

# DNA polymerase $\eta$ contributes to genome-wide lagging strand synthesis

Katrin Kreisel<sup>1</sup>, Martin K.M. Engqvist<sup>1,2</sup>, Josephine Kalm<sup>1</sup>, Liam J. Thompson<sup>1</sup>, Martin Boström<sup>1</sup>, Clara Navarrete<sup>1</sup>, John P. McDonald<sup>3</sup>, Erik Larsson<sup>1</sup>, Roger Woodgate<sup>3</sup> and Anders R. Clausen<sup>1,\*</sup>

<sup>1</sup>Department of Medical Biochemistry and Cell Biology, University of Gothenburg, 40530 Gothenburg, Sweden,

<sup>2</sup>Department of Biology and Biological Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden and

<sup>3</sup>Laboratory of Genomic Integrity, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA

Received October 19, 2018; Revised December 11, 2018; Editorial Decision December 13, 2018; Accepted December 14, 2018

## ABSTRACT

**DNA polymerase  $\eta$  (pol  $\eta$ ) is best known for its ability to bypass UV-induced thymine–thymine (T–T) dimers and other bulky DNA lesions, but pol  $\eta$  also has other cellular roles. Here, we present evidence that pol  $\eta$  competes with DNA polymerases  $\alpha$  and  $\delta$  for the synthesis of the lagging strand genome-wide, where it also shows a preference for T–T in the DNA template. Moreover, we found that the C-terminus of pol  $\eta$ , which contains a PCNA-Interacting Protein motif is required for pol  $\eta$  to function in lagging strand synthesis. Finally, we provide evidence that a pol  $\eta$  dependent signature is also found to be lagging strand specific in patients with skin cancer. Taken together, these findings provide insight into the physiological role of DNA synthesis by pol  $\eta$  and have implications for our understanding of how our genome is replicated to avoid mutagenesis, genome instability and cancer.**

## INTRODUCTION

The nuclear genome is replicated primarily by three DNA polymerases (1): DNA polymerase  $\alpha$  (pol  $\alpha$ ) and  $\delta$  (pol  $\delta$ ) synthesize the lagging strand, while DNA polymerase  $\epsilon$  (pol  $\epsilon$ ) continuously synthesizes most of the leading strand (2,3). This phenomenon has also been described as the ‘division of labor’ among replicative polymerases. The stringent discrimination by replicative DNA polymerases limits their ability to replicate DNA at sites where the template strand is difficult to replicate, or damaged (4). These sites in the genome may cause the replication fork to stall. Alternatively, helicases can uncouple from the stalled DNA polymerase and continue to unwind the double strand further allowing DNA synthesis on one strand to continue

but leaving stretches of unreplicated single strand DNA in front of the stalled polymerase (5). To prevent DNA breaks and replication fork collapse, cells can synthesize across such sites through a mechanism called translesion synthesis (TLS). TLS is facilitated by a number of specialized DNA polymerases equipped to handle a variety of perturbations (6).

DNA polymerase  $\eta$  (pol  $\eta$ ) is best characterized by its ability to accurately bypass UV-induced thymine–thymine (T–T) dimers (7). All Y-family polymerases, such as pol  $\eta$ , possess a typical ‘right-hand’ structure that can adapt to the encountered DNA substrate. Its protective and beneficial role in normal cells is apparent when one considers the strongly increased predisposition for cancer in *Xeroderma Pigmentosum* Variant patients, which is caused by the impairment of TLS across DNA lesions due to loss of pol  $\eta$  function (7,8). It is evident, however, that in addition to TLS of UV-induced lesions, pol  $\eta$  also has several other cellular roles, such as promoting somatic hypermutation (9,10), replication of common fragile sites (11,12), replication of recombination intermediates (13,14), replication of cohesion-bound DNA (15) and the maintenance of telomere DNAs in undamaged cells (16,17).

Given that TLS pol  $\eta$  also participates in a variety of DNA transactions in undamaged cells, it is of interest to know where in the genome pol  $\eta$  is active and if there is a strand bias (leading vs. lagging) for pol  $\eta$ 's activities as observed for the replicative polymerases (1). As previously shown for the replicative DNA polymerases  $\alpha$ ,  $\delta$  and  $\epsilon$ , ribonucleotide incorporation by their respective steric gate variants can be used to track DNA polymerase activity with high resolution. Using the HydEn-seq method, which maps incorporated ribonucleotides genome-wide at single nucleotide resolution level (3), a footprint of the polymerase is revealed. To understand where pol  $\eta$  may gain access to the DNA and compete with replicative polymerases we studied its function in *Saccharomyces cerevisiae*. We uti-

\*To whom correspondence should be addressed. Tel: +46 766 18 50 79; Email: anders.clausen@gu.se

lized a steric gate mutant of pol  $\eta$  (*rad30-F35A*) that has an increased propensity to incorporate ribonucleotides into DNA, but also possesses an increased deoxyribonucleotide base-selection fidelity on undamaged DNA (18). To enhance the detection of pol  $\eta$ -dependent ribonucleotide incorporation, we made use of strains where *RNH201* was deleted and are therefore defective in ribonucleotide excision repair (RER), which is the major pathway for ribonucleotide removal. Since nucleotide excision repair (NER) serves as a back-up pathway for the removal of ribonucleotides incorporated into DNA in the absence of RER (19), NER was abolished by deletion of *RAD1* to further increase the detection of pol  $\eta$ -dependent incorporated ribonucleotides. Lastly, the strains lacked the catalytic subunit of DNA polymerase  $\zeta$  (pol  $\zeta$ ) through deletion of *REV3*, which is responsible for the majority of spontaneous mutagenesis in yeast (20,21) and which might compete with pol  $\eta$  for access to the undamaged DNA (22).

Our data reveal that pol  $\eta$ 's activities can be tracked across the entire yeast genome and mapped to the lagging strand, where it competes with pol  $\alpha$  and  $\delta$  during genome replication. Furthermore, our data suggests that pol  $\eta$  prefers T–T template sequences and that pol  $\eta$ 's strand-specificity is mediated by its PCNA-Interacting Protein (PIP) motif. Finally, in human melanomas we detected a weak strand bias for WA>WG transitions attributed to pol  $\eta$ , suggesting an evolutionary conservation of strand specificity.

## MATERIALS AND METHODS

### Experimental model and subject details

All yeast strains used for this study can be found in the Supplementary Table S1. All strains were freshly thawed from frozen stocks and grown to mid-log phase ( $OD_{600} \approx 0.5$ ) at 30°C in YPDA medium containing 0.1 mg/ml adenine.

### *S. cerevisiae* strains

*RAD1* and *REV3* deficient strains were constructed by replacing the *RAD1* and *REV3* gene with a *LEU2* cassette and *TRP1* cassette, respectively, using HR-mediated integration of PCR-generated fragments according to standard methods. *S. cerevisiae* yeast strains used were isogenic derivatives of strain *pol2-M644G rnh201 $\Delta$* , *pol2-M644L rnh201 $\Delta$* , *pol3-L612M rnh201 $\Delta$* , *pol3-L612G rnh201 $\Delta$* , *pol1-L868M rnh201 $\Delta$* , *pol1-Y869A rnh201 $\Delta$*  and *rnh201 $\Delta$*  (3). The *rad30-F35A rnh201 $\Delta$  rad1 $\Delta$  rev3 $\Delta$*  and *rad30-F35A rnh201 $\Delta$*  strains are derived from W303 and were previously described (18). The double mutant strains used in the experiments described were obtained by mating appropriate mutants and sporulating the resulting diploids. Freshly derived spore clones were frozen and subsequently genotyped by PCR verification. *RAD1* and *REV3* deficient strains (including *rad30-F35A(1-622)*) were constructed by TopGene Technologies, Inc., Montreal, Canada.

### HydEn-seq protocol

The HydEn-seq protocol was performed essentially as previously described (3). Briefly, mid-log phase yeast cells from

independent cultures were harvested and genomic DNA (gDNA) isolated using the MasterPure Yeast Purification Kit (Epicentre, no RNase A treatment). 1  $\mu$ g of gDNA was hydrolyzed with 0.3 M KOH for 2 h at 55°C. DNA fragments were purified using ethanol precipitation, denatured and phosphorylated with 10 U of 3'-phosphatase-minus T4 polynucleotide kinase (New England BioLabs) for 30 min at 37°C, followed by heat inactivation at 65°C for 20 min. DNA was purified using CleanPCR beads (CleanNA), denatured and ligated to oligo ARC140 overnight at room temperature with 10 U T4 RNA ligase (New England BioLabs), 25% PEG 8000 and 1 mM  $CoCl_3(NH_3)_6$ . After DNA purification with CleanPCR beads and denaturing, the ARC76-ARC77 adaptor was annealed for 5 min at room temperature. Second-strand synthesis was performed using 4 U of T7 DNA polymerase (New England BioLabs) and DNA purified with CleanPCR beads. Libraries were PCR amplified with primer ARC49 and index primers ARC79 to ARC107 using the KAPA HiFi Hotstart ReadyMix (KAPA Biosystems). Libraries were purified with CleanPCR beads and pooled for sequencing. 75-bp paired-end sequencing was performed on an Illumina NextSeq500 instrument, to locate 5'-ends generated by alkaline hydrolysis.

### Sequence trimming, filtering and alignment

Trimming for quality and adaptor sequence of all reads was performed with cutadapt 1.12 (23). Pairs containing one or both reads shorter than 15 nt were discarded. Bowtie 1.2 (24) was used to align Mate 1 of all remaining pairs to the list of index primers used to prepare the libraries; all matching pairs were discarded. All remaining pairs were aligned to the *sacCer3 Saccharomyces cerevisiae* reference genome with bowtie (-v2 -X2000-best). Single-end alignments were then performed for mate 1 of all unaligned pairs (-m1, -v2). The count of 5'-ends of all unique paired-end and single-end alignments was determined for all samples and shifted one base upstream to the location of the hydrolyzed ribonucleotide.

### End count scaling and background subtraction

For visual comparison of individual libraries (Heatmaps and meta-analyses) end counts were normalized to counts per million uniquely mapped reads (divided by the total of uniquely mapped ends and multiplied by 1 000 000). For analyses of strand bias maps requiring weighted averaging of multiple libraries and background subtraction, end counts were scaled as previously described (3). For analyses, where individual replicates were combined to create an average dataset, samples were averaged after the normalization to per million reads mapped. To remove reads stemming from free 5'-ends, scaled end counts from strains with wild type polymerases were subtracted from the steric gate mutants. For the *pol2-M644G* variant, we used the *pol2-M644L rnh201 $\Delta$  rad1 $\Delta$  rev3 $\Delta$*  variant as a control background, while for the pol  $\alpha$  and  $\delta$  variants we used the *rnh201 $\Delta$  rad1 $\Delta$  rev3 $\Delta$*  variants as a control background. For each position where both the forward and reverse strands presented numeric values, the forward strand value was divided by the sum of the forward and reverse strand values. If no reads

were encountered in both the forward and reverse reads, a no value was returned. If the forward read was numeric, but the reverse read was not numeric, then 1 was returned. For the inverse scenario where the reverse read was numeric but the forward read not, then a 0 was returned.

### Statistical analysis

Statistical comparison of the bedgraph files and Braid plots (Supplementary Figures S1 and S2) was performed as follows: All reads in each bedgraph file were binned into 200 bp bins. For every chromosome, a Spearman correlation coefficient was calculated between every binned library. To obtain an average sample  $r$  value, a Fisher transformation was performed after which the mean across all chromosomes for each library was obtained.  $Z$ -scores were converted back into  $r$  values and this was plotted as a heatmap. Coding, plotting and statistics were performed using Pandas, Seaborn and Numpy.

### Meta-analyses and heatmaps

Total counts of the per-strain 5'-ends intersecting same- and opposite-strand bins centered on genomic features of interest (ACS, nucleotide dimers) were determined with custom tools, excluding all mitochondrial and telomeric annotations. For origins, reads from  $-2$  kb to  $+2$  kb around each of the known OriDB ACS sites (using liftOver to sacCer3) were scaled to per million reads and aggregated into bins of 50 bp (Figure 3 and Supplementary Figures S3 and S4). Heatmaps generated with the R 'image' function depict counts in all bins. For origin meta-analyses the sum across all features was then obtained for both sample and control, and sample data was then divided by control data and plotted using R and ggplot2. Reads for each of the 16 possible dimers were gathered, scaled to per million reads mapped and positions 1 and 2 of the dimer were averaged (Figure 4). Reads present in telomeric or ribosomal regions were removed and replicates of the steric gate (*rad30-F35A*), non-steric gate (*RAD30+*) were averaged, and finally the fraction of steric gate reads was divided by the non-steric gate reads and plotted using ggplot2.

To determine whether sequence context was present in the distribution of included ribonucleotides, the number of true dimers present in the reference genome were counted. Dimers were defined as being flanked by bases dissimilar to those bases of the dimer, e.g. dimer TG could have A/C/G at the 5', but not T and A/C/T/ at the 3' end but not G. All other combinations were discarded and dimers in the telomeric and ribosomal regions were removed. The sum of each of the 16 possible dimers was divided by the total number of all true dimers in the genome to obtain a percentage which was plotted as a bar diagram.

To determine whether the pol  $\eta$  steric gate mutant incorporates ribonucleotides at preferred dimer sequences as opposed to purely random incorporation, a paired  $t$ -test was performed between the pol- $\eta$  relative ratio values for the 16 dimers for each lagging, leading and none designated dimer sub-groups against a hypothetical value of 1 for each dimer which would have been the expected value if the steric gate mutant did not incorporate ribonucleotides due to sequence preference.

### Analysis of TA→TG mutations in melanoma

Coordinates of origins of replication were taken from sequence-derived 'N-domains' (25) and lifted over from hg17 to hg19 using liftOver (26), resulting in 1047 origins. Whole genome somatic mutation calls from the Australian Melanoma Genome Project (AMGP) cohort (27), downloaded from the International Cancer Genome Consortium's (ICGC) database (28), were pooled with whole genome mutations from The Cancer Genome Atlas (TCGA) melanoma cohort (29), called as described previously (30). Mutations overlapping with population variants (dbSNP v138) as well as duplicate samples from the same patient were removed, resulting in a total of 221 samples. The number of TA→TG mutations in the samples were counted and aggregated in 100 kb bins around the origins, and strand bias was calculated as the  $\log_2$  ratio between the mutation counts of the top and bottom strands.

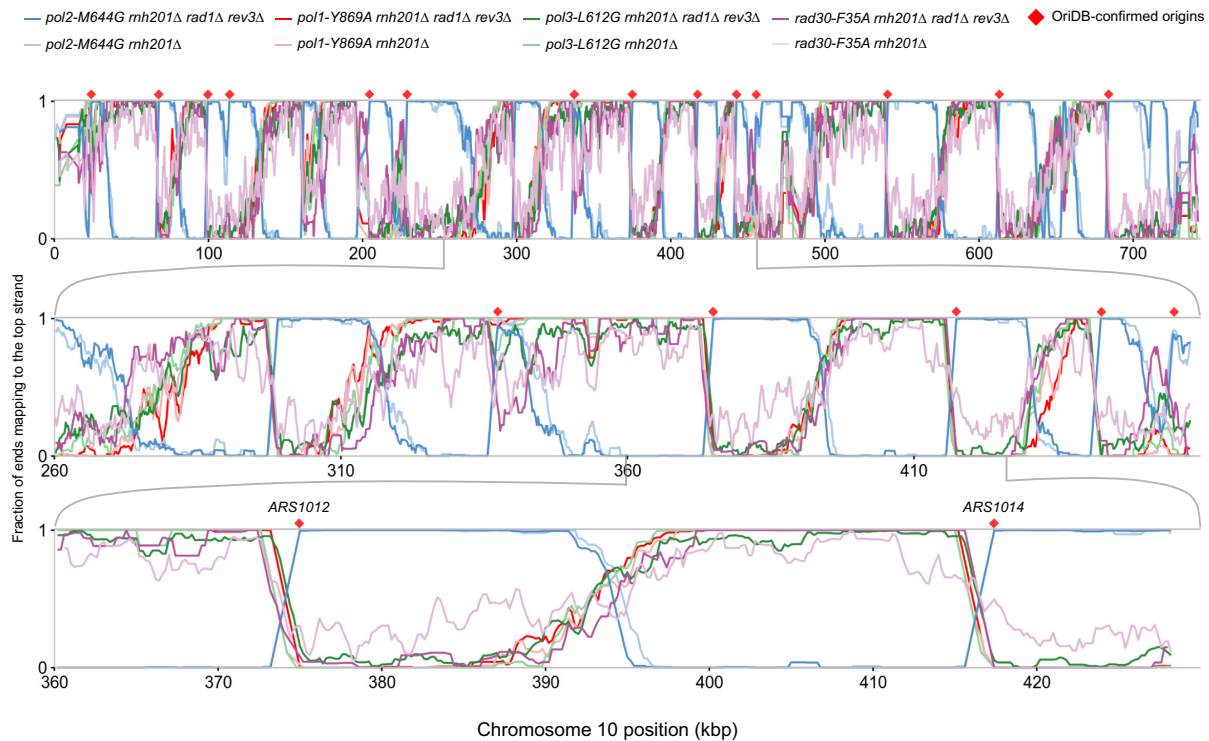
## RESULTS

### Using ribonucleotides to track pol $\eta$ enzymology genome-wide

To understand where pol  $\eta$  may gain access to undamaged DNA and compete with DNA replicases, we utilized a *RNH201* deficient yeast strain expressing the pol  $\eta$ -F35A variant that has an increased propensity to incorporate ribonucleotides into DNA (18). We mapped the locations of incorporated ribonucleotides by HydEn-seq, hydrolyzing the genomic DNA with alkali and sequenced the resulting single strand fragments using 75-bp paired-end sequencing. The reads were aligned to the sacCer3 genome and we determined the location of ribonucleotides genome-wide. We deduced the identity of the ribonucleotides for this strain and calculated the fraction of ends that mapped to the top strand and compared those data to previously published data for the three replicative DNA polymerases, DNA pol  $\alpha$ ,  $\delta$  and  $\epsilon$  (Figure 1, Supplementary Figures S1 and S2) (3). Specifically, we calculated the fraction of generated 5'-ends mapping to the top strand in bins of 200 bp for chromosome 10 to investigate the strand specificity of pol  $\eta$  more closely. A representative graph of chromosome 10 and magnifications are shown in Figure 1. DNA fragments from the *rad30-F35A rnh201*  $\Delta$  strain aligned to the two DNA strands in an alternating pattern (Figure 1, light purple lines). We compared the 5'-end read counts from libraries of the pol  $\eta$  steric gate mutant to those generated by steric gate mutants of replicative pols  $\alpha$ ,  $\delta$  and  $\epsilon$ . We found that the strand specificity switches at origins of replication (Figure 1, red diamonds) and that the alignment weakly resembles the strand-specific patterns of pol  $\delta$  (Figure 1, light green lines) and  $\alpha$  (Figure 1, light red lines), which primarily synthesize the lagging strand and show a behavior contrary to the localization of pol  $\epsilon$  (Figure 1, light blue lines), which primarily synthesizes the leading strand during replication.

### Strand specific tracking of replicative DNA polymerases in the absence of NER and pol $\zeta$

Based upon previous studies in *Escherichia coli*, NER serves as a back-up pathway to remove ribonucleotides incorpo-



**Figure 1.** Strand specific tracking of DNA polymerases  $\epsilon$ ,  $\alpha$ ,  $\delta$  and  $\eta$  on chromosome 10 in yeast deficient or proficient in *RAD1* or *REV3*. Fraction of ends from the *pol2-M644G rnh201Δ rad1Δ rev3Δ* strain mapping to the top strand (blue). Fraction of ends from *pol1-Y869A rnh201Δ rad1Δ rev3Δ* strain mapping to the top strand (red). Fraction of ends from *pol3-L612G rnh201Δ rad1Δ rev3Δ* strain mapping to the top strand (green). Fraction of ends from *rad30-F35A rnh201Δ rad1Δ rev3Δ* strain mapping to the top strand (purple). Light blue, light red, light green and light purple lines correspond to the *RAD1* and *REV3* proficient parent strains. Confirmed origins of replication are depicted as red diamonds. Top: map of chromosome 10 showing the fraction of reads mapping to the top strand in bins of 200 bp. Middle: 200 kb region of chromosome 10. Bottom: 70 kb region of chromosome 10. All lines show average values from three independent libraries.

rated into DNA in the absence of RER (19). For this reason, a yeast strain also deficient in NER (by deletion of *RAD1*) was employed. To potentially increase the access of pol  $\eta$  to undamaged DNA, the strains were also deficient in the low-fidelity DNA pol  $\zeta$ , which plays a major role in promoting spontaneous mutagenesis in yeast. To determine whether this genetic background still reflects the previously described division of labor among replicative polymerases, we used steric gate variants of pols  $\alpha$ ,  $\delta$  and  $\epsilon$  in this genetic background and mapped the location of the incorporated ribonucleotides. The eight investigated strains included one wild-type strain, two variants of pol  $\epsilon$  (*pol2-M644G* and *pol2-M644L*), two variants of pol  $\alpha$  (*pol1-Y869A* and *pol1-L868M*), two variants of pol  $\delta$  (*pol3-L612G* and *pol3-L612M*) and one pol  $\eta$  variant (*rad30-F35A*). Previously, it was shown that in a *RNH201* deficient yeast background, the *pol2-M644G* strain incorporates more ribonucleotides than the wild-type enzyme on the leading strand, the *pol2-M644L* strain incorporates less ribonucleotides than the wild-type enzyme, the *pol1-Y869A* strain incorporates more ribonucleotides than *pol1-L868M* on the lagging strand and the *pol3-L612G* strain also incorporates more ribonucleotides than the *pol3-L612M* strain on the lagging strand (3). A summary of strains and genotypes used in this work can be found in Supplementary Table S1. We analyzed three independent libraries for each of the *RAD1* and *REV3* deficient strains and calculated the

mean Spearman correlation coefficient for each library, for all 16 nuclear chromosomes of the same genotype on both forward and reverse strand (Supplementary Figure S3). Differences between different steric gate mutants of the same polymerase (e.g. pol  $\alpha$ -869A and pol  $\alpha$ -L868M) can be explained by distinct ribonucleotide incorporation signatures, resulting in distinct ratios of incorporated ribonucleotides (Supplementary Table S2) and possibly differing behavior depending on the sequence context, while a high correlation between replicates of the same strain was observed.

We plotted average values from three replicates of chromosome 10 for each of the steric gate variants deficient in *RAD1* and *REV3*: maps of *pol2-M644G*, *pol1-Y869A*, *pol3-L612G* (Figure 1, blue lines, red lines, green lines, respectively) were compared to the maps that were previously made using strains proficient in *RAD1* and *REV3* (Figure 1, light blue lines, light red lines and light green lines, respectively) (3). 5'-ends from the *pol2-M644G rnh201Δ rad1Δ rev3Δ* strain (Figure 1, blue lines) aligned with the *pol2-M644G rnh201Δ* strain (Figure 1, light blue lines). 5'-ends from the *pol1-Y869A rnh201Δ rad1Δ rev3Δ* strain (Figure 1, red lines) aligned with the *pol1-Y869A rnh201Δ* strain (Figure 1, light red lines). 5'-ends from the *pol3-L612G rnh201Δ rad1Δ rev3Δ* strain (Figure 1, green lines) aligned with 5'-ends from the *pol3-L612G rnh201Δ* strain (Figure 1, light green lines). These observations were also made for the two other variants of pol  $\alpha$  and  $\delta$  (Sup-

plementary Figure S4): The 5'-ends from the *pol1-L868M rnh201Δ rad1Δ rev3Δ* strain (red lines) aligned with the 5'-ends from the *pol1-L868M rnh201Δ* strain (light red lines) and the 5'-ends from the *pol3-L612M rnh201Δ rad1Δ rev3Δ* strain (green lines) aligned with the 5'-ends from the *pol3-L612M rnh201Δ* strain (light green lines). Equally, the Spearman correlation coefficients of these libraries reflect this behavior (Supplementary Figures S3 and S5). Interestingly, the fraction of ribonucleotides incorporated by pol  $\eta$  showed a higher strand-specificity for the lagging-strand in the *RADI REV3* deficient background (purple lines). Based on these observations, it is possible that pol  $\eta$  may also be involved in some leading strand synthesis during normal DNA synthesis, when pol  $\zeta$  is present. The contribution of pol  $\eta$  and  $\zeta$  on the leading strand maybe be necessary at difficult to replicate regions where pol  $\epsilon$  may stall. Further studies are needed to clarify the role of  $\eta$  and  $\zeta$  on the leading strand (such as a mediating role for pol  $\zeta$ ), for example by tracking steric gate variants of DNA polymerase  $\zeta$  when pol  $\eta$  is absent or present. Alternatively, a bias of NER for the lagging strand, as was recently found in human time series excision repair sequencing data (31), could contribute to a lower level of detectable ribonucleotides on the lagging strand in the *RADI REV3* proficient background, therefore leading to a less clear detection of pol  $\eta$  in this genetic background. It also should be noted that it is possible that DNA synthesis on the lagging strand by pol  $\eta$  may be less frequent in a wild type strain. It is conceivable that the lagging strand specificity observed in this study may be due to strand-displacement DNA synthesis of DNA polymerase  $\eta$  during Okazaki fragment maturation, an activity that has been observed on homologous recombination, e.g. in humans (13) rather than during replication.

To assess pol  $\eta$  usage genome-wide in the *RADI REV3* deficient background, 5'-end read counts of the pol  $\eta$  steric gate mutant were plotted for all 16 nuclear chromosomes and the same alternating pattern of alignment to both strands was observed. Figure 2 shows representative graphs for all chromosomes of all four polymerases combined (note that the fraction of ends mapping to the top strand was used for pol  $\epsilon$ , while the fraction mapping to the bottom strand was used for the other polymerases. Again, pol  $\eta$  acts on the lagging strand, following the pattern of pol  $\alpha$  and  $\delta$ . Taken together, these results support the idea of strand-specific pol  $\eta$  usage and its coupling to lagging strand replication in an undamaged cell.

We conclude that deletion of *RADI* and *REV3* does not disrupt the division of labor previously observed for replicative polymerases during normal growth. This indicates that, in a yeast strain deficient in *RADI* and *REV3*, pol  $\eta$  primarily synthesizes DNA on the lagging strand, where it competes with pols  $\alpha$  and  $\delta$  (Figure 1, bottom panel). Furthermore, these data suggest a direct role of pol  $\eta$  in general lagging strand DNA synthesis.

### Pol $\eta$ usage at replication origins

Previously, it has been shown that switching between DNA strands of replicative polymerases occurs over several hundred base pairs centered on the autonomously replicating sequence (ARS) consensus sequence (ACS) (3). Moreover,

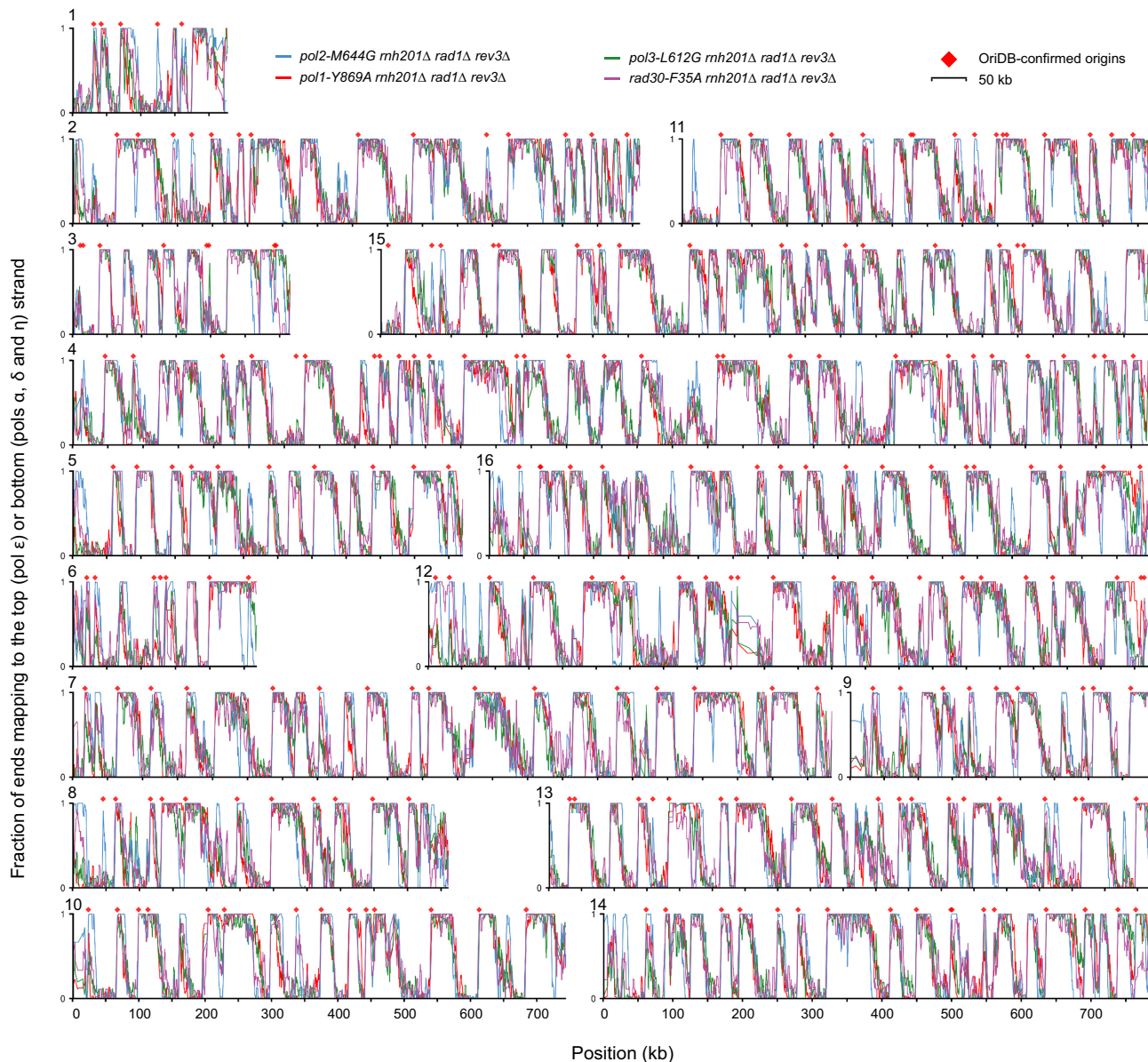
the site of the observed transitions of pols  $\alpha$ ,  $\delta$  and  $\epsilon$  between the two strands, coincide with confirmed replication origins (Figures 1 and 2, Supplementary Figures S1 and S4, red diamonds) (32). Similarly, we also made this observation for pol  $\eta$ : Heat maps (Figure 3, upper panels) and meta-analysis (Figure 3, lower panels) of 5'-ends in 50-bp bins near ACS revealed that pol  $\epsilon$  has a relatively sharp transition between the strands, while the transition for pol  $\alpha$  and  $\delta$  seems less rapid. We also observed a pattern for the transition of pol  $\eta$  at ACS that shows the same directionality as the transitions of pol  $\alpha$  and  $\delta$ . It is however less sharp than for pol  $\alpha$  and  $\delta$ . This observation of strand transition alongside pol  $\alpha$  and  $\delta$  further supports the idea of pol  $\eta$  being linked to the lagging strand replication.

Lagging strand synthesis involves the synthesis of thousands of Okazaki fragments and each 3'-end may be a substrate for pol  $\eta$ . Moreover, pol  $\eta$  may be recruited to the lagging strand by PCNA or Rev1 via the C-terminus of pol  $\eta$ , which contains a PIP and Rev1-interacting region (RIR). To determine if pol  $\eta$  is recruited to the lagging strand through this domain, we deleted the ten C-terminal amino acid residues containing the PIP and RIR motif (33,34) in pol  $\eta$ -F35A and tracked where this polymerase variant, *rad30-F35A(1-622)*, incorporated ribonucleotides. We plotted the strand bias across origins of replication and this variant did not display any lagging strand preference, thus we conclude that pol  $\eta$  is recruited to the lagging strand through its 10 C-terminal amino acid residues containing the conserved PIP and RIR motifs (Figure 3).

To characterize how pol  $\eta$ -F35A incorporates ribonucleotides *in vivo*, we wanted to determine if pol  $\eta$ -F35A incorporates single ribonucleotides, or multiple consecutive ribonucleotides, by replacing the *rnh201Δ* allele with the *rnh201-P45D-Y219A* allele. This allele retains the ability to nick at polyribonucleotide tracts, but not at single ribonucleotides (18). Reads were again centred across a 4-kb window of the origins and we determined the strand specificity of pol  $\eta$ -F35A, when only stretches of ribonucleotides are repaired (Figure 3). We observed that the lagging strand specificity was completely abolished and in contrast, we observed an increase in ribonucleotides in the region of 0–0.5 kb on the top strand and –0.5 to 0 kb on the bottom strand. This suggests that the majority of ribonucleotides incorporated by  $\eta$ -F35A are short stretches of multiple consecutive ribonucleotides.

### Sequence context-dependent template usage by pol $\eta$

To further characterize the behavior of pol  $\eta$ -F35A, we identified the dinucleotide sequences containing incorporated ribonucleotides in the *rad30-F35A rnh201Δ rad1Δ rev3Δ* strain and compared it to the *rnh201Δ rad1Δ rev3Δ* strain. Based on the strand bias maps for the three replicative polymerases, we binned the dinucleotide template sequences corresponding to lagging strand sequences (Figure 4, upper panel), neither lagging strand nor leading strand (Figure 4, middle panel), or leading strand dinucleotide sequences (Figure 4, lower panel). Again, we found that pol  $\eta$ -F35A incorporates more ribonucleotides on the lagging strand as signified by ratios greater than 1 (Figure 4, upper panel), while ratios below 1 were calculated for the leading



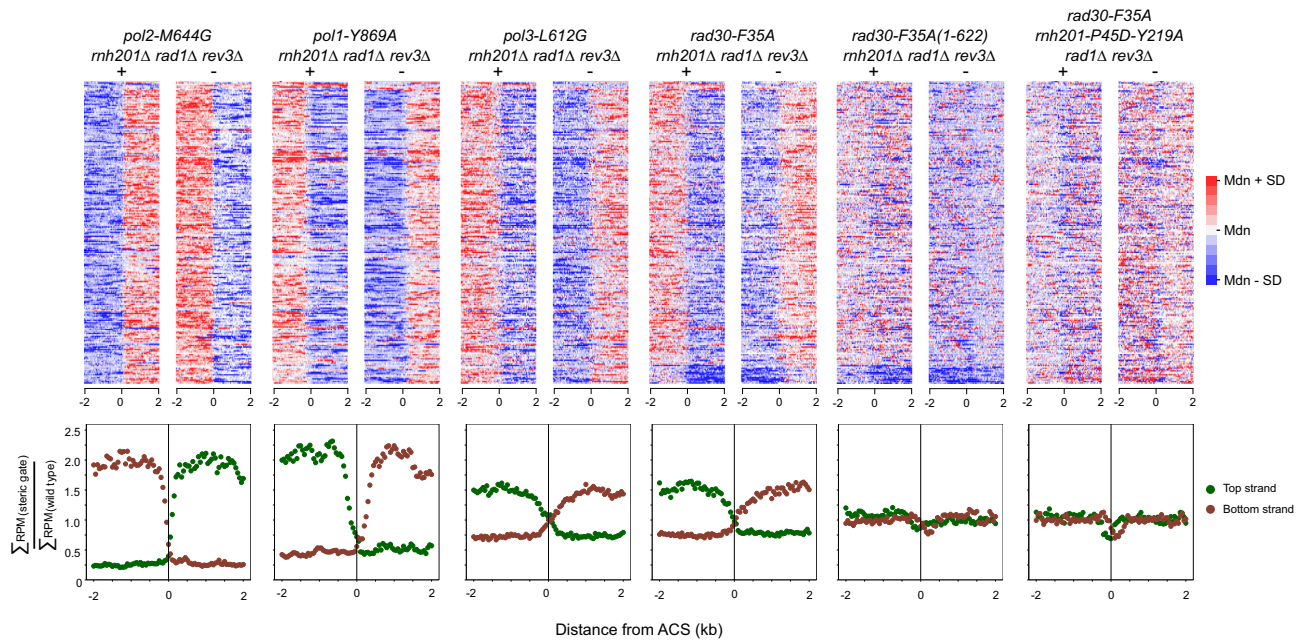
**Figure 2.** Genome-wide tracking of DNA polymerases  $\epsilon$ ,  $\alpha$ ,  $\delta$  and  $\eta$  in yeast deficient in *RAD1* and *REV3*. Fraction of reads mapping to the top (*pol2-M644G rnh201Δ rad1Δ rev3Δ* strain, (blue)) or bottom (*pol1-Y869A rnh201Δ rad1Δ rev3Δ* strain (red), *pol3-L612G rnh201Δ rad1Δ rev3Δ* strain (green) and *rad30-F35A rnh201Δ rad1Δ rev3Δ* (purple)) strand genome-wide. Confirmed origins of replication are depicted as red diamonds. All lines show average values from 3 independent libraries.

strand (Figure 4, lower panel). Without any sequence context dependence, the expected enrichment of dimers compared to the observed dimer bases on the DNA context (Supplementary Figure S6) will result in a ratio of 1 and a paired two-tailed *t*-test comparing the observed versus expected enrichment factors gives *p*-values of  $3.8 \times 10^{-6}$ ,  $3.1 \times 10^{-2}$ ,  $5.1 \times 10^{-16}$  for dimers replicated as lagging, none and leading strand, respectively. Of the 16 possible dimers, the seven with greatest increase in reads all contained a template T. Most ribonucleotide incorporation was detected for a T–T template, where a 2.2-fold increase in reads was observed. Assuming that consecutive rAs can be incorporated opposite template T–T as seen *in vitro* (18), the ratio for rA–rA incorporation may be underestimated

by our method, because it does not allow a distinction between one, or multiple consecutive ribonucleotides.

### Deletion of T:A base pairs in *CAN1*

Previously, spontaneous mutation types caused by pol  $\eta$ -F35A were analyzed using the *CAN1* forward mutation assay (18). Expression of *rad30-F35A* in the *rnh201Δ rad1Δ rev3Δ* background increased the overall mutation rate 2.5-fold compared to the *RAD30 rnh201Δ rad1Δ rev3Δ* strain. The overall rate of one base pair substitutions and 2–5 base pair deletions are similar, but surprisingly there was a 17-fold increase in mutation rate for single base pair deletions. Single base pair deletions were observed 73 times at 7 differ-



**Figure 3.** Tracking of DNA polymerases  $\epsilon$ ,  $\alpha$ ,  $\delta$  and  $\eta$  near replication origins in *rad1* $\Delta$  and *rev3* $\Delta$  strains. Heat maps (upper panel) for the top (+) and bottom (–) strands of the nuclear genome for steric gate variants of pol  $\epsilon$ ,  $\alpha$ ,  $\delta$ ,  $\eta$ ,  $\eta$  (1–622) and pol  $\eta$  with *rnh201-P45D-Y219A* allele. Counts are scaled per million reads and centred across a 4 kb window of the yeast replication origins (ACS (32)). Meta-analysis (lower panels) of strand specific ribonucleotides in steric gate mutant strains of pol  $\epsilon$ ,  $\alpha$ ,  $\delta$  and  $\eta$  at ACS, scaled as reads per million (RPM) and in bins of 50 bp.

ent positions (positions 165–167, 387–389, 540–543, 936–939, 1022–1025, 1048, Figure 5A) on the *CAN1* gene. These deletion events corresponded to 69 T deletions, 3 G deletions and 1 C deletion on the coding strand. The 69 T deletions can either result from deletion of a template T on the coding strand, or deletion of template A on the non-coding strand. When the pol  $\eta$ -F35A variant incorporated ribonucleotides *in vitro*, only the incorporation opposite a template T was efficient (18). Using single base pair deletion events in the *CAN1* gene as a marker of replication enzymology combined with the *in vitro* assays demonstrating that the steric gate variant preferably uses T as a template, suggests that the T:A base pair deletions are a result of pol  $\eta$  replicating A on the non-coding strand, but these data do not show if this strand bias is related to specific leading, or lagging, strand synthesis.

*CAN1* is located on chromosome 5 at position 31,694–33,466. A confirmed origin, ARS504.2, is located 20 kb upstream of the *CAN1* reporter gene, while the closest confirmed origin downstream is ARS507 and located 26 kb away. Based on the two origins' distance from the *CAN1* reporter gene, it is not possible to determine if the non-coding strand is replicated as the leading, or lagging, strand. To determine on which strand pol  $\eta$  is replicating *CAN1* using ribonucleotides as a marker of replication enzymology, we used the HydEn-seq method and plotted the fraction of reads that mapped to the top strand for *rad30-F35A rnh201* $\Delta$  *rad1* $\Delta$  *rev3* $\Delta$  in a region spanning 0–80 kb of chromosome 5 (Figure 5B). From the HydEn-seq data, we observe that at ARS507, pol  $\eta$ -F35A demonstrates an abrupt change in strand preference, while at ARS504.2, pol  $\eta$  does not demonstrate a clear transition between the strands. As a consequence of the clear, abrupt change of strand prefer-

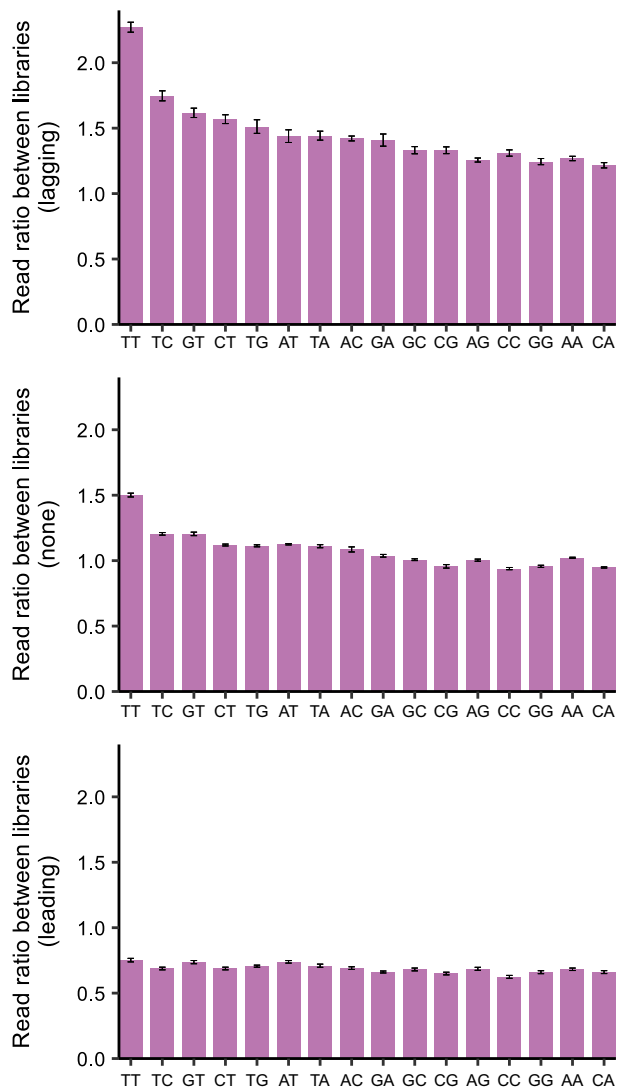
ence at ARS507, we observe that the non-coding strand of *CAN1* is replicated by pol  $\eta$ -F35A. These data imply that single base pair deletions found in the *CAN1* reporter gene are a result of ribonucleotide incorporation by pol  $\eta$ -F35A on the non-coding strand and HydEn-seq data confirms that pol  $\eta$  is replicating the non-coding strand of *CAN1* as the lagging strand.

### Strand preference for pol $\eta$ induced mutations in melanomas

To investigate if the strand bias of pol  $\eta$  found in *S. cerevisiae* is evolutionarily conserved in humans, we studied the replication strand bias of likely pol  $\eta$ -induced mutations in melanoma. Human pol  $\eta$  preferentially causes WA $\rightarrow$ WG transitions that arise when a template T is mispaired with a G (35,36). Among these transitions, the TA $\rightarrow$ TG substitutions are particularly useful when assessing strand asymmetry since this dinucleotide is present in equal numbers on both strands. Recently, the majority of substitutions at A:T base pairs in melanoma have been estimated to be caused by pol  $\eta$  acting on undamaged DNA (37). In a cohort of 221 melanomas subjected to whole genome sequencing (27,29), we identified 313 751 TA $\rightarrow$ TG mutations and they showed a weak strand bias around predicted origins of replication (25) consistent with a likely lagging strand preference for human pol  $\eta$  (Figure 5C). This finding was recently further supported in the analysis of damage sequencing data in UV-treated human fibroblasts, that suggested TLS of bulky lesions is giving rise to mutations during replication (31).

## DISCUSSION

DNA polymerase  $\eta$  is best characterized as a TLS polymerase. *In vitro* studies in *S. cerevisiae* suggest that pol  $\eta$



**Figure 4.** Dinucleotide template sequences associated with HydEn-seq sequencing reads of pol  $\eta$ -F35A. Bars show the relative ratio of read counts from the *rad30-F35A rnh201 $\Delta$  rad1 $\Delta$  rev3 $\Delta$*  strain compared to the *rnh201 $\Delta$  rad1 $\Delta$  rev3 $\Delta$*  strain by template dimer from either the lagging strand (upper panel), neither leading nor lagging strand (middle panel) or leading strand (lower panel). Values are from three independent libraries. Error bars show the standard error of the mean.

has an ability to bypass a variety of DNA lesions including *cis-syn* T–T dimers (7,8), C–C and T–C photoproducts (38) and 7,8-dihydro-8-oxoguanine (8-oxoG) (39,40). Moreover, *in vivo* studies support pol  $\eta$ 's relevance in the bypass of *cis-syn* T–T dimers (41,42) and 8-oxoG (43). In addition to its role in specific lesion bypass, pol  $\eta$  has been shown to be involved in other cellular processes, such as somatic hypermutation, replication of common fragile sites and maintenance of telomere DNA (12,17,37).

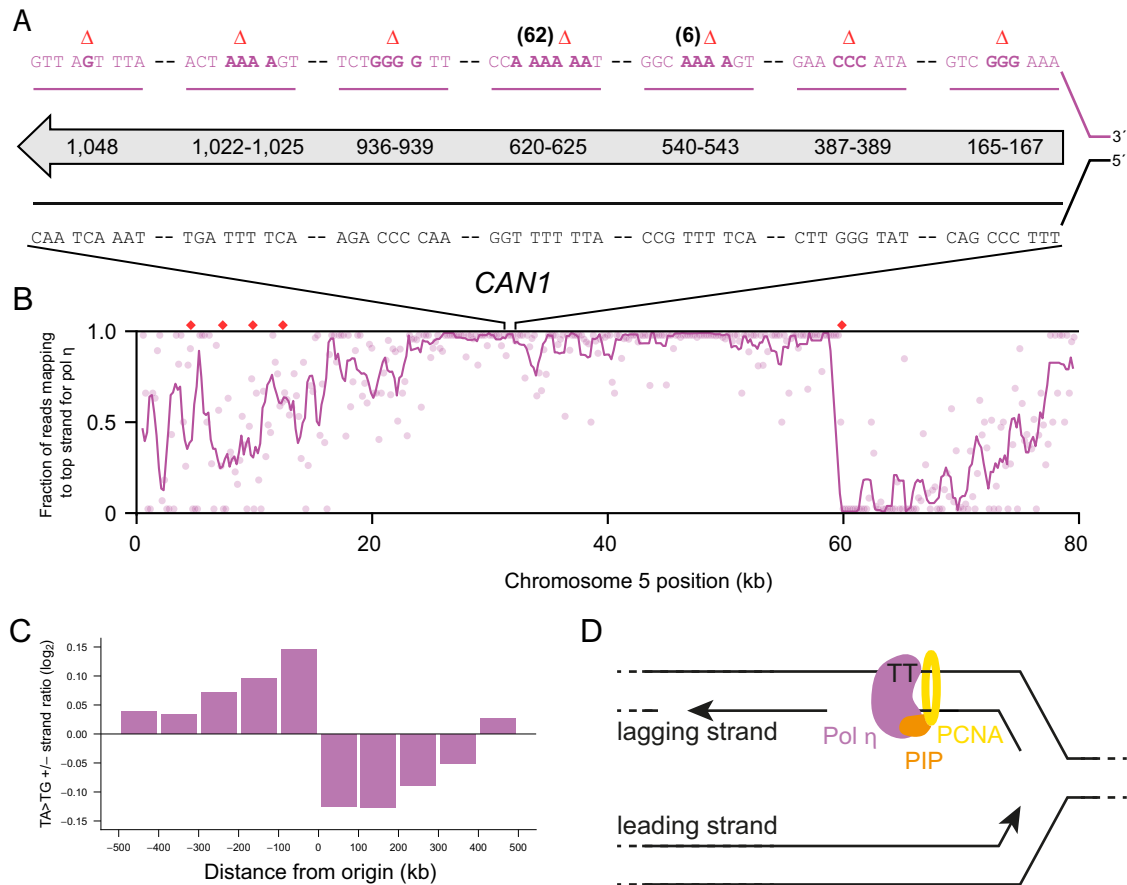
As previously shown for the whole yeast genome, replicative polymerases perform DNA synthesis in a pattern reflecting the course of leading and lagging strand synthesis during replication (3). The results of our current study illustrate that under normal growth conditions and in the ab-

sence of NER and the TLS polymerase  $\zeta$ , pol  $\eta$  is mainly active on the lagging strand (Figure 5D) and the recruitment of pol  $\eta$  to the lagging strand has a mutational cost in humans (Figure 5C). The observed increase in rA incorporation at undamaged template T–T is most likely related to pol  $\eta$ 's main role in accurately replicating past *cis-syn* T–T dimers (7). The lagging strand specificity of pol  $\eta$  is consistent with studies of the related Y-family TLS polymerases, pol IV and pol V, in *E. coli* (reviewed in (44)). In this case, pol IV and pol V may gain access to abandoned primer termini generated by pol I during Okazaki fragment processing, and thereby become recruited to the lagging strand (45–47). Our results are also consistent with studies in budding yeast, where pol  $\eta$  was mainly found to be involved in bypass of 8-oxoG lesions on the lagging strand in a forward mutation assay (48). Our findings are furthermore consistent with structure-function studies of pol  $\eta$ , where it was demonstrated that the PIP motif in pol  $\eta$  is essential for interaction with proliferating cell nuclear antigen (PCNA) (33). Accessibility of specialized TLS polymerases to undamaged DNA therefore appears to be conserved throughout evolution and TLS polymerases may have increased accessibility at lagging strand templates which are not efficiently repaired by NER.

Our findings suggest that the concept of the division of labor among replicative DNA polymerases may extend to the specialized DNA pol  $\eta$ . It seems reasonable to hypothesize that this is connected to the different mechanisms of leading and lagging strand synthesis: After pol  $\alpha$  has synthesized a primer, pol  $\delta$  performs the majority of DNA synthesis of the Okazaki fragment. In contrast to the lagging strand mechanism, leading strand replication involves pol  $\epsilon$ , which performs the majority of continuous DNA synthesis. Current models suggest that the stalling of pol  $\epsilon$  still allows continuous unwinding of the template strands so that lagging strand synthesis may continue (5). Uncoupling of helicase and pol  $\epsilon$  may thus expose a relatively large stretch of single stranded DNA. Post-replicative gap filling by specialized TLS polymerases other than pol  $\eta$ , or template switching, are strategies to tolerate such lesions. In both cases, the extension after successful bypass of the replication stalling lesion may be performed more efficiently by another (TLS) DNA polymerase. The genetic background of the strains investigated in our study suggests, however, that pol  $\zeta$  is not essential for this process.

Why the recruitment of pol  $\eta$  to the lagging strand is facilitated remains to be fully elucidated. The ring-shaped homotrimeric PCNA, a protein belonging to the sliding clamp family (49), was identified as an auxiliary protein of pol  $\delta$  (50). It plays a role in modulating replication fork assembly and replicative polymerase activities (51–53). It serves as a platform to mediate a number of cellular processes (54) including DNA repair (55) as well as an inheritance and formation of chromatin structures (56,57). Furthermore, links between PCNA and TLS have been discovered (58,59) and PCNA may act as a ‘tool belt’ to facilitate polymerase switching. Our data supports the idea, that when pol  $\delta$  is stalled, it may be exchanged for pol  $\eta$  via its PIP motif (58). Pol  $\eta$  can then bind to the template DNA and bypass





**Figure 5.** Strand specific mapping of pol  $\eta$  mutational signatures using T–A deletions in yeast or TA→TG mutations in humans. (A) Preferential replication of the lagging strand template by pol  $\eta$ -F35A. The orientation of the *CAN1* reporter is indicated by the direction of the arrow. Deletion hotspots are identified numerically within the arrow. The deletion hotspot sequences are shown in bold. (B) Fraction of reads mapping to the top strand on chromosome 5 from 0–80 kb and origins of replication around *CAN1*. (C) Ratio of likely pol  $\eta$ -induced mutations (TA>TG) between the top and bottom strand in 100 kb bins around 1047 origins of replication in human melanomas ( $n = 221$ ). (D) Model of lagging strand specific recruitment of pol  $\eta$ . Pol  $\eta$  is recruited to the lagging strand through the PCNA binding motif (PIP) and has a preference for T–T dinucleotide sequences in the template.

replication-blocking lesions, thereby providing tolerance to DNA damage. In particular, PCNA was shown to physically interact with the C-terminus of pol  $\eta$ , to stimulate its activity together with replication factor C and replication protein A *in vitro* and to be essential for pol  $\eta$ -mediated UV resistance *in vivo* (33). Whether the ubiquitination status of PCNA has an effect on the TLS efficiency or if it is merely relevant to the recruitment and polymerase switch remains controversial (59,60). It would be interesting to track the genome-wide effects of PCNA dysregulation on *in vivo* polymerase utilization. It is possible that pol  $\eta$ 's strand-specific usage would switch to also include the leading strand when cells are challenged with UV-light, DNA-modifying agents, or when the competing TLS polymerase, pol  $\zeta$ , is available. These possibilities can be determined tracking enzyme activity through ribonucleotide incorporation as outlined here.

#### DATA AVAILABILITY

The sequencing data has been deposited in the Gene Expression Omnibus database under accession no.

GSE110241. Custom scripts are available upon request to the corresponding author.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

#### ACKNOWLEDGEMENTS

We would like to thank the GeneCoreSU NGS platforms, at the Sahlgrenska Academy and Sahlgrenska Hospital.

*Authors contributions:* K.K., J.K. and C.N. performed experiments. K.K., M.K.M.E., L.J.T., M.B., E.L., R.W. and A.R.C. analyzed data. R.W. and A.R.C. designed experiments. J.P.M. provided reagents. All authors were involved in data analysis. K.K. and A.R.C. wrote the manuscript with input from all authors. Correspondence and requests for materials should be addressed to anders.clausen@gu.se.

#### FUNDING

Swedish Research Council ([www.vr.se](http://www.vr.se)) [2014-6466 to A.R.C.]; Swedish Foundation for Strategic Research ([www.sfr.se](http://www.sfr.se)).

stratresearch.se) [ICA14-0060 to A.R.C.]; Genomics Core Facility at University of Gothenburg for sequencing; National Institute of Child Health and Human Development, National Institutes of Health, Intramural Research Program (to R.W. and J.P.M.); Chalmers University of Technology provided financial support (to M.K.M.E.); Swedish Research Council (to E.L.); Swedish Cancer Foundation (to E.L.); Knut and Alice Wallenberg Foundation (to E.L.); Swedish Foundation for Strategic Research (to E.L.); The results published here are in part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI, and upon data generated by the Australian Melanoma Genome Project and hosted by the International Cancer Genome Consortium. We are most grateful to the patients, investigators, clinicians, technical personnel, and funding bodies who contributed to these projects, thereby making this study possible. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Funding for open access charge: Vetenskapsrådet.  
*Conflict of interest statement.* None declared.

## REFERENCES

- Burgers,P.M.J. and Kunkel,T.A. (2017) Eukaryotic DNA replication fork. *Ann. Rev. Biochem.*, **86**, 417–438.
- Pursell,Z.F., Isoz,I., Lundstrom,E.B., Johansson,E. and Kunkel,T.A. (2007) Yeast DNA polymerase  $\epsilon$  participates in leading-strand DNA replication. *Science*, **317**, 127–130.
- Clausen,A.R., Lujan,S.A., Burkholder,A.B., Orebaugh,C.D., Williams,J.S., Clausen,M.F., Malc,E.P., Mieczkowski,P.A., Fargo,D.C., Smith,D.J. *et al.* (2015) Tracking replication enzymology *in vivo* by genome-wide mapping of ribonucleotide incorporation. *Nat. Struct. Mol. Biol.*, **22**, 185–191.
- McCulloch,S.D., Kokoska,R.J., Chilkova,O., Welch,C.M., Johansson,E., Burgers,P.M.J. and Kunkel,T.A. (2004) Enzymatic switching for efficient and accurate translesion DNA replication. *Nucleic Acids Res.*, **32**, 4665–4675.
- Hedglin,M. and Benkovic,S.J. (2017) Eukaryotic translesion DNA synthesis on the leading and lagging Strands: Unique detours around the same obstacle. *Chem. Rev.*, **117**, 7857–7877.
- Sale,J.E., Lehmann,A.R. and Woodgate,R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.*, **13**, 141–152.
- Johnson,R.E., Prakash,S. and Prakash,L. (1999) Efficient bypass of a thymine–thymine dimer by yeast DNA polymerase, pol  $\eta$ . *Science*, **283**, 1001–1004.
- Masutani,C., Kusumoto,R., Yamada,A., Dohmae,N., Yokoi,M., Yuasa,M., Araki,M., Iwai,S., Takio,K. and Hanaoka,F. (1999) The *XPV* (xeroderma pigmentosum variant) gene encodes human DNA polymerase  $\eta$ . *Nature*, **399**, 700–704.
- Kim,S., Davis,M., Sinn,E., Patten,P. and Hood,L. (1981) Antibody diversity: Somatic hypermutation of rearranged VH genes. *Cell*, **27**, 573–581.
- Di Nola,J.M. and Neuberger,M.S. (2007) Molecular mechanisms of antibody somatic hypermutation. *Ann. Rev. of Biochem.*, **76**, 1–22.
- Rey,L., Sidorova,J.M., Puget,N., Boudsocq,F., Biard,D.S.F., Monnat,R.J., Cazaux,C. and Hoffmann,J.S. (2009) Human DNA polymerase  $\eta$  is required for common fragile site stability during unperturbed DNA replication. *Mol. Cell Biol.*, **29**, 3344–3354.
- Bergoglio,V., Boyer,A.S., Walsh,E., Naim,V., Legube,G., Lee,M., Rey,L., Rosselli,F., Cazaux,C., Eckert,K.A. *et al.* (2013) DNA synthesis by pol  $\eta$  promotes fragile site stability by preventing under-replicated DNA in mitosis. *J. Cell Biol.*, **201**, 395–408.
- McIlwraith,M.J., Vaisman,A., Liu,Y.L., Fanning,E., Woodgate,R. and West,S.C. (2005) Human DNA polymerase  $\eta$  promotes DNA synthesis from strand invasion intermediates of homologous recombination. *Mol. Cell*, **20**, 783–792.
- Sneeden,J.L., Grossi,S.M., Tappin,I., Hurwitz,J. and Heyer,W.D. (2013) Reconstitution of recombination-associated DNA synthesis with human proteins. *Nucleic Acids Res.*, **41**, 4913–4925.
- Enervald,E., Lindgren,E., Katou,Y., Shirahige,K. and Strom,L. (2013) Importance of pol  $\eta$  for damage-induced cohesion reveals differential regulation of cohesion establishment at the break site and genome-wide. *PLoS Genet.*, **9**, 17.
- Garcia-Exposito,L., Bourneque,E., Bergoglio,V., Bose,A., Barroso-Gonzalez,J., Zhang,S.F., Roncaioli,J.L., Lee,M., Wallace,C.T., Watkins,S.C. *et al.* (2016) Proteomic Profiling reveals a specific role for translesion DNA polymerase  $\eta$  in the alternative lengthening of telomeres. *Cell Rep.*, **17**, 1858–1871.
- Pope-Varsalona,H., Liu,F.J., Guzik,L. and Opreko,P.L. (2014) Polymerase  $\eta$  suppresses telomere defects induced by DNA damaging agents. *Nucleic Acids Res.*, **42**, 13096–13109.
- Donigan,K.A., Cerritelli,S.M., McDonald,J.P., Vaisman,A., Crouch,R.J. and Woodgate,R. (2015) Unlocking the steric gate of DNA polymerase  $\eta$  leads to increased genomic instability in *Saccharomyces cerevisiae*. *DNA Repair (Amst.)*, **35**, 1–12.
- Vaisman,A., McDonald,J.P., Huston,D., Kuban,W., Liu,L.L., Van Houten,B. and Woodgate,R. (2013) Removal of misincorporated ribonucleotides from prokaryotic genomes: An unexpected role for nucleotide excision repair. *PLoS Genet.*, **9**, 15.
- Quah,S.K., Vonborstel,R.C. and Hastings,P.J. (1980) The origin of spontaneous mutation in *Saccharomyces cerevisiae*. *Genetics*, **96**, 819–839.
- Harfe,B.D. and Jinks-Robertson,S. (2000) DNA polymerase  $\zeta$  introduces multiple mutations when bypassing spontaneous DNA damage in *Saccharomyces cerevisiae*. *Mol. Cell*, **6**, 1491–1499.
- Livneh,Z., Ziv,O. and Shachar,S. (2010) Multiple two-polymerase mechanisms in mammalian translesion DNA synthesis. *Cell Cycle*, **9**, 729–735.
- Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, **17**, 10–12.
- Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
- Huvet,M., Nicolay,S., Touchon,S., Audit,B., d'Aubenton-Carafa,Y., Arneodo,A. and Thermes,C. (2007) Human gene organization driven by the coordination of replication and transcription. *Genome Res.*, **17**, 1278–1285.
- Hinrichs,A.S., Karolchik,D., Baertsch,R., Barber,G.P., Bejerano,G., Clawson,H., Diekhans,M., Furey,T.S., Harte,R.A., Hsu,F. *et al.* (2006) The UCSC genome browser Database: Update 2006. *Nucleic Acids Res.*, **34**, D590–D598.
- Hayward,N.K., Wilmott,J.S., Waddell,N., Johansson,P.A., Field,M.A., Nones,K., Patch,A.M., Kakavand,H., Alexandrov,L.B., Burke,H. *et al.* (2017) Whole-genome landscapes of major melanoma subtypes. *Nature*, **545**, 175–180.
- Livneh,Z., Anderson,W., Aretz,A., Barker,A.D., Bell,C., Bernabe,R.R., Bhan,M.K., Calvo,F., Eerola,I., Gerhard,D.S. *et al.* (2010) International network of cancer genome projects. *Nature*, **464**, 993–998.
- Akbani,R., Akdemir,K.C., Aksoy,B.A., Albert,M., Ally,A., Amin,S.B., Arachchi,H., Arora,A., Auman,J.T., Ayala,B. *et al.* (2015) Genomic classification of cutaneous melanoma. *Cell*, **161**, 1681–1696.
- Fredriksson,N.J., Ny,L., Nilsson,J.A. and Larsson,E. (2014) Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nat. Genet.*, **46**, 1258–1263.
- Septyarskiy,V.B., Akkuratov,E.E., Akkuratova,N., Andrianova,M.A., Nikolaev,S.I., Bazykin,G.A., Adameyko,I. and Sunyaev,S.R. (2019) Error-prone bypass of DNA lesions during lagging-strand replication is a common source of germline and cancer mutations. *Nat. Genet.*, **51**, 36–41.
- Siow,C.C., Nieduszynska,S.R., Muller,C.A. and Nieduszynski,C.A. (2012) OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res.*, **40**, D682–D686.
- Haracska,L., Kondratyck,C.M., Unk,I., Prakash,S. and Prakash,L. (2001) Interaction with PCNA is essential for yeast DNA polymerase  $\eta$  function. *Mol. Cell*, **8**, 407–415.
- Boehm,E.M., Powers,K.T., Kondratyck,C.M., Spies,M., Houtman,J.C.D. and Washington,M.T. (2016) The proliferating cell nuclear antigen (PCNA)-interacting Protein (PIP) motif of DNA

- polymerase mediates its interaction with the C-terminal domain of Rev1. *J. Biol. Chem.*, **291**, 8735–8744.
35. Matsuda, T., Bebenek, K., Masutani, C., Hanaoka, F. and Kunkel, T.A. (2000) Low fidelity DNA synthesis by human DNA polymerase  $\eta$ . *Nature*, **404**, 1011–1013.
  36. Rogozin, I.B., Pavlov, Y.I., Bebenek, K., Matsuda, T. and Kunkel, T.A. (2001) Somatic mutation hotspots correlate with DNA polymerase  $\eta$  error spectrum. *Nat. Immunol.*, **2**, 530–536.
  37. Supek, F. and Lehner, B. (2017) Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. *Cell*, **170**, 534–547.
  38. Yu, S.L., Johnson, R.E., Prakash, S. and Prakash, L. (2001) Requirement of DNA polymerase  $\eta$  for error-free bypass of UV-induced CC and TC photoproducts. *Mol. Cell. Biol.*, **21**, 185–188.
  39. Haracska, L., Yu, S.L., Johnson, R.E., Prakash, L. and Prakash, S. (2000) Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase  $\eta$ . *Nat. Genet.*, **25**, 458–461.
  40. McCulloch, S.D., Kokoska, R.J., Garg, P., Burgers, P.M. and Kunkel, T.A. (2009) The efficiency and fidelity of 8-oxo-guanine bypass by DNA polymerases  $\delta$  and  $\eta$ . *Nucleic Acids Res.*, **37**, 2830–2840.
  41. Abdulovic, A.L. and Jinks-Robertson, S. (2006) The *in vivo* characterization of translesion synthesis across UV-induced lesions in *Saccharomyces cerevisiae*: Insights into pol  $\zeta$ - and pol  $\eta$ -dependent frameshift mutagenesis. *Genetics*, **172**, 1487–1498.
  42. Gibbs, P.E.M., McDonald, J., Woodgate, R. and Lawrence, C.W. (2005) The relative roles *in vivo* of *Saccharomyces cerevisiae* pol  $\eta$ , pol  $\zeta$ , Rev1 protein and Pol32 in the bypass and mutation induction of an abasic site, T–T (64) photoadduct and T–T *cis-syn* cyclobutane dimer. *Genetics*, **169**, 575–582.
  43. Rodriguez, G.P., Song, J.B. and Crouse, G.F. (2013) *In vivo* bypass of 8-oxodG. *PLoS Genet.*, **9**, e1003682.
  44. Fijalkowska, I.J., Schaaper, R.M. and Jonczyk, P. (2012) DNA replication fidelity in *Escherichia coli*: A multi-DNA polymerase affair. *FEMS Microbiol. Rev.*, **36**, 1105–1121.
  45. Fijalkowska, I.J., Jonczyk, P., Tkaczyk, M.M., Bialoskorska, M. and Schaaper, R.M. (1998) Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 10020–10025.
  46. Kuban, W., Banach-Orlowska, M., Bialoskorska, M., Lipowska, A., Schaaper, R.M., Jonczyk, P. and Fijalkowska, I.J. (2005) Mutator phenotype resulting from DNA polymerase IV overproduction in *Escherichia coli*: Preferential mutagenesis on the lagging strand. *J. Bacteriol.*, **187**, 6862–6866.
  47. Tang, M.J., Pham, P., Shen, X., Taylor, J.S., O'Donnell, M., Woodgate, R. and Goodman, M.F. (2000) Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature*, **404**, 1014–1018.
  48. Mudrak, S.V., Welz-Voegel, C. and Jinks-Robertson, S. (2009) The polymerase  $\eta$  translesion synthesis DNA polymerase acts independently of the mismatch repair system to limit mutagenesis caused by 7,8-dihydro-8-oxoguanine in yeast. *Mol. Cell. Biol.*, **29**, 5316–5326.
  49. Moldovan, G.L., Pfander, B. and Jentsch, S. (2007) PCNA, the maestro of the replication fork. *Cell*, **129**, 665–679.
  50. Bravo, R., Frank, R., Blundell, P.A. and Macdonaldbravo, H. (1987) Cyclin/PCNA is the auxiliary protein of DNA polymerase  $\delta$ . *Nature*, **326**, 515–517.
  51. Acharya, N., Klassen, R., Johnson, R.E., Prakash, L. and Prakash, S. (2011) PCNA binding domains in all three subunits of yeast DNA polymerase  $\delta$  modulate its function in DNA replication. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 17927–17932.
  52. Chilkova, O., Stenlund, P., Isoz, I., Stith, C.M., Grabowski, P., Lundstrom, E.B., Burgers, P.M. and Johansson, E. (2007) The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res.*, **35**, 6588–6597.
  53. Schauer, G.D. and O'Donnell, M.E. (2017) Quality control mechanisms exclude incorrect polymerases from the eukaryotic replication fork. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 675–680.
  54. Naryzhny, S.N. (2008) Proliferating cell nuclear antigen: A proteomics view. *Cell. Mol. Life Sci.*, **65**, 3789–3808.
  55. Umar, A., Buermeier, A.B., Simon, J.A., Thomas, D.C., Clark, A.B., Liskay, R.M. and Kunkel, T.A. (1996) Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell*, **87**, 65–73.
  56. Shibahara, K. and Stillman, B. (1999) Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. *Cell*, **96**, 575–585.
  57. Zhang, Z.G., Shibahara, K. and Stillman, B. (2000) PCNA connects DNA replication to epigenetic inheritance in yeast. *Nature*, **408**, 221–225.
  58. Freudenthal, B.D., Gakhar, L., Ramaswamy, S. and Washington, M.T. (2010) Structure of monoubiquitinated PCNA and implications for translesion synthesis and DNA polymerase exchange. *Nat. Struct. Mol. Biol.*, **17**, 479–U123.
  59. Garg, P. and Burgers, P.M. (2005) Ubiquitinated proliferating cell nuclear antigen activates translesion DNA polymerases  $\eta$  and REV1. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 18361–18366.
  60. Haracska, L., Unk, I., Prakash, L. and Prakash, S. (2006) Ubiquitylation of yeast proliferating cell nuclear antigen and its implications for translesion DNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 6477–6482.