1   Dec. 12, 2018

2

3   Selection for synchronized replication of genes encoding the same protein

4   complex during tumorigenesis

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

6   [1] State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental

7   Biology, Chinese Academy of Sciences, Beijing 100101, China

8   [2] Key Laboratory of Genetic Network Biology, Institute of Genetics and

9   Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

10   [3] Department of Experimental and Health Sciences, Universitat Pompeu Fabra, 08003

11   Barcelona, Spain.

12   [4] University of Chinese Academy of Sciences, Beijing 100049, China

13   [5] Co-first author

14

15   [*] Correspondence to:

16   Wenfeng Qian

17   Institute of Genetics and Developmental Biology

18   Chinese Academy of Sciences

19   No. 1 West Beichen Road

20   Beijing 100101, China

21   Phone: 86-10-64806550

22   Email: wfqian@genetics.ac.cn

23

24

25   **Running title:** The synchronized replication hypothesis

26

1

1 ABSTRACT

2 DNA replication alters the dosage balance among genes; at the mid-S phase, early-
3 replicating genes have doubled their copies while late-replicating genes have not.
4 Dosage imbalance among proteins, especially within members of a protein complex,
5 is toxic to cells. Here, we propose the synchronized replication hypothesis: genes
6 sensitive to stoichiometric relationships will be replicated simultaneously to maintain
7 stoichiometry. In support of this hypothesis, we observe that genes encoding the same
8 protein complex have similar replication timing, but surprisingly, only in fast-
9 proliferating cells such as embryonic stem cells and cancer cells. The synchronized
10 replication observed in cancer cells, but not in slow-proliferating differentiated cells,
11 is due to convergent evolution during tumorigenesis that restores synchronized
12 replication timing within protein complexes. Collectively, our study reveals that the
13 selection for dosage balance during S phase plays an important role in the
14 optimization of the replication-timing program; that this selection is relaxed during
15 differentiation as the cell cycle is elongated, and restored as the cell cycle shortens
16 during tumorigenesis.

17

18 **Keywords:** Dosage balance; protein complex; replication timing; cellular evolution

19

20

1 INTRODUCTION

2   The balance hypothesis asserts that the stoichiometric relationship among
3 subunits of a protein complex is essential for the survival and proliferation of cells;
4 the disruption of this relationship perturbs functions of protein complexes and
5 sometimes even causes cytotoxicity (1-8). The balance hypothesis provides a unique
6 framework for understanding a variety of biological phenomena, especially the
7 proliferation rate of aneuploid cells and the fate of duplicated genes. Aneuploidy,
8 defined as a karyotype that is not a multiple of the haploid complement, generates
9 dosage imbalance among genes on different chromosomes. Consistent with the
10 balance hypothesis, aneuploidy often results in a more severe growth defect than a
11 whole genome duplication that keeps the dosage balance among genes (5, 9).
12 Furthermore, the addition of a larger chromosome, which leads to a dosage imbalance
13 among more genes, often results in a greater reduction in fitness (6, 10-13). Gene
14 duplication confers the second type of dosage imbalance, between duplicate genes and
15 singletons. Consistent with the balance hypothesis, genes often reduce their
16 expression soon after duplication (14), through which the dosage balance is restored.
17 Furthermore, genes encoding protein complexes exhibit a higher retention rate after
18 the whole genome duplication so that the dosage balance among subunits is
19 maintained (2, 15).

20   A probably more prevalent but less studied source of dosage imbalance is caused
21 by DNA replication that occurs each cell cycle. During the DNA synthesis phase (S
22 phase) of a cell cycle, the genome is replicated in a defined temporal order known as
23 the replication-timing program (16, 17). In the middle of S phase, early-replicating
24 genes have doubled their copy number, but late-replicating genes have not, leading to
25 a dosage imbalance between early and late-replicating genes (**Fig. 1A**). Such dosage
26 imbalance likely causes a growth defect especially among genes sensitive to dosage
27 relationship such as those encoding the same protein complex (2, 6). Although
28 acetylated histones (H3K56ac) can incorporate into newly replicated DNA regions
29 and partly suppress the expression of newly replicated genes in yeast (18, 19), this
30 compensatory mechanism cannot completely restore the dosage balance; the mRNA
31 levels of early-replicating genes still exhibited a ~20% increase compared to late-
32 replicating genes during the mid-S phase (18). Consistent with this, a GFP reporter
33 inserted into early-replicating regions in yeast exhibits higher expression (20).

34   The dosage imbalance during S phase could be severer in mammalian cells where

1   DNA replication lasts longer (~8 hours) each cell cycle. An exacerbating factor is that

2   H3K56ac may not mark newly replicated DNA in mammalian cells (21). Consistently,

3   in mouse embryonic stem cells (ESCs), the transcription rates of *Oct4* and *Nanog*

4   increased by 28% and 50%, respectively, upon DNA replication (22). These data show

5   that replication can cause dosage imbalance during S phase, and suggest that

6   additional mechanisms should exist in mammalian cells to solve the problem. Here,

7   we proposed a hypothesis that the replication of genes encoding the same protein

8   complex is synchronized during S phase so that the dosage balance is warranted.

9   Indeed, we observed a synchronized replication within protein complexes, but

10 surprisingly, only in fast-proliferating cells such as various tumor cells, indicating a

11 convergent evolution towards synchronized replication during tumorigenesis.

12

13 **RESULTS**

14 **Genes encoding subunits of the same protein complex tend to replicate**

15 **simultaneously in HeLa cells**

16   The synchronized replication hypothesis predicts a reduced variation in

17 replication timing among genes encoding the same protein complex. Indeed, genes

18 encoding some protein complexes are replicated almost simultaneously in HeLa cells

19 as exemplified in **Fig. 2A**. To test this prediction at the genomic scale, we retrieved

20 the components of 1,521 protein complexes from the Human Protein Reference

21 Database (23) and the replication-timing program of HeLa cells (24). For each protein

22 complex, we calculated the standard deviation of replication timing of all genes

23 encoding the protein complex (**Fig. 2B**). As a control, we randomly sampled genes

24 from the genome to constitute "pseudo" protein complexes, keeping the number of

25 complexes and the number of subunits in each complex unchanged (**Fig. 2B**). We

26 performed the random sampling 1,000 times. The median of observed standard

27 deviations is significantly smaller than the random expectation ($P < 0.001$,

28 permutation test, **Fig. 2B**), indicating synchronized replication within protein

29 complexes. The same conclusion can be reached when we shuffled among genes

30 encoding protein complexes to constitute "pseudo" protein complexes ($P < 0.001$,

31 permutation test, **Fig. 2C**).

32   It is worth noting that genes encoding the same protein complex tend to form

33 clusters on chromosomes (25, 26), which are likely to simultaneously replicate

4

1 because they have similar physical distances to the closest replication origin. To

2 determine if the smaller variation in replication timing within a protein complex is

3 fully explained by such gene clusters, we discarded protein complexes of which at

4 least two subunits are encoded by the genes on the same chromosome. The smaller

5 standard deviation of replication timing within a protein complex remained observed

6 ($P = 0.002$, permutation test, **Fig. S1**).

7 **Synchronized replication occurs only in fast-proliferating cells**

8   The replication-timing program varies among cell types. To determine if

9 synchronized replication occurs uniformly among various human cells, we retrieved

10 the replication-timing programs previously reported in 17 cell lines/types (24, 27).

11 They include 6 human ESC lines, 5 cancer cell lines, and 6 differentiated cell types

12 such as liver and pancreas cells derived from human ESCs (**Fig. 3A**). The

13 proliferation of ESCs and cancer cells is fast whereas that of differentiated cells is

14 slow (28). These cell lines/types exhibit various levels of synchronized replication

15 within protein complexes (as exemplified in **Fig. 3B**). To assess synchronized

16 replication at the genomic scale, we randomly shuffled genes encoding protein

17 complexes (**Fig. S2**, with three examples shown in **Fig. 3C**). We used the P value of

18 the permutation test to infer the level of synchronized replication in each cell line/type

19 and labeled synchronized replication for those with $P < 0.05$ (**Fig. 3E**). Surprisingly,

20 synchronized replication was exclusively observed in the 11 fast-proliferating cell

21 lines (**Fig. 3D-E**, $P = 8\times10^{-5}$, the Fisher's exact test).

22   We hypothesized that the loss of synchronized replication in differentiated cells

23 was caused by the reduced power of natural selection for synchronized replication in

24 slow-proliferating cells that spend a greater fraction of time in G_0 phase and a smaller

25 fraction of time in S phase (S%, **Fig. 1B**). To test this hypothesis, we estimated the

26 proliferation rate from the average expression level of 11 proliferation marker genes

27 (such as proliferating cell nuclear antigen, *PCNA*) for each cell line/type (**Fig. 3E**)

28 (29). As expected, the proliferation rate was positively correlated with the proportion

29 of cells in S phase among the 7 cell lines/types where flow-cytometry data were

30 available ($r = 0.8$, $P = 0.03$, Pearson's correlation, **Fig. 3F**) (27). The proliferation rate

31 was positively correlated with the level of synchronized replication among 17 cell

32 lines/types ($r = 0.83$, $P = 3\times10^{-5}$, Pearson's correlation, **Fig. 3G**), suggesting an S%-

33 dependent optimization of the replication-timing program in human cells.

34 **Synchronized replication is evolved convergently during tumorigenesis**

1   Since cancer cells are "evolved" from various differentiated cells rather than

2   ESCs. Consistently, ESCs and cancer cells did not form a monophyletic group (**Fig**

3   **3D**) when we clustered cell lines/types based on the similarity of their replicating-

4   timing programs. Instead, four out of the five cancer cell lines were clustered with

5   primitive hepatocytes and pancreatic endoderm cells, echoing their evolutionary

6   origin of differentiated cells. Nevertheless, the replication-timing programs of these 4

7   cancer cell lines permit synchronized replication whereas those of primitive

8   hepatocytes and pancreatic endoderm cells do not (**Fig. 3D**). More intriguingly, a

9   human colon cancer cell line, HCT116, has a very different replication-timing

10   program from all other cell types but exhibits the pattern of synchronized replication

11   (**Fig. 3D**). Collectively, these observations suggest convergent cellular evolution

12   during tumorigenesis to optimize the replication-timing program for fast proliferation.

13   **Differentiated cells lose synchronized replication mainly through a replication**

14   **delay**

15   To investigate the molecular mechanism by which synchronized replication is lost

16   in differentiated cells, we identified 165 protein complexes in which the standard

17   deviation of replication timing among subunits was significantly increased during

18   differentiation ($P < 0.05$ in the t-tests, an example is shown in **Fig. 3B**). For each of

19   the 491 genes encoding these protein complexes, we calculated the average

20   replication timing among ESCs and differentiated cells, respectively (**Fig. 4A**).

21   Among these genes, 92% exhibited a delay in replication in differentiated cells (top in

22   **Fig. 4B**), likely through a postponement of firing time of the closest replication

23   origins (an example is shown in **Fig. 4C**). In contrast, only 73% of non-complex

24   encoding genes exhibited delays in replication timing (top in **Fig. 4B**, odds ratio =

25   4.1, $P < 2.2 \times 10^{-26}$, the Fisher's exact test), suggesting that the differentiated cells lose

26   synchronized replication preferentially through a replication delay.

27   **Cancer cells reverse the change in replication timing during cell differentiation**

28   **to restore synchronized replication**

29   Among the 165 protein complexes in which synchronized replication is lost in

30   differentiated cells, 79 protein complexes restored synchronized replication in cancer

31   cells (bottom in **Fig. 4B**, $P < 0.05$ in the t-tests for the standard deviations between

32   differentiated and cancer cells; an example is shown in **Fig. 3B**). Presumably, the

33   restoration could occur either through i) reversing the change in replication timing

34   during cell differentiation or ii) through an intergenic suppression that the replication

6

1   timing of a second gene in the same protein complex follows that of the first. We

2   found that during tumorigenesis, 82% genes reversed the changes during cell

3   differentiation (an example is shown in **Fig. 4C**), significantly higher than the fraction

4   (72%) among genes not encoding protein complexes (bottom in **Fig. 4B**, odds ratio =

5   1.7, $P = 7 \times 10^{-4}$, the Fisher's exact test). Consistently, when we clustered the 17 cell

6   types/lines with the replication timing of the genes encoding these 79 protein

7   complexes, ESCs and cancer cells become closer in the dendrogram (**Fig. S3**).

8   **DISCUSSION**

9   Abnormal replication-timing programs have been known to be related to disease

10   and cancer (30-32). Our study provides a mechanism why a proper regulation of the

11   replication-timing program is essential, especially for fast-proliferating cells: to

12   maintain the dosage balance between early and late-replicating genes during S phase.

13   We observed a convergent cancer evolution of replication-timing program toward

14   ESCs, echoing previous analyses on the evolution of tumor cells at different levels,

15   such as those at the transcriptome or the amino acid usage level (33-35).

16   We showed that the demand for dosage balance during S phase could cause

17   synchronized replication of genes encoding the same protein complex. However, such

18   synchronized replication could also have evolved under other selection pressures. For

19   example, it may evolve to meet the demand for similar expression levels of genes

20   encoding the same protein complex because replication timing is associated with gene

21   expression level (36, 37). Nevertheless, this mechanism cannot explain why

22   synchronized replication is lost in differentiated cells, where the genes encoding

23   protein complexes remain expressed (**Fig. S4**) and the dosage balance among subunits

24   remains important (**Fig. S5**). The selection for dosage balance during S phase

25   uniquely predicts the loss of synchronized replication in slow-proliferating cells and

26   the re-gain of it in cancer cells.

27   The synchronized replication of complex members is restored in cancer cells,

28   although the number of mutations bared in each cancer cell is usually small (38). It is

29   therefore unlikely that each complex restores the synchronized replication through

30   individual mutations on its members during tumorigenesis. Master regulators of the

31   replication-timing program exist which control the firing of multiple replication

32   origins (39). In principle, the "switching" back of such master regulators to the ESC

33   status could make cancer cells rapidly restored synchronized replication. Consistently,

34   the re-gain of synchronized replication in cancer cells is mainly through reversing the

1   changes in replication timing during cell differentiation (bottom in **Fig. 4B**).

2   Whereas the dosage imbalance during S phase can be partly relieved by the

3   H3K56ac-associated transcription repression of newly replicated genes in yeast, the

4   balance is not completely restored; early-replicating genes still exhibit a ~20% higher

5   expression level during the mid-S phase (18). In cancer cells, the challenge of the

6   dosage imbalance during S phase is likely greater, for two reasons. First, H3K56ac

7   may play a less important role in repressing the expression of newly replicated genes

8   in mammalian cells (21). Second, S phase lasts for a much longer time in mammalian

9   cells than in yeast. For example, a HeLa cell divides every 24 hours and its S phase

10   lasts for ~8 hours (40); HeLa cells need to suffer from the imbalance between early-

11   and late-replicating genes for a few hours every 24 hours. By contrast, yeast has

12   adapted to the life cycle of 24~48 hours per generation in the fermentation industry

13   (41) or in nature which can be mimicked by a synthetic oak exudate medium (42, 43).

14   However, the S phase lasts for less than 1 hour even in a poor carbon source (44),

15   likely because the total time for DNA replication is mainly determined by the

16   elongation rate of the DNA polymerase. The imbalance between early and late-

17   replicating genes lasts for only dozen of minutes every 24~48 hours in yeast.

18   Collectively, the natural selection for synchronized replication is likely stronger in

19   cancer cells.

20   Some of the genes encoding the same protein complex form clusters on

21   chromosomes (25, 26). This observation is often explained by the demand for

22   coordinated gene expression (26), reduction in expression noise (45), and by the

23   positive epistatic relationship among these genes (46). We showed that the

24   synchronized replication hypothesis remained supported after controlling for such

25   gene clusters (**Fig. S1**), yet the gene cluster itself could, in turn, be an evolutionary

26   outcome of the selection for the dosage balance during S phase. Replication origins

27   fire stochastically at the single-cell level (47). Therefore, the synchronization of

28   replication timing is not robust in individual cells when these genes are interspersed in

29   the genome and use the different replication origins. Forming a gene cluster is a more

30   robust strategy for maintaining a dosage-sensitive relationship among genes.

31   Our results also have implications for DNA sequence evolution. For example,

32   since replication timing is a major determinant of mutation rate (48, 49), we predict

33   that genes encoding the same protein complex will have similar mutation rates due to

34   synchronized replication. Consistently, it was reported that the evolutionary rate

8

1 coevolves between a pair of genes that share a biological function or are co-expressed

2 (50). Collectively, our study not only identifies the driving forces underlying the

3 evolution of the replication-timing program but also provides new insights into the

4 evolution of DNA sequences.

5

6 **METHODS**

7 **Data retrievals**

8 The information of 1,521 protein complexes in humans was downloaded from the

9 Human Protein Reference Database (HPRD) release 9 (www.hprd.org). Among them,

10 1,317 were annotated completely and were used in this study. Chromosomal locations

11 of these genes were retrieved from Ensembl release 87 (www.ensembl.org).

12 The replication-timing profiles used in this study were downloaded from the

13 ReplicationDomain database (24) (https://www2.replicationdomain.com/) and are

14 listed in **Table S1**. Below we briefly describe how the experiments were done to

15 obtain the replication timing data. Detailed methods are in two previous studies (51,

16 52). Growing cells were pulse-labeled with bromodeoxyuridine (BrdU) for 1-2 hours,

17 fixed, and then labeled with propidium iodide. Labeled cells were separated by

18 fluorescence-activated cell sorter (FACS) into early and late-S bins. DNA with BrdU-

19 incorporation was immunoprecipitated, differentially labeled, and co-hybridized to a

20 microarray. The $\log_2$-transformed (early/late) ratio of the intensity of each probe was

21 used to generate a replication-timing profile for the entire genome.

22 Gene expression data used in this study were downloaded from NCBI Gene

23 Expression Omnibus (GEO) and are listed in **Table S1**.

24 **Estimation of gene replication timing**

25 The list of 19,805 protein-coding genes was retrieved from human GRCh38.p12

26 annotation file that was downloaded from Ensembl

27 (http://asia.ensembl.org/index.html). The average replication-timing ratio of probes

28 having overlap with each gene was defined as the replication timing of this gene.

29 **Estimation of the proliferation rate with gene expression profiles**

30 PCNA is a component of DNA polymerase δ and its expression level is a reporter

31 of DNA synthesis. We defined a panel of 11 meta-PCNA genes whose expression is

1 positively correlated with *PCNA*. Specifically, we normalized the average expression

2 level of genes ($\log_2$(RPKM+1)) in 17 cell lines/types and calculated the Pearson's

3 correlation between PCNA and each of 131 previously identified candidate meta-

4 PCNA genes. Genes with correlation coefficients greater than 0.9 were defined as

5 meta-PCNA genes in this study (*PCNA, ZWINT, RFC3, LBR, TFDP1, SNRPB, SMC4,*

6 *NUSAP1, BIRC5, UBE2C,* and *TROAP*). The proliferation rate was inferred from the

7 average expression level ($\log_2$(RPKM+1)) of the 11 meta-PCNA genes (29).

8 **Estimation of the fraction of cells in S phase with flow-cytometry**

9 Flow-cytometry data of propidium iodide-stained cells in 7 cell lines/types (listed

10 in **Table S1**) were generated in a previous study (27). We estimated the fraction of

11 cells belongs to one of the three stages in the cell cycle (G_1, S, and G_2/M) using

12 FlowJo.

13 **Clustering of cell lines/types**

14 The clustering of cell lines was performed with the function hclust in *R*. The

15 Ward's method was used.

16 **Loess smoothed replication-timing profiles**

17 We combined the $\log_2$-transformed ratios (early/late) of the intensity of each

18 probe for each of the three cell types (ESCs, differentiated cells, and cancer cells).

19 The function loess.smooth in *R* was used to generate a smoothed profile (**Fig. 4C**)

20 with the parameters span = 1/200 and evaluation = 2,000.

21 **Code availability**

22 All codes to analyze the data and generate figures are available at

23 https://github.com/YingChen10/Synchronized-replication-during-S-phase.

24

25 **AUTHOR CONTRIBUTIONS**

26 Y. C., K. L. and W. Q. conceived the research; Y. C., K. L., and X. C. analyzed the

27 data; Y. C., L. B. C., and W. Q. wrote the manuscript.

28 **ACKNOWLEDGMENTS**

29 We thank Tokyo Sasaki and David M Gilbert for providing the flow-cytometry data of

4   **CONFLICT OF INTEREST**
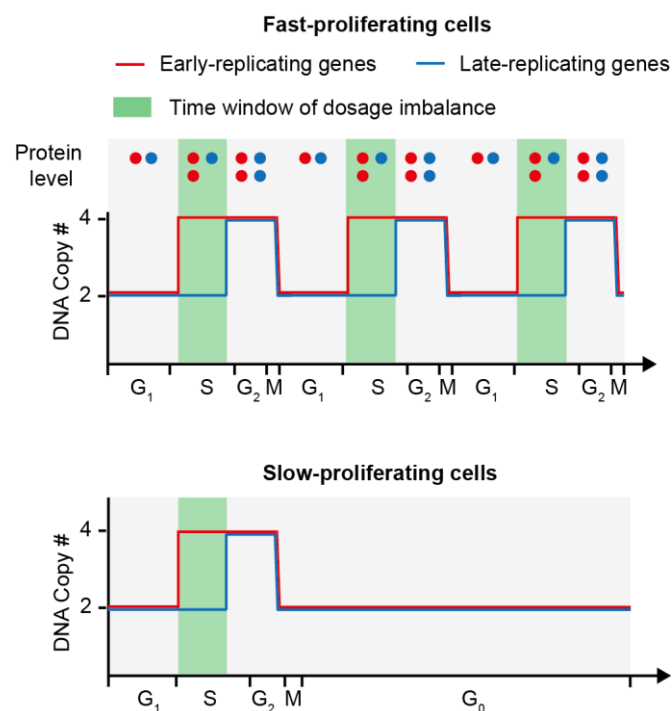
5   The authors declare no conflict of interest.

REFERENCES

1. Veitia RA (2002) Exploring the etiology of haploinsufficiency. *BioEssays : news and reviews in molecular, cellular and developmental biology* 24(2):175-184.
2. Papp B, Pal C, & Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424(6945):194-197.
3. Veitia RA (2005) Gene dosage balance: deletions, duplications and dominance. *Trends in genetics : TIG* 21(1):33-35.
4. Birchler JA & Veitia RA (2007) The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* 19(2):395-402.
5. Birchler JA & Veitia RA (2010) The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *The New phytologist* 186(1):54-62.
6. Birchler JA & Veitia RA (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proceedings of the National Academy of Sciences of the United States of America* 109(37):14746-14753.
7. Oromendia AB, Dodgson SE, & Amon A (2012) Aneuploidy causes proteotoxic stress in yeast. *Genes Dev* 26(24):2696-2708.
8. Sheltzer JM & Amon A (2011) The aneuploidy paradox: costs and benefits of an incorrect karyotype. *Trends in genetics : TIG* 27(11):446-453.
9. Otto SP & Whitton J (2000) Polyploid incidence and evolution. *Annual review of genetics* 34:401-437.
10. Stingele S*, et al.* (2012) Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Mol Syst Biol* 8:608.
11. Torres EM*, et al.* (2010) Identification of aneuploidy-tolerating mutations. *Cell* 143(1):71-83.
12. Torres EM*, et al.* (2007) Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* 317(5840):916-924.
13. Pavelka N*, et al.* (2010) Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* 468(7321):321-325.
14. Qian W, Liao BY, Chang AY, & Zhang J (2010) Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends in genetics : TIG* 26(10):425-430.
15. Qian W & Zhang J (2008) Gene dosage and gene duplicability. *Genetics* 179(4):2319-2324.
16. Pope BD & Gilbert DM (2013) The replication domain model: regulating replicon firing in the context of large-scale chromosome architecture. *J Mol Biol* 425(23):4690-4695.
17. Yamazaki S*, et al.* (2012) Rif1 regulates the replication timing domains on the

human genome. *The EMBO journal* 31(18):3667-3677.

18. Voichek Y, Bar-Ziv R, & Barkai N (2016) Expression homeostasis during DNA replication. *Science* 351(6277):1087-1090.

19. Bar-Ziv R, Voichek Y, & Barkai N (2016) Dealing with Gene-Dosage Imbalance during S Phase. *Trends in genetics : TIG* 32(11):717-723.

20. Chen X & Zhang J (2016) The Genomic Landscape of Position Effects on Protein Expression Level and Noise in Yeast. *Cell systems* 2(5):347-354.

21. Stejskal S, *et al.* (2015) Cell cycle-dependent changes in H3K56ac in human cells. *Cell cycle* 14(24):3851-3863.

22. Skinner SO, *et al.* (2016) Single-cell analysis of transcription kinetics across the cell cycle. *Elife* 5:e12175.

23. Keshava Prasad TS, *et al.* (2009) Human Protein Reference Database--2009 update. *Nucleic acids research* 37(Database issue):D767-772.

24. Weddington N, *et al.* (2008) ReplicationDomain: a visualization tool and comparative database for genome-wide replication timing data. *BMC Bioinformatics* 9:530.

25. Albig W & Doenecke D (1997) The human histone gene cluster at the D6S105 locus. *Hum Genet* 101(3):284-294.

26. Hurst LD, Pal C, & Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5(4):299-310.

27. Rivera-Mulia JC, *et al.* (2015) Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. *Genome research* 25(8):1091-1103.

28. Ruijtenberg S & van den Heuvel S (2016) Coordinating cell proliferation and differentiation: Antagonism between cell cycle regulators and cell type-specific gene expression. *Cell cycle* 15(2):196-212.

29. Venet D, Dumont JE, & Detours V (2011) Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput Biol* 7(10):e1002240.

30. Ryba T, *et al.* (2012) Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome research* 22(10):1833-1844.

31. Smith L, Plug A, & Thayer M (2001) Delayed replication timing leads to delayed mitotic chromosome condensation and chromosomal instability of chromosome translocations. *Proceedings of the National Academy of Sciences of the United States of America* 98(23):13300-13305.

32. Amiel A, *et al.* (2001) The influence of cytogenetic aberrations on gene replication in chronic lymphocytic leukemia patients. *Cancer Genet Cytogenet* 125(2):81-86.

33. Chen H & He X (2016) The Convergent Cancer Evolution toward a Single Cellular Destination. *Molecular biology and evolution* 33(1):4-12.

34. Chen H, Lin F, Xing K, & He X (2015) The reverse evolution from multicellularity to unicellularity during carcinogenesis. *Nature communications* 6:6367.

35. Zhang H, *et al.* (2018) Biosynthetic energy cost for amino acids decreases in cancer evolution. *Nature communications* 9(1):4124.

36. Schubeler D, *et al.* (2002) Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing. *Nature genetics* 32(3):438-442.

37. Woodfine K, *et al.* (2004) Replication timing of the human genome. *Hum Mol Genet* 13(2):191-202.

38. Martincorena I, *et al.* (2017) Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* 171(5):1029-1041 e1021.

39. Rivera-Mulia JC, Kim S, Gabr H, Kahveci T, & Gilbert DM (2017) Replication Timing Networks: a novel class of gene regulatory networks. *bioRxiv*.

40. 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(3):456-458.

41. Gallone B, *et al.* (2016) Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts. *Cell* 166(6):1397-1410 e1316.

42. Murphy HA, Kuehne HA, Francis CA, & Sniegowski PD (2006) Mate choice assays and mating propensity differences in natural yeast populations. *Biol Lett* 2(4):553-556.

43. Qian W, Ma D, Xiao C, Wang Z, & Zhang J (2012) The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell reports* 2(5):1399-1410.

44. Leitao RM & Kellogg DR (2017) The duration of mitosis and daughter cell size are modulated by nutrients in budding yeast. *J Cell Biol* 216(11):3463-3470.

45. Batada NN & Hurst LD (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nature genetics* 39(8):945-949.

46. Yang YF, Cao W, Wu S, & Qian W (2017) Genetic Interaction Network as an Important Determinant of Gene Order in Genome Evolution. *Molecular biology and evolution* 34(12):3254-3266.

47. Bechhoefer J & Rhind N (2012) Replication timing and its emergence from stochastic processes. *Trends in genetics : TIG* 28(8):374-381.

48. Stamatoyannopoulos JA, *et al.* (2009) Human mutation rate associated with DNA replication timing. *Nature genetics* 41(4):393-395.

49. Woo YH & Li WH (2012) DNA replication timing and selection shape the landscape of nucleotide variation in cancer genomes. *Nature communications* 3:1004.

50. Clark NL, Alani E, & Aquadro CF (2012) Evolutionary rate covariation reveals shared functionality and coexpression of genes. *Genome research* 22(4):714-720.

51. Hiratani I, Leskovar A, & Gilbert DM (2004) Differentiation-induced replication-timing changes are restricted to AT-rich/long interspersed nuclear element (LINE)-rich isochores. *Proceedings of the National Academy of Sciences of the United States of America* 101(48):16861-16866.

52. Hiratani I, *et al.* (2008) Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol* 6(10):e245.

1 **FIGURES**

2

3 **Figure 1. Dosage imbalance between early and late-replicating genes during S**
4 **phase in fast-proliferating (A) and slow-proliferating (B) cells.** The demand for
5 dosage balance during S phase is higher in fast-proliferating cells.
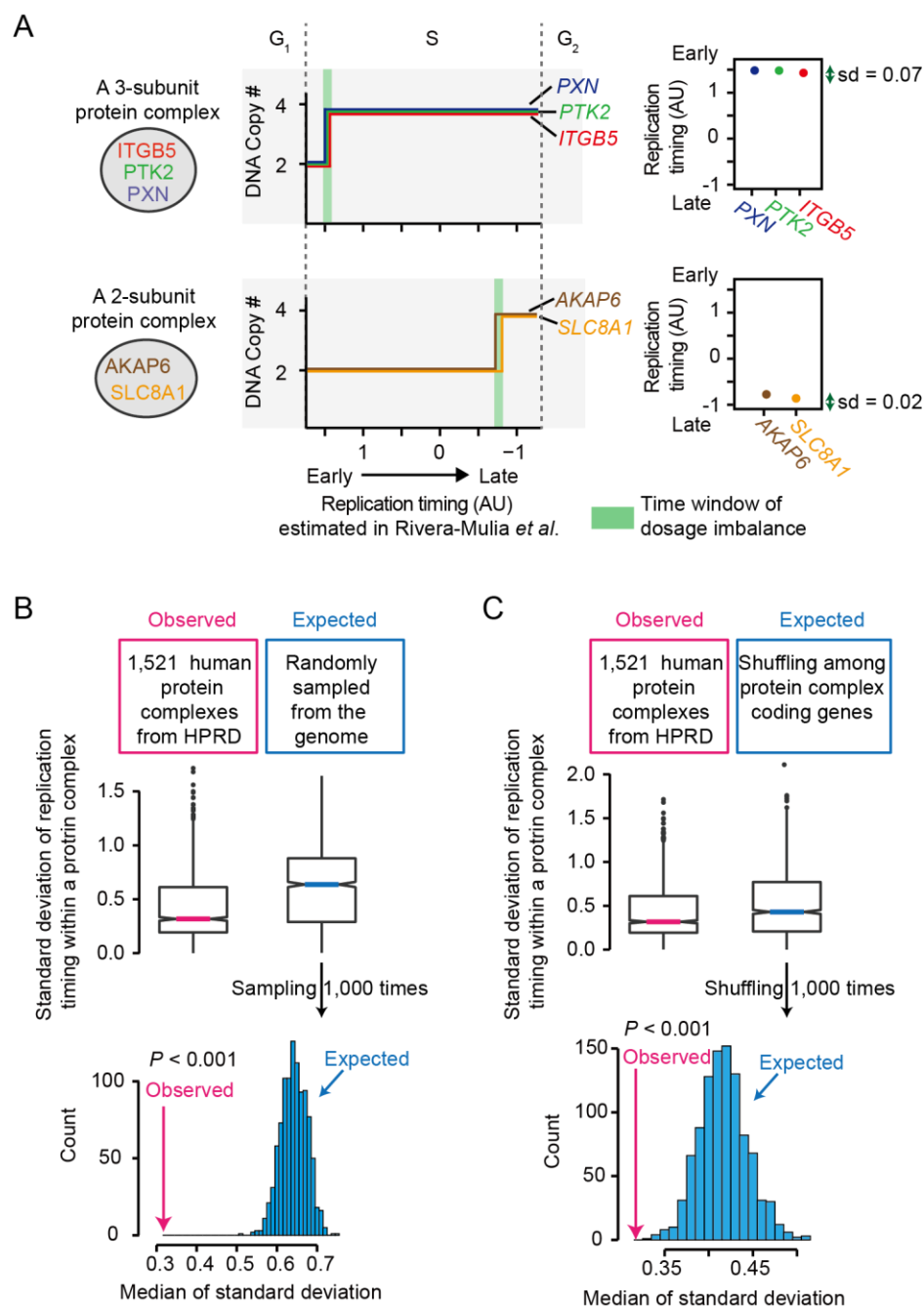
6

16

1

Figure 2. Genes encoding the same protein complex are replicated simultaneously in HeLa cells.

(A) Two examples showcase the synchronized replication of genes encoding the same protein complex. The standard deviation of replication timing within a protein complex is shown on the right.

(B-C) The observed standard deviation of replication timing within a protein complex is significantly smaller than the random expectation where the protein complex-coding genes were randomly sampled from the genome (B) or shuffled (C). The protein complex information was retrieved from the Human Protein Reference Database (HPRD).
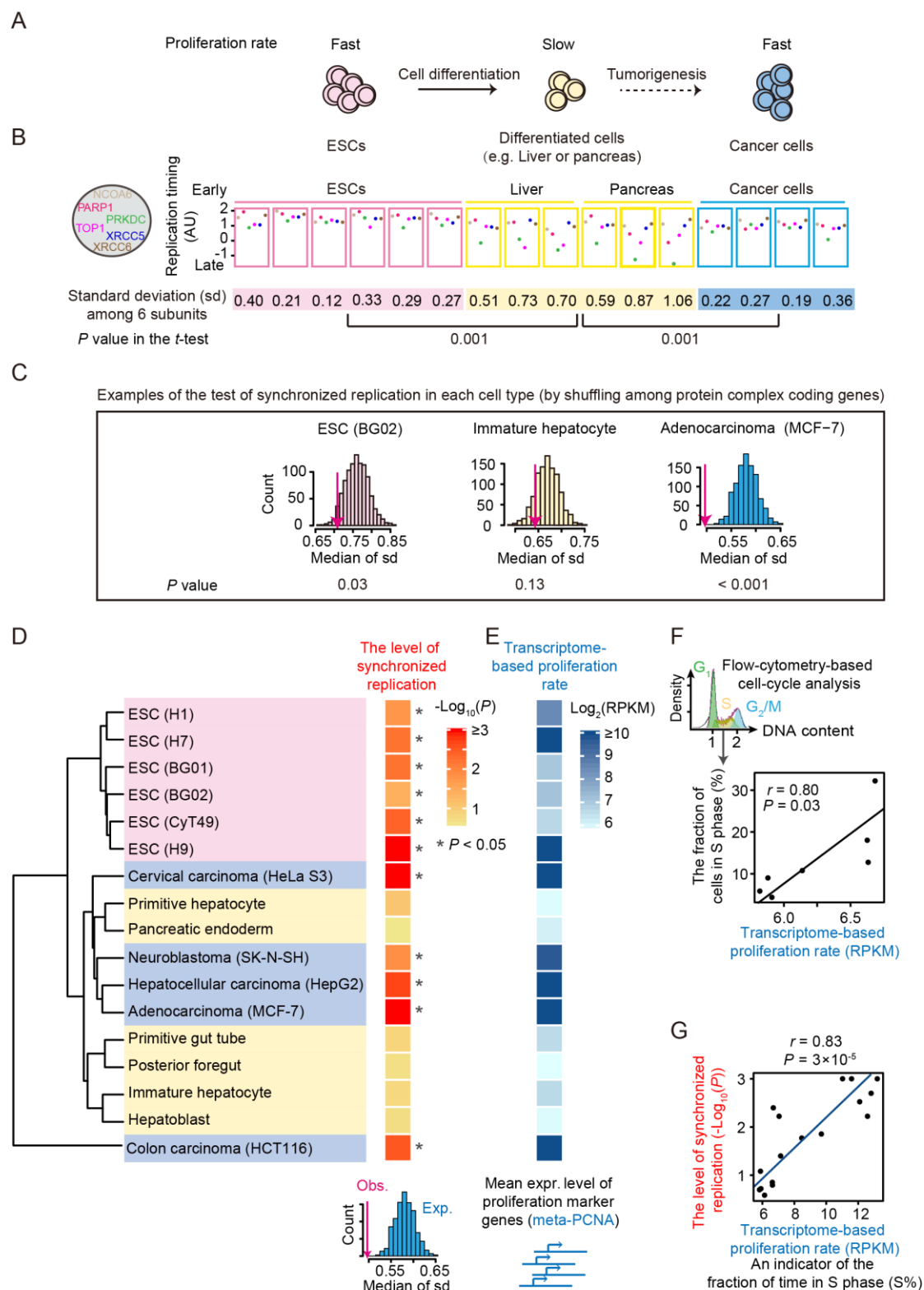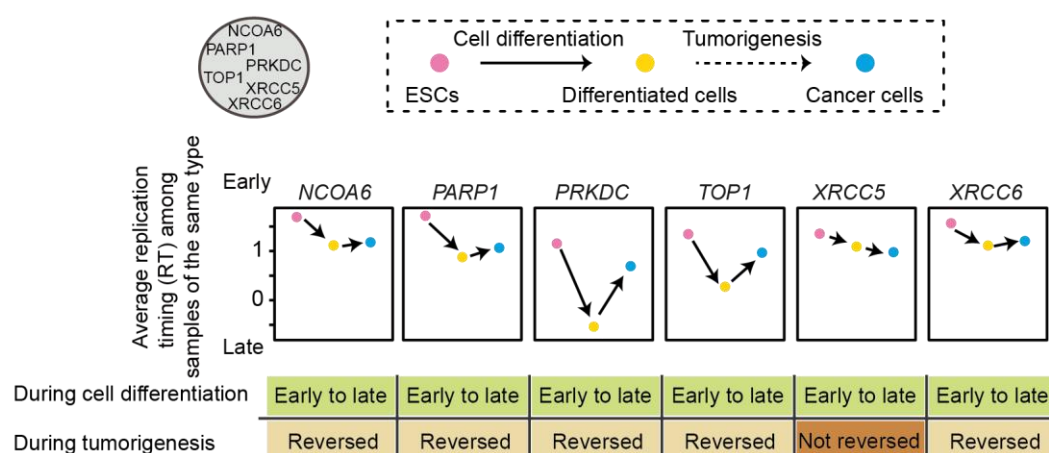
17

1

2 **Figure 3. Genes in the same protein complex are replicated simultaneously only**
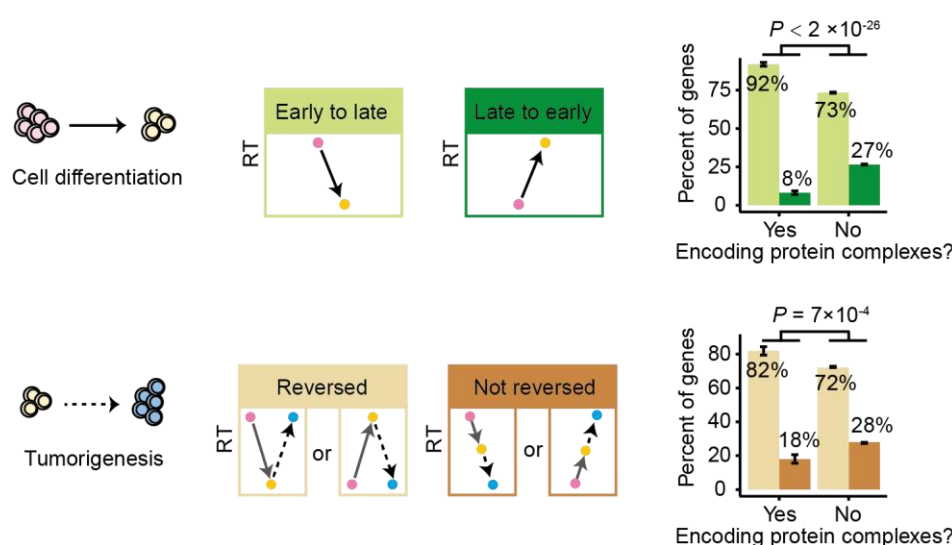
3 **in fast-proliferating cells.**

4 (A) A schematic diagram shows the changes in cell proliferation rate during cell

5 differentiation and tumorigenesis.

1   (B) Replication timings of 6 genes encoding the same protein complex. Significant

2   changes in standard deviation were estimated from P values in the t-tests.

3   (C) Three examples of the tests for synchronized replication.

4   (D) Synchronized replication occurs exclusively in ESCs and cancer cells. The

5   dendrogram (left) shows the clustering of 14 cell lines/types based on the replication-

6   timing profile of all genes in the genome. The heat map on the right shows the level of

7   synchronized replication estimated from the permutation test. The asterisk represents

8   significant synchronized replication ($P < 0.05$) in the corresponding cell line/type.

9   (E) The proliferation rate was estimated from the average expression level of 11

10   proliferation marker genes (meta-PCNA) in the mRNA-seq data.

11   (F) The fraction of cells in S phase (estimated by flow-cytometry) can be predicted

12   from the transcriptome-based proliferation rate. Therefore, the transcriptome-based

13   proliferation rate can be used as an indicator of the fraction of time in S phase (S%).

14   (G) The proliferation rate and the level of synchronized replication are positively
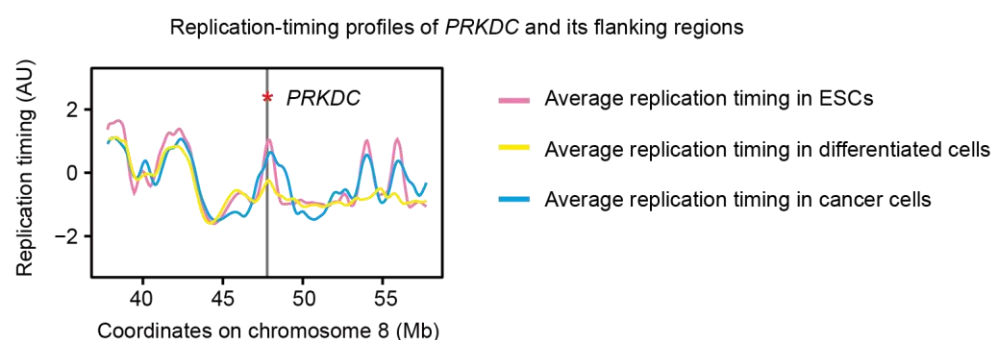
15   correlated.

Figure 4. Mechanisms by which synchronized replication is lost and restored.
(A) Calculation of the average replication timing (RT) among cells in the same group (ESCs, differentiated cells, or cancer cells).
(B) The fraction of genes in each category during cell differentiation or tumorigenesis.
(C) An example of changes in the replication-timing program. The loess-smoothed curves of replication timing in each cell group are shown. The vertical line and the asterisk indicate the position of the gene.