Supplementary Information

Small-Molecule-Based Human Genome G4 Profiling Reveals Potential Gene Regulation Activity

Weiwu Zeng[1], Fan Wu[1], Chaoxing Liu[1], Yan Yang[3], Bingyao Wang[3], Yushu Yuan[1], Jiaqi Wang[1], Yuqi Chen[1], Boshi Fu[1], Zhiguo Wu[2,*], and Xiang Zhou[1,3,*]

[1]College of Chemistry and Molecular Sciences, Key Laboratory of Biomedical Polymers of Ministry of Education, Wuhan University, Wuhan 430072, Hubei Province, China.

[2]College of Life Sciences, Wuhan University, Wuhan 430072, Hubei Province, China.

[3]Institude of Advanced Studies, Wuhan University, Wuhan 430072, Hubei Province, China.

List of Contents:

1. General materials and methods
2. Synthetic scheme and procedure
3. CD and Molecular simulation
4. Correlation between biological replicates of input and pulldown libraries
5. GO and KEGG analysis of G4 peaks
6. Verification of G4 peaks
7. Genomic view of representative peaks
8. MTT assay
9. Correlation heatmap between RNA-seq biological replicates
10. GO and KEGG analysis of DEGs
11. ODNs for qRT-PCR, CD and asorbance
12. References

1. General materials and methods

Materials

^{1}H NMR, ^{13}C NMR spectra were collected on Bruker Advance II HD 400 MHz. HRMS was acquired with Thermo Scientific™ Dionex Ultimate 3000 hybrid LTQ Orbitrap Elite Velos Pro (Thermo Scientific, USA). qRT-PCR experiments were performed with CFX96TM Real-Time System. DNA and chromatin fragments concentrations were quantified by Qubit 3.0 (Thermo Scientific, USA). LC-MS data were collected with the Agilent™ 1220 Infinity LC combined with the 6120 Single Quadrupole

mass spectrometer (Agilent Technologies). Chromatins were fragmented with S220 Focused-ultrasonicator (Covaris). All the oligonucleotides were synthesized and purified by Tsingke (Wuhan, China). Proteinase K and Hieff™ qPCR SYBR® Green Master Mix were purchased from YEASEN Biotechnology Co. Ltd., (Shanghai, China). DNA libraries were prepared with ThruPLEX® DNA-seq Kit (Rubicon Genomics, USA) and purified using AMPure XP beads (Beckman Coulter, USA). RNA libraries were constructed with VAHTS Stranded mRNA-seq Library Prep Kit for Illumina®(Vazyme, China). Dynabeads™ MyOne™ Streptavidin C1 (Thermo Scientific, USA) was used for pull-down. Protease Inhibitor Cocktail was purchased from Sigma-Aldrich Corporation®. Cell culture was performed with DMEM (GE Healthcare), FBS (Zhejiang Tianhang Biotechnology) and penicillin-streptomycin (Biosharp), respectively. (+)-Biotin-PEG_{24}-NHS Ester was purchased from Aladdin (Shanghai, China). CD melting studies were performed with Applied Photophysics Chirascan (UK). EasyPure PCR Purification Kit was purchased from TransGen Biotech (Beijing, China).

Determination of the number of counter anions

Trifluoroacetic acid (TFA) samples of five different concentrations, namely 0.0015 mg/mL, 0.0020 mg/mL, 0.0027 mg/mL, 0.0036 mg/mL, 0.0049 mg/mL, resolved in 1‰ amonia acetonitrile (v/v), were subjected to mass spectrometry in negative mode (30% acetonitrile, 70% water, 0.300 mL/min). The corresponding peak area of each TFA concentration was recorded. A linear equation of peak area vs TFA concentration can be acquired by fitting the data. Product samples of four different concentrations, namely 0.057 mg/mL, 0.043 mg/mL, 0.029 mg/mL, 0.025 mg/mL, resolved in 1‰ amonia acetonitrile (v/v), were subjected to mass spectrometry in the same condition as mentioned above. The corresponding peak area of each product concentration was recorded. The number of counter anions can be deduced to be 2 by calculation.

Circular dichroism.

For G4 ODNs and non-G4 ssDNA, CD melting studies were performed using 10 μM oligonucleotides, 100 mM KCl, with 0 μM, 10 μM, 30 μM PDP-PEG_{24}-biotin and 10 μM PDP, respectively. For dsDNA, asorbance studies were performed at 260 nm using 2 μM oligonucleotides, 100 mM KCl, with 0 μM, 2 μM and 6 μM PDP-PEG24-biotin, respectively.

Cell culture

For pull-down experiment, cells were cultured in DMEM, supplemented with 10% FBS and 1 × penicillin-streptomycin. For cell incubation, 1 μM PDP were added to the medium for 3 h, 24 h and 72 h, respectively. The same volume of DMSO was added to control medium.

MTT

In general, cells were seeded in 96-well plates at a density of 10^3-10^4 cells per well and cultured for 12 h. Then the primary cell culture medium was replaced by 100 μL fresh medium with different concentration of PDP and incubated for another 3 h, 24 h and 72 h, respectively. After that, 20 μL of MTT solution (5 mg/mL in PBS) was added into each well maintained at 37 °C for 4 hours. Then the medium containing MTT in each well was removed, followed by adding 100 μL of DMSO to dissolve the purple crystals. The optical density readings at 492 nm were recorded using a plate reader. Each experiment was performed with 5 wells parallelly. The cell viability was expressed by the average values ± standard deviation (SD).

Pulldown procedure

Cells in 10 mL medium were fixed with 1 mL crosslinking buffer (11% formaldehyde; 100 mM NaCl; 1 mM EDTA; 50 mM Hepes, pH 8.0) for 10 min at room temperature followed by quenching with 0.12 M glycine at final concentration at room temperature for 5 min. After centrifugation at 1000 g for 10 minutes, supernatant was discarded with the cell pellet washed by cold PBS for once. Add 10 uL protease inhibitors to 1 mL L1 buffer (50 mM Hepes, pH 8.0; 140 mM NaCl; 1 mM EDTA; 10% glycerol; 0.5% NP-40; 0.25% Triton-X100), or 1 mL L2 buffer (200 mM NaCl; 1 mM EDTA; 10 mM Tris pH 8.0). Resuspend crosslinked cell pellets in 1 mL cold L1 buffer and rotate at 4 degree for 10 minutes. After 5 min of centrifugation at 10000 rpm, 4 °C, supernatant was discarded. Cell pellets were resuspended in 1 mL L2 buffer, rotated at room temperature for 10 min, following which centrifuation was performed as previously mentioned and supernatant was discarded. Wash the cell pellets with 0.5 mL sonication buffer (1×TE + 0.1% SDS), and resuspend them in 1 mL sonication buffer. Then chromatins were sonicated to yield fragments ranging from 100-250 bp, digested by proteinase K and purified. 100 μL of streptavidin beads were incubated with 0.35 μL PDP-PEG_{24}-Biotin (11 mmol/L) in 1 mL PBS for 1 h at room temperature, washed with 2 × 1 mL PBS. 40 μg of DNA fragments (pre-annealed in 100 mM KCl, 10 mM Tris, 1 mM EDTA at 95 °C) were incubated with beads in 1 mL volume (100 mM KCl, 10 mM Tris-HCl, 1 mM EDTA) overnight. Beads were washed with washing buffer (100 mM KCl, 10 mM Tris, 1 mM EDTA, pH 8.0) for 3 times, subjected to decrosslinking to digest the streptavidin, thus enabling G4s to fall off beads. After purification, the DNA was eluted in 35 μL 1×TE.

NGS

Pull-down and RNA-seq libraries were sequenced on the Illumina HiSeq X Ten platform.

Bioinformatic data analysis

For pulldown libraries, raw fastq reads were trimmed by Trim Galore to remove adapters and low quality bases and mapped to the human reference genome GRCh38.p7 using Bowtie2,[1] respectively. Reads with mapping quality less than 10 were

discarded by SAMtools,[2] and duplicates were discarded by Picard. HOMER[3] was employed for further process, including peak calling (makeTagDirectory, findPeaks), annotation (annotatePeaks.pl), motif analysis (findMotifsGenome.pl) and enrichment of Gene Ontology and KEGG Pathway (annotatePeaks.pl). Tag directories were created and subjected to peak calling in factor mode. Only peaks occuring across replicates (mergePeaks) were retained for further analysis (fold change cutoff: 4). For motif analysis, random sequences grabbed from human genome were used as backgroud. After trimming adapters, RNA-seq data were mapped to GRCh38.p7 using Hisat2[4] and only reads with MAPQ over 30 were retained. Differential genes were decided with cufflinks (FDR: 0.05, fold change cutoff: 1.5).[5] The GO and KEGG analysis of RNA-seq was performed using KOBAS 3.0.[6-7] Results of enrichment of Gene Ontology and KEGG, correlation heatmap and scattter plot were plotted using R.

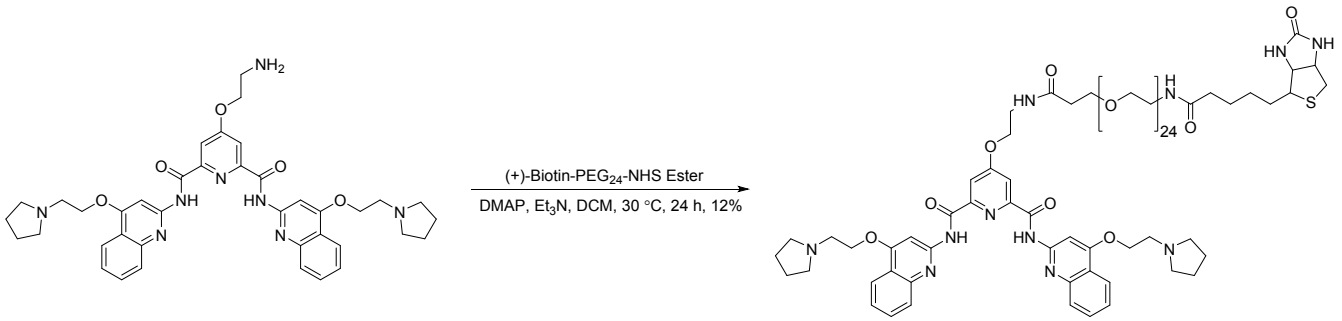
qRT-PCR

For pulldown samples, the input genomic DNA and enriched DNA fragments were used to determine enrichment efficency by qRT-PCR. 1 μL template was added to a mixture of Hieff qPCR SYBR Green Master Mix (5 μL), forward primer (0.5 μL), reverse primer (0.5 μL) and water (3 μL) to give a final volume of 10 μL. A reference sequence (REF) was chosen for calculating the fold of enrichment. Cycling conditions were 95.0 °C for 5 mins, followed by 40 cycles of 10 s at 95.0 °C, 20 s at 63.0 °C and 20 s at 72 °C. For PDP-treated samples, PSMB4 was chosen as the reference gene, and the Cycling conditions were 95.0 ˚°C for 5 mins, followed by 40 cycles of 10 s at 95.0 °C, 20 s at 60.0 °C and 20 s at 72 °C.

2. Synthetic scheme and procedure

PDP·3CF_3COOH was synthesized according to the protocol described in previous report.[8] PDP·3CF_3COOH (35 mg, 0.033 mmol), (+)-Biotin-PEG_{24}-NHS Ester (25 mg, 0.017 mmol), DMAP (0.62 mg, 0.0050 mmol) were stirred in 10 mL DCM at 30 °C in argon atmosphere, after which triethylamine was added dropwise until PDP·3CF_3COOH was completely disolved. The reaction lasted for 24 h, and solvent was removed in vacuo to yield oil liquid which was purified by HPLC (gradient: 10% MeCN/90% H_2O, 0.1% TFA, to 60% MeCN/40% H_2O, 0.1% TFA, over 45 min, Rt=26.5-28.0 min). TFA salt of product (PDP-PEG_{24}-biotin•2CF_3COOH) was acquired as a white solid (5 mg, 0.002 mmol, 12%). ^{1}H NMR (400 MHz, CD_3OD) δ 8.34 (d, J = 8.0 Hz, 2H), 8.17 (s, 2H), 8.08 (s, 2H), 7.95 (d, J = 8.4 Hz, 2H), 7.80 (t, J = 7.3 Hz, 2H), 7.58 (t, J = 7.5 Hz, 2H), 4.82 – 4.71 (m, 4H), 4.48 (dd, J = 7.8, 4.6 Hz, 1H), 4.40 (t, J = 5.2 Hz, 2H), 4.28 (dd, J = 7.9, 4.5 Hz, 1H), 3.98 – 3.92 (m, 4H), 3.80 – 3.67 (m, 6H), 3.64 – 3.56 (m, 96H), 3.52 (t, J = 5.4 Hz, 2H), 3.34 (t, J = 5.5 Hz, 2H), 3.22 – 3.15 (m, 1H), 2.89 (d, J = 4.3 Hz, 1H), 2.72 – 2.66 (m, 1H), 2.50 (t, J = 6.0 Hz, 2H), 2.38 – 1.95 (m, 12H), 1.77 – 1.36 (m, 6H). ^{13}C NMR (100 MHz, CD_3OD) δ 174.72, 173.18, 168.59, 165.59, 164.68, 163.03, 150.99, 149.54, 141.29, 133.14, 126.58, 122.74, 122.58, 118.63, 113.03, 94.09, 70.07, 69.93,

69.86, 69.19, 67.90, 66.82, 65.50, 61.97, 60.22, 55.63, 54.71, 53.21, 39.68, 38.96, 38.17, 36.18, 35.32, 28.40, 28.13, 25.48, 22.61. HRMS (ESI+) calculated for $C_{100}H_{159}N_{11}O_{32}S$ ($[M+2H]^{2+}$): 1030.55262, found 1030.55245.



Scheme S1. Synthesis of PDP-PEG_{24}-biotin

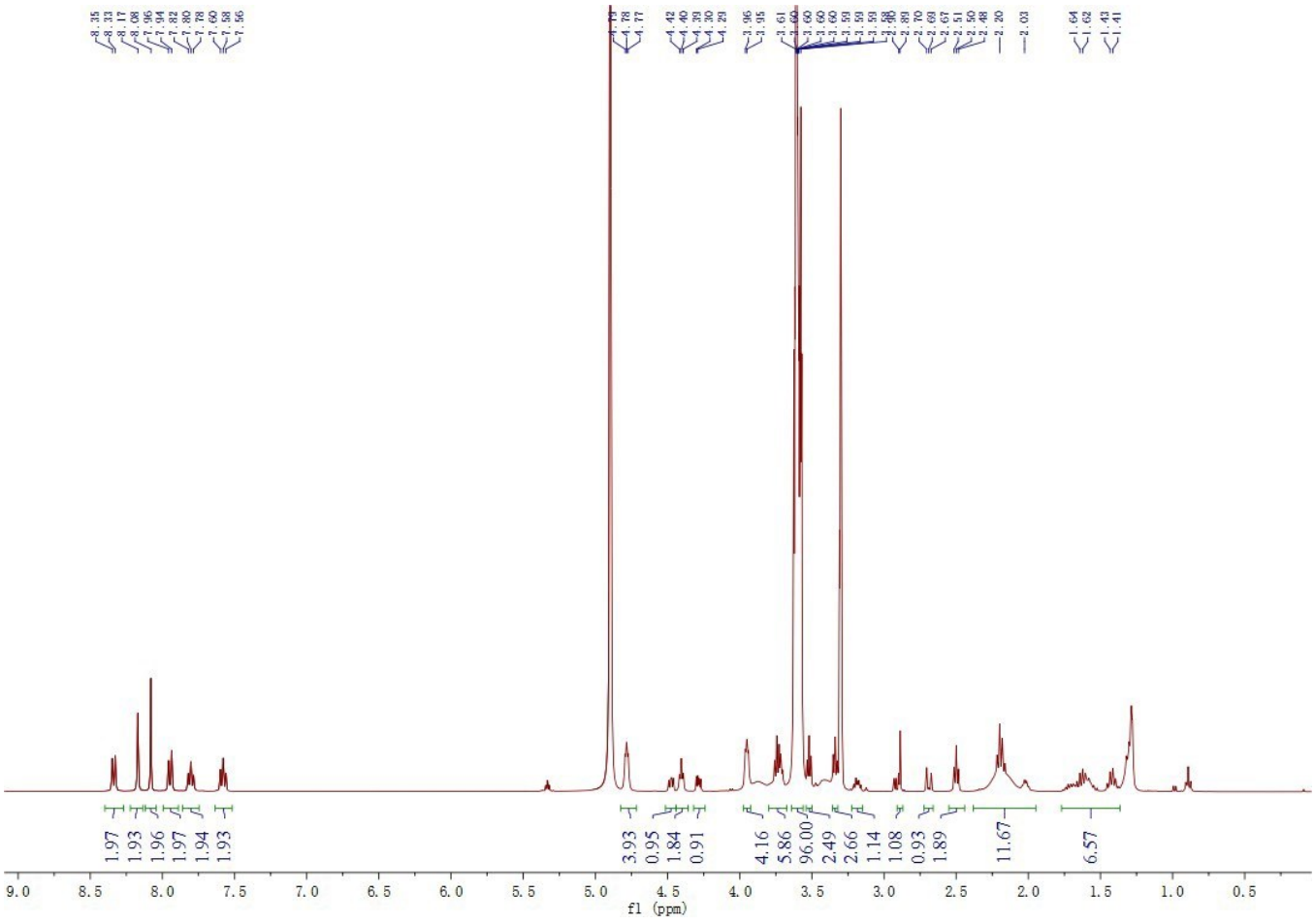


Figure S1. ^{1}HNMR spectrum of PDP-PEG_{24}-biotin

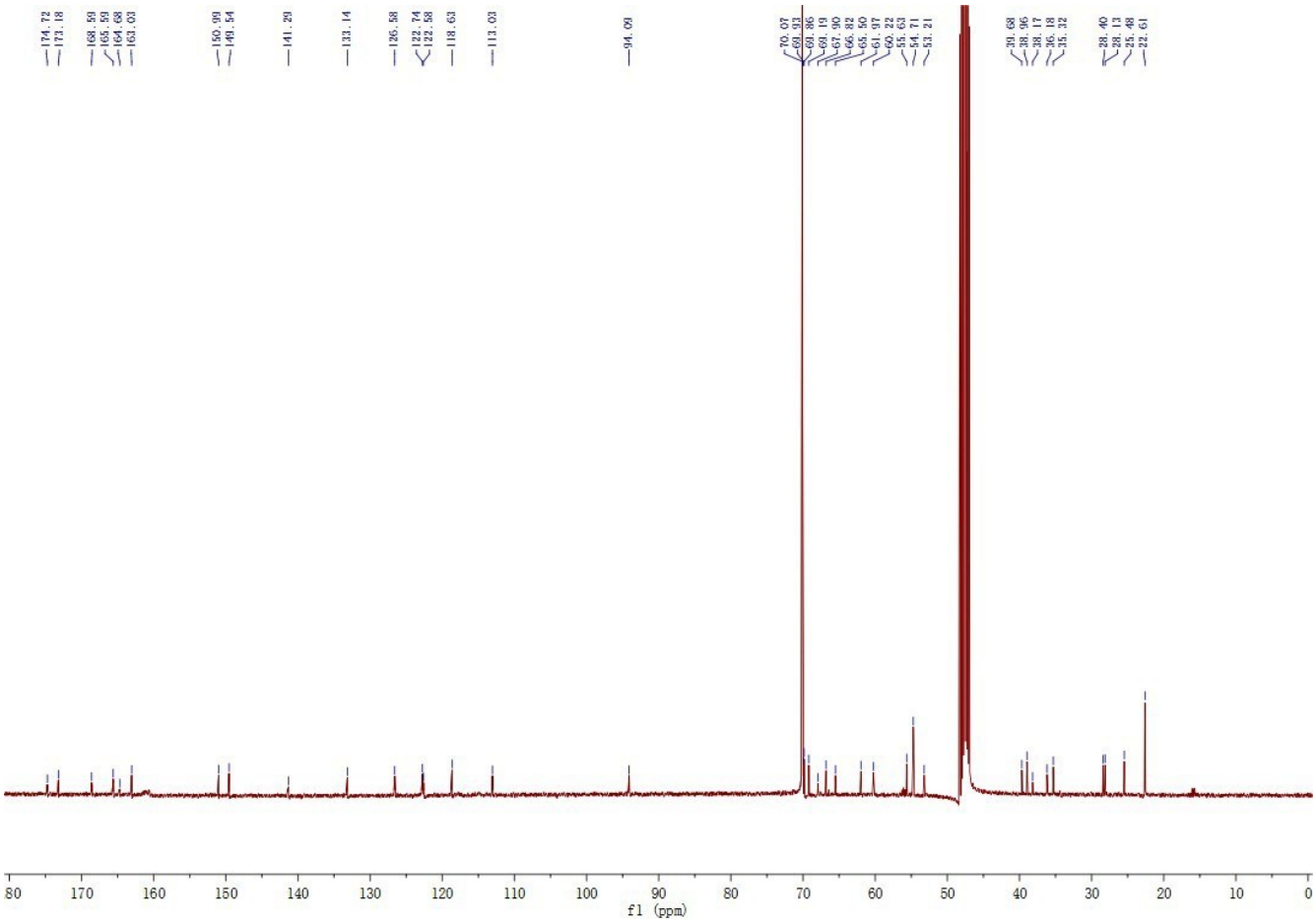


Figure S2. ^{13}C NMR spectrum of PDP-PEG_{24}-biotin

3. CD and molecular simulation

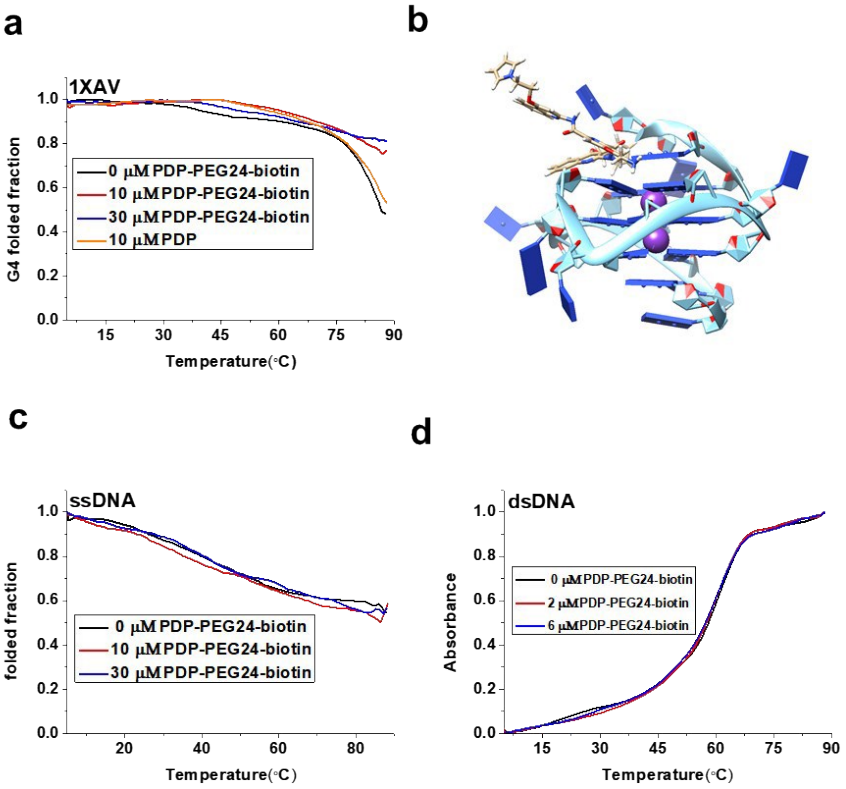


Figure S3. (a) Stabilization effect of PDP-PEG$_{24}$-biotin and PDP on G-quadruplex oligo 1XAV. (b) Molecular simulation reveals the interaction pattern between 1XAV and PDP. (c)(d) PDP-PEG$_{24}$-biotin showed no significant stabilization for non-G4 ssDNA (c) and non-G4 dsDNA (d).

4. Correlation between biological replicates of input and pulldown libraries

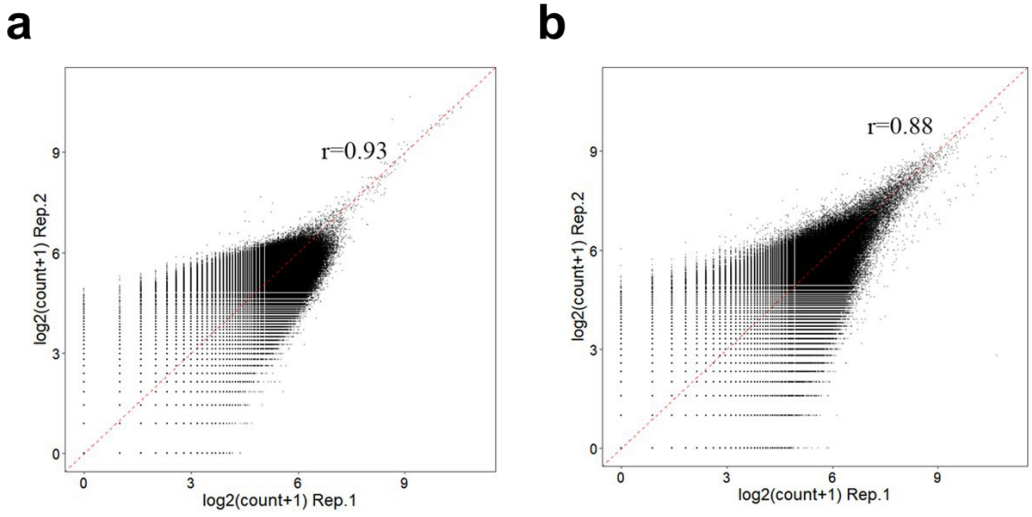


Figure S4. Scatter plots showing correlation between biological replicates of input and pulldown libraries, respectively. Reads were counted in windows of 1 kb. (a) Correlation between input biological replicates. (b) Correlation between pulldown biological replicates.

5. GO and KEGG analysis of G4 peaks

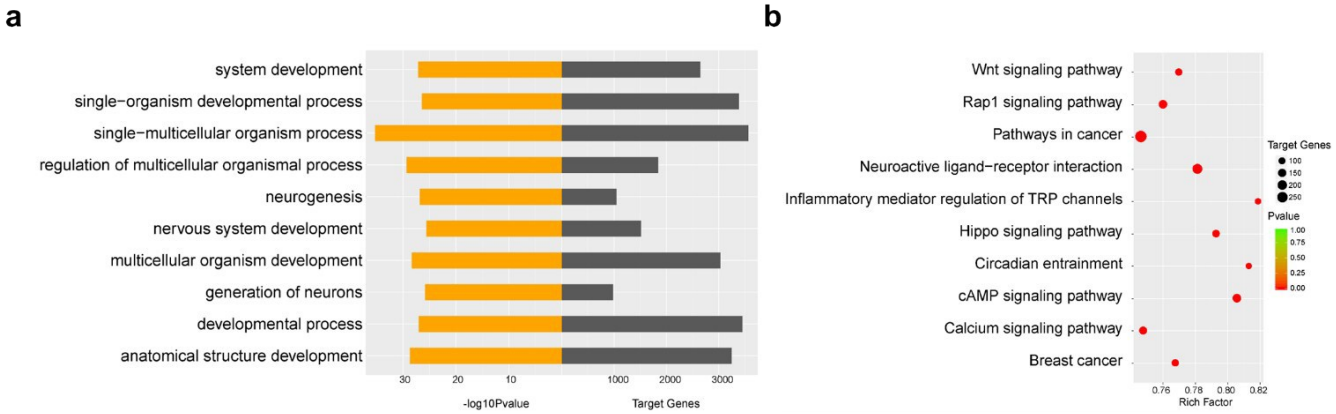


Figure S5. GO and KEGG analysis of G4 peaks

6. Verification of G4 peaks

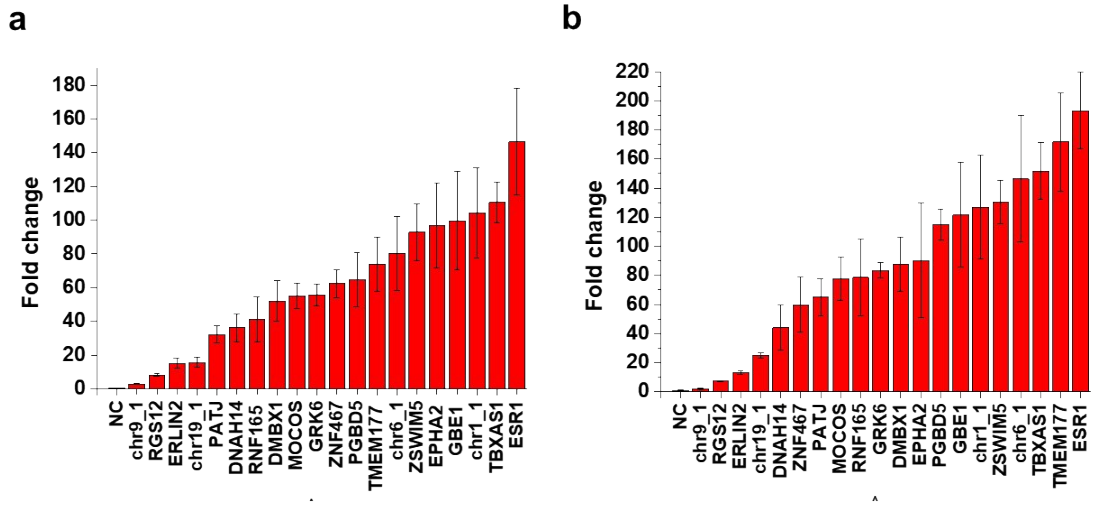


Figure S6. Fold enrichment of G4 peaks using 1.3 μM (a) and 3.85 μM (b) of PDP-PEG24-biotin calculated by qRT-PCR. NC, negative control.

7. Genomic view of representative peaks

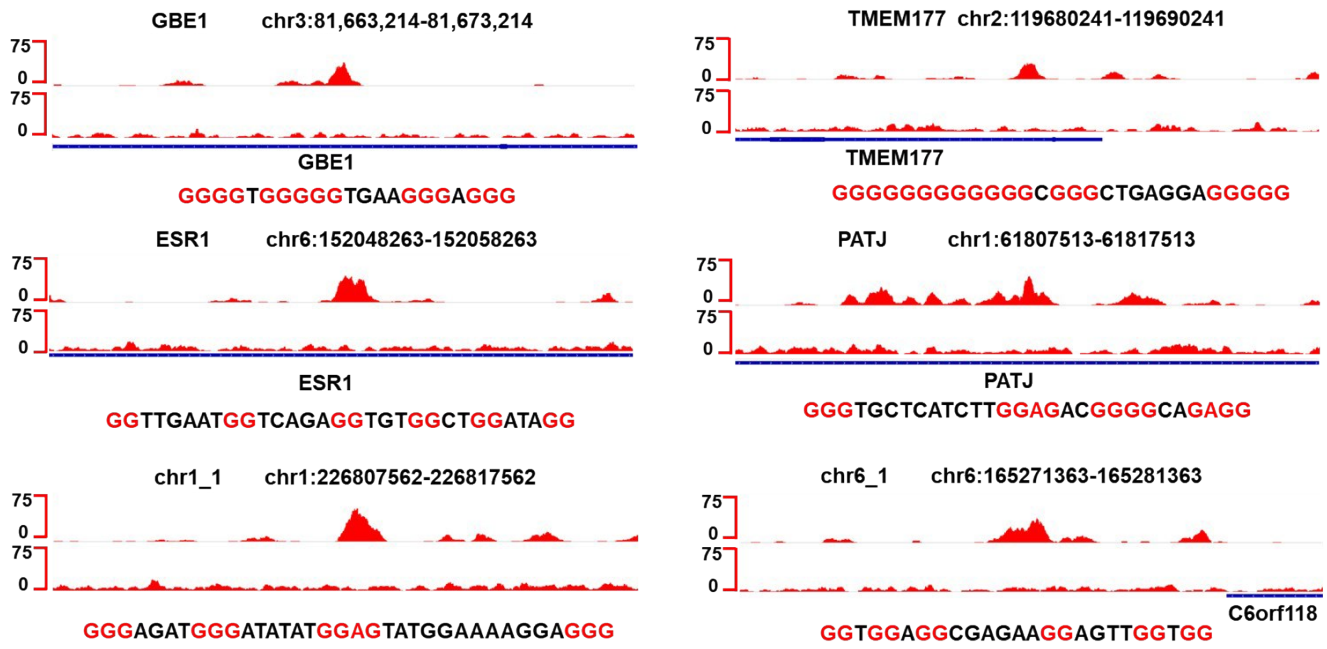


Figure S7. Visualization of G4 peaks in IGV (Intergrative Genomics Viewer). Tracks, from top to bottom, are pulldown, input and genomic annotation, respectively. Potential G4s in corresponding peaks are marked in red. The genomic coordinate shows the range of tracks (10 kb).

8. MTT assay

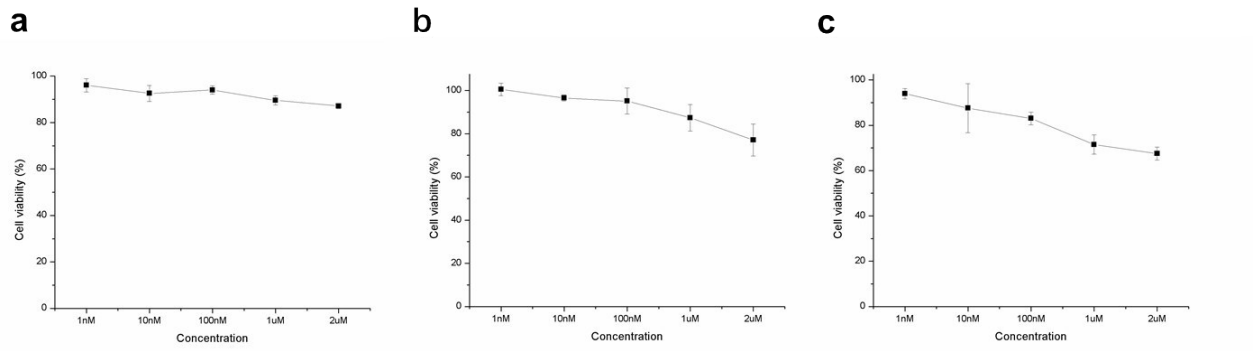


Figure S8. MTT assay for cell incubation of 3 h (a), 24 h (b) and 72 h (c).

9. Correlation heatmap between RNA-seq biological replicates

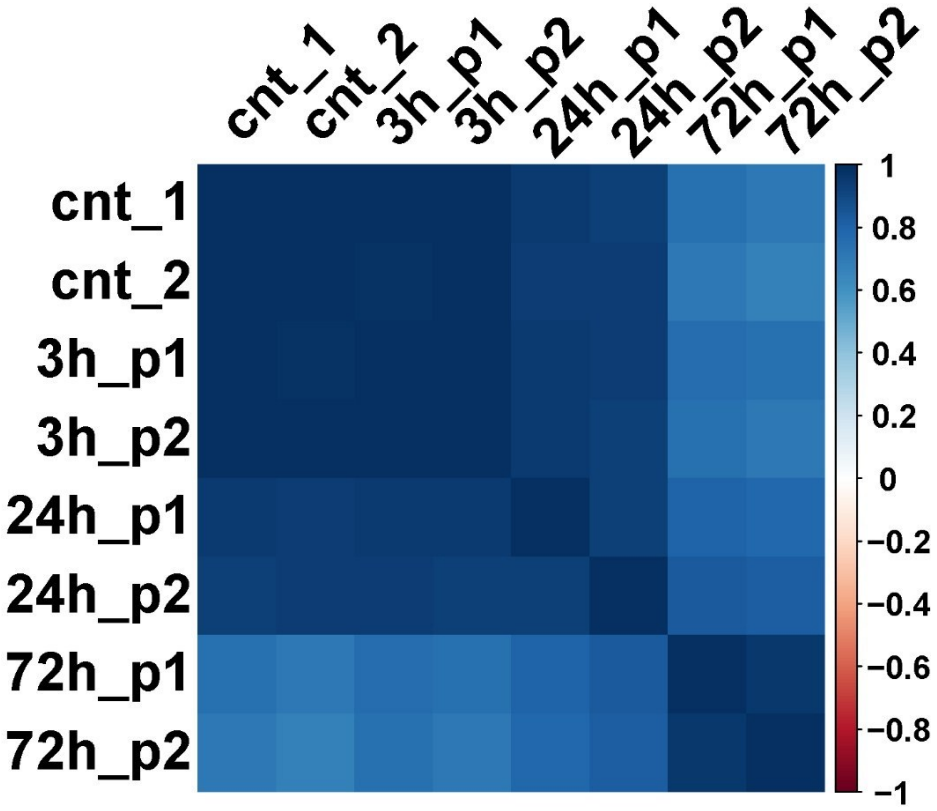


Figure S9. Correlation heatmap between RNA-seq biological replicates

10. GO and KEGG analysis of DEGs

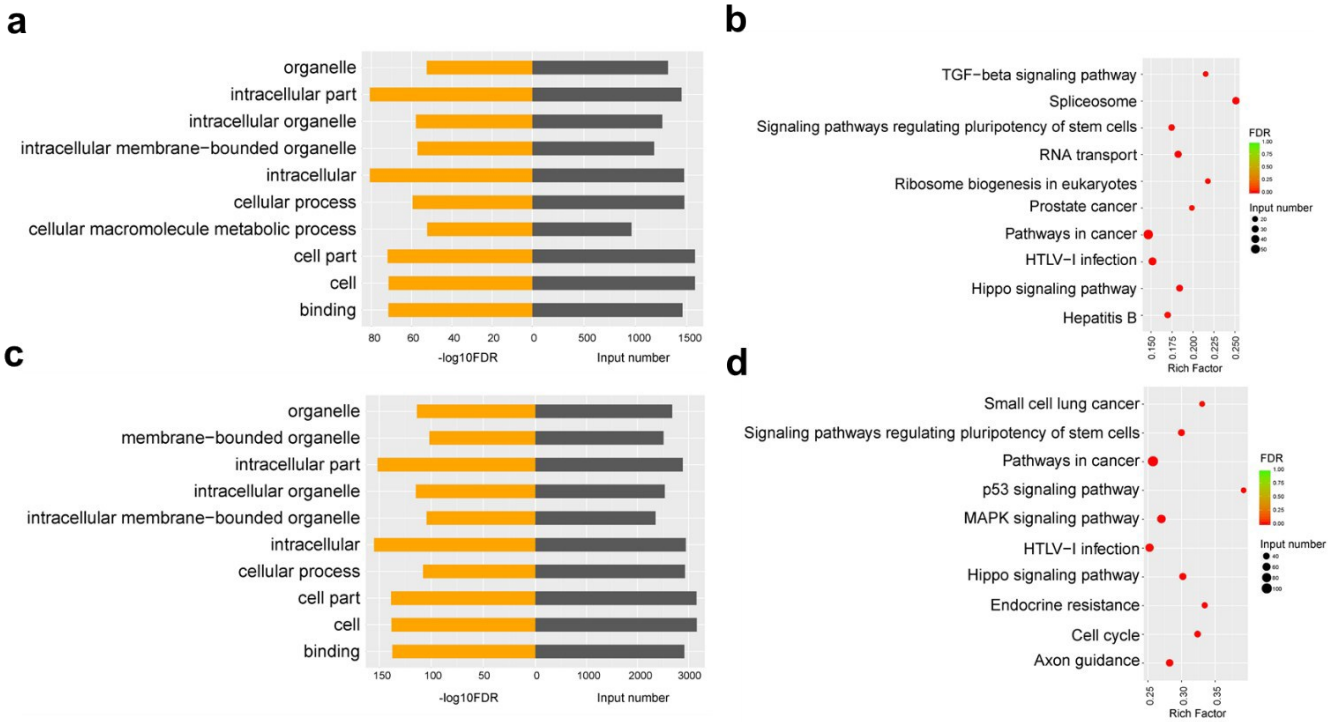


Figure S10. (a), (b) GO and KEGG analysis of DEGs of 24 h incubation. (c), (d), GO and KEGG analysis of DEGs of 72 h incubation. Only the top ten terms are showed in each panel.

11. ODNs for qRT-PCR, CD and asorbance

Table S1. qRT-PCR primers and amplicons for validation of G4 peaks

chr1_1	Primer	F: AACCCAGGAGCCCTTCATGATCC R: GGTGACCTGGCATGAATTGGGTG
	Sequence	GGTGACCTGGCATGAATTGGGTGATCAGAGAAGGGTGGATGGGGGAGACTGGCTGGGAGATGGG ATATATGGAGTATGGAAAAGGAGGGATCATGAAGGGCTCCTGGGTT
chr6_1	Primer	F: CTCCACTTACAATCAGTACCACCA R: TACAGGTAGGGAGAGCAGGAGT
	Sequence	TACAGGTAGGGAGAGCAGGAGTTGAAGGTACAGATATCAGGTGGAGGCGAGAAGGAGTTGGTGG TACTGATTGTAAGTGGAG
PATJ	Primer	F: ATTGGTGGGTGCTCATCTTGGA R: AAAAGTGAAGCAAATCAGAGGTGGT
	Sequence	ATTGGTGGGTGCTCATCTTGGAGACGGGGCAGAGGTACAGTTCTGTAAGAGTCAGACCACCTCTGA TTTGCTTCACTTTT
MOCOS	Primer	F: ACGAATTTTTCAGGCAGTCACAGGC R: CCTCTAAACCCTTCTCTGAGCCCAT
	Sequence	ACGAATTTTTCAGGCAGTCACAGGCGGGATGGTTACCAGGGAGCAGCATGGCAGGGCTTGGCATTC

		ATGGGCTCAGAGAAGGGTTTAGAGG
ERLIN2	Primer	F: TACTTCAAAGTTGCCCTTCTCTGTT R: TTAAAATCCATCACTCCCAGGCACA
	Sequence	TACTTCAAAGTTGCCCTTCTCTGTTCTGACTCCTGGGACTTCTGGTCCTGGGCACACTTTTTGCAGGCA ACAAAATGTGCCTGGGAGTGATGGATTTTAA
chr19_1	Primer	F: ATGTTTTCTGAGCTGAACTGGGACC R: AGAAACCTTTGTGAGATTGATGCTGA
	Sequence	ATGTTTTCTGAGCTGAACTGGGACCTGAAGAAGGTTCACAGCTGGTAAGGTGTTTGGGTCCTATTAC TGGTATCAGCATCAATCTCACAAAGGTTTCT
GRK6	Primer	F: CAGATGGGATGGACGGACAGAGA R: TCTCCTCTCTCTTGCCTTACCCAAA
	Sequence	CAGATGGGATGGACGGACAGAGAGGAGGGGGACCTTAGGCAGGGGATTAGGCAAGGATAGGGCA CCAAGTTTGGGTAAGGCAAGAGAGAGGAGA
TBXAS1	Primer	F: TAATGGGAAAGAAGTGTAGCTGGGTT R: ATTTTCAGGTCTAGTCCATTCCCAC
	Sequence	TAATGGGAAAGAAGTGTAGCTGGGTTGACGGAGGGGTGGAAAATGCTAGATCCATTGTGTGGTGG GAATGGACTAGACCTGAAAAT
ESR1	Primer	F: AAATGGTTGAATGGTCAGAGGTGTG R: AGTCCCTCCCTTAGTCTATCTCACT
	Sequence	AAATGGTTGAATGGTCAGAGGTGTGGCTGGATAGGAGCGACCAAGATGTTGAATGTCAGAAGGTCC AAAGGAAAAGATGGTAGAGTGAGATAGACTAAGGGAGGGACT
chr9_1	Primer	F: TCCCTAGCAGGGGCTGGTTCTA R: CATTCCCATCACACACATTCCCACC
	Sequence	TCCCTAGCAGGGGCTGGTTCTAGGCTCCGAGCAAGCAGGCAGTGGGAGGCGGGTGGGAATGTGTG TGATGGGAATG
DMBX1	Primer	F: TTGATCTCCCTCTTCCTGTCTCTGT R: GACAAGGAGAAAGATCAAGGGTGGA
	Sequence	ACAAGGAGAAAGATCAAGGGTGGAGAATACAGAGGCTGGGCCGAGACGGGGGCAGGGACCAGGA CAGAGACAGGAAGAGGGAGATCAA
RGS12	Primer	F:AATGGGTGGGGTAGGAAGGACTTTT R: GACACACCCCATCATGTCCTCAGAC
	Sequence	AATGGGTGGGGTAGGAAGGACTTTTCAGTTGAGGAGGAGCAGCGCAGTGCAGGGAGACCCAGGC GGGAATGGGTCTGAGGACATGATGGGGTGTGT
ZSWIM5	Primer	F: GGACCCTGTCTGTGAAAGAAAGAGAG R: CTAAACACCAACTACTCCCTCCCTCC
	Sequence	GGACCCTGTCTGTGAAAGAAAGAGAGAGAGAGAGAGACAGAGGGAGAGAGGAAGGAAAGGAGG GAGGGAGTAGTTGGTGTTTA
PGBD5	Primer	F: ATGCCATGGAGATAAGGTAGGTGGA R: CACACTCAGCTCTTTGGGAACCATA
	Sequence	TGCCATGGAGATAAGGTAGGTGGAGTGGGGATGCTAGAGGACTATCCCGGGCCATCGATATGGTTC CCAAAGAGCTGAGTGTG
DNAH14	Primer	F: GGTTTGCTGCTTCCATCAAAACCAT R: AACAAATGGACACAGAGAGGGGAAC
	Sequence	AACAAATGGACACAGAGAGGGGAACATCACACACTGGGGCCTGTTGCGGGGTGGGGGCAAGGGG AGGGAGAGCATTAGGACAAATACCTAATGCATGCGGGGCTGAAAACCCAGATGATGGTTTTGATGG AAGCAGCAAAC
EPHA2	Primer	F: GTTGGCGGGTAGACGAGTGAATG R: TTTCACATGCACATGGTAGGTTTGC
	Sequence	GTTGGCGGGTAGACGAGTGAATGGATGGATGATGGATGGATGGATGGATGGATGGATGGATGAAT GGATGGATGGATGTAGGGGTGGATGGGTGGAGGGGTGGATGGATGGATGAGTGGATGGGTGGAC AGGTGGATGGAAGAGTGAGTGGGAGAGAGAATGGGTGGCAAACCTACCATGTGCATGTGAA
ZNF467	Primer	F: AAAGGAGAAGGAAGGGAAACCGGG R: CCTCCTCTCAGCTCTTTTCCCCAAC
	Sequence	AAGGAGAAGGAAGGGAAACCGGGGTGGGGGAGGAGAGGAGGGAGGGGAGAGGAGAACTGGGT TGGGGAAAAGAGCTGAGAGGAGG

TMEM177	Primer	F: CTCTGAGACTCCCCTCCCTGTATTC R: AACTACAACACACAAAGACTCCCCC
	Sequence	TCTGAGACTCCCCTCCCTGTATTCTGGCTGTATCCAGATGTGTGTGTGTGTGTTTGTGTGTGTGTGTG TGTGCCTGCAAAGGGGGGGGGGGGCGGGCTGAGGAGGGGGAGTCTTTGTGTGTTGTAGTT
GBE1	Primer	F: ACAATGAGAATACACGGGGAGGAGA R: CTGTACCTACTGACCCGTTCTCTGA
	Sequence	ACAATGAGAATACACGGGGAGGAGAACAACACACACCAGGGCCTGTTGTGGGGTGGGGGTGAAG GGAGGGAACTCAGAGAACGGGTCAGTAGGTACA
RNF165	Primer	F: CATTTGTTCCAAGCTGCTGCTTCTC R: GAACTGGAAAAGGTGGAAGTGGAGG
	Sequence	AACTGGAAAAGGTGGAAGTGGAGGGGAAGGGGAAGAGGAGAAGGAGGAGAAAGAGGAAGAGG AAGAGGAGTGAGAAGGTCAGAGAAGAATAGGAGAAAGAGAGAGAAGAGAAGCAGCAGCTTGGAA CAAATG
REF	Primer	F: GAAGACAAGATTCTGACCCTGAA R: TAGCAAGGCAGGAAACGTGAA
	Sequence	GAAGACAAGATTCTGACCCTGAAATAAAGGCTAAGCCTCTTGGCAAGCTCAGACGTGGTTTGTCAAG TTCACGTTTCCTGCCTTGCTA
NC	Primer	F: AGAGTCATGGAGATTATGAAGCCT R: TCCTGCATACCGCTGTCAAT
	Sequence	TCCTGCATACCGCTGTCAATTATTCGCAGATGGATTCCAGAGTAGAAGGCTTCATAATCTCCATGACT CT

Table S2. qRT-PCR primers for validation of DEGs

	Forward	Reverse
ZSWIM5	TACTGACAGCACCCTGCTCAAC	CGCACACCCACTTCTGTAGCA
ZNF467	CTGGGTTACCATGAGAGAGACCTTG	CCCAGTGCTCTCTCTTCCCTAGAA
GBE1	ACGCTCACAAACAGGCGCTA	TGCCAGCGACTTATCCCCAAC
TMEM177	AGCTTCTGTCGGGCAACCTG	CGATGTTCCCGCTGGGTGTA
EPHA2	TGGGACCTGATGCAGAACATC	CTCAGCCTCTCCTCGGTACA
RNF165	GCAGCCGGAGCATGTTAGCTT	TTGGGGCTGTCGTTGGTTCG
PSMB4	TCGGCGTTAAGTTCGAGGGC	TTGCGGAAACGAGCCAAGGA

Table S3. Sequences of G4-TMEM177、G4-GBE1, ssDNA and dsDNA

G4-TMEM177	GGGGGGGGGGGGCGGGCTGAGGAGGGGG
G4-GBE1	GGGCCTGTTGTGGGGTGGGGGTGAAGGGAGGG
ssDNA	CTGTACCTACTGACCCGTTCTCTGA
dsDNA	F: AAGACAAGATTCTGACCCTGAAATAAA R: TTTATTTCAGGGTCAGAATCTTGTCTT

12. References

1. B. Langmead and S. L. Salzberg, *Nat. Methods*, 2012, **9**, 357-359.
2. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and S. Genome Project Data Processing, *Bioinformatics*, 2009, **25**, 2078-2079.
3. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass, *Mol. Cell*, 2010, **38**, 576-589.
4. D. Kim, B. Langmead and S. L. Salzberg, *Nat. Methods*, 2015, **12**, 357-360.
5. C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn and L. Pachter, *Nat. Protoc*., 2012, **7**, 562-578.
6. J. Wu, X. Mao, T. Cai, J. Luo and L. Wei, *Nucleic Acids Res*., 2006, **34**, W720-W724.
7. C. Xie, X. Mao, J. Huang, Y. Ding, J. Wu, S. Dong, L. Kong, G. Gao, C.-Y. Li and L. Wei, *Nucleic Acids Res*., 2011, **39**, W316-W322.
8. S. Müller, S. Kumari, R. Rodriguez and S. Balasubramanian, *Nat. Chem.*, 2010, **2**, 1095-1098.