Synchronized replication of genes encoding the same protein complex in fast-proliferating cells

Ying Chen,[1,2,3,6] Ke Li,[1,2,6] Xiao Chu,[1,2,3] Lucas B. Carey,[4,5] and Wenfeng Qian[1,2,3]

[1]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China; [2]Key Laboratory of Genetic Network Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China; [3]University of Chinese Academy of Sciences, Beijing 100049, China; [4]Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona 08003, Spain; [5]Center for Quantitative Biology and Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

DNA replication perturbs the dosage balance among genes; at mid-S phase, early-replicating genes have doubled their copies while late-replicating ones have not. Dosage imbalance among genes, especially within members of a protein complex, is toxic to cells. However, the molecular mechanisms that cells use to deal with such imbalance remain not fully understood. Here, we validate at the genomic scale that the dosage between early- and late-replicating genes is imbalanced in HeLa cells. We propose the synchronized replication hypothesis that genes sensitive to stoichiometric relationships will be replicated simultaneously to maintain stoichiometry. In support of this hypothesis, we observe that genes encoding the same protein complex have similar replication timing but mainly in fast-proliferating cells such as embryonic stem cells and cancer cells. We find that the synchronized replication observed in cancer cells, but not in slow-proliferating differentiated cells, is due to convergent evolution during tumorigenesis that restores synchronized replication timing within protein complexes. Taken together, our study reveals that the demand for dosage balance during S phase plays an important role in the optimization of the replication-timing program; this selection is relaxed during differentiation as the cell cycle prolongs and is restored during tumorigenesis as the cell cycle shortens.

[Supplemental material is available for this article.]

The balance hypothesis asserts that the stoichiometric relationship among subunits of a protein complex is essential for the survival and proliferation of cells; the disruption of this relationship perturbs functions of protein complexes and sometimes even causes cytotoxicity (Papp et al. 2003; Veitia 2005; Birchler and Veitia 2010, 2012). The balance hypothesis provides a unique framework for understanding a variety of biological phenomena, especially the proliferation rate of aneuploid cells, the fate of duplicated genes, and the X-Chromosome inactivation. Aneuploidy, defined as a karyotype that is not a multiple of the haploid complement, generates dosage imbalance among genes on different chromosomes (Birchler et al. 2001; Chen et al. 2019). Consistent with the balance hypothesis, aneuploidy often results in a more severe growth defect than polyploidy, which keeps the dosage balance among genes (Birchler and Newton 1981; Otto and Whitton 2000; Birchler and Veitia 2010). Furthermore, the addition of a larger chromosome, which leads to a dosage imbalance among more genes, often results in a greater reduction in fitness (Torres et al. 2007, 2010; Pavelka et al. 2010; Birchler and Veitia 2012; Stingele et al. 2012). Individual gene duplication confers the second type of dosage imbalance, between duplicate genes and singletons. Consistent with the balance hypothesis, genes often reduce their expression soon after duplication (Qian et al. 2010), through which the dosage balance is restored. Furthermore, genes encoding protein complexes exhibit a higher retention rate after the whole genome duplication so that the dosage balance among sub-

units is maintained (Papp et al. 2003; Qian and Zhang 2008; Freeling 2009; Tasdighian et al. 2017). In addition, dosage balance in a protein complex also explains the evolution of gene size (Chen et al. 2009) as well as mammalian X-Chromosome inactivation (Lin et al. 2012).

A probably more prevalent but less studied source of dosage imbalance is caused by DNA replication that occurs in each cell cycle. During the DNA synthesis phase (S phase) of a cell cycle, the genome is replicated in a defined temporal order known as the replication-timing program (Pope and Gilbert 2013; Sima et al. 2019). In the middle of S phase, early-replicating genes have doubled their copy number, but late-replicating genes have not, leading to a dosage imbalance between early- and late-replicating genes. Such dosage imbalance likely causes a growth defect, especially among genes sensitive to dosage relationships, such as those encoding the same protein complex (Papp et al. 2003; Birchler and Veitia 2012). Although acetylated histones (H3K56ac) can incorporate into newly replicated DNA regions and partly suppress the expression of newly replicated genes in yeast (Voichek et al. 2016), this compensatory mechanism cannot completely restore the dosage balance; the mRNA levels of early-replicating genes still exhibited an ~20% increase compared to late-replicating genes during mid-S phase (Voichek et al. 2016). Consistently, higher expression was observed when a GFP reporter was inserted into the early-replicating regions in budding yeast (Chen and Zhang 2016).

The dosage imbalance during S phase could be more severe in mammalian cells, where H3K56ac may not mark newly replicated

29:1929–1938 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/19; www.genome.org
Genome Research **1929**
www.genome.org

DNA (Stejskal et al. 2015). Consistently, in mouse embryonic stem cells (ESCs), the transcription rates of *Pou5f1* and *Nanog* increased by 28% and 50%, respectively, upon DNA replication (Skinner et al. 2016). A similar phenomenon was also observed in *UBC* in human primary fibroblasts (Padovan-Merhar et al. 2015). These data show that replication can cause dosage imbalance during S phase and suggest that additional mechanisms should exist in mammalian cells to solve the problem.

In this study, we aim to explore the specific mechanisms by which mammalian cells maintain dosage balance within a protein complex and to investigate if such mechanism plays a role in tumorigenesis as the fraction of time cells spend in S phase increases.

Results

The single-cell transcriptomes of HeLa cells through the cell cycle

Increased gene expression after DNA replication has been observed in individual genes in mammalian cells using single-molecule fluorescence in situ hybridization. To determine if this phenomenon is generalizable to other genes in the genome, we characterized the transcriptomes of HeLa, a cell line derived from cervical cancer cells, through the cell cycle. Taking advantage of single-cell RNA sequencing (scRNA-seq) technology, we obtained the transcriptomes of 884 individual cells. These cells are neither synchronized with drugs nor sorted by flow cytometers and therefore are more likely to represent the native cell-cycle status (Cooper 2003). We performed principal component analysis (PCA) using the expression levels of 151 previously identified cell-cycle phase-specific genes (Supplemental Table S1; Whitfield et al. 2002). We observed that cells formed a circle in the scatterplot of the first two principal components (Fig. 1A). To determine if this circle is related to the cell cycle, we assigned each cell to one of the five cell-cycle phases (G_1/S, S, G_2, G_2/M, or M/G_1) by scoring for the expression specificity of the phase-specific genes (Fig. 1B; Supplemental Table S1; Macosko et al. 2015). Cells assigned to the same phase are clustered in the PCA plot, and five clusters are in the order of the known cell-cycle progress in a clockwise direction (Fig. 1A), underscoring the competence of the single-cell transcriptome analysis in recognizing the cell-cycle statuses of individual cells.

The overexpression of early-replicating genes relative to late-replicating ones is predicted to take place during mid-S phase when the replication of early-replicating genes has finished and that of late-replicating genes has not yet started. To test this prediction, we further divided S-phase HeLa cells into six subgroups based on their order in the clockwise direction in the PCA plot (Fig. 1C). Two lines of evidence suggest that subgroups 1–6 reflect genuine S-phase progression. First, genes known to be specifically expressed in a certain stage of S phase were precisely expressed in the cells at this stage. For example, CCNE1, a protein required for the G_1/S transition (Koff et al. 1992), was more highly expressed in the cells of subgroup 1 (Fig. 1D). In contrast, CENPA, a protein that plays a role in centromere assembly (Zeitlin et al. 2001), was more highly expressed in the cells of subgroups 4–6 (Fig. 1D). Second, we inferred the mRNA abundance in each cell from the total number of unique molecular identifiers (UMI) of all genes in the scRNA-seq data. mRNA abundance in each cell increased from subgroup 1 to 6 (Supplemental Fig. S1), consistent with previous findings of the increase of cell volume and mRNA abundance along the progress of the cell cycle (Boye and Nordström 2003; Mitchison 2003; Padovan-Merhar et al. 2015). Both observations indicate that we successfully obtained the transcriptomes at different S-phase stages.

Early-replicating genes are overexpressed during S phase in HeLa cells

To determine if early-replicating genes are overexpressed during mid-S phase (Fig. 2A), we compared the expression of early- versus late-replicating genes during S phase. We retrieved DNA replication-timing profiles of HeLa cells (Weddington et al. 2008; Oda et al. 2012) and defined 2000 (~10%) genes with the highest or lowest replication timing values as early- or late-replicating genes, respectively (Fig. 2B). We generated all possible gene pairs in which one is from the 2000 early-replicating genes and the other is from the 2000 late-replicating ones. We estimated the expression ratio of early-/late-replicating genes for each gene pair and calculated the mean of all gene pairs for each S-phase
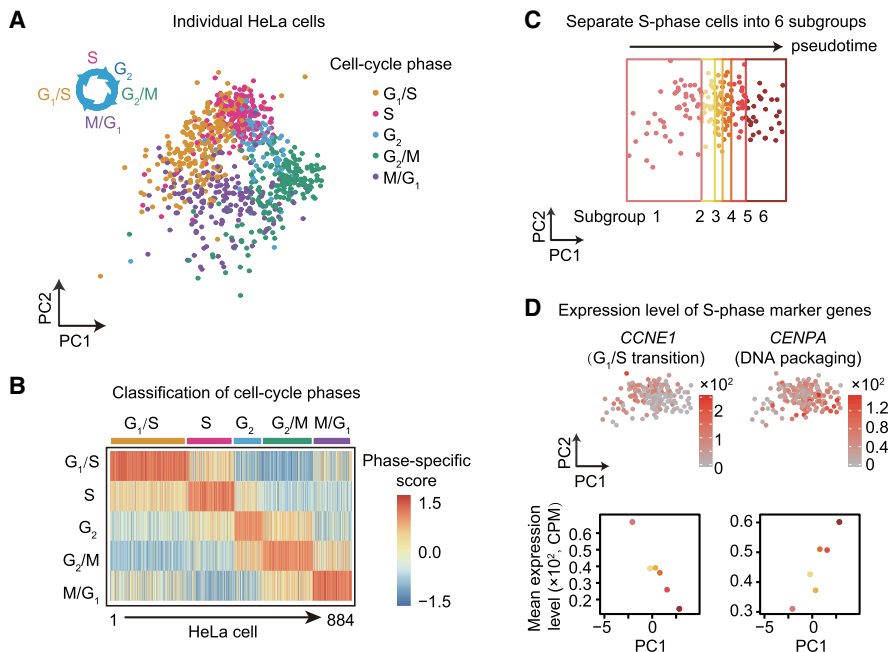

Figure 1. Cell-cycle analysis of the HeLa scRNA-seq data. (*A*) The PCA plot of 884 individual HeLa cells shows asynchrony in the cell cycle. (*B*) The cell-cycle status of individual HeLa cells. Each cell is classified into one of the five cell-cycle phases based on the phase-specific scores; 259, 172, 86, 201, and 166 cells were classified into G_1/S, S, G_2, G_2/M, and M/G_1 phases, respectively. The phase-specific scores were standardized among five phases for each cell. (*C*) S-phase cells were separated into six subgroups based on the PC1 values in the PCA plot. Subgroups 1–6 represent the six stages of S phase from early to late. (*D*) The expression level of S-phase marker genes among cells. CCNE1 (*left*) and CENPA (*right*) are known to be highly expressed in the early- and late-S phases, respectively. The average expression level among cells is shown for each S-phase subgroup (*bottom*).
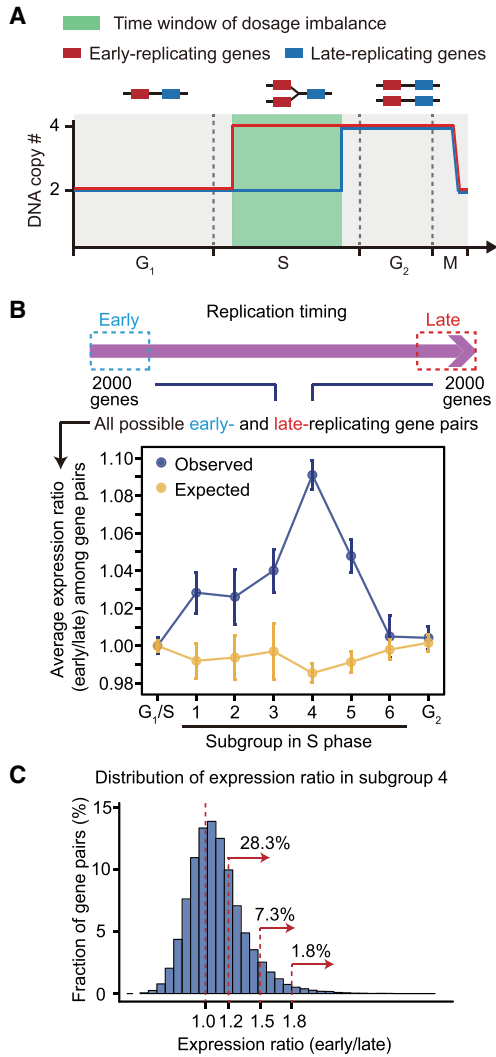
Figure 2. Early-replicating genes are overexpressed during S phase. (*A*) The dosage between early- and late-replicating genes is predicted to be imbalanced during mid-S phase because early-replicating genes have finished replication and late-replicating ones have not yet started. (*B*) Early-replicating genes are overexpressed during mid-S phase. The observed (blue) average expression ratio of all gene pairs in each S-phase subgroup was normalized by that in G_1/S. The random expectation (orange) was obtained from 2000 "pseudo" early- or "pseudo" late-replicating genes that were randomly sampled from the genome. Error bars represent the standard errors among gene pairs in each subgroup. (*C*) The distribution of the expression ratio of early- to late-replicating gene in each gene pair in subgroup 4.

subgroup. The overexpression of early-replicating genes was observed in almost all S-phase subgroups and peaked in subgroup 4 (by ∼10%, $P = 8 \times 10^{-8}$, bootstrap hypothesis testing) (Fig. 2B, blue dots). As a control, we randomly assigned 2000 genes from the genome as "pseudo" early- or late-replicating genes; the overexpression was no longer observed (Fig. 2B, orange dots). Note that this observation cannot be due to the effect of a small number of genes exempt from dosage compensation (Müller and Nieduszynski 2017), as the same conclusion can be reached when we randomly dropped 10% of the 2000 early- and 2000 late-replicating genes (Supplemental Fig. S2A). These results revealed that, although the dosage imbalance was partly compensated during S

phase, an average ∼10% overexpression of early-replicating genes remained.

Such imbalance was stronger when we considered individual gene pairs. We observed that 28.3%, 7.3%, and 1.8% of gene pairs with the relative expression of early-replicating genes increased by at least 20%, 50%, and 80% in subgroup 4, respectively (Fig. 2C). Genes replicating at different times show variation in the relative expression level during S phase, and vice versa. A pair of genes with a large variance of expression ratio through S phase tended to replicate at different times (Supplemental Fig. S2B). Collectively, the dosage balance between early- and late-replicating genes is disrupted during S phase.

Genes encoding subunits of the same protein complex tend to replicate simultaneously in HeLa cells

When dosage imbalance occurs among genes sensitive to the stoichiometric relationship, such as those encoding protein complexes, fitness is likely reduced (Papp et al. 2003; Birchler and Veitia 2012; Oromendia and Amon 2014). Nevertheless, the reduction in fitness can be avoided if genes encoding the same protein complex are replicated simultaneously. Following this logic, we proposed the synchronized replication hypothesis that the replication of genes encoding the same protein complex is synchronized.

Genes encoding some protein complexes are replicated almost simultaneously in HeLa cells, as exemplified in Figure 3A. To test the synchronized replication hypothesis at the genomic scale, we retrieved the components of 1521 protein complexes from the Human Protein Reference Database (Peri et al. 2003). For each protein complex, we calculated the standard deviation of replication timing of all genes encoding this protein complex (Fig. 3B). As a control, we shuffled among genes encoding protein complexes to constitute "pseudo" protein complexes, keeping the number of complexes and the number of subunits in each complex unchanged (Fig. 3B). We performed the shuffling 1000 times. The median of observed standard deviations was significantly smaller than the random expectation ($P < 0.001$, permutation test) (Fig. 3B), indicating synchronized replication within protein complexes.

The synchronized replication hypothesis makes two more predictions. First, genes encoding the same protein complex should exhibit better mRNA dosage balance through S phase than the random expectation; this was indeed observed (Supplemental Fig. S3A). Second, the S-phase–expressed protein complexes should exhibit better synchronized replication than the unexpressed ones. Consistently, the standard deviations of replication timing of the expressed protein complexes were significantly smaller than those of the unexpressed ones ($P = 0.0036$, Mann–Whitney U test) (Fig. 3C), indicating that synchronized replication within a protein complex is driven by the expression of this complex in S phase.

There are several confounding factors that should be controlled for. First, genes encoding the same protein complex tend to form clusters on chromosomes (Albig and Doenecke 1997; Hurst et al. 2004), which are likely to simultaneously replicate because they have similar physical distances to the closest replication origin. We discarded protein complexes of which at least two subunits are encoded by the genes on the same chromosome; synchronized replication remained observed ($P = 0.002$, permutation test) (Supplemental Fig. S3B). Second, to test if the size of protein complexes affects the level of synchronized replication, we classified the protein complexes into three categories according to its subunit number (<4, =4, or >4) and tested if replication is
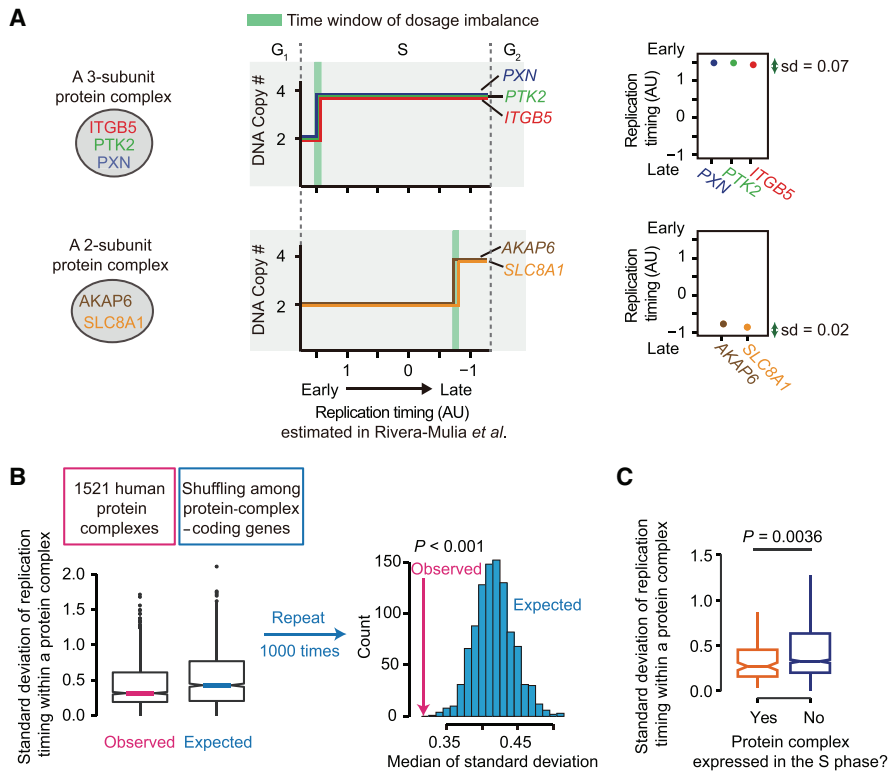
Figure 3. Genes encoding the same protein complex are replicated simultaneously in HeLa cells. (*A*) Two protein complexes exemplify the synchronized replication of genes encoding the same protein complex. The standard deviation (sd) of replication timing within a protein complex is shown on the *right*. (*B*) The observed median standard deviation of replication timing within a protein complex is significantly smaller than the random expectation in which the protein-complex–coding genes were shuffled. (*C*) The comparison of the similarity in replication timing between the S-phase–expressed and –unexpressed protein complexes. The expressed protein complexes are defined as the complexes with at least two-thirds of subunits expressed; the expressed subunits are defined as the subunits expressed in at least half of the S-phase cells (CPM > 0). Outliers are omitted in the box plot. *P*-value was given by the Mann–Whitney *U* test.

synchronized by shuffling genes within each category. Synchronized replication remained observed in all three categories (Supplemental Fig. S3C). Third, we tested the synchronized replication hypothesis in 14 most validated protein complexes that were identified by multiple methods, including the transcription factor II D (TF$_{II}$D) complex, exon junction complex, and cyclin-dependent kinase 8 (CDK8) subcomplex; synchronized replication remained observed (*P* = 0.007, permutation test) (Supplemental Fig. S3D).

Synchronized replication mainly occurs in fast-proliferating cells

The replication-timing program varies among cell types. To determine if synchronized replication occurs uniformly among various human cells, we retrieved the replication-timing programs in 19 cell lines/types (Supplemental Table S2) from the ReplicationDomain database (Weddington et al. 2008) or NCBI Gene Expression Omnibus (GEO) (Edgar et al. 2002). They include six human ESC (hESC) lines, five cancer cell lines, and eight differentiated cell types derived from hESCs, such as liver, pancreas, smooth muscle, and mesothelial cells (Fig. 4A); the proliferation of ESCs and cancer cells is fast whereas that of differentiated cells is slow. These cell lines/types exhibit various levels of synchronized replication within protein complexes (as exemplified in Fig. 4B and Supplemental Fig. S4A). We further used the *P*-value of the permutation test, as in

Figure 3B, to infer the level of synchronized replication for each cell line/type (Supplemental Fig. S4B, with three examples shown in Fig. 4C) and labeled synchronized replication for those with *P* < 0.05 (Fig. 4D). Synchronized replication was mainly observed in fast-proliferating cells (*P* = 0.001, Fisher's exact test) (Fig. 4D).

The absence of synchronized replication in six differentiated cells may be due to the fact that the protein-complex–coding genes are no longer expressed or that the dosage balance among them is no longer important. To test these possibilities, we performed two additional analyses. First, we retrieved the mRNA levels of protein-complex–coding genes (Rivera-Mulia et al. 2015) and found them highly correlated between ESCs and differentiated cells (Supplemental Fig. S5A). Second, we tested the mRNA dosage balance, as for HeLa cells in Supplemental Figure S3A, for all cell lines/types; the mRNA dosage balance remained in differentiated cells (Supplemental Fig. S5B). These results showed that neither explanation worked.

We then speculated that the loss of synchronized replication in differentiated cells was caused by the reduced power of natural selection for synchronized replication in slow-proliferating cells that spend a greater fraction of time in G$_0$ phase and a smaller fraction of time in S phase (S%) (Fig. 5A). To test it, we estimated the proliferation rate of these cell lines/types from proliferation marker genes. The expression of the gene encoding proliferating cell nuclear antigen (*PCNA*), which promotes DNA replication in actively proliferating cells, has been widely used for estimating cell proliferation rate (Kubben et al. 1994; Moldovan et al. 2007; Bologna-Molina et al. 2013). We calculated the proliferation rate for each cell line/type (Fig. 4E) as the average expression level of *PCNA* and 29 more genes whose expression is positively correlated with it (meta-PCNA) (Supplemental Table S3). As expected, the proliferation rate was positively correlated with the proportion of cells in S phase among the eight cell lines/types for which flow-cytometry data were available (*r* = 0.8, *P* = 0.03, Pearson's correlation) (Supplemental Fig. S6). The proliferation rate was positively correlated with the level of synchronized replication among 19 cell lines/types (*r* = 0.7, *P* = 4 × 10^{-4}, Pearson's correlation) (Fig. 5B), and the positive correlation remained after the normalization of the replication-timing data (*r* = 0.8, *P* = 2.3 × 10^{-4}) (Supplemental Fig. S7), suggesting an S%-dependent optimization of the replication-timing program in human cells.

Genes causing changes in synchronized replication reside in long interspersed nuclear element (LINE)-rich regions

How does synchronized replication get lost in differentiated cells? To answer this question, we identified 165 protein complexes (as
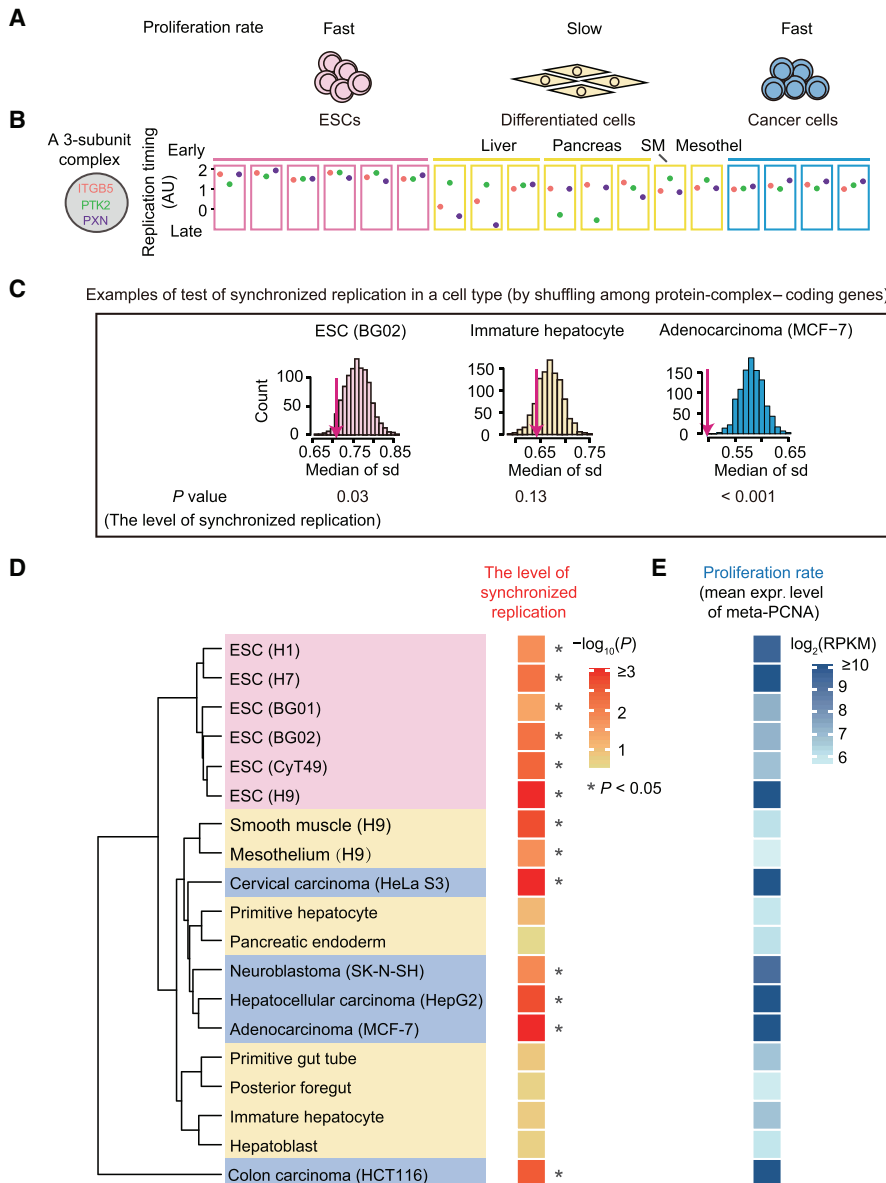
Figure 4. Synchronized replication of genes in the same protein complex occurs mainly in fast-proliferating cells. (*A*) Nineteen cell lines/types across ESCs, differentiated cells, and cancer cells are used to detect the synchronized replication. Differentiated cells include liver, pancreas, smooth muscle (SM), and mesothelial cells (mesothel) derived from hESCs. (*B*) Replication timing of three genes encoding a three-subunit protein complex in each cell line/type. HCT116 (shown in Supplemental Fig. S4A) is not shown here because it deviates from others in replication timing. (*C*) Three examples of the tests for synchronized replication. The observed median standard deviation (sd) of replication timing of genes encoding the same protein complex (magenta) is showed on the distribution of the random expectations (blue). (*D*) Synchronized replication occurs in 11 fast-proliferating cell lines and two slow-proliferating cell types. The dendrogram (*left*) shows the clustering of 19 cell lines/types based on the replication-timing profile of all protein-coding genes and was generated with Ward's method in the hclust function of *R* (R Core Team 2018). The heat map on the *right* shows the level of synchronized replication estimated from the permutation test. The asterisk indicates that replication is synchronized (*P* < 0.05) in the corresponding cell line/type. (*E*) The proliferation rate was estimated from the average expression level of 30 meta-PCNA genes.

exemplified in Fig. 6A) in which the similarity of replication timing within a protein complex was significantly reduced during differentiation (*P* < 0.05 in the *t*-tests for the standard deviations of replication timing in a protein complex, between ESCs and differentiated cells). Among the 491 genes encoding these protein complexes, 92% exhibited a delay in replication in differentiated cells

(top in Fig. 6B). In contrast, only 73% of noncomplex encoding genes exhibited delays in replication timing ($P < 2.2 \times 10^{-16}$, Fisher's exact test), suggesting that the differentiated cells lose synchronized replication preferentially through a replication delay.

Among the 165 protein complexes in which synchronized replication is lost in differentiated cells, 79 protein complexes restored synchronized replication in cancer cells (*P* < 0.05 in the *t*-tests for the standard deviations of replication timing in a protein complex, between differentiated and cancer cells). Presumably, the restoration could occur either through (1) reversing the change in replication timing during cell differentiation, or (2) through an inter-genic suppression such that the change in replication timing of a second gene in the same protein complex follows that of the first. We found that during tumorigenesis, 82% of genes encoding these 79 protein complexes reversed the changes during cell differentiation (bottom in Fig. 6B), significantly higher than the fraction (72%) among genes not encoding protein complexes ($P = 7 \times 10^{-4}$, Fisher's exact test).

That the changes in replication timing during cell differentiation were often reversed during tumorigenesis suggested that genes affecting synchronized replication reside in the transition or variable regions of the genome. It has been reported that replication-timing transitions during cell differentiation often occur in LINE-rich regions (Hiratani et al. 2004, 2008). The 79 complexes with synchronized replication restored in cancer cells were also encoded by genes located in the genomic region with higher LINE density ($P = 2.3 \times 10^{-5}$, Mann–Whitney *U* test) (Fig. 6C), suggesting that the change in synchronized replication during cell differentiation or tumorigenesis is caused by the variable firing times of replication origins in the transition regions of the genome.

Discussion

Abnormal replication-timing programs have been known to be related to disease and cancer (Smith et al. 2001; Ryba et al. 2012). Our study provides a new mechanism that explains why a proper regulation of the replication-timing program is essential, especially for fast-proliferating cells: to maintain the dosage balance among stoichiometry-sensitive genes during S phase. Synchronized replication was observed in ESCs and cancer cells but not in most differentiated cells.
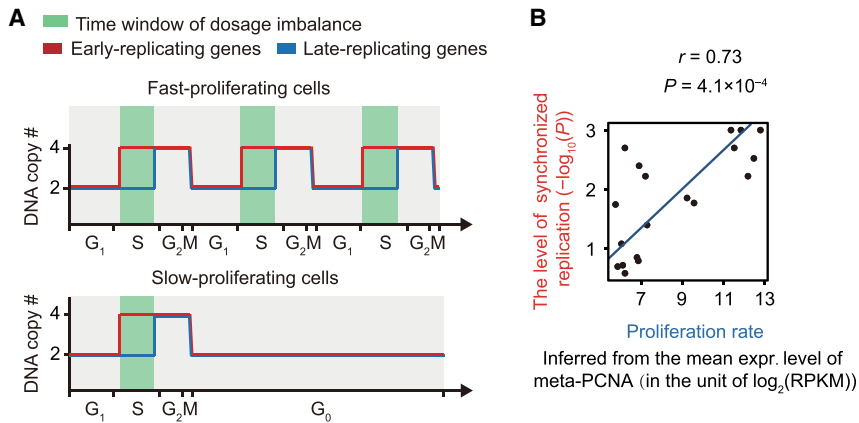
Figure 5. Synchronized replication of genes in the same protein complex is S%-dependent. (A) Dosage imbalance between early- and late-replicating genes during S phase in fast- (*top*) and slow-proliferating (*bottom*) cells. (B) The proliferation rate and the level of synchronized replication are positively correlated among 19 cell lines/types.

Since cancer cells are "evolved" from differentiated cells rather than ESCs, regaining synchronized replication in cancer cells implies convergent evolution of the replication-timing program during tumorigenesis. These observations echo previous analyses on the evolution of tumor cells at different levels, such as those at the transcriptome or the amino-acid usage level (Chen et al. 2015; Chen and He 2016; Zhang et al. 2018).

We showed that the demand for dosage balance during S phase could cause synchronized replication of genes encoding the same protein complex. Presumably, such synchronized replication could also have evolved under other selection pressures. For example, it may evolve to meet the demand for similar expression levels of genes encoding the same protein complex because replication timing is associated with gene expression level (Schübeler et al. 2002; Woodfine et al. 2004). Nevertheless, this mechanism cannot explain why synchronized replication is lost in differentiated cells, where the genes encoding protein complexes remain expressed (Supplemental Fig. S5A) and the dosage balance among subunits remains important (Supplemental Fig. S5B). The selection for dosage balance during S phase uniquely predicts the loss of synchronized replication in slow-proliferating cells and the regain of it in cancer cells.

Whereas the dosage imbalance during S phase can be partly relieved by the H3K56ac-associated transcription repression of newly replicated genes in yeast, the balance is not completely restored; early-replicating genes still exhibit an ~20% higher expression level during mid-S phase (Voichek et al. 2016). In mammalian cells, the challenge of the dosage imbalance during S phase is likely greater because S phase lasts for a much longer time. For example, a HeLa cell divides every 24 h and its S phase lasts for ~8 h (Volpe and Eremendo-Volpe 1970); HeLa cells need to suffer from the imbalance between early- and late-replicating genes for a few hours every 24 h. In contrast, yeast has adapted to the life cycle of 24–48 h per generation in the fermentation industry (Gallone et al. 2016) or in nature, which can be mimicked by a synthetic oak exudate medium (Murphy et al. 2006; Qian et al. 2012). However, S phase lasts for <1 h even in a poor carbon source (Leitao and Kellogg 2017), likely because the total time for DNA replication is mainly determined by the elongation rate of the DNA polymerase. Therefore, the imbalance between early- and late-replicating genes lasts for only dozens of minutes every 24–48 h in yeast.

Collectively, the natural selection for synchronized replication is likely stronger in mammalian cells.

The components of protein complexes may be heterogeneous among cell types (Long et al. 2017), which in principle can drive changes in replication-timing programs during development. For example, protein A may form a complex with protein B in one cell type but form another complex with protein C in a second cell type. To maintain the dosage balance in multiple cell types, the replication times of genes A and B should be more similar in the former cell type, while those of genes A and C should be more similar in the latter. It would be of great interest to test this possibility when the genome-wide heterogeneity data of protein complexes become available.

Genes encoding the same protein complex sometimes form clusters on chromosomes (Albig and Doenecke 1997), which is often explained by the demand for coordinated gene expression (Hurst et al. 2004), reduction in expression noise (Batada and Hurst 2007), the positive epistatic relationship among genes (Yang et al. 2017), and the cofluctuation of stochastic gene expression (Sun and Zhang 2019; Xu et al. 2019). We showed that the synchronized replication hypothesis remained supported after controlling for such gene clusters (Supplemental Fig. S3B); yet the gene cluster itself could, in turn, be an evolutionary outcome of the selection for the dosage balance during S phase. Replication origins fire stochastically at the single-cell level (Bechhoefer and Rhind 2012), and therefore, the synchronization of replication timing is not robust in individual cells when these genes are interspersed in the genome and use different replication origins. Forming a cluster on chromosomes is a more robust strategy for maintaining dosage balance among genes.

Our results also have implications for evolution at the nucleotide level. For example, since replication timing is a major determinant of mutation rate (Stamatoyannopoulos et al. 2009; Woo and Li 2012), we predict that genes encoding the same protein complex will have similar mutation rates due to their synchronized replication. Consistently, similar evolutionary rates have been reported between a pair of genes that share a biological function or are coexpressed (Clark et al. 2012). Collectively, our study not only identifies the driving forces underlying the evolution of the replication-timing program but also provides new insights into the evolution of DNA sequences.

Methods

Cell culture and the preparation of the single-cell suspension

HeLa cells were cultured in Dulbecco's Modified Eagle Medium containing 10% fetal calf serum and 2 mM L-glutamine at 37°C in a 5% CO_2 incubator. HeLa cells were grown to 70% confluence and were digested by 0.25% trypsin-EDTA solution. The digestion was stopped by adding culture medium. Cells were gently mixed and were washed with 1 × PBS with 0.04% bovine serum albumin
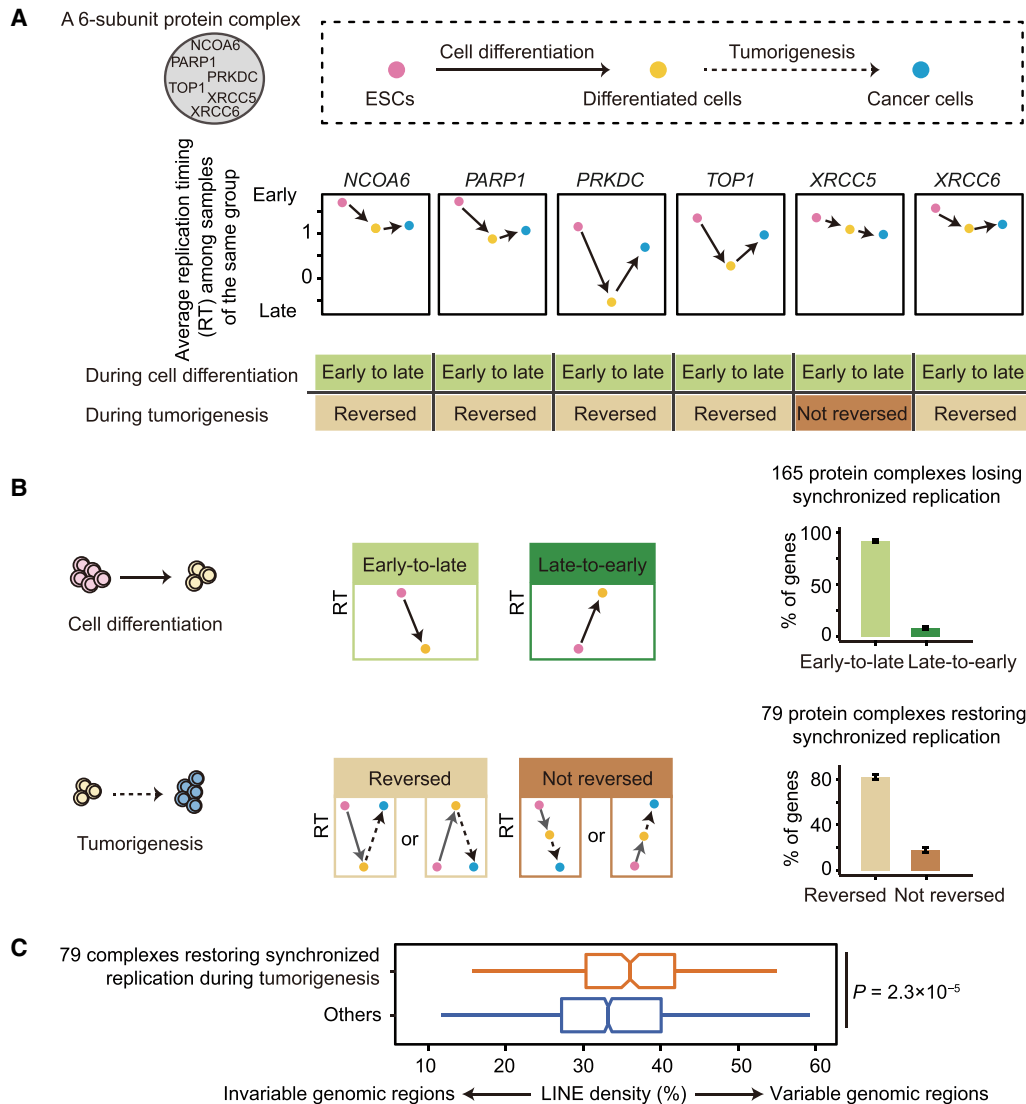
Figure 6. Genes causing changes in synchronized replication reside in the genomic regions with higher LINE density. (*A*) The average replication timing (RT) among cells in the same group: ESCs, six differentiated cells that do not exhibit synchronized replication (liver and pancreas cells), or cancer cells. A six-subunit protein complex as an example shows the change in replication timing among three groups. (*B*) Protein complexes losing synchronized replication during cell differentiation or restoring synchronized replication during tumorigenesis were identified. Genes encoding these protein complexes were classified into categories according to the change in replication timing during these two processes. The fraction of genes in each category is shown. Error bars represent the standard deviation estimated from bootstrapping. (*C*) The comparison of the LINE densities between genes encoding 79 protein complexes that restore synchronized replication during tumorigenesis (orange) and genes encoding other complexes (blue). Outliers are omitted in the box plot.

twice. Finally, cells were filtered with a cell strainer (BD, #352340) and were assessed for viability with Trypan blue. The final cell viability was >95%.

Preparation of single-cell sequencing libraries

We performed scRNA-seq to analyze transcriptomes of "natural" cells through the cell cycle, to avoid the stresses imposed on cells by conventional experimental strategies such as cell-cycle synchronization or fluorescence-activated cell sorting (FACS); agents that prevent DNA synthesis or inhibit the formation of mitotic spindles not only arrest the cell cycle at certain points but also kill important fractions of the cells due to their toxic effect (Cooper 2003).

Single-cell suspensions were loaded onto 10x Genomics Single Cell 3′ Chips (Single Cell A Chip kit, #120236) and were cap-

tured along with the barcoded gel beads. Single-cell libraries were constructed using Chromium Single Cell 3′ Library & Gel Bead kit v2 (10x Genomics, #120237) and Chromium i7 Multiplex kit (10x Genomics, #120262), following the manufacturer's protocol (#CG00052). The libraries were sequenced with the Illumina HiSeq 2500 System.

Initial processing of the sequencing data

Cell Ranger (v2.0.0) was used to process initial sequencing data. The FASTQ files were aligned to the reference genome (GRCh38-1.2.0). UMIs were counted for each bead barcode, and the number of cells was estimated from the barcode count distribution. The transcriptome data of a total of 884 cells were obtained in this study, and the average sequencing depth of them was 378,473 reads per cell.

Cell-cycle analyses for HeLa scRNA-seq data

The list of genes that are specifically expressed in each of the five phases of the cell cycle (G_1/S, S, G_2/M, M, and M/G_1) was retrieved from a previous study (Whitfield et al. 2002). The cell-cycle status of each cell was inferred following Macosko et al. (2015). Briefly, we defined a total of 151 genes as phase-specific genes (Supplemental Table S1), the expression levels (in the unit of counts per million, CPM) of which were used for PCA. The cell-cycle status of each cell was inferred from the five phase-specific scores calculated as follows. The expression level ($\log_2[\text{CPM}+1]$) of each of the 151 phase-specific gene was standardized (z-score) among the 884 single cells. For each cell-cycle phase, the phase-specific score of a cell was calculated as the average z-score of all marker genes of this phase. Each cell was assigned to the cell-cycle phase with the greatest score.

Calculating the expression ratio between early- and late-replicating genes

Early- or late-replicating genes were defined as the 2000 genes with highest or lowest replication-timing values, respectively. Genes specifically expressed (Supplemental Table S1) and unexpressed in S phase were excluded. The average expression level of each gene among all cells in an S-phase subgroup was normalized by dividing its average expression level among all cells in G_1/S phase. For each gene pair, in which one gene from the 2000 early-replicating genes and the other from the 2000 late-replicating ones, the expression ratio of the early-replicating gene versus the late-replicating one was calculated in each S-phase subgroup.

Data sources

The information of 1521 protein complexes in humans was downloaded from the Human Protein Reference Database release 9 (www.hprd.org) (Peri et al. 2003). Among them, 1317 were annotated with complete information and were used in this study. The information of protein complexes was also retrieved from the Database of Interacting Proteins (DIP, dip.doe-mbi.ucla.edu), which documents experimentally determined protein complexes with the methods of purification (Xenarios et al. 2000).

The replication-timing profiles used in this study were downloaded from the ReplicationDomain database (www.replicationdomain.org) (Weddington et al. 2008) or NCBI GEO (www.ncbi.nlm.nih.gov/geo/) (Edgar et al. 2002). The accession numbers are listed in Supplemental Table S2 (Thurman et al. 2007; Weddington et al. 2008; Hansen et al. 2010; Oda et al. 2012; Rivera-Mulia et al. 2015). Specifically, all available replication-timing profiles of ESCs and cancer cell lines were retrieved, together with eight hESC-derived differentiated cell types generated in Rivera-Mulia et al. (2015).

Gene expression data used in this study were downloaded from NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/), and the accession numbers are listed in Supplemental Table S2.

Estimation of the replication timing for genes

The chromosomal locations of 19,805 protein-coding genes were retrieved from the human GRCh38.p12 annotation file, which was downloaded from Ensembl release 87 (www.ensembl.org) (Zerbino et al. 2018). The average intensity ratio of probes (when measuring replication timing) that overlap with a gene was defined as its replication timing.

Estimation of the proliferation rate with gene expression profiles

PCNA is a component of DNA polymerase delta, and its expression level is a reporter of DNA synthesis. We defined a panel of 30 meta-PCNA genes whose expression is positively correlated with *PCNA*. Specifically, we normalized the average expression level of genes ($\log_2[\text{RPKM}+1]$) in 19 cell lines/types and calculated the Pearson's correlation between *PCNA* and each of 131 previously identified candidate meta-PCNA genes (Venet et al. 2011). Genes with a correlation coefficient >0.8 were defined as the meta-PCNA genes of this study. The proliferation rate was inferred from the average expression level of the 30 meta-PCNA genes, which are listed in Supplemental Table S3.

Estimation of the fraction of cells in S phase with flow cytometry

Flow-cytometry data of propidium iodide-stained cells in eight cell lines/types (listed in Supplemental Table S2) were generated in a previous study (Rivera-Mulia et al. 2015). We estimated the fraction of cells that belong to each of the three stages (G_1, S, and G_2/M) using the tool for cell-cycle analysis in FlowJo.

Calculation of LINE density

The human LINE annotation file (rmsk.txt) was downloaded from the UCSC Genome Browser (www.genome.ucsc.edu, hg38) (Kent et al. 2002). LINE density of a gene was defined following a previous study (Hiratani et al. 2004) as the percentage of repetitive sequences in the 400-kb region surrounding the transcription start site of this gene.

Data access

The raw scRNA-seq data reported in this study have been submitted to the Genome Sequence Archive (Wang et al. 2017) in the Beijing Institute of Genomics (BIG) Data Center (https://bigd.big.ac.cn/gsa) under accession number CRA001055. All codes to analyze the data and generate figures are available at www.github.com/YingChen10/Synchronized-replication-during-S-phase and as Supplemental Code.

References

Albig W, Doenecke D. 1997. The human histone gene cluster at the D6S105 locus. *Hum Genet* **101:** 284–294. doi:10.1007/s004390050630

Batada NN, Hurst LD. 2007. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet* **39:** 945–949. doi:10.1038/ng2071

Bechhoefer J, Rhind N. 2012. Replication timing and its emergence from stochastic processes. *Trends Genet* **28:** 374–381. doi:10.1016/j.tig.2012.03.011

Birchler JA, Newton KJ. 1981. Modulation of protein levels in chromosomal dosage series of maize: the biochemical basis of aneuploid syndromes. *Genetics* **99:** 247–266.

Birchler JA, Veitia RA. 2010. The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol* **186:** 54–62. doi:10.1111/j.1469-8137.2009.03087.x

Birchler JA, Veitia RA. 2012. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci* **109:** 14746–14753. doi:10.1073/pnas.1207726109

Birchler JA, Bhadra U, Bhadra MP, Auger DL. 2001. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Dev Biol* **234:** 275–288. doi:10.1006/dbio.2001.0262

Bologna-Molina R, Mosqueda-Taylor A, Molina-Frechero N, Mori-Estevez AD, Sanchez-Acuna G. 2013. Comparison of the value of PCNA and Ki-67 as markers of cell proliferation in ameloblastic tumors. *Med Oral Patol Oral Cir Bucal* **18:** e174–e179. doi:10.4317/medoral.18573

Boye E, Nordström K. 2003. Coupling the cell cycle to cell growth. *EMBO Rep* **4:** 757–760. doi:10.1038/sj.embor.embor895

Chen H, He X. 2016. The convergent cancer evolution toward a single cellular destination. *Mol Biol Evol* **33:** 4–12. doi:10.1093/molbev/msv212

Chen X, Zhang J. 2016. The genomic landscape of position effects on protein expression level and noise in yeast. *Cell Syst* **2:** 347–354. doi:10.1016/j.cels.2016.03.009

Chen X, Shi S, He X. 2009. Evidence for gene length as a determinant of gene coexpression in protein complexes. *Genetics* **183:** 751–754. 751SI-755SI. doi:10.1534/genetics.109.105361

Chen H, Lin F, Xing K, He X. 2015. The reverse evolution from multicellularity to unicellularity during carcinogenesis. *Nat Commun* **6:** 6367. doi:10.1038/ncomms7367

Chen Y, Chen S, Li K, Zhang Y, Huang X, Li T, Wu S, Wang Y, Carey LB, Qian W. 2019. Overdosage of balanced protein complexes reduces proliferation rate in aneuploid cells. *Cell Syst* **9:** 129–142.e5. doi:10.1016/j.cels.2019.06.007

Clark NL, Alani E, Aquadro CF. 2012. Evolutionary rate covariation reveals shared functionality and coexpression of genes. *Genome Res* **22:** 714–720. doi:10.1101/gr.132647.111

Cooper S. 2003. Rethinking synchronization of mammalian cells for cell cycle analysis. *Cell Mol Life Sci* **60:** 1099–1106. doi:10.1007/s00018-003-2253-2

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30:** 207–210. doi:10.1093/nar/30.1.207

Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu Rev Plant Biol* **60:** 433–453. doi:10.1146/annurev.arplant.043008.092122

Gallone B, Steensels J, Prahl T, Soriaga L, Saels V, Herrera-Malaver B, Merlevede A, Roncoroni M, Voordeckers K, Miraglia L, et al. 2016. Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* **166:** 1397–1410.e16. doi:10.1016/j.cell.2016.08.020

Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, Dorschner MO, Gartler SM, Stamatoyannopoulos JA. 2010. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci* **107:** 139–144. doi:10.1073/pnas.0912402107

Hiratani I, Leskovar A, Gilbert DM. 2004. Differentiation-induced replication-timing changes are restricted to AT-rich/long interspersed nuclear element (LINE)-rich isochores. *Proc Natl Acad Sci* **101:** 16861–16866. doi:10.1073/pnas.0406687101

Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, Chang CW, Lyou Y, Townes TM, Schübeler D, Gilbert DM. 2008. Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol* **6:** e245. doi:10.1371/journal.pbio.0060245

Hurst LD, Pál C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* **5:** 299–310. doi:10.1038/nrg1319

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The Human Genome Browser at UCSC. *Genome Res* **12:** 996–1006. doi:10.1101/gr.229102

Koff A, Giordano A, Desai D, Yamashita K, Harper JW, Elledge S, Nishimoto T, Morgan DO, Franza BR, Roberts JM. 1992. Formation and activation of a cyclin E-cdk2 complex during the G_1 phase of the human cell cycle. *Science* **257:** 1689–1694. doi:10.1126/science.1388288

Kubben FJ, Peeters-Haesevoets A, Engels LG, Baeten CG, Schutte B, Arends JW, Stockbrugger RW, Blijham GH. 1994. Proliferating cell nuclear antigen (PCNA): a new marker to study human colonic cell proliferation. *Gut* **35:** 530–535. doi:10.1136/gut.35.4.530

Leitao RM, Kellogg DR. 2017. The duration of mitosis and daughter cell size are modulated by nutrients in budding yeast. *J Cell Biol* **216:** 3463–3470. doi:10.1083/jcb.201609114

Lin F, Xing K, Zhang J, He X. 2012. Expression reduction in mammalian X chromosome evolution refutes Ohno's hypothesis of dosage compensation. *Proc Natl Acad Sci* **109:** 11752–11757. doi:10.1073/pnas.1201816109

Long Y, Stahl Y, Weidtkamp-Peters S, Postma M, Zhou W, Goedhart J, Sánchez-Pérez MI, Gadella TWJ, Simon R, Scheres B, et al. 2017. *In vivo* FRET–FLIM reveals cell-type-specific protein interactions in *Arabidopsis* roots. *Nature* **548:** 97–102. doi:10.1038/nature23317

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161:** 1202–1214. doi:10.1016/j.cell.2015.05.002

Mitchison JM. 2003. Growth during the cell cycle. *Int Rev Cytol* **226:** 165–258. doi:10.1016/S0074-7696(03)01004-0

Moldovan GL, Pfander B, Jentsch S. 2007. PCNA, the maestro of the replication fork. *Cell* **129:** 665–679. doi:10.1016/j.cell.2007.05.003

Müller CA, Nieduszynski CA. 2017. DNA replication timing influences gene expression level. *J Cell Biol* **216:** 1907–1914. doi:10.1083/jcb.201701061

Murphy HA, Kuehne HA, Francis CA, Sniegowski PD. 2006. Mate choice assays and mating propensity differences in natural yeast populations. *Biol Lett* **2:** 553–556. doi:10.1098/rsbl.2006.0534

Oda M, Kanoh Y, Watanabe Y, Masai H. 2012. Regulation of DNA replication timing on human chromosome by a cell-type specific DNA binding protein SATB1. *PLoS One* **7:** e42375. doi:10.1371/journal.pone.0042375

Oromendia AB, Amon A. 2014. Aneuploidy: implications for protein homeostasis and disease. *Dis Model Mech* **7:** 15–20. doi:10.1242/dmm.013391

Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu Rev Genet* **34:** 401–437. doi:10.1146/annurev.genet.34.1.401

Padovan-Merhar O, Nair GP, Biaesch AG, Mayer A, Scarfone S, Foley SW, Wu AR, Churchman LS, Singh A, Raj A. 2015. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol Cell* **58:** 339–352. doi:10.1016/j.molcel.2015.03.005

Papp B, Pál C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424:** 194–197. doi:10.1038/nature01771

Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, Sanderson BW, Hattem GL, Li R. 2010. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* **468:** 321–325. doi:10.1038/nature09529

Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjan V, Muthusamy B, Gandhi TK, Gronborg M, et al. 2003. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res* **13:** 2363–2371. doi:10.1101/gr.1680803

Pope BD, Gilbert DM. 2013. The replication domain model: regulating replicon firing in the context of large-scale chromosome architecture. *J Mol Biol* **425:** 4690–4695. doi:10.1016/j.jmb.2013.04.014

Qian W, Zhang J. 2008. Gene dosage and gene duplicability. *Genetics* **179:** 2319–2324. doi:10.1534/genetics.108.090936

Qian W, Liao BY, Chang AY, Zhang J. 2010. Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends Genet* **26:** 425–430. doi:10.1016/j.tig.2010.07.002

Qian W, Ma D, Xiao C, Wang Z, Zhang J. 2012. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Rep* **2:** 1399–1410. doi:10.1016/j.celrep.2012.09.017

R Core Team. 2018. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Rivera-Mulia JC, Buckley Q, Sasaki T, Zimmerman J, Didier RA, Nazor K, Loring JF, Lian Z, Weissman S, Robins AJ, et al. 2015. Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. *Genome Res* **25:** 1091–1103. doi:10.1101/gr.187989.114

Ryba T, Battaglia D, Chang BH, Shirley JW, Buckley Q, Pope BD, Devidas M, Druker BJ, Gilbert DM. 2012. Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome Res* **22:** 1833–1844. doi:10.1101/gr.138511.112

Schübeler D, Scalzo D, Kooperberg C, van Steensel B, Delrow J, Groudine M. 2002. Genome-wide DNA replication profile for *Drosophila melanogaster*: a link between transcription and replication timing. *Nat Genet* **32:** 438–442. doi:10.1038/ng1005

Sima J, Chakraborty A, Dileep V, Michalski M, Klein KN, Holcomb NP, Turner JL, Paulsen MT, Rivera-Mulia JC, Trevilla-Garcia C, et al. 2019. Identifying *cis* elements for spatiotemporal control of mammalian DNA replication. *Cell* **176:** 816–830.e18. doi:10.1016/j.cell.2018.11.036

Skinner SO, Xu H, Nagarkar-Jaiswal S, Freire PR, Zwaka TP, Golding I. 2016. Single-cell analysis of transcription kinetics across the cell cycle. *eLife* **5:** e12175. doi:10.7554/eLife.12175

Smith L, Plug A, Thayer M. 2001. Delayed replication timing leads to delayed mitotic chromosome condensation and chromosomal instability of chromosome translocations. *Proc Natl Acad Sci* **98:** 13300–13305. doi:10.1073/pnas.241355098

Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. 2009. Human mutation rate associated with DNA replication timing. *Nat Genet* **41:** 393–395. doi:10.1038/ng.363

Stejskal S, Stepka K, Tesarova L, Stejskal K, Matejkova M, Simara P, Zdrahal Z, Koutna I. 2015. Cell cycle-dependent changes in H3K56ac in human cells. *Cell Cycle* **14:** 3851–3863. doi:10.1080/15384101.2015.1106760

Stingele S, Stoehr G, Peplowska K, Cox J, Mann M, Storchova Z. 2012. Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Mol Syst Biol* **8:** 608. doi:10.1038/msb.2012.40

Sun M, Zhang J. 2019. Chromosome-wide co-fluctuation of stochastic gene expression in mammalian cells. *PLoS Genet* **15:** e1008389. doi:10.1371/journal.pgen.1008389

Tasdighian S, Van Bel M, Li Z, Van de Peer Y, Carretero-Paulet L, Maere S. 2017. Reciprocally retained genes in the angiosperm lineage show the hallmarks of dosage balance sensitivity. *Plant Cell* **29:** 2766–2785. doi:10.1105/tpc.17.00313

Thurman RE, Day N, Noble WS, Stamatoyannopoulos JA. 2007. Identification of higher-order functional domains in the human ENCODE regions. *Genome Res* **17:** 917–927. doi:10.1101/gr.6081407

Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, Amon A. 2007. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317:** 916–924. doi:10.1126/science.1142210

Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker CA, Gygi SP, Dunham MJ, Amon A. 2010. Identification of aneuploidy-tolerating mutations. *Cell* **143:** 71–83. doi:10.1016/j.cell.2010.08.038

Veitia RA. 2005. Gene dosage balance: deletions, duplications and dominance. *Trends Genet* **21:** 33–35. doi:10.1016/j.tig.2004.11.002

Venet D, Dumont JE, Detours V. 2011. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput Biol* **7:** e1002240. doi:10.1371/journal.pcbi.1002240

Voichek Y, Bar-Ziv R, Barkai N. 2016. Expression homeostasis during DNA replication. *Science* **351:** 1087–1090. doi:10.1126/science.aad1162

Volpe P, Eremendo-Volpe T. 1970. A method for measuring the length of each phase of the cell cycle in spinner's cultures. *Exp Cell Res* **60:** 456–458. doi:10.1016/0014-4827(70)90542-2

Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, Tang B, Dong L, Ding N, Zhang Q, et al. 2017. GSA: Genome Sequence Archive. *Genomics Proteomics Bioinformatics* **15:** 14–18. doi:10.1016/j.gpb.2017.01.001

Weddington N, Stuy A, Hiratani I, Ryba T, Yokochi T, Gilbert DM. 2008. ReplicationDomain: a visualization tool and comparative database for genome-wide replication timing data. *BMC Bioinformatics* **9:** 530. doi:10.1186/1471-2105-9-530

Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, et al. 2002. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* **13:** 1977–2000. doi:10.1091/mbc.02-02-0030

Woo YH, Li WH. 2012. DNA replication timing and selection shape the landscape of nucleotide variation in cancer genomes. *Nat Commun* **3:** 1004. doi:10.1038/ncomms1982

Woodfine K, Fiegler H, Beare DM, Collins JE, McCann OT, Young BD, Debernardi S, Mott R, Dunham I, Carter NP. 2004. Replication timing of the human genome. *Hum Mol Genet* **13:** 191–202. doi:10.1093/hmg/ddh016

Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. 2000. DIP: the Database of Interacting Proteins. *Nucleic Acids Res* **28:** 289–291. doi:10.1093/nar/28.1.289

Xu H, Liu JJ, Liu Z, Li Y, Jin YS, Zhang J. 2019. Synchronization of stochastic expressions drives the clustering of functionally related genes. *Sci Adv* **5:** eaax6525. doi:10.1126/sciadv.aax6525

Yang YF, Cao W, Wu S, Qian W. 2017. Genetic interaction network as an important determinant of gene order in genome evolution. *Mol Biol Evol* **34:** 3254–3266. doi:10.1093/molbev/msx264

Zeitlin SG, Shelby RD, Sullivan KF. 2001. CENP-A is phosphorylated by Aurora B kinase and plays an unexpected role in completion of cytokinesis. *J Cell Biol* **155:** 1147–1157. doi:10.1083/jcb.200108125

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, et al. 2018. Ensembl 2018. *Nucleic Acids Res* **46:** D754–D761. doi:10.1093/nar/gkx1098

Zhang H, Wang Y, Li J, Chen H, He X, Zhang H, Liang H, Lu J. 2018. Biosynthetic energy cost for amino acids decreases in cancer evolution. *Nat Commun* **9:** 4124. doi:10.1038/s41467-018-06461-1