

RESEARCH

Open Access



Global early replication disrupts gene expression and chromatin conformation in a single cell cycle

Miguel M. Santos^{1,2†} , Mark C. Johnson^{1,2†} , Lukáš Fiedler^{1,2}  and Philip Zegerman^{1,2*} 

[†]Miguel M. Santos and Mark C. Johnson contributed equally to this work.

*Correspondence: paz20@cam.ac.uk

¹ Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK

² Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK

Abstract

Background: The early embryonic divisions of many organisms, including fish, flies, and frogs, are characterized by a very rapid S-phase caused by high rates of replication initiation. In somatic cells, S-phase is much longer due to both a reduction in the total number of initiation events and the imposition of a temporal order of origin activation. The physiological importance of changes in the rate and timing of replication initiation in S-phase remains unclear.

Results: Here we assess the importance of the temporal control of replication initiation using a conditional system in budding yeast to drive the early replication of the majority of origins in a single cell cycle. We show that global early replication disrupts the expression of over a quarter of all genes. By deleting individual origins, we show that delaying replication is sufficient to restore normal gene expression, directly implicating origin firing control in this regulation. Global early replication disrupts nucleosome positioning and transcription factor binding during S-phase, suggesting that the rate of S-phase is important to regulate the chromatin landscape.

Conclusions: Together, these data provide new insight into the role of the temporal control of origin firing during S-phase for coordinating replication, gene expression, and chromatin establishment as occurs in the early embryo.

Background

Eukaryotic genomes are replicated from multiple start sites or origins. To ensure genome stability, these origins must initiate replication only once during the cell cycle. Strict initiation control first involves the loading of the inactive replicative helicase (Mcm2-7) onto DNA only in late M/G1 phase, in a process called licensing. Initiation at these licensed origins can then only occur during S-phase, due to the activation of S-phase cyclin-dependent kinase (S-CDK) and Dbf4-dependent kinase (DDK). DDK phosphorylates the loaded Mcm2-7 helicase, which drives the recruitment of Sld7/Sld3 (MTBP/Treslin in humans) and the helicase activating protein Cdc45, while CDK



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

phosphorylates Sld3 and another protein Sld2 (RecQL4 in humans), driving interactions with Dpb11 (TopBP1 in humans), DNA polymerases, and other factors to form active replisomes and initiate replication [1].

In most eukaryotic cells, origins do not all fire at the beginning of S-phase. Although all origins are licensed in G1 phase, only a fraction of these (10% in mammalian cells) will be activated during a normal S-phase [2]. For those origins that fire, their activation occurs as a continuum throughout S-phase, resulting in differences in the timing of genome replication. This temporal order of genome duplication is known as the replication timing (RT) programme. Importantly, this RT programme is evolutionarily conserved between related species [3], is consistent within particular cell types [4], and changes during development [5], differentiation [6], and disease [7].

In budding yeast and metazoa, the signal for early or late replication is established in G1 phase during licensing [8, 9]. Several studies suggest that the establishment of RT is dependent on the regulation of the chromatin context of origins [10–13] and early origins differentially bind to the origin recognition complex (ORC) [14, 15] and load more Mcm2-7 [16]. We and others have shown that a critical determinant of origin timing is the ability of licensed origins to compete for a limiting pool of initiation factors [17, 18]. The activatory subunit of DDK, Dbf4, and the two critical CDK targets, Sld3 and Sld2, and their binding partner, Dpb11 are stoichiometrically limiting for initiation in budding yeast, and over-expression of these factors, together with the Sld3 partner proteins, Sld7 and Cdc45, is sufficient to allow the early firing of late and dormant origins in yeast [18]. Importantly, the vertebrate orthologues of these factors are also rate-limiting for replication initiation during early embryonic divisions [19]. Direct recruitment and inhibition of these limiting initiation factors can also influence replication timing, for example kinetochores and the forkhead box transcription factors Fkh1/Fkh2 can drive the early firing of subsets of origins by directly recruiting Dbf4 [20, 21], while conversely Rif1 counteracts DDK activity to delay RT in late-replicating domains [22].

Although differential replication timing of the genome was identified over 50 years ago [23] and RT is dramatically altered during differentiation and development [5, 6], the functional importance of the RT programme is still very poorly understood. Over-expression of limiting replication initiation factors, causing the early firing of many origins in budding yeast, results in cell lethality [18], strongly suggesting that ordered origin firing is important. In many organisms including humans, mouse, and *Drosophila* there is a positive correlation between early replication and the probability of gene expression [24], while in budding yeast only the highest expressed genes tend to be earlier replicated and the lowest expressed genes tend to be late replicated [25]. Despite this, whether early replication is a cause or a consequence of gene expression is not clear. Recently, it was shown that perturbations in RT caused by loss of Rif1 in human cells is coupled with alterations in histone modifications and 3D chromatin compartments, but only causes a limited effect on gene expression [26]. In yeast however, early replication of the histone genes is required for their maximal expression in S-phase [27].

Here we take advantage of our system that conditionally over-expresses limiting replication initiation factors in budding yeast [18] to determine the relationship between perturbations in DNA replication timing and gene expression in a single cell cycle. By driving the early replication of the majority of the genome, we observe dramatic changes

in the expression of over a quarter of yeast genes during S-phase. Early genomic replication causes concomitant changes in chromatin dynamics and transcription factor binding events during S-phase. This study provides new insight into the importance of ordered origin firing, with implications for how replication rate can influence gene expression and the chromatin landscape.

Results

Over-expression of limiting replication initiation factors causes global early replication

To perturb replication timing genome-wide in a single cell cycle, we took advantage of our conditional system in budding yeast that over-expresses six limiting replication factors Sld2, Sld3, Dpb11, Dbf4, Cdc45, and Sld7 (abbreviated to SSDDCS) upon the addition of galactose [18]. Galactose induction of the SSDDCS strain speeds up S-phase, as measured by flow cytometry, but also leads to the depletion of the dNTP pool, which causes activation of the DNA damage checkpoint [18]. To advance replication timing without activating the checkpoint, we performed all experiments with strains containing a null mutation of the ribonucleotide reductase inhibitor *SML1*. The *sml1Δ* mutation increases the dNTP pool and overcomes the galactose-induced checkpoint activation in the SSDDCS strain [18].

To determine the genome-wide impact of over-expression of the limiting replication factors on DNA replication timing in a single cell cycle, we analyzed replication progression by DNA sequencing and copy number analysis. For this, the *sml1Δ* and the *sml1Δ* SSDDCS strains were synchronized in G1 using the mating pheromone alpha factor; galactose was added to these G1 arrested cells for 30 min, and then cells were released into a synchronous S-phase in the presence of galactose, whereby samples were collected every 5 min to generate libraries for whole genome sequencing. To determine the replication dynamics of these strains, the ratio of mapped reads at each timepoint compared to the G1 sample (copy number) was calculated for each genomic 1-kb bin, as described [28]. The median replication time (T_{rep}) for each 1-kb bin was determined by fitting a sigmoidal curve to the copy number profile from each bin and extracting the time (in minutes after G1 release) at which each bin is half-way between one and two copies. Replication profiles were generated by plotting the T_{rep} values along the corresponding chromosome position, followed by smoothing using a moving average (Fig. 1A—only Chromosome VIII is shown). From such a profile, the peaks (local minima of T_{rep}) represent origins of replication and troughs (local maxima of T_{rep}) are termination zones (note that for these analyses the y-axis is inverted so that origins are peaks—Fig. 1A). As expected [18, 29], SSDDCS over-expression advances the replication timing of all origins across the whole chromosome, including stimulating the firing of normally passively replicated origins, such as ARS810 and ARS821 (Fig. 1A).

To address whether all origins genome-wide were equally affected by over-expression of limiting factors, replication origins were divided into quintiles according to their T_{rep} in the control strain (Fig. 1B). This analysis demonstrated that all groups of origins replicate earlier on average upon SSDDCS over-expression, with later origins having a greater advance in replication timing than early origins (Fig. 1B). This is consistent with our previous data that there is less capacity for the earliest origins to fire even earlier when limiting factors are over-expressed [18], as the transition from G1 to S-phase and

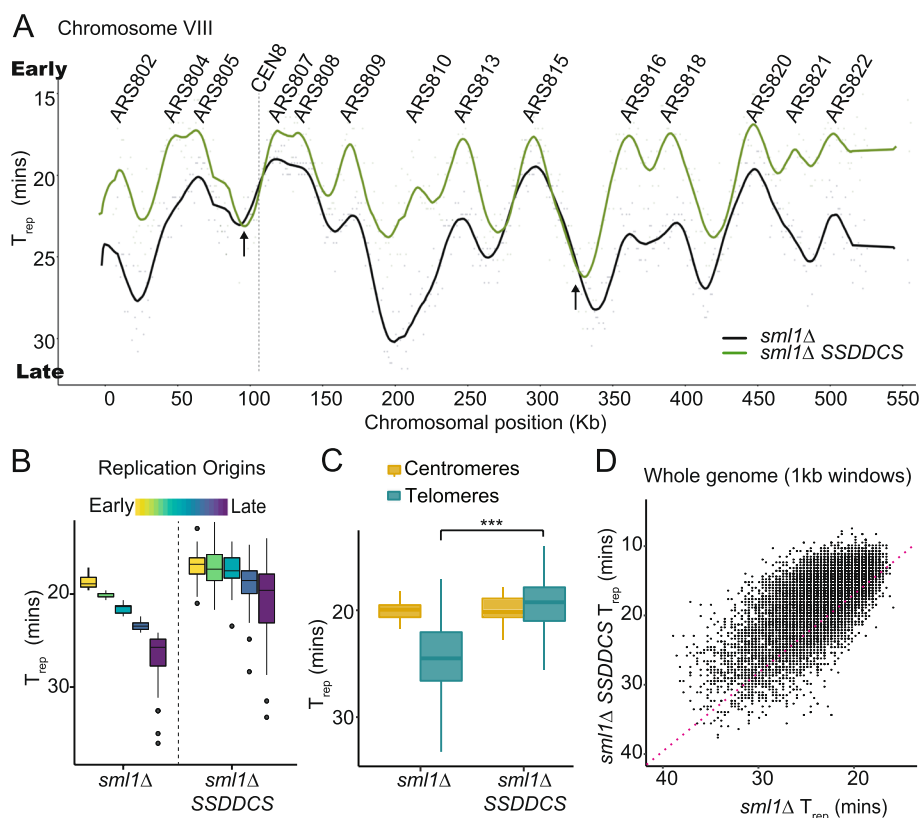


Fig. 1 Over-expression of limiting initiation factors causes global early replication. **A** T_{rep} values were plotted from the indicated strains along the corresponding chromosome positions to generate genome-wide replication profiles and smoothed using a moving average. Chromosome VIII is shown here as an example. The y-axis is flipped so that early replicating regions are at the top of the plot and late-replicating regions at the bottom. The location of annotated origins (ARS) is shown above. Overall, all origins fire earlier in the *SSDDCS* strain. Arrows indicate termination zones that have a delayed T_{rep} in the *SSDDCS* strain. **B** Distribution of T_{rep} values for all origins divided into quintiles according to T_{rep} values from the *sm11Δ* strain. **C** Box and whisker plot of the T_{rep} values for all centromeres and telomeres (within 50kb of chromosome ends).*** p -value = $< 2.2e-16$, Welch two-sample t -test. Centromeres were not significantly affected by *SSDDCS* over-expression (p -value = 0.9578, Welch two-sample t -test). **D** Scatterplot of the T_{rep} values for the whole genome in 1-kb bins. Red dashed line is the line of equal T_{rep} between the *sm11Δ* and *sm11Δ SSDDCS* strains

activation of DDK and CDK is still required before an origin can fire. In the *SSDDCS* strain, the first three quintiles of origins fire at the same time, showing that at least 60% of all origins fire at the very earliest time in S-phase (Fig. 1B). The two latest quintiles of origins, although not advanced as much as the earliest origins, still fire as early as the majority of early origins in the wild type strain (see for example ARS821 in Fig. 1A). Therefore, *SSDDCS* over-expression drives the early firing of almost all origins in yeast.

In budding yeast, the telomeres are among the latest replicating regions of the genome, while the centromeres are among the earliest replicating due to the direct recruitment of limiting replication factors to centromeres [21]. Consistently with Fig. 1B, *SSDDCS* over-expression does not induce centromeres to replicate even earlier, but does drive the early replication of telomeres, such that these regions now replicate at the same time as centromeres (Fig. 1C). Analysis of the whole genome as 1-kb bins reveals that the vast majority of the genome is replicated earlier upon *SSDDCS* over-expression (Fig. 1D,

above magenta line). Despite this, some sites are replicated later after *SSDDCS* over-expression (Fig. 1D, below magenta line). We have previously shown that high rates of initiation cause topological defects, due to the overwhelming of topoisomerase activities, resulting in delays in termination [30]. Such delayed termination sites are indeed visible on the T_{rep} plots (black arrows, Fig. 1A). Apart from these small number of termination zones, the *SSDDCS* over-expression strain clearly induces the early replication of the majority of the genome in a single cell cycle.

Global early replication perturbs the transcription of over a quarter of the genome in a single cell cycle

To analyze the impact of advanced replication on gene expression during a single cell cycle, samples were collected for whole transcriptome sequencing (RNA-seq) from synchronized cells as in Fig. 2A. Notably, the advanced S-phase in the *SSDDCS* strain was not caused by differences in the G1-S transition, which are highly similar between strains as determined by budding index and the expression of G1 and S-phase cyclins (Additional file 1: Fig. S1), but instead are likely due to the advanced global replication in this strain (Fig. 1). A principal component analysis (PCA) of the gene expression changes during advanced replication timing revealed the absence of batch effects between biological replicates (Additional file 1: Fig. S2). To address a role for replication timing in gene expression control, each timepoint was compared between the two strains using DESeq2 [31]. Note that due to the over-expression of the six factors in the *SSDDCS* strain, these six genes were excluded from all downstream analyses. Genes were considered up- or downregulated in the *SSDDCS* strain if the \log_2 normalized fold-change (*SSDDCS sml1Δ / sml1Δ*) was above or below 0 respectively and if the adjusted p -value or false discovery rate (FDR) was below 0.01 (DESeq2 Wald test). Using this approach, a large number of genes were differentially expressed (DE) in the *SSDDCS* strain in a single cell cycle (Fig. 2B). Importantly, the majority of these expression changes occur after 20 min, suggesting that S-phase is required for these differences (Fig. 2B). Notably, the number of DE genes was greatly reduced by the end of the time course, suggesting that any changes in expression are resolved before the next cell cycle (Fig. 2B).

Taking advantage of the temporal resolution of this dataset, we analyzed the expression pattern of each gene during the cell cycle, using the DESeq2 likelihood ratio test (LRT) [31]. A significant FDR from this test identifies genes that show a difference between the two strains at one or more timepoints after G1 phase. Using this approach, a total of 1771 genes were identified as differentially expressed, representing ~27% of the genome. From this list, k-means clustering was used to group genes with similar expression profiles (Fig. 2C,D). This clustering predominantly segregated genes by their normal periodic expression [32] (Fig. 2E), for example Cluster 2 genes are significantly over-represented in S/G2 expressed genes. As the k-means clusters indicated that the genes that are differentially expressed in the *SSDDCS* strain are dynamically regulated even in a wild type strain (Fig. 2D), we wondered whether constitutively expressed genes, which do not change their expression under different conditions [33], might be less affected than inducible genes by advanced replication timing. Figure 2F shows that constitutively expressed genes show much less differential expression than inducible genes after early

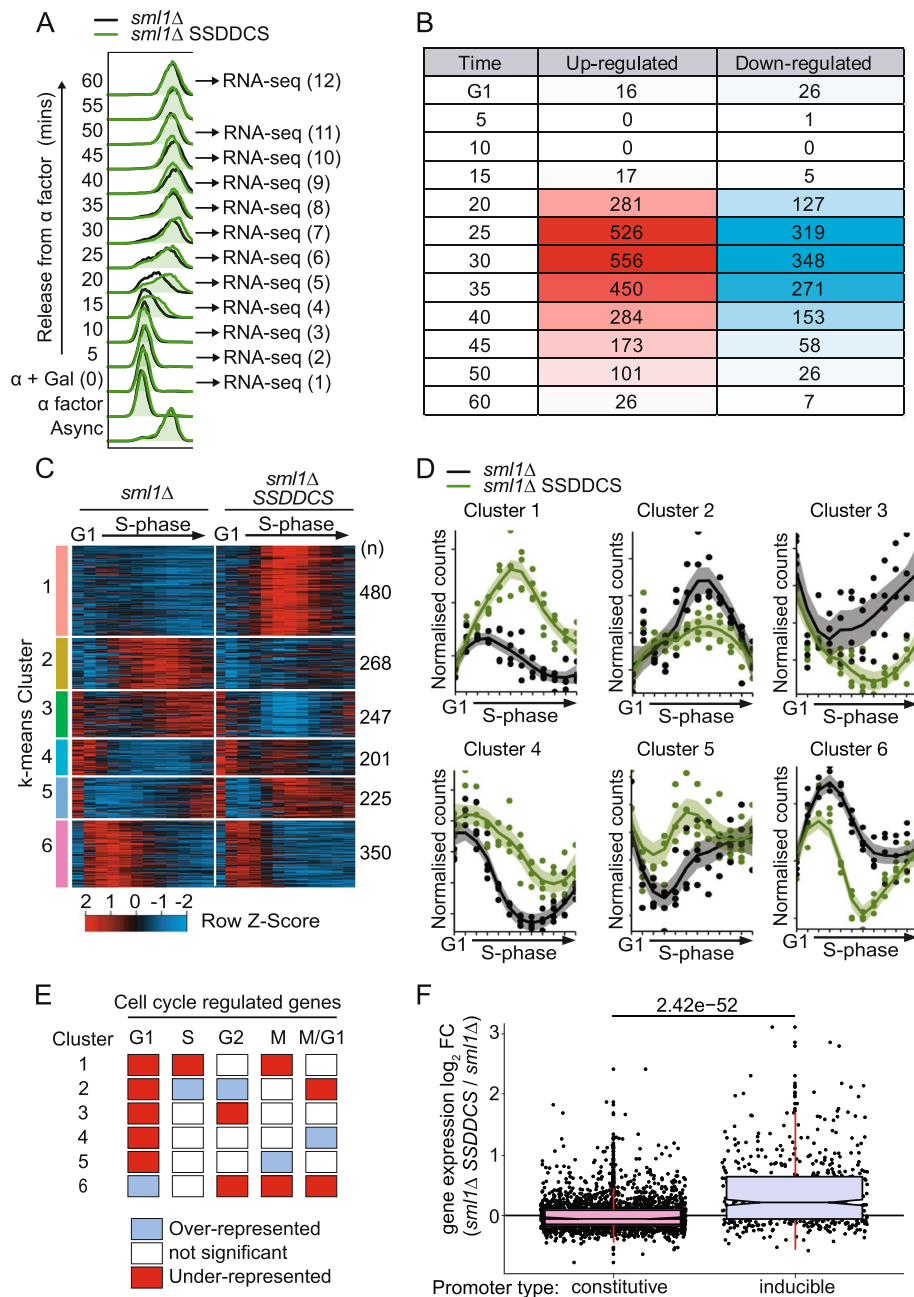


Fig. 2 Global early replication perturbs the expression of 27% of genes in one cell cycle. **A** Overlay of the flow cytometry between the *sml1Δ* and *sml1Δ SSDDCS* strains arrested in a factor and released into S-phase in galactose. The timepoints (1-12) used to generate time-resolved RNA-seq libraries are indicated. **B** Table showing the number of differentially expressed genes at each of the 12 RNA-seq timepoints between the *sml1Δ* and *sml1Δ SSDDCS* strains. Cells are color-coded based on the number of DE genes on each timepoint. **C** Heatmap of 1771 genes that are differentially expressed (DE) between the two strains in one or more timepoints after G1. Each row corresponds to one individual gene and each column to one timepoint (G1 to 60 min from left to right, respectively). Genes were clustered according to their expression profiles using k-means clustering. The expression levels were z-scored and normalized by row. The maximal and minimum expression are colored in red and blue, respectively. **D** Normalized read counts per timepoint using DESeq2 size factors for one illustrative gene from each DE cluster from **C**. **E** Table showing k-means clusters from **C** with the statistically significant (p -value < 0.05) over- (blue) or under- (red) representation of genes designated to be cell cycle regulated according to [32]. p -values were calculated using binomial probability tests. **F** Box and whisker plot of the \log_2 time course-averaged expression fold-change for *sml1Δ SSDDCS* / *sml1Δ* strains of genes separated into constitutively expressed or inducible according to [33]

replication, consistent with the observation that these constitutive genes are refractory to environmental stimuli [33].

We wondered to what extent replication timing changes might be important for the gene expression changes that we observed (Fig. 2B–D). As *SSDDCS* over-expression causes the earlier replication of the majority of the genome (Fig. 1D), it was not surprising that all DE gene clusters, as well as non-DE genes are on average earlier replicated in the *sml1Δ SSDDCS* strain compared the *sml1Δ* strain (Additional file 1: Fig. S3A). Analysis of the proximity of DE genes to origins revealed that clusters 1, 4, and 5 are on average more origin proximal than non-DE genes, and vice versa for the remaining clusters (Additional file 1: Fig. S3B). Clusters 1 and 4 show enrichment for genes that are proximal to the normally late-replicating telomeres, although the majority of DE genes are not telomeric (Additional file 1: Fig. S3C). Interestingly, given that centromeres do not show replication timing changes in the *SSDDCS* strain (Fig. 1C), centromeres are not enriched for DE genes from any of the k-means clusters (Additional file 1: Fig. S3D). This data suggests that earlier replication is likely to be necessary for the gene expression changes that we observe in the *SSDDCS* strain, but advanced timing is not sufficient for these changes as non-DE genes are also earlier replicating in the *SSDDCS* strain (Additional file 1: Fig. S3A).

The expression of some genes is directly affected by an increase in copy number during replication, such as the histone genes [27]. It has also been suggested that to maintain expression homeostasis during S-phase, the earliest S-phase genes are buffered against copy number changes through the histone acetyltransferase Rtt109 [34]. As early replication of the genome in the *SSDDCS* strain might affect gene expression by increasing the copy number of a gene in early S-phase, we wondered to what extent this might influence the differential expression we observe during advanced replication (Fig. 2C). Importantly, we observed no correlation between the genes that are DE in the *SSDDCS* strain and the *rtt109Δ* sensitive or insensitive gene sets [34] (data not shown). In addition, the k-means clusters that are enriched for genes that are normally expressed in S-phase and G1/S-phase (clusters 2 and 6 respectively, Fig. 2E) show reduced expression after early replication (Fig. 2C,D), which is the opposite of what we would expect if copy number was affecting their expression. Finally, the differences in expression between *sml1Δ* and *sml1Δ SSDDCS* were significantly greater than 2-fold for many genes (in particular for Cluster 1, e.g., see Fig. 3A), again suggesting that the expression changes of these genes are not simply due to copy number changes. Together these data show that genome-wide early replication perturbs the normal expression of a large proportion of the genome in a single cell cycle.

Single-cell analysis of early replication-induced transcription

From the RNA-seq analysis, we conclude that altering replication timing globally affects the expression of many genes across a population of synchronized cells. We wondered whether the effect of replication timing on gene transcription was of sufficient magnitude to observe these expression changes in individual cells in a single cycle. For this, we replaced the *THI12* gene with *GFP* as a fluorescent proxy for *THI12* expression. *THI12* is a cluster 1 gene (Fig. 2C,D) that is highly expressed in the *SSDDCS* strain (Fig. 3A). We replaced *THI12* with a version of *GFP* that has a PEST degron from *Cln2* to ensure

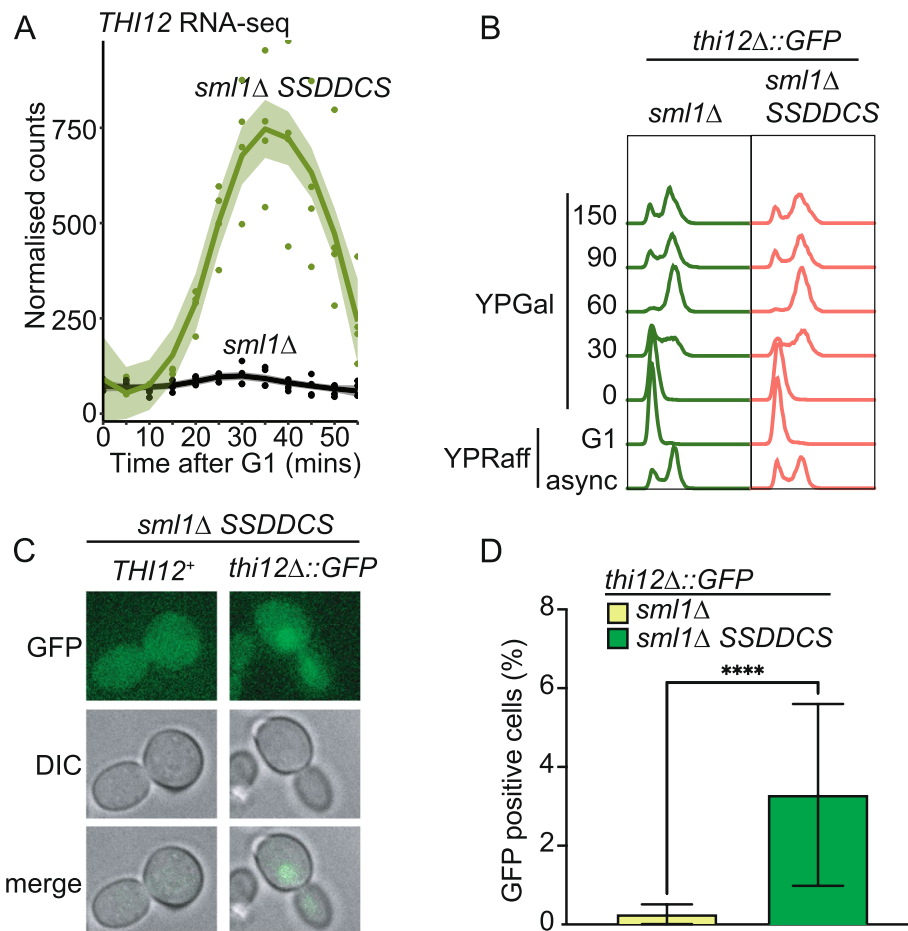


Fig. 3 Single-cell analysis of early replication-induced gene expression. **A** RNA-seq data for *THI12*, a cluster 1 DE gene, from the data in Fig. 2C. **B** Flow cytometry of the time course for analysis of GFP expression from the *THI12* locus. **C** Representative images from the analysis of GFP expression at the 150-min timepoint. **D** Quantitation of GFP expression from the 150-min timepoint as in **C** for 3 independent experiments. $n=755$ and 1156 for *sm1Δ* and *sm1ΔSSDDCS* respectively. Error bars are SD, p -value <0.001 from an unpaired T -test

that the protein is short-lived and an NLS to concentrate the protein in the nucleus and enhance the microscopy analysis [35]. We performed the same experiment as with the RNA-seq (Fig. 3B), except that we took later timepoints, due to the delay between GFP expression, translation and folding [35]. Using this system, we observed that GFP was not expressed in the *sm1Δ* strain in galactose as expected, as *THI12* is not expressed in these conditions (Fig. 3A, D). However, in the *SSDDCS* strain, GFP was significantly expressed in approximately 3% of cells in a single cell cycle (Fig. 3C,D). Although only a small percentage of cells appeared positive, we do not know what threshold of GFP expression is necessary for a positive signal in this assay. Despite this, it is clear that early replication indeed affects gene transcription in individual cells in a single cell cycle.

Delaying replication rescues aberrant transcriptional activation

If early replication is required for the aberrant gene expression that we observe (Fig. 2), then we predicted that if we could delay the replication of individual genes in the *SSDDCS* strain, we might be able to restore their normal expression. For this, we

identified two cluster 1 genes, *IME2* and *NDT80*, that are highly expressed in S-phase in the *SSDDCS* strain, similar to *THI12* (Fig. 3A) and are also proximal to active origins (Fig. 4A,B). Taking advantage of the fact that replication origins are defined genetic elements in budding yeast (autonomously replicating sequences, ARS), we deleted the two neighboring ARS elements from the *IME2* locus (ARS1008/1009, Fig. 4A) and from the *NDT80* locus (ARS816/818, Fig. 4B). Importantly, these ARS deletions specifically delayed the replication of the target locus (Fig. 4A,B), but the replication timing of all other origins remained similar between the strains (Additional file 1: Fig. S4). Analysis of the expression of the *IME2* and the *NDT80* genes by qPCR showed that these genes are not expressed in the *sm11Δ* strain but are expressed in S-phase in the *sm11Δ SSDDCS* strain, as expected (Fig. 4C,D). Significantly, deletion of the neighboring origins, which delays replication specifically of the *IME2* or the *NDT80* loci, completely abrogated the aberrant expression of these two genes in the *SSDDCS* strain (Fig. 4C,D). This experiment strongly suggests that it is the early replication of these loci that is critical for their aberrant S-phase expression.

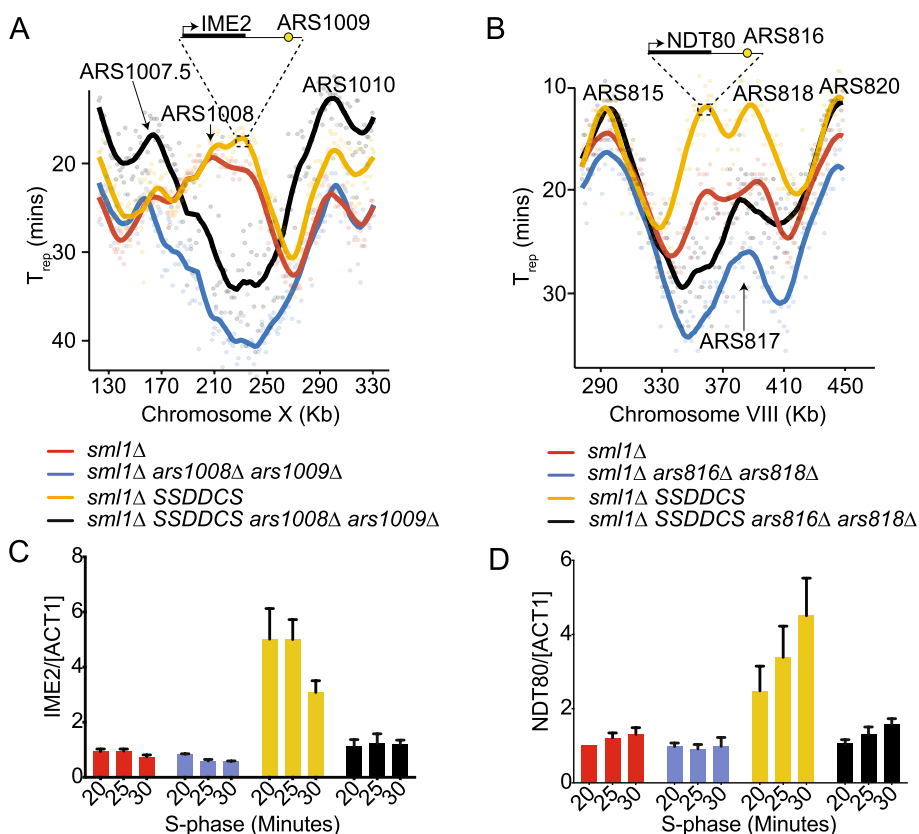


Fig. 4 Delaying replication rescues aberrant transcriptional activation. **A** T_{rep} profile from the indicated strains as in Fig. 1A, from a segment of Chr X. The position of the *IME2* locus is indicated on the *sm11Δ SSDDCS* profile. **B** As **A**, except for the *NDT80* locus on Chr VIII. **C** qPCR data of *IME2* expression from the indicated timepoints for the strains as in **A**. Error bars are SD, $n=3$. Gene expression was normalized to actin. **D** As **C**, except for *NDT80* expression

Global early replication perturbs chromatin conformation in a single cell cycle

To address how galactose induction of the *SSDDCS* strain might lead to dramatic changes in transcription in a single cell cycle, we set out to analyze whether early replication of the genome affects chromatin organization. For this, we analyzed nucleosome positioning and occupancy across the genome by sequencing after micrococcal nuclease digestion (MNase-seq). Most genes contain a well-positioned nucleosome (+1 nucleosome), just downstream of the transcriptional start site (TSS), which is dynamically repositioned during transcription [36, 37]. From our MNase-seq data, we observed that the +1 nucleosome remains largely static relative to the TSS in the *sml1Δ* strain, regardless of whether the genes are expressed or not during the time course (Fig. 5A). However, in the *SSDDCS* strain that induces global early replication, all DE and even non-DE genes displayed greater mobility of the +1 nucleosome towards the gene body (Fig. 5A). Importantly this mobility of the +1 nucleosome was transient and was restored to the normal position by the end of the time course (Fig. 5A).

To analyze nucleosome organization in the gene body, the pattern of the first 4 nucleosomes was assessed using the autocorrelation function (ACF), as recently described [39]. Genes with higher ACF values have well-phased and organized nucleosomes, while lower values represent poorly organized chromatin (Fig. 5B, top). ACF values were calculated for all genes in each timepoint and plotted as a heatmap (Fig. 5B, bottom). This analysis was restricted to genes larger than 700bp (as 4 nucleosomes are needed for ACF calculation), which corresponds to approximately 70% of all genes. In the *sml1Δ* control strain, there is a transient decrease in nucleosome organization during S-phase, most likely caused by the passage of replication forks, but once replication is completed the nucleosome organization is re-established (Fig. 5B). Strikingly, greater nucleosome disorganization was observed genome-wide in the *SSDDCS* strain (Fig. 5B), which is in agreement with the +1 nucleosome data (Fig. 5A). This decreased nucleosome organization in the *SSDDCS* strain was not

(See figure on next page.)

Fig. 5 Global early replication perturbs chromatin conformation in a single cell cycle. **A** Position of the +1 nucleosome relative to the transcriptional start site (TSS) represented as a heatmap for the DE genes separated into their k-means cluster (Fig. 2C) plus 250 random non-DE genes with an annotated +1 nucleosome. Blue represents the most upstream position of the +1 nucleosome, while green represents the most downstream position (into the gene body). Data is normalized by row as done for previous heatmaps. Each row corresponds to a single gene and each column to a single timepoint (G1 to 60 from left to right, respectively). **B** Top: The nucleosome occupancy and organization of the first 4 nucleosomes after the TSS was analyzed using an autocorrelation function (ACF) analysis: nucleosomes in green are well phased (i.e., well defined peak and valleys) and well positioned (similar inter-nucleosome distance), corresponding to high ACF values. On the other hand, nucleosomes in black are poorly phased and disorganized, resulting in low ACF. Bottom: Heatmap of ACF values for the DE genes and 350 random non-DE genes longer than 700bp. Each row corresponds to a single gene and each column to a single timepoint (G1 to 60 from left to right, respectively). The two strains are separated by a white vertical line. White and green represents the timepoints at which the chromatin is mostly disorganized and organized, respectively. **C** Plot of statistical significance versus effect size for Kruskal-Wallis tests applied to chromatin mutant gene expression data from [38]. Red dashed horizontal and vertical lines represent a Bonferroni-corrected p -value threshold of 0.001 and a high-effect threshold ($\eta^2 = 0.14$), respectively. **D** Box and whisker plot of the averaged gene expression change in the *SSDDCS* strain for the k-means clusters from the RNA-seq data in Fig. 2C. p -values are from Holm-corrected Wilcoxon rank sum tests relative to 300 random non-DE genes (black). **** $p < 0.0001$. **E** Box and whisker plot of the RNA expression changes in the *cac1Δ* strain (data from [38]). p -values are from Holm-corrected Wilcoxon rank sum tests relative to 300 random non-DE genes (black). **** $p < 0.0001$

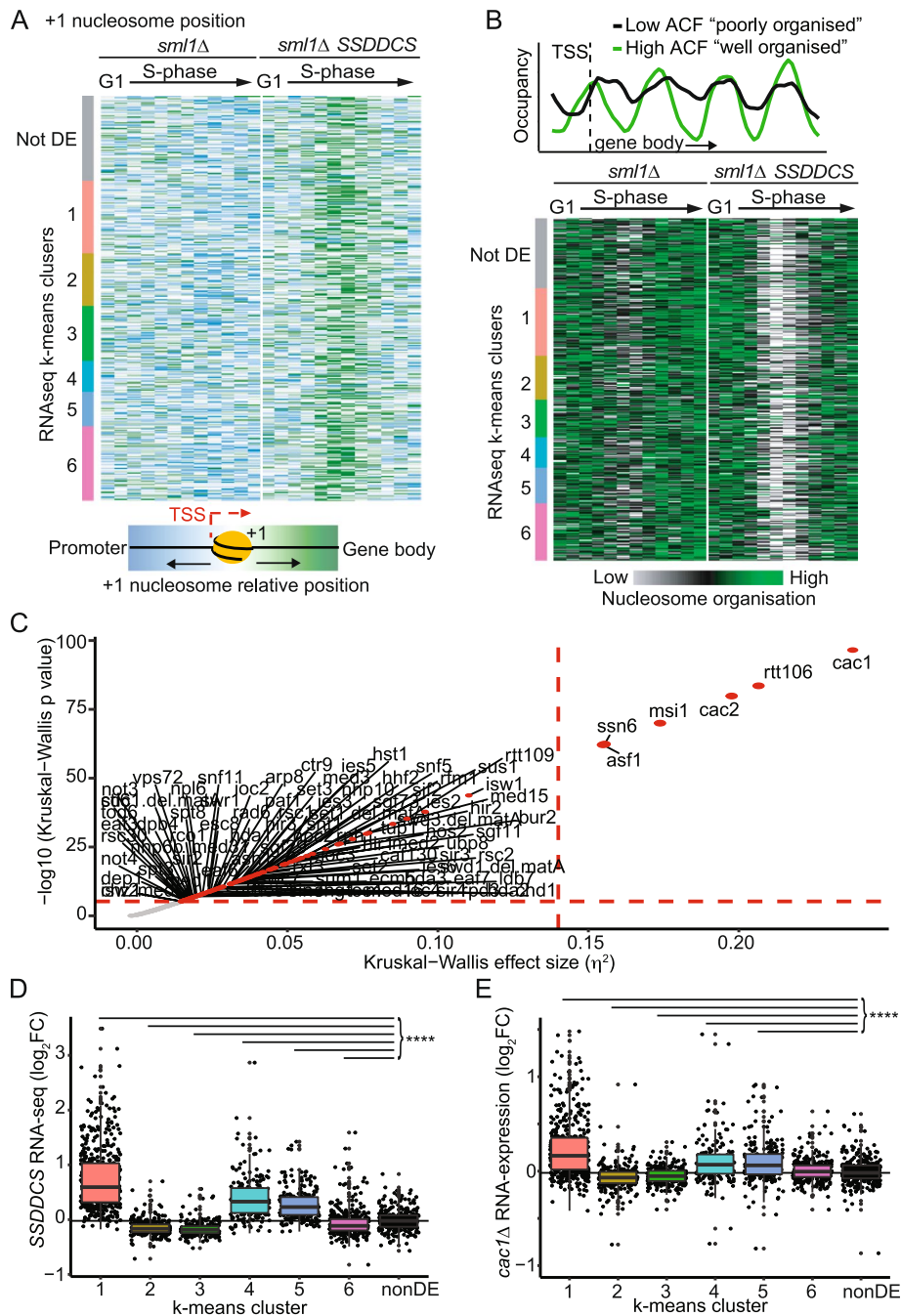


Fig. 5 (See legend on previous page.)

specific to the DE genes and was restored back to normal levels of organization by the end of the time course (Fig. 5B). Together, this data shows that genome-wide early replication perturbs the nucleosome positioning and organization of many genes.

Figure 5 A and B demonstrate that driving early replication genome-wide perturbs chromatin organization during S-phase. One mechanism that might cause this defect could be that the high rate of replication early in S-phase overwhelms the histone supply and assembly machinery, leading to a transient period when new nucleosome

deposition/positioning cannot keep pace with replication. If such transient chromatin assembly defects are responsible for the global gene expression changes that we observe in the *SSDDCS* strain, then it would follow that strains that are defective in chromatin assembly should mimic the gene expression changes in the *SSDDCS* strain. To investigate the relationship between chromatin defects and expression of the *SSDDCS*-dependent genes, we compared our data with the gene expression changes in a compendium of 165 chromatin machinery deletion mutants, which includes nucleosome remodelers, histone chaperones, histone modifiers, and transcription co-regulators [38]. As the gene expression data for these chromatin machinery deletion datasets is from only a single asynchronous timepoint, we converted our gene expression data (Fig. 2C) to an average gene expression value for the entire time course (Fig. 5D). Importantly, even after averaging the gene expression time course data, each k-means cluster remained highly significantly different to the non-differentially expressed control group (Fig. 5D). This data was then used to quantify the similarity between the *SSDDCS* data and the 165 chromatin machinery mutants using Kruskal-Wallis tests. A plot of statistical significance against effect size clearly identified six mutants with highly significantly similar expression profiles to our *SSDDCS* data (Fig. 5C): the CAF-1 histone chaperone complex (*Cac1*, *Cac2*, *Msi1*), the *Rtt106* and *Asf1* histone chaperones, and the *Ssn6* transcriptional co-repressor. Direct comparison of the *cac1Δ* dataset with our *SSDDCS* data mimicked the significant upregulation of the clusters 1/4/5 and the downregulation of clusters 2/3 (Fig. 5D versus E). The magnitude of the fold-change in the *cac1Δ* data was smaller than for our *SSDDCS* data (Fig. 5D versus E), but this might reflect methodological differences from comparing our RNA-sequencing data with the *cac1Δ* microarray data from asynchronous mid-log yeast populations [38]. Importantly, the similarity between the gene expression changes in the *cac1Δ* strain and the *SSDDCS* over-expression experiment is not because the CAF-1 complex is downregulated by global early replication (Additional file 1: Fig. S5). Together, the dramatic defect in chromatin organization in mid-S-phase (Fig. 5A,B) combined with the correlation of our gene expression data with chromatin assembly mutants (Fig. 5C–E) strongly suggests that earlier replication of the majority of the genome causes a transient chromatin assembly defect, which may explain to some extent the changes in gene expression that we observe.

High rates of initiation disrupt the TF landscape

Although there is a correlation between chromatin disruption and changes in gene expression (Fig. 5), it is not clear why changes in chromatin should affect the expression of certain genes. Newly synthesized DNA must reassemble nucleosomes and transcription factors (TFs), which compete for binding to nascent DNA [40]. We therefore wondered whether chromatin disruption caused by global early replication, might also perturb the TF landscape. Analysis of nucleosome positioning at the *NDT80* locus (Fig. 6A), which is a cluster 1 gene that is highly expressed in S-phase in the *SSDDCS* strain (Fig. 4), shows not only that the +1 nucleosome moves into the gene body during S-phase, as expected (Fig. 5A), but that the nucleosome positions around the binding sites for the key *NDT80* regulatory TF Ume6 also change during S-phase, specifically in the *SSDDCS* strain (Fig. 6A). To address whether nucleosome positioning was altered at other Ume6 binding sites (URS1 sites), these sites were

extracted from the JASPAR database [41] and the locations of all URS1 sequences in the budding yeast genome were identified using Find Individual Motif Occurrences (FIMO) [42]. Of the 2875 URS1 sequences identified, 89 were located 1kb upstream of the TSS of genes from cluster 1. Analysis of the average nucleosome profile centered around these 89 locations showed that nucleosome positioning and occupancy was identical between the strains in G1 phase (Fig. 6B, top). In S-phase however, the *SSDDCS* strain exhibited reduced occupancy of nucleosomes at the Ume6 binding sites and movement of the nucleosomes away from these sites (Fig. 6B, bottom). This suggests that when chromatin is perturbed by global early replication the interplay between TF binding and nucleosome positioning also changes.

To analyze TF binding dynamics in an unbiased way, we used an optimized version of MNase-seq that allows the analysis of DNA binding proteins such as TFs that are below the size of single nucleosomes [39]. Using this approach and analyzing only fragments smaller than 100bp, we identified 7493 sub-nucleosomal peaks. To pinpoint differential TF binding events that could explain the observed differences in gene expression, only peaks that were located 1kb upstream of a TSS and that had a 2-fold difference in the *SSDDCS* strain in at least 1 timepoint outside G1 phase were selected (2950 peaks). These 2950 sites represent 61% of all the peaks that we identified in gene promoters suggesting that early replication affects a majority of TF binding events during S-phase.

In order to verify that the sub-nucleosomal peaks we identified correspond to genuine TF binding events, peaks were mapped to annotated TF binding sites [43]. Only peaks within 200bp of an annotated site were considered for downstream analyses (1538 peaks). As the *SSDDCS* strain contains extra copies of the *GALI-10* promoter (which are used to over-express the six factors), we analyzed TF binding at this promoter to confirm that the sub-nucleosomal MNase-seq isolated genuine TF binding locations. As expected, the read coverage at the *GALI-10* promoter was significantly higher in the *SSDDCS* strain compared to the *sml1Δ* strain (Fig. 6C), suggesting that

(See figure on next page.)

Fig. 6 High rates of initiation disrupt the transcription factor landscape. **A** Nucleosome positioning heatmaps of the *NDT80* locus, which has two Ume6 binding sites in the promoter region (vertical solid red lines). The vertical dashed red line marks the TSS and each row represents one timepoint, from G1 to 60 from top to bottom. Yellow corresponds to a high density of MNase-seq reads corresponding to nucleosome peaks, while blue corresponds to nucleosome depleted regions. **B** Average nucleosome profiles for the 89 URS1 locations identified in promoters of cluster 1 genes in G1 and at 25 min. The profiles are centered on the URS1 sequence (the Ume6 binding site - dotted line). **C** Sub-nucleosomal MNase-seq read coverage around the *GALI-10* promoter in G1. **D** Sub-nucleosomal MNase-seq reads of the *NDT80* locus 5 min after release from G1 phase. Y-axis shows the sub-nucleosomal read coverage in this genomic region. Top track shows the location of genes. Peak annotated to the Ume6 binding site is highlighted. **E** Each sub-nucleosomal MNase-seq peak was assigned to a known TF binding site from [43]. For each TF, the ratio of peaks that change in the *SSDDCS* strain over total number of peaks for that TF (x-axis "TF binding changes") was plotted against the ratio of the differential TF peaks that are in the promoters of genes that are DE or non-DE in the *SSDDCS* strain (y-axis "TF effect on gene expression"). The size of the points is proportional to the total number of sub-nucleosomal peaks identified for each TF. The dashed lines mark the ratios when considering all sub-nucleosomal peaks. **F** In wild type cells, nascent DNA is loaded with old and new nucleosomes (yellow) in competition with TFs (blue) as the replication fork progresses. Post-replicative chromatin matures via a variety of mechanisms. In the *SSDDCS* strain, which drives global early replication, lots of concurrent replication forks increase the rate of nascent DNA production, altering the balance of competition between nucleosomes and TFs, leading to aberrant chromatin organization and TF positioning. This increased S-phase rate provides a window of opportunity to change the pattern of gene expression in a single cell cycle

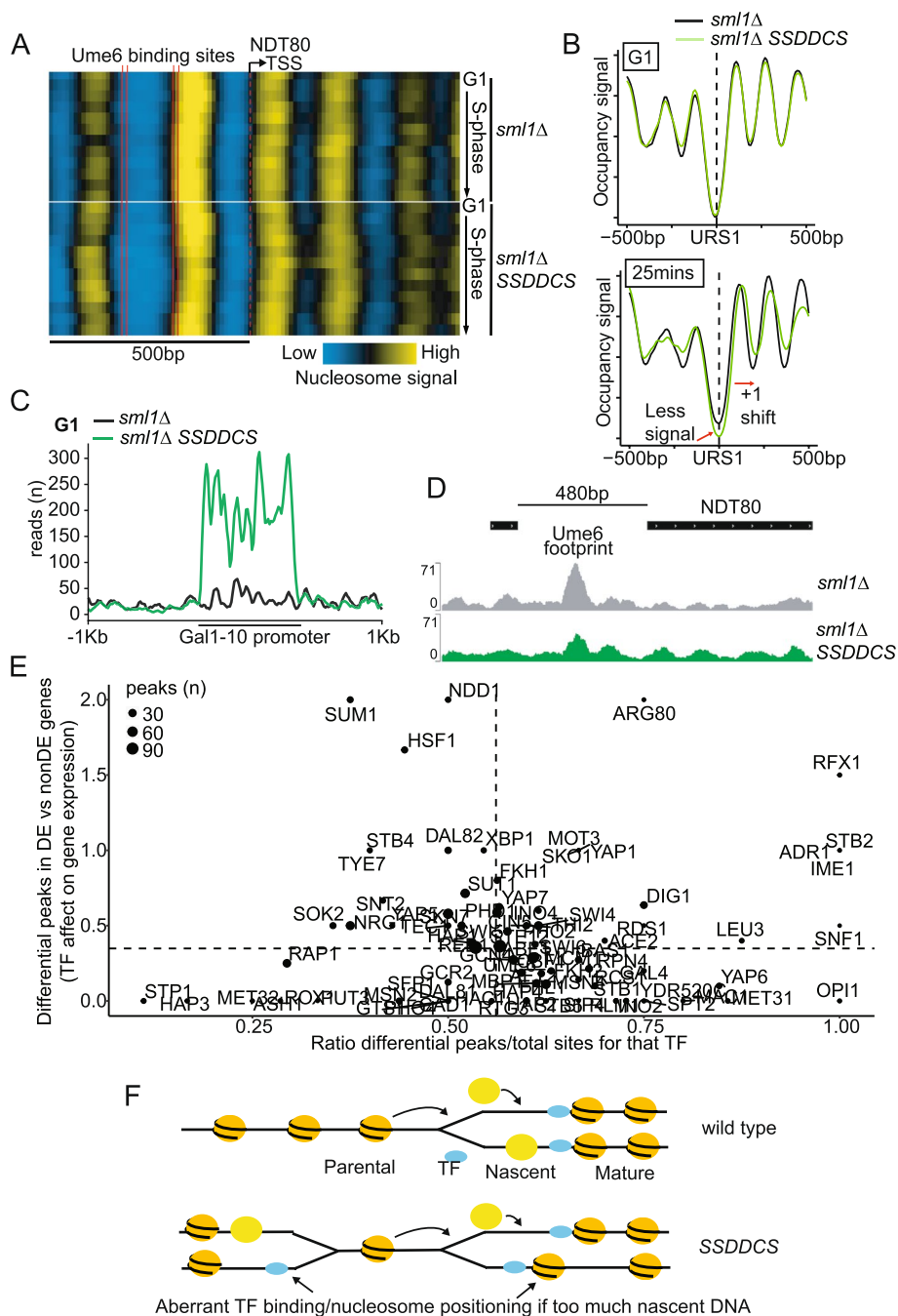


Fig. 6 (See legend on previous page.)

TF binding events can be identified using this method. In line with the nucleosome mapping data (Fig. 6A,B), we also observed changes in sub-nucleosomal peak abundance in the *SSDDCS* strain in the promoter region of *NDT80*, which overlap with the Ume6 binding sites (Fig. 6D). As Ume6 represses *NDT80* expression in mitotic cycles [44], this reduced Ume6 footprint (Fig. 6D) is consistent with the increased *NDT80* expression in the *SSDDCS* strain (Fig. 4). To analyze all sub-nucleosomal peaks and their relationship to changes in gene expression in the *SSDDCS* strain, we plotted the

fraction of differential peaks for a particular TF (as a ratio of that TF's total number of genomic sites, x -axis, Fig. 6E) versus the number of those differential peaks that occur in the promoters of DE genes (y -axis, Fig. 6E). These two ratios allow the simultaneous comparison of the effect of advancing replication timing on TF binding (x -axis) and gene expression for each particular TF (y -axis). From this analysis, we observe that the binding of many TFs to gene promoters is affected in the *SSDDCS* strain that advances replication timing (Fig. 6E). Despite this, altered TF binding is not sufficient for gene expression changes, as differential TF binding also occurs in the promoters of genes that are not differentially expressed after advanced replication timing (Fig. 6E, below dotted line).

Discussion

Here we have used a conditional system to over-express six limiting initiation factors in budding yeast and advance replication timing genome-wide in a single cell cycle. This global advance in replication timing was accompanied by a change in the transcription of approximately 27% of all genes (Fig. 2), as well as by dramatic changes in chromatin organization (Fig. 5) and transcription factor binding (Fig. 6) during S-phase. A potential explanation for these data is that the high rate of replication initiation induced in the *SSDDCS* strain causes a defect in nucleosome positioning/occupancy because the histone supply and/or the chromatin assembly machinery cannot keep pace with the rapid accumulation of nascent DNA (Fig. 6F). Importantly, mutation of the chromatin assembly factor CAF-1 (*cac1Δ*) not only mimics the gene expression changes caused by global early replication (Fig. 5C–E), but also mimics the reduced occupancy and altered positioning of nucleosomes on nascent chromatin (Fig. 5A,B and [45]). From this, it is likely that a significant fraction of the transcriptional changes that we observe are therefore a consequence of the chromatin perturbations that arise from defects in nucleosome deposition on nascent DNA [45]. This does not exclude that some genes may have altered expression directly due to replication timing changes via other mechanisms, but such genes may be masked by the global defects in chromatin assembly.

Transcription factors and nucleosomes compete for binding to nascent DNA [40]. Since we show that global early replication causes dramatic changes in chromatin organization, it is not surprising that we also identify significant changes in TF dynamics (Fig. 6). From our experiments, we cannot determine whether TF dynamics are affected by chromatin perturbation or vice versa. It is clear however that maintaining the balance between replication rate and re-assembly of the chromatin and TF landscape on nascent DNA is likely to be an important aspect of why origin firing is distributed throughout S-phase in wild type cells (Fig. 6F) and may be an explanation for why the *SSDDCS* strain loses viability in galactose [18].

A striking feature of our experiments is that the massive perturbations of chromatin and gene expression that we induce with global early replication return to normal after S-phase (Figs. 2 and 5). Although deposition of new histones is replication fork-dependent, the maturation of chromatin and the re-establishment of nucleosome/TF positioning is dependent on post-replicative factors such as chromatin remodelling [46]. As we performed our experiments in a single cell cycle, it seems that despite the dramatic replication-dependent perturbations in chromatin and TF binding that we induce in the

SSDDCS strain, the post-replicative maturation pathways can restore the normal regulatory landscape before mitosis. Significantly, post-replicative chromatin maturation pathways also restore the normal nucleosomal landscape in *cac1Δ* mutants, which mimic the chromatin and gene expression changes we observe in the *SSDDCS* strain (Fig. 5) [45]. Although the rate of chromatin maturation differs between loci [39], we did not detect any correlation between the genes that are differentially expressed during early replication and sites of rapid or slow maturation kinetics (data not shown). In addition, delaying replication at individual loci in the *SSDDCS* strain by deleting local origins (Fig. 4) was sufficient to prevent abnormal expression of these loci. This suggests that the transient perturbation of chromatin positioning/TF binding caused during replication (Fig. 5) is necessary to create the permissive window for changes in gene expression. This fits well with previous work showing that the complete absence of replication causes very little change in gene expression during S-phase [47], as we would predict that the lack of replication also removes the window of opportunity for chromatin/TF binding changes on nascent DNA (Fig. 6F).

The potential for increased rates of replication to create a window of opportunity for changes in gene expression described in this study has implications for understanding the phenotypic changes in cells in which the rate of replication is regulated or mis-regulated. For example, during erythroid differentiation the shortening of S-phase favors changes in cell fate [48]. In addition, during senescence extra origins become activated [49] and senescent cells experience large-scale changes in chromatin organization and gene expression [50, 51]. As oncogene-activation can also drive senescence, increased origin firing rates, and chromatin/gene expression changes [24, 52], this study may provide a new mechanistic link between changes in the rate of replication and the chromatin/gene expression perturbations in cancers.

The embryonic divisions of many organisms, such as fish, flies, and frogs, are associated with very high rates of replication initiation, but these early cell divisions lack zygotic transcription and the chromatin landscape remains immature during this period [53]. Maintaining high rates of replication delays the normal timing of genome activation [19], but the direct role of replication in the establishment of the chromatin landscape and the transcriptionally competent state of the zygote is poorly understood [54]. This study provides new links between replication control and gene expression changes, which may be relevant to understand how key factors such as Rif1 coordinate the events of early embryonic development, through regulating both chromatin organization and replication timing [22, 55].

Conclusions

Our findings provide new explanations for why all origins do not fire simultaneously at the beginning of S-phase in most cell types. A high rate of replication initiation causes dramatic changes in gene expression, chromatin organization, and transcription factor binding, which is likely to be due to an inability for chromatin assembly to keep pace with the rapid production of nascent DNA. These results have consequences for understanding how cells respond to changes in the rate of origin firing, during development, differentiation, and disease.

Methods

Yeast strains

All strains are *W303a ade2-1 ura3-1 his3-11, 15 trp1-1 leu2-3*, and *112 can1-100 rad5-535*.

Strain	Relevant genotype
PZ356	<i>sml1Δ::URA3</i>
PZ523	<i>leu2::Sld7-P_{GAL1-10}⁻Cdc45::LEU2 his3::SLD3-A-P_{GAL1-10}⁻Dbf4-A::HIS3 trp1::Sld2-P_{GAL1-10}⁻Dpb11::TRP1 sml1Δ::HphMX</i>
PZ1407	As strain PZ523, but <i>ARS1008Δ ARS1009Δ</i>
PZ1435	As strain PZ356, but <i>ARS1008Δ ARS1009Δ</i>
PZ3004	As strain PZ356, but <i>ARS816Δ ARS818Δ::KanMX</i>
PZ3005	As strain PZ523, but <i>ARS816Δ ARS818Δ::KanMX</i>
PZ4505	<i>sml1Δ::HphMX thi12Δ::GFP-PEST-NLS ADE2⁺</i>
PZ4504	<i>leu2::Sld7-P_{GAL1-10}⁻Cdc45::LEU2 his3::SLD3-A-P_{GAL1-10}⁻Dbf4-A::HIS3 trp1::Sld2-P_{GAL1-10}⁻Dpb11::TRP1 sml1Δ::HphMX thi12Δ::GFP-PEST-NLS ADE2⁺</i>

Block and release time course

The *sml1Δ* and *sml1Δ SSDDCS Saccharomyces cerevisiae* strains were grown overnight in YP-raffinose at room temperature. After ensuring that cultures were growing exponentially, 100 ml was transferred to 30°C shaking water bath for one cell cycle. At 1×10^7 cells/ml, 90 μl of stock solution of alpha factor was added to 100ml of culture (1:900 dilution), and after 90 min, 45 μl of stock solution of alpha factor (5mg/ml) was added. To confirm the G1 arrest, cells were analyzed under microscope using a hemocytometer and the arrest was considered successful if more than 95% of cells had the G1 characteristic shape (“shmoo”) or were unbudded. Upon arrest, 10ml of 20% galactose was added to the cultures to induce the over-expression of the six factors. Thirty minutes post galactose addition, G1 samples were collected. Then cultures were washed twice with fresh YP-galactose to release cells from G1 arrest and resuspended in 100ml of YP-galactose. Cultures were maintained at the 30°C shaking water bath for 60 min, and 8ml was taken every 5 min for MNase-seq, 10ml for DNA or RNA profiling, and 500 μl for flow cytometry.

Flow cytometry of yeast with sodium citrate buffer

In total, 500 μl of yeast culture was spun down, then fixed in 500 μl of cold 70% ethanol for 2 h at room temperature (~20°C) or overnight at 4°C. After fixation, cells were centrifuged at 13,300 rpm for 2 min and the pellet was washed with 1ml of 50 mM sodium citrate. Cells were then centrifuged and resuspended in 1ml of 50 mM sodium citrate with 10 μg/ml of RNase and incubated at 37°C for 4 h. After the RNase treatment, cells were centrifuged and resuspended in 50 mM HCl with 5 mg/ml of pepsin and incubated at 37°C for 30 min. Cells were then washed with 1ml of 50 mM sodium citrate. After centrifugation, cells were resuspended in 1ml of 50 mM Tris pH 7.4 with

0.5 µg/ml of propidium iodide (PI). Finally, tubes were vortexed and 100µl was added to FACS tubes with 1ml of 50 mM Tris pH 7.4 with 0.5 µg/ml of PI. Before processing in the cytometer, cells were sonicated for 8 s at 40% amplitude.

Collection of samples for whole genome sequencing (replication profiles)

Yeast genomic DNA was extracted using the smash and grab method (<https://fangman-brewer.genetics.washington.edu/smash-n-grab.html>). DNA was sonicated using the Bioruptor Pico sonicator (Diagenode), and the libraries were prepared according to the TruSeq Nano sample preparation guide from Illumina.

Collection of samples for whole transcriptome sequencing (RNA-seq)

Total RNA was extracted using the RiboPure RNA Purification Kit (Thermo Fisher). Quality and integrity of the RNA was evaluated using an Agilent TapeStation. Libraries were prepared from RNA samples exceeding a RIN quality score of 9.0, according to the TruSeq RNA Library Prep Kit v2 sample preparation guide from Illumina.

GFP experiments

The *sml1Δ* and *sml1Δ SSDDCS thi12Δ::GFP S. cerevisiae* strains were grown as previously described in the block and release experiment, with the modification that cultures were maintained in the 30°C shaking water bath for 150 min. Samples were then plated onto 35-mm glass bottom plates (MatTek) precoated with Concanavalin A (Sigma). After 5 min, the medium was changed to SC medium and cells were imaged on a Delta-vision widefield fluorescent microscope (GE Healthcare), based on an inverted fluorescence microscope (IX70; Olympus) with an oil immersion Plan-Apochromat ×60 NA 1.4 lens (Olympus) for imaging of live cells. Images were acquired, deconvoluted, and projected using SoftWoRx (GE Healthcare). Cells with GFP signal were measured using FIJI.

Digestion of chromatin with MNase for mono-nucleosome analysis (adapted from [36])

Day 1

Eight milliliters of yeast culture collected in each timepoint was centrifuged for 2 min at 4000 rpm and resuspended in 40ml of 1× PBS, 1% formaldehyde. Samples were mixed and left shaking gently for 10 min at room temperature on gyro-rocker to crosslink DNA and proteins. Crosslinking was quenched by adding 5ml of 2.5M glycine and left shaking for 10 mins on orbital shaker at room temperature. Samples were left on ice until all samples have been collected. Then, samples were spun for 5 min at 3200 rpm in 50-ml tubes and washed with 50 ml of sterile ddH₂O. Pellets were resuspended and transferred to 1.7ml Axygen tubes, and spun down at top speed in table top centrifuge. Liquid was carefully aspirated and pellets vortexed. Pellets were resuspended in 950µL of zymolyase digestion buffer (ZDB: 50 mM Tris Cl at pH 7.5, 1 M sorbitol, 10 mM β-mercaptoethanol) to remove the cell wall. Then, 100 µL of freshly prepared zymolyase solution (10mg/ml dissolved in ZDB) was added to each sample, and digestion was performed for 60 min at 30°C shaking gently in water bath. Efficiency of digestion was assessed by checking cell morphology under the microscope: cells with digested cell walls will appear spherical. A second test is to take 1µl and dilute to 20µl with ddH₂O.

As cells no longer have a cell wall, the osmotic shock will burst them. So absence of cells means zymolyase treatment was successful (spheroplasting). Spheroplasts were pelleted in a microfuge, at 5000 rpm for 5 min at 4°C and washed with 1 ml of ZDB. Pellets were then resuspended in 1 ml of spheroplast digestion buffer (SDB: 1 M sorbitol, 50 mM NaCl, 10 mM Tris at pH 8, 5 mM MgCl₂, 1 mM CaCl₂, 1 mM β-mercaptoethanol, 0.15% NP40). Samples were spun down in the microcentrifuge and gently resuspended in 0.5 ml of SDB. Then 90U of MNase (9μl of 10U/μl MNase solution) was added, and tubes were well mixed and left with gentle agitation for 3 min at 37°C. The amount of MNase needs to be experimentally determined by titration with every batch of MNase. MNase digestion was stopped with the addition of 50 μl of 0.5 M EGTA. Tubes were vortexed after adding EGTA. Samples were treated with RNase by adding 2 μl of RNase I (100units/μL) for at least 1 h at 37°C. Ten microliters of freshly made up stock of 10mg/ml Proteinase K was added, and samples were left at 42°C for at least 3 h. The formaldehyde crosslinks were reversed by incubating samples for > 6 h at 65°C.

Day 2

Samples were transferred to 2-ml rubber-sealed screw-cap tubes. One volume of Phenol-Chloroform pH 8 (~570μl) was added to samples and vortexed and spun for 5 min. Aqueous phases were collected to a new 2-ml lo-bind Eppendorf tube, 5 μl of glycogen and 190 μL of 3M sodium acetate were added, and samples were ethanol-precipitated with 1250 μL of cold absolute ethanol (2.5×) and vortexed. Samples were incubated at –20°C overnight and centrifuged at 13,000 rpm for 30 min at 4°C.

Day 3

Pellets were gently washed by adding 1ml of freshly made 70% ethanol. Then pellets were pulsed down quickly and most volume was carefully aspirated, the wash was repeated with 1ml of 70% ethanol, and then tubes were spun down for 15 min at 4°C. Ethanol was carefully aspirated and pellets were air dried for 15 min at room temperature, 100μl of Illumina Resuspension Buffer was added, and samples were incubated for 1 h at 37°C to redissolve DNA. Size range and relative molarity were determined on a Tapestation using D1000 and Genomic screen tape, and total yield was quantified using Qubit broad range dsDNA kit.

Digestion of chromatin with MNase for transcription factor binding analysis

This is the same protocol as the one used for mono-nucleosome analysis described above, with the following modifications in the MNase digestion step: after spheroplasting and resuspension in SDB, 5U of MNase (5μl of 1U/μl MNase solution) was added and tubes were incubated on the bench (room temperature) for 20 min.

Library preparation—MNase-seq mono-nucleosome analysis

A total of 250ng of MNase-digested DNA from each sample was end-repaired using the Illumina TruSeq DNA nano kit. AMPure XP beads were added (1.8× volume of DNA, DNA >100bp on beads, <100bp in supernatant) to each reaction to purify the mono-nucleosomal fragments. A-tailing and adapter ligation was performed using the Illumina TruSeq DNA nano kit. Two subsequent steps of beads purification (1.4× volume of

DNA) were performed in order to remove adapter dimers. Based on tests using hyperladder V (25bp bands) the beads can selectively retain DNA of ~270bp (mono-nucleosomal + adapters) from free adapters (60–120bp) if used at a 1.4× ratio to the volume of DNA sample. PCR cycle quantitation was performed for each sample using KAPA Syber Fast reagents, and libraries were PCR amplified using the Illumina TruSeq DNA nano kit, followed by another step of bead purification (1.4× volume of DNA). Library quality and quantity were validated on a TapeStation using D1000 screen tape, Qubit broad range dsDNA kit, and NEBNext library quantitation kit for Illumina. Libraries were pooled to final 100nM molarity, and one step of bead purification (1× volume of DNA) was performed to completely remove adapter dimers. Finally, 20µl of the pooled libraries at a final 20nM molarity was sequenced in an Illumina HiSeq 1500 platform by the Gurdon Institute Core NGS sequencing facility using 50 bp paired-end reads.

Library preparation—subMNase-seq transcription factor reads

This is the same as the previous with the following modifications: after end-repair and before A-tailing, samples were cleaned by performing a phenol-chloroform precipitation, in order to reduce the volume of the samples without using AMPure XP beads. For adapter ligation, 20% of the amount recommended by Illumina was used (adapters were diluted 1:10 in RSB), in order to minimize adapter dimer formation. This was done because adapter dimers cannot be removed using AMPure XP beads as their size is very similar to TF binding fragments, and smaller fragments are preferentially amplified during library preparation, which means we would be wasting sequencing depth with adapters. After ligation, samples were cleaned using 1.8× AMPure XP beads (DNA >100bp on beads), so TF binding events corresponding to 10–80bp footprint + two adapters (120bp) will bind to the beads.

Quality control and mapping

Sample quality was assessed using FastQC High Throughput Sequence QC Report version 0.11.4. All samples were mapped using bowtie2 (version 2.2.6) to the budding yeast reference genome (strain S288C, version R64-2-1), which was indexed using bowtie2-build. SAM files were then converted to BAM and sorted and indexed using samtools (version 0.1.19). The quality control of the alignments was assessed using Qualimap (version 2.2.1).

Replication profiles

Before generating the replication profiles, sequencing depth was normalized for each timepoint using a bulk value derived from the fraction of the genome that has been replicated at that timepoint (a value between 1 and 2). These values were derived using fit-Sigmoid (<https://dzmitry.shinyapps.io/flowfit/>).

To generate replication timing profiles, the ratio of uniquely mapped reads in the replicating samples to the non-replicating samples was calculated following [28]. Then, this ratio was plotted for each timepoint, and a sigmoid line was fitted. T_{rep} was determined as the time of half-maximal replication (ratio = 1.5). Replication profiles were generated

by plotting T_{rep} values for each chromosome location using ggplot2 and smoothed using a moving average in R. All downstream analysis was performed in R.

RNA-seq analysis

Read counts for each gene were extracted using genomic ranges, and differential expression analysis was performed using DESeq2 [31]. Pair-wise analysis of each timepoint was performed using the Wald test. For the time course analysis, the likelihood ratio test (LRT) was used. Genes were considered differentially expressed if the p -value adjusted value from these tests was < 0.01 . PCA analysis was performed using the variance stabilization transformation (vst command from DESeq2) and plotted using ggplot2. For the k-means clustering, the gene expression data was normalized by row using the scale command and 6 clusters were generated using the k-means command with a maximum of 50 iterations. Heatmaps were generated in R using heatmap.2. Distance to origins, centromeres, and telomeres was calculated using HOMER (v4.10.1) [56]. Statistical analyses were performed using R.

Mono-nucleosome MNase-seq analysis

Nucleosome calls were identified by processing the BAM files using DANPOS (version 2.2.2) [57]: danpos dpos was used to generate wig files. danpos profile was used to generate the files required for the nucleosome profiles, which were plotted in R. The files with the genomic coordinates of the locations where the heatmaps should be centered were generated using UCSC Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgTables>). Peaks in these profiles represent nucleosome dyads and valleys linker DNA or nucleosome depleted regions.

The +1 nucleosome was identified by calculating the distance of nucleosomes to promoters using HOMER. Nucleosomes within -20 to 80bp of the TSS were classified as +1. The +1 relative position to the TSS was calculated by subtracting the genomic position of the nucleosome to the TSS of the corresponding gene. ACF analysis was performed following [39]. The autocorrelation function was used to determine the pattern of organization of the first four nucleosomes (+1, +2, +3, and +4) within each gene, using the nucleosome sized reads between 140 and 180bp overlapping each gene.

Identification of TF binding motifs genome-wide

To identify TF binding regions genome-wide, the sequence motif was extracted in meme format from the JASPAR database [41], which was used as an input to FIMO [42] to identify all genomic locations where this motif is present. These regions were annotated to genes using HOMER.

Sub-nucleosomal MNase-seq analysis

BAM files were converted into bigWig files for visualization of the results using IGV. For this purpose, bamCoverage (version 3.0.2) was used with the following argument: `--bin-Size 1 --minFragmentLength 0 --maxFragmentLength 100`.

Identification of high confidence sub-nucleosomal peaks and calculation of fold-change differences between the strains was performed following [39]: DNA fragments with less than 100bp (sub-nucleosomal events) were selected and the same number of

reads was sampled for each fragment size using the timepoint with the minimum number of reads for that fragment size. Then, all samples were merged to call all possible peaks in the data. A peak was considered high confidence if the sum of reads mapping to that peak (log₂ normalized) across all samples was higher than 75. The log₂ ratio of normalized reads occupying each peak between the two strains was calculated for each timepoint. Sub-nucleosomal peaks were annotated to TSS and TF binding sites using HOMER as described previously.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-022-02788-7>.

Additional file 1. Contains Figures S1-5, which are supplementary data in support of the main figures.

Additional file 2. Review history.

Acknowledgements

We thank members of the Zegerman lab for comments and technical support. We also thank Charles Bradshaw (Gurdon Institute, Cambridge, UK) for bioinformatic support, Kay Harnish (Gurdon Institute) for next-generation sequencing, the Gurdon Institute Media Kitchen and the Gurdon Institute Imaging Facility. We are grateful to Naama Barkai (Weizmann Institute of Science, Israel), Yoav Voicheck (Gregor Mendel Institute of Molecular Plant Biology, Austria), and Craig Peterson (University of Massachusetts Medical School, USA) for sharing data and we thank Mónica Gutiérrez and David MacAlpine (Duke University, USA) for sharing protocols and advice for MNase-seq analysis.

Review history

The review history is available as Additional file 2.

Peer review information

Stephanie McClelland was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions

MS, LF, and MJ performed and designed the experiments and analyses. PZ managed the project and wrote the paper. All author(s) read and approved the final manuscript.

Funding

Work in the PZ lab was supported by AICR 10-0908, Wellcome Trust 107056/Z/15/Z, Cancer Research UK C15873/A12700. MS was funded by the BBSRC BB/M011194/1 and a Cambridge Trust scholarship.

Availability of data and materials

All sequencing data has been deposited in GEO under the accession code GSE199450 [58]. Public datasets used are as follows:

Figure 2E—list of cell cycle regulated genes from [32].

Figure 2F—constitutive / inducible promoters from [33], dataset citation [59].

Figure 5C,E—compendium of 165 chromatin machinery deletion mutants from [38], dataset citation [60].

All materials are available on request.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 6 May 2022 Accepted: 10 October 2022

Published online: 17 October 2022

References

1. Bell SP, Labib K. Chromosome duplication in *Saccharomyces cerevisiae*. *Genetics*. 2016;203:1027–67.
2. Rivera-Mulia JC, Gilbert DM. Replicating large genomes: divide and conquer. *Mol Cell*. 2016;62:756–65.
3. Muller CA, Nieduszynski CA. Conservation of replication timing reveals global and local regulation of replication origin activity. *Genome Res*. 2012;22:1953–62.
4. Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res*. 2010;20:761–70.

5. Kermi C, Lo Furno E, Maiorano D. Regulation of DNA replication in early embryonic cleavages. *Genes (Basel)*. 2017; 8(1):42.
6. Rivera-Mulia JC, Buckley Q, Sasaki T, Zimmerman J, Didier RA, Nazor K, et al. Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. *Genome Res*. 2015;25:1091–103.
7. Rivera-Mulia JC, Sasaki T, Trevilla-Garcia C, Nakamichi N, Knapp D, Hammond CA, et al. Replication timing alterations in leukemia affect clinically relevant chromosome domains. *Blood Adv*. 2019;3:3201–13.
8. Raghuraman MK, Brewer BJ, Fangman WL. Cell cycle-dependent establishment of a late replication program. *Science*. 1997;276:806–9.
9. Dimitrova DS, Gilbert DM. The spatial position and replication timing of chromosomal domains are both established in early G1 phase. *Mol Cell*. 1999;4:983–93.
10. Rhind N, Gilbert DM. DNA replication timing. *Cold Spring Harb Perspect Biol*. 2013;5:a010132.
11. Ferguson BM, Fangman WL. A position effect on the time of replication origin activation in yeast. *Cell*. 1992;68:333–9.
12. Yoshida K, Bacal J, Desmarais D, Padioleau I, Tsaponina O, Chabes A, et al. The histone deacetylases sir2 and rpd3 act on ribosomal DNA to control the replication program in budding yeast. *Mol Cell*. 2014;54:691–7.
13. Vogelauer M, Rubbi L, Lucas I, Brewer BJ, Grunstein M. Histone acetylation regulates the time of replication origin firing. *Mol Cell*. 2002;10:1223–33.
14. Belsky JA, MacAlpine HK, Lubelsky Y, Hartemink AJ, MacAlpine DM. Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. *Genes Dev*. 2015;29:212–24.
15. Hoggard T, Shor E, Muller CA, Nieduszynski CA, Fox CA. A Link between ORC-origin binding mechanisms and origin activation time revealed in budding yeast. *PLoS Genet*. 2013;9:e1003798.
16. Dukaj L, Rhind N. The capacity of origins to load MCM establishes replication timing patterns. *PLoS Genet*. 2021;17:e1009467.
17. Tanaka S, Nakato R, Katou Y, Shirahige K, Araki H. Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Curr Biol*. 2011;21:2055–63.
18. Mantiero D, Mackenzie A, Donaldson A, Zegerman P. Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *EMBO J*. 2011;30:4805–14.
19. Collart C, Allen GE, Bradshaw CR, Smith JC, Zegerman P. Titration of four replication factors is essential for the *Xenopus laevis* midblastula transition. *Science*. 2013;341:893–6.
20. Fang D, Lengronne A, Shi D, Forey R, Skrzypczak M, Ginalska K, et al. Dbf4 recruitment by forkhead transcription factors defines an upstream rate-limiting step in determining origin firing timing. *Genes Dev*. 2017;31:2405–15.
21. Natsume T, Muller CA, Katou Y, Retkute R, Gierlinski M, Araki H, et al. Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. *Mol Cell*. 2013;50:661–74.
22. Richards L, Das S, Nordman JT. Rif1-dependent control of replication timing. *Genes (Basel)*. 2022;13(3):550.
23. Lima-de-Faria A, Jaworska H. Late DNA synthesis in heterochromatin. *Nature*. 1968;217:138–42.
24. Rivera-Mulia JC, Gilbert DM. Replication timing and transcriptional control: beyond cause and effect-part III. *Curr Opin Cell Biol*. 2016;40:168–78.
25. Fraser HB. Cell-cycle regulated transcription associates with DNA replication timing in yeast and human. *Genome Biol*. 2013;14:R111.
26. Klein KN, Zhao PA, Lyu X, Sasaki T, Bartlett DA, Singh AM, et al. Replication timing maintains the global epigenetic state in human cells. *Science*. 2021;372:371–8.
27. Muller CA, Nieduszynski CA. DNA replication timing influences gene expression level. *J Cell Biol*. 2017;216:1907–14.
28. Batrakou DG, Muller CA, Wilson RHC, Nieduszynski CA. DNA copy-number measurement of genome replication dynamics by high-throughput sequencing: the sort-seq, sync-seq and MFA-seq family. *Nat Protoc*. 2020;15:1255–84.
29. McGuffee SR, Smith DJ, Whitehouse I. Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol Cell*. 2013;50:123–35.
30. Morafraila EC, Hanni C, Allen G, Zeisner T, Clarke C, Johnson MC, et al. Checkpoint inhibition of origin firing prevents DNA topological stress. *Genes Dev*. 2019;33:1539–54.
31. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
32. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell*. 1998;9:3273–97.
33. Lu Z, Lin Z. Pervasive and dynamic transcription initiation in *Saccharomyces cerevisiae*. *Genome Res*. 2019;29:1198–210.
34. Voichek Y, Bar-Ziv R, Barkai N. Expression homeostasis during DNA replication. *Science*. 2016;351:1087–90.
35. Osborne EA, Hiraoka Y, Rine J. Symmetry, asymmetry, and kinetics of silencing establishment in *Saccharomyces cerevisiae* revealed by single-cell optical assays. *Proc Natl Acad Sci U S A*. 2011;108:1209–16.
36. Nocetti N, Whitehouse I. Nucleosome repositioning underlies dynamic gene expression. *Genes Dev*. 2016;30:660–72.
37. Jiang C, Pugh BF. A compiled and systematic reference map of nucleosome positions across the *Saccharomyces cerevisiae* genome. *Genome Biol*. 2009;10:R109.
38. Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NA, Margaritis T, et al. The specificity and topology of chromatin interaction pathways in yeast. *Mol Cell*. 2011;42:536–49.
39. Gutierrez MP, MacAlpine HK, MacAlpine DM. Nascent chromatin occupancy profiling reveals locus- and factor-specific chromatin maturation dynamics behind the DNA replication fork. *Genome Res*. 2019;29:1123–33.
40. Ramachandran S, Henikoff S. Transcriptional regulators compete with nucleosomes post-replication. *Cell*. 2016;165:580–92.
41. Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res*. 2004;32:D91–4.
42. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011;27:1017–8.

43. Maclsaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, Fraenkel E. An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics*. 2006;7:113.
44. Pak J, Segall J. Regulation of the premiddle and middle phases of expression of the NDT80 gene during sporulation of *Saccharomyces cerevisiae*. *Mol Cell Biol*. 2002;22:6417–29.
45. Fennessy RT, Owen-Hughes T. Establishment of a promoter-based chromatin architecture on recently replicated DNA can accommodate variable inter-nucleosome spacing. *Nucleic Acids Res*. 2016;44:7189–203.
46. Stewart-Morgan KR, Petryk N, Groth A. Chromatin replication and epigenetic cell memory. *Nat Cell Biol*. 2020;22:361–71.
47. Omberg L, Meyerson JR, Kobayashi K, Drury LS, Diffley JF, Alter O. Global effects of DNA replication and DNA replication origin activity on eukaryotic gene expression. *Mol Syst Biol*. 2009;5:312.
48. Hwang Y, Hidalgo D, Socolovsky M. The shifting shape and functional specializations of the cell cycle during lineage development. *WIREs Mech Dis*. 2021;13:e1504.
49. Rivera-Mulia JC, Schwerer H, Besnard E, Desprat R, Trevilla-Garcia C, Sima J, et al. Cellular senescence induces replication stress with almost no affect on DNA replication timing. *Cell Cycle*. 2018;17:1667–81.
50. Sati S, Bonev B, Szabo Q, Jost D, Bensadoun P, Serra F, et al. 4D genome rewiring during oncogene-induced and replicative senescence. *Mol Cell*. 2020;78(522–538):e529.
51. Hu Z, Chen K, Xia Z, Chavez M, Pal S, Seol JH, et al. Nucleosome loss leads to global transcriptional up-regulation and genomic instability during yeast aging. *Genes Dev*. 2014;28:396–408.
52. Chandra T, Ewels PA, Schoenfelder S, Furlan-Magaril M, Wingett SW, Kirschner K, et al. Global reorganization of the nuclear landscape in senescent cells. *Cell Rep*. 2015;10:471–83.
53. Vastenhouw NL, Cao WX, Lipshitz HD. The maternal-to-zygotic transition revisited. *Development*. 2019;146:dev161471.
54. Hug CB, Vaquerizas JM. The birth of the 3D genome during early embryonic development. *Trends Genet*. 2018;34:903–14.
55. Sella CA, O'Farrell PH. Rif1 prolongs the embryonic S phase at the *Drosophila* mid-blastula transition. *PLoS Biol*. 2018;16:e2005687.
56. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–89.
57. Chen K, Xi Y, Pan X, Li Z, Kaestner K, Tyler J, et al. DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res*. 2013;23:341–51.
58. Santos MM, Johnson MC, Fiedler L, Zegerman P. Global early replication disrupts gene expression and chromatin conformation in a single cell cycle. *Gene Expression Omnibus*. 2022. [GSE199450](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE199450).
59. Lu Z, Lin Z. Pervasive and Dynamic Transcription Initiation in *Saccharomyces cerevisiae*. NCBI BioProject database. 2018. [PRJNA483730](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA483730).
60. Lenstra T, Benschop J, Kim T, Schulze J, Brabers N, Margaritis T, van der Pasch L, van Heesch S, Groot Koerkamp M, Ko C, van Leenen D, Sameith K, van Hooff S, Lijnzaad P, Kemmeren P, Henrich T, Kobor M, Buratowski S, Holstege F. Gene deletion expression profiles of yeast chromatin modifiers. *Gene Expression Omnibus*. 2011. [GSE25909](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25909).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

