1 **Incomplete Reprogramming of DNA Replication Timing in Induced Pluripotent**
2 **Stem Cells**

3

4 Matthew M. Edwards[1]; Ning Wang[2,3]; Dashiell J. Massey [1]; Dieter Egli [2,3#]; Amnon Koren [1#]

5

6 [#] corresponding authors: de2220@columbia.edu & koren@cornell.edu

7

8 [1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA
9 [2]Department of Pediatrics and Naomi Berrie Diabetes Center, Columbia University, New York, New York
10 10032, USA
11 [3]Columbia University Stem Cell Initiative, New York, New York 10032, USA

12

13

Abstract

15   Induced pluripotent stem cells (iPSC) are a widely used cell system and a foundation for
16 cell therapy. Differences in gene expression, DNA methylation, and chromatin conformation,
17 which have the potential to affect differentiation capacity, have been identified between iPSCs
18 and embryonic stem cells (ESCs). Less is known about whether DNA replication timing – a
19 process linked to both genome regulation and genome stability – is efficiently reprogrammed to
20 the embryonic state. To answer this, we profiled and compared genome-wide replication timing
21 between ESCs, iPSCs, and cells reprogrammed by somatic cell nuclear transfer (NT-ESCs). While
22 NT-ESCs replicated their DNA in a manner indistinguishable from ESCs, a subset of iPSCs exhibit
23 delayed replication at heterochromatic regions containing genes downregulated in iPSC with
24 incompletely reprogrammed DNA methylation. DNA replication delays were not the result of
25 gene expression and DNA methylation aberrations and persisted after differentiating cells to
26 neuronal precursors. Thus, DNA replication timing can be resistant to reprogramming and lead
27 to undesirable phenotypes in iPSCs, establishing it as an important genomic feature to consider
28 when evaluating iPSC lines.

29

30

Introduction

32   Reprogrammed pluripotent stem cells are widely used in research and provide the
33 foundation for cell therapy (Bragança et al., 2019; Shi et al., 2017). Stem cells derived via
34 somatic nuclear transfer to an enucleated oocyte (NT-ESCs) appear to be fully reprogrammed to
35 the embryonic state and both mouse and human NT-ESCs can be reliably differentiated to
36 specialized cell types (Kim et al., 2010; Ma et al., 2014; Sui et al., 2018). However, they are
37 difficult to derive, in part because it requires access to human oocytes and highly specialized
38 expertise. In contrast, induced pluripotent stem cells (iPSCs) can be readily derived by
39 expressing defined factors in somatic cells, however their use is complicated by substantial
40 variation in differentiation potential among iPSC lines and compared to their ESC counterparts
41 (Koyanagi-Aoi et al., 2013; Sui et al., 2018). For example, some iPSC lines fail to undergo
42 complete neural differentiation and give rise to teratomas following transplantation (Koyanagi-
43 Aoi et al., 2013). Similarly, iPSCs show highly variable and reduced ability to differentiate to

1 beta cells compared to ESC or reprogrammed NT-ESC (Sui et al., 2018). This raises important
2 questions regarding the molecular basis of these differences, as well as regarding quality
3 controls required for both research and cell therapy.
4   iPSC differentiation potential in both humans and mice is impacted by the incomplete
5 reprogramming of primary cells to the stem cell state (Kim et al., 2010; Lister et al., 2011; Ma et
6 al., 2014; Marchetto et al., 2009; Ohi et al., 2011). Studies in mice have reported that epigenetic
7 memory can cause iPSCs to maintain epigenetic hallmarks of their cells-of-origin, predisposing
8 them to differentiate back to the original cell type (Bar-Nur et al., 2011; Kim et al., 2010; Polo et
9 al., 2010). Mouse iPSCs derived from beta cells, for example, show a propensity to differentiate
10 to insulin producing cells (Bar-Nur et al., 2011), while fibroblast-derived iPSCs are resistant to
11 differentiation into non-fibroblast lineages such as haematopoietic cells (Kim et al., 2010). This
12 effect of source cell type on differentiation capacity, however, has not been observed in human
13 cells.
14   Another possible source of variation in iPSCs are aberrant epigenetic changes acquired
15 during the reprogramming process itself. These are typically sites of DNA methylation that are
16 not seen in somatic donor cells or in ESCs (Lister et al., 2011; Panopoulos et al., 2017; Ruiz et al.,
17 2012; Stadtfeld et al., 2010), and that can adversely affect developmental potential. These
18 aberrations are both non-random and generally consistent across studies (Panopoulos et al.,
19 2017). The most common form of such differentially methylated regions (DMRs) between iPSCs
20 and ESCs are short (1-11Kb) regions that impact cytosines in the context of CpG dinucleotides
21 (CG DMRs). At the majority of these DMRs, iPSCs are hypomethylated compared to ESCs. A
22 second type of DMRs involve megabase-scale non-CG sites (methylation at CH sites, where H =
23 A, T, or C) (non-CG DMRs)(Lister et al., 2011). Ultimately, both epigenetic memory and aberrant
24 changes due to reprogramming have been reported at many genomic loci (Lister et al., 2011;
25 Panopoulos et al., 2017), and deserve attention when considering epigenetic variation in iPSCs.
26   In addition to epigenetic alterations in iPSC lines, gene expression alterations have been
27 characterized, with genes such as TCERG1L, FAM19A5, and COL22A1 reported across several
28 studies to be differentially expressed between iPSCs and ESCs (Koyanagi-Aoi et al., 2013; Kyttälä
29 et al., 2016; Lister et al., 2011). Such gene expression alterations, and the general heterogeneity
30 of gene expression in iPSCs, pose challenges for using iPSCs for disease modeling (Rouhani et
31 al., 2014). Similarly, some genes, such as TCERG1L, show consistent DNA methylation
32 differences between ESCs and iPSCs (Koyanagi-Aoi et al., 2013; Lister et al., 2011). Another
33 property that has been shown to resist reprogramming to the ESC state is chromosome
34 conformation: some iPSCs derived from neuronal precursor cells (NPCs) fail to establish ESC-like
35 chromosomal conformations on a small (sub-Mb) scale (Beagan et al., 2016), although
36 topologically associating domains (TADs) were faithfully reconstructed in iPSCs (Krijger et al.,
37 2016).
38   Notwithstanding these proposed molecular differences between iPSCs and ESCs, It
39 remains unclear whether they fully account for the reduced differentiation potential of iPSCs.
40 While expression alterations between iPSCs are frequently observed, it has been consistently
41 shown that gene expression differences between iPSCs due to cell-of-origin (epigenetic
42 memory) are subtle compared to differences brought about by inter-individual genetic variation
43 (Burrows et al., 2016; Guenther et al., 2010; Kilpinen et al., 2017; Koyanagi-Aoi et al., 2013;
44 Kyttälä et al., 2016). Aside from the handful of well-described genes, differential gene

1 expression varies substantially between studies, suggesting that consistent reprogramming-
2 associated gene expression alterations may be limited. Similarly, it has been shown that DNA
3 methylation differences in iPSCs compared to ESCs, as well as between iPSCs from different
4 primary cell types, are more subtle than those resulting from interindividual genetic variation
5 (Burrows et al., 2016; Choi et al., 2015). Clustering based on either gene expression or DNA
6 methylation groups cell lines by donor (genetic background) and fails to separate iPSCs from
7 ESCs (Bock et al., 2011; Johannesson et al., 2014; Koyanagi-Aoi et al., 2013; Kyttälä et al., 2016).
8 The insufficiency of gene expression and DNA methylation aberrations in accounting for the
9 differentiation potential of iPSCs raises the question of whether other epigenetic properties
10 may underly iPSC differences from ESCs.
11 A less explored epigenetic property that must undergo genome-wide changes as
12 somatic nuclei dedifferentiate back to a stem-cell-like state is DNA replication timing. The
13 coordinated replication of the genome throughout S-phase through the firing of replication
14 origins is a highly regulated process unique to a given cell type (Hansen et al., 2010; Ryba et al.,
15 2010). However, our knowledge of the ability of cells to reprogram DNA replication timing, and
16 the differences in DNA replication timing between ESCs and reprogrammed cell types such as
17 iPSCs and NT-ESCs, remain limited.
18 Examination of DNA replication dynamics with florescent *in situ* hybridization (FISH) has
19 shown that both iPSCs and NT-ESCs fully reprogram DNA replication timing at several
20 developmentally regulated loci (such as Rex1) as well as imprinted genes (such as SNRPN and
21 IGF2R) (Shufaro et al., 2010). However, this approach is limited both in its genomic resolution
22 (specific loci) and its temporal resolution. Genome-scale analysis of iPSC reprogrammed from
23 fibroblasts has previously suggested that their DNA replication timing is "virtually
24 indistinguishable" from ESCs (Hiratani et al., 2008; Ryba et al., 2010), however this analysis was
25 done at the level of large replication "domains" and on a small number of predominantly
26 mouse cell lines. In contrast, a more recent study using a single molecule approach has
27 reported incomplete reprogramming of DNA replication timing at the FXN and NANOG loci in
28 iPSCs, but not NT-ESCs (Paniza et al., 2020), although only three loci were examined with a
29 small sample size. A high-resolution, genome-wide analysis of iPSC and NT-ESC DNA replication
30 timing compared to ESC is currently lacking, and necessary to resolve replication dynamics in
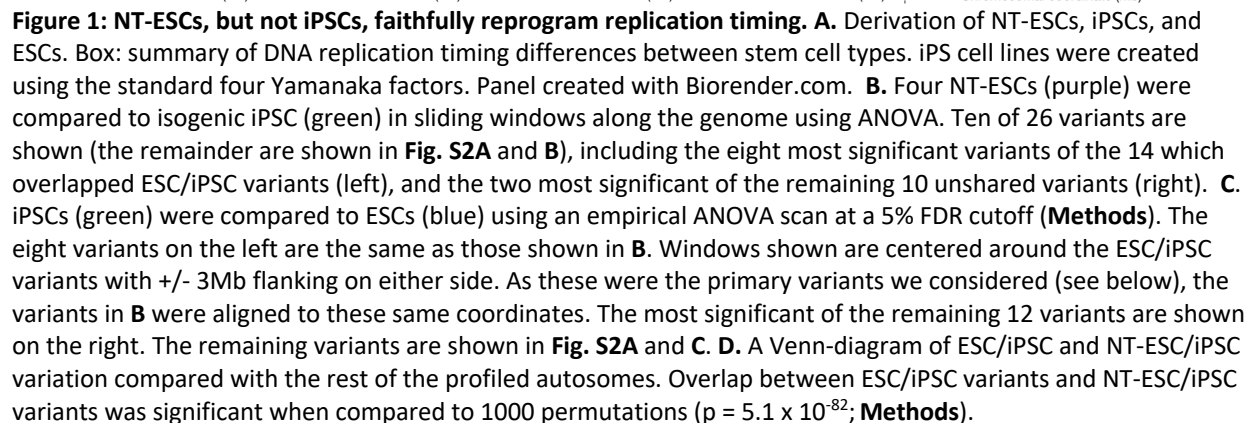31 reprogrammed stem cells.
32 Here, we generated and compared genome-wide DNA replication timing profiles of
33 multiple ESC, iPSC, and NT-ESC lines. While NT-ESCs faithfully recapitulated the DNA replication
34 timing program of ESCs, we identified 26 genomic regions with altered replication timing in a
35 subset of iPSCs, most of them replication delays. iPSC replication timing aberrations were
36 located primarily in centromere- and telomere-proximal regions and were enriched for
37 H3K9me3 and zinc finger repeats. These regions were related to previously identified DMRs and
38 genes with altered expression in iPSCs, though the replication timing alterations appear distinct
39 from these other two epigenetic properties. Finally, we demonstrate that replication timing
40 alterations persist in NPCs derived from aberrantly replicating iPSCs, which carries important
41 implications for their use in research and cell therapy.

Results

Aberrant DNA replication timing in iPSCs

To comprehensively compare the DNA replication timing program across pluripotent stem cells derived by different methods (**Fig. 1A**), we concomitantly profiled genome-wide DNA replication timing in 28 ESC lines, four NT-ESC lines, and 19 iPSC lines (Supplementary Table 1). Of the four NT-ESCs, three (NT5, NT6 and NT8) were derived from a neonatal fibroblast cell line (BJ) and one was derived from adult dermal fibroblasts (1018); two iPSCs were also derived from neonatal BJ cells and two from adult 1018 cells, allowing for isogenic NT-SC/iPSC comparison. We also profiled neuronal precursor cell lines derived from either ESCs or iPSCs to examine differentiation in different stem cell types (see further below).

Replication timing was profiled at high-resolution using a whole genome sequencing (WGS)-based approach which we have previously described (Koren et al., 2014, 2021). Briefly, earlier-replicating regions of the genome contribute more DNA in a population of proliferating cells than do later-replicating regions, which can be observed in the form of fluctuating DNA copy number along chromosomes in WGS data. Accordingly, we calculated the number of sequencing reads in uniquely-alignable genomic windows, corrected for the influence of GC content, and filtered copy number variations and outliers. The remaining DNA copy number profiles were smoothed and normalized to standard deviation units to derive the replication timing profiles of each cell line and compare them across samples. The replication timing profiles were of high quality, as evidenced by high correlations within ESCs (r = 0.96), iPSCs (r = 0.97), and NT-ESCs (r = 0.98). The replication profiles of all cell lines also resembled previously profiled ESCs (Ding et al., 2021) (mean r = 0.84), more so than they resembled a different cell type (lymphoblastoid cell lines, mean r = 0.71). One iPSC line was removed based on correlation to other iPSCs and principal component analysis (**Fig. S1A,B**).

Figure 1: NT-ESCs, but not iPSCs, faithfully reprogram replication timing. A. Derivation of NT-ESCs, iPSCs, and ESCs. Box: summary of DNA replication timing differences between stem cell types. iPS cell lines were created using the standard four Yamanaka factors. Panel created with Biorender.com. **B.** Four NT-ESCs (purple) were compared to isogenic iPSC (green) in sliding windows along the genome using ANOVA. Ten of 26 variants are shown (the remainder are shown in **Fig. S2A** and **B**), including the eight most significant variants of the 14 which overlapped ESC/iPSC variants (left), and the two most significant of the remaining 10 unshared variants (right). **C.** iPSCs (green) were compared to ESCs (blue) using an empirical ANOVA scan at a 5% FDR cutoff (**Methods**). The eight variants on the left are the same as those shown in **B**. Windows shown are centered around the ESC/iPSC variants with +/- 3Mb flanking on either side. As these were the primary variants we considered (see below), the variants in **B** were aligned to these same coordinates. The most significant of the remaining 12 variants are shown on the right. The remaining variants are shown in **Fig. S2A** and **C**. **D.** A Venn-diagram of ESC/iPSC and NT-ESC/iPSC variation compared with the rest of the profiled autosomes. Overlap between ESC/iPSC variants and NT-ESC/iPSC variants was significant when compared to 1000 permutations (p = 5.1 x 10^{-82}; **Methods**).

1   Correlations between stem cell types (ESC/NT-ESC: r = 0.97; ESC/iPSC: r = 0.96 ; NT-
2 ESC/iPSC: r = 0.97) were comparable to those within stem cell types, indicating that globally,
3 replication timing is consistent between reprogrammed and embryonic stem cells. Nonetheless,
4 we observed notable differences in the replication profiles at the local scale. To quantify this
5 systematically, we first compared NT-ESCs and their isogenic iPSC counterparts to identify
6 genomic regions where replication timing differed between reprogramming methods. We used
7 an approach we employed previously (Edwards et al., 2021), where candidate regions were first
8 identified using ANOVA tests in approximately 200Kb-wide overlapping windows, and then
9 were filtered to remove regions unlikely to contain true variation (see **Methods**). We thus
10 identified 24 regions of replication timing variation (**Fig. 1B; Fig. S2A, B**).
11   To further understand the source of this variation, we compared cell lines derived by
12 both reprogramming methods to ESCs. Because of the larger number of samples available for
13 this analysis, we used permutations to empirically set a p-value threshold corresponding to a
14 5% false discovery rate (FDR; **Methods**). This identified 26 ESC/iPSC variants, covering a total of
15 2.63% of the autosomes (**Fig. 1C; Fig. S2A, C**), at a p-value threshold of 1×10^{-53}. We found that
16 these ESC/iPSC variants had a significant and considerable degree of overlap with NT-ESC/iPSC
17 variation (14/26, or 53.8% of variants overlapped, p = 5.2×10^{-82}; **Fig. 1D**). We also directly
18 tested NT-ESC/iPSC variants using all 18 iPSCs (not just isogenic lines) and the genome-wide
19 Bonferroni-corrected cutoff from the filtering-based ANOVA scan, and confirmed that 15/24
20 (62.5%) variants were identified in this larger set (**Fig. S3A**). In contrast, the ESCs/NT-ESC
21 comparison did not reveal any variation at a 5% FDR. A down-sampling analysis confirmed that
22 the smaller sample size when examining NT-ESCs vs isogenic iPSCs did not bias these results
23 (**Methods**). Consistent with iPSCs being the outlier stem cell type, the 18 iPSCs showed lower
24 correlation to NT-ESCs at variant regions (r = 0.93) compared to non-variant regions (r= 0.99),
25 while ESCs were highly correlated (r = 0.99) to NT-ESCs at both variant and non-variant regions
26 (**Fig. S3B**). Given the high correspondence between NT-ESCs and ESCs, and the larger number of
27 ESC samples, we focused our subsequent analyses on replication timing differences between
28 ESCs and iPSCs.
29   Notably, 22/26 (84.6%) of the ESC/iPSC variants consisted of replication delays in iPSCs,
30 with mean iPSC replication timing later than the means of both ESCs and NT-ESCs (**Fig. 1B-C,**
31 **Fig. S2D**). Replication was already late in ESCs in these regions, with an average replication
32 timing of 0.73 SD units later than the mean, while iPSCs replicated even later, at an average of
33 0.91 SD units later than the mean (**Fig. S2E**). Many of these regions showed prominent
34 replication timing peaks in ESCs, indictive of replication origins or origin clusters. These peaks
35 appeared diminished or even lost in some iPSCs (see for examples variants on chromosome 10,
36 12, and the strongest variant on chromosome 8 in **Fig. 1C**), suggesting that replication origin
37 firing may be inhibited in iPSCs.
38   To further validate the ESC/iPSC replication timing variants that we identified, we turned
39 to a larger, independent dataset of replication timing in 108 ESCs and 300 iPSCs (Ding et al.,
40 2021). We were able to test 17 of the 26 variants in this dataset (**Methods**) and found that the
41 iPSCs in the larger dataset showed the same direction of variation (delayed/advanced) in
42 accordance with expectation at 14 of 17 variants (82%). Furthermore, nine of the 17 testable
43 variants (53%) showed similar replication timing alterations as those initially observed in the
44 smaller dataset (**Fig. S4; Methods**). We did not attempt to identify new variants using these

1 larger sample sizes because of concerns of batch effects, as the two cell types were sequenced
2 in separate projects.
3   Taken together, these results suggest that NT-ESCs closely resemble ESCs in their
4 replication dynamics, while iPSCs show replication timing alterations, largely consisting of
5 delays, at specific genomic regions when compared to either type of ESCs.
6
7 **Replication timing delays at TCERG1L**
8   One of the most significant variant region we identified in both the ESC/iPSC and
9 NT/iPSC comparisons (Chr10:131,222,871-133,380,323; **Fig. 1B-C**) contained the gene
10 TCERG1L, identified in several previous studies as having differential methylation and decreased
11 gene expression in iPSCs compared to ESCs (Lister et al., 2011; Nishino et al., 2011; Ruiz et al.,
12 2012). Specifically, iPSCs replicated later than ESCs at this locus, consistent with decreased gene
13 expression. This is in contradiction with the results of Single Molecule Analysis of Replicated
14 DNA (SMARD), which suggested similar replication dynamics between ESCs and iPSCs at
15 TCERG1L (Paniza et al., 2020). SMARD also suggested incomplete replication timing
16 reprogramming of the FXN and NANOG loci, however neither locus showed significant variation
17 in our data between the two pluripotent cell types (p = 0.77 and 0.045 respectively, compared
18 to p = 1 x 10^{-53} required to achieve genome-wide significance as detailed above; **Fig. S5A**).
19 Similar results were observed when testing the exact cell lines used in the previous study (the
20 four NT-ESCs and their isogenic iPSCs; **Fig. S5B**). SMARD also reported that iPSCs derived from
21 adult cells (1018A/E) have lower initiation frequency at the FXN locus than those derived from
22 neonatal cells (BJiPSM/BJiPSO), however we did not observe this either (**Fig. S5B**). We ascribe
23 these inconsistencies to the different methodologies and the limited genomic scale of analysis,
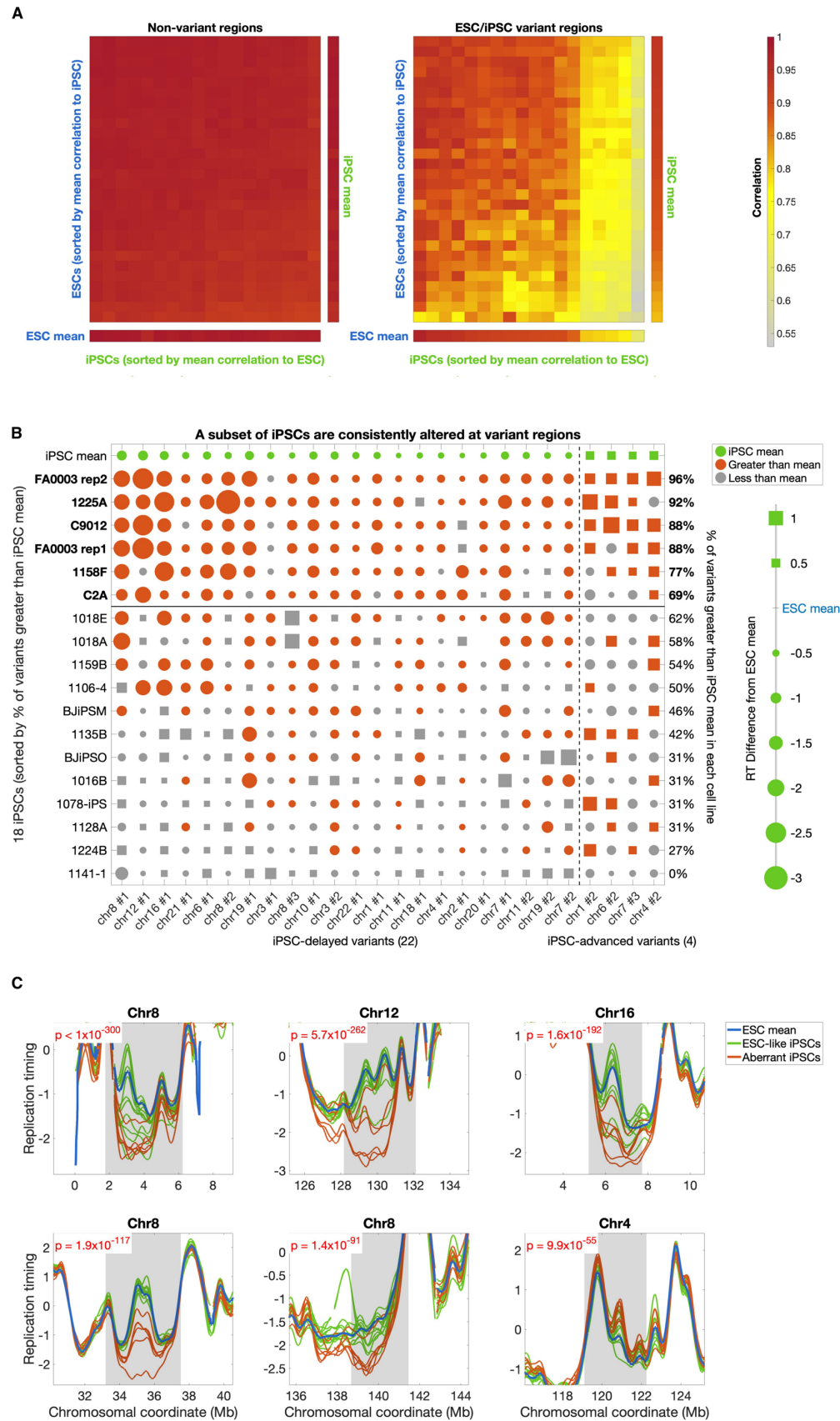24 as well as the limited number of samples and fibers used by SMARD.
25
26 **Replication timing alterations identify a subset of aberrant iPSC lines**
27   While replication was altered in iPSCs, there appeared to be substantial heterogeneity in
28 replication timing alterations among individual iPSC lines. Specifically, in several variant regions
29 some iPSC lines resembled ESC replication timing while others had delayed replication. To
30 quantify this heterogeneity, we compared replication timing for each ESCs and iPSCs cell line at
31 both variant and non-variant regions. We observed high correspondence in replication timing at
32 non-variant regions for both iPSCs (correlation to the ESC mean ranged from r = 0.96 to 0.99)
33 and ESCs (correlation to the iPSC mean ranged from r = 0.95 to 0.99). At variant regions, iPSCs
34 and ESCs line showed slightly lower correspondence, but were nonetheless still similar to each
35 other (median r = 0.91 and 0.90 to the ESC and iPSC means, respectively). However, iPSCs
36 exhibited much greater variability (correlation to the ESC mean ranged from r = 0.66 to 0.97)
37 compared to ESCs (correlation to the iPSC mean ranged from r = 0.81 to 0.94; **Fig. 2A**),
38 suggesting heterogeneity among iPSCs at variant regions.
39   To determine if this heterogeneity originated from across iPSC lines or resulted from a
40 subset of them, we quantified the extent of replication timing aberration in each individual iPSC
41 line, at the region of maximum variation (maximum absolute replication timing difference
42 between each iPSC and the ESC mean) in each variant (**Fig. 2B**). This separated the iPSC lines
43 into two groups: those that resembled the replication timing of ESCs, and others that were
44 altered at nearly every variant. Accordingly, we identified the iPSCs that were more aberrant

1   than the iPSC mean at each variant (more delayed at iPSC-delayed variants, more advanced at
2   iPSC-advanced variants), and defined those that were consistently more aberrant in at least 18
3   of 26 variants (>= 69%; binomial p = 0.038) as "aberrant iPSCs". The six cell lines thus identified
4   had similar replication timing alterations at each variant region (**Fig. 2C)**. A genome-wide scan
5   for replication timing variation (empirical ANOVA scan, 5% FDR) between the remaining 12
6   iPSCs and ESCs did not identify any significant variation, indicating that variation was driven
7   primarily by the aberrant iPSCs. Accordingly, we term the remaining iPSCs as "ESC-like".
8   We also observed heterogeneity among individual iPSC (but not ESC) lines in the larger
9   dataset of 300 iPSC and 108 ESCs that we used above for validation (**Fig. S6A**). Once again, we
10 were able to identify two opposing subsets of iPSC lines based on the correlation of each cell
11 line's replication profile to ESCs and the maximum replication timing alteration (across the
12 seven iPSC-delayed regions validated in this dataset; **Fig. S4**) relative to ESCs. Using a cutoff of
13 the 10% most outlying cell lines in both of the above criteria, we defined 16 aberrant iPSCs and
14 14 ESC-like iPSCs (**Fig. S6B**). When comparing replication timing between these two groups,
15 there were clear differences at variant regions, with the ESC-like iPSCs largely resembling ESCs
16 and the aberrant cell lines showing replication delays (**Fig. S6C**).

1　**Figure 2**. **Heterogenous replication delays in iPSCs. A**. Correlation of replication timing between ESCs (rows) and
2　iPSCs (columns) at non-variant regions (left) and at ESC/iPSC variant regions (right). **B.** Replication timing difference
3　from the ESC mean for each individual iPSC line. Values are shown as squares for replication timing earlier than the
4　ESC mean, and circles for replication timing later than the mean; the size of each marker reflects the extent of
5　advance/delay at the region of maximum variation (scale to the right). Replication timing is shown for each of the
6　26 variant regions, first separated into iPSC delayed (left) and iPSC advanced (right), and then ordered from left to
7　right by significance (based on the p-values from **Fig. 1C** and **S2A,C**; variants are labeled by chromosome first and
8　then numbered in increasing p-value order; i.e., chr8 #2 is the second lowest p-value variant on chr8). The iPSC
9　mean (green) is shown in the top row, and values for each cell line are indicated as more variant than the mean
10　(orange; e.g., later than the mean at a delayed variant), or less variant (grey; e.g., earlier than the mean at a
11　delayed variant). Cell lines are ordered by the percentage of regions which are more variant than the mean (values
12　on the right). The six cell lines which were more variant than expected by chance (binomial test, see text) are
13　indicated in bold and classified as aberrant iPSCs, while the remaining twelve were classified as ESC-like. **C.** Six
14　example variant regions, with the six aberrant iPSCs shown in orange. The first five variants were iPSC-delayed,
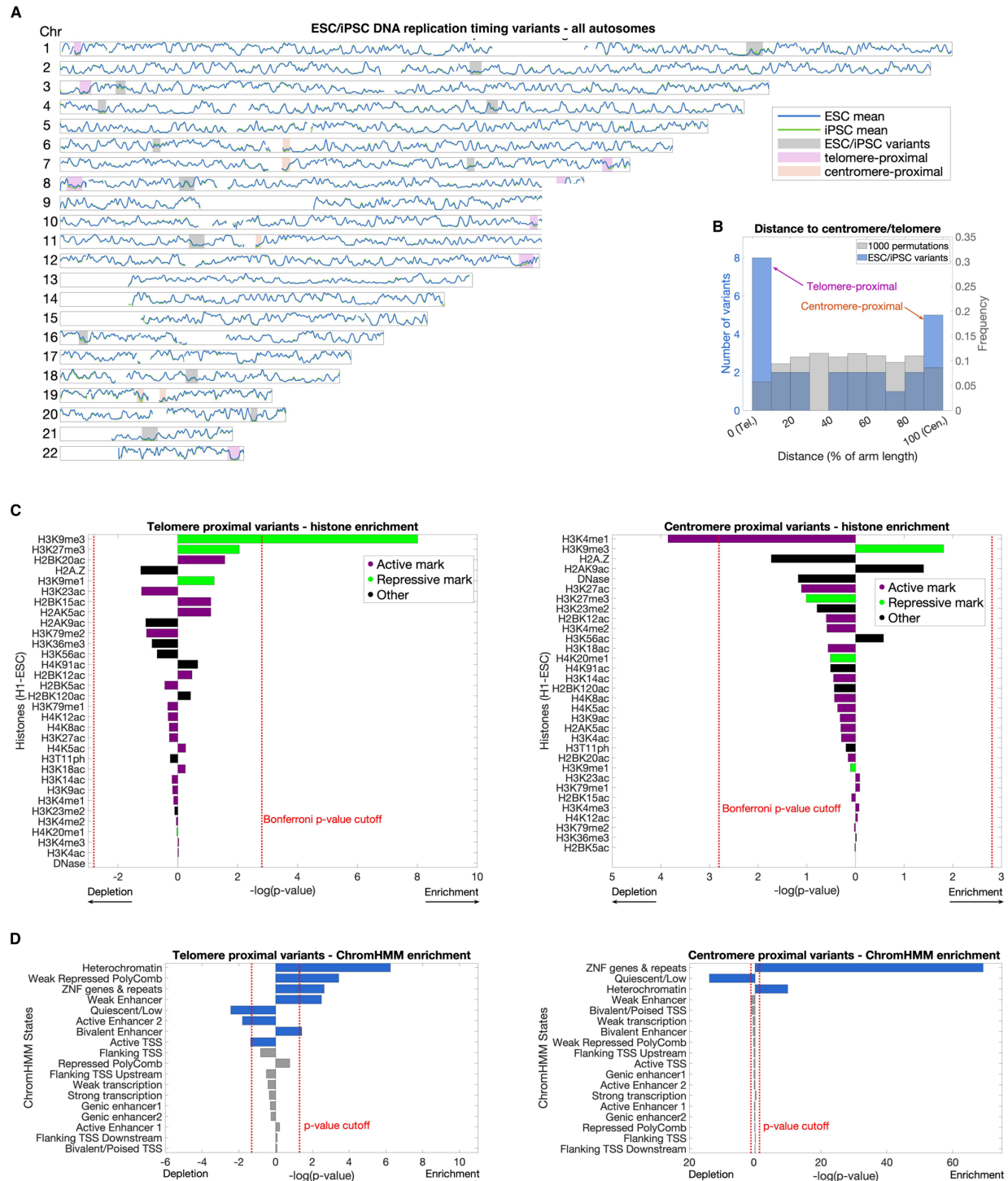15　while the bottom right variant is iPSC-advanced.

Figure 3: Replication timing variants are located in heterochromatic regions near centromeres and telomeres. A. Whole genome plot of the mean ESC and iPSC replication timing profiles. Variant regions are shaded, and variants within 10% of chromosome arm length of telomeres (purple) and centromeres (orange) are indicated. **B.** Histogram of the relative locations along the chromosome arm of the 26 variants (blue; left y scale) compared with 1000 random permutations of variant region locations across the genome (grey; right y scale). Distance is calculated between the middle of each variant and the nearest telomere and normalized by the length of the chromosome arm. Variants within 10% of telomeres or centromeres were considered as centromere- or telomere-

1 proximal, respectively. **C.** Analysis of histones in telomere-proximal variants (left) and centromere-proximal
2 variants (right) based on ChIP-seq data (ENCODE). Bars indicate enrichment/depletion two-sided t-test p-value of
3 each chromatin mark (purple: active marks, green: repressive marks) in variants compared to 1000 replication
4 timing matched permutations. Each chromatin mark was considered to be independent, and the Bonferroni
5 corrected p-value threshold (.05/32) is shown in red. Telomere-proximal variants are significantly enriched for the
6 histone mark H3K9me3, while centromere-proximal variants are nominally enriched for H3K9me3 and significantly
7 depleted of H3K4me1. **D.** Same as **C** for ChromHMM states in telomere-proximal variants (left) and centromere
8 proximal variants (right). Because these states are by definition mutually exclusive, we considered a cutoff of p =
9 0.05 for significance (red line). ChromHMM state counts were defined by the number of base pairs belonging to
10 each state in variant regions and permutations. Both sets of variants were enriched for heterochromatin and
11 depleted for quiescent regions.

12

13 **iPSC replication timing variants are associated with repressed genes and heterochromatin**

14   Interestingly, 31% (8/26) of ESC/iPSC variants were located in the 10% of chromosome
15 arms closest to telomeres (telomere-proximal), while another 19% (5/26) were located in the
16 10% closest to centromeres (centromere-proximal) (**Fig. 3A,B**). Since both centromere- and
17 telomere-proximal regions are heterochromatic (Benetti et al., 2007; Gonzalo et al., 2006;
18 Saksouk et al., 2015), we separately characterized epigenetic features of those variants
19 proximal to centromeres or telomeres, as well as the remaining variants. We tested for
20 enrichment of 31 histone marks as well as DNase hypersensitivity (Hoffman et al., 2013) and
21 chromatin states defined by ChromHMM (Ernst & Kellis, 2017). The histone mark H3K9me3, an
22 established marker of constitutive heterochromatin, was greatly enriched specifically in
23 telomere-proximal variants (**Fig. 3C**), while H3K27me3, a marker of facultative
24 heterochromatin, was not. Centromere-proximal variants were also significantly depleted of
25 the active mark H3K4me1 (**Fig. 3C**). The remaining variant regions that were not near
26 centromeres or telomeres did not show any histone mark enrichment (**Fig. S7A**). The
27 ChromHMM state "heterochromatin" was found to be enriched in all three classes of variants
28 (**Fig. 3D, Fig. S7B**), while there was a substantial enrichment of zinc finger (ZNF) genes &
29 repeats in centromere-proximal regions (**Fig. 3D**). Although the statistical enrichment of ZNF
30 genes & repeats was largely due to two ZNF clusters at the two variants on either side of the
31 chr19 centromere (**Fig. 1C and S2A**), this chromatin state was found in all five centromere-
32 proximal variants.
33   There were 202 genes located within the variant regions, in line with what is expected
34 for similarly late-replicating regions (p=0.64 when compared to 1000 permutations; **Methods**).
35 Gene ontology analysis revealed a single enriched category, "*cis*-regulatory region sequence
36 specific binding", which was predominantly attributed to a cluster of zinc finger protein-
37 encoding genes at a single variant (3.23-fold enrichment, FDR = 0.0282, 15 of 18 genes located
38 at chr19:21,405,159-23,345,945). Among these genes were FAM19A5 (TAFA5), FZD10,
39 TMEM132D, COL22A1, and TCERG1L (**Fig. 4C**), which have all been shown to be differentially
40 expressed in iPSCs (Kyttälä et al., 2016; Liang & Zhang, 2013; Lister et al., 2011; Ruiz et al.,
41 2012).

42

43 **Replication timing variation in iPSCs is related to DNA methylation aberrations**

44   It has previously been shown that DNA methylation is altered in iPSCs compared to ESCs
45 due to both incomplete reprogramming to the stem cell state at certain genomic regions, and
46 aberrant methylation at other regions (Lister et al., 2011). We thus asked whether replication

1   timing variants are associated with regions of differential methylation. Lister et al. identified
2   1,175 short CG-DMRs between ESCs and iPSCs, among which 11 were consistently
3   hypermethylated in all iPSCs and 119 were consistently hypomethylated (the remainder were
4   not consistent across tested cell lines). In addition to these short CG-DMR, Lister et al. identified
5   29 long non-CG DMRs, seven of which were hypermethylated in iPSCs relative to ESCs, while
6   the remaining 22 were hypomethylated and considered not fully epigenetically reprogrammed.
7   Interestingly, non-CG DMRs were enriched near centromeres and telomeres, similar to
8   replication timing variants.
9   Sixty-four (5.45%) of the CG-DMRs overlapped our ESC/iPSC replication timing variant
10   regions, significantly more than expected by chance (p = 6.2 x 10^{-18}; **Fig. 4A,C**). Notably among
11   those, seven of the 11 (63.6%) hypermethylated CG-DMRs overlapped replication timing
12   variants, a strong enrichment compared to chance (p = 1.9 x 10^{-31}; **Fig. 4A,C**), while there was
13   no enrichment for the more abundant hypomethylated CG-DMRs in variant regions (p = 0.47;
14   **Fig. 4A**). In contrast to CG-DMRs, hypomethylated non-CG DMRs showed a strong
15   correspondence to the ESC/iPSC variants (p = 3.7 x 10^{-24}; **Fig. 4B,C**), while hypermethylated non-
16   CG DMRs did not overlap with replication timing variation (p = 0.54; **Fig. 4B**). Specifically, of 19
17   non-CG DMR hypomethylated regions that we were able to test (successfully lifted over to
18   hg19), nine (47%) at least partially overlapped nine of our 26 variant regions (**Fig. 4B-C**), all of
19   which were iPSC-delayed. Moreover, directly testing these regions revealed that 14/19 (73.7%)
20   DMRs (including eight that overlapped ESC/iPSC variant regions) showed ESC/iPSC replication
21   timing variation that was significant at a genome-wide Bonferroni-corrected p-value threshold
22   of 8.63 x 10^{-7} (**Fig. S8**; note that six of these were not identified as replication timing variants in
23   our original scan because the individual constituent windows were not significantly different).
24   Taken together, replication timing differences between iPSCs and ESCs correspond to aberrant
25   methylation in iPSCs involving both short, hypermethylated CG DMRs and long,
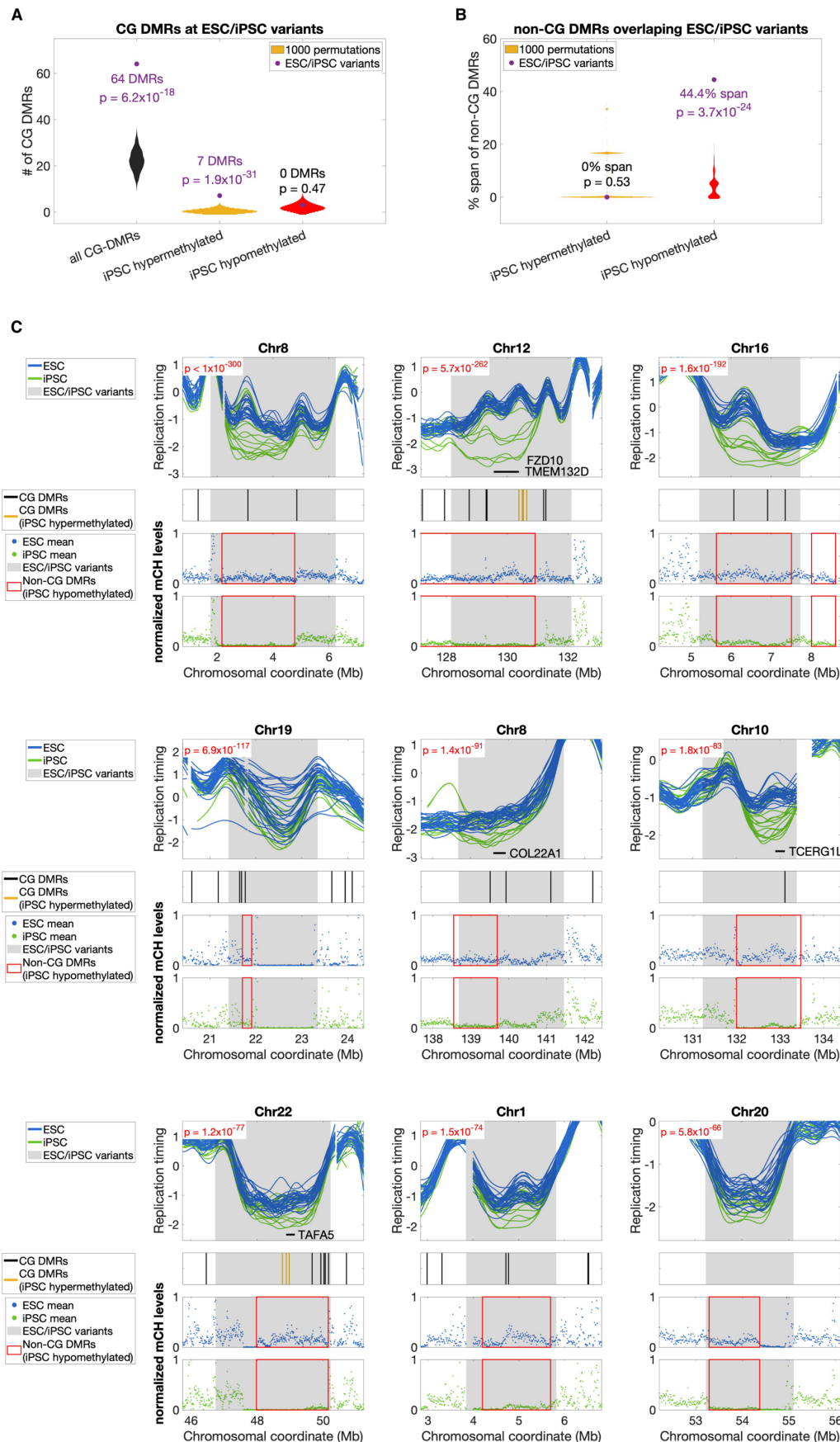26   hypomethylated non-CG DMRs.

1

Figure 4: Regions of iPSC replication timing variation correspond to differential CG and non-CG DNA methylation. A. Distributions of CG DMRs in 1000 permutations of ESC/iPSC variant regions (Violin plots; see **Methods**). Distributions were smoothed using the ksdensity MATLAB function with robust kernel estimation (via the DistributionPlot function). The total number of the CG DMRs in variant regions is represented as a purple dot; two-tailed p-value calculated from z-score compared to the distribution is indicated (purple font indicates a significant p-value). **B.** Similar to **A**, for regions of non-CG hypomethylation in iPSCs. Because all non-CG DMRs were either hypermethylated or hypomethylated (unlike in **A** where most DMRs were unclassified), only these two sets of DMRs were compared to variant regions. Additionally, due to the larger size of these non-CG DMRs, overlap was calculated as cumulative percentage of each replication timing variant spanned by the DMR (see **Methods**). Because the iPSC hypermethylated non-CG DMRs (orange) were on average smaller than replication timing variants (mean sizes of 526Kb and 2.92Mb, respectively), the distribution of overlaps follows several discrete modes. The hypomethylated non-CG DMRs (red; mean size of 1.38Mb) were larger, and thus the distribution of overlaps was smoother. **C.** The nine ESC/iPSC variant regions that overlapped a non-CG iPSC hypomethylated regions are shown with corresponding differentially expressed genes (DEGs) and DNA methylation data. iPSC CG DMRs are shown as vertical black lines (and are enriched in variant regions, see **A**), while those consistently hypermethylated in iPSCs are indicated in orange. Non-CG methylation levels (ranging from 0 to 1 after normalization) are plotted for the mean of two ESCs (blue) and the mean of five iPSCs (green); red boxes: non-CG DMRs hypomethylated in iPSCs (see **Methods** for methylation data processing)

Aberrant replication timing in iPSCs persists after neuronal differentiation

The replication timing alterations we identified suggest that some iPSCs may be incompletely reprogrammed or have newly acquired epigenetic aberrations; this observation is important for evaluating the utility of such iPSC lines, for instance as models for cellular differentiation. To test whether iPSC replication timing alterations persist after differentiation, we differentiated three iPSC lines that were identified as aberrant and three that were ESC-like, as well as eleven ESC lines, to NPCs (**Fig. S1C**), and profiled replication timing together with the source stem cell lines. Two cell lines derived from ESC, and one from an ESC-like iPSC were removed based on correlation to other NPCs and principal component analysis (**Fig. S1A,B**).

As expected for different cell types, undifferentiated and differentiated cells were less correlated with each other (r = 0.82) compared to the correlations among cell lines within each differentiation state (ESC: r = 0.96, iPSC: r = 0.97, NPC: r = 0.96) or compared to the correlations between ESCs and iPSCs (r = 0.96; **Fig. S1A**).

Correlations among ESC-derived NPCs (r = 0.97), among iPSC-derived NPCs (r = 0.95) and between the two groups of NPCs (r = 0.96) were similar, suggesting that the type of stem cells used to derive NPCs has little impact on the global replication timing program. ESC-derived and iPSC-derived NPCs also appeared similar visually at the chromosomal scale (**Fig. 5A**). In contrast, at ESC/iPSC replication timing variant regions, we also observed variation in those NPCs derived from aberrant iPSCs (herein referred to as aberrant-derived NPCs; **Fig. 5A,B**). Systematic testing of replication timing variant regions using ANOVA revealed that aberrant-derived NPCs indeed showed variation compared to ESC-derived NPCs, which was both significant at a genome-wide Bonferroni-corrected significance threshold and in the same direction as the ESC/iPSC variation. Aberrant iPSCs were variant at 16/26 (58%) regions (**Fig. 5B**), an enrichment compared to permutations (9.7 x 10^{-7}). We also called variation between ESC-derived and iPSC-derived NPCs using an unbiased genome-wide ANOVA scan, and found 27 regions of variation, spanning 3.2% of the autosomes, at a 5% FDR. Of these, nine (33.3%) overlapped nine ESC/iPSC variants (**Fig. 5B**; p = 2.6 x 10^{-12} when compared to 1000 permutations; **Methods**), suggesting that at least some of the replication timing alterations in iPSCs are maintained through differentiation. In

1 certain cases, e.g., on chromosomes 12 and 16 (**Fig. 5**), replication delays appeared to become
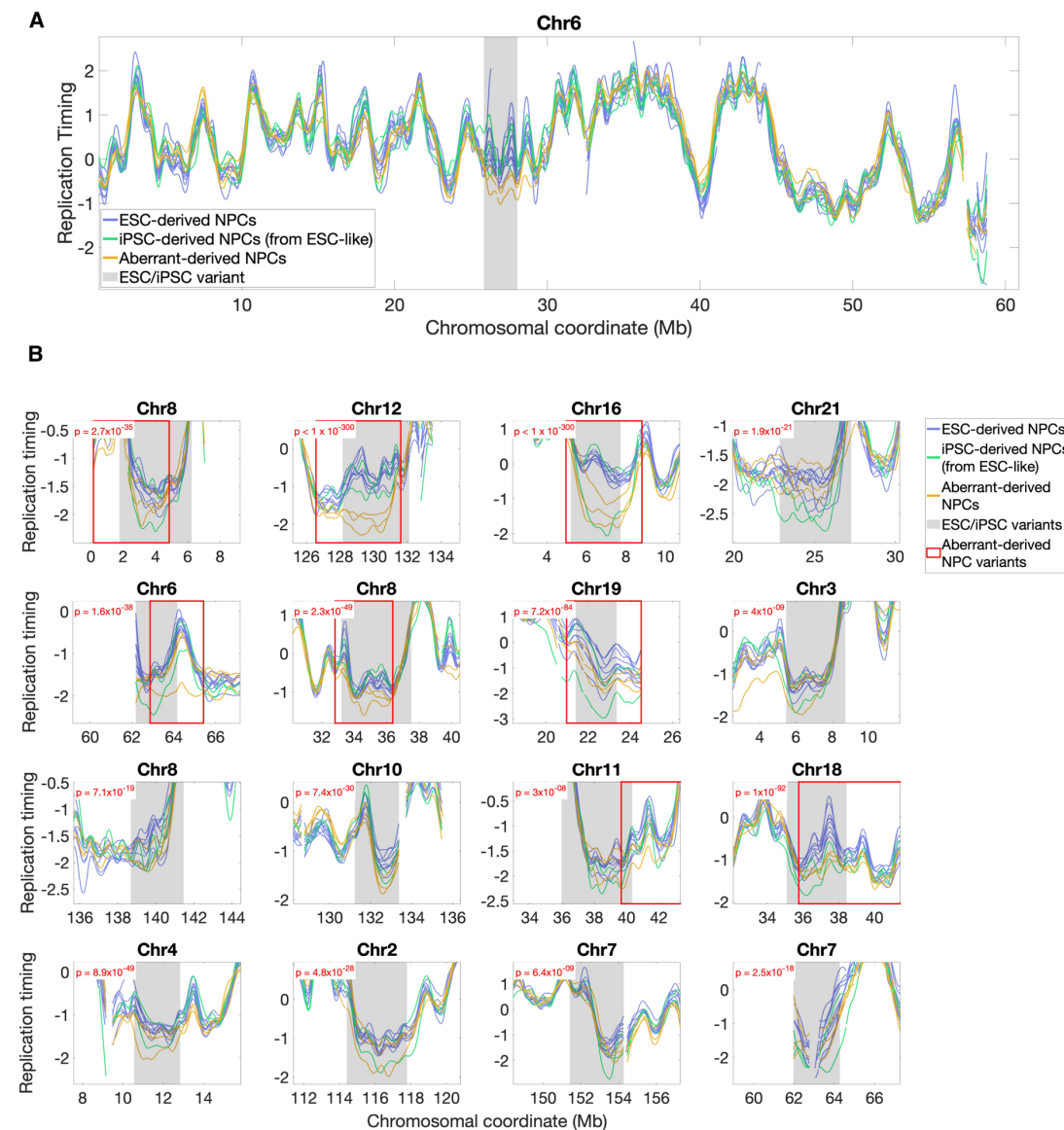2 even more extreme after NPC differentiation.
3
4

5
6 **Figure 5: Replication timing aberrations persist after differentiation to NPCs. A.** Replication timing profiles for
7 ESC-derived NPCs (blue), ESC-like iPSC-derived NPCs (green), and aberrant-derived NPCs (orange) on the p arm of
8 chromosome 6. An iPSC/ESC replication timing variant region (grey shaded area) is maintained through
9 differentiation. **B.** Replication timing profiles for NPCs at the sixteen regions where aberrant-derived iPSCs were
10 significantly different from ESC-derived NPCs at the genome-wide Bonferroni threshold ($p = 9.58 \times 10^{-7}$). Significant
11 regions were also required to show the same direction of effect (delayed in all validated cases) as the ESC/iPSC
12 variant (this removed two regions of variation).

Discussion

Reprogrammed stem cells provide a necessary alternative to ESCs due to limitations on the ability to generate new ESC lines and their ability to be genetically and immunologically matched to individual patients. However, molecular variations involving genetic and/or epigenetic aberrations can greatly impact the use of iPSCs and exacerbate phenotypic variation including cells' ability to differentiate to somatic cells of interest. Therefore, epigenetic deviations of reprogrammed cells compared to ESCs as well as genetic change compared to the somatic cell donor, raise concerns for both research and therapeutics. Here, we report that NT-ESCs provide a faithful recapitulation of the ESC DNA replication timing program. In contrast, iPSCs are prone to replication defects. We identify a total of 26 genomic regions, covering 2.63% of the autosomes, where a subset of iPSCs consistently deviate in their replication timing compared with ESCs. We validated these regions by comparing isogenic iPSCs and NT-ESCs, by comparing a larger set of iPSCs to NT-ESC, and by examining a yet larger dataset encompassing hundreds of ESC and iPSC cell lines. Evaluating replication dynamics across many cell lines controlled for natural and technical variation between samples, while a genome-wide analysis that considers the chromosomal vicinity of each locus controlled for any potential localized noise at specific sites; in these respects, our analysis is superior to previous methodologies and studies.

Given the tendency of ESC/iPSC variants to be near telomeres and centromeres, it is likely that we are missing additional variants in unmapped genomic regions including centromeres and telomeres proper. We also identified six additional variants when testing non-CG DMRs, suggesting that there may be additional variant regions we did not identify using our stringent genome-wide criteria. Similarly, there were 18 regions of variation specifically observed between aberrant-derived NPCs and ESC-derived NPCs, which could represent undiscovered ESC/iPSC variants which were only significant after differentiation, and/or regions at which replication timing variation emerged after differentiation.

Delayed DNA replication timing in iPSCs and their differentiated derivatives has important implications for their use, as it could lead to increase in mutation rates in the affected regions, to aberrant cell cycle progression, or possibly to alterations in genome regulation (Koren, 2014). Indeed, iPSCs are known to be prone to acquisition of somatic mutations, copy number alterations, and chromosomal rearrangements (Ben-David et al., 2011; Hussein et al., 2011; Lu et al., 2014; Pasi et al., 2011; Robinton & Daley, 2012), possibly due in part to delayed replication timing.

Replication timing defects in iPSCs associate with, but are not fully explained by gene expression and DNA methylation alterations

iPSCs are known to exhibit altered gene expression, DNA methylation, and chromatin conformation compared to ESCs. Here, we have shown that iPSCs are also prone to alterations in the DNA replication timing program, which co-occur with, but are distinct from, these other epigenomic alterations. While several downregulated genes are located in regions of delayed replication in iPSCs, replication timing alterations are concentrated in late-replicating, gene-poor regions of the genome. These regions tend to be heterochromatic, enriched for H3K9me3, and are often localized near centromeres and telomeres. The larger number (at least 26) of replication timing variants compared to gene expression changes (only a handful consistent

1 among studies) between ESCs and iPSCs suggests that DNA replication timing alterations are
2 more abundant, span a greater fraction of the genome, and are at least partially independent of
3 gene expression alterations. This raises important questions regarding the mechanism causing
4 these differences, in particular the possibility of aberrant *de novo* changes that are neither
5 present in the donor cells nor in embryonic stem cells.
6   Compared to gene expression changes, differences in DNA methylation in iPSCs relative
7 to ESCs were more abundant at replication timing variant loci. Replication timing alterations
8 coincided with both short CG-DMRs (particularly iPSC hypermethylation) and longer regions of
9 non-CG hypomethylation in iPSCs. Previous studies have suggested that DNA replication timing
10 is influenced by DNA methylation in some contexts but independent of it in other contexts (Du
11 et al., 2019, 2021; Gribnau et al., 2003; Takebayashi et al., 2021). Consistently, we identified
12 cases of replication timing alterations in iPSCs that occurred independently of DNA methylation
13 aberrations, including near centromeres and telomeres.
14
15 **Cellular sources of altered DNA replication timing**
16   Epigenetic alterations in iPSC compared to ESCs have been shown to arise from either
17 epigenetic memory of the cells from which the iPSCs were derived (Bar-Nur et al., 2011; Kim et
18 al., 2010; Polo et al., 2010), or alterations due to the reprogramming process (Lister et al.,
19 2011). Similarly, abnormal replication timing in iPSCs could be due to a failure to sufficiently
20 reprogram somatic cell (fibroblast in this case) replication timing, or aberrations due to
21 reprogramming itself. Understanding whether replication defects are a remnant from
22 fibroblasts, or a product of reprogramming could potentially be resolved by studying DNA
23 replication timing in fibroblasts. However, fibroblasts proliferate significantly slower than stem
24 cells and are not amenable to the WGS-based replication profiling utilized here. An alternative
25 approach is to use single cell DNA sequencing data, for instance those available for the
26 fibroblast cell line BJ (isogenic to NT5/6/8, BJiPSM, and BJiPSO; 10xgenomics.com), together
27 with our recently published approach for inferring DNA replication timing from proliferating
28 single cells (**Methods**; Massey & Koren, 2022). Such an analysis suggested that at four
29 replication timing variants, iPSC replication resembled fibroblasts, possibly reflecting epigenetic
30 memory of the source cell type; while at two other variants, fibroblasts resembled ESCs but not
31 iPSCs, more consistent with a possible reprogramming defect in iPSCs (**Fig. S9**). The remaining
32 variants were not informative, as fibroblasts did not resemble either stem cell type (see
33 **Methods**; although this may itself be interpreted as refuting epigenetic memory). It is
34 important to note that we cannot rule out coincidental similarity in either case given the small
35 number of examples and the methodological differences. Thus, further research will be
36 required in order to resolve the cellular sources of altered DNA replication timing in iPSC lines.
37
38 **Replication timing as a screen for differentiation potential**
39   We have shown that reprogramming defects occur in a heterogeneous manner across
40 cell lines, with variation being confined to a subset of iPSC lines. This parallels commonly
41 observed heterogeneity across iPSCs in gene expression and DNA methylation, as well as
42 phenotypes such as growth rate and cellular morphology (Volpato & Webber, 2020). There
43 have been many efforts to evaluate the quality of individual iPSC lines. In mouse iPSCs, chimera
44 formation can be used to determine the ability of cells to form adult tissue. Ultimately, the

1   most stringent test is tetraploid complementation, where cells must lead to development of a
2   whole organism (Robinton & Daley, 2012). Complete reprogramming to reconstitute an entire
3   organism has been shown through the cloning of numerous animals in different species
4   (Gurdon & Byrne, 2004), consistent with the high developmental potential of human NT-ESCs.
5   In contrast, only a few studies reported full developmental competence of mouse iPSC lines
6   reprogrammed from adult cells in tetraploid embryo complementation assays (Stadtfeld et al.,
7   2012). Whether these differences between nuclear transfer ES cells and iPSCs are due to a
8   selection for developmental potential during early development, or due to other mechanisms,
9   is not known. Specific challenges arise in the human system, as stringent developmental assays
10   are not available for human iPSCs. Instead, criteria for selection of human iPSCs must rely on *in*
11   *vitro* differentiation potential and detailed molecular characterization. However, assays for
12   gene expression and DNA methylation largely fail to separate ESCs and iPSCs (Bock et al., 2011;
13   Johannesson et al., 2014; Koyanagi-Aoi et al., 2013; Kyttälä et al., 2016), suggesting there is not
14   an iPSC-specific signature of these properties. While some genes and DMRs can serve as
15   markers for reduced iPSC differentiation potential, they represent a small number of loci
16   (Koyanagi-Aoi et al., 2013). Moreover, even signatures of epigenetic memory for iPSCs derived
17   from a given cell type are inconsistent across studies (Robinton & Daley, 2012), leaving a need
18   for a more consistent and readily attainable molecular signature. To integrate multiple
19   molecular measurements, the Pluritest (Müller et al., 2011) and a "score card" for deviation
20   from ES cells (Bock et al., 2011) bioinformatic tools have been developed, however they only
21   measure pluripotency and not differentiation capacity (Liu & Zheng, 2019) and rely on labor-
22   and cost-intensive data collection that may not be easily scalable. Furthermore, these methods
23   often disagree with one another (Allison et al., 2018; Bouma et al., 2017).
24   While further work is needed to examine the links between replication timing and stem
25   cell pluripotency, we suggest that replication timing could potentially be used as a screen for
26   iPSC quality in parallel to other quality controls and without the obligatory need for RNA-seq or
27   bisulfite analysis. Our results suggest that reproducible replication timing aberrations are
28   broader than gene expression and DNA methylation signatures in iPSCs. Furthermore, DNA
29   replication timing profiling can be performed using the same data (WGS) used for genotyping
30   and identification of *de novo* mutations, making it a compelling molecular signature at minimal
31   added costs for identifying low quality iPSCs prior to use in research or cell therapy.
32
33

34   **Methods**
35
36   **Stem cell derivation and culture**
37   All experiments with human pluripotent stem cells were reviewed and approved by the
38   Columbia Institutional Review Board (IRB) and the Columbia University Embryonic Stem Cell
39   Committee. Human pluripotent stem cell lines were cultured on geltrex-coated plates in
40   StemFlex Medium (A3349401, Gibco) to 80-90% confluence and passaged with TrypLE
41   (12605036; Life Technologies). iPS cell lines were generated using mRNA mediated
42   reprogramming (Johannesson et al., 2014). The cell line 1000-5 (which was excluded from the
43   analysis) was generated using Sendai viral vectors. Previous studies had not observed significant

1   differences between iPSCs generated using different non-integrating reprogramming methods
2   (Schlaeger et al., 2015).

3

4   **NPC differentiation**
5   Cells were differentiated to NPCs from human pluripotent stem cell lines using a
6   published protocol with modifications (Wang et al., 2016). Briefly, human pluripotent stem cells
7   were cultured in 6-well geltrex-coated dishes with StemFlex medium. When each well reached
8   95-100% confluence, StemFlex medium was replaced with KSR differentiation medium; this was
9   considered day 0 of differentiation. The KSR differentiation medium was changed daily for 4
10   days. From day 5 to 7, cells were incubated with KSR differentiation medium mixed with N2
11   differentiation medium at ratios of 3:1 (day 5), 1:1 (day 6), and 1:3 (day 7). On day 8, detection
12   of EdU staining and PAX6 expression by FACS was performed (**Fig. S1C**), and cell pellets were
13   collected for whole genome sequencing.
14   KSR differentiation medium consisted of knockout DMEM (10829-018, Gibco)
15   supplemented with 15% KSR (10828-028, Gibco), 1% GlutaMAX (35050061, Thermo Fisher), 1%
16   NEAA (11140-050, Gibco), 0.1 mM 2-Mercaptoethanol (21985-023, Gibco), 10,000 U/ml
17   penicillin–streptomycin (15070-063, Thermo Fisher), 10 µM SB 431542 (S1067, Selleckchem),
18   and 2.5 µM LDN 193189 (S2618; Selleckchem). N2 differentiation medium was based on
19   DMEM/F12 (10565042, Thermo Fisher) by supplementing with 1% N2 supplement (17502-048,
20   Thermo Fisher), 1% GlutaMAX, 1% NEAA, 0.4 N ascorbic acid (A-5960, Sigma), 1% D-glucose
21   with 16% (w/v) (G8270, Sigma) 10,000 U/ml penicillin–streptomycin, 10 µM SB431542 (S1067,
22   Selleckchem), and 2.5 µM LDN193189 (S2618; Selleckchem).

23

24   **FACS**
25   EdU staining was performed by Click-iT^{TM} EdU Alexa Fluor 488 Imaging Kit (C10337),
26   samples were washed with cold PBS and centrifuged for 5 min at 1000 rpm at 4 °C. Samples
27   were incubated with PAX6 antibody (901302, Biolegend; **Fig. S1C**) at room temperature (RT) for
28   30 min, followed by 3 washes with PBST (3% BSA in PBS with 0.1% Triton) and a secondary
29   Alexa Fluor 647 donkey anti-rabbit IgG (H+L) (A31573, Thermo Fisher) for another 30 min at RT.
30   The cell suspension was analyzed on a BD Fortessa flow cytometer, and data was analyzed using
31   FlowJo v.10.

32

33   **Whole genome sequencing**
34   Samples were sequenced in two batches: batch 1 included 26 of the 28 ESCs and 17 of
35   the 19 iPSCs, and batch 2 included the remaining two ESCs and two iPSCs, and all 17 NPCs. For
36   both batches, DNA was extracted with the MasterPure DNA purification kit (Lucigen). Libraries
37   were prepared using the Illumina TruSeq PCR-free library preparation kit, and sequencing was
38   performed on the Illumina HiSeq X Ten with 150-bp paired-end reads at GeneWiz (South
39   Plainfield, NJ). Reads were then aligned to the hg19 human reference genome using BWA-MEM
40   (Li & Durbin, 2010).

41

42   **Replication timing profiles**
43   DNA replication timing was inferred as described in (Edwards et al., 2021). Genome
44   STRiP (Handsaker et al., 2015; Koren et al., 2014) was used to infer DNA replication timing

1 across the genome. Sequence read depth was calculated in 2.5-kb windows along the genome,
2 corrected for alignability and GC content. Copy number values were normalized to an average
3 DNA copy number of two. These copy number values were then filtered as follows:

4   1. Windows spanning gaps in the reference genome were removed.

5   2. Windows with copy number greater than one above or below the median
6   copy number were removed.

7   3. In order to remove extreme data points, the data was segmented using the
8   MATLAB function *segment*, which groups consecutive data points into
9   segments based on a tolerance threshold. This analysis was done twice using
10   two different segmentations parameters of 0.5 (less strict) and 0.1 (more
11   strict). By using two different parameters, both shorter and larger genomic
12   regions that deviate from the median can be captured. Segments falling
13   above or below the median by a threshold of 0.45 copies were removed.

14   4. Genomic regions that were further than 30 kb from other data points and
15   that were <300 kb long were removed.

16   5. Regions shorter than 100 kb between removed data points were removed.

17   6. Regions shorter than 500 kb between three or more removed data points
18   were removed.

19 Data were then smoothed using the MATLAB function *csaps* with smoothing parameter of
20 10^{-17}, and then normalized to an autosomal median of 0 and a standard deviation of 1, such
21 that positive values represent early replication and negative values represent late replication.
22 After the above filtering steps, the iPSC cell line BJiPSM still showed a reduced copy number on
23 a portion of the p arm of chromosome 18 (chr18:1-10063413) and an elevated copy number on
24 the entire q arm of chromosome 9 (chr9: 70938816-141135844); these regions are suspected to
25 be subclonal aneuploidies and were removed from further analysis.
26   The mean correlation of replication timing in ESC and iPSC samples from the second
27 batch (with the exception of cell line 1000-5 that was removed due to low correlations with
28 other cell lines and being an outlier by PCA) with the first batch was r=0.96, equal to the
29 internal correlation of batch 1 (r=0.96); this suggests little if any batch effects. PCA analysis also
30 clustered all ESC and iPSC samples together, as well as NPCs together (PC1 = 89.1% explained,
31 PC2 = 5.7% explained), with the only clear visual outliers being the removed samples: 1000-5, as
32 well as the three NPCs labeled in **Fig. S1A**.
33
34 **Replication timing variation identified using filtering-based ANOVA-scan**
35   To call replication timing variation between NT-ESCs and isogenic iPSCs, we used a
36 modified version of the ANOVA-based scan with filtering that we used previously (Edwards et
37 al., 2021). Briefly, regions of 76 replication timing windows (covering 190kb of uniquely
38 alignable sequence) were tested, with a slide of a quarter region (such that each genomic locus

1   was tested four times) along the entire genome using a one-way ANOVA test between the

2   relevant groups (e.g., NT-ESCs versus isogenic iPSCs). A Bonferroni corrected p-value cutoff of

3     $p < 9.58 \times 10^{-7}$ (0.05/52,184 regions tested) was employed, and regions that were overlapping

4   or separated by short (< 750kb) gaps were merged. Because there were more samples than in

5   our previous work (Edwards et al., 2021), the filtering step was modified to require that

6   pairwise comparisons of NT-ESC and iPSC profiles showed a consistent direction of effect

7   (delayed or advanced relative to the mean of the other group) in at least 3 samples (75%) for

8   each group (i.e., at least 3 NT-ESC were required to be earlier/later than the mean of iPSCs, and

9   conversely iPSCs were reciprocally later/earlier with respect to the NT-ESC mean). As with the

10   original scan, windows where profiles overlapped were removed, and only those regions

11   greater than 76 windows in size were maintained. Finally, to capture the full region of variation,

12   all variant regions were extended bidirectionally as long as the two sample groups remained

13   separated and the ANOVA p-value was still below the significance threshold. This resulted in

14   the final set of 24 NT-ESC/iPSC variants and identified no variants in the ESC/NT-ESC

15   comparison.

16

17   **Down-sampling analysis for filtering-based ANOVA-scan**

18   There were 24 replication timing variants between the four NT-ESCs and four isogenic

19   iPSCs, while a comparison of the four NT-ESCs with all 28 ESCs (using the more appropriate

20   empirical scan for larger samples) did not identify any variation. To test that this was not a

21   function of the number of cell lines, or the scan used, the following down-sampling analysis was

22   performed:

23   Four randomly selected samples from each stem cell group (ESC, NT-ESC, iPSC) were

24   compared using the initial filtering-based ANOVA scan. This was performed 100 times for each

25   comparison, and the mean variation for each comparison was calculated. The mean extent of

26   variation called between NT-ESCs/down-sampled iPSCs (0.79% autosomal span) was identical to

27   that identified between NT-ESCs/isogenic iPSCs (0.79% autosomal span; **Fig. 1B**), indicating that

28   the full set of iPSCs exhibited similar variation to the isogenic subset. Down-sampled ESC/NT-

29   ESC mean variation was lower (0.4%), consistent with the lack of variation seen when

30   comparing the full set of ESCs to NT-ESCs. Additionally, intra-cell-type comparisons of ESCs (4

31   ESCs v 4 ESCs) and iPSCs (4 iPSC vs 4 iPSC) both revealed 0.38% variation, indicating that

32   ESC/NT-ESC variation is akin to comparisons between cells of the same type. Finally, the mean

33   down-sampled ESC/iPSC variation was greater than all other down-sampled comparisons

34   (1.2%), consistent with the greater variation we identified between these samples. Thus,

35   regardless of the scan methodology used (the filtering approached used here, or the empirical

36   approach used to identify ESC/iPSC variants) and the sample size, we still observe minimal

37   ESC/NT-ESC variation, and notable iPSC/ESC variation.

38

39   **Replication timing variation identified using empirical ANOVA-scan**

40   In order to identify replication timing variation using a larger number of samples

41   (ESC/NT-ESC and ESC/iPSC comparisons), a permutation-based approach was employed using

42   the genome-wide ANOVA scan to establish an empirical p-value cutoff. Because of the

43   heterogeneity of the iPSC samples, permuting ESC and iPSC samples together would result in

44   high levels of background variation, and therefore only ESC samples were permuted to quantify

1 the expected level of variation in cell lines without cell type differences. One thousand
2 permutations of the 28 ESC samples were performed, comparing a random selection of 14
3 samples to the remaining 14 in each permutation using the ANOVA scan (see above description
4 of filtering based scan, and (Edwards et al., 2021) for a detailed description; the filtering steps
5 for direction of effect and profile overlap were excluded from this empirical analysis). One
6 additional modification was made, wherein initial windows were tested with the one-way
7 ANOVA, and then corrected with a q-value approach (Storey, 2002). Adjacent windows with q-
8 values <0.05 were merged, and then still subjected to a one-way ANOVA which was required to
9 pass the genome-wide Bonferroni-corrected threshold. By using this multistep approach, large
10 regions with subtle variation that may have escape the initial cutoff in 190kb windows, but
11 which may be significant, can be identified.
12   The final p-value cutoff was determined by an empirically determined threshold at a
13 False discovery rate (FDR) < 0.05. FDR was calculated by dividing the total variation found (in
14 Mb) by the average variation across 1000 permutations. This was then performed at every p-
15 value cutoff (lowering by 10^{-1} each time) until the false discovery rate of 0.05 was reached at p
16 = 1×10^{-53}. Regions in the true comparison under this threshold were then extended in both
17 directions until profiles overlapped or the p-value dropped below the threshold to form the
18 final variant list.
19   To ensure that this estimate of false variation was not influenced by the sample size
20 differences (14 vs 14 in permutations compared to 18 vs 28 in ESC/iPSC comparisons), a
21 permutation analysis was also performed on the 108 validation ESCs. Nine of the 108 ESCs were
22 previously shown to have autosomal replication delays linked to putative X reactivation
23 (Edwards et al., 2021) and were excluded from this analysis. A hundred permutations were
24 performed comparing 14 vs 14 of these ESCs, and separately 100 permutations comparing 18
25 and 28 ESCs. At a cutoff of p = 1×10^{-53}, the 18 vs 28 comparison resulted in an FDR of 0.054,
26 while the 14 vs 14 comparison resulted in FDR = 0.024. Thus, while the larger sample size may
27 result in slightly more background variation, false discovery rates near 0.05 are nonetheless
28 observed at this p-value cutoff even with larger numbers of samples.
29   NT-ESCs were also compared to the full set of 17 iPSCs using the empirical approach.
30 While this did not produce any variation at 5% FDR (likely due to a high degree of false positives
31 resulting from iPSC variability, see **Fig. 2**), the full set of iPSCs directly validated the variants
32 found using the isogenic samples (**Fig. S2**).
33
34 **ESC/iPSC variant validation**
35   Replication timing profiles (corrected using principal component analysis) for the 108
36 ESCs and 300 iPSCs were obtained from Ding et al. 2021. Because ESC profiles were generated
37 using 10kb replication timing windows, while iPSC profiles were generated using 2.5kb profiles,
38 the iPSC profiles were interpolated to the ESC coordinated prior to analysis. In order to test
39 whether iPSCs exhibited variation in these regions compared to ESCs in the validation set, we
40 first ensured that these regions were consistent between datasets by calculating the mean
41 Pearson correlations between the validation ESCs and the original ESCs for each variant and all
42 non-variant regions. In non-variant regions the average correlation was r = 0.83. Testable
43 regions were defined as those where ESC/ESC correlation was r > 0.70, leaving 15 of the 22
44 iPSC-delayed and two of four iPSC-advanced variants for testing.

Twelve of the 15 iPSC-delayed variants and both iPSC-advanced variants tested showed the expected direction of effect (delayed/advanced) across 50% or more of the variant region. Suspected batch effects precluded testing for significant variation using ANOVA. However, these regions were tested by asking whether replication timing difference between the iPSC mean and ESC mean in each region was more extreme than flanking regions of equal size to the variant. Flanking regions were required to contain at least half as many data points as the variant region, i.e., be less than 50% spanned by a reference genome gap. Variants where the maximum replication timing difference (iPSC vs. ESC) was more extreme (more positive for advanced regions or more negative for delayed regions) were considered validated.

Permutation methodology

In order to determine the significance of overlap between replication timing variants and various other genomic features, replication timing variant locations were permuted 1000 times, and significance was determined by comparing the overlap between variants and the tested feature of interest to permutations.

For generating permuted variant regions, we required the following:

1. Each permutation consisted of a number of permuted windows equal to the number of variants, and each permuted window was the same size as the variant from which it was derived.

2. Replication timing in the middle of the permuted windows was required to be within ± 0.2 SD of the variant from which it was derived.

3. Permuted regions could not overlap variants or each other. Permuted regions were ordered by the p-value of the corresponding variant, so the most significant variant was permuted first, and later regions within this same permutation could not overlap the regions that were determined prior.

4. Permuted regions could not overlap gaps in the reference genome.

5. Regions had to have data in at least 50% of the replication timing bins.

Overlap was determined in one of two ways. For comparisons with features of large (Mb-scale) size (comparisons of ESC/iPSC variants to NT-ESC/iPSC variants, non-GC DMRs, and NPC variants) the overlap was considered as a cumulative percentage of each variant. For example, if 40% of a replication timing variant coincided with a region of interest, the contribution of this region to the total overlap would be 0.4. Summed over all the variants, this meant that the total possible overlap for n variants ranged from 0 to n. Doing so normalized the contribution of each variant equally, regardless of size. Overlap comparison for smaller (sub Mb scale) features (those between ESC/iPSC variants and CG-DMRs, genes, or ChrommHMM states which were considered as a series of single base pair states) were simply a cumulative count of the number of features which fell inside variant regions.

Fibroblast replication timing profiles

Using single cell DNA sequencing data of BJ fibroblast (isogenic to NT5/6/8, BJiPSM, and BJiPSO; 10xgenomics.com), we identified cells in S phase and aggregated their DNA sequence data to generate a profile of fibroblast replication timing as previously described (Massey & Koren, 2022). Briefly, G1/G2 cells and S phase cells were identified based on the fluctuations in read depth along chromosomes in the sequence data. Genomic windows were constructed using the G1/G2 cells, and for each window, each S phase cell was categorized as either having replicated or non-replicated DNA. Aggregating the single cell profiles created an ensemble profile that was used for analysis.

Accordingly, the single-cell fibroblast profile was used for comparison with ESC and iPSC replication timing. Six variants were tested that had a correlation of r > 0.3 between the fibroblast and ESC and iPSC means, and where the ESC or iPSC mean showed replication timing similar to fibroblasts (+/- 0.2 SD) at the region of maximum variation between ESCs and iPSCs. These variants were then categorized based on whether the difference in replication timing was smaller between the iPSC mean and fibroblast (epigenetic memory) or between the ESC mean and fibroblast (reprogramming-linked aberration). In all cases, fibroblasts were only within 0.2 SD of either ESCs or iPSCs, but not both.

DNA methylation data and DMRs

All CG DMRs (Lister et al., 2011) were successfully converted from hg18 to hg19 using the UCSC LiftOver tool, while nineteen of 22 non-CG iPSC hypomethylated, and six of seven non-CG iPSC hypermethylated regions were successfully converted to hg19. In order to obtain raw methylation data for plotting purposes (**Fig. 4C**), the 26 variant regions as well as 1Mb upstream and downstream were first lifted from hg19 to hg18. Four of these regions failed to be lifted over, leaving 22 variant regions for analysis. Methylation data in these regions was then obtained and lifted back to hg19. All data in **Fig. 4C** was plotted using hg19 coordinates. A flanking region size of 1Mb was chosen (instead of 3Mb as in the remainder of our analyses) because larger flanking regions failed LiftOver.

Non-CG methylation data was processed according to methods described previously(Lister et al., 2011), differing only in the size of the flanking region. Each variant region +/- 1Mb was divided into 10 Kb windows, and the methylation level of all CHH and CHG context sites was calculated. These window values were then normalized to the maximum value across the region, and data was then averaged for ESCs and iPSC before plotting.

Data availability

All raw sequencing data generated in this study have been submitted to the NCBI Database of Genotypes and Phenotypes (dbGaP; https://www.ncbi.nlm.nih.gov/gap/) under accession number phs001957.

Declaration of interests

The authors declare no competing interests

References Cited

Allison, T. F., Andrews, P. W., Avior, Y., Barbaric, I., Benvenisty, N., Bock, C., Brehm, J., Brüstle, O., Damjanov, I., Elefanty, A., Felkner, D., Gokhale, P. J., Halbritter, F., Healy, L. E., Hu, T. X., Knowles, B. B., Loring, J. F., Ludwig, T. E., Mayberry, R., … Yamanaka, S. (2018). Assessment of established techniques to determine developmental and malignant potential of human pluripotent stem cells. *Nature Communications*, *9*(1). https://doi.org/10.1038/S41467-018-04011-3

Bar-Nur, O., Russ, H. A., Efrat, S., & Benvenisty, N. (2011). Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. *Cell Stem Cell*, *9*(1), 17–23. https://doi.org/10.1016/J.STEM.2011.06.007

Beagan, J. A., Gilgenast, T. G., Kim, J., Plona, Z., Norton, H. K., Hu, G., Hsu, S. C., Shields, E. J., Lyu, X., Apostolou, E., Hochedlinger, K., Corces, V. G., Dekker, J., & Phillips-Cremins, J. E. (2016). Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. *Cell Stem Cell*, *18*(5), 611–624. https://doi.org/10.1016/J.STEM.2016.04.004

Ben-David, U., Mayshar, Y., & Benvenisty, N. (2011). Large-scale analysis reveals acquisition of lineage-specific chromosomal aberrations in human adult stem cells. *Cell Stem Cell*, *9*(2), 97–102. https://doi.org/10.1016/J.STEM.2011.06.013

Benetti, R., García-Cao, M., & Blasco, M. A. (2007). Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. *Nature Genetics 2007 39:2*, *39*(2), 243–250. https://doi.org/10.1038/ng1952

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z. D., Ziller, M., Croft, G. F., Amoroso, M. W., Oakley, D. H., Gnirke, A., Eggan, K., & Meissner, A. (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell*, *144*(3), 439–452. https://doi.org/10.1016/J.CELL.2010.12.032

Bouma, M. J., van Iterson, M., Janssen, B., Mummery, C. L., Salvatori, D. C. F., & Freund, C. (2017). Differentiation-Defective Human Induced Pluripotent Stem Cells Reveal Strengths and Limitations of the Teratoma Assay and In Vitro Pluripotency Assays. *Stem Cell Reports*, *8*(5), 1340–1353. https://doi.org/10.1016/J.STEMCR.2017.03.009

Bragança, J., Lopes, J. A., Mendes-Silva, L., & Santos, J. M. A. (2019). Induced pluripotent stem cells, a giant leap for mankind therapeutic applications. *World Journal of Stem Cells*, *11*(7), 421. https://doi.org/10.4252/WJSC.V11.I7.421

Burrows, C. K., Banovich, N. E., Pavlovic, B. J., Patterson, K., Gallego Romero, I., Pritchard, J. K., & Gilad, Y. (2016). Genetic Variation, Not Cell Type of Origin, Underlies the Majority of

Identifiable Regulatory Differences in iPSCs. *PLOS Genetics*, *12*(1), e1005793. https://doi.org/10.1371/JOURNAL.PGEN.1005793

Choi, J., Lee, S., Mallard, W., Clement, K., Tagliazucchi, G. M., Lim, H., Choi, I. Y., Ferrari, F., Tsankov, A. M., Pop, R., Lee, G., Rinn, J. L., Meissner, A., Park, P. J., & Hochedlinger, K. (2015). A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs. *Nature Biotechnology 2015 33:11*, *33*(11), 1173–1181. https://doi.org/10.1038/nbt.3388

Ding, Q., Edwards, M. M., Wang, N., Zhu, X., Bracci, A. N., Hulke, M. L., Hu, Y., Tong, Y., Hsiao, J., Charvet, C. J., Ghosh, S., Handsaker, R. E., Eggan, K., Merkle, F. T., Gerhardt, J., Egli, D., Clark, A. G., & Koren, A. (2021). The genetic architecture of DNA replication timing in human pluripotent stem cells. *Nature Communications 2021 12:1*, *12*(1), 1–18. https://doi.org/10.1038/s41467-021-27115-9

Du, Q., Bert, S. A., Armstrong, N. J., Caldon, C. E., Song, J. Z., Nair, S. S., Gould, C. M., Luu, P.-L., Peters, T., Khoury, A., Qu, W., Zotenko, E., Stirzaker, C., & Clark, S. J. (2019). Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nature Communications*, *10*(1), 1–15. https://doi.org/doi:10.1038/s41467-019-08302-1

Du, Q., Smith, G. C., Luu, P. L., Ferguson, J. M., Armstrong, N. J., Caldon, C. E., Campbell, E. M., Nair, S. S., Zotenko, E., Gould, C. M., Buckley, M., Chia, K. M., Portman, N., Lim, E., Kaczorowski, D., Chan, C. L., Barton, K., Deveson, I. W., Smith, M. A., … Clark, S. J. (2021). DNA methylation is required to maintain both DNA replication timing precision and 3D genome organization integrity. *Cell Reports*, *36*(12). https://doi.org/10.1016/J.CELREP.2021.109722

Edwards, M. M., Zuccaro, M. V., Sagi, I., Ding, Q., Vershkov, D., Benvenisty, N., Egli, D., & Koren, A. (2021). Delayed DNA replication in haploid human embryonic stem cells. *Genome Research*, *31*(12), 2155–2169. https://doi.org/10.1101/GR.275953.121

Ernst, J., & Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. *Nature Protocols*, *12*(12), 2478–2492. https://doi.org/doi:10.1038/nprot.2017.124

Gonzalo, S., Jaco, I., Fraga, M. F., Chen, T., Li, E., Esteller, M., & Blasco, M. A. (2006). DNA methyltransferases control telomere length and telomere recombination in mammalian cells. *Nature Cell Biology 2006 8:4*, *8*(4), 416–424. https://doi.org/10.1038/ncb1386

Gribnau, J., Hochedlinger, K., Hata, K., Li, E., & Jaenisch, R. (2003). Asynchronous replication timing of imprinted loci is independent of DNA methylation, but consistent with differential subnuclear localization. *Genes Dev*, *17*(6), 759–773. https://doi.org/10.1101/gad.1059603

Guenther, M. G., Frampton, G. M., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., & Young, R. A. (2010). Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell*, *7*(2), 249–257. https://doi.org/10.1016/J.STEM.2010.06.015

Gurdon, J. B., & Byrne, J. A. (2004). The first half-century of nuclear transplantation. *Bioscience Reports*, *24*(4–5), 545–557. https://doi.org/10.1007/S10540-005-2744-5

1   Handsaker, R. E., Van Doren, V., Berman, J. R., Genovese, G., Kashin, S., Boettger, L. M., &
2   Mccarroll, S. A. (2015). Large multiallelic copy number variations in humans. *Nature*
3   *Genetics 2015 47:3*, *47*(3), 296–303. https://doi.org/10.1038/ng.3200
4   Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C. W., Lyou, Y., Townes, T. M.,
5   Schübeler, D., & Gilbert, D. M. (2008). Global reorganization of replication domains during
6   embryonic stem cell differentiation. *PLoS Biology*, *6*(10).
7   https://doi.org/10.1371/journal.pbio.0060245
8   Hoffman, M. M., Ernst, J., Wilder, S. P., Kundaje, A., Harris, R. S., Libbrecht, M., Giardine, B.,
9   Ellenbogen, P. M., Bilmes, J. A., Birney, E., Hardison, R. C., Dunham, I., Kellis, M., & Noble,
10   W. S. (2013). Integrative annotation of chromatin elements from ENCODE data. *Nucleic*
11   *Acids Research*, *41*(2). https://doi.org/10.1093/nar/gks1284
12 Hussein, S. M., Batada, N. N., Vuoristo, S., Ching, R. W., Autio, R., Narvä, E., Ng, S., Sourour, M.,
13   Hämälä, R., Olsson, C., Lundin, K., Mikkola, M., Trokovic, R., Peitz, M., Brüstle, O., Bazett-
14   Jones, D. P., Alitalo, K., Lahesmaa, R., Nagy, A., & Otonkoski, T. (2011). Copy number
15   variation and selection during reprogramming to pluripotency. *Nature*, *471*(7336), 58–62.
16   https://doi.org/10.1038/NATURE09871
17 Johannesson, B., Sagi, I., Gore, A., Paull, D., Yamada, M., Golan-Lev, T., Li, Z., LeDuc, C., Shen, Y.,
18   Stern, S., Xu, N., Ma, H., Kang, E., Mitalipov, S., Sauer, M. V., Zhang, K., Benvenisty, N., &
19   Egli, D. (2014). Comparable frequencies of coding mutations and loss of imprinting in
20   human pluripotent cells derived by nuclear transfer and defined factors. *Cell Stem Cell*,
21   *15*(5), 634–642. https://doi.org/10.1016/J.STEM.2014.10.002
22 Kilpinen, H., Goncalves, A., Leha, A., Afzal, V., Alasoo, K., Ashford, S., Bala, S., Bensaddek, D.,
23   Casale, F. P., Culley, O. J., Danecek, P., Faulconbridge, A., Harrison, P. W., Kathuria, A.,
24   McCarthy, D., McCarthy, S. A., Meleckyte, R., Memari, Y., Moens, N., … Gaffney, D. J.
25   (2017). Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature*,
26   *546*(7658), 370–375. https://doi.org/10.1038/NATURE22403
27 Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M. J., Ji, H., Ehrlich, L. I. R.,
28   Yabuuchi, A., Takeuchi, A., Cunniff, K. C., Hongguang, H., McKinney-Freeman, S., Naveiras,
29   O., Yoon, T. J., Irizarry, R. A., Jung, N., … Daley, G. Q. (2010). Epigenetic memory in induced
30   pluripotent stem cells. *Nature*, *467*(7313), 285–290.
31   https://doi.org/10.1038/NATURE09342
32 Koren, A. (2014). DNA replication timing: Coordinating genome stability with genome regulation
33   on the X chromosome and beyond. *BioEssays : News and Reviews in Molecular, Cellular*
34   *and Developmental Biology*, *36*(10), 997–1004. https://doi.org/10.1002/BIES.201400077
35 Koren, A., Handsaker, R. E., Kamitaki, N., Karlić, R., Ghosh, S., Polak, P., Eggan, K., & McCarroll,
36   S. A. (2014). Genetic Variation in Human DNA Replication Timing. *Cell*, *159*(5), 1015–1026.
37   https://doi.org/https://doi.org/10.1016/j.cell.2014.10.025
38 Koren, A., Massey, D. J., & Bracci, A. N. (2021). TIGER: inferring DNA replication timing from
39   whole-genome sequence data. *Bioinformatics*, *btab166*.
40   https://doi.org/10.1093/bioinformatics/btab166
41 Koyanagi-Aoi, M., Ohnuki, M., Takahashi, K., Okita, K., Noma, H., Sawamura, Y., Teramoto, I.,
42   Narita, M., Sato, Y., Ichisaka, T., Amanoa, N., Watanabe, A., Morizane, A., Yamada, Y., Sato,
43   T., Takahashi, J., & Yamanaka, S. (2013). Differentiation-defective phenotypes revealed by
44   large-scale analyses of human pluripotent stem cells. *Proceedings of the National Academy*

of Sciences of the United States of America, *110*(51), 20569–20574. https://doi.org/10.1073/PNAS.1319061110/SUPPL_FILE/ST02.DOCX

Krijger, P. H. L., Di Stefano, B., De Wit, E., Limone, F., Van Oevelen, C., De Laat, W., & Graf, T. (2016). Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming. *Cell Stem Cell*, *18*(5), 597. https://doi.org/10.1016/J.STEM.2016.01.007

Kyttälä, A., Moraghebi, R., Valensisi, C., Kettunen, J., Andrus, C., Pasumarthy, K. K., Nakanishi, M., Nishimura, K., Ohtaka, M., Weltner, J., Van Handel, B., Parkkonen, O., Sinisalo, J., Jalanko, A., Hawkins, R. D., Woods, N. B., Otonkoski, T., & Trokovic, R. (2016). Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential. *Stem Cell Reports*, *6*(2), 200–212. https://doi.org/10.1016/J.STEMCR.2015.12.009

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *26*(5), 589–595. https://doi.org/10.1093/BIOINFORMATICS/BTP698

Liang, G., & Zhang, Y. (2013). Genetic and Epigenetic Variations in iPSCs: Potential Causes and Implications for Application. *Cell Stem Cell*, *13*(2), 149–159. https://doi.org/10.1016/J.STEM.2013.07.001

Lister, R., Pelizzola, M., Kida, Y. S., Hawkins, R. D., Nery, J. R., Hon, G., Antosiewicz-Bourget, J., Ogmalley, R., Castanon, R., Klugman, S., Downes, M., Yu, R., Stewart, R., Ren, B., Thomson, J. A., Evans, R. M., & Ecker, J. R. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*, *471*(7336), 68–73. https://doi.org/10.1038/NATURE09798

Liu, L. P., & Zheng, Y. W. (2019). Predicting differentiation potential of human pluripotent stem cells: Possibilities and challenges. *World Journal of Stem Cells*, *11*(7), 375. https://doi.org/10.4252/WJSC.V11.I7.375

Lu, J., Li, H., Hu, M., Sasaki, T., Baccei, A., Gilbert, D. M., Liu, J. S., Collins, J. J., & Lerou, P. H. (2014). The Distribution of Genomic Variations in Human iPSCs Is Related to Replication-Timing Reorganization during Reprogramming. *Cell Reports*, *7*(1), 70–78. https://doi.org/10.1016/J.CELREP.2014.03.007

Ma, H., Morey, R., O'Neil, R. C., He, Y., Daughtry, B., Schultz, M. D., Hariharan, M., Nery, J. R., Castanon, R., Sabatini, K., Thiagarajan, R. D., Tachibana, M., Kang, E., Tippner-Hedges, R., Ahmed, R., Gutierrez, N. M., Van Dyken, C., Polat, A., Sugawara, A., … Mitalipov, S. (2014). Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature*, *511*(7508), 177–183. https://doi.org/10.1038/NATURE13551

Marchetto, M. C. N., Yeo, G. W., Kainohana, O., Marsala, M., Gage, F. H., & Muotri, A. R. (2009). Transcriptional Signature and Memory Retention of Human-Induced Pluripotent Stem Cells. *PLoS ONE*, *4*(9), 7076. https://doi.org/10.1371/JOURNAL.PONE.0007076

Massey, D. J., & Koren, A. (2022). High-throughput analysis of single human cells reveals the complex nature of DNA replication timing control. *Nature Communications*, *13*(1), 2402. https://doi.org/10.1038/S41467-022-30212-Y

Müller, F. J., Schuldt, B. M., Williams, R., Mason, D., Altun, G., Papapetrou, E. P., Danner, S., Goldmann, J. E., Herbst, A., Schmidt, N. O., Aldenhoff, J. B., Laurent, L. C., & Loring, J. F. (2011). A bioinformatic assay for pluripotency in human cells. *Nature Methods*, *8*(4), 315–317. https://doi.org/10.1038/NMETH.1580

Nishino, K., Toyoda, M., Yamazaki-Inoue, M., Fukawatase, Y., Chikazawa, E., Sakaguchi, H., Akutsu, H., & Umezawa, A. (2011). DNA Methylation Dynamics in Human Induced Pluripotent Stem Cells over Time. *PLoS Genetics*, *7*(5), 1002085. https://doi.org/10.1371/JOURNAL.PGEN.1002085

Ohi, Y., Qin, H., Hong, C., Blouin, L., Polo, J. M., Guo, T., Qi, Z., Downey, S. L., Manos, P. D., Rossi, D. J., Yu, J., Hebrok, M., Hochedlinger, K., Costello, J. F., Song, J. S., & Ramalho-Santos, M. (2011). Incomplete DNA methylation underlies a transcriptional memory of the somatic cell in human iPS cells. *Nature Cell Biology*, *13*(5), 541. https://doi.org/10.1038/NCB2239

Paniza, T., Deshpande, M., Wang, N., O'Neil, R., Zuccaro, M. V, Smith, M. E., Madireddy, A., James, D., Ecker, J., Rosenwaks, Z., Egli, D., & Gerhardt, J. (2020). Pluripotent stem cells with low differentiation potential contain incompletely reprogrammed DNA replication. *The Journal of Cell Biology*, *219*(9). https://doi.org/10.1083/jcb.201909163

Panopoulos, A. D., Smith, E. N., Arias, A. D., Shepard, P. J., Hishida, Y., Modesto, V., Diffenderfer, K. E., Conner, C., Biggs, W., Sandoval, E., D'Antonio-Chronowska, A., Berggren, W. T., Izpisua Belmonte, J. C., & Frazer, K. A. (2017). Aberrant DNA Methylation in Human iPSCs Associates with MYC-Binding Motifs in a Clone-Specific Manner Independent of Genetics. *Cell Stem Cell*, *20*(4), 505-517.e6. https://doi.org/10.1016/J.STEM.2017.03.010

Pasi, C. E., Dereli-Öz, A., Negrini, S., Friedli, M., Fragola, G., Lombardo, A., Van Houwe, G., Naldini, L., Casola, S., Testa, G., Trono, D., Pelicci, P. G., & Halazonetis, T. D. (2011). Genomic instability in induced stem cells. *Cell Death and Differentiation*, *18*(5), 745–753. https://doi.org/10.1038/CDD.2011.9

Polo, J. M., Liu, S., Figueroa, M. E., Kulalert, W., Eminli, S., Tan, K. Y., Apostolou, E., Stadtfeld, M., Li, Y., Shioda, T., Natesan, S., Wagers, A. J., Melnick, A., Evans, T., & Hochedlinger, K. (2010). Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nature Biotechnology*, *28*(8), 848–855. https://doi.org/10.1038/NBT.1667

Robinton, D. A., & Daley, G. Q. (2012). The promise of induced pluripotent stem cells in research and therapy. *Nature 2012 481:7381*, *481*(7381), 295–305. https://doi.org/10.1038/nature10761

Rouhani, F., Kumasaka, N., de Brito, M. C., Bradley, A., Vallier, L., & Gaffney, D. (2014). Genetic background drives transcriptional variation in human induced pluripotent stem cells. *PLoS Genetics*, *10*(6). https://doi.org/10.1371/JOURNAL.PGEN.1004432

Ruiz, S., Diep, D., Gore, A., Panopoulos, A. D., Montserrat, N., Plongthongkum, N., Kumar, S., Fung, H. L., Giorgetti, A., Bilic, J., Batchelder, E. M., Zaehres, H., Kan, N. G., Schoĺer, H. R., Mercola, M., Zhang, K., & Belmonte, J. C. I. (2012). Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(40), 16196–16201. https://doi.org/10.1073/PNAS.1202352109

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T. C., Robins, A. J., Dalton, S., & Gilbert, D. M. (2010). *Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types*. https://doi.org/10.1101/gr.099655.109

Saksouk, N., Simboeck, E., & Déjardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & Chromatin*, *8*(1). https://doi.org/10.1186/1756-8935-8-3

Schlaeger, T. M., Daheron, L., Brickler, T. R., Entwisle, S., Chan, K., Cianci, A., DeVine, A., Ettenger, A., Fitzgerald, K., Godfrey, M., Gupta, D., McPherson, J., Malwadkar, P., Gupta, M., Bell, B., Doi, A., Jung, N., Li, X., Lynes, M. S., … Daley, G. Q. (2015). A comparison of non-integrating reprogramming methods. *Nature Biotechnology*, *33*(1), 58–63. https://doi.org/10.1038/NBT.3070

Shi, Y., Inoue, H., Wu, J. C., & Yamanaka, S. (2017). Induced pluripotent stem cell technology: a decade of progress. *Nature Reviews. Drug Discovery*, *16*(2), 115–130. https://doi.org/10.1038/NRD.2016.245

Shufaro, Y., Lacham-Kaplan, O., Tzuberi, B. Z., Mclaughlin, J., Trounson, A., Cedar, H., & Reubinoff, B. E. (2010). Reprogramming of DNA Replication Timing. *STEM CELLS*, *28*(3), 443–449. https://doi.org/10.1002/STEM.303

Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., Shioda, T., & Hochedlinger, K. (2010). Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature*, *465*(7295), 175–181. https://doi.org/10.1038/NATURE09017

Stadtfeld, M., Apostolou, E., Ferrari, F., Choi, J., Walsh, R. M., Chen, T., Ooi, S. S. K., Kim, S. Y., Bestor, T. H., Shioda, T., Park, P. J., & Hochedlinger, K. (2012). Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all–iPS cell mice from terminally differentiated B cells. *Nature Genetics 2012 44:4*, *44*(4), 398–405. https://doi.org/10.1038/ng.1110

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(3), 479–498. https://doi.org/10.1111/1467-9868.00346

Sui, L., Danzl, N., Campbell, S. R., Viola, R., Williams, D., Xing, Y., Wang, Y., Phillips, N., Poffenberger, G., Johannesson, B., Oberholzer, J., Powers, A. C., Leibel, R. L., Chen, X., Sykes, M., & Egli, D. (2018). β-Cell Replacement in Mice Using Human Type 1 Diabetes Nuclear Transfer Embryonic Stem Cells. *Diabetes*, *67*(1), 26–35. https://doi.org/10.2337/DB17-0120

Takebayashi, S. I., Ryba, T., Wimbish, K., Hayakawa, T., Sakaue, M., Kuriya, K., Takahashi, S., Ogata, S., Hiratani, I., Okumura, K., Okano, M., & Ogata, M. (2021). The Temporal Order of DNA Replication Shaped by Mammalian DNA Methyltransferases. *Cells*, *10*(2), 1–18. https://doi.org/10.3390/CELLS10020266

Volpato, V., & Webber, C. (2020). Addressing variability in iPSC-derived models of human disease: guidelines to promote reproducibility. *Disease Models & Mechanisms*, *13*(1). https://doi.org/10.1242/DMM.042317

Wang, L., Egli, D., & Leibel, R. L. (2016). Efficient Generation of Hypothalamic Neurons from Human Pluripotent Stem Cells. *Current Protocols in Human Genetics*, *90*(1), 21.5.1-21.5.14. https://doi.org/10.1002/CPHG.3