



The Distribution of Genomic Variations in Human iPSCs Is Related to Replication-Timing Reorganization during Reprogramming

Citation

Lu, Junjie, Hu Li, Ming Hu, Takayo Sasaki, Anna Baccei, David M. Gilbert, Jun S. Liu, James J. Collins, and Paul H. Lerou. 2014. "The Distribution of Genomic Variations in Human iPSCs Is Related to Replication-Timing Reorganization During Reprogramming." *Cell Reports* 7 (1) (April): 70–78. doi:10.1016/j.celrep.2014.03.007.

Published version

<https://doi.org/10.1016/j.celrep.2014.03.007>

Link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:27002570>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

The Distribution of Genomic Variations in Human iPSCs Is Related to Replication-Timing Reorganization during Reprogramming

Junjie Lu,^{1,2,8} Hu Li,^{3,8} Ming Hu,^{4,9} Takayo Sasaki,⁵ Anna Baccei,^{1,2} David M. Gilbert,⁵ Jun S. Liu,⁴ James J. Collins,^{6,7} and Paul H. Lerou^{1,2,*}

¹Department of Pediatric Newborn Medicine and Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA

²Harvard Medical School, Harvard Stem Cell Institute, Boston, MA 02115, USA

³Center for Individualized Medicine, Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905, USA

⁴Department of Statistics, Harvard University, Cambridge, MA 02138, USA

⁵Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA

⁶Howard Hughes Medical Institute, Department of Biomedical Engineering and Center of Synthetic Biology, Boston University, Boston, MA 02215, USA

⁷Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA

⁸These authors contributed equally to this work

⁹Present address: Division of Biostatistics, Department of Population Health, New York University School of Medicine, New York, NY 10016, USA

*Correspondence: plerou@partners.org

<http://dx.doi.org/10.1016/j.celrep.2014.03.007>

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

SUMMARY

Cell-fate change involves significant genome reorganization, including changes in replication timing, but how these changes are related to genetic variation has not been examined. To study how a change in replication timing that occurs during reprogramming impacts the copy-number variation (CNV) landscape, we generated genome-wide replication-timing profiles of induced pluripotent stem cells (iPSCs) and their parental fibroblasts. A significant portion of the genome changes replication timing as a result of reprogramming, indicative of overall genome reorganization. We found that early- and late-replicating domains in iPSCs are differentially affected by copy-number gains and losses and that in particular, CNV gains accumulate in regions of the genome that change to earlier replication during the reprogramming process. This differential relationship was present irrespective of reprogramming method. Overall, our findings reveal a functional association between reorganization of replication timing and the CNV landscape that emerges during reprogramming.

INTRODUCTION

The genome is topologically organized in three-dimensional space within the nucleus and is highly dynamic as cell fate changes during normal development, as well as in disease states, such as cancer (Ryba et al., 2012). The significant cell-fate

changes that occur during embryonic stem cell (ESC) differentiation and reprogramming of somatic cells to pluripotency also involve genome-wide resetting of the higher-order chromatin structure (Watanabe et al., 2013). In both scenarios, genome reorganization precedes gene expression changes, suggesting a potential causal relationship (Apostolou et al., 2013; Phillips-Cremens et al., 2013; Wei et al., 2013; Zhang et al., 2013).

Higher-order genome organization has been shown to significantly influence the distribution of genomic aberrations in both immortalized somatic cells and cancer cells, but the underlying mechanism remains elusive (De and Michor, 2011; Fudenberg et al., 2011; Koren et al., 2012; Schuster-Böckler and Lehner, 2012). One method of mapping genomic organization is by segmenting the genome into domains based on replication timing, which occurs in a tightly regulated cell-type-specific manner (Gilbert et al., 2010). Domains with similar replication timing tend to colocalize within the nucleus, and as a result, genome organization defined by replication timing overlaps with other methods of defining chromatin topology, including DNase I-hypersensitivity profiles, various epigenetic markers, and genome-wide chromatin-interaction mapping (e.g., Hi-C) (Dixon et al., 2012; Hu et al., 2013; Ryba et al., 2010). Several studies have revealed that the cancer mutational landscape is closely associated with replication timing by demonstrating that copy-number variation (CNV) boundaries tend to have similar replication timing, and gains and losses distribute differentially with respect to replication timing (De and Michor, 2011; Liu et al., 2013; Schuster-Böckler and Lehner, 2012). These studies share one important limitation: the replication-timing maps were not generated from the same cell types in which the CNVs were analyzed. This limitation is significant because cellular identity is closely coupled to genome organization; in fact, only half of

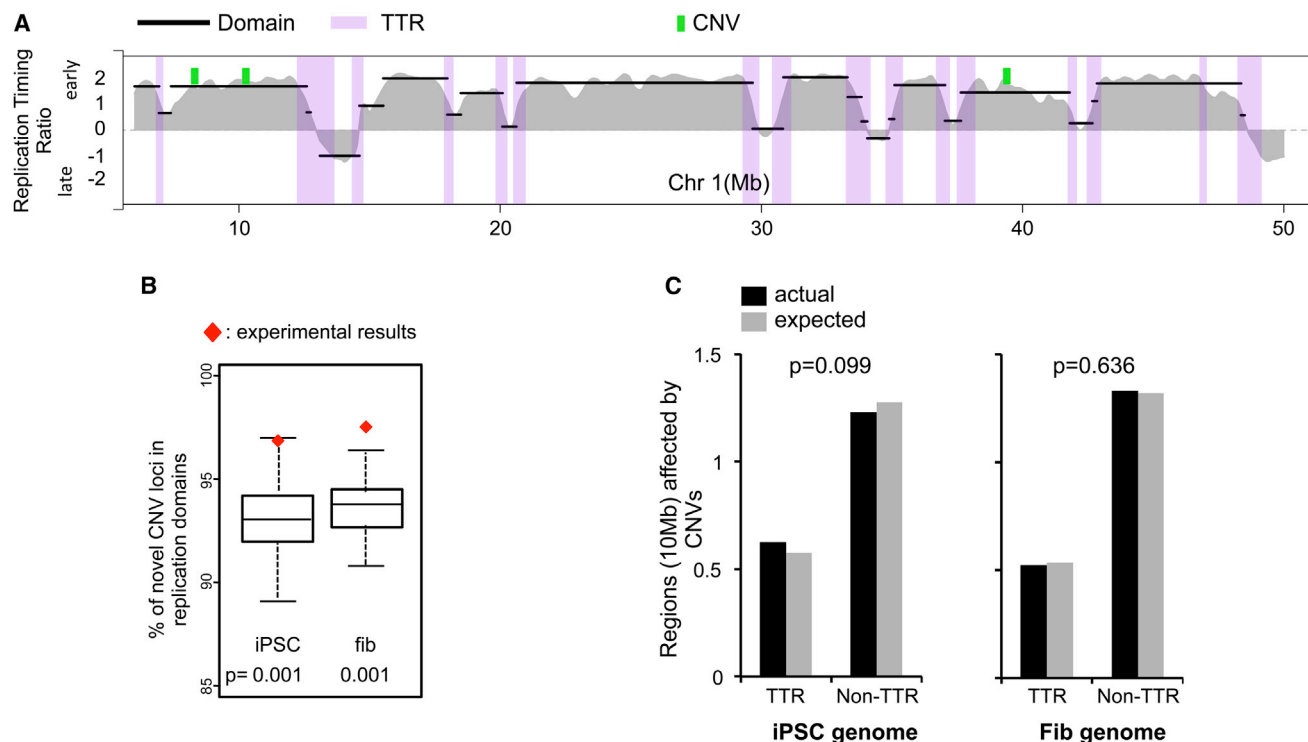


Figure 1. iPSC CNVs Reside within Replication-Timing Domains and Are Distributed Nonrandomly

(A) Newly replicated DNA from early and late S phase is differentially labeled and hybridized to a whole-genome oligonucleotide microarray to produce a replication-timing profile. Loess-smoothed replication-timing ratio (filled gray) of the first 50 Mb of chromosome (Chr) 1 is shown as an example. Segmentation algorithms are used to define replication domains (black segments) as regions with similar replication-timing ratio. TTRs are defined as genomic regions with a significant rate of replication-timing ratio change (purple). TTRs cover approximately one-third of the genome. CNVs (green segments) are plotted above the replication profile, and they can fall either entirely within individual domains or cross boundaries and involve more than one domain. A CNV may or may not involve a TTR.

(B) Permutation analysis demonstrates that CNVs are significantly more likely to reside within individual replication domains than expected by chance in both iPSCs and parental fibroblasts (fib). Boxes display the percentages obtained from 1,000 simulations, and red diamonds represent the actual percentages.

(C) TTRs are not enriched within CNVs. The actual amount (black) of TTR and non-TTRs affected by CNVs, compared to expected numbers (gray) based on random distribution, is shown with p values (Fisher's exact test) above each plot.

See also [Figure S1](#).

the genome has stable replication timing across all cell types, with the remainder being organized in a highly cell-type-specific manner ([Hansen et al., 2010](#); [Hiratani et al., 2008, 2010](#)). Thus, when an unmatched replication-timing profile is used, up to 50% of the genome may not be accounted for during the analysis due to the cell-type specificity of genome organization.

How cell fate change and its associated genome reorganization are related to genome variation has not been examined. We overcome the limitation of unmatched replication-timing maps by generating replication-timing profiles of both induced pluripotent stem cells (iPSCs) and their parental fibroblasts in order to explore the link between the CNV landscape and genome reorganization due to cell fate change. We show that nuclear reprogramming results in dramatic replication-timing reorganization that influences the observed CNV landscape in iPSCs.

RESULTS

CNVs Reside within Replication-Timing Domains

Primary dermal fibroblasts from one healthy volunteer (wild-type, WT) were reprogrammed using retroviral transduction of the

standard Yamanaka factors to generate multiple iPSC lines ([Takahashi et al., 2007](#)) ([Figure S1](#)). Replication-timing profiles were generated from an iPSC line and the parental fibroblasts. Newly replicated DNA from early and late S phase was differentially labeled and hybridized to a whole-genome oligonucleotide microarray. The ratio of the abundance of each probe in the early versus late fractions generates a replication profile, which reveals clear demarcation between megabase-sized regions of coordinated replication called replication domains ([Figure 1A](#)). Computational segmentation algorithms can be used to define the replication domains ([Ryba et al., 2011](#)). The iPSC profile we generated is identical to previously published profiles of both human ESCs and iPSCs, reflecting human pluripotent stem cell-specific genome organization ([Figure S1D](#)) ([Ryba et al., 2010](#)).

To map the corresponding genome-variation landscape of iPSCs, we analyzed the fibroblast and low-passage (p6) iPSC lines using high-resolution Affymetrix SNP Array 6.0 ([Tables S1 and S2](#)). CNV calls were validated using quantitative PCR with an overall validation rate of 91% ([Supplemental Experimental Procedures](#)). Raw data from the replication-timing profiles also

independently validated all the homozygous losses detected by SNP arrays in the same cell lines (Figure S1E) (Ryba et al., 2012). Most CNVs of iPSCs were not detected in the parental fibroblasts and are heretofore referred to as “iPSC manifested” (Tables S1 and S2). iPSC-manifested CNVs in low-passage iPSCs could represent clonal expansion of low-frequency genetic aberrations in the tissue of origin or de novo mutations during the reprogramming process (Liang and Zhang, 2013).

We found that, whereas two-thirds of CNVs detected in iPSCs involve genic regions (including exons and introns), the affected genes do not fall into any clear functional groups by Gene Ontology analysis (Table S3). The CNV-affected genes did not show any particular relationship to pluripotency, nor did we find any enrichment of tumor suppression genes or oncogenes. Our data are similar to other published reports that failed to demonstrate consistent functional consequences due to genomic aberrations in iPSCs (Amps et al., 2011; Hussein et al., 2011; Martins-Taylor et al., 2011).

We mapped the iPSC CNVs to the newly generated replication-timing profiles and found that more than 95% of CNVs are contained within individual replication domains, irrespective of whether the iPSC or the parental fibroblast profile was used. Permutation analysis using size-controlled random CNV lists demonstrates that the detected CNVs are significantly more likely to reside completely within individual replication domains and do not span domain boundaries than would be expected to occur by chance (Figure 1B).

Timing transition regions (TTRs), defined by genomic regions with a significant rate of replication-timing ratio change, have different replication kinetics compared to the center of the replication domain. TTRs probably represent sequences vulnerable to DNA damage (Labib and Hodgson, 2007; Watanabe et al., 2002, 2004). However, we did not find an enrichment of TTR sequence in CNV-affected genomic regions in iPSCs (Figure 1C).

Impact of Replication-Timing Dynamics on the CNV Landscape

Reprogramming and ESC differentiation involve global changes in chromosome structure that include temporal reorganization of replication domains and spatial reorganization of the genome; in fact, one-half of the genome changes organization at some stage during development (Hiratani et al., 2008; Orkin and Hochedlinger, 2011; Ryba et al., 2010). Comparison of fibroblast and iPSC replication-timing profiles revealed that during reprogramming, 40% of the genome changes replication timing, as defined by a greater than 0.5 difference in the early/late ratio between fibroblast and iPSC profiles (Figure 2A). This cutoff corresponds to the amount of time (about 1 hr) required to complete a replication factory (Leonhardt et al., 2000) and is the same cutoff for significance for transcription (D.M.G., unpublished data). This change is nearly double in magnitude compared to the change observed when ESCs are differentiated into neural precursor cells (Hiratani et al., 2008), reflecting the extensive genome reorganization during nuclear reprogramming.

To test how genome reorganization influences the genome-variation landscape, we determined the fibroblast and iPSC replication-timing values of genomic loci affected by iPSC-manifested CNVs using cell-type-specific replication-timing

profiles. The distribution of replication-timing values of CNV gain loci is statistically different in fibroblasts and iPSCs; they are predominantly late replicating in fibroblasts but early replicating in iPSCs. In contrast, loci affected by CNV losses are generally late replicating in both cell types. (Figure 2B).

To better understand the dynamics of the CNV landscape, a genome-wide representation of replication-timing dynamics during reprogramming was generated by plotting fibroblast versus iPSC replication-timing values for all 620,000 probes on the array (Figure 2C). In this plot, the gray diagonal region represents genomic loci with no change in replication timing during reprogramming, defined as a less than 0.5 difference in replication-timing ratio between fibroblasts and iPSCs. The portions of the genome that replicate either earlier (Figure 2C, red) or later (blue) in iPSCs than in fibroblasts fall outside the diagonal. When the replication-timing values of CNVs were plotted on this genome background, we show that CNV gains are enriched in the genome compartment that changes to earlier replication during reprogramming. This scatterplot presentation reveals the replication-timing dynamics underlying the distribution shift of gains seen in Figure 2B. In contrast, the majority of losses reside in late replicating regions in both fibroblasts and iPSCs (Figure 2C), consistent with the absence of a shift in the CNV loss distribution (Figure 2B). Together, these findings further support the strong relationship between replication timing and genome variation by demonstrating that the distribution of CNVs observed in low-passage iPSCs is influenced by replication-timing reorganization associated with cell fate change.

Human iPSC genomic variation has been studied using a variety of approaches. Hussein et al. and Laurent et al. generated iPSC using the same retroviral factors as we did here, but whereas Hussein et al. also used the Affymetrix 6.0 SNP array, Laurent et al. used the Illumina OmniQuad v.1 SNP array (Hussein et al., 2011; Laurent et al., 2011). More recently, Abyzov et al. used whole-genome sequencing to identify CNVs in retroviral reprogrammed iPSC lines (Abyzov et al., 2012). None of these studies considered the role of higher-order chromatin organization on genomic variation, and their data sets serve as de facto controls to exclude potential lab-specific, CNV detection platform-specific, or parental fibroblast-specific effects in our experiments. Therefore, we analyzed these data sets relative to our newly generated replication-timing profiles. To maintain consistency with our experimental approach, we limited the analysis to data sets of low-passage iPSCs generated from dermal fibroblasts. An overall similar trend was seen using data sets from different laboratories: CNV distribution is nonrandom, with gains predominating in regions changing to earlier replication in iPSCs (Figure S2B).

Nonrandom, Differential CNV Distribution Is a Generalized Feature of Reprogramming

To exclude the possibility that the nonrandom replication-timing distribution of the CNVs is due to a direct effect of the retroviral vectors (RVs) used for reprogramming, we applied our analysis to iPSCs generated using other reprogramming methods. To this end, we generated iPSCs using episomal (EP) factors that do not integrate into the genome. In low-passage EP iPSCs, like RV iPSCs, CNV gains are significantly enriched in regions

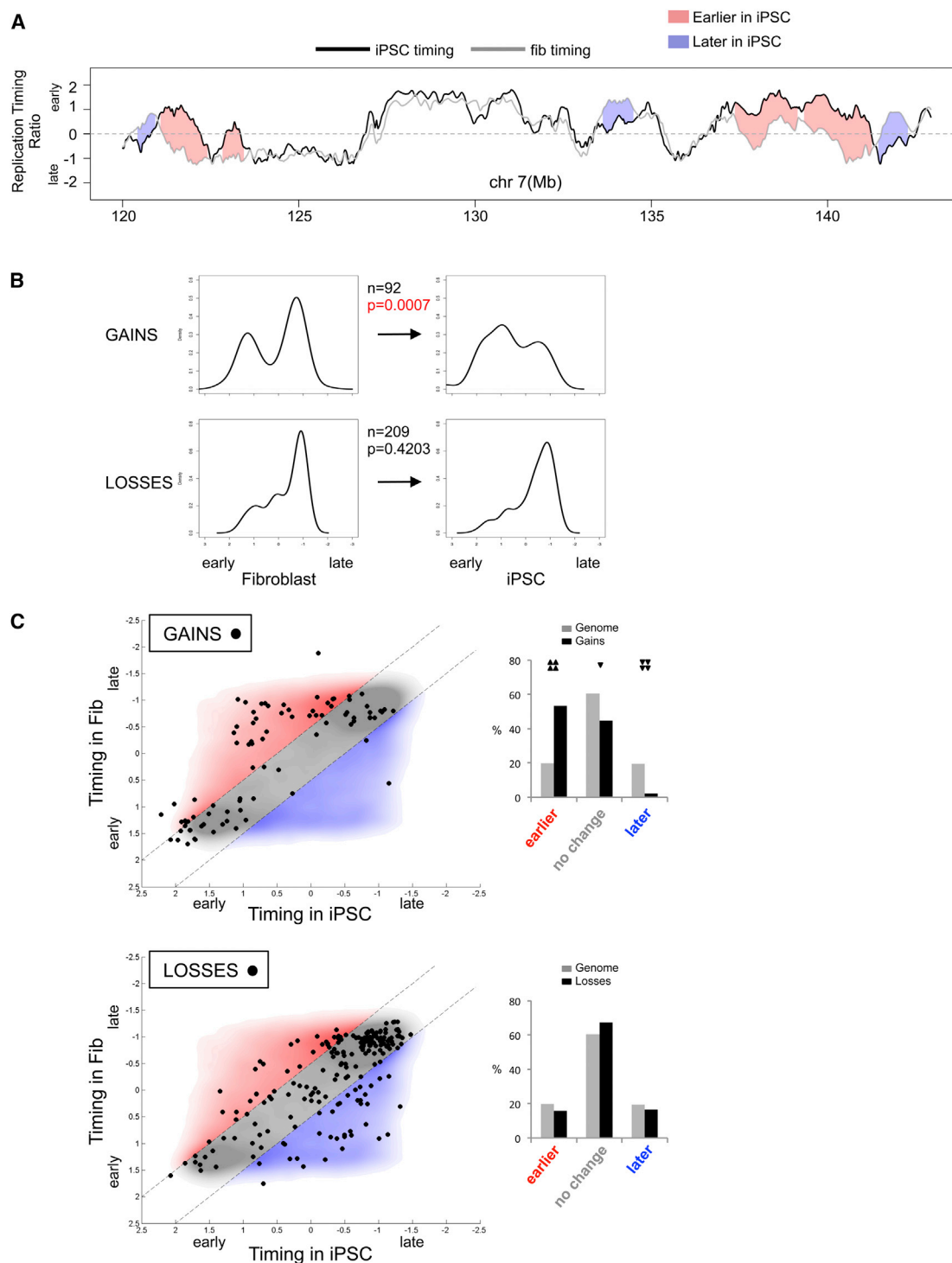


Figure 2. Reorganization of Replication-Timing Program during Reprogramming Affects CNV Distribution

(A) Comparison between replication-timing profiles of iPSCs (black) and parental fibroblasts (gray) of chr 7 120–143 Mb shows changes during reprogramming. Pink indicates regions changing to earlier replication: greater than +0.5 difference in replication-timing ratio between fibroblasts and iPSCs. Blue indicates regions changing to later replication: greater than –0.5 difference in replication-timing ratio between fibroblasts and iPSCs.

(B) Kernel density plots show the replication-timing distribution of CNV-affected regions in fibroblasts (left) and in iPSCs (right). CNV gains have earlier replication timing in iPSCs than fibroblasts. CNV losses are late replicating in both iPSCs and fibroblasts. p values from the K-S tests between fibroblast and iPSC distribution are shown.

(legend continued on next page)

of the genome whose replication timing changed to earlier during reprogramming. Again, we found that the majority of CNV losses occur at regions that do not significantly change replication timing during reprogramming (Figures 3A and S3).

Next, we analyzed Hussein et al.'s data on sister iPSC lines derived from the same parental fibroblast line (human foreskin fibroblasts, HFFs) using either retroviruses or piggyBac transposons (PBs) (Hussein et al., 2011). Distribution plots similar to Figure 2B from multiple data sets were directly compared to each other (Figures 3B and S3). Gains and losses from these PB and RV iPSCs had a similar nonrandom distribution with respect to replication-timing dynamics during reprogramming: gains were likely to have earlier replication in iPSCs compared to fibroblasts, whereas losses maintain late replication. Together, these findings demonstrate that the trend for nonrandom, differential CNV gain and loss accumulation as a function of change in replication timing is a generalized feature of reprogramming.

DISCUSSION

Improved mapping and understanding of the higher-order genome organization is providing a clearer picture of how genome functions, such as gene expression and genome integrity, are related to its spatial organization (Cavalli and Misteli, 2013). A static association between genome organization and genome aberration has been described, but how the mutagenic landscape is affected by changes in genome organization has not been studied. Here, by leveraging reprogramming to alter cell fate and analyzing the genome aberration landscape before and after using cell-type-matched genome-wide replication-timing profiles, we show that the CNV landscape of iPSCs is spatially constrained by chromatin organization, and the distribution of CNVs is influenced by changes in replication timing due to reprogramming.

Some studies have reported that reprogramming results in genetic aberrations such as point mutations and CNVs in human iPSCs (Gore et al., 2011; Hussein et al., 2011; Laurent et al., 2011; Martins-Taylor et al., 2011; Mayshar et al., 2010). However, more recently, others have demonstrated that the majority of iPSC CNVs preexisted in cells within the parental population (Abyzov et al., 2012; Cheng et al., 2012; Quinlan et al., 2011). Consistent with these reports, we found shared iPSC-manifested CNVs among the iPSC clones in each of our reprogramming experiments, albeit at low frequency. Interestingly, we also found that about 50% of iPSC-manifested iPSC CNVs have overlap with known CNVs in the Database of Genomic Variation (MacDonald et al., 2014), and these likely represent previously existing low-frequency CNVs in the starting cell population.

Ultimately, the accumulation of genomic changes in any genome will result from a combination of mutation rates, repair

rates, and selection. In a completely stochastic process, akin to genetic drift, CNV gains and losses would be distributed randomly. Our results show that this is not the case because CNVs are nonrandomly distributed in iPSCs as a function of replication-timing change. iPSC-manifested gains predominate in loci that change to earlier replication during reprogramming, whereas losses occur in loci that remain late replicating. The mechanism and timing of this relationship are not immediately apparent, but it is conceivable that both preexisting and de novo CNVs are subject to the influence of the extensive genome reorganization (Figure 4).

Switching to earlier replication involves epigenetic resetting that results in increased active chromatin modifications, such as histone acetylation (Hiratani et al., 2008; Schwaiger et al., 2009) and an overall open chromatin state. These regions contain genes essential to the cell's new functions and are therefore more likely to be under increased selection during reprogramming. Although this functional selection is plausible, it is difficult to reconcile with the inability to link CNVs in iPSCs to gene expression changes. Neither our gene expression analysis nor the published literature has consistently identified gene expression changes that correlate with genomic aberration to prove that this mechanism is dominant (Amps et al., 2011; Hussein et al., 2011; Martins-Taylor et al., 2011). In fact, in a recent analysis of 125 human pluripotent stem cell lines, the most commonly identified aberration was a minimal amplicon at 20q11.21 (affecting >20% of lines), yet there was no correlation between the presence of this amplicon and the expression of the genes it contains (Amps et al., 2011).

Variable efficiency of DNA damage response (DDR) pathways in different chromatin compartments could contribute to the differential distribution of gains and losses. It is possible that DNA damage occurring in early S phase is more accurately repaired due to higher-fidelity repair mechanisms, including copying from the correct template (Chang and Cimprich, 2009). Not only is DDR more efficient in open chromatin, but also heterochromatin appears to retard repair processes and could potentially lead to accumulation of deletions in the late replicating compartment (Goodarzi et al., 2008; Murga et al., 2007; Rübe et al., 2011). Another mechanism to explain differential distribution of de novo mutations is that replication-timing reorganization could adversely affect the interaction between replication and transcription machineries, leading to replication and transcription collisions, and consequently to double-strand breaks and CNVs (Helmrich et al., 2013).

Interestingly, we noticed subtle differences in CNV occurrence among the three reprogramming methods. In contrast to RV iPSCs, PB and EP iPSCs overall had significantly more CNV gains compared to losses (Figure 3B), suggesting that there

(C) Replication-timing values of CNV gains (top) or losses (bottom) were determined using the profiles of both fibroblasts and iPSCs and were shown as a scatterplot with density-smoothed genome-wide replication-timing values (RTV) of over 620,000 probes serving as background. The gray central region represents ~60% of the genome with no change in replication timing defined as iPSC RTV – fibroblast RTV $\leq \pm 0.5$. Regions with greater than +0.5 change (iPSC RTV – fibroblast RTV) are classified as “earlier in iPSC” (red), and those with less than –0.5 change (iPSC RTV – fibroblast RTV) are classified as “later in iPSCs” (blue). The bar graphs depict the distribution of CNV gains or losses compared to the genome-wide distribution of RTVs among these regions. Upward-pointing arrowheads indicate statistically significant enrichment of CNVs in the region, whereas downward-pointing arrowheads indicate statistically significant depletion of CNVs compared to the genome distribution in corresponding regions based on binomial test. Number of arrowheads correlates with p value significance (see Figure S2).

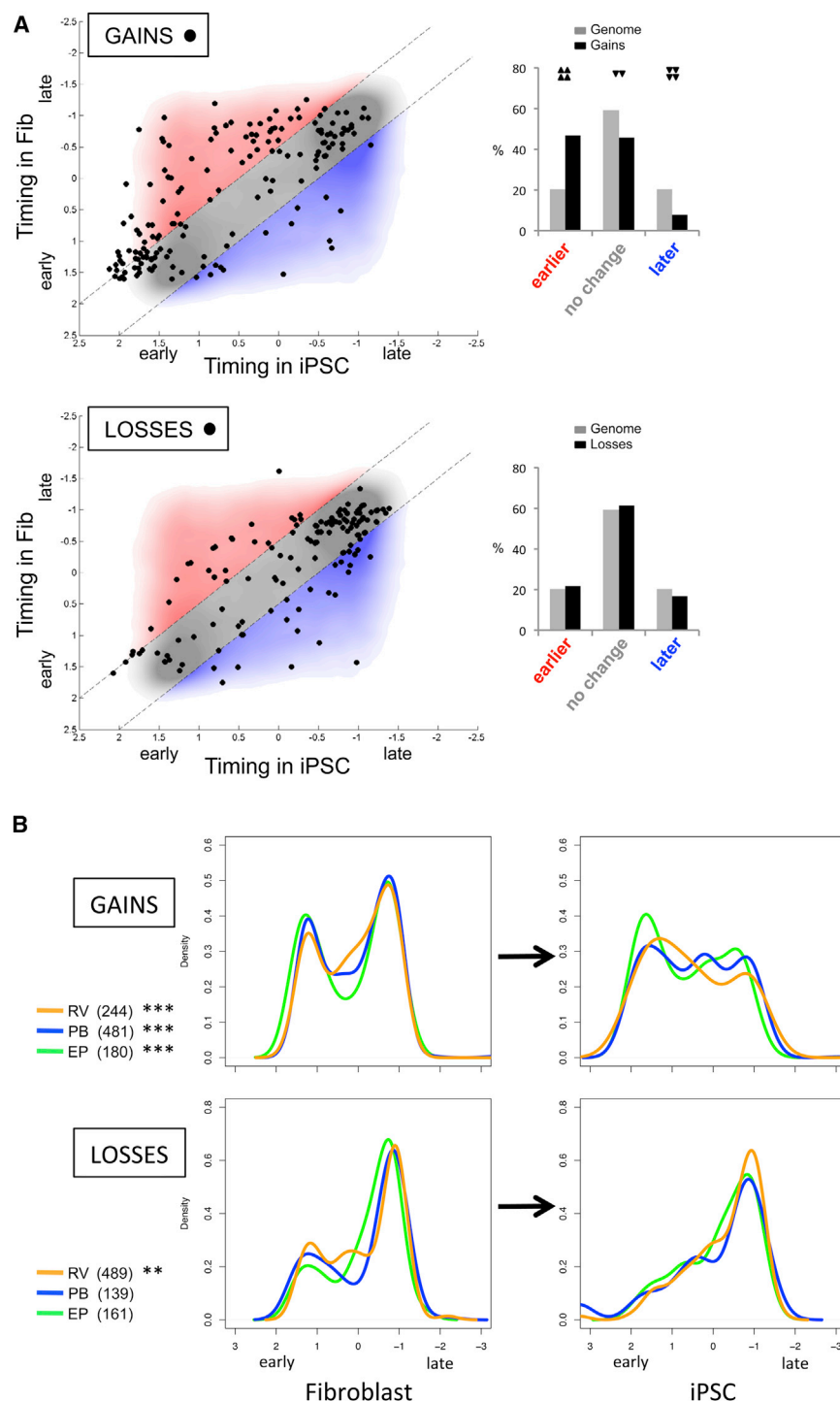


Figure 3. Nonrandom Differential CNV Distribution Is a Generalized Feature of Reprogramming

(A) Replication-timing dynamics of CNVs from EP reprogrammed iPSCs. iPSC-manifested CNVs in low-passage iPSCs are presented in the same format as in Figure 2D. Matched BJ fibroblast replication-timing profile is used. p values are shown in Figure S3A.

(B) Kernel density plots show the replication-timing distribution of CNV gains and losses detected in iPSCs derived using retroviral (purple), PB (blue), or EP (green) vectors. RV and PB iPSCs were derived from the same parental lines (Hussein et al., 2011). EP iPSCs were reprogrammed from HFF line BJ. Gains enrich in regions that change to earlier replication during reprogramming in all cases. (Numbers of CNVs in parentheses. p values from K-S tests are shown in Figure S3B.)

In sum, genome-wide analysis of iPSC CNVs using the replication-timing profiles of both fibroblasts and iPSCs demonstrates an important functional association between genome variations and replication-timing reorganization resulting from reprogramming. We have focused on CNVs and replication timing during reprogramming, but in the future, genomic aberrations identified by other techniques (e.g., whole-genome sequencing) as well as epigenomic changes can be similarly mapped to dynamic genome organization landscapes. Understanding the link between genome organization and function will be enhanced by improved biochemical techniques and computational approaches to allow higher throughput and improved resolution (Misteli, 2012; Wang et al., 2012). Because iPSC technology readily allows the manipulation of cell fate change, it will serve as a powerful experimental platform to further study the relationship between genome organization and function.

EXPERIMENTAL PROCEDURES

Generation, Characterization, and Genome-wide Analyses of iPSCs

WT fibroblasts hFib2 were reprogrammed with RVs as described by Park et al. (2008). For EP

may be a differential impact of distinct reprogramming methods on different genomic compartments and the iPSC CNV landscape. This finding warrants further investigation and is likely intertwined in the complex interplay of various other internal and external genomic stresses, the cell's DDR network, and selective pressure exerted by the cell's environment.

vector-mediated reprogramming, vectors were obtained from Addgene (pCXLE-hUL, pCXLE-hSK, and pCXLE-hOCT3/hOCT4) and introduced into BJ fibroblasts by electroporation, and reprogrammed colonies were picked and expanded. Immunofluorescent staining of pluripotency markers Oct4, Nanog, SSEA-4, Tra-1-60, and Tra-1-81 and teratoma formation assays demonstrated the pluripotency of all the established iPSC lines. Genome-wide expression profiling with Affymetrix Human Genome U133 Plus 2.0

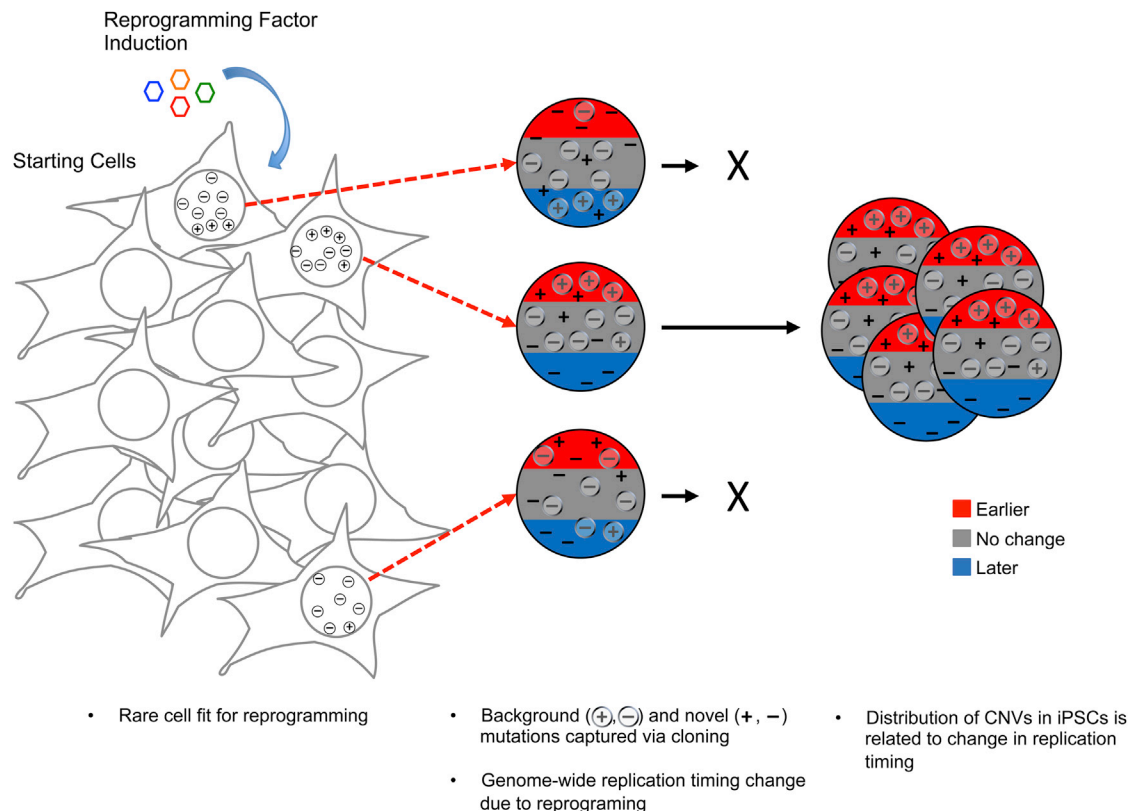


Figure 4. The Distribution of Genomic Variations in Human iPSCs Is Related to Replication-Timing Reorganization during Reprogramming

Individual cells in the starting population have a unique set of preexisting background CNVs (CNV gains are indicated with a circled +; CNV losses are indicated with a circled -), which can be captured by the clonal nature of reprogramming. iPSC-manifested mutations may also lead to de novo CNVs that occur during the reprogramming process (noncircled +, -). Preexisting and de novo mutations may contribute to reprogramming fitness. The extensive genomic and epigenomic changes associated with reprogramming can in part be reflected by changes in replication timing (gray indicates no change during reprogramming, red shows earlier in iPSCs, and blue indicates later in iPSCs). Early and late replicating domains in iPSCs are differentially affected by copy-number gains and losses, during reprogramming. Specifically, CNV gains accumulate in regions of the genome that change to early replication during the reprogramming process. In contrast, CNV losses tend to be late replicating in both fibroblasts and iPSCs. This specific CNV distribution, like individual mutations, contributes to reprogramming fitness.

microarrays and CNV analysis with Affymetrix SNP 6.0 arrays were performed according to manufacturer's instructions at the Genotyping and Microarray Center of Coriell Institute for Medical Research. Replication-timing profiling was performed and analyzed as in Ryba et al. (2011).

Computational Analyses

To statistically analyze if CNVs fall within replication-timing domains (Figure 1B), we performed permutation analysis. For each simulation, we generated a new list of CNVs by assuming uniform distribution of CNV start positions along chromosomes (omitting centromeric regions) while keeping their size distribution. For each new list, we mapped to the replication domain profile and obtained the percentage of new CNVs falling within domains. We repeated the permutation 1,000 times. The permutation p value is defined as the number of permutations yielding a higher percentage than the actual experimental result. In Figures 2 and 3, the Kolmogorov-Smirnov (K-S) test was used to detect differences between distributions. Binomial tests were used to assess the statistical significance of CNV enrichment in the genomic regions. The null hypothesis is random distribution, i.e., no CNV gain or loss enrichment in the regions of interest.

ACCESSION NUMBERS

The GEO accession number for the replication-timing microarray, SNP array, and gene expression array reported in this paper is GSE46848.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2014.03.007>.

ACKNOWLEDGMENTS

We thank R. Zhao, S. Agarwal, P. Cahan, and S. Sunyaev for critical reading and discussion of the manuscript and T. Ryba for computational help. This work was supported by NICHD grant R00HD061981 to P.H.L., NIH grant 5R01GM080625 to J.S.L., and R21CA161666 and P01GM085354 to D.M.G. J.J.C. and H.L. are supported by the Howard Hughes Medical Institute, SysCODE (Systems-based Consortium for Organ Design & Engineering), and NIH grant RL1DE019021. H.L. is supported by Mayo Clinic Center for Individualized Medicine and Mayo Clinic Center for Regenerative Medicine.

Received: June 10, 2013

Revised: February 18, 2014

Accepted: March 5, 2014

Published: March 27, 2014

REFERENCES

- Abyzov, A., Mariani, J., Palejev, D., Zhang, Y., Haney, M.S., Tomasini, L., Ferrandino, A.F., Rosenberg Belmaker, L.A., Szekely, A., Wilson, M., et al. (2012). Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* **492**, 438–442.
- Amps, K., Andrews, P.W., Anyfantis, G., Armstrong, L., Avery, S., Baharvand, H., Baker, J., Baker, D., Munoz, M.B., Beil, S., et al.; International Stem Cell Initiative (2011). Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. *Nat. Biotechnol.* **29**, 1132–1144.
- Apostolou, E., Ferrari, F., Walsh, R.M., Bar-Nur, O., Stadtfeld, M., Cheloufi, S., Stuart, H.T., Polo, J.M., Ohsumi, T.K., Borowsky, M.L., et al. (2013). Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. *Cell Stem Cell* **12**, 699–712.
- Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. *Nat. Struct. Mol. Biol.* **20**, 290–299.
- Chang, D.J., and Cimprich, K.A. (2009). DNA damage tolerance: when it's OK to make mistakes. *Nat. Chem. Biol.* **5**, 82–90.
- Cheng, L., Hansen, N.F., Zhao, L., Du, Y., Zou, C., Donovan, F.X., Chou, B.K., Zhou, G., Li, S., Dowey, S.N., et al.; NISC Comparative Sequencing Program (2012). Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. *Cell Stem Cell* **10**, 337–344.
- De, S., and Michor, F. (2011). DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. *Nat. Biotechnol.* **29**, 1103–1108.
- 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.
- Fudenberg, G., Getz, G., Meyerson, M., and Mirny, L.A. (2011). High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. *Nat. Biotechnol.* **29**, 1109–1113.
- Gilbert, D.M., Takebayashi, S.I., Ryba, T., Lu, J., Pope, B.D., Wilson, K.A., and Hiratani, I. (2010). Space and time in the nucleus: developmental control of replication timing and chromosome architecture. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 143–153.
- Goodarzi, A.A., Noon, A.T., Deckbar, D., Ziv, Y., Shiloh, Y., Löbrich, M., and Jeggo, P.A. (2008). ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Mol. Cell* **31**, 167–177.
- Gore, A., Li, Z., Fung, H.L., Young, J.E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M.A., Kiskinis, E., et al. (2011). Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63–67.
- 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.
- Helmrich, A., Ballarino, M., Nudler, E., and Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. *Nat. Struct. Mol. Biol.* **20**, 412–418.
- Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.W., Lyou, Y., Townes, T.M., Schübeler, D., and Gilbert, D.M. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, e245.
- Hiratani, I., Ryba, T., Itoh, M., Rathjen, J., Kulik, M., Papp, B., Fussner, E., Bazett-Jones, D.P., Plath, K., Dalton, S., et al. (2010). Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome Res.* **20**, 155–169.
- Hu, M., Deng, K., Qin, Z., Dixon, J., Selvaraj, S., Fang, J., Ren, B., and Liu, J.S. (2013). Bayesian inference of spatial organizations of chromosomes. *PLoS Comput. Biol.* **9**, e1002893.
- Hussein, S.M., Batada, N.N., Vuoristo, S., Ching, R.W., Autio, R., Närvä, E., Ng, S., Sourour, M., Hämmäläinen, R., Olsson, C., et al. (2011). Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58–62.
- Koren, A., Polak, P., Nemesh, J., Michaelson, J.J., Sebat, J., Sunyaev, S.R., and McCarroll, S.A. (2012). Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am. J. Hum. Genet.* **91**, 1033–1040.
- Labib, K., and Hodgson, B. (2007). Replication fork barriers: pausing for a break or stalling for time? *EMBO Rep.* **8**, 346–353.
- Laurent, L.C., Ulitsky, I., Slavin, I., Tran, H., Schork, A., Morey, R., Lynch, C., Harness, J.V., Lee, S., Barrero, M.J., et al. (2011). Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. *Cell Stem Cell* **8**, 106–118.
- Leonhardt, H., Rahn, H.P., Weinzierl, P., Sporbert, A., Cremer, T., Zink, D., and Cardoso, M.C. (2000). Dynamics of DNA replication factories in living cells. *J. Cell Biol.* **149**, 271–280.
- Liang, G., and Zhang, Y. (2013). Genetic and epigenetic variations in iPSCs: potential causes and implications for application. *Cell Stem Cell* **13**, 149–159.
- Liu, L., De, S., and Michor, F. (2013). DNA replication timing and higher-order nuclear organization determine single-nucleotide substitution patterns in cancer genomes. *Nat. Commun.* **4**, 1502.
- MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L., and Scherer, S.W. (2014). The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* **42**, D986–D992.
- Martins-Taylor, K., Nisler, B.S., Taapken, S.M., Compton, T., Crandall, L., Montgomery, K.D., Lalande, M., and Xu, R.H. (2011). Recurrent copy number variations in human induced pluripotent stem cells. *Nat. Biotechnol.* **29**, 488–491.
- Mayshar, Y., Ben-David, U., Lavon, N., Biancotti, J.C., Yakir, B., Clark, A.T., Plath, K., Lowry, W.E., and Benvenisty, N. (2010). Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* **7**, 521–531.
- Misteli, T. (2012). Parallel genome universes. *Nat. Biotechnol.* **30**, 55–56.
- Murga, M., Jaco, I., Fan, Y., Soria, R., Martinez-Pastor, B., Cuadrado, M., Yang, S.M., Blasco, M.A., Skoultschi, A.I., and Fernandez-Capetillo, O. (2007). Global chromatin compaction limits the strength of the DNA damage response. *J. Cell Biol.* **178**, 1101–1108.
- Orkin, S.H., and Hochedlinger, K. (2011). Chromatin connections to pluripotency and cellular reprogramming. *Cell* **145**, 835–850.
- Park, I.H., Lerou, P.H., Zhao, R., Huo, H., and Daley, G.Q. (2008). Generation of human-induced pluripotent stem cells. *Nat. Protoc.* **3**, 1180–1186.
- Phillips-Cremens, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295.
- Quinlan, A.R., Boland, M.J., Leibowitz, M.L., Shumilina, S., Pehrson, S.M., Baldwin, K.K., and Hall, I.M. (2011). Genome sequencing of mouse induced pluripotent stem cells reveals retroelement stability and infrequent DNA rearrangement during reprogramming. *Cell Stem Cell* **9**, 366–373.
- Rübe, C.E., Lorat, Y., Schuler, N., Schanz, S., Wennemuth, G., and Rübe, C. (2011). DNA repair in the context of chromatin: new molecular insights by the nanoscale detection of DNA repair complexes using transmission electron microscopy. *DNA Repair (Amst.)* **10**, 427–437.
- 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.
- Ryba, T., Battaglia, D., Pope, B.D., Hiratani, I., and Gilbert, D.M. (2011). Genome-scale analysis of replication timing: from bench to bioinformatics. *Nat. Protoc.* **6**, 870–895.
- Ryba, T., Battaglia, D., Chang, B.H., Shirley, J.W., Buckley, Q., Pope, B.D., Devidas, M., Druker, B.J., and Gilbert, D.M. (2012). Abnormal developmental

- control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome Res.* **22**, 1833–1844.
- Schuster-Böckler, B., and Lehner, B. (2012). Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507.
- Schwaiger, M., Stadler, M.B., Bell, O., Kohler, H., Oakeley, E.J., and Schübeler, D. (2009). Chromatin state marks cell-type- and gender-specific replication of the *Drosophila* genome. *Genes Dev.* **23**, 589–601.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872.
- Wang, P., Zhang, W., Yang, J., Qu, J., and Liu, G.H. (2012). Higher-order genomic organization in pluripotent stem cells. *Protein Cell* **3**, 483–486.
- Watanabe, Y., Fujiyama, A., Ichiba, Y., Hattori, M., Yada, T., Sakaki, Y., and Ikemura, T. (2002). Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions. *Hum. Mol. Genet.* **11**, 13–21.
- Watanabe, Y., Ikemura, T., and Sugimura, H. (2004). Amplicons on human chromosome 11q are located in the early/late-switch regions of replication timing. *Genomics* **84**, 796–805.
- Watanabe, A., Yamada, Y., and Yamanaka, S. (2013). Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120292.
- Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., Wang, K., and Lu, W. (2013). Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. *Cell Stem Cell* **13**, 36–47.
- Zhang, H., Jiao, W., Sun, L., Fan, J., Chen, M., Wang, H., Xu, X., Shen, A., Li, T., Niu, B., et al. (2013). Intrachromosomal looping is required for activation of endogenous pluripotency genes during reprogramming. *Cell Stem Cell* **13**, 30–35.