Genomic Variations in SARS-CoV-2 Strains at the Target Sequences of Nucleic Acid Amplification Tests

Canhui Cao

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Ruidi Yu

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Shaoqing Zeng

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Dan Liu

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Wenjian Gong

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Ruyuan Li

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Siyuan Wang

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Yuan Yuan

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Jianhua Chi

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Jiahao Liu

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Yang Yu

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Xiaofei Jiao

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Guangyao Cai

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Ning Jin

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Fei Ye

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Qinglei Gao (✉ qingleigao@hotmail.com)

Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Keywords: SARS-CoV-2, nucleic acid tests, genome variations, COVID-19, primers of PCR

Posted Date: October 26th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-70133/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Archives of Medical Science on March 21st, 2021. See the published version at <https://doi.org/10.5114/aoms/133120>.

Abstract

Background

Nucleic acid amplification is the main method used to detect infections of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). However, the false-negative rate of nucleic acid tests cannot be ignored.

Methods

Herein, we demonstrated genomic variations at the target sequences for the tests and the geographical distribution of the variations across countries by analyzing the whole-genome sequencing data of SARS-CoV-2 strains from the 2019 Novel Coronavirus Resource (2019nCoV) database.

Results

Among the 21 pairs of primer sequences in regions ORF1ab, S, E, and N, the total length of primer and probe target sequences was 938 bp, with 131 (13.97%) variant loci in 2415 (38.96%) isolates. Primer targets in the N region contained the most variations that were distributed among the most isolates, and the E region contained the least. Single nucleotide polymorphisms were the most frequent variation, with C to T transitions being detected in the most variant loci. G to A transitions and G to C transversions were the most common and had the highest isolate density. Genomic variations at the three mutation sites N: 28881, N: 28882, and N: 28883 were the most commonly detected, including in 608 SARS-CoV-2 strains from 33 countries, especially in the United Kingdom, Portugal, and Belgium.

Conclusions

Our work comprehensively analyzed genomic variations on the target sequences of the nucleic acid amplification tests, offering evidence to optimize primer and probe target sequence selection, thereby improving the performance of the SARS-CoV-2 diagnostic test.

Background

The coronavirus disease 2019 (COVID-19) outbreak has resulted in a Public Health Emergency of International Concern since 30 January 2020, and has been classified as an ongoing pandemic since 11 March 2020 [1, 2]. There have been 5,000,561 confirmed infected cases and 328,172 deaths by as of May 21, 2020, involving 229 countries and territories worldwide [3]. COVID-19 is an infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus [4], which transmits rapidly via human respiratory droplets, direct contact, and aerosol transmission [5]. The widespread popularity and rapid spread of COVID-19 poses an urgent need for fast and accurate diagnosis methods.

Currently, the diagnosis of COVID-19 relies on positive pathogenetic, epidemiological exposure history, CT imaging features, and serological characteristics [6, 7], as well as symptoms that are atypical and might be

confused with influenza, common cold, and pneumonia caused by other pathogens [8–11]. Positive pathogenicity is commonly confirmed by nucleic acid testing via reverse transcription polymerase chain reaction (RT-PCR), virus isolation and sequencing, or specific antibody detection. Currently, nucleic acid testing in various bodily fluids is the primary method of diagnosing COVID-19 [9, 12]. This is similar to the diagnostic approach for SARS and MERS, the two other coronavirus diseases [13–15]. Hence, the sensitivity and specificity of nucleic acid testing are crucial. Although it is highly specific, the sensitivity is not satisfactory [16]. There are various factors that may interfere with test sensitivity. The sampling days from onset of symptoms, different sites used to obtain a specimen, insufficient viral material in the specimen, laboratory error during sampling, and restrictions on sample transportation could all influence the nucleic testing and cause false-negative results [17–20]. However, few studies mention false negatives due to variations in primer target sequences.

The RT-PCR kits have been designed to detect SARS-CoV-2 genetically, involving the reverse transcription of SARS-CoV-2 RNA into complementary DNA (cDNA) strands, followed by amplification of specific regions of the cDNA with paired primers and probes [21]. Among the SARS-related viral genomes, there are three regions with conserved sequences: 1) the RdRP gene (RNA-dependent RNA polymerase gene) in the open reading frame ORF1ab region, 2) the E gene, and 3) the N gene [22]. Both the RdRP and E genes had high analytical sensitivity for detection, whereas the N gene provided poorer analytical sensitivity [9, 22]. Thus, some assays were designed as a two-target system, where one primer set universally detected coronaviruses, and a second set only detected SARS-CoV-2 [9]. As of 21 May 2020, 315 commercialized molecular assays had been designed to detect SARS-CoV-2 genetically, worldwide, according to information provided by The Foundation for Innovative New Diagnostics, some PCR protocols had been shared [23, 24]. The sensitivity of nucleic acid detection kits relies on the binding of primers and probes to the virus genome [9, 25]. High variation frequency, especially the variation near the 3' end, might influence the primers or probes binding to the virus genome [26]. Thus, variations in the target sequence and how these variations affect the accuracy of the tests need to be elucidated.

SARS-CoV-2 is a positive-sense single-stranded RNA virus, with a genome that has typical betacoronavirus organization: a 5' untranslated region (UTR), replicase complex (orf1ab), S (spike) gene, E (envelope) gene, M (membrane) gene, N (nucleocapsid) gene, 3' UTR, and several unidentified nonstructural open reading frames [27–30]. The RNA virus's mutation rate is dramatically high, and this high rate is correlated with virulence modulation and evolvability for viral adaptation [31]. It has been reported that the most frequent variations are found within open-reading frame (ORF) regions [32], however, some studies suggest that overall variation in ORF regions is low [33]. Therefore, the variation frequency in the aforementioned regions of the genome needs further study, especially in the primer and probe target sequence.

Although most primer and probe sets were designed to focus on specific and conserved sequences, there were still some variations in the target sequences of the nucleic acid test assays of SARS-CoV-2 strains. A number of studies compared the sensitivity and efficiency of different SARS-CoV-2 detection assays [34, 35]. But few people pay attention to the variations and geographic distribution that may affect the PCR performance of detection kits used on a large scale in the population. Herein, we comprehensively analyzed genomic variations in the target sequences of the nucleic acid amplification tests and their geographic

distribution. Our results are expected to provide evidence for optimizing the selection of detection kits used on a large scale in the population, thereby improving the diagnostic accuracy of the tests.

Methods

2019 Novel Coronavirus Resource

We used the 2019 Novel Coronavirus Resource [36, 37], constructed by the Chinese Academy of Sciences. It integrates genomic and proteomic sequences as well as their metadata from the Global Initiative on Sharing All Influenza Data, National Center for Biotechnology Information, China National GeneBank, National Microbiology Data Center, and China National Center for Bioinformation/National Genomics Data Center. The 2019 nCoV-2 offers visualization functionalities for genomic variation analysis results based on all SARS-CoV-2 strains collected.

Source of primers and probes used in nucleic acid detection kits.

The WHO offered molecular assay protocols for SARS-CoV-2 that have been shared (WHO, 2020), including seven institutes: the Chinese Center for Disease Control and Prevention (China CDC); Institute Pasteur, Paris, France (IP France); Centers for Disease Control and Prevention, the United States (US CDC); National Institute of Infectious Diseases, Japan (NIID Japan); Charité, Germany; School of Public Health, Hong Kong University (HKU); and the National Institute of Health, Thailand (NIH Thailand). Meanwhile, 21 pairs of primers and probes from the above protocols were mapped to the reference genome.

Mapping of genomic variations

The sequences of primers and probes were mapped to the SARS-CoV-2 isolate Wuhan-Hu-1, complete genome (NCBI Reference Sequence: NC_045512.2) [36]. The positions of the primers and probes used in the nucleic acid testing kits were identified. Then, the sequences of the primers and probes were compared with the variant loci offered by 2019nCoV-2, with the variant loci and corresponding isolates downloaded from the database.

Variation curve

Virus isolates with variations at each site divided by country at each time point were counted. The curve shows the trends of isolates from different countries over time at some variant loci.

Variation density and isolate density calculation

In ORF1ab, S, E, and R regions of the SARS-CoV-2 genome, the length of primers and probes, and the number of variant loci covered by the primers and probes in each region were calculated first. The variation density of a region was the number of variant loci divided by the length of primers and probes. In addition, the number of isolates presenting variations in the primer and probe target sequence was calculated. Isolate density was the number of isolates divided by the length of the target sequence.

The variance of time and regional variations

First the frequency of population occurrence at each variant loci over time or country was identified. Time variance was the variance based on the frequency of population occurrence at each time point, to assess the dispersion of changes at that site. Taking the country as a unit, we calculated regional variance based on the frequency of population occurrence in each country, to evaluate the dispersion of variations at that location.

Statistical analysis

The majority of statistical analyses were performed using SPSS software package version 22. The chi-square test was used to compare the differences in counting data between groups. Variance was used to evaluate the degree of dispersion as described above. A two-sided P value less than 0.005 was considered statistically significant.

Results

Variation landscape of primer and probe target sequences in the SARS-CoV-2 strains

Twenty-one pairs of primers and probes obtained from seven institutes worldwide were analyzed. The primers and probes were mainly focused on the replicase complex (ORF1ab), spike (S), envelope (E), and nucleocapsid (N) regions. The total length of primer and probe target sequence was 938 bp, with 131 (variation density = 13.97%) variant loci. In addition, 2415 (38.96%, isolate density = 2.57 per locus) of the 6198 SARS-CoV-2 strains from the 2019 Novel Coronavirus Resource (2019nCoV) database had variations in the target sequence (Table 1).

Table 1
Polymerase Chain Reaction kits for SARS-CoV-2 and Variations

No.	Region	Sequence	Variation Sites	Variation	Isolates	Institute	
1	ORF1ab	F	TTCGGATGCTCGAACTGCACC	3	SNP	37	NIID Japan
		R	CTTTACCAGCACGTGCTAGAAGG	3	SNP	24	
2	ORF1ab	F	CTCGAACTGCACCTCATGG	3	Deletion, SNP	5	NIID Japan
		R	CAGAAGTTGTTATCGACATAGC	5	SNP	17	
3	ORF1ab	F	ACCTCATGGTCATGTTATGG	7	SNP, Deletion	93	NIID Japan
		R	GACATAGCGAGTGTATGCC	1	SNP	1	
4	ORF1ab	F	ATGAGCTTAGTCCTGTTG	5	SNP	10	NIID Japan
		R	CTCCCTTTGTTGTGTTGT	0	SNP	0	
		P	AGATGTCTTGTGCTGCCGGTA	1	SNP	2	
5	ORF1ab	F	CCCTGTGGGTTTTACACTTAA	1	SNP	1	China CDC
		R	ACGATTGTGCATCAGCTGA	1	SNP	1	
		P	CCGTCTGCGGTATGTGGAAAGGTTATGG	3	SNP	53	
6	ORF1ab	F	GGTAACTGGTATGATTTTCG	0	SNP	0	IP France
		R	CTGGTCAAGGTTAATATAGG	1	SNP	2	
		P	TCATACAAACCACGCCAGG	2	SNP	3	
7	ORF1ab	F	GTGARATGGTCATGTGTGGCGG	2	SNP	2	Charité, Germany
		R	CARATGTTAAASACACTATTAGCATA	0	SNP	0	
		P	CCAGGTGGWACRTCATCMGGTGATGC	2	SNP	3	
		P	CAGGTGGAACCTCATCAGGAGATGC	2	SNP	3	
8	ORF1ab	F	GTGARATGGTCATGTGTGGCGG	2	SNP	2	Charité, Germany
		R	CARATGTTAAASACACTATTAGCATA	0	SNP	0	
		P	CAGGTGGAACCTCATCAGGAGATGC	2	SNP	3	

China CDC: the Chinese Center for Disease Control and Prevention; IP France: Institute Pasteur, Paris, France; US CDC: Centers for Disease Control and Prevention, the United States; NIID Japan: National Institute of Infectious Diseases, Japan; HKU: School of Public Health, Hong Kong University; NIH Thailand: the National Institute of Health, Thailand.

No.	Region	Sequence	Variation Sites	Variation	Isolates	Institute	
9	ORF1ab	F	TGGGGYTTTACRGGTAACCT	1	SNP	11	HKU
		R	AACRCGCTTAACAAAGCACTC	4	SNP	8	
		P	TAGTTGTGATGCWATCATGACTAG	1	SNP	1	
10	S	F	TTGGCAAAATTCAAGACTCACTTT	5	SNP	19	NIID Japan
		R	TGTGGTTCATAAAAATTCCTTTGTG	1	SNP	1	
11	S	F	TCAAGACTCACTTTCTTCCAC	10	SNP	31	NIID Japan
		R	ATTTGAAACAAAGACACCTTCAC	1	SNP	2	
12	S	F	AAGACTCACTTTCTTCCACAG	9	SNP	30	NIID Japan
		R	CAAAGACACCTTCACGAGG	1	SNP	2	
13	E	F	ACAGGTACGTTAATAGTTAATAGCGT	1	SNP	1	Charité, Germany
		R	ATATTGCAGCAGTACGCACACA	1	SNP	1	
		P	ACACTAGCCATCCTTACTGCGCTTCG	3	SNP	6	
14	E	F	ACAGGTACGTTAATAGTTAATAGCGT	1	SNP	1	IP France
		R	ATATTGCAGCAGTACGCACACA	1	SNP	1	
		P	ACACTAGCCATCCTTACTGCGCTTCG	3	SNP	6	
15	N	F	GACCCCAAAATCAGCGAAAT	5	SNP	7	US CDC
		R	TCTGGTTACTGCCAGTTGAATCTG	3	SNP	25	
		P	ACCCCGCATTACGTTTGGTGGACC	3	SNP	42	
16	N	F	CGTTTGGTGGACCCTCAGAT	3	SNP	10	NIH Thailand
		R	CCCCACTGCGTTCTCCATT	6	SNP	8	

China CDC: the Chinese Center for Disease Control and Prevention; IP France: Institute Pasteur, Paris, France; US CDC: Centers for Disease Control and Prevention, the United States; NIID Japan: National Institute of Infectious Diseases, Japan; HKU: School of Public Health, Hong Kong University; NIH Thailand: the National Institute of Health, Thailand.

No.	Region	Sequence	Variation Sites	Variation	Isolates	Institute
		P CAACTGGCAGTAACCA	1	SNP	14	
17	N	F GGGAGCCTTGAATACACCAAAA	2	SNP	90	US CDC
		R TGTAGCACGATTGCAGCATTG	4	SNP	7	
		P AYCACATTGGCACCCGCAATCCTG	8	SNP	13	
18	N	F GGGGAACTTCTCCTGCTAGAAT	8	SNP	1838	China CDC
		R CAGACATTTTGCTCTCAAGCTG	7	SNP	9	
		P TTGCTGCTGCTTGACAGATT	0	SNP	0	
19	N	F AAATTTTGGGGACCAGGAAC	2	SNP	14	NIID Japan
		R TGGCAGCTGTGTAGGTCAAC	3	SNP	3	
		P ATGTCGCGCATTGGCATGGA	6	SNP	8	
20	N	F TAATCAGACAAGGAACTGATTA	2	SNP	6	HKU
		R CGAAGGTGTGACTTCCATG	7	SNP	20	
		P CAAATTGTGCAATTTGCGG	3	SNP	5	
21	N	F TTACAAACATTGGCCGCAAA	3	SNP	7	US CDC
		R GCGCGACATTCCGAAGAA	0	SNP	0	
		P ACAATTTGCCCCAGCGCTTCAG	3	SNP	6	
China CDC: the Chinese Center for Disease Control and Prevention; IP France: Institute Pasteur, Paris, France; US CDC: Centers for Disease Control and Prevention, the United States; NIID Japan: National Institute of Infectious Diseases, Japan; HKU: School of Public Health, Hong Kong University; NIH Thailand: the National Institute of Health, Thailand.						

Among the four regions, most primers and probes were located in the ORF1ab region. They covered 484–18909, including 42 variant loci and 271 SARS-CoV-2 strains, with variations at these sites. The most common variant loci in the ORF1ab region were 514 (87 virus strains, 32.10% of the 271 strains), followed by 13402, and 490 (Fig. 1A). As for the S region, there were only three pairs of primers covered from 24354 to 24900. There were 12 variant loci and 34 SARS-CoV-2 strains, with 24368 (14 virus strains, 41.18% of the 34 strains) being the most common variant loci (Fig. 1B). Interestingly, there were two pairs of primers and probes provided by two different institutes located in the E region, and the two pairs were the same. It covered 26269 to 26381. Barely five variant loci and eight SARS-CoV-2 strains were discovered. The most common variant loci were located at 26340, with four virus strains that had variation at this site, occupying 50% of the eight virus strains (Fig. 1C). Although the N region only had seven pairs of primers and probes covered from 28287 to 29282, it had the most variations (72 loci), and involved virus strains (2102, 5.8 per locus, 33.91% of the 6198 SARS-CoV-2 virus strains from the database) among the four regions. The most common variant loci in the N region were 28881 with 609 virus strains (28.97% of the 2102 virus strains), 28882 with 608 virus

strains (28.92%), and 28883 with 608 virus strains (28.92%). These three variant loci had the most related isolates among the four regions (Fig. 1D).

The length of primer and probe target sequence in the ORF1ab region was 415 bp, which is the longest among the four regions, with 42 variant loci (variation density = 0.10). The N region, the most mutative region, had the most variant loci (72 loci) and the highest variation density (0.20), with a 365 bp target sequence length. The primer or probe target sequences in the S region were only 84 bp in length. However, only 12 variant loci were identified with high variation density (0.14). The target sequence of the E region was the shortest (74 bp) and the least mutative, with the least variant loci (5 loci) and the lowest variation density (0.07, $p = 0.024$) (Figs. 1E and 1F).

Among these four regions, there were 2415 (isolate density = 2.57 per locus) virus strains with variant loci among primer or probe target sequences. In ORF1ab, 271 (isolate density = 0.65 per locus) isolates had variant loci involved in the primers or probes, and 34 isolates (isolate density = 0.40 per locus) in the S region. Significantly, the least isolates and isolate density were demonstrated in the E region, with 8 strains and 0.11 isolates per locus. Conversely, the N region had the most isolates and the highest density, with 2102 isolates and 5.78 isolates per locus (Fig. 1G).

Variation analysis of SARS-CoV-2 strains in the target sequence of nucleic acid amplification tests.

To further illustrate the variations, we investigated the specific variation types. Among the variant loci covered by paired primers and probes, the most common variations were single nucleotide polymorphisms (SNPs), accounting for 96% of all variations observed. Only 3% of the variations were deletions, and 1% of the variations had SNPs and deletions (Fig. 2A). Regarding the involved SNPs, 58% were transitions and 35% were transversions. A total of 37% of SNP variant loci were C to T transitions, the most common type of SNP, followed by G to T transversions and T to C transitions (Fig. 2B).

In the 6198 SARS-CoV-2 virus strains, the G to A transition involved the most isolates (1218) with the highest isolate density (121.8 per locus) in all types of SNPs, followed by G to C transversion and T to C transition. T to G transversion and C to T transitions showed the opposite trends. The former were found in fewer virus strains, but at a higher density (53 strains, 17.67 strains per site), while the latter were found in more virus strains, but at a lower density (172 strains, 3.25 strains per site) (Fig. 2C).

The C to T transition was the most common SNP type, accounting for 42% of the SNPs in the ORF1ab region, 38% in the S region, 60% in the E region, and 39% in the N region. The ORF1ab region contained 12 SNPs. Following the C to T transition, the common SNPs were G to T transversion, T to C transition, and A to G transition. There were eight SNP types in the S region. In addition to the C to T transition and A to G transition (15%), the other six types uniformly accounted for 7%-8% separately. The least SNP types were in the E region, with only three types of SNPs; C to T transition, T to A transversion (20%), and T to C transition (20%). In the N region, there were 16 types of SNPs, the most abundant. Besides the C to T transition, the most common type of SNPs were G to T transversion and T to C transition (Fig. 2D).

Variation analysis of N:28881, N: 28882, and N:28883 loci.

To investigate the variations at the most common variant loci, we analyzed the occurrence of related primers, isolates, and variations over time and within different geographical regions. The variant loci that involved the most isolates among the primers or probes covered region were N: 28881 (609), N: 28882 (608), and N:28883 (608). The primers covering these three variant loci were, F: GGGGAACCTTCTCCTGCTAGAAT, and R: CAGACATTTTGCTCTCAAGCTG, recommended by China CDC. Besides 28881–28883, some variant loci had many corresponding mutant virus strains, such as 28887, 28896, and 28969 (Fig. 3A). The virus strains with variations at 28881, 28882, and 28883 had 608 overlapping strains. Only one strain was unique, shown at 28881 (Fig. 3B). The variation tendencies of the isolates involved in the above three variant loci were almost the same. At the end of February, major mutant isolates were obtained from Switzerland, at the end of March and the beginning of April, the United Kingdom (UK) provided a large portion of the mutant isolates, followed by Belgium and Portugal. The time variances of the tree sites were the same (Fig. 3C). In detail, the unique strain, at 28881, came from Australia, with 45 isolates at this site. In other countries, except Australia, the trend was the same at 28881, 28882, and 28883. Most isolates came from the UK, with 140 mutant isolates at these three sites separately. The following were from Belgium, Netherlands, Portugal, and Iceland. The variation frequencies of the three sites were also the same, showing different geographical preferences with isolate distribution. The highest variation frequency was from Mexico (1.00), whose mutant isolates were extremely limited. Greece and Portugal also had high variation frequencies at the three sites. The regional variances of the three sites equaled 0.0424 (Fig. 3D).

Isolate analysis of common variant loci.

Geographical distributions of several other common variant loci that involved many isolates were also demonstrated. At ORF1ab:490, there were 21 (60.00%) mutant isolates from Australia, and 11 (31.43%) from the United States. The highest variation frequencies were in New Zealand and Georgia (0.125, regional variance = 0.0007) (Fig. 4A). And the forward primer of the first pair of primers and probes from Japan National Institute of Infectious Diseases targeted this variant locus (Fig. 1A and Table 1). At ORF1ab:514, most mutant isolates were from the Netherlands (66 isolates, 75.86%) with a high variation frequency (0.12). The highest variation frequencies were observed in Slovakia (Fig. 4B). And this variant locus was targeted by the forward primer of No.3 pair of primers and probes from Japan National Institute of Infectious Diseases (Fig. 1A and Table 1). Prominently, 96.08% (49 isolates) mutant isolates at ORF1ab:13402 were from Belgium with the second-highest variation frequency. The highest variation frequency was from Denmark (0.29, regional variance = 0.0020) (Fig. 4C). As for N:28311, targeted by No.5 pair of primers and probes provided by China CDC (Fig. 1D and Table 1), 20 (57.14%), isolates were from Australia, which was the highest. However, the variation frequency of strains from Australia was not high. The highest variation frequencies occurred in Saudi Arabia, South Korea, and Malaysia (regional variance = 0.0053) (Fig. 4D). The probe of No.15 pair of primers and probes provided by US CDC targeted this variant locus (Fig. 1D and Table 1). The distribution of mutant isolates at N:28688 was decentralized. It was targeted by the forward primer of No.17 pair of primers and probes, which was recommended by US CDC (Fig. 1D and Table 1). There were 31 (36.47%) isolates from Australia, 18 (21.18%) from China, and 11 (12.94%) from India. High variation frequencies were observed in Turkey, India, and Kuwait (regional variance = 0.0273) (Fig. 4E).

Discussion

In the present study, our fundamental purpose is to reveal the variations in the target sequence of primers and probes in COVID19 detection kits used on a large scale in the population. We compared the primer and probe sequences used to detect SARS-CoV-2 infection to those exported from the 2019nCoV database to gain important insights into virus variations in primer and probe target regions.

By integrating 21 pairs of primer sequences in regions ORF1ab, S, E, and N from 6198 SARS-CoV-2 strains, we found 31(13.97%) variant loci, involving 2415 (38.96%) isolates, especially in the N region, which had the highest variation density and the highest isolate density. The most variant loci and the most isolates were also located in N region (Figs. 1F and 1G). The high variation frequency and included isolate density implied that the N region might not be suitable for primer design, which might elevate the false-negative rate. Therefore, merely nucleic acid testing alone is not enough to suppose a negative case, it should be combined with other tests such as antibody detection [9]. In contrast, the lowest variation density and the lowest isolate density were observed in the E region. In addition, it had the least variant loci and the least isolates. Thus, the E region might be the most conserved region, suitable for primer design [26]. This is consistent with the findings of Buddhisha *et al.* and Corman *et al.* that the RdRP and E genes had high analytical sensitivity for detection, whereas the N gene provided poorer analytical sensitivity [9, 22, 38]. However, only two identical pairs of primers and probes were from the E region, which might cause an incomplete understanding of this region in the present study.

The variations involved in the primers and probes were mainly SNPs, which were abundant in the SARS-CoV-2 genome [39–41]. Recent research has demonstrated that SNPs in SARS-CoV-2 are capable of substantially changing their pathogenicity [41]. Importantly, a small number of SNPs in the primer target sequence might not cause primer binding failure [26]. The C to T transition was the most common type of SNP in the primer and probe target sequence. N:28881, N:28882, and N:28883 were the most influential variant loci, involving 608 or 609 viral isolates. It would be better to avoid designing primers or probes whose target sequences contain these three sites. As only the No. 18 primer pair targeted this sequence in the 21 pairs, verification with different nucleic acid detection kits or ddPCR could elevate sensitivity [42, 43]. As viruses evolve during outbreaks, SNP in primer or probe binding regions could alter the sensitivity of PCR assays [35]. We analyzed the variations in the target sequence of primers and probes in COVID19 detection kits used on a large scale in the population and their geographic distribution. Thereby, providing a reference for optimizing the selection of detection kits used on a large scale in the population, improving the diagnostic accuracy of the tests. Sapoval *et al.* figured out that each probe/primer sequence that was available on the WHO website contained 2.529 iSNV and/or 2.477 SNPs, suggesting the potential for a drop in the sensitivity of the affected probes and primers. They also noted that the potential impact of iSNVs on primer and probe binding should not be overlooked given the possibility of iSNVs establishing as SNPs [44]. And there were already some researches about how would SNPs in SARS-CoV-2 affect the biological function of the virus [45–49].

The 17th pair of primers and probes is used to be recommended by the US CDC, involving 110 isolates, the second most among 21 pairs of primers and probes. N: 28688 is the most common mutation sites among the above sequences, and at this site, the United States has the third most isolates since May 12th, 2020. And this pair of primers and probes was removed in the US CDC recommended protocol updated in July [50]. Whereas at other common mutation sites like N:28881- N:28883, ORF1ab:490, ORF1ab: 514, ORF1ab: 13402 and N: 28311, the region where mutant strains often appear does not overlap with the region where this kit is used

(Figure S1). And in the selection and application of detection kits in the future, the 18th pair of primers and probes provided by China CDC would better not be used in the United Kingdom, Belgium, Netherlands and Portugal (Fig. 3). Japan National Institute of Infectious Diseases recommended No.1 and No.3 pairs of primers and probes. The first pair should be avoided in the US and the third one should be avoided in Netherlands, the UK and the US (Fig. 4A, 4B, S1A and S1B). It would be better not to use the 5th pair in Belgium (Fig. 4C). Singapore, India and Australia would better avoid using the 15th pair of primers and probes (Fig. 4D and S1C).

There were some limitations to this study. Initially, the sequences of primers and probes used in the nucleic acid detection kits were provided by the WHO. However, not all kits used worldwide were available, some of which were not open access. Although the virus strain data is updated frequently, the strains we used merely presented the situation before our research ended. In addition, it has been reported that variations in primer binding sites could affect PCR efficiency [51, 52]. Thus, we are supposed to further validate the relationship between variations and PCR efficiency in the future. In addition, how SNPs influence virus detection and virus pathogenicity needs to be further investigated.

Conclusions

By integrating 21 pairs of primer target sequences in regions ORF1ab, S, E, and N from 6198 SARS-CoV-2 strains, the total length of primer and probe target sequences was 938 bp, with 131 (13.97%) variant loci, involving 2415 (38.96%) isolates. The E region had the highest priority in target sequence selection, with N region the least priority. SNP was the most frequent variation with C to T transitions involving the most variant loci. G to A transitions and G to C transversions were the most common with the highest isolate density. Genomic variations at the three mutation sites N: 28881, N: 28882, and N: 28883 were the most commonly detected variations, including 608 SARS-CoV-2 strains from 33 countries, especially in the UK, Portugal, and Belgium. Herein, we comprehensively analyzed the genomic variations on the target sequences of the nucleic acid amplification tests, offering evidence to the optimization of primer and probe target sequence selection, providing implications for the identification of infected patients via the tests.

Abbreviations

SARS-CoV-2

severe acute respiratory syndrome coronavirus 2

2019nCoV-R

the 2019 Novel Coronavirus Resource

COVID-19

the coronavirus disease 2019

RT-PCR

reverse transcription polymerase chain reaction

cDNA

complementary DNA

RdRP

RNA-dependent RNA polymerase
UTR
untranslated region
S
spike
E
envelope
M
membrane
N
nucleocapsid
ORF
open-reading frame
China CDC
the Chinese Center for Disease Control and Prevention
IP France
Institute Pasteur, Paris, France
US CDC
Centers for Disease Control and Prevention, the United States
NIID Japan
National Institute of Infectious Diseases, Japan
HKU
Charité, Germany; School of Public Health, Hong Kong University
NIH Thailand
the National Institute of Health, Thailand
SNP
single nucleotide polymorphism
UK
the United Kingdom

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and analysed during the current study are available in the 2019nCoV, <https://bigd.big.ac.cn/ncov>.

Competing interests

The authors declare that they have no competing interests

Funding

The study was supported by the National Science and Technology Major Sub-Project (2018ZX10301402-002), the Technical Innovation Special Project of Hubei Province(2018ACA138), the National Natural Science Foundation of China (81772787, 81873452, and 81974405), and the Fundamental Research Funds for the Central Universities (2019kfyXMBZ024).

Authors' contributions

CC and RY contributed equally to this work. They analyzed and interpreted the data, and wrote the original draft. QG and FY conceptualized and designed the study, and revised the manuscript. SZ and DL advised on the conception and design of the study. WG, RL, SW, YY and JC helped with the collection of resources. JL, YY, XJ, GC and NJ helped with editing assistance. All authors vouch for the respective data and analysis, revised, approved the final version, and agreed to publish the manuscript.

Acknowledgments

The authors thank all health-care workers and people involved in fighting against COVID-19. We would like to acknowledge the platform 2019nCoV provided by the Chinese Academy of Sciences. We thank the seven institutes for the disclosure of nucleic acid test kit contents. We appreciate the WHO providing SARS-CoV-2 relevant data.

References

1. **WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020** [https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020]
2. Hui DS, E IA, Madani TA, Ntoumi F, Kock R, Dar O, Ippolito G, McHugh TD, Memish ZA, Drosten C *et al*: **The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China.** *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 2020, **91**:264-266.
3. **Template:COVID-19 pandemic data** [https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data]
4. **The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2.** *Nature microbiology* 2020, **5**(4):536-544.
5. Liu Y, Ning Z: **Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals.** 2020.
6. **New Coronavirus Pneumonia Diagnosis and Treatment Guidelines (Trial Version 7)** [http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml]
7. Cheng MP, Papenburg J, Desjardins M, Kanjilal S, Quach C, Libman M, Dittrich S, Yansouni CP: **Diagnostic Testing for Severe Acute Respiratory Syndrome–Related Coronavirus-2.** *Annals of Internal Medicine*

2020.

8. Bordi L, Nicastrì E, Scorzolini L, Di Caro A, Capobianchi MR, Castilletti C, Lalle E, On Behalf Of Inmi Covid-Study G, Collaborating C: **Differential diagnosis of illness in patients under investigation for the novel coronavirus (SARS-CoV-2), Italy, February 2020.** *Euro Surveill* 2020, **25**(8):2000170.
9. Udugama B, Kadhiresan P, Kozłowski HN, Malekjahani A, Osborne M, Li VYC, Chen H, Mubareka S, Gubbay JB, Chan WCW: **Diagnosing COVID-19: The Disease and Tools for Detection.** *ACS Nano* 2020.
10. Guan WJ, Liang WH, Zhao Y, Liang HR, Chen ZS, Li YM, Liu XQ, Chen RC, Tang CL, Wang T *et al*: **Comorbidity and its impact on 1590 patients with Covid-19 in China: A Nationwide Analysis.** *The European respiratory journal* 2020.
11. Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, Liu L, Shan H, Lei C-l, Hui DSC *et al*: **Clinical Characteristics of Coronavirus Disease 2019 in China.** *New England Journal of Medicine* 2020.
12. Beeching NJ, Fletcher TE, Beadsworth MJB: **Covid-19: testing times.** *BMJ* 2020, **369**:m1403.
13. Al Johani S, Hajeer AH: **MERS-CoV diagnosis: An update.** *Journal of infection and public health* 2016, **9**(3):216-219.
14. Zumla A, Hui DS, Perlman S: **Middle East respiratory syndrome.** *Lancet (London, England)* 2015, **386**(9997):995-1007.
15. Hui DS, Wong PC, Wang C: **SARS: clinical features and diagnosis.** *Respirology (Carlton, Vic)* 2003, **8 Suppl**(Suppl 1):S20-24.
16. Xie X, Zhong Z: **Chest CT for Typical 2019-nCoV Pneumonia: Relationship to Negative RT-PCR Testing.** 2020:200343.
17. Li Y, Yao L, Li J, Chen L, Song Y, Cai Z, Yang C: **Stability issues of RT-PCR testing of SARS-CoV-2 for hospitalized patients clinically diagnosed with COVID-19.** *Journal of medical virology.*
18. Wenling W, Yanli X, Ruqin G, Roujian L, Kai H, Guizhen W, Wenjie T: **Detection of SARS-CoV-2 in Different Types of Clinical Specimens.** *JAMA* 2020, **323**(18):-.
19. Pan Y, Long L, Zhang D, Yan T, Cui S, Yang P, Wang Q, Ren S: **Potential false-negative nucleic acid testing results for Severe Acute Respiratory Syndrome Coronavirus 2 from thermal inactivation of samples with low viral loads.** *Clin Chem* 2020.
20. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, Yu J, Kang M, Song Y, Xia J *et al*: **SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients.** *N Engl J Med* 2020, **382**(12):1177-1179.
21. Chu DKW, Pan Y, Cheng SMS, Hui KPY, Krishnan P, Liu Y, Ng DYM, Wan CKC, Yang P, Wang Q *et al*: **Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia.** *Clinical Chemistry* 2020, **66**(4):549-555.
22. Corman VB, T.; Brünink, S.; Zambon, M. : **Diagnostic Detection of Wuhan Coronavirus 2019 by Real-Time RT-PCR.** In: *World Health Organization.* Geneva; 2020.
23. **Summary table of available protocols in this document** [https://www.who.int/docs/default-source/coronaviruse/whoinhouseassays.pdf?sfvrsn=de3a76aa_2]
24. **SARS-COV-2 DIAGNOSTIC PIPELINE** [<https://www.finddx.org/covid-19/pipeline/>]
25. Freeman WM, Walker SJ, Vrana KE: **Quantitative RT-PCR: Pitfalls and Potential.** *BioTechniques* 1999, **26**(1):112-125.

26. Dieffenbach CW, Lowe TM, Dveksler GS: **General concepts for PCR primer design.** *PCR methods and applications* 1993, **3**(3):S30-37.
27. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, Haagmans BL, Lauber C, Leontovich AM, Neuman BW *et al.*: **The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2.** *Nature microbiology* 2020, **5**(4):536-544.
28. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R *et al.*: **A Novel Coronavirus from Patients with Pneumonia in China, 2019.** *New England Journal of Medicine* 2020, **382**(8):727-733.
29. Bosch BJ, van der Zee R, de Haan CA, Rottier PJ: **The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex.** *Journal of virology* 2003, **77**(16):8801-8811.
30. Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y, Wang Q, Xu Y, Li M, Li X *et al.*: **Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods.** *Acta pharmaceutica Sinica B* 2020.
31. Duffy S: **Why are RNA virus mutation rates so damn high?** *PLoS biology* 2018, **16**(8):e3000003.
32. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, Masciovecchio C, Angeletti S, Ciccozzi M, Gallo RC *et al.*: **Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant.** *Journal of Translational Medicine* 2020, **18**(1):179.
33. Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, Zhang Z: **The establishment of reference sequence for SARS-CoV-2 and variation analysis.** 2020.
34. Nalla AK, Casto AM, Huang M-LW, Perchetti GA, Sampoleo R, Shrestha L, Wei Y, Zhu H, Jerome KR, Greninger AL: **Comparative Performance of SARS-CoV-2 Detection Assays Using Seven Different Primer-Probe Sets and One Assay Kit.** *Journal of clinical microbiology* 2020, **58**(6):e00557-00520.
35. Vogels CBF, Brito AF, Wyllie AL, Fauver JR, Ott IM, Kalinich CC, Petrone ME, Casanovas-Massana A, Catherine Muenker M, Moore AJ *et al.*: **Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT-qPCR primer-probe sets.** *Nature microbiology* 2020.
36. Zhao WM, Song SH, Chen ML, Zou D, Ma LN, Ma YK, Li RJ, Hao LL, Li CP, Tian DM *et al.*: **The 2019 novel coronavirus resource.** *Yi chuan = Hereditas* 2020, **42**(2):212-221.
37. **2019 nCoV** [<https://bigd.big.ac.cn/ncov>]
38. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brünink S, Schneider J, Schmidt ML *et al.*: **Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR.** *Euro Surveill* 2020, **25**(3):2000045.
39. Yin C: **Genotyping coronavirus SARS-CoV-2: methods and implications.** *Genomics* 2020.
40. Chang TJ, Yang DM, Wang ML, Liang KH, Tsai PH, Chiou SH, Lin TH, Wang CT: **Genomic Analysis and Comparative Multiple Sequence of SARS-CoV2.** *Journal of the Chinese Medical Association : JCMA* 2020.
41. Yao H, Lu X, Chen Q, Xu K, Chen Y, Cheng L, Liu F, Wu Z, Wu H, Jin C *et al.*: **Patient-derived mutations impact pathogenicity of SARS-CoV-2.** *medRxiv* 2020:2020.2004.2014.20060160.
42. Suo T, Liu X, Feng J, Guo M, Hu W, Guo D, Ullah H, Yang Y, Zhang Q, Wang X *et al.*: **ddPCR: a more accurate tool for SARS-CoV-2 detection in low viral load specimens.** *Emerging Microbes & Infections*

2020:1-30.

43. Liu X, Feng J, Zhang Q, Guo D, Zhang L, Suo T, Hu W, Guo M, Wang X, Huang Z *et al*: **Analytical comparisons of SARS-CoV-2 detection by qRT-PCR and ddPCR with multiple primer/probe sets.** 2020:1-12.
44. Sapoval N, Mahmoud M, Jochum MD, Liu Y, Leo Elworth RA, Wang Q, Albin D, Ogilvie H, Lee MD, Villapol S *et al*: **Hidden genomic diversity of SARS-CoV-2: implications for qRT-PCR diagnostics and transmission.** 2020.
45. Treangen TJ: *bioRxiv*.
46. Asselta R, Paraboschi EM, Mantovani A, Duga S: **ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy.** *Aging* 2020, **12**(11):10087-10098.
47. Becerra-Flores M, Cardozo T: **SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate.** 2020:e13525.
48. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, Masciovecchio C, Angeletti S, Ciccozzi M, Gallo RC *et al*: **Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant.** *J Transl Med* 2020, **18**(1):179.
49. Chen J, Wang R, Wang M, Wei GW: **Mutations Strengthened SARS-CoV-2 Infectivity.** *Journal of molecular biology* 2020.
50. **CDC's Diagnostic Test for COVID-19 Only and Supplies** [<https://www.cdc.gov/coronavirus/2019-ncov/lab/virus-requests.html>]
51. Kalle E, Kubista M, Rensing C: **Multi-template polymerase chain reaction.** *Biomolecular Detection and Quantification* 2014, **2**:11-29.
52. Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsky JJ: **Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies.** *Nucleic acids research* 1990, **18**(4):999-1005.

Figures

Figure 1

Variation landscape of primer and probe target sequences in SARS-CoV-2 strains. Variant loci and numbers of mutant strains in ORF1ab (A), S (B), E (C), and N region (D). (E) Comparison of the number of variant loci in the four regions. (F) The number of variant loci (gray) and variation density (blue) in the four regions. (G) Mutant strains (light green) and strain density (green) of the four regions.

Figure 1

Variation landscape of primer and probe target sequences in SARS-CoV-2 strains. Variant loci and numbers of mutant strains in ORF1ab (A), S (B), E (C), and N region (D). (E) Comparison of the number of variant loci in the four regions. (F) The number of variant loci (gray) and variation density (blue) in the four regions. (G) Mutant strains (light green) and strain density (green) of the four regions.

Figure 1

Variation landscape of primer and probe target sequences in SARS-CoV-2 strains. Variant loci and numbers of mutant strains in ORF1ab (A), S (B), E (C), and N region (D). (E) Comparison of the number of variant loci in the four regions. (F) The number of variant loci (gray) and variation density (blue) in the four regions. (G) Mutant strains (light green) and strain density (green) of the four regions.

Figure 2

Variation analysis of SARS-CoV-2 strains in the target sequence of the nucleic acid amplification tests. (A) Pie chart of variation types. (B) Pie chart of SNP types. (C) Strain number and strain density of each SNP type. (D) Pie chart of the SNP types in ORF1ab, S, E, and N regions.

Figure 2

Variation analysis of SARS-CoV-2 strains in the target sequence of the nucleic acid amplification tests. (A) Pie chart of variation types. (B) Pie chart of SNP types. (C) Strain number and strain density of each SNP type. (D) Pie chart of the SNP types in ORF1ab, S, E, and N regions.

Figure 2

Variation analysis of SARS-CoV-2 strains in the target sequence of the nucleic acid amplification tests. (A) Pie chart of variation types. (B) Pie chart of SNP types. (C) Strain number and strain density of each SNP type. (D) Pie chart of the SNP types in ORF1ab, S, E, and N regions.

Figure 3

Variation analysis of N:28881, N: 28882, and N:28883 loci. (A) Variation landscape of primers F-18 and R-18. (B) Venn of the mutant strains containing variation in N:28881, N: 28882, or N:28883. (C) Variation curve of N:28881, N: 28882, and N:28883 over time and by country. (D) The number of isolates and the variation frequency of N:28881, N: 28882, and N:28883 by country. Regional variances are shown at the top; blue histogram indicates isolate number; and the black line chart indicates variation frequency.

Figure 3

Variation analysis of N:28881, N:28882, and N:28883 loci. (A) Variation landscape of primers F-18 and R-18. (B) Venn of the mutant strains containing variation in N:28881, N:28882, or N:28883. (C) Variation curve of N:28881, N:28882, and N:28883 over time and by country. (D) The number of isolates and the variation frequency of N:28881, N:28882, and N:28883 by country. Regional variances are shown at the top; blue histogram indicates isolate number; and the black line chart indicates variation frequency.

Figure 3

Variation analysis of N:28881, N:28882, and N:28883 loci. (A) Variation landscape of primers F-18 and R-18. (B) Venn of the mutant strains containing variation in N:28881, N:28882, or N:28883. (C) Variation curve of N:28881, N:28882, and N:28883 over time and by country. (D) The number of isolates and the variation frequency of N:28881, N:28882, and N:28883 by country. Regional variances are shown at the top; blue histogram indicates isolate number; and the black line chart indicates variation frequency.

Figure 4

Isolates analysis of common variant loci. The number of isolates and the variation frequency of ORF1ab: 490 (A), ORF1ab: 514 (B), ORF1ab: 13402 (C), N: 28311 (D), and N: 28688 (E). Regional variances are shown at the top; blue histogram indicates isolate number; and the black line chart indicates variation.

Figure 4

Isolates analysis of common variant loci. The number of isolates and the variation frequency of ORF1ab: 490 (A), ORF1ab: 514 (B), ORF1ab: 13402 (C), N: 28311 (D), and N: 28688 (E). Regional variances are shown at the top; blue histogram indicates isolate number; and the black line chart indicates variation.

Figure 4

Isolates analysis of common variant loci. The number of isolates and the variation frequency of ORF1ab: 490 (A), ORF1ab: 514 (B), ORF1ab: 13402 (C), N: 28311 (D), and N: 28688 (E). Regional variances are shown at the top; blue histogram indicates isolate number; and the black line chart indicates variation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S1Table.xlsx](#)
- [S1Table.xlsx](#)
- [S1Table.xlsx](#)
- [S2table.xlsx](#)
- [S2table.xlsx](#)
- [S2table.xlsx](#)
- [S3tabel.xlsx](#)
- [S3tabel.xlsx](#)
- [S3tabel.xlsx](#)
- [S4Table.xlsx](#)
- [S4Table.xlsx](#)
- [S4Table.xlsx](#)
- [FigureS1.pdf](#)
- [FigureS1.pdf](#)
- [FigureS1.pdf](#)
- [SupplementaryMaterial.docx](#)
- [SupplementaryMaterial.docx](#)
- [SupplementaryMaterial.docx](#)