Delayed DNA replication in haploid human embryonic stem cells

Matthew M. Edwards,[1] Michael V. Zuccaro,[2,3] Ido Sagi,[4,5] Qiliang Ding,[1] Dan Vershkov,[4] Nissim Benvenisty,[4] Dieter Egli,[2,3] and Amnon Koren[1]

[1]*Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA;* [2]*Department of Pediatrics and Naomi Berrie Diabetes Center, Columbia University, New York, New York 10032, USA;* [3]*Columbia University Stem Cell Initiative, New York, New York 10032, USA;* [4]*The Azrieli Center for Stem Cells and Genetic Research, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University, Jerusalem 91904, Israel*

Haploid human embryonic stem cells (ESCs) provide a powerful genetic system but diploidize at high rates. We hypothesized that diploidization results from aberrant DNA replication. To test this, we profiled DNA replication timing in isogenic haploid and diploid ESCs. The greatest difference was the earlier replication of the X Chromosome in haploids, consistent with the lack of X-Chromosome inactivation. We also identified 21 autosomal regions that had delayed replication in haploids, extending beyond the normal S phase and into G2/M. Haploid-delays comprised a unique set of quiescent genomic regions that are also underreplicated in polyploid placental cells. The same delays were observed in female ESCs with two active X Chromosomes, suggesting that increased X-Chromosome dosage may cause delayed autosomal replication. We propose that incomplete replication at the onset of mitosis could prevent cell division and result in re-entry into the cell cycle and whole genome duplication.

[Supplemental material is available for this article.]

Embryonic stem cells are typically derived from in vitro fertilization of human oocytes. Alternatively, oocytes can be artificially activated to develop into blastocysts from which parthenogenetic stem cells, containing only maternally derived chromosomes, can be derived. Recently, fluorescence-activated cell sorting (FACS) has been used to isolate and maintain haploid parthenogenetic embryonic stem cells (h-pESCs) (Sagi et al. 2016). Haploid cells hold great promise as a tool for conducting loss-of-function genetic screens (Leeb et al. 2014; Yilmaz et al. 2020); for studying the stability of cell ploidy in development and disease, tolerance to ploidy changes, X-Chromosome inactivation and parental imprinting; and potentially for applications in regenerative and reproductive medicine (Yilmaz et al. 2016; Li and Shuai 2017; Zhang et al. 2020). However, haploid cells are naturally unstable, experiencing high levels of spontaneous diploidization (Tarkowski et al. 1970; Yaguchi et al. 2018). Human h-pESCs diploidize at a rate of 3%–9% every cell cycle (Sagi et al. 2016), which poses a major limitation for their use in genetic studies. This diploidization also raises fundamental questions regarding the stability of the haploid state in mammals.

Oocyte activation without fertilization can also occur in vivo, in which case it results in the development of ovarian teratomas (Stevens and Varnum 1974; Linder et al. 1975) that consist mostly of diploid cells (Baker et al. 1998; Stelzer et al. 2011; Heskett et al. 2020), suggesting that diploidization occurs early in their development. Diploidization is not specific to parthenogenicity, as hydatidiform moles forming from the loss of the nucleus in a fertilized egg are found to be diploid as well (Fan et al. 2002). Similarly, most androgenetic stem cells derived from haploid eggs with only a paternal genome are also diploid (Sagi et al. 2019).

It has previously been shown that diploidization of mouse haploid stem cells occurs through two rounds of DNA replication without an intervening cell division—rather than by cell fusion (Leeb et al. 2012; Takahashi et al. 2014). Although previous studies have focused on mitotic progression for explaining diploidization (Guo et al. 2017; Leng et al. 2017; Li et al. 2017), the fundamental reasons for the failure of haploid cells to normally progress through the cell cycle remain unknown. Haploid human stem cells show several notable differences compared to their diploid counterparts, including their smaller size, a larger surface area-to-volume ratio, and a higher proportion of mitochondrial relative to nuclear DNA. In addition, haploid cells have one active X Chromosome (Xa), whereas diploid cells have one active and one inactive X Chromosome (Xi). Therefore, relative to the autosomes, haploids have an approximately twofold higher expression of X-linked genes when compared to diploids (Sagi et al. 2016). Any of these differences, as well as the lack of homologous chromosomes per se, could underlie the instability of mammalian haploid cells.

Diploidization resembles polyploidization, which also involves whole genome duplication. Polyploid cells normally arise in the human liver, bone marrow, and placenta, as well as across tissues of *Drosophila melanogaster*, in many plants, and in other organisms (Schoenfelder and Fox 2015; Sagi and Benvenisty 2017). Polyploidy is also common in cancer (Bielski et al. 2018). Polyploidization is often accompanied by genomic regions of reduced relative DNA copy number. In *Drosophila* polyploid cells,

[5]Present address: Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
Corresponding author: koren@cornell.edu
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.275953.121.

underreplication is due to active inhibition of replication fork progression in a subset of late-replicating genomic regions. At least two negative regulators of DNA replication, Suppressor of Under-Replication (SuUR) and Rif1, have been implicated in this process (Makunin et al. 2002; Nordman et al. 2011, 2014; Munden et al. 2018). More recently, the highly polyploid trophoblast giant cells (TGCs) of the mouse placenta have also been shown to harbor unique regions of reduced copy number (Hannibal et al. 2014).

Here, we hypothesized that spontaneous diploidization of haploid stem cells, and potentially also physiological polyploidization of other cell types, could be related to abberant DNA replication that propagates to aborted mitoses. We thus profiled DNA replication dynamics genome-wide in isogenic haploid and diploid human embryonic stem cells and aimed to study the causes and consequences of any potential ploidy-related differences in replication timing by characterizing their genomic and epigenomic properties.

Results

Haploid embryonic stem cells show replication timing profiles characteristic of pluripotent stem cells

We hypothesized that defects in DNA replication could be the fundamental cause of haploid cell diploidization, for instance, by incomplete DNA replication carrying over to mitotic failure. To test for differences in DNA replication timing between haploid and diploid cells, we generated genome-wide DNA replication timing profiles for isogenic parthenogenetic haploid (h) and diploid (d) cultures of two cell lines, pES10 and pES12 (Sagi et al. 2016) along with long-term stably diploid pES10 and pES12 (unsorted for ploidy) (see Methods) and two control ESC lines derived by in vitro fertilization (CU-ES4 and CU-ES5). Haploid and diploid cells were both highly proliferating and had similar cell cycle distributions (Supplemental Fig. S1).

DNA replication results in differential DNA copy number across the genome, with earlier-replicating loci showing an increased DNA content. We previously showed that these fluctuations in copy number can be detected from whole-genome sequencing of cell populations, whereas nonproliferating cells exhibit a relatively uniform profile of DNA copy number along chromosomes. This principle can be used to generate high-resolution profiles of DNA replication timing that match or exceed the quality of replication profiles measured using S/G1 sorting or Repli-seq. The replication profiles obtained by whole-genome sequencing of proliferating cell cultures has been used to identify differences between cell lines from different individuals and link them to genetic polymorphisms (Koren et al. 2014, 2021; Ding et al. 2020). Accordingly, we sequenced genomic DNA and calculated DNA copy number (sequencing read depth) in 2.5-kb windows of uniquely alignable sequence, normalized by local GC content (Koren et al. 2014). We filtered out copy number variants (CNVs) and outliers, then smoothed the data to generate DNA replication timing profiles (Methods). This approach is particularly applicable to embryonic stem cells, which are highly proliferative and contain a large fraction of cells in S phase (Ding et al. 2020). In contrast to other replication profiling methods, it does not rely on sorting cells within predetermined stages of S phase and thus provides a less biased view of genome replication.

To assess the quality of the data, we first analyzed the autocorrelation of the unsmoothed replication timing profiles, which is a measure of profile continuity and represents the replicative activity of the samples. Long-range autocorrelation was observed in haploid, isogenic diploidized (hereafter referred to as "diploid" for simplicity) and control cells across all chromosomes (Supplemental Fig. S2A). We then compared the replication timing profiles of the haploid, diploid, and control cell lines, as well as 57 separately sequenced ESCs with indications of normal X Chromosome inactivation (karyotypically XX and henceforth referred to as female) (Ding et al. 2020). The correlations within and between haploid and diploid cell lines ranged from $r = 0.92$ to $r = 0.96$, whereas the correlations to control ESCs ranged from $r = 0.77$ to $r = 0.88$ when compared to the separately-sequenced ESCs, $r = 0.81$ to $r = 0.93$ compared to the concurrently sequenced control ESCs, and $r = 0.94$ to $r = 0.95$ compared to the isogenic unsorted diploid ESCs (Supplemental Fig. S2B). Thus, the replication profiles were highly reproducible and consistent with controls and previous measurements, with expected minor differences likely due to experimental protocols and genetic background. In particular, control cell lines are biparental and may have specific differences from parthenogenetic cell lines at imprinted regions; our analyses below focus on comparisons between the isogenic parthenogenetic haploid and diploid cell lines. Finally, visual inspection of the replication timing profiles showed a high degree of similarity in the profiles of haploid, diploid, and control ESCs along the vast majority of the genome (Fig. 1A), with some notable differences explored further below.

Significant replication timing variation between isogenic haploid and diploid human embryonic stem cells

Notwithstanding the high overall similarity in replication timing between haploid and diploid cells, we also observed sites of replication timing variation between the two cell ploidies or in individual haploid cell lines (Fig. 1A). To systematically identify and categorize such replication timing variants, we utilized one-way ANOVA tests on consecutive regions across the genome to identify consistent variation between the two haploid and the two diploid cell lines and Student's *t*-tests to identify regions in which only one haploid cell line was different from diploid cells (Methods). Although differences specific to a single cell line could arise from genetic differences (Koren et al. 2014), somatic copy number alterations (see further below), or measurement noise, shared differences specifically reflect associations between cell ploidy and DNA replication timing.

To evaluate the specificity of the ANOVA test, we compared the extent of identified haploid-diploid replication timing variation across the autosomes to that of a sample permutation that disrupted the ploidy and genotype relationships (h-pES10 and d-pES12 compared to h-pES12 and d-pES10) (Methods), revealing a 14.2-fold greater haploid-diploid variation than expected by default. The most extensive variation encompassed 82.7% of the X Chromosome across 19 distinct regions (86.8% when excluding PAR1 and evolutionary strata 4 and 5). Although replication timing was highly correlated between haploid and diploid cells across the autosomes ($r = 0.92$ to $r = 0.94$), their correlations on the X Chromosome were substantially lower ($r = 0.73$ to $r = 0.76$) (Supplemental Fig. S2B). Autocorrelation of the X Chromosome was also significantly higher in haploids compared to diploids (Supplemental Fig. S2A). The mean replication timing of the X Chromosome was much earlier in haploids (−0.29) compared to diploids (−1.6) (Fig. 1A). These results can be readily explained by X-Chromosome inactivation in diploid cells: diploid pES10 and pES12 have an inactive X Chromosome (Sagi et al. 2016), which was

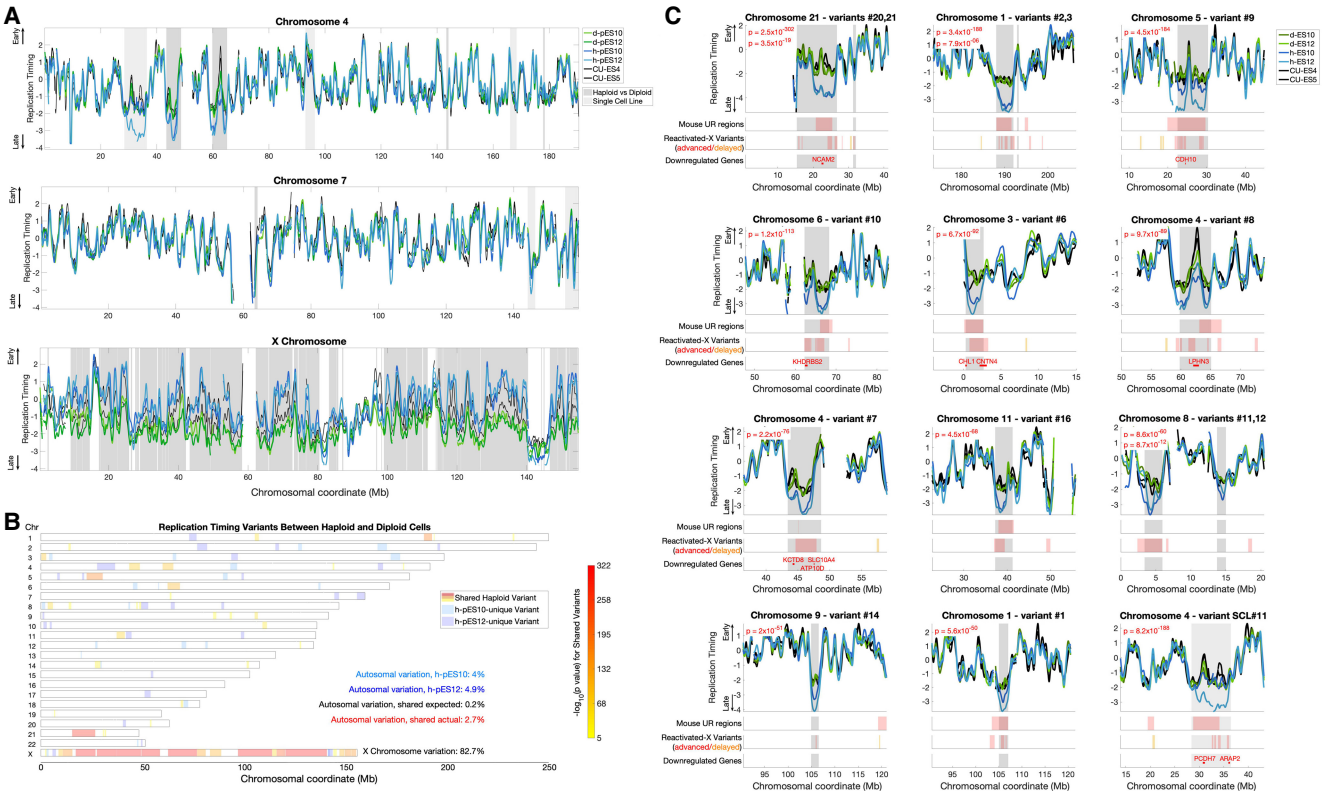

Figure 1. Significant replication timing variation between haploid and diploid ESCs. (*A*) Representative replication timing profiles for haploid and diploid pES10 and pES12 and control ESCs. Profiles are z-score-normalized to a genome-wide average of zero and a standard deviation of one, such that positive and negative values indicate replication timing that is earlier and later than average, respectively, and the units correspond to standard deviations (SD). Gray: regions of variation between haploid and diploid cells (light: one cell line; dark: shared in both cell lines). (*B*) A genome-wide view of all replication timing variants. Shared variants are color-coded by *P*-value. The degree of shared haploid variation is significantly greater than expected given the extent of single cell line variation (χ^2 test $P << 10^{-16}$). (*C*) The 11 most significant haploid-delayed variants, numbered by genomic location and ordered by *P*-value. Variants #21, #3, and #12 are each proximal to another, larger variant (represented by the second-listed *P*-value in each panel). Variant SCL#11 (single cell line variant #11) is only delayed in a h-pES12 but nonetheless shows features common to other haploid-delayed variants (see text). Shown *below* each variant are the locations of all mouse UR regions (those found in all stages of development) and reactivated-X variants in the plotted interval of each panel. Reactivated-X delays are shown in red, and advanced regions are in orange. Down-regulated genes are shown only within the replication timing variant borders (gray shades). Supplemental Figure S3 shows all remaining variants.

shown before to replicate much later than the active X Chromosome and without a well-defined replication timing program (Koren and McCarroll 2014); this pattern of X Chromosome replication is also presently observed in human ESCs (see below) (Supplemental Fig. S7A,C). Haploid cells, in contrast, only carry a single, active X Chromosome. These results reaffirm the quality of our replication timing measurements and indicate that X-Chromosome inactivation, which leads to the largest gene expression differences between haploids and diploids (Sagi et al. 2016), is also reflected in DNA replication timing differences in haploids compared to diploids.

In addition to the large-scale variation on the X Chromosome, we also found 31 replication timing variants on the autosomes. These variants ranged from 392 kb to 11.3 Mb in length (mean 2.3 Mb), cumulatively covered 2.7% of the autosomes, and ranged in significance from $P = 7.9 \times 10^{-6}$ to $P = 2.5 \times 10^{-302}$ (Fig. 1B,C; Supplemental Fig. S3A,B). We also identified 19 variants unique to h-pES10 and 24 variants that were specific to h-pES12. These cell line–specific variants provided a useful way to further evaluate the significance of having found as much shared haploid-diploid variation as we observed. Specifically, h-pES10 was variant from diploid cells across a total of 4.0% of the genome (1.3% specific to h-pES10 and 2.7% shared in both haploid cell lines), whereas h-pES12 was variant across 4.9% of the genome (2.2% cell line–specific). Therefore, coincidental shared variation is expected to encompass 0.2% of the genome by chance (4.9% × 4%). However, we observed 2.7% of the genome to harbor shared replication timing variation between the two cell lines, a significant, 13.5-fold enrichment compared to expectation (χ^2 test, $P << 10^{-16}$) (Fig. 1B). In addition, single cell line variants were less significant on average and showed much less correlation with other biological properties (see below). A notable exception is a large variant in h-pES12 on Chromosome 4 (Fig. 1C, variant SCL#11), which was larger (by 2.55-fold) and much more significant than all other single cell line variants ($P = 2.7 \times 10^{-149}$, compared to the next strongest *P*-value of 2×10^{-42}). Notwithstanding this exceptional variant, the strong enrichment of variation shared by both haploid cell lines suggests that the haploid state per se, and not sporadic variation between cell lines, is driving most of the replication timing variation that we identified. We conclude that human haploid ESCs have significant replication timing differences from their diploid counterparts on both the X Chromosome and the autosomes. These results were reproduced in two additional replicate experiments (see further below).

The replication of a subset of normally late-replicating genomic regions is further delayed in haploid cells

Of the 31 shared autosomal replication timing variant regions, 21 replicated later in haploid cells compared to diploids ("haploid-delayed"), whereas only 10 replicated earlier in haploids ("haploid-advanced") (Fig. 2A). The haploid-delayed variants were larger than haploid-advanced variants (median size 2.01 Mb compared to 836 kb, rank-sum $P = 0.019$), showed a greater replication timing difference compared to diploid cells (median of 1.58 SD from the mean, compared to 0.80, rank-sum $P = 4.2 \times 10^{-4}$) and were more significant (median $P = 5.57 \times 10^{-50}$ compared to $P = 8.87 \times 10^{-10}$) (Fig. 2B).

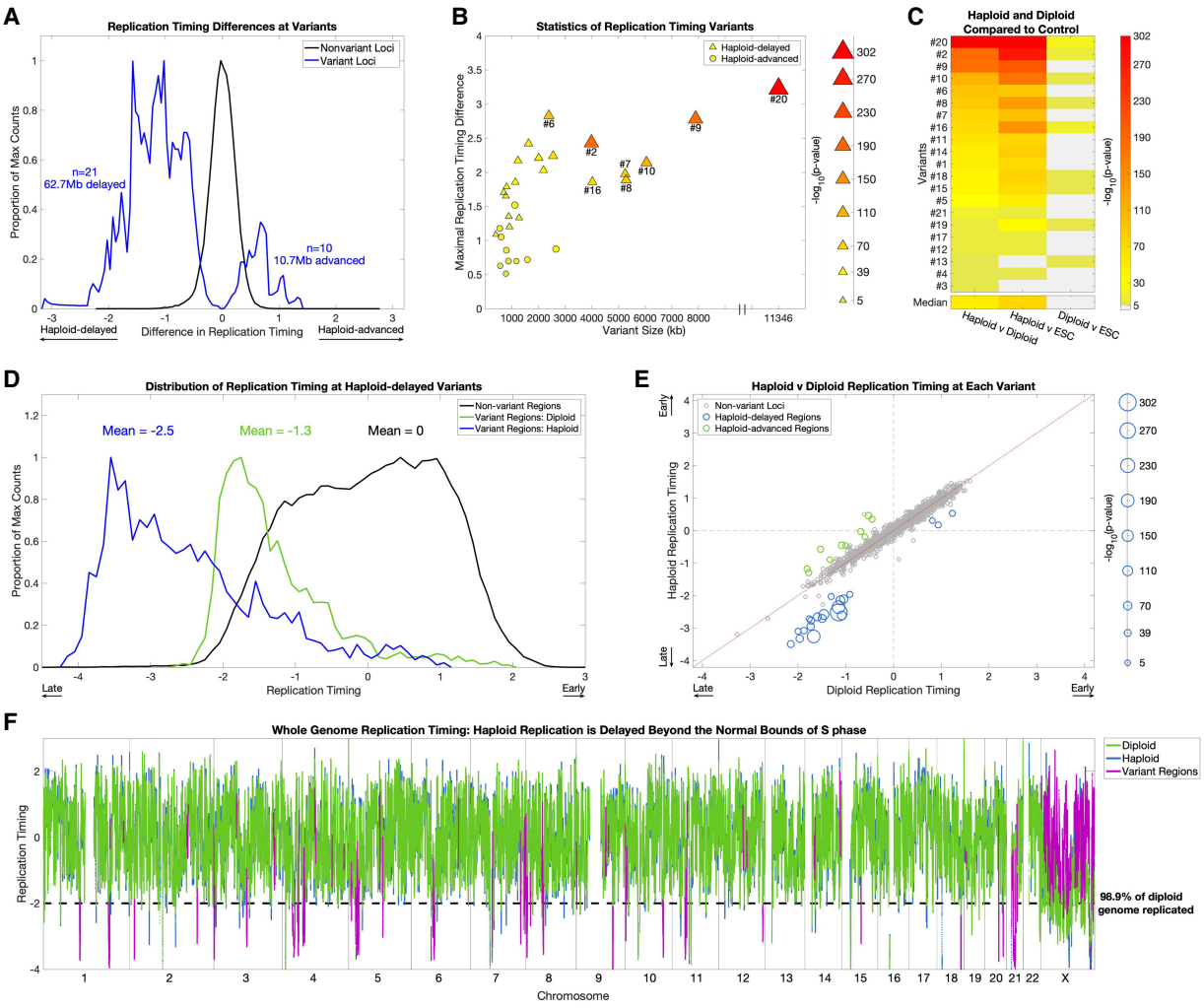

Figure 2. Replication timing in haploid cells is delayed beyond the normal S-phase. (*A*) Haploid-delays are the predominant form of replication timing variation. Distributions of the replication timing differences between haploid and diploid cells (mean of each pair of cell lines) in nonvariant and variant regions. The distributions represent individual windows (2.5 kb) in each category. (*B*) The size (scaled by the $-\log_{10}$[*P*-value]), absolute replication timing difference, and variation *P*-value of all 31 haploid-diploid replication timing variants. Haploid-delayed variants segregate from haploid-advanced variants on all three metrics, shown by the clustering of haploid-advanced (circles) in the *bottom left*, as well as their weaker significance. Conversely, haploid-delayed variants (triangles) are longer, have a greater replication timing difference from diploids, and are more significant. (*C*) Replication timing variation is attributed to alterations in haploid cells. One-way ANOVA *P*-values between haploid, diploid, and control ESC lines (CU-ES4 and CU-ES5, shown as "ESC") at haploid-delayed variants, sorted by *P*-value of haploid versus diploid comparisons (*leftmost* column). Light gray: not significant. Nineteen of the 21 haploid-delayed variants are also significantly different from controls (*middle* column), in contrast to only nine variants that vary between diploid cells and ESCs (*rightmost* column). In all but one case (variant #13), haploid-ESC variation is more significant than diploid-ESC variation, thus haploids, not diploids, have the exceptional replication timing values. (*D*) Replication timing variants are normally late-replicating regions that are further and severely delayed in haploids. Distributions of replication timing for diploid cells in nonvariant regions (i.e., normal replication timing distribution; black), and for diploid (green) and haploid (blue) cells in haploid-delayed variant regions. The distributions represent individual windows (2.5 kb) in each category. Haploid-delayed variants replicate late (on average, 1.3 SD later than the genome mean) in diploid cells and considerably later (2.5 SD later than the mean and beyond the bounds of normal S phase) in haploid cells. (*E*) The majority of haploid-delayed variants are already late-replicating in diploid cells. Scatterplot of haploid versus diploid replication timing across the genome. Nonvariant loci were defined by first removing variants, then binning the genome into 2.3-Mb regions (mean size of all replication timing variants). Haploid-delayed variants (blue) replicate late in diploid cells and even later in haploids. The size of data points is scaled by $-\log_{10}$(*P*-value). The three dots (corresponding to variants #3, #4, and #13) with both early haploid and diploid replication had the poorest *P*-values of all haploid-delayed variants ($P = 7.9 \times 10^{-6}$, 2.3×10^{-10}, and 2.8×10^{-10}). (*F*) Replication in haploid cells is delayed beyond the bounds of S-phase. The dashed line indicates a replication timing of 2 SDs below the mean, when 98.9% of the genome has already completed replication. Eighteen of the 21 replication timing delays (all but the three least significant variants #3, #4, and #13) extended beyond this value.

Because we detect regions with strongly delayed replication, we considered whether these could be due to deletions of the underlying sequence. Even though our analysis filters for regions with DNA copy number significantly different than a sample's ploidy, it is still possible that the replication variants are influenced by deletions that passed this filtering. However, several aspects of the data argue against the replication variants being chromosomal deletions. First, we expect deletions in haploid cells to have a copy number near zero. In support of this, we found that 46.2% of loci with copy number below 0.05 overlap deletions identified by the 1000 Genomes Project (Sudmant et al. 2015). In contrast, at haploid-delayed variants, although the absolute copy number was lower than the rest of the genome, it was still much greater than zero (mean of 0.86, compared to a genome average copy number of 1) (Supplemental Fig. S4A). Within haploid-delayed variants, the frequency of suspected deletions (copy number < 0.05) was extremely small (0.0037%), and the only deletions within haploid-delayed variants (variants #9 and #20) were found in both haploid and the corresponding isogenic diploid cell lines, indicating that they are not ploidy-dependent copy number variations (Supplemental Fig. S4B). These analyses, however, do not rule out the possibility of subclonal deletions, in which only a subset of the cells have a deleted region, thus giving rise to a copy number value intermediate between zero and one when analyzing a population of cells. If this were the case, we would expect different subclonal deletions to have a range of copy number values between zero and one. In contrast, all haploid-delayed variants were much closer to a copy number of one than they were to zero (Supplemental Fig. S4A). Subclonal deletions would also be expected to appear as discontinuous copy number changes compared to their flanking regions; however, DNA copy number at haploid-delays changed gradually and continuously, more consistent with replication timing changes (Supplemental Fig. S4B). Our finding that the copy number at variant regions is highly consistent between the two haploid cell lines (Fig. 1D) also argues against the replication variants being subclonal deletions, as it would be extremely unlikely for these to occur in the same location in two separate cell lines and to have a similar level of subclonality (i.e., similar copy number).

To resolve whether variation was attributed to replication timing changes in the haploid cells, the diploid cells, or both, we compared the replication timing profiles to control ESCs. The haploid cell lines showed significant differences from control ESCs at 19 out of 21 haploid-delayed variants (P-values ranging from 1.9×10^{-6} to 2.5×10^{-302}) (Fig. 2C), compared to just nine variants for the diploid cell lines (P-values ranging from 1.5×10^{-6} to 5.7×10^{-29}), which, in every case, were substantially weaker than the corresponding variation between haploid and diploid cells (Fig. 2C). The greater difference from controls of haploid compared to diploid cells suggests that replication timing delays are largely due to haploid-specific replication timing changes (justifying their designation as "haploid-delayed"). This comparison had a less clear interpretation at haploid-advanced variants, where we found a mixture of effects with greater ambiguity and more subtle differences between haploid and diploid cells (Supplemental Fig. S5A,B). Taken together, these results indicate that delayed replication in haploid cells is the predominant replication timing difference between haploid and diploid cells.

Importantly, replication timing at haploid-delayed variants was already late in diploid cells, with a mean replication timing of 1.2 SDs later than the autosome-wide mean. In haploid cells, replication was further delayed to an average of 2.5 SDs, and a maximum of 4.1 SDs, later than the mean (Fig. 2D). This brings the average replication timing of these variant regions in haploids cells to later than 99.8% of all other autosomal loci (Fig. 2D–F) and later than previously described heterochromatic regions such as centromeres (Massey et al. 2019) and the inactive X Chromosome (Koren and McCarroll 2014). Given that 98.6% of the diploid genome replicates by 2 SDs later than the mean, a delay of 4.1 SDs extends S-phase by an estimated 52.5%. In ESCs, the duration of S-phase is ~8 h, whereas G2 spans ~4 h (Becker et al. 2006); an extension of 52.5% translates into a delay of 4.2 h, potentially extending replication into late G2 phase and even into mitosis. Only three haploid-delayed variants (#3, #4, and #13) replicated earlier than the autosomal mean in haploid (as well as diploid) cells, and these were the three least significant variants (Figs. 1C, 2E; Supplemental Fig. S3A). Thus, replication timing variation between haploid and diploid ESCs comprises predominantly delays in already late-replicating regions, rendering these regions extremely late-replicating and greatly extending the bounds of S-phase, possibly into mitosis.

Replication delays in haploid cells occur in quiescent, unorganized heterochromatin

Given that haploid-delayed variants are late-replicating and show severly delayed replication, we considered whether they correspond to fragile sites—genomic regions that are prone to double-stranded breaks and are often late-replicating. However, neither haploid- delayed nor haploid-advanced variants significantly overlapped various classes of fragile sites (Methods; Supplemental Fig. S6). Haploid-delayed variants had a lower gene density than expected, even after accounting for the general sparsity of genes in late-replicating genomic regions ($P = 0.011$) (Fig. 3A). The 181 genes within haploid-delayed variants were enriched for the Gene Ontology terms keratinization (23 genes, FDR = 1.49×10^{-19}) (Supplemental Table S1), which was entirely attributed to the *KRTAP* gene cluster in variant #21, and cell-cell adhesion via plasma-membrane adhesion molecules (nine genes across six variants, FDR = 3.34×10^{-3}). There was a significant enrichment ($P = 0.03$) (Figs. 1C, 3A) within haploid-delayed variants for genes previously found to be down-regulated in haploid cells (Sagi et al. 2016), as well as a strong correlation (r = 0.72) between the magnitude of haploid replication delay and the extent of gene expression down-regulation (Fig. 3B). Down-regulated genes within haploid-delayed variants were modestly enriched for cell-cell adhesion genes (FDR = 0.014) and components of the membrane (FDR = 0.033) (Supplemental Table S1). Similarly, haploid-advanced regions were enriched for genes up-regulated in haploid cells (10 genes compared to an expected 0.48; $P = 6.6 \times 10^{-34}$) (Supplemental Fig. S3B).

We next compared the haploid replication timing variant locations to ESC data for 18 ChromHMM states, 31 histone marks, sites of DNase I hypersensitivity (Roadmap Epigenomics Consortium et al. 2015), and chromosome conformation (topological domains and chromatin loops) (Dekker et al. 2017). Haploid-delayed variants were significantly enriched for the heterochromatin and quiescent (chromatin lacking histone marks) (Hoffman et al. 2013) states ($P = 1.5 \times 10^{-5}$ and 2.8×10^{-3}, respectively) (Fig. 3D) and otherwise significantly depleted for most other ChromHMM states. In total, 84.4% of haploid-delayed variant regions were quiescent, whereas 8.4% were heterochromatic. Of the 32 specific chromatin marks we tested, 31 were nominally depleted in haploid-delayed variants, much more than expected

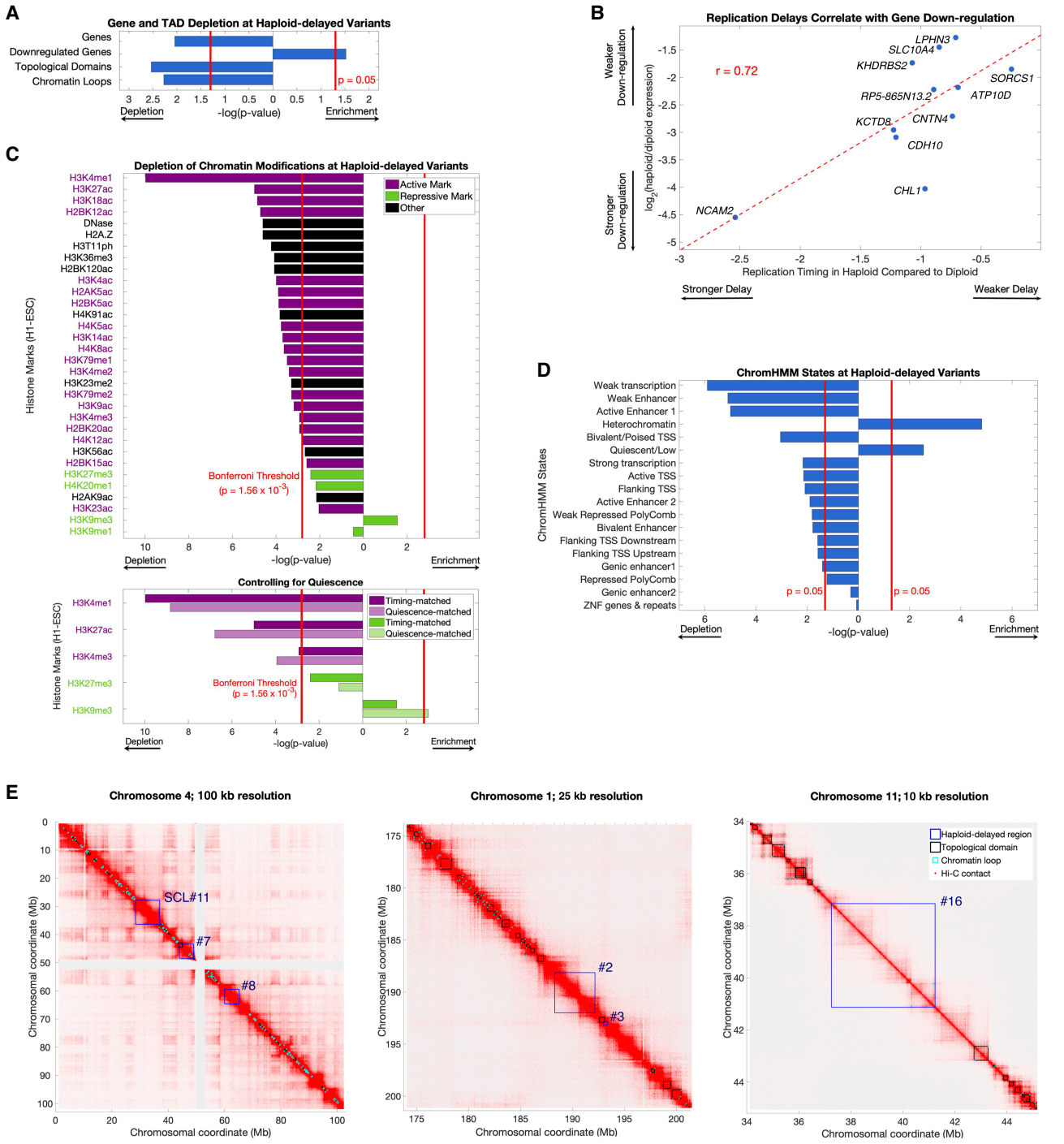

Figure 3. Haploid-delayed variants are located in regions of unorganized, quiescent heterochromatin depleted of genes, histone marks, and 3D contacts. (*A*) Haploid-delayed variants show a significant depletion of genes, topologically-associated domains (TADs), and chromatin loops, and an enrichment of genes down-regulated in haploid ESCs. Bars indicate enrichment/depletion two-sided *t*-test *P*-value of each tested feature in replication timing variants compared to 1000 matched permutations. (*B*) The extent of haploid replication delay correlates with the magnitude of gene expression differences between haploid and diploid cells. Spearman's rank correlation was similar (rho = 0.65); thus, this correlation is not driven by *NCAM2*. (*C*) Haploid-delayed variants are depleted of chromatin marks. Bars shown as in *A*. Each chromatin mark was considered to be independent, and the Bonferroni-corrected *P*-value 0.0016 (0.05/32) is shown as a solid red line. Histone mark activation/repression information was obtained from Schiltz et al. (1999) and Zhao and Garcia (2015). *Bottom*: Three of five tested histone marks (chosen due to their similarity with UR regions) remain depleted when controlling for quiescence, whereas H3K27me3 remained nominally depleted, and H3K9me3 became significantly enriched. *P*-values for matched permutations (dark bars) are identical to the *upper* panel, with *P*-values for corresponding quiescence-matched permutations (light bars) shown *below* them. (*D*) As in *A* and *C*, for ChromHMM states. Overlap was calculated by the number of base pairs of each state at variant regions. ChromHMM states are mutually exclusive (i.e., not independent); thus, a *P*-value cutoff of 0.05 was used for significance. (*E*) Hi-C chromosome interaction maps in H1-ESC (Dekker et al. 2017) at three different genomic locations and resolutions. Haploid-delayed variants (dark blue) have little overlap with chromosome domains (black) or loops (light blue). This trend is observed across all haploid-delayed variants (panel *A*).

(binomial $P = 7.45 \times 10^{-9}$) (Fig. 3C). These depletions were statistically significant for 25 of the marks, with a particularly prominent depletion of the active histone mark H3K4me1 ($P = 1.04 \times 10^{-10}$). The only nominally enriched mark at haploid-delayed variants was the repressive mark H3K9me3, which is typically associated with constitutive herterochromatin ($P = 0.04$; not significant after Bonferroni correction).

Haploid-delayed variants were also significantly depleted for both chromosome conformation topological domains ($P = 0.0037$) and chromatin loops ($P = 0.0067$) compared to permutations (Fig. 3A). Haploid-delayed variants occupied regions devoid of topological domains (Fig. 3E). Such regions are stratified based on size into small regions (<50 kb) that constitute topological boundaries and larger regions (>50 kb) that are considered to be "unorganized heterochromatin" (Dixon et al. 2012). The large size (1.62 Mb to 14.7 Mb) of the regions containing haploid-delayed variants suggests that they all fall into the latter category.

Neither replication timing nor the quiescent state explained the observed depletions, as genes ($P = 0.033$), toplogical domains ($P = 0.028$), and chromatin loops ($P = 0.029$) remained significantly depleted from haploid-delayed variants after controlling for quiescence (Methods), and down-regulated genes remained enriched ($P = 0.002$). Histone mark enrichment patterns were also independent of the quiescent state (Fig. 3C). In summary, haploid-delayed variants are late-replicating, depleted of genes, devoid of almost all chromatin marks, and show reduced chromatin contacts and generally unorganized heterochromatin. They thus represent a set of genomic locations with a strong replication phenotype yet are distinct from previously characterized genomic fragile sites.

Haploid replication delays correspond to sites of DNA underreplication in mouse placenta polyploid cells

We showed above that a prominent replication aberration in haploid cells is severe delays at 21 regions throughout the autosomes. An attractive possibility is that the severity of these delays, which extends DNA replication well beyond the normal bounds of S-phase, may be related to the frequent diploidization of haploid cells. Furthermore, it is possible that similar replication aberrations occur during physiological polyploidization. To begin to test this, we considered analogous instances of replication abnormalities in polyploid tissues. In particular, trophoblast giant cells of the mammalian placenta undergo successive rounds of genome duplication, giving rise to cells with ploidies as high as 1000 N (Zybina and Zybina 1996). These cells do not exhibit uniform DNA copy number; instead, a study in mouse TGCs found large genomic regions with reduced DNA copy number in polyploid cells that was thought to be due to DNA underreplication during the multiple replication cycles (to our knowledge, equivalent studies in human placental cells have not been carried out). Underrepresented (UR) regions gradually accumulate over time during TGC development, becoming both larger and more numerous in successive cell cycles. Similar to the delayed variants in haploid ESCs, UR regions are late-replicating regions in trophoblast stem cells (Hannibal et al. 2014).

As DNA replication timing is largely conserved between human and mouse (Ryba et al. 2010; Yaffe et al. 2010), we compared 45 UR regions found across mouse TGC development to the 21 regions with replication timing delays in haploid human ESCs (Methods). Eleven of these UR regions each overlapped a separate haploid-delayed variant (Fig. 1C; Supplemental Fig. S3A). When compared to matched random permutations (which maintained the number, size, and replication timing of the haploid-delayed variants) (Methods), this overlap was highly significant ($P = 1.2 \times 10^{-5}$ permuting haploid-delayed variants; $P = 4.0 \times 10^{-3}$ controlling for quiescence; $P = 1.4 \times 10^{-9}$ permuting UR regions) (Fig. 4A). The similarity between mouse TGC UR regions and haploid-delayed variants was even more substantial considering that the 11 UR regions corresponded to 13 of the most significant ($P < 10^{-30}$) haploid-delayed variants (Fig. 1C; Supplemental Fig. S3A). In addition, when considering the same UR regions but using the broader genomic coordinates present at late stages of TGC development (Hannibal et al. 2014), we found a substantially increased span of overlap: six of the 11 haploid-delayed variants showed complete overlap with a UR region, whereas the other five showed substantial (>48%) overlap along their lengths. This amounted to 64% of the total length of haploid-delayed variants showing co-occurance with UR regions (Fig. 4B). The most significant single cell line variant (SCL#11) also encompassed an entire UR region (Fig. 1C). Mouse UR regions did not significantly overlap haploid-advanced variants ($P = 0.94$); however, the strongest haploid-advanced variant (#28) (Supplemental Fig. S3B) showed near complete correspondence to a UR region (100% overlapped by UR region, and 91% coverage of the UR region). Thus, there is a compelling correspondence of mouse TGC UR regions and haploid ESC-delayed variants, despite the different cell types and species.

UR regions in mouse TGCs are depleted for H3K4me1, H3K27ac, H3K4me3, and H3K27me3 and enriched for H3K9me3 (Hannibal et al. 2014)—the exact same trends observed for haploid-delayed variants (Fig. 3C). Furthermore, there was a correlation in the level of enrichment of these five histone marks between haploid-delayed and mouse UR regions (Fig. 4C). In further similarity to haploid-delayed variants (Fig. 3A), UR regions were depleted of genes compared to the rest of the genome (Hannibal et al. 2014). The genes that were found in mouse UR regions were enriched for cell adhesion and neurogenesis annotations, similar to haploid-delayed variants, which contained both cell-adhesion genes and several categories of genes related to the nervous system (GABA-A receptor activity, neuron part, and post-synaptic membrane annotations) (Supplemental Table S1).

Taken together, there is a strong correspondence of mouse polyploid TGC underreplicated regions and haploid-delayed variants in terms of genomic location, replication timing, histone composition, and gene content. These similarities could represent common biological mechanisms linked to whole genome duplication. A key difference is that delayed replication in haploid cells cannot possibly be a consequence of diploidization, because it precedes it.

Replication delay in haploid cells is linked to X-Chromosome dosage

We next considered possible mechanisms leading to the replication defects in haploid cells. In particular, we considered whether the relative increase in X-linked gene expression in haploid cells (Sagi et al. 2016) could influence autosomal replication timing and account for the replication delays we observed. To test this, we first sought to find other instances of increased X Chromosome expression dosage in ESCs. Female ESCs occasionally undergo partial or complete reactivation of the inactive X Chromosome (Patel et al. 2017). We therefore searched for cell lines with evidence of a reactivated X Chromosome by utilizing the replication profiles of 116 human ESCs (Ding et al. 2020). Although replication timing was highly correlated among cell lines along the autosomes (mean $r = 0.9$) as well as the X Chromosome in male

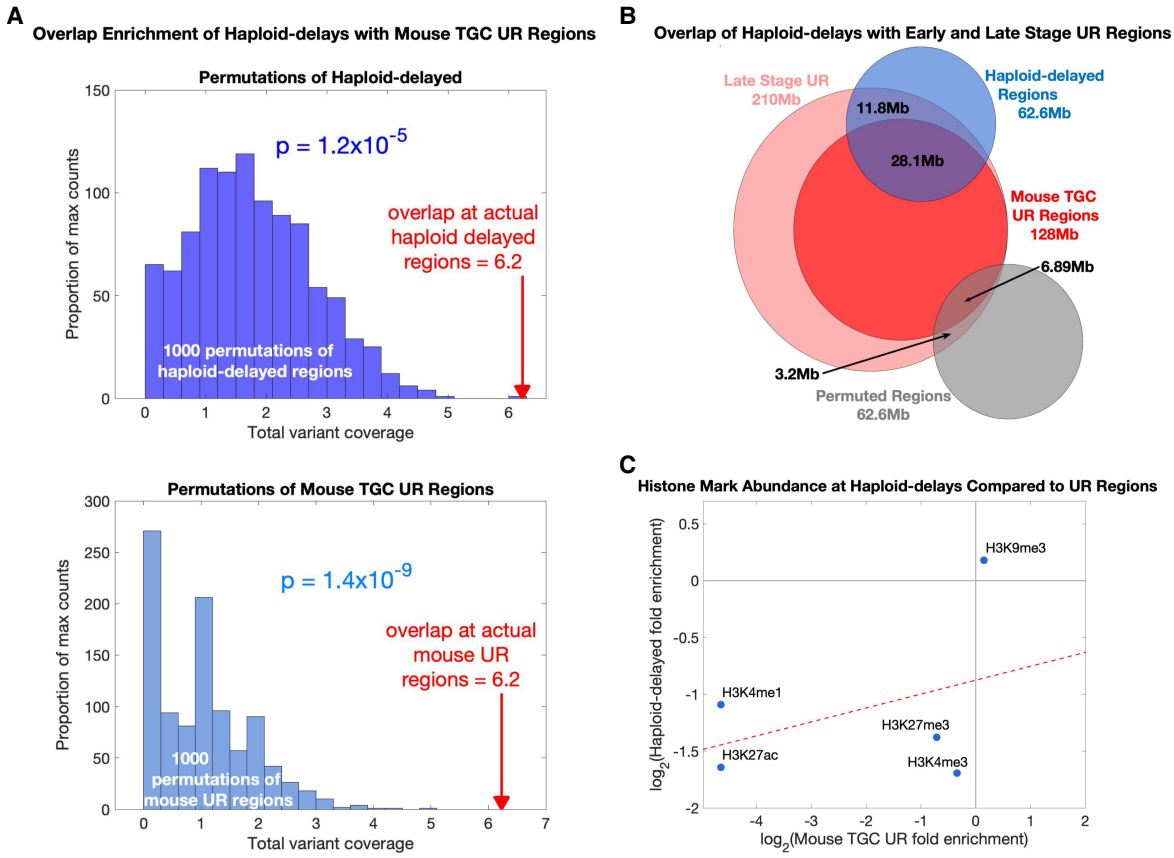

Figure 4. Regions of replication delay in haploid cells corresponds to underreplication in polypoid cells. (*A*) Haploid-delayed variants significantly overlap UR regions in mouse TGCs compared to matched permutations. *Top* histogram (dark blue): degree of overlap between mouse UR regions and 1000 permutations of haploid-delayed variants; *bottom* histogram (light blue): degree of overlap between haploid-delayed regions and 1000 permutations of UR regions. Red arrows: overlaps between actual haploid-delayed variants and actual mouse UR regions. Each overlap was counted as the fraction of the haploid-delayed variant that overlapped a UR region (i.e., a value between 0 and 1 for each variant, with a maximum overlap value of 21 corresponding to the 21 haploid-delayed variants). (*B*) Venn diagram comparing haploid-delayed variants (blue) to mouse UR regions (red). Dark red: shared regions found in all mouse placenta; light red: regions only identified at later stages of development. Gray: randomly selected permutation; that is, one of the 1000 permutations in the *top* histogram of *A*. (*C*) Histone mark enrichments at haploid-delays resemble those in mouse TGCs at UR regions.

(karyotypically XY) samples (r = 0.91), it was significantly less correlated on the X Chromosome of female samples (mean r = 0.83) (Supplemental Fig. S7A), as expected due to the random replication of the inactive X Chromosome (Koren and McCarroll 2014) (similar patterns were observed when comparing haploid and diploid cells) (see Supplemental Fig. S2B). However, we identified nine female samples with X Chromosome replication timing patterns that resembled male more than they resembled other female cell lines. These cell lines were identified by their high correlations to male samples on the X Chromosome (those with r > 0.89) (Supplemental Fig. S7B), yet also showed higher correlations among them (0.86–0.92; similar to autosomes or the X Chromosome in male samples) and lower correlations to other female samples (0.78–0.84). Furthermore, the X Chromosomes of these cell lines replicated earlier than other female samples (–0.30 compared to –0.87) (Supplemental Fig. S7C). We suspect that these female cell lines may have undergone partial or complete reactivation of the inactive X Chromosome.

When comparing the nine cell lines with suspected X Chromosome reactivation (designated "XaXa") to the other 107 ESCs, we identified 214 autosomal regions with subtle, yet significant replication timing variation between the two (Methods). This suggests that the dosage of active X Chromosomes may be linked to replication timing alterations genome-wide. We found that 36 of the 214 reactivated-X variants overlapped 16 of the 21 haploid-delayed variants (Fig. 1C). This represents a highly significant enrichment compared to expectation ($P = 6.4 \times 10^{-32}$ when permuting haploid-delayed variants; $P = 6.4 \times 10^{-33}$ when controlling for quiescence; $P = 2.5 \times 10^{-63}$ when permuting reactivated-X variants) (Fig. 5A). Furthermore, six reactivated-X variants overlapped five different haploid-advanced variants, again much more than expected by chance ($P = 9.0 \times 10^{-12}$ and $P = 0.031$ when permuting haploid-advanced variants or reactivated-X variants, respectively) (Fig. 5B). All of the 42 overlapping regions showed a consistent direction of replication timing change in XaXa ESCs and in haploid ESCs, such that when haploids were delayed (or advanced) relative to diploids, XaXa ESCs were also delayed (or advanced) relative to XaXi ESCs. Across all these genomic regions, XaXa and haploid replication timing differences were well correlated (r = 0.32) (Fig. 5C). These results are consistent with our premise by which replication timing alterations in haploid cells could be related to the elevated dosage of X Chromosome gene expression.

To more directly test whether X Chromosome activity could be causing the replication timing alterations in haploid cells, we

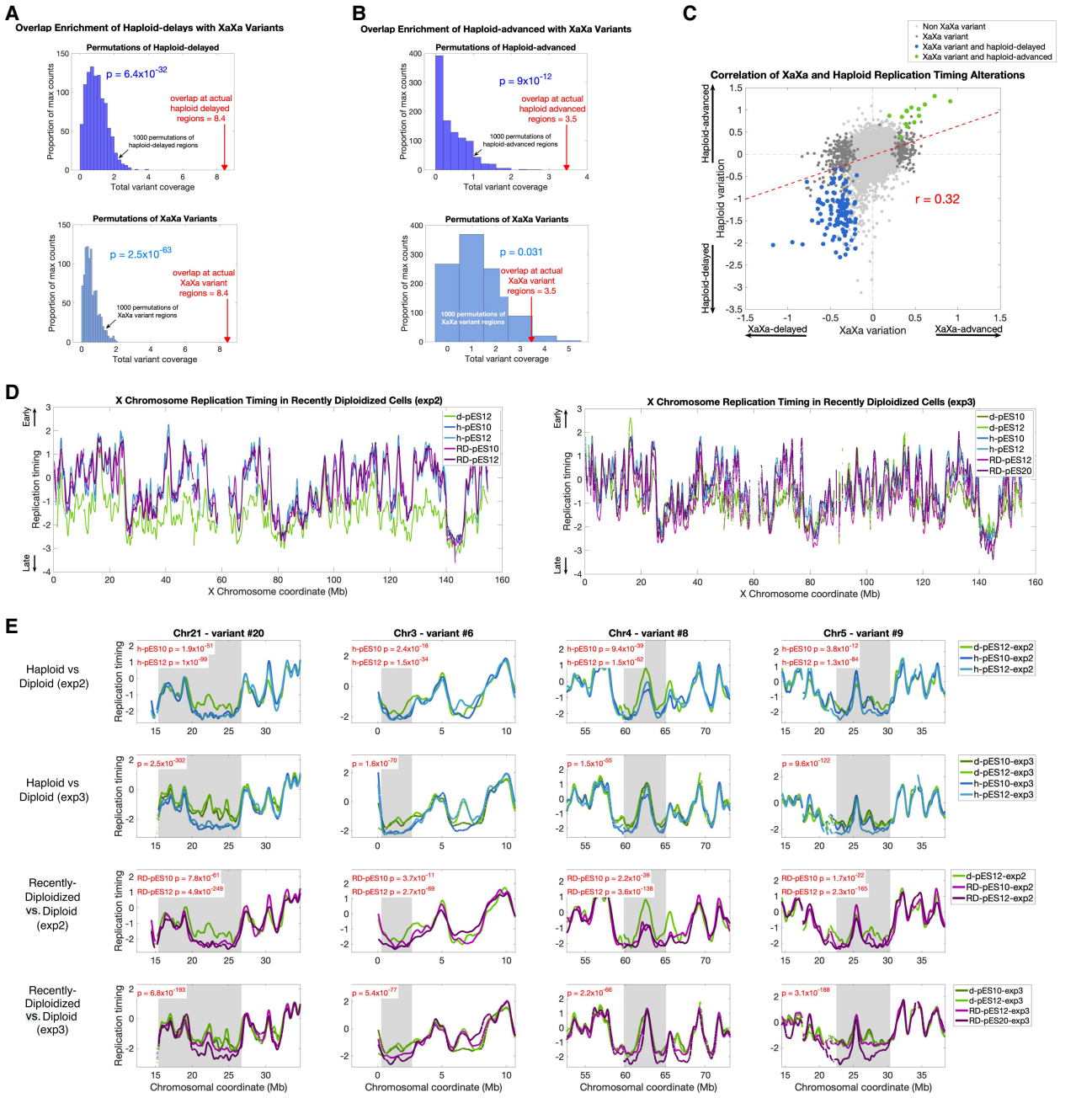

Figure 5. Regions of replication delay in haploid cells are correlated with the level of X Chromosome activity. (*A*) Haploid-delayed variants significantly overlap XaXa variant regions compared to matched permutations. Same as in Figure 4A; dark blue: permutations of haploid-dealyed regions, light blue: permutations of XaXa variant regions. (*B*) Haploid-advanced variants significantly overlap XaXa variant regions. Same as *A*, for haploid-advanced variants. (*C*) Replication timing differences between XaXa females and other ESCs (*x*-axis) compared to haploid versus diploid replication timing variants (*y*-axis). For presentation purposes, data was downsampled to one window every 200 kb of uniquely alignable sequence. (*D*) X Chromosome replication timing profiles for haploid, recently diploidized, and diploid ESCs. Haploid and recently diploidized cell lines show comparable replication timing along the X Chromosome that is earlier than diploid cells. This trend is more pronounced in experiment 2 (*left*) than in experiment 3 (*right*). (*E*) Replication timing in recently diploidized ESCs resembles haploids more than it resembles diploids, thus tracking X Chromosome status rather than ploidy. For experiment 2, *P*-values (*t*-test) are shown for each cell line compared to the single diploid cell line in that experiment; for experiment 3, a single *P*-value (ANOVA) is shown for the joint comparison of the two haploid cell lines with the two diploid cell lines. See Supplemental Table S2 for statistics at other haploid-delayed variants.

profiled replication timing in recently diploidized (RD; haploid cell cultures shortly after becoming diploid) alongside haploid and diploid (XaXi) cells. X-Chromosome inactivation occurs only several cell divisions after cells become diploid; thus, shortly after diploidization, the two X Chromosomes are both active (Sagi et al. 2016). This was supported by our data, as X Chromosome replication timing in recently diploidized cells was similar to haploid cells and much earlier than in later-stage

diploid cells, consistent with the presence of two active X Chromosomes (Fig. 5D).

Across two separate experiments (experiments "2" and "3"), we reproduced between 15 and 18 (71%–86%, three delays only validated in h-pES12) of the 21 variants in experiment 2, and 18 (86%) in experiment 3 when comparing haploid and stably diploid cells (Fig. 5E; Supplemental Table S2). The replication delays in these repeat experiments were dampened compared to the original experiment; we suspect that this is related to "culture adaptation," in which late-passage, serially sorted haploid cell lines become more stable in the haploid state and undergo reduced rates of diploidization.

Across the majority of variant regions, RD cell lines showed the same replication delays as haploid cells and were distinct from stably diploid cells. Of the 21 haploid-delayed variant regions, between 14 and 19 (67%–90%; experiment 2; five variants found only in RD-pES12) and 19 (90%; experiment 3) were also significantly delayed in RD cell lines compared to later-stage diploid cell lines, despite both being diploid (Fig. 5E; Supplemental Table S2). In particular, of the 15 regions delayed in all haploid cells across all three experiments, all were also delayed in all RD cell lines. One of the RD cell lines (RD-pES12-exp3) showed comparatively weaker delays than other RD cells (Fig. 5E) and also showed X Chromosome replication timing more consistent with later-stage diploids; this may indicate that this cell line partially inactivated one X Chromosome, which is consistent with it having been cultured for many (34) passages. Taken together, haploid-delays correlate more strongly with the dosage of active X Chromosomes than with ploidy per se. This provides compelling support to the notion that X Chromosome dosage could underlie the DNA replication delays in haploid ESCs.

Finally, the repeat experiments and inclusion of recently diploidized cell lines further clarified the haploid-diploid differences on the X Chromosome. Specifically, although the X Chromosome generally replicated earlier in haploid compared to diploid cells, we noticed two regions that seemed not to exert this difference (Fig. 1A). Although this observation in the initial experiment was suspected to result from technical noise, these regions reproducibly showed this same effect in both repeat experiments as well as in the RD cell lines (Supplemental Fig. S7D). We thus propose that these are likely regions of haploid replication delays on the X Chromosome. This possibility is also consistent with these two regions being long and late-replicating, similar to autosomal haploid-delayed regions. Thus, we suggest that there are a total of 23 regions across the genome identified here as having delayed replication in haploid ESCs.

Discussion

Haploid human stem cells provide a powerful system for human genetic studies as well as a unique opportunity for investigating the biology of cell ploidy. However, their spontaneous diploidization is a major limitation to their use and remains a poorly understood phenomenon. Here, we show that DNA replication is delayed well beyond the bounds of S-phase at multiple regions in human haploid embryonic stem cells. The replication delays we describe are among the most profound alterations to the otherwise highly stable eukaryotic DNA replication timing program. The ability to identify replication delays that transcend S-phase was enabled by our approach of directly sequencing DNA from proliferating cell cultures; previous replication profiling approaches that utilize FACS sorting of S-phase cells (Hulke et al. 2020) would likely have missed these extreme delays, as G2 cells are not included and thereby the valleys of very late replicating regions are not fully captured.

Haploid-delayed regions have a distinctive genetic and epigenetic signature, characterized by late replication, a paucity of genes, limited histone modifications, and reduced chromatin contacts. They differ from previously described chromosomal fragile sites. In contrast, they show profound similarity to UR regions in mouse placenta trophoblast giant cells. Thus, these regions may represent a novel class of coregulated genomic sites that are susceptible to abnormal replication and linked to polyploidization in the placenta and potentially other cell types. Finally, we provide compelling evidence implicating the dosage of X Chromosome activity, rather than the haploid state per se, in replication delays. Similar replication delays were observed in ESCs with evidence of a reactivated X Chromosome and, more importantly, in diploidized ESCs with two active X Chromosomes.

What causes replication delays in haploid cells?

We can envision two mechanisms by which X Chromosomes could cause replication delays. First, the mere presence of an inactive X Chromosome may be the critical factor. An inactive X Chromosome may recruit heterochromatin factors, sequestering them from other chromosomes. An absence of an inactive X, on the other hand, may release such factors to bind to specific autosomal regions and delay their replication. This is similar to the "chromatin sink" model suggested for the highly heterochromatic *Drosophila* Y Chromosome (Francisco and Lemos 2014). This hypothesis can be tested with human cell lines lacking an inactive X Chromosome, such as Turner syndrome (X0) cells. We profiled replication timing in an XX H9 ESC line and an X0 progeny that was cloned after spontaneous loss of one of the X Chromosomes. Consistent with absence of the inactive X Chromosome, we observed earlier replication of the X Chromosome in X0 cells (after multiplying copy number by two prior to normalization to enable comparison) compared to XX (Supplemental Fig. S7E). We only identified four of the 21 haploid-delayed regions (as well SCL#11) as delayed in X0 compared to XX ESCs (Supplemental Fig. S7E). Although notable, the lack of more substantial correpsondence between X0 delays and haploid-delayed regions suggests that the absence of an Xi is insufficient to explain the full complement of haploid replication delays.

An alternative hypothesis is that overexpression of an X-linked gene(s) contributes to the observed replication delays. Thus, a single active X Chromosome in haploid cells, or two active X Chromosomes in diploid cells, could produce an elevated level of a transcript(s) compared to normal XaXi diploid cells. This overexpression could then potentially cause replication delays at specific loci across the genome. Several X-linked genes are possible candidates for mediating autosomal replication timing delays. For instance, *ELK1* encodes an X-linked transcription factor primarily expressed in placenta and ovarian tissue (Uhlén et al. 2015) that was shown to cause transcriptional changes of autosomal genes in reactivated-X ESCs (Bruck et al. 2013); *PPP2R3B* encodes a protein phosphatase 2A subunit that delays the firing of replication origins throughout the genome by stabilizing the CDC6-CDT1 interaction (van Kempen et al. 2016); *BCOR*, a Polycomb-group repressive complex gene, is required for normal placental development (Hamline et al. 2020); and *NAP1L2*, which encodes a nucleosome assembly protein, is induced during differentiation of mouse trophoblast stem cells to TGCs (Attia et al. 2007) and is up-regulated in human haploid ESCs (Sagi et al. 2016). The role of these and other genes in haploid replication

delays could be tested by knockout, knockdown, or overexpression in haploid and diploid ESCs.

Can delayed replication cause diploidization?

The mechanisms causing haploid cells to diploidize, or diploid cells to become polyploid, are not fully understood. It is intriguing to consider whether the severe replication delays we described in haploid ESCs could induce diploidization and whether similar mechanisms could be operative in other instances of polyploidization. It is known that incomplete DNA replication leads to an ATR-dependent activation of the S-M checkpoint that prevents cells from prematurely entering mitosis (Enoch et al. 1992; Eykelenboom et al. 2013). Thus, it is possible that replication delays ultimately lead haploid cells to avert mitosis and re-enter the cell cycle, making them diploid. In support of this possibility, it was previously shown that in human cancer cell lines, chromosome-scale replication delays and associated delays in mitotic chromosome condensation activate the S-M checkpoint and result in endoreplication in a subset of cells (Chang et al. 2007). Disruption of the replication initiation factors ORC2 or GINS2 has also been shown to induce polyploidization in human cells (Rantala et al. 2010; Huang et al. 2016). Similarly, it was recently shown that ~2% of the maize genome is delayed in its replication in endocycles compared to normal mitotic cycles (Wear et al. 2020). Polyploidization is also very common in cancer (Bielski et al. 2018), and so is replication stress; it is intriguing to consider whether these two phenomena are related to each other. To further test these links, we induced replication delays in haploid ESCs using aphidicolin, a DNA polymerase inhibitor. Over three independent experiments, we observed a significant increase in diploidization rates 48 h after aphidicolin treatment, consistent with incomplete replication promoting diploidization of these cells (Supplemental Fig. S7F). Another observation supporting a general restructuring of the cell cycle is that not every chromosome contained regions of replication delay in haploid cells. This suggests that diploidization is not due to replication delays directly causing mitotic chromosome segregation defects on the chromosomes on which they occur, but rather that one or more delayed regions can activate a global cellular response (e.g., a checkpoint) that could lead to diploidization.

The fate of unreplicated DNA during mitotic entry has been studied before—for instance, in the context of common fragile sites following replication stress. A mechanism, termed mitotic DNA synthesis (MiDAS), has been described in which DNA repair synthesis is initiated at chromosomal gaps or breaks during mitosis. MiDAS utilizes a pathway resembling RAD52-dependent break induced replication (BIR) (Minocherhomji et al. 2015; Bhowmick et al. 2016). This raises the possibility that lack of homologous chromosomes renders haploid cells particularly vulnerable to incomplete replication because they are unable to perform MiDAS. Delayed replication in haploid cells would thus induce a robust checkpoint response. In contrast, diploid cells do not suffer the replication delays that require replication completion in mitosis, whereas recently diploid cells have replication delays similar to haploids but are competent at MiDAS given the presence of homologous chromosomes. Indeed, we do not observe frequent tetraploidization of recently diploid cells despite them having replication delays. This model would also imply that haploid cells are able to eventually complete genome replication and/or repair any associated DNA damage following diploidization. Consistently, we do not observe any gross or recurrent genome rearrangements in diploidized cells. Such repair synthesis could potentially occur in TP53BP1 (also known as 53BP1) nuclear bodies during G1-phase following diploidization (Harrigan et al. 2011; Lukas et al. 2011; Minocherhomji et al. 2015; Bhowmick et al. 2016).

In support of the idea that haploid-delays are linked to whole genome duplication, we observed a remarkable correspondence between these delays and UR regions in mouse TGCs. Haploid delays and URs are found in corresponding genomic locations, have similar replication timing, contain nearly identical histone patterns, and are enriched for some of the same gene categories. The replication factor Rif1, which is required for underreplication in *Drosophila* (Munden et al. 2018), was also suggested to be important for underreplication in mouse TGCs (Hannibal and Baker 2016). Furthermore, the inactive X Chromosome in mouse TGCs is unusual in harboring both the heterochromatic mark H3K27me3 and euchromatic marks such as H3K4me2 (Corbel et al. 2013) and a high fraction of genes that escape X-Chromosome inactivation (Schoenfelder and Fox 2015). However, in male embryo pregnancies, placental cells only carry a single X Chromosome, thus the mechanism of polyploidization in TGCs is likely independent of X-Chromosome dosage, although it may still be related to a similar gene circuitry as the one putatively disrupted in human haploid ESCs. Mouse UR regions were originally identified using microarray genomic DNA hybridization (Hannibal et al. 2014). However, TGC profiling using high-throughput sequencing suggests that DNA copy number decreases gradually rather than sharply at UR regions (see Fig. 2 in Hannibal et al. 2014 and Fig. 1 in Hannibal and Baker 2016). This, and the absence of evidence for chromosomal deletions at UR regions using paired-end sequencing (Hannibal et al. 2014) may be more consistent with the interpretation that UR regions represent severely delayed replication, similar to haploid delays, rather than DNA loss.

Further work is required in order to understand the interplay between X-Chromosome dosage, transcriptome remodeling, replication dynamics, and whole genome duplication. Key remaining questions are why haploid cells become diploid at high rates whereas the diploid state is much more stable, and whether similar replication-related mechanisms, whether dependent on X Chromosome activity or not, contribute to polyploidization in TGCs and other tissues. Extending our analysis to mouse haploid cells, androgenetic embryonic stem cells, differentiated cells, and genetically manipulated cells will shed light on the fundamental links between genome regulation and ploidy control and could ultimately enable the stabilization of the haploid state in human ESCs. These efforts would be greatly facilitated by identifying potential sequence elements that control replication timing at haploid-delayed regions (Ding et al. 2020).

Our results suggest that replication dynamics in S-phase have the potential to influence the entire cell cycle. It is thus critical for cells to maintain their temporal order of DNA replication. As a corollary, certain DNA sequences can have a physiological role by virtue of their replication properties rather than their actual coding potential. This could ascribe a developmental function to late DNA replication, in particular by various mammalian cell types (Sagi and Benvenisty 2017) exploiting site-specific DNA replication delays in order to become polyploid.

Methods

Cell culture

Haploid pESCs, diploid pESCs, and control ES cell lines were cultured and maintained as previously described (Sagi et al. 2016).

Briefly, we used StemFlex media on Geltrex matrix at 37°C with 5% CO_2 and atmospheric oxygen concentration. Passaging was performed with trypsination via incubation with TryplE and culturing newly passaged cells with 10 µM ROCK inhibitor (Y-27632) for 24 h. Freezing was performed with media consisting of 10% DMSO and 40% FBS. Haploid and diploid cell lines were FACS-sorted in G1-phase and plated and cultured for 3–7 d in order to ensure the desired ploidy as well as the cell cycle asynchrony of the cultures (which is a prerequisite for the replication timing assay).

Flow cytometry

For cell cycle analysis with EdU and Phospho-Histone H3 antibody staining, cells were cultured in StemFlex media (Thermo Fisher Scientific A3349401) on Geltrex (Thermo Fisher Scientific A1413302) until 70% confluency. To label S-phase, cells were incubated in warm medium with 10 µM EdU (Thermo Fisher Scientific C10337) for 30 min followed by a warm medium change without EdU. Cells were then rinsed with PBS (Thermo Fisher Scientific 14190-250), trypsinized with TrypLE Express (Thermo Fisher Scientific 12605036) for 5 min, and suspended in 4% PFA in PBS (Santa Cruz Biotechnology sc-281692) for 15 min. Cells were pelleted and suspended in 0.1% Triton X-100 (Sigma-Aldrich T8787) for 30 min. After the EdU Click-iT protocol was followed as per kit instructions, cells were pelleted and suspended in 10% FBS (Gemini Bio-Products 900-108) in PBS for 30 min, followed by incubation with primary antibody against Phospho-Histone H3 (PH3) (Thermo Fisher Scientific 06-570-MI) at 1:1000 for 4 h. Cells were pelleted and washed in 10% FBS in PBS and incubated with secondary antibody (Thermo Fisher Scientific A-21245) at 1:500 and Hoechst 33342 (Invitrogen H3570) at 1:10,000 for 1 h. Cells were pelleted and washed in 10% FBS in PBS and analyzed by flow cytometry with the FACS-Aria machine at the Columbia University Stem Cell Initiative flow cytometry core. Populations were gated first for cells, followed by gating for single cells. EdU was observed through AlexaFluor 488 and mitotic cells exhibiting PH3 were observed through AlexaFluor 647. Analysis was performed using FlowJo software (BD Biosciences).

Whole-genome sequencing

DNA was extracted with the MasterPure DNA purification kit (Lucigen). Libraries were prepared using the Illumina TruSeq PCR-free library preparation kit, and sequencing was performed on the Illumina NextSeq 500 with 75-bp single-end reads (first experiment), or the Illumina HiSeq X Ten with 150-bp paired-end reads (second and third experiments). For backwards-compatibility with existing hESC replication timing data, reads were then aligned to the hg19 human reference genome using BWA-MEM (Li and Durbin 2010). Haploid-delayed regions were not substantially altered in the hg38 reference genome compared to hg19.

In the first experiment, haploid and diploid pES10 and pES12 were sequenced alongside three control ESCs. One control ESC (B123) showed poorer correlations and autocorrelations than other samples and was removed from further analysis.

In the second experiment, h-pES10, h-pES12, RD-pES10, RD-pES12, and d-pES12 were sequenced. In the third experiment, we again sequenced h-pES10, h-pES12, d-pES10, and d-pES12 (two separate samples), as well as RD-pES12 and RD-pES20. RD-pES12 was sequenced at passage 9, whereas RD-pES20 was sequenced at 34 passages. One of the two d-pES12 samples was removed from further analysis due to poor data quality.

Generation of DNA replication timing profiles

In order to infer replication timing, we first used GenomeSTRiP to infer DNA copy number across the genome (Koren et al. 2014; Handsaker et al. 2015). Sequence read depth was calculated in 2.5-kb windows along the genome, corrected for alignability and GC content. Copy number values for both haploid and diploid cells were normalized to an average DNA copy number of two. These copy number values were then filtered as follows:

1. Windows spanning gaps in the reference genome were removed.
2. Windows with copy number greater than one above or below the median copy number were removed.
3. In order to remove extreme data points, the data was segmented using the MATLAB function *segment*, which groups consecutive data points into segments based on a tolerance threshold. This analysis was done twice using two different segmentations parameters of 0.5 (less strict) and 0.1 (more strict). By using two different parameters, both shorter and larger genomic regions that deviate from the median can be captured. Segments falling above or below the median by a threshold of 0.45 copies were removed.
4. Genomic regions that were further than 30 kb from other data points and that were <300 kb long were removed.
5. Regions shorter than 100 kb between removed data points were removed.
6. Regions shorter than 500 kb between three or more removed data points were removed.

Data were then smoothed using the MATLAB function *csaps* with smoothing parameter of 10^{-17}, and then normalized to an autosomal median of 0 and a standard deviation of 1, such that positive values represent early replication and negative values represent late replication.

Identifying replication timing variations

In order to identify replication timing differences between groups of samples (e.g., haploid and diploid cell lines), we used ANOVA across "regions" tiling each chromosome. ANOVA tests the null hypothesis that all samples come from a population with the same mean versus the alternative that each group (haploid or diploid in this case) is drawn from populations with differing means. ANOVA was applied on filtered, raw (unsmoothed) replication timing values. Both the region size and overlap between adjacent regions were optimized by finding the false discovery rate (FDR) for a given set of parameters. To determine the FDR, ANOVA was repeated comparing permuted samples to disrupt the haploid-diploid comparison; that is, h-pES10 and d-pES12 were compared to h-pES12 and d-pES10. Because this permuted scan compares neither cell ploidy nor genetic background (pES10 vs. pES12), we considered significant regions arising from this test to be false. By dividing the total length of these false regions (after several filtering steps—see further below) by the total size of autosomal variants (after the same filtering steps) found between the haploid and diploid cell lines, we determined an FDR for regions of the chosen length. We chose a region size of 76 replication timing data windows (covering 190 kb of uniquely alignable sequence), with a slide of a quarter region (such that each genomic locus was tested four times) in order to optimize the specificity and sensitivity of variant detection; this resulted in an FDR of 0.07. Using this FDR, we identified 14.2-fold more autosomal variation in the haploid-diploid comparison compared to the permuted sample comparison.

Regions with a Bonferroni-corrected ANOVA $P < 9.58 \times 10^{-7}$ (0.05/52,184 regions tested; note that we stringently regard each

region as independent) were merged into continuous replication timing variants, of which an initial 57 were identified. Then, we removed variants in which pairwise comparisons of haploid and diploid profiles showed differences in the direction of effect (i.e., one haploid-diploid pair showed earlier haploid replication whereas another showed earlier diploid replication), trimmed variants in which any pair of a haploid and a diploid cell line had overlapping replication timing profiles (i.e., one pair was not variant in a given region), and removed any variants that, after being trimmed, were shorter than the original tested region size (190 kb) or that fell below the significance *P*-value threshold. This resulted in 53 autosomal variants. These regions were then extended in both directions as long as all haploid and diploid profiles remained separated (i.e., were still variable in the same direction). Any variants that were overlapping, or were nearby (<750 kb) or separated by small gaps and appeared to result from the same region of variation, were merged. Additonally, one variant region on Chromosome 4 that occupied a portion (Chr 4: 30,583,544–31,202,423) of variant SCL#11 was removed, as this variant region was both much smaller and weaker (in terms of both *P*-value and extent of delay) than the encompassing SCL#11 variation in pES12 alone. This resulted in the final set of 21 haploid-delayed variants and 10 haploid-advanced variants.

In order to identify regions in which only one haploid cell line had different replication timing compared to diploid cells, we performed a genome-wide scan similar to the one above but utilizing a *t*-test, rather than ANOVA, and comparing one pair of samples at a time (this approach was also employed for experiment 2, as we were only comparing each haploid cell line to a single diploid cell line). In each 190-kb region, replication timing in a given haploid cell line was compared to the mean diploid replication timing in that region. Candidate replication timing variants were filtered as above, with the exception that the filter for consistent direction of effect was no longer applicable with only one haploid sample. In addition, we removed any part of these variants that overlapped the shared variants found using ANOVA (with the exception of SCL#11, in which the smaller nested shared variant was removed instead of the single cell variant [see previous paragraph]). This *t*-test approach was also used to identify replication timing variants in recently diploidized cells, where we compared the mean of recently diploidized cells to the single d-pES12 sample.

For identification of replication timing variants between the nine reactivated-X individuals and the other 107 individuals among the 116 ESCs, we used 20 instead of 76 windows for the ANOVA tests, because these data were in 10-kb instead of 2.5-kb windows.

In order to test whether haploid-delays, identified in the initial experiment, were also present in the second and third experiments, and to test for the presence of these variations in the recently diploidized ESCs, we performed ANOVA specifically at the haploid-delayed variant regions. Filtered, unsmoothed replication timing data from haploid and recently diploidized pESCs were each compared to the concurrently sequenced diploid pESCs in each experiment, and *P*-values $<2.4 \times 10^{-3}$ (Bonferonni correction of .05/21 for each variant tested) were considered significant (Supplemental Table S2).

Permutation methodology

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

For comparison to mouse UR regions and reactivated-X replication timing variants, we also performed reciprocal permutations, in which replication timing and size were retained relative to the those regions, rather than to the haploid replication timing variants. For quiescence-controlled permutations, we also required that each permutation was roughly the same percentage (±5%) quiescent (defined by the 18-state ChromHMM model) as the matched replication timing variant.

In order to determine overlap, we considered the span of each replication timing 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 between variants and a given data set were then compared to the distribution of 1000 permutations for both variant region permutations and reciprocal permutations, and a two-tailed *P*-value was calculated from the z-score of the real overlap. For genes and histone marks, we considered these as discrete calls rather than regions; therefore overlap distributions considered the number of occurrences within variants rather than the variant span. For ChromHMM states, overlap was not normalized by variant size and was instead calculated as the total number of base pairs in a given ChromHMM state; this was done to avoid the possibility that the contribution of important but physically small chromatin regions were deflated in large variants.

Gene Ontology

We used the enrichment analysis tool from The Gene Ontology Consortium (2019) to examine the genes in each variant class for biological process, molecular function, and cellular component enrichment. A Fisher's exact test was used to calculate a false discovery rate, which we report for all ontologies.

Mouse underrepresented regions

Mouse underrepresented regions in TGCs were obtained from Hannibal et al. (2014) and lifted-over from the mm9 mouse reference genome (as was used in the publication) to the human genome reference hg19 using the UCSC liftOver tool. There were 47 regions found in all mouse placenta cells. We took both the minimal region found across all six TGC samples and the maximal regions which were found in the TGC sample taken at the latest stage of development. In both cases, 45 of the 47 regions were successfully lifted-over. We do not expect that analyzing data based

on mm9 sequences rather than the more recent version of the mouse genome assembly would significantly affect the conclusions, as the regions analyzed have not meaningfully changed between the mm9 and mm10 references.

Data access

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.

Competing interest statement

N.B. is CSO of NewStem Ltd.

Acknowledgments

N.B. is the Herbert Cohn Chair in Cancer Research. This work was partially supported by the Israel Science Foundation (grant no. 494/17), by the Rosetrees Trust, and by the Azrieli Foundation (to N.B.), by New York State Stem Cell Science (NYSTEM IDEA) award #C32564GG (to D.E.), and by the Cornell Stem Cell Program and startup funds from Cornell University (to A.K.).

References

Attia M, Rachez C, De Pauw A, Avner P, Rogner UC. 2007. Nap1l2 promotes histone acetylation activity during neuronal differentiation. *Mol Cell Biol* **27:** 6093–6102. doi:10.1128/MCB.00789-07

Baker BA, Frickey L, Yu IT, Hawkins EP, Cushing B, Perlman EJ. 1998. DNA content of ovarian immature teratomas and malignant germ cell tumors. *Gynecol Oncol* **71:** 14–18. doi:10.1006/gyno.1998.5102

Becker KA, Ghule PN, Therrien JA, Lian JB, Stein JL, van Wijnen AJ, Stein GS. 2006. Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase. *J Cell Physiol* **209:** 883–893. doi:10.1002/jcp.20776

Bhowmick R, Minocherhomji S, Hickson ID. 2016. RAD52 facilitates mitotic DNA synthesis following replication stress. *Mol Cell* **64:** 1117–1126. doi:10.1016/j.molcel.2016.10.037

Bielski CM, Zehir A, Penson AV, Donoghue MTA, Chatila W, Armenia J, Chang MT, Schram AM, Jonsson P, Bandlamudi C, et al. 2018. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat Genet* **50:** 1189–1195. doi:10.1038/s41588-018-0165-1

Bruck T, Yanuka O, Benvenisty N. 2013. Human pluripotent stem cells with distinct X inactivation status show molecular and cellular differences controlled by the X-linked ELK-1 gene. *Cell Rep* **4:** 262–270. doi:10.1016/j.celrep.2013.06.026

Chang BH, Smith L, Huang J, Thayer M. 2007. Chromosomes with delayed replication timing lead to checkpoint activation, delayed recruitment of Aurora B and chromosome instability. *Oncogene* **26:** 1852–1861. doi:10.1038/sj.onc.1209995

Corbel C, Diabangouaya P, Gendrel AV, Chow JC, Heard E. 2013. Unusual chromatin status and organization of the inactive X chromosome in murine trophoblast giant cells. *Development* **140:** 861–872. doi:10.1242/dev.087429

Dekker J, Belmont AS, Guttman M, Leshyk VO, Lis JT, Lomvardas S, Mirny LA, O'Shea CC, Park PJ, Ren B, et al. 2017. The 4D nucleome project. *Nature* **549:** 219–226. doi:10.1038/nature23884

Ding Q, Edwards MM, Hulke ML, Bracci AN, Hu Y, Tong Y, Zhu X, Hsiao J, Charvet CJ, Ghosh S, et al. 2020. The genetic architecture of DNA replication timing in human pluripotent stem cells. bioRxiv doi:10.1101/2020.05.08.085324

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485:** 376–380. doi:10.1038/nature11082

Enoch T, Carr AM, Nurse P. 1992. Fission yeast genes involved in coupling mitosis to completion of DNA replication. *Genes Dev* **6:** 2035–2046. doi:10.1101/gad.6.11.2035

Eykelenboom JK, Harte EC, Canavan L, Pastor-Peidro A, Calvo-Asensio I, Llorens-Agost M, Lowndes NF. 2013. ATR activates the S-M checkpoint during unperturbed growth to ensure sufficient replication prior to mitotic onset. *Cell Rep* **5:** 1095–1107. doi:10.1016/j.celrep.2013.10.027

Fan JB, Surti U, Taillon-Miller P, Hsie L, Kennedy GC, Hoffner L, Ryder T, Mutch DG, Kwok PY. 2002. Paternal origins of complete hydatidiform moles proven by whole genome single-nucleotide polymorphism haplotyping. *Genomics* **79:** 58–62. doi:10.1006/geno.2001.6676

Francisco FO, Lemos B. 2014. How do Y-chromosomes modulate genome-wide epigenetic states: genome folding, chromatin sinks, and gene expression. *J Genomics* **2:** 94–103. doi:10.7150/jgen.8043

The Gene Ontology Consortium. 2019. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47:** D330–D338. doi:10.1093/nar/gky1055

Guo A, Huang S, Yu J, Wang H, Li H, Pei G, Shen L. 2017. Single-cell dynamic analysis of mitosis in haploid embryonic stem cells shows the prolonged metaphase and its association with self-diploidization. *Stem Cell Reports* **8:** 1124–1134. doi:10.1016/j.stemcr.2017.03.025

Hamline MY, Corcoran CM, Wamstad JA, Miletich I, Feng J, Lohr JL, Hemberger M, Sharpe PT, Gearhart MD, Bardwell VJ. 2020. OFCD syndrome and extraembryonic defects are revealed by conditional mutation of the Polycomb-group repressive complex 1.1 (PRC1.1) gene *BCOR*. *Dev Biol* **468:** 110–132. doi:10.1016/j.ydbio.2020.06.013

Handsaker RE, Van Doren V, Berman JR, Genovese G, Kashin S, Boettger LM, McCarroll SA. 2015. Large multiallelic copy number variations in humans. *Nat Genet* **47:** 296–303. doi:10.1038/ng.3200

Hannibal RL, Baker JC. 2016. Selective amplification of the genome surrounding key placental genes in trophoblast giant cells. *Curr Biol* **26:** 230–236. doi:10.1016/j.cub.2015.11.060

Hannibal RL, Chuong EB, Rivera-Mulia JC, Gilbert DM, Valouev A, Baker JC. 2014. Copy number variation is a fundamental aspect of the placental genome. *PLoS Genet* **10:** e1004290. doi:10.1371/journal.pgen.1004290

Harrigan JA, Belotserkovskaya R, Coates J, Dimitrova DS, Polo SE, Bradshaw CR, Fraser P, Jackson SP. 2011. Replication stress induces 53BP1-containing OPT domains in G1 cells. *J Cell Biol* **193:** 97–108. doi:10.1083/jcb.201011083

Heskett MB, Sanborn JZ, Boniface C, Goode B, Chapman J, Garg K, Rabban JT, Zaloudek C, Benz SC, Spellman PT, et al. 2020. Multiregion exome sequencing of ovarian immature teratomas reveals 2N near-diploid genomes, paucity of somatic mutations, and extensive allelic imbalances shared across mature, immature, and disseminated components. *Mod Pathol* **33:** 1193–1206. doi:10.1038/s41379-019-0446-y

Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, Giardine B, Ellenbogen PM, Bilmes JA, Birney E, et al. 2013. Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res* **41:** 827–841. doi:10.1093/nar/gks1284

Huang C, Cheng J, Bawa-Khalfe T, Yao X, Chin YE, Yeh ETH. 2016. SUMOylated ORC2 recruits a histone demethylase to regulate centromeric histone modification and genomic stability. *Cell Rep* **15:** 147–157. doi:10.1016/j.celrep.2016.02.091

Hulke ML, Massey DJ, Koren A. 2020. Genomic methods for measuring DNA replication dynamics. *Chromosome Res* **28:** 49–67. doi:10.1007/s10577-019-09624-y

Koren A, McCarroll SA. 2014. Random replication of the inactive X chromosome. *Genome Res* **24:** 64–69. doi:10.1101/gr.161828.113

Koren A, Handsaker RE, Kamitaki N, Karlić R, Ghosh S, Polak P, Eggan K, McCarroll SA. 2014. Genetic variation in human DNA replication timing. *Cell* **159:** 1015–1026. doi:10.1016/j.cell.2014.10.025

Koren A, Massey DJ, Bracci AN. 2021. TIGER: inferring DNA replication timing from whole-genome sequence data. *Bioinformatics* 10.1093/bioinformatics/btab166

Leeb M, Walker R, Mansfield B, Nichols J, Smith A, Wutz A. 2012. Germline potential of parthenogenetic haploid mouse embryonic stem cells. *Development* **139:** 3301–3305. doi:10.1242/dev.083675

Leeb M, Dietmann S, Paramor M, Niwa H, Smith A. 2014. Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell Stem Cell* **14:** 385–393. doi:10.1016/j.stem.2013.12.008

Leng L, Ouyang Q, Kong X, Gong F, Lu C, Zhao L, Shi Y, Cheng D, Hu L, Lu G, et al. 2017. Self-diploidization of human haploid parthenogenetic embryos through the Rho pathway regulates endomitosis and failed cytokinesis. *Sci Rep* **7:** 4242. doi:10.1038/s41598-017-04602-y

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26:** 589–595. doi:10.1093/bioinformatics/btp698

Li Y, Shuai L. 2017. A versatile genetic tool: haploid cells. *Stem Cell Res Ther* **8:** 197. doi:10.1186/s13287-017-0657-4

Li H, Guo A, Xie Z, Tu W, Yu J, Wang H, Zhao J, Zhong C, Kang J, Li J, et al. 2017. Stabilization of mouse haploid embryonic stem cells with combined kinase and signal modulation. *Sci Rep* **7:** 13222. doi:10.1038/s41598-017-13471-4

Linder D, McCaw BK, Hecht F. 1975. Parthenogenic origin of benign ovarian teratomas. *N Engl J Med* **292:** 63–66. doi:10.1056/NEJM197501092920202

Lukas C, Savic V, Bekker-Jensen S, Doil C, Neumann B, Pedersen RS, Grøfte M, Chan KL, Hickson ID, Bartek J, et al. 2011. 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nat Cell Biol* **13:** 243–253. doi:10.1038/ncb2201

Makunin IV, Volkova EI, Belyaeva ES, Nabirochkina EN, Pirrotta V, Zhimulev IF. 2002. The *Drosophila* suppressor of underreplication protein binds to late-replicating regions of polytene chromosomes. *Genetics* **160:** 1023–1034. doi:10.1093/genetics/160.3.1023

Massey D, Kim D, Brooks K, Smolka M, Koren A. 2019. Next-generation sequencing enables spatiotemporal resolution of human centromere replication timing. *Genes (Basel)* **10:** 269. doi:10.3390/genes10040269

Minocherhomji S, Ying S, Bjerregaard VA, Bursomanno S, Aleliunaite A, Wu W, Mankouri HW, Shen H, Liu Y, Hickson ID. 2015. Replication stress activates DNA repair synthesis in mitosis. *Nature* **528:** 286–290. doi:10.1038/nature16139

Munden A, Rong Z, Sun A, Gangula R, Mallal S, Nordman JT. 2018. Rif1 inhibits replication fork progression and controls DNA copy number in *Drosophila*. *eLife* **7:** e39140. doi:10.7554/eLife.39140

Nordman J, Li S, Eng T, Macalpine D, Orr-Weaver TL. 2011. Developmental control of the DNA replication and transcription programs. *Genome Res* **21:** 175–181. doi:10.1101/gr.114611.110

Nordman JT, Kozhevnikova EN, Verrijzer CP, Pindyurin AV, Andreyeva EN, Shloma VV, Zhimulev IF, Orr-Weaver TL. 2014. DNA copy-number control through inhibition of replication fork progression. *Cell Rep* **9:** 841–849. doi:10.1016/j.celrep.2014.10.005

Patel S, Bonora G, Sahakyan A, Kim R, Chronis C, Langerman J, Fitz-Gibbon S, Rubbi L, Skelton RJ, Ardehali R, et al. 2017. Human embryonic stem cells do not change their X inactivation status during differentiation. *Cell Rep* **18:** 54–67. doi:10.1016/j.celrep.2016.11.054

Rantala JK, Edgren H, Lehtinen L, Wolf M, Kleivi K, Vollan HK, Aaltola AR, Laasola P, Kilpinen S, Saviranta P, et al. 2010. Integrative functional genomics analysis of sustained polyploidy phenotypes in breast cancer cells identifies an oncogenic profile for GINS2. *Neoplasia* **12:** 877–888, IN4–IN14. doi:10.1593/neo.10548

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518:** 317–330. doi:10.1038/nature14248

Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, Schulz TC, Robins AJ, Dalton S, Gilbert DM. 2010. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res* **20:** 761–770. doi:10.1101/gr.099655.109

Sagi I, Benvenisty N. 2017. Haploidy in humans: an evolutionary and developmental perspective. *Dev Cell* **41:** 581–589. doi:10.1016/j.devcel.2017.04.019

Sagi I, Chia G, Golan-Lev T, Peretz M, Weissbein U, Sui L, Sauer MV, Yanuka O, Egli D, Benvenisty N. 2016. Derivation and differentiation of haploid human embryonic stem cells. *Nature* **532:** 107–111. doi:10.1038/nature17408

Sagi I, De Pinho JC, Zuccaro MV, Atzmon C, Golan-Lev T, Yanuka O, Prosser R, Sadowy A, Perez G, Cabral T, et al. 2019. Distinct imprinting signatures and biased differentiation of human androgenetic and parthenogenetic embryonic stem cells. *Cell Stem Cell* **25:** 419–432.e9. doi:10.1016/j.stem.2019.06.013

Schiltz RL, Mizzen CA, Vassilev A, Cook RG, Allis CD, Nakatani Y. 1999. Overlapping but distinct patterns of histone acetylation by the human coactivators p300 and PCAF within nucleosomal substrates. *J Biol Chem* **274:** 1189–1192. doi:10.1074/jbc.274.3.1189

Schoenfelder KP, Fox DT. 2015. The expanding implications of polyploidy. *J Cell Biol* **209:** 485–491. doi:10.1083/jcb.201502016

Stelzer Y, Yanuka O, Benvenisty N. 2011. Global analysis of parental imprinting in human parthenogenetic induced pluripotent stem cells. *Nat Struct Mol Biol* **18:** 735–741. doi:10.1038/nsmb.2050

Stevens L, Varnum D. 1974. The development of teratomas from parthenogenetically activated ovarian mouse eggs. *Dev Biol* **37:** 369–380. doi:10.1016/0012-1606(74)90155-9

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH, et al. 2015. An integrated map of structural variation in 2,504 human genomes. *Nature* **526:** 75–81. doi:10.1038/nature15394

Takahashi S, Lee J, Kohda T, Matsuzawa A, Kawasumi M, Kanai-Azuma M, Kaneko-Ishino T, Ishino F. 2014. Induction of the G2/M transition stabilizes haploid embryonic stem cells. *Development* **141:** 3842–3847. doi:10.1242/dev.110726

Tarkowski AK, Witkowska A, Nowicka J. 1970. Experimental parthenogenesis in the mouse. *Nature* **226:** 162–165. doi:10.1038/226162a0

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjöstedt E, Asplund A, et al. 2015. Proteomics. Tissue-based map of the human proteome. *Science* **347:** 1260419. doi:10.1126/science.1260419

van Kempen LC, Redpath M, Elchebly M, Klein KO, Papadakis AI, Wilmott JS, Scolyer RA, Edqvist PH, Ponten F, Schadendorf D, et al. 2016. The protein phosphatase 2A regulatory subunit PR70 is a gonosomal melanoma tumor suppressor gene. *Sci Transl Med* **8:** a177. doi:10.1126/scitranslmed.aai9188

Wear EE, Song J, Zynda GJ, Mickelson-Young L, LeBlanc C, Lee TJ, Deppong DO, Allen GC, Martienssen RA, Vaughn MW, et al. 2020. Comparing DNA replication programs reveals large timing shifts at centromeres of endocycling cells in maize roots. *PLoS Genet* **16:** e1008623. doi:10.1371/journal.pgen.1008623

Yaffe E, Farkash-Amar S, Polten A, Yakhini Z, Tanay A, Simon I. 2010. Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. *PLoS Genet* **6:** e1001011. doi:10.1371/journal.pgen.1001011

Yaguchi K, Yamamoto T, Matsui R, Tsukada Y, Shibanuma A, Kamimura K, Koda T, Uehara R. 2018. Uncoordinated centrosome cycle underlies the instability of non-diploid somatic cells in mammals. *J Cell Biol* **217:** 2463–2483. doi:10.1083/jcb.201701151

Yilmaz A, Peretz M, Sagi I, Benvenisty N. 2016. Haploid human embryonic stem cells: half the genome, double the value. *Cell Stem Cell* **19:** 569–572. doi:10.1016/j.stem.2016.10.009

Yilmaz A, Braverman-Gross C, Bialer-Tsypin A, Peretz M, Benvenisty N. 2020. Mapping gene circuits essential for germ layer differentiation via loss-of-function screens in haploid human embryonic stem cells. *Cell Stem Cell* **27:** 679–691.e6. doi:10.1016/j.stem.2020.06.023

Zhang XM, Wu K, Zheng Y, Zhao H, Gao J, Hou Z, Zhang M, Liao J, Zhang J, Gao Y, et al. 2020. In vitro expansion of human sperm through nuclear transfer. *Cell Res* **30:** 356–359. doi:10.1038/s41422-019-0265-1

Zhao Y, Garcia BA. 2015. Comprehensive catalog of currently documented histone modifications. *Cold Spring Harb Perspect Biol* **7:** a025064. doi:10.1101/cshperspect.a025064

Zybina EV, Zybina TG. 1996. Polytene chromosomes in mammalian cells. *Int Rev Cytol* **165:** 53–119. doi:10.1016/S0074-7696(08)62220-2

Received June 30, 2021; accepted in revised form October 20, 2021.