# Online Supporting Information for

# "Polymerase η Recruits DHX9 Helicase to Promote Replication across Guanine Quadruplex Structures"

Feng Tang[1], Yinan Wang[1], Zi Gao[1], Shiyuan Guo[2] and Yinsheng Wang[1,2,*]

[1]*Department of Chemistry, University of California, Riverside, CA 92521-0403, United States*

[2]*Genetics, Genomics and Bioinformatics Graduate Program, University of California, Riverside, CA 92521-0403, United States*

*Corresponding authors: Yinsheng Wang.

**E-mail:** Yinsheng.Wang@ucr.edu

## Table of Contents:

## Materials and Methods

### Cell lines

HEK293T cells were purchased from ATCC. The SV40-transformed Pol η-deficient XP30RO fibroblasts (XPV) and the corresponding cells reconstituted with wild-type human Pol η (XPV + Pol η) were kindly provided by Professor James E. Cleaver.[1] All cell lines used in this study were examined for a mycoplasma contamination test using LookOut Mycoplasma PCR Kit (Sigma, MP0035).

Cells were maintained in DMEM (Life Technologies) supplemented with 10% FBS (Invitrogen) and 1% penicillin/streptomycin (v/v) at 37°C in a humidified incubator with 5% $CO_2$. For genomic DNA extraction, cells were harvested at 80% confluence.

### Targeted Integration of Flag Tag using CRISPR-Cas9

Genome editing-based integration of three tandem repeats of the Flag epitope tag (3× Flag tag) into the C-terminus of endogenous Pol η was conducted following the previously reported procedures.[2] DNA sequence for the production of sgRNA targeting the *POLH* gene was inserted into the hSpCas9 plasmid pX330 (Addgene, Cambridge, MA, USA). The template donor plasmid for tagging *POLH* was synthesized (gBlock, IDT) and inserted into pUC19. The constructed Cas9 plasmid and the donor plasmid were transfected into HEK293T cells using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA). At 24 h following plasmid transfection, the cells were diluted and seeded into a 96-well plate at a concentration of approximately 0.5 cell per well. After 2 weeks, single colonies were transferred to a 24-well plate. Genomic DNA was extracted for PCR amplification and agarose gel electrophoresis to screen for clones with successful insertion of the Flag tag. Sanger sequencing and Western blot were performed to confirm the correct insertion. The guide RNAs were designed according to a previously described method,[3] and their sequences are listed in Table S2.

### APEX-labeling Experiments

The human *POLH* gene was amplified from pRK7–Pol η vector[4] to introduce a 5′ Not1 site and a 3′ BamHI site and subcloned into a mito-V5-APEX2 vector to replace the mitochondria-targeting sequence with the coding sequence of human Pol η.[5] The APEX2-NLS vector was a gift from Prof. Alice Ting.[6] The APEX-labeling experiments were conducted following the previously published procedures with minor modifications.[7] The constructed APEX2 plasmids were transfected into HEK293T cells using Lipofectamine 2000. At 24 h following plasmid transfection, the cells were incubated in complete medium with 0.5 mM biotin-phenol for 30 min at 37°C. Hydrogen peroxide was subsequently added to a final concentration of 1 mM and the cells were incubated at room temperature for 1 min. The labeling medium was aspirated quickly and the cells were washed five times with a quenching solution, which contained PBS supplemented with 10 mM sodium ascorbate, 10 mM sodium azide, and 5 mM Trolox.

For LC-MS/MS and Western blot analyses, RIPA buffer (50 mM Tris/HCl, pH 7.5, 150 mM NaCl, 0.1% SDS (w/v), 0.5% sodium deoxycholate (w/v), 1% Triton X-100 (v/v), and 10% glycerol (v/v)) supplemented with 1× protease inhibitor cocktail, 1 mM PMSF and quenchers (10 mM sodium ascorbate, 10 mM sodium azide, and 5 mM Trolox) was immediately added to the cell pellet. After resuspension, the sample was placed on the rotator at 4°C for 30 min. The cell lysate was cleared by centrifugation at 16,000× g for 15 min. The protein concentrations were measured using Bradford Quick Start Protein Assay kit (Bio-Rad), and the lysate was used immediately for further analyses or stored at −80°C. One mg of whole-cell lysate was incubated with 50 μL high-capacity streptavidin agarose beads (Thermo Fisher Scientific), which was pre-washed twice with RIPA buffer, on a rotator at room temperature for 1 h. After the incubation, the beads were washed twice with RIPA buffer, once with 1 M KCl, once with 0.1 M $Na_2CO_3$, twice with 2 M urea in 10 mM Tris-HCl (pH 8.0), and twice with RIPA buffer. After washing, the bound proteins were eluted by boiling for 10 min in 3× protein loading buffer (Bio-rad) supplemented with 2 mM biotin and 20 mM DTT. The resulting mixture was

centrifuged, and the supernatant was loaded onto a 15% SDS-PAGE gel and ran for a very short time (120 V, ~10 min). The gel was washed to remove the running buffer and stained with Coomassie blue (VWR) for 0.5 h. The gel was subsequently destained overnight with a destaining solution (ethanol:$H_2O$:acetic acid, 4.5:4.5:1, v/v). Gel slices containing the proteins were excised and cut into small cubes (1 × 1 mm).

The eluted proteins were digested in-gel following a previously described protocol.[8] Briefly, excess SDS in the gel was removed by incubating with 500 µL of 100 mM ammonium bicarbonate/acetonitrile (1:1, v/v) in a thermomixer at 37°C with interval mixing at 1,200 rpm. The supernatant was removed and the gel pieces were dehydrated with acetonitrile for 5 min. Gel pieces were further dehydrated in a vacuum centrifuge for 1-2 min. Proteins were then reduced with 10 mM dithiothreitol (DTT, Sigma) in 25 mM ammonium bicarbonate at 55°C for 1 h, and subsequently alkylated by incubating with 55 mM iodoacetamide (IAA, Sigma) in the dark for 1 h. Gel pieces were dehydrated with acetonitrile and washed three times with 1 mL of 25 mM ammonium bicarbonate. Proteins were then digested with trypsin at 37°C overnight. After digestion, the peptides were eluted from the gel by incubating, with vigorous shaking for 15 min at 37°C, first in 200 µL 5% formic acid (v/v) in 25 mM $NH_4HCO_3$, then in 200 µL 5% formic acid in 25 mM $NH_4HCO_3$ and 50% acetonitrile (v/v), and finally in 200 µL 5% acetic acid in 25 mM $NH_4HCO_3$ and 70% acetonitrile (v/v). After elution, the peptide fractions were combined, evaporated to dryness, desalted using OMIX C18 Tips (Agilent), and analyzed by LC-MS/MS.

**Affinity pull-down of Pol η-binding proteins**

For pull-down assay, Pol η-Flag and parental HEK293T cells were collected and lysed in CelLytic M cell lysis reagent (Sigma) with 1× protease inhibitor cocktail (Sigma). After lysis, the cell lysate was incubated with 30 µL pre-equilibrated anti-Flag M2 affinity gel (Sigma) at 4 °C overnight. The beads were washed for five times with a washing buffer containing 50 mM Tris (pH 8.0), 150 mM NaCl, and

0.1% Tween 20 (v/v). After washing, the bound proteins were eluted with 30 μL of 2× SDS-PAGE loading buffer (Bio-Rad) with 5 min of boiling. The resulting proteins were loaded onto an SDS-PAGE gel, which was run for a short duration, and the proteins digested in-gel following the aforementioned procedures and subjected to LC-MS/MS analysis.

**LC-MS/MS**

LC-MS/MS experiments were performed as previously described with minor modifications.[9] Briefly, the peptides were analyzed using an EASY-nLC 1200 system coupled with a Q Exactive Plus quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). The HPLC separation was conducted using a trapping column followed by a separation column, both packed in-house with ReproSil-Pur C18-AQ resin (3 μm, Dr. Maisch HPLC GmbH, Germany). The peptides were separated using a 200-min linear gradient of 2-40% acetonitrile in 0.1% formic acid at a flow rate of 300 nL/min. The mass spectrometer was set up in the positive-ion mode, and the spray voltage was 1.8 kV. MS/MS was recorded in a data-dependent acquisition mode, in which one full MS scan was followed with 25 MS/MS scans.

**Data Analysis**

LC-MS/MS data were processed by MaxQuant version 1.6.14 using default parameters unless otherwise specified (http://www.maxquant.org).[10] Database searches were performed using the Andromeda search engine included with MaxQuant with the UniProt human sequence database (UP000005640). The mass tolerances for precursor and fragment ions were set at 4.5 ppm and 20 ppm, respectively. Digestion enzyme specificity was set to trypsin with a maximum of 2 missed cleavages. A minimum peptide length of 7 residues was required for identification. Protein acetylation and oxidation were set as variable modifications, and cysteine carboamidomethylation was set as the fixed modification. “Match between runs” based on accurate m/z and retention time was enabled with a 5-min alignment time window. Label-free quantitation (LFQ) was performed using the MaxLFQ

algorithm in MaxQuant.[11] Protein LFQ intensities were calculated from the median of pairwise intensity ratios of peptides identified in two or more samples and adjusted to the cumulative intensity across samples. Quantification was performed using razor and unique peptides, including those modified by acetylation (protein N-terminal) and oxidation (Met). A minimum peptide ratio of 1 was required for protein intensity normalization.

Data processing was performed using Perseus version 1.6.13.0 (http://www.perseus-framework.org). Contaminants and protein groups identified by a single peptide were filtered from the data set. Protein group LFQ intensities were $\log_2$-transformed to reduce the effect of outliers. For statistical comparisons between proteomes, protein groups missing LFQ values were assigned values using imputation. Missing values were assumed to be biased toward low-abundance proteins that were below the MS detection limit, referred to as "missing not at random", an assumption that is frequently made in proteomics studies. The missing values were replaced with random values taken from a median downshifted Gaussian distribution to simulate low-abundance LFQ values. Imputation was performed separately for each sample from a distribution with a width of 0.3 and a downshift of 1.8. Log ratios were calculated as the difference in $\log_2$ LFQ intensity averages between experimental and control groups. A two-tailed, unpaired, Student's t-test was employed to identify differentially expressed proteins. Visualization of the results was performed with volcano plots and Venn diagrams using the R libraries ggplot2 and VennDiagram.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[12] partner repository with the dataset identifier PXD033885.

**Pull-down and Western blot**

Protein pull-down was performed with anti-Flag M2 affinity gel following the aforementioned procedures. Western blot analysis was conducted following previously published procedures with minor modifications,[13] where protein samples were separated on a SDS-PAGE gel and transferred to a nitrocellulose membrane (Bio-Rad). After blocking with a blotting-grade blocker (Bio-Rad), the

membrane was incubated in a solution containing primary antibody and 5% BSA (w/v) at 4°C overnight, and then incubated with the HRP-conjugated secondary antibody in a 5% blotting-grade blocker (w/v) at room temperature for 1 h. The immunoblots were detected using ECL Western blotting detection reagent (Amersham). Primary antibodies used in this study included anti-Flag (Cell Signaling Technology, 2368S; 1:2000), DHX9 (Proteintech, 67153-1-Ig; 1:2000), and DDX5 (Proteintech, 67025-1-Ig; 1:2000).

**Chromatin Immunoprecipitation (ChIP) and Next-Generation Sequencing**

Chromatin immunoprecipitation was performed as previously described with a few modifications.[14] Briefly, $2\times10^7$ cells were cross-linked by adding formaldehyde dropwise directly to the media for a final concentration of 1% (w/v) and rotating gently at room temperature for 10 min. The reaction was quenched by adding 2.5 M glycine to the media until its final concentration reached 125 mM and the mixture was incubated at room temperature with shaking for 5 min. After washing with 1× PBS three times, the cells were resuspended in PBS. After centrifugation, the cell pellet was resuspended in RIPA buffer [50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% NP40 (v/v), 0.5% sodium deoxycholate (w/v), 0.1% SDS (w/v) and a protease inhibitor cocktail] at 4 °C for 30 min with rotation. Chromatin was sheared using a Covaris S220 sonicator at 4°C for 10 min with a peak incident power of 140, a duty cycle of 10%, and 200 cycles per burst. After centrifugation at 16,000g for 15 min, the supernatant was incubated with pre-blocked anti-Flag M2 affinity gel (Sigma) on a rotator at 4°C overnight. After washing, DNA was eluted from the beads with 120 μL 100 mM $NaHCO_3$ and 1% SDS (w/v) at 68 °C for 2 h. Cross-links were subsequently reversed by adding 4.8 μL of 5 M NaCl and 2 μL RNase A (10 mg/mL) and incubated with shaking at 65°C overnight. Proteins in the resulting DNA samples were removed with 2 μL proteinase K (20 mg/mL) and incubated while shaking at 60°C for 2 h. Finally, the DNA was purified using Cycle Pure Kit (omega).

Purified DNA was quantified and verified on an Agilent 2100 Bioanalyzer. The library was then constructed using NEBNext Ultra DNA Library Prep Kit (NEB, E7103S) following the manufacturer's

instructions.[15] Subsequently, the purified DNA libraries were assessed using an Agilent 2100 Bioanalyzer and multiplexed for sequencing on a HiSeq4000 system (Illumina).

The sequencing reads of ChIP-seq were checked with FastQC and aligned to the human hg19 reference genome using Bowtie2 with the default configuration.[16] For genome-wide identification of Pol η-binding sites, peak calling was performed using the model-based analysis of ChIP-seq (MACS2)[17] with the following configuration: macs2 callpeak -t treat.bam -c ctrl.bam -f BAM -g 2.7e+9 -n --ourdir macs2. The Integrative Genomics Viewer[18] was used to visualize the mapping results. Overlap of two biological replicates was defined as high-confidence peaks, which were employed in further analysis. Genomic annotations were performed using PAVIS.[19] The intersection between bed files was performed using BEDTools.[20] Plotprofile was performed using deepTools.[21] The ChIP-Seq data generated in this study have been deposited into the NCBI Gene Expression Omnibus (GEO) with the accession number of GSE 203015.

**Chromatin immunoprecipitation-quantitative PCR analysis of DHX9**

The ChIP-qPCR analysis of DHX9 was performed by following a previous report with some modifications.[22] Cells ($2 \times 10^7$) were crosslinked with 1% (w/v) formaldehyde at room temperature for 10 min. The crosslinking reaction was quenched by adding 2.5 M glycine to the media for a final concentration of 125 mM and incubated with shaking at room temperature for 5 min. Cells were washed and harvested in cold PBS. After centrifugation, the cell pellet was resuspended in RIPA buffer at 4°C for 30 min with rotation. Chromatin was sheared using a Covaris S220 sonicator at 4°C for 10 min. Sheared chromatin lysates were pre-cleared by incubation with 30 μL of Protein A/G PLUS-Agarose (Santa Cruz) followed by incubation for 4 h with 20 μg of anti-DHX9 (Santa Cruz) at 4 °C with rotation. Thirty μL of Protein A/G PLUS-Agarose was then added and incubated overnight at 4 °C with rotation. Antibody was omitted in the control sample. The beads were washed successively twice with 1 ml of RIPA buffer, followed by 1 mL of washing buffer 1 [20 mM Tris–HCl, pH 8.0, 500 mM NaCl, 1 mM EDTA, 1% sodium deoxycholate (w/v), 1% NP40 (v/v)], and 1 ml of washing

buffer 2 [20 mM Tris–HCl, pH 8.0, 500 mM LiCl, 2 mM EDTA, 1% Triton-100 (v/v)]. After washing, DNA was eluted from the beads with 120 μL 100 mM $NaHCO_3$ and 1% SDS (w/v) at 68°C for 2 h. Cross-links were subsequently reversed by adding 4.8 μL of 5 M NaCl and 2 μL RNase A and incubated with shaking at 65°C overnight. After cross-link reversed, proteins in the resulting DNA samples were removed with 2 μL proteinase K (20 mg/mL) and incubated while shaking at 60°C for 2 h. Finally, the DNA was purified using Cycle Pure Kit (omega). Quantitative PCR was performed using standard protocols using Luna® Universal qPCR Master Mix (NEB, M3003X). Analysis was carried out using the *ΔΔC*t method (fold enrichment).[13]

**Immunocapture of BrdU-labeled DNA and qPCR analysis**

The 5-bromo-2′-deoxyuridine (BrdU) incorporation assay was performed according to previously reported procedures with minor modifications.[23] Cells were incubated in 30 μM BrdU for 2 h, and washed twice with cold PBS gently. Genomic DNA was subsequently extracted by using a Monarch Genomic DNA Purification Kit (NEB, T3010S) following the manufacturer's instructions. The genomic DNA was sheared to an average size of approximately 300 bp using a Covaris S220 sonicator at 4°C. The enrichment of BrdU-labeled DNA was performed as previously described.[24] Briefly, the genomic DNA was heat-denatured at 95°C for 5 min and incubated on ice for 2 min. Two μL of anti-BrdU antibody (BD Biosciences) was added to the denatured DNA and incubated at room temperature for 30 min with constant rotation. Twenty μL of Protein A/G PLUS-Agarose was then added and incubated at 4 °C for 1 h with constant rotation. After washing, DNA was eluted from the beads with 2.5 μL Proteinase K (NEB, P8107S) in 200 μL digestion buffer [50 mM Tris-HCl (pH 8.0), 10 mM EDTA and 0.5% SDS (w/v)] at 55°C overnight. Finally, the DNA was purified with Cycle Pure Kit (omega) by following the manufacturer's guidelines. Quantitative PCR was performed using standard protocols using Luna® Universal qPCR Master Mix (NEB, M3003X).

**G4 ChIP-qPCR**

G4 ChIP-qPCR of XPV, XPV + Pol η and the corresponding DHX9 siRNA knockdown cells were performed using BG4 antibody purified in house, and ChIP was conducted following previously described procedures with minor modifications.[25] Approximately $1 \times 10^7$ cells were crosslinked with 1% formaldehyde (w/v) at room temperature for 10 min, followed by quenching with 0.125 M glycine for 10 min. After washing with ice-cold 1 × PBS three times, cells were resuspended in LB1 buffer [50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol (v/v), 0.5% NP-40 (v/v), and 0.25% Triton X-100 (v/v)] and incubated at 4 ºC for 10 min. After centrifugation, the cells were resuspended in LB2 buffer [10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA], centrifuged, and the cell pellet collected. The cells were subsequently resuspended in LB3 buffer [10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate and 0.5% N-lauroylsarcosine (w/v)] and incubated for at least 15 min before sonication. Protease inhibitor cocktail solution was added in each buffer before use. Sonication was conducted using a Covaris S220 sonicator at a peak incident power of 160, a duty cycle of 10%, 200 cycles per burst for 500 sec to achieve the average chromatin size of 100-500 bp. After centrifugation at 16000g for 10 min, chromatin samples were diluted 50-fold in blocking buffer [25 mM HEPES, pH 7.5, 10.5 mM NaCl, 110 mM KCl, 1 mM $MgCl_2$, 1% BSA (w/v)] and treated with RNase A. The chromatin sample was incubated with BG4 antibody under rotation at 1400 rpm at 16 ºC for 1.5 h. To the mixture was added 5 μL of blocked Anti-Flag M2 affinity gel (Sigma) and the resulting mixture incubated under the same conditions for 1 h. After washing, the captured DNA was eluted in TE buffer (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA) with Proteinase K, where the sample was rotated at 1400 rpm at 65ºC for 4 h. The eluted DNA were then purified by Zymo DNA Clean and Concentrator-5. The immunoprecipitated samples and the input were used to quantify G4 enrichment by using qPCR, where the primers are listed in Table S2.

## References


1. Thakur, M.; Wernick, M.; Collins, C.; Limoli, C. L.; Crowley, E.; Cleaver, J. E., DNA polymerase eta undergoes alternative splicing, protects against UV sensitivity and apoptosis, and suppresses Mre11-dependent recombination. *Genes Chromosomes Cancer* **2001,** *32*, 222-235.
2. Williams, P.; Li, L.; Dong, X.; Wang, Y., Identification of SLIRP as a G quadruplex-binding protein. *J. Am. Chem. Soc.* **2017,** *139*, 12426-12429.
3. Ran, F. A.; Hsu, P. D.; Wright, J.; Agarwala, V.; Scott, D. A.; Zhang, F., Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **2013,** *8*, 2281-2308.
4. Dai, X.; You, C.; Wang, Y., The functions of serine 687 phosphorylation of human DNA polymerase h in UV damage tolerance. *Mol. Cell. Proteomics* **2016,** *15*, 1913-1920.
5. Lam, S. S.; Martell, J. D.; Kamer, K. J.; Deerinck, T. J.; Ellisman, M. H.; Mootha, V. K.; Ting, A. Y., Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **2015,** *12*, 51-54.
6. Kaewsapsak, P.; Shechner, D. M.; Mallard, W.; Rinn, J. L.; Ting, A. Y., Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *Elife* **2017,** *6*, e29224.
7. Hung, V.; Udeshi, N. D.; Lam, S. S.; Loh, K. H.; Cox, K. J.; Pedram, K.; Carr, S. A.; Ting, A. Y., Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat. Protoc.* **2016,** *11*, 456-475.
8. He, X.; Wang, P.; Wang, Y., Mitochondrial transcription factor A binds to and promotes mutagenic transcriptional bypass of $O^4$-alkylthymidine lesions. *Anal. Chem.* **2020,** *93*, 1161-1169.
9. Qi, T. F.; Tang, F.; Yin, J.; Miao, W.; Wang, Y., Parallel-reaction monitoring revealed altered expression of a number of epitranscriptomic reader, writer, and eraser proteins accompanied with colorectal cancer metastasis. *Proteomics* **2022**, e2200059.
10. Tyanova, S.; Temu, T.; Cox, J., The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **2016,** *11*, 2301-2319.
11. Shah, A. D.; Goode, R. J. A.; Huang, C.; Powell, D. R.; Schittenhelm, R. B., LFQ-analyst: an easy-to-use interactive web platform to analyze and visualize label-free proteomics data preprocessed with MaxQuant. *J. Proteome Res.* **2020,** *19*, 204-211.
12. Perez-Riverol, Y.; Bai, J.; Bandla, C.; Garcia-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D. J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; Walzer, M.; Wang, S.; Brazma, A.; Vizcaino, J. A., The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **2022,** *50*, D543-D552.
13. Tang, F.; Yuan, J.; Yuan, B.-F.; Wang, Y., DNA–protein cross-linking sequencing for genome-wide mapping of thymidine glycol. *J. Am. Chem. Soc.* **2022**, 144, 1, 454–462.
14. Li, L.; Wang, Y., Cross-talk between the H3K36me3 and H4K16ac histone epigenetic marks in DNA double-strand break repair. *J. Biol. Chem.* **2017,** *292*, 11951-11959.
15. Tang, F.; Liu, S.; Li, Q. Y.; Yuan, J.; Li, L.; Wang, Y.; Yuan, B. F.; Feng, Y. Q., Location analysis of 8-oxo-7,8-dihydroguanine in DNA by polymerase-mediated differential coding. *Chem. Sci.* **2019,** *10*, 4272-4281.
16. Langmead, B.; Salzberg, S. L., Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012,** *9*, 357-9.
17. Feng, J.; Liu, T.; Qin, B.; Zhang, Y.; Liu, X. S., Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **2012,** *7*, 1728-1740.

18. Nicol, J. W.; Helt, G. A.; Blanchard, S. G., Jr.; Raja, A.; Loraine, A. E., The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* **2009,** *25*, 2730-2731.
19. Huang, W.; Loganantharaj, R.; Schroeder, B.; Fargo, D.; Li, L., Pavis: a tool for p eak a nnotation and vis ualization. *Bioinformatics* **2013,** *29*, 3097-3099.
20. Quinlan, A. R.; Hall, I. M., BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010,** *26*, 841-842.
21. Ramírez, F.; Ryan, D. P.; Grüning, B.; Bhardwaj, V.; Kilpert, F.; Richter, A. S.; Heyne, S.; Dündar, F.; Manke, T., deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **2016,** *44* (W1), W160-W165.
22. Chakraborty, P.; Huang, J. T. J.; Hiom, K., DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. *Nat. Commun.* **2018,** *9*, 1-14.
23. Ryba, T.; Battaglia, D.; Pope, B. D.; Hiratani, I.; Gilbert, D. M., Genome-scale analysis of replication timing: from bench to bioinformatics. *Nat. Protoc.* **2011,** *6*, 870-895.
24. Urbach, E.; Vergin, K. L.; Giovannoni, S. J., Immunochemical detection and isolation of DNA from metabolically active bacteria. *Appl. Environ. Microbiol.* **1999,** *65*, 1207-1213.
25. Hansel-Hertsch, R.; Spiegel, J.; Marsico, G.; Tannahill, D.; Balasubramanian, S., Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat. Protoc.* **2018,** *13*, 551-564.

**Table S2.** Oligonucleotides used in this study.

| **Primers used for ligation and RT-qPCR** | |
|---|---|
| Sequence Names | Sequences |
| PMS2-G4-F | 5′- AGCTGAGAGCTCGAGGTGAG-3′ |
| PMS2-G4-R | 5′- GAGATCGCTGCAACACTGAG-3′ |
| PMS2-NG4-F | 5′-GCCAGACGTTGAGGAAGAAG-3′ |
| PMS2-NG4-F | 5′- ATCAACCCTTCCACTGCTTG-3′ |
| CLSPN-G4-F | 5′- GGCTGAGGGAATCAGAGACA-3′ |
| CLSPN-G4-R | 5′- GGGCGTGTGCATAAACTCA-3′ |
| CLSPN-NG4-F | 5′- GGCACACAGCTTGGAATGTA-3′ |
| CLSPN-NG4-R | 5′- CCCCAGCAACTCTGAGACAG-3′ |
| NEAT1-G4-F | 5′- TAGTTGTGGGGGAGGAAGTG-3′ |
| NEAT1-G4-R | 5′- ACCCTGCGGATATTTTCCAT-3′ |
| NEAT1-NG4-F | 5′- AGAGGGAGGGAGAGCTGAAG-3′ |
| NEAT1-NG4-R | 5′- GCATGAAGTCAGACCAGCAA-3′ |
| JUN-G4-F | 5′- GGGTGACATCATGGGCTATT-3′ |
| JUN-G4-R | 5′- TCGGACTATACTGCCGACCT-3′ |
| JUN-NG4-F | 5′- AAATAAGCAGGCTGGGGAAT-3′ |
| JUN-NG4-R | 5′- CAATCAAGCATGGGGATAGG-3′ |
| **Sequence used for CRISPR/Cas9 editing** | |
| POLH-sg-top | 5′- CACCGAATCCTACAGGCAAGCCTGA -3′ |
| POLH-sg-bot | 5′- AAACTCAGGCTTGCCTGTAGGATTC -3′ |
| POLH-PCR-F | 5′- AGCTTCTTGAGCAGTGACCC -3′ |
| POLH -PCR-R | 5′- GCCTGTGAAGAGATGGGACC -3′ |
| POLH-donor | 5′-<br>CGCCGCGGATCCAAACTTGCAGGGCCCACAACAGCCAAAGCATGCACGCCT<br>CTTCAGCTTCCAAATCTGTGCTGGAGGTGACTCAGAAAGCAACCCCAAATCC<br>AAGTCTTCTAGCTGCTGAGGACCAAGTGCCCTGTGAGAAGTGTGGCTCCCT<br>GGTACCGGTATGGGATATGCCAGAACACATGGACTATCATTTTGCATTGGAGT<br>TGCAGAAATCCTTTTTGCAGCCCCACTCTTCAAACCCCCAGGTTGTTTCTGC<br>CGTATCTCATCAAGGCAAAAGAAATCCCAAGAGCCCTTTGGCCTGCACTAATA<br>AACGCCCCAGGCCTGAGGGCATGCAAACATTGGAATCATTTTTTAAGCCATTA<br>ACACATGGTTCCGCCGGCGACTACAAGGACCACGACGGCGATTATAAGGATC<br>ACGACATCGACTACAAAGACGACGATGACAAGGGCGCCAGCAGCGCCTGGT<br>CCCACCCTCAGTTTGAGAAGGGCGGAGGCTCTGGCGGCGGAAGCGGAGGA<br>TCTGCTTGGAGCCACCCCCAGTTCGAAAAGTAGTGCTGCCCTCAGGCTTGC<br>CTGTAGGATTTAATATTTTTTATCTTTACAGATCTTTATCTTTAATATTTTATCTTTA<br>CAGATTTCCCTGAGAAAGGGAATTATGAAATTTTTAATACAAAAAATAATCCATT<br>TAGGTGCTGAGTTACGGTCCCATCTCTTCACAGGCATGGATTCTAATCCCACT |

| | |
|---|---|
| | GCTGACAGAGATGTAAAAATTCATCCTACCAGAGTTTTTAATCTTTAGCATTTA<br>GGGAGGCAGTGTCATAAAGTAAAAAGTGTGTGGGCCTTGGAGTCTAAGAGAC<br>GTGGTTGCAAACTTAGCTCTGGTTATTGCAATGAGGGCCTTGAACAAGTCTTG<br>GTGGGGAATTCCGGCGG -3′ |

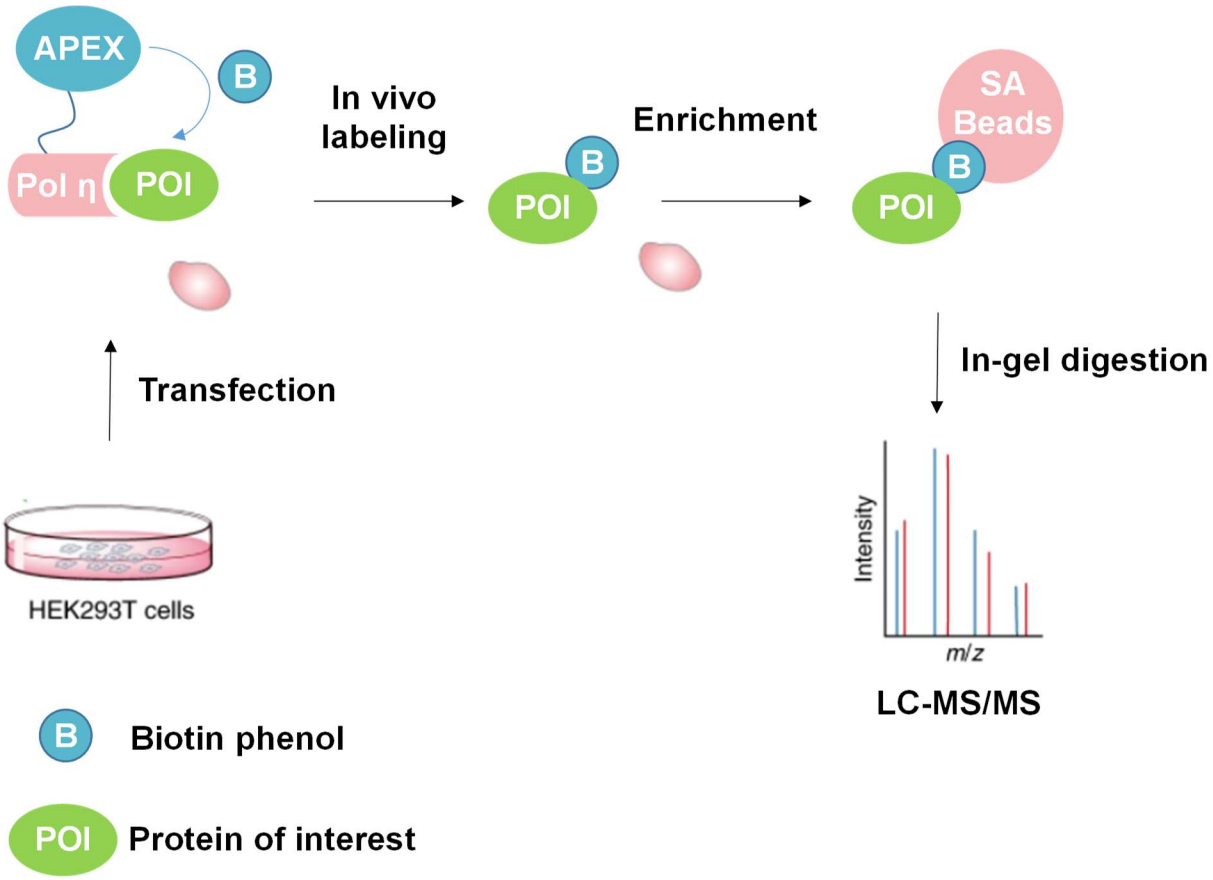


**Figure S1.** A schematic diagram showing the APEX workflow for identifying Pol η-interacting proteins in HEK293T cells.

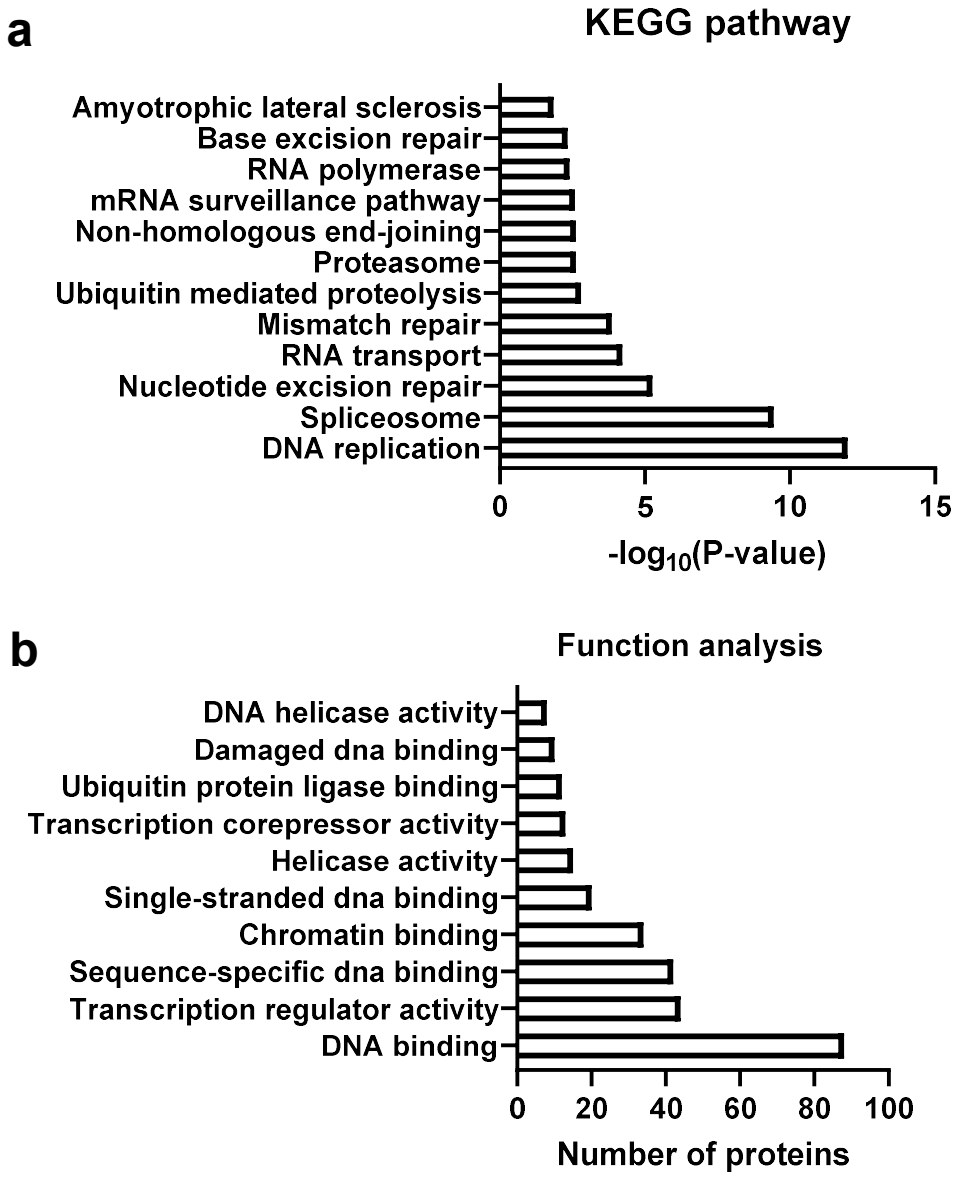


**Figure S2.** KEGG pathway (a) and GO analysis (b) of Pol η-interacting proteins from APEX proteomics data.

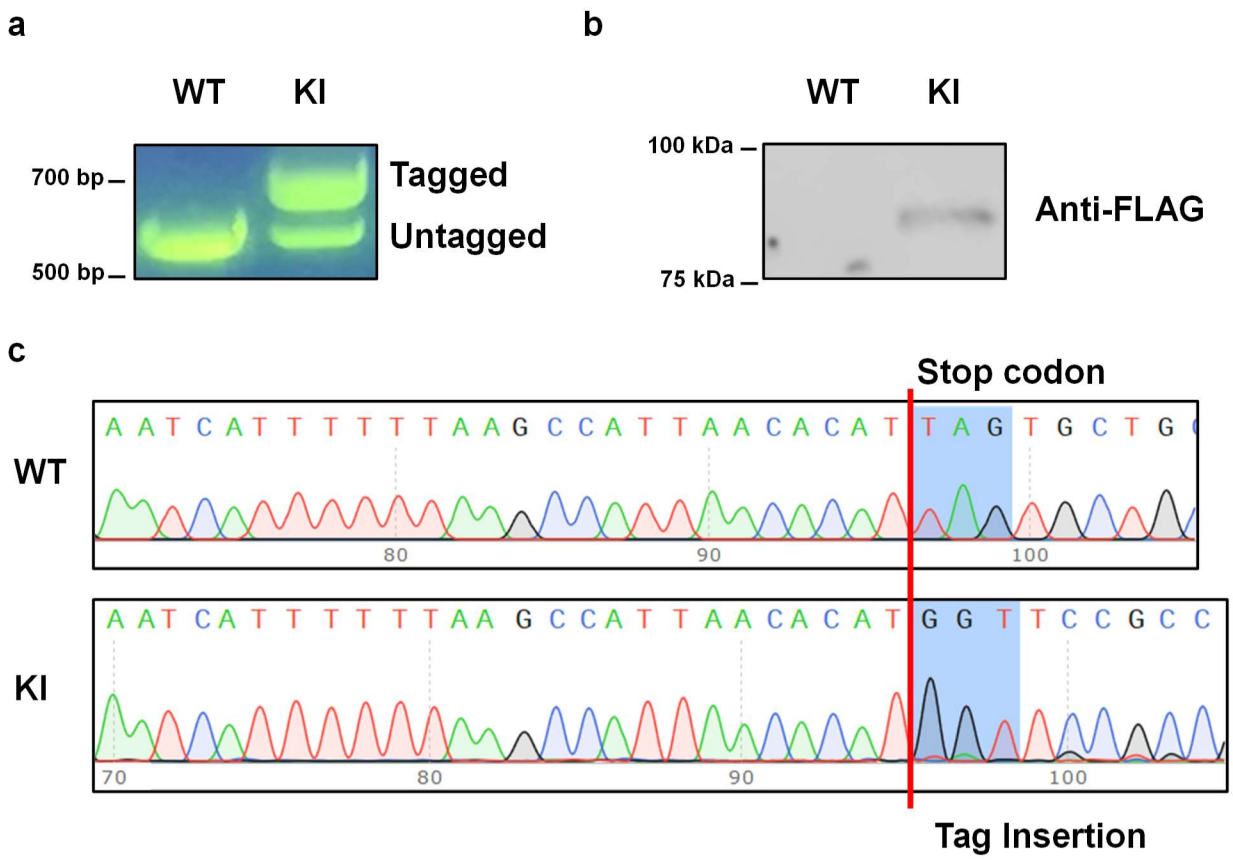


**Figure S3.** (a) Agarose gel electrophoresis for detecting targeted insertion of 3× Flag tag sequence to the C-terminus of Pol η in HEK293T cells. PCR primers (Table S2) are designed to target outside the homology arms and yield a longer PCR product if the tag inserted. The expected amplicon lengths of parental HEK293T cells and the knock-in clone are 584 and 761 bp, respectively. Shown also are the Western blot (b) and Sanger sequencing (c) results for confirming the successful knockin (KI) of 3× Flag Tag to the C-terminus of endogenous Pol η.

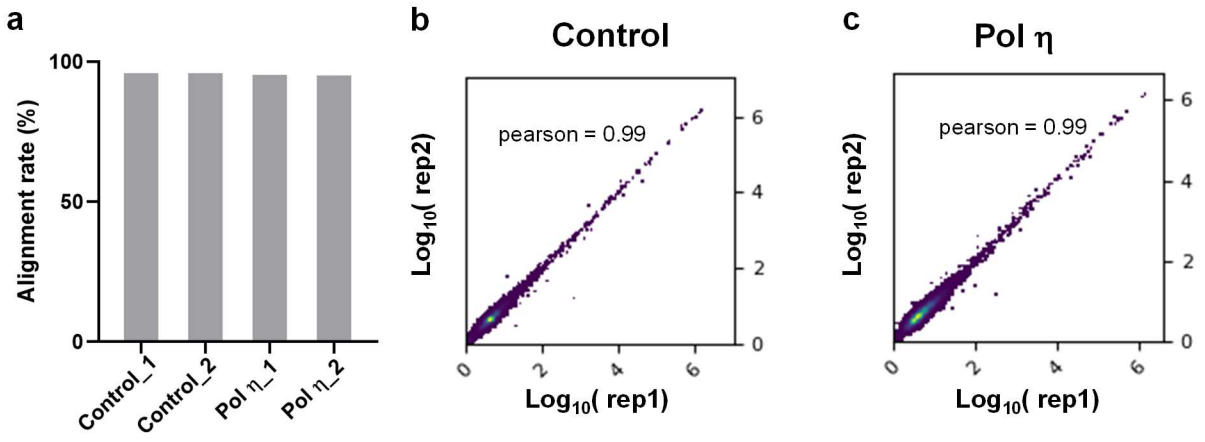


**Figure S4.** Alignment rates and correlation analyses of ChIP-Seq data. (a) The percentages of the sequencing reads that could be aligned with the hg19 human genome. Correlation analysis of sequencing data obtained from the two biological replicates of Control (b) and Pol η (c).

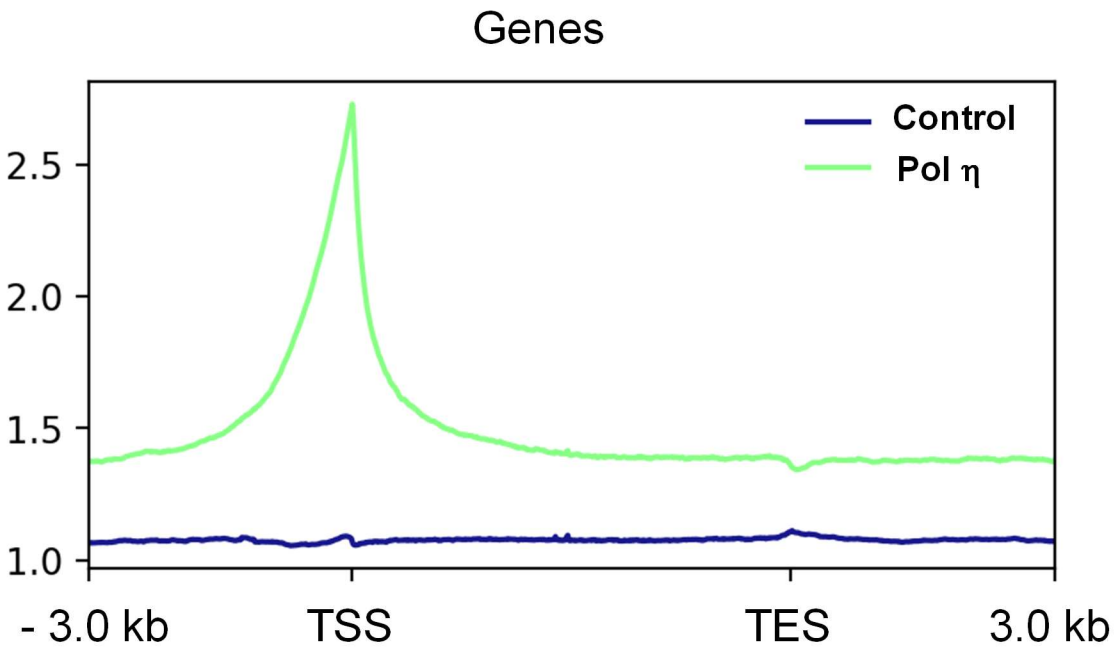


**Figure S5.** ChIP-Seq signal intensity meta profiles for CRISPR-engineered Pol η-Flag HEK293T cells (Pol η) and parental HEK293T cells (Control).

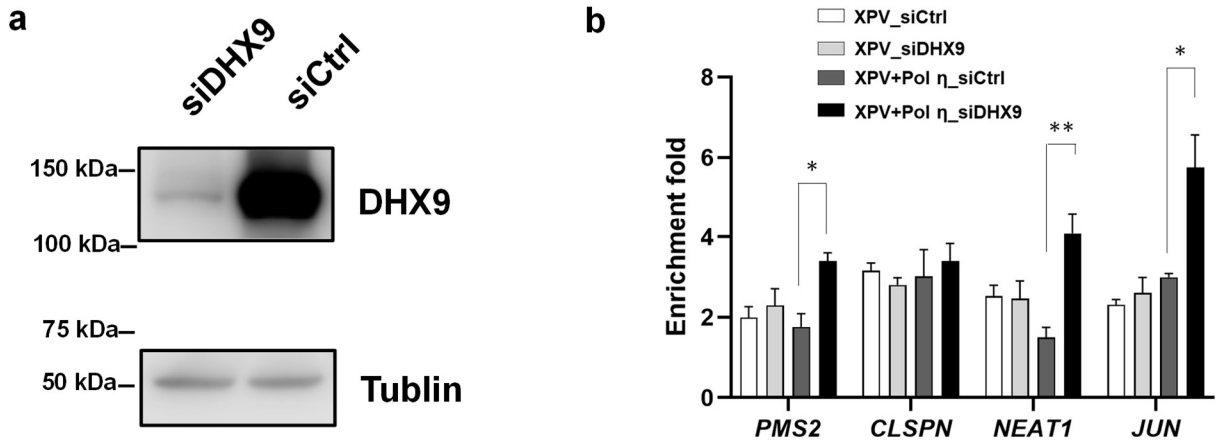


**Figure S6.** (a) Western blot for confirming the successful knockdown of DHX9. (b) BG4 ChIP-qPCR analyses of the corresponding G4 loci. The data are represented as fold enrichment relative to control. The data in (b) represent mean ± S.D. from four experiments (two technical replicates each for two biological replicates). The *p* values were calculated by using two-tailed, unpaired Student's *t*-test: *, $p < 0.05$; **, $p < 0.01$.