Electronic Supplementary Material (ESI) for Analyst.

Supplementary Information

Efficient Large-scale Screening of Viral Pathogen by Fragment Length Identification of Pooled Nucleic Acid Samples (FLIPNAS)

Xianzhen Feng[a], Xinyu Zhuang[a], Grace Lui[b.] and I-Ming Hsing[*a]

[a.] Department of Chemical and Biological Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China. E-mail: kehsing@ust.hk

[b.] Department of Medicine & Therapeutics, The Chinese University of Hong Kong, Hong Kong, China. Email: gracelui@cuhk.edu.hk

* To whom correspondence should be addressed.

Table. S1 Target domains and the lengths of de novo designed unique primers, PCR primers.

Name	Detailed sequence(5’ to 3’)
Synthetic RNA(EURO-019)	GGGAGACGAAUUGGGCCCUCUAGAUGCAUGCUCGAGCGGCCGCCAGUG UGAUGGAUAUCUGCAGAAUUCGCCCUUAUUCAAGUAUUGAGUGAAAU GGUCAUGUGUGGCGGUUCACUAUAUGUUAAACCAGGUGGAACCUCAU CAGGAGAUGCCACAACUGCUUAUGCUAAUAGUGUUUUUAACAUUUGU CAAGCUGUCCGGAAGAGACAGGUACGUUAAUAGUUAAUAGCGUACUU CUUUUUCUUGCUUUCGUGGUAUUCUUGCUAGUUACACUAGCCAUCCU UACUGCGCUUCGAUUGUGUGCGUACUGCUGCAAUAUUGUUAACGUAU AAUGGACCCCAAAAUCAGCGAAAUGCACCCCGCAUUACGUUUGGUGGA CCCUCAGAUUCAACUGGCAGUAACCAGAAUGGAGAACGCAUUGCAACU GAGGGAGCCUUGAAUACACCAAAAGAUCACAUUGGCACCCGCAAUCCU GCUAACAAUGCUGCAAUCGUGCUACAACUUCCUCAAGGAAAUUUUGGG GACCAGGAACUAAUCAGACAAGGAACUGAUUACAAACAUUGGCCGCAA AUUGCACAAUUUGCCCCCAGCGCUUCAGCGUUCUUCGGAAUGUCGCGC AUUGGCAUGGAAGUCACACCUUCGGGAACGUGGUUGACCUACACAGGU GCCAUCAAAUUGGAGUGUGACAUACCCAUUGGUGCAGGUAUAUGCGC UAGUUAUCAGACUCAGACUAAUUCUCCUCGGCGGGCACGUAGUGUAGC UAGUCAACCUGCUUUUGCUCGCUUGGAUCCGAAUUCAAAGGUGAAAU UGUUAUCCGCUCACAAUUCCACACAACAUACGAGCCGGAAGCAUAAAG UGUAAAGCCUGGGGUGCCUAAUGA
Unique barcode primer 1	GCAGAGTCGGCCTACAGGTTTTCTGGTTACTGCCAGTTGAAT
Unique barcode primer 2	TGCCTACTACTCCACATAGCACACTGTCTGGTTACTGCCAGTTGAAT
Unique barcode primer 3	CTCTCAACCTCCACCCCTTCAACTACTTCACTCTGGTTACTGCCAGTTGAAT
Unique barcode primer 4	CTGCTGTATGCATCCAAGGCGCTCGAAAATATAAAGTCTGGTTACTGCCAG TTGAAT
Unique barcode primer 5	CGAGAGCCAGGTAACGAATGGAGTTACGAGTTAAGAATGAATCTGGTTAC TGCCAGTTGAAT
Unique barcode primer 6	CTGAGACTCGGACGAACTGCTACTTTTGCTAACACATCACCTTATATCTGGT TACTGCCAGTTGAAT
Unique barcode primer 7	TGCACCGGACCGTCTGAAATGATGATGACTACCTAACTAATACTCCTGCTAT CTGGTTACTGCCAGTTGAAT
Unique barcode primer 8	CGTCACGACCGGCTGAAATACCTCTACTAACTCTACCTCCATCATCTTATCCT GTCTCTGGTTACTGCCAGTTGAAT
Reverse Primer 1	GCAGAGTCGGCCTACAGGTTT
Reverse Primer 2	TGCCTACTACTCCACATAGCA
Reverse Primer 3	CTCTCAACCTCCACCCCTTCA
Reverse Primer 4	CTGCTGTATGCATCCAAGGCG
Reverse Primer 5	CGAGAGCCAGGTAACGAATGG
Reverse Primer 6	CTGAGACTCGGACGAACTGCT
Reverse Primer 7	TGCACCGGACCGTCTGAAATG
Reverse Primer 8	CGAGTGAATGGAGCAGCCAGC
SARS-CoV-2 N Gene Pseudovirus	AUGUCUGAUAAUGGACCCCAAAAUCAGCGAAAUGCACCCCGCAUUACG

	UUUGGUGGACCCUCAGAUUCAACUGGCAGUAACCAGAAUGGAGAACGC AGUGGGGCGCGAUCAAAACAACGUCGGCCCCAAGGUUUACCCAAUAAU ACUGCGUCUUGGUUCACCGCUCUCACUCAACAUGGCAAGGAAGACCUU AAAUUCCCUCGAGGACAAGGCGUUCCAAUUAACACCAAUAGCAGUCCA GAUGACCAAAUUGGCUACUACCGAAGAGCUACCAGACGAAUUCGUGGU GGUGACGGUAAAAUGAAAGAUCUCAGUCCAAGAUGGUAUUUCUACUA CCUAGGAACUGGGCCAGAAGCUGGACUUCCCUAUGGUGCUAACAAAGA CGGCAUCAUAUGGGUUGCAACUGAUGGGAGCCUUGAAUACACCAAAAG AUCACAUUGGCACCCGCAAUCCUGCUAACAAUGCUGCAAUCGUGCUAC AACUUCCUCAAGGAACAACAUUGCCAAAAGGCUUCUACGCAGAAGGGA GCAGAGGCGGCAGUCAAGCCUCUUCUCGUUCCUCAUCACGUAGUCGCA ACAGUUCAAGAAAUUCAACUCCAGGCAGCAGUAGGGGAACUUCUCCUG CUAGAAUGGCUGGCAAUGGCGGUGAUGCUGCUCUUGCUUUGCUGCUG CUUGACAGAUUGAACCAGCUUGAGAGCAAAAUGUCUGGUAAAGGCCAA CAACAACAAGGCCAAACUGUCACUAAGAAAUCUGCUGCUGAGGCUUCU AAGAAGCCUCGGCAAAAACGUACUGCCACUAAAGCAUACAAUGUAACAC AAGCUUUCGGCAGACGUGGUCCAGAACAAACCCAAGGAAAUUUUGGGG ACCAGGAACUAAUCAGACAAGGAACUGAUUACAAACAUUGGCCGCAAA UUGCACAAUUUGCCCCCAGCGCUUCAGCGUUCUUCGGAAUGUCGCGCA UUGGCAUGGAAGUCACACCUUCGGGAACGUGGUUGACCUACACAGGU GCCAUCAAAUUGGAUGACAAAGAUCCAAAUUUCAAAGAUCAAGUCAUU UUGCUGAAUAAGCAUAUUGACGCAUACAAAACAUUCCCACCAACAGAG CCUAAAAAGGACAAAAAGAAGAAGGCUGAUGAAACUCAAGCCUUACCG CAGAGACAGAAGAAACAGCAAACUGUGACUCUUCUUCCUGCUGCAGAU UUGGAUGAUUUCUCCAAACAAUUGCAACAAUCCAUGAGCAGUGCUGAC UCAACUCAGGCCUAA
CDC approved N1 gene forward primer	GACCCCAAAATCAGCGAAAT
CDC approved N1 gene reverse primer	TCTGGTTACTGCCAGTTGAATCTG
N1 gene TaqMan probe	FAM_ACCCCGCATTACGTTTGGTGGACC_BHQ1

Table S2. BLAST result of target sequence of FLPNAS and the sequence of SARS-CoV-2 variants monitored by US CDC.

Variant	GeneBank ID	Consensus of the targeted region of our work (%)
Alpha (B.1.1.7)	GenBank: OW998408.1	100%
Beta (B.1.351)	GenBank: OX008586.1	100%
Gamma (P.1)	GenBank: OX000832.1	100%
Delta (B.1.617.2)	GenBank: OW998779.1	100%
Eta (B.1.525)	GenBank: OX014268.1	100%
Kappa (B.1.617.1)	GenBank: MZ157006.1	100%

Mu (B.1.621)	GenBank: PA544053.1	100%
Zeta (P.2)	GenBank: OW996226.1	100%
Omicron (BA.1)	GenBank: OX315743.1	100%

Table S3. The details of 40 clinical sample.

No.	Name	COVID-19 PCR result	Specimen type
1	UST Mouth G. - 1	Negative	Mouth Gargle
2	UST Mouth G. - 2	Negative	Mouth Gargle
3	UST Mouth G. - 3	Negative	Mouth Gargle
4	UST Mouth G. - 4	Negative	Mouth Gargle
5	UST DTS - 1	Negative	DTS
6	UST DTS - 2	Negative	DTS
7	UST DTS - 3	Negative	DTS
8	UST DTS - 4	Negative	DTS
9	UST NPSTS - 1	Negative	NPS + TS
10	UST NPSTS - 2	Negative	NPS + TS
11	UST NPSTS - 3	Negative	NPS + TS
12	UST NPSTS - 4	Negative	NPS + TS
13	UST NPSTS - 5	Negative	NPS + TS
14	UST NPSTS - 6	Negative	NPS + TS
15	UST NPSTS - 7	Negative	NPS + TS
16	UST NPSTS - 8	Negative	NPS + TS
17	UST NPSTS - 9	Negative	NPS + TS
18	UST NPSTS - 10	Negative	NPS + TS
19	UST NPSTS - 11	Negative	NPS + TS
20	UST NPSTS - 12	Negative	NPS + TS
21	UST NPSTS - 13	Negative	NPS + TS
22	UST NPSTS - 14	Negative	NPS + TS
23	UST NPSTS - 15	Negative	NPS + TS
24	UST NPSTS - 16	Negative	NPS + TS
25	UST NPSTS - 17	Negative	NPS + TS
26	UST NPSTS - 18	Negative	NPS + TS
27	UST NPSTS - 19	Negative	NPS + TS
28	UST NPSTS - 20	Negative	NPS + TS
29	UST NPSTS - 21	Negative	NPS + TS

30	UST NPSTS - 22	Negative	NPS + TS
31	UST NPSTS - 23	Negative	NPS + TS
32	UST NPSTS - 24	Negative	NPS + TS
33	UST NPSTS - 25	Negative	NPS + TS
34	UST NPSTS - 26	Negative	NPS + TS
35	UST NPSTS - 27	Negative	NPS + TS
36	UST NPSTS - 28	Negative	NPS + TS
37	UST NPSTS - 29	Negative	NPS + TS
38	UST NPSTS - 30	Negative	NPS + TS
39	UST NPSTS - 31	Negative	NPS + TS
40	UST NPSTS - 32	Negative	NPS + TS

Note: DTS: deep throat saliva. NP+TS: pooled nasopharyngeal and throat swabs.

Table S4. The expected lengths of amplicons using the synthetic RNA as template with the help of different length of unique primers (UPs). The peak positions of amplicons are converted based on the elution time from the capillary arrays. The reasonable difference was calculated based on the separation resolution provided by the qualitative DNA kits.

The types of Ups used	Expected length (bp)	Peak position observed in Fig S5 (bp)	Reasonable difference (bp)
1	102	100	± 5.1
2	107	105	± 5.35
3	112	110	± 5.6
4	117	119	± 5.85
5	122	122	± 6.1
6	127	128	±6.35
7	132	133	±6.6
8	137	138	±6.85

Note: The sizing accuracy of the used qualitative kit is ± 5%.

Table S5. The expected lengths of amplicons using the SARS-CoV-2 pseudovirus as template with the help of different lengths of unique primers (UPs). The peak positions of amplicons are converted based on the elution time from the capillary arrays. And the reasonable difference was calculated based on the separation resolution provided by the qualitative DNA kits.

The type of added UP	Expected length (bp)	Peak position observed in Fig 2b (bp)	Reasonable difference (bp)
1	93	94	±4.65
2	98	99	±4.9
3	103	103	±5.15

4	108	109	±5.4
5	113	113	±5.65
6	118	119	±5.9
7	123	122	±6.15
8	128	128	±6.4

Note: The sizing accuracy of the used qualitative kit is ± 5%. And the composition of amplicons would be slightly different between using SARS-CoV-2 pseudovirus and synthetic RNA.

Table S6. The estimated economy of FLIPNAS compared with individual PCR test and Dorfman pooling strategy based on the data collected from one practical testing program in Rio de Janeiro State.

Single-tube detection		
	qRT-PCR testing cost per run	US$21.2635
	Total cost	US$129,622.19
Pooling strategy (pool size = 4)		
	People tested	**6096**
	Population prevalence (%)	**430 (7.0)**
	Number of pools tested	**1524**
	Number of positive pools (%)	**365 (24)**
	Total qRT-PCR runs	**2984**
	Total cost	US$63,450.23
	Cost savings (%)	**US$66,171.96 (51.1)**
FLIPNAS assay		
	Total qRT-PCR runs	762
	qRT-PCR testing cost per run	US$21.2635
	Cost of fragment analysis per PCR run	US$0.91
	Total cost	US$16,896.19
	Cost savings compared with individual PCR test (%)	US$112,726.00 (86.97)
	Cost savings compared with Dorfman pooling strategy (%)	US$46,554.04 (73.37)

Note: The information in bold was obtained in the practical work between April and May 2020.

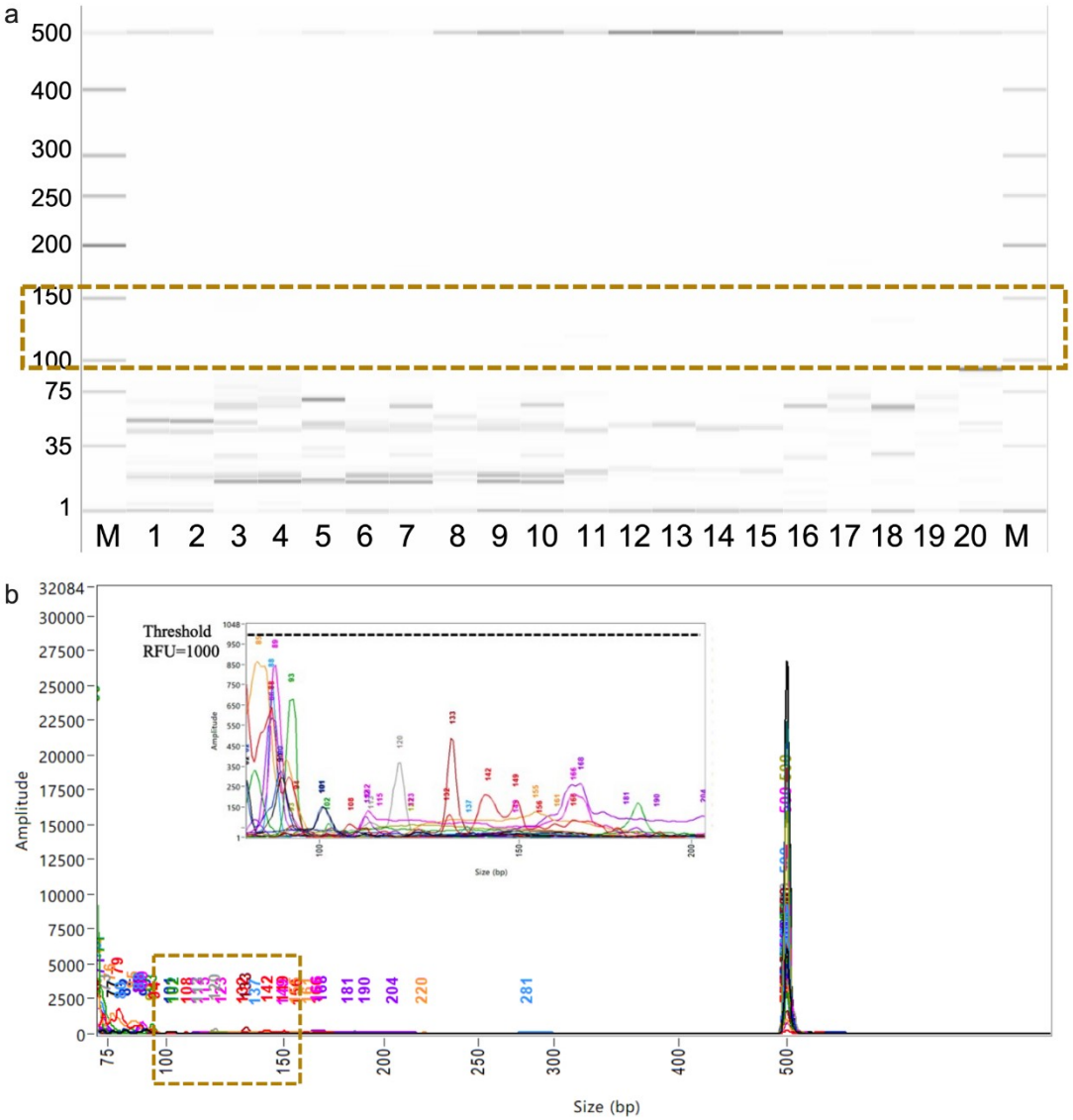

Fig S1. The threshold verification of FLIPNAS. Due to the residues of DNA fragment among repetitive runs and the unexpected primer multimer, the threshold is important for the accuracy of detection. In this study, the threshold was set as 1000 RFU which was decided by 20 blank samples parallelly or after running with mock positive samples. (a) The gel image of blank samples (lanes 1 to 20, M: marker). (b) Electropherograms of 20 blank samples. Inset: the enlarged electropherograms from 90 to150 bp.

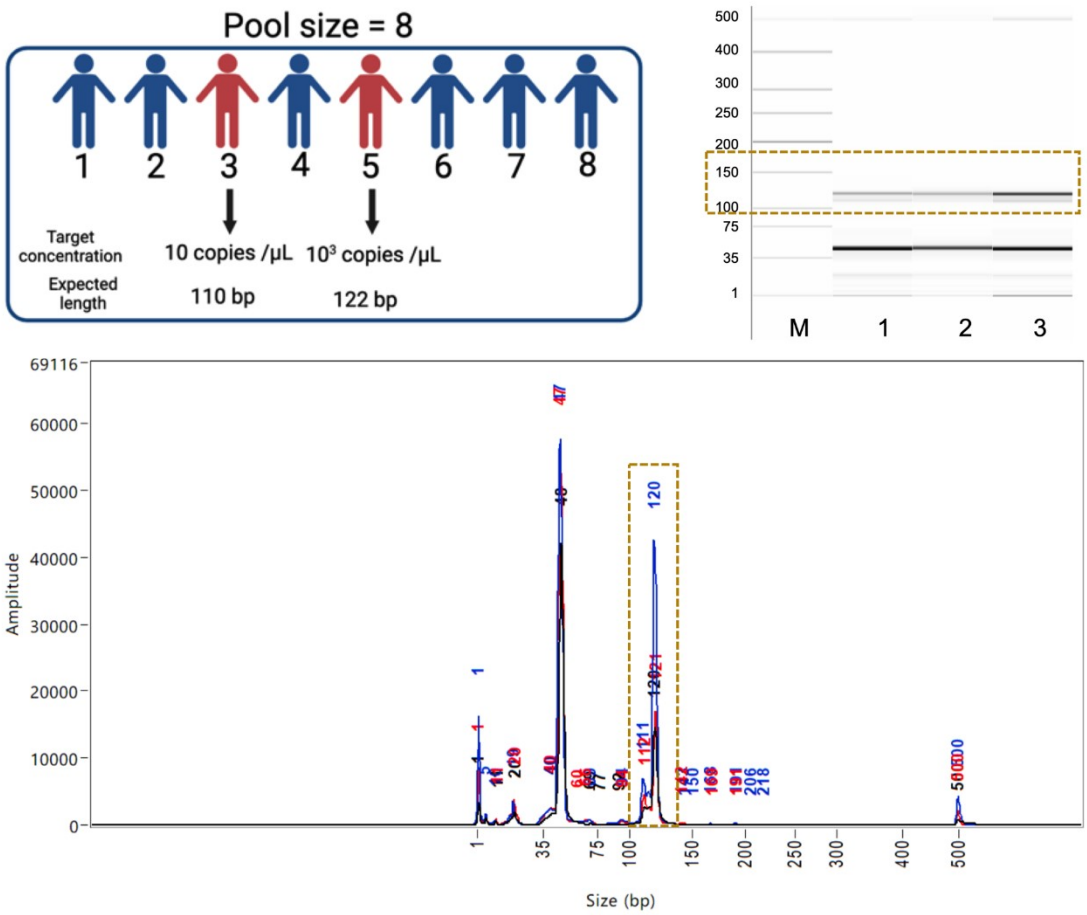

Fig S2. The difficult identification of pooled samples containing two positive samples with large difference in RNA target concentration. The electropherogram shows the result of an 8-in-1 detection with FLIPNAS containing two positive subsamples. The amplicon lengths and target concentration of the two positive subsamples are illustrated in the left-upper panel. The gold dashed rectangles indicate the electrophoresis bands and peaks generated by the two positive subsamples in the fragment analysis.

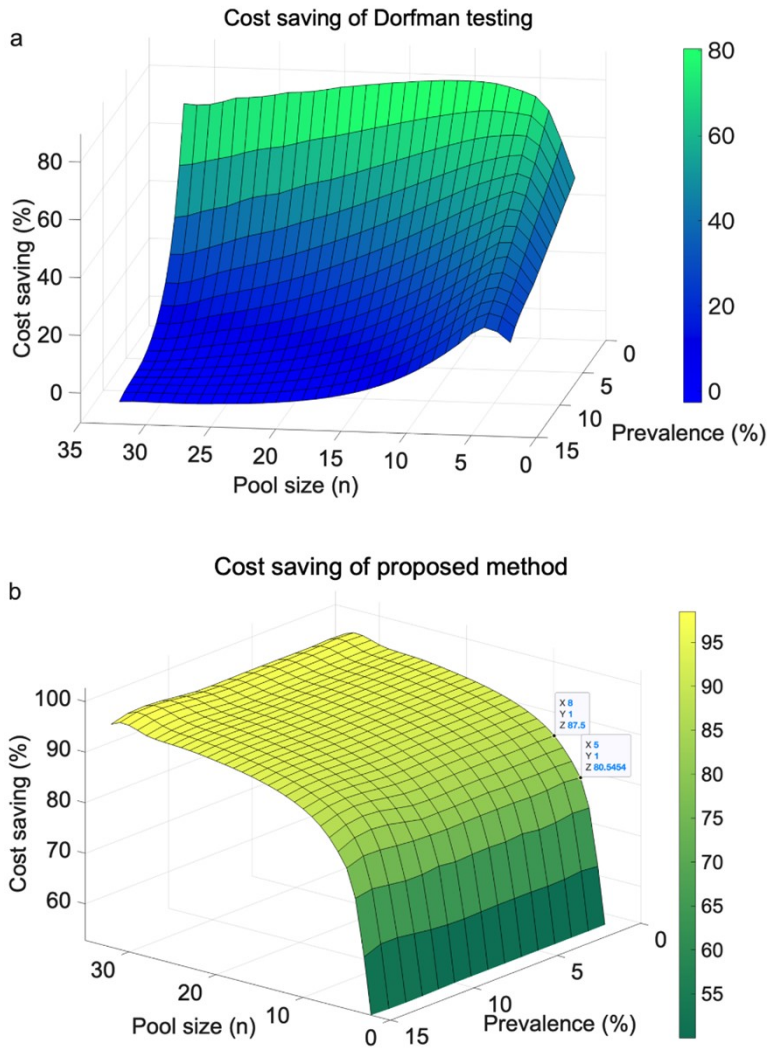

Fig S3 The cost saving model of two group testing methods. Prevalence and pool size were main parameters and considered in the model. (a) The cost-saving model of Dorfman group testing against individual PCR tests, which has been standardized and applied in practice. The economy of pooling testing significantly decreases with the increasing pool size under high prevalence. (b) The cost saving model of FLIPNAS against individual PCR test. With the removal of re-testing of positive pooled samples, theoretically, the economy would keep stable with the increasing pool size and prevalence. However, the impact of dilution effect of pooling testing and enzyme activity would be severely obvious in practical use. Thus, the pool size is chosen to be 8 in this study.

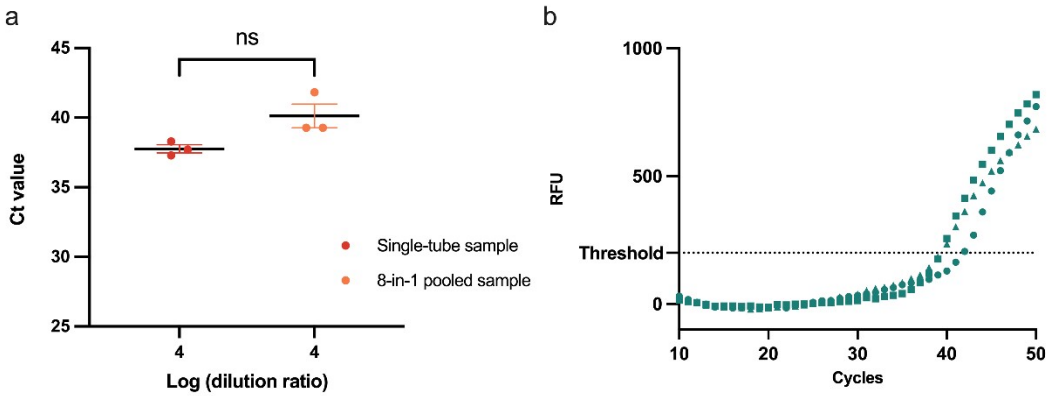

Fig S4. (a) The dilution effect of 8-in-1 pooled samples against single-tube PCR test using one-step RT-qPCR method. The samples were 10^4-fold diluted from purchased stock solution. In individual PCR test, the average Ct value was 37.76. While in 8-in-1 pooled sample detection, the average Ct value was 40.12. The increase in average Ct values was caused by the dilution effect. Although no significantly difference between individual PCR test and 8-in-1 pooling testing was observed (P = 0.0587. *P*-values > 0.05 using Two-tailed Student's *t*-test), the standard deviation was clearly enlarged, hinting the difficulty in pooling testing for samples with high Ct values. 3 technical replicates were used. (b) The amplification curves of 8-in-1 pooled samples containing one positive sample shown in S4(a) (Ct value in individual test=37.76). Currently, the cycle threshold in most standard PCR assays is 40, thus these pooled samples could be considered as negative following current threshold standard, suggesting the need to increase the number of cycles in pooling detection. Threshold was 200 RFU which was automatically set by CFX Opus 96 real-time PCR system and kept the same in all following experiments.

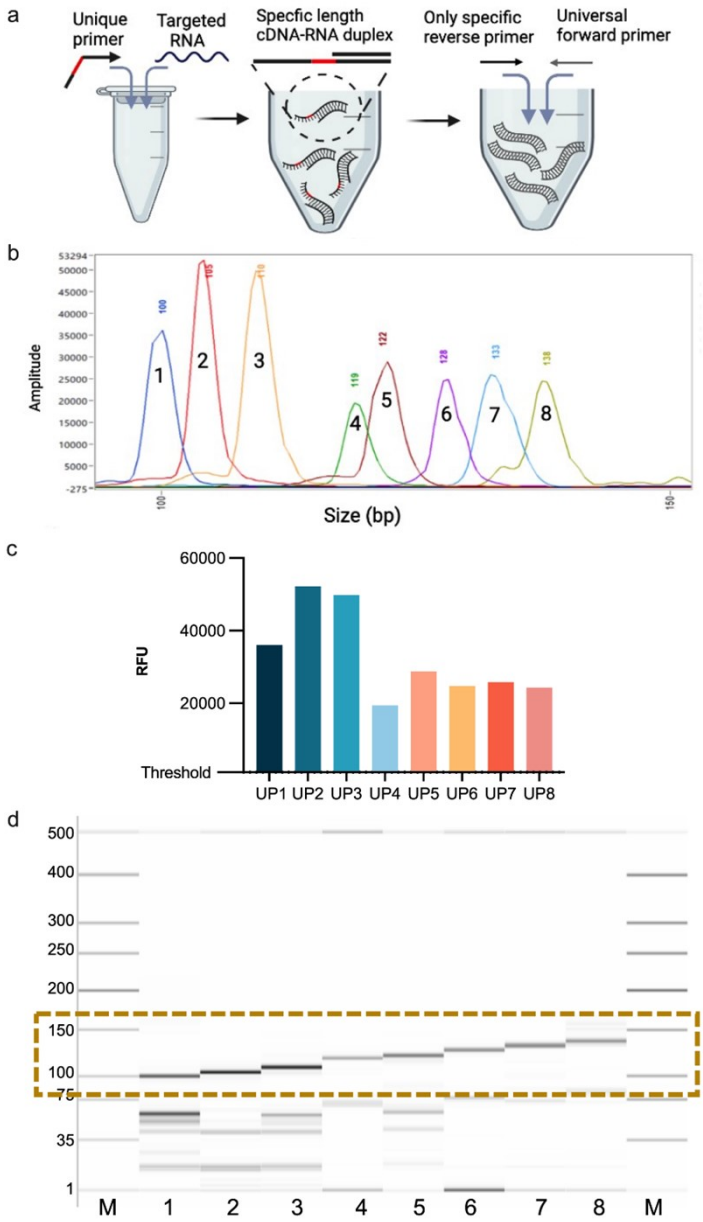

Fig S5. Characterization of the fragment lengths with the help of the unique primer (UP) design of FLIPNAS using synthetic RNA as templates. (a) The illustration of individual samples detection only added the corresponding reverse primer. The in vitro transcribed (IVT) SARS-CoV-2 RNA concentration is 10^3 copies/µL. (b) The superimposition analysis of 8 individual samples' electropherograms. Only designed lengths of amplicons were observed, which were 100, 105, 110, 119, 122, 128, 133, 138 bp, respectively. The *X* axis and *Y* axis indicate the size of PCR amplicons and fluorescent signal, respectively. Separate electropherograms for each target are found in Fig S6. (c) The relative fluorescence units (RFU) of the fragments generated by different UPs in individual samples. 8 fragments with different lengths were recognized as positive comparing the RFU with the threshold. (d) The gel image of individual samples with the addition of different UPs and the corresponding reveres primers. The gold dashed rectangle indicates the electrophoresis bands generated by the individual samples in the fragment analysis. The serial number of the lane indicates the UP used, for example, 1 means that the UP 1 was added in the RT step. M: marker.

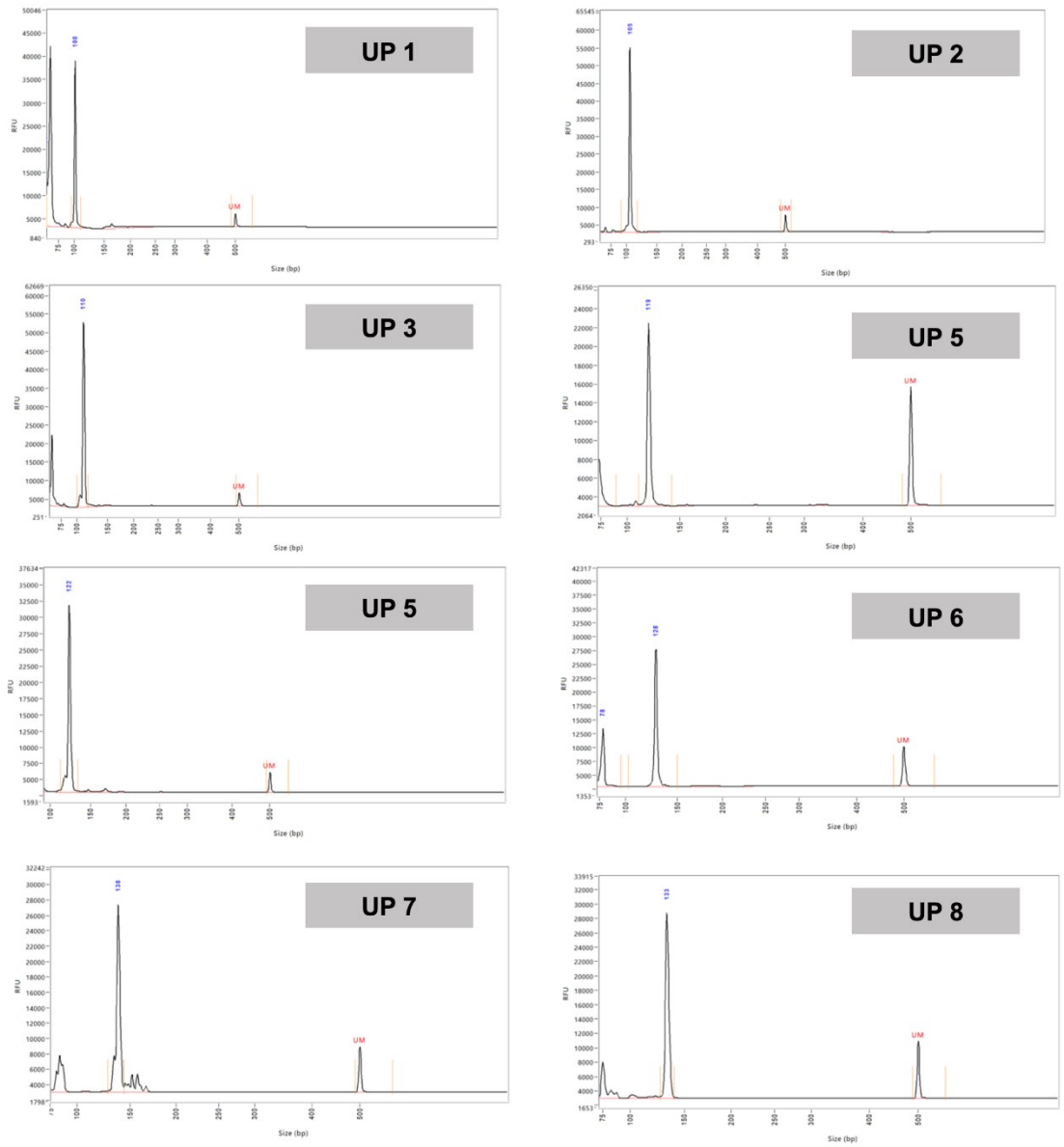

Fig S6. The individual electropherograms of synthetic RNA as templates.

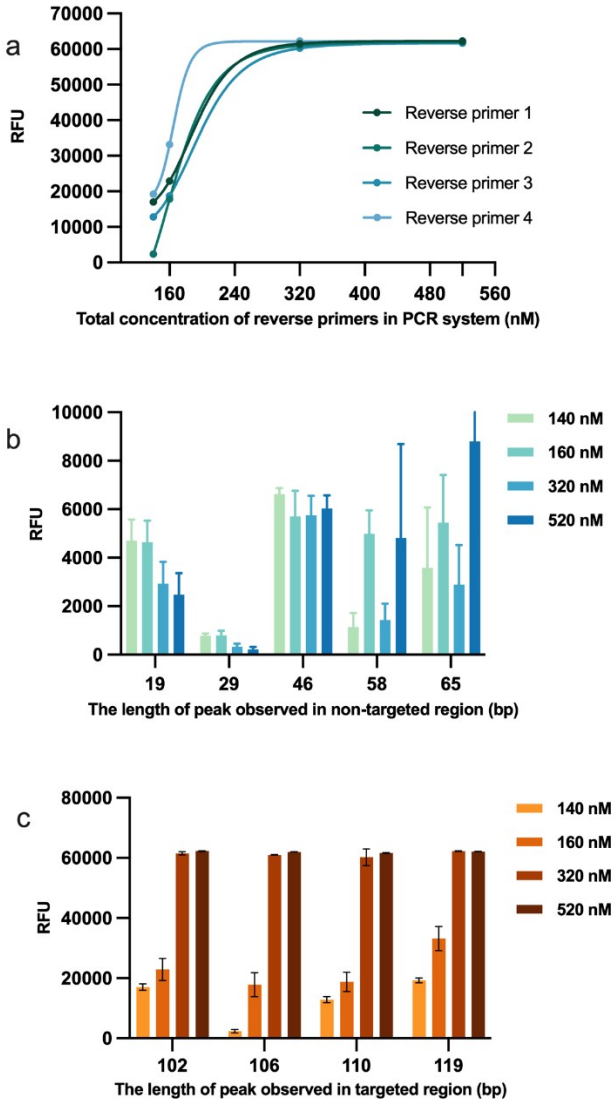

Fig S7. The optimization of the total concentration of this PCR system using synthetic RNA as template. (a) The sigmoidal model of the RFU of fragment with expected length, which was detected by fragment analyzer, against the total concentration of reverse primer in PCR system. According to the types of unique primer added, corresponding reverse primer 1, 2, 3, 4 were respectively added and its concentration was tuned to be 140 nM, 160 nM, 320 nM, and 520 nM in 25 µL reaction system. The RFU of fragment analysis was saturated when the total concentration of reverse primer was 320 nM. The possible reasons may be the restricted amplification efficiency in high primer concentration; limited fluorescent response of fragment analyzer for high concentration fragments. In this study, the optimal total reverse primer was 320 nM, and each reverse primer concentration was 40 nM in the 25 µL reaction system. (b) The RFU of fragments observed in non-targeted region with increasing total concentration of reverse primer in PCR system. (c) The RFU of fragments observed in the targeted region with increasing total concentration of reverse primer in PCR system. RFU: relative fluorescence units. Bars indicate the mean of RFU values, and error bars indicate s.d. n=3.

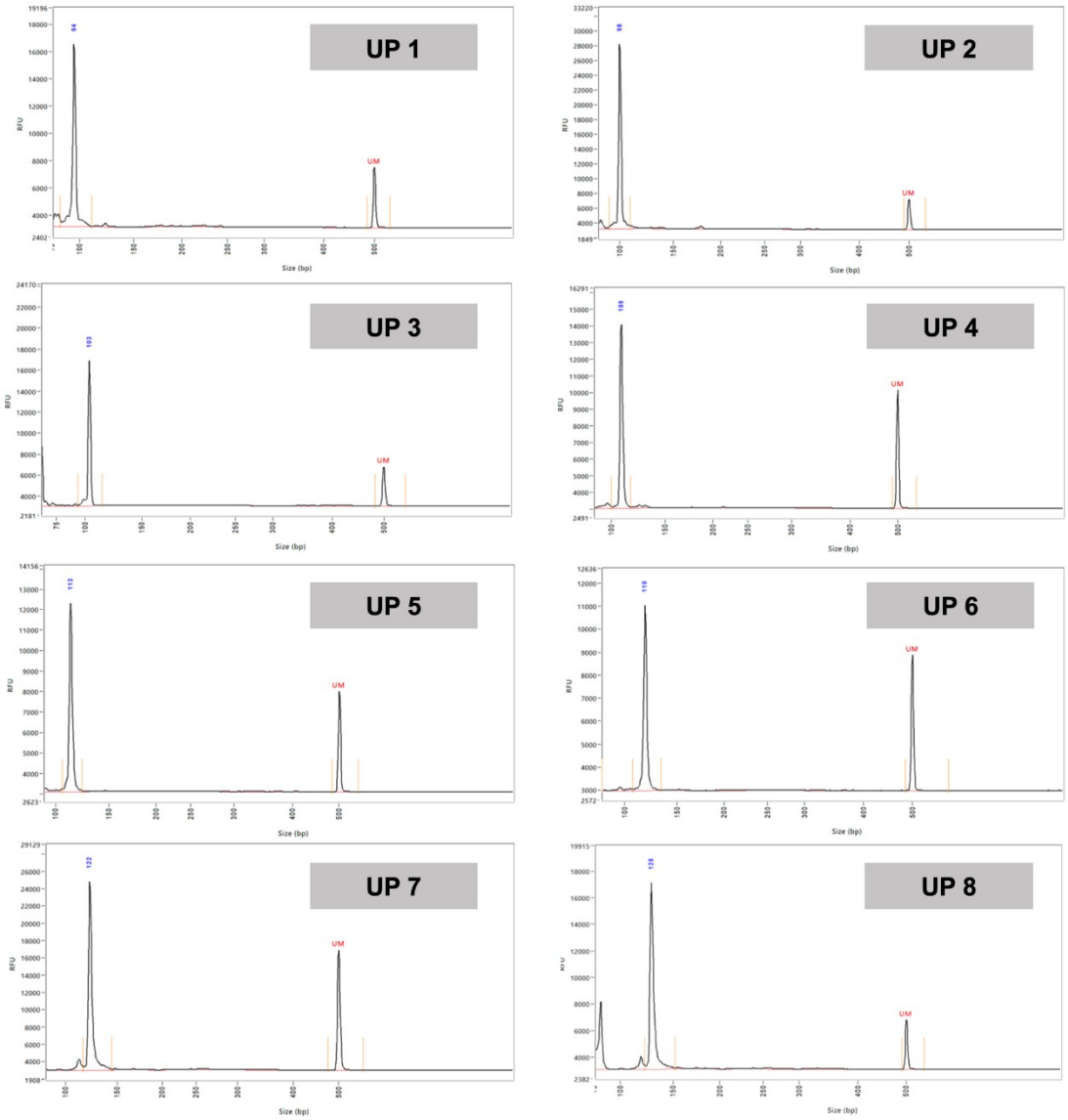

Fig S8. The individual electropherograms of mock SARS-CoV-2 pseudovirus as samples.

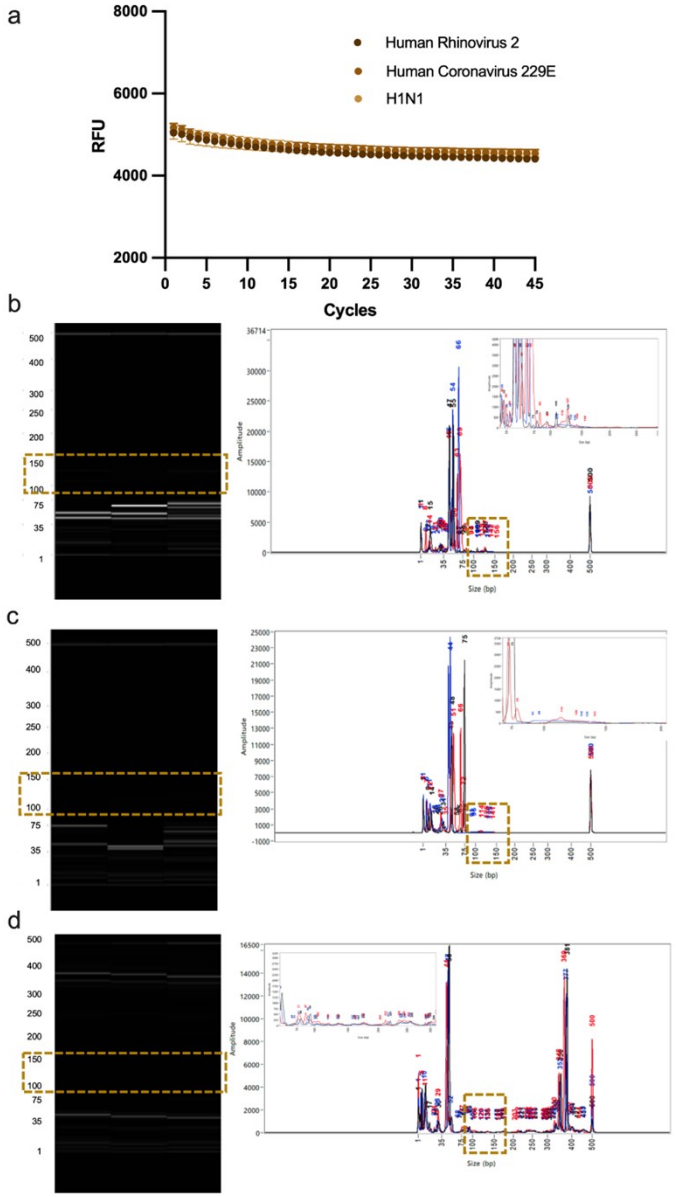

Fig S9. (a) The amplification plots of the H1N1, Human Coronavirus 229E, and Human Rhinovirus 2 using the same protocol of FLIPNAS targeting the SARS-CoV-2 *N1* gene. No increment in RFU was observed, indicating no amplification was conducted with the primers set used in FLIPNAS. The electropherograms of (b) Human Rhinovirus 2, (c) Human Coronavirus 229E, and (d) H1N1. The black circle of gel images and the gold dashed rectangle of electropherograms indicated the targeted region in this study. Inset: the enlarged electropherograms from 90 to150 bp.

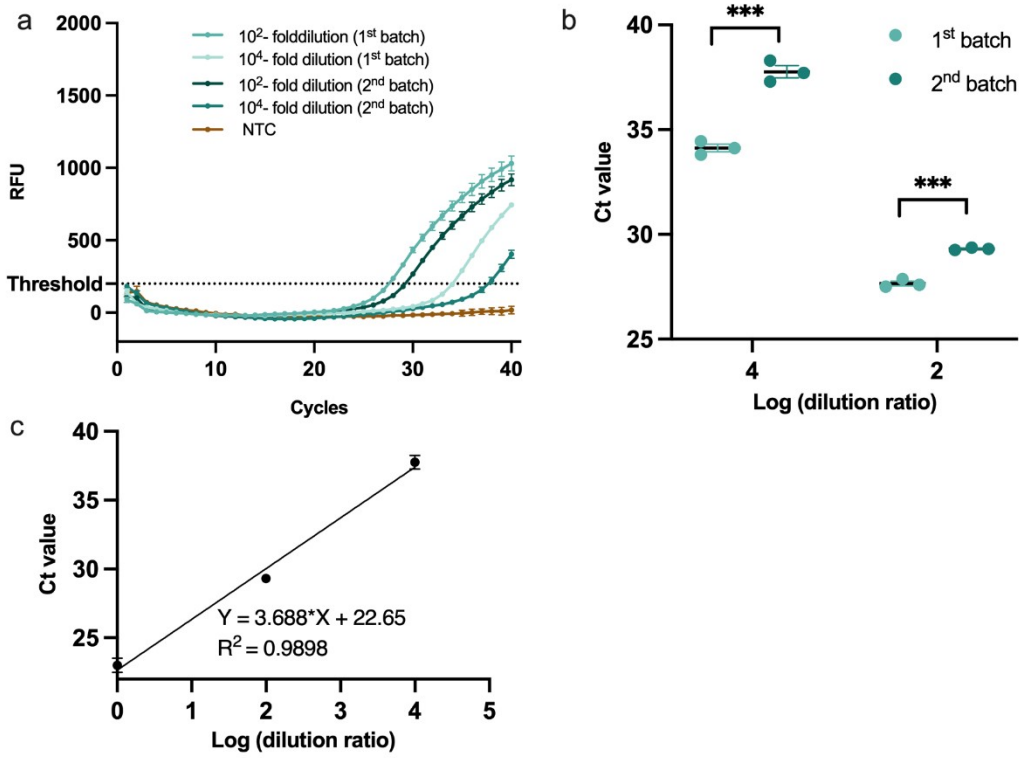

Fig S10. The confirmation of the Ct values and the linear relationship of SARS-CoV-2 pseudovirus used in this work. (a) The amplification curve of SARS-CoV-2 pseudovirus which were 10^2 and 10^4-fold diluted from two batch of stock solution purchased. The first batch (1^{st}) of SARS-CoV-2 pseudovirus was used in another work of our group. The second batch (2^{nd}) of SARS-CoV-2 pseudovirus was used in this study. There is obvious difference in target concentration between two batch of samples. (b) The comparison of Ct values with two batches of SARS-CoV-2 pseudovirus. The concentration of 2^{nd} batch of SARS-CoV-2 pseudovirus was significantly lower than the 1^{st} batch both low and high dilution ratio (P =0.0001 and 0.0004, respectively). This difference possibly came from the batch deviation in manufacture, the difference in storage and transportation conditions. (*P-values < 0.05; ** P-values < 0.005; *** P-values < 0.0005 using Two-tailed Student's t-test) (c) Liner relationship between the SARS-CoV-2 pseudovirus loads and the Ct values which were confirmed via gold standard one-step RT-qPCR assay. The linear regression equation: Y = 3.688*X +22.65 (R^2=0.9898), Y means Ct value, X means the Log of dilution ratio of stock solution. The SARS-CoV-2 pseudovirus loads used in this work were described as Ct values according to this linear relationship.

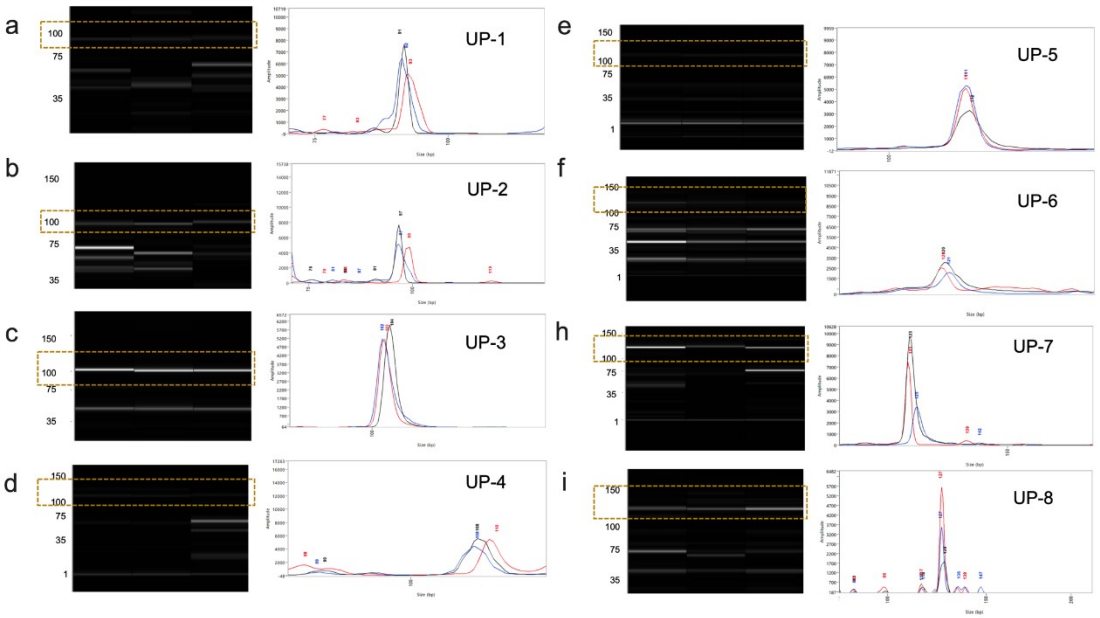

Fig S11. The gel images and electropherograms of pooled samples contained one low viral load of the positive subsample. The Ct value of mock positive samples was 37.4. The targeted fragment of these 8-in-1 pooled samples was generated mediated with (a) unique primer 1, (b) unique primer 2, (c) unique primer 3, (d) unique primer 4, (e) unique primer 5, (f) unique primer 6, (h) unique primer 7, (i) unique primer 8. 3 technical replicates were used in (a-i). The RFU of fragments within the targeted region was higher than the threshold, which can be considered positive.

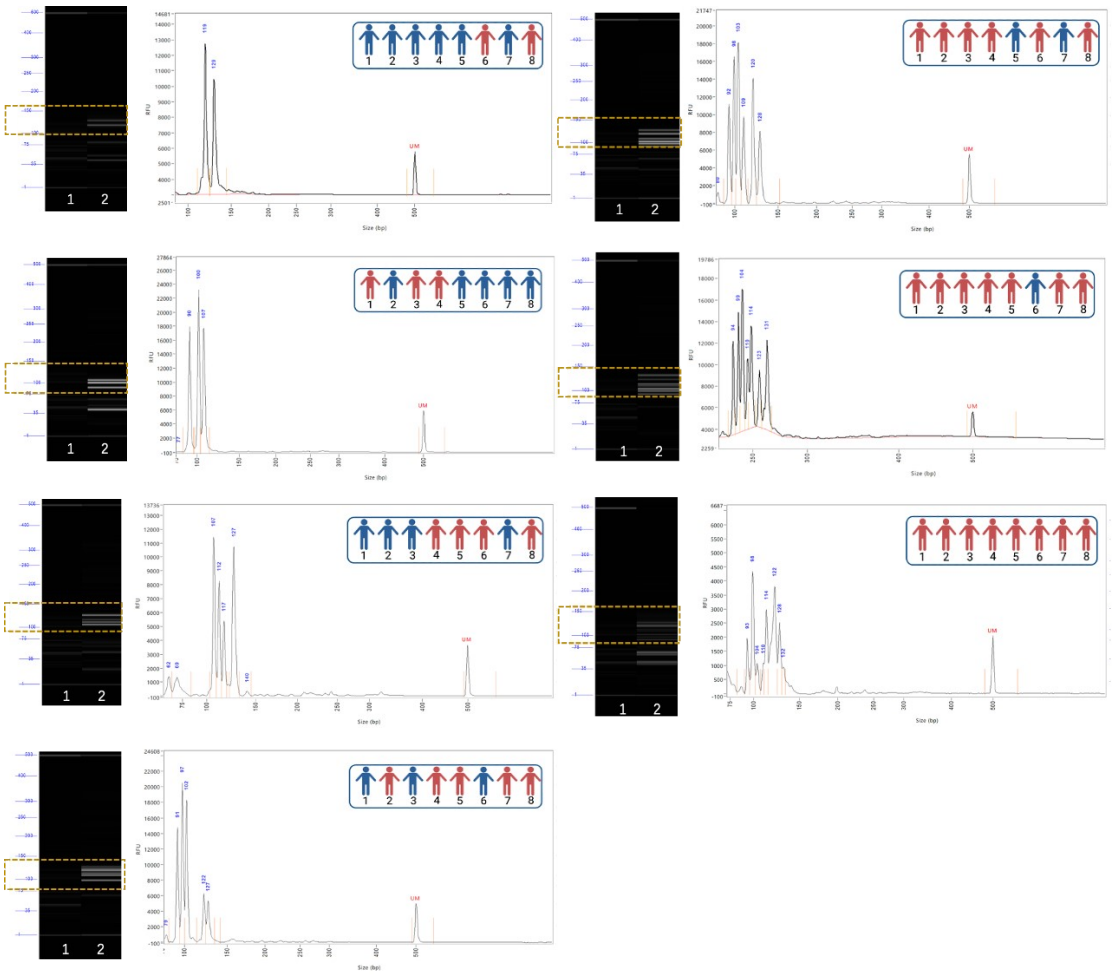

Fig S12. The robustness of FLIPNAS when multiple positive subsamples with similar viral load (Ct = 26.3) were encountered in one pooled sample. In one pooled sample, the mock positive samples can be directly and specifically recognized from gel images when the number of positive samples is not more than 4. The results of positive mock samples can still be identified from the positions in electropherograms with the increasing number of positive subsamples.

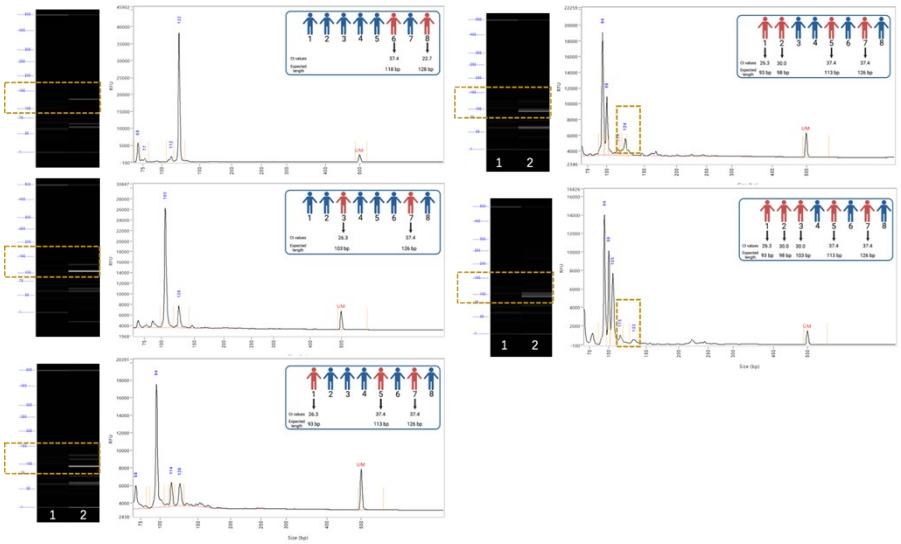

Fig S13. The robustness of FLIPNAS when multiple positive subsamples with different viral loads were encountered in one pooled sample.

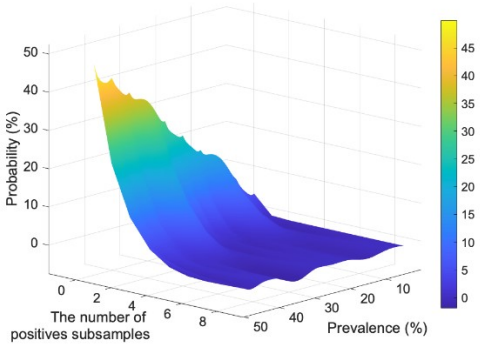

Fig S14. The probability of multiple positive subsamples in pooled samples against disease prevalence. With the prevalence range of 4%~ 15%, the probability of pooled samples, containing more than 2 positive subsamples, is less than 0.5%.

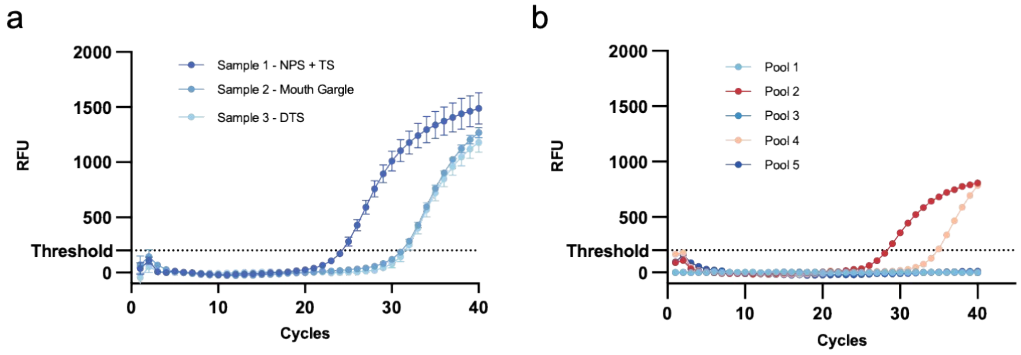

Fig S15. (a) The amplification curve of three mock positive samples spiked using gold standard one-step RT-qPCR assay. Different loads of SARS-CoV-2 pseudovirus were spiked in negative clinical samples. The clinical sample types include mouth gargle, deep throat saliva (DTS), and pooled nasopharyngeal and throat swabs (NP+TS). (b) The amplification curve of 5 pooled samples. No amplification was observed in negative pool 1, 3, 5. Noticeable amplification was observed in pool 2 and 4. Moreover, the increased Ct values were also observed compared with individual RT-qPCR detection.

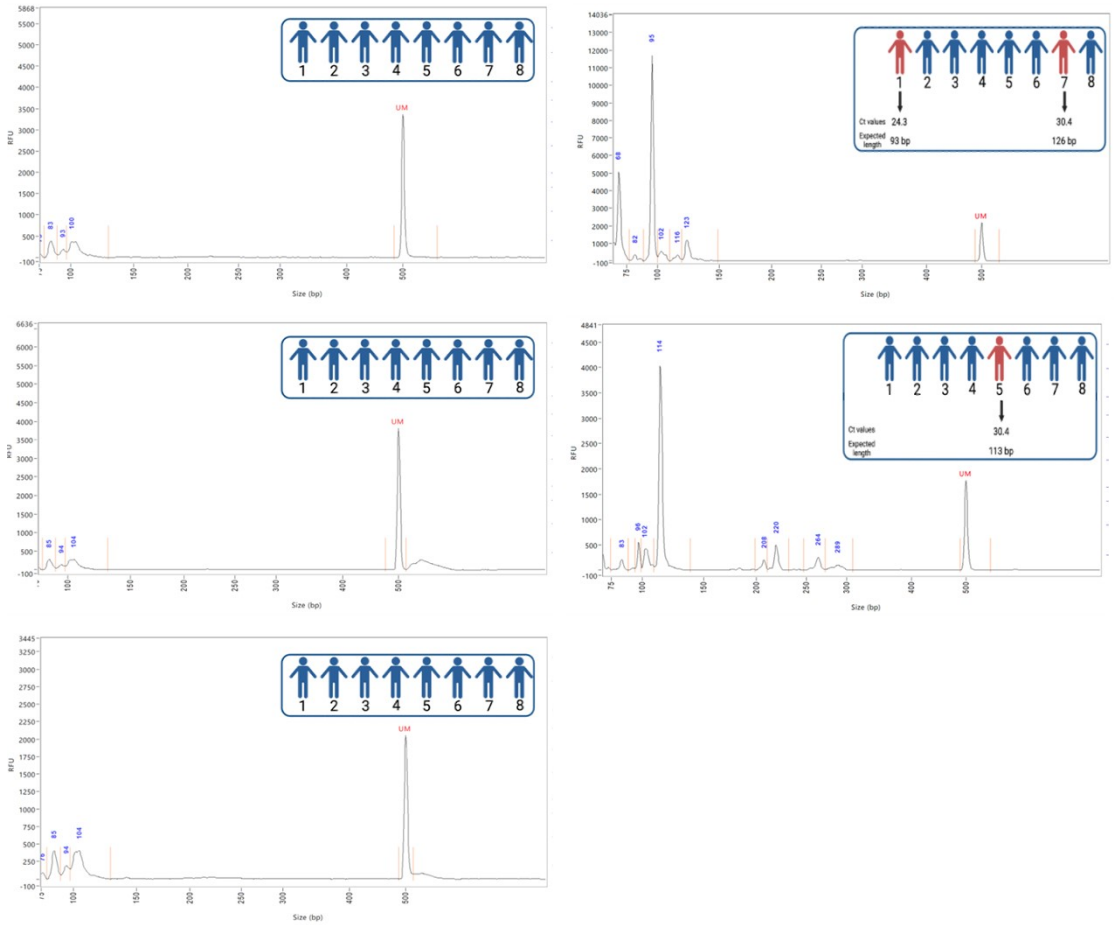

Fig S16. The results of fragment analysis of 5 pooled samples.

Text S1. Unique length region design algorism in NUPACK.

material = dna

trials = 5

temperature = 42.0

material = dna

magnesium[mM] = 0.0

trials = 5

sodium[M] = 0.5

#

target structures

#

structure ID1 = U42

structure ID2 = U47

structure ID3 = U52

structure ID4 = U57

structure ID5 = U62

#

domain a = N5

domain b = N10

domain c = N15

domain d = N20

domain q = TCTGGTTACTGCCAGTTGAAT

domain RP1 = GCAGAGTCGGCCTACAGGTTT

domain RP2 = TGCCTACTACTCCACATAGCA

domain RP3 = CTCTCAACCTCCACCCCTTCA

domain RP4 = CTGCTGTATGCATCCAAGGCG

domain RP5 = CGAGAGCCAGGTAACGAATGG

#

strands (optional, used for threading sequence information

and for displaying results)

#

strand J = RP1 q

strand M = RP2 a q

strand Y = RP3 b q

strand o = RP4 c q

strand u = RP5 d q

#

thread strands onto target structures

#

ID1.seq = J

ID2.seq = M

ID3.seq = Y

ID4.seq = o

ID5.seq = u

#

prevent sequence patterns

(stringency reduced due to constraints imposed by source sequence)

#

prevent = AAAA, CCCC, GGGG, TTTT, AAA, CCC, GGG, TTT, KKKKKK, MMMMMM, RRRRRR, SSSSSS, WWWWWW, YYYYYY

Note: Because the maximum design workload of NUPACK is five at one time, after first round design, another three unique primers can be designed with little change of above algorism.

Text S2. The estimate of optimal pool size with the modified model

Mathematical model of Dorfman group testing benefit, compared with single test, was built up based on a practical application.[1] Based on this research, our proposed method benefit, compared with the Dorfman group testing, can be simply modified by removal the term (in red).

$$p\ (neg) = (1 - p)^n$$

$$V\ (v, P, n) = \frac{v}{n} + v\ (1 - p(neg))$$

The p is the possibility of obtained positive result in single test. $p\ (neg)$ means the possibility of negative results obtained from in Dorfman group testing. n is the pool size conducted in group testing. v is the cost of single test which were needed in that large-scale screening of SARS-CoV-2 for industrial workers of Rio de Janeiro State.

After analyzing the data from abovementioned screening and modification of the mathematical model, the term of x, y, z (following) can be obtained and input in MATLAB, then the benefit model can be obtained.

Matlab code:

```
x_max = max(x);
x_min = min(x);
y_max = max(y);
y_min = min(y);
[X,Y]= meshgrid(x_min:0.1:x_max, y_min:1:y_max)
Z = griddata(x,y,z,X,Y,'v4')
surf(X,Y,Z)
shading interp
colorbar
```

Text S3. The script of random sampling used in the part of clinical validation.

```
operation = [zeros(1,37) 1 2 3];
operation = operation(:,(randperm(size(operation,2))));
B = reshape(operation,5,8);
for i = 1:1:3
    a = (randi(4,1,1)==(1:3))*[24.27 30.43 35.92]';
    B(B==i)=a;
End
```

Reference:

(1) Correa, I. A.; de Souza Rodrigues, T.; Queiroz, A.; de França Nascimento, L.; Wolff, T.; Akamine, R. N.; Kuriyama, S. N.; da Costa, L. J.; Fidalgo-Neto, A. A. Boosting SARS-CoV-2 Detection Combining Pooling and Multiplex Strategies. Sci Rep 2022, 12 (1). https://doi.org/10.1038/s41598-022-12747-8.