



Article

# SARS-CoV-2 Consensus-Sequence and Matching Overlapping Peptides Design for COVID19 Immune Studies and Vaccine Development

Alex Olvera <sup>1,2,†</sup> , Marc Noguera-Julian <sup>1,3,†</sup> , Athina Kilpelainen <sup>1</sup> , Luis Romero-Martín <sup>1</sup>, Julia G. Prado <sup>1,4,\*,‡</sup> and Christian Brander <sup>1,3,5,\*,‡</sup>

<sup>1</sup> IrsiCaixa AIDS Research Institute-HIVACAT, Hospital Universitari Germans Trias i Pujol, 08916 Badalona, Spain; aolvera@irsicaixa.es (A.O.); mnoguera@irsicaixa.es (M.N.-J.); akilpelainen@irsicaixa.es (A.K.); lromero@irsicaixa.es (L.R.-M.)

<sup>2</sup> Faculty of Sciences and Technology, Universitat de Vic-Central de Catalunya (UVic-UCC), 08500 Vic, Spain

<sup>3</sup> Faculty of Medicine, Universitat de Vic-Central de Catalunya (UVic-UCC), 08500 Vic, Spain

<sup>4</sup> Germans Trias i Pujol Research Institute (IGTP), 08196 Barcelona, Spain

<sup>5</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

\* Correspondence: jgarcia@irsicaixa.es (J.G.P.); cbrander@irsicaixa.es (C.B.)

† A.O. and M.N.-J. contributed equally to this paper.

‡ J.G.P. and C.B. contributed equally to this paper.

Received: 22 June 2020; Accepted: 31 July 2020; Published: 6 August 2020



**Abstract:** Synthetic antigens based on consensus sequences that represent circulating viral isolates are sensitive, time saving and cost-effective tools for in vitro immune monitoring and to guide immunogen design. When based on a representative sequence database, such consensus sequences can effectively be used to test immune responses in exposed and infected individuals at the population level. To accelerate immune studies in SARS-CoV-2 infection, we here describe a SARS-CoV-2 2020 consensus sequence (CoV-2-cons) which is based on more than 1700 viral genome entries in NCBI and encompasses all described SARS-CoV-2 open reading frames (ORF), including recently described frame-shifted and length variant ORF. Based on these sequences, we created curated overlapping peptide (OLP) lists containing between 1500 to 3000 peptides of 15 and 18 amino acids in length, overlapping by 10 or 11 residues, as ideal tools for the assessment of SARS-CoV-2-specific T cell immunity. In addition, CoV-2-cons sequence entropy values are presented along with variant sequences to provide increased coverage of the most variable sections of the viral genome. The identification of conserved protein fragments across the coronavirus family and the corresponding OLP facilitate the identification of T cells potentially cross-reactive with related viruses. This new CoV-2-cons sequence, together with the peptides sets, should provide the basis for SARS-CoV-2 antigen synthesis to facilitate comparability between ex-vivo immune analyses and help to accelerate research on SARS-CoV-2 immunity and vaccine development.

**Keywords:** COVID-19; SARS-CoV-2; consensus sequence; T cell immunity; overlapping peptides set

## 1. Introduction

Since the start of the COVID-19 pandemic in December 2019, researchers around the world have put major efforts towards a better understanding of the immune response to its causative agent, the SARS-CoV-2. Although an impressive amount of scientific information has been generated in a very short period of time, there remain significant gaps in our understanding of SARS-CoV-2 immune control. In particular, it remains unclear what kind of adaptive immunity should be triggered by vaccination in order to achieve sterile immunity, or at least lead to an ameliorated disease course, in cases where

vaccination cannot provide absolute protection from infection. We know from the available literature on other coronaviruses (mainly SARS-CoV-1 and MERS) that antibodies can neutralize the infection, although these humoral responses are short lived in many individuals, and that long-lived T cell responses are present in people with less severe disease outcomes [1–5]. The emerging data on the immune response to SARS-CoV-2 demonstrate the essential contribution of the virus-specific T-cell responses, possibly in addition to the action of neutralizing antibodies, in viral control [3,6–13]. Thus, improved tools to assess host T cell immunity in detail are urgently needed to better identify these responses and to define their role in the outcome of SARS-CoV-2 infection.

Ex-vivo immune analyses of samples from infected individuals can identify T cell responses to specific pathogens like viruses. Such analyses can help to better understand the role of host immunity in virus control and to guide successful vaccine development. However, they rely on the use of the correct recall antigens that can elicit specific responses in vitro. The urgency of the current SARS-CoV-2 pandemic has led researchers to tackle the problem of screening the 10,000 amino acids of the SARS-CoV-2 proteome for T cell responses by selecting viral sequences based on different criteria: (i) bioinformatically predicted epitopes, (ii) homology of SARS-CoV-2 sequences with epitopes defined in other coronaviruses (mainly SARS-CoV) or (iii) selecting some specific SARS-CoV-2 proteins over others [5,7,9,11,14–19]. However, all these approaches have intrinsic limitations. Bioinformatic prediction tools are trained on sets of previously described epitopes, but since the available epitope repertoire for many human leukocyte antigen (HLA) alleles is limited, its prediction capacity is also limited [20,21]. Inferences based on epitope sequence homology with other coronaviruses are hampered because past studies on SARS-CoV-1 and MERS only included few selected viral proteins. This is of concern, since screening only a part of the SARS-CoV-2 proteome will potentially miss an important portion of the virus-specific T cell response. Indeed, recent data indicate the existence of T cell responses against structural and non-structural proteins [5,9] for SARS-CoV-2 and other viral infections [22]. Finally, no study has considered the existence of T cell responses to epitopes encoded by open-reading frames (ORF) in alternative frames, as reported for other viral infections [23–26].

In order to reliably measure total virus-specific T cell immunity, the recall antigens used need to be as representative as possible of the worldwide viral sequences, even for genetically more stable viruses like coronaviruses. T cell recognition of epitopes is very sensitive to mismatches and not matching the recall antigen with the autologous virus can lead to missed responses [27]. For this reason, different test antigen design strategies, trying to cope with the diversity of circulating viral isolates in a single sequence, have been developed in the past. These strategies include central sequence designs such as Center of Tree (COT) [28–32], Ancestral [33–36] or Consensus sequences [29–32,35,37–43]; which may (Ancestral, COT) or may not (Consensus) represent naturally occurring sequences of replication competent viruses. All these designs are sensitive to the underlying sequence database and may change over time as new sequence information on additional isolates becomes available. Direct comparisons of these different central sequence approaches have been performed for a highly variable pathogen (human immunodeficiency virus, HIV) and shown that the different designs yielded comparable results when synthetic peptides covering these sequences were used to measure virus-specific T cell responses [42,43]. However, the additional costs in terms of peptide synthesis and cells needed for ex-vivo experiments, may not warrant inclusion of all the different variants into a single test set.

Thus, the characterization of the complete T cell responses to SARS-CoV-2 urgently needs T cell antigens that cover the whole SARS-CoV-2 proteome while covering sequence diversity, and which can be combined in different experimental set-ups and immune assays. To this end, we created a consensus sequence to cover the genetic diversity of SARS-CoV-2 (CoV-2-cons) for all ORF, including those described in alternative open reading frames. Given the computational ease for its initial generation and periodic updates, we designed a consensus sequence using more than 1700 CoV-2 full-genome sequences and designed overlapping peptide (OLP) sets as recall antigens in T cell assays. The CoV-2-cons OLP sets are presented here in different designs, balancing costs for synthesis with the

sensitivity of detecting T cell responses and with the intention to provide a common test antigen that will allow data comparability across laboratories.

## 2. Methods

### 2.1. Consensus Sequence ORF Generation and Entropy Calculation

A total of 1731 full-length SARS-CoV-2 sequences were downloaded from NCBI (30 April 2020, txid2697049, minimum length = 29,000 bp) and aligned using MAFFT [44]. The alignment was visually inspected and curated using Genbank NC\_045512.2 as a coordinate reference [45]. A nucleotide consensus sequence was generated by keeping all nucleotides present in at least 25% of the sequences in the alignment. The amino acid consensus sequence was then created by using NC\_045512.2 annotated Open Reading Frames (ORFs) plus additional ORFs described in Finkel et al. [46] using the Biostrings R package. Mixed nucleotide positions were either resolved if they were synonymous or flagged for downstream analysis. Positional entropy was calculated at the amino acid level both as the standard and 22-aminoacid-normalized Shannon entropy for every ORF using Bio3d R package on the alignment [47], and afterward, the mean OLP normalized entropy was calculated.

### 2.2. Overlapping Peptide Set Design and Variability Plots

For the automated design of overlapping peptides with variable length, we used the previously described Peptgen algorithm available at the Los Alamos National Laboratories HIV Immunology database [48]. This OLP generator allows predefining peptide length and level of the desired overlap between adjacent OLP. Peptgen is also set up to exclude from the C-terminal end of OLP certain “forbidden” amino acids (G, P, E, D, Q, N, T, S and C) that are rarely seen to serve as the C-terminal anchor position of HLA class I presented epitopes [49]. Using this optional modification can lead to length variation in the OLP set, which can be controlled by limiting the maximal length of an OLP in regions with numerous serial “forbidden” residues. The settings used for the present SARS-CoV-2 consensus OLP design were a) OLP length of 15 or 18 amino acids, with maximal extension or truncation of up to  $\pm 3$  residues to avoid forbidden C-terminal residues. In addition, the overlap between adjacent OLP was set at 10 or 11 residues. The no-glutamine at N-terminal setting was applied to prevent OLP starting with a glutamine residue as this can lead to complications with peptide synthesis. For positions where two or more amino acids were present above 25% of the sequences in the alignment, two or more sequence variants for those OLPs were generated. Sequence logos were generated for these cases with the ggseqlogo R package [50].

### 2.3. Detection of Conserved Peptides Among Coronavirus

In an attempt to detect protein fragments that are conserved across a wide range of members of the coronavirus family, full-length consensus ORF from SARS-CoV-2 were aligned with other coronavirus sequences. Three alignments were performed based on different sequence selection criteria: (i) 50 reference sequences (RefSeq) with the lowest E-values resulting from a pBLAST search [51] using the ORF-specific consensus sequences (pan-coronavirus alignment) (ii) homologous proteins from 17 viruses representing the Betacoronavirus taxon (beta-coronavirus alignment) or, (iii) homologous proteins from the 7 full-genome sequenced human coronaviruses (including SARS-CoV, MERS-CoV, and common cold species OC43, NL63, 229E, HKU1, human-coronavirus alignment). Selected sequences were aligned using the MUSCLE algorithm in MEGA X [52]. Conserved protein fragments were identified using BioEdit with the following criteria: minimum length of 8 amino acid, maximum average entropy of 0.25, maximum entropy per position of 1 and limiting the search to 1 gap per segment. Sequence logos were generated for the aligned peptides on Weblogo [53].

#### 2.4. Identification of Previously Described Epitopes in CoV-2 Conserved Regions

To identify previously reported epitopes in the conserved regions of coronaviruses (pan-coronavirus, betacoronaviruses, and human coronaviruses), and match them with the SARS-CoV-2 consensus sequence, searches for experimentally described epitopes were carried out in the Immune Epitope Database [54]. The search criteria were as follows: “linear peptide; blast option: 90%; Host: Homo sapiens; Any MHC restriction; Positive assays only; All assays; Any disease”. The search yielded 141 epitopes, of which 14 B-cell epitopes and 2 epitopes from a hypothetical protein were removed. The remaining identified epitopes were subsequently used to generate an epitope map of the respective conserved regions.

### 3. Results

#### 3.1. Open Reading Frames and Sequence Isolates for CoV-2-Cons Sequence Creation

For creation of the CoV-2 Consensus sequence, nucleotide sequences from 1731 SARS-CoV-2 genomes were aligned and a full genome nucleotide consensus was created, 23 open reading frames (ORF) were then located in the alignment using the NC\_045512.2 and the Finkel et al. [46] coordinates and translated to amino acids. Of the 23 ORF, 12 were canonical ORF as annotated in NC\_045512.2 and 11 in alternative reading frames described by Finkel et al. [46] (Table 1). In addition, the membrane protein glycoprotein (M), is completely embedded inside an extended ORF (exORFM) without any frameshifts and was not used for separate OLP set design.

**Table 1.** Canonical and alternative open reading frames (ORF) in SARS-CoV-2. iORF: internal ORF, extORF: extended ORF, upORF: upstream ORF.

Gene	Start	End	Protein	Protease Products	Frame
ORF1a.iORF1.ext	59	136	upORF1a1	-	Alternative
ORF1a.iORF2.ext	163	264	upORF1a2	-	Alternative
ORF1ab	266	13483	pp1a	leader protein nsp2 nsp3 nsp4 3C-like proteinase nsp6 nsp7 nsp8 nsp9 nsp10 nsp11	Canonical
ORF1ab	13468	21555	pp1ab	RNA-dependent RNA polymerase helicase 3'-to-5' exonuclease endoRNase 2'-O-ribose methyltransferase	Canonical
S	21563	25384	surface glycoprotein	S1 S2	Canonical
ORFS.iORF1	21744	21863	inORFS	-	Alternative
ORF3a	25393	26220	ORF3a protein	-	Canonical
ORF3a.iORF1	25457	25582	inORF3a1	-	Alternative
ORF3a.iORF2	25596	25697	inORF3a2	-	Alternative
E	26245	26472	envelope protein	-	Canonical
ORFM.ext	26484	27191	exORFM	-	Alternative
M	26523	27191	membrane glycoprotein	-	Canonical
ORFM.iORF	27151	27195	inORFM	-	Alternative
ORF6	27202	27387	ORF6 protein	-	Canonical

**Table 1.** *Cont.*

Gene	Start	End	Protein	Protease Products	Frame
ORF7a	27394	27759	ORF7a protein	-	Canonical
ORF7b	27756	27887	ORF7b protein	-	Canonical
ORF7b.iORF2	27862	27897	inORF7b	-	Alternative
ORF8	27894	28259	ORF8 protein	-	Canonical
ORF8.iORF	27965	27994	inORF8	-	Alternative
N	28274	29533	nucleocapsid phosphoprotein	-	Canonical
ORFN.iORF1	28284	28577	ORF9b	-	Alternative
ORF10.upORF	29538	29570	upORF10	-	Alternative
ORF10	29558	29674	ORF10 protein	-	Canonical

ORF position is referred to the NC\_045512.2 reference sequence.

### 3.2. Overlapping Peptides (OLP) Sets Design

In order to achieve a balance between the number of peptides needed to cover the whole SARS-CoV-2 proteome, the costs for peptide synthesis and the design of peptide sets that allow for detecting T cell responses with high sensitivity, three OLP sets were designed (Table 2). Shorter peptides (15 mers) with longer sequence overlap between adjacent OLP (11 amino acids) offer high resolution detection of responses, thus lowering the risk of missing longer epitopes located in the OLP overlap. The consequence, however, will be a higher number of peptides to synthesize and screen, in this case a set of 2821 OLP. When the overlap between OLP was reduced from 11 amino acids to 10, the sensitivity of OLP testing is maintained, but some longer epitopes located in the overlap of two OLP may be missed. With this caveat in mind, an OLP set of 15-mers overlapping by 10 residues helped reduce the number of peptides needed by 560 OLP (total number OLP required 2262). Similarly, longer peptides (18 mers) significantly reduce the number of OLP to be synthesized, but tend to reduce in vitro sensitivity [55]. This approach, with an 11 mer overlap, reduced the number of needed OLP to 1561. The final decision for a specific design may also be driven by the assay system used for screening, an a-priori focus on fewer or more viral proteins and the available cells and funding to test immunogenicity. The three full OLP sets with their entropies are included in Table S1. Of note, the 15–11 OLP sequences were subjected to a search for homologies in the human genome to predict molecular mimicry events related to the autoimmune process. A blastp search (>8aa consecutive identical amino acids per OLP) of the whole set against the human genome yielded no hits.

**Table 2.** Description of the three CoV-2 OLP sets.

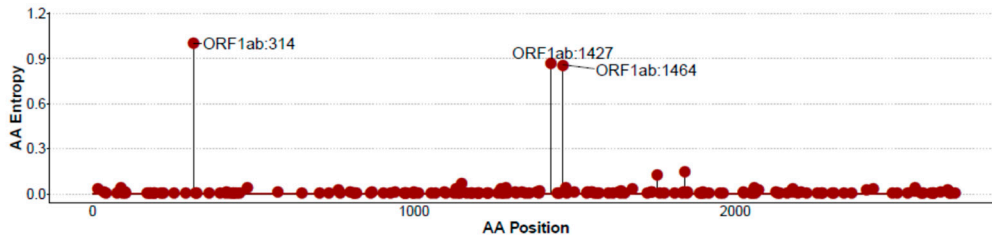
Set	Length	Overlapp	Number	Variants
15–11	15	11	2821	31
15–10	15	10	2262	23
18–11	18	11	1561	22

### 3.3. CoV-2-Cons Variability Analysis by Entropy Scores across the Full Genome

Mismatches between the sequence of in vitro antigen sets and the autologous virus in an infected individual can lead to missed responses. This has been described for highly variable pathogens, such as HCV and HIV, and showed a direct relationship between sequence entropy and the frequency of detected responses [56,57]. Even though the variability of SARS-CoV-2 reported is substantially lower than for HIV and HCV, the sequence entropy was calculated at the amino acid level and as the mean OLP entropy in order to identify positions and OLP that may escape detection in T cell screening assays.

Amino acid positional Shannon entropies were generally highly conserved, although specific more variable positions were identified (Figure S1), linked to specific amino acid variants. The ORF1ab

protein, including three of the most variable positions, is shown in Figure 1. In the CoV-2-cons 15–11 OLP set, mean OLP normalized entropies were overall low (Range: 0.947–0.758) and comparable between OLP covering the canonical ORF (Range: 0.947–0.879) and OLP matching the alternative frameshift ORF (Range: 0.932–0.758).



**Figure 1.** Standard Shannon entropy plot by amino acid position for ORF1ab. Zero entropy indicates total conservation at each specific position.

### 3.4. Variant OLP Sequences to Cover CoV-2 Sequence Diversity

Based on the SARS-CoV-2 alignment used to design the consensus, only nine amino acid positions in the entire SARS-CoV-2 genome showed two amino acids present in at least 25% of the sequences (Figure 2). Three of them were located in ORF1ab, one in the RNA polymerase and two in the Helicase sub-proteins. None of them were located close enough to each other to affect the same OLP. Still, the synthesis of a single consensus peptide could miss T cell responses in individuals exposed to the virus with the subdominant sequence variant. To prevent missing responses, a small number of additional OLP containing each of the variants were generated to cover the variability of these OLP, creating an additional set of 31 different variant OLP in the 15–11 OLP set (Table 2).



**Figure 2.** Sequence Logos for epitopes encompassing variable (>25%) positions. Protein location and starting amino acid positions are indicated on top of the logo.

### 3.5. Conserved Protein Sequences Matching Other Coronavirus Family Member and Identification of Pan-Coronavirus Sequences

In addition to variable positions, we also evaluated the presence of protein regions conserved among coronavirus species, as these may support the design of immunogen sequences for pan-coronavirus vaccines. A total of 26 regions, ranging from 8 to 23 amino acids, were identified as being conserved in at least one of the three different sequence alignments (Table 3). Fifteen fragments were identified in the pan-coronavirus alignment, 17 in the beta-coronavirus alignment and 12 in the human coronavirus alignment. Seven of them were detected in all three alignments. To identify potential T cell epitopes in these conserved regions, we searched the IEDB for described T-cell epitopes similar (>90% sequence identity) to the conserved peptides present in the CoV-2 consensus sequence. Interestingly, the majority of the conserved regions contained several matches, most of which were described epitopes derived from SARS-CoV. In total, 125 similar epitopes were identified, from all but two of the conserved regions (Table 3). The similar epitopes were found to be derived from the following organisms; SARS-CoV: 71, Human coronavirus 229E: 1, Alphacoronavirus 1: 1, Unknown origin: 3, and Homo sapiens: 47. Interestingly, 24 out of 26 fragments contained the described SARS-CoV T cell epitopes, indicating that these regions are immunogenic in humans and reinforcing the idea that some degree of cross-reactivity among coronavirus can be expected [11,58]. Also, the majority, i.e., 40 of the 47 human epitopes, clustered around one single region conserved in the beta-coronavirus alignment (QGPPGTGKSH). Several conserved peptides have thus been identified, which could potentially contain epitopes cross-reactive among different Coronavirus species. These conserved peptides can thus provide valuable information to understand if the immune response to SARS-CoV-2 is affected by previous infection with other coronaviruses and for pan-coronavirus vaccine design (Figure S2).

**Table 3.** Conserved sequences among different coronavirus. I: Pan-coronavirus, II: Betacoronavirus, III: Human coronavirus alignment. The black squares that indicated which alignments contained the conserved sequences.

Consensus Sequence	ORF	Consensus Start Position	Alignment Hit				Epitopes		
			I	II	III	Unknown	SARS-CoV	Human	Other Coronavirus
VGVLTLDNQDLNG	ORF1b	193	■	■	■	1	4	-	-
TQMNLKYAISAKNRARTVAGVSI	ORF1b	530	■	■	■	-	5	2	-
VIGTSKFYGGW	ORF1b	580	■	■	■	-	3	-	-
LMGWDYPKCDRAMPN	ORF1b	605	■	■	■	1	3	-	-
LANECAQVL	ORF1b	646	■	■	■	-	1	-	-
YVKPGCTSSGDATTA	ORF1b	665	■	■	■	-	3	-	-
KHFSMMILSDDAVVCFN	ORF1b	743	■	■	■	-	2	1	-
LYYQNNVFMS	ORF1b	778	■	■	■	-	-	-	-
GPHEFCSQHT	ORF1b	800	■	■	■	-	2	-	-
LPYPDPSRIL	ORF1b	820	■	■	■	-	2	3	-
ERFVSLAIDAYPL	ORF1b	849	■	■	■	-	5	-	1
SQTSLRCG	ORF1b	934	■	■	■	-	1	-	-
LYLGGMSYY	ORF1b	986	■	■	■	-	3	-	-
LKLFAAET	ORF1b	1054	■	■	■	-	4	-	-
QGPPGTGKSH	ORF1b	1205	■	■	■	1	2	40	-
TACSHAAVDALCEKA	ORF1b	1231	■	■	■	-	1	-	-
GDPAQLPAPR	ORF1b	1324	■	■	■	-	3	-	-
AVFISPYNSQN	ORF1b	1432	■	■	■	-	4	1	-
NRFNVAITRA	ORF1b	1483	■	■	■	-	2	-	-
CNLGGAVC	ORF1b	2002	■	■	■	-	1	-	-
KYTQLCQYLN	ORF1b	2443	■	■	■	-	3	-	-
RSFIEDLLF	Spike	815	■	■	■	-	2	-	-
QIDRLITGRL	Spike	993	■	■	■	-	5	-	1
KWPWYIWL	Spike	1211	■	■	■	-	-	-	-
WSFNPETN	M	110	■	■	■	-	3	-	-
PRWYFYLLGTGP	N	106	■	■	■	-	7	-	-

#### 4. Discussion

We here report the design of a CoV-2-cons sequence and the matched OLP sets for the comprehensive analysis of the adaptive T cell immune response against SARS-CoV-2. Three sets of OLP reported here provide enough flexibility to balance exhaustive screening for T cell responses and available resources. Ideally, the wide use of such a CoV-2-cons sequence and a specific OLP set (ideally 15 mer with 11 overlap) would ensure the comparability and reproducibility of immunological data across laboratories worldwide to accelerate SARS-CoV-2 immunological studies.

Fifteen-mer designs allow sensitive screens for both, CD4+ and CD8+ T cell responses while 18 mer allow for cheaper peptide synthesis and require less cells for comprehensive screenings. However, longer test peptides tend to yield fewer responses and imply bigger efforts for subsequent epitope mapping. For the 15 mer design, an alternative 10 amino acid overlap was proposed to reduce peptide synthesis, while maintaining the sensitivity. This approach may be valuable, but may miss epitopes restricted by HLA class I molecules known to presented longer peptides (such as HLA-B\*27, -B\*57 and others). Regardless of the final OLP design, the use of large OLP data sets for immune screening raises several challenges. How to pool peptides in suitable numbers may depend on the downstream analyses, whether or not subsequent epitope identification are planned, on the experimental setup and whether long incubation periods will be required. The latter may be especially important as pooling of a large number of peptides will possibly require lyophilization of the pooled peptides to eliminate dimethyl sulfoxide (DMSO) as this can be toxic for the cells during culture [11]. Also, as we gain more insights into the distribution of virus-specific T cell responses across the full proteome, more or less reactive regions can be pooled based on expected reactivity, protein expression level, and/or degree of conservation [46].

Canonical and alternative frame ORF were considered in the present CoV-2-consensus sequence design to ensure an as broad as possible screening for all potentially expressed protein sequences. Whether all these putative ORF are indeed expressed remains to be confirmed. If shown that not all these sequences are indeed expressed, the OLP set could be reduced by some 65 peptides, focusing exclusively on the canonical ORF. Consensus sequence design is highly dependent on the sequences included in the alignments used to construct them. We used publicly available sequences in the growing SARS-CoV-2 NCBI repository as a representative set of worldwide sequences. As noted, coverage of sequence diversity for in-vitro antigen test sets is critical as responses to autologous viral variants may be missed if these variant sequences are not matched [27]. This may be most critical for highly variable pathogens, such as HCV and HIV, where it has been shown that sequence entropy was directly related to the frequency of OLP reactivity in vitro and essential to identify the potential emergence of immune escape variants [59,60]. However, even genetically more stable pathogens such DNA viruses (for instance Epstein Barr Virus, EBV) have been reported to exist as a swarm of quasi-species and to lose specific T cell epitopes over time [61,62]. This is also supported by recent data showing some degree of adaptation to host immunity and sequence variability for SARS-CoV-2 as it moves through the global human population [63]. To cover these variant sites, variant OLP can be synthesized. An alternative approach to the synthesis of individual variant peptide sequences is the use of “toggled peptides”, where the sequence variation is directly incorporated into the peptide synthesis. To achieve this, peptide synthesis uses mixes of amino acids at variable positions, so that the resulting OLP resembles a mini-peptide library that can achieve an a-priori set coverage of circulating viral variants [64]. This would readily allow to cover more sequence diversity beyond the 25% frequency cut-off that was applied in the present study.

The existence of protein fragments conserved among different coronavirus species has several implications. For the interpretation of T cell responses, it has to be taken into account that some degree of cross-reactivity can exist among human coronavirus [5,65]. This implies that responses to these regions could be associated with previous infections by other human coronaviruses, some of them triggering much milder infections that can pass unnoticed, like those by coronaviruses causing a common cold. This observation will need to be taken into consideration when interpreting immune



data on SARS-CoV-2. On the other hand, the existence of conserved sequences among beta- or even the whole coronavirus family suggests that T cell responses to these regions could provide broad protection and that the creation of a pan-coronavirus vaccine may be feasible. Such a vaccine could allow to prevent infection not only with SARS-CoV-2, but also with other, clinically relevant coronavirus like SARS-CoV-1 and MERS, and even with new coronaviruses jumping the species barrier to humans. However, the design of a pan-coronavirus vaccine will critically depend on the identification of epitopes shared among them. These pan-coronavirus epitopes are likely to exist in conserved sequences, but need to be experimentally validated. At the same time, the existence of SARS-CoV-2 homologous regions in the human genome, together with the existence of described epitopes in these regions raise some concern that coronaviruses could be involved in a molecular mimicry process triggering autoimmune diseases like the Guillain-Barré syndrome [66–69].

The present study is currently limited to the design of the CoV-2 consensus sequence, without functional immune analyses of the OLP sets in samples from infected individuals. However, the principal aim here was to provide a SARS-CoV-2 T cell test reagent, including all described ORF and covering as much viral variability as possible, for its implementation in future screening efforts. In addition, the OLP sets will certainly elicit T cell responses *in vitro* as partial evaluation has been performed by others in studies using peptides spanning some of the regions covered by the present consensus sequence [5,9,11] and since the current peptide designs (length, overlap) has been shown to be effective in the past [55,70]. Thus, the present peptide designs will afford a high-resolution analysis of the T cell response to SARS-CoV-2, the nature of the targeted epitopes and the functionality and T cell receptor use of the T cells targeting these epitopes, thereby increasing our knowledge of factors that drive COVID-19 disease progression and which could be implemented in vaccine development.

## 5. Conclusions

We here present the first SARS-CoV-2 Consensus sequence for all described SARS-CoV-2 ORF, including those in alternative frames covering the SARS-CoV-2 sequence variability represented by 1700 available sequences. The description of this sequence and of the matching OLP sets will aid the further immune analyses in SARS-CoV-2 infection and ensure reproducibility between laboratories. In light of recent studies, the T cell response to SARS-CoV-2 can be crucial to control SARS-CoV-2 infection. To date, published studies are generally limited to a few viral proteins, using recall antigens that do not reflect sequence diversity nor alternative ORFs. To overcome these limitations, the description of the global landscape of T cell responses to SARS-CoV-2 urgently needs unbiased, comparable, full-proteome screens for virus-specific T cell responses. The CoV-2-cons and matched OLP sets described here will allow to integrate data globally, generating crucial information for vaccine development. We also include measures of sequence entropy to identify the most variable segments and design additional OLP sequences that cover these sites. Of note, these entropy analyses, together with sequence alignments across a wide range of coronaviruses, also allowed the identification of highly conserved regions among different coronaviruses. These regions may be targeted by T cells, which could target a wide range of coronaviruses and may be relevant targets for T cell vaccine design.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2076-393X/8/3/444/s1>, Figure S1: Shannon entropy plot by amino acid position for all canonical and alternative frame ORF of SARS-CoV-2, Figure S2: SARS-CoV-2 ORF fragments containing conserved regions, Table S1: Overlapping peptide lists.

**Author Contributions:** Conceptualization, A.O.; Data curation, A.O., M.N.-J., A.K. and L.R.-M.; Formal analysis, A.O., M.N.-J., A.K. and L.R.-M.; Funding acquisition, J.G.P. and C.B.; Investigation, A.K. and J.G.P.; Methodology, A.O., M.N.-J., A.K., L.R.-M., J.G.P. and C.B.; Supervision, J.G.P. and C.B.; Visualization, A.O., M.N.-J., A.K., L.R.-M., J.G.P. and C.B.; Writing—original draft, A.O., M.N.-J., A.K., L.R.-M., J.G.P. and C.B.; Writing—review & editing, A.O., M.N.-J., A.K., L.R.-M., J.G.P. and C.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported in part by grants from the National Health Institute Carlos III (ISCIII) COV20/00660, PI17/000164 and RETIC RD16/0025/0041 (Co-funded by European Regional Development Fund/European Social Fund) for J.G.P. The funders had no role in study design, data collection and analysis, the

decision to publish or drafting of the manuscript. This study has received partial funding from Grifols and the crowdfunding initiative YoMeCorono.

**Conflicts of Interest:** The authors declare that a patent application (application number 63051925) has been submitted that covers the CoV-2-cons sequence.

## References

1. Channappanavar, R.; Zhao, J.; Perlman, S. T cell-mediated immune response to respiratory coronaviruses. *Immunol. Res.* **2014**, *59*, 118–128. [[CrossRef](#)]
2. Zhao, J.; Alshukairi, A.N.; Baharoon, S.A.; Ahmed, W.A.; Bokhari, A.A.; Nehdi, A.M.; Layqah, L.A.; Alghamdi, M.G.; Al Gethamy, M.M.; Dada, A.M.; et al. Recovery from the Middle East respiratory syndrome is associated with antibody and T cell responses. *Sci. Immunol.* **2017**, *2*. [[CrossRef](#)]
3. Vabret, N.; Britton, G.J.; Gruber, C.; Hegde, S.; Kim, J.; Kuksin, M.; Levantovsky, R.; Malle, L.; Moreira, A.; Park, M.D.; et al. Immunology of COVID-19: Current State of the Science. *Immunity* **2020**, *52*, 910–941. [[CrossRef](#)]
4. Liu, W.J.; Zhao, M.; Liu, K.; Xu, K.; Wong, G.; Tan, W.; Gao, G.F. T-cell immunity of SARS-CoV: Implications for vaccine development against MERS-CoV. *Antiviral Res.* **2017**, *137*, 82–92. [[CrossRef](#)]
5. Le Bert, N.; Tan, A.T.; Kunasegaran, K.; Tham, C.Y.L.; Hafezi, M.; Chia, A.; Chng, M.H.Y.; Lin, M.; Tan, N.; Linster, M.; et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **2020**, 1–10. [[CrossRef](#)]
6. Wu, F.; Wang, A.; Liu, M.; Wang, Q.; Chen, J.; Xia, S.; Ling, Y.; Zhang, Y.; Xun, J.; Lu, L.; et al. Neutralizing Antibody Responses to SARS-CoV-2 in a COVID-19 Recovered Patient Cohort and Their Implications. *medRxiv* **2020**. [[CrossRef](#)]
7. Sekine, T.; Perez-Potti, A.; Rivera-Ballesteros, O.; Straling, K.; Gorin, J.-B.; Olsson, A.; Llewellyn-Lacey, S.; Kamal, H.; Bogdanovic, G.; Muschiol, S.; et al. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *bioRxiv* **2020**. [[CrossRef](#)]
8. Robbiani, D.; Gaebler, C.; Muecksch, F.; Lorenzi, J.; Wang, Z.; Cho, A.; Agudelo, M.; Barnes, C.; Gazumyan, A.; Finkin, S.; et al. Convergent Antibody Responses to SARS-CoV-2 Infection in Convalescent Individuals. *bioRxiv* **2020**. [[CrossRef](#)]
9. Peng, Y.; Mentzer, A.J.; Liu, G.; Yao, X.; Yin, Z.; Dong, D.; Dejnirattisai, W.; Rostron, T.; Supasa, P.; Liu, C.; et al. Broad and strong memory CD4 + and CD8 + T cells induced by SARS-CoV-2 in UK convalescent COVID-19 patients. *bioRxiv* **2020**. [[CrossRef](#)]
10. Ju, B.; Zhang, Q.; Ge, J.; Wang, R.; Sun, J.; Ge, X.; Yu, J.; Shan, S.; Zhou, B.; Song, S.; et al. Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* **2020**. [[CrossRef](#)]
11. Grifoni, A.; Weiskopf, D.; Ramirez, S.I.; Mateus, J.; Dan, J.M.; Rydyznski Moderbacher, C.; Rawlings, S.A.; Sutherland, A.; Premkumar, L.; Jadi, R.S.; et al. Journal Pre-Proof Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **2020**, *181*. [[CrossRef](#)] [[PubMed](#)]
12. Gallais, F.; Velay, A.; Wendling, M.-J.; Nazon, C.; Partisani, M.; Sibilia, J.; Candon, S.; Fafi-Kremer, S. Intrafamilial Exposure to SARS-CoV-2 Induces Cellular Immune Response without Seroconversion. *medRxiv* **2020**. [[CrossRef](#)]
13. Seow, J.; Graham, C.; Merrick, B.; Acors, S.; Steel, K.J.A.; Hemmings, O.; O’Byrne, A.; Kouphou, N.; Pickering, S.; Galao, R.; et al. Longitudinal evaluation and decline of antibody responses in SARS-CoV-2 infection. *medRxiv* **2020**. [[CrossRef](#)]
14. Ahmed, S.F.; Quadeer, A.A.; McKay, M.R. Preliminary identification of potential vaccine targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses* **2020**, *12*, 254. [[CrossRef](#)] [[PubMed](#)]
15. Baruah, V.; Bose, S. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. *J. Med. Virol.* **2020**, *92*, 495–500. [[CrossRef](#)]
16. Bhattacharya, M.; Sharma, A.R.; Patra, P.; Ghosh, P.; Sharma, G.; Patra, B.C.; Lee, S.S.; Chakraborty, C. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. *J. Med. Virol.* **2020**, *92*, 618–631. [[CrossRef](#)]

17. Gao, A.; Chen, Z.; Segal, F.P.; Carrington, M.M.; Streeck, H.; Chakraborty, A.K.; Juelg, B.; Julg, B. Predicting the Immunogenicity of T cell epitopes: From HIV to SARS-CoV-2. *bioRxiv* **2020**. [[CrossRef](#)]
18. Grifoni, A.; Sidney, J.; Zhang, Y.; Scheuermann, R.H.; Peters, B.; Sette, A. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* **2020**, *27*, 671–680.e2. [[CrossRef](#)]
19. Lucchese, G. Epitopes for a 2019-nCoV vaccine. *Cell. Mol. Immunol.* **2020**, *17*, 539–540. [[CrossRef](#)]
20. Silva-Arrieta, S.; Goulder, P.J.R.; Brander, C. In silico veritas? Potential limitations for SARS-CoV-2 vaccine development based on T-cell epitope prediction. *PLoS Pathog.* **2020**, *16*, e1008607. [[CrossRef](#)]
21. Rivino, L.; Tan, A.T.; Chia, A.; Kumaran, E.A.P.; Grotenbreg, G.M.; MacAry, P.A.; Bertoletti, A. Defining CD8 + T Cell Determinants during Human Viral Infection in Populations of Asian Ethnicity. *J. Immunol.* **2013**, *191*, 4010–4019. [[CrossRef](#)] [[PubMed](#)]
22. Carragher, D.M.; Kaminski, D.A.; Moquin, A.; Hartson, L.; Randall, T.D. A Novel Role for Non-Neutralizing Antibodies against Nucleoprotein in Facilitating Resistance to Influenza Virus. *J. Immunol.* **2008**, *181*, 4168–4176. [[CrossRef](#)] [[PubMed](#)]
23. Cardinaud, S.; Consiglieri, G.; Bouziat, R.; Urrutia, A.; Graff-Dubois, S.; Fourati, S.; Malet, I.; Guergnon, J.; Guihot, A.; Katlama, C.; et al. CTL escape mediated by proteasomal destruction of an HIV-1 cryptic epitope. *PLoS Pathog.* **2011**, *7*, e1002049. [[CrossRef](#)] [[PubMed](#)]
24. Bansal, A.; Carlson, J.; Yan, J.; Akinsiku, O.T.; Schaefer, M.; Sabbaj, S.; Bet, A.; Levy, D.N.; Heath, S.; Tang, J.; et al. CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *J. Exp. Med.* **2010**, *207*, 51–59. [[CrossRef](#)] [[PubMed](#)]
25. Carlson, T.L.; Green, K.A.; Green, W.R. Alternative translational reading frames as a novel source of epitopes for an expanded CD8 T-cell repertoire: Use of a retroviral system to assess the translational requirements for CTL recognition and lysis. *Viral Immunol.* **2010**, *23*, 577–583. [[CrossRef](#)]
26. Berger, C.T.; Carlson, J.M.; Brumme, C.J.; Hartman, K.L.; Brumme, Z.L.; Henry, L.M.; Rosato, P.C.; Piechocka-Trocha, A.; Brockman, M.A.; Harrigan, P.R.; et al. Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences. *J. Exp. Med.* **2010**, *207*, 61–75. [[CrossRef](#)]
27. Altfeld, M.; Addo, M.M.; Shankarappa, R.; Lee, P.K.; Allen, T.M.; Yu, X.G.; Rathod, A.; Harlow, J.; O’Sullivan, K.; Johnston, M.N.; et al. Enhanced Detection of Human Immunodeficiency Virus Type 1-Specific T-Cell Responses to Highly Variable Regions by Using Peptides Based on Autologous Virus Sequences. *J. Virol.* **2003**, *77*, 7330–7340. [[CrossRef](#)]
28. Nickle, D.C.; Rolland, M.; Jensen, M.A.; Kosakovsky Pond, S.L.; Deng, W.; Seligman, M.; Heckerman, D.; Mullins, J.I.; Jojic, N. Coping with viral diversity in HIV vaccine design. *PLoS Comput. Biol.* **2007**, *3*, 754–762. [[CrossRef](#)]
29. Rolland, M.; Manochewewa, S.; Swain, J.V.; Lanxon-Cookson, E.C.; Kim, M.; Westfall, D.H.; Larsen, B.B.; Gilbert, P.B.; Mullins, J.I. HIV-1 Conserved-Element Vaccines: Relationship between Sequence Conservation and Replicative Capacity. *J. Virol.* **2013**, *87*, 5461–5467. [[CrossRef](#)]
30. Malhotra, U.; Nolin, J.; Mullins, J.I.; McElrath, M.J. Comprehensive epitope analysis of cross-clade Gag-specific T-cell responses in individuals with early HIV-1 infection in the US epidemic. *Vaccine* **2007**, *25*, 381–390. [[CrossRef](#)]
31. Kesturu, G.S.; Colleton, B.A.; Liu, Y.; Heath, L.; Shaikh, O.S.; Rinaldo, C.R.; Shankarappa, R. Minimization of genetic distances by the consensus, ancestral, and center-of-tree (COT) sequences for HIV-1 variants within an infected individual and the design of reagents to test immune reactivity. *Virology* **2006**, *348*, 437–448. [[CrossRef](#)] [[PubMed](#)]
32. Rolland, M.; Jensen, M.A.; Nickle, D.C.; Yan, J.; Learn, G.H.; Heath, L.; Weiner, D.; Mullins, J.I. Reconstruction and Function of Ancestral Center-of-Tree Human Immunodeficiency Virus Type 1 Proteins. *J. Virol.* **2007**, *81*, 8507–8514. [[CrossRef](#)] [[PubMed](#)]
33. Ross, H.A.; Nickle, D.C.; Liu, Y.; Heath, L.; Jensen, M.A.; Rodrigo, A.G.; Mullins, J.I. Sources of variation in ancestral sequence reconstruction for HIV-1 envelope genes. *Evol. Bioinform. Online* **2007**, *2*, 53–76. [[CrossRef](#)] [[PubMed](#)]
34. Bansal, A.; Gough, E.; Ritter, D.; Wilson, C.; Mulenga, J.; Allen, S.; Goepfert, P.A. Group M-based HIV-1 Gag peptides are frequently targeted by T cells in chronically infected US and Zambian patients. *AIDS* **2006**, *20*, 353–360. [[CrossRef](#)] [[PubMed](#)]

35. Arenas, M.; Posada, D. Computational Design of Centralized HIV-1 Genes. *Curr. HIV Res.* **2011**, *8*, 613–621. [[CrossRef](#)] [[PubMed](#)]
36. Kothe, D.L.; Li, Y.; Decker, J.M.; Bibollet-Ruche, F.; Zammit, K.P.; Salazar, M.G.; Chen, Y.; Weng, Z.; Weaver, E.A.; Gao, F.; et al. Ancestral and consensus envelope immunogens for HIV-1 subtype C. *Virology* **2006**, *352*, 438–449. [[CrossRef](#)]
37. Rutebemberwa, A.; Currier, J.R.; Jagodzinski, L.; McCutchan, F.; Birx, D.; Marovich, M.; Cox, J.H. HIV-1 MN Env 15-mer peptides better detect HIV-1 specific CD8 T cell responses compared with consensus subtypes B and M group 15-mer peptides. *AIDS* **2005**, *19*, 1165–1172. [[CrossRef](#)]
38. De Groot, A.S.; Bishop, E.A.; Khan, B.; Lally, M.; Marcon, L.; Franco, J.; Mayer, K.H.; Carpenter, C.C.J.; Martin, W. Engineering immunogenic consensus T helper epitopes for a cross-clade HIV vaccine. *Methods* **2004**, *34*, 476–487. [[CrossRef](#)]
39. Koita, O.A.; Dabitaou, D.; Mahamadou, I.; Tall, M.; Dao, S.; Tounkara, A.; Guiteye, H.; Noumsi, C.; Thiero, O.; Kone, M.; et al. Confirmation of immunogenic consensus sequence HIV-1 T-cell epitopes in Bamako, Mali and Providence, Rhode Island. *Hum. Vaccin.* **2006**, *2*, 119–128. [[CrossRef](#)]
40. Almeida, R.R.; Rosa, D.S.; Ribeiro, S.P.; Santana, V.C.; Kallás, E.G.; Sidney, J.; Sette, A.; Kalil, J.; Cunha-Neto, E. Broad and Cross-Clade CD4+ T-Cell Responses Elicited by a DNA Vaccine Encoding Highly Conserved and Promiscuous HIV-1 M-Group Consensus Peptides. *PLoS ONE* **2012**, *7*. [[CrossRef](#)]
41. Fonseca, S.G.; Coutinho-Silva, A.; Fonseca, L.A.M.; Segurado, A.C.; Moraes, S.L.; Rodrigues, H.; Hammer, J.; Kallás, E.G.; Sidney, J.; Sette, A.; et al. Identification of novel consensus CD4 T-cell epitopes from clade B HIV-1 whole genome that are frequently recognized by HIV-1 infected patients. *AIDS* **2006**, *20*, 2263–2273. [[CrossRef](#)] [[PubMed](#)]
42. Frahm, N.; Nickle, D.C.; Linde, C.H.; Cohen, D.E.; Zuniga, R.; Lucchetti, A.; Roach, T.; Walker, B.D.; Allen, T.M.; Korber, B.T.; et al. Increased detection of HIV-specific T cell responses by combination of central sequences with comparable immunogenicity. *AIDS* **2008**, *22*, 447–456. [[CrossRef](#)] [[PubMed](#)]
43. Brander, C.; Self, S.; Korber, B. Capturing viral diversity for in-vitro test reagents and HIV vaccine immunogen design. *Curr. Opin. HIV AIDS* **2007**, *2*, 183–188. [[CrossRef](#)] [[PubMed](#)]
44. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
45. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.M.; Wang, W.; Song, Z.G.; Hu, Y.; Tao, Z.W.; Tian, J.H.; Pei, Y.Y.; et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265–269. [[CrossRef](#)]
46. Finkel, Y.; Mizrahi, O.; Nachshon, A.; Weingarten-Gabbay, S.; Yahalom-Ronen, Y.; Tamir, H.; Achdout, H.; Melamed, S.; Weiss, S.; Israely, T.; et al. The coding capacity of SARS-CoV-2. *bioRxiv* **2020**. [[CrossRef](#)]
47. Grant, B.J.; Rodrigues, A.P.C.; ElSawy, K.M.; McCammon, J.A.; Caves, L.S.D. Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics* **2006**, *22*, 2695–2696. [[CrossRef](#)]
48. PeptGen Peptide Generator. Available online: <https://www.hiv.lanl.gov/content/sequence/PEPTGEN/peptgen.html> (accessed on 3 August 2020).
49. Llano, A.; Cedeño, S.; Silva Arrieta, S.; Brander, C.; Theoretical Biology and Biophysics Group. The 2019 Optimal HIV CTL epitopes update: Growing diversity in epitope length and HLA restriction. In *HIV Molecular Immunology*; Yusim, K., Korber, B., Brander, C., Barouch, D., de Boer, R., Haynes, B.F., Koup, R., Moore, J.P., Walker, B., Eds.; Los Alamos National Laboratory: Los Alamos, NM, USA, 2019.
50. Wagih, O. Ggseqlogo: A versatile R package for drawing sequence logos. *Bioinformatics* **2017**, *33*, 3645–3647. [[CrossRef](#)]
51. BLAST: Basic Local Alignment Search Tool. Available online: <https://blast.ncbi.nlm.nih.gov/Blast.cgi> (accessed on 3 August 2020).
52. Kumar, S.; Stecher, G.; Li, M.; Nnyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)]
53. WebLogo. Available online: <http://weblogo.berkeley.edu/> (accessed on 3 August 2020).
54. IEDB.org: Free Epitope Database and Prediction Resource. Available online: <http://www.iedb.org/> (accessed on 3 August 2020).
55. Draenert, R.; Altfeld, M.; Brander, C.; Basgoz, N.; Corcoran, C.; Wurcel, A.G.; Stone, D.R.; Kalams, S.A.; Trocha, A.; Addo, M.M.; et al. Comparison of overlapping peptide sets for detection of antiviral CD8 and CD4 T cell responses. *J. Immunol. Methods* **2003**, *275*, 19–29. [[CrossRef](#)]

56. Frahm, N.; Korber, B.T.; Adams, C.M.; Szinger, J.J.; Draenert, R.; Addo, M.M.; Feeney, M.E.; Yusim, K.; Sango, K.; Brown, N.V.; et al. Consistent cytotoxic-T-lymphocyte targeting of immunodominant regions in human immunodeficiency virus across multiple ethnicities. *J. Virol.* **2004**, *78*, 2187–2200. [[CrossRef](#)] [[PubMed](#)]
57. Yerly, D.; Heckerman, D.; Allen, T.M.; Chisholm, J.V.; Faircloth, K.; Linde, C.H.; Frahm, N.; Timm, J.; Pichler, W.J.; Cerny, A.; et al. Increased Cytotoxic T-Lymphocyte Epitope Variant Cross-Recognition and Functional Avidity Are Associated with Hepatitis C Virus Clearance. *J. Virol.* **2008**, *82*, 3147–3153. [[CrossRef](#)]
58. Braun, J.; Loyal, L.; Frentsch, M.; Wendisch, D.; Georg, P.; Kurth, F.; Hippenstiel, S.; Dingeldey, M.; Kruse, B.; Fauchere, F.; et al. Presence of SARS-CoV-2 reactive T cells in COVID-19 patients and healthy donors. *medRxiv* **2020**. [[CrossRef](#)]
59. Prado, J.G.; Honeyborne, I.; Brierley, I.; Puertas, M.C.; Martinez-Picado, J.; Goulder, P.J.R. Functional Consequences of Human Immunodeficiency Virus Escape from an HLA-B\*13-Restricted CD8+ T-Cell Epitope in p1 Gag Protein. *J. Virol.* **2009**, *83*, 1018–1025. [[CrossRef](#)] [[PubMed](#)]
60. Honeyborne, I.; Codoñer, F.M.; Leslie, A.; Tudor-Williams, G.; Luzzi, G.; Ndung'u, T.; Walker, B.D.; Goulder, P.J.; Prado, J.G. HLA-Cw\*03-Restricted CD8+ T-Cell Responses Targeting the HIV-1 Gag Major Homology Region Drive Virus Immune Escape and Fitness Constraints Compensated for by Intracodon Variation. *J. Virol.* **2010**, *84*, 11279–11288. [[CrossRef](#)]
61. De Campos-Lima, P.O.; Gavioli, R.; Zhang, Q.J.; Wallace, L.E.; Dolcetti, R.; Rowe, M.; Rickinson, A.B.; Masucci, M.G. HLA-A11 epitope loss isolates of Epstein-Barr virus from a highly A11 + population. *Science* **1993**, *260*, 98–100. [[CrossRef](#)]
62. Gutiérrez, M.I.; Spangler, G.; Kingma, D.; Raffeld, M.; Guerrero, I.; Misad, O.; Jaffe, E.S.; Magrath, I.T.; Bhatia, K. Epstein-Barr virus in nasal lymphomas contains multiple ongoing mutations in the EBNA-1 gene. *Blood* **1998**, *92*, 600–606. [[CrossRef](#)]
63. Li, X.; Giorgi, E.E.; Honnayakanahalli Marichann, M.; Foley, B.; Xiao, C.; Kong, X.-P.; Chen, Y.; Korber, B.; Gao, F. Emergence of SARS-CoV-2 through Recombination and Strong Purifying Selection Short Title: Recombination and origin of SARS-CoV-2 One Sentence Summary: Extensive Recombination and Strong Purifying Selection among coronaviruses from different hosts facilita. *bioRxiv* **2020**. [[CrossRef](#)]
64. Frahm, N.; Kaufmann, D.E.; Yusim, K.; Muldoon, M.; Kesmir, C.; Linde, C.H.; Fischer, W.; Allen, T.M.; Li, B.; McMahon, B.H.; et al. Increased sequence diversity coverage improves detection of HIV-specific T cell responses. *J Immunol* **2007**, *179*, 6638–6650. [[CrossRef](#)]
65. Ng, K.; Faulkner, N.; Cornish, G.; Rosa, A.; Earl, C.; Wrobel, A.; Benton, D.; Roustan, C.; Bolland, W.; Thompson, R.; et al. Pre-existing and de novo humoral immunity to SARS-CoV-2 in humans. *bioRxiv* **2020**. [[CrossRef](#)]
66. Lascano, A.M.; Epiney, J.B.; Coen, M.; Serratrice, J.; Bernard-Valnet, R.; Lalive, P.H.; Kuntzer, T.; Hübers, A. SARS-CoV-2 and Guillain-Barré syndrome: AIDP variant with favorable outcome. *Eur. J. Neurol.* **2020**. [[CrossRef](#)] [[PubMed](#)]
67. Bigaut, K.; Mallaret, M.; Baloglu, S.; Nemoz, B.; Morand, P.; Baicry, F.; Godon, A.; Voulleminot, P.; Kremer, L.; Chanson, J.-B.; et al. Guillain-Barré syndrome related to SARS-CoV-2 infection. *Neurol. Neuroimmunol. Neuroinflammation* **2020**, *7*. [[CrossRef](#)] [[PubMed](#)]
68. Riva, N.; Russo, T.; Falzone, Y.M.; Strollo, M.; Amadio, S.; Del Carro, U.; Locatelli, M.; Filippi, M.; Fazio, R. Post-infectious Guillain-Barré syndrome related to SARS-CoV-2 infection: A case report. *J. Neurol.* **2020**. [[CrossRef](#)] [[PubMed](#)]
69. Chan, J.L.; Ebadi, H.; Sarna, J.R. Guillain-Barré syndrome with facial diplegia related to SARS-CoV-2 infection. *Can. J. Neurol. Sci.* **2020**, 1–10. [[CrossRef](#)]
70. Draenert, R.; Brander, C.; Yu, X.G.; Altfeld, M.; Verrill, C.L.; Feeney, M.E.; Walker, B.D.; Goulder, P.J.R. Impact of intrapeptide epitope location on CD8 T cell recognition: Implications for design of overlapping peptide panels. *AIDS* **2004**, *18*, 871–876. [[CrossRef](#)]

