vaccines

MDPI

Perspective

On the Origins of Omicron's Unique Spike Gene Insertion

A. J. Venkatakrishnan [1], **Praveen Anand** [2], **Patrick J. Lenehan** [1], **Rohit Suratekar** [2], **Bharathwaj Raghunathan** [3], **Michiel J. M. Niesen** [1] **and Venky Soundararajan** [1,2,3,*]

[1] nference, Cambridge, MA 02139, USA
[2] nference Labs, Bengaluru 560017, Karnataka, India
[3] nference, Toronto, ON M5V 1M1, Canada
* Correspondence: venky@nference.net

Abstract: The emergence of a heavily mutated SARS-CoV-2 variant (Omicron; Pango lineage B.1.1.529 and BA sublineages) and its rapid spread to over 75 countries raised a global public health alarm. Characterizing the mutational profile of Omicron is necessary to interpret its clinical phenotypes which are shared with or distinctive from those of other SARS-CoV-2 variants. We compared the mutations of the initially circulating Omicron variant (now known as BA.1) with prior variants of concern (Alpha, Beta, Gamma, and Delta), variants of interest (Lambda, Mu, Eta, Iota, and Kappa), and ~1500 SARS-CoV-2 lineages constituting ~5.8 million SARS-CoV-2 genomes. Omicron's Spike protein harbors 26 amino acid mutations (23 substitutions, 2 deletions, and 1 insertion) that are distinct compared to other variants of concern. While the substitution and deletion mutations appeared in previous SARS-CoV-2 lineages, the insertion mutation (ins214EPE) was not previously observed in any other SARS-CoV-2 lineage. Here, we consider and discuss various mechanisms through which the nucleotide sequence encoding for ins214EPE could have been acquired, including local duplication, polymerase slippage, and template switching. Although we are not able to definitively determine the mechanism, we highlight the plausibility of template switching. Analysis of the homology of the inserted nucleotide sequence and flanking regions suggests that this template-switching event could have involved the genomes of SARS-CoV-2 variants (e.g., the B.1.1 strain), other human coronaviruses that infect the same host cells as SARS-CoV-2 (e.g., HCoV-OC43 or HCoV-229E), or a human transcript expressed in a host cell that was infected by the Omicron precursor.

Keywords: COVID-19; Omicron; template switching

1. Introduction

A new SARS-CoV-2 variant with an extensively mutated Spike protein was first reported to the World Health Organization (WHO) from South Africa on 24 November 2021, with the first sample collected on 9 November 2021. This strain was subsequently denoted as the Omicron variant (WHO nomenclature) and B.1.1.529 (Pango lineage) [1]. The rapid assessment of the variant by *The Technical Advisory Group on SARS-CoV-2 Virus Evolution* and classification of Omicron as a variant of concern by the WHO within 48 h facilitated timely epidemiological surveillance. After its initial discovery, this variant rapidly spread across the globe and was detected in over 75 countries across 6 continents by 16 December 2021 [2,3]. After this, multiple Omicron sublineages emerged (Pango lineages BA.1, BA.2, BA.3, BA.4, BA.5, and descendants thereof) and drove case surges around the world.

Thoroughly characterizing the mutational profile of Omicron is a necessary step to interpret its shared or distinctive clinical phenotypes with respect to other variants, its sensitivity or resistance to existing vaccines, and whether Omicron-like variants that evolve in the future may have heightened virulence. Indeed, SARS-CoV-2 has evolved into different variants of concern and variants of interest through a combination of missense, deletion, and insertion mutations. For example, the D614G substitution in the Spike (S) protein,

which emerged early and has been detected in nearly all SARS-CoV-2 genomes in GISAID since mid-2020, increases the replication capacity and infectivity of SARS-CoV-2 [4,5]. Other substitutions (e.g., E484K and E484A) have led to significant changes in the Spike–ACE2 binding affinity, and deletions (e.g., ΔY144) have modulated the effects of neutralizing anti-Spike antibodies [6–13]. Insertion mutations have been less prevalent in the evolution of SARS-CoV-2 [14]. However, one of the most functionally consequential mutations in the evolutionary history of SARS-CoV-2 to date was the "PRRA" Spike protein insertion in the S1/S2 cleavage site, which introduced the polybasic furin cleavage site that mimics the RRARSVAS peptide in human ENaC-alpha [15–19]. This insertion plays an important role in the transmission of SARS-CoV-2, at least in part by facilitating an endosome-independent entry pathway into respiratory epithelial cells that bypasses important innate antiviral responses [18]. It is also mechanistically required for the syncytium-mediated death of lymphocytes, which may contribute to the lymphopenia that is often clinically observed in COVID-19 patients [20–23]. The availability of 5.8 million SARS-CoV-2 genomes covering ~1500 lineages from over 203 countries/territories in the GISAID database since the beginning of the pandemic provides an opportunity to characterize the genomic landscape of the Omicron variant in comparison to other SARS-CoV-2 variants.

In this study, we compare the mutational profiles of early Omicron genomes (primarily sublineage BA.1, hereafter referred to as "Omicron") with all other SARS-CoV-2 lineages, including the variants of concern and variants of interest. We highlight that Omicron's Spike protein harbored an insertion mutation ins214EPE that was absent in all other SARS-CoV-2 lineages at the time of its emergence. Given the salience of viral genetic recombination and the debated plausibility of host genome integration by SARS-CoV-2 [24–26], we considered a variety of host–viral and interviral genomic matter exchange scenarios that may have contributed to the adoption of this insertion mutation in the precursor variant of Omicron. We discuss potential sources for the origin of ins214EPE and highlight the need to experimentally characterize the role of ins214EPE in viral transmission and immune evasion.

2. Methods

2.1. Analysis of Mutations Defining the Omicron Lineage

Core mutations were derived for parental lineages from the Coronavirus Antiviral Research Database (CoV-RDB; covdb.standford.edu; accessed 10 December 2021) for each variant of interest or variant of concern: Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), Lambda (C.37), Mu (B.1.621), Eta (B.1.525), Iota (B.1.526), Kappa (B.1.617.1), and Omicron (BA.1) [27]. All SARS-CoV-2 genome sequences corresponding to the Omicron variant were directly derived from the GISAID database [2,3]. There were 1448 SARS-CoV-2 genomes annotated as B.1.1.529/BA.1/BA.2 as of 13 December 2021. These genomes were compared to other (non-Omicron) genomes deposited in GISAID before the same date to determine the number of lineages in which each Omicron Spike protein mutation was previously observed.

To understand which SARS-CoV-2 mutations correlate with COVID-19 test positivity, we identified "surge-associated mutations" using genomic data from GISAID (5,781,715 sequences from 203 countries/territories between December 2019 and December 2021) and epidemiology data from Our World in Data (OWID) [28], as described in our previous study [29]. A mutation is considered to be "surge associated" if it satisfies the following criteria: (1) it is present in at least 100 SARS-CoV-2 sequences in GISAID; and (2) in a period of three consecutive months during which there was a monotonic increase in PCR positivity by at least 5% in a given country, the prevalence of the mutation in that country also monotonically increased by at least 5%.

To assess the prevalence of ins214EPE in each Omicron lineage (B.1.1.529 and all BA sublineages), SARS-CoV-2 genomes (n = 11,945,950) were collected from GISAID on 18 July 2022 along with their annotated metadata, including lineage assignment and mutational profiles. To identify other insertions in this region that were observed prior to the emergence

of Omicron, we filtered these 11,945,950 genomes to those meeting the following criteria: (i) high coverage, (ii) complete sequence, (iii) collection date on or before 13 December 2020 (i.e., the data cutoff date for our initial analysis of Omicron mutations), (iv) not assigned to the Omicron lineage (B.1.1.529 or a BA sublineage), (v) containing an insertion between positions 210 and 218 of the Spike protein, (vi) the inserted sequence does not contain a stop codon or a low-confidence alignment (i.e., amino acid indicated as "X"), and (vii) the insertion is not ins214EPE.

2.2. Nucleotide 9-mer Search to Identify Candidate Viral and Human Templates for ins214EPE

An exact 9-mer search for all three possible inserts (5′-AGCCAGAAG-3′, 5′-GAGCCAGAA-3′, and 5′-GCCAGAAGA-3′) and their reverse-complement sequences was performed across three different databases: (i) the human transcriptome, (ii) SARS-CoV-2 genomes from GISAID, and (iii) human-infecting *Coronaviridae* family viruses. For a reference Omicron genome (EPI_ISL_6640916) [30], we also performed a modified search to identify all nine-nucleotide sequences that differed from the possible insertion sequences by only one nucleotide (i.e., allowing for a single mismatch). The human transcript sequences (n = 244,939) were downloaded from the GENCODE database [31] (version 39; GRCh38.p13 of the human genome). Coding sequences (CDSs) for 5,781,715 SARS-CoV-2 genomes were accessed from GISAID [2] on 13 December 2021. All available sequences from human-infecting *Coronaviridae* family (taxid:11118) viruses were accessed from the NCBI virus database (https://www.ncbi.nlm.nih.gov/labs/virus, accessed on 13 December 2021; n = 574,178 complete genomes with 4,028,478 CDSs) [32]. In the final analysis, we excluded genomes that were labeled as SARS-CoV, SARS-CoV-2, or Middle East respiratory syndrome (MERS)-CoV (or otherwise named strains of these viruses) so as to consider only seasonal and enteric coronavirus strains—human coronavirus (HCoV)-229E, HCoV-OC43, HCoV-NL63, and human enteric coronavirus strain 4408.

2.3. Assessment of Homology between Regions Flanking Insertion and Origin Sites

For a given insertion, we defined a putative origin as the matched sequence in a viral genome, a viral anti-genome, or a human transcript. We then assessed the homology between the 35 nucleotides upstream of the insertion and the 35 nucleotides upstream of the origin, and similarly we assessed the homology between the 35 nucleotides downstream of the insertion and the 35 nucleotides downstream of the origin. To assess the similarity between any given pair of nucleotide sequences, we defined a function using the Bio.pairwise2 module from Biopython v1.76 (https://biopython.org/wiki/Documentation accessed on 10 January 2022), which performs a global alignment of nucleotide sequences using a custom scoring scheme (i.e., +5 for a match, −4 for a mismatch, 0 for gap start and extension). This score ranges from 0 (no matches) to 175 (perfect match for all 35 nucleotides). We further defined a normalized homology score (NHS), ranging from 0 to 100, which is calculated by simply dividing the homology score by 175 and multiplying the result by 100. We also applied a similar protocol to assess the homology between shorter upstream and downstream sequences (7 nucleotides), where the NHS was calculated by dividing the homology score by 35 and multiplying the result by 100.

A set of "positive control" template-switch-mediated insertions was obtained from Supplementary Table S4 from the prior analysis by Garushyants et al. [14]. We filtered this table to include only those insertions that were assigned a mechanism of "Template switch" and which were 12 or more nucleotides long. This set of insertions is shown in Table S5. For each insertion, the genomic positions of the origin and insertion sequences are provided in this table, along with the Pango lineage assigned to the genome(s) harboring the insertion. We searched SARS-CoV-2 genomes in GISAID to identify those genomes that contained the provided origin and insertion sequences at approximately the provided genomic positions. The GISAID identifiers for the corresponding insertion-containing genomes are provided in Table S5. For each of these genomes, we then obtained the 35 nucleotide sequences upstream and downstream of the insertion and origin. Finally, we calculated the NHS for

each relevant pair of sequences: (i) the 35 nucleotides upstream of the insertion versus the 35 nucleotides upstream of the origin, and (ii) the 35 nucleotides downstream of the insertion versus the 35 nucleotides downstream of the origin. These scores, and the local alignments contributing to them, are shown for each individual genome in Table S11.

As a "negative control" analysis, we calculated the NHSs between 10,000 pairs of randomly selected non-overlapping *n*-mers (35-mers or 7-mers) from the original SARS-CoV-2 genome (NC_0545512.2; https://www.ncbi.nlm.nih.gov/nuccore/NC_045512, accessed on 10 January 2022). To generate each pair of *n*-mers, we randomly selected two genomic coordinates (i.e., nucleotide positions between 1 and 29,903) as the starting positions for an *n*-mer nucleotide sequence. If the two positions were within *n* nucleotides of one another, a new pair of coordinates was selected because these would, by definition, generate overlapping *n*-mers. Similarly, if either position was less than *n* nucleotides from the end of the SARS-CoV-2 genome sequence, then a new pair of coordinates was selected.

2.4. Single-Cell Analysis of Coronavirus Receptor Co-Expression

Publicly available single-cell RNA sequencing datasets were obtained and processed as described previously [33,34]. The data are hosted at https://academia.nferx.com/dv/202011/singlecell/, accessed on 10 January 2022, and include approximately 2.8 million cells derived from dozens of independent studies covering most major human tissues. We determined the total numbers and percentages of cells (of these 2.8 million total cells) expressing each gene encoding a coronavirus receptor of interest (e.g., ACE2, ANPEP, DPP4), along with the numbers and percentages of cells co-expressing ACE2 and ANPEP or ACE2 and DPP4. We also performed similar analyses on two specific studies of interest: (i) a study of nasopharyngeal and bronchial samples from COVID-19 patients and healthy controls (approximately 33,000 cells) [35], and (ii) a study of ileal biopsies from Crohn's disease patients (approximately 136,000 cells) [36]. For these studies, we also evaluated co-expression in the cell types that showed the strongest expression of these genes. To evaluate whether the degree of co-expression for ACE2 and ANPEP or DPP4 was more than expected by chance, we calculated the observed-to-expected ratio of co-expression, assuming that the expression of each gene is distributed randomly across the analyzed cells. Specifically, this value was calculated by dividing the co-expressing percentage by the product of the individual expression percentages and multiplying the result by 100.

3. Results

3.1. Comparison of Mutations in Omicron to Previous SARS-CoV-2 Lineages Shows the Presence of a Unique Insertion Mutation in Omicron's Spike Protein

Omicron harbors 37 mutations in the Spike protein, which include 6 deletion mutations, 1 insertion mutation, and 30 substitution mutations [27]. Of the 37 mutations, 16 were associated with regional case surges prior to the Omicron era (Table S1; see Methods) [29]. Comparing these Spike protein mutations in Omicron with pre-existing variants of concern (VOCs: Alpha, Beta, Gamma, and Delta) shows that 26 mutations are distinct to Omicron, while the remaining 11 are shared with at least one other VOC (Figure 1A, Table S1). When compared to the Delta variant, the mutational load of Omicron is particularly high in the Spike protein sequence, with more similar rates of mutation in regions of the viral genome encoding other proteins (Figure S1).

We next analyzed which of these 26 mutations (i.e., Spike mutations present in Omicron but not in any prior VOCs) appeared in the prior variants of interest (VOIs: Lambda, Mu, Eta, Iota, and Kappa) or other prior SARS-CoV-2 lineages by comparing them with mutations from 5,781,715 genomes corresponding to ~1500 lineages from the GISAID database. Two of these mutations were present in VOIs (A67V in Eta, T95I in Mu and Iota) (Table S2), while twenty-three others appeared in previously collected genomes assigned to SARS-CoV-2 lineages that were not classified as VOIs or VOCs (Figure 1B, Table S2). Interestingly, only the insertion mutation ins214EPE had not been previously observed in any SARS-CoV-2 lineages (Figure 1B, Table S2), although other insertions at or near

this position had been observed in several lineages before the emergence of Omicron (Table S3) [14,37]. Specifically, of the 1168 SARS-CoV-2 genomes in GISAID harboring this insertion at the time of this analysis, 1164 were classified as Omicron. Of the remaining four genomes, three had not yet been assigned a Pango lineage, while the other one (which was deposited on 24 November 2021) was labeled as B.1 (a parent lineage of Omicron).
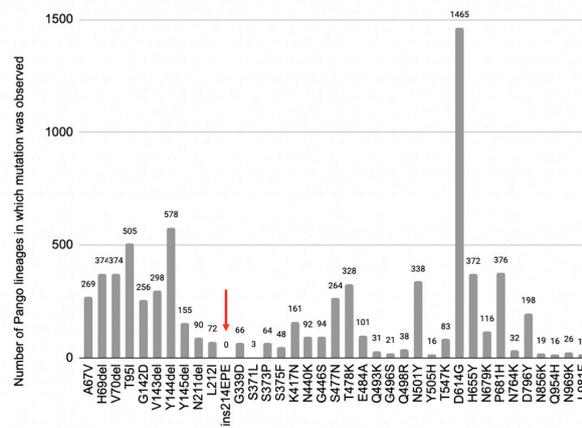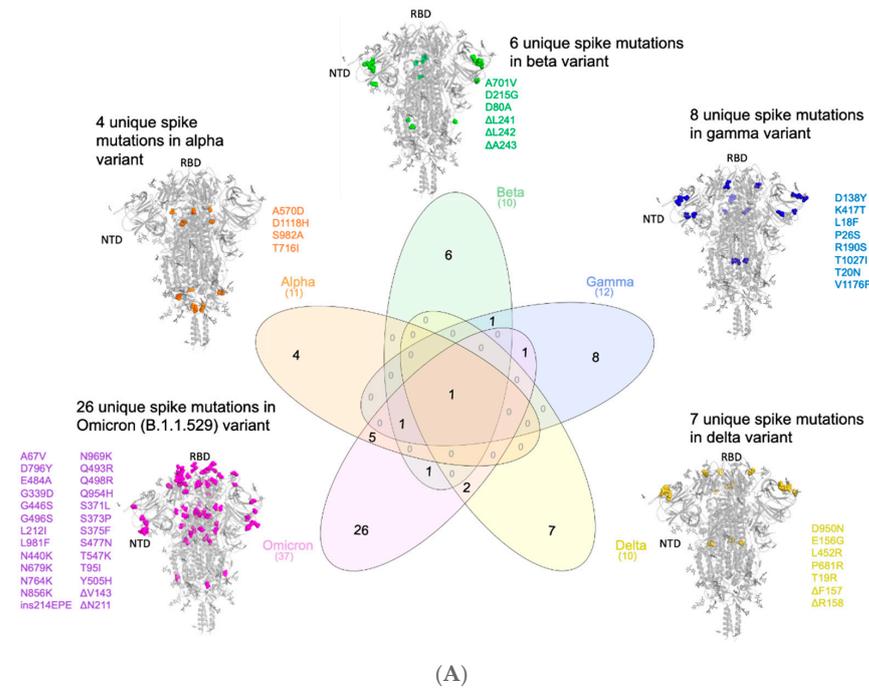

(A)

(B)

Figure 1. Specificity of Omicron's Spike mutations relative to SARS-CoV-2 variants of concern and other lineages. (**A**) Comparing the lineage-specific Spike protein mutations in the SARS-CoV-2 variants of concern. The unique mutations observed in the Spike protein for each of the variants are highlighted (spheres) on the homotrimeric Spike protein of SARS-CoV-2. The Omicron (B.1.1.529/BA.1/BA.2) variant has the highest number (26) of unique mutations in the Spike protein from this perspective, making its emergence a "step function" in the evolution of SARS-CoV-2 strains. (**B**) Prevalence of Omicron's Spike mutations in other SARS-CoV-2 lineages. Bar plot denoting the number of SARS-CoV-2 lineages (besides Omicron) in which the mutations present in Omicron are observed. The red arrow highlights the insEPE214 mutation, which is absent from all other SARS-CoV-2 lineages.

The EPE insertion (ins214EPE) on Omicron maps to the Spike protein's N-terminal domain (NTD) distal from the antibody-binding supersite [11]. However, the loop where the insertion is present maps to a known human T-cell epitope on SARS-CoV-2 [38]. Further studies will be necessary to understand whether this insertion may help SARS-CoV-2 to escape T-cell immunity [14]. A recent study also suggests that insertions in the NTD, including ins214EPE in Omicron, may increase viral transmissibility by enhancing sialic acid receptor binding [39]. Such a functional consequence may be consistent with the observation that multiple SARS-CoV-2 lineages have acquired insertions of 1–6 additional amino acids at or near this same site (i.e., between positions 210 and 218) (Table S3) [37]. Given the potential for insertions to impact SARS-CoV-2's virulence (e.g., the PRRA insertion giving rise to a polybasic furin cleavage site in the original SARS-CoV-2 strain), it is important to understand the functional significance and evolutionary origins of ins214EPE in the Omicron variant [15–17,19].

It should be noted that BA.1 was the predominant Omicron sublineage at the time of these initial characterizations in December 2021. Multiple other sublineages (BA.2, BA.3, BA.4, and BA.5) sequentially emerged, and became the dominant (i.e., most prevalent) strains of SARS-CoV-2. While ins214EPE is present in over 75% of BA.1 sequences collected in GISAID as of July 2022, it is almost uniformly absent from all of the subsequent sublineages (Table S4). In the remainder of this perspective, we consider possible mechanisms that could have led to the generation of this insertion, which was a defining feature of the earliest circulating Omicron variant, BA.1.

3.2. Template Switching Is a Plausible Mechanism for the Origin of ins214EPE in Omicron

Ins214EPE is an in-frame insertion of nine nucleotides that occurs between nucleotide positions 22204 and 22207. Based on the local sequence alignment, there are three candidate insertions: (1) 5′-GAGCCAGAA-3′ between 22204 and 22205, (2) 5′-AGCCAGAAG-3′ between 22205 and 22206, and (3) 5′-AGCCAGAAG-3′ between positions 22206 and 22207 (Figure S2). According to a recent analysis of the secondary structure of the reference SARS-CoV-2 genome, the nucleotides between positions 22205 and 22208 (5′-GAUC-3′) constitute an RNA loop [40], which is notable given that RNA loops appear to be more prone to insertions than stems.

A previous analysis of sequences deposited in GISAID concluded that insertions in the SARS-CoV-2 genome most likely arise from one of three mechanisms: (1) local duplications, (2) polymerase slippage, or (3) template switching [14]. While local duplications were found to explain some short (fewer than nine nucleotides) and long (nine or more nucleotides) insertions, the observation that these groups had distinct nucleotide compositions, phyletic patterns, and genomic localizations led to the conclusion that short and long insertions typically arise from distinct mechanisms [14]. Specifically, it was suggested that polymerase slippage is the best explanation for most short insertions, which typically have an excess of uracil nucleotides and are non-monophyletic. It is proposed that slippage is most likely to occur during runs of uracils, owing to the slow processing of polyU tracts that has been demonstrated for the RNA-dependent RNA polymerase (RdRp) of SARS-CoV-1 and is hypothesized to be true for the RdRp of SARS-CoV-2 as well [14,41]. On the other hand, it was posited that long insertions typically arise from template switching, given that their nucleotide composition is consistent with the SARS-CoV-2 genome, they are typically monophyletic, and they tend to occur at or near sites that were previously identified as hotspots for template switching [14,42]. It is notable that template switching is a normal part of the life cycle for *Coronaviridae*, as discontinuous transcription via template switching is responsible for the synthesis of subgenomic RNAs (sgRNAs) [43,44]. In this light, we asked which mechanism is the most fitting explanation for the origin of ins214EPE in Omicron.

As mentioned above, exact duplication of an adjacent nucleotide sequence has been observed previously as a mechanism for both short and long insertions in the SARS-CoV-2 genome [14]. For example, in three GISAID sequences of the B.1.429 lineage, there is a duplication of 24 nucleotides (5′-AAAAGAAGAAGGCTGATGAAACTC-3′) resulting in the se-

quence 5′-*AAAAGAAGAAGGCTGATGAAACTC***AAAAGAAGAAGGCTGATGAAACTC**-3′ at position 29387 (corresponding to the nucleocapsid protein amino acid position 372) [14]. However, the inserted nucleotide sequence resulting in the Omicron ins214EPE is not a result of such a local duplication, as it is not identical or closely homologous to the preceding or subsequent nucleotide sequences in the original reference genome sequence of SARS-CoV-2, nor to that of the Omicron variant (Figure S2).

We thus asked whether polymerase slippage or template switching was a more plausible explanation for the origin of ins214EPE. For several reasons, it appears that template switching is the more plausible hypothesis. First, this is a long insertion per the definition described previously (nine or more nucleotides), and long insertions are more likely to arise from template switching than from polymerase slippage [14]. That said, we recognize that with exactly nine nucleotides, this is a borderline case between short and long insertions, and so its length alone may have limited value in distinguishing between these mechanisms. Second, this insertion has no uracil nucleotides, contrary to the expected excess uracils in slippage-mediated insertions [14]. Third, this insertion is monophyletic, although it is worth noting that other insertions at the same location have been observed in several other SARS-CoV-2 lineages (Table S3) [14,37]. Finally, the insertion occurs near previously described sites of potential non-canonical template switching. Specifically, there were non-canonical junctions observed 30 nucleotides upstream and 60 nucleotides downstream of this site (at positions 22183 and 22276, respectively) [42].

3.3. Candidate Templates for the Origin of ins214EPE in Omicron

While it is not certain that ins214EPE was generated by template switching, the points above illustrate that this is a plausible mechanism. If true, it would be of interest to determine candidate template RNA molecules from which this insertion could have arisen. We reasoned that there are three broad categories of most likely templates: (1) genomic material of SARS-CoV-2 itself (i.e., the positive-sense genomic RNA or the negative-sense anti-genomic RNA); (2) genomic or anti-genomic material of other viruses that have the capacity to co-infect the same cells as SARS-CoV-2; and (3) human transcripts that are expressed in cells infected with SARS-CoV-2 (Figure 2A,B). Although the latter category is not a well-described method of template switching [45], it has been suggested previously that insertions in SARS-CoV-2 genomes could be derived from the host transcriptome [46]. We identified exact matches for the forward and/or reverse-complement sequences in all three categories (Table 1, Figure 2C).

Table 1. Numbers of viral genomes and human transcripts with forward or reverse-complement matches to the three potential insertion sequences. Because the insertion sequences occur by definition in Omicron genomes, counts are shown separately for total SARS-CoV-2 genomes from GISAID and for genomes that are not assigned to the Omicron lineage. We confirmed that no sequences that were assigned to the Omicron lineage contained the insertion sequence or its reverse complement at any sites other than at the insertion itself. For the human *Coronaviridae* genomes, counts were first obtained by considering the available sequences for all human-infecting coronaviruses, and then filtered to retain only the viruses that are known to cause common colds or enteric illness (i.e., severe acute respiratory syndrome and Middle East respiratory syndrome virus sequences were excluded).

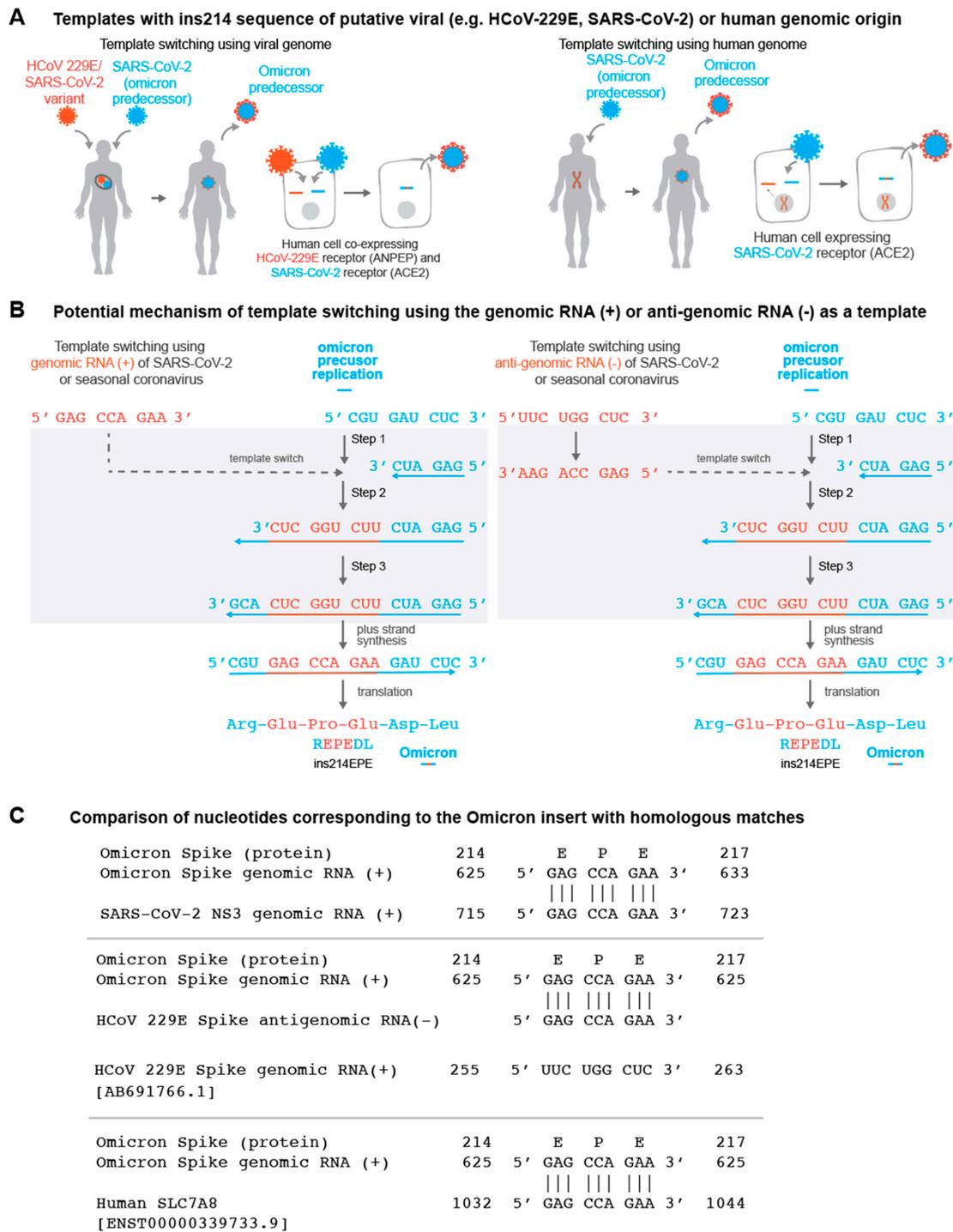| Candidate Insertion Sequence	Human Transcriptome Transcripts (Genes)		SARS-CoV-2 Genomes from GISAID Total (Non-Omicron)		Human *Coronaviridae* Genomes Any (Seasonal or Enteric)	
	Forward	Reverse-Complement	Forward	Reverse-Complement	Forward	Reverse-Complement
5′-GAGCCAGAA-3′	4677 (1534)	3264 (1220)	2100 (931)	27 (27)	18 (8)	17 (6)
5′-AGCCAGAAG-3′	6190 (2008)	4293 (1591)	1275 (106)	269 (269)	3 (0)	12 (0)
5′-GCCAGAAGA-3′	5210 (1564)	3146 (1144)	1319 (150)	201,632 (201,632)	13 (7)	4 (2)

A   **Templates with ins214 sequence of putative viral (e.g. HCoV-229E, SARS-CoV-2) or human genomic origin**

B   **Potential mechanism of template switching using the genomic RNA (+) or anti-genomic RNA (-) as a template**

C   **Comparison of nucleotides corresponding to the Omicron insert with homologous matches**

```
Omicron Spike (protein)              214          E    P    E         217
Omicron Spike genomic RNA (+)        625    5' GAG  CCA  GAA 3'        633
                                            ||| ||| |||
SARS-CoV-2 NS3 genomic RNA (+)       715    5' GAG  CCA  GAA 3'        723


Omicron Spike (protein)              214          E    P    E         217
Omicron Spike genomic RNA (+)        625    5' GAG  CCA  GAA 3'        625
                                            ||| ||| |||
HCoV 229E Spike antigenomic RNA(-)          5' GAG  CCA  GAA 3'

HCoV 229E Spike genomic RNA(+)       255    5' UUC  UGG  CUC 3'        263
[AB691766.1]


Omicron Spike (protein)              214          E    P    E         217
Omicron Spike genomic RNA (+)        625    5' GAG  CCA  GAA 3'        625
                                            ||| ||| |||
Human SLC7A8                        1032    5' GAG  CCA  GAA 3'       1044
[ENST00000339733.9]
```

Figure 2. (**A**) Schematic representation of Omicron's evolution through template switching involving viral (e.g., seasonal coronavirus or SARS-CoV-2) or human RNA. (**B**) Potential mechanism of template switching using viral genomic RNA (positive sense) or anti-genomic RNA (negative sense) as a template. Step 1: Negative-strand synthesis begins using an Omicron predecessor's genomic RNA as template. Step 2: Negative-strand synthesis temporarily uses the genomic or anti-genomic RNA of SARS-CoV-2 or a co-infecting virus. Step 3: Negative-strand synthesis resumes using the Omicron predecessor's genomic RNA as template. (**C**) Examples of matches identical to the nucleotide sequence "GAG CCA GAA" in the SARS-CoV-2 genome, the HCoV-229E anti-genome, and a human SLC7A8 transcript are shown.

There are no exact matches in any Omicron genomes collected to date for the three candidate sequences outside of the insertion site itself. This is notable because in previous template-switch-mediated insertions in the SARS-CoV-2 genome, the putative insertion template ("origin") has typically been present in the insertion-containing genome (Table S5) [14]. For the Omicron insertion, on the other hand, the only SARS-CoV-2 sequences in GISAID containing exact forward or reverse-complement matches are assigned to other SARS-CoV-2 lineages (or were not assigned to any lineage at the time of this analysis). There are several possible implications or interpretations of this finding.

First, substitutions can be introduced within the inserted sequence during template switching itself, or during subsequent rounds of viral replication, which would result in imperfect matching between the insertion and template sequences. This is particularly relevant for Omicron, which represents a long phylogenetic branch that arose after presumably several months of unobserved evolution [30,47]. We indeed found that the reference Omicron genome (EPI_ISL_6640916) [30] harbors several nucleotide 9-mers that differ from the candidate insertions by only a single nucleotide (Table S6), and additional 9-mers that differ by two or three nucleotides. These should be considered as possible templates for ins214EPE. Second, it is possible that the utilized template was derived from a co-infecting SARS-CoV-2 variant that does harbor the exact inserted sequence. Recombination between SARS-CoV-2 lineages in the context of simultaneous co-infection has been described previously, with particularly high recombination rates seen in the Spike protein sequence [43,48]. The distributions of lineages comprising the matched genomes for each candidate insertion are shown in Table S7. Finally, we noticed that several of the non-Omicron genomes with exact matches to one or more of the candidate insertions have been assigned to the B.1.1 Pango lineage (or sublineages thereof) (Table S7). Given that Omicron (originally Pango lineage B.1.1.529) is a phylogenetic descendant of B.1.1, this suggests that the ancestral genome that evolved into Omicron could have provided the necessary template for this insertion.

We also identified several genomes (or anti-genomes) of seasonal or enteric human-infecting coronaviruses that contain one or more of the putative insertion sequences. For example, the genomes of multiple human coronavirus OC43 (HCoV-OC43) and human enteric coronavirus strain 4408 both contain 5′-GCCAGAAGA-3′ and 5′-GAGCCAGAA-3′ in their nucleocapsid and replicase polyprotein genes, respectively (Table S8). Furthermore, 5′-GAGCCAGAA-3′ was present in 33 HCoV-229E anti-genomes, and 5′-GCCAGAAGA-3′ was present in 2 HCoV-NL63 anti-genomes (Table S8). The importance of recombination between coronaviruses has been highlighted recently [49], and its potential is supported by clinical reports showing that COVID-19 patients are co-infected with other respiratory pathogens, including non-SARS-CoV-2 viruses of the *Coronaviridae* family, at relatively high frequencies [50–52]. Furthermore, host receptors utilized by other coronaviruses (e.g., ANPEP and DPP4) are co-expressed with the SARS-CoV-2 receptor (ACE2) at the single-cell level in respiratory and/or gastrointestinal epithelial cells (Tables S9 and S10) [33], which could facilitate co-infection at the cellular level (a prerequisite for genomic recombination). Intestinal co-expression is relevant given the evidence that SARS-CoV-2 and other coronaviruses can infect enterocytes [53–55]. That said, it is worth noting that if template switching was indeed the mechanism by which this insertion arose, it is possible that the genomic material of respiratory pathogens outside of the *Coronaviridae* family (e.g., influenza, respiratory syncytial virus, parainfluenza, human metapneumovirus) could also serve as substrates in the context of such cellular co-infection. Genetic recombination between co-infecting viruses has been described, but this is more likely to occur between viruses in the same family with a high degree of genomic homology [56–58].

Finally, there were 4677 human transcripts (from 1534 genes) containing the forward sequence 5′-GAGCCAGAA-3′, and 3264 human transcripts (from 1220 genes) containing the reverse-complement sequence. Similar summary statistics are shown for the two other potentially inserted nine-nucleotide sequences in Table 1.

3.4. Consideration of Local Homology for the Candidate Templates

This landscape of possible templates, particularly among human transcripts, is expectedly quite vast given the total space of possible 9-mer nucleotide combinations (4^9 = 262,144). This raises the question of whether the most likely candidates may be those transcripts that have more homology surrounding the inserted sequence, as complementary base pairing resulting from such local similarity can increase the likelihood of serving as a template for recombination [45]. Indeed, in the normal process of coronavirus genomic replication, the prevailing model is that subgenomic RNAs are generated during negative-strand synthesis via a template-switching mechanism that relies on homology between conserved transcription regulatory sequence (TRS) elements dispersed strategically throughout the genome [42,44,59,60]. However, in the context of SARS-CoV-2, non-canonical transcripts with junctions that are not derived from TRS sequences and that share little homology between the 5′ and 3′ sites suggest that template switching guided by partial complementarity or other mechanisms may also play a role [42,61].

To study the homology between the regions surrounding template-switch-mediated insertions and their origins (i.e., the putative template that was copied to generate the insertion), we first considered a "positive control" set of four 12–15-nucleotide insertions in SARS-CoV-2 that were previously attributed to template switching with high confidence (Table S5) [14]. We calculated a normalized homology score (NHS) between the 7 or 35 nucleotides upstream or downstream of the insertion and origin sequences in these genomes, respectively (see Methods). Surprisingly, the degree of homology observed was generally not higher than expected by chance, as assessed via the NHS distribution of 10,000 randomly paired non-overlapping 7-mer or 35-mer nucleotide sequences from the SARS-CoV-2 genome (Figure S3, Table S11). This suggests that local homology may not be a prerequisite for the generation of template-switch-mediated insertions in SARS-CoV-2 [42].

Nevertheless, we still assessed the homology between the 35 nucleotides upstream or downstream of the Omicron insertion and the 35 nucleotides upstream or downstream of all candidate templates (see Methods). The NHS distributions for these candidates were generally similar to the distributions observed previously for randomly selected SARS-CoV-2 35-mers (Figure S4). Candidates in each category with the highest degrees of homology in the flanking upstream or downstream sequences included SARS-CoV-2 genomes from the lineages B.1.609 (NHS = 69) and AY.103 (NHS = 66), the HCoV-229E Spike protein (NHS = 63), and human transcripts of ACTN1 (NHS = 74) and EMC4 (NHS = 71). Some candidate templates also showed more homology in shorter sequences directly upstream and/or downstream of the inserted sequence. For example, in the reverse complement of the human TMEM245 transcript, there is a 17-nucleotide stretch with exact homology to the insertion-containing region of the Omicron genome (i.e., the nine-nucleotide inserted sequence plus eight exactly matched flanking nucleotides) (Figure S5).

4. Discussion

Omicron is more highly transmissible than prior variants [62,63], is less susceptible to neutralization by monoclonal antibodies and sera of vaccinated individuals [64–70], and is more likely to cause re-infections and vaccine breakthrough infections [71,72]. Among the many mutations in its Spike protein, ins214EPE is the only one that was not observed in other lineages prior to the emergence of Omicron. Whether this insertion, alone or in concert with other mutations, contributes to heightened transmissibility or lower susceptibility to neutralization by antibodies warrants further investigation.

While we cannot definitively determine the mechanism that gave rise to ins214EPE, we propose that template switching is a plausible explanation (Figure 2). Although the RNA-dependent RNA polymerases of SARS-CoV-2 and other coronaviruses do normally utilize template switching to generate subgenomic RNAs [42,59], it appears that template-switch-mediated insertions may derive from a non-canonical form of this process, in which a high degree of local homology (e.g., homologous TRS core sequences in the leader and body regions of the genome) is not essential [42,61]. It is not clear why a given sequence

would be utilized as a template in the absence of local homology, but this mechanism of "template selection" could involve secondary RNA structures or other unappreciated aspects of the SARS-CoV-2 replication machinery [73]. Here, we highlight several possible sources of the template for this insertion, including the SARS-CoV-2 genome itself, along with the genomes of other viruses or human transcripts. The use of SARS-CoV-2 genomic material as the template is supported by the finding that SARS-CoV-2 genome replication occurs in organelles that spatially concentrate the viral genomic material and replication machinery [74]. That said, it might indeed be possible for non-SARS-CoV-2 viral genomes or host mRNAs to be aberrantly included in these organelles, rendering them accessible for utilization during template switching as well.

There may be additional mechanisms that contribute to the acquisition of insertions by SARS-CoV-2 beyond those considered here. Saltational viral evolution in immuno-compromised patients has been suggested to underlie the emergence of highly mutated variants [75], and it is possible that ins214EPE (along with other Omicron-defining mutations) emerged in this context. Such individuals may be more prone to simultaneous co-infection with multiple SARS-CoV-2 variants, or with SARS-CoV-2 and other respiratory pathogens. It is also evident that the evolution of SARS-CoV-2 can occur in non-human species such as mice, deer, and mink or other mustelids [76–79], in which case other viral genomes and transcripts should be considered as possible templates. A recent analysis suggests that the proofreading exoribonuclease (encoded in nonstructural protein 14, or nsp14) is required for at least some of the genetic recombination observed in SARS-CoV-2 [80], but the potential mechanisms described here do not account for this. Finally, it is reasonable to question whether there is a relationship between ins214EPE and the three-nucleotide deletion (ΔN211) that occurs shortly upstream of it in most Omicron BA.1 sequences. However, because most sequences in GISAID with other insertions at position 214 do not possess such neighboring deletions (Figure S6), we believe that this proximal deletion was not a mechanistic prerequisite for the generation of the Omicron insertion.

It is not clear whether any one of these insertion-generating mechanisms would have more far-reaching consequences than the others. Template switching offers the intriguing possibility of new SARS-CoV-2 lineages borrowing protein domains or subdomains from previous variants, other viruses, or human proteins. However, it is worth noting that several of the possible templates that we identified for ins214EPE were derived from viral anti-genomes (i.e., from HCoV-229E or SARS-CoV-2 anti-genomes) or the reverse-complement sequences of human transcripts (e.g., TMEM245). In such cases, the polypeptide encoded in the Omicron genome would differ from that encoded by the positive-sense strand of the template. The genomic location and/or amino acid content of the inserted sequence may be more important than its origin, and we thus highlight the need to characterize the functional impact of ins214EPE on the clinical and epidemiological properties of the Omicron variant. Importantly, the data and hypotheses presented here are not sufficient to make inferences about properties such as transmissibility, immune evasion, or disease severity. Even if Omicron did acquire an insertion by utilizing a host transcript or the genome of a common-cold-causing coronavirus (e.g., HCoV-OC43, HCoV-229E), we do not propose that this would explain the reduced severity of COVID-19 observed in patients infected with Omicron compared to prior VOCs [81–84].

Multiple studies have demonstrated reduced effectiveness of COVID-19 vaccines against the Omicron variant compared to prior variants, including both primary vaccination series and booster doses [85–89]. Mechanistic characterization of the immune-evasive properties of Omicron have highlighted several substitutions in the Spike protein that confer antibody resistance, but have not revealed a role for ins214EPE [69,90–92]. The absence of this insertion in Omicron sublineages that emerged after BA.1 (Table S4) raises the question of whether this insertion was critical for the initial rapid global spread of Omicron, or if it was rather acquired as a non-essential "passenger mutation" during its evolution. It is also important to recognize that throughout the pandemic, different SARS-CoV-2 variants have variably impacted populations in distinct regions around the

globe. For example, the Gamma variant spread predominantly in South America, while the Beta variant was more prominent in South Africa, Europe, and Asia. Omicron has demonstrated the capacity to spread globally, but whether and how prior exposure to these different variants impacts its transmission is worth further exploration. Finally, the lack of clinical annotation associated with publicly deposited viral genomic data limits our ability to assess how the trajectory of viral evolution (including the acquisition of ins214EPE and other mutations) may be impacted by features such as vaccination status and immune competence. Future analyses of clinically annotated viral samples could help to address these questions.

5. Conclusions

The rapid rise in COVID-19 cases attributed to the Omicron variant, including among fully vaccinated individuals, raised alarm globally. In this context, it is important to better understand both the origins and the consequences of new genomic alterations that distinguish Omicron from prior VOCs and VOIs. Here, we begin to address the former by providing several plausible hypotheses on the origins of a nine-nucleotide insertion in the N-terminal domain of the Spike protein of the initially identified Omicron BA.1 variant. We suggest that genomic surveillance strategies should include an emphasis on sequencing SARS-CoV-2 genomes from immunocompromised patients and individuals with viral co-infections (including co-infections with multiple SARS-CoV-2 variants), as such individuals may provide unique contexts for genomic recombination and the evolution of new variants.

References

1. Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern. Available online: https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern (accessed on 1 December 2021).
2. Shu, Y.; McCauley, J. GISAID: Global initiative on sharing all influenza data—From vision to reality. *Euro Surveill.* **2017**, *22*, 30494. [CrossRef] [PubMed]
3. GISAID—hCov19 Variants. Available online: https://www.gisaid.org/hcov19-variants/ (accessed on 13 December 2021).
4. Plante, J.A.; Liu, Y.; Liu, J.; Xia, H.; Johnson, B.A.; Lokugamage, K.G.; Zhang, X.; Muruato, A.E.; Zou, J.; Fontes-Garfias, C.R.; et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **2020**, *592*, 116–121. [CrossRef] [PubMed]
5. Daniloski, Z.; Jordan, T.X.; Ilmain, J.K.; Guo, X.; Bhabha, G.; tenOever, B.R.; Sanjana, N.E. The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *eLife* **2021**, *10*, e65365. [CrossRef]
6. Wang, P.; Nair, M.S.; Liu, L.; Iketani, S.; Luo, Y.; Guo, Y.; Wang, M.; Yu, J.; Zhang, B.; Kwong, P.D.; et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **2021**, *593*, 130–135. [CrossRef] [PubMed]
7. Uriu, K.; Kimura, I.; Shirakawa, K.; Takaori-Kondo, A.; Nakada, T.-A.; Kaneda, A.; Nakagawa, S.; Sato, K. Genotype to Phenotype Japan (G2P-Japan) Consortium. Neutralization of the SARS-CoV-2 Mu Variant by Convalescent and Vaccine Serum. *N. Engl. J. Med.* **2021**, *22*, 942–943. [CrossRef]
8. Collier, D.A.; De Marco, A.; Ferreira, I.A.T.M.; Meng, B.; Datir, R.P.; Walls, A.C.; Kemp, S.A.; Bassi, J.; Pinto, D.; Silacci-Fregni, C.; et al. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **2021**, *593*, 136–141. [CrossRef] [PubMed]
9. McCarthy, K.R.; Rennick, L.J.; Nambulli, S.; Robinson-McCarthy, L.R.; Bain, W.G.; Haidar, G.; Duprex, W.P. Recurrent deletions in the SARS-CoV-2 Spike glycoprotein drive antibody escape. *Science* **2021**, *371*, 1139–1142. [CrossRef]
10. Motozono, C.; Toyoda, M.; Zahradnik, J.; Saito, A.; Nasser, H.; Tan, T.S.; Ngare, I.; Kimura, I.; Uriu, K.; Kosugi, Y.; et al. SARS-CoV-2 Spike L452R variant evades cellular immunity and increases infectivity. *Cell Host Microbe* **2021**, *29*, 1124–1136. [CrossRef]
11. McCallum, M.; De Marco, A.; Lempp, F.A.; Tortorici, M.A.; Pinto, D.; Walls, A.C.; Beltramello, M.; Chen, A.; Liu, Z.; Zatta, F.; et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **2021**, *184*, 2332–2347. [CrossRef]
12. Harvey, W.T.; Carabelli, A.M.; Jackson, B.; Gupta, R.K.; Thomson, E.C.; Harrison, E.M.; Ludden, C.; Reeve, R.; Rambaut, A.; Peacock, S.J.; et al. SARS-CoV-2 variants, Spike mutations and immune escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409–424. [CrossRef]
13. Ku, Z.; Xie, X.; Davidson, E.; Ye, X.; Su, H.; Menachery, V.D.; Li, Y.; Yuan, Z.; Zhang, X.; Muruato, A.E.; et al. Molecular determinants and mechanism for antibody cocktail preventing SARS-CoV-2 escape. *Nat. Commun.* **2021**, *12*, 469. [CrossRef] [PubMed]
14. Garushyants, S.K.; Rogozin, I.B.; Koonin, E.V. Template switching and duplications in SARS-CoV-2 genomes give rise to insertion variants that merit monitoring. *Commun. Biol.* **2021**, *4*, 1–9. [CrossRef] [PubMed]
15. Anand, P.; Puranik, A.; Aravamudan, M.; Venkatakrishnan, A.J.; Soundararajan, V. SARS-CoV-2 strategically mimics proteolytic activation of human ENaC. *Elife* **2020**, *9*, e58603. [CrossRef] [PubMed]
16. Coutard, B.; Valle, C.; de Lamballerie, X.; Canard, B.; Seidah, N.G.; Decroly, E. The Spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antivir. Res.* **2020**, *176*, 104742. [CrossRef] [PubMed]
17. Jaimes, J.A.; Millet, J.K.; Whittaker, G.R. Proteolytic Cleavage of the SARS-CoV-2 Spike Protein and the Role of the Novel S1/S2 Site. *iScience* **2020**, *23*, 101212. [CrossRef]
18. Peacock, T.P.; Goldhill, D.H.; Zhou, J.; Baillon, L.; Frise, R.; Swann, O.C.; Kugathasan, R.; Penn, R.; Brown, J.C.; Sanchez-David, R.Y.; et al. The furin cleavage site in the SARS-CoV-2 Spike protein is required for transmission in ferrets. *Nat. Microbiol.* **2021**, *6*, 899–909. [CrossRef]
19. Walls, A.C.; Park, Y.J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281–292. [CrossRef]
20. Liu, Z.; Long, W.; Tu, M.; Chen, S.; Huang, Y.; Wang, S.; Zhou, W.; Chen, D.; Zhou, L.; Wang, M.; et al. Lymphocyte subset (CD4+, CD8+) counts reflect the severity of infection and predict the clinical outcomes in patients with COVID-19. *J. Infect.* **2020**, *81*, 318. [CrossRef]
21. Diao, B.; Wang, C.; Tan, Y.; Chen, X.; Liu, Y.; Ning, L.; Chen, L.; Li, M.; Liu, Y.; Wang, G.; et al. Reduction and Functional Exhaustion of T Cells in Patients with Coronavirus Disease 2019 (COVID-19). *Front. Immunol.* **2020**, *11*, 827. [CrossRef]

22. Tavakolpour, S.; Rakhshandehroo, T.; Wei, E.X.; Rashidian, M. Lymphopenia during the COVID-19 infection: What it shows and what can be learned. *Immunol. Lett.* **2020**, *225*, 31. [CrossRef]

23. Zhang, Z.; Zheng, Y.; Niu, Z.; Zhang, B.; Wang, C.; Yao, X.; Peng, H.; Franca, D.N.; Wang, Y.; Zhu, Y.; et al. SARS-CoV-2 Spike protein dictates syncytium-mediated lymphocyte elimination. *Cell Death Differ.* **2021**, *28*, 2765–2777. [CrossRef]

24. Zhang, L.; Richards, A.; Barrasa, M.I.; Hughes, S.H.; Young, R.A.; Jaenisch, R. Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2105968118. [CrossRef] [PubMed]

25. Parry, R.; Gifford, R.J.; Lytras, S.; Ray, S.C.; Coin, L.J.M. No evidence of SARS-CoV-2 reverse transcription and integration as the origin of chimeric transcripts in patient tissues. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2109066118. [CrossRef] [PubMed]

26. Zhang, L.; Richards, A.; Barrasa, M.I.; Hughes, S.H.; Young, R.A.; Jaenisch, R. Response to Parry et al.: Strong evidence for genomic integration of SARS-CoV-2 sequences and expression in patient tissues. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2109497118. [CrossRef] [PubMed]

27. Tzou, P.L.; Tao, K.; Nouhin, J.; Rhee, S.-Y.; Hu, B.D.; Pai, S.; Parkin, N.; Shafer, R.W. Coronavirus Antiviral Research Database (CoV-RDB): An Online Database Designed to Facilitate Comparisons between Candidate Anti-Coronavirus Compounds. *Viruses* **2020**, *12*, 1006. [CrossRef] [PubMed]

28. Mathieu, E.; Ritchie, H.; Ortiz-Ospina, E.; Roser, M.; Hasell, J.; Appel, C.; Giattino, C.; Rodés-Guirao, L. A global database of COVID-19 vaccinations. *Nat. Hum. Behav* **2021**, *5*, 947–953. [CrossRef] [PubMed]

29. Venkatakrishnan, A.J.; Anand, P.; Lenehan, P.; Ghosh, P.; Suratekar, R.; Siroha, A.; Chowdhury, D.R.; O'Horo, J.C.; Yao, J.D.; Pritt, B.S.; et al. Antigenic minimalism of SARS-CoV-2 is linked to surges in COVID-19 community transmission and vaccine breakthrough infections. *medRxiv* **2021**. [CrossRef]

30. Kandeel, M.; Mohamed, M.E.M.; Hm, A.E.-L.; Venugopala, K.N.; El-Beltagi, H.S. Omicron variant genome evolution and phylogenetics. *J. Med. Virol.* **2021**, *94*, 1627–1632. [CrossRef] [PubMed]

31. Frankish, A.; Diekhans, M.; Jungreis, I.; Lagarde, J.; Loveland, J.E.; Mudge, J.M.; Sisu, C.; Wright, J.C.; Armstrong, J.; Barnes, I.; et al. GENCODE 2021. *Nucleic Acids Res.* **2021**, *49*, D916–D923. [CrossRef]

32. Brister, J.R.; Ako-Adjei, D.; Bao, Y.; Blinkova, O. NCBI viral genomes resource. *Nucleic Acids Res.* **2015**, *43*, D571–D577. [CrossRef]

33. Venkatakrishnan, A.J.; Puranik, A.; Anand, A.; Zemmour, D.; Yao, X.; Wu, X.; Chilaka, R.; Murakowski, D.K.; Standish, K.; Raghunathan, B.; et al. Knowledge synthesis of 100 million biomedical documents augments the deep expression profiling of coronavirus receptors. *Elife* **2020**, *9*, e58040. [CrossRef] [PubMed]

34. Doddahonnaiah, D.; Lenehan, P.J.; Hughes, T.K.; Zemmour, D.; Garcia-Rivera, E.; Venkatakrishnan, A.J.; Chilaka, R.; Khare, A.; Kasaraneni, A.; Garg, A.; et al. A Literature-Derived Knowledge Graph Augments the Interpretation of Single Cell RNA-seq Datasets. *Genes* **2021**, *12*, 898. [CrossRef] [PubMed]

35. Chua, R.L.; Lukassen, S.; Trump, S.; Hennig, B.P.; Wendisch, D.; Pott, F.; Debnath, O.; Thürmann, L.; Kurth, F.; Völker, M.T.; et al. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* **2020**, *38*, 970–979. [CrossRef]

36. Martin, J.C.; Chang, C.; Boschetti, G.; Ungaro, R.; Giri, M.; Grout, J.A.; Gettler, K.; Chuang, L.S.; Nayar, S.; Greenstein, A.J.; et al. Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. *Cell* **2019**, *178*, 1493–1508. [CrossRef] [PubMed]

37. Gerdol, M.; Dishnica, K.; Giorgetti, A. Emergence of a recurrent insertion in the N-terminal domain of the SARS-CoV-2 spike glycoprotein. *Virus Res* **2022**, *310*, 198674. [CrossRef]

38. Tarke, A.; Sidney, J.; Kidd, C.K.; Dan, J.M.; Ramirez, S.I.; Yu, E.D.; Mateus, J.; da Silva Antunes, R.; Moore, E.; Rubiro, P.; et al. Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. *Cell Rep. Med.* **2021**, *2*, 100204. [CrossRef]

39. Lam, S.D.; Waman, V.P.; Orengo, C.; Lees, J. Insertions in the SARS-CoV-2 Spike N-Terminal Domain May Aid COVID-19 Transmission. *bioRxiv* **2021**. [CrossRef]

40. Huston, N.C.; Wan, H.; Strine, M.S.; de Cesaris Araujo Tavares, R.; Wilen, C.B.; Pyle, A.M. Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol. Cell* **2021**, *81*, 584. [CrossRef]

41. te Velthuis, A.J.W.; Arnold, J.J.; Cameron, C.E.; van den Worm, S.H.E.; Snijder, E.J. The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Res.* **2009**, *38*, 203–214. [CrossRef]

42. Kim, D.; Lee, J.Y.; Yang, J.S.; Kim, J.W.; Kim, V.N.; Chang, H. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, *181*, 914–921. [CrossRef]

43. Jackson, B.; Boni, M.F.; Bull, M.J.; Colleran, A.; Colquhoun, R.M.; Darby, A.C.; Haldenby, S.; Hill, V.; Lucaci, A.; McCrone, J.T.; et al. Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. *Cell* **2021**, *184*, 5179–5188. [CrossRef] [PubMed]

44. Sawicki, S.G.; Sawicki, D.L. Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands. *Adv. Exp. Med. Biol.* **1995**, *380*, 499–506.

45. Simon-Loriere, E.; Holmes, E.C. Why do RNA viruses recombine? *Nat. Rev. Microbiol.* **2011**, *9*, 617–626. [CrossRef] [PubMed]

46. Peacock, T.P. Putative Host Origins of RNA Insertions in SARS-CoV-2 Genomes. *Virological*. 2021. Available online: https://virological.org/t/putative-host-origins-of-rna-insertions-in-sars-cov-2-genomes/761 (accessed on 10 January 2022).

47. Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* **2018**, *34*, 4121–4123. [CrossRef]

48. Turkahia, Y.; Thornlow, B.; Hinrichs, A.; McBroome, J.; Ayala, N.; Ye, C.; De Maio, N.; Haussler, D.; Lanfear, R.; Corbett-Detig, R. Pandemic-Scale Phylogenomics Reveals Elevated Recombination Rates in the SARS-CoV-2 Spike Region. *bioRxiv* **2021**. [CrossRef]

49. Morens, D.M.; Taubenberger, J.K.; Fauci, A.S. Universal Coronavirus Vaccines—An Urgent Need. *N. Engl. J. Med.* **2021**, *386*, 297–299. [CrossRef]

50. Lau, S.K.P.; Lung, D.C.; Wong, E.Y.M.; Aw-Yong, K.L.; Wong, A.C.P.; Luk, H.K.H.; Li, K.S.M.; Fung, J.; Chan, T.T.Y.; Tang, J.Y.M.; et al. Molecular Evolution of Human Coronavirus 229E in Hong Kong and a Fatal COVID-19 Case Involving Coinfection with a Novel Human Coronavirus 229E Genogroup. *mSphere* **2021**, *6*, e00819-20. [CrossRef]

51. 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. [CrossRef] [PubMed]

52. Musuuza, J.S.; Watson, L.; Parmasad, V.; Putman-Buehler, N.; Christensen, L.; Safdar, N. Prevalence and outcomes of co-infection and superinfection with SARS-CoV-2 and other pathogens: A systematic review and meta-analysis. *PLoS ONE* **2021**, *16*, e0251170. [CrossRef] [PubMed]

53. Zhou, J.; Li, C.; Zhao, G.; Chu, H.; Wang, D.; Yan, H.H.; Poon, V.K.; Wen, L.; Wong, B.H.; Zhao, X.; et al. Human intestinal tract serves as an alternative infection route for Middle East respiratory syndrome coronavirus. *Sci. Adv.* **2017**, *3*, eaao4966. [CrossRef]

54. Lamers, M.M.; Beumer, J.; van der Vaart, J.; Knoops, K.; Puschhof, J.; Breugem, T.I.; Ravelli, R.B.G.; van Schayck, J.P.; Mykytyn, A.Z.; Duimel, H.Q.; et al. SARS-CoV-2 productively infects human gut enterocytes. *Science* **2020**, *369*, 50–54. [CrossRef] [PubMed]

55. Bein, A.; Kim, S.; Goyal, G.; Cao, W.; Fadel, C.; Naziripour, A.; Sharma, S.; Swenor, B.; LoGrande, N.; Nurani, A.; et al. Enteric Coronavirus Infection and Treatment Modeled with an Immunocompetent Human Intestine-On-A-Chip. *Front. Pharmacol.* **2021**, *12*, 718484. [CrossRef] [PubMed]

56. Kumar, N.; Sharma, S.; Barua, S.; Tripathi, B.N.; Rouse, B.T. Virological and immunological outcomes of coinfections. *Clin. Microbiol. Rev.* **2018**, *31*, e00111-17. [CrossRef]

57. Saade, G.; Deblanc, C.; Bougon, J.; Marois-Créhan, C.; Fablet, C.; Auray, G.; Belloc, C.; Leblanc-Maridor, M.; Gagnon, C.A.; Zhu, J.; et al. Coinfections and their molecular consequences in the porcine respiratory tract. *Vet. Res.* **2020**, *51*, 80. [CrossRef]

58. Meurens, F.; Keil, G.M.; Muylkens, B.; Gogev, S.; Schynts, F.; Negro, S.; Wiggers, L.; Thiry, E. Interspecific recombination between two ruminant alphaherpesviruses, bovine herpesviruses 1 and 5. *J. Virol.* **2004**, *78*, 9828–9836. [CrossRef] [PubMed]

59. Sola, I.; Almazán, F.; Zúñiga, S.; Enjuanes, L. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu. Rev. Virol.* **2015**, *2*, 265. [CrossRef]

60. Yang, Y.; Yan, W.; Hall, A.B.; Jiang, X. Characterizing Transcriptional Regulatory Sequences in Coronaviruses and Their Role in Recombination. *Mol. Biol. Evol.* **2020**, *38*, 1241–1248. [CrossRef]

61. Nomburg, J.; Meyerson, M.; DeCaprio, J.A. Pervasive generation of non-canonical subgenomic RNAs by SARS-CoV-2. *Genome Med.* **2020**, *12*, 108. [CrossRef]

62. HKUMed Finds Omicron SARS-CoV-2 can Infect Faster and Better than Delta in Human Bronchus but with Less Severe Infection in Lung. Available online: https://www.med.hku.hk/en/news/press/20211215-omicron-sars-cov-2-infection (accessed on 5 January 2022).

63. Callaway, E.; Ledford, H. How bad is Omicron? What scientists know so far. *Nature* **2021**, *600*, 197–199. [CrossRef] [PubMed]

64. Garcia-Beltran, W.F.; St Denis, K.J.; Hoelzemer, A.; Lam, E.C.; Nitido, A.D.; Sheehan, M.L.; Berrios, C.; Ofoman, O.; Chang, C.C.; Hauser, B.M.; et al. mRNA-based COVID-19 vaccine boosters induce neutralizing immunity against SARS-CoV-2 Omicron variant. *Cell* **2022**, *185*, 457–466. [CrossRef]

65. Pajon, R.; Doria-Rose, N.A.; Shen, X.; Schmidt, S.D.; O'Dell, S.; McDanal, C.; Feng, W.; Tong, J.; Eaton, A.; Maglinao, M.; et al. SARS=CoV-2 Omicron Variant Neutralization after mRNA-1273 Booster Vaccination. *N Engl J Med* **2022**. [CrossRef] [PubMed]

66. Dejnirattisai, W.; Shaw, R.H.; Supasa, P.; Liu, C.; Stuart, A.S.V.; Pollard, A.J.; Liu, X.; Lambe, T.; Crook, D.; Stuart, D.I.; et al. Reduced neutralisation of SARS-COV-2 omicron B.1.1.529 variant by post-immunisation serum. *Lancet* **2022**, *399*, 234–236. [CrossRef]

67. Wilhelm, A.; Widera, M.; Grikscheit, K.; Toptan, T.; Schenk, B.; Pallas, C.; Metzler, M.; Kohmer, N.; Hoehl, S.; Helfritz, F.A.; et al. Limited neutralisation of the SARS-CoV-2 Omicron subvariants BA.1 and BA.2 by convalescent and vaccine serum and monoclonal antibodies. *EBioMedicine* **2022**, *82*, 104158. [CrossRef]

68. Pfizer and BioNTech Provide Update on Omicron Variant. Available online: https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-provide-update-omicron-variant (accessed on 10 December 2021).

69. Liu, L.; Iketani, S.; Guo, Y.; Chan, J.F.-W.; Wang, M.; Liu, L.; Luo, Y.; Chu, H.; Huang, Y.; Nair, M.S.; et al. Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature* **2022**, *602*, 676–681. [CrossRef]

70. Zhang, L.; Li, Q.; Liang, Z.; Li, T.; Liu, S.; Cui, Q.; Nie, J.; Wu, Q.; Qu, X.; Huang, W.; et al. The significant immune escape of pseudotyped SARS-CoV-2 Variant Omicron. *Emerg. Microbes Infect.* **2021**, *11*, 1–5. [CrossRef]

71. Pulliam, J.R.C.; van Schalkwyk, C.; Govender, N.; von Gottberg, A.; Cohen, C.; Groome, M.J.; Dushoff, J.; Mlisana, K.; Moultrie, H. Increased risk of SARS-CoV-2 reinfection associated with emergence of the Omicron variant in South Africa. *Science* **2022**, *376*, eabn4947. [CrossRef]

72. Varrelman, T.J.; Rader, B.M.; Astley, C.M.; Brownstein, J.S. Syndromic Surveillance-Based Estimates of Vaccine Efficacy Against COVID-Like Illness from Emerging Omicron and COVID-19 Variants. *medRxiv* **2021**. [CrossRef]

73. Chrisman, B.S.; Paskov, K.; Stockham, N.; Tabatabaei, K.; Jung, J.-Y.; Washington, P.; Varma, M.; Sun, M.W.; Maleki, S.; Wall, D.P. Indels in SARS-CoV-2 occur at template-switching hotspots. *BioData Min.* **2021**, *14*, 20. [CrossRef]

74. Snijder, E.J.; Rwal, L.; de Wilde, A.H.; de Jong, A.W.M.; Zevenhoven-Dobbe, J.C.; Maier, H.J.; Ffga, F.; Koster, A.J.; Bárcena, M. A unifying structural and functional model of the coronavirus replication organelle: Tracking down RNA synthesis. *PLoS Biol.* **2020**, *18*, e3000715. [CrossRef]

75. Corey, L.; Beyrer, C.; Cohen, M.S.; Michael, N.L.; Bedford, T.; Rolland, M. SARS-CoV-2 variants in patients with immunosuppression. *N. Engl. J. Med.* **2021**, *385*, 562–566. [CrossRef]

76. Kupferschmidt, K. Where did "weird" Omicron come from? *Science* **2021**, *374*, 1179. [CrossRef] [PubMed]

77. Wei, C.; Shan, K.J.; Wang, W.; Zhang, S.; Huan, Q.; Qian, W. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. *J. Genet. Genomics* **2021**, *48*, 1111–1121. [CrossRef] [PubMed]

78. Oude Munnink, B.B.; Sikkema, R.S.; Nieuwenhuijse, D.F.; Molenaar, R.J.; Munger, E.; Molenkamp, R.; van der Spek, A.; Tolsma, P.; Rietveld, A.; Brouwer, M.; et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* **2021**, *371*, 172–177. [CrossRef] [PubMed]

79. European Food Safety Authority and European Centre for Disease Prevention and Control; Boklund, A.; Gortázar, C.; Pasquali, P.; Roberts, H.; Nielsen, S.S.; Stahl, K.; Stegeman, A.; Baldinelli, F.; Broglia, A.; et al. Monitoring of SARS-CoV-2 infection in mustelids. *EFSA J.* **2021**, *19*, e06459.

80. Gribble, J.; Stevens, L.J.; Agostini, M.L.; Anderson-Daniels, J.; Chappell, J.D.; Lu, X.; Pruijssers, A.J.; Routh, A.L.; Denison, M.R. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog.* **2021**, *17*, e1009226. [CrossRef]

81. Madhi, S.A.; Kwatra, G.; Myers, J.E.; Jassat, W.; Dhar, N.; Mukendi, C.K.; Nana, A.J.; Blumberg, L.; Welch, R.; Ngorima-Mabhena, N.; et al. Population Immunity and COVID-19 Severity with Omicron Variant in South Africa. *N. Engl. J. Med.* **2022**, *386*, 1314–1326. [CrossRef]

82. Wolter, N.; Jassat, W.; Walaza, S.; Welch, R.; Moultrie, H.; Groome, M.; Amoako, D.G.; Everatt, J.; Bhiman, J.N.; Scheepers, C.; et al. Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in South Africa: A data linkage study. *Lancet* **2022**, *399*, 437–446. [CrossRef]

83. Ulloa, A.C.; Buchan, S.A.; Daneman, N.; Brown, K.A. Estimates of SARS-CoV-2 Omicron Variant Severity in Ontario, Canada. *JAMA* **2022**, *327*, 1286–1288. [CrossRef]

84. Maslo, C.; Friedland, R.; Toubkin, M.; Laubscher, A.; Akaloo, T.; Kama, B. Characteristics and Outcomes of Hospitalized Patients in South Africa During the COVID-19 Omicron Wave Compared with Previous Waves. *JAMA* **2022**, *327*, 583–584. [CrossRef]

85. Andrews, N.; Stowe, J.; Kirsebom, F.; Toffa, S.; Rickeard, T.; Gallagher, E.; Gower, C.; Kall, M.; Groves, N.; O'Connell, A.-M.; et al. COVID-19 vaccine effectiveness against the omicron (B.1.1.529) variant. *N. Engl. J. Med.* **2022**, *386*, 1532–1546. [CrossRef]

86. Link-Gelles, R.; Levy, M.E.; Gaglani, M.; Irving, S.A.; Stockwell, M.; Dascomb, K.; DeSilva, M.B.; Reese, S.E.; Liao, I.-C.; Ong, T.C.; et al. Effectiveness of 2, 3, and 4 COVID-19 mRNA vaccine doses among immunocompetent adults during periods when SARS-CoV-2 omicron BA.1 and BA.2/BA.2.12.1 sublineages predominated—VISION network, 10 states, December 2021–June 2022. *MMWR Morb. Mortal. Wkly. Rep.* **2022**, *71*, 931. [CrossRef] [PubMed]

87. Higdon, M.M.; Baidya, A.; Walter, K.K.; Patel, M.K.; Issa, H.; Espié, E.; Feikin, D.R.; Knoll, M.D. Duration of effectiveness of vaccination against COVID-19 caused by the omicron variant. *Lancet Infect. Dis.* **2022**, *22*, 1114–1116. [CrossRef]

88. Altarawneh, H.N.; Chemaitelly, H.; Ayoub, H.H.; Tang, P.; Hasan, M.R.; Yassine, H.M.; Al-Khatib, H.A.; Smatti, M.K.; Coyle, P.; Al-Kanaani, Z.; et al. Effects of previous infection and vaccination on symptomatic omicron infections. *N. Engl. J. Med.* **2022**, *387*, 21–34. [CrossRef] [PubMed]

89. Gray, G.; Collie, S.; Goga, A.; Garrett, N.; Champion, J.; Seocharan, I.; Bamford, L.; Moultrie, H.; Bekker, L.-G. Effectiveness of Ad26.COV2.S and BNT162b2 vaccines against omicron variant in South Africa. *N. Engl. J. Med.* **2022**, *386*, 2243–2245. [CrossRef]

90. Iketani, S.; Liu, L.; Guo, Y.; Liu, L.; Chan, J.F.-W.; Huang, Y.; Wang, M.; Luo, Y.; Yu, J.; Chu, H.; et al. Antibody evasion properties of SARS-CoV-2 Omicron sublineages. *Nature* **2022**, *604*, 553–556. [CrossRef]

91. Cao, Y.; Wang, J.; Jian, F.; Xiao, T.; Song, W.; Yisimayi, A.; Huang, W.; Li, Q.; Wang, P.; An, R.; et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **2022**, *602*, 657–663. [CrossRef]

92. Cele, S.; Jackson, L.; Khoury, D.S.; Khan, K.; Moyo-Gwete, T.; Tegally, H.; San, J.E.; Cromer, D.; Scheepers, C.; Amoako, D.G.; et al. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature* **2022**, *602*, 654–656. [CrossRef]