nature structural & molecular biology

Tracking replication enzymology *in vivo* by genome-wide mapping of ribonucleotide incorporation

Anders R Clausen[1], Scott A Lujan[1], Adam B Burkholder[2], Clinton D Orebaugh[1], Jessica S Williams[1], Maryam F Clausen[3], Ewa P Malc[3], Piotr A Mieczkowski[3], David C Fargo[2], Duncan J Smith[4] & Thomas A Kunkel[1]

Ribonucleotides are frequently incorporated into DNA during replication in eukaryotes. Here we map genome-wide distribution of these ribonucleotides as markers of replication enzymology in budding yeast, using a new 5′ DNA end–mapping method, hydrolytic end sequencing (HydEn-seq). HydEn-seq of DNA from ribonucleotide excision repair–deficient strains reveals replicase- and strand-specific patterns of ribonucleotides in the nuclear genome. These patterns support the roles of DNA polymerases α and δ in lagging-strand replication and of DNA polymerase ε in leading-strand replication. They identify replication origins, termination zones and variations in ribonucleotide incorporation frequency across the genome that exceed three orders of magnitude. HydEn-seq also reveals strand-specific 5′ DNA ends at mitochondrial replication origins, thus suggesting unidirectional replication of a circular genome. Given the conservation of enzymes that incorporate and process ribonucleotides in DNA, HydEn-seq can be used to track replication enzymology in other organisms.

Among the many eukaryotic DNA polymerases (Pols), for example, the 17 in humans and eight in budding yeast, three replicate the bulk of the nuclear genome[1,2]. Synthesis at replication origins is initiated when an RNA primase synthesizes an RNA primer that is extended by limited DNA synthesis by Pol α (ref. 3). Pol ε has been proposed to then catalyze the majority of leading-strand replication[4–6], in a largely continuous manner. In contrast, the nascent lagging strand is synthesized as a series of ~180-nucleotide Okazaki fragments that are initiated by RNA primase, and this is followed by limited synthesis by Pol α, extensive synthesis catalyzed by Pol δ (refs. 6–8) and subsequent maturation of Okazaki fragments into a continuous nascent lagging strand[9]. The exact locations of polymerase switching during leading- and lagging-strand replication are under investigation[10,11] but remain uncertain. Equally uncertain is polymerase use after replication forks encounter difficult circumstances that may require switching to a different replicase or a more specialized DNA polymerase, for example, to copy unusual DNA sequences or to bypass lesions[12,13]. Replication enzymology differs for the mitochondrial genome, in which both DNA strands are replicated by the same replicase, Pol γ, by mechanisms that also remain uncertain[14–16].

We have been using mutator alleles of yeast Pols ε, α and δ to infer their roles in nuclear DNA replication *in vivo*. These mutator alleles, *pol2-M644G* (Pol ε), *pol1-L868M* (Pol α) and *pol3-L612M* (Pol δ), generate single-base replication errors at higher rates than in their wild-type parents. In the absence of mismatch repair (MMR), these errors remain in the genome and mark where each replicase synthesized DNA during replication. The results (ref. 6 and references therein) imply that in unstressed yeast cells, Pol ε is the primary leading-strand replicase, and Pols α and δ are primarily responsible for lagging-strand replication. However, the resolution of this approach for tracking replication enzymology *in vivo* is limited by the high fidelity of replication. For example, the average genome-wide replication error rates of the mutator replicases are 1×10^{-7} to 2×10^{-7} (ref. 6), such that single-base replication errors are low-density markers of replication enzymology.

In the present study, we set out to track replication enzymology *in vivo* at much higher resolution by using ribonucleotides rather than mutations. This approach takes advantage of several facts. The presence of an oxygen atom on the 2′ position of a ribose increases the sensitivity of the phosphodiester bond in nucleic acids to alkaline hydrolysis by five orders of magnitude. The active sites of Pols α, δ and ε can be engineered to increase the probability of ribonucleotide incorporation into DNA, to frequencies as high as 10^{-2} to 10^{-3} (J.S.W., A.R.C, L. Marjavaara, A.B. Clark, S.A.L. *et al.*, unpublished data). Disabling ribonucleotide excision repair (RER) prevents removal of ribonucleotides from both the nascent leading strand and the nascent lagging strand (J.S.W., A.R.C, L. Marjavaara, A.B. Clark, S.A.L. *et al.*, unpublished data, and refs. 17,18). RER-defective yeast cells are viable, including those encoding replicases that are promiscuous for ribonucleotide incorporation. These facts led us to propose[19] that ribonucleotides can be used as high-density markers of DNA-polymerization reactions *in vivo*. Here we demonstrate that this is indeed the case, using a newly developed method to map ribonucleotides in the yeast genome at single-nucleotide resolution.

[1]Genome Integrity & Structural Biology Laboratory, National Institute of Environmental Health Sciences, National Institute of Health (NIH), Research Triangle Park, North Carolina, USA. [2]Integrative Bioinformatics, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, North Carolina, USA. [3]Department of Genetics, High Throughput Sequencing Facility, University of North Carolina, Chapel Hill, North Carolina, USA. [4]Center for Genomics and Systems Biology, Department of Biology, New York University, New York, New York, USA. Correspondence should be addressed to T.A.K. (kunkel@niehs.nih.gov).

Received 10 November 2014; accepted 18 December 2014; published online 26 January 2015; doi:10.1038/nsmb.2957

Initial results support the strand assignments for the nuclear replicases, confirm nuclear replication origins and identify new origins, reveal the locations of replication termination zones, quantify ribonucleotide incorporation for each of the four bases, establish that the distribution of ribonucleotides across the genome is nonuniform and provide new information that is likely to be relevant to mitochondrial DNA replication.

RESULTS

Genome-wide mapping of ribonucleotides in DNA by HydEn-seq

We used a new genome-wide mapping method, which we call HydEn-seq (**Fig. 1a** and **Supplementary Table 1**), to map ribonucleotides in five pairs of RER-deficient (*rnh201*Δ) versus RER-proficient (*RNH201*) yeast strains (**Supplementary Table 2**). One pair encodes wild-type Pols α, δ and ε. A second pair encodes *pol2-M644L*, a Pol ε variant that incorporates fewer ribonucleotides than does a wild-type strain[17]. A third pair encodes *pol2-M644G*, a Pol ε variant that is promiscuous for ribonucleotide incorporation[17,20]. A fourth pair encodes a *pol3-L612G* variant in which Leu612 in the Pol δ active site is replaced with glycine, on the basis of the prediction that, like the analogous *pol2-M644G* (Pol ε) variant[17], the *pol3-L612G* variant would be even more promiscuous for ribonucleotide incorporation than our previously studied *pol3-L612M* allele[21]. The fifth pair encodes a *pol1-Y869A* variant with alanine substituted for the 'steric gate' tyrosine in the Pol α active site that normally prevents ribonucleotide incorporation[22]. We used this allele to increase the frequency of ribonucleotide incorporation by Pol α over that observed in our previously studied *pol1-L868M* variant (J.S.W., A.R.C., L. Marjavaara, A.B. Clark, S.A.L. *et al.*, unpublished data).

Alkaline hydrolysis of genomic DNA and subsequent electrophoresis in an alkaline agarose gel revealed that the genomes of all five *rnh201*Δ-mutant strains contain more alkali-sensitive sites than do their *RNH201*$^+$ parents (**Fig. 1b**). Importantly, the genomes of the double-mutant *pol1-Y869A rnh201*Δ, *pol2-M644G rnh201*Δ and *pol3-L612G rnh201*Δ strains contain many more alkali-sensitive sites than do the strains with either single mutation alone (**Fig. 1b–d**), such that most of the 5′ DNA ends in these strains result from alkaline hydrolysis of ribonucleotides incorporated during replication by the variant derivatives of Pols α, δ or ε. This contrasts with the *pol2-M644L rnh201*Δ mutant strain, which contains fewer ribonucleotides than the other *rnh201*Δ strains with variant replicases.

We mapped the locations of the 5′ DNA ends in the genomes of these strains by HydEn-seq (**Fig. 1a**) by hydrolyzing genomic DNA samples with 0.3 M KOH[23], preparing libraries from the resulting single-stranded DNA fragments and performing 50-base paired-end sequencing on an Illumina HiSeq2500 instrument to identify the location of the 5′ DNA ends. Ribonucleotides were located immediately adjacent to the 5′ DNA ends (**Fig. 1a**). We analyzed two or more independent libraries for each strain (**Supplementary Table 3**) and found that replicate libraries yielded similar results (**Supplementary Table 4**). Alignment of the fragments to a well-annotated reference genome[6] identified the DNA strand to which fragments align and the location and identity of ribonucleotides in the genome. Read counts, scaled according to the number of 5′ ends at the ends of chromosomes (Online Methods), confirmed the relative ribonucleotide densities anticipated by agarose gel electrophoresis.

Strand specificity and origin identification

We found that DNA fragments from the *pol2-M644G rnh201*Δ strain aligned with the two DNA strands in the nuclear genome in an alternating pattern complementary to alignments for the *pol1-Y869A rnh201*Δ and *pol3-L612G rnh20*Δ strains (**Fig. 2**, chromosome 10; **Supplementary Fig. 1**, 5′ DNA end read counts corresponding to **Fig. 2**, bottom; **Fig. 3**, all 16 chromosomes; **Fig. 4a**, heat maps). In contrast, we did not observe a strand-specific pattern in the *pol2-M644L rnh201*Δ strain (**Fig. 4a**) or in RER-proficient strains (**Supplementary Fig. 2**). Thus the majority of 5′ DNA ends in the *pol2-M644G rnh201*Δ, *pol1-Y869A rnh201*Δ and *pol3-L612G rnh20*Δ strains are due to ribonucleotides incorporated during replication that are not removed because RER is defective. Comparison of ribonucleotide maps in these three strains revealed numerous strand-specific transitions (**Figs. 2** and **3**, diamonds). Among these are 294 transitions that correspond to confirmed replication origins in the yeast origin database[24]. We also observed transitions at 72 locations (**Fig. 3**; listed in **Supplementary Table 5**) that have not yet been reported to be origins but may be origins used in some cells in the population.

The ribonucleotide maps in the three *rnh201*Δ strains encoding the variant replicases support earlier interpretations, based on replication errors[4–7], that Pol ε synthesizes the majority of the nascent leading strand, and Pols α and δ synthesize the majority of the nascent lagging

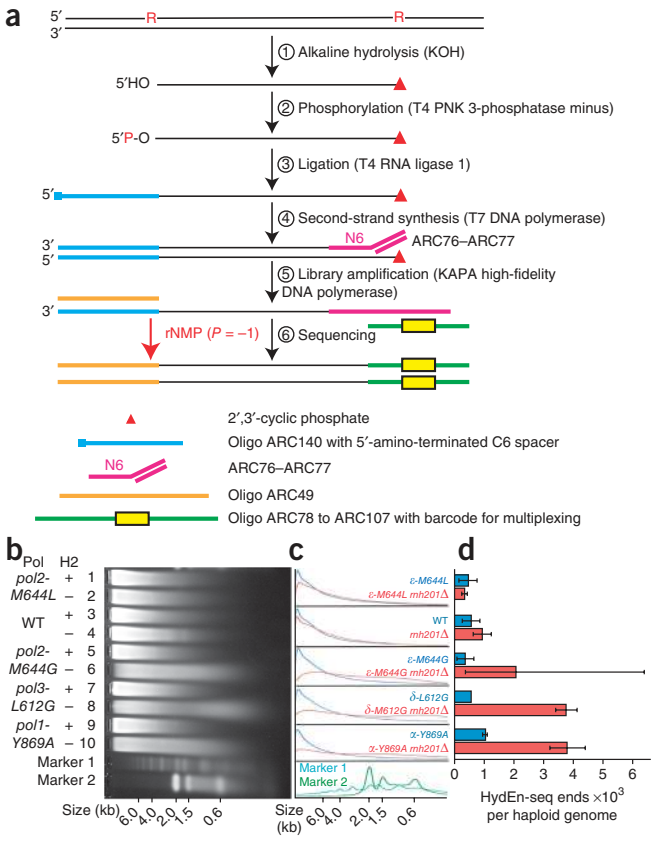

Figure 1 Mapping ribonucleotides by HydEn-seq. (**a**) HydEn-seq protocol. The procedure was performed as described in Online Methods, with the oligonucleotides listed in **Supplementary Table 1**. (**b**) Alkaline agarose gel electrophoresis. The analysis was performed as previously described[17]. Genomic DNA samples from the indicated yeast strains (lanes 1–10) were treated with alkali, separated by 1% alkaline agarose gel electrophoresis and imaged after staining with SYBR Gold. Migration positions of two DNA size standards are indicated. WT, wild type. (**c**) Densitometry scans of the gel image in **b**. The *y* axis is scaled to maximum intensity for each pair of lanes. (**d**) Mean HydEn-seq end counts per haploid genome (N_{ends} calculation described in Online Methods; bars represent ranges of 2–4 independent measurements). Replicate counts from top to bottom: 2, 2, 2, 2, 3, 4, 2, 2, 2 and 3.

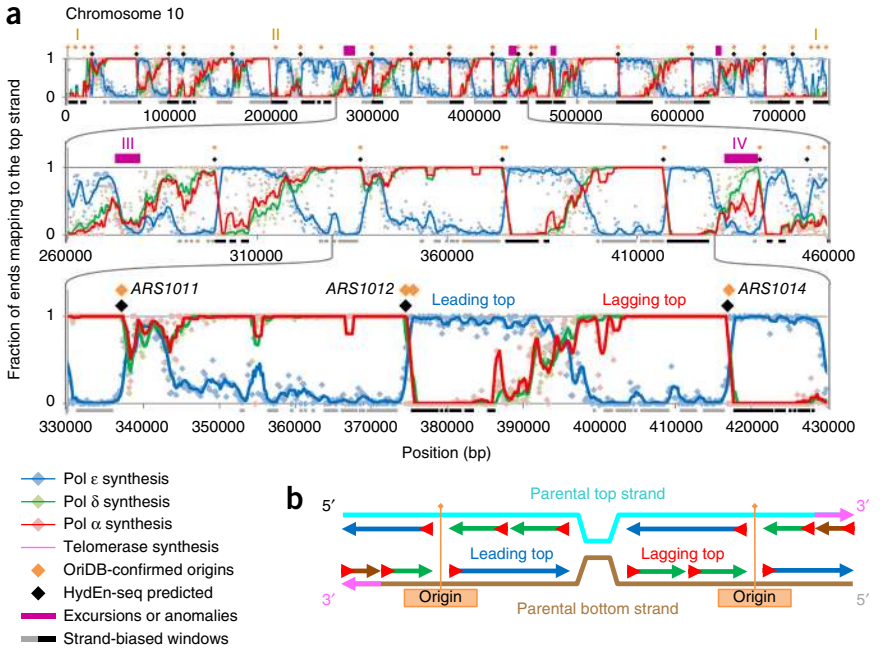

Figure 2 Strand-specific ribonucleotide mapping of chromosome 10. (**a**) Top, map of chromosome 10 showing the fraction of end reads mapped to the top strand in bins of 200 bp, after background subtraction (Online Methods). Origin prediction was complicated at chromosome ends (I) and in other highly repetitive regions (II). Middle, an expanded 200-kb region of chromosome 10. Excursions (in magenta) from the simplest polymerase division of labor (Pol α or δ lagging, Pol ε leading) fall into two classes: unexpected Pol α, δ or ε correspondence (III) and Pol α or δ divergence (IV). Bottom, 100-kb region of chromosome 10. Inter-origin regions are more (for example, *ARS1012-ARS1014*) or less (for example, *ARS1011-ARS1012*) symmetrical, depending on fork progression rates and origin firing times. Some origins in the origin database have little effect on ribonucleotide strand bias (for example, *ARS1013*); this indicates either an incorrect call, minority participation in normally growing cells, unidirectional origin firing or simply later firing, such that forks proceeding from adjacent origins approach to within current detection thresholds. (**b**) A stylized chromosome with two replication origins, showing the division of polymerase labor, as predicted from the direction of strand-bias transitions at origins (cf. the *ARS1012-ARS1014* region above). Roughly three-quarters of previously confirmed replication origins (orange diamonds in **a**; *S. cerevisiae* OriDB) align with abrupt transitions in strand preference (quantitation in Online Methods). This allows algorithmic prediction of origins (black diamonds). *ARS1013* was not detected via HydEn-seq; it is indicated (orange diamond) but not labeled.

strand of the budding-yeast nuclear genome. Thus, HydEn-seq confirms a fundamental aspect of yeast replication enzymology. The evolutionary conservation among eukaryotic nuclear replicases and among type 2 RNases H in all three kingdoms of life (**Supplementary Fig. 3**) suggests that the HydEn-seq strategy used here may be applicable to tracking replication enzymology and identifying origins in other organisms. The ribonucleotide map in the *pol1-Y869A rnh201Δ* strain further demonstrates that when RER is deficient, some DNA synthesized by Y869A Pol α survives Okazaki-fragment maturation and resides in the mature lagging strand. This same conclusion was reached in earlier studies[6,7,25–27] that monitored replication errors rather than ribonucleotides by using *pol3-L612M* strains that were either proficient or deficient in MMR but that were RER proficient. It remains to

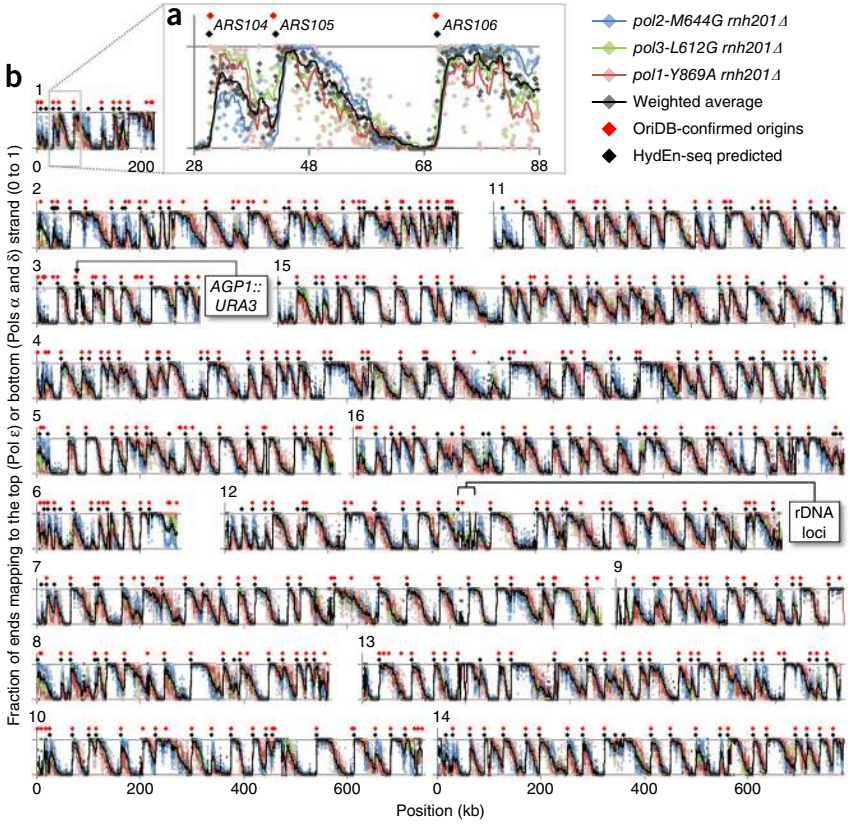

Figure 3 Genome-wide replication origins located by HydEn-seq. (**a**) A 60-kb segment of chromosome 1 showing the fraction of ends, after background subtraction, that mapped to the top strand from Pol ε data (*pol2-M644G rnh201Δ*; blue) and the fraction mapped to the bottom strand for Pols α and δ data (*pol1-Y869A* and *pol3-L612G*, in red and green, respectively). Gray points are the weighted average of the other three data sets in each bin (Online Methods). All curves are trend lines smoothed over ten bins. (**b**) As in **a**, but for all 16 *S. cerevisiae* chromosomes. Shown for reference are the locations of the *URA3* mutational reporter gene (near *ARS306*, used in our previous studies of leading- and lagging-strand replication fidelity[5], and the rDNA locus in chromosome 12 (not drawn to scale; the highly repetitive sequence precludes read mapping).

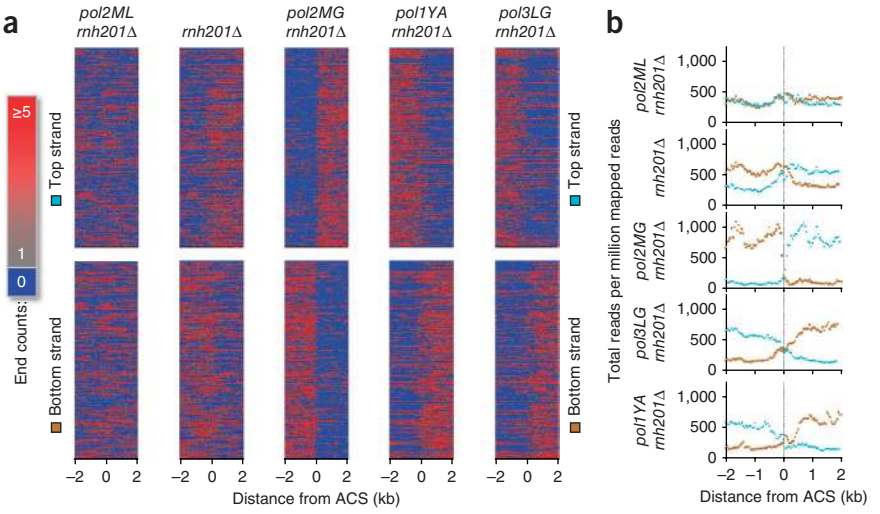

Figure 4 Distribution of ribonucleotides near origins in RER-deficient strains. (**a**) Heat maps for the top and bottom strands of the nuclear genome in five different *rnh201Δ* strains, scaled per million reads and centered across a 4-kb window of the 394 replication origins reported in the yeast origin database[24]. (**b**) Meta-analysis of strand-specific ribonucleotides at 214 replication origins analyzed in a previous study[10], again scaled per million reads, in bins of 50 bp.

be determined whether DNA synthesized by Pol α survives Okazaki-fragment maturation in wild-type yeast.

Polymerase use at replication origins and termination zones

Heat maps (**Fig. 4a**) and meta-analyses of 5′ DNA ends in 50-bp bins (**Fig. 4b**) revealed where strand switches at origins occur in all three replicase-variant backgrounds. These transitions occurred over several hundred base pairs centered on the autonomously replicating sequence (ARS) consensus sequence (ACS; **Fig. 4b**, orange line). The results in the *pol1-Y869A rnh201Δ* strain are consistent with a role for Pol α in initiating synthesis on both strands at origins. The breadth of the strand transition at origins in this strain suggests that in a cell population, initiation occurs within a zone rather than at a single base pair. Deeper coverage of 5′ DNA ends in the future should allow higher-resolution mapping, to investigate whether initiation occurs at a single base pair at some origins, as previously reported for one origin on chromosome 4 (ref. 28). The strand transition at origins is much sharper in the *pol2-M644G rnh201Δ* strain than in the *pol3-L612G rnh201Δ* strain. Investigating this difference may eventually provide information that complements recent biochemical studies of initiation of leading- and lagging-strand replication (ref. 2 and references therein).

HydEn-seq also reveals where mergers occur between forks arriving from adjacent origins and moving in opposite directions. The results suggest that termination occurs in zones that vary in location and breadth. In some cases, the termination zone is broad and equidistant from adjacent origins, whereas in other cases the zone is narrower and or closer to one origin than the other (**Fig. 2a**). HydEn-seq offers the opportunity to explore the mechanisms and genetic controls underlying these variations.

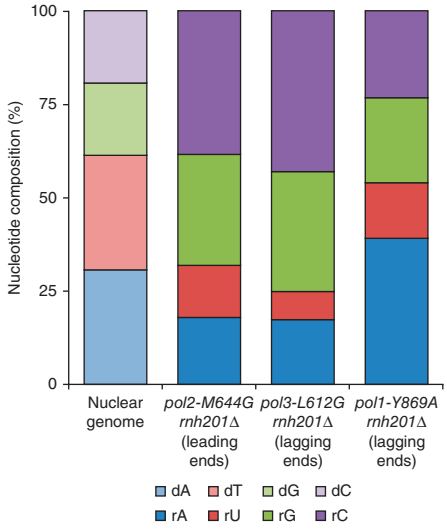

Figure 5 Ribonucleotide base identity. The proportion of each ribonucleotide base present in the nuclear genome of the three *rnh201Δ* strains encoding the indicated variant replicases. The base composition of the genome is shown on the left. Ribonucleotide proportions were calculated from the most highly strand-biased 10% of the genome (i.e., windows near replication origins; examples in **Fig. 2a**).

Ribonucleotide incorporation in wild-type yeast

Studies of ribonucleotide incorporation *in vitro* by wild-type Pols α, δ and ε have predicted that there should be 2.3 times more ribonucleotides incorporated into the nascent leading strand as compared to the nascent lagging strand[23]. This prediction is supported by our results in the RER-defective (*rnh201Δ*) strain encoding wild-type replicases. In this strain, the strand-specific heat map (**Fig. 4a**) and the transition from one strand to the other, as analyzed by meta-analysis (**Fig. 4b**), match those of the *pol2-M644G rnh201Δ* strain and are opposite to those in the *pol3-L612G 201Δ* or *pol1-Y869A rnh201Δ* strains.

The observation that ribonucleotides are preferentially incorporated into the nascent leading strand in the strain encoding wild-type replicases is relevant to the genome instability reported in the wild-type replicase background when RER is defective. In this strain[29], the specificity of 2- to 5-bp deletion mutations resulting from topoisomerase1 (Top1) cleavage at unrepaired ribonucleotides is indistinguishable from the specificity of 2- to 5-bp deletions in the *pol2-M644G rnh201Δ* strain[17] that primarily contains ribonucleotides in the nascent leading strand[5,20,21]. The fact that ribonucleotides preferentially map to the nascent leading strand is also relevant to recent studies[21,30] suggesting that nicks generated by RNase H2 at ribonucleotides in the continuously replicated nascent leading strand may direct MMR to correct replication errors in that strand. This idea, when combined with the nonuniform distribution of ribonucleotides in the genome discussed below, implies that the potential contribution of this MMR signaling mechanism may vary across the genome. The preferential presence of ribonucleotides in the nascent leading strand may also be relevant to other suggested signaling functions for ribonucleotides in DNA[19,23].

Variations in ribonucleotide incorporation by base identity

Wild-type Pols α, δ and ε have different preferences for incorporating each of the four different ribonucleotides *in vitro*[23]. To determine whether this is also true during replication *in vivo*, we analyzed fragments close to replication origins where (as explained previously[6]) leading- and lagging-strand assignments can be made with the greatest

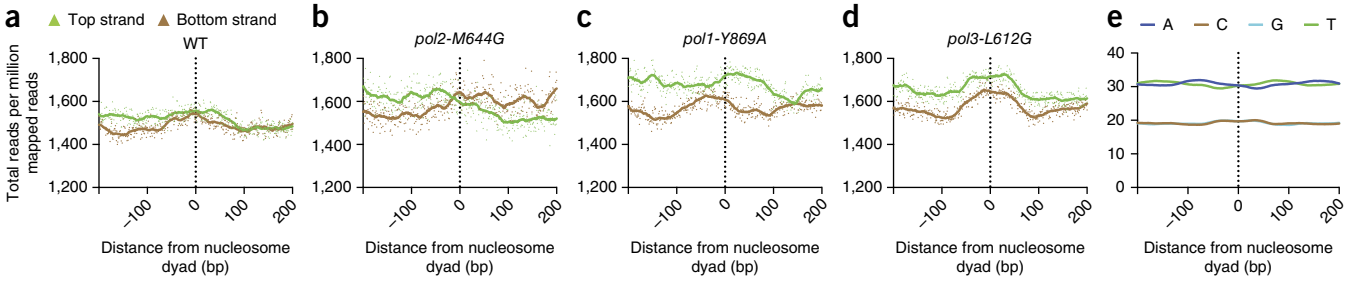

Figure 6 Meta-analysis of ribonucleotides at the nucleosome dyad. (**a**) Meta-analysis of strand-specific ribonucleotide mapping at 37,888 nucleosome dyads[6] for the *rnh201*Δ strain, scaled per million reads and centered within a 400-bp window. Each dot indicates the number of 5′ DNA end reads at one base pair. The vertical dotted line indicates the dyad. (**b**) As in **a**, but for the *pol2-M644G rnh201*Δ strain. (**c**) As in **a**, but for the *pol1-Y869A rnh201*Δ strain. (**d**) As in **a**, but for the *pol3-L612G rnh201*Δ strain. The solid lines are the smoothed averages for a sliding window. (**e**) The base composition surrounding the nucleosome dyad.

confidence. Although that the 12 million–bp budding-yeast genome is 62% A + T, the most abundant ribonucleotide present in the genome of the *pol2-M644G rnh201*Δ strain is rC, and this is followed by rG, then rA and then rU (**Fig. 5**). These preferences recapitulate the rank order for ribonucleotide incorporation by M644G Pol ε *in vitro*[23]. We observed the same rank order for ribonucleotide incorporation (rC > rG > rA > rU) in the *pol3-L612G rnh201*Δ strain, but the proportions of the four rNTPs incorporated are different. For example we observed more rC and less rU in the *pol3-L612G rnh201*Δ genome as compared to the *pol2-M644G rnh201*Δ genome (**Fig. 5**). We observed a different rank order in the *pol1-Y869A rnh201*Δ strain, where after correction for genome composition, the preference in the *pol1-Y869A* strains was rA ≈ rC ≈ rG > rU. In this strain, the non-rU rankings changed slightly among the three replicates examined.

The low abundance of rU seen in all three genomes is consistent with the fact that, among the four dNTPs, dTTP is present at the highest concentration in strains encoding either wild-type replicases[23] or the *pol2-M644G* variant[17], thereby reducing the probability of incorporating rU more than the other ribonucleotides. However, the dATP/rATP ratio is the lowest among the four ratios[23], yet rATP is the most frequent ribonucleotide in only one of the three strains. Thus, in addition to competition for incorporation within the polymerase active site, on the basis of mass action, other parameters may modulate ribonucleotide incorporation probability during replication *in vivo*. This includes the effect of DNA sequence context, as predicted by sequence-context effects of ribonucleotide incorporation probability during DNA synthesis *in vitro*[23,31,32].

Nonuniform distribution of ribonucleotide in the genome

Several HydEn-seq libraries contained an average of less than one 5′ DNA end read per base pair in the nuclear genome (**Supplementary Table 3**). It is therefore striking that end read counts varied from zero at many base pairs to more than 1,000 at others. This nonuniform distribution of ribonucleotides in the genome has implications for MMR signaling, mentioned above, and for a second mechanism of genome instability wherein Top1 incises ribonucleotides in DNA to initiate the deletion of 2–5 bp within repetitive sequences[17,33]. This instability is highly dependent on the DNA strand and sequence context in which the ribonucleotide resides[17]. Variations in the location and density of ribonucleotides in DNA may also be relevant to recombination[34] and gross chromosomal rearrangements in yeast[35] and to chromosomal abnormalities in RNase H2–defective mouse cells[36,37].

In certain regions of the genome, strand-specific ribonucleotide density also deviated from the expectations of a simple division of labor among the three replicases. Initial analyses indicated that these 'excursions' fall into at least two classes: those that show unexpected Pol α, δ or ε correspondence (III in **Fig. 2a**) and those that show unexpected Pol α or δ divergence (IV in **Fig. 2a**). These excursions may result from ribonucleotides remaining in the genomes of these RER-deficient cells, and these ribonucleotides can lead to replicase pausing during DNA synthesis[17,38,39]. Such pausing may elicit template switching or bypass synthesis (for example, Pol ζ or Pol η) in a subsequent round of replication[40] or DNA synthesis associated with DNA repair or

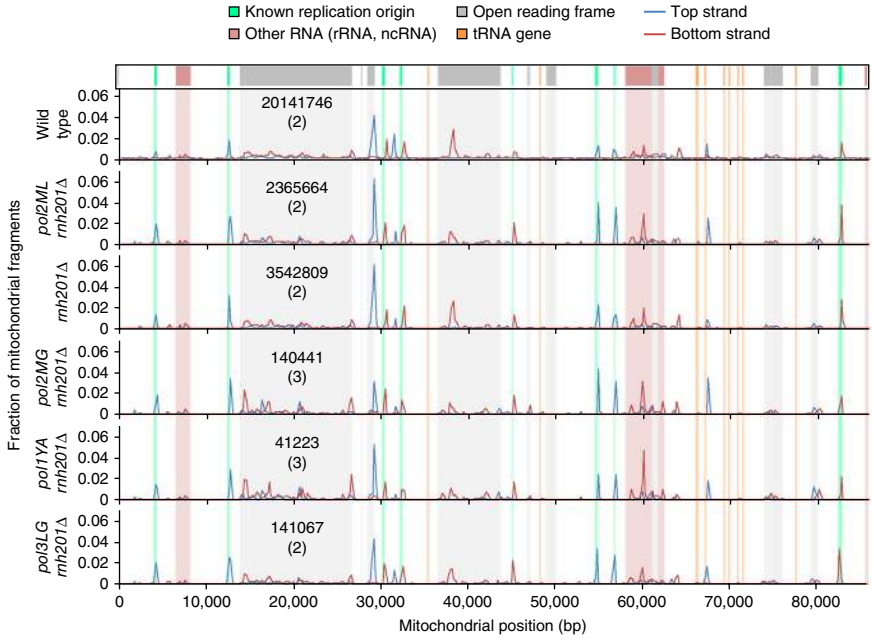

Figure 7 HydEn-seq maps of mitochondrial DNA. Mitochondrial genomes for six strains are shown to indicate the base pair locations and proportions of strand-specific 5′ DNA ends detected by HydEn-seq (blue for plus strand, red for minus). Previously assigned replication origins[43] are shaded in green, coding sequences in gray, tRNA genes in orange and genes for other noncoding RNAs in pink. Total mitochondrial end counts are shown for each strain with the number of replicate HydEn-seq libraries for each in parentheses.

recombination after Top1 incision at ribonucleotides[20,33,34]. An example is in *Schizosaccharomyces pombe*, in which mating-type switching occurs by recombination posited to be initiated by pausing of leading-strand replication upon encountering a diribonucleotide imprint[41]. Additional possibilities for some excursions detected by HydEn-seq include events unrelated to ribonucleotides, such as encounters of replication forks with transcription complexes or the presence of a bulky lesion, repetitive DNA or non-B-form DNA or tightly bound proteins. There is no obvious limitation to monitoring the distance over which a newly recruited DNA polymerase may operate, for example, within a short repair or lesion-bypass patch or to the end of a chromosome during break-induced recombination[42].

Ribonucleotide distribution relative to nucleosome dyads

Meta-analysis with nucleosome-positioning data[6] revealed that ribonucleotide densities are elevated at positions corresponding to the nucleosome dyad (**Fig. 6a–d**). The elevations were subtle (hence the scale on the *y* axis). They may partly reflect a bias in sequence composition, because nucleosome dyads are slightly enriched for G and C content (**Fig. 6e**), the preferred ribonucleotides incorporated during replication in the *pol2-M644G rnh201Δ and pol3-L612G rnh201Δ* strains. However, this may not be the sole explanation because (i) the peaks at the dyad are more prominent in the *pol3-L612G rnh201Δ* strain (**Fig. 6d**) as compared to the *pol2-M644G rnh201Δ* strain (**Fig. 6b**), yet these two strains have similar G + C versus A + T ribonucleotide-incorporation preferences (**Fig. 5**); (ii) the peaks are more prominent in the *pol3-L612G rnh201Δ* and *pol1-Y869A rnh201Δ* strains as compared to the *pol2-M644G rnh201Δ* strain; and (iii) the peaks in both strands in the *pol3-L612G rnh201Δ* strain are symmetrical around the dyad, whereas the peaks in the two strands in the *pol1-Y869A rnh201Δ* strain are offset and on opposite sides of the dyad. In the latter cases, lagging-strand replicase features could be signatures of polymerization by Pol α and Pol δ during Okazaki-fragment maturation, a process that is proposed to preferentially occur at the nucleosome dyad and to be phased according to the nucleosome repeat[10].

Ribonucleotides at mitochondrial DNA replication origins

HydEn-seq also revealed that the yeast mitochondrial genome contains large numbers of 5′-DNA ends generated by alkaline hydrolysis (**Fig. 7**). Most of these ends were in discrete, strand-specific peaks that span multiple base pairs. Eight of these peaks correspond to previously identified[43] mitochondrial replication origins. Interestingly, the relative proportions of 5′ DNA ends in the major peaks were similar in all yeast strains examined. Thus the peaks are independent of the status of the nuclear replicases, which have no known role in mitochondrial replication, and they are also independent of the status of RNase H2, which has not been found in mitochondria[44].

These observations are consistent with at least three hypotheses. The peaks may represent the ends of linear chromosomes, similar to the high density of 5′ ends observed at the ends of linear nuclear chromosomes. This possibility cannot yet be eliminated, but it seems unlikely because the most prominent peaks largely map to either the plus or the minus strand but not to both. Also, when we rearranged the mitochondrial genome *in silico* (Online Methods) to join the chromosome 'ends' that were arbitrarily assigned and numbered when the genome was sequenced[43], we observed no drop in the depth of coverage of fragments at the junction relative to immediately adjacent regions. Thus, like mammalian mitochondrial genomes, the *Saccharomyces cerevisiae* mitochondrial genome may largely, albeit not necessarily exclusively, be circular.

The second hypothesis is that some mitochondrial DNA (mtDNA) fragments generated may be due to lesions other than ribonucleotides, such as strand-specific nicks or alkali-sensitive abasic sites resulting from oxidative stress. We cannot exclude this possibility, but it is currently disfavored by the fact that the 5′-DNA ends are distributed in a highly nonuniform and largely strand-specific manner.

The third hypothesis stems from previous studies showing that mammalian mtDNA contains ribonucleotides[45,46] and that human mitochondrial replicase (Pol γ) incorporates ribonucleotides during DNA synthesis *in vitro*[47]. Moreover, eight of the most prominent, strand-biased 5′-DNA-end peaks in the yeast mitochondrial genome correspond to previously identified[43] mitochondrial replication origins. These peaks, and perhaps similar strand-specific peaks detected within open reading frames and in sequences encoding RNAs, could reflect the presence of unrepaired residues of RNA primers made by mtRNA polymerase and used to initiate mtDNA replication, as has been reported in mammalian cells[16,48–51]. The results suggest that the terminal ribonucleotides of RNA primers for mtDNA replication may not always be removed, either by mitochondrial RNase H1 (ref. 52), which cannot incise at an RNA-DNA junction[44], or by strand displacement and flap cleavage, as for Okazaki-fragment maturation during nuclear DNA replication[9]. If this explanation holds, then the fact that 5′-DNA ends at the origins of mtDNA replication preferentially map to one strand or the other favors a unidirectional replication model for mtDNA in budding yeast.

DISCUSSION

This study demonstrates that ribonucleotides can be used to track replication enzymology at high resolution, using a simple five-step library preparation procedure involving minimal use of enzymes and requiring under 2 d to execute. Although HydEn-seq is used here to map 5′ DNA ends primarily generated by alkaline hydrolysis at ribonucleotides, it can also be used to study other lesions in DNA, and it is not limited to spontaneous chemical hydrolysis but can be adapted to map 5′ and 3′ DNA ends generated by enzymatic hydrolysis. HydEn-seq should be useful to study polymerization changes in response to endogenous and exogenous environmental stress in addition to normal replication enzymology. The ability of HydEn-seq to identify replication origins, termination zones and polymerase usage during replication should be applicable to other organisms in which replicases can be engineered to enhance ribonucleotide incorporation and RER can be inactivated. Polymerase structure-function studies have advanced to the point that it is now feasible to engineer replicases (e.g., **Supplementary Fig. 3a**), and more-specialized polymerases in most polymerase families, so that they retain catalytic efficiency yet are rendered promiscuous for ribonucleotide incorporation. Theoretically, this may permit a variety of DNA synthesis reactions in cells to be studied by HydEn-seq.

Because high-density peaks in the mitochondrial genome may be due to unrepaired residues of RNA primers made by mtRNA polymerase, HydEn-seq may also be useful to study RNA primers synthesized by RNA polymerases, RNA primases or enzymes with combined primase and polymerase activity. The ability to map genomic locations that contain a high density of ribonucleotides can be used to explore the idea that ribonucleotides in DNA provide selective advantages to cells[23]. Relevant here are ribonucleotides that may persist in certain locations even in RER-proficient cells, as exemplified by the diribonucleotide imprint used for mating-type switching in *S. pombe*[53].

Note added in proof: Three other recent articles also illustrate the value of using ribonucleotides as markers of replication enzymology in budding yeast[54–56].

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Sequencing data have been deposited in the Gene Expression Omnibus database under accession code GSE62181.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank M. Young and M. Longley for helpful comments on the manuscript.
This work was supported by the Division of Intramural Research of the US National Institutes of Health (NIH), National Institute of Environmental Health Sciences (project Z01 ES065070 to T.A.K.) and by NIH grant 2R01GM052319-16A1 to P.A.M.

AUTHOR CONTRIBUTIONS

A.R.C., D.J.S. and T.A.K. designed the experiments, A.R.C., C.D.O., J.S.W., M.F.C. and E.P.M. performed the experiments, A.R.C., S.A.L., A.B.B., D.C.F., P.A.M. and T.A.K. analyzed the data, T.A.K. wrote the manuscript, and all authors edited the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Johansson, E. & Dixon, N. Replicative DNA polymerases. *Cold Spring Harb. Perspect. Biol.* **5**, a012799 (2013).
2. Georgescu, R.E. *et al.* Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat. Struct. Mol. Biol.* **21**, 664–670 (2014).
3. Burgers, P.M. Polymerase dynamics at the eukaryotic DNA replication fork. *J. Biol. Chem.* **284**, 4041–4045 (2009).
4. Pursell, Z.F., Isoz, I., Lundstrom, E.B., Johansson, E. & Kunkel, T.A. Yeast DNA polymerase ε participates in leading-strand DNA replication. *Science* **317**, 127–130 (2007).
5. Lujan, S.A. *et al.* Mismatch repair balances leading and lagging strand DNA replication fidelity. *PLoS Genet.* **8**, e1003016 (2012).
6. Lujan, S.A. *et al.* Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res.* **24**, 1751–1764 (2014).
7. Nick McElhinny, S.A., Gordenin, D.A., Stith, C.M., Burgers, P.M. & Kunkel, T.A. Division of labor at the eukaryotic replication fork. *Mol. Cell* **30**, 137–144 (2008).
8. Larrea, A.A. *et al.* Genome-wide model for the normal eukaryotic DNA replication fork. *Proc. Natl. Acad. Sci. USA* **107**, 17674–17679 (2010).
9. Balakrishnan, L. & Bambara, R.A. Okazaki fragment metabolism. *Cold Spring Harb. Perspect. Biol.* **5**, a010173 (2013).
10. Smith, D.J. & Whitehouse, I. Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature* **483**, 434–438 (2012).
11. McGuffee, S.R., Smith, D.J. & Whitehouse, I. Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol. Cell* **50**, 123–135 (2013).
12. Yeeles, J.T., Poli, J., Marians, K.J. & Pasero, P. Rescuing stalled or damaged replication forks. *Cold Spring Harb. Perspect. Biol.* **5**, a012815 (2013).
13. Ghosal, G. & Chen, J. DNA damage tolerance: a double-edged sword guarding the genome. *Transl. Cancer Res.* **2**, 107–129 (2013).
14. Wanrooij, S. & Falkenberg, M. The human mitochondrial replication fork in health and disease. *Biochim. Biophys. Acta* **1797**, 1378–1388 (2010).
15. Gerhold, J.M., Aun, A., Sedman, T., Joers, P. & Sedman, J. Strand invasion structures in the inverted repeat of *Candida albicans* mitochondrial DNA reveal a role for homologous recombination in replication. *Mol. Cell* **39**, 851–861 (2010).
16. Reyes, A. *et al.* Mitochondrial DNA replication proceeds via a 'bootlace' mechanism involving the incorporation of processed transcripts. *Nucleic Acids Res.* **41**, 5837–5850 (2013).
17. Nick McElhinny, S.A. *et al.* Genome instability due to ribonucleotide incorporation into DNA. *Nat. Chem. Biol.* **6**, 774–781 (2010).
18. Sparks, J.L. *et al.* RNase H2-initiated ribonucleotide excision repair. *Mol. Cell* **47**, 980–986 (2012).
19. Williams, J.S. & Kunkel, T.A. Ribonucleotides in DNA: origins, repair and consequences. *DNA Repair (Amst.)* **19**, 27–37 (2014).
20. Williams, J.S. *et al.* Topoisomerase 1-mediated removal of ribonucleotides from nascent leading-strand DNA. *Mol. Cell* **49**, 1010–1015 (2013).
21. Lujan, S.A., Williams, J.S., Clausen, A.R., Clark, A.B. & Kunkel, T.A. Ribonucleotides are signals for mismatch repair of leading-strand replication errors. *Mol. Cell* **50**, 437–443 (2013).
22. Pavlov, Y.I., Shcherbakova, P.V. & Kunkel, T.A. *In vivo* consequences of putative active site mutations in yeast DNA polymerases α, ε, δ, and ζ. *Genetics* **159**, 47–64 (2001).
23. Nick McElhinny, S.A. *et al.* Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc. Natl. Acad. Sci. USA* **107**, 4949–4954 (2010).
24. Siow, C.C., Nieduszynska, S.R., Muller, C.A. & Nieduszynski, C.A. OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res.* **40**, D682–D686 (2012).
25. Niimi, A. *et al.* Palm mutants in DNA polymerases α and η alter DNA replication fidelity and translesion activity. *Mol. Cell. Biol.* **24**, 2734–2746 (2004).
26. Nick McElhinny, S.A., Stith, C.M., Burgers, P.M. & Kunkel, T.A. Inefficient proofreading and biased error rates during inaccurate DNA synthesis by a mutant derivative of *Saccharomyces cerevisiae* DNA polymerase δ. *J. Biol. Chem.* **282**, 2324–2332 (2007).
27. Nick McElhinny, S.A., Kissling, G.E. & Kunkel, T.A. Differential correction of lagging-strand replication errors made by DNA polymerases α and δ. *Proc. Natl. Acad. Sci. USA* **107**, 21070–21075 (2010).
28. Bielinsky, A.K. & Gerbi, S.A. Chromosomal ARS1 has a single leading strand start site. *Mol. Cell* **3**, 477–486 (1999).
29. Clark, A.B., Lujan, S.A., Kissling, G.E. & Kunkel, T.A. Mismatch repair-independent tandem repeat sequence instability resulting from ribonucleotide incorporation by DNA polymerase ε. *DNA Repair (Amst.)* **10**, 476–482 (2011).
30. Ghodgaonkar, M.M. *et al.* Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. *Mol. Cell* **50**, 323–332 (2013).
31. Clausen, A.R., Zhang, S., Burgers, P.M., Lee, M.Y. & Kunkel, T.A. Ribonucleotide incorporation, proofreading and bypass by human DNA polymerase δ. *DNA Repair (Amst.)* **12**, 121–127 (2013).
32. Williams, J.S. *et al.* Proofreading of ribonucleotides inserted into DNA by yeast DNA polymerase ε. *DNA Repair (Amst.)* **11**, 649–656 (2012).
33. Kim, N. *et al.* Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science* **332**, 1561–1564 (2011).
34. Potenski, C.J., Niu, H., Sung, P. & Klein, H.L. Avoidance of ribonucleotide-induced mutations by RNase H2 and Srs2-Exo1 mechanisms. *Nature* **511**, 251–254 (2014).
35. Allen-Soltero, S., Martinez, S.L., Putnam, C.D. & Kolodner, R.D. A *Saccharomyces cerevisiae* RNase H2 interaction network functions to suppress genome instability. *Mol. Cell. Biol.* **34**, 1521–1534 (2014).
36. Reijns, M.A. *et al.* Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* **149**, 1008–1022 (2012).
37. Hiller, B. *et al.* Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity. *J. Exp. Med.* **209**, 1419–1426 (2012).
38. Watt, D.L., Johansson, E., Burgers, P.M. & Kunkel, T.A. Replication of ribonucleotide-containing DNA templates by yeast replicative polymerases. *DNA Repair (Amst.)* **10**, 897–902 (2011).
39. Clausen, A.R., Murray, M.S., Passer, A.R., Pedersen, L.C. & Kunkel, T.A. Structure-function analysis of ribonucleotide bypass by B family DNA replicases. *Proc. Natl. Acad. Sci. USA* **110**, 16802–16807 (2013).
40. Lazzaro, F. *et al.* RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA. *Mol. Cell* **45**, 99–110 (2012).
41. Dalgaard, J.Z. Causes and consequences of ribonucleotide incorporation into nuclear DNA. *Trends Genet.* **28**, 592–597 (2012).
42. Anand, R.P., Lovett, S.T. & Haber, J.E. Break-induced DNA replication. *Cold Spring Harb. Perspect. Biol.* **5**, a010397 (2013).
43. Foury, F., Roganti, T., Lecrenier, N. & Purnelle, B. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett.* **440**, 325–331 (1998).
44. Cerritelli, S.M. & Crouch, R.J. Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* **276**, 1494–1505 (2009).
45. Grossman, L.I., Watson, R. & Vinograd, J. The presence of ribonucleotides in mature closed-circular mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **70**, 3339–3343 (1973).
46. Yang, M.Y. *et al.* Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. *Cell* **111**, 495–505 (2002).
47. Kasiviswanathan, R. & Copeland, W.C. Ribonucleotide discrimination and reverse transcription by the human mitochondrial DNA polymerase. *J. Biol. Chem.* **286**, 31490–31500 (2011).
48. Shadel, G.S. & Clayton, D.A. Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* **66**, 409–435 (1997).
49. Bowmaker, M. *et al.* Mammalian mitochondrial DNA replicates bidirectionally from an initiation zone. *J. Biol. Chem.* **278**, 50961–50969 (2003).
50. Fusté, J.M. *et al.* Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Mol. Cell* **37**, 67–78 (2010).
51. Holt, I.J. & Reyes, A. Human mitochondrial DNA replication. *Cold Spring Harb. Perspect. Biol.* **4**, a012971 (2012).
52. Cerritelli, S.M. *et al.* Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Mol. Cell* **11**, 807–815 (2003).
53. Vengrova, S. & Dalgaard, J.Z. The wild-type *Schizosaccharomyces pombe* mat1 imprint consists of two ribonucleotides. *EMBO Rep.* **7**, 59–65 (2006).
54. Koh, K.D., Balachander, S., Hesselberth, J.R., & Storici, F. Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. *Nat. Methods* doi:10.1038/nmeth.3259 (26 January 2015).
55. Reijns, M.A.M. *et al.* Lagging strand replication shapes the mutational landscape of the genome. *Nature* doi:10.1038/nature14183 (26 January 2015).
56. Daigaku, Y. *et al.* A global profile of replicative polymerase usage. *Nat. Struct. Mol. Biol.* (in the press).

ONLINE METHODS

Materials. Oligonucleotides and yeast strains used in this study are listed in **Supplementary Tables 1** and **2**, respectively. The *pol1-Y869A* and *pol3-L612G* strains and their *rnh201*Δ derivatives were constructed as described earlier for *pol1-L868M* and *pol3-L612M* strains[7].

HydEn-seq protocol. Yeast strains were grown to mid-log phase (OD_{600} = 0.6) at 30 °C in YPDA medium supplemented with 0.25 mg/ml adenine. DNA was isolated with the MasterPure Yeast DNA Purification Kit (Epicentre) without RNase A treatment. HydEn-seq (**Fig. 1**) was performed by hydrolysis of 1 µg of genomic DNA with 0.3 M KOH for 2 h at 55 °C (ref. 23). After ethanol precipitation, the DNA fragments were treated for 3 min at 85 °C, phosphorylated with 10 U of 3′-phosphatase-minus T4 polynucleotide kinase (New England BioLabs) for 30 min at 37 °C, heat inactivated for 20 min at 65 °C and purified with HighPrep PCR beads (MagBio). Phosphorylated products were treated for 3 min at 85 °C, ligated to oligo ARC140 (**Supplementary Table 1**) overnight at room temperature with 10 U of T4 RNA ligase, 25% PEG 8000 and 1 mM $CoCl_3(NH_3)_6$, and purified with HighPrep PCR beads (MagBio). Ligated products were treated for 3 min at 85 °C. The ARC76–ARC77 adaptor was annealed to the second strand for 5 min at room temperature. The second strand was synthesized with 4 U of T7 DNA polymerase (New England BioLabs) and purified with HighPrep PCR beads (MagBio). Libraries were PCR amplified with primer ARC49 and primer ARC79 or ARC84 to ARC107, with KAPA HiFi Hotstart ReadyMix (KAPA Biosystems). Libraries were then purified with HighPrep PCR beads (MagBio) and pooled for sequence analysis. Paired-end sequencing was performed on a HiSeq2500 sequencer (Illumina) to identify the location of the 5′ DNA ends generated by alkaline hydrolysis.

HydEn-seq trimming, filtering and alignment. All reads were trimmed for quality and adaptor sequence with cutadapt 1.2.1 (-m 15 -q 10–match-read-wildcards)[57]. Pairs with one or both reads shorter than 15 nt were discarded. Mate 1 of the remaining pairs was aligned to an index containing the sequence of all oligos used in the preparation of these libraries with bowtie 0.12.8 (-m1 -v2), and all pairs with successful alignments were discarded. Pairs passing this filter were subsequently aligned to the L03 *S. cerevisiae* reference genome[6] (-m1 -v2 -X10000–best). Single-end alignments were then performed with mate 1 of all unaligned pairs (-m1 -v2). The count of 5′ ends of all unique paired-end and single-end alignments were determined for all samples, per strand, across all chromosomes, combining all technical replicates, and shifted one base upstream to the location of the hydrolyzed ribonucleotide as summarized (**Supplementary Table 3**). These counts were converted to bigWig format for visualization on the UCSC browser. The distributions of counts per nucleotide were determined with these values.

End count scaling and background subtraction. Two modes of end count scaling were used, depending on several factors. For analyses resulting in visual comparisons of individual libraries (i.e., heat maps and meta-analyses), end counts were normalized to counts per million uniquely mapped reads (divided by the values listed in **Supplementary Table 3** under 'uniquely mapped ends' and then multiplied by 1,000,000). For analyses that required weighted averaging of multiple libraries and background subtraction (i.e., strand bias maps, origin predictions and genomic ribonucleotide density estimates), end counts were scaled with *n* chromosomal 5′-end counts as internal standards (with counts divided by the values listed in **Supplementary Table 3** under 'telomere end-derived scale factor'; described below). To ensure that ends in these latter analyses originated only from replicase-inserted genomic ribonucleotides, scaled end counts from polymerase 'L' variant strains (*rnh201*Δ, *rnh201*Δ, and *pol2-M644L rnh201*Δ) were subtracted from the scaled end counts of corresponding promiscuous-replicase strains (*pol1-Y869A rnh201*Δ, *pol3-L612G rnh201*Δ, and *pol2-M644G rnh201*Δ, respectively).

Calculating telomere end-derived scale factors and genomic ribonucleotide densities. The genomic ribonucleotide density (R_{bulk}) is

$$R_{bulk} = \frac{N_{bulk}}{2 \times L_{genome} - L_{telomere}}$$

where N_{bulk} is the bulk end count, L_{genome} is the length of the genome and $L_{telomere}$ is the total length of all telomeric repeats in the reference genome

$$L_{telomere} = \sum_{i=1}^{2 \times N_C} \max(0, L_{i,telomere} - L_{read})$$

where N_C is the chromosome count (16 in *S. cerevisiae*). The mean number of 5′ chromosome ends per telomere ($N_{\overline{telomere}}$) is similar, but it should be corrected for R_{bulk} in order to account for ribonucleotides found in telomeric repeats but not at chromosome ends

$$N_{\overline{telomere}} = \frac{N_{telomere} - L_{telomere} \times R_{bulk}}{2 \times N_C}$$

where $N_{telomere}$ is the unadjusted total telomeric end count. (This correction never amounted to more than 2% of the final value.) $N_{\overline{telomere}}$ serves as a scaling factor, allowing conversion of end counts in any bin into counts per position per genome (**Supplementary Table 3**). Where the bin is the whole genome, this results in an estimate of the mean fragment size ($L_{\overline{fragment}}$; always larger than the median fragment size as reported in refs. 20,21)

$$L_{\overline{fragment}} = \frac{N_{\overline{telomere}}}{R_{bulk}}$$

and thence the mean number of ends per genome (N_{ends}; always smaller than the median count per genome as reported in refs. 20,21)

$$N_{ends} = \frac{2 \times L_{genome}}{L_{\overline{fragment}}}.$$

Predicting replication origins from HydEn-seq maps. Replication origins were predicted from the weighted-average fraction of scaled and background-subtracted ends (described above) mapping to the top strand in the Pol α, δ or ε variant strains. In each bin, the weighted-average top-strand-fraction (f) was

$$\bar{f} = \begin{cases} Null, \alpha_f + \alpha_r + \delta_f + \delta_r + \varepsilon_f + \varepsilon_r = 0 \\ \dfrac{\left(\begin{cases} 0, \alpha_f + \alpha_r = 0 \\ 1 - \dfrac{\alpha_f}{\alpha_f + \alpha_r}, otherwise \end{cases}\right) + \left(\begin{cases} 0, \delta_f + \delta_r = 0 \\ 1 - \dfrac{\delta_f}{\delta_f + \delta_r}, otherwise \end{cases}\right) + \left(\begin{cases} 0, \varepsilon_f + \varepsilon_r = 0 \\ \dfrac{2\varepsilon_f}{\varepsilon_f + \varepsilon_r}, otherwise \end{cases}\right)}{\left(\begin{cases} 0, \alpha_f + \alpha_r = 0 \\ 1, otherwise \end{cases}\right) + \left(\begin{cases} 0, \delta_f + \delta_r = 0 \\ 1, otherwise \end{cases}\right) + \left(\begin{cases} 0, \varepsilon_f + \varepsilon_r = 0 \\ 2, otherwise \end{cases}\right)}, otherwise \end{cases}$$

where α, δ and ε are the background-subtracted end counts from *pol1-Y869A rnh201*Δ, *pol3-L612G rnh201*Δ and *pol2-M644G rnh201*Δ strains, respectively (each of which is itself the average of scaled counts from all replicate libraries). Parameters for origin calling were set on the basis of results from a training set of OriDB-confirmed replication origins on chromosome 11. Predicted origins were defined as regions where the bias changed abruptly over a defined distance in the weighted-average curve in **Figure 4** (black; 200-bp bins; smoothed over nine bins). In order for a position to be called an origin, either the average slope

doi:10.1038/nsmb.2957

(the derivative) of the black curve had to exceed 0.00011 fractional units per bp in an 11-bin window (2.2 kb) or 0.00016 fractional units per base pair in at least three of five surrounding bins (≥600 bp out of 1 kb). These parameters attempt to define a sufficiently abrupt bias change over a region wide enough to exclude random noise.

Meta-analyses and preparation of heat maps. Total counts of the per-strain 5′ ends intersecting same- and opposite-strand bins centered on genomic features of interest were determined with custom tools, excluding all mitochondrial annotations. Heat maps, generated with the Partek Genomics Suite depict counts in all bins (normalized to ends per 1,000,000 uniquely mapped reads; **Supplementary Table 3**), whereas meta-analyses depict the sum across all features.

Ribonucleotide frequencies. The composition of uniquely mapped ends was tallied in defined windows on each genomic strand. Windows were set in regions of high strand bias (≥99%) to ensure that nearly all ends represented ribonucleotides inserted by a particular replicase, depending on strand (leading versus lagging; for example, gray or black bars, respectively in **Fig. 3**). The frequency of nucleotides occurring at 5′ ends intersecting these windows was determined with custom tools.

***In silico* mapping of the mitochondrial genome.** The mitochondrial genome sequence was rearranged, such that the first 42,888 bp were removed from the start and appended to the end. Read pairs from sample WT.1 (**Supplementary Table 3**) were aligned to this reordered mitochondrial genome, with bowtie 0.12.8 (-m1 -v2 -X10000). On the basis of start and end coordinates of the paired-end alignments, per-nucleotide coverage of HydEn-seq fragments was determined with genomeCoverageBed (-d).

Analysis of polymerase and RNase H2 conservation. PSI-BLAST searches[58] (parameters in **Supplementary Table 6**) were conducted to find sequences homologous to *S. cerevisiae* replicases (catalytic subunit sequences from the *Saccharomyces* Genome Database) and the predicted *RNH201* of *S. pombe* 972h- (GI 19114596). Environmental sequences were excluded. For the replicases, PSI-BLAST was iterated until no new eukaryotic sequences were found. For *RNH201*, PSI-BLAST was iterate until >5,000 hits were acquired. The top hits were selected until the cumulative *e* value exceeded 1.47. In all cases, partial sequences were culled and the remainder were aligned with CLUSTAL X (2.0) default parameters[59]. Sequences with obvious deletions spanning active sites were culled, the remainder realigned and trees built from the results (neighbor-joining, default parameters). The tree in **Supplementary Figure 3** was constructed in part with the Interactive Tree of Life version 2.2.2 (http://itol.embl.de/)[60].

57. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10–2 (2011).
58. Altschul, S.F. *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
59. Larkin, M.A. *et al*. Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
60. Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**, W475–W478 (2011).