Next-Generation Sequencing (NGS) in COVID-19

Next-generation sequencing (NGS) is a technology used by countless laboratories across the world for investigating the genetic makeup of all forms of living beings, but its utilization in infectious disease diagnostics is relatively scarce at the present moment. Information gleaned from NGS, whereby the pathogen's genome sequence is determined, yields a much greater trove of knowledge than the data produced by standard testing procedures, including information for the development of therapeutics and vaccines, the monitoring of changes in the virus as it circulates through the population, and deeper insights into patterns of transmission across time and geography.

1. Introduction

The outbreak of COVID-19 was first identified in Wuhan, China, a city in the Hubei province, in December 2019. A cluster of unexplained pneumonia cases led to the designation of COVID-19 as novel pneumonia, and work began immediately to identify the pathogen responsible for the outbreak and to delineate its genomic sequence [1][2][3][4].

The virus has now spread to all corners of the globe, and innovators in academia, government, and private industry are moving at an unprecedented scale and pace to bring forward solutions to mitigate and resolve the crisis. Of the several measures taken to control the spread of disease, testing for SARS-CoV-2 in the population is a primary measure that has been implemented globally. Most importantly, testing provides people with evidence of infection, allowing them and those they have encountered to take necessary precautions, such as quarantining, to reduce community exposure. Additionally, standard, widespread testing yields data for researchers and public health officials to utilize for transmission modeling and public policy decision making regarding issues such as social distancing and mask use, among others.

2. Next-Generation Sequencing (NGS)

NGS is an emergent technology that has the power to sequence billions of nucleic acid fragments simultaneously, with recent advances rendering dramatically reduced time and cost of sequencing. The power of high throughput NGS technology presents the scientific community with many promising applications, and the current pandemic is providing an enormous sense of urgency to push some of these applications into widespread usage. One application opportunity for NGS is in the field of clinical diagnosis of infectious disease. Using NGS technology for the diagnosis of infectious diseases offers an unbiased approach detecting pathogens that does not rely on culturing or the need for clinical hypotheses. While standard testing procedures require clinicians to identify possible explanations for a patient's symptoms and employ tests aimed at those specific pathogens, NGS testing can reveal the presence of all types of microorganisms present in a sample, including bacteria, viruses, fungi, and parasites.

The massive additional benefit of the widespread adoption of NGS for routine diagnostics is the wealth of information provided by the test results. Sequencing data are vital information for those involved in the fight against infectious disease, aiding in vaccine and antiviral development, phylogenetic analysis, viral spread tracing, the monitoring of the evolution of a pathogen, the development of other diagnostic tests, and the identification of any primary and intermediate zoonotic hosts [5]. Before examining the implementation of NGS for clinical diagnosis and the SARS-CoV-2 pandemic in the literature, we describe the basic nature of the technology, its bioinformatic processing pipelines, and challenges facing its widespread application.

2.1. NGS Fundamentals

2.1.1. NGS Methods

The development of NGS subsequently occurred in the early 2000s and rapidly advanced the scientific community's ability to investigate the genomes of living organisms via its massively high throughput capability. The methods of next-

generation sequencing vary in their technical mechanisms, but all share the basic defining features. They rely on samples being broken down into fragment libraries that are each amplified and sequenced independently, generating millions of fragment reads (small sequences) that can be pieced together to generate a readout of the genome. Pyro-sequencing, sequencing by synthesis, sequencing by ligation, and ion-semiconductor sequencing are all subtypes of NGS technology, each coming with its technical variations as well as advantages and weaknesses. Sequencing by synthesis (SBS) is a widely used method and is utilized by several commercial companies.

In the SBS method, the generated library fragments are modified with adapter motifs which include sequence binding sites, indices, and regions complementary to oligomer linkers on a chip surface. The fragments are subsequently attached to a chip via binding with oligomer linker sequences, and each fragment is then amplified at a distinct point on the chip, creating its cluster. The amplification is accomplished by a process involving bridge amplification, resulting in distinct clusters of fragments with identical sequences, present in both the forward and reverse orientations. The clusters are modified by washing processes so that only the forward sequences remain, and sequencing is commenced. After reading the forward sequences, the process is inverted and the reverse sequences are read. The sequencing process itself is accomplished by successive rounds of the addition of fluorescently tagged nucleotides whose base can be identified in real-time by their light emission profile. The device can analyze each cluster's data simultaneously, effectively reading millions of fragment sequences in a short amount of time [6].

2.1.2. Bioinformatic Processing of NGS Data

NGS technology generates a massive amount of raw data that requires substantial amounts of computational processing to yield significant and actionable results. An NGS bioinformatic pipeline refers to a series of algorithms that the data are run through to generate a useful interpretable output [7]. Due to the differences among proprietary NGS technological methods and differences in commercial laboratory processes and goals, there is inevitable variation in the software and computational tools utilized by these groups. All NGS bioinformatics pipelines share several major features including sequence generation, assembly and alignment, variant identification, variant annotation, and variant prioritization and visualization [8][7][9]. Bioinformatic pipelines also require rigorous quality control methodology to guarantee the validity of results.

Sequence generation is the process by which the NGS platform converts sensor information into base calls, thereby generating the raw nucleotide sequence of the fragment at each cluster. Technical methods vary among manufacturers. The Illumina sequence-by-synthesis platform accomplishes this task by measuring fluorescence output at each oligonucleotide cluster to identify the nucleotide being incorporated into the strand during each round [8].

The next component of the bioinformatic pipeline involves the assembly and alignment of sequencing fragments. Assembly refers to the creation of *contigs*, which are longer consensus sequences pieced together by computational analysis of overlapping sequences generated from different clusters. Alignment is also described as *mapping* and refers to the arrangement of fragment reads or contigs along a reference genome. The relative ordering of alignment and assembly in the pipeline workflow is variable and depends on individual project aims. The two processes can occur concurrently, or a de novo sequence can be constructed, followed by a comparison to a reference genome. The de novo construction of a genome is a method by which scientists can create a reference genome for newly studied organisms [8].

An important consideration in the alignment step is the concept of stringency, which refers to how strictly the read sequences or contigs must match the reference sequence. Low stringency would allow for possible strand misalignment, while high stringency could allow variants to be lost due to their lack of an exact match with the reference sequence. Short-read sequencing (<250 bp) can also face difficulties in alignment due to the presence of large regions of homology in the reference genome that restrict the ability of short fragments to be appropriately mapped [10].

Variant identification, also known as *variant calling*, is the practice of identifying regions where the sample and reference sequences diverge [7]. The most basic example of this would be a single nucleotide variant (SNV), where one nucleotide is mismatched amid otherwise identical sequences. Other variants include small insertions and deletion mutations (INDELs), copy number variations, or larger structural changes such as inversions or translocations. There are a variety of computational tools available for use in the variant calling step of the NGS bioinformatics pipeline.

Annotation is a process whereby identified variants are further characterized in the context of associated metadata. Essentially, programs query assorted databases to link variants to useful information that helps put them into context. Aspects such as the variant's location in the genome, predicted changes to cDNA or amino acid sequences, or how commonly they show up in common variant databases are probed [7]. These annotations are subsequently used to prioritize the variants. Insignificant findings can be filtered out to help scientists focus on salient findings; this is important

to help researchers or clinicians avoid being inundated with an unmanageable flood of information. An example of an insignificant finding is an SNV that has been widely established as benign [7]. Laboratories can utilize hard filters that remove all variants except for those with relevance to their research or clinical interest.

2.1.3. Quality Control for Bioinformatics Pipelines

Another universal and vital aspect of all NGS bioinformatic pipelines is quality control. Multiple steps are taken throughout the data processing operation to ensure that the results are of high quality and clinically useful. Some common quality control metrics are described below.

A simple yet important measure of validity is the number of reads performed by the sequencer. Coverage depth refers to the average number of reads there are for a reference locus and helps determine the degree of confidence of the validity of a variant discovery. Horizontal coverage refers to the percentage of the reference genome that is aligned to NGS-generated sequences. Higher coverage lowers the likelihood that a false positive or false negative variant will be obtained [11].

When multiple samples are run together or *multiplexed*, each sample is identified by a distinct molecular bar code. Demultiplexing involves the separation of samples by barcode after base calls are made, and the success of this process yields a useful quality metric [8]. During sequence generation, the NGS platform assigns a quality score to each "base call" to indicate the statistical likelihood of correct identification of each nucleotide. Similarly, during the alignment step, a mapping-quality score is assigned to each of the fragment reads and is used as an indication of the likelihood of accurate alignment to the reference genome. The alignment phase allows for the calculation of the horizontal coverage and depth of sequencing, useful indicators of statistical significance, as mentioned previously [8][7]. Before variant calls can be made, more computational processing is often necessary to ensure that the base calling and genome-alignment mapping data are of high quality. An example of one of these prevariant identification processes is local sequence realignment around loci where INDEL mutations are expected to have occurred. *Variant filtering* is a post-variant identification step used to screen identified genomic variants for likely false positives using metadata such as base-call quality scores, mapping quality scores, and read depth, among others [8].

2.2. NGS and the SARS-CoV-2 Pandemic

2.2.1. NGS for Pathogen Detection: Prepandemic

The development of sequencing technology for diagnostic and pathogen surveillance was an urgent undertaking even before the SARS-CoV-2 pandemic erupted, as the cause of many infections often goes undiagnosed. Jain et al. found that of 2,488 hospitalized patients with community-acquired pneumonia, only 853 cases (38%) were diagnosed with a causative pathogen [12]. The problem extends beyond the respiratory system; clinicians are often similarly unable to pinpoint the etiology for CNS infections. A study conducted by Glaser et al. of 1570 patients in California found no etiology in 63% of cases of encephalitis [13]. These unexplained infections often lead to inadequate treatment and poor outcomes, while simultaneously contributing to the widespread overuse of antibiotics as unsure providers use antibiotics liberally when an infection is unexplained.

Using NGS technology for the identification of infectious diseases promises an unbiased approach that does not rely on culturing, and NGS has already been shown in various case reports and preliminary studies to be capable of identifying pathogens in samples taken from the respiratory system, central nervous system, gastrointestinal system, and the eyes [14]. Studies have demonstrated the utility and practicality of NGS in diagnosis, showing that results can be obtained from "sample-to-answer" in 48 h [15], similar to the wait times experienced by those being tested by standard RT-PCR COVID-19 tests around the country. The ability to run many samples together by multiplexing should allow laboratories to accommodate the high number of samples for testing to clear the backlog. Although there is currently limited data available on the use of NGS for high volume COVID-19 testing, we have examined reports from labs across the world that are using NGS technology to aid in the fight against the SARS-CoV-2 virus.

Key Findings:
1. A large percentage of commonly clinical diseases are due to infections of unknown etiology [12][13].
2. NGS has been proven to be capable of identifying infectious microorganisms from various patient sample types [14].
3. NGS has been shown to provide clinically practical turnaround times [15].

2.2.2. NGS for Detection of SARS-CoV-2

In late January 2020, Lu et al. reported SARS-CoV-2 genomic data from nine patients presenting with pneumonia of unknown origin at three hospitals in Wuhan, China [1]. BAL and cultured isolates were used as samples. The patients' samples were negative for known respiratory pathogens, with five tested by the Chinese CDC and four by the BGI group in Beijing, China. NGS technology was used to sequence and identify the causative pathogen, with the BGI and CDC labs differing slightly in their sequencing techniques and bioinformatic processing pipelines. In both groups, gaps between contigs were connected using Sanger sequencing and terminal genome regions were identified via rapid amplification of cDNA ends (RACE).

At the BGI group, RNA extraction of BALF samples was carried out with a QIAamp Viral RNA Mini Kit, and a probe-captured technique was used to remove human nucleic acid material. Next, RNA was reverse transcribed to cDNA, second-strand synthesis was performed, and a DNA library was constructed. The DNA library was quantified with a Qubit method and transformed into a single strand circular library. Rolling circle amplification was used to construct DNA nanoballs, and they were subsequently qualified. The DNBSEQ-T7 high throughput sequencer from MGI was used with paired-end, 100 bp read lengths. High quality reads were filtered for human reads against the hg19 human reference genome with Burrow-Wheeler alignment software. The remaining data were aligned with published data on coronaviruses from the US National Center for Biotechnology Information. Mapped reads were assembled with SPAdes software to create a consensus genome sequence.

The Chinese CDC sequencing protocol similarly used the QIAamp Viral RNA Mini Kit to extract viral RNA from the clinical samples, followed by cDNA synthesis and second-strand synthesis. cDNA libraries were generated and then purified with Agencourt AMPure XP beads to remove contaminants. Following quantitation, the sequencing was carried out on MiSeq or iSeq platforms from Illumina. The terminal genome regions were identified by the use of Rapid amplification of cDNA ends (RACE) system from Invitrogen. Assembled genomes were confirmed with traditional Sanger sequencing. The raw sequencing reads were filtered via the same protocol used by the BGI group, and CLCBio software version 11.0.1 was used for de novo assembly, variant calling, and alignment. The bat-SL-CoVZC45 virus (containing 87.99% sequence similarity) was also used to perform a mapped assembly.

Sequencing yielded eight full genomes and two partial genomes (one patient's BALF sample was used to isolate the virus, which was also sequenced, yielding 10 total samples). The sequences were used to generate PCR-based assays, that were then used to confirm the presence of the SARS-CoV-2 virus, and cycle threshold (Ct) values ranged from 22.85 to 34.23.

The results of sequencing the viral genome in this study yielded highly useful information during the early stages of the SARS-CoV-2 outbreak. Genomic analyses led to the revelation that, while the whole genome sequence of SARS-CoV-2 is highly similar to bat-SL-CoVZC45 (87.99% similarity) and bat-SL-CoVZXC21 (87.23%), the receptor binding domain (S1) sequence of the spike protein (S), was more similar to that of SARS-CoV, the virus responsible for the first SARS outbreak in the early 2000s. This evidence supports the suggestion that SARS-CoV-2 uses the ACE-2 receptor to gain entry into cells, the same route utilized by SARS-CoV. The utilization of ACE-2 receptors by SARS-CoV-2 has also been demonstrated in infectivity studies by Zhou et al. [3]. The phylogenetic analysis, made possible by the assembled sequences, allowed the classification of the virus, showing that the virus belongs to the subgenus Sarbecovirus, a member of the Betacoronavirus genus. The high sequence similarity (over 99.9%) among viral samples obtained from the nine patients in Wuhan provides evidence of very recent entry into the human population.

Other laboratories in China conducted parallel investigations at the onset of the outbreak, such as Zhu et al. [4], who used a similar combination of Illumina and nanopore sequencing, RACE, and Sanger sequencing to identify and characterize the SARS-CoV-2 genomes extracted from three patient samples in Wuhan, China. Their bioinformatics pipeline included CLC Genomics software, version 4.6.1; Muscle; and RAxML (13) for phylogenetic analysis. Their sequencing protocol yielded more than 20,000 viral reads per sample, obtaining one full-length genome and two nearly full-length genomes. They similarly noted that contigs aligned with high similarity with bat-SL-CoVZC45. Published 24 January 2020, they reached similar conclusions to Lu et al. regarding the phylogenetic characterizations of the virus and used their de novo generated sequences to design primers for PCR-based diagnostic assays.

Groups all over the world are now investigating possible diagnostic interventions made possible by NGS technology. Campos et al. reported the use of metatranscriptomic next-generation sequencing technology in the detection of SARS-CoV-2 in a nasopharyngeal swab specimen from a patient in Feira de Santana-Bahia, Brazil [16]. They used the Ion S5 platform from ThermoFisher with an Ion 540™ chip and the Ion Total RNA-Seq kit v2. This platform uses an ion-semiconductor sequencing process, and they implemented the Low Input RiboMinus™ Eukaryote System v2 from ThermoFisher to remove rRNA from one sample. The rRNA-depleted library contained human transcripts as 77.29% of

total reads, while the whole RNA library had 84.49% of total reads as human transcripts. Contigs from the rRNA-depleted library provided 29.9% genome coverage, while contigs from the non-depleted sample yielded only 5.4% genome coverage. Total genome coverage from all viral reads in the rRNA depleted sample was 59.9%. These results indicate that rRNA-depletion strategies may play a role in improving NGS diagnostic abilities.

2.2.3. Co-Infection in COVID-19 Patients

Co-infection data relating to SARS-CoV-2 is of high interest to clinicians around the world, as rates of co-infection differ among viruses, and their presence is associated with poor patient outcomes in many common viruses [17]. One early investigation of 191 inpatients with COVID-19 in China in December 2019–January 2020 reported that 50% (27/54) of non-survivors had a secondary infection [18]. The possibility of co-infection is also important for clinical diagnosis. If co-infection is found to be rare, a positive test for a common respiratory pathogen could indicate a lack of SARS-CoV-2 infection, while higher chances of co-infection would not allow clinicians to rule out such a possibility. Reports on the rates of co-infection in COVID-19 patients have varied widely, ranging from approximately 1–20%, with one early study in Wuhan reporting a co-infection rate of 57.3% [19][20]. Consequently, larger-scale studies are needed to fully assess the role of secondary infections in these patients.

2.2.4. NGS as a Tool for Understanding SARS-CoV-2: Additional Benefits of Adoption

The adoption of NGS for widespread use yields the additional benefit of generating massive amounts of data that can be used to study pathogens. Researchers in academia, government, and industry have also been using sequencing technology and data to aid in understanding the attributes, processes, and phylogenetics of the SARS-CoV-2 virus. Discoveries in these areas are important for the development of vaccines, antivirals, and novel diagnostics, as well as generating useful information for public health authorities such as data on transmission and viral tracing. The widespread adoption of NGS in diagnosis will lead to major gains in scientists' and governments' ability to monitor emerging variants of infectious diseases such as SARS-CoV-2. The United States CDC is leading what it calls the "SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology, and Surveillance (SPHERES)" program. The program is aimed at coordinating large-scale sequencing of the SARS-CoV-2 virus to facilitate these previously mentioned goals [21].

2.2.5. Understanding Physical and Chemical Properties of the Virus

As previously noted, Lu et al. used sequencing results generated in their lab to provide evidence that the structure of the receptor binding domain (S1) of the spike (S) protein was highly similar to that of the original SARS-CoV virus [1], indicating that the protein would likely bind the ACE-2 receptor. This type of genetic analysis is a highly useful byproduct enabled by sequencing infectious diseases, as it can reveal structural parameters that guide pharmaceutical developments.

2.2.6. SARS-CoV-2 Phylogenetics and Mutational Characteristics

Scientists in Italy, a country especially hard hit by the outbreak, began performing NGS analyses early in the pandemic to gain a better understanding of how the virus has spread across the country. Lorusso et al. sequenced 46 samples from patients in the Abruzzo region of central Italy between 16 March and 23 March 2020 [22]. They chose these 46 samples for NGS among 839 SARS-CoV-2 positive samples based on their low Ct scores during RT-PCR testing. Their protocol utilized the MiniSeq Mid Output Kit (300-cycles) with 150 bp paired-end reads, and they used trimmomatic bowtie2 and samtools software products for bioinformatic processing.

2.3. Challenges Related to NGS Adoption

The opportunities made available by the widespread adoption of NGS in clinical laboratories are numerous. However, several challenges impede the routine use of NGS in these settings. To implement NGS in clinical laboratories, validation studies with diagnostic performance characteristics such as limit of detection, sensitivity, and specificity for intended microorganisms need to be determined. Apart from validation, logistical challenges include the management of contamination and the implementation of rigorous, standardized controls at all stages of the testing workflow. Furthermore, challenges exist in ensuring database validity and assessing the clinical significance of results. Finally, cost–benefit analyses must be considered relating to the economic realities of which testing methods are most appropriate in different instances.

References

1. Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu, N.; et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. Lancet 2020, 395, 565–574.

2. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; et al. A new coronavirus associated with human respiratory disease in China. Nature 2020, 579, 265–269.

3. Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020, 579, 270–273.

4. 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. N. Engl. J. Med. 2020, 382, 727–733.

5. Udugama, B.; Kadhiresan, P.; Kozlowski, H.N.; Malekjahani, A.; Osborne, M.; Li, V.Y.C.; Chen, H.; Mubareka, S.; Gubbay, J.B.; Chan, W.C.W. Diagnosing COVID-19: The Disease and Tools for Detection. ACS Nano 2020, 14, 3822–3835.

6. Illumina. Illumina Sequencing Technology. Available online: https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf (accessed on 5 May 2021).

7. Roy, S.; Coldren, C.; Karunamurthy, A.; Kip, N.S.; Klee, E.W.; Lincoln, S.E.; Leon, A.; Pullambhatla, M.; Temple-Smolkin, R.L.; Voelkerding, K.V.; et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. J. Mol. Diagn. 2018, 20, 4–27.

8. Laura, J.; Tafe, M.E.A. Genomic Medicine: A Practical Guide; Springer: Cham, Switzerland, 2020.

9. Wadapurkar, R.M.; Vyas, R. Computational analysis of next generation sequencing data and its applications in clinical oncology. Inform. Med. Unlocked 2018, 11, 75–82.

10. Mandelker, D.; Schmidt, R.J.; Ankala, A.; Gibson, K.M.; Bowser, M.; Sharma, H.; Duffy, E.; Hegde, M.; Santani, A.; Lebo, M.; et al. Navigating highly homologous genes in a molecular diagnostic setting: A resource for clinical next-generation sequencing. Genet. Med. 2016, 18, 1282–1289.

11. Illumina. Coverage Depth Recommendations. Available online: https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/coverage.html (accessed on 5 May 2021).

12. Jain, S.; Self, W.H.; Wunderink, R.G.; Fakhran, S.; Balk, R.; Bramley, A.M.; Reed, C.; Grijalva, C.G.; Anderson, E.J.; Courtney, D.M.; et al. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. N. Engl. J. Med. 2015, 373, 415–427.

13. Glaser, C.A.; Honarmand, S.; Anderson, L.J.; Schnurr, D.P.; Forghani, B.; Cossen, C.K.; Schuster, F.L.; Christie, L.J.; Tureen, J.H. Beyond viruses: Clinical profiles and etiologies associated with encephalitis. Clin. Infect. Dis. 2006, 43, 1565–1577.

14. Gu, W.; Miller, S.; Chiu, C.Y. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. Annu. Rev. Pathol. Mech. Dis. 2019, 14, 319–338.

15. Wilson, M.R.; Naccache, S.N.; Samayoa, E.; Biagtan, M.; Bashir, H.; Yu, G.; Salamat, S.M.; Somasekar, S.; Federman, S.; Miller, S.; et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N. Engl. J. Med. 2014, 370, 2408–2417.

16. Campos, G.S.; Sardi, S.I.; Falcao, M.B.; Belitardo, E.M.M.A.; Rocha, D.J.P.G.; Rolo, C.A.; Menezes, A.D.; Pinheiro, C.S.; Carvalho, R.H.; Almeida, J.P.P.; et al. Ion torrent-based nasopharyngeal swab metatranscriptomics in COVID-19. J. Virol. Methods 2020, 282, 113888.

17. Joseph, C.; Togawa, Y.; Shindo, N. Bacterial and viral infections associated with influenza. Influ. Other Respi. Viruses 2013, 7, 105–113.

18. Zhou, F.; Yu, T.; Du, R.; Fan, G.; Liu, Y.; Liu, Z.; Xiang, J.; Wang, Y.; Song, B.; Gu, X.; et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. Lancet 2020, 395, 1054–1062.

19. Wee, L.E.; Ko, K.K.K.; Ho, W.Q.; Kwek, G.T.C.; Tan, T.T.; Wijaya, L. Community-acquired viral respiratory infections amongst hospitalized inpatients during a COVID-19 outbreak in Singapore: Co-infection and clinical outcomes. J. Clin. Virol. 2020, 128, 104436.

20. Kim, D.; Quinn, J.; Pinsky, B.; Shah, N.H.; Brown, I. Rates of Co-infection Between SARS-CoV-2 and Other Respiratory Pathogens. JAMA 2020, 323, 2085–2086.

21. CDC. A National Open Genomics Consortium for the COVID-19 Response. Available online: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/spheres.html (accessed on 5 May 2021).

22. Lorusso, A.; Calistri, P.; Mercante, M.T.; Monaco, F.; Portanti, O.; Marcacci, M.; Cammà, C.; Rinaldi, A.; Mangone, I.; Di Pasquale, A.; et al. A "One-Health" approach for diagnosis and molecular characterization of SARS-CoV-2 in Italy. One Health 2020, 10, 100135.