Molecular Cell
Resource

Quantitative, Genome-Wide Analysis of Eukaryotic Replication Initiation and Termination

Sean R. McGuffee,[1,2] Duncan J. Smith,[1,2] and Iestyn Whitehouse[1,*]
[1]Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA
[2]These authors contributed equally to this work
*Correspondence: whitehoi@mskcc.org
http://dx.doi.org/10.1016/j.molcel.2013.03.004

SUMMARY

Many fundamental aspects of DNA replication, such as the exact locations where DNA synthesis is initiated and terminated, how frequently origins are used, and how fork progression is influenced by transcription, are poorly understood. Via the deep sequencing of Okazaki fragments, we comprehensively document replication fork directionality throughout the *S. cerevisiae* genome, which permits the systematic analysis of initiation, origin efficiency, fork progression, and termination. We show that leading-strand initiation preferentially occurs within a nucleosome-free region at replication origins. Using a strain in which late origins can be induced to fire early, we show that replication termination is a largely passive phenomenon that does not rely on *cis*-acting sequences or replication fork pausing. The replication profile is predominantly determined by the kinetics of origin firing, allowing us to reconstruct chromosome-wide timing profiles from an asynchronous culture.

INTRODUCTION

Chromosomal replication origins in *S. cerevisiae* have been mapped at high resolution with a variety of techniques; *S. cerevisiae* is unusual among eukaryotes in possessing consensus sequences sufficient for origin activity (Kearsey, 1984; Marahrens and Stillman, 1992; Rao et al., 1994). Most recent approaches to origin mapping and replication profiling rely on pulse-chase experiments in which cultures are synchronously released into S phase, harvested along a time course, and assayed for incorporation of detectable nucleotides (Fachinetti et al., 2010; Raghuraman et al., 2001), increased copy number (Yabuki et al., 2002), single-stranded DNA (Feng et al., 2006), or occupancy of replication fork proteins (Sekedat et al., 2010). These approaches provide a wealth of data describing average replication behavior across a population of cells, thus informing mathematical models of genome replication (de Moura et al., 2010; Yang et al., 2010). Pioneering work by Raghuraman et al. (2001) provided evidence that distinct subsets of replication origins fire with predictable timing and efficiencies at defined intervals during S phase. However, in contrast to the somewhat deterministic view of genome-wide replication obtained from such studies, detailed single-molecule analysis of *S. cerevisiae* chromosome VI via DNA combing (Czajkowsky et al., 2008) provided evidence that no two cells have identical patterns of origin use, implying a globally stochastic pattern of independent origin firing.

Current replication profiling techniques are limited by the lack of a direct readout of replication fork directionality in regions more than a few kilobases from efficient origins. The extent of variation within a population has therefore proved difficult to assay directly—assignment of a signal to a given origin becomes difficult once signals arising from convergent forks have merged. Because one cannot clearly distinguish between incoming replication forks from either direction and the firing of inefficient origins that are predominantly passively replicated, quantitative analysis of origin efficiencies has only been carried out in a low-throughput fashion; e.g., as in Friedman et al. (1997) and Yamashita et al. (1997).

Relative to replication initiation, comparatively little is known about how and where convergent replication forks terminate. In some regions of eukaryotic genomes, the sites of fork convergence are precisely determined by *cis*-acting barriers analogous to the Tus-Ter system in *E. coli* (Hill and Marians, 1990). For example, a polar barrier within the *S. cerevisiae* ribosomal DNA (rDNA) repeat, comprising the Fob1 protein in complex with the replication-fork-blocking sequence RFB, impedes the passage of replication forks moving in one direction (Brewer and Fangman, 1988; Kobayashi and Horiuchi, 1996) and, thus, ensures unidirectional replication of the repeat region. However, termination at RFBs most likely accounts for a tiny fraction of all regions of termination. Genomic regions with high occupancy of nonnucleosomal proteins, such as centromeres and highly transcribed genes, are known to be problematic for replication fork progression and can elicit stable replisome pausing (Deshpande and Newlon, 1996; Greenfeder and Newlon, 1992; Ivessa et al., 2003). Passage of the replication fork through potential pausing elements is promoted by the action of the Rrm3 helicase (Ivessa et al., 2003), but recent work has postulated the existence of specific termination zones (TERs) in which chromosomal features that mediate fork pausing can slow replication fork progression to the extent that converging forks will likely meet in their vicinity, restricting termination to a defined region (Fachinetti et al., 2010).

Here, we analyze Okazaki fragments by deep sequencing to generate a high-resolution view of the *S. cerevisiae* replication program. We provide detailed measurements of the efficiencies

of all replication origins and regions of termination and demonstrate a preference for leading-strand initiation within the nucleosome-free region generally found at origins. In addition, we present evidence that S phase follows a temporal program dominated by replication origins firing with high probability within distinct time intervals. Contrary to expectation, we find that centromeres and highly transcribed regions are not strong determinants of replication termination; rather, termination generally occurs midway between two adjacent replication origins at positions dictated by their relative firing times. Therefore, sites of termination are indicative of origin-firing time, allowing us to reconstruct the temporal dynamics of the replication program with the use of data from an asynchronous culture.

RESULTS

Deep Sequencing Okazaki Fragments for Replication Profiling

We have developed methods for the purification and deep sequencing of Okazaki fragments from *S. cerevisiae* (Smith and Whitehouse, 2012). DNA ligase I is degron-tagged and placed under the control of a doxycycline-repressible promoter. After ligase repression, short single-stranded DNA fragments are purified and sequenced using the Illumina HiSeq or Ion Torrent platforms. An important aspect of our strategy is the preservation of strand identity, which allows us to unambiguously distinguish Okazaki fragments replicated as the Watson or Crick strand (arising, respectively, from leftward- or rightward-moving replication forks). Replication origins are readily detected as sharp transitions from leftward- to rightward-moving replication forks; origin efficiency—a measure of the likelihood that a replication origin is used during S phase—is proportional to the magnitude of the transition at the origin. A unique advantage of our methodology is that regions of termination, arising from the convergence of two oppositely oriented replication forks, can also be detected as transitions with the opposite strand bias to origins (Figure 1).

We reasoned that transiently repressing DNA ligase I in an asynchronous culture would provide a "snapshot" of Okazaki fragments produced throughout S phase, and we were able to obtain coverage of the entire genome from a single library after 2.5 hr of repression. We developed a computational framework to systematically identify replication origins and sites of termination from our data; our method compares the density of Okazaki fragments on the Watson and Crick strands within a four-part sliding window composed of equally sized, strand-specific 10 kb quadrants around each base pair of each chromosome (see Experimental Procedures). The upper two quadrants (WL and WR) measure Okazaki fragment density in the left and right quadrants on the Watson strand; whereas the lower two quadrants (CL and CR) measure Okazaki fragment density in the left and right quadrants on the Crick strand (Figures 1A and 1C). All quadrant scores were normalized with respect to total Okazaki fragment density within the left or right side of the sliding window, as appropriate, such that $WL_n = WL_{raw}/(WL_{raw} + CL_{raw})$. At an idealized origin that fires in each cell in the population, left quadrants would entirely indicate leftward fork motion ($WL_n = 1$, $CL_n = 0$), and right quadrants would entirely indicate rightward fork motion ($WR_n = 0$, $CR_n = 1$). We used quadrant values to calculate an origin efficiency metric (OEM, defined as $OEM = WL_n - WR_n$) ranging from −1 to 1 for each base in the genome. Localized maxima in the OEM score represent replication origins, and OEM (from zero to one) is proportional to origin-firing efficiency (Figures 1B and 1D). Regions of termination are captured on our OEM plot as localized minima; the degree of termination at each position can be measured from 0 to −1 (where −1 theoretically represents a point terminator between two origins that invariably fire; Figure 1D). Since sites of termination often span several KB, we term them fork merger zones (FMZs).

Global Analysis of Replication Origin Location and Efficiency

To assay the robustness of our origin location and efficiency calls, we applied our algorithm to replicate data sets of a *lig4Δ, rad9Δ* strain (hereafter referred to as wild-type) used previously (Smith and Whitehouse, 2012). Of the 302 and 318 origins meeting the minimum criteria in replicate data sets A and B, respectively, 283 were shared between the two sets. An essentially complete list of sequences capable of acting as replication origins in *S. cerevisiae* is available in the OriDB (Nieduszynski et al., 2007), and each origin is classified as "confirmed," "likely," or "dubious" depending on the number and type of experimental approaches validating its use. For origins predicted in both data sets, 221 (213 unambiguously) corresponded to (defined as lying within 2.5 kb) a confirmed OriDB origin, 44 (42 unambiguously) to a likely origin, and 4 (all unambiguously) to a dubious origin; a further 14 could not be assigned to any origin (Figure 2A). Ambiguous calls result from OriDB origins that are closer together than can be distinguished by our matching protocol.

The correspondence between our predictions and pre-existing origin identifications in *S. cerevisiae* indicates that our methodology represents an effective way to locate replication origins; those identified in only one data set were significantly less likely to match with confirmed origins (Figure 2A). Poor correspondence with OriDB and low efficiency scores (Figure S4A available online) suggest either that these calls represent extremely inefficient origins that fall below the detection limit across a variety of data sets or that they are false positives resulting from noise in our experimental data. Furthermore, our data provide high spatial resolution; the median distance between replicate origin midpoint calls was 90 bp, and over 90% of midpoints fell within a ±1.5 kb range (Figure 2B). Of 253 origin ARS consensus sequence (ACS) sites mapped by Eaton et al. (2010), 186 lie within 2.5 kb of an active origin in both data sets with a median distance of ~200 bp to the ACS midpoint (Figure 2B).

The efficiency of equivalent origins in replicate data sets was highly correlated (Figure 2C; $r^2 = 0.85$); across chromosome 6 (Figure 2D), our measured origin efficiencies agree well with those previously obtained via two-dimensional gels (Friedman et al., 1997; Yamashita et al., 1997). Additionally, we find that origins that replicate within the first half of S phase are significantly more efficient than those generally replicated in the second half (Figure 2E), although we note that many examples of inefficient early origins and efficient late origins do exist.

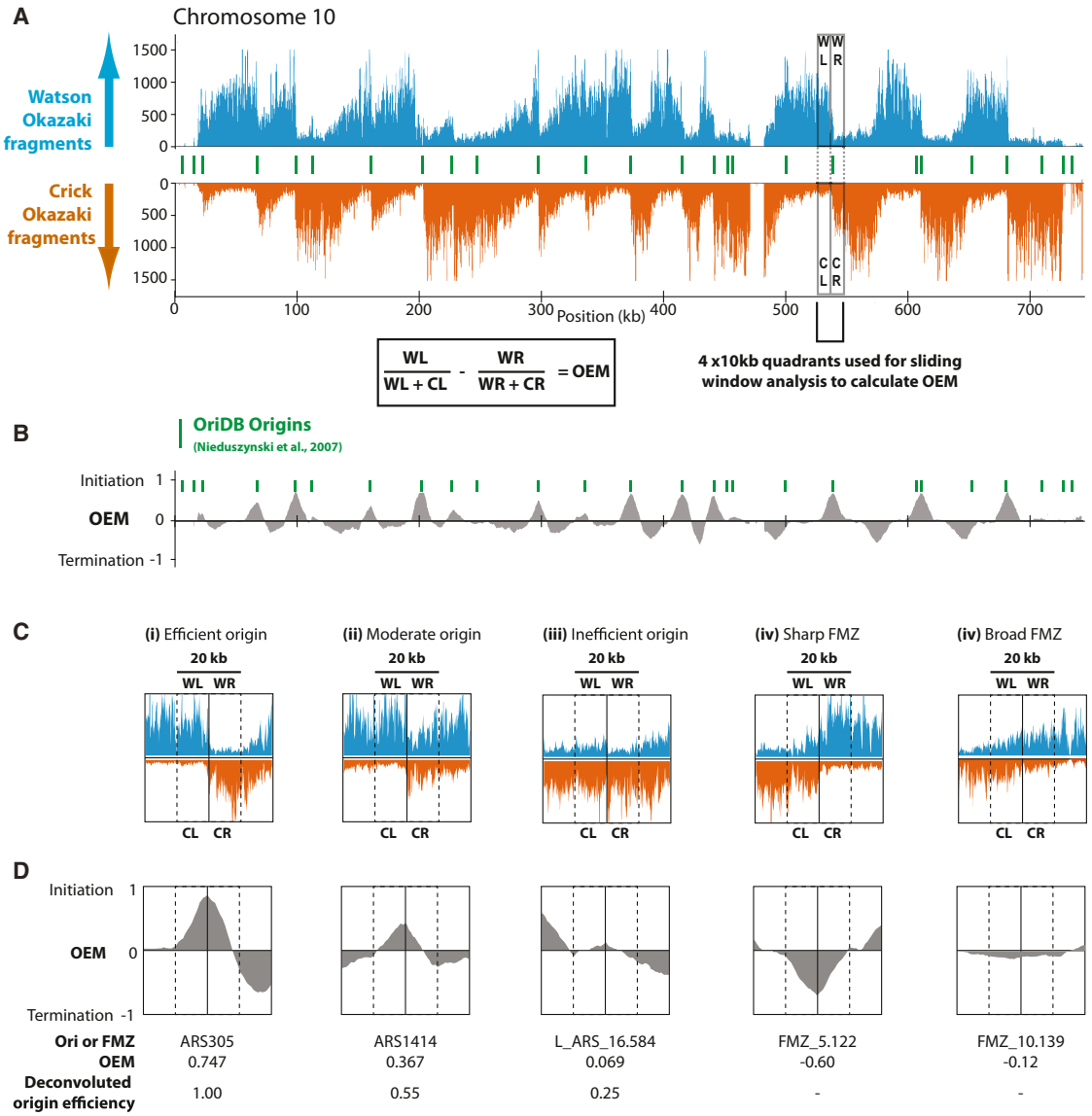

Figure 1. Quantitative Identification of Replication Origins and Fork Merger Zones from Okazaki Fragment Maps

(A) Sequenced Okazaki fragments mapping to the Watson (blue) or Crick (orange) strand on *S. cerevisiae* chromosome 10. Watson or Crick strand fragments indicate leftward- or rightward-moving replication forks, respectively. Both replication origins and fork merger zones (FMZs) can be readily identified with the use of a sliding window to generate an origin efficiency metric (OEM).

(B) Across a 20 kb window, the proportion of Watson strand hits in the right half is subtracted from that in the left, generating peaks at origins whose height is proportional to their firing efficiency and analogous troughs at FMZs. Positions of replication origins listed as "confirmed" in the OriDB yeast origin database (Nieduszynski et al., 2007) are indicated in green.

Other chromosomes are shown in Figure S6.

(C and D) Okazaki fragment density and OEM across a 40 kb window for representative origins of (i) high, (ii) moderate, and (iii) low efficiency, and for representative sharp (iv) and broad (v) FMZs. Raw OEM score and calculated origin efficiency differ because of the deconvolution process described in Experimental Procedures and Figures S1–S3.

Interestingly, we find no correlation between origin efficiency and origin recognition complex (ORC) or minichromosome maintenance (MCM) levels determined by chromatin immunoprecipitation (ChIP) (Xu et al., 2006) (Figures 2F and S4B), suggesting that the presence of multiple pre-RCs does not significantly contribute to origin efficiency, as previously proposed (Yang et al., 2010).

Leading-Strand Synthesis Is Preferentially Initiated within the Nucleosome-Free Region at Origins

As well as providing a global view of replication origin use, our high-resolution sequence data allow us to investigate whether initiation occurs at locations specified by the ACS that are asymmetrically located within a nucleosome-free region (NFR) at most active replication origins in *S. cerevisiae* (Eaton et al.,

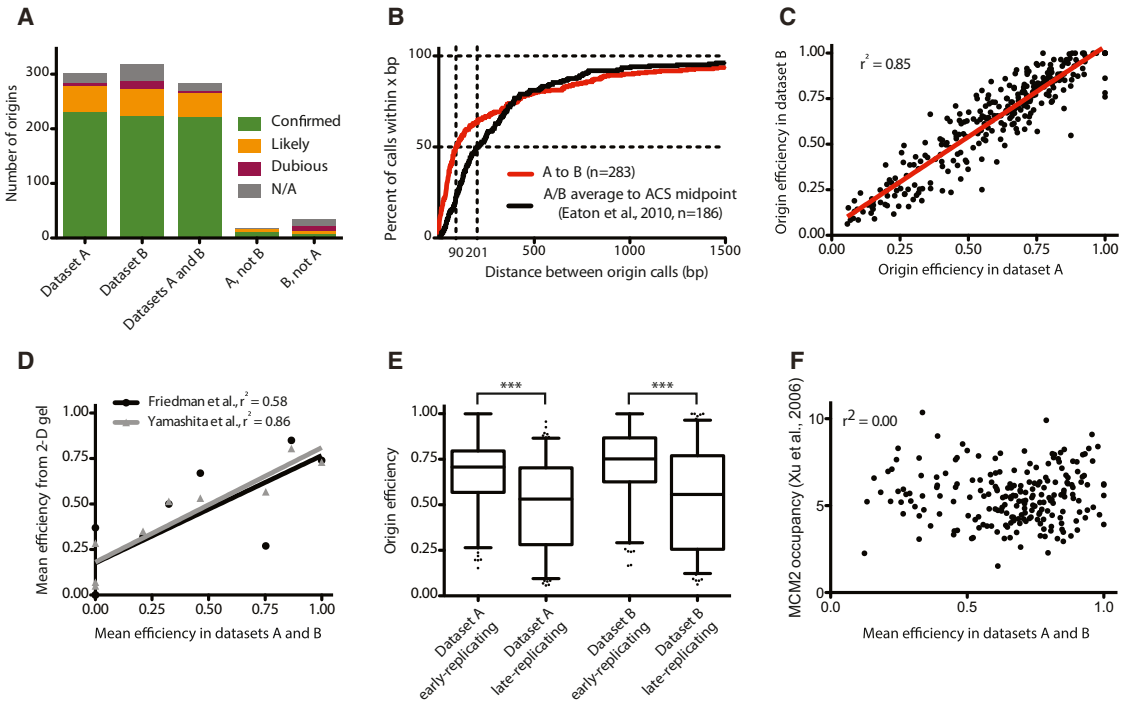

Figure 2. Validation of Replication Origins Identified from Okazaki Fragment Sequencing and Global Evaluation of Origin Efficiency across Replicate Data Sets

(A) Identified origins correspond well with previous reports. Column graphs indicate the number of origins identified in replicate data sets lying within ±2.5 kb of OriDB (Nieduszynski et al., 2007) origins listed as "confirmed," "likely," or "dubious." N/A indicates an origin that is not within 2.5 kb of any OriDB entry. Origins within 5 kb of one another in replicate data sets were deemed to be the same.

(B) Origins are accurately identified in replicate data sets. The percentage of the 284 matched origin pairs called less than x bp apart in replicate data sets is plotted in red on the y axis. The median distance between matched origins is 90 bp. A total of 186 origins lie within ±2.5 kb of an ACS sequence corresponding to an active origin (Eaton et al., 2010). The black line represents the percentage of these origins within x bp of the ACS midpoint and has a median of 201 bp.

(C) Origin efficiencies are consistent in replicate data sets. The red line corresponds to a linear regression.

(D) Origin efficiencies agree well with values determined by two-dimensional gel electrophoresis. Mean efficiencies calculated biochemically for origins across chromosome 6 (Friedman et al., 1997; Yamashita et al., 1997) are compared to mean efficiencies for the corresponding origins across our two data sets. Origins not detected in our analysis are assigned an efficiency of zero. Lines correspond to linear regressions.

(E) Origins in early-replicating regions of the genome are significantly more efficient than those in late-replicating regions. Origins were ranked by t_{rep} (Raghuraman et al., 2001) and divided into bins above and below the median value. $p < 0.0001$ for each data set. The box represents lower, median, and upper quartile; whiskers denote the 5th and 95th percentiles.

(F) Origin efficiency does not correlate with Mcm2 ChIP. The mean origin efficiency for origins identified in replicate wild-type data sets and within 2.5 kb of Mcm2 ChIP peaks (Xu et al., 2006) is plotted against Mcm2 occupancy.

See also Figure S4.

2010). At each origin, the 5′ end of the leading strand will be juxtaposed with the 3′ end of the first Okazaki fragment synthesized by the oppositely oriented replication fork. Although our spatial resolution is limited by degradation of the RNA primer and a variable amount of DNA from the 5′ end of the leading strand by RNase H, Pol δ, and associated nucleases (Stith et al., 2008), we can infer leading strand initiation sites by mapping Okazaki fragment 3′ ends at replication origins. We aligned Okazaki fragment termini around the 186 ACS sequences predicted to be used as origins in both of our data sets. Replication forks moving away from the ACS in either direction generate an overlapping pair of Okazaki fragments whose 5′ and 3′ ends cluster, respectively, around the −1 and +1 nucleosomes flanking the NFR (Figures 3A and 3B); Okazaki fragment end density falls rapidly to background levels in regions that are replicated on the leading strand. Thus, much like the rest of the genome, the ends of the Okazaki fragments we map at replication origins are generally positioned by nucleosome-inhibited strand-displacement synthesis by the lagging-strand polymerase Pol δ (Smith and Whitehouse, 2012). We note that this analysis indicates a preference for initiation within the NFR but does not rule out initiation outside this region at some origins.

Global Analysis of Mergers between Convergent Replication Forks

Recently, replication TERs have been proposed to play a role in the replication of the yeast genome. TERs were operationally defined as genomic regions that are unreplicated late in S phase when replication is slow (in the presence of hydroxyurea or at temperatures below 16°C) (Fachinetti et al., 2010). TERs were

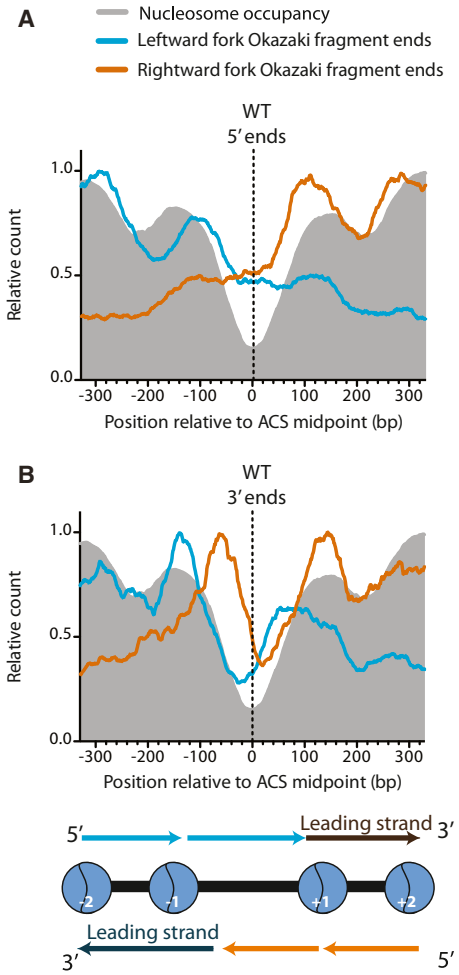

Figure 3. Leading-Strand Initiation Preferentially Occurs within the Nucleosome-Free Region at Replication Origins

(A and B) Distribution of 5′ (A) and 3′ (B) ends observed for Okazaki fragments arising from leftward- (blue) or rightward-moving (orange) forks. Data are oriented such that the T-rich strand of the ACS runs from left to right, and the data are aligned to the ACS midpoint (Eaton et al., 2010). As expected, if leading-strand initiation occurs within the NFR and Pol δ is removed during strand displacement by nascent nucleosomes (Smith and Whitehouse, 2012), a pair of overlapping Okazaki fragments spans the NFR with their termini clustering over the flanking +1 and −1 nucleosomes.

shown to be correlated with genomic features, such as centromeres and sites of high Pol II or Pol III transcription, which can induce replication fork pausing. Our data, which report population-wide replication fork directionality, make a systematic genome-wide analysis of replication fork convergence possible.

We divided the yeast genome into segments, each comprising two replication origins and their corresponding FMZ. We calculated FMZ midpoint using a folded cumulative probability distribution (Monti, 1995) (see Experimental Procedures); a probability was calculated for the FMZ in each segment as the product of the efficiency of each flanking origin and the probability of skipping each intervening origin between the two (Figure 4A; see Experimental Procedures). Of the 714 FMZs shared between our replicate wild-type data sets, most were predicted to be rarely used. Applying a probability cutoff of 0.1 in each data set produced 346 shared FMZs and reduced the median distance between calculated FMZ midpoints to ~1,800 bp (Figure 4B). FMZs with a probability greater than 0.1 were used for all subsequent analyses, because applying a more stringent cutoff of 0.2 did not substantially improve the correspondence between replicate midpoints (Figure 4B). Contrary to the previously reported association between termination regions and high transcription, we observed no enrichment of either RNA Pol II occupancy (Figure 4C) or transfer RNA (tRNA) genes (Figure 4D) around FMZ midpoints as compared to equal numbers of random genomic locations. Although our technique is not sufficiently sensitive to detect pausing at a specific site by a small percentage of forks, the lack of enrichment of highly transcriptionally active loci about FMZs is inconsistent with a widespread role for such loci in termination.

Of our FMZ midpoints, 60 were located within ±5 kb of a region identified as a TER (Fachinetti et al., 2010), 56 of 71 TERs were matched, and 4 TERs were located within 5 kb of more than one FMZ. In these ambiguous situations, both FMZs were assigned TER status (Figure 4E). FMZs corresponding to TER regions are hereafter referred to as TER-FMZs. We noticed several significant differences between TER- and non-TER-FMZs: the former have disproportionately high probabilities (Figure S5A) and replicate early in S phase (Figure S5B). Additionally, we found TERs to be "sharper"—i.e., to represent abrupt transitions from Crick to Watson strand Okazaki fragments, which is indicative of termination occurring over a narrow genomic range (Figure 4F; cf. Figures 1C and 1D). Because our OEM score reports directly on merger sharpness—more negative OEM score indicates sharper mergers—we can assess the range over which convergent forks merge. TER-FMZs showed significantly more negative OEM scores than other FMZs (Figure 4F; $p < 0.0001$ in each data set). Sharp transitions from rightward- to leftward-moving forks could arise because of replication fork stalling at *cis* elements, but, in light of the observation that TERs are generally flanked by early-firing efficient origins whose firing times presumably overlap, we speculated that sharp terminators could simply arise from the near-synchronous firing of adjacent origins. Consistent with our hypothesis, termination tends to occur at the midpoint between two origins (see below), and, assuming that t_{rep}—the time of half-maximal replication (Raghuraman et al., 2001)—can be used to approximate firing time for efficient origins, the origins flanking a sharp FMZ are more likely to fire at similar times to one another than those flanking a broad FMZ (Figure 4G; $p < 0.0001$) (Raghuraman et al., 2001).

Altering the Replication Program and FMZs

To extend our analysis of the replication program and test the hypothesis that the locations of FMZs are passively determined by origin firing kinetics, we wished to sequence Okazaki fragments from a strain in which the normal replication profile could be altered. We used a strain recently described by the Zegerman laboratory, in which additional copies of Sld2, Sld3, Dbf4, Dpb11, Cdc45, and Sld7 (SSDDCS)—six factors limiting for

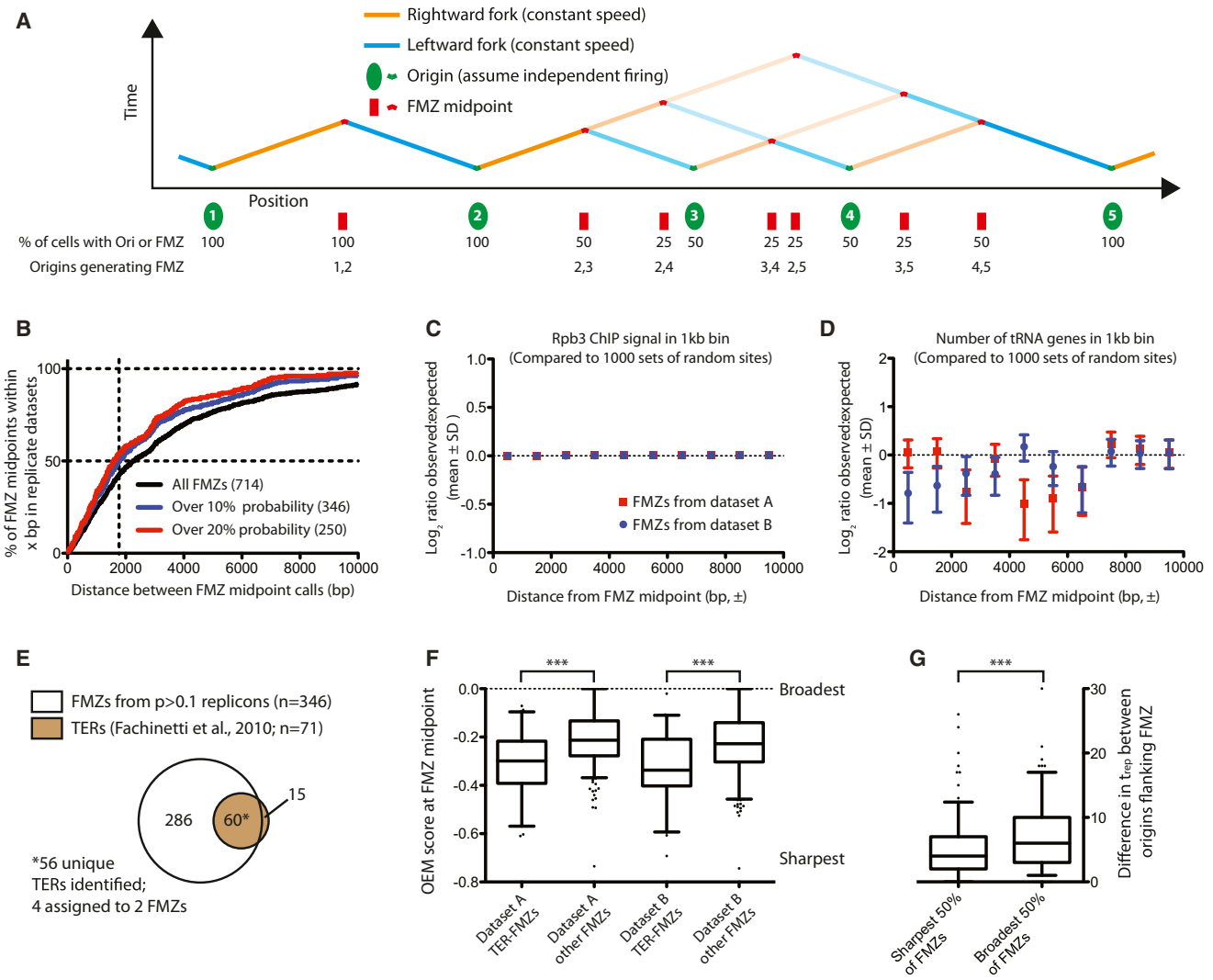

Figure 4. Global Analysis of Replication Fork Mergers

(A) A schematic depiction of FMZ midpoint location and probability for a hypothetical chromosomal region containing origins with various efficiencies.

(B) The midpoints of moderate and high-probability FMZs can be reproducibly identified. For FMZs identified in replicate data sets (within 5 kb), the percentage of replicate FMZ midpoints less than x bp apart is plotted on the y axis. Probability cutoffs of 0% (black line), 10% (blue line), or 20% (red line) are shown. The 346 FMZs with efficiency over 10% in each replicate data set (i.e., present in > 10% of cells in the population) were used for subsequent analyses.

(C) Sites of high transcription are not enriched around FMZs. For FMZs in either data set A or data set B calculated to have probability >10% (n = 511 for data set A, 536 for data set B), the total Rbp3 ChIP signal from Fachinetti et al. (2010) was calculated in 1 kb bins up- and down-stream of the FMZ midpoint; an identical analysis was carried out for an equal number (i.e., 511 or 536) of random genomic sites, and the ratio of signal around FMZs to signal around random sites (i.e., observed:randomly expected) was calculated in each bin for each sample. This process was iterated 1,000 times. The Log2-transformed ratio of observed:expected is plotted (mean ± SD) for the 1,000 iterations of random sites. FMZs from data set A are plotted in red, and those from data set B are in blue. Error bars are too small to be visualized in this plot. Significant deviation above zero, which would indicate enrichment of RNA PolII around FMZs, is not observed.

(D) tRNA genes are not enriched around FMZs. Analysis is identical to that in (C) but follows the presence of tRNA genes around FMZs or random genomic sites. As in (C), significant deviation above zero is not observed.

(E) FMZs identified by our analysis encompass the majority of TER regions previously identified (Fachinetti et al., 2010). Of the 71 TERs, 56 were located within ±5 kb of at least one calculated FMZ midpoint, and 4 TERs were assigned to two FMZs, because they could not be unambiguously assigned to a single one.

(F) FMZs previously identified as TERs are generally sharper than bulk FMZs. The OEM score at the FMZ midpoint is plotted ($p < 0.0001$ for each data set). OEM negatively correlates with FMZ sharpness; a theoretical point terminator between origins that invariably fire will give rise to an OEM of −1, whereas 0 indicates no termination or origin activity (cf. Figures 1C and 1D).

(G) Origins flanking sharp FMZs fire at approximately the same time. FMZs were binned by mean OEM score across two data sets. The difference between t_{rep} values for each origin in the replicon is plotted ($p < 0.0001$, t test). Note that, for origins with an efficiency >0.5, half-maximal replication time (i.e., t_{rep}) is equivalent to average firing time.

See also Figure S5.

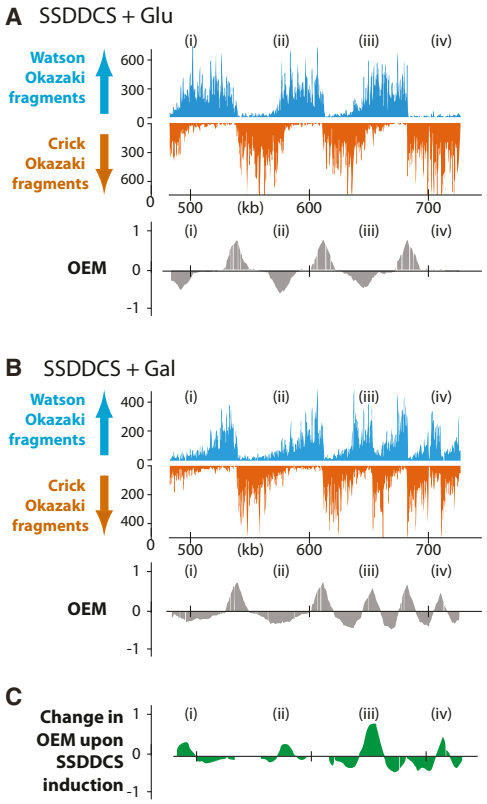

Figure 5. Alteration of the Global Replication Profile by the Overexpression of Limiting Factors

(A–C) Sequenced Okazaki fragments mapping to the Watson or Crick strands in a 250 kb region of chromosome 10 in the SSDDCS strain, in which additional copies of Sld2, Sld3, Dbf4, Dpb11, Cdc45, and Sld7 are expressed from galactose-inducible promoters (Mantiero et al., 2011). Cells were grown in YEP + glucose or galactose for 4 hr prior to ligase inactivation for 2 hr. OEM is shown for each growth condition, along with a difference map in (C), which was calculated by subtracting the OEM signal in (A) from the OEM signal in (B). Several differences are highlighted: (i) FMZ moves and becomes broader; (ii) FMZ becomes broader; (iii) a latent origin is activated, resulting in FMZ movement; and (iv) a latent origin is activated, generating new FMZs.

See Table S1 for genotypes of strains used in this work.

origin activation—could be overexpressed by induction with galactose and, thereby, cause the premature firing of many normally late-firing origins (Mantiero et al., 2011). We obtained single-end sequencing reads of nucleosome-sized Okazaki fragments (Smith and Whitehouse, 2012) from cultures grown in YEP with either glucose (one data set) or galactose (two replicate data sets) prior to and during ligase shutoff. Consistent with our expectations, many origins become highly efficient when cells were grown in inducing conditions (Figures 5A–5C; see Figure S6 for all chromosomes). Our algorithm identified 245 common origins in the three SSDDCS data sets, 224 of which were shared among all five data sets analyzed in this work (Figure 6A; details of origins and FMZs identified in each data set are provided in Supplemental Information). Importantly, the spatial precision with which origins were identified by our algorithm was almost entirely unaffected by strain background and SSDDCS overexpression (Figure 6B).

As expected, global origin efficiency was substantially higher upon SSDDCS overexpression (Figure 6C). Strikingly, the strong correlation between origin efficiencies in the wild-type strain and the SSDDCS strain was completely abolished upon galactose induction (Figure 6D), indicating that the global efficiency increase is not simply due to a constant activation of each origin. Indeed, 31 of the 224 shared origins were more efficient in the wild-type strain than under conditions of origin hyperactivation—in large part because of increased passive replication from hyperactivated nearby origins. An obvious conclusion from these data is that origin efficiency is a composite of many factors, including, but not necessarily limited to, MCM loading and activation (Sheu and Stillman, 2006), SSDDCS activity, and the presence of nearby origins (which tends to suppress firing by causing the origin to be passively replicated). With SSDDCS at saturating concentrations, other properties will, by definition, become limiting for firing efficiency, thus changing the global profile of origin use somewhat independently of normal origin efficiency.

We note that the origin efficiencies observed in the SSDDCS overexpression strain grown in glucose are generally slightly higher than those observed in the (lig4Δrad9Δ) wild-type strain (Figure 6C). Checkpoint abrogation via *RAD9* deletion (the SSDDCS overexpression strain contains a wild-type *RAD9* gene) may lead to a small amount of fork stalling and the completion of replication via the use of inefficient cryptic origins (Doksani et al., 2009). Alternatively, low levels of leaky SSDDCS expression under glucose repression could produce a partial hyperactivation phenotype, given that global firing efficiency is likely to be affected by even very small changes in SSDDCS protein levels. It is clear, however, that global differences are minimal (cf. chromosome-wide profiles for WT and SSDDCS + glucose in Figure S6) and that DNA replication can proceed robustly upon DNA ligase I depletion to allow replication profiling without the use of checkpoint mutants—further simplifying our approach.

If the effect of replication fork pause sites on the location of fork mergers is small relative to that of origin firing, then a change in the global distribution of origin-firing times should alter the location of FMZs genome wide. Origin usage changes dramatically upon SSDDCS overexpression (Figures 5C and 6C), providing a means to test our hypothesis directly. Upon origin hyperactivation, we observed a pronounced and highly reproducible global alteration in the location of FMZ midpoints when compared to either the wild-type or the uninduced SSDDCS strain (meta-analysis is shown in Figure 6E; anecdotal examples are highlighted in Figure 5). Assuming roughly constant fork speed throughout the genome (Sekedat et al., 2010), temporally coordinated origin firing should give rise to FMZs that lie halfway between origins. Indeed, FMZ midpoints for both TER- and non-TER FMZs are normally distributed about the interorigin midpoint (Figure 6F; data not shown). Consistent with the global early activation of normally late-firing origins, FMZ midpoints move substantially and are tightly clustered about the midpoint of the interorigin range after galactose induction (Figure 6F; $p < 0.0001$ for SSDDCS glucose versus either galactose replicate), confirming that origin-firing time is the dominant determinant of FMZ location. Interestingly, FMZs generally appear less sharp

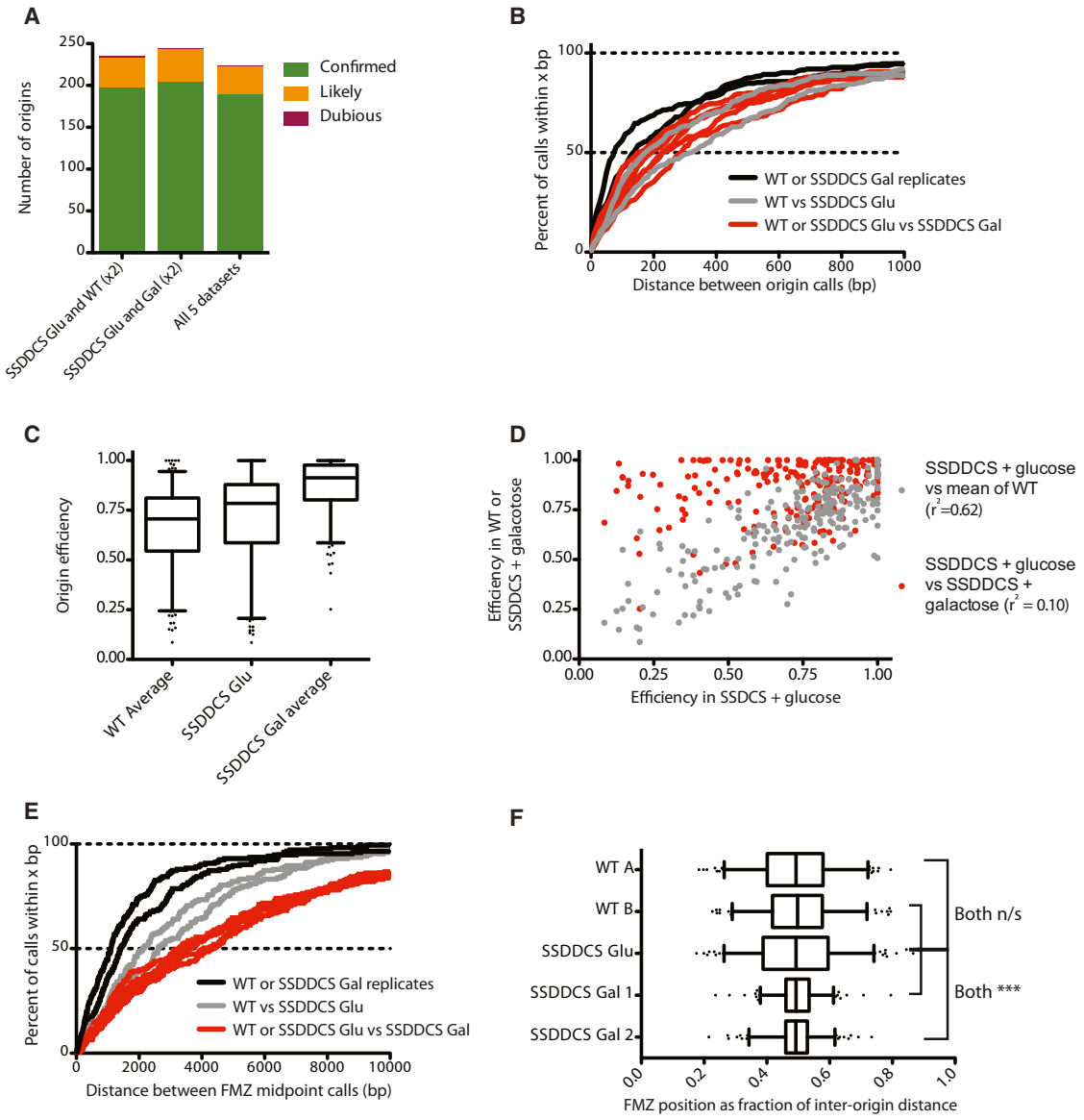

Figure 6. The Location of FMZs Is Predominantly Determined by Origin Firing

(A) As in Figure 2A, column graphs indicate the number of origins lying within ±2.5 kb of OriDB entries in data sets from wild-type or SSDDCS strains.

(B) The position of origin calls does not change substantially in different strain backgrounds or growth conditions. The distance between origin calls was calculated for all pairwise combinations of wild-type (two replicates) and SSDDCS (two replicates grown in galactose and one grown in glucose). The percentage of origin pairs called within x bp of one another in paired data sets is plotted on the y axis. Comparisons between replicate data sets are shown as black lines, those between wild-type and glucose-grown SSDDCS are shown as gray lines, and those between galactose-grown SSDDCS and any other sample are shown as red lines.

(C) The overexpression of SSDDCS substantially increases global origin efficiency.

(D) Origin efficiency in the SSDDCS strain correlates well with that in wild-type strains only in the absence of limiting factor overexpression. Efficiencies in wild-type and SSDDCS + galactose (mean value from two replicates in each case) are plotted in gray and red, respectively.

(E) Globally, FMZ midpoints move substantially upon origin hyperactivation. As in (B), the distance between FMZ calls was calculated for all pairwise combinations. Comparisons between replicate data sets are shown as black lines, those between wild-type and glucose-grown SSDDCS are shown as gray lines, and those between galactose-grown SSDDCS and any other sample are shown as red lines. Analysis in (E) and (F) was conducted using the 173 FMZs with a probability >0.1 in each of the five data sets.

(F) FMZ midpoints converge on interorigin midpoints upon origin hyperactivation. FMZ midpoints are normally distributed about the midpoint of the interorigin range. The variance of the distribution becomes significantly smaller upon SSDDCS overexpression (***, $p < 0.0001$ for each SSDDCS galactose data set versus the uninduced strain; t test).

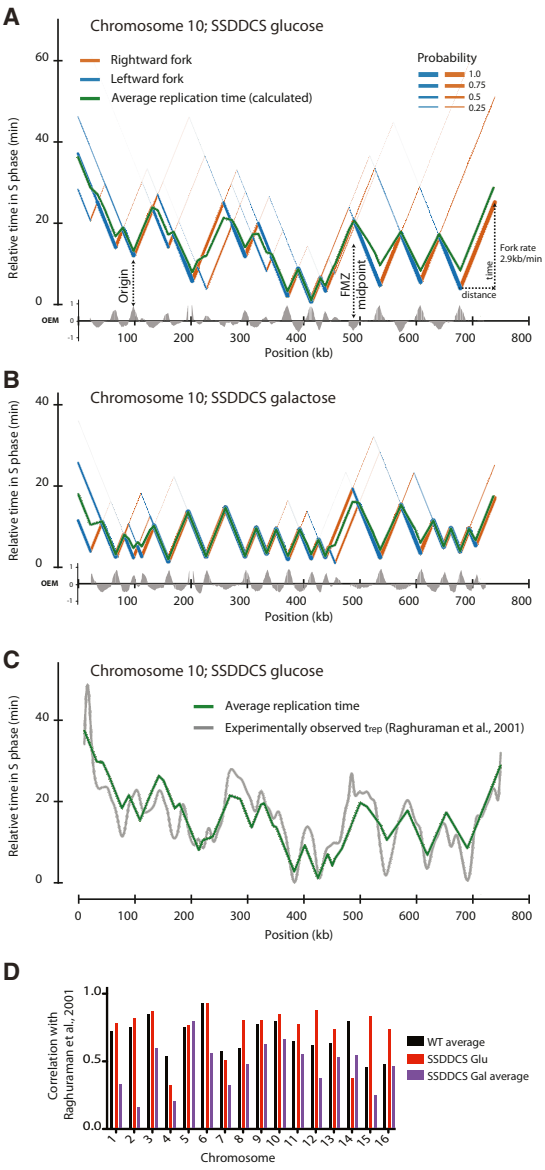

Figure 7. Reconstruction of Chromosome-Wide Replication Profiles from Okazaki Fragment Sequencing Data Obtained from Asynchronous Cultures

(A and B) Using the algorithm described in Experimental Procedures and assuming a constant fork speed of 2.9 kb/min (Raghuraman et al., 2001), we modeled the replication dynamics of each chromosome. Data is shown for chromosome 10 in the SSDDCS strain without (A) or with (B) galactose induction: leftward- and rightward-moving forks are colored blue and orange, respectively, with thickness proportional to probability. Replication forks diverge from origins (valleys) and progress at a constant rate until they converge at FMZs (peaks). The average replication time of each base pair is indicated by a green line. The earliest origin is set to t = 0 in each case. Other chromosomes are shown in Figure S7.

(C) Chromosome-wide comparison of our calculated average replication time and previously determined t_{rep} values (Raghuraman et al., 2001) for chromosome 10 in the uninduced SSDDCS strain.

(D) A column graph indicating the Pearson correlation between calculated timing profiles and previously determined t_{rep} values (Raghuraman et al., 2001) for wild-type (average of two data sets) and noninducing or inducing (average of two data sets) SSDDCS strains. Reported correlation (RMSD) is for the entire chromosome calculated at 2.5 kb intervals.

when SSDDCS are overexpressed (Figures 5A and 5B; see Figure S6); a plausible explanation for this observation is that, upon SSDDCS induction, adjacent origins fire at similar times to one another but with much less precision than in normal conditions, suggesting that origin hyperactivation may disrupt or reduce the effect of origin clustering and lead to a reduction in temporal coupling (see Discussion). An alternative—although not mutually exclusive—explanation for FMZ broadening is increased variability in replication fork speed as a result of the dNTP depletion observed when SSDDCS are overexpressed (Mantiero et al., 2011).

Reconstructing a Replication Map of the Yeast Genome

If the position of a given FMZ is indicative of the relative times at which two adjacent origins fire, we reasoned that we could use FMZ positions to determine whether one origin typically fires in advance of another; for example, if forks from adjacent origins travel the same distance, we assume their origins were activated at the same time. However, forks traveling different distances (i.e., the FMZ midpoint lies asymmetrically between the two origins) indicate that the origins fired at different times, the degree of asymmetry being indicative of the difference in time. Thus, if the replication program is governed by origin firing, we should be able to reconstruct a replication profile from our asynchronous data set.

We divided each chromosome into segments consisting of pairs of all possible firing origins. Initially, we considered only origin pairs in which each origin has a high firing probability (OEM > 0.5); then, each replication origin was assigned to its most probable neighbors (one each side). After establishing a baseline set of high-probability segments, the relative times of origin firing were determined. To do this, we divided the distance between flanking origins and their associated median merger position by a 2.9 kb/min linear rate of replication (Raghuraman et al., 2001). Then, baseline segments were adjusted relative to one another as units, ensuring that the same replication origin in adjacent segments was assigned the same time. Next, the lower probability segments were added to this baseline map.

The replication map for a representative chromosome (chromosome 10) from the uninduced SSDDCS strain is shown in Figure 7A, along with the calculated average replication time at each position. As expected, the same chromosome from galactose-grown cells shows a substantially "flatter" profile with earlier average replication times (Figure 7B), consistent with a global shift toward early origin activation upon the overexpression of limiting factors. The profiles generated from wild-type cells and from the SSDDCS strain when grown in glucose (Figure 7A; data not shown) contain distinct regions of early and late replication and reveal that chromosome replication is dominated by the activity of efficient origins firing within distinct time windows.

To confirm the validity of our contention that the replication program is governed by origin firing and executed by replication

forks that proceed at a uniform rate, we compared our calculated timing profiles with t_{rep} determined directly from density-transfer time course experiments (Raghuraman et al., 2001). Timing profiles from wild-type and uninduced SSDDCS cells closely resemble the experimental t_{rep} data (Figure 7C) and show correspondingly strong correlations (as judged by Pearson correlation coefficient; Figure 7D). As anticipated, SSDDCS induction generally reduces the strength of this correlation.

DISCUSSION

Our genome-wide analysis of both replication origin efficiency and fork termination allows us to analyze both replication origins and termination regions and, thus, to reconstruct the global replication profile of an asynchronous population of cells.

We observe that leading-strand initiation is biased toward the nucleosome-free region present at most *S. cerevisiae* replication origins. However, even highly efficient origins, or those whose locations we predict with high precision relative to the ACS, do not have a single sharp transition from the leading to the lagging strand. The heterogeneity of initiation sites may represent variability in the precise site of leading-strand initiation, in differences in the amount of leading strand displaced by Pol δ during synthesis of the first Okazaki fragment, or in both processes. Nevertheless, the clustering of initiation sites within the NFR indicates that the initial DNA unwinding event, as well as subsequent primer synthesis, is most likely to occur within this region. The structure of the NFR has been shown to be important for origin function, and mutations that either restrict or expand NFR width diminish origin activity (Lipford and Bell, 2001; Simpson, 1990). DNA-bound MCM2-7 double hexamers are known to slide along DNA (Remus et al., 2009); thus, the positioned nucleosomes on either side of the ~125 bp origin NFR may function to position a single MCM2-7 double hexamer (which occupies ~70 bp) at the initiation site.

Highly transcribed regions appear to have little overall impact on fork progression and termination; however, we note that these findings do not contradict the large body of data that show an apparent conflict between RNA transcription and DNA replication. Replication fork pauses appear to be simply infrequent and/or of such a sufficiently short duration that they have little effect on the population as a whole. In keeping with this, the orientation of transcription across the yeast genome is not biased with respect to replication fork direction, suggesting that, under normal circumstances, there is little selective pressure to avoid collisions between the transcription and replication machineries; moreover, our genome-wide data reveal that a large fraction of the genome can be replicated in either orientation (Figure S6). Uni- or bi-directional blocks affecting a substantial proportional of replication forks for an appreciable amount of time would be expected to generate discrete FMZs whose location and sharpness are insensitive to SSDDCS overexpression; our data are inconsistent with the existence of a significant number of such blocks. The stable pausing of a small proportion of replisomes, as observed by two-dimensional gel (Ivessa et al., 2003) or genome-wide ChIP (Sekedat et al., 2010), would not affect the median behavior of the population significantly, and, thus, cannot be detected by our method. For the special case of the rDNA repeat, we can infer the existence of the well-characterized RFB that specifically impedes rightward-moving forks from the strong strand bias of Okazaki fragments observed in this region, 85%–95% of hits arising from leftward-moving forks (see the rDNA panel of Figure S6). However, because of variation in sequence coverage, the repeat nature of this region, and the unknown efficiency of rDNA origins, our data do not allow de novo determination of RFB location.

Our data dispute a strictly deterministic replication program; because of its implicit inflexibility, absolute determinism seems unlikely to give rise to behavior that is sufficiently robust to be evolutionarily successful. The use of invariant patterns of origins and/or strong replication terminators genome-wide would almost certainly lead to an increase in incomplete replication and, thus, be subject to strong negative selective pressure. Instead, budding yeast appears to have adopted a somewhat flexible program in which active mechanisms exist to ensure that certain replication origins generally fire efficiently within distinct periods of time, whereas origins outside early-firing clusters remain competent to fire if they are not passively replicated before they have recruited all the factors necessary for firing. Ultimately, therefore, although the overall pattern of origin activation shows somewhat deterministic behavior, the only point at which active regulation is required is the initiation of the earliest firing origins.

Pre-RCs are assembled at essentially all replication origins in G1 phase, yet a subset will ultimately be used during the subsequent S phase. It has been suggested that origin usage and time of activation throughout S phase is governed by an origin's sensitivity to initiation factors whose concentration increases during S phase. Thus, an early-firing origin will be more sensitive and fire in a narrower time period than a later-firing origin, which will fire over a much larger time window. The fact that most origins fire early and efficiently when Sld2, Sld3, Dbf4, Dpb11, Cdc45, and Sld7 are overexpressed is in apparent agreement with this model. However, multiple initiator models require that later-firing origins fire stochastically and inefficiently (Yang et al., 2010). Although we do find that overall origin efficiency decreases later in S phase (Figure 2E), the trend is relatively weak, and there are numerous examples of late origins that are very efficient and appear to fire within a discrete time period.

Our data are consistent with a model in which most replication origins are competent to fire if bound by factors required for initiation; but the temporal order of origin firing is at least partially conferred by the spatial distribution of replication origins and limiting factors within the three-dimensional architecture of the nucleus. The replication program may be initiated by the physical association of select replication origins into replication factories; high local concentration of factors being necessary for replication initiation (Cook, 1999; Meister et al., 2006). Origin participation within an early cluster may be promoted by proximity to centromeres (Pohl et al., 2012) or association with forkhead (Knott et al., 2012) but may be inhibited by "repressive" chromatin structure or proximity to telomeres. Indeed, early origins and centromeres are known to cluster (Duan et al., 2010), and recent reports show that FKH-activated origins may be in close proximity to one another (Knott et al., 2012). Origin firing would be initially limited to these clusters—ensuring that, although

most origins within the cell are competent to fire, only a select few do so in the earliest stages of S phase. Inspection of replication profiles reveals that origin activation time tends to progressively increase as a function of one-dimensional distance from an early-firing origin (higher-order V shapes can be observed in timing profiles in Figure 7A; timing profiles can be observed in Figure S7), suggesting that the positioning of an origin along chromosomes is an important determinant of origin timing. Indeed, initiation time has been shown to depend upon chromosomal context and is not an inherent property of replication origins (Ferguson and Fangman, 1992). Thus, an early-firing origin could stimulate the firing of nearby origins, reminiscent of cascade effects observed in human cells (Guilbaud et al., 2011). Such a pattern of activation may be governed by the subdiffusive motion of factors limiting for activation that are recycled from recently fired replication origins (or moving replication forks) to proximal origins (Sporbert et al., 2002) and/or by the motion of proximal origins to active replication clusters (Gauthier and Bechhoefer, 2009). Overexpression of SSDDCS could override the normal replication program in two ways: first, a greater number of origins could participate in early clusters; second, the activation of origins outside these clusters would be less reliant on the recycling of initiation factors and, thus, no longer spatially restricted to the vicinity of recently fired origins.

The high competence of *S. cerevisiae* origins and the general absence of specific, *cis*-acting termination sequences indicate that the replication program is dominated by origin firing. However, few origins are 100% efficient, implying that almost all will occasionally fail to fire prior to being replicated by an incoming fork. The temporal firing window for the majority of origins appears to be relatively broad—an assertion supported both by the widespread passive replication of origins and by the observation that fork mergers occur over a large proportion of the interorigin distance for most origin pairs. Instances (e.g., at many centromeres or in the right arm of chromosome 10, as shown in Figure 1) where sharp mergers exist halfway between a pair of origins are indicative of tight temporal coupling between the pair and are not necessarily indicative of a narrowing of either origin's firing window relative to the population as a whole. Although they are not the norm, we observe numerous such isolated sharp merger regions. Coupled origins are generally the earliest to fire within each chromosome and are sensitive to Fkh1 and Fkh2 depletion (Knott et al., 2012; data not shown), consistent with physical juxtaposition. Our approach can detect apparent coordination between adjacent origins but not putative longer-range and/or interchromosomal coupling, almost certainly leading to an underestimate of the extent of this phenomenon. The biological significance of such tight coupling remains unclear: it may reflect a mechanism to ensure that certain regions are replicated early or a passive means to ensure that termination occurs in a distinct location. However, coordinated firing also imposes predictable asymmetry in replication patterns, giving rise to daughters with distinct chromosomal regions replicated exclusively on the leading or lagging strand. If chromatin is assembled differentially on the two strands, then such replication patterns could facilitate the asymmetric propagation of epigenetic information.

EXPERIMENTAL PROCEDURES

Okazaki fragments were purified and sequenced as described previously (Smith and Whitehouse, 2012), except Ion Torrent 318 sequencing was used for the SSDDCS strain. Genotypes of all strains used in this study are listed in Table S1.

Quadrants and Origin Efficiency Metric

For each position in the genome, we summed the cumulative fragment count within 10 kb to the left and right; this was performed on each strand. We define four quadrants as follows: Watson strand left (WL), Watson strand right (WR), Crick strand left (CL), and Crick strand right (CR) at each base pair on each chromosome (see Figure 1). To account for differences in read depth due to base composition, etc., and because the total amount of replication on both strands should be constant across the genome, we normalized to the total signal on either side of the sliding window—i.e., (WL + CL) or (WR + CR).

To find origins and FMZs, we converted quadrant data into an OEM defined by Equation 1 and dependent on Equations 2 and 3:

$$OEM = WL_n - WR_n \quad \text{(Equation 1)}$$

$$WL_n = \frac{WL}{(WL + CL)} \quad \text{(Equation 2)}$$

$$WR_n = \frac{WR}{(WR + CR)} \quad \text{(Equation 3)}$$

WL_n and WR_n range from one to zero. In the case of an ideal 100% efficient point origin, from which two forks invariably diverge, $WL_n = 1$ and $WR_n = 0$; therefore, from Equation 1, OEM = 1. In the analogous case of an ideal point merger between 100% efficient origins, at which two forks invariably converge, $WL_n = 0$, $WR_n = 1$, and OEM = −1.

OEM provides an averaged readout of the heterogeneous behavior within a population. For an origin that fires in 50% of cells but is passively replicated by an incoming fork in the remaining 50%, the observed OEM will comprise a triangular signal from the 50% of cells in which the origin fires and a flat signal from the 50% in which it does not, generating a weighted average of 0.5. Theoretical model origins of varying efficiencies are shown schematically in Figure S1, and real origins are shown in Figures 1C and 1D.

Origin Positions and Efficiencies

Origin positions were calculated using a three-point method whose algorithm identifies localized maxima in OEM within 10 Kb ranges.

Deconvolution

The presence of FMZs in a subpopulation of cells near origins fired in another subpopulation convolutes the OEM signal for both origins and FMZs, given that the observed OEM is a composite of the subpopulations (Figure S2). Because origins in *S. cerevisiae* are small and discrete whereas FMZs are large and diffuse, we can generally deconvolute our signal by assuming origins to be point entities and modeling ideal origin behavior about this point. A signal arising because of origins is deconvoluted first and subsequently used to correct OEM at FMZs, as described below. Deconvolution of theoretical model data is shown in Figure S2; examples of the process using real data are shown in Figure S3.

Because OEM is calculated from a sliding window, it will be triangular in shape at origins and span 20 kb. If an origin's OEM signal is convoluted by an underlying FMZ signal, then its OEM is systematically reduced, and some of or the entire signal will lie below zero, causing the measured height of the OEM peak to provide an underestimate of true origin efficiency.

To correct for underestimated origin efficiencies, we measured the gradient of OEM at positions 5 kb and 10 kb from either side of the origin and used the numerical mean of the larger two of these four values to fit an idealized triangle to each origin OEM. Fitted triangles were then used to calculate a deconvoluted origin (OEM_{origin}) which forces the OEM signal to be above zero throughout the 20 Kb range spanned by the triangle. All origin efficiencies

reported in this work represent OEM_{origin}, and only origins in which OEM was originally > 0.05 were considered in our analysis.

Finally, to remove origin signal from OEM at FMZs, we subtracted the OEM_{origin} calculated above from the original OEM to give OEM_{FMZ} for each FMZ; analogous to those reported for origins, reported values of OEM at FMZs represent OEM_{FMZ}. Additional experimental details can be found in Supplemental Experimental Procedures.

ACCESSION NUMBERS

Sequencing data and timing analyses have been deposited in the Gene Expression Omnibus under accession numbers 33786 and 40696.

SUPPLEMENTAL INFORMATION

Supplemental Information contains Supplemental Experimental Procedures, seven figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2013.03.004.

ACKNOWLEDGMENTS

We thank Philip Zegerman (Gurdon Institute) for the SSDDCS strain and Ken Marians, Dirk Remus, Toshi Tsukiyama (Fred Hutchinson Cancer Research Center), and members of the Molecular Biology Program and Whitehouse Lab for discussions and comments on the manuscript. This work was supported by the National Institutes of Health (grant R01 GM102253) and an Alfred Bressler Scholars Endowment Award to I.W. D.J.S. is a Howard Hughes Medical Institute fellow of the Damon Runyon Cancer Research Foundation (DRG-#2046-10).

Received: August 31, 2012
Revised: November 20, 2012
Accepted: March 1, 2013
Published: April 4, 2013

REFERENCES

Brewer, B.J., and Fangman, W.L. (1988). A replication fork barrier at the 3′ end of yeast ribosomal RNA genes. Cell *55*, 637–643.

Cook, P.R. (1999). The organization of replication and transcription. Science *284*, 1790–1795.

Czajkowsky, D.M., Liu, J., Hamlin, J.L., and Shao, Z. (2008). DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. J. Mol. Biol. *375*, 12–19.

de Moura, A.P., Retkute, R., Hawkins, M., and Nieduszynski, C.A. (2010). Mathematical modelling of whole chromosome replication. Nucleic Acids Res. *38*, 5623–5633.

Deshpande, A.M., and Newlon, C.S. (1996). DNA replication fork pause sites dependent on transcription. Science *272*, 1030–1033.

Doksani, Y., Bermejo, R., Fiorani, S., Haber, J.E., and Foiani, M. (2009). Replicon dynamics, dormant origin firing, and terminal fork integrity after double-strand break formation. Cell *137*, 247–258.

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., and Noble, W.S. (2010). A three-dimensional model of the yeast genome. Nature *465*, 363–367.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. (2010). Conserved nucleosome positioning defines replication origins. Genes Dev. *24*, 748–753.

Fachinetti, D., Bermejo, R., Cocito, A., Minardi, S., Katou, Y., Kanoh, Y., Shirahige, K., Azvolinsky, A., Zakian, V.A., and Foiani, M. (2010). Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. Mol. Cell *39*, 595–605.

Feng, W., Collingwood, D., Boeck, M.E., Fox, L.A., Alvino, G.M., Fangman, W.L., Raghuraman, M.K., and Brewer, B.J. (2006). Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. Nat. Cell Biol. *8*, 148–155.

Ferguson, B.M., and Fangman, W.L. (1992). A position effect on the time of replication origin activation in yeast. Cell *68*, 333–339.

Friedman, K.L., Brewer, B.J., and Fangman, W.L. (1997). Replication profile of Saccharomyces cerevisiae chromosome VI. Genes Cells *2*, 667–678.

Gauthier, M.G., and Bechhoefer, J. (2009). Control of DNA replication by anomalous reaction-diffusion kinetics. Phys. Rev. Lett. *102*, 158104.

Greenfeder, S.A., and Newlon, C.S. (1992). Replication forks pause at yeast centromeres. Mol. Cell. Biol. *12*, 4056–4066.

Guilbaud, G., Rappailles, A., Baker, A., Chen, C.L., Arneodo, A., Goldar, A., d'Aubenton-Carafa, Y., Thermes, C., Audit, B., and Hyrien, O. (2011). Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. PLoS Comput. Biol. *7*, e1002322.

Hill, T.M., and Marians, K.J. (1990). Escherichia coli Tus protein acts to arrest the progression of DNA replication forks in vitro. Proc. Natl. Acad. Sci. USA *87*, 2481–2485.

Ivessa, A.S., Lenzmeier, B.A., Bessler, J.B., Goudsouzian, L.K., Schnakenberg, S.L., and Zakian, V.A. (2003). The Saccharomyces cerevisiae helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. Mol. Cell *12*, 1525–1536.

Kearsey, S. (1984). Structural requirements for the function of a yeast chromosomal replicator. Cell *37*, 299–307.

Knott, S.R., Peace, J.M., Ostrow, A.Z., Gan, Y., Rex, A.E., Viggiani, C.J., Tavaré, S., and Aparicio, O.M. (2012). Forkhead transcription factors establish origin timing and long-range clustering in S. cerevisiae. Cell *148*, 99–111.

Kobayashi, T., and Horiuchi, T. (1996). A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities. Genes Cells *1*, 465–474.

Lipford, J.R., and Bell, S.P. (2001). Nucleosomes positioned by ORC facilitate the initiation of DNA replication. Mol. Cell *7*, 21–30.

Mantiero, D., Mackenzie, A., Donaldson, A., and Zegerman, P. (2011). Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. EMBO J. *30*, 4805–4814.

Marahrens, Y., and Stillman, B. (1992). A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science *255*, 817–823.

Meister, P., Taddei, A., and Gasser, S.M. (2006). In and out of the replication factory. Cell *125*, 1233–1235.

Monti, K.L. (1995). Folded Empirical Distribution Function Curves-Mountain Plots. Am. Stat. *49*, 342–345.

Nieduszynski, C.A., Hiraga, S., Ak, P., Benham, C.J., and Donaldson, A.D. (2007). OriDB: a DNA replication origin database. Nucleic Acids Res. *35*(Database issue), D40–D46.

Pohl, T.J., Brewer, B.J., and Raghuraman, M.K. (2012). Functional centromeres determine the activation time of pericentric origins of DNA replication in Saccharomyces cerevisiae. PLoS Genet. *8*, e1002677.

Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., and Fangman, W.L. (2001). Replication dynamics of the yeast genome. Science *294*, 115–121.

Rao, H., Marahrens, Y., and Stillman, B. (1994). Functional conservation of multiple elements in yeast chromosomal replicators. Mol. Cell. Biol. *14*, 7643–7651.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell *139*, 719–730.

Sekedat, M.D., Fenyö, D., Rogers, R.S., Tackett, A.J., Aitchison, J.D., and Chait, B.T. (2010). GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. Mol. Syst. Biol. *6*, 353.

Sheu, Y.J., and Stillman, B. (2006). Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. Mol. Cell *24*, 101–113.

Simpson, R.T. (1990). Nucleosome positioning can affect the function of a cis-acting DNA element in vivo. Nature *343*, 387–389.

Smith, D.J., and Whitehouse, I. (2012). Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature *483*, 434–438.

Sporbert, A., Gahl, A., Ankerhold, R., Leonhardt, H., and Cardoso, M.C. (2002). DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters. Mol. Cell *10*, 1355–1365.

Stith, C.M., Sterling, J., Resnick, M.A., Gordenin, D.A., and Burgers, P.M. (2008). Flexibility of eukaryotic Okazaki fragment maturation through regulated strand displacement synthesis. J. Biol. Chem. *283*, 34129–34140.

Xu, W., Aparicio, J.G., Aparicio, O.M., and Tavaré, S. (2006). Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. BMC Genomics *7*, 276.

Yabuki, N., Terashima, H., and Kitada, K. (2002). Mapping of early firing origins on a replication profile of budding yeast. Genes Cells *7*, 781–789.

Yamashita, M., Hori, Y., Shinomiya, T., Obuse, C., Tsurimoto, T., Yoshikawa, H., and Shirahige, K. (1997). The efficiency and timing of initiation of replication of multiple replicons of Saccharomyces cerevisiae chromosome VI. Genes Cells *2*, 655–665.

Yang, S.C., Rhind, N., and Bechhoefer, J. (2010). Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. Mol. Syst. Biol. *6*, 404.