Supporting Information for

G-quadruplex on Chromosomal DNA Negatively Regulates Topoisomerase 1 Activity

Hui-ting Liang [1,2#], Jiang-yu Yan [1#], Hao-jun Yao [1], Xue-nan Zhang [1], Zhi-ming Xing [1], Lin Liu [3], Yao-qing Chen [3], Guo-rui Li [1], Jing Huang [1], Yi-de He [2*], Ke-wei Zheng [1*]

[1]School of Biomedical Sciences, Hunan University, Changsha 410082, China
[2]School of Pharmaceutical Sciences, Shenzhen Campus of Sun Yat-Sen University, Shenzhen 518107, China
[3]School of Public Health (Shenzhen), Shenzhen Campus of Sun Yat-sen University, Shenzhen 518107, China

[#]Contributed equally to the work.

*Correspondence should be addressed to Yi-de He (Heyd8@mail.sysu.edu.cn) or Ke-wei Zheng (zhengkewei@hnu.edu.cn)

Supplementary Figures S1-S14, Table S1-S4.

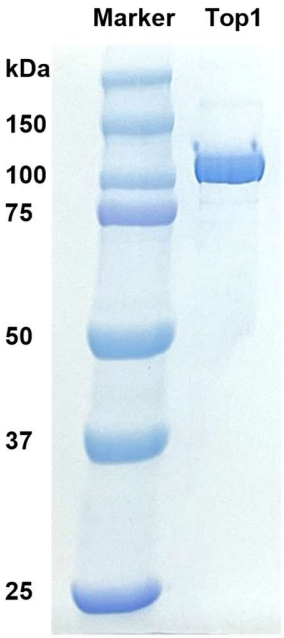

Figure S1 Purified recombinant Top1. Gel was stained by Coomassie brilliant blue G250.

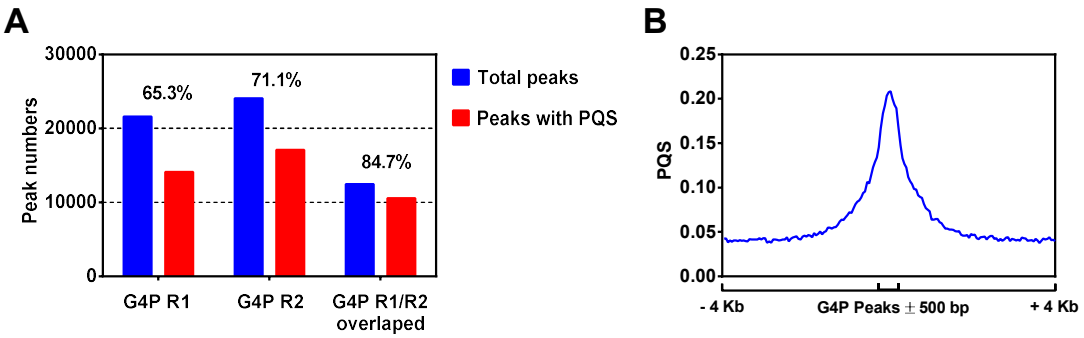

Figure S2. Enrichment of putative G4-forming sequences (PQS) in the G4P peaks. (A) The number of G4P peaks containing PQS in HCT116 cells. The percentage number in each group represents the proportion of G4P peaks that overlapped with at least one PQS by one or more nucleotides. The bed file of PQS was obtained from the NCBI Gene Expression Omnibus (GEO) under accession code GSE133379. (B) Distribution of PQS around G4P peaks.

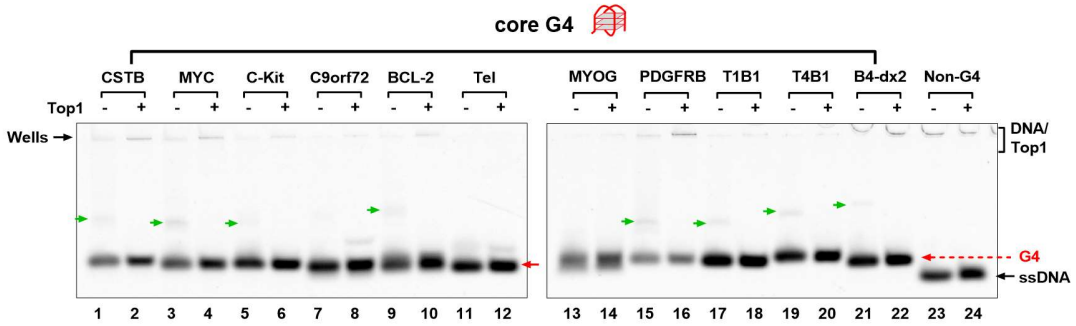

Figure S3. Evaluation of the interaction between Top1 and core G4s using EMSA. The core G4s (**Table** S2) were formed from single-stranded DNA (ssDNA) sequences capable of G-quadruplex formation. A non-G4 ssDNA served as a control. The concentrations of Top1 and DNA were 1 μM and 0.1 μM, respectively. The band marked by the red arrow represents intramolecular G4, while the band indicated by the green arrow represents intermolecular G4.

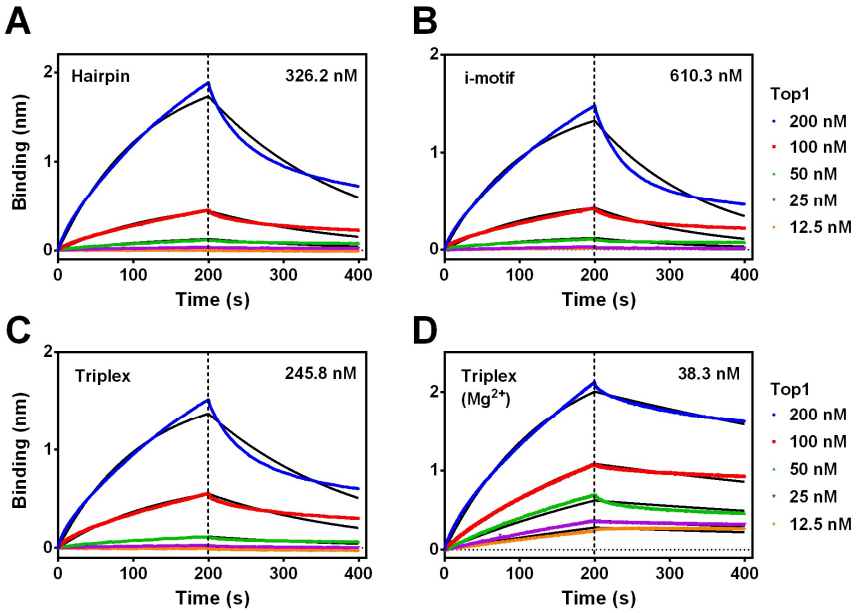

Figure S4. Determination of Top1 affinity for hairpin, i-motif, and triplex DNA by biolayer interferometry (BLI). DNAs (**Table** S2) were biotin-labeled at the 3′ end and immobilized on biosensor tips through the biotin-streptavidin interaction. Indicated concentrations refer to Top1 in the association phase. The association phase was 200 s, followed by a 200-s dissociation phase. Binding is indicated by a wavelength shift (measured in nanometers (nm)) as detected by the Octet instrument. The numerical value on the upper right side of each figure represents the binding constant. (A-C) The binding constants between Top1 and hairpin, i-motif, and triplex DNA were determined in a buffer comprising 10 mM Tris−HCl (pH 7.4), 75 mM KCl, 0.5 mM EDTA, 0.1% Triton X-100, 1 mg/mL BSA, and 0.1 mg/mL fish sperm DNA. (D) The binding constant between Top1 and triplex DNA was determined as in (C), except that 10 mM $MgCl_2$ was added to the buffers.

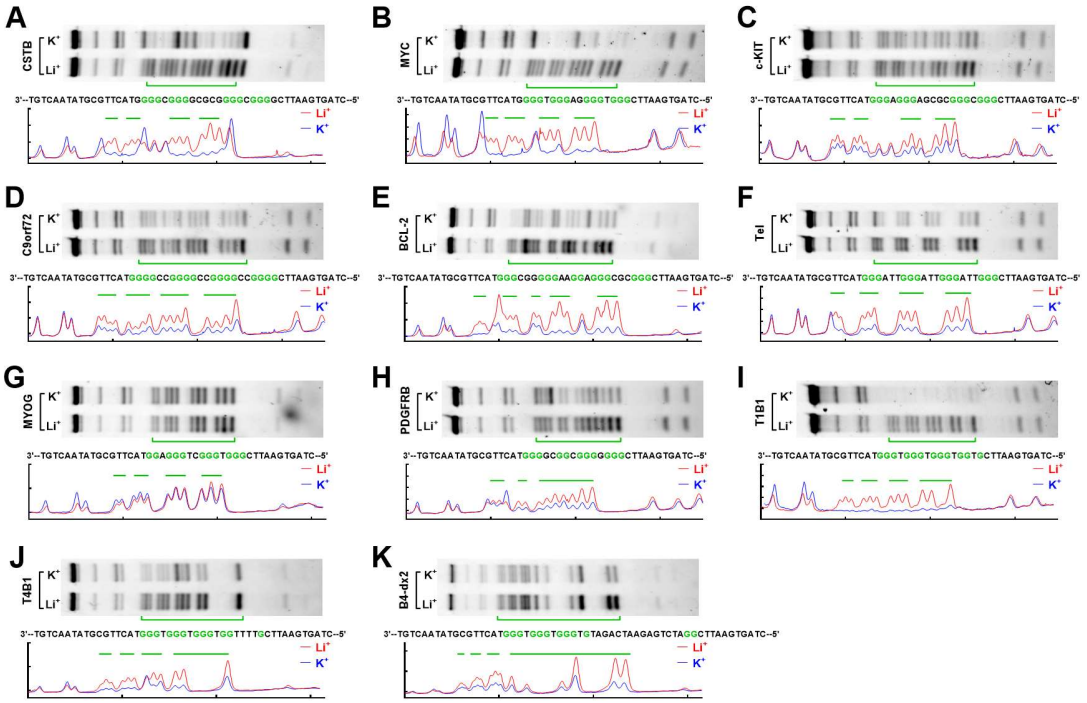

Figure S5. Identification of G4 formation (A-K) in double-stranded DNA (dsDNA) using DMS footprinting. 5′-FAM labeled G-rich DNA and a complementary DNA (**Table** S3) were dissolved at 20 nM in a buffer containing 20 mM Tris–HCl (pH 7.4), 75 mM LiCl or KCl, 1 mM EDTA, denatured at 95°C for 5 min, and slowly cooled down to 25°C. The sample was then treated with 1% dimethyl sulfate (DMS) for 6 min at room temperature and subjected to DMS footprinting as previously described (1). Cleavage fragments were resolved by denaturing gel electrophoresis (up) and digitized (down) for comparison. The G in green font in the DNA sequence indicates the bases potentially involved in the formation of G4.

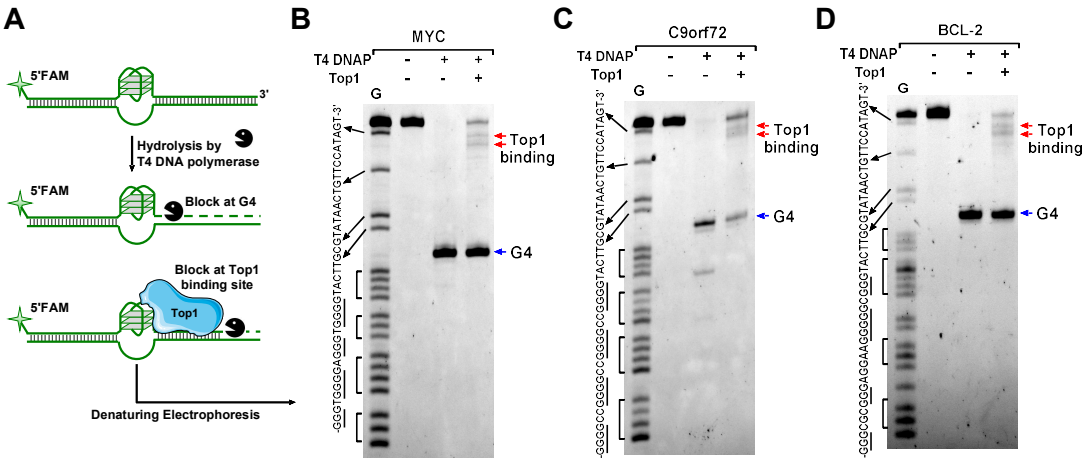

Figure S6. Identification of G4 recognition by Top1 in dsDNA using exonuclease digestion. (A) Schematic diagram of the G-quadruplex and Top1 protein in double-stranded DNA protecting DNA from hydrolysis by T4 DNA polymerase (T4 DNAP). Exonuclease digestion using T4 DNAP was carried out as described (2). (B-D) Top1 binds with MYC, C9orf72, and BCL-2 G4 in dsDNA and protects the DNA from being hydrolyzed from the 3′ end by T4 DNAP. The concentrations of DNA and Top1 were 0.1 and 0.2 μM, respectively. The band indicated by the blue arrow is the cleavage product blocked by G4, and the bands indicated by the red arrows are the cleavage products blocked by the DNA/Top1 binding complex. The lane marked G above is the ladder of base G prepared by DMS footprinting experiment.

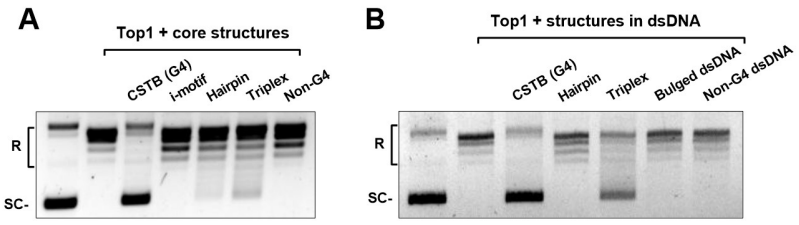

Figure S7. Effects of different types of DNA secondary structures on Top1 activity. (A) Detection of Top1 activity using a plasmid relaxation assay in the presence of 5 μM of a CSTB core G4, an i-motif of hTel, a hairpin DNA, a triplex DNA, and a non-G4 ssDNA (**Table** S2). A negatively supercoiled pGL3-basic plasmid serves as a substrate for Top1. The final concentration of purified Top1 in the reaction was 80 pM. (B) Detection of Top1 activity using a plasmid relaxation assay in the presence of 1 μM of dsDNA containing a CSTB G4, a hairpin, a triplex, and a bulge (**Table** S3). A non-G4 dsDNA was used as control. A negatively supercoiled pGL3-basic plasmid serves as the substrate for Top1. The relaxed and supercoiled plasmids were labeled as R and SC-, respectively.

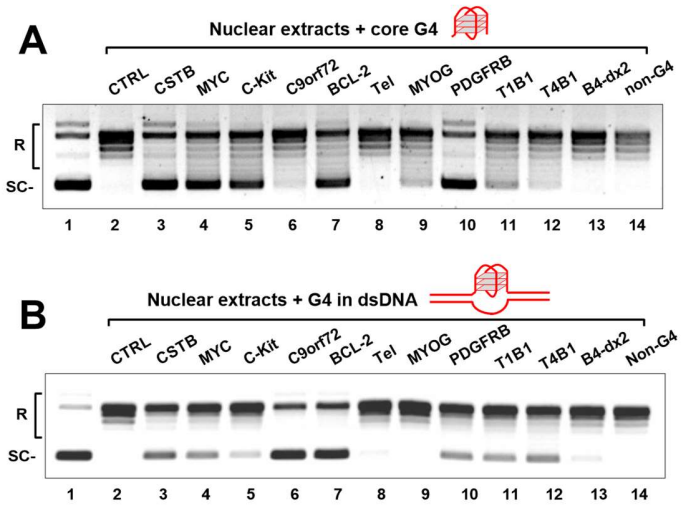

Figure S8. G4 inhibits the unwinding activity of Top1 in nuclear extracts. (A) Detection of Top1 activity of nuclear extracts from 293T cell using a plasmid relaxation assay in the presence of 5 μM of 11 core G4s (CSTB, MYC, C-Kit, C9orf72, BCL-2, Tel, MYOG, PDGFRB, T1B-1, T4B-1, B4-dx2) and a non-G4 ssDNA (**Table** S2). (B) Detection of Top1 activity of nuclear extracts from 293T cell using a plasmid relaxation assay in the presence of 1 μM of 11 G4s (CSTB, MYC, C-Kit, C9orf72, BCL-2, Tel, MYOG, PDGFRB, T1B-1, T4B-1, B4-dx2) in dsDNA or a non-G4 dsDNA (**Table** S3). A negatively supercoiled pGL3-basic plasmid serves as the substrate for Top1. The relaxed and supercoiled plasmids were labeled as R and SC-, respectively.

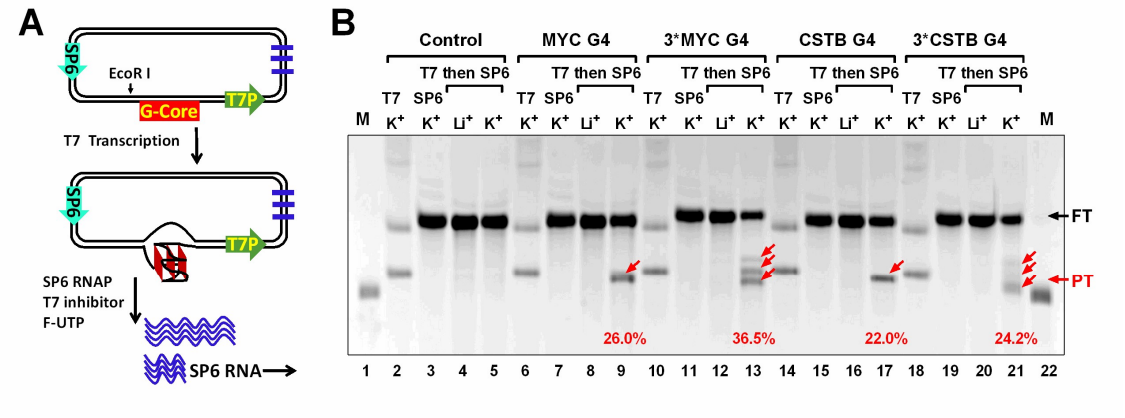

Figure S9. Identification of G4 formation in plasmid DNA using a transcription arrest assay. A) The plasmid used in the assay contained G-core sequences of MYC G4, 3*MYC G4, CSTB G4, 3*CSTB G4, or a non-G4 control sequence, along with a T7 promoter and an SP6 promoter. The G-core sequence was positioned upstream of the T7 promoter and downstream of the SP6 promoter. The plasmid underwent transcription with T7 RNA polymerase to induce G4 formation. Subsequently, SP6 RNA polymerase, along with fluorescein-UTP and a T7 inhibitor, was introduced to generate fluorescent SP6 RNA transcripts. The T7 inhibitor prevented further transcription by the T7 RNA polymerase. (B) G4s in the plasmids were detected through the premature termination of SP6 RNA polymerase. A marker (M) was used to indicate termination sites at *Eco*R I located before the G-core sequences. G-quadruplex formation in the G-core is indicated by the premature termination (PT) band (red arrows) below the full-length transcript (FT) (black arrow). The percentage values in lane 9, 13, 17, and 21 indicate the proportion of PT bands (PT/(PT+FT)).

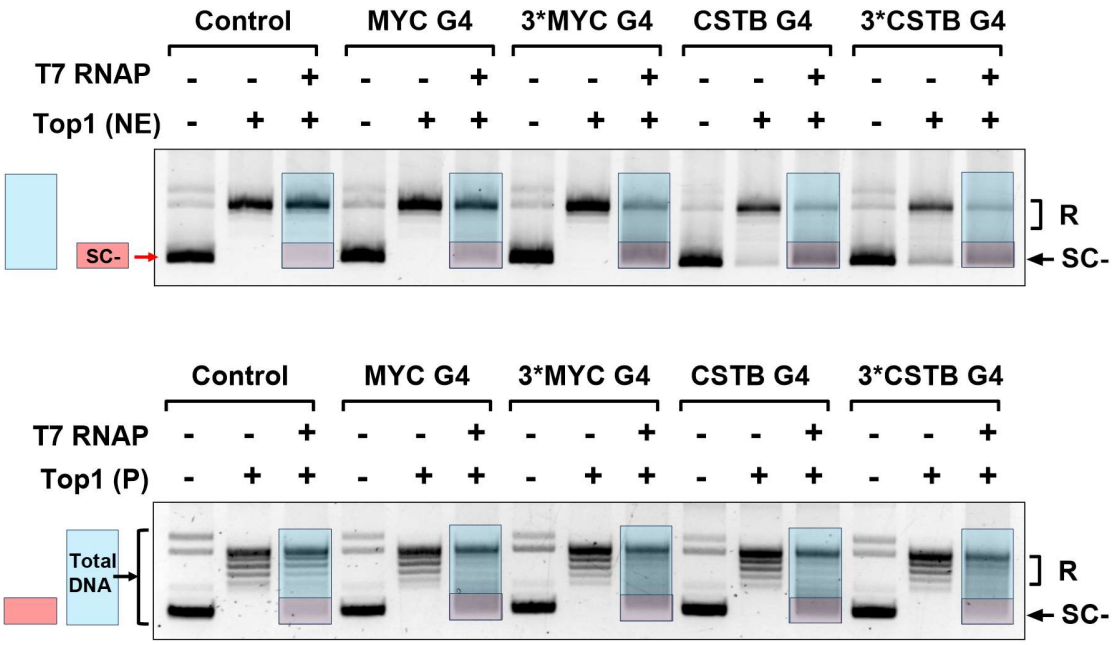

Figure S10. Method used to quantify the DNA relaxing activity of Top1 in **Figure** 4. The gel image was processed with ImageQuant 5.2 software. The signal density in the area selected by the red box represents the amount of unrelaxed negatively supercoiled plasmid, and the signal density in the area selected by the blue box represents the amount of all DNA (including negatively supercoiled plasmid and relaxed plasmid). Local average mode was selected for background subtraction. The fraction of unrelaxed plasmid was calculated by dividing the signal density of the red box by the signal density of the blue box.

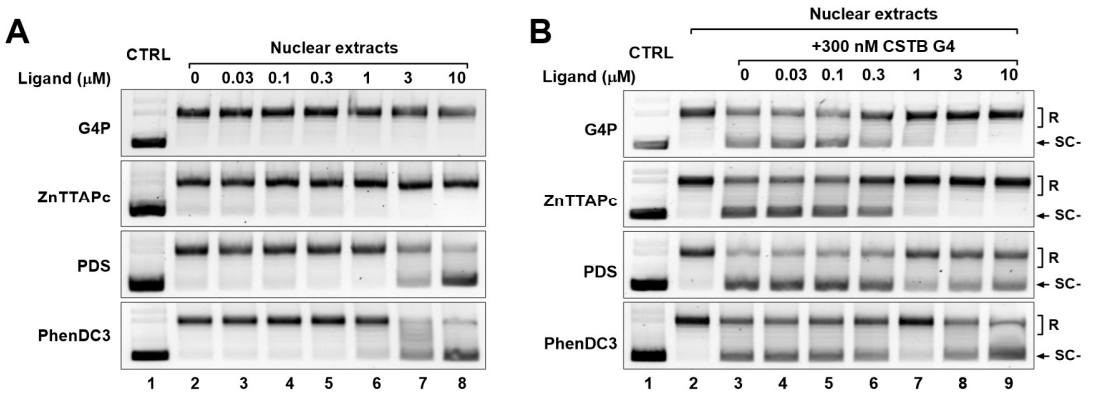

Figure S11. G4 ligands alter the ability of G4 to inhibit Top1 activity in nuclear extracts. A negatively supercoiled pGL3-basic plasmid serves as the substrate for Top1. The relaxed and supercoiled plasmids were labeled as R and SC-, respectively. (A) The effect of G4 ligands on the activity of human Top1 in nuclear extracts. (B) The effect of G4 ligands on the activity of human Top1 in nuclear extracts in the presence of 300 nM CSTB core G4.

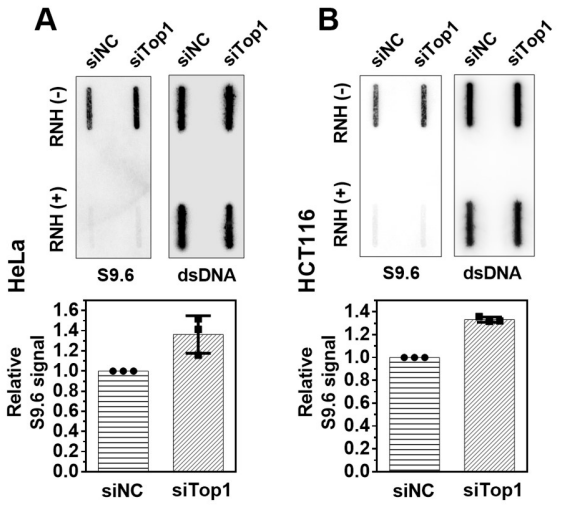

Figure S12. R-loop levels are elevated following the knockdown of Top1 expression. (A-B) Comparison of R-loop levels in genomic DNA from HeLa and HCT116 cells before and after Top1 gene knockdown using slot blot.

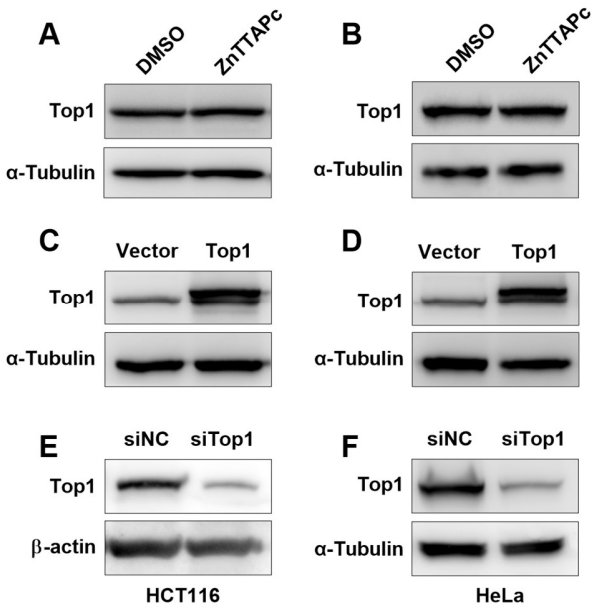

Figure S13. Detection of the protein levels of Top1 using western blot. (A-B) Effect of ZnTTAPc on Top1 protein levels in HeLa and HCT116 cells. (C-D) Detection of the expression level of Top1 protein in HeLa and HCT116 cells following transient transfection of the plasmid encoding the human Top1 gene. (E-F) Detection of the expression level of Top1 protein in HeLa and HCT116 cells after transient transfection of the siRNA targeting human Top1 gene.

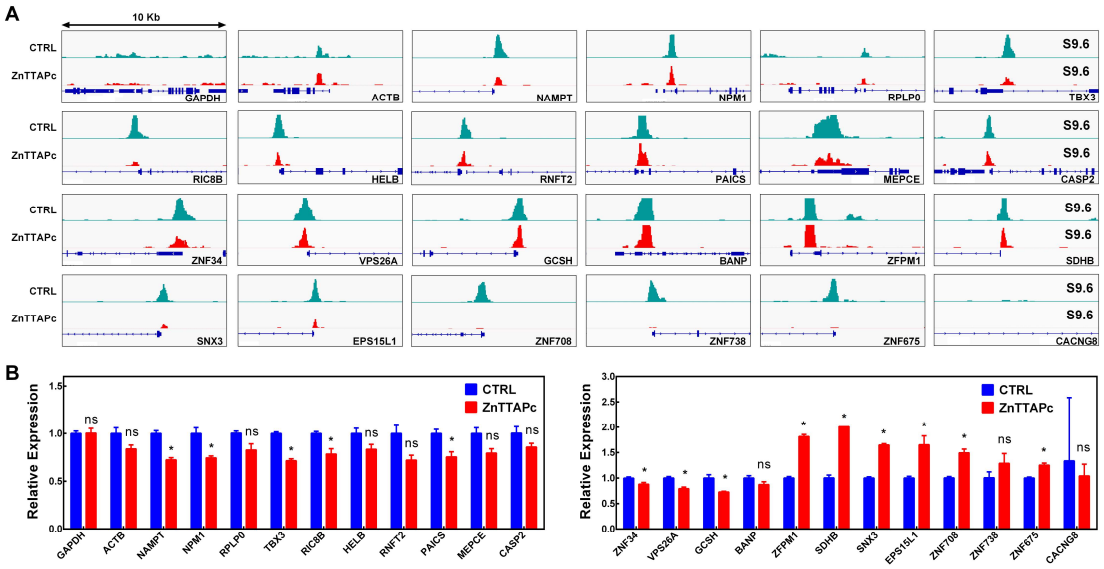

Figure S14. The reduction in R-loop caused by ZnTTAPc was not associated with changes in gene transcription efficiency. (A) Comparison of R-loop signals of 24 gene loci before and after ZnTTAPc treatment. (B) Changes in gene transcription of the 24 genes in (A) after treatment with 10 μM ZnTTAPc. HCT116 cells were cultured in 6-well plates and treated with 10 μM ZnTTAPc for 24h. Total RNA was isolated from cells using Total RNA extraction kit (Magen, Guangzhou, China). Then, reverse transcription of RNA to cDNA was performed according to the HiScriptII one step qRT-PCR kit (TransGen Biotech). The expression of the indicated genes was analyzed using real-time PCR with SYBR Green qPCR mix (Hunan Ruoyu Biotech, China). The primer sequences are shown in **Table** S4. Relative expression levels were determined using the $2^{-\Delta\Delta Cq}$ method and normalized against the reference gene GAPDH. The data presented represent the mean values from three independent experiments, and the error bars represent the standard deviation (SD). P values were calculated using two-tailed unpaired student's t-test (not significant (ns), $P > 0.05$; * $P<0.05$).

Table S1. Oligonucleotides used for plasmids construction.

Name	Sequence (5′ to 3′)
Control-F	AATTCATAGATGTCCGCTTATGTGAAAGTTAGTTAAGAACCTAGGTGTAGTAGGTCACA
Control-R	GATCTGTGACCTACTACACCTAGGTTCTTAACTAACTTTCACATAAGCGGACATCTATG
MYC-F	AATTCATAGATGTCCGCTTATGTTCCCCACCCTCCCCACCCTAGGTGTAGTAGGTCACA
MYC-R	GATCTGTGACCTACTACACCTAGGGTGGGGAGGGTGGGGAACATAAGCGGACATCTATG
3*MYC-F	AATTCATAGATGCCCCACCCTCCCCACCCTAAAGTCCCCACCCTCCCCACCCTATGTTCCCCACCCTCCCCACCCTAGGTGTAGTAGGTCACA
3*MYC-R	GATCTGTGACCTACTACACCTAGGGTGGGGAGGGTGGGGAACATAGGGTGGGGAGGGTGGGGACTTTAGGGTGGGGAGGGTGGGGCATCTATG
CSTB-F	AATTCATAGATGTCCGCTTATGTTCCCCGCCCCGCGCCCCGCCCCTAGGTGTAGTAGGTCACA
CSTB-R	GATCTGTGACCTACTACACCTAGGGGCGGGGCGCGGGGCGGGGAACATAAGCGGACATCTATG
3*CSTB-F	AATTCATAGATGCCCCGCCCCGCGCCCCGCCCCTAAAGTCCCCGCCCCGCGCCCCGCCCCTATGTTCCCCGCCCCGCGCCCCGCCCCTAGGTGTAGTAGGTCACA
3*CSTB-R	GATCTGTGACCTACTACACCTAGGGGCGGGGCGCGGGGCGGGGAACATAGGGGCGGGGCGCGGGGCGGGGACTTTAGGGGCGGGGCGCGGGGCGGGGCATCTATG

Table S2. Sequences of core G4s and other structures.

Name	Sequence (5′ to 3′)
CSTB (G4)	TTTGGGGCGGGGCGCGGGGCGGGGTTT
C9orf72 (G4)	TTTGGGGCCGGGGCCGGGGCCGGGGTTT
MYC (G4)	TTTGGGTGGGGAGGGTGGGGTTT
BCL-2 (G4)	TTTGGGCGCGGGAGGAAGGGGGCGGGTTT
C-Kit (G4)	TTTGGGCGGGCGCGAGGGAGGGTTT
MYOG (G4)	TTTGGGTGGGCTGGGAGGTTT
PDGFRB (G4)	TTTGGGGGGGCGGCGGGGTTT
T1B1 (G4)	TTTGTGGTGGGTGGGTGGGTTT
T4B1 (G4)	TTTGTTTTGGTGGGTGGGTGGGTTT
Tel (G4)	TTTGGGTTAGGGTTAGGGTTAGGGTTT
B4-dx2 (G4)	TTGGATCTGAGAATCAGATGTGGGTGGGTGGGT
non-G4	GCCATTCGCCATTCAGG
Hairpin	TCAGCTGCGCAGTACTTTTGTACTGCGCAGCTGA
i-motif (hTelo)	CCCTAACCCTAACCCTAACCCT
Triplex (R1)	GGGAGGGGCGCTTATGGGGAGGGTTTTTCCCTCCCC

Table S3. Sequences of G4s and other structures in dsDNA. Core sequences are marked in red font.

Name	Sequences
CSTB (G4)	5′-CCTGAAGCAGACAGCTAGTGAATTCGGGGCGGGGCGCGGGGCGGGGTACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAGTTTTTTTTTTTTTTTTTTTTTATGAACGCATATTGACAAGGTATCA-5′
C-MYC (G4)	5′-CCTGAAGCAGACAGCTAGTGAATTC--GGGTGGGGAGGGTGGGG--TACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAGTTTTTTTTTTTTTTTTTTTTATGAACGCATATTGACAAGGTATCA-5′
C-Kit (G4)	5′-CCTGAAGCAGACAGCTAGTGAATTC-GGGCGGGCGCGAGGGAGGG-TACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAGTTTTTTTTTTTTTTTTTTTTATGAACGCATATTGACAAGGTATCA-5′
c9orf72 (G4)	5′-CCTGAAGCAGACAGCTAGTGAATTCGGGGCCGGGGCCGGGGCCGGGGTACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAG-TTTTTTTTTTTTTTTTTTTTATGAACGCATATTGACAAGGTATCA-5′
BCL2 (G4)	5′-CCTGAAGCAGACAGCTAGTGAATTCGGGCGCGGGAGGAAGGGGGCGGGTACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAG-TTTTTTTTTTTTTTTTTTTT-ATGAACGCATATTGACAAGGTATCA-5′
Tel (G4)	5′-CCTGAAGCAGACAGCTAGTGAATTCGGGTTAGGGTTAGGGTTAGGGTACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAGTTTTTTTTTTTTTTTTTTTTATGAACGCATATTGACAAGGTATCA-5′
MYOG (G4)	5′-CCTGAAGCAGACAGCTAGTGAATTC---GGGTGGGCTGGGAGG---TACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAGTTTTTTTTTTTTTTTTTTTTATGAACGCATATTGACAAGGTATCA-5′
PDGFRB (G4)	5′-CCTGAAGCAGACAGCTAGTGAATTC---GGGGGGGCGGCGGGG---TACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAGTTTTTTTTTTTTTTTTTTTTATGAACGCATATTGACAAGGTATCA-5′
T1B1 (G4)	5′-CCTGAAGCAGACAGCTAGTGAATTC---GTGGTGGGTGGGTGGG--TACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAGTTTTTTTTTTTTTTTTTTTTATGAACGCATATTGACAAGGTATCA-5′
T4B1 (G4)	5′-CCTGAAGCAGACAGCTAGTGAATTC-GTTTTGGTGGGTGGGTGGG-TACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAGTTTTTTTTTTTTTTTTTTTTATGAACGCATATTGACAAGGTATCA-5′
B4-dx2 (G4)	5′-CCTGAAGCAGACAGCTAGTGAATTCGGATCTGAGAATCAGATGTGGGTGGGTGGGTACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAG-----TTTTTTTTTTTTTTTTTTTT----ATGAACGCATATTGACAAGGTATCA-5′
non-G4 dsDNA	5′-CCTGAAGCAGACAGCTAGTGAATTCAAAAAAAAAAAAAAAAAAAAATACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAGTTTTTTTTTTTTTTTTTTTTATGAACGCATATTGACAAGGTATCA-5′
Hairpin	5′-CCTGAAGCAGACAGCTAGTGAATTCTCAGCTGCGCAGTACTTTTGTACTGCGCAGCTGATACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAG----TTTTTTTTTTTTTTTTTTTT----ATGAACGCATATTGACAAGGTATCA-5′
i-motif (hTel)	5′-CCTGAAGCAGACAGCTAGTGAATTCCCCTAACCCTAACCCTAACCCTTACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAG-TTTTTTTTTTTTTTTTTTTTATGAACGCATATTGACAAGGTATCA-5′
Triplex (R1)	5′-CCTGAAGCAGACAGCTAGTGAATTCGGGAGGGGCGCTTATGGGGAGGGTTTTTCCCTCCCCTACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAG----TTTTTTTTTTTTTTTTTTTT----ATGAACGCATATTGACAAGGTATCA-5′
Bulged dsDNA	5′-CCTGAAGCAGACAGCTAGTGAATTCAGGTCATACTTGCGTATAACTGTTCCATAGT-3′ 3′-GGACTTCGTCTGTCGATCACTTAAG------ATGAACGCATATTGACAAGGTATCA-5′

Table S4. The forward (F) and reverse (R) primer sequences used in real-time PCR.

Name	Sequence (5′ to 3′)
GAPDH-F	GGAGCGAGATCCCTCCAAAAT
GAPDH-R	GGCTGTTGTCATACTTCTCATGG
ACTB-F	AGATGACCCAGATCATGTTTGAGA
ACTB-R	CAGAGGCGTACAGGGATAGCA
NAMPT-F	CAGACCTGATTCTGGAAACCCTC
NAMPT-R	GTGGCAGCAACTTGTAACCCT
NPM1-F	AGAGGCAATGAATTACGAAGGC
NPM1-R	CAAGGGAAACCGTTGGCTGT
RPLP0-F	GCACCATTGAAATCCTGAGTGA
RPLP0-R	CGAACACCTGCTGGATGACC
TBX3-F	CCCGAAGAAGAGGTGGAGGA
TBX3-R	CCAGCCCAGAACATCTCACTTTA
RIC8B-F	GCTTTGACTTGCGCTTGCTC
RIC8B-R	CTCATCGGTCCACTTGATGCTA
HELB-F	TGTAGGAATCTTCAAATCGGTCTT
HELB-R	TTCAGCAAGTTACCAGGTTCAATAC
RNFT2-F	ACATTCCACCTGAAAGAAACCG
RNFT2-R	CCTGATAAGCCCGAGAAGAGC
PAICS-F	GGATGATGCCAATAATGACCCA
PAICS-R	GCCTATAAGAAGTCCAGCAAAGC
MEPCE-F	TGACCCGCTCAGTCTCAATACTT
MEPCE-R	CGGTGTCTATGCCGCTTCC
CASP2-F	CATGCTGGTTAAGGTGAACGC
CASP2-R	GTCCTGGGAACAGGTAGAGGTG
ZNF34-F	CAGGGCACCTCTGGGAAAG
ZNF34-R	CTGATGTGGTCACAGGCTTCTACT
VPS26A-F	GCCAATGTCCGCTTGAGGTAT

VPS26A-R	TGCCCACTTCCATCTTAATAGAGTT
GCSH-F	AGAAACACGAATGGGTAACAACAG
GCSH-R	CCAACTTCAGGGAGACTACAATAAAC
BANP-F	TTGGCATCACAGAATCCGACT
BANP-R	CCGCGAGAAGCTCTTGACC
ZFPM1-F	GCAAGGACTGTGGCATCTGG
ZFPM1-R	TGGGCTTCTCGTCTGTGGC
SDHB-F	CAGGAAGGCAAGCAGCAGTAT
SDHB-R	TGTCTCCGTTCCACCAGTAGC
SNX3-F	TACGAAATCAGGGTCAAGACAAA
SNX3-R	TCTAATTCACTTCGCAGCCATT
EPS15L1-F	ATCAAGTCCCTGAAGTCAACGC
EPS15L1-R	CATCGAGCACCTGGTCATACTG
ZNF708-F	CCTGGGTATTGCTGTGTCTAATTTA
ZNF708-R	GTTTGGCTGCCATCTCGTGT
ZNF738-F	TTACATATTCTCATTTTGCCCAGG
ZNF738-R	CTCATTCACACTTTTACAGCCTTTT
ZNF675-F	TGTTGACATTTAGGGATGTGGC
ZNF675-R	CAGCAATACCCAGGAAGACCAG
CACNG8-F	GGAGGACACGGACTACGACC
CACNG8-R	AGGATGGCGCTAAGGATGG

Reference

1. Li, X.M., Zheng, K.W., Zhang, J.Y., Liu, H.H., He, Y.D., Yuan, B.F., Hao, Y.H. and Tan, Z. (2015) Guanine-vacancy-bearing G-quadruplexes responsive to guanine derivatives. *P Natl Acad Sci USA*, **112**, 14581-14586.
2. Chen, J.N., He, Y.D., Liang, H.T., Cai, T.T., Chen, Q. and Zheng, K.W. (2021) Regulation of PDGFR-beta gene expression by targeting the G-vacancy bearing G-quadruplex in promoter. *Nucleic Acids Res*, **49**, 12634-12643.