Non-specific Primers Reveal False-negative Risk in Detection of COVID-19 Infections

Wei Liu*

School of Life Sciences, Zhengzhou University, Zhengzhou, China.

*Correspondence to: Prof. Wei Liu (Ph.D.), weiliu@zzu.edu.cn.

Keywords: COVID-19, False-negative detection, Non-specific primers, SARS-Cov-2

Summary

Background A novel coronavirus disease 2019 (COVID-19) broke out in Wuhan of Hubei province and had spread throughout the world since December 2019. Because the clinically diagnosed cases in Hubei province were reported for the first time on February 13, 2020, a very high peak of new cases in China was observed. The reason why so many clinically diagnosed cases appeared was not clear.

Methods All data of new cases in China were acquired from WHO situation reports. Linear fitting was used to infer the ability to detect COVID-19 infections. Primer-BLAST and nucleotide blast were applied to check the specificity of primers. Expression data of human mRNA in different tissues was obtained from Human Protein Atlas.

Findings Based on the data and analysis of changes of new laboratory-confirmed cases and new clinically diagnosed cases, it was inferred that there were many false-negative results in all clinically diagnosed cases in Hubei province. There were eight non-specific primers in dozens of primers used in clinical or research detection of COVID-19. Among them, a pair of primer for the ORF1ab regions of SARS-CoV-2 genome, which widely applied to detect SARS-CoV-2 virus in China, well matched some human mRNAs such as Cathepsin C transcripts. Compared to other transcripts, Cathepsin C mRNA had a high abundance in tonsil, lung and small intestine.

Interpretation Some non-specific RT-PCR primers could cause the serious interference during RT-PCR amplification so as to increase the risk of false-negative diagnoses for COVID-19 infections.

Funding Key Research Project of the Higher Education of Henan Province

Introduction

The World Health Organization (WHO) had declared coronavirus disease 2019 (COVID-19) a pandemic on March 11, 2020, in which SARS-CoV-2 virus had spread around the world.[1] SARS-CoV-2 virus was first isolated from patients in Wuhan and identified as a novel coronavirus in late December, 2019.[2] It caused an outbreak of COVID-19 in China in late January, 2020. Although so far WHO had reported more than 82,000 confirmed coronavirus infections in China, there was few new native cases in late March, 2020. However, there were more than 1 million confirmed coronavirus infections around world.[3] As an easy and rapid method to diagnose coronavirus virus, including SARS-CoV-2, SARS and so on, reverse transcription-polymerase chain reaction (RT-PCR) was commonly used to confirm COVID-19 infections worldwide. Some institutes or universities provided several RT-PCR assays to diagnose 2019-nCoV (previous name of COVID-19) and the protocols.[4]

Because of occurrence of abnormal results, several articles from Chinese research groups had referred to the RT-PCR diagnoses for COVID-19 infections according to China CDC guideline.[5] False-negative RT-PCR test results had been reported by a previous study.[6] Five in 167 patients had presented negative results in first RT-PCR detection for 2019 novel coronavirus with positive chest CT finding. All of them were finally confirmed with COVID-19 infection by second, third or multiple repeated RT-PCR detections. A correspondence on viral load in upper respiratory samples showed that some patients received negative or positive results alternately and repeatedly by using the primers targeted the ORF1ab region of novel coronavirus genome, and similar situation was also described in another correspondence.[7, 8] A recent letter discussed positive RT-PCR test results after two consecutively negative results in four patients recovered from

COVID-19, and could not rule out false-negative results.[9]

The investigation had paid attention to the above phenomenon and tried to understand the extent of false-negative results from changes of new cases of the COVID-19 infections in China. Based on the primers provided by research institutes from different countries, especially primers from China, detailed analysis of non-specificity of primer sequences had been conducted, and interference of human mRNA targeted by the primer was discussed deeply for RT-PCR detection of COVID-19 infections.

Methods

Data source

Numbers of new cases of COVID-19 infections were released by WHO.[3] Data spanned from January 22, 2020 to March 6, 2020. Seventeen specific primers for detection of Covid-19, as shown in Table S1, were provided by several institutes or universities from different countries.[4] Among them, two sets of primers, including forward primer (F), reverse primer (R) and fluorescence probe, were designed by National Institute for Viral Disease Control and Prevention (IVDC), China CDC.[5] Since the onset of the COVID-19 outbreak, two sets of primers had recommended to guide disease prevention and control in China.[10,11] All mRNA expressions in different human organs were obtained from Human Protein Atlas (HPA, https://www.proteinatlas.org/).[12]

Analysis of primer specificity

Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi) was applied to determine whether specific primers for SARS-CoV-2 virus show significant matches with human

RNA or not.[13] Two "Exclusion" options should be selected to avoid PREDICTED results, and other options were set as default. Nucleotide blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was also used to check the specificity of primers, in which options Refseq RNA was chosen as database, word size was set to 7, and expect threshold was set to 1000.[14]

Results

Figure 1 showed the new cases of the COVID-19 infections in China from January 22 to March 6, 2020. Blue, green and red lines represented laboratory-confirmed cases, clinically diagnosed cases and total new cases, respectively. Numbers of new laboratory-confirmed cases and new clinically diagnosed was not reported from February 17 to February 19, 2020 and clinically diagnosed cases in Hubei province was removed by WHO from February 20, 2020. Except for positive results in RT-PCR detection, clinically diagnosed patients met all clinical diagnostic conditions including chest CT images.[15, 16] Since China reported 12,289 clinically diagnosed cases in Hubei province for the first time, number of total new cases grew explosively on February 13, 2020, whereas number of clinically diagnosed cases decreased dramatically to 888 from February 14 to February 16, 2020. Meanwhile number of new laboratory-confirmed cases decreased gradually from less than 2,000 to about 1,000, which were obviously less than 3,893 new laboratory-confirmed cases on February 5, 2020. The inset of Figure 1 showed the linear fitting of the number of laboratory-confirmed cases to the number of days from January 22 to February 5, 2020. About 5512 detections could be carried out according to linear fitting formula (the number of detection = 260.32 × days - 475.44) on February 13, 2020, while there were 1820 laboratory-confirmed cases.

Figure 2 showed the position of all primers on the genome of SARS-CoV-2 Wuhan-Hu-1 strain.

There were seventeen primers, thirteen for universal RT-PCT and four for two sets of nested RT-PCR. The lengths of all primers were between 17 and 26. The numbers of primers targeted ORF1ab, spike protein S, envelop protein E and nucleoprotein N nucleotide sequences were 7, 2, 1 and 7, respectively. While three primers overlapped to some extent in *RdRp* gene in ORF1ab region, two and three primers overlapped to some degree in nucleoprotein N gene, respectively. Except for nested RT-PCR with amplicon size of 346 to 547 nt, the lengths of all RT-PCR amplification products were between 57 and 158 nt.

The specificity of all primers was checked with Primer-BLAST and nucleotide blast to ascertain whether it were a primer for one or more human RNA transcripts or not. It is important to note that more meaningful matches could be obtained using Primer-BLAST than nucleotide Blast, and choosing Refseq mRNA database could acquire a little more matches than choosing Refseq RNA database with Primer-BLAST.

Table 1 listed the non-specific primers originally designed for SARS-CoV-2 and their unintended human mRNA targets found by using Primer-BLAST. Besides those non-specific primers, all other primers in Figure 2 were also checked with Primer-BLAST. Three new combinations of the forward and reverse sequences of primer 1 (for ORF1ab regions) matched many human mRNA transcripts to a higher degree, whereas there were no any primer matching human mRNAs for primer 2 (for N gene). Cross combinations of the forward and reverse sequences of primer 1 and 2 did not match more human mRNAs than the above combinations. Several consecutive nucleotides at 3’ ends of four new primers composed of the reverse sequence of primer 1 perfectly matched those of *CTSC*, *ZNF7*, *FYCO1* and *RRAGB* transcripts, and its probe (5'-FAM-CCGTCTGCGGTATGTGGAAAGGTTATGG-BHQ1-3') could not match *CTSC*,

ZNF7, *FYCO1* and *RRAGB* transcripts more than nine nucleotides.

Furthermore, all primers from NIID, MPH and IP and one primer from HKU in Table S1 were non-specific primers. Primer 6, 7B, 8B, 9, 13, 14 and 15 matched 3, 2, 29, 1, 13, 30 and 1 human mRNAs or their transcript variants, respectively. Although there were different clinical respiratory symptoms between SARS-CoV-2 and SARS virus, both of them belonged to beta coronavirus. Nevertheless, it was a remarkable fact that there were no any SARS primers matched human mRNAs.[21] Otherwise, it was noted that no pathogen mRNAs were considered except for human host.

Expression of target human mRNAs of primer 1 was shown in Figure 3 and those of other primers were also listed in Table S2. The result indicted that Cathepsin C (*CTSC*) mRNA had a much higher abundance than other mRNAs in the respective tissues, but difference of *CTSC* transcripts between individuals was not evident in tonsil. Relative to house-keeping gene glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), protein-transcripts per million (pTPM) values of *CTSC* mRNA in normal tonsil, lung and small intestine were 7.2%, 50.8% and 4.9% of *GAPDH* mRNA, respectively. The pTPM values gave a quantification of the mRNA abundance, which could be compared between different genes and samples. Except for *CTSC*, *SON*, *THOC2* and *AGPAT3* transcripts, all other transcripts shown in Table S3 had smaller average pTPM values. It was noted that the latter three transcripts were targeted by same primer. Relative to house-keeping gene *GAPDH*, the total pTPM values of *SON*, *THOC2* and *AGPAT3* transcripts in normal tonsil, lung and small intestine were 6.5%, 18.0% and 8.3% of *GAPDH* mRNA, respectively.

Discussion

Although many patients tested one or more negative before receiving positive results for SARS-CoV-2 virus in China, it was difficult to understand the extent to which this abnormal phenomenon occurs. However, the WHO reports on new laboratory-confirmed cases and new clinically diagnosed cases in China during the COVID-19 epidemic were helpful to deduce that extent. Rapid raise and drop of total new cases in Figure 1 indicated that the amazing number of clinically diagnosed cases was just an accumulation of those in Hubei province earlier than February 13, 2020. Successive revision in the diagnosis criterion for the novel coronavirus pneumonia in Hubei province had resulted in a sudden surge in the number of total new cases and the removal of clinically diagnosed cases in the province.[16, 17] There was no reports on whether the clinically diagnosed patients tested or not, but it was possible to infer general situation through evaluating the ability to detect COVID-19 infections. The ability to detect COVID-19 infections was steadily enhanced along with the outbreak of the epidemic and then exhibited a great accelerated growth. China National Biotec Group could produce novel coronavirus detection kits for 200,000 people each day on February 18, 2020.[18] In the light of the linear fitting of laboratory-confirmed cases from January 22 to February 5, 2020 in the inset of Figure 1, the ability to detect COVID-19 infections for more than 14,400 persons besides laboratory-confirmed patients could be deduced from February 6 to February 13, 2020, and that basically met the requirements for 12,289 detections for clinically diagnosed patients on February 13, 2020. Hence it was inferred that a considerable number of clinically diagnosed patients tested and obtained negative results.

Furthmore, the number of recovered cases in China moderately increased from 261 on February 5, 2020 to 3,622 on February 27, 2020, which was shown in Figure S1. It represented that

cross-infection in hospital could be well controlled. Because there was no rapid increase of recovered cases after February 13, 2020, it could be concluded that most of more than 17,410 clinically diagnosed patients had finally diagnosed COVID-19 infections. The slowly decline of number of existing confirmed cases as shown in Figure S2 also supported the conclusion. Therefore, it was considered that false-negative results in RT-PCR detection of COVID-19 infections usually occurred among clinically diagnosed patients.

RT-PCR detection was applied to diagnose SARS virus in 2003, but little false-negative diagnosis was reported in more than 8000 confirmed cases. Moremore, even if so many RT-PCR detections were repeated in laboratories and hospitals in China, general operation problems resulting in false-negative detections were not found. Consequently, the specificity of RT-PCR primers was investigated. A pair of specific primers used to detect SARS-CoV-2 virus should only amplify the virus sequence, but not any unintended RNA, including those of human host and other pathogens. In contrast to pathogens, human host was usually omitted.

To eliminate the host interference in RT-PCR detection of COVID-19 infections, specific amplification of the intended target of SARS-Cov-2 sequence required that primers matched as little as possible to any human RNA transcript. Besides primers matched *ZNF7*, *FYCO1* and *RRAGB* transcripts, it was noted that primer1 with sixteen consecutive nucleotides at 3' ends, which perfectly matched those of *CTSC* transcripts except one nucleotide, could cause the serious interference during RT-PCR amplification. The technical guidance, including the primer 1 and 2, was employed to instruct centers for disease control at all levels in China.[6, 7, 9, 19, 20] Although there was no interference with primer 2 for N gene, only both positive PT-PCR detection for COVID-19 infections with primer 1 and 2 were obtained in the same specimen could a positive result be

confirmed in term of the clinical diagnostic criterion for nucleotide detection in China. Hence it was not difficult to draw a conclusion that non-specificity of the primer designed for SARS-CoV-2 virus might be an important factor resulting in so many false-negative diagnoses for COVID-19 infections in China.

Moreover, there were seven non-specific primers besides primer 1. Although primer 7B and 8B were inner primers of nested RT-PCR, second round RT-PCR operation would be susceptible to contamination. Therefore, it was believed that false-negative diagnoses for COVID-19 infections might occur when using the above primers.

Besides the specificity of primers, the amount of the unintended target mRNA was another key factor in determining whether human host contamination could happen in RT-PCR detection. RT-PCR detection samples were collected usually from nasopharynx or oropharynx swap, sputum or fluid in trachea or alveoli and stool, which corresponded to tonsil, lung and intestine. The high abundance of *CTSC* transcripts in tonsil, lung and small intestine tissues and good match of *CTSC* transcripts to primer 1 could lead to the contamination of clinical samples, which were used for RT-PCR detection of COVID-19 infections, and might trigger the false-negative diagnosis. Furthermore, the higher total abundance of *SON*, *THOC2* and *AGPAT3* transcripts in tonsil, lung and small intestine tissues and better matches of the above three transcripts to primer 14 could also result in the same problems. Although other transcripts had smaller pTPM values, expression of the transcripts might become very complex when inflammatory and oxidative stress occurred during viral infections.

According to the above mentioned analysis and data, there could be no doubt that non-specificity of RT-PCR primers obviously increased the risk of false-negative diagnoses for

COVID-19 infections. Since there were so many false-negative diagnoses for COVID-19 infections in Hubei province at the early stage of epidemic and the reason was that some primers were nonspecific, Table 2 gave the suggestion on the RT-PCR primers in order to reduce false-negative results in RT-PCR detection for SARA-CoV-2 virus during COVID-19 pandemic.

Declaration of interests

We declare no competing interests.

References

1. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020. Mar 11, 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (accessed Mar 7, 2020).

2. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020; **382**: 727-33.

3. World Health Organization. Coronavirus disease (COVID-2019) situation reports. April 4, 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200404-sitrep-75-covid-19.pdf?sfvrsn=99251b2_4 (accessed Mar 7, 2020).

4. World Health Organization. Current list of WHO reference laboratories providing confirmatory testing for COVID-19. Mar 2, 2020. https://www.who.int/docs/default-source/coronaviruse/who-reference-laboratories-providing-confirmatory-testing-for-covid-19.pdf?sfvrsn=a03a01e6_2 (accessed Mar 7, 2020).

5. National Microbiology Data Center. Novel Coronavirus National Science and Technology Resource Service System. Jan 24, 2020. http://nmdc.cn/nCoV/en (accessed Mar 7, 2020).

6. Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J. Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. *Radiology* 2020; published online Feb 12. DOI: 10.1148/radiol.2020200343.

7. Zou L, Ruan F, Huang M, et al. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med* 2020; published online Feb 19. DOI: 10.1056/NEJMc2001737.

8. Pan Y, Zhang D, Yang P, Poon LLM. Viral load of SARS-CoV-2 in clinical samples. *Lancet Infect Dis* 2020; published online Feb 24.DOI: 10.1016/S1473-3099(20)30113-4.

9. Lan L, Xu D, Ye, G, et al. Positive RT-PCR test results in patients recovered from COVID-19. *JAMA* 2020; published online Feb 27. DOI: 10.1001/jama.2020.2783.

10. Technical guidance for laboratory testing of 2019-nCoV infection (third edition). *Biosafety and Health*. published online Feb 5. DOI: 10.1016/j.bsheal.2020.02.001.

11. New coronavirus pneumonia prevention and control program (2nd edtion.) (in Chinese). Jan 22, 2020. http://www.nhc.gov.cn/jkj/s3577/202001/c67cfe29ecf1470e8c7fc47d3b751e88.shtml (accessed Mar 7, 2020).

12. Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* 2010; **28**(12):1248-50.

13. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 2012; **13**: 134.

14. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool (BLAST). *J Mol Biol* 1990; **215**(3): 403-410.
15. Shi H, Han X, Jiang N, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis* 2020; published online Feb 24. DOI: 10.1016/S1473-3099(20)30086-4.
16. New coronavirus pneumonia prevention and control program (5th edtion.) (in Chinese). Feb 21, 2020. http://www.nhc.gov.cn/jkj/s3577/202002/a5d6f7b8c48c451c87dba14889b30147.shtml (accessed Mar 7, 2020).
17. New coronavirus pneumonia prevention and control program (6th edtion.) (in Chinese). Mar 7, 2020. http://www.nhc.gov.cn/jkj/s3577/202003/4856d5b0458141fa9f376853224d41d7.shtml (accessed Mar 7, 2020).
18. The State Council of the People's Republic of China. China's central SOEs mobilize to fight against coronavirus. Feb 19, 2020. http://english.www.gov.cn/statecouncil/ministries/202002/19/content_WS5e4c9d14c6d0595e03c210a0. html (accessed Mar 7, 2020).
19. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 2020; published online Feb 7. DOI: 10.1001/jama.2020.1585.
20. Guan W, Ni Z, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020; published online Feb 28. DOI: 10.1056/NEJMoa2002032.
21. World Health Organization. PCR primers for SARS developed by WHO network laboratories. April 17, 2003. https://www.who.int/csr/sars/primers/en/ (accessed Mar 7, 2020).

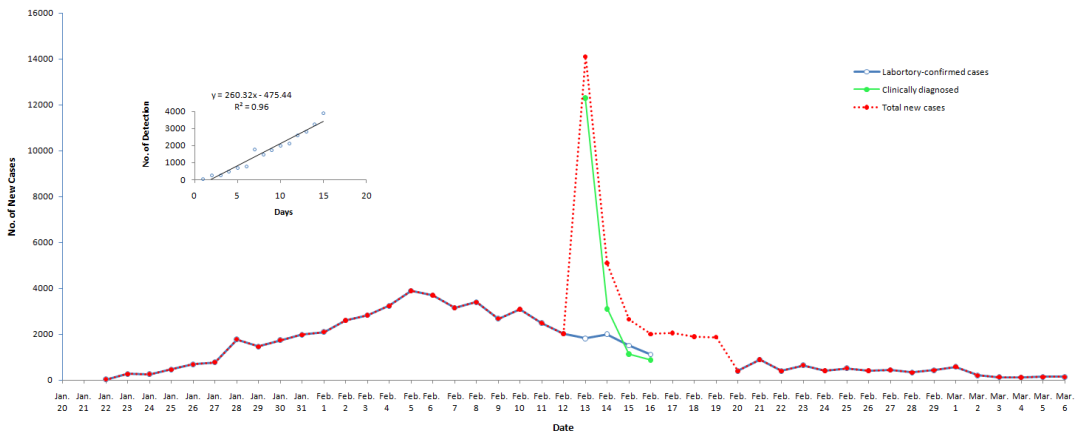

Figure 1: **New cases of the Covid-19 infections in China from January 22 to March 6, 2020.**

Inset showed the number of detection vs. days from January 22 to February 5, 2020. Linear fitting formula was as follows: $y = 260.32\ x - 475.44$.

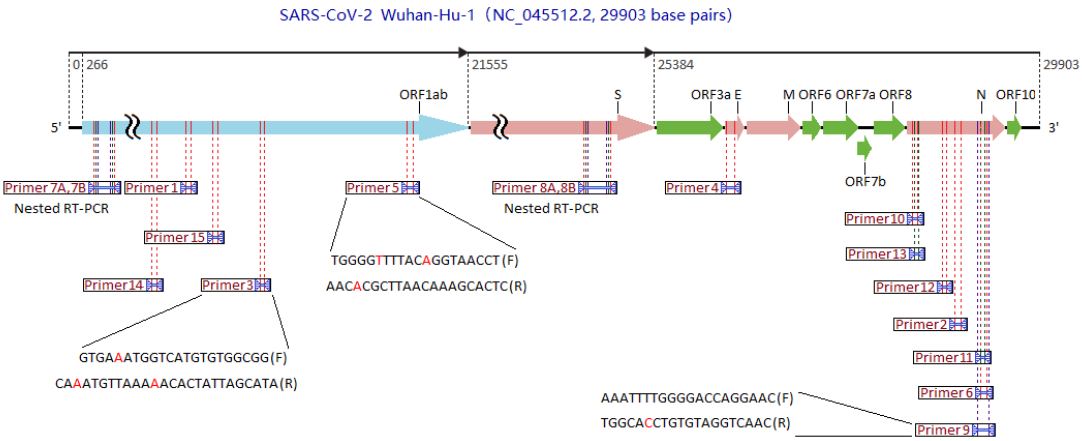

Figure 2: **Schematic diagram of prime positions located on the SARS-CoV-2 genome.**

Bases in red color were determined according to SARS-CoV-2 Wuhan-Hu-1 sequence. A portion of the respective length of ORF1ab and S was omitted.

Non-specific primer		Unintended target		
No.*	New pair	Human mRNA	Template matched by forward primer#	Template matched by reverse primer#
1	F, R	*RNF207*	GCTTCTGGGTTTTACTCTTGA	GCTGTTGTGCATCCGCTGC
	F, F	*CTSC* transcript variant 2, 3	TGTTTTGGGTTTTACATTTAA	AATTTTGGGTTTTACATTTAA
	R, R	*ZNF7* transcript variant 1, 2, 4, 6-10	TCCAGTGTGAATCAGCTGA	CTGAGTCAGCATCAGCTGA
	R, R	FYCO1	ATGATTTTGGAGCAGCTGA	ACCCTTGTGCAGCAGCTGA
	R, R	*RRAGB* transcript variant RAGBl, RAGBs, 3	AACATTGGGAATCAGCTGA	AGCATAGTTTATCAGCTGA
6	R, R	*GRIK3*	AGCCGGTGGGACTTCCATG	CAGAGGATTCACTTCCATG
	R, R	*CNTN2* transcript variant 1, 2	AGAAGGTGGGATCTCCAAG	CCAGGCAGTAACTTCCATG
7B	F, F	*JRK* transcript variant 1	TCCACACTGCACCTCAGGG	CTCCCACTGAGCCTCATGG
	F, F	*SPEN*	CTGCTCCTGCCCCTCATGG	CACAATCTGCACCCCATGG
8B	F, F	*RIMS2* transcript variant 1-3, 5-30	TTAAAATTTACTTTCTTCCAG	TAAGAATTCACTTTCTTCCTC
9	F, R	*SDR42E1*	GAAGCTTGGAGCCCAGGAAC	TGTCATCTTTGGAGGTCAAG
13	F, R	*PEG10* 4 transcript variants	CGCCTGGTTGACCCTCACAT	TCCCGCAGGGCTCTCCATT
	F, R	*COL6A6*	TTTTTGGTGGACAGCTCTGAT	CCTCACAGCTTACTCCATT
	R, R	*POLR1A*	CCCCACTGTCCTCTCCACT	CTCCACTGCAACCTCCATC

	R, R	*LRRC28* transcript variant 1-4, 6-8	ACTTTCTGCCTTCTCCATT	CCTCAATGAGCTCTCCATG
14	F, R	*LRRK1*	CTGAACTTGGACCTGTTG	TGCCCTTTGAAATGTTGT
	F, R	*MCHR1*	AAACGCTTGGTCCTGTCG	CTCCCGGTGGCGTGTTGT
	F, R	*SON*	CTGGTACTAGTCCTGTTG	ATCCAGTTCTTGTGTTGA
	F, R	*CDRT1*	GTGATCAGAGTCCTGTTT	TGCCACTTGTTGTGCTGT
	F, R	*AGPAT3* transcript variant 1-5	CTGAGCTTAATTCAGTTG	CTTACATTGTTGTTTTGT
	F, R	*MET* transcript variant 1,2,4	AAGTGTGTAGTCCTGTTG	AACACTTTTGTGTGTTGT
	F, F	*EPPK1*	CTGAACTTGATCCTGTTG	ATGAGCTCCTTCCTGCTG
	R, R	*C8orf34* transcript variant 2,7	TTAATTTTATTGTGTTGT	CACTCTTTGTTCTGTTTT
	R, R	*BNC2* transcript variant 1-3	ATGGCTTTGTTGTGTTGC	TTCTCTGTGTTGTGCTGT
	R, R	*PEAK1*	CTAACTTTTTTGTGTTTT	CACCCTTTGGTTTGCTGT
	R, R	*CHRM3* transcript variant 1-10	ATAACTTTGGAGTGTTGT	TTTTCTTTGGTGTGTTGT
	R, R	*THOC2*	TTCCCTTTCTTGAGTTGA	TTACGATGGTTGTGTTGT
15	F, F	*MED13L*	AATAATTGATATGATTTTG	GGTAAGAGTTCTGATTTCT

* Number of primer in Figure 2.

\# Nucleic acid bases in red color were different with that of original primers.

Table 1: **Non-specific primer and the unintended target**

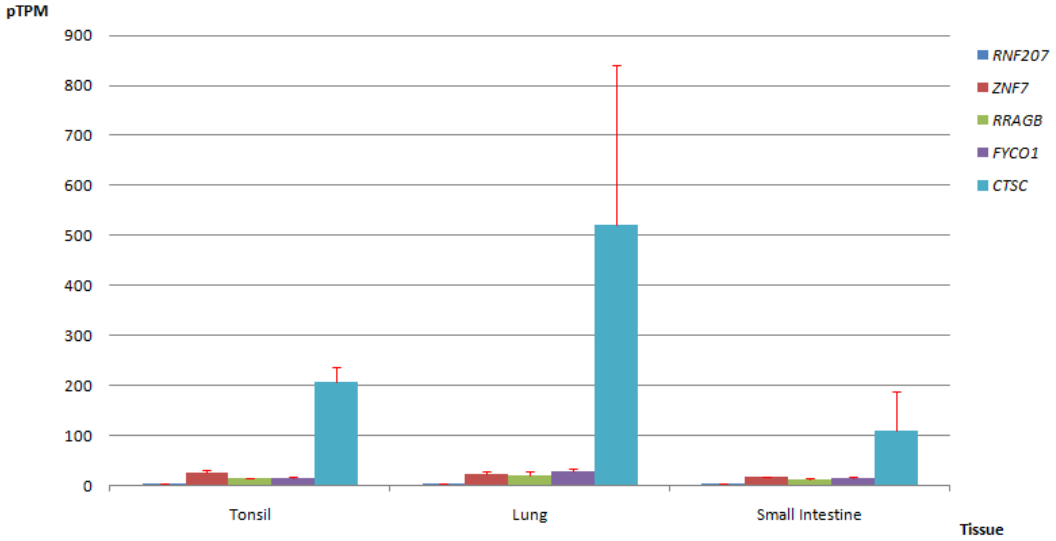

Figure 3: **Expression of some target human mRNAs in different normal tissue.**

Not Recommended	Primer 1	Primer 6	Primer 9	Primer 13	Primer 14	Primer 15	
	Primer 7A, 7B		Primer 8A, 8B				
Recommended	Primer 2	Primer 3	Primer 4	Primer 5	Primer 10	Primer 11	Primer 12

Table 2: **Suggestion on the RT-PCR primers for detection of SARA-CoV-2 virus**