

RESEARCH ARTICLE

CD8 T cell epitope generation toward the continually mutating SARS-CoV-2 spike protein in genetically diverse human population: Implications for disease control and prevention

Elisa Guo¹, Hailong Guo^{2*}

1 Mounds View High School, Arden Hills, Minnesota, United States of America, **2** Independent Scientist, St Paul, Minnesota, United States of America

* jhguo2000@gmail.com

Abstract

The ongoing pandemic of SARS-CoV-2 has brought tremendous crisis on global health care systems and industrial operations that dramatically affect the economic and social life of numerous individuals worldwide. Understanding anti-SARS-CoV-2 immune responses in population with different genetic backgrounds and tracking the viral evolution are crucial for successful vaccine design. In this study, we reported the generation of CD8 T cell epitopes by a total of 80 alleles of three major class I HLAs using NetMHC 4.0 algorithm for the SARS-CoV-2 spike protein, which can be targeted by both B cells and T cells. We found diverse capacities of S protein specific epitope presentation by different HLA alleles with very limited number of predicted epitopes for HLA-B*2705, HLA-B*4402 and HLA-B*4403 and as high as 132 epitopes for HLA-A*6601. Our analysis of 1000 S protein sequences from field isolates collected globally over the past few months identified three recurrent point mutations including L5F, D614G and G1124V. Differential effects of these mutations on CD8 T cell epitope generation by corresponding HLA alleles were observed. Finally, our multiple alignment analysis indicated the absence of seasonal CoV induced cross-reactive CD8 T cells to drive these mutations. Our findings suggested that individuals with certain HLA alleles, such as B*44 are more prone to SARS-CoV-2 infection. Studying anti-S protein specific CD8 T cell immunity in diverse genetic background is critical for better control and prevention of the SARS-CoV-2 pandemic.

OPEN ACCESS

Citation: Guo E, Guo H (2020) CD8 T cell epitope generation toward the continually mutating SARS-CoV-2 spike protein in genetically diverse human population: Implications for disease control and prevention. PLoS ONE 15(12): e0239566. <https://doi.org/10.1371/journal.pone.0239566>

Editor: Feng Wen, Foshan University, CHINA

Received: September 3, 2020

Accepted: November 23, 2020

Published: December 10, 2020

Copyright: © 2020 Guo, Guo. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting information](#) files.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The coronavirus (CoV) is an enveloped, positive-stranded RNA virus that can cause respiratory and enteric diseases in wide range of hosts including human, numerous animals, birds and fish [1]. Four genera (Alpha, Beta, Gamma, and Delta) of CoVs have been classified with human CoVs designated within Alpha and Beta groups. Their genome, about 30kb in length, is the largest found in RNA viruses and encodes more than 20 putative proteins, including

four major structural proteins: spike (S), envelope (E), membrane (M), and nucleocapsid (N) [1]. The human seasonal CoVs are endemic throughout the world, causing 15–30% respiratory tract infections that are typically mild and self-limiting. However, the outbreak of severe acute respiratory syndrome (SARS) in 2003 [2] and Middle East respiratory syndrome (MERS) in 2012 [3] that were caused by the infection of SARS-CoV and MERS-CoV had led to a mortality rate of about 10% and 40%, respectively [1, 4].

In December 2019, a novel coronavirus disease (COVID-19), caused by 2019 novel coronavirus (2019-nCoV) and later renamed as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was identified in Wuhan City, Hubei Province, China from patients with severe pneumonia [5–7]. Subsequently, this novel viral infection has rapidly spread into nearly all countries over the world, leading to the declaration of the first-known coronavirus global pandemic by the World Health Organization (WHO) on March 11, 2020 [8]. As of November 8, 2020, the COVID-19 pandemic has resulted in over 49 million confirmed cases and more than 1.2 million deaths globally according to WHO's situation report released on Nov 10th, 2020. Like SARS-CoV and MERS-CoV, SARS-CoV-2 was identified as betacoronavirus through genome sequencing and bioinformatic analysis [6, 7] and has been formally classified by the International Committee on Taxonomy of Viruses (ITVC) [9].

The S protein of SARS-CoV contains neutralization epitopes, which are critical for eliciting anti-viral B cell immune responses. Therefore, S protein appears to be a promising antigen target for vaccine development [4, 10]. Similarly, a growing body of research has identified neutralization epitopes on SARS-CoV-2 S protein [11–13]. Although SARS-CoV neutralization antibody is critical, the neutralizing antibody titers and memory B cell response were short-lived in SARS-recovered patients [14]. Additionally, the antibodies including the neutralization antibody induced by SARS-CoV-2 infection could fade quickly [15, 16]. In addition to the anti-S protein humoral response, S protein specific CD8 T cells have been shown with protective immunity during SARS-CoV infection and are critical for vaccine efficacy [17, 18]. Therefore, identifying SARS-CoV-2 S protein specific CD8 T cell epitopes and fully utilizing CD8 T cell immunity in addition to neutralization response are urgently needed for studying anti-COVID-19 immunity and developing effective vaccines. Since its initial outbreak, SARS-CoV-2 has undergone genetic mutations [19]. Further, mutations on S proteins that may affect human ACE2 receptor binding, viral transmission and infectivity have been reported [20–22]. However, the potential effects of these mutations on CD8 T cell epitope generation have not been analyzed.

In this study, we used the NetMHC 4.0 prediction algorithm to analyze a panel of classical human MHC I (HLA-A, B and C) restricted CD8 epitopes on the S protein of SARS-CoV-2. Our data showed a significant variation of CD8 T cell epitope repertoires for different HLA-A, HLA-B and HLA-C alleles. In addition, our analysis using a large number of S protein sequences derived from SARS-CoV-2 genomes revealed important mutations including L5F, D614G and G1124V that are although less likely driven by seasonal CoV derived cross-reactive CD8 T cells, may differentially impact CD8 T cell epitope generation by different HLA alleles. Our results indicate genetic variability of three major class I HLAs may regulate S protein specific CD8 T cell responses, leading to differential susceptibility to and severity of the infection of SARS-CoV-2 that keeps mutating.

Material & methods

S protein sequence of reference virus

Full-length S protein amino acid sequence (accession number QHD43416.1) of the reference SARS-CoV-2 Wuhan-Hu-1 isolate (accession number MN908947) was downloaded in FASTA format from the NCBI GenBank.

CD8 T cell epitope prediction

CD8 T cell epitope prediction was performed on NetMHC4.0 Web Server that utilizes artificial neural networks (ANNs) [23, 24]. A total of 36 HLA-A, 34 HLA-B and 10 HLA-C alleles were used for prediction. 9-mer peptides (epitopes) with rank score $\leq 2.0\%$ were selected as positive HLA-binder. As the output format for the epitopes derived from NetMHC4.0 begins with 0, one was added to the resulting positions of all epitopes presented in this study to match the S protein sequence numbering.

Mutational analysis

To identify mutations on SARS-CoV-2 S protein, we selected a total of 1000 complete or near-complete S protein sequences of viruses isolated from countries in North America, Europe, Asia, Oceania, Africa and South America. These sequences were deposited in the NCBI database with collection dates ranging from January to June 2020 and showed at least 95% of query coverage for the reference S protein (accession number QHD43416.1). The list of the accession numbers for these sequences was provided in [S1 Table](#). Multiple alignment of these 1000 protein sequences against the reference S protein was conducted using NCBI BLASTP program under default conditions.

Comparison of CoV S protein sequences

The S protein sequences of four common seasonal CoVs: HCoV-229E (AAG48592.1), HCoV-OC43 (CAA83661.1), HCoV-NL63 (APF29063.1), and HCoV-HKU1 (BBA20983.1) were used for comparison with the reference S protein of SARS-CoV 2 by multiple alignment in MEGA-X using MUSCLE under default conditions. Phylogenetic tree analysis of the S protein of SARS-CoV-2 and four seasonal human CoVs was performed using the maximum likelihood method and JTT model with a bootstrap of 1000 replicates within the Mega-X software.

Statistical analysis

One-way analysis of variance (ANOVA) and Tukey's multiple comparison test were used for assessment of the means of predicted CD8 T cell epitopes presented by three HLA allele groups. Statistical analysis was performed using GraphPad Prism version 5.0 (GraphPad Software, Inc., San Diego, CA). A p-value < 0.05 was considered statistically significant.

Results

S protein specific CD8 T cell epitopes of different HLA alleles

The NetMHC 4.0 algorithm has been widely utilized for predicting CD8 T cell epitopes of human and animal pathogens. By using this tool, our analysis for the S protein of reference Wuhan-Hu-1 virus generated a total of 5044 CD8 T cell epitope pool for three major class I HLAs with a total of 80 different alleles ([S2 Table](#)). Among these, there are 2282 epitopes for HLA-A, 1816 for HLA-B and 946 for HLA-C ([Table 1](#)). Within HLA-A, A*6601 allele showed 132 epitopes, the highest among all, indicating a potentially broad and strong CD8 T cell responses against SARS-CoV-2 from individuals bearing this particular allele. However, there were several HLA-A alleles (*0201, A*0205, A*0207, A*0301, A*3101, A*3301) that only had about 30 epitopes. Other HLA-A alleles showed epitopes ranging from 40 to around 100 ([Table 1](#)). Similarly, within HLA-B, the total number of epitopes for individual alleles differed substantially, ranging from 16 to 118. The lowest number of epitopes among all the alleles analyzed was generated by HLA-B*2705. Additionally, HLA-B*4402 and HLA-B*4403 also had very low numbers of CD8 T cell epitopes (17 and 19 each) ([Table 1](#)). This observation may

Table 1. Summary of the total number of predicted SARS-CoV-2 S protein specific CD8 T cell epitopes for each HLA alleles.

HLA-A		HLA-B		HLA-C	
Alleles	Total	Alleles	Total	Alleles	Total
HLA-A*0101	43	HLA-B*0702	25	HLA-C*0303	84
HLA-A*0201	30	HLA-B*0801	27	HLA-C*0401	90
HLA-A*0202	42	HLA-B*0802	90	HLA-C*0501	75
HLA-A*0203	41	HLA-B*0803	79	HLA-C*0602	79
HLA-A*0205	32	HLA-B*1402	88	HLA-C*0701	101
HLA-A*0206	60	HLA-B*1501	50	HLA-C*0702	86
HLA-A*0207	25	HLA-B*1502	93	HLA-C*0802	89
HLA-A*0211	94	HLA-B*1503	118	HLA-C*1203	121
HLA-A*0212	71	HLA-B*1509	85	HLA-C*1402	104
HLA-A*0216	96	HLA-B*1517	94	HLA-C*1502	117
HLA-A*0217	36	HLA-B*1801	36		
HLA-A*0219	77	HLA-B*2705	16		
HLA-A*0250	99	HLA-B*2720	69		
HLA-A*0301	31	HLA-B*3501	75		
HLA-A*1101	41	HLA-B*3503	43		
HLA-A*2301	56	HLA-B*3801	31		
HLA-A*2402	43	HLA-B*3901	80		
HLA-A*2403	78	HLA-B*4001	24		
HLA-A*2501	77	HLA-B*4002	24		
HLA-A*2601	44	HLA-B*4013	77		
HLA-A*2602	70	HLA-B*4201	46		
HLA-A*2603	88	HLA-B*4402	17		
HLA-A*2902	51	HLA-B*4403	19		
HLA-A*3001	58	HLA-B*4501	22		
HLA-A*3002	43	HLA-B*4601	42		
HLA-A*3101	30	HLA-B*4801	71		
HLA-A*3201	72	HLA-B*5101	43		
HLA-A*3207	109	HLA-B*5301	46		
HLA-A*3215	101	HLA-B*5401	34		
HLA-A*3301	32	HLA-B*5701	24		
HLA-A*6601	132	HLA-B*5801	47		
HLA-A*6801	45	HLA-B*5802	26		
HLA-A*6802	62	HLA-B*7301	58		
HLA-A*6823	107	HLA-B*8301	97		
HLA-A*6901	96				
HLA-A*8001	70				

<https://doi.org/10.1371/journal.pone.0239566.t001>

explain why B*44 positive individuals are more prone to SARS-CoV-2 infection [25]. Furthermore, the total number of epitopes the HLA-C alleles could present varied from 75 to 121, although its mean CD8 epitopes is statistically higher than these of HLA-A and HLA-B ($p \leq 0.01$ and $p \leq 0.001$, respectively). These results indicate the capacity to present SARS-CoV-2 S protein specific CD8 T cell epitopes among different HLA alleles could be of great difference.

Mutations on S protein of SARS-CoV-2

SARS-CoV-2 has evolved since its initial outbreak by changing its viral genome, which may cause mutations in viral proteins including S protein, resulting in the establishment of better

1	M <u>F</u> V <u>F</u> <u>L</u> V <u>L</u> L <u>P</u> L <u>V</u> S <u>S</u> Q <u>C</u> V <u>N</u> L <u>T</u> T <u>R</u> T <u>Q</u> L <u>P</u> P <u>A</u> Y <u>T</u> N <u>S</u> F <u>T</u> R <u>G</u> V <u>Y</u> Y <u>P</u> D <u>K</u> V <u>F</u> R <u>S</u> S <u>V</u> L <u>H</u> S <u>T</u> Q <u>D</u> L <u>F</u> L <u>P</u> F <u>F</u> S	60
61	N <u>V</u> T <u>W</u> F <u>H</u> A <u>I</u> H <u>V</u> S <u>G</u> T <u>N</u> G <u>T</u> K <u>R</u> F <u>D</u> N <u>P</u> V <u>L</u> P <u>F</u> N <u>D</u> G <u>V</u> Y <u>F</u> A <u>S</u> T <u>E</u> K <u>S</u> N <u>I</u> I <u>R</u> G <u>W</u> I <u>F</u> G <u>T</u> T <u>L</u> D <u>S</u> K <u>T</u> Q <u>S</u> L <u>L</u> I <u>V</u>	120
121	N <u>N</u> A <u>T</u> N <u>V</u> I <u>K</u> V <u>C</u> E <u>F</u> Q <u>F</u> C <u>N</u> D <u>P</u> F <u>L</u> G <u>V</u> Y <u>Y</u> H <u>K</u> N <u>N</u> K <u>S</u> W <u>M</u> E <u>S</u> E <u>F</u> R <u>V</u> Y <u>S</u> S <u>A</u> N <u>N</u> C <u>T</u> F <u>E</u> Y <u>V</u> S <u>Q</u> P <u>F</u> L <u>M</u> D <u>L</u> E	180
181	G <u>K</u> Q <u>G</u> N <u>F</u> K <u>N</u> L <u>R</u> E <u>F</u> V <u>F</u> K <u>N</u> I <u>D</u> G <u>Y</u> F <u>K</u> I <u>Y</u> S <u>K</u> H <u>T</u> P <u>I</u> N <u>L</u> V <u>R</u> D <u>L</u> P <u>Q</u> G <u>F</u> S <u>A</u> L <u>E</u> P <u>L</u> V <u>D</u> L <u>P</u> I <u>G</u> I <u>N</u> I <u>T</u> R <u>F</u> Q <u>T</u>	240
241	L <u>L</u> A <u>L</u> H <u>R</u> S <u>Y</u> L <u>T</u> P <u>G</u> D <u>S</u> S <u>S</u> G <u>W</u> T <u>A</u> G <u>A</u> A <u>Y</u> Y <u>V</u> G <u>Y</u> L <u>Q</u> P <u>R</u> T <u>F</u> L <u>L</u> K <u>Y</u> N <u>E</u> N <u>G</u> T <u>I</u> T <u>D</u> A <u>V</u> D <u>C</u> A <u>L</u> D <u>P</u> L <u>S</u> E <u>T</u> K	300
301	C <u>T</u> L <u>K</u> S <u>F</u> T <u>V</u> E <u>K</u> G <u>I</u> Y <u>Q</u> T <u>S</u> N <u>F</u> R <u>V</u> Q <u>P</u> T <u>E</u> S <u>I</u> V <u>R</u> F <u>P</u> N <u>I</u> T <u>N</u> L <u>C</u> P <u>F</u> G <u>E</u> V <u>F</u> N <u>A</u> T <u>R</u> F <u>A</u> S <u>V</u> Y <u>A</u> W <u>N</u> R <u>K</u> R <u>I</u> S	360
361	C <u>V</u> A <u>D</u> Y <u>S</u> V <u>L</u> Y <u>N</u> S <u>A</u> S <u>F</u> S <u>T</u> F <u>K</u> C <u>Y</u> G <u>V</u> S <u>P</u> T <u>K</u> L <u>N</u> D <u>L</u> C <u>F</u> T <u>N</u> V <u>Y</u> A <u>D</u> S <u>F</u> V <u>I</u> R <u>G</u> D <u>E</u> V <u>R</u> Q <u>I</u> A <u>P</u> G <u>Q</u> T <u>G</u> K <u>I</u> A <u>D</u>	420
421	Y <u>N</u> Y <u>K</u> L <u>P</u> D <u>D</u> F <u>T</u> G <u>C</u> V <u>I</u> A <u>W</u> N <u>S</u> N <u>N</u> L <u>D</u> S <u>K</u> V <u>G</u> G <u>N</u> Y <u>N</u> L <u>Y</u> R <u>L</u> F <u>R</u> K <u>S</u> N <u>L</u> K <u>P</u> F <u>E</u> R <u>D</u> I <u>S</u> T <u>E</u> I <u>Y</u> Q <u>A</u> G <u>S</u> T <u>P</u> C	480
481	N <u>G</u> V <u>E</u> G <u>F</u> N <u>C</u> Y <u>F</u> P <u>L</u> Q <u>S</u> Y <u>G</u> F <u>Q</u> P <u>T</u> N <u>G</u> V <u>G</u> Y <u>Q</u> P <u>Y</u> R <u>V</u> V <u>V</u> L <u>S</u> F <u>E</u> L <u>L</u> H <u>A</u> P <u>A</u> T <u>V</u> C <u>G</u> P <u>K</u> K <u>S</u> T <u>N</u> L <u>V</u> K <u>N</u> K <u>C</u> V	540
541	F <u>N</u> F <u>N</u> G <u>L</u> T <u>G</u> T <u>G</u> V <u>L</u> T <u>E</u> S <u>N</u> K <u>K</u> F <u>L</u> P <u>F</u> Q <u>Q</u> F <u>G</u> R <u>D</u> I <u>A</u> D <u>T</u> T <u>D</u> A <u>V</u> R <u>D</u> P <u>Q</u> T <u>L</u> E <u>I</u> L <u>D</u> I <u>T</u> P <u>C</u> S <u>F</u> G <u>G</u> V <u>S</u> V <u>I</u> T	600
601	G <u>T</u> N <u>T</u> S <u>N</u> Q <u>V</u> A <u>V</u> L <u>Y</u> Q <u>D</u> V <u>N</u> C <u>T</u> E <u>V</u> P <u>V</u> A <u>I</u> H <u>A</u> D <u>Q</u> L <u>T</u> P <u>T</u> W <u>R</u> V <u>Y</u> S <u>T</u> G <u>S</u> N <u>V</u> F <u>Q</u> T <u>R</u> A <u>G</u> L <u>I</u> G <u>A</u> E <u>H</u> V <u>N</u> N <u>S</u> Y	660
661	E <u>C</u> D <u>I</u> P <u>I</u> G <u>A</u> G <u>I</u> C <u>A</u> S <u>Y</u> Q <u>T</u> Q <u>T</u> N <u>S</u> P <u>R</u> R <u>A</u> R <u>S</u> V <u>A</u> S <u>Q</u> S <u>I</u> I <u>A</u> Y <u>T</u> M <u>S</u> L <u>G</u> A <u>E</u> N <u>S</u> V <u>A</u> Y <u>S</u> N <u>N</u> S <u>I</u> A <u>I</u> P <u>T</u> N <u>F</u> T <u>I</u>	720
721	S <u>V</u> T <u>T</u> E <u>I</u> L <u>P</u> V <u>S</u> M <u>T</u> K <u>T</u> S <u>V</u> D <u>C</u> T <u>M</u> Y <u>I</u> C <u>G</u> D <u>S</u> T <u>E</u> C <u>S</u> N <u>L</u> L <u>L</u> Q <u>Y</u> G <u>S</u> F <u>C</u> T <u>Q</u> L <u>N</u> R <u>A</u> L <u>T</u> G <u>I</u> A <u>V</u> E <u>Q</u> D <u>K</u> N <u>T</u> Q <u>E</u>	780
781	V <u>F</u> A <u>Q</u> V <u>K</u> Q <u>I</u> Y <u>K</u> T <u>P</u> P <u>I</u> K <u>D</u> F <u>G</u> G <u>F</u> N <u>F</u> S <u>Q</u> I <u>L</u> P <u>D</u> P <u>S</u> K <u>P</u> S <u>K</u> R <u>S</u> F <u>I</u> E <u>D</u> L <u>L</u> F <u>N</u> K <u>V</u> T <u>L</u> A <u>D</u> A <u>G</u> F <u>I</u> K <u>Q</u> Y <u>G</u> D	840
841	L <u>G</u> D <u>I</u> A <u>A</u> R <u>D</u> L <u>I</u> C <u>A</u> Q <u>K</u> F <u>N</u> G <u>L</u> T <u>V</u> L <u>P</u> P <u>L</u> L <u>T</u> D <u>E</u> M <u>I</u> A <u>Q</u> Y <u>T</u> S <u>A</u> L <u>L</u> A <u>G</u> T <u>I</u> T <u>S</u> G <u>W</u> T <u>F</u> G <u>A</u> G <u>A</u> L <u>Q</u> I <u>P</u> F <u>A</u> M	900
901	Q <u>M</u> A <u>Y</u> R <u>F</u> N <u>G</u> I <u>G</u> V <u>T</u> Q <u>N</u> V <u>L</u> Y <u>E</u> N <u>Q</u> K <u>L</u> I <u>A</u> N <u>Q</u> F <u>N</u> S <u>A</u> I <u>G</u> K <u>I</u> Q <u>D</u> S <u>L</u> S <u>S</u> T <u>A</u> S <u>A</u> L <u>G</u> K <u>L</u> Q <u>D</u> V <u>V</u> N <u>Q</u> N <u>A</u> Q <u>A</u> L <u>N</u>	960
961	T <u>L</u> V <u>K</u> Q <u>L</u> S <u>S</u> N <u>F</u> G <u>A</u> I <u>S</u> S <u>V</u> L <u>N</u> D <u>I</u> L <u>S</u> R <u>L</u> D <u>K</u> V <u>E</u> A <u>E</u> V <u>Q</u> I <u>D</u> R <u>L</u> I <u>T</u> G <u>R</u> L <u>Q</u> S <u>L</u> Q <u>T</u> Y <u>V</u> T <u>Q</u> Q <u>L</u> I <u>R</u> A <u>E</u> I <u>R</u> A	1020
1021	S <u>A</u> N <u>L</u> A <u>A</u> T <u>K</u> M <u>S</u> E <u>C</u> V <u>L</u> G <u>Q</u> S <u>K</u> R <u>V</u> D <u>F</u> C <u>G</u> K <u>G</u> Y <u>H</u> L <u>M</u> S <u>F</u> P <u>Q</u> S <u>A</u> P <u>H</u> G <u>V</u> V <u>F</u> L <u>H</u> V <u>T</u> Y <u>V</u> P <u>A</u> Q <u>E</u> K <u>N</u> F <u>T</u> T <u>A</u> P <u>A</u>	1080
1081	I <u>C</u> H <u>D</u> G <u>K</u> A <u>H</u> F <u>P</u> R <u>E</u> G <u>V</u> F <u>S</u> N <u>G</u> T <u>H</u> W <u>F</u> V <u>T</u> Q <u>R</u> N <u>F</u> Y <u>E</u> P <u>Q</u> I <u>I</u> T <u>T</u> D <u>N</u> T <u>F</u> V <u>S</u> G <u>N</u> C <u>D</u> V <u>V</u> I <u>G</u> I <u>V</u> N <u>N</u> T <u>V</u> Y <u>D</u> P	1140
1141	L <u>Q</u> P <u>E</u> L <u>D</u> S <u>F</u> K <u>E</u> E <u>L</u> D <u>K</u> Y <u>F</u> K <u>N</u> H <u>T</u> S <u>P</u> D <u>V</u> D <u>L</u> G <u>D</u> I <u>S</u> G <u>I</u> N <u>A</u> S <u>V</u> V <u>N</u> I <u>Q</u> K <u>E</u> I <u>D</u> R <u>L</u> N <u>E</u> V <u>A</u> K <u>N</u> L <u>N</u> E <u>S</u> L <u>I</u> D <u>L</u>	1200
1201	Q <u>E</u> L <u>G</u> K <u>Y</u> E <u>Q</u> Y <u>I</u> K <u>W</u> P <u>W</u> Y <u>I</u> W <u>L</u> G <u>F</u> I <u>A</u> G <u>L</u> I <u>A</u> I <u>V</u> M <u>V</u> T <u>I</u> M <u>L</u> C <u>C</u> M <u>T</u> S <u>C</u> C <u>S</u> C <u>L</u> K <u>G</u> C <u>C</u> S <u>C</u> G <u>S</u> C <u>K</u> F <u>D</u> E <u>D</u> D	1260
1261	S <u>E</u> P <u>V</u> L <u>K</u> G <u>V</u> K <u>L</u> H <u>Y</u> T	1273

Fig 1. Distribution of mutations over the entire reference S protein. Multiple alignment analysis of 1000 S protein sequences of SARS-CoV-2 with the reference S protein was performed as described. The observed amino acids with mutations were highlighted in red and underlined on the reference S protein. Mutation of the boxed amino acid at position 5 (L to F) was observed 16 times, at position 614 (D to G) 632 times and at position 1124 (G to V) 13 times, respectively. All other mutations occurred less than 10 times each, typically 1–2 times.

<https://doi.org/10.1371/journal.pone.0239566.g001>

fitness [19, 26]. We analyzed 1000 full length or near-full length S protein sequences (S1 Table) through multiple alignment and comparison with the reference S protein sequence to identify potential recurrent mutations and assess the effect on CD8 T cell epitope generation for different HLAs. The result in Fig 1 showed the location map of massive mutations across entire S protein that were observed from the 1000 S protein sequences. It appears that the hot spot targeted for mutation is located on the S1 subunit composed of the signal peptide region (Position 1–13), N terminal domain, NTD (Position 16–317), receptor binding domain, RBD (Position 330–521) and the subsequent SD1 and SD2 domains (Position 522–680) that is ahead of the S1/S2 cleavage motif PRRAR [27]. Although there are plenty of positions where the mutation can occur, the frequency of these mutations is usually 0.1–0.2%, except for three amino acids at position 5, 614 and 1124 that showed a frequency of 1.6%, 63% and 1.3%, respectively (Fig 1 and S1 Fig). These mutation rates are all beyond the preferred 0.5% mutation threshold that is used in influenza research [28] and the 0.3% tracking threshold for SARS-CoV-2 surveillance [29]. In addition, these mutations are exclusively L to F at position 5, D to G at position 614 and G to V at position 1124. A single mutation on CD8 T cell epitope that is recognized and bound by HLA is one of the tactics several viruses, such as HIV, HCV and Influenza can use to escape host immunity [30–32]. Therefore, we further analyzed the potential impact of these individual mutations on CD8 T cell epitope generation by different HLA alleles.

Effect of L5F mutation on CD8 T cell epitope generation

The amino acid L on position 5 is part of the S protein signal peptide [27]. Signal peptides are important targets of CD8 T cell responses against viral and tumor antigens [33, 34]. From the

list of CD8 T cell epitopes for the 80 class I HLA alleles (S2 Table), we identified a total of 47 epitopes containing the L5 amino acid with the same sequence of FVFLVLLPL that could be presented by 22 of 36 HLA-As, 16 of 34 HLA-Bs and 9 of 10 HLA-Cs we analyzed (Table 2). Other potential epitopes containing the L5 are very limited and thus not evaluated. To examine whether a single mutation of L to F at position 5 on the SARS-CoV-2 S protein would affect FVFLVLLPL epitope presentation by different HLA alleles, we replaced the L with F at position 5 on the reference S protein sequence and re-analyzed the 9-mer CD8 epitopes for all the 80 HLA-A, B and C alleles using NetMHC 4.0 as described. The resulting epitopes of each allele were compared to the corresponding epitopes derived from the original reference S protein (S2 Table). The comparison data in Table 2 showed that L5F mutation increased the epitope binding affinity for 37 different HLA alleles, meanwhile only 10 other alleles had decreased binding affinity for the mutated epitope FVFFVLLPL. In addition, the mutated epitopes could be presented by 5 more HLA alleles (Table 2).

Effect of G1124V mutation on CD8 T cell epitope generation

The amino acid G on position 1124 is within the connector domain of the S protein [27] that is important for S protein trimerization and critical for stabilizing S protein conformational structure during pre- and post-fusion with host cell membrane [35]. By inspecting the epitopes on the S2 Table, we found a total of 19 of the exact same epitopes starting at position 1121 (FVSGNCDVV) that could be presented by 12 HLA-As (HLA-A*0201, HLA-A*0202, HLA-A*0203, HLA-A*0206, HLA-A*0207, HLA-A*0211, HLA-A*0212, HLA-A*0216, HLA-A*0219, HLA-A*0250, HLA-A*6802 and HLA-A*6901) and 7 HLA-Cs (HLA-C*0303, HLA-C*0501, HLA-C*0602, HLA-C*0701, HLA-C*0802, HLA-C*1203 and HLA-C*1502). Only one additional epitope starting at a different position that contains the G1124 was identified for just one individual allele, indicating FVSGNCDVV is the predominantly presentable epitope in the connector domain of the S protein. Re-analyzing the epitopes of the 80 HLA alleles with the mutated reference S protein containing the V1124 identified 13 HLA alleles with decreased binding affinity for the mutated epitope FVSVNCDVV (Table 3). We also observed 6 other HLA alleles lost the ability to present the mutant epitope. No alleles showed increased binding affinity due to this mutation. Further there were no other alleles capable of presenting this mutated epitope (Table 3).

Effect of D614G mutation on CD8 T cell epitope generation

The mutation at position 614 with a sole D to G switch at a 63% frequency is especially alarming. The result revealed that SARS-Cov-2 isolates with G614 mutation has been adapted and

Table 2. Effect of L5F mutation on the generation of SARS-CoV-2 S protein specific CD8 T cell epitope FVFFVLLPL by different HLA alleles.

Effect	HLA alleles
Affinity increased	HLA-A*0201, HLA-A*0202, HLA-A*0203, HLA-A*0206, HLA-A*0211, HLA-A*0212, HLA-A*0216, HLA-A*0250, HLA-A*2601, HLA-A*2*603, HLA-A*3201, HLA-A*3207, HLA-A*3215, HLA-A*6601, HLA-A*6802, HLA-A*6823, HLA-A*6901, HLA-A*8001, HLA-B*1402, HLA-B*1517, HLA-B*2720, HLA-B*3501, HLA-B*4013, HLA-B*4201, HLA-B*4601, HLA-B*4801, HLA-B*5301, HLA-B*7301, HLA-B*8301, HLA-C*0303, HLA-C*0401, HLA-C*0602, HLA-C*0702, HLA-C*0802, HLA-C*1203, HLA-C*1402, HLA-C*1502
Affinity decreased	HLA-A*0207, HLA-A*0217, HLA-A*0219, HLA-A*2501, HLA-B*0801, HLA-B*0803, HLA-B*1509, HLA-B*3503, HLA-B*3901, HLA-C*0701
Epitope gained	HLA-A*2301, HLA-A*2602, HLA-A*2902, HLA-B*1502, HLA-B*5101
Epitope lost	None

<https://doi.org/10.1371/journal.pone.0239566.t002>

Table 3. Effect of G1124V mutation on the generation of SARS-CoV-2 S protein specific CD8 T cell epitope FVSVNCDVV by different HLA alleles.

Effect	HLA alleles
Affinity increased	None
Affinity decreased	HLA-A*0202, HLA-A*0206, HLA-A*0212, HLA-A*0216, HLA-A*0219, HLA-A*0250, HLA-A*6802, HLA-C*0303, HLA-C*0501, HLA-C*0602, HLA-C*0802, HLA-C*1203, HLA-C*1502
Epitope gained	None
Epitope lost	HLA-A*0201, HLA-A*0203, HLA-A*0207, HLA-A*0211 HLA-A*6901, HLA-C*0701

<https://doi.org/10.1371/journal.pone.0239566.t003>

spread efficiently within human population. To our knowledge, the D614 amino acid is not within the essential receptor binding domain or a residue of any validated neutralization epitopes for SARS-CoV-2. In addition, whether it is involved in CD8 T cell response is not known. Through screening the epitopes listed on the [S2 Table](#), we found that a panel of 22 HLA alleles including HLA-A*0101, HLA-A*0201, HLA-A*0206, HLA-A*0207, HLA-A*0211, HLA-A*0212, HLA-A*0216, HLA-A*0219, HLA-A*0250, HLA-A*2603, HLA-A*6601, HLA-A*6802, HLA-A*6901, HLA-B*1509, HLA-B*2720, HLA-B*3901, HLA-B*4801, HLA-C*0501, HLA-C*0602, HLA-C*0802, HLA-C*1203 and HLA-C*1402 had at least one CD8 T cell epitope containing the D614 amino acid ([Table 4](#)). Unlike the L5 and G1124 containing epitopes, D614 containing epitopes could start at several different positions including 606, 607, 610, 611, 612 and 614 with the epitope of YQDVNCTEV most frequently observed ([Table 4](#)).

Similarly, we reanalyzed the 9-mer CD8 T cell epitope for the reference S protein with a single D614G mutation. The resulting epitopes of each allele were compared to the corresponding epitopes derived from the original reference S protein ([S2 Table](#)). The comparison data was summarized in [Table 5](#). With a single D to G switch, there were seven HLA alleles including HLA-A*0203, HLA-A*0205, HLA-A*0206, HLA-A*2403, HLA-A*2501, HLA-A*2601 and HLA-C*1203 that obtained at least one CD8 epitope containing the replaced amino acid G614. Interestingly, nine HLA alleles including HLA-A*0101, HLA-A*0207, HLA-A*6802, HLA-A*6901, HLA-B*1509, HLA-B*3901, HLA-C*0501, HLA-C*0802 and HLA-C*1203 lost one CD8 epitopes each that were predicted using unmodified reference S protein sequence ([Table 5](#)). In addition to the direct gain or loss of CD8 T cell epitopes, several HLA alleles had their epitope binding affinities changed. These include the G614 containing epitopes with increased affinity for HLA-A*0211, HLA-A*0212, HLA-A*0216, HLA-A*0219, HLA-A*0250, HLA-A*6601, HLA-A*6802, HLA-A*6901, HLA-B*2720 and HLA-C*1402 ([Table 5](#)). Finally, we also observed there were other G614 containing epitopes with decreased affinity for alleles including HLA-A*0201, HLA-A*0206, HLA-A*0211, HLA-A*0212, HLA-A*0216, HLA-B*4801, HLA-C*0602 ([Table 5](#)). These results illustrated differential effects of D614G mutation on the capability to present S protein specific CD8 T cell epitopes by different HLA alleles.

Mutations on CD8 T cell epitopes are unlikely driven by seasonal CoVs

Although we identified major mutations on CD8 T cell epitopes including L5F, D614G and G1124V for the SARS-CoV-2 S protein, the detailed mechanisms driving these mutations are not known. One possibility is that existing cross-reactive CD8 T cell immunity elicited by human seasonal CoVs could promote the virus to mutate and escape immune recognition in general population that are frequently targeted by seasonal CoVs. To test this, we went ahead

Table 4. List of CD8 T cell epitopes containing the amino acid D614 on the reference S protein of SARS-CoV-2.

HLA alleles	Epitope	
	Position	Sequence
HLA-A*0101	612	YQDVNCTEV
HLA-A*0201	612	YQDVNCTEV
HLA-A*0206	612	YQDVNCTEV
HLA-A*0207	612	YQDVNCTEV
HLA-A*0211	610	VLYQDVNCT
	612	YQDVNCTEV
	614	DVNCTEVPV
HLA-A*0212	610	VLYQDVNCT
	612	YQDVNCTEV
	614	DVNCTEVPV
HLA-A*0216	610	VLYQDVNCT
	612	YQDVNCTEV
	614	DVNCTEVPV
HLA-A*0219	610	VLYQDVNCT
	612	YQDVNCTEV
	614	DVNCTEVPV
HLA-A*0250	607	QVAVLYQDV
	610	VLYQDVNCT
	612	YQDVNCTEV
HLA-A*2603	614	DVNCTEVPV
HLA-A*6601	612	YQDVNCTEV
HLA-A*6802	607	QVAVLYQDV
	614	DVNCTEVPV
HLA-A*6901	607	QVAVLYQDV
	614	DVNCTEVPV
HLA-B*1509	612	YQDVNCTEV
HLA-B*2720	612	YQDVNCTEV
HLA-B*3901	612	YQDVNCTEV
HLA-B*4801	606	NQVAVLYQD
HLA-C*0501	612	YQDVNCTEV
HLA-C*0602	612	YQDVNCTEV
HLA-C*0802	612	YQDVNCTEV
HLA-C*1203	612	YQDVNCTEV
HLA-C*1402	611	LYQDVNCTE

<https://doi.org/10.1371/journal.pone.0239566.t004>

to evaluate if there are shared CD8 T cell epitopes containing L5, D614 and G1124 from four representative seasonal CoVs [1, 4]. However, phylogenetic analysis indicated the reference S protein was not clustered together with the S proteins of seasonal CoVs (Fig 2). Our pairwise alignment of the reference S protein with each of the four seasonal CoVs showed very low percent of identities (27% for HCoV-NL63, 31% for HCoV-HKU1, 28% for HCoV-229E and 33% for HCoV-OC-43). Further, multiple alignment of these S protein sequences failed to identify any identical or similar CD8 T cell epitope motif to these we described above that include: FVFLVLLPL, FVSVNCDVV and several epitopes containing D614 (Table 4). From these results, it was concluded that there were unlikely any cross-reactive CD8 T cells induced by seasonal CoVs that could promote the key mutations we observed.

Table 5. Effect of D614G mutation on the generation of SARS-CoV-2 S protein specific CD8 T cell epitopes by different HLA alleles.

Alleles	Epitopes gained		Epitopes lost		Affinity decreased		Affinity increased	
	Pos	Sequence	Pos	Sequence	Pos	Sequence	Pos	Sequence
HLA-A*0101		None	612	YQDVNCTEV		None		None
HLA-A*0201		None		None		None	612	YQGVNCTEV
HLA-A*0203	612	YQGVNCTEV		None		None		None
HLA-A*0205	607	QVAVLYQGV		None		None		None
	612	YQGVNCTEV						
HLA-A*0206	607	QVAVLYQGV		None		None	612	YQGVNCTEV
	614	GVNCTEVPV						
HLA-A*0207		None	612	YQDVNCTEV		None		None
HLA-A*0211		None		None	610	VLYQGVNCT	612	YQGVNCTEV
					614	GVNCTEVPV		
HLA-A*0212		None		None	610	VLYQGVNCT	612	YQGVNCTEV
					614	GVNCTEVPV		
HLA-A*0216		None		None	610	VLYQGVNCT	612	YQGVNCTEV
					614	GVNCTEVPV		
HLA-A*0219		None		None	610	VLYQGVNCT		None
					612	YQGVNCTEV		
					614	GVNCTEVPV		
HLA-A*0250		None		None	607	QVAVLYQGV		None
					610	VLYQGVNCT		
					612	YQGVNCTEV		
HLA-A*2403	611	LYQGVNCTE		None		None		None
HLA-A*2501	607	QVAVLYQGV		None		None		None
HLA-A*2601	607	QVAVLYQGV		None		None		None
HLA-A*2603		None		None		None		None
HLA-A*6601		None		None	612	YQGVNCTEV		None
HLA-A*6802		None	614	DVNCTEVPV	607	QVAVLYQGV		None
HLA-A*6901		None	614	DVNCTEVPV	607	QVAVLYQGV		None
HLA-B*1509		None	612	YQDVNCTEV		None		None
HLA-B*2720		None		None	612	YQGVNCTEV		None
HLA-B*3901		None	612	YQDVNCTEV		None		None
HLA-B*4801		None		None		None	606	NQVAVLYQG
HLA-C*0501		None	612	YQDVNCTEV		None		None
HLA-C*0602		None		None		None	612	YQGVNCTEV
HLA-C*0802		None	612	YQDVNCTEV		None		None
HLA-C*1203	610	VLYQGVNCT	612	YQDVNCTEV		None		None
HLA-C*1402		None		None	611	LYQGVNCTE		None

<https://doi.org/10.1371/journal.pone.0239566.t005>

Discussion

CD8 T cell response specific for S protein of SARS-CoV has been well characterized, yet most of the data has primarily focused on HLA-A*0201 [18, 36–39]. The epitopes of CD8 T cells and their specific responses against SARS-CoV in population with other HLA alleles are scarcely investigated. During SARS-CoV-2 infection, reduction of CD8 T cells is associated with worse prognosis and systemic inflammation [40]. Yet, the specificities of these CD8 T cell and associated HLA genetic background were not provided.

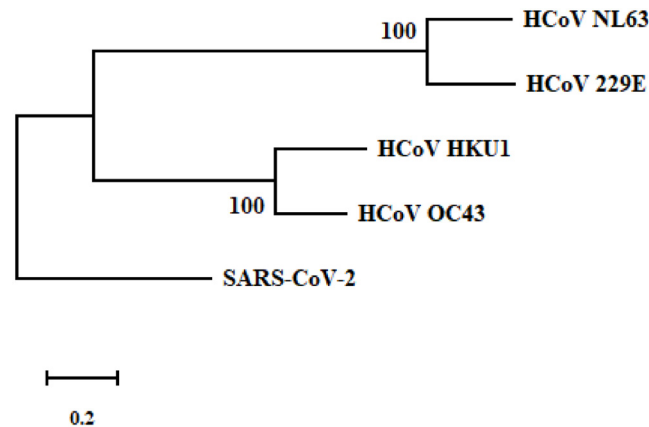


Fig 2. Phylogenetic analysis of the reference S protein of SARS-CoV-2 with four seasonal human CoVs (NL63, HKU1, 229E and OC43). The phylogenetic tree was performed using the maximum likelihood method and JTT model with a bootstrap of 1000 replicates within the Mega-X software. Scale bar indicates the number of substitutions per site for the trees. The accession number for S proteins of SARS-CoV-2, HCoV-NL63, HCoV-HKU1, HCoV-229E and HCoV-OC43 are QHD43416.1, APF29063.1, BBA20983.1, AAG48592.1 and CAA83661.1, respectively.

<https://doi.org/10.1371/journal.pone.0239566.g002>

In this study, we intended to provide a broad representation of S protein specific CD8 T cell epitopes of 80 different human HLAs for better understanding anti-SARS-CoV2 T cell immunity. Our results indicated a potential differential anti-S protein CD8 T cell response during COVID-19 infection in individuals with unique HLA alleles as the capacity of these alleles to present potential epitopes varies dramatically (Table 1). Among them, HLA-B*2705 is only capable of presenting 16 epitopes. This allele is highly associated with various forms of arthritis [41]. It's known that SARS-CoV-2 infection in aged population and people with underlying health conditions including immune disorders tend to be much more severe and lethal [42]. Additionally, during acute influenza infection and chronic HIV progression, viral escape mutations could occur on HLA-B*2705 restricted CD8 T cell epitopes [43, 44]. Although we didn't find evidence that seasonal CoV-2 could induce cross-reactive CD8 T cells toward the three key mutations we identified, other mechanism may promote SARS-CoV-2 to mutate and escape HLA-B*2705 restricted CD8 T cell immunity. Population with Caucasian origins in several countries including Spain, Belgium, Austria, Netherlands and USA have from about 6% to 14% high frequency for this allele according to the Allele Frequency Net Database (AFND). Some of these countries have experienced substantially high COVID-19 case numbers and/or mortalities. Therefore, SARS-CoV-2 infection rate, disease severity and immune responses in HLA-B*2705 individuals warrant thorough investigations.

In addition to HLA-B*2705, B44 alleles also had very low number of epitopes (17 for B4402 and 19 for B4403, Table 1). A recent epidemiologic study reported that HLA-B*44 and HLA-C*01 positive individuals were more susceptible to SARS-CoV-2 infections in Italy when compared with HLA-A*25, B*08, B*15:01, B*51, B*14, B*18, B*49, and C*03 [25]. Although we weren't able to predict the epitopes for C*01 allele as it hasn't been enlisted in NetMHC 4.0, each of A*25, B*08, B*15:01, B*51, B*14, B*18 and B*49 alleles we analyzed showed higher numbers of CD8 T cell epitopes than B*44 (Table 1), indicating broad anti-S protein CD8 T cell responses may provide better protection against SARS-CoV-2 infection.

Our data lacks experimental support, however, some of the S protein specific CD8 T cell epitopes presented in this study have been validated in SARS-CoV infection and utilized for monitoring human anti-SARS specific CD8 T cell responses using immunological techniques such as ELISPOT and tetramer staining [18, 36–39]. One of these epitopes is FIAGLIAIV that

was identified in HLA-A*0201 patient infected with SARS-CoV [36]. This epitope was predicted not only for HLA-A*0201 on SARS-CoV2 S protein, but also a total of other 21 alleles analyzed in this report including HLA-A*0202, HLA-A*0203, HLA-A*0205, HLA-A*0206, HLA-A*0207, HLA-A*0211, HLA-A*0212, HLA-A*0216, HLA-A*0217, HLA-A*0219, HLA-A*0250, HLA-A*2501, HLA-A*2601, HLA-A*2602, HLA-A*6802, HLA-A*6901, HLA-B*3901, HLA-B*4601, HLA-C*1203, HLA-C*0802, and HLA-C*1502 (S2 Table). The identified epitope RLNEVAKNL on SARS-CoV [39] was shared with 8 HLA-A, 4 HLA-B and 3 HLA-C alleles for SARS-CoV-2 in our study. Another SARS-CoV S protein specific CD8 T cell epitope, VLNDILSRL [38] were predicted on the SARS-CoV-2 S protein for 9 HLA-A alleles, 1 HLA-B and 3 HLA-C alleles. The epitope NLNESLIDL we identified for 10 HLA-A alleles was also characterized as a valid HLA-A*0201 restricted SARS-CoV CD8 T cell epitope in a published study [37].

Although recent studies agreed that D614G mutation could enhance SARS-CoV-2 infectivity and promote its transmission [29, 45], regarding its effect on virulence, one study reported that G614 virus infection was associated with higher mortality [46], while the other study concluded no obvious effect on disease severity [29]. One potential explanation for these disparate clinical findings is that the prevalent HLAs of the infected subjects in the two studies differ, which leads to divergent anti-S protein CD8 T cell responses, either toward the epitopes containing the G614 mutation alone and/or in combination with other epitopes we identified, as D614G mutation can occur simultaneously with other mutations on the S protein (S1 Fig). This is because first, our prediction result showed only about 25% of the total HLA alleles we analyzed could mount CD8 T cell responses targeting the epitopes containing D614 (Table 4). Secondly, among these HLA alleles that could generate D614 containing CD8 T cell epitopes, the epitopes they recognize and present to TCR differ (Table 5), which may lead to different TCR clonotypes and anti-viral efficacy [47, 48]. Third, our mutational analysis on a panel of CD8 T cell epitopes that have G614 also suggested possible different control outcomes by various HLA alleles as their bindings to the mutated epitope could be altered differently (Table 5). We believe pairing HLA typing with SARS-CoV-2 sequencing and testing anti-S protein CD8 T cell responses could allow precise assessment of clinical outcome of D614/G614 virus infection on individuals with different HLA alleles.

Our data of mutational effect of L5F suggests that the evolution of SARS-CoV2 virus targeting L5 might be eventually unsuccessful as the mutated epitope could enhance CD8 T cell recognition and killing through the enhanced interaction with most of the HLA alleles (Table 2). In contrast, the G1124V mutation appears to favor the virus to escape immune recognition as this mutation reduces epitope binding affinity of HLA alleles to a level that some are no longer able to bind (Table 3). However, the V1124 variant has not become as dominant as the G614. This could be due to other viral and host factors, such as viral structure stability, peptide-MHC stability and innate responses that prevent this variant to stand out further. Since the pandemic SARS-CoV-2 outbreak in several major countries hasn't yet been controlled, the continual monitoring of these mutations on S protein are still necessary.

Based on the data and discussions provided here, we would like to encourage the research labs to carry out further validation and characterization of these candidate S protein epitopes and study their role in protecting SARS-CoV-2 infection and vaccine immunity. These epitopes should include not only the ones that had been identified for the S protein of SARS-CoV, but also uncharacterized epitopes such as these containing L/F5, D/G614 and G/V1124. We also recommend the clinical labs to organize and utilize resources to combine SARS-CoV-2 sampling and viral genome sequencing with HLA typing and S protein specific CD8 T cell immune testing to gather more useful data for better identifying risk groups and implementing policies that are suited for different geographical locations, resulting in more effective

transmission control. Ultimately, we believe these efforts will provide more solid data for effective vaccine development and elimination of SARS-CoV-2 from human population.

Supporting information

S1 Table. List of the accession numbers of 1000 S protein sequences of SARS-CoV-2 deposited in the NCBI protein database.

(XLSX)

S2 Table. List of the predicted human CD8 T cell epitopes for the reference S protein of SARS-CoV-2.

(XLSX)

S1 Fig. Multiple sequence alignment of 1000 S protein sequences with reference S protein of SARS-CoV-2 (QHD43416.1) showing position 601 to 720 that contains D614G mutation with high frequency.

(PDF)

Acknowledgments

The health care professionals, researchers and laboratories around the world who have helped with collecting samples, sequencing and depositing the data for SARS-CoV-2 into NCBI database are sincerely appreciated.

Author Contributions

Conceptualization: Elisa Guo.

Data curation: Elisa Guo.

Formal analysis: Elisa Guo, Hailong Guo.

Supervision: Hailong Guo.

Writing – original draft: Elisa Guo, Hailong Guo.

Writing – review & editing: Elisa Guo, Hailong Guo.

References

1. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol.* 2015; 1282: 1–23. https://doi.org/10.1007/978-1-4939-2438-7_1 PMID: 25720466
2. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, et al. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med.* 2003; 348: 1953–1966. <https://doi.org/10.1056/NEJMoa030781> PMID: 12690092
3. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med.* 2012; 367: 1814–1820. <https://doi.org/10.1056/NEJMoa1211721> PMID: 23075143
4. Enjuanes L, Zuñiga S, Castaño-Rodríguez C, Gutierrez-Alvarez J, Canton J, Sola I. Molecular Basis of Coronavirus Virulence and Vaccine Development. *Adv Virus Res.* 2016; 96: 245–286. <https://doi.org/10.1016/bs.aivir.2016.08.003> PMID: 27712626
5. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med.* 2020; 382: 727–733. <https://doi.org/10.1056/NEJMoa2001017> PMID: 31978945
6. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. *Nature.* 2020; 579: 265–269. <https://doi.org/10.1038/s41586-020-2008-3> PMID: 32015508

7. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020; 579: 270–273. <https://doi.org/10.1038/s41586-020-2012-7> PMID: 32015507
8. Liu Y-C, Kuo R-L, Shih S-R. COVID-19: The first documented coronavirus pandemic in history. *Biomed J*. 2020; 43: 328–333. <https://doi.org/10.1016/j.bj.2020.04.007> PMID: 32387617
9. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*. 2020; 5: 536–544. <https://doi.org/10.1038/s41564-020-0695-z> PMID: 32123347
10. Du L, He Y, Zhou Y, Liu S, Zheng B-J, Jiang S. The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nat Rev Microbiol*. 2009; 7: 226–236. <https://doi.org/10.1038/nrmicro2090> PMID: 19198616
11. Poh CM, Carissimo G, Wang B, Amrun SN, Lee CY-P, Chee RS-L, et al. Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat Commun*. 2020; 11: 2806. <https://doi.org/10.1038/s41467-020-16638-2> PMID: 32483236
12. Rogers TF, Zhao F, Huang D, Beutler N, Burns A, He W-T, et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science*. 2020; 369: 956–963. <https://doi.org/10.1126/science.abc7520> PMID: 32540903
13. Cao Y, Su B, Guo X, Sun W, Deng Y, Bao L, et al. Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. *Cell*. 2020; 182: 73–84.e16. <https://doi.org/10.1016/j.cell.2020.05.025> PMID: 32425270
14. Channappanavar R, Zhao J, Perlman S. T cell-mediated immune response to respiratory coronaviruses. *Immunol Res*. 2014; 59: 118–128. <https://doi.org/10.1007/s12026-014-8534-z> PMID: 24845462
15. Xu X, Sun J, Nie S, Li H, Kong Y, Liang M, et al. Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China. *Nat Med*. 2020; 26: 1193–1195. <https://doi.org/10.1038/s41591-020-0949-6> PMID: 32504052
16. Yin S, Tong X, Huang A, Shen H, Li Y, Liu Y, et al. Longitudinal anti-SARS-CoV-2 antibody profile and neutralization activity of a COVID-19 patient. *J Infect*. 2020; 81: e31–e32. <https://doi.org/10.1016/j.jinf.2020.06.076> PMID: 32622905
17. Channappanavar R, Fett C, Zhao J, Meyerholz DK, Perlman S. Virus-specific memory CD8 T cells provide substantial protection from lethal severe acute respiratory syndrome coronavirus infection. *J Virol*. 2014; 88: 11034–11044. <https://doi.org/10.1128/JVI.01505-14> PMID: 25056892
18. Zhao K, Yang B, Xu Y, Wu C. CD8+ T cell response in HLA-A*0201 transgenic mice is elicited by epitopes from SARS-CoV S protein. *Vaccine*. 2010; 28: 6666–6674. <https://doi.org/10.1016/j.vaccine.2010.08.013> PMID: 20709007
19. Wang M, Li M, Ren R, Li L, Chen E-Q, Li W, et al. International Expansion of a Novel SARS-CoV-2 Mutant. *J Virol*. 2020; 94. <https://doi.org/10.1128/JVI.00567-20> PMID: 32269121
20. Saha P, Banerjee AK, Tripathi PP, Srivastava AK, Ray U. A virus that has gone viral: amino acid mutation in S protein of Indian isolate of Coronavirus COVID-19 might impact receptor binding, and thus, infectivity. *Biosci Rep*. 2020; 40. <https://doi.org/10.1042/BSR20201312> PMID: 32378705
21. Lau S-Y, Wang P, Mok BW-Y, Zhang AJ, Chu H, Lee AC-Y, et al. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerg Microbes Infect*. 2020; 9: 837–842. <https://doi.org/10.1080/22221751.2020.1756700> PMID: 32301390
22. Laha S, Chakraborty J, Das S, Manna SK, Biswas S, Chatterjee R. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infect Genet Evol*. 2020; 85: 104445. <https://doi.org/10.1016/j.meegid.2020.104445> PMID: 32615316
23. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci*. 2003; 12: 1007–1017. <https://doi.org/10.1110/ps.0239403> PMID: 12717023
24. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*. 2016; 32: 511–517. <https://doi.org/10.1093/bioinformatics/btv639> PMID: 26515819
25. Correale P, Mutti L, Pentimalli F, Baglio G, Saladino RE, Sileri P, et al. HLA-B*44 and C*01 Prevalence Correlates with Covid19 Spreading across Italy. *Int J Mol Sci*. 2020; 21. <https://doi.org/10.3390/ijms21155205> PMID: 32717807
26. Phan T. Genetic diversity and evolution of SARS-CoV-2. *Infect Genet Evol*. 2020; 81: 104260. <https://doi.org/10.1016/j.meegid.2020.104260> PMID: 32092483
27. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020; 367: 1260–1263. <https://doi.org/10.1126/science.abb2507> PMID: 32075877

28. Xue KS, Bloom JD. Linking influenza virus evolution within and between human hosts. *Virus Evol.* 2020; 6: veaa010. <https://doi.org/10.1093/ve/veaa010> PMID: 32082616
29. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell.* 2020; 182: 812–827.e19. <https://doi.org/10.1016/j.cell.2020.06.043> PMID: 32697968
30. Sun X, Shi Y, Akahoshi T, Fujiwara M, Gatanaga H, Schönbach C, et al. Effects of a Single Escape Mutation on T Cell and HIV-1 Co-adaptation. *Cell Rep.* 2016; 15: 2279–2291. <https://doi.org/10.1016/j.celrep.2016.05.017> PMID: 27239036
31. Timm J, Walker CM. Mutational escape of CD8+ T cell epitopes: implications for prevention and therapy of persistent hepatitis virus infections. *Med Microbiol Immunol.* 2015; 204: 29–38. <https://doi.org/10.1007/s00430-014-0372-z> PMID: 25537849
32. Price GE, Ou R, Jiang H, Huang L, Moskophidis D. Viral escape by selection of cytotoxic T cell-resistant variants in influenza A virus pneumonia. *J Exp Med.* 2000; 191: 1853–1867. <https://doi.org/10.1084/jem.191.11.1853> PMID: 10839802
33. Hombach J, Pircher H, Tonegawa S, Zinkernagel RM. Strictly transporter of antigen presentation (TAP)-dependent presentation of an immunodominant cytotoxic T lymphocyte epitope in the signal sequence of a virus protein. *J Exp Med.* 1995; 182: 1615–1619. <https://doi.org/10.1084/jem.182.5.1615> PMID: 7595234
34. El Hage F, Stroobant V, Vergnon I, Baurain J-F, Echchakir H, Lazar V, et al. Preprocalcitonin signal peptide generates a cytotoxic T lymphocyte-defined tumor epitope processed by a proteasome-independent pathway. *Proc Natl Acad Sci U S A.* 2008; 105: 10119–10124. <https://doi.org/10.1073/pnas.0802753105> PMID: 18626012
35. Walls AC, Tortorici MA, Snijder J, Xiong X, Bosch B-J, Rey FA, et al. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc Natl Acad Sci U S A.* 2017; 114: 11157–11162. <https://doi.org/10.1073/pnas.1708727114> PMID: 29073020
36. Wang Y-D, Sin W-YF, Xu G-B, Yang H-H, Wong T, Pang X-W, et al. T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus spike protein elicit a specific T-cell immune response in patients who recover from SARS. *J Virol.* 2004; 78: 5612–5618. <https://doi.org/10.1128/JVI.78.11.5612-5618.2004> PMID: 15140958
37. Chen Y-Z, Liu G, Senju S, Wang Q, Irie A, Haruta M, et al. Identification of SARS-COV spike protein-derived and HLA-A2-restricted human CTL epitopes by using a new muramyl dipeptidederivative adjuvant. *Int J Immunopathol Pharmacol.* 2010; 23: 165–177. <https://doi.org/10.1177/039463201002300115> PMID: 20378004
38. Lv Y, Ruan Z, Wang L, Ni B, Wu Y. Identification of a novel conserved HLA-A*0201-restricted epitope from the spike protein of SARS-CoV. *BMC Immunol.* 2009; 10: 61. <https://doi.org/10.1186/1471-2172-10-61> PMID: 19958537
39. Wang B, Chen H, Jiang X, Zhang M, Wan T, Li N, et al. Identification of an HLA-A*0201-restricted CD8 + T-cell epitope SSp-1 of SARS-CoV spike protein. *Blood.* 2004; 104: 200–206. <https://doi.org/10.1182/blood-2003-11-4072> PMID: 15016646
40. Urra JM, Cabrera CM, Porras L, Ródenas I. Selective CD8 cell reduction by SARS-CoV-2 is associated with a worse prognosis and systemic inflammation in COVID-19 patients. *Clin Immunol.* 2020; 217: 108486. <https://doi.org/10.1016/j.clim.2020.108486> PMID: 32479985
41. McMichael A, Bowness P. HLA-B27: natural function and pathogenic role in spondyloarthritis. *Arthritis Res.* 2002; 4 Suppl 3: S153–158. <https://doi.org/10.1186/ar571> PMID: 12110134
42. Banerjee A, Pasea L, Harris S, Gonzalez-Izquierdo A, Torralbo A, Shallcross L, et al. Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study. *Lancet.* 2020; 395: 1715–1725. [https://doi.org/10.1016/S0140-6736\(20\)30854-0](https://doi.org/10.1016/S0140-6736(20)30854-0) PMID: 32405103
43. Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, Nowak MA, et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat Med.* 1997; 3: 212–217. <https://doi.org/10.1038/nm0297-212> PMID: 9018241
44. Berkhoff EGM, Boon ACM, Nieuwkoop NJ, Fouchier R a. M, Sintnicolaas K, Osterhaus ADME, et al. A mutation in the HLA-B*2705-restricted NP383-391 epitope affects the human influenza A virus-specific cytotoxic T-lymphocyte response in vitro. *J Virol.* 2004; 78: 5216–5222. <https://doi.org/10.1128/jvi.78.10.5216-5222.2004> PMID: 15113903
45. Zhang L, Jackson CB, Mou H, Ojha A, Rangarajan ES, Izzard T, et al. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv.* 2020. <https://doi.org/10.1101/2020.06.12.148726> PMID: 32587973
46. Becerra-Flores M, Cardozo T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *Int J Clin Pract.* 2020; 74: e13525. <https://doi.org/10.1111/ijcp.13525> PMID: 32374903

47. Lissina A, Chakrabarti LA, Takiguchi M, Appay V. TCR clonotypes: molecular determinants of T-cell efficacy against HIV. *Curr Opin Virol.* 2016; 16: 77–85. <https://doi.org/10.1016/j.coviro.2016.01.017> PMID: [26874617](https://pubmed.ncbi.nlm.nih.gov/26874617/)
48. Lima NS, Takata H, Huang S-H, Haregot A, Mitchell J, Blackmore S, et al. CTL Clonotypes with Higher TCR Affinity Have Better Ability to Reduce the HIV Latent Reservoir. *J Immunol.* 2020; 205: 699–707. <https://doi.org/10.4049/jimmunol.1900811> PMID: [32591402](https://pubmed.ncbi.nlm.nih.gov/32591402/)