Supporting Information

Direct sequencing of 2'-deoxy-2'-fluoroarabinonucleic acid (FANA) using Nanopore Induced Phase-Shift Sequencing (NIPSS)

Shuanghong Yan[#,[1,3]], Xintong Li[#,*,[4]], Panke Zhang[[1,2,3]], Yuqin Wang[[1,3]], Hong-Yuan Chen[[1,2,3]], Shuo Huang[*,[1,2,3]] and Hanyang Yu[*,[4]]

[1] State Key Laboratory of Analytical Chemistry for Life Sciences, Nanjing University, 210023, Nanjing, China

[2] Collaborative Innovation Centre of Chemistry for Life Sciences, Nanjing University, 210023, Nanjing, China

[3] School of Chemistry and Chemical Engineering, Nanjing University, 210023, Nanjing, China

[4] College of Engineering and Applied Sciences, Nanjing University, 210023, Nanjing, China

[#] These authors contribute equally to this work

Corresponding Author:

Prof. Shuo Huang: shuo.huang@nju.edu.cn

Miss Xintong Li: dg1634021@smail.nju.edu.cn

Prof. Hanyang Yu: hanyangyu@nju.edu.cn

Table of Contents

Materials

Streptavidin, phi29 DNA polymerase (phi29 DNAP) and dNTPs were purchased from New England Biolabs. Hexadecane, pentane, ethylenediaminetetraacetic acid (EDTA) and Genapol X-80 were from Sigma-Aldrich. Potassium chloride, sodium chloride, sodium hydrogen phosphate and sodium dihydrogen phosphate were from Aladdin (China). 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) was supplied by Avanti Polar Lipids. E. coli strain BL21 (DE3) was obtained from Biomed (China). Dioxane-free isopropyl-propthiogalactopyranoside (IPTG), kanamycin sulfate, DL-Dithiothreitol (DTT) and imidazole were purchased from Solarbio (China). 4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid (HEPES) was from Shanghai Yuanye Bio-Technology (China). Ammonium sulfate [$(NH4)_2SO_4$] was from XiLong Scientific. Magnesium chloride was from Macklin. LB broth and LB agar were from Hopebio (China). Hydrochloric acid was from Sinopharm (China). The RT521K polymerase was expressed and purified as described previously.[1]

DNA oligonucleotides such as DNA30, primer and blocker (Supporting Table S1) were purchased from Genscript (New Jersey, USA). 5'-phosphate modified DNA oligonucleotides were from Sangon Biotechnology (Shanghai, China). Monomeric DNA, monomeric FANA, biotin and abasic phosphoramidites for nucleic acid synthesis were obtained from Biosyntech (Suzhou, China).

Supporting Methods

1. Operating buffers. 1M KCl buffer (1 M KCl, 10 mM HEPES at pH 8.0) is used for static pore blockage experiments. 0.3 M KCl buffer (0.3 M KCl, 10 mM HEPES at pH 7.5,

10 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$, and 4 mM DTT) is used for nanopore sequencing. Both aqueous buffer are prepared with Milli-Q water and membrane filtered (0.2 μm, Whatman) prior to use.

2. M2 MspA preparation. The M2 MspA nanopres[2] are expressed with E Coli. BL21 (DE3) and purified with nickel affinity chromatography as previously published[3].

3. Streptavidin tethered FANA homopolymers. Chimeric DNA-FANA with a 3' biotin modification is mixed with streptavidin with equal molar ratio at room temperature for 10 min to create streptavidin-tethered FANA complex prior to use[4].

4. Synthesis and purification the long chimeric FANA-DNA. Chimeric DNA-FANAs including FANA polyA, polyC, polyU and FANA42 (Supporting Table S1) were synthesized on an automated ABI 349 DNA synthesizer. Parts of FANAx (51 nucleotides from the 5' end) and FANA30 (48 nucleotides from the 5' end) were also synthesized on an automated ABI 349 DNA synthesizer, de-protected in concentrated NH4OH at 55 ℃ for 16 h, precipitated with ethanol and purified by denaturing polyacrylamide gel electrophoresis (PAGE). Full-length FANAx and FANA30 (Supporting Table S1, Supporting Figure S3) were prepared by T4 ligase mediated ligation of chimeric FANA-DNA and 5'-phosphate modified DNA. The full-length chimeric FANA-DNA product was purified by denaturing PAGE.

5. FANA sequencing library construct. Chimeric DNA-FANA templates, primer and blockers are mixed with 1:1:2 molar ratio and thermal annealed (95 ℃ for 3 min and followed by the programmed temperature drop to 25 ℃ within 14 min.) on a PCR thermal cycler (ABI 2720, Applied Biosystems) as reported[2].

6. Nanopore measurements. All electrophysiology measurements were

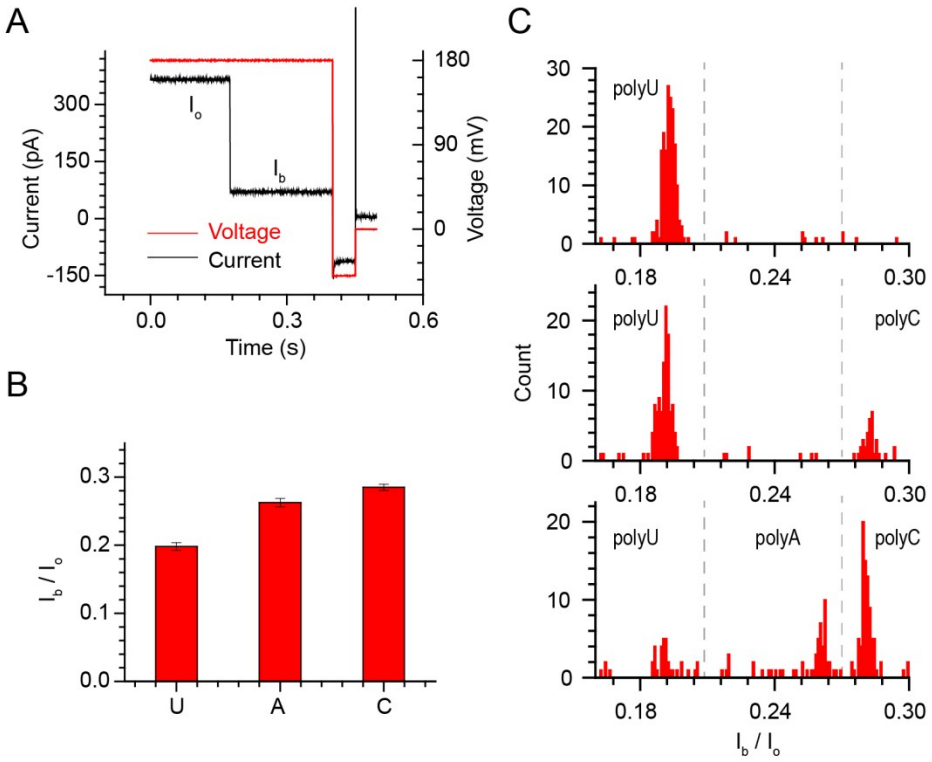
performed with single MspA nanopores inserted in a self-assembled lipid bilayer formed by 1,2-diphytanoyl-sn-glycero- 3-phosphocholine (DPhPC, Avanti Polar Lipids) as reported previously.[3] This lipid bilayer separates two compartments containing electrolyte buffers (Supporting Methods). The compartment which is electrically grounded during the electrophysiology recording, is termed the *cis* side and the other compartment is named the trans side. A voltage protocol is applied across the membrane to drive directional electrophoretic flow of ions and analytes through the pore. Conventionally, MspA nanopore is added to the *cis* side. Spontaneous pore insertions can be monitored via the electrophysiology recording. A membrane with a single MspA nanopore could be formed by manually exchanging the *cis* compartment with fresh buffer immediately after the first pore insertion. All single channel recording traces were acquired with an Axon 200B patch clamp amplifier and digitized by a Digidata 1550A plus digitizer (Molecular Devices).

For static pore blockage experiments, streptavidin-tethered homo FANA was added to the *cis* compartment to reach a 20 nM final concentration. Cyclic recordings for open pore levels (I_0), blockage levels (I_b) were performed 1000 times with a pre-defined voltage protocol (+180 mV for 900 ms, -50 mV for 100 ms, 0 mV for 100 ms) and automatically analyzed by a previously published custom program.[4] The trace is recorded with 10 kHz sampling rate and filtered with a 1 kHz cut-off frequency.

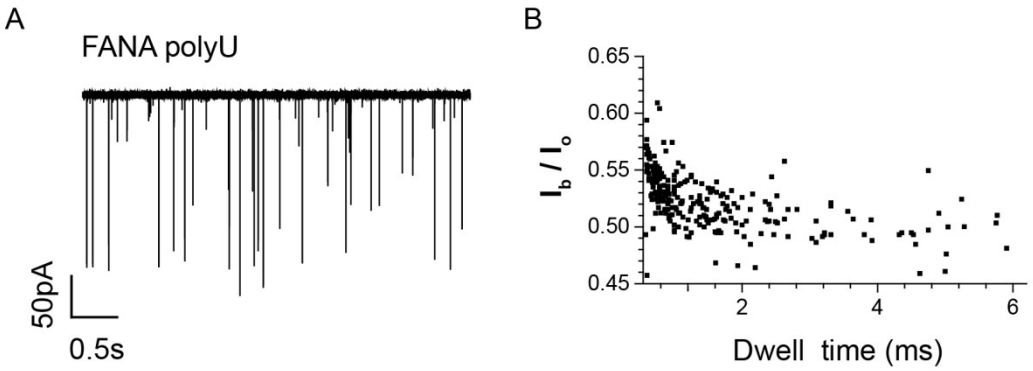
For nanopore sequencing measurements, sequencing libraries thermally annealed at 95 ℃ for 3 min followed by slow cooling to 25 ℃ within 14 min, were added to *cis* to reach a 5 nM final concentration. Prior to data acquisition, 5 nM of phi29 DNAP and 1 mM of dNTP mixture were added to the *cis* side. The trace was continuously recorded at +180 mV with 100 kHz sampling rate and filtered with a

10 kHz cut off frequency following published procedures.[5] Nanopore sequencing events were extracted by Clampfit 10.7, saved as separate .abf files and further analyzed by a custom program written by LabView **(Supporting Figure S10)**.

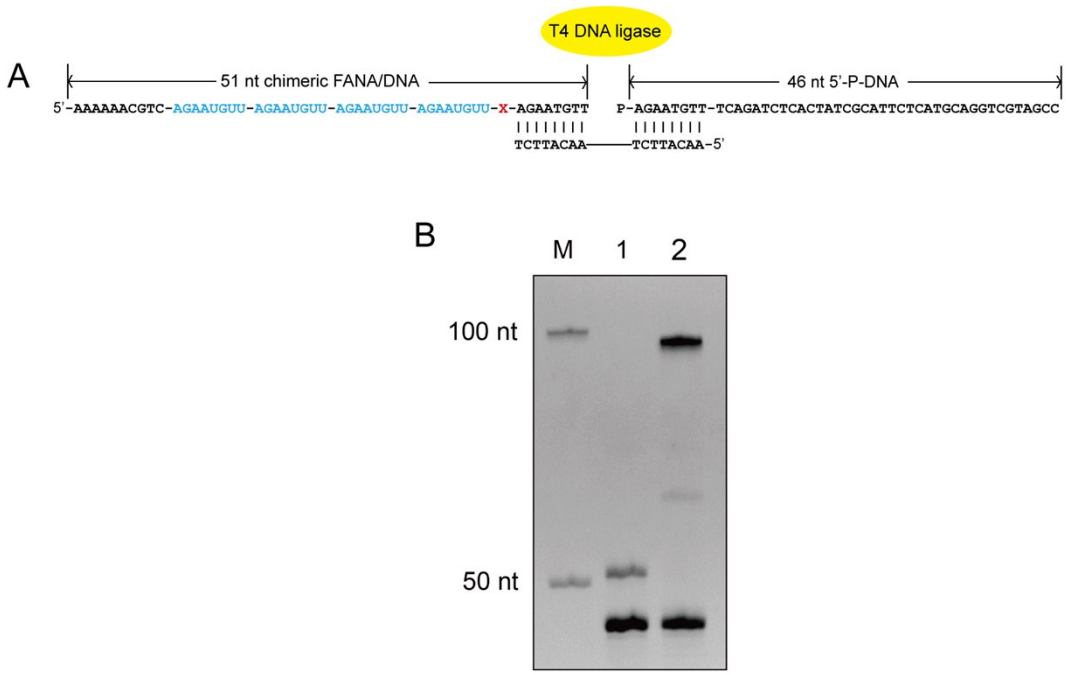
7. Reverse transcription. The phi29 DNA polymerase mediated FANA reverse transcription was carried out with 1 U/μl enzyme in 1X phi29 DNA polymerase reaction buffer (50 mM Tris-HCl, 10 mM MgCl2, 10 mM (NH4)2SO4, 4 mM DTT, pH 7.5). The RT521K polymerase mediated FANA reverse transcription was carried out with 0.1 mg/ml enzyme or 1x ThermoPol buffer (20 mM Tris-HCl, 10 mM (NH4)2SO4, 10 mM KCl, 2 mM MgSO4, 0.1% Triton X-100, pH 8.8). The reaction mixture also contained 10 nM RT primer, 50 nM template (Supporting Table S1), 50 μM of each dNTP, and was heated at 94 ℃ for 4 min then cooled on ice for 15 min. The reactions were then incubated at room temperature, 30 ℃ or 65 ℃ for 1 h. The reaction was terminated by addition of 10 μl stop solution (6X TBE, 8 M urea) and heating at 90 ℃ for 5 min. The reaction products were analyzed by 12% denaturing PAGE.



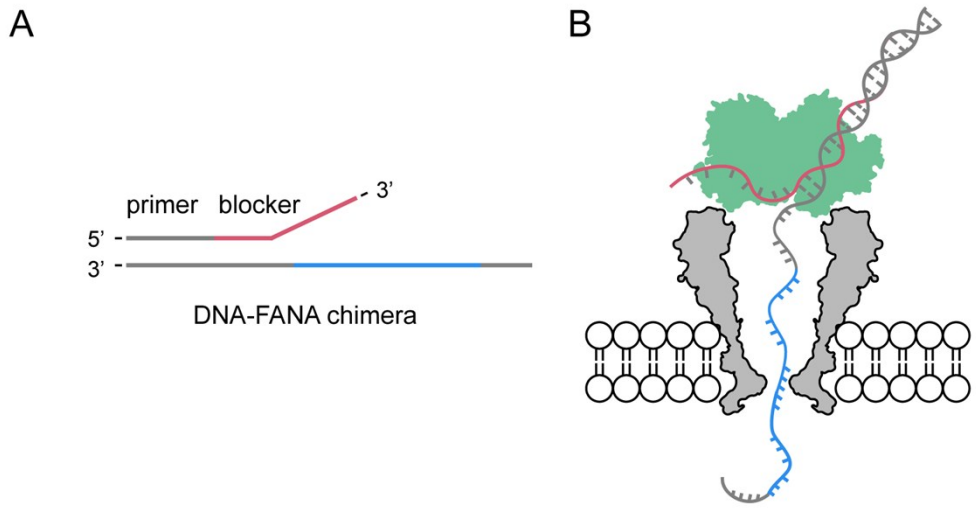
Supporting Figure S1. Static pore blockage by FANA homopolymer. (a) A representative cycle of static pore blockage by FANA homopolymer. At +180 mV, an open nanopore (I_o, open pore current) could be blocked by negatively charged FANA, generating a reduced, sequence specific signal (I_b, blocked current,) for FANA discrimination. After the acquisition, the applied voltage is reversed (-50 mV) to unload the FANA in the pore and switched to 0 mV for a new cycle. **(b)** Statistics (mean±standard deviation) of I_b/I_o of different FANA homopolymers (N=3 for each homopolymer). **(c)** Homopolymer identity determination. The homopolymers are sequentially added (polyU, polyC and polyA) and measured with the same pore. The homopolymer identities are determined according to the order of appearance for corresponding peaks. PolyA is added in the end since spontaneous hybridization between polyA and polyU will results in reduced measurement counts from these two types of analytes. The experiment was carried out in the electrolyte buffered solution of 1 M KCl, 10 mM HEPES at pH 8.0.



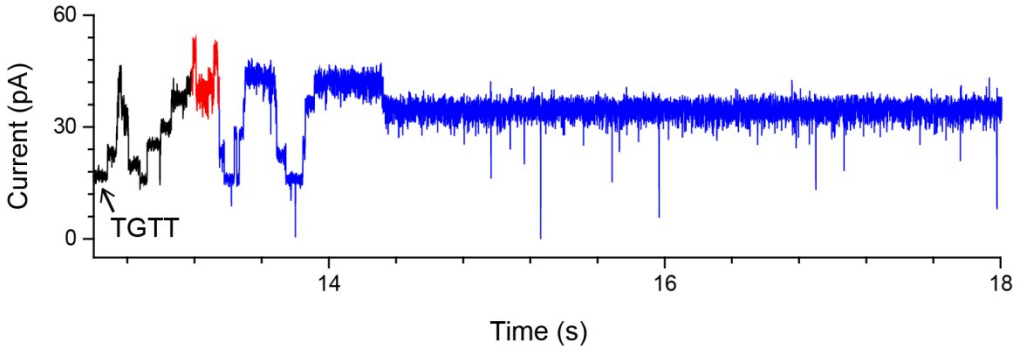
Supporting Figure S2. Free translocation of FANA polyU. (A) Free translocation events of FANA polyU. The demonstrated data is acquired at 100 kHz and filtered at 10 kHz. **(B)** Scatter plots of I_b/I_o vs dwell time for FANA polyU translocation events. The duration time for most events are below 1 ms. The experiment was carried out in the electrolyte buffered solution of 1 M KCl, 10 mM HEPES at pH 8.0.



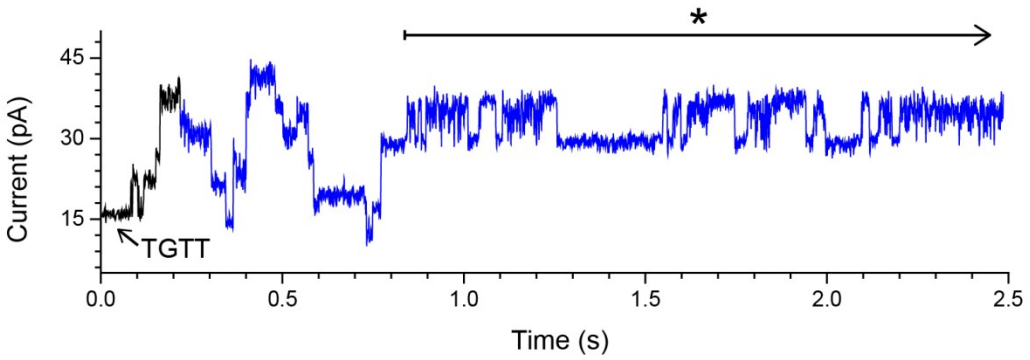
Supporting Figure S3. FANAx polymer preparation by T4 DNA ligase mediated ligation. (A) Diagram of T4 DNA ligase mediated ligation of 51-nt chimeric FANA/DNA and 46-nt 5' phosphate modified DNA. FANA was show in cyan. The abasic marker was represented by a red X. Chimeric FANA/DNA vs. 5'-P-DNA ratio was 1:5 in the reaction. **(B)** Ligation products analyzed by denaturing PAGE. Lane M: DNA maker. Lane 1: before ligation. Lane 2: after ligation. Full-length ligation product should be 97 nt.



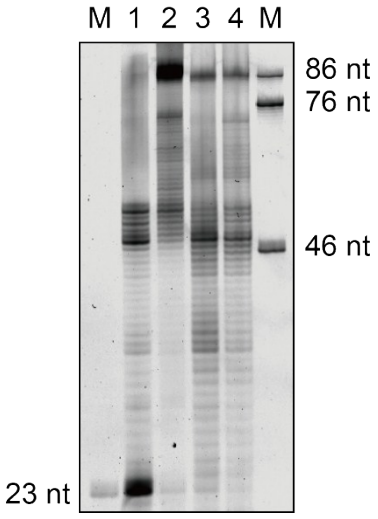
Supporting Figure S4. Sequencing library strategy. (A) The sequencing library is composed of three parts: DNA-FANA chimera (bottom), primer (top, grey) and blocker (red). The blocker protects the substrate from enzymatic extension in solution. **(B)** When a sequencing library is captured, the electric force draws the chimera through the phi29 DNA polymerase, the blocker is then released from its substrate during unzipping, which triggers DNA synthesis by the phi29 DNA polymerase.



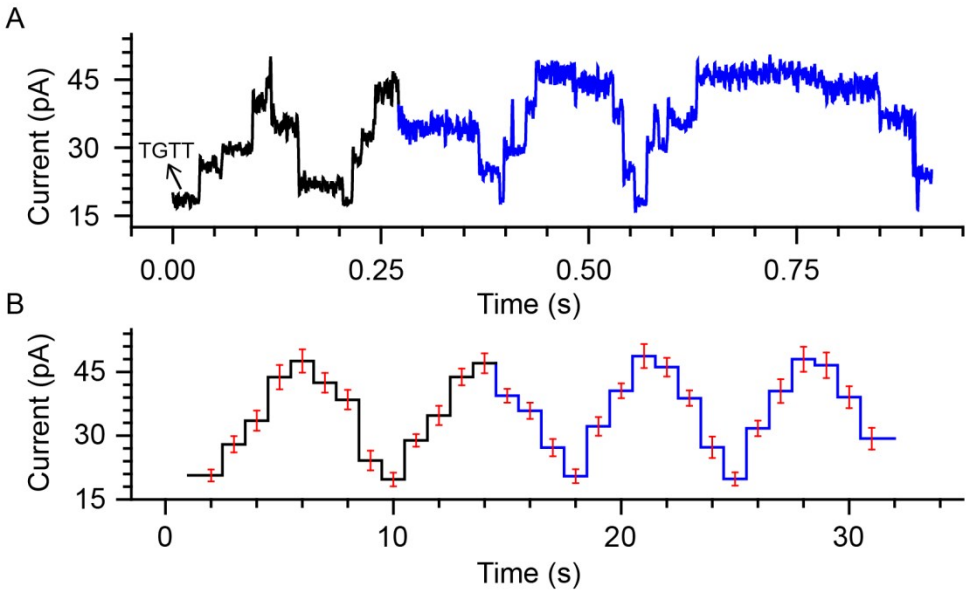
Supporting Figure S5. Representative nanopore sequencing trace from FANAx. The FANAx strand (**Supporting Table S1**) halts when the abasic site reaches the binding pocket of the phi29 DNAP. The blockage level stays unchanged (14.2 s to 18 s) until the applied voltage is reversed. Sequencing trace corresponds to DNA (black), the abasic site (red) and FANA (blue) in the restriction of MspA. The initial step with an arrow is produced by TGTT in the restriction.



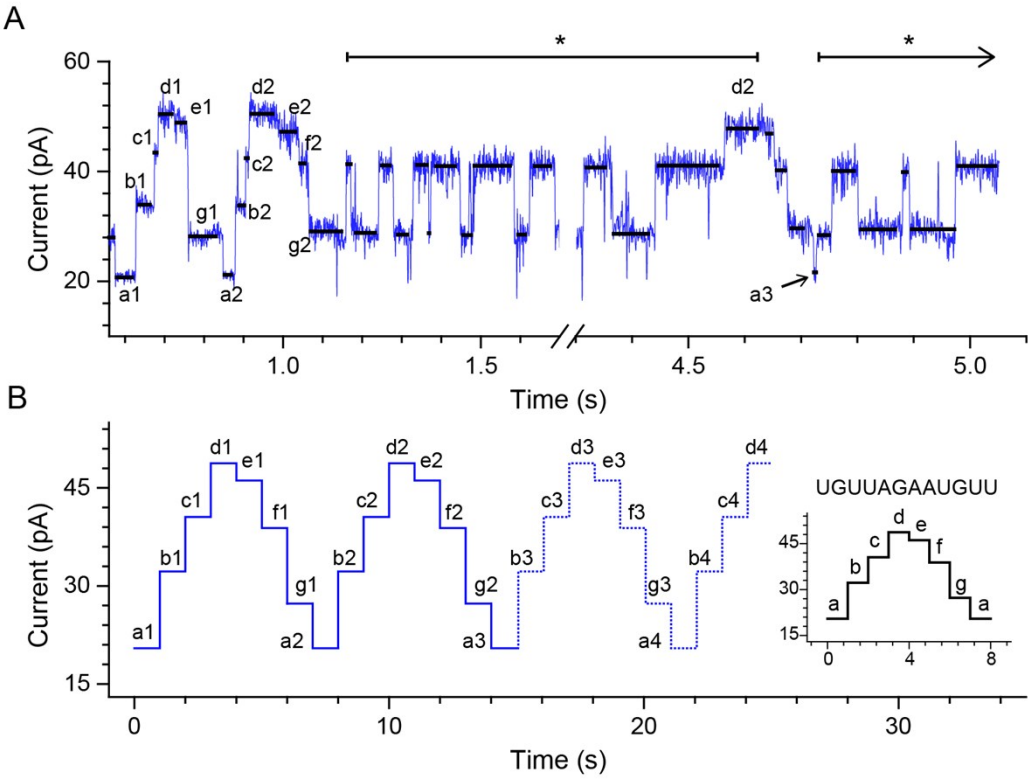
Supporting Figure S6. Representative nanopore sequencing trace from FANA30. Sequencing trace corresponds to DNA (black), and FANA (blue) in the restriction of MspA. The initial step with an arrow is produced by TGTT in the restriction. The measured signal show 'back and forth' movement between two levels when FANA reaches the binding pocket of the phi29 DNAP. The "back and forth" movement is marked with *.



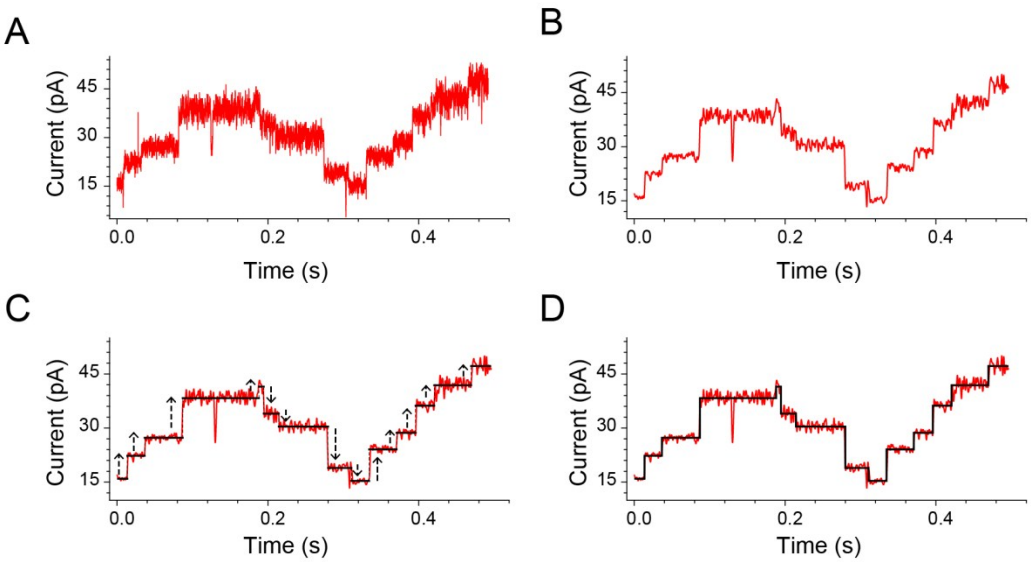
Supporting Figure S7. RT521K and phi29 reverse transcription. Reverse transcription analyzed by denaturing PAGE. Lane M: DNA marker with indicated chain length. Lane 1 and 2: RT521K-mediated FANA30 reverse transcription at room temperature and 65 ℃, respectively. Few full-length products appeared at room temperature (25 ℃), but lots of full-length product (86-nt) appeared at 65 ℃. Lane 3 and 4: phi29 DNAP-mediated FANA30 reverse transcription at 25 ℃ in different trials. Full length product (86-nt) are clearly observable, which confirms that the phi29 DNAP is a reverse transcriptase of FANA.



Supporting Figure S8. Representative nanopore sequencing trace of FANA42 (primer extension). (A) A representative sequencing trace for FANA42 (**Supporting Table S1**) when phi29 DNAP is bound on the DNA part of the strand. Sequencing steps of DNA (black) and FANA (blue) are marked respectively. The initial step with an arrow is produced by TGTT blockage in the pore restriction. **(B)** Mean current levels extracted from N=20 independent events using a level detection algorithm. Error bars (red) represents standard deviations of each sequencing step.



Supporting Figure S9. Representative nanopore sequencing trace of FANA42 (hysteresis motion). (A) A representative sequencing trace of FANA42. Letters represent different current levels judged from B. Hysteresis motion of the phi29 DNAP is marked with *. **(B)** A predicted signal pattern of the FANA part of FANA42 based on its periodic signal pattern which has appeared. The graph on the right represents a single sequencing period originates from the sequence UGUUAGAAUGUU.



Supporting Figure S10. Example of plateau extraction from raw sequencing data. **(A)** An example raw nanopore sequencing data. The trace is acquired with 100 kHz sampling rate and filtered with 10 kHz corner frequency on an Axon 200B patch clamp amplifier and digitized by a Digidata 1550A plus digitizer (Molecular Devices). **(B)** Low pass Bessel filtered (100 Hz) trace from **A**. **(C)** The digitally filtered trace is analyzed by the transition detection function within the custom LabView program. Upward and downward transitions within the trace could be automatically detected and be utilized to fragment the sequencing trace. **(D)** Mean currents levels are calculated from fragmented sequencing traces between transitions. The calculated mean currents are overlaid with the sequencing trace.

Supporting Table S1. Strand Sequences

Strand	Sequence	Length（nt）
FANA polyA	5’- GCC GTC CCT CTG TCC GCC GTC CCT CTG TCC **AAA AAA AAA AAA AAA AAA AA** /biotin/ -3’	50
FANA polyC	5’- GCC GTC CCT CTG TCC GCC GTC CCT CTG TCC **CCC CCC CCC CCC CCC CCC CC** /biotin/ -3’	50
FANA polyU	5’- GCC GTC CCT CTG TCC GCC GTC CCT CTG TCC **UUU UUU UUU UUU UUU UUU UU** /biotin/ -3’	50
FANAx	5’- AAA AAA CGT C**AG AAU GUU AGA AUG UUA GAA UGU UAG AAU GUU** XAG AAT GTT AGA ATG TTT CAG ATC TCA CTA TCG CAT TCT CAT GCA GGT CGT AGC C -3’	97
FANA42	5’- **AAA AAA CGU CAG AAU GUU AGA AUG UUA GAA UGU UAG AAU GUU** AGA ATG TTA GAA TGT TTC AGA TCT CAC TAT CGC ATT CTC ATG CAG GTC GTA GCC -3’	96
FANA30	5’- AAA AAA CGT C**UG UUA CAU GCA AGC UUG GCG UAA UCA UGU U**AG AAT GTT TCA GAT CTC ACT ATC GCA TTC TCA TGC AGG TCG TAG CC -3’	86
DNA30	5’- AAA AAA CGT CTG TTA CAT GCA AGC TTG GCG TAA TCA TGT TAG AAT GTT TCA GAT CTC ACT ATC GCA TTC TCA TGC AGG TCG TAG CC -3’	86
Primer	5’- GCG TAC GCC TAC GGT TTT CCG TAG GCG TAC GCG GCT ACG ACC TGC ATG AGA ATG C -3’	55
Blocker	5’- GAT AGT GAG ATC TGA TTT CCC AAA TTT AAA /cholesterol/ -3’	30
RT primer	5’- Cy5.5-GGC TAC GAC CTG CAT GAG AAT GC-3’	23

Notes:

1. The abasic site is indicated by ‘X’.

2. The FANA sequence is shown in red and underlined.

Supporting Table S2. Statistics of I_b/I_o for FANA homopolymers.

Homopolymers	I_b/I_0/ Standard Deviation		
	1	2	3
FANA polyU	0.20/0.05	0.19/0.05	0.20/0.06
FANA polyA	0.26/0.07	0.27/0.06	0.26/0.06
FANA polyC	0.29/0.09	0.29/0.08	0.28/0.08

Supporting Table S3. Statistics of nanopore sequencing levels.

FANAx		FANA42		FANA30		DNA30	
Sequence	**Mean Current (pA)/ Standard Deviation**	**Sequence**	**Mean Current (pA)/ Standard Deviation**	**Sequence**	**Mean Current (pA)/ Standard Deviation**	**Sequence**	**Mean Current (pA)/ Standard Deviation**
TGTT	17.55/1.18	TGTT	20.67/1.40	TGTT	18.88/2.55	TGTT	25.30/2.15
ATGT	23.54/0.98	ATGT	27.97/1.93	ATGT	26.07/2.91	ATGT	33.05/2.55
AATG	28.08/1.18	AATG	33.57/2.37	AATG	30.91/3.36	AATG	38.57/2.71
GAAT	38.95/1.77	GAAT	43.77/2.83	GAAT	42.21/5.38	GAAT	51.38/2.96
AGAA	42.93/2.03	AGAA	47.56/2.75	AGAA	46.50/4.96	AGAA	54.36/2.68
TAGA	36.43/1.60	TAGA	42.45/2.34	UAGA	38.29/4.04	TAGA	47.19/2.64
TTAG	32.99/1.52	TTAG	38.47/2.30	UUAG	35.08/3.60	TTAG	41.31/2.27
GTTA	21.00/1.39	GTTA	24.18/2.30	GUUA	25.30/3.08	GTTA	28.40/2.14
TGTT	16.84/0.97	TGTT	19.73/1.61	UGUU	17.49/2.32	TGTT	24.62/2.04
ATGT	25.04/0.90	ATGT	28.92/1.47	AUGU	28.58/4.43	ATGT	34.73/2.36
AATG	29.99/1.03	AATG	34.77/2.26	CAUG	41.45/4.08	CATG	40.43/3.00
GAAT	37.60/1.48	GAAT	43.81/1.95	UCAU	46.04/4.73	TCTA	45.69/2.60
AGAA	41.10/1.15	AGAA	47.03/2.33	AUCA	40.29/4.48	ATCA	41.30/2.29
XAGA	43.87/1.52	UAGA	39.41/1.64	AAUC	33.12/3.58	AATC	34.07/2.35
UXAG	50.25/2.11	UUAG	35.87/1.87	UAAU	38.87/4.26	TAAT	43.80/2.58
UUXA	39.95/2.66	GUUA	27.21/2.02	GUAA	31.84/3.88	GTAA	39.77/2.42
GUUX	24.67/1.93	UGUU	20.46/1.63	CGUA	25.53/6.84	CGTA	30.86/2.77
UGUU	16.67/0.78		32.21/2.19	GCGU	19.93/4.61	GCGT	40.20/2.69
	26.70/1.42		40.56/1.73	GGCG	27.91/5.20	GGCG	43.91/2.34

	35.80/1.74	AGAA	48.75/2.84	UGGC	25.40/4.74	TGGC	36.03/1.77
AGAA	43.52/1.56	UAGA	46.12/2.21	UUGG	33.51/3.32	TTGT	34.19/2.18
UAGA	41.39/1.73	UUAG	38.84/1.82	CUUG	40.00/3.91	CTTG	31.39/2.30
UUAG	34.80/1.35	GUUA	27.30/2.53				
GUUA	24.34/1.75	UGUU	19.84/1.55				
UGUU	16.89/0.93		31.74/1.84				
	24.98/1.33		40.51/2.80				
	35.40/1.43	AGAA	48.00/2.95				
AGAA	42.52/1.47	UAGA	46.55/3.00				
UAGA	40.92/1.90	UUAG	39.06/2.57				
UUAG	35.38/1.42	GUUA	29.32/2.55				
GUUA	26.55/3.51						

Note：Only 4 plateau transitions are detected when reading UGUU to AGAA within the FANA part which is supposed to be 5 plateau transitions. So a signal amplitude degeneracy exists among AUGU, AAUG and GAAU, and that's why there are some cells with no sequence data..

References

1. V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S. Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn and P. Holliger, *Science*, 2012, **336**, 341-344.
2. T. Z. Butler, M. Pavlenok, I. M. Derrington, M. Niederweis and J. H. Gundlach, *Proc. Natl. Acad. Sci. U.S.A.*, 2008, **105**, 20647-20652.
3. Y. Wang, S. Yan, P. Zhang, Z. Zeng, D. Zhao, J. Wang, H. Chen and S. Huang, *ACS Appl. Mater. Interfaces*, 2018, **10**, 7788-7797.
4. S. Huang, M. Romero-Ruiz, O. K. Castell, H. Bayley and M. I. Wallace, *Nat. Nanotechnol.*, 2015, **10**, 986-991.
5. E. A. Manrao, I. M. Derrington, A. H. Laszlo, K. W. Langford, M. K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis and J. H. Gundlach, *Nat. Biotechnol.*, 2012, **30**, 349-353.