with heat inactivated fetal bovine serum (10%, GIBCO, #16000-044) at 37°C and 5% CO_2. Cells at 70%–80% confluency were rinsed in PBS (1x Phosphate-buffered saline, GIBCO, #14190-144) and trypsinised in Trypsin-EDTA (0.05%, GIBCO, #15400054). Trypsin was inactivated using equal volume growth medium and cells were pelleted at 250xg for 5 min. Cells were resuspended in growth medium and typically split 1:8. HCT116 with double knockouts (KO) in DNMT1 and DNMT3B (DKO1) (Rhee et al., 2002) were selected in growth medium supplemented with hygromycin (0.05 mg/mL, GIBCO, #10687-010) and geneticin (0.1 mg/mL, GIBCO, #10131) for 1 week after thawing to ensure a pure KO population, then further cultured without selection. DKO1 cells were subcultured following the same protocol as for HCT116 above, typically at a 1:4 split. HCT116 and DKO1 cells were validated for double knock out of *DNMT1* and *DNMT3B* using western blot and expression qRT-PCR (Figures S1I–S1K).

METHOD DETAILS

Expression qRT-qPCR of DNMT genes

RNA was extracted from cultured cells using TRIzol (Life Technologies, #15596018) and DNaseI treated (NEB, #M0303) as per manufacturer's instructions. cDNA was generated from 500 ng of RNA using SuperScript III reverse transcriptase (Invitrogen, #18080-044) and random hexamers according to manufacturer's instructions. 1 μL of RNA sequins (Hardwick et al., 2016) (mix A, 1:100 dilution) was spiked in to RNA samples prior to making cDNA to be used as negative controls for qRT-PCR. Primers to *DNMT1* were designed to exons present in both the full transcript and the hypomorphic product present in DKO1 cells (See Key resources table). Primers to *DNMT3A* and *-3B* were designed to capture the majority of transcripts annotated by the GENCODE genes v19 track in UCSC Genome Browser (Kent et al., 2002) (See Key resources table). All primer pairs were designed over an intron to avoid genomic products.

Western blotting for DNMT proteins

Whole cell protein lysates were prepared by resuspending scaped cells in modified RIPA buffer (50 mM Tris-HCL pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM NaF, 1 mM Na_3VO_4, 1% Igepal, 0.25% Sodium-deoxycholate) with protease inhibitors and incubated on ice for 30 min with vortexing every 5 min. Lysate was sonicated for 15 s at 25% amplitude with a microtip using the QSonica (Q55) and stored at −80°C. Protein concentration was determined using a BCA assay (Pierce, #23225) according to manufacturer's instructions. Protein lysate was prepared using NuPAGE® LDS sample buffer (Life Technologies, NP0007) and NuPAGE® sample reducing agent (Life Technologies, NP0004) followed by heating at 95°C for 10 min. Samples were resolved by gel electrophoresis using the NuPage Bis-Tris 4%–12% precast gel system according to the manufacturer's instructions (Life Technologies). Western blot transfer was carried out according to the manufacturer's instructions with the SureLock X-cell system (ThermoFisher Scientific) using transfer buffer containing 20% methanol content. Antibodies used are as follows: N-terminal DNMT1 – Sigma-Aldrich #D4692; C-terminal DNMT1 – Abcam #ab92314; GAPDH – Invitrogen #AM4300. Western blots (WBs) were then treated with Western Lightning Plus-ECL (Perkin Elmer, #NEL103E001) before developing on Super Rx Fuji Medical X-Ray Film (Fujifilm, #4741019236) using the Konica Tabletop X-Ray Film Processor (#SRX101A). Developed film was scanned using the Epson Perfection V800/850 scanner.

ImageJ was used to quantitate WB films following the *Gel Analysis* method outlined in their documentation. Band densities of the protein of interest were divided by the loading control and converted to a percentage of the highest ratio (most dense band compared to loading control). Average and standard error of the mean were calculated for replicates.

Nanopore sequencing and base calling

HCT116 and DKO1 DNA (1 μg) was sheared using the Covaris g-TUBE spun at 3400 x g in 2 × 60 s spins. Sheared DNA was prepared for Nanopore sequencing using the Ligation 1D kit (SQK-LSK109) according to manufacturer's instructions. Each cell line was sequenced on one PromethION flow cell. Reads were base called with Guppy (v3.3.0) on a GPU-enabled Sun Grid Engine high performance computing server (parameters "–chunks_per_runner 1500–gpu_runners_per_device 1–cpu_threads_per_caller 4 -x "cuda:0 cuda:1 cuda:2 cuda:3" -r" and configuration "dna_r9.4.1_450bps_hac_prom.cfg").

Calling structural variants

Structural variants (SVs) were called using both Nanopore and single cell DNA sequencing (scCNV) datasets. Analysis was limited to autosomes and chrX. For Nanopore, fastq files were mapped with NGMLR (v0.2.7, (Sedlazeck et al., 2018)) to hg19. DKO1 bam file was downsampled to match median coverage of HCT116. Sniffles (v1.0.9, (Sedlazeck et al., 2018)) was used to call structural variants (SVs) using parameters '—genotype –cluster'. SVs were filtered as follows: 'FILTER = PASS', larger than 50bp, allele frequency over 30%. Due to the imprecise breakpoint mapping of Nanopore data, SVs were considered the same between HCT116 and DKO1 if they reciprocally overlapped by 50% and were of the same SV type (Zhou et al., 2019). For single cell DNA sequencing datasets, consensus copy number variations for HCT116 and DKO1 G1 cells were generated by the Chromium 10X Cell Ranger DNA pipeline (v1.1.0) ('node_unmerged_cnv_calls.bed'). Bins (20kb) with any copy number change and with confidence scores greater than or equal to 5 were counted as variable between HCT116 and DKO1. SVs that differ between HCT116 and DKO1 from both Nanopore and scDNA datasets were merged. Due to the scale of replication timing (50kb sliding windows) and Hi-C data (20kb bins), SV regions were considered for removal if they were equal to or over 10kb in width.

Whole genome bisulphite sequencing and processing

200 ng of DNA was bisulphite converted using the EZ DNA Methylation-Gold Kit (Zymo, #D5005) according to manufacturer's instructions. Input DNA was spiked with unmethylated lambda DNA (0.5%) (Promega, #D1521). Replicate bisulphite libraries were generated with the CEGX TrueMethyl® Whole Genome Kit (CEGX, #CEGXTMWG, v3.1) according to manufacturer's instructions. Libraries were sequenced on the Illumina X Ten. Sequencing reads from WGBS data were aligned to the human genome using v1.2 of an internally developed pipeline Meth10X (Nair et al., 2018). This is publicly available and can be downloaded from https://github.com/luuloi/Meth10X. The pipeline backbone is built based on workflow control Bpipe (v0.9.9.2) (Sadedin et al., 2012). Briefly, adaptor sequences were removed using in-house bash script in paired-end mode following library prep kit guide. Bwa-meth (v0.20) (Pedersen et al., 2014) was then used to align reads to hg19 using default parameters. The generated bam files were marked for duplicates using Picard (v2.3.0) (http://broadinstitute.github.io/picard). Bam files were then quality checked using Qualimap (v2.2.1) (Okonechnikov et al., 2016). Count tables of the number of methylated and unmethylated bases sequenced at each CpG site in the genome were constructed using MethylDackel (https://github.com/dpryan79/MethylDackel) and Biscuit (https://github.com/zhou-lab/biscuit). Biscuit was used to call SNPs that were discounted from the final table. Sequencing metrics can be found in Table S2. Downstream analyses were limited to autosomes and chrX. Partially methylated domains (PMDs) were called using MethPipe (v3.4.2) (Song et al., 2013). Total methylation levels were calculated by dividing the sum of all C calls with the sum of all C+T calls and CpGs with a minimum coverage of 5 were used for downstream analyses.

Repli-Seq data generation and processing

Repli-Seq was performed in duplicate for each cell line as previously described with slight modifications (Du et al., 2019). Briefly, cells were labeled with BrdU (50 μM, Sigma, #B5002) for two hours. Labeled cells were sorted into 6 fractions across the cell cycle (G1b, S1, S2, S3, S4, G2M) as per protocol on the FACS Aria III. DNA extraction and BrdU-labeled DNA immunoprecipitation were performed with anti-BrdU antibody (40 μL of 25 μg mL^{-1}, BD PharMingen, #555627). Validation of BrdU immunoprecipitation was carried out using qRT-PCR on known Early (*BMP1*) and Late (*DPPA2*) loci, and compared to a fractionation negative control (MITO) (Figures S1A and S1B). As the mitochondrial genome (MITO) replicates independently of the nuclear genome, it should not show differences between S-phase fractions. 10 ng of ssDNA was used as input for the Epicenter EpiGnome Methyl-Seq Kit (Illumina, EGMK81312, now called the TruSeq® DNA Methylation kit) and processed according to manufacturer's instructions. The ssDNA was not bisulphite converted prior to library preparation. Libraries were sequenced on the HiSeq 2500 as 50bp single-end reads. Full sequencing outputs can be found in Table S3.

Replication timing weighted average (WA) scores were calculated as previously described (Du et al., 2019). Bins overlapping SVs were removed prior to downstream analyses. WA values for replicates of HCT116 and DKO1 were highly correlated (r^2 values > 0.99) (Figure S1C). The distributions of WA scores were comparable to the WA distributions in other normal and cancer cell Repli-Seq datasets (n = 16) (Figure S1D). To get a single score per cell line, replicate HCT116 and DKO1 weighted average (WA) replication timing scores were quantile normalized and the replicates averaged. Early- and late-replicating regions were defined as those regions in the top and bottom 10% of WA scores in both cell lines. This definition gives upper and lower limits of 77.61 and 16.10 respectively

(i.e., early regions have WA > 77.76 and late regions WA < 16.07). The limits were rounded to 78 and 16 for downstream analyses. WA thresholds for a change in timing were were calculated as previously described (Du et al., 2019). Differences in WA that are larger than ± 15 ΔWA are therefore considered to show a robust change in replication timing (Figures S1E and S1F). To identify domains of loci with changed replication timing, we merged all loci within 50kb that had |ΔWA| > 15. This approach gave 169 *earlier* domains and 116 *later* domains called between HCT116 and DKO cells.

Defining partially methylated domain boundary shifts

Partially methylated domains (PMDs) were called from WGBS data using the package MethPipe (v3.4.2) (Song et al., 2013). To call PMD regions for HCT116 and DKO1, first, replicate PMD regions were merged and any regions smaller than 50kb were removed. Then regions were merged if they were within 50kb of each other. The cut-off of 50kb was used because this is the resolution of the replication timing data. SVs were removed from PMDs prior to downstream analyses. To define PMD boundary shifts, we first calculated which PMD regions overlap between HCT116 and DKO1, then calculated the distance between the 5′ start coordinate of the HCT116 PMD to the 5′ start coordinate of the DKO1 PMD. The same was repeated for the 3′ end coordinate. The 5′ and 3′ coordinates were then categorised by the degree of shift inward from the HCT116 to DKO1 into 4 categories, < 50kb, 50-200kb, 200-500kb, > 500kb. Due to the low numbers of regions within the larger shift categories, 5′ and 3′ regions were pooled prior to plotting the changes in replication timing and DNA methylation.

Hi-C library preparation

Hi-C data in triplicate was generated using the Arima-HiC kit, according to the manufacturers protocols (Cat. #A510008). Briefly, cells were cross-linked with 2% formaldehyde to obtain 1-5 μg of DNA per Hi-C reaction. The Arima kit uses two restriction enzymes: ^GATC (DpnII), G^ANTC (N can be either of the 4 genomic bases) (HinfI), which after ligation of DNA ends generates 4 possible ligation junctions in the chimeric reads: GATC-GATC, GANT-GATC, GANT-ANTC, GATC-ANTC. Hi-C libraries were prepared using the KAPA Hyper Library Prep Kit with a modified protocol provided by Arima with 12 PCR cycles for library amplification as required. Hi-C libraries were sequenced on Illumina HiSeq X Ten in 150bp paired-end mode.

Hi-C data processing

HiC-Pro (Servant et al., 2015) (v2.11.4) was used to align and filter the Hi-C data, identify chromatin interactions, and generate Hi-C heatmaps. To generate filtered Hi-C contact matrices, the Hi-C reads were aligned against the human reference genome (hg19) and corrected using the ICE "correction" algorithm (Imakaev et al., 2012) built into HiC-Pro. Statistics on the number of read pairs, valid interactions and interactions in *cis* are presented in Table S4. Contact matrices used in down-stream analysis were Knight-Ruiz (KR)-normalized using JuiceBox tools (Durand et al., 2016a, 2016b) using Hi-C contact matrices in .hic format generated by hicpro2juicebox script in HiC-Pro as input. Obtained Hi-C matrices and Pearson correlation matrices were visualized in JuiceBox (Durand et al., 2016a). Public mouse Hi-C datasets from Nothjunge et al. (2017) was mapped to mouse reference genome mm10 and processed as above. Hi-C datasets from patient-derived xenograft (PDX) tumor mouse models were mapped to human reference genome hg38 and processed as above.

TADs

KR-normalized contact matrices were retrieved from Juicer for all chromosomes at 40kb resolution and TADs were identified using TADtool with the insulation score algorithm (Kruse et al., 2016). Analyses were limited to autosomes and chrX. We called TADs with a window size value of 103kb and a TAD cutoff of 30. We found that these parameters show good agreement between identified TADs and visual inspection of Hi-C datasets in JuiceBox and TADtool. The exact SV region was removed from TADs prior to downstream analyses.

A/B compartments

Compartment analysis was performed using the Homer pipeline (Heinz et al., 2010) (v4.6) with Hi-C KR-normalized contact matrices as input. Homer performs a principal component analysis of the normalized interaction matrices and uses the PC1 values to predict regions of active (A-compartments) and inactive chromatin (B-compartments). Homer works under the assumption that gene-rich regions have similar PC1 values, while gene-poor regions show differing PC1 values and assigns compartment status based on genome-wide gene density. Bins overlapping SVs were removed prior to downstream analyses.

Defining domains of genome reorganization

We initially tried to define genome reorganization through A/B compartment switching using compartments defined by HOMER. However, we observed that a large proportion of compartment switches were centered around PC1 values of zero, close to the A/B boundary; 46.94% of A to B switches occur where HCT116 PC1 < 0.5 and DKO1 PC1 > −0.5, and 42.87% of B to A switches occur where HCT116 PC1 > −0.5 and DKO1 PC1 < 0.5 (Figure S2E). This indicates that although compartment switches do cross the midline that defines A versus B compartments, a large proportion of these switches show little difference in PC1 values between HCT116 and DKO1 (Figure S2F). Therefore, we defined regions of genome organization change by their ΔPC1 score, with a cut-off of $\Delta PC1 \geq |1|$. PC1 values were quantile normalized and then the cutoff was defined by first examining ΔPC1 values within

replicates of HCT116 or DKO1 Hi-C datasets, then determining the ΔPC1 values where less than 5% (2.5% either tail) of the genome would be called 'differential' among the replicates (Figure S2G). Called domains were merged if within 50kb.

Hi-C compartment strength calculation

A/B compartment strengths were calculated as previously described (Stadhouders et al., 2018). Briefly, 100kb iterative corrected matrices were generated by HiC-Pro using the hicpro2juicebox utility. Bins overlapping SVs were removed prior to iterative correction. 100kb bins were grouped into 50 percentile groups based on their PC1 (1st eigenvector) value. Within pairwise combinations of the 50 percentile groups, average contact enrichments (obs/exp) between bins were calculated using GENOVA (v1.0.0, https://github.com/robinweide/GENOVA). Log_2 of the contact enrichment scores were plotted as a heat saddle plot. Summarized A-A and B-B compartment strengths were calculated as the mean log_2 contact enrichment between the top (A-A) or bottom (B-B) 20% of PC1 percentiles, and between the top and bottom 20% of PC1 percentiles for A-B compartment strength, excluding chrY and chrM. The compartmentalisation score was calculated as previously described using mean contact enrichments for A-A, B-B and A-B following this formula: $log(A\text{-}A*B\text{-}B/A\text{-}B^2)$ (Alavattam et al., 2019).

Weighted variance calculation for Repli-Seq fractions

The 6-fraction Percent-normalized Density Values (PNDVs) were used for this calculation. Briefly, PNDV values for one fraction represent the % of replication occurring within that timing fraction at any given 1kb locus. For example, the PNDV values for a locus are G1b = 40, S1 = 38, S2 = 9, S3 = 3, S4 = 4, G2M = 6. This means that 40% of this locus is replicated in the G1 fraction, 38% of this locus is replicated in the S1 fraction etc. This locus is biased or 'weighted' toward early replication timing. The sum of all fractions for any locus adds up to 100%. We made the assumption that if replication occurs evenly throughout S-phase, then each fraction from G1b to G2M should increment by 16.67% (i.e., 100%/6 fractions) to give G1 = 16.67, S1 = 33.33, S2 = 50.00, S3 = 66.67, S4 = 83.33, G2 = 100. This represents a 'neutral' locus that is unbiased or unweighted toward either early or late replication timing. The incremental nature of the 'neutral' locus informs the order of the S-phase fractions. A real locus is biased or 'weighted' toward early or late replication. To perform the weighted variation calculation, the PNDV values for each loci is used as the 'weights' against the pseudo 'neutral' locus, giving a measure of how the locus deviates from the 'neutral' locus. Formulae can be found in Figure S4C and S4D and a table of calibration tests can be found in Table S5. The weighted variance score was discretised for visualization purposes (Figure 5) into 4 bins from 0 to 0.081 in 0.02 intervals and called var1-4.

Single cell replication timing

Cell sorting and library generation using Chromium 10X single cell CNV solution. We stained HCT116 and DKO1 cells using a live cell double stranded DNA dye, Vybrant DyeCycle Violet Ready Flow (Invitrogen, #R37172), according to manufacturers' instructions. Cells were sorted (FACS Aria III) into 4-fractions: G1, Early, Mid and Late (Figure S5A). Equal numbers of Early, Mid and Late cells were pooled prior to use in the Single Cell CNV system to meet the minimum recommended cell recovery number (250 cells). We aimed for 500 recovered cells. As we did not need as many G1 cells or a specific number of G1 cells, G1 cells were loaded below minimum recommended cell stock concentration and we aimed for 50 recovered cells. Single cell capture, library generation and sequencing were performed by the Garvan-Weizmann Centre for Cellular Genomics (GWCCG). Libraries were sequenced on an Illumina NovaSeq 6000 S4 flowcell (200 cycles).

Read mapping and filtering. Data was mapped and processed using the 10X Genomics Cell Ranger DNA (v1.1.0) software with default parameters, using the hg19/GRCh37 reference genome. Bam files generated by Cell Ranger DNA was split into individual cells/barcodes using SAMtools (Li et al., 2009) (v1.10) and filtered to remove duplicates and MAPQ < 10 reads with SAMtools, and reads overlapping hg19 blacklist (DAC) regions with BEDtools (Quinlan and Hall, 2010) (v2.26.0). Cells/barcodes with less than 1 M reads were discarded. A summary of sequencing metrics can be found in Table S6.

scRepli-Seq processing. We used a modification of previously described single cell Repli-Seq (scRepli-Seq) methods ((Dileep and Gilbert, 2018; Takahashi et al., 2019) & https://github.com/kuzobuta/scRepliseq-Pipeline) with the following adjustments. Analysis was limited to autosomes.

1) Cells were filtered using median-absolute-deviation (MAD) scores, where G1 cells < 0.3 and S-phase cells > 0.4 and < 0.8 were kept for further analysis. MAD scores were calculated in non-overlapping 200kb bins. Reads were counted using *binReads* command from scCNV R package AneuFinder (Bakker et al., 2016) (v1.14.0). 'mappable_regions.bed' output from Cell Ranger DNA was used to generate a merged unmappable bed file of HCT116 and DKO1 for further read filtering within the *binReads* command. This applies to all further use of the *binReads* command.
2) A control dataset representing baseline copy number variations (CNVs) and mappability for S-phase cell comparison was created by merging high-quality G1 cells. CNVs in G1 cells were identified using AneuFinder *findCNVs* in 500kb bins as described by Takahashi et al. (2019). 'spikiness' and 'bhattacharyya' quality measures was obtained using the *clusterByQuality* command in AneuFinder. Cells were removed if spikiness > = 0.21 and bhattacharyya $\leq$ 1. Cells were also removed if deemed 'noisy' by Cell Ranger DNA. 48/67 HCT116 G1 and 55/84 DKO1 G1 cells passed QC and were merged for further use. To obtain a CNV baseline representative of all cells within the cell population, regions with CNV heterogeneity were removed from further analysis. Heterogeneous regions were defined as those with heterogeneity score > 0.2, calculated from the 500kb CNV data

using the *karyotypeMeasures* command in AneuFinder. *karyotypeMeasures* was modified to output heterogeneity score per bin. Heterogeneous regions and SVs from HCT116 and DKO1 were merged and removed from all further analyses.

3) Single cell data was normalized against the merged G1 control data. G1 and S-phase cell reads were binned in non-overlapping 80kb bins and in 200kb bins at 40kb sliding intervals using AneuFinder *binReads.* Read counts were normalized using the *correctMappability* command from AneuFinder (copied from v1.5.0), using the merged G1 data as reference control.
4) Single cell RT scores were generated from the mappability-corrected 200kb bin-40kb sliding window data as described in Takahashi et al. (2019). Single cell replication timing profiles are similar to whole population Repli-Seq timing profile (Figure S5B). Single cell RT scores were used to generate Pearson's correlation matrix, hierarchical clustering (ward.d2) and tSNE plot as described in Takahashi et al. (2019). At this point, any S-phase cells that cluster with G1 in the hierarchical clustering or in the tSNE plot were removed from further analysis. 322/646 HCT116 S and 208/582 DKO1 S cells were used for further analysis.
5) Mappability-corrected 80kb bin data was binarized using the *findCNVs* command in AneuFinder. We used the following parameters: method = "HMM," max.iter = 3000, states = c("zero-inflation," "0-somy," "1-somy," "2-somy"), eps = 0.01, most.frequent.state = ("1-somy" or "2-somy"). The *findCNVs* command outputs a 2-state HMM model were states '1' and '2′ indicates un-replicated and replicated, respectively. Within the *findCNVs* command, Takahashi et al. (2019) specifies whether the most common state is 1-somy (unreplicated) or 2-somy (replicated) depending on FACS gating to reduce HMM calling ambiguity, i.e., a Early cell would have majority state '1' and a Late cell would have majority state '2′. However, as Early, Mid and Late S-phase cells were pooled here, we could not directly assign whether 1-somy or 2-somy was the most common state. Therefore, we generated two HMM models per cell, specifying either 1-somy or 2-somy as the most common state. For most cells, the two HMM models had little to no difference in binarisation. These tended to be Mid-S cells (Figure S5C). We determined that a cell had 'evenHMM' if the absolute differences in bin numbers between 1-somy state '1' and 2-somy state '1' is less than 1,500 bins, and the same for state '2′. A cell was 'unevenHMM' if the absolute differences were greater than 1,500 bins for both states. The cutoff of 1,500 bins was conservatively chosen to separate the two groups based on the distribution of bin number differences between 1-somy and 2-somy calls of the same state. This distribution was bimodal, with one group centered around 0 ('evenHMM cells) and the other peak centered around 15,000-20,000 ('unevenHMM' cells) (Figure S5D). 'unevenHMM' cells were then assigned as Early or Late through visual comparison of the HMM bed file against the earliest and latest 'evenHMM' cells. For final HMM calls, we used the 2-somy calls for 'evenHMM' cells, the 1-somy calls for 'unevenHMM' cells that were assigned as Early and the 2-somy calls for 'unevenHMM' cells that were assigned as Late. Binarized 80kb data was then used to generate the % replication score per cell, single cell RT value per bin, cell-to-cell variability scores and RT sigmoid modeling (gain and M values) (Dileep and Gilbert, 2018; Takahashi et al., 2019). For calculation of the slope of the sigmoid model (gain), we used the value 100 as the maximum value of the sigmoid (Dileep and Gilbert, 2018). The M-value is the x-intercept at the sigmoid's midpoint and represents when 50% of the cell population has replicated that locus.

Identifying biphasically replicating loci

Loci were called as biphasic if their weighted variance score was $\geq$ 0.081. Calibration tests of the weighted variance score showed that a score of $\sim$0.081 is achieved for a locus with even replication timing across all 6 fractions (Table S5, 'even'), and a score above $\sim$0.081 represents loci where there were high PNDV values in non-adjacent fractions separated by in-between fractions of low PNDV values (Table S5). The cut-off of 0.081 does miss some biphasic regions with smaller separations between high PNDV fractions, hence, the use of the Hansen score described below. Biphasic loci were also identified according to Hansen et al. (2010). Briefly, the 6-fraction PNDV scores were reduced to 5 fractions by pairwise addition of adjacent fractions (G1b+S1, S1+S2, S2+S3, S3+S4, S4+G2M). A 1kb locus was deemed biphasic if more than 40% of the 5-fraction score was in non-adjacent fractions. For example, a locus is biphasic if $\geq$ 40% of the score is in G1+S1 and $\geq$ 40% is in S2+S3, S3+S4 or S4+G2, with S1+S2 < 40%.

Biphasic loci are defined as *maintained* where at least one HCT116 replicate and one DKO1 replicate are biphasic. Biphasic loci are defined as *gained* where no HCT116 replicates are biphasic and at least one DKO1 replicate is biphasic. Biphasic loci are defined as *lost* where at least one HCT116 replicate is biphasic and no DKO1 replicates are biphasic.

Allelic replication timing

Nanopore read alignment, variant calling and phasing. HCT116 and DKO1 Nanopore base called reads (fastq) were aligned to hg19 using minimap2 (Li, 2018) (v.2.17-r943-dirty) with parameters "-ax map-ont." Mapped reads were sorted and indexed with SAMtools (v1.9). Variants were called and phased with medaka_variant (v0.11.4, https://nanoporetech.github.io/medaka) with the options "-t 36 -s r941_prom_high_g330 -m r941_prom_high_g330 -p -b 100."

Variant filtering, haplotype mapping and Repli-Seq processing. Medaka variants were filtered as follows using bcftools (v1.9): i) occurs in both HCT116 and DKO1 datasets; ii) quality score above 20; iii) within each dataset, only phased heterozygote single nucleotide variants (SNVs) were used. hg19 reference genome fasta files were generated for each haplotype per cell line using bcftools consensus (parameters "-H 1pIu" and "-H 2pIu"). Repli-Seq fractions were mapped to each haplotype reference genome using bowtie (Langmead et al., 2009) (v1.1.0, parameters "-v 0 -m 1–tryhard–best–strata–time–trim5 6"). Mapped bam files were filtered for reads that overlapped phased SNVs using SAMtools view (v1.9, option "-L"). Weighted average scores were calculated from filtered bam files as described above with the following modifications: i) Reads were counted in 50kb sliding windows at 1kb intervals,

excluding chrY and chrM. The 50kb sliding windows were modified so that only reads within each haplotype block were counted for each block. Bins overlapping SVs were removed prior to downstream analyses; ii) The low coverage threshold for 50kb windows was set to at least 4 reads per fraction per 50kb loci.

Calling allelic differentially timed regions. Allelically replicating regions were called if there was a WA difference of less than 10 between replicates and more than 30 between alleles. Due to the sparseness of allelically mapped regions, allelic regions were merged if within 1 Mb of each other. The 1 Mb merged region was used to identify overlapping genes.

Imprinted gene and cancer-related gene annotation. A list of human imprinted genes was obtained from https://www.geneimprint.com (Luedi et al., 2007). Genes located in allele-specific replication regions were also checked against the Candidate Cancer Gene Database (Table S1) (http://ccgd-starrlab.oit.umn.edu) (Abbott et al., 2015).

Profile plots

We used SeqPlots (Stempor and Ahringer, 2016) to calculate average scores over regions of interest, then used ggplots (Wickham, 2016) to plot the average scores across all regions for each bin with standard error and confidence intervals.

ChIP-seq processing

ChIP-seq datasets were processed as previously described (Bert et al., 2013; Taberlay et al., 2014). Briefly, ChIP-seq reads were aligned to hg19 using bowtie (Langmead et al., 2009) (v1.1.0) allowing up to 3 mismatches, discarding ambiguous and clonal reads. All histone ChIP-seq peaks were called using PeakRanger (Feng et al., 2011) (v1.16). For the distribution of histone mark occupancy across replication timing, consensus peaks were used where replicates were available. Broad domains of histone mark enrichment, H3K4me3 and H3K9me3, were processed with Enriched Domain Detector (EDD) (Lund et al., 2014) (settings: required_fraction_of_informative_bins = 0.9, p_hat_CI_method = normal). Replicate-shared domain for H3K9me3 were called using the 'intersect' function (R, GenomicRanges v1.22.4 (Lawrence et al., 2013)), before calling regions of maintenance ('intersect' between HCT116 and DKO1), gain and loss ('setdiff' between HCT116 and DKO1). H3K4me3 domains that exist only in DKO1 (replicate merged) were intersected with H3K9me3 domain regions and used for further analyses. The exact SV region was removed from histone domains prior to downstream analyses.

ChromHMM analysis

15-state ChromHMM tracks for HCT116 and DKO1 were called based on the Roadmap Epigenomics 15-state chromHMM model (Kundaje et al., 2015) using the chromHMM program (v1.10) (Ernst and Kellis, 2012). ChIP-seq data was prepared for segmentation by first using 'bamToBed', followed by 'BinarizeBed'. Replicates were pooled at the bamToBed stage. The Roadmap 15-state model parameters were then applied to produce 15-state segmentations for HCT116 and DKO1. Analysis of ChromHMM change-of-state is based on Fiziev et al. (2017). To calculate ChromHMM state change enrichment scores, we divided the number of observed state changes by the number of expected changes as outputted by the chisq.test in R. Two-sided p values were calculated from the chisq.test standard residuals (similar to the z-score) and FDR corrected. To control for reciprocal state changes (i.e., Het to TssA versus TssA to Het in the direction of HCT116 to DKO1), the enrichment scores of Het-TssA was divided by the enrichment score of TssA-Het. A count of 1 was added to both observed and expected to avoid divisions by 0. Only transitions where both scores were significant are shown.

RNA-seq data generation and processing

Total RNA in triplicates (different passages) was extracted from cultured cells using TRIzol (Life Technologies, #15596018). Libraries were constructed with the Illumina TruSeq Stranded mRNA library preparation kit (Illumina, #RS-122-2102) and sequenced on the Illumina HiSeq X Ten. Paired-end reads were processed as previously described (Du et al., 2019) using Trim Galore (v0.4.0, parameter settings:–fastqc–paired–retain_unpaired–length 16) and STAR (Dobin et al., 2013) (v2.5.3a, parameter settings:–quantMode TranscriptomeSAM) for mapping reads to the hg19 human transcriptome build (GENCODE 19 (Harrow et al., 2012)). Mapped reads where counted into genes using rsem (v1.2.21) (Li and Dewey, 2011). TMM normalization was applied using edgeR (v3.12.1) (Robinson et al., 2010). Fold changes (FC) were computed as the log2 ratio of normalized reads per gene using edgeR. Genes with fold change $\pm$ 1.5 and FDR $<$ 0.01 were considered as significantly altered. Promoters were defined as the region from $-$2000 bp to +100 bp around the transcriptional start site. Genes overlapping SV regions were removed. The GSEA desktop application was used to perform gene set enrichment analysis against the MolSigDB v7.4 gene sets (Liberzon et al., 2011; Subramanian et al., 2005). TPM values per replicate per gene were used to generate the Gene Cluster Text file and analysis was performed using the following parameters: gene set permutation, weighted enrichment statistic and Signal2Noise ranking. Significant pathways (FDR $<$ 0.05) relevant to the study are shown.

Genomic annotation

Promoters were defined as $-$2000 bp to +100 bp around the transcription start site from the GENCODE 19 reference transcriptome (Harrow et al., 2012). Exons and introns were called per transcript using the GenomicFeatures (v1.22.13) package in R

(Lawrence et al., 2013), and merged if overlapping. Intron regions were retained if they did not intersect an exonic region. 5′ and 3′ UTRs were also called using the GenomicFeatures package, and merged if overlapping. Intergenic regions were defined as the gaps between the other elements.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical details can be found in the figure legends and relevant methods sections, including statistical test, significance cut-offs, multiple testing correction, number of replicates and precision measures (e.g., error bars). Statistical tests were performed using R (v3.2.3; R Development Core Team, 2015). Generally, for genomic interval overlaps and genomic rearrangement overlaps, a modified LOLA (Sheffield and Bock, 2016) package was used to perform a two-sided log-odds ratio test which reports significance using "BH" FDR values. The Mann-Whitney-Wilcoxon test was used for 2-group non-parametric comparisons, and the one-tailed test was used where a directional difference between the groups was of interest. Unless otherwise stated, statistical tests were two-sided. Permutation tests were used to calculate one-sided *p-value*s for the difference in means of scRepli-Seq scores between HCT116 and DKO1. n = 10,000 permutations were performed.