Electronic Supplementary Material (ESI) for Chemical Science.

Supplementary information for "Locus-patterned sequence oriented enrichment for multi-dimensional gene analysis"

Yue Zhao[a], Xiaoxing Fang[a], Feng Chen[a], Min Bai[a], Chunhai Fan[b] and Yongxi Zhao[a,*]

[a] The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xianning West Road, Xi'an, Shaanxi 710049, P. R. China

[b] School of Chemistry and Chemical Engineering, and Institute of Molecular Medicine, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200240, P. R. China.

* To whom correspondence should be addressed. Tel: +86-029-82668908; Fax: +86-029-82668908; Email: yxzhao@mail.xjtu.edu.cn

Table of Contents

Experimental Procedures

Apparatus. The real-time fluorescence measurements were performed with LightCycler 96 (Roche Applied Science, Mannheim, Germany). The end point fluorescent images with PCR tubes were taken with Amersham™ Imager 600 instrument and software (GE Lifesciences). Excitation light was provided by a blue channel (460nm), a green channel (520nm) and a red channel (630nm). Confocal fluorescence images were obtained using a commercial high resolution laser confocal microscope (TCS SP8 STED 3X, Leica) with 100× oil-immersion objective. Centrifugation were carried out with ThermoSorvallST8R centrifuge (Thermo Fisher Scientific Inc, USA).

Materials and reagents. All chemicals were obtained from commercial sources and used without further purification. Bst2.0 WarmStart® DNA polymerase and 100mM magnesium sulfate solution were purchased from New England Biolabs Ltd. Betaine was obtained from Macklin Biochemical Company (Shanghai, China). dNTPs were ordered from Generay Biotechnology Co., Ltd (Shanghai, China). SYTOX Orange nucleic acid stain was ordered from Invitrogen (Thermo Fisher Scientific Inc, USA). All oligonucleotides were ordered from Sangon Biotechnology Co., Ltd. (Shanghai, China). Oligonucleotide sequences are summarized in Table S1. The 380 bp fragments of C gene from hepatitis B virus (HBV), 400 bp fragments of S gene from type B or type C HBV were respectively screened and cloned into a pUC57 vector by Sangon Biotechnology Co., Ltd. (Shanghai, China). Viral gene was isolated from peripheral serum using Viral genome DNA/RNA rapid extraction kit (Bioteke, Beijing, China), and the sequencing was completed by Sangon Biotechnology Co., Ltd. (Shanghai, China).

Thermodynamics calculations for strand migration reaction

According to the *van't Hoff* equation, the Δ*G* of locus-oriented reaction are given by

$$\Delta G_T = \Delta G^0_{FT} + \Delta G^0_{SQ} - \Delta G^0_{ST} - \Delta G^0_{FQ} + RT \ln Q \quad (1)$$

$$\Delta G_A = \Delta G^0_{FA} + \Delta G^0_{SQ} - \Delta G^0_{SA} - \Delta G^0_{FQ} + RT \ln Q \quad (2)$$

Where Q is the reaction quotient, R is the gas constant (8.314 J/mol·K), T is the Kelvin temperature), and G^0_X is the standard free energy of species X under 338K (reaction temperature for LAMP), which could be calculated by using NUPACK software (http://www.nupack.org/), giving

$$\Delta G^0_{FT} + \Delta G^0_{SQ} - \Delta G^0_{ST} - \Delta G^0_{FQ} = -0.71 \text{ kcal/mol} \quad (3)$$

$$\Delta G^0_{FA} + \Delta G^0_{SQ} - \Delta G^0_{SA} - \Delta G^0_{FQ} = 6.02 \text{ kcal/mol} \quad (4)$$

When the reaction is at equilibrium, which means Δ*G* = 0, the *Q* for ***T*** and ***A*** could be calculated as 2.88 and 1.28E-4 according to eqs 1-4. So the Discrimination Factor (***DF***) defined as the ratio of reaction quotient can be calculated to be 2.26 E4

Whereas in basic reaction, the Δ*G* are given by

$$\Delta G_T = \Delta G^0_{FT} - \Delta G^0_{FQ} + RT \ln Q \quad (5)$$

$$\Delta G_A = \Delta G^0_{FA} - \Delta G^0_{FQ} + RT \ln Q \quad (6)$$

Where

$$\Delta G^0_{FT} - \Delta G^0_{FQ} = -3.97 \text{ kcal/mol} \quad (7)$$

$\Delta G^{0}_{FA} - \Delta G^{0}_{FQ} = 0.26$ kcal/mol (8)

When $\Delta G = 0$, the Q for ***T*** and ***A*** could be calculated as 370 and 0.68 with a ***DF*** of 5.45E2.

Experimental Figures, Table and Data

Table S1 Sequences of oligonucleotides involved in this method.

Name	Sequences (5' to 3')	Usage
HBVC FIP	GCGGTGTCGAGRAGATCYCGAATTTGGAGCTTCTGTGGAGT	Forward inner primer for C gene
HBVC BIP	TCWGCTCTGTATCGGGAGGCCAGAATAGCTTGCCTGAGTGC	Backward inner primer for C gene
HBVC LF	AGGAAAGAAGTCAGAAGGCA	Forward loop primer for C gene
HBVC LB	CTCCGGAACATTGTTCWCCT	Backward loop primer for C gene
HBVC F3	GCATGGACATTGACCCGTAT	Forward outer primer for C gene
HBVC B3	TCATCAACTCACCCCAACAC	Backward outer primer for C gene
HBVS FIP	GCCGCAGACACATCCAGCGATAGCCAAAATTCGCAGTCCC	Forward inner primer for S gene
HBVS BIPB	CTTCTTGTTGGTTCTTCTGGACTAATGGTCCGGTGCTGGT	Backward inner primer for S gene of type B
HBVS BIPC	CTTCTTGTTGGTTCTTCTGGACTACATGGTCCTGTGCTGG	Backward inner primer for S gene of type C
HBVS LFB	AAATTGGAGGACAACAGGTTGGTGAGTGAC	Forward loop primer for S gene of type B
HBVS LFC	AAATTGGAGGACAAGAGGTTGGTGAGTGAT	Forward loop primer for S gene of type C
HBVS LBB	ATGTTGCCCGTTTGTCCTCTAATTCCAGGAT	Backward loop primer for S gene of type B
HBVS LBC	ATGTTGCCCGTTTGTCCTCTACTTCCAGGAA	Backward loop primer for S gene of type C
HBVS F3	CTCTCAATTTTCTAGGGGGA	Forward outer primer for S gene
HBVS B3B	AGCAGGAGTTGTGCAGGT	Backward outer primer for S gene of type B
HBVS B3C	AGCAGGAATCGTGCAGGT	Backward outer primer for S gene of type C
HBVS FIP Hex	BHQ2-GATGTGTCTGCGGC/iSp18/GCCGCAGACACA/iHEXdT/CCAGCGATAGCCAAAATTCGCAGTCCC	Fluorescence labeled forward inner primer for S gene
HBVC BIP Tex	BHQ2-CGATACAGAGCAGA/iSp18/TCTGCTCTGTA/iTexRddT/CGGGAGGCCAGAATAGCTTGCCTGAGTGC	Fluorescence labeled backward outer primer for C gene
HBVS LFB FAM	FAM-AAATTGGAGGACAACAGGTTGGTGAGTGAC	Fluorescence labeled forward loop primer for S gene of type B
HBVS LBC Cy5	Cy5-ATGTTGCCCGTTTGTCCTCTACTTCCAGGAA	Fluorescence labeled backward loop primer for S gene of type C
LF Block BHQ2	CACCAACCTGTTGTCCTCCAATTT-BHQ2	Quencher labeled primer complementary with forward loop primer for S gene of type B

LB Block BHQ2	GAAGTAGAGGACAAACGGGCAACAT-BHQ2	Quencher labeled primer complementary with backward loop primer for S gene of type C
LF template of type B	TGGCCAAAATTCGCAGTCCCAAATCTCCAGTCACTCACCAA CCTGTTGTCCTCCAATTT	The sequences from type B for recognition of LF
LF template of type C	TGGCCAAAATTCGCAGTCCCCAACCTCCAATCACTCACCAA CCTCTTG TCCTCCAATTT	The sequences from type C for recognition of LF
FP-ASPCR	CCACAGAGTCTAGACTCGTG	Forward PCR primer for S gene
RPB-ASPCR	GACAACAGGTTGGTGAGTGAC	Backward PCR primer for S gene of type B
RPC-ASPCR	GACAAGAGGTTGGTGAGTGAT	Backward PCR primer for S gene of type C

The underlined regions in all oligonucleotides represents the stem sequence of hairpin probes.

Figure S1. Sequences alignment of S gene is partially shown with representative sequences of each genotype (A-F). The isolates numbers are obtained from National Center for Biotechnology Information (NCBI). The variant locus of type B and type C are highlighted in red.

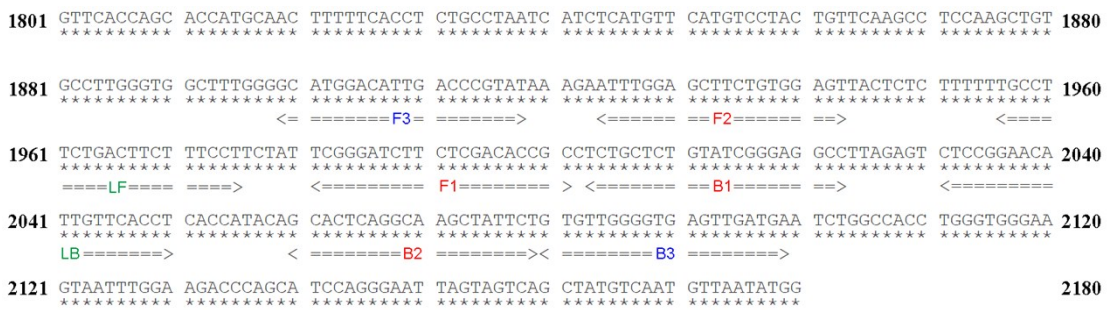

Figure S2. The gene fragment sequences and positions of primers used for C gene. The sequence was obtained from GenBank with the accession number KX660678. Two inner primers (FIP: F1-F2 and BIP: B1-B2), outer primers (F3 and B3) and loop primers (LF and LB) are underlined respectively.

Figure S3. The gene fragment sequences and positions of primers used for S gene (type B). The sequence was obtained from GenBank with the accession number X97851. Two inner primers (FIP: F1-F2 and BIP: B1-B2), outer primers (F3 and B3) and loop primers (LF and LB) are underlined respectively.

Figure S4. The gene fragment sequences and positions of primers used for S gene (type C). The sequence was obtained from GenBank with the accession number D23680.1. Two inner primers (FIP: F1-F2 and BIP: B1-B2), outer primers (F3 and B3) and loop primers (LF and LB) are underlined respectively.

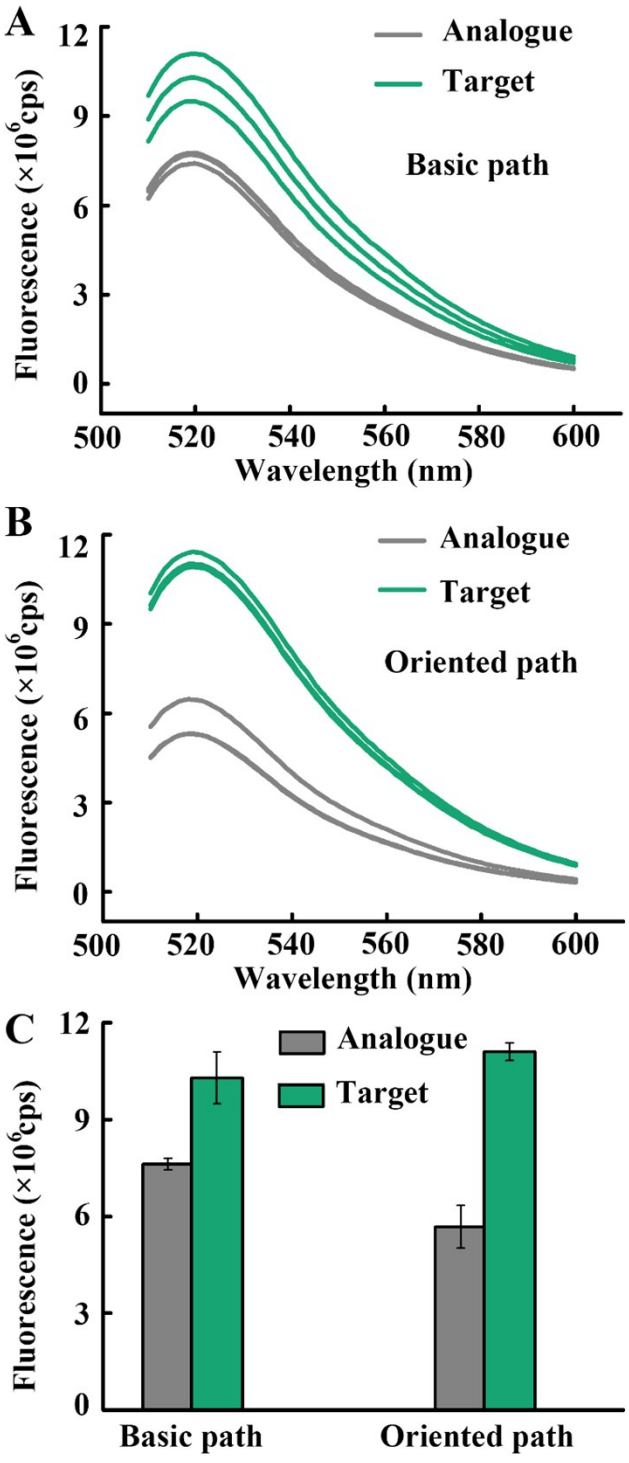

Figure S5. The experimental verification of the thermodynamics discrimination via basic reaction path and oriented sequence enrichment reaction path. The error bar indicates the SD from three independent experiments.

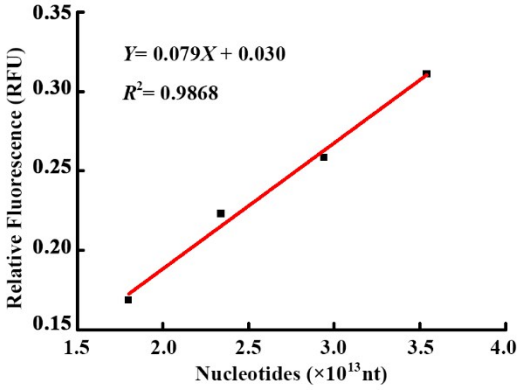

Figure S6. HRM dyes presents a linear dose-dependent fluorescence response for duplex DNA synthesis.

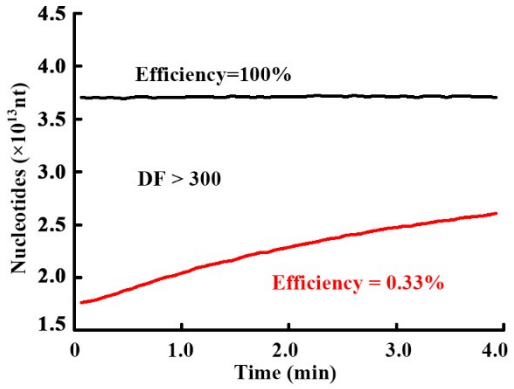

Figure S7. Comparison of extension between matched primer and mismatched primer. The extension efficiency of matched primer (black line) with rate above 60×10^{13} nt/min is defined as 100%, and mismatched primer (red line) with rate only at 0.2×10^{13} nt/min can be calculated as 0.33%. The discrimination factor (DF) is greater than 300.

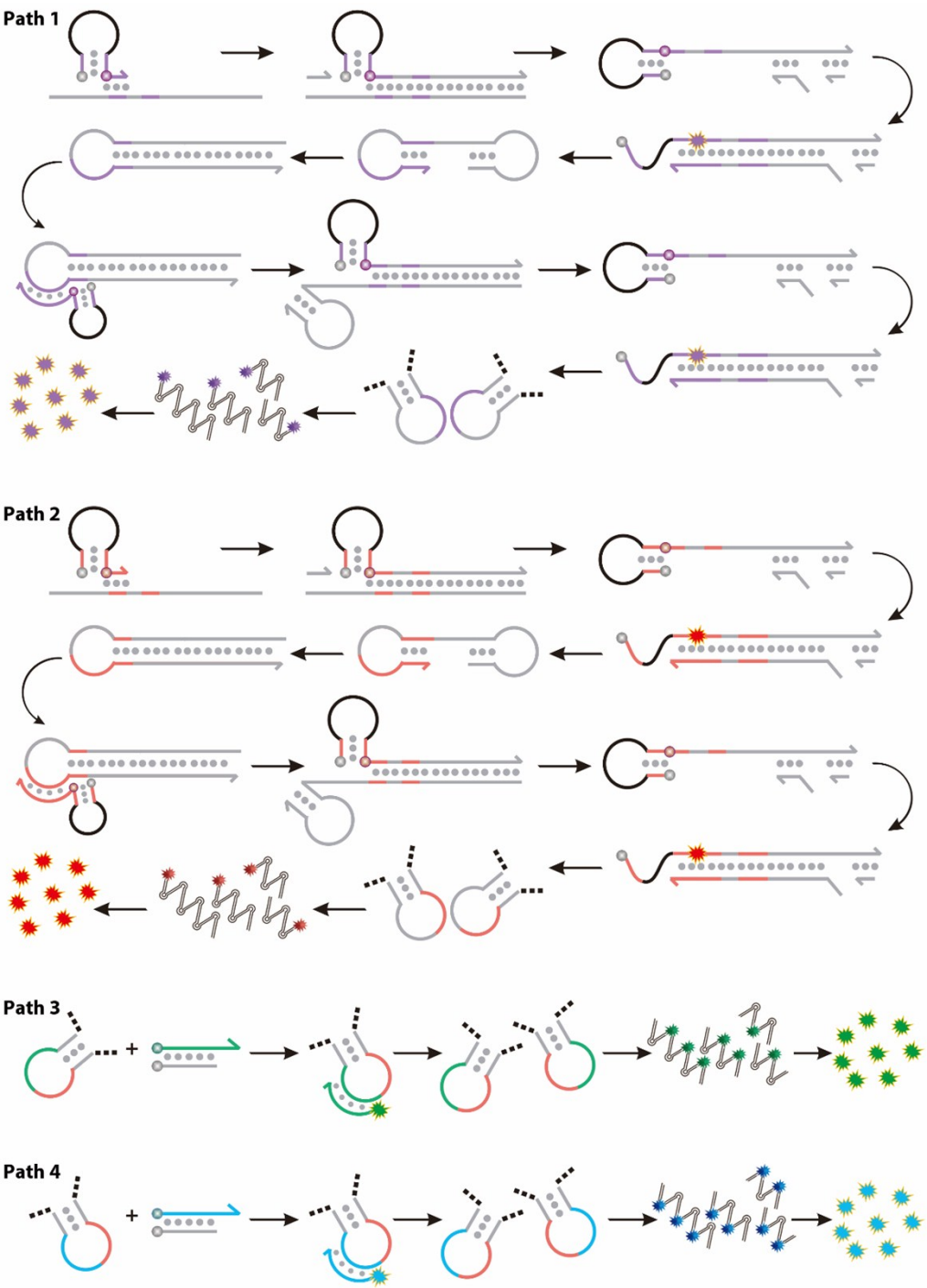

Figure S8. Schematic illustration of locus-patterned sequence oriented enrichment.

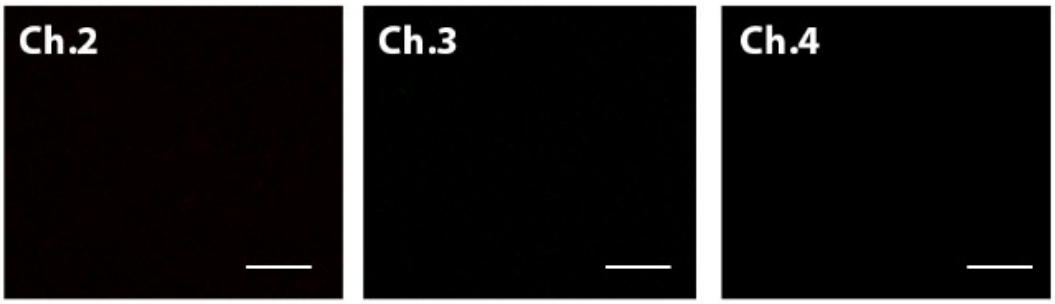

Figure S9. Confocal fluorescence image of locus-patterned sequences enrichment without template. Scale bars=20 μm

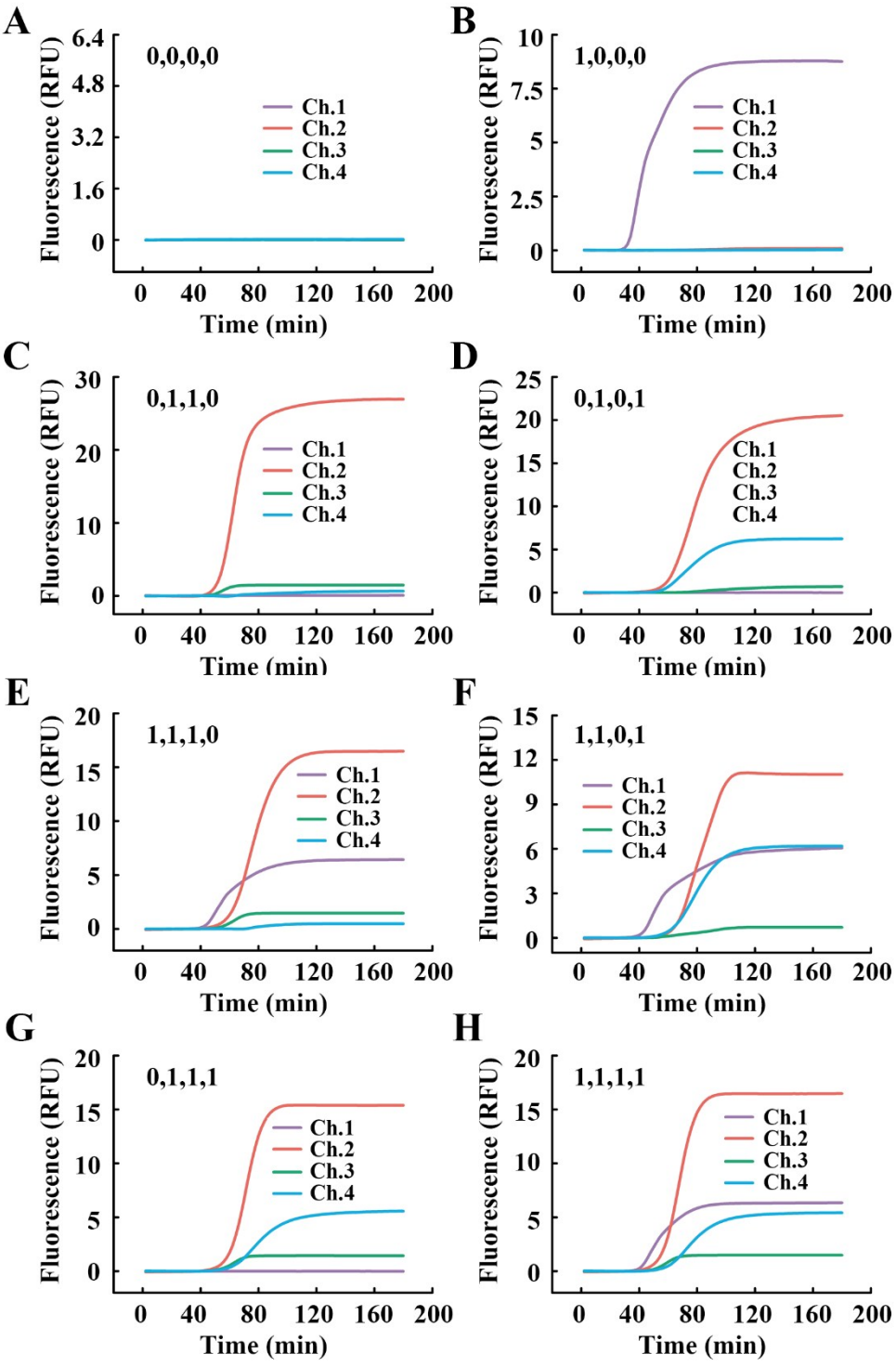

Figure S10. Real-time fluorescence curves obtained from locus-patterned sequence enrichment of heterogeneous genes.

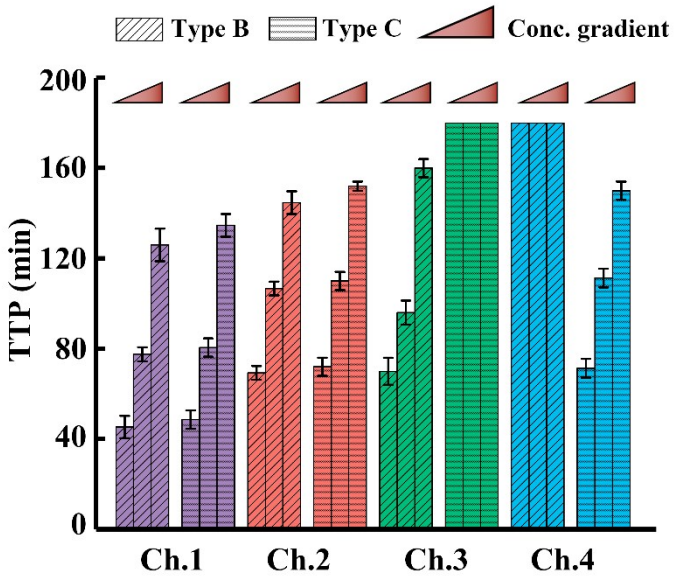

Figure S11. The effect of hairpin primers (HBVS FIP Hex and HBVC BIP Tex) concentration on the performance of proposed method. Primers concentration are involved at **0.2 μM**, 0.4 μM and 0.6 μM. The diagonal column and shaded column represent type B and type C respectively. Ch.1, 2, 3 and 4 correspond to fluorescence signal from Texas Red, Hex, FAM and Cy5 channels. The time to positive (TTP) of each channels are determined by the default threshold and TTP equals 180 min represents negative signal. The bold concentration is selected for subsequent experiments due to its earliest TTP. The error bar indicates the standard error from three independent experiments.

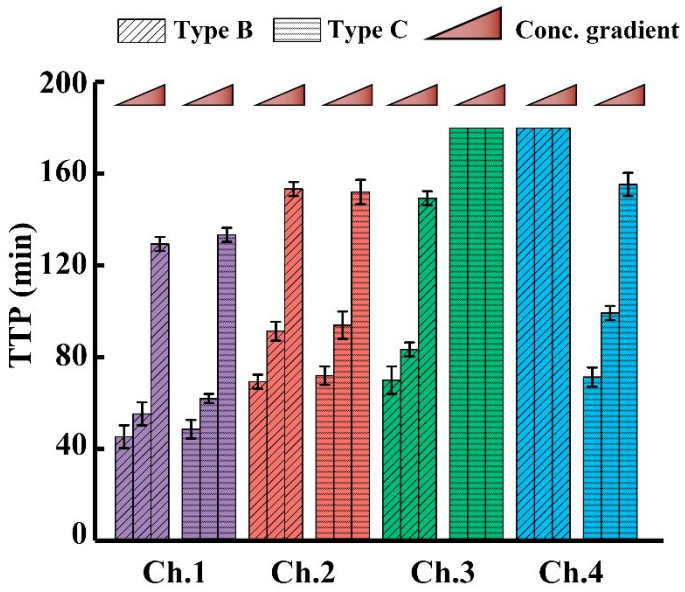

Figure S12. The effect of labelled complex primers (HBVS LFB FAM-LF Block BHQ2 and HBVS LBC Cy5-LB Block BHQ2) concentration on the performance of proposed method. Primers concentration are involved at **0.4 μM**, 0.8 μM and 1.2 μM. The diagonal column and shaded column represent type B and type C respectively. Ch.1, 2, 3 and 4 correspond to fluorescence signal from Texas Red, Hex, FAM and Cy5 channels. The time to positive (TTP) of each channels are determined by the default threshold and TTP equals 180 min represents negative signal. The bold concentration is selected for subsequent experiments as its earliest TTP. The error bar indicates the standard error from three independent experiments.

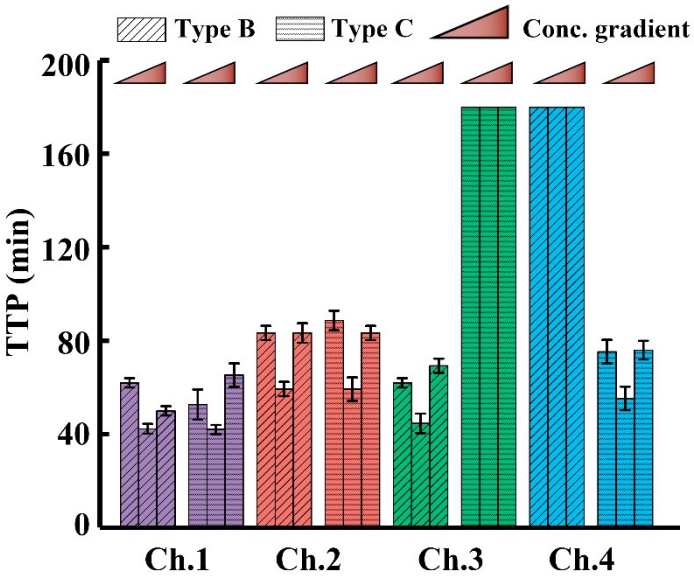

Figure S13. The effect of dNTP concentration on the performance of proposed method. dNTPs concentration are involved at 0.4 mM, **0.8 mM** and 1.2 mM. The diagonal column and shaded column represent type B and type C respectively. Ch.1, 2, 3 and 4 correspond to fluorescence signal from Texas Red, Hex, FAM and Cy5 channels. The time to positive (TTP) of each channels are determined by the default threshold and TTP equals 180 min represents negative signal. The bold concentration is selected for subsequent experiments as its earliest TTP. The error bar indicates the standard error from three independent experiments.

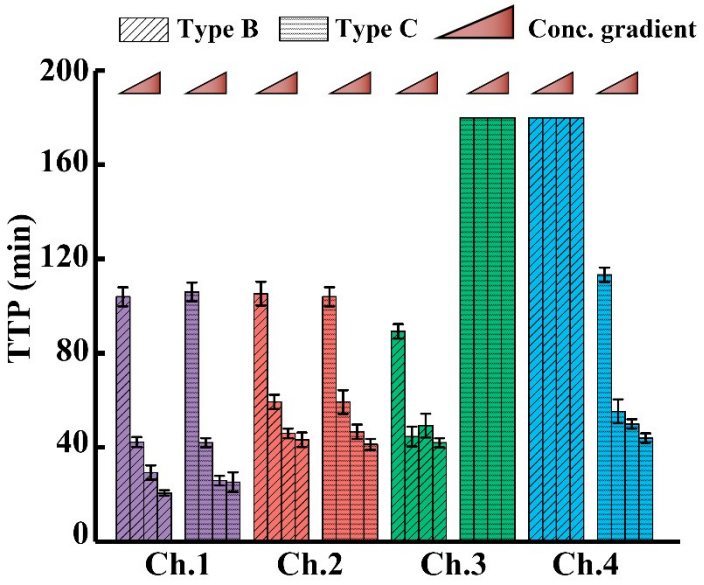

Figure S14. The effect of enzyme concentration on the performance of proposed method. Enzyme concentration are involved at 8 U, 12 U, **16U** and 20 U. The diagonal column and shaded column represent type B and type C respectively. Ch.1, 2, 3 and 4 correspond to fluorescence signal from Texas Red, Hex, FAM and Cy5 channels. The time to positive (TTP) of each channels are determined by the default threshold and TTP equals 180 min represents negative signal. The bold concentration is selected for subsequent experiments as its TTP reaches plateau. The error bar indicates the standard error from three independent experiments.

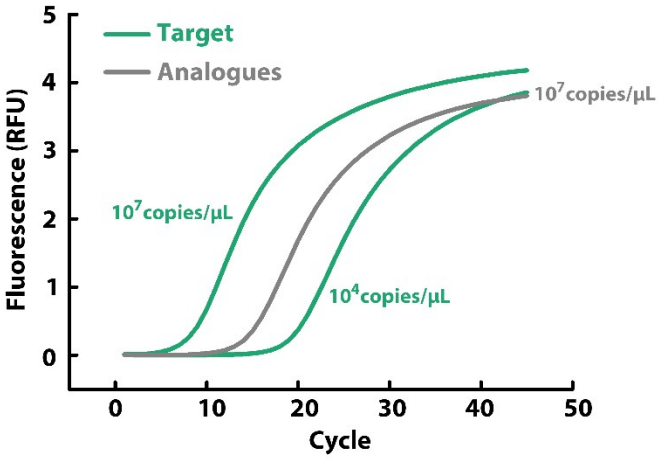

Figure S15. Real-time fluorescence PCR curves of target type B and analogues type C template via allele specific-PCR.

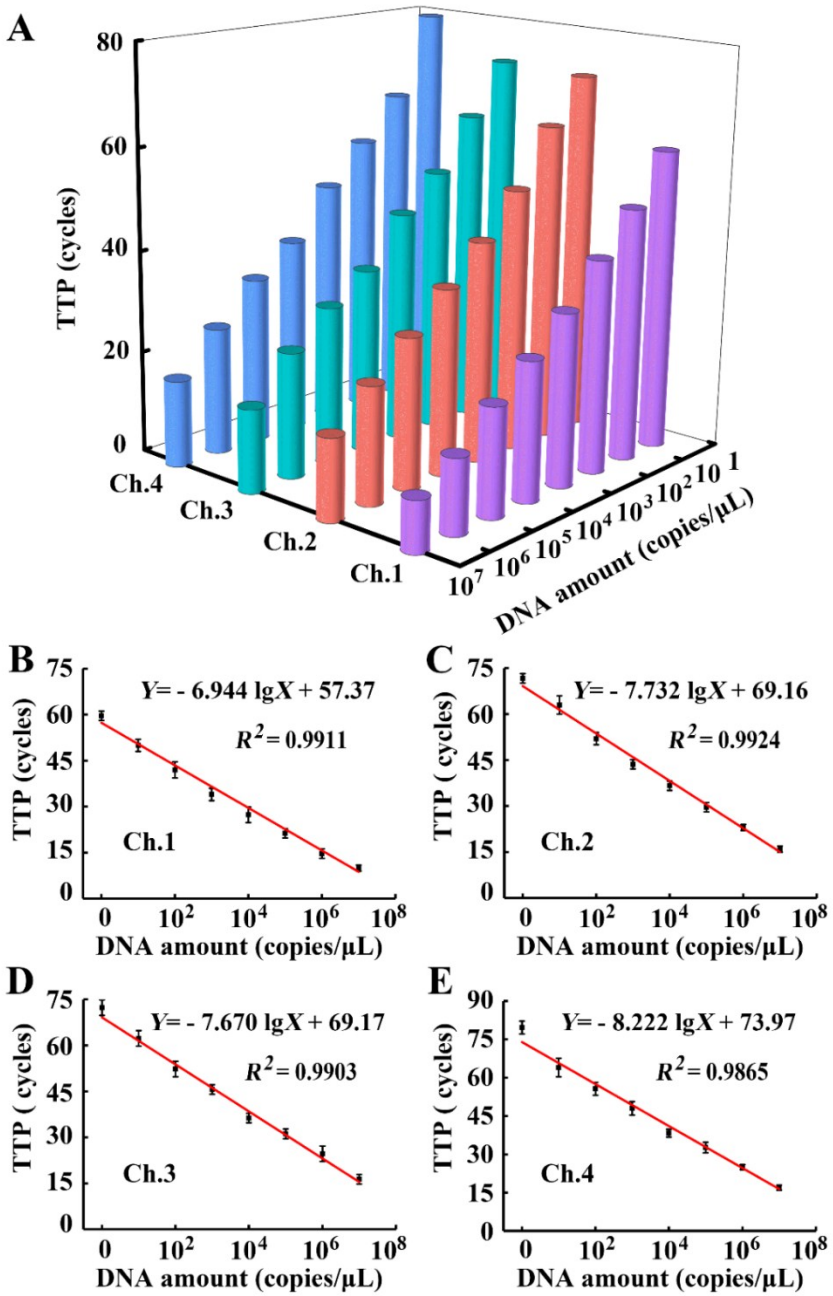

Figure S16. (A) Fluorescence responses from multiple channels of the locus-patterned sequence oriented enrichment. The amounts of standard template DNA is 10^7, 10^6, 10^5, 10^4, 10^3, 100, 10 and 1 copies/μL. The reactions were incubated at a constant temperature of 65°C, with fluorescence read steps at intervals of 2 min/cycle. (B) The linear correlation between time to positive (TTP) and the logarithm of template amount in channel 1 (Texas Red). (C) The linear correlation between TTP and the logarithm of template amount in channel 2 (Hex). (D) The linear correlation between TTP and the logarithm of template amount in channel 3 (FAM). (E) The linear correlation between TTP and the logarithm of template amount in channel 4 (Cy5). The error bar indicates the standard error from three independent experiments.

No.	Code	No.	Code	No.	Code	No.	Code
1	[1, 1,0, 0]	2	[1, 1, 1,0]	3	[1, 1,0, 1]	4	[1, 1,0, 1]
5	[1, 1,0, 1]	6	[1, 1,0, 1]	7	[1, 1,0, 1]	8	[1, 1,0, 1]
9	[1, 1,0, 1]	10	[1, 1,0, 1]	11	[1, 1,0, 0]	12	[1, 1,0, 1]
13	[1, 1,0, 1]	14	[1, 1, 1,0]	15	[1, 1,0, 1]	16	[1, 1,0, 1]
17	[1, 1, 0, 1]	18	[1, 1, 0, 1]	19	[1, 1, 0, 1]	20	[1, 1, 0, 1]
21	[1, 1, 0, 1]	22	[1, 1, 0, 1]	23	[1, 1, 0, 1]	24	[1, 1, 0, 1]

Figure S17. The binary codes generated from clinical samples.

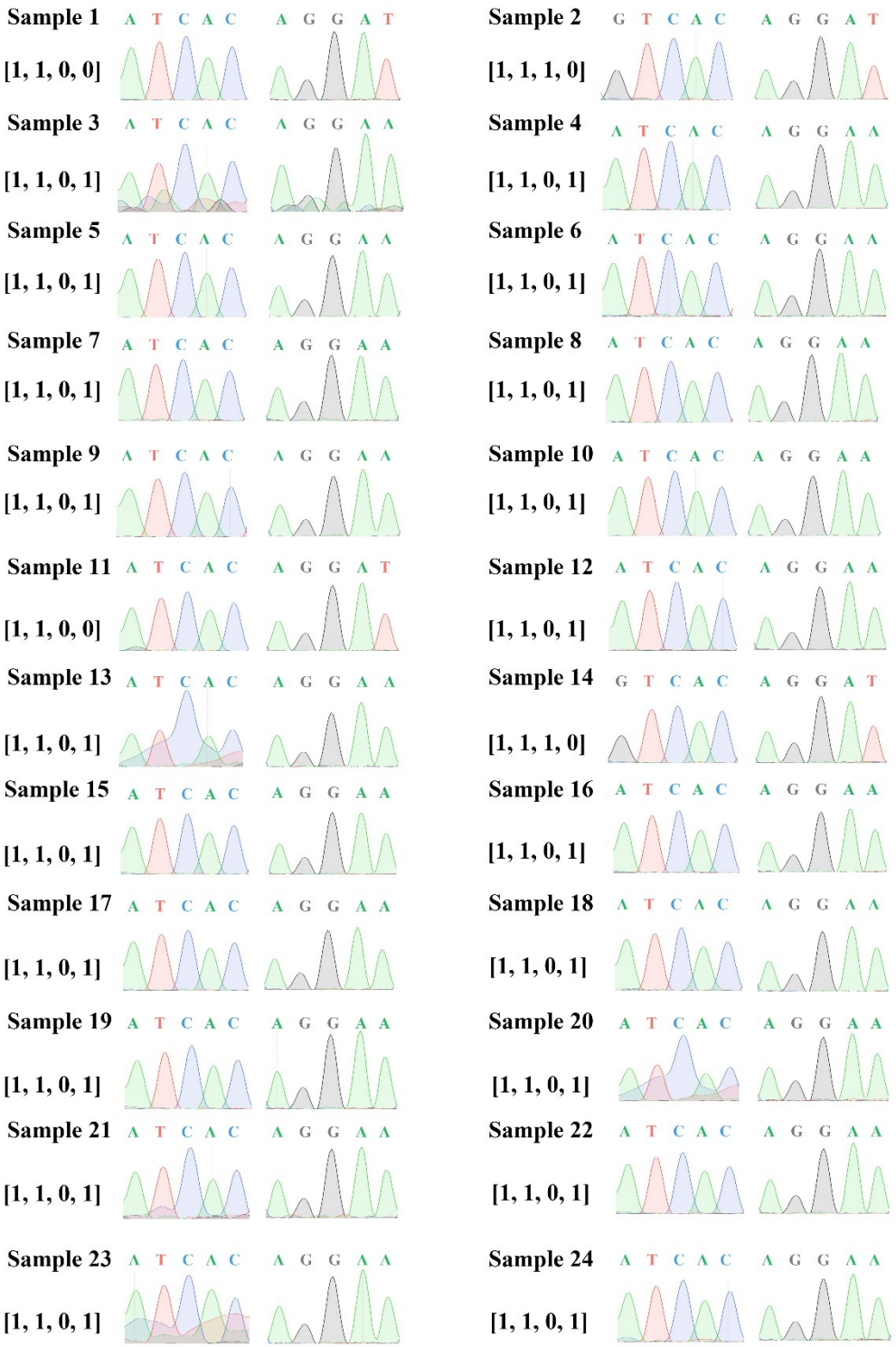

Figure S18. Sequencing results of S gene extracted from clinical samples.

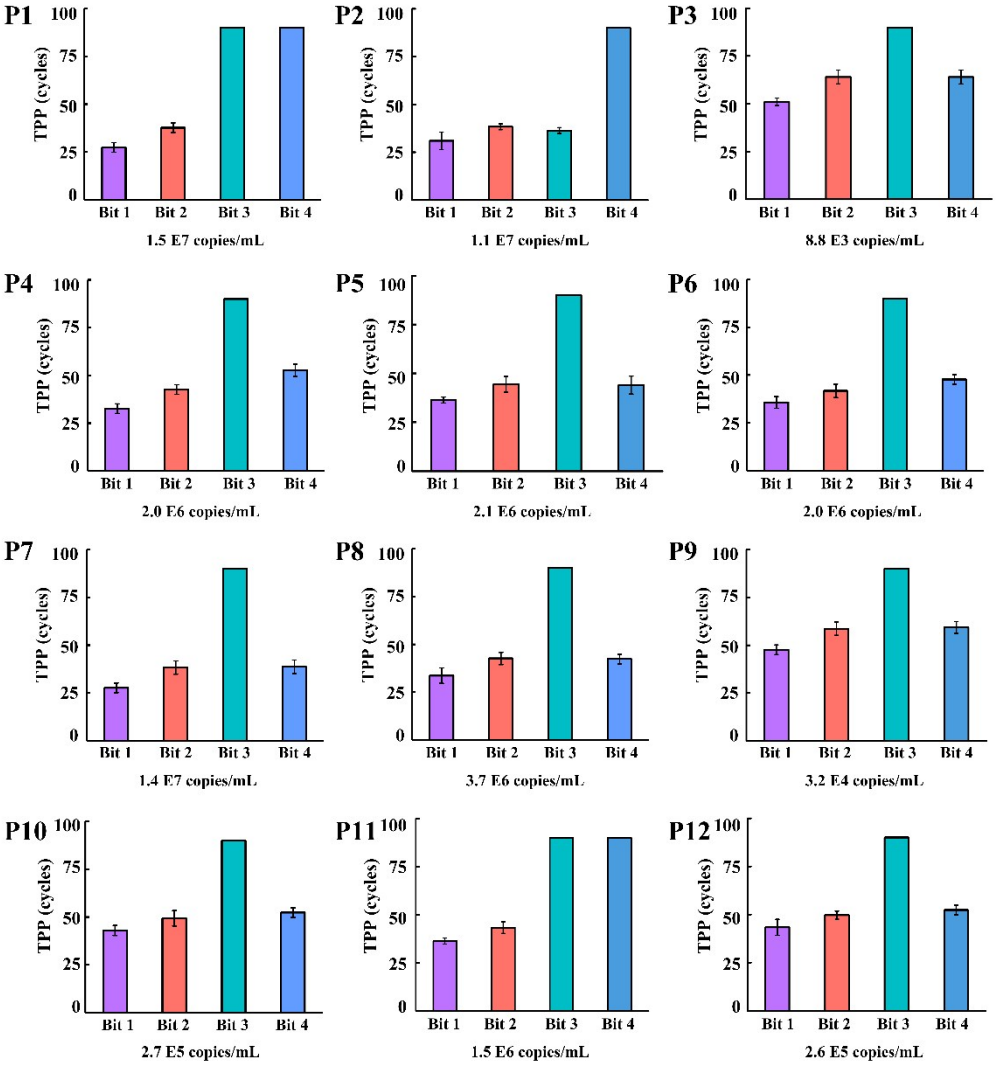

Figure S19. Clinical performance on sample No.1 to 12. The quantification results present the amount of viral in per mL blood.

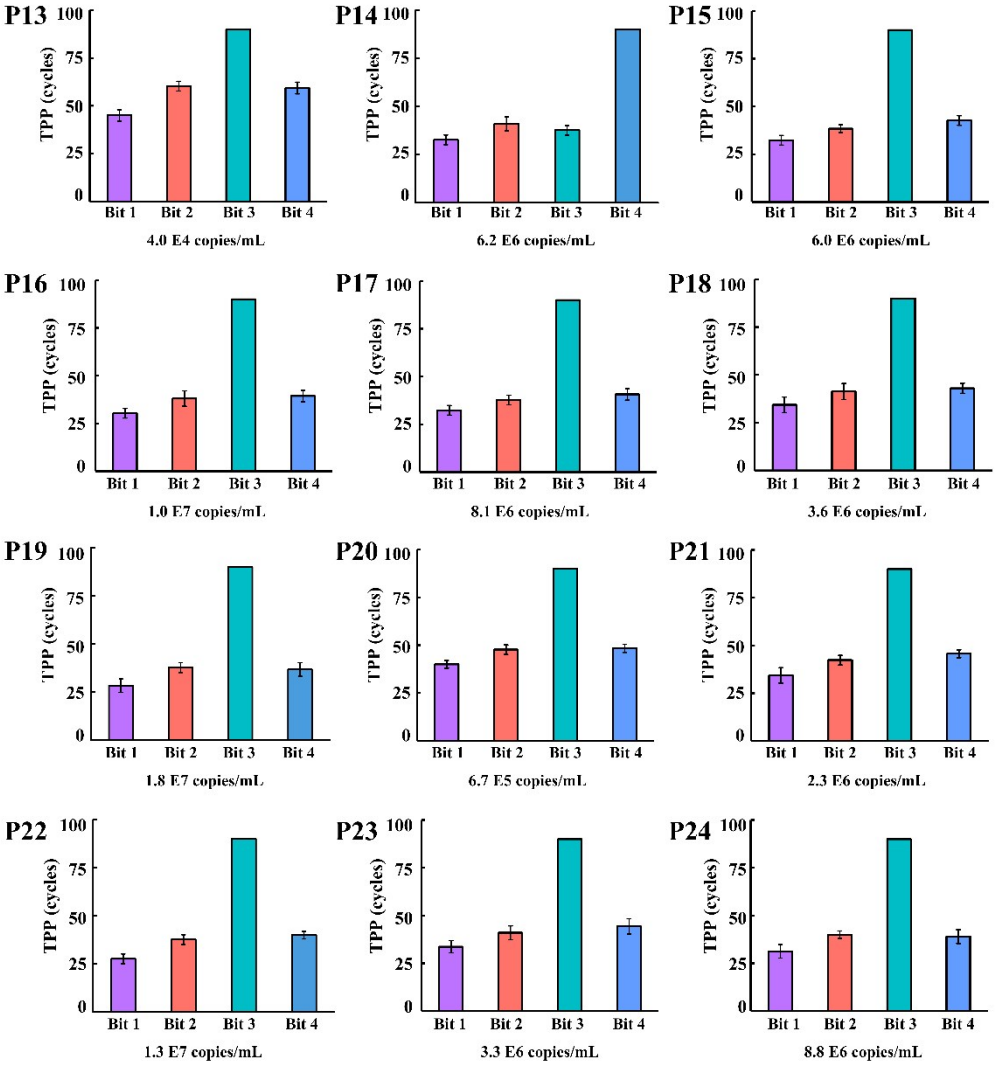

Figure S20. Clinical performance on sample No.13 to 24. The quantification results present the amount of viral in per mL blood.