



The next phase of SARS-CoV-2 surveillance: real-time molecular epidemiology

Bas B. Oude Munnink , Nathalie Worp, David F. Nieuwenhuijse , Reina S. Sikkema, Bart Haagmans , Ron A. M. Fouchier  and Marion Koopmans  

The current coronavirus disease 2019 (COVID-19) pandemic is the first to apply whole-genome sequencing near to real time, with over 2 million severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) whole-genome sequences generated and shared through the GISAID platform. This genomic resource informed public health decision-making throughout the pandemic; it also allowed detection of mutations that might affect virulence, pathogenesis, host range or immune escape as well as the effectiveness of SARS-CoV-2 diagnostics and therapeutics. However, genotype-to-phenotype predictions cannot be performed at the rapid pace of genomic sequencing. To prepare for the next phase of the pandemic, a systematic approach is needed to link global genomic surveillance and timely assessment of the phenotypic characteristics of novel variants, which will support the development and updating of diagnostics, vaccines, therapeutics and nonpharmaceutical interventions. This Review summarizes the current knowledge on key viral mutations and variants and looks to the next phase of surveillance of the evolving pandemic.

The COVID-19 pandemic has put the use of pathogen genomic sequencing to support public health decision-making on center stage. Rapid sharing of the first viral genome sequences of SARS-CoV-2 (ref. ¹) showed that this virus is a member of the species *Severe acute respiratory syndrome-related coronavirus* in the family *Coronaviridae*, subfamily *Orthocoronavirinae*, genus *Betacoronavirus*, subgenus *Sarbecovirus*², and is closely related to SARS-CoV and a diverse group of SARS-like coronaviruses identified in bats. The recent World Health Organization (WHO) mission to search for the origin of SARS-CoV-2 described the jump of the virus from bats, either directly or through an intermediate animal host, to humans as the most likely route by which the virus caused the pandemic³. Interestingly, SARS-CoV-2 also clusters with sequences obtained from pangolin^{4,5}, although the time to the most recent common ancestor of SARS-CoV-2 and the related pangolin viruses dates to around 150 years ago⁶. Timely sharing of the first viral genome sequences also enabled the establishment of diagnostic tools⁷, including the development of specific SARS-CoV-2 whole-genome sequencing protocols⁸ and the rapid development of vaccines⁹.

With current massive genomic sequencing efforts, virological epidemiological surveillance is being performed near to real time. Mutations in the viral genome are also detected and shared near to real time, leaving the interpretation of their relevance for future work. Mutations and other genome changes are part of the normal replication and evolution process, and most mutations will not result in increased viral fitness. This Review summarizes the genomic surveillance efforts as well as the current nomenclature and detection of variants of concern (VOC) and variants of interest (VOI). In addition, the emergence of these variants is discussed, and future needs for faster genotype-to-phenotype prediction are described.

SARS-CoV-2 genomic surveillance

Virus genome sequencing has been increasingly used in recent years for outbreak research in the emerging disease field, as seen during the recent Ebola virus outbreak in Africa¹⁰ and the arbovi-

rus outbreaks in South America^{11–14}; however, the scale of genomic surveillance undertaken during the current pandemic is unprecedented. During the first year of the pandemic, a large number of SARS-CoV-2 whole-genome sequences were generated from all around the world and shared, mostly through GISAID. As of 5 July 2021, 25,284 whole-genome sequences from Africa (0.32% of all reported SARS-CoV-2-positive cases from that continent), 146,562 from Asia (0.30% coverage), 1,292,415 from Europe (2.35% coverage), 692,704 from North America (1.75% coverage), 37,913 from South America (0.12% coverage) and 20,613 from Oceania (25% coverage) had been generated (ref. ¹⁵ and WHO Coronavirus (COVID-19) Dashboard (<https://covid19.who.int/table>)). Although the number of genomes is unprecedented, the coverage is still heavily biased toward regions and countries with specialized genomic facilities, programs and research projects^{16,17}. During the current SARS-CoV-2 outbreak, virus genome sequencing combined with metadata has been used to further study the origins of the pandemic¹⁸, to determine main routes of introduction and spread for outbreak investigations in hospitals¹⁹, nursing homes²⁰, schools²¹ and mink farms²², to analyze regional, national and international epidemiological trends and to study potential immune escape^{23–25}.

With the expanding scale of genomic sequencing, new analytical challenges arise. The massive spread of the virus, with over 183 million human individuals infected as of 5 July 2021 (ref. ²⁶), led to the accumulation of mutations within the viral genome. This is all part of the game: the virus replication process is not 100% error proof, leading to the generation of progeny genomes with small numbers of mutations or, occasionally, insertions or deletions²⁷. Currently, close to 50,000 nonsynonymous mutations have been observed²⁸. With ongoing transmission, such mutations can be replicated in subsequent rounds of infection, evolving into a unique fingerprint. When sufficient diversity is observed, the fingerprints can be used for epidemiological analyses at different levels of resolution, for instance, to link cases to a cluster, to track the origins of outbreaks, to understand the seeding of pandemic waves and to monitor the effects of control measures^{19,29–34}.

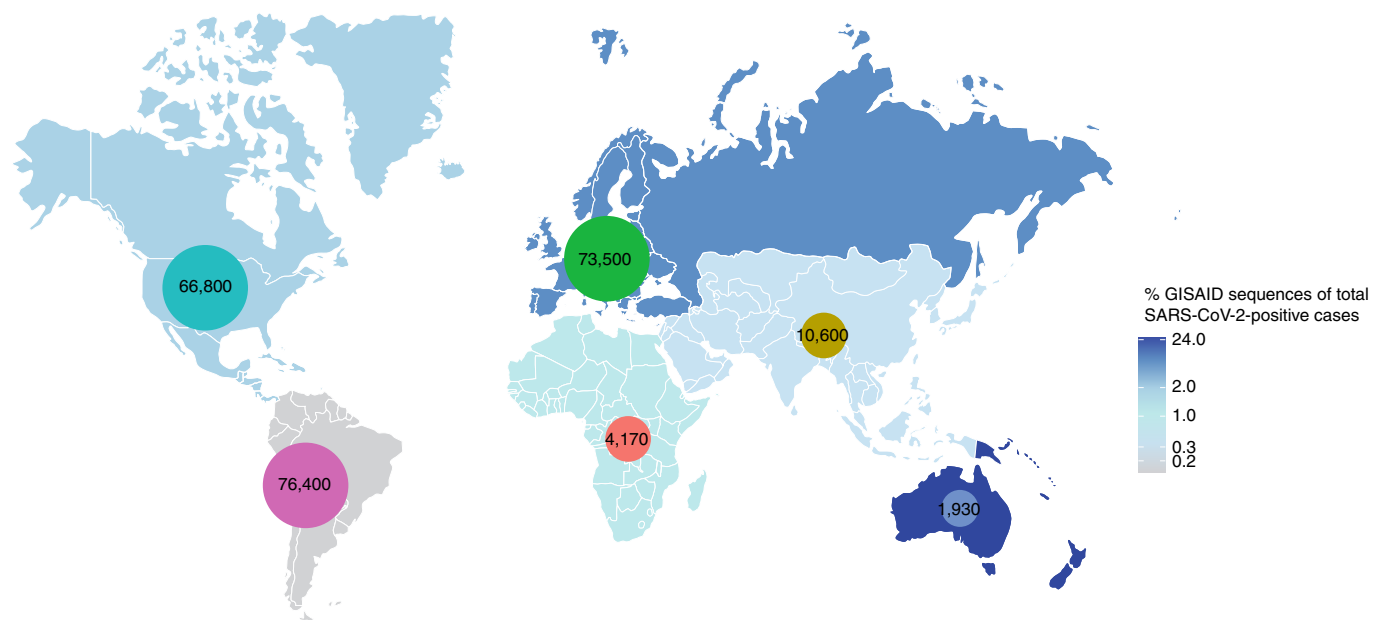


Fig. 1 | Overview of the percentage of whole-genome sequences generated and shared on GISAID compared to the total number of COVID-19 cases per continent as of 5 July 2021. The number in each circle indicates the number of diagnosed SARS-CoV-2 infections per 1 million people.

It is important to consider the biases implicit in generation of genomic data. The current genomic effort is biased toward a limited number of countries with high sequencing capacity. An overview of the percentage of SARS-CoV-2 sequences generated and shared on GISAID¹⁵ as compared to the total number of SARS-CoV-2 infections diagnosed as of 5 July 2021 (WHO Coronavirus (COVID-19) Dashboard, <https://covid19.who.int/table>) is shown in Fig. 1.

Nomenclature and classification tools

During the pandemic, a plethora of bioinformatics tools have been developed, and the open sharing of genomic data has triggered a massive stream of publications analyzing local, regional or global datasets for a broad range of questions³⁵. For such applications, a key issue has been the need for standardized, downsized reference datasets and for standardization of lineage nomenclature, which has been challenging as this needed to be developed during the evolving pandemic. The most frequently used lineage assignment and data visualization tools such as Pangolin³⁶, Nextstrain³⁷ and GISAID¹⁵ have greatly aided this process, but continued reassessment is needed as new challenges arise³⁸. Recently, the WHO published a nomenclature system using Greek alphabet letters to label VOIs and VOCs to make the names more easy to remember and more practical³⁹.

The most well-known systems are the Nextstrain SARS-CoV-2 clade naming strategy and Pango. In Pango, the earliest sequences from Wuhan were designated as lineage A (represented by Wuhan/WH04/2020; sampled 5 January 2020; GISAID accession EPI_ISL_406801) and lineage B (represented by Wuhan-Hu-1; sampled 26 December 2019; GenBank accession MN908947). Subsequent lineages were assigned a number, for instance, B.1, B.2 and so on, or letters, depending on the system used³⁴. To make tracking of strains accessible for providers of genetic data, GISAID collaborated with bioinformaticians, using interactive visualization software that provides rough overviews of the distribution of virus lineages across the world based on typical amino acid substitutions³⁵. The Pango nomenclature tool uses a numerical system to classify lineages in more detail³⁶ and seems to have gained the most traction in public health communications, in combination with the WHO classification that is limited to specific variants (for instance, Alpha variant, Pango lineage B.1.1.7).

VOCs

The implications of mutations, insertions or deletions in SARS-CoV-2 genomes are hard to determine from sequence data alone. While most mutations are silent or might result in phenotypic differences that are neutral or detrimental to viral fitness, some genomic changes may affect properties that are relevant for our ability to detect, treat, control or prevent infections or disease⁴⁰. Our understanding of the effects of certain genomic signatures is currently limited, as translating genotypes into phenotypes requires carefully designed experimental studies that may require months to complete.

Genomic tracking and data analysis has helped to identify virus variants that have drawn attention because of their epidemiological behavior. A first example was the emergence and global dispersal of viruses with an amino acid substitution (aspartic acid to glycine) in the spike protein at position 614 (ref. ⁴¹). This substitution was first described in B.1 lineage viruses identified toward the end of January 2020 in Guangzhou, Sichuan and Shanghai and subsequently in viruses from the same lineage identified in early cases of the pandemic in Germany, linked to a traveler from Shanghai⁴². This initial cluster was controlled, but viruses with the same substitution have been introduced on multiple occasions, seeding the pandemic in Europe. At that stage, it was not possible to determine whether the substitution reflected a founder event in the country of origin; however, since then, this mutation has been fixed in the genome and is now—as of 19 May 2021—present in 99.27% of the genomes sequenced since the start of 2021. Incursion into the United Kingdom allowed comparison of the spread of B.1 viruses with the 614G substitution over 614D viruses in the same epidemiological background, and displacement of 614D-encoding viruses over time was observed. Subsequent testing of the effect of the substitution on the infectivity of the virus in different cell types (using lentiviral vectors with SARS-CoV-2 spike protein on the viral surface) suggested that the D614G substitution caused an increase in infectivity^{43,44}, while structural analysis suggested a conformational change in the spike protein affecting binding and/or fusion⁴⁴. In addition, enhanced replication in the upper respiratory tract in hamsters⁴⁵ and somewhat enhanced transmission in animals were observed⁴⁶. Considering this in combination with the observed global displace-

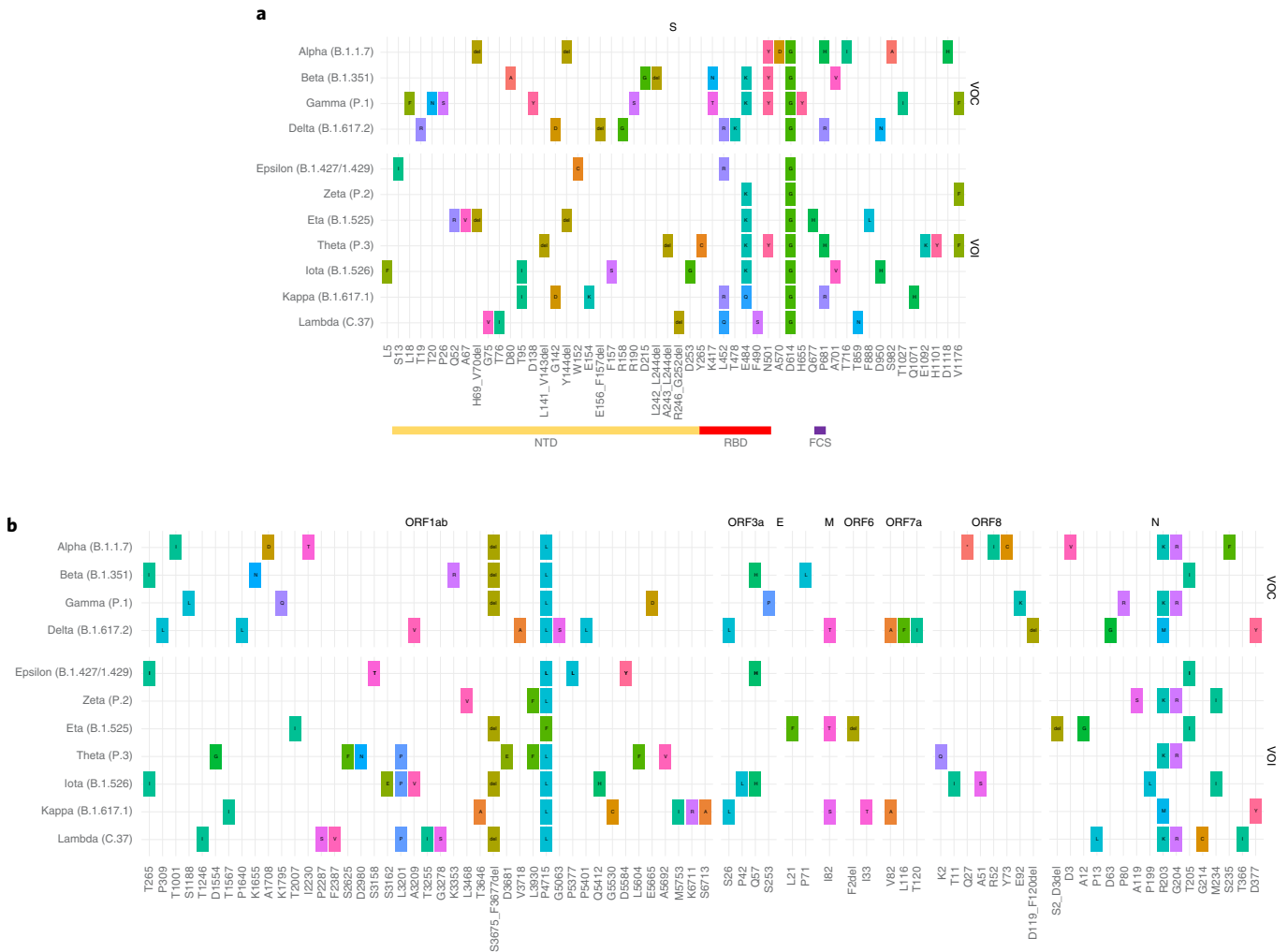


Fig. 2 | Overview of amino acid changes in specific proteins of VOCs and currently detected and former VOIs. a, Amino acid changes in the spike (S) protein of the indicated variants in comparison to the Wuhan-Hu-1 strain (NC_045512.2). **b**, Amino acid changes in the ORF1ab, ORF3a, envelope (E), membrane (M), ORF6, ORF7a, ORF8 and nucleocapsid (N) proteins of the indicated variants in comparison to the Wuhan-Hu-1 strain (NC_045512.2). NTD, N-terminal domain; RBD, receptor-binding domain; FCS, furin cleavage site; * indicates a stop codon.

ment of D614-encoding viruses, Hou et al. concluded that the virus had adapted to increased transmissibility, possibly through a shift toward more efficient upper respiratory tract infection⁴⁶.

A more recent phenomenon is the detection of new SARS-CoV-2 variants with multiple mutations across the genome that appear to have undergone a process of natural selection, resulting in an evolutionary jump in comparison to previous circulating viruses (Fig. 2a,b). These variants are declared VOCs when phenotypic traits of relevance to public health are attributed to them^{23,47,48}. The first variant with such an unusual number of mutations (Alpha (B.1.1.7)) was first noted in mid-November 2020 in the United Kingdom, a country that has stood out because of its massive sequencing effort. This VOC differed in 22 nucleotide positions from previously sequenced viruses, including at least 8 nonsynonymous changes mapping to the spike protein⁴⁸. One consequence of the genetic changes was that one of the three PCR targets used in the routine screening of cases in large test facilities failed, making it relatively easy to track the emergence and spread of the Alpha variant by monitoring the proportion of positive cases with target failure in the spike gene^{49,50}. The Alpha variant rapidly increased in prevalence in large parts of the United Kingdom and beyond and was associated with rapidly expanding community epidemics in different regions. UK scientists, on the basis of phylodynamic analyses

and modeling, have suggested that the variant strain may be more transmissible^{50,51}. This conclusion was based on their analyses of virus-lineage-specific trends in COVID-19 reporting, combined with data on social contacts and mobility information^{48,52}. These analyses led to the conclusion that the observed pattern of spread was best explained by assuming that the Alpha variant had increased transmissibility, increasing the reproduction number by 0.4 or more in comparison to previous circulating variants. Studies in hamsters showed higher viral shedding of Alpha variant viruses^{53,54}, and it is possible that increased viral load might partly explain the increased rates of transmission between humans as well. Although previously acquired natural or vaccine-induced immunity to SARS-CoV-2 provides protection against severe disease upon infection with the Alpha variant, the possibility that immune escape may explain its rapid spread cannot be excluded as antibody cross-reactivity was variable^{54–56}. Thus, a combination of factors ranging from neutral drift and seeding events to viral shedding patterns, immune escape and increased transmissibility may have contributed to the rapid spread of the Alpha variant around the globe.

In a separate event, another VOC was first detected in South Africa (Beta (B.1.351)). Like the Alpha variant, this variant has undergone an unusually large number of mutations, some of which are shared with the Alpha variant. The Beta variant is characterized by at least

eight nonsynonymous changes in the spike protein, including three that affect key residues in the receptor-binding domain (K417N, E484K and N501Y), which potentially affect receptor binding or antigenicity, or both. As observed with the Alpha variant, Beta variant viruses have rapidly increased in prevalence, with initial modeling suggesting that these viruses have increased transmissibility⁵⁷. In addition, reduced sensitivity to neutralizing antibodies elicited by either natural infection or vaccination was observed for this variant, which is in line with its first emergence in a region with high seroprevalence due to the first pandemic wave^{56,58–63}.

A third highly divergent variant was detected in Japan, traced back to travelers from Brazil⁶⁴. Subsequent analyses by a sequencing consortium in Brazil confirmed circulation of this variant, referred to as the Gamma variant, in a region that had been hit particularly hard earlier in the pandemic²³. The Gamma variant has also been reported to transmit more easily and might be associated with a higher case fatality ratio among young and middle-aged adults⁶⁵.

More recently, a fourth variant emerged in India, the so-called Delta (B.1.617.2) variant, and was declared a VOC. The Delta variant is characterized by L452R, T478K and P681R substitutions in the spike protein, of which P681R is located in the S1–S2 furin cleavage site, which is an essential site enabling the virus to infect target cells⁶⁶. It has been speculated that the specific combination of L452R, E484Q and P681R substitutions may result in increased ACE2 binding and a higher rate of S1–S2 cleavage, which could lead to increased transmissibility of variant viruses, but experimental evidence is lacking⁶⁶. The Delta variant was already identified in December 2020, but has received increased attention recently owing to a rapid surge in COVID-19 cases in India and the United Kingdom caused by this variant since February 2021 (ref. ⁶⁷). Since then, the Delta variant has rapidly spread across different continents and increased spread as compared to the Alpha variant has been observed⁶⁸. Additionally, reduced neutralization was observed after vaccination, although vaccination most likely still protects against severe disease and hospitalization^{69,70}.

VOIs

Next to these VOCs, an expanding list of other variants have been identified that might be associated with phenotypic changes but have not yet been demonstrated to circulate widely and/or negatively affect transmissibility, virulence and immune escape or result in decreased effectiveness of available vaccines, diagnostics and therapeutics. These variants are so-called VOIs and need careful monitoring to determine their possible impact on public health.

VOIs might harbor similar mutations as some of the VOCs and have been found in multiple countries or have caused multiple COVID-19 cases. For example, in December 2020, VOI Eta (B.1.525) was detected both in Nigeria and the United Kingdom. This variant shares mutations with the Alpha variant (deletions at positions 69, 70 and 144 of the spike protein) and has the E484K substitution that is found in the Beta and Gamma variants. This specific substitution is monitored because it has been associated with reduced sensitivity to neutralizing antibodies elicited by natural infection or vaccination⁷¹. Other examples of VOIs that carry the E484K substitution within the receptor-binding domain are the former VOI Zeta (P.2), former VOI Theta (P.3) and VOI Iota (B.1.526) variants that emerged in Brazil, the Philippines and the United States, respectively. VOI Kappa (B.1.617.1), which was identified in India together with VOC Delta, also has a substitution at position 484 in the spike protein but encodes a glutamine at this position, which is also associated with reduced susceptibility to neutralization with convalescent sera⁷². Another VOI was first detected in July 2020 in California and subsequently spread rapidly throughout the United States. This variant, former VOI Epsilon (1.427/1.429), is characterized by a set of substitutions in the spike protein, of which the L452R substitution in the receptor-binding domain is

also thought to increase infectivity, and has the potential to escape antibodies⁷³. A variant circulating widely in South America, VOI Lambda (C.37), which was first identified in Peru in August 2020, encodes a substitution in the receptor-binding domain at position 452 as well. Instead of a change from a lysine to an arginine as seen in VOC Delta and VOIs Epsilon and Kappa, a glutamine residue occupies this potentially important site⁷⁴. An overview of the currently detected VOCs and currently detected and former VOIs and their substitutions in the spike protein as well as the rest of the virally encoded proteins is shown in Fig. 2a,b, but the number of variants is rapidly expanding and their categorization is constantly being updated on the basis of ongoing risk assessments.

Where did these VOIs and VOCs emerge?

All currently recognized variants were first identified in countries with considerable capacity for genomic surveillance, which does not mean that they also first developed in those countries. At the moment, despite the massive surveillance effort where around 0.93% of all SARS-CoV-2-positive cases around the world are sequenced, the origin of these VOCs has not been found. One hypothesis is that accumulation of multiple mutations may occur within a single specific patient, as several case reports have described the identification of mutations shared with the current VOCs. For instance, deletion of amino acids 141–144 and the E484K and N501Y substitutions in the spike protein were observed in an immunocompromised patient who received plasma therapy in Hong Kong⁷⁵. In another case report, deletion of amino acids 141–144 in the spike protein was observed in an immunocompromised patient with cancer⁷⁶, while deletion of positions 69 and 70 in the spike protein, a hallmark of the Alpha variant, was observed in a chronically infected patient⁷⁷.

A second hypothesis is that the virus mutated in an animal reservoir, as SARS-CoV-2 has been shown to be able to infect many different animal species. Large-scale outbreaks of SARS-CoV-2 have for instance been identified in mink farms²². In the Netherlands, only limited spillback to the human population was observed, while in France and Denmark there seemed to be temporal transmission from humans to animals and vice versa^{78,79}. SARS-CoV-2 infection has also been demonstrated in wild mink⁸⁰, making it not unlikely that mink can serve as a reservoir host. Other animal species that have been shown to be susceptible to SARS-CoV-2, some of which can also transmit the virus, are hamsters, ferrets, cats, dogs, lions, deer, monkeys and fruit bats, among others^{81–87}. Of note, newly emerging VOCs may have an extended host range, as the Alpha and Gamma variants have been shown to be able to infect mice⁸⁸. Taken together, these findings demonstrate that SARS-CoV-2 has a wide host range and that the role of animals as reservoir hosts and as a source for the emergence of new variants needs to be investigated.

A third possibility is that a particular virus variant may have evolved gradually in parts of the world where there is less genomic surveillance but widespread circulation. Whereas in some countries a substantial amount of SARS-CoV-2-positive cases have been sequenced, this is not true for all regions of the world. Given that variants with large numbers of mutations have been detected more frequently later in the pandemic and in countries with relatively high seroprevalence due to intense early pandemic waves, it is possible that natural selection of variants with immune escape is occurring during virus circulation in populations with little genomic surveillance.

When does a variant become a VOC?

As genomic monitoring continues to increase in volume, new variants will continue to be detected. A key challenge is to predict and flag VOIs that might be of concern. This requires in-depth knowledge of the genomic profile and possible biological implications of these VOIs. The WHO's working definition of a VOI is that a variant should be phenotypically different or have a genome with mutations

that lead to amino acid changes with established or suspected phenotypic implications. Another requirement is epidemiological evidence of sustained and possibly increased community transmission in one or several countries. The WHO has convened a working group to assess evidence from multiple sources to underpin the assignment of VOIs and VOCs⁸⁹. A SARS-CoV-2 variant is currently classified as a VOC if it has been demonstrated that this variant is associated with an increase in transmissibility, an increase in virulence or changes in clinical disease presentation, or a decrease in the effectiveness of public health and social measures or available diagnostics, vaccines or therapeutics or when a variant is assessed to be a VOC by the WHO in consultation with the WHO SARS-CoV-2 working group⁸⁹.

Future needs in genomics: fast genotype-to-phenotype prediction

Observed changes in epidemiological patterns can be explained by multiple mechanisms not necessarily related to the observed mutations in viral genomes. To draw conclusions about specific variants, epidemiological observations need to be combined with experimental data to assess virus properties, such as infectivity, transmissibility, tropism, virulence and immune escape. To develop a robust knowledge base for monitoring of viral evolution in relation to pandemic preparedness, the rapidly expanding viral genomic sequencing network needs to include reference centers for virus characterization, working toward a suite of standardized assays, reagents, strain collections and serum samples, none of which is currently available for SARS-CoV-2 (ref. ⁹⁰). The devil is in the details. For instance, propagation of the viruses *in vitro* can result in cell culture-adaptive mutations⁹¹. This can be overcome by using specific cell lines or organoids and by developing a consensus mechanism for standardization and auditing of cell lines and protocols; however, such harmonization efforts are challenging and their implementation may take years. Currently, such standardization efforts are not part of molecular surveillance work and still remain to be developed by the field.

A key question is what the priorities are for genotype-to-phenotype prediction, based on lessons learned so far⁹². On the basis of the criteria for assignment of VOCs, it would make sense to focus on virus traits that can provide information about key properties of concern: transmissibility, virulence and immune escape or decreased effectiveness of available vaccines, diagnostics and therapeutics. However, this is a wide scope, and further prioritization may be needed. For instance, the inferred increased transmissibility observed for several VOCs thus far has not translated to fundamental changes in baseline public health strategies⁹³. That could be different if variants emerge for which new traits, such as changes in the age groups predominantly affected or modes of transmission, would warrant updating of interventions on the basis of solid experimental or field data. Arguably, the most urgent question is whether vaccine-derived immunity is affected by variant emergence, for which assessment of both humoral and cellular immune responses will be needed^{94–97}. This assessment has been performed for the Alpha and Beta variants, where it was shown that, although these variants can partially escape humoral immunity, CD4-specific T cell activation was not affected⁹⁷. The turnaround time for such assays, however, has to be improved for informed public health decision-making. Alternatively, reduced neutralization of VOCs may become an important screening assay, as neutralizing antibodies are a likely correlate of protection and neutralization assays can be performed relatively quickly once a high-quality virus isolate has been obtained^{98,99}.

It is likely that SARS-CoV-2 will continue to circulate and evolve and that a system analogous to the global influenza virus surveillance network will be needed^{100–102}. This is a network of national influenza centers and WHO collaborating centers that

Box 1 | Key recommendations

- Genomic analysis networks need to include reference centers for virus characterization.
- Epidemiological observations combined with experimental data are needed to assess virus properties.
- Experiments have to be performed using live virus and standardized for use in rapid risk assessment.
- It will be necessary to work toward a suite of standardized assays, reagents, strain collections and serum samples.

collect data on influenza-like illness trends and provide genetic and antigenic characterization data on a representative selection of viruses circulating in a more or less standardized manner. This information is aggregated globally and is used to decide whether and when the vaccine composition for the next season needs to be adapted. However, whereas the global influenza surveillance system has been largely reactive by selecting newly emerging antigenic variants identified during epidemics to generate vaccines for the next season, high-throughput global virus genome sequencing efforts also allow more forward-looking approaches. When robust assays are developed to quantify immune responses to SARS-CoV-2 after infection and vaccination, such assays may be used to test the effect of all amino acid substitutions observed in global surveillance studies on immune escape, in real time. Examples of such studies are already available for immune escape changes resulting from substitutions in the receptor-binding domain of the spike protein⁷¹, but additional assays can be developed, including assays for other antibody targets in the spike protein (for example, the N-terminal domain) and for T cell immunity. Robust, standardized and validated assays to measure viral immune escape based on genome sequence data and population immunity studies can provide information on vaccine effectiveness against emerging variants. The development of these assays would allow for timely risk assessment and a more immediate response in the case of emergence of VOCs with increased diversity of responses to vaccines.

Other potential indicators of increased public health risk are changes in transmissibility, changes in disease severity, reduced detection by diagnostic assays and reduced susceptibility to drugs and/or therapeutics. For each of these parameters, several assays and study types are available that should be further developed, standardized and assessed for their suitability for use in rapid risk assessment. Examples include the use of panels of viral antigens to screen for potentially reduced sensitivity of widely used rapid tests and the development of well-characterized reference sera for neutralization assays and reference viruses to be used in competition assays for each of the different variants, as mutations will continue to occur in VOCs and VOIs after their initial detection. Given the fact that SARS-CoV-2 may rapidly acquire mutations mapping to the spike protein upon inoculation in animals, human organoids potentially represent a powerful tool to further characterize SARS-CoV-2 variants. Recent studies, for example, have indicated that the Alpha variant, in comparison to an ancestral SARS-CoV-2 clade B virus, produces higher levels of infectious virus late in infection and has higher replicative fitness in human airway, alveolar and intestinal organoid models¹⁰³. These assays should also be performed in a timely fashion because the increasing volume of sequencing data otherwise has the potential to become a burden instead of a valuable source of information⁹⁰.

Whenever possible, this work should be conducted without the use of full-length infectious SARS-CoV-2, for example, by using virus pseudotypes. However, some phenotypes, such as virulence and transmission, cannot be investigated without infectious

viruses. Conclusive evidence for other experiments (for example, assays for immune escape) may also require use of infectious viruses. These experiments, conducted under the appropriate biosafety level 3 conditions, are crucial to keep intervention strategies up to date in the interest of public health and animal health. Key recommendations are summarized in Box 1. In combination, globally representative genomic surveillance linked with experimental data to validate signals from genomic data will provide a critical step forward in surveillance of current and potential pandemic threats¹⁰⁴.

Received: 26 May 2021; Accepted: 20 July 2021;
Published online: 9 September 2021

References

- Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
- Gorbalenya, A. E. et al. The species *Severe acute respiratory syndrome-related coronavirus*: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* **5**, 536–544 (2020).
- World Health Organization. *WHO-Convened Global Study of Origins of SARS-CoV-2: China Part* (WHO, 2021); <https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part>
- Han, G. Z. Pangolins harbor SARS-CoV-2-related coronaviruses. *Trends Microbiol.* **28**, 515–517 (2020).
- Lam, T. T. Y. et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* **583**, 282–285 (2020).
- Boni, M. F. et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat. Microbiol.* **5**, 1408–1417 (2020).
- Corman, V. M. et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT–PCR. *Eurosurveillance* **25**, 2000045 (2020).
- Oude Munnink, B. B. et al. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat. Med.* **26**, 1405–1410 (2020).
- World Health Organization. *COVID-19 Vaccine Tracker and Landscape* (WHO, 2021); <https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines>
- Quick, J. et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).
- Adelino, T. É. R. et al. Field and classroom initiatives for portable sequence-based monitoring of dengue virus in Brazil. *Nat. Commun.* **12**, 2296 (2021).
- Faria, N. R. et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* **546**, 406–410 (2017).
- Grubaugh, N. D., Faria, N. R., Andersen, K. G. & Pybus, O. G. Genomic insights into Zika virus emergence and spread. *Cell* **172**, 1160–1162 (2018).
- Faria, N. R. et al. Zika virus in the Americas: early epidemiological and genetic findings. *Science* **352**, 345–349 (2016).
- Shu, Y. & McCauley, J. GISAID: Global Initiative on Sharing All Influenza Data—from vision to reality. *Eurosurveillance* **22**, 30494 (2017).
- Wu, S. L. et al. Substantial underestimation of SARS-CoV-2 infection in the United States. *Nat. Commun.* **11**, 4507 (2020).
- Mohanani, M., Malani, A., Krishnan, K. & Acharya, A. Prevalence of SARS-CoV-2 in Karnataka, India. *JAMA* **325**, 1001–1003 (2021).
- Banerjee, A., Doxey, A. C., Mossman, K. & Irving, A. T. Unraveling the zoonotic origin and transmission of SARS-CoV-2. *Trends Ecol. Evol.* **36**, 180–184 (2021).
- Meredith, L. W. et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect. Dis.* **20**, 1263–1272 (2020).
- Ladhani, S. N. et al. Investigation of SARS-CoV-2 outbreaks in six care homes in London, April 2020. *EClinicalMedicine* **26**, 100533 (2020).
- Ismail, S. A., Saliba, V., Lopez Bernal, J., Ramsay, M. E. & Ladhani, S. N. SARS-CoV-2 infection and transmission in educational settings: a prospective, cross-sectional analysis of infection clusters and outbreaks in England. *Lancet Infect. Dis.* **21**, 344–353 (2021).
- Oude Munnink, B. B. et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* **371**, 172–177 (2020).
- Faria, N. R. et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **371**, 815–821 (2021).
- Wibmer, C. K. et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* **27**, 622–625 (2021).
- Frampton, D. et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *Lancet Infect. Dis.* [https://doi.org/10.1016/S1473-3099\(21\)00170-5](https://doi.org/10.1016/S1473-3099(21)00170-5) (2021).
- Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1) (2020).
- Domingo, E. & Holland, J. J. RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* **51**, 151–178 (1997).
- Singer, J. B., Gifford, R. J., Cotten, M. & Robertson, D. L. CoV-GLUE: a web application for tracking SARS-CoV-2 genomic variation. *Preprint at* <https://doi.org/10.20944/preprints202006.0225.v1> (2020).
- Volz, E. M. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* **184**, 64–75 (2021).
- Lu, J. et al. Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* **181**, 997–1003 (2020).
- Candido, D. S. et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* **369**, 1255–1260 (2020).
- Alm, E. et al. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Eurosurveillance* **25**, 2001410 (2020).
- du Plessis, L. et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* **371**, 708–712 (2021).
- Giandhari, J. et al. Early transmission of SARS-CoV-2 in South Africa: an epidemiological and phylogenetic report. *IJID* **103**, 234–241 (2021).
- Hufsky, F. et al. Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research. *Brief. Bioinform.* **22**, 642–663 (2021).
- Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).
- Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
- Callaway, E. 'A bloody mess': confusion reigns over naming of new COVID variants. *Nature* **589**, 339 (2021).
- Konings, F. et al. SARS-CoV-2 variants of interest and concern naming scheme conducive for global discourse. *Nat. Microbiol.* **6**, 821–823 (2021).
- Peacock, T. P., Penrice-Randal, R., Hiscox, J. A. & Barclay, W. S. SARS-CoV-2 one year on: evidence for ongoing viral adaptation. *J. Gen. Virol.* **102**, 001584 (2021).
- Volz, E. et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* **184**, 64–75 (2021).
- Böhmer, M. M. et al. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *Lancet Infect. Dis.* **20**, 920–928 (2020).
- Korber, B. et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**, 812–827 (2020).
- Yurkovetskiy, L. et al. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell* **183**, 739–751 (2020).
- Plante, J. A. et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).
- Hou, Y. J. et al. SARS-CoV-2 D614G variant exhibits efficient replication *ex vivo* and transmission *in vivo*. *Science* **370**, 1464–1468 (2021).
- Tegally, H. et al. Emergence of a SARS-CoV-2 variant of concern with mutations in spike glycoprotein. *Nature* **592**, 438 (2021).
- Public Health England. Investigation of novel SARS-CoV-2 variant Variant of Concern 202012/01—technical briefing. <https://www.gov.uk/government/publications/phe-investigation-of-novel-sars-cov-2-variant-of-concern-20201201-technical-briefing-3-6-january-2021> (6 January 2021).
- Vogels, C. B. F. et al. Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. *PLoS Biol.* **19**, e3001236 (2021).
- Volz, E. et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021).
- New and Emerging Respiratory Virus Threats Advisory Group. NERVTAG/SPI-M extraordinary meeting on SARS-CoV-2 variant of concern 202012/01 (variant B.1.1.7): note of meeting. <https://app.box.com/s/3lkcbxepqixkg4mv640dpvvg978ixjtf/file/756964987830> (2020).
- Davies, N. G. et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021).
- Mohandas, S. et al. Comparison of the pathogenicity and virus shedding of SARS-CoV-2 VOC 202012/01 and D614G variant in hamster model. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.02.25.432136> (2021).
- Núñez, I. A. et al. SARS-CoV-2 B.1.1.7 infection of Syrian hamster does not cause more severe disease, and naturally acquired immunity confers protection. *mSphere* **6** (2021).
- Fischer, R. J. et al. ChAdOx1 nCoV-19 (AZD1222) protects hamsters against SARS-CoV-2 B.1.351 and B.1.1.7 disease. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.11.435000> (2021).
- Wang, P. et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130–135 (2021).
- Pearson, C. A. B. et al. Estimates of severity and transmissibility of novel SARS-CoV-2 variant 501Y.V2 in South Africa. Preprint at *CMMID* <https://cmmid.github.io/topics/covid19/sa-novel-variant.html> (2021).

58. Li, R. et al. Differential efficiencies to neutralize the novel mutants B.1.1.7 and 501Y.V2 by collected sera from convalescent COVID-19 patients and RBD nanoparticle-vaccinated rhesus macaques. *Cell. Mol. Immunol.* **18**, 1058–1060 (2021).
59. Tada, T. et al. Convalescent-phase sera and vaccine-elicited antibodies largely maintain neutralizing titer against global SARS-CoV-2 variant spike. *mBio.* **12** (2021).
60. Hoffmann, M. et al. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. *Cell* **9**, 2384–2393 (2021).
61. Planas, D. et al. Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. *Nat. Med.* **27**, 917–924 (2021).
62. Garcia-Beltran, W. F. et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372–2383 (2021).
63. Zhou, D. et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* **184**, 2348–2361 (2021).
64. National Institute of Infectious Diseases, Japan. Brief report: new variant strain of SARS-CoV-2 identified in travelers from Brazil. <https://www.niid.go.jp/niid/en/2019-ncov-e/10108-covid19-33-en.html> (National Institute of Infectious Diseases, Japan, 12 January 2021).
65. Mendes Coutinho, R. et al. Model-based estimation of transmissibility and reinfection of SARS-CoV-2 P.1 variant. Preprint at *medRxiv* <https://doi.org/10.1101/2021.03.03.21252706> (2021).
66. Cherian, S. et al. Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.04.22.440932> (2021).
67. Public Health England. SARS-CoV-2 variants of concern and variants under investigation in England: technical briefing. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993879/Variants_of_Concern_VOC_Technical_Briefing_15.pdf (11 June 2021).
68. Campbell, F. et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Eurosurveillance* **26**, 2100509 (2021).
69. Davis, C. et al. Reduced neutralisation of the Delta (B.1.617.2) SARS-CoV-2 variant of concern following vaccination. Preprint at *medRxiv* <https://doi.org/10.1101/2021.06.23.21259327> (2021).
70. Planas, D. et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* <https://doi.org/10.1038/s41586-021-03777-9> (2021).
71. Greaney, A. J. et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* **29**, 44–57 (2021).
72. Greaney, A. J. et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463–476 (2021).
73. McCallum, M. et al. SARS-CoV-2 immune evasion by variant B.1.427/B.1.429. *Science* **373**, 648–654 (2021).
74. Romero, P. E. et al. The emergence of SARS-CoV-2 variant Lambda (C.37) in South America. Preprint at *medRxiv* <https://doi.org/10.1101/2021.06.26.21259487> (2021).
75. Choi, B. et al. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *N. Engl. J. Med.* **383**, 2291–2293 (2020).
76. Avanzato, V. A. et al. Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer. *Cell* **183**, 1901–1912 (2020).
77. Kemp, S. A. et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **592**, 277–282 (2021).
78. Fournier, P.-E. et al. Emergence and outcomes of the SARS-CoV-2 ‘Marseille-4’ variant. *Int. J. Infect. Dis.* **106**, 228–236 (2021).
79. Boklund, A. et al. SARS-CoV-2 in Danish mink farms: course of the epidemic and a descriptive analysis of the outbreaks in 2020. *Animals* **11**, 164 (2021).
80. ProMED. COVID-19 update (536): animal, USA (UT) wild mink, 1st case. <https://promedmail.org/promed-post/?id=8015608> (13 December 2020).
81. Haagmans, B. L. et al. SARS-CoV-2 neutralizing human antibodies protect against lower respiratory tract disease in a hamster model. *J. Infectious Diseases* **223**, 2020–2028 (2021).
82. Shi, J. et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* **368**, 1016–1020 (2020).
83. Halfmann, P. J. et al. Transmission of SARS-CoV-2 in domestic cats. *N. Engl. J. Med.* **383**, 592–594 (2020).
84. Rockx, B. et al. Comparative pathogenesis of COVID-19, MERS, and SARS in a nonhuman primate model. *Science* **368**, 1012–1015 (2020).
85. Richard, M. et al. SARS-CoV-2 is transmitted via contact and via the air between ferrets. *Nat. Commun.* **11**, 3496 (2020).
86. Sit, T. H. C. et al. Infection of dogs with SARS-CoV-2. *Nature* **586**, 776–778 (2020).
87. Sailleau, C. et al. First detection and genome sequencing of SARS-CoV-2 in an infected cat in France. *Transbound. Emerg. Dis.* **67**, 2324–2328 (2020).
88. Montagutelli, X. et al. The B.1.351 and P.1 variants extend SARS-CoV-2 host range to mice. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.18.436013> (2021).
89. WHO. Coronavirus disease (COVID-19): virus evolution. <https://www.who.int/news-room/q-a-detail/sars-cov-2-evolution> (30 December 2020).
90. Bedford, J. et al. A new twenty-first century science for effective epidemic response. *Nature* **575**, 130–136 (2019).
91. Lamers, M. M. et al. Human airway cells prevent SARS-CoV-2 multibasic cleavage site cell culture adaptation. *eLife* **10**, e66815 (2021).
92. Bakhshandeh, B. et al. Mutations in SARS-CoV-2: consequences in structure, function, and pathogenicity of the virus. *Microb. Pathog.* **154**, 104831 (2021).
93. Grubaugh, N. D., Hodcroft, E. B., Fauver, J. R., Phelan, A. L. & Cevik, M. Public health actions to control new SARS-CoV-2 variants. *Cell* **184**, 1127–1132 (2021).
94. Weisblum, Y. et al. Escape from neutralizing antibodies 1 by SARS-CoV-2 spike protein variants. *eLife* **9**, e61312 (2020).
95. Weiskopf, D. et al. Phenotype and kinetics of SARS-CoV-2-specific T cells in COVID-19 patients with acute respiratory distress syndrome. *Sci. Immunol.* **5**, 2071 (2020).
96. Liu, Z. et al. Landscape analysis of escape variants identifies SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.11.06.372037> (2020).
97. Geers, D. et al. SARS-CoV-2 variants of concern partially escape humoral but not T-cell responses in COVID-19 convalescent donors and vaccinees. *Sci. Immunol.* **6**, eabj1750 (2021).
98. Earle, K. A. et al. Evidence for antibody as a protective correlate for COVID-19 vaccines. *Vaccine* **39**, 4423–4428 (2021).
99. Khoury, D. S. et al. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat. Med.* **27**, 1205–1211 (2021).
100. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868 (2020).
101. Petrova, V. N. & Russell, C. A. The evolution of seasonal influenza viruses. *Nat. Rev. Microbiol.* **16**, 47–60 (2018).
102. Ampofo, W. K. et al. Strengthening the influenza vaccine virus selection and development process. Report of the 3rd WHO Informal Consultation for Improving Influenza Vaccine Virus Selection held at WHO headquarters, Geneva, Switzerland, 1–3 April 2014. *Vaccine* **33**, 4368–4382 (2015).
103. Lamers, M. M. et al. Human organoid systems reveal in vitro correlates of fitness for SARS-CoV-2. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.05.03.441080> (2021).
104. Gardy, J. L. & Loman, N. J. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* **19**, 9–20 (2018).

Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 874735 (VEO) and from ZonMW (grant agreement no. 10150062010005). The authors acknowledge the technical collaboration and financial support of the Health Emergencies Programme of the WHO.

Author contributions

B.B.O.M. and M.K. drafted the initial manuscript. N.W., D.F.N., R.S.S., B.H. and R.A.M.F. critically read and contributed to the manuscript. N.W. and D.F.N. produced the figures.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Marion Koopmans.

Peer review information *Nature Medicine* thanks Estée Török and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Karen O'Leary was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature America, Inc. 2021, corrected publication 2021