SUPPLEMENTARY INFORMATION

Normalization

Our experiments utilize Agilent microarrays that contain 10,807 oligonucleotide probes, covering 6,256 \textit{S. cerevisiae} ORFs with extensive replicates and Eurogentec microarrays that contain 11,520 PCR amplified probes covering 4,976 \textit{S. pombe} ORFs with replicates. In our microarray experiments, we sought a quantitative measurement of the ratio of single stranded DNA between a time point \( N \) (denoted TP\(_N\) ) and a reference time point (denoted TP\(_0\), for time point zero). After collection of DNA for TP\(_N\) and TP\(_0\) and isolation of genomic DNA for time points \( N \) and 0, we differentially labeled DNA complementary to the single stranded DNA, using Cy-3 and Cy-5 dyes, respectively. The labeled samples were hybridized to the microarray. Using GenePix 4.0 software (Axon), we converted the Cy-5 and Cy-3 fluorescence intensities from TIFF files into numerical intensity data for each labeled sample and extracted the background subtracted median feature pixel intensity numbers for each spot on the array. Averaging over duplicated features, we obtained TP\(_N\) Cy-3 signals (denoted \( a_i^* \)) and TP\(_0\) Cy-5 signals (denoted \( b_i^* \)), where \( i \) indexes the set of all yeast ORFs. We arrived at

Equation (1): observed ratio of ssDNA at ORF\(_i\) = \( a_i^*/b_i^* \), \( i \) indexing the ORFs.

This calculation led to the chromosomal observed single stranded ratio profile by plotting the data points \( (c_j, \ a_j^*/b_j^* ) \), where \( j \) runs over the ORFs for a given chromosome. Although informative, this profile suffers from the fact it is not
quantitative: the TIFF files produced by core facilities arbitrarily adjust the gain in both Cy channels to achieve a roughly equal balance of total signal.

To get around this problem, the data were normalized and made quantitative by applying the scheme laid out in Collingwood et al. (manuscript in preparation). The normalization uses an external measurement of ssDNA content in S vs. G1 (slot blot) to correct one of the Cy channels, so as to restore to the microarray data the S/G1 ratio seen in the external calibration. The key result asserts the existence of a computable constant g, such that

Equation (2): actual ratio of ssDNA at ORF$_i$ = g $a_i^*/b_i^*$.

The constant g = n/m, where

n = (total TP$_N$ ss DNA)/(total TP$_0$ ss DNA),
m = (total Cy-3 signal)/(total Cy-5 signal).

The constant m is directly computed from the array data. Because we isolated equal amounts of TP$_N$ and TP$_0$ DNA,

n = (total TP$_N$ ssDNA)/(total TP$_0$ ssDNA)

= [(total TP$_N$ ssDNA)/(total TP$_N$ DNA)]/ [(total TP$_0$ ssDNA)/(total TP$_N$ DNA)]

= [(total TP$_N$ ssDNA)/total TP$_N$ DNA]/ [(total TP$_0$ ssDNA)/total TP$_0$ DNA]

= (% TP$_N$ ssDNA)/(% TP$_0$ ssDNA).
We were able to experimentally compute this ratio of percentages, thus allowing us to compute the normalization constant \( g \). We then obtained the *actual single stranded ratio profile* by plotting the data points \( (c_j, g a_j^*/b_j^*) \), where \( j \) runs over the ORFs for a given chromosome.

**Smoothing**

We transformed the raw data of ssDNA ratio by using the Fourier convolution smoothing technique previously introduced\(^5\) to obtain a smoothed profile. The smoothed profiles offered the advantage of prominently identifying local extrema in the data. In our application of smoothing, a window of 4 kb was specified and a moving average with this window size was constructed. We used this moving average as a target and selected the closest Fourier smoothing among a large family of smoothings. See the supporting online text to Raghuraman *et al.*\(^5\) for full details of this procedure.

**Extrema detection**

Given any discrete dataset of points \((x_i, y_i)\), we detected local extrema as follows: First, for each data point, we calculated the numbers

\[
S^i_L = \frac{(y_i - y_{i-1})}{(x_i - x_{i-1})} \quad \text{and} \quad S^i_R = \frac{(y_{i+1} - y_i)}{(x_{i+1} - x_i)}.
\]

If \( S^i_L > 0 \) and \( S^i_R < 0 \), then we flagged the point \((x_i, y_i)\) as a local maximum, whereas if \( S^i_L < 0 \) and \( S^i_R > 0 \), then we flagged the point \((x_i, y_i)\) as a local minimum.
Identification of significant ssDNA peaks

In order to calculate standard deviation in the background level of ssDNA labeling in all timed samples, we first identified and flagged those data points with values above the median value in each data set, thus removing prominent peaks from our estimation of background variation. We then removed those data points that were flagged in any timed sample data from all timed sample data in all further calculations. Averaging the three median values (Own Median), we obtained the Average Median and normalized the remaining data points by multiplying each value with the constant M (M=Average Median/Own Median). Standard deviation was calculated as the square root of variance (variance = \(\sum_i \{1/3[(X_{i,1} - X_{i,2})^2+(X_{i,1} - X_{i,3})^2+(X_{i,2} - X_{i,3})^2]\} / \text{total number of data points}\), where \(i\) indexes the data points and the subscripts 1, 2, and 3 index time points 1, 2, and 3 hr).

We present a sample calculation (numbers have been rounded here to two decimal places for clarity):

Calculate median for \textit{rad53} at 1, 2 and 3 hour

Median\_rad53\_1hr = 1.90
Median\_rad53\_2hr = 2.28
Median\_rad53\_3hr = 2.51

Calculate average of the Medians above

Average of Median = 2.23
Calculate the normalization factor (constant $M$) for $rad53$ at 1, 2 and 3 hr

$M_{rad53\_1hr} = 1.90/2.23 = 0.85$

$M_{rad53\_2hr} = 2.28/2.23 = 1.02$

$M_{rad53\_3hr} = 2.51/2.23 = 1.13$

Normalize each data set by multiplying the smoothed ratio of ssDNA (e.g., $rad\_xhr$ in the spreadsheet below) with the normalization factor constant $M$ and arrive at a normalized value (e.g., $rad\_xhr\_nm$ in the spreadsheet below). Shown below is a spreadsheet of calculations for a portion of chromosome 1.

| chr coord | rad_1hr | rad_1hr_nm | rad_2hr | rad_2hr_nm | rad_3hr | rad_3hr_nm |
|-----------|---------|------------|---------|------------|---------|------------|
| 1         | 48      | 1.85       | 2.16    | 2.02       | 1.98    | 2.27       | 2.02       |
| 1         | 49      | 1.53       | 1.79    | 1.76       | 1.73    | 2.00       | 1.77       |
| 1         | 50      | 1.44       | 1.69    | 1.66       | 1.62    | 2.01       | 1.78       |
| 1         | 51      | 1.48       | 1.73    | 1.63       | 1.60    | 2.17       | 1.93       |
| 1         | 52      | 1.57       | 1.84    | 1.65       | 1.62    | 2.36       | 2.10       |
| 1         | 53      | 1.66       | 1.94    | 1.68       | 1.65    | 2.51       | 2.23       |
| 1         | 58      | 1.73       | 2.03    | 1.74       | 1.70    | 2.45       | 2.18       |
| 1         | 59      | 1.64       | 1.93    | 1.65       | 1.62    | 2.28       | 2.02       |
| 1         | 60      | 1.57       | 1.84    | 1.59       | 1.56    | 2.13       | 1.89       |
| 1         | 61      | 1.54       | 1.80    | 1.60       | 1.57    | 2.02       | 1.79       |
| 1         | 62      | 1.56       | 1.83    | 1.67       | 1.64    | 1.97       | 1.74       |
| 1         | 63      | 1.65       | 1.94    | 1.79       | 1.76    | 2.02       | 1.80       |
| 1         | 64      | 1.77       | 2.08    | 1.92       | 1.89    | 2.17       | 1.92       |
| 1         | 65      | 1.88       | 2.20    | 2.00       | 1.96    | 2.27       | 2.02       |
| 1         | 80      | 1.83       | 2.15    | 1.98       | 1.94    | 2.39       | 2.12       |

Calculate variance and standard deviation
Variance = \Sigma_i \left\{ \frac{1}{3} [ (X_{i,1} - X_{i,2})^2 + (X_{i,1} - X_{i,3})^2 + (X_{i,2} - X_{i,3})^2] \right\} / \text{total number of data points},
where \(i\) indexes the data points and the subscripts 1, 2, and 3 index time points 1, 2, and 3 hr.

Standard deviation = square root of variance

For example, at the first coordinate (48 kb on chr 1), calculate
\[
\frac{1}{3} [(2.1621-1.9770)^2+(2.1622-2.0161)^2+(1.9770-2.0161)^2] = 0.0571.
\]
Repeat this calculation for all the coordinates in the genome and average all of them and arrive at
Variance = 0.0558; Standard deviation = 0.2362.

We next identified the local maxima and minima in each timed sample and calculated the difference between every local maximum and its two flanking minima (note that for some of the telomeric points, not every local maximum is flanked by two local minima). Those local maxima with values that are above 3 standard deviations from both its flanking local minima were considered significant ssDNA peaks. Those telomeric local maxima that only have a single flanking local minimum but are above 3 standard deviation of the said local minimum are also considered significant ssDNA peaks.
**Figure S1.** Overlay of smoothed ssDNA profiles for *S. cerevisiae*. The time course for both WT and *rad53* strains are shown in series of increasing color intensity. WT cell profiles at 1, 2 and 3-hour post release are shown as light purple, magenta and dark purple curves, respectively; *rad53* cell profiles at 1, 2 and 3-hour post release are shown as yellow, orange and red curves, respectively. WT cell profiles are plotted on the Y1 (left) axes and *rad53* cell profiles are plotted on the Y2 (right) axes. Positions of Pro-ARSs\(^{13}\) are shown as green diamonds. Positions of clustered origins (those that appear in at least two of the timed samples) from ssDNA profiles of *rad53* cells are shown as filled blue circles. Positions of singleton origins (those that appear in only one of the three timed sample) are shown as filled red circles.
Figure S2. Comparison between *S. cerevisiae* Rad53-unchecked origins and early origins identified by copy number change detection at 90 minutes in HU\textsuperscript{15}. The smoothed data of ratios of ssDNA (S/G1) for WT cells at 1 hour post release (green curves) are overlaid with changes in copy number of genomic DNA in the presence of HU (orange shaded curves). The ratios of ssDNA are plotted on the Y1 (left) axes and the copy number change is plotted on the Y2 (right) axes. The positions of all the clustered ssDNA peaks in rad53 cells representing all origins are indicated by filled blue circles; the peaks that occur in a single timed sample are indicated by filled red circles. The ssDNA peaks that meet the statistical criteria for local maxima in the WT profile are indicated by open diamonds: those that match the clustered ssDNA peaks in *rad53* cells are shown in blue and those that match the singleton ssDNA peaks in *rad53* cells are shown in red. The four ssDNA peaks that only appear in WT cells but are not identified as clustered or singleton ssDNA peaks in *rad53* cells are shown as solid black diamonds.
**Figure S3.** Overlay of ssDNA profiles for *S. pombe* WT (purple) and Δcds1 (orange) cells. WT cell profiles are plotted on the Y1 (left) axes and Δcds1 cell profiles are plotted on the Y2 (right) axes. Positions of significant ssDNA peaks in WT and Δcds1 cells are indicated by purple and orange filled circles respectively. Positions of AT-rich islands reported by Segurado *et al.*\textsuperscript{19} are shown as green filled circles. Positions of previously mapped origins from Segurado *et al.*\textsuperscript{19} and references therein are shown as red filled circles. Those previously mapped origins that have also been identified as significant ssDNA peaks in Δcds1 cells are labeled above the graphs. The two origins that eluded our analysis are boxed.