

Prediction of SARS-CoV2 spike protein epitopes reveals HLA-associated susceptibility

Romero-López JP (✉ pablorolo30@gmail.com)

Programa MEDICI, Carrera de Médico Cirujano, Facultad de Estudios Superiores Iztacala, UNAM; Laboratorio de Inmunología 1, Posgrado en Ciencias Químico-biológicas, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional

Carnalla-Cortés M

3. Centro de Investigación en Salud Poblacional, Instituto Nacional de Salud Pública.

Pacheco-Olivera DL

Laboratorio de Inmunología 1, Posgrado en Ciencias Químico-biológicas, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional; 4. Unidad Médica de Investigación en Inmunología, Hospital de Especialidades, Centro Médico Nacional Siglo XXI, IMSS.

Ocampo M

Programa MEDICI, Carrera de Médico Cirujano, Facultad de Estudios Superiores Iztacala, UNAM; Laboratorio de Inmunología 1, Posgrado en Ciencias Químico-biológicas, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional

Oliva-Ramírez J

6. Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Campus Estado de México

Moreno-Manjón J

7. Laboratorio de Infectología, Microbiología e Inmunología Clínicas, Unidad de Investigación en Medicina Experimental, Facultad de Medicina, UNAM.

Bernal-Alferes B

Laboratorio de Inmunología 1, Posgrado en Ciencias Químico-biológicas, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional; 8. Escuela Superior de Medicina, Instituto Politécnico Nacional.

García-Latorre E

Laboratorio de Inmunología 1, Posgrado en Ciencias Químico-biológicas, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional

Domínguez-López ML

Laboratorio de Inmunología 1, Posgrado en Ciencias Químico-biológicas, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional

Arturo Reyes- Sandoval

9. The Jenner Institute, Nuffield Department of Medicine, University of Oxford, The Henry Wellcome Building for Molecular Physiology.

Jiménez-Zamudio L

10. Laboratorio de Inmunología Clínica 1, Posgrado en Ciencias Químico-biológicas, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional. Carpio y Plan de Ayala SN, Colonia Santo Tomás, Alcaldía Miguel Hidalgo

Research Article

Keywords: SARS-CoV2, antigen-presenting HLA-class I and II alleles, probable susceptibility associations, bioinformatic prediction of T cell epitopes

Posted Date: April 28th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-25844/v1>

Abstract

SARS-CoV2 infection is causing a pandemic disease that is reflected in important public health problems worldwide. HLA-based epitope prediction and association with susceptibility provides an important base for treatment design. Hence the aim of this study is to predict the best antigen-presenting HLA-class I and II alleles in top affected populations and determine probable susceptibility associations. A bioinformatic prediction of T cell epitopes and their restricted HLA class I and II alleles was performed to predict immunogenic epitopes and HLA alleles from the spike protein of the SARS-CoV-2 virus, together with molecular modeling analysis and a correlation with the cumulative incidence of COVID-19 infection in 14 countries. Here, we describe a set of ten highly immunogenic epitopes, together with different HLA alleles that can efficiently present these epitopes to T cells. Most of these epitopes are located within the S1 subunit of the spike protein, suggesting that this area is highly immunogenic. A statistical correlation was found between the frequency of HLA-A*02:03 and HLA-A*31:01 and a low cumulative incidence in the selected countries.

Introduction

The coronavirus disease (COVID-19) has been declared a pandemic by the World Health Organization (WHO) and the number of cases and deaths associated with this disease has increased significantly.¹ It is estimated that there are over 932,000 confirmed cases and 47,000 deaths, representing a severe global emergency on March 31st, 2020. To date, the US, Italy, and Spain are the most affected countries with the highest total number of confirmed cases and number of deaths.² Even though the incidence among the total population would be a better estimate to evaluate the pandemic burden in each country.

COVID-19 is a disease generated by the novel severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) that during its initial outbreak in China in 2020 presented with a wide range of clinical manifestations, like fever (88·7%), cough (67·8%), fatigue (38·1%), and acute respiratory distress syndrome (ARDS) in severe cases.³ Interestingly, the molecular and clinical manifestations of the disease vary between asymptomatic, mild-symptomatic, and severe patients, requiring in some cases hospitalization to prevent fatal outcomes.⁴

Currently, the SARS-CoV-2 genome has been characterized as a new Betacoronavirus, which shares around 87% of genomic identity to bat-SL-CoVZC45 and bat-SL-CoVZXC21 viruses.⁵ A more recent analysis by Zhou *et al.* reported that there is a 96·2% identity with CoV BatCoV RaTG13 and a 79·5% identity with SARS-CoV.⁶ The genomic characterization of the virus not only provides information about its taxonomy and probable origin but also offers opportunities to perform deeper analysis using bioinformatic tools.

The angiotensin-converting enzyme-2 (ACE-2) receptor and the transmembrane serine protease 2 (TMPRSS2) are essential components of the human host for the virus entry into the upper respiratory epithelial cells. The virus recognizes ACE-2 through the viral spike glycoprotein (S), event that leads to the virus-cell membrane fusion.⁷ The S glycoprotein is found as a homotrimer of three identical monomers, each one of which is divided into two subunits: S1 and S2. The first subunit folds in four domains: A, B, C, and D. The B domain possesses a receptor-binding domain (RBD) that recognizes ACE2, hence its importance for the viral entry.⁸ The S2 subunit sequence has two tandem domains, named HR1 and HR2, that play an essential role in the viral fusion to the membrane.⁹ Furthermore, analysis of the spike protein showed that it is conserved among SARS-CoV and SARS-CoV-2 with 76·3% of identity and 87·3% of similarity.¹⁰

Several viral diseases studies have shown that disease severity is closely associated with some individual factors, such as genetic background and immune response. The human leukocyte antigen HLA is responsible for the antigen presentation to T cells and, therefore, a key component for adaptive immune response initiation. The HLA genes are the most polymorphic genes in the human genome, as these polymorphisms influence the ability to present different sets of epitopes to T cells. Some HLA molecules are more efficient than others presenting certain antigens, what may lead to a better induction of immune responses. This fact has already been proven for some viral diseases like A H1N1 influenza ¹¹ and HIV.¹²

It has been reported an association between infection and HLA-B*07:03,¹³ HLA-Cw*08:01,¹⁴ HLA-B*46:01, and HLA-B*54:01 for SARS-CoV. Specifically, it has been reported that the individuals who are HLA-B*46:01 positive have a higher risk of severe infection,¹⁵ whereas the frequency of HLA-DRB2*03:01 is lower among SARS-CoV patients.¹³ The relationship between viral

infection, HLA and disease susceptibility has led to vaccine development and molecular epidemiology research that can contribute to novel therapies.

So far, the control of the COVID-19 pandemic remains a challenge, resulting in thousands of new cases and deaths reported daily. It is necessary to find prophylaxis and specific treatment to control the unbridled infection and to diminish the global morbidity and mortality. The generation of a vaccine that targets this virus remains the primary solution.¹⁶ However, the lack of knowledge regarding the immune response, such as the HLA-virus interactions, makes it a challenging task.¹⁷

Additionally, the genetic variations among different populations and their possible link with SARS-CoV-2 viral responses remain unknown. In this report, we analyzed which epitopes of the spike protein of SARS-CoV2 are highly immunogenic and able to be presented by HLA class I and II in different populations using bioinformatic tools. We also demonstrate an epidemiological correlation between HLA allele frequency and cumulative incidence of COVID-19 among countries.

Methods

Study design

A bioinformatic epitope prediction of the spike glycoprotein was performed. This gave information about the most immunogenic peptide-HLA matches and the HLA alleles that more likely present these epitopes in an efficient way. An ecological study was performed for the evaluation of HLA allele types and the cumulative incidences among countries from the day on which the first case was reported to March 31st, 2020.

Variables definition

Cumulative incidence was calculated as the total of confirmed cases from day one per one million inhabitants using the total population of each country in 2020 reported by the United Nations Population Division.

Bioinformatic epitope prediction

Bioinformatic analyses were performed to predict HLA class I and II epitopes using the sequence of the SARS-CoV-2 Spike protein. The sequence for SARS-CoV-2 Spike Glycoprotein was obtained from the GenBank with the accession number QHR63290.2 in FASTA format. This sequence was then submitted to the TepiTool server from the IEDB Analysis Resource database (<http://tools.iedb.org/tepitool/>).¹⁸ The epitope prediction was performed for the 27 most frequent HLA-A and B alleles that cover for most populations (table 1).¹⁹ Once the total epitope list was obtained, it was submitted to the T cell class I pMHC immunogenicity predictor server (<http://tools.iedb.org/immunogenicity/>) to get the immunogenicity score, which is predicted according to the aminoacid residues of the peptide.²⁰

The peptide-HLA pairs with a positive immunogenicity score and a predicted IC50 level lower than the established cut-off from the complete list were selected (table 2).²¹ Considering that the lower the IC50 value is, the higher the binding affinity. The ten more immunogenic pMHC combinations from this list were selected.

The epitopes for HLA class II molecules were also predicted. To do so, the same sequence as before was used and submitted to the IEDB MHC class II epitope prediction tool (<http://tools.iedb.org/mhcii/>) using the IEDB recommended 2·2 algorithm and the most common HLA-DP, DQ and DR alleles (table 1).²² The predicted epitopes that had an SMM-predicted IC50 value higher than 50 were excluded and the sequences were ordered by the percentile rank.²³ The MHC-II prediction tools use a core of nine amino acids to predict the best peptide binding affinity, even when the class II molecules bind peptides with 15 amino acid length, so the ten SMM cores with the minor percentile rank (what means the highest affinity binding) were selected.

Structural modeling

The Cryo-EM 3D model of the SARS-CoV2 spike glycoprotein crystal was downloaded from the RSCB Protein Data Bank (PDB) (<https://www.rcsb.org/pdb/home/sitemap.do>) (ID: 6VYB).²⁴

The 3D structure was obtained and analyzed using PyMOL® software (Schrödinger LLC. Molecular Graphics System (PyMOL). Version 1.8.0 LLC, New York, NY. 2015). The basic local alignment search tool online (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to assess the position of the predicted peptides in the glycoprotein and the protein sequence was adjusted manually using the PyMOL tools.

Modeling structural analysis of SARS-CoV-2 epitopes and HLA

Modeling of protein-peptide interactions between an HLA class I and two SARS-CoV-2 spike protein epitopes was performed using the CABS-dock server (<http://biocomp.chem.uw.edu.pl/CABSdock>) to predict whether the most possible binding site would correspond to the HLA groove. To perform this analysis, the 3D structure of HLA-A*02:01 (accession number 3HLA) was used, since it was the only class I HLA molecule with a full structure available that did not have any pre-loaded peptides in the RSCB PDB database (RSCB Protein Data Bank) (<https://www.rcsb.org/pdb/home/sitemap.do>). The YLQPRTFLL and FIAGLIAIV peptides were selected from the epitope prediction results, as well as A2-ALW (ALWGPDDPAA), which is a well-known weak binding epitope for the HLA-A*02:01 that has been related to autoreactive CD8+ cells in Diabetes Type 1 and constitutes a good control element for the analysis.²⁵ The last epitopes were re-analyzed using the NetMHC 4.0 server (<http://www.cbs.dtu.dk/services/NetMHC/index.php>) to obtain a prediction of the properties for each peptide-MHC-class I complex. The HLA-A*02:01 flexibility was determined by the PDBFlex server (<http://pdfflex.org/>) to select the best level of flexibility to introduce into the CABS-dock tool. Two regions were also removed for docking analysis, such as the β 2-microglobulin and a fragment of the A chain that was outside of the pocket. This was done by selecting the specific rank of amino acids that constitute them and selecting 1:B-99: B to remove β 2-microglobulin and 182:A- 270: A for A chain. Finally, the analysis was run using ten simulation cycles per assay and using the full peptide flexibility.

HLA alleles frequency and cumulative incidence analysis.

The top 14 countries with the highest number of confirmed cases by March 31st, 2020 were selected to correlate the HLA allele frequency and the cumulative incidence rates. The database consulted provided information since January 22th, 2020 and did not have data from China. Hence, China was excluded. Japan was included despite having performed a relatively low number of COVID tests compared with other countries (the proportion of positive cases among tests performed was lower (7%) than in the US and Italy (21%))^{26,27} because it has a large population.

HLA allele frequencies were searched from our epitope prediction results in the open database Allele Frequencies in Worldwide populations and the samples were restricted to studies of blood donors and bone marrow registries.²⁸ The population that better represented the country (minorities or subpopulations were not included) and had information on HLA alleles was the one considered. Ten countries were selected (the database name of the population is shown in parentheses): the US (Caucasian pop 2), Italy (pop 5), Spain (Catalunya, Navarra...), Germany (pop 8), France (French bone marrow donor registry), Belgium (Belgium), Netherlands (Leiden), Austria (Austria), South Korea (pop 8), and Japan (pop 3).

For each country, day one was determined as the day when the first confirmed case was reported. To March 31st, 2020, the day of analysis, Netherlands was on day 34, Austria on day 36, Belgium on day 57, Spain on day 60, and the rest of the countries on day 61 and higher. The cumulative incidence on days 34, 36, 57, and 60 was calculated.

Statistical analysis

The correlation between each HLA allele frequency and the cumulative incidence among countries was evaluated with the Spearman's rank correlation. Statistical significance was considered when the p-value was <0.10. The analysis was performed in the software GraphPad Prism® (GraphPad Software version 6.0, La Jolla California USA, www.graphpad.com).

Results

In order to assess the best Spike protein epitope-HLA matches, the sequence was analyzed looking for epitope predictions in the most frequent HLA-A and HLA-B alleles. The ten most immunogenic peptides with a higher affinity binding to its restricted HLA are shown in table 3.

Although the most immunogenic peptide from this list is GTHWFVTQR, the match with the highest affinity was between the peptide FIAGLIAIV and HLA-A*02:03. Of note, here we analyzed the most frequent class I A and B alleles, so this analysis not only reveals epitopes that can be used for vaccine development but also the HLA alleles that best present epitopes of this particular protein.

The best epitopes and HLA class II alleles were also predicted, as shown in table 4. The prediction tool for HLA class II uses a core of nine amino acids to predict the binding efficiency of peptides to the pocket of the molecules, even if this core is in the middle of different peptides of 15 aa. Of note, between this whole set of peptides only seven HLA molecules resulted with a high binding affinity: HLA-DPA1*01:03/DPB1*02:01, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*03:01/DPB1*04:02, HLA-DQA1*05:01/DQB1*03:01, HLA-DRB1*01:01, HLA-DRB1*07:01, and HLA-DRB1*09:01.

To track down the specific location of the peptides in the SARS-CoV2 spike glycoprotein, the corresponding 3D model was obtained. In this model, different predicted epitopes were searched for (tables 2 and 3) in the protein structure considering its subunits and domains (figure 1). Notably, HLA class I peptides WTAGAAAYY, SANNCTFEY, and YLQPRTFLL (7, 8, and 9) are located in the A domain, which is highly conserved among other coronavirus species 24, suggesting that these could be also epitopes for other viruses. On the other hand, it was found that the class II epitopes FELLHAPAT, VVLSFELL, FLVLLPLVS, VLSFELLHA, and FTISVTTEI (a, b, c, d, and h) and the HLA class I EVFNATRFA (4) are preferentially founded in the B domain.

The most probable junctions among the HLA-A*02:01 (table 3) against two of the most immunogenic SARS-CoV-2 epitopes (peptides FIAGLIAIV and YLQPRTFLL) were predicted. They were also compared with one well-known weak binding epitope for the HLA-A*02:01, the A2-ALW (ALWGPDDPAA), as a control.²⁵

In order to perform a docking analysis, the last epitopes were re-analyzed using the NetMHC 4.0 server to obtain a prediction of the properties of each peptide-MHC-class I complex.³⁴ The YLQPRTFLL epitope showed the strongest affinity and the highest binding percentage level for HLA-A*02:01, with values of 5.36 nM and 0.04% respectively. Conversely, the FIAGLIAIV epitope had less affinity and percentage of binding level (10.29 nM and 0.12% respectively), as well as the control peptide A2-ALW (117.15 nM and 1.20%).

Despite the differences in affinities between the two SARS-CoV-2 epitopes, both showed to fit the groove of HLA-A*02:01 maintaining their unfolded conformations. A contact map of the interface between the amino acids of HLA-A*02:01 and the epitopes was made, finding that each amino acid of FIAGLIAIV peptide was interacting at least with one amino acid of the HLA-A*02:01 groove (figure 2). Our data complemented the previous analysis of the epitopes and demonstrated that all of them have a great chance of being successfully presented by their respective HLA-restricted allele into a pMHC-TCR complex and trigger an immunological response against SARS-CoV-2. As the other HLA class I and II alleles structures are not found free of a pre-loaded peptide, they were not submitted to docking analysis.

HLA allele analysis and correlation with the country cumulative incidence

The population behavior of the HLA alleles that resulted in our prediction analysis was looked for considering the cumulative incidence of COVID infection in the most affected countries by March 31st, 2020. Table 5 shows the cumulative incidence per one million inhabitants on days 36, 57, 60, and 65 after the report of the first case in each country. Austria had the highest cumulative incidence on day 34 and 36 (976 and 1130 per one million inhabitants, respectively).

There was no association between any HLA allele frequencies and the cumulative incidence on days 34 and 36 (data not shown). On day 57, HLA-A*02:03 frequency was significantly associated with a lower cumulative incidence per one million inhabitants ($R = -0.75$, $p \text{ value} = 0.08$) and remained the same on day 60.

In addition, the frequency of HLA-A*31:01 was significantly associated with a lower cumulative incidence on day 57 ($R = -0.69$, p value=0.05), and on day 60 the association increased ($R = -0.85$, p value=0.01). Strikingly, HLA- A*30:02 frequency was significantly associated with a higher cumulative incidence on day 57 ($R = 0.75$, p value=0.05) and day 60 ($R = 0.71$, p value=0.11) (Figure 3).

Discussion

Determining HLA interactions with epitopes for optimal presentation is crucial for the understanding of immunological response to the SARS-CoV-2. Here we present a group of epitopes of the S protein that can be efficiently presented to CD8 and CD4 T cells and are probably related to the immune-mediated elimination of the virus. These peptides can be either used for peptide-based design of vaccines or for further analysis of the immunogenicity and structure of this relevant protein.

Remarkably, our structural analysis of the protein shows a higher abundance of epitopes in the A and B domains of the S1 subunit of the virus, indicating that, in the case of this part of the protein being processed by the host cells, it could represent a highly immunogenic region. In this analysis, we did not look for B cell epitopes in the structure of the protein, and we cannot confirm that the specific target of the presented epitopes could interfere with its viral function, as would be the case of neutralizing antibodies.

HLA peptide groove sequence determines which epitopes from an antigen are presented to the immune system to elicit an effective response. The high rate of polymorphisms in the HLA locus can indicate a different ability to respond to certain antigens by different individuals, and some HLA alleles can be more efficient presenting certain antigens, thus protecting from certain infections¹². Our analysis from the most representative HLA alleles revealed those that present more effectively spike protein antigens of SARS-CoV-2, hence, one can hypothesize that their presence in an individual might confer an enhanced ability to defend against the virus.

To assess this, we analyzed the frequency of these alleles and their relation to the disease dynamics in some of the countries that are currently most affected by the COVID-19 epidemic. We found a negative correlation between HLA- A*02:03 and A*31:01 frequencies and the cumulative incidence per 1 million inhabitants, confirming the hypothesis that these alleles that were predicted to be good antigen presenters are associated to a better immunity against the infection. On the other hand, a positive correlation was found with HLA-A*03:02, suggesting that this allele can be considered as a risk factor.

Our results differ from other studies exploring the association of HLA alleles and different outcomes of infection with SARS-CoV. Often, the studies compare HLA alleles within specific populations: HLA-B*07:03 in China¹³, or HLA- B*46:0115 and HLA-Cw*08 in Taiwan,²⁹ therefore, the variability of HLA allele frequencies might not be as high as between populations.³⁰ We did not find any association of the studied alleles on day 34 and day 36, suggesting that the possible effect of the different allele frequencies in the cumulative incidence might be evident in the later exponential stages of the epidemic so far. It would be interesting to explore this association once the pandemic is over and we have a better understanding of the epidemic curve.

We acknowledge the limitations of the ecological association and its risk of bias. First, the association is at a population level and we cannot apply it at an individual level. The frequency of the HLA alleles is determined for the resident population of each country, and it could be subject to migration bias due to foreign people with COVID 19 infection that might have different HLA alleles frequencies compared to the native populations. Second, several factors influence the disease transmission rate, such as the actions implemented to contain the epidemic and its right execution. Although much of the information will not be available until the pandemic is over. Third, the new cases reported depend on the number of tests performed, so we might be underestimating the cumulative incidence in some countries.

Declarations

Acknowledgements: MCC, DLPO, MO, JMM and BBA receive grants from Consejo Nacional de Ciencia y Tecnología (CONACyT), EGL and MLDL, receive scholarships from EDI and COFAA- IPN. JLOL, EGL, MLDL and ARS receive a scholarship from Sistema Nacional de Investigadores-CONACyT . Figures were created with Biorender (Biorender.com).

Author contribution: JPRL conceived the idea and directed the project. MC performed all the epidemiological analysis and correlations. DLPO performed the molecular modeling and epitope localization, MO analyzed the molecular interactions by docking, JMM, JLOR and BBA participated in manuscript writing and revision, EGL and MLDL revised the manuscript, ARS and LJZ participated in results discussion.

The authors did not receive any funding for this study.

Conflict of interests: All authors declare not to have any conflict of interests.

References

1. Sohrabi, C. *et al.* World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *J. Surg.* **76**, 71–76 (2020).
2. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real *Lancet. Infect. Dis.* **0**, (2020).
3. Guan, W. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *Engl. J. Med.* NEJMoa2002032 (2020). doi:10.1056/NEJMoa2002032
4. Siddiqi, H. K. & Mehra, M. R. COVID-19 Illness in Native and Immunosuppressed States: A Clinical- Therapeutic Staging Proposal. (2020). doi:10.1016/j.healun.2020.03.012
5. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020).
6. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
7. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **0**, (2020).
8. Tortorici, M. A. & Velesler, D. Structural insights into coronavirus entry. in *Advances in Virus Research* **105**, 93–116 (Academic Press Inc., 2019).
9. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *Virology* **1–9** (2020). doi:10.1128/jvi.00127-20
10. Baruah, V. & Bose, S. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. *Med. Virology* **495–500** (2020). doi:10.1002/jmv.25698
11. Falfán-Valencia, R. *et al.* An increased frequency in HLA class I alleles and haplotypes suggests genetic susceptibility to influenza A (H1N1) 2009 pandemic: A case-control study. *Immunol. Res.* **2018**, (2018).
12. Miura, T. *et al.* HLA-B57/B*5801 Human Immunodeficiency Virus Type 1 Elite Controllers Select for Rare Gag Variants Associated with Reduced Viral Replication Capacity and Strong Cytotoxic T-Lymphocyte Recognition. *Virology* **83**, 2743–2755 (2009).
13. Ng, M. H. L. *et al.* Association of Human-Leukocyte-Antigen Class I (B*0703) and Class II (DRB1*0301) Genotypes with Susceptibility and Resistance to the Development of Severe Acute Respiratory Syndrome. *J. Infect. Dis.* **190**, 515–518 (2004).
14. Chen, Y. M. A. *et al.* Epidemiological and genetic correlates of severe acute respiratory syndrome coronavirus infection in the hospital with the highest nosocomial infection rate in Taiwan in 2003. *Clin. Microbiol.* **44**, 359–365 (2006).
15. Lin, M. *et al.* Association of HLA class I with severe acute respiratory syndrome coronavirus *BMC Med. Genet.* **4**, 1–7 (2003).
16. Shi, J. *et al.* Epitope-based vaccine target screening against highly pathogenic MERS-CoV: An In Silico approach applied to emerging infectious diseases. *PLoS One* **10**, 1–16 (2015).
17. Zheng, M. *et al.* Functional exhaustion of antiviral lymphocytes in COVID-19 patients. *Mol. Immunol.* **7–9** (2020). doi:10.1038/s41423-020-0402-2
18. Paul, S., Sidney, J., Sette, A. & Peters, B. TepiTool: A pipeline for computational prediction of T cell epitope candidates. *Protoc. Immunol.* **2016**, 18.19.1-18.19.24 (2016).
19. Weiskopf, D. *et al.* Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Natl. Acad. Sci. U. S. A.* **110**, E2046–E2053 (2013).

20. Calis, J. J. A. *et al.* Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLoS Comput. Biol.* **9**, (2013).
21. Paul, S. *et al.* HLA Class I Alleles Are Associated with Peptide-Binding Repertoires of Different Size, Affinity, and Immunogenicity. *Immunol.* **191**, 5831–5839 (2013).
22. Wang, P. *et al.* Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics* **11**, 568 (2010).
23. Nielsen, M., Lundegaard, C. & Lund, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* **8**, (2007).
24. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (80-)*. **367**, 1260–1263 (2020).
25. Bulek, A. M. *et al.* Structural basis for the killing of human beta cells by CD8 + T cells in type 1 *Nat. Immunol.* **13**, 283–289 (2012).
26. 新型コロナウイルス感染症. Available at: https://www.mhgo.jp/stf/seisakunitsuite/bunya/0000164708_00001.html. (Accessed: 3rd April 2020)
27. COVID-19 tests per day. Available at: <https://ourworldindata.org/grapher/full-list-covid-19-tests-per-day>. (Accessed: 3rd April 2020)
28. González-Galarza, F. F. *et al.* Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res.* **43**, (2015).
29. Chen, Y. M. A. *et al.* Epidemiological and genetic correlates of severe acute respiratory syndrome coronavirus infection in the hospital with the highest nosocomial infection rate in Taiwan in 2003. *Clin. Microbiol.* **44**, 359–365 (2006).
30. dos Santos Francisco, R. *et al.* HLA supertype variation across populations: new insights into the role of natural selection in the evolution of HLA-A and HLA-B polymorphisms. *Immunogenetics* **67**, 651–663 (2015).

Tables

Table 1. HLA alleles analyzed for epitope prediction

Class I: HLA A and B	Class II: HLA-DP / HLA-DQ / HLA-DR
A*01:01, A*02:01, A*02:03, A*02:06,	DPA1*01:03-DPB1*02:01, DPA1*01:03-
A*03:01, A*11:01, A*23:01, A*24:02,	DPB1*04:01, DPA1*02:01-DPB1*01:01,
A*26:01, A*30:01, A*30:02, A*31:01,	DPA1*02:01-DPB1*05:01, DPA1*02:01-
A*32:01, A*33:01, A*68:01, A*68:02,	DPB1*14:01, DPA1*03:01-DPB1*04:02,
B*07:02, B*08:01, B*15:01, B*35:01,	DQA1*01:01-DQB1*05:01, DQA1*01:02-
B*40:01, B*44:02, B*44:03, B*51:01,	DQB1*06:02, DQA1*03:01-DQB1*03:02,
B*53:01, B*57:01, B*58:01.	DQA1*04:01-DQB1*04:02, DQA1*05:01-
	DQB1*02:01, DQA1*05:01-DQB1*03:01,
	DRB1*01:01, DRB1*03:01, DRB1*04:01,
	DRB1*04:05, DRB1*07:01, DRB1*08:02,
	DRB1*09:01, DRB1*11:01, DRB1*12:01,
	DRB1*13:02, DRB1*15:01, DRB3*01:01,
	DRB3*02:02, DRB4*01:01, DRB5*01:01.

Table 2. HLA class 1 cut-off values

Allele	Affinity cut-off (IC50 nM)	Allele	Affinity cut-off (IC50 nM)
A*0101	884	A*3101	329
A*0201	255	A*3201	131
A*0203	92	A*3301	606
A*0206	60	A*6801	197
A*0301	602	A*6802	259
A*1101	382	B*0702	687
A*2301	740	B*0801	663
A*2402	849	B*1501	528
A*2601	815	B*3501	348
A*3001	109	B*4402	904
A*3002	674	B*4403	780
A*3101	329	B*5301	538
A*3201	131	B*5701	716
..	..	B*5801	446

Table 3. HLA Class I epitope prediction

	Peptide/ protein residues (Predicted immunogenicity score)	HLA Restriction			
1	GTHWFVTQR /1096-1104 (0.3513)	HLA-A*31:01	HLA-A*68:01	HLA-A*11:01	HLA-A*03:01
	Predicted IC50	9.5	14.5	29.6	379.3
2	RSFIEDLLF/ 813-821 (0.2744)	HLA-B*58:01	HLA-B*57:01	HLA-A*32:01	
	Predicted IC50	7.5	24.6	62.7	
3	FIAGLIAIV/ 1218-1224 (0.272)	HLA-A*02:03	HLA-A*02:06	HLA-A*02:01	HLA-A*68:02
	Predicted IC50	3.2	6.3	8.5	13.7
4	EVFNATRFA / 338-346 (0.2182)	HLA-A*68:02			
	Predicted IC50	12			
5	QYIKWPWYI / 1205-1213 (0.2162)	HLA-A*23:01	HLA-A*24:02		
	Predicted IC50	4.3	6.9		
6	NTQEVFAQV / 775-783 (0.1788)	HLA-A*68:02			
	Predicted IC50	5.2			
7	WTAGAAAYY /256- 264	HLA-A*26:01	HLA-A*68:01	HLA-A*01:01	HLA-A*30:02

	(0.1525)							
	Predicted IC50	9.9	27.4	31.1	36.4			
8	SANNCTFEY /160- 168 (0.1327)	HLA-B*35:01						
	Predicted IC50	14.1						
9	YLQPRTFLL /267- 275 (0.1305)	HLA-A*02:01	HLA-A*02:03	HLA-A*02:06	HLA-B*08:01	HLA-A*23:01	HLA-A*24:02	HLA-A*32:01
	Predicted IC50	4.1	7.8	9.1	23.9	125.3	201.3	202.7
10	VVFLHVTVV /1057- 1065 (0.1278)	HLA-A*02:03	HLA-A*02:06	HLA-A*02:01	HLA-A*68:02			
	Predicted IC50	9.3	11.9	21.2	24.5			

Table 4. HLA Class II epitope prediction

	SMM Core	Peptides	HLA- Restriction	Percentil e rank
a	FELLHAPAT	LSFELLHAPATVCGP	HLA-DRB1*01:01	0.03
		VLSFELLHAPATVC G	HLA-DRB1*01:01	0.03
		VVLSFELLHAPATV C	HLA-DRB1*01:01	0.03
		SFELLHAPATVCGPK	HLA-DRB1*01:01	0.09
		VVLSFELLHAPAT V	HLA-DRB1*01:01	0.09
		FELLHAPATVCGPK K	HLA-DRB1*01:01	0.71
b	VVVLSFELL	QPYRVVLSFELLH A	HLA- DPA1*03:01/DPB1*04:0 2	0.24
		PYRVVLSFELLHAP	HLA- DPA1*03:01/DPB1*04:0 2	0.25
		YRVVLSFELLHAP A	HLA- DPA1*03:01/DPB1*04:0 2	0.25
		PYRVVLSFELLHAP	HLA- DPA1*02:01/DPB1*01:0 1	0.3
		QPYRVVLSFELLH A	HLA- DPA1*02:01/DPB1*01:0 1	0.3
		YQPYRVVLSFELL H	HLA- DPA1*02:01/DPB1*01:0 1	0.3
		YRVVLSFELLHAP A	HLA- DPA1*02:01/DPB1*01:0 1	0.3
		PYRVVLSFELLHAP	HLA- DPA1*01:03/DPB1*02:0 1	0.36
		QPYRVVLSFELLH A	HLA- DPA1*01:03/DPB1*02:0 1	0.36
		YQPYRVVLSFELL H	HLA- DPA1*01:03/DPB1*02:0 1	0.36

		YRVVLSFELLHAP A	HLA- DPA1*01:03/DPB1*02:0 1	0.36
		RVVLSFELLHAPA T	HLA- DPA1*02:01/DPB1*01:0 1	0.63
		VVVLSFELLHAPAT V	HLA- DPA1*02:01/DPB1*01:0 1	0.68
		RVVLSFELLHAPA T	HLA- DPA1*03:01/DPB1*04:0 2	0.85
		YQPYRVVLSFELL H	HLA- DPA1*03:01/DPB1*04:0 2	2.2
c	FLVLLPLVS	FVFLVLLPLVSSQCV	HLA-DRB1*01:01	0.24
		MFVFLVLLPLVSSQC	HLA-DRB1*01:01	0.24
		VFLVLLPLVSSQCVN	HLA-DRB1*01:01	1.3
		FLVLLPLVSSQCVNL	HLA-DRB1*01:01	1.8
d	VLSFELLHA	RVVLSFELLHAPA T	HLA-DRB1*01:01	0.24
e	GYQPYRVVV	GYQPYRVVLSFEL L	HLA- DPA1*02:01/DPB1*01:0 1	0.3
f	FGAGAALQI	SGWTFGAGAALQIP F	HLA-DRB1*09:01	0.33
		TSGWTFGAGAALQI P	HLA-DRB1*09:01	0.34
		GWTFGAGAALQIPF A	HLA-DRB1*09:01	0.35
		WTFGAGAALQIPFA M	HLA-DRB1*09:01	0.67
		WTFGAGAALQIPFA M	HLA- DQA1*05:01/DQB1*03:0 1	1.6
g	FVFLVLLPL	MFVFLVLLPLVSSQC	HLA- DPA1*03:01/DPB1*04:0 2	0.34
		FVFLVLLPLVSSQCV	HLA- DPA1*03:01/DPB1*04:0 2	0.36
		MFVFLVLLPLVSSQC	HLA- DPA1*01:03/DPB1*02:0 1	5.2

h	RVVLSFEL	GYQPYRVVLSFEL L	HLA- DPA1*01:03/DPB1*02:0 1	0.36
		GYQPYRVVLSFEL L	HLA- DPA1*03:01/DPB1*04:0 2	6.2
i	FTISVTTEI	AIPTNFTISVTTEIL	HLA-DRB1*07:01	0.4
		PTNFTISVTTEILPV	HLA-DRB1*07:01	0.51
		IPNFTISVTTEILP	HLA-DRB1*07:01	0.52
		TNFTISVTTEILPVS	HLA-DRB1*07:01	0.52
		NFTISVTTEILPVSM	HLA-DRB1*07:01	2.5
		FTISVTTEILPVSM	HLA-DRB1*07:01	2.6
j	TNFTISVTT	IAIPTNFTISVTTEI	HLA-DRB1*07:01	0.47

Country	Day 34	Day 36	Day 57	Day 60	Day 65
Germany	0.9	1.9	346.8	524.4	857.1
France	0.3	0.9	193.7	305.6	578.6
US	0.2	0.2	23.5	77.0	253.3
South Korea	16.2	24.6	164.1	171.6	180.2
Japan	1.3	1.5	7.0	8.0	11.0
Spain	5.5	10.7	1566.7	2051.6	-
Italy	51.1	76.7	1430.6	1682.7	-
Belgium	17.3	23.0	1102.3	-	-
Austria	975.8	1130.3	-	-	-
Netherlands	739.3	-	-	-	-

Bold refers to the highest cumulative incidence for the given day.

Figures

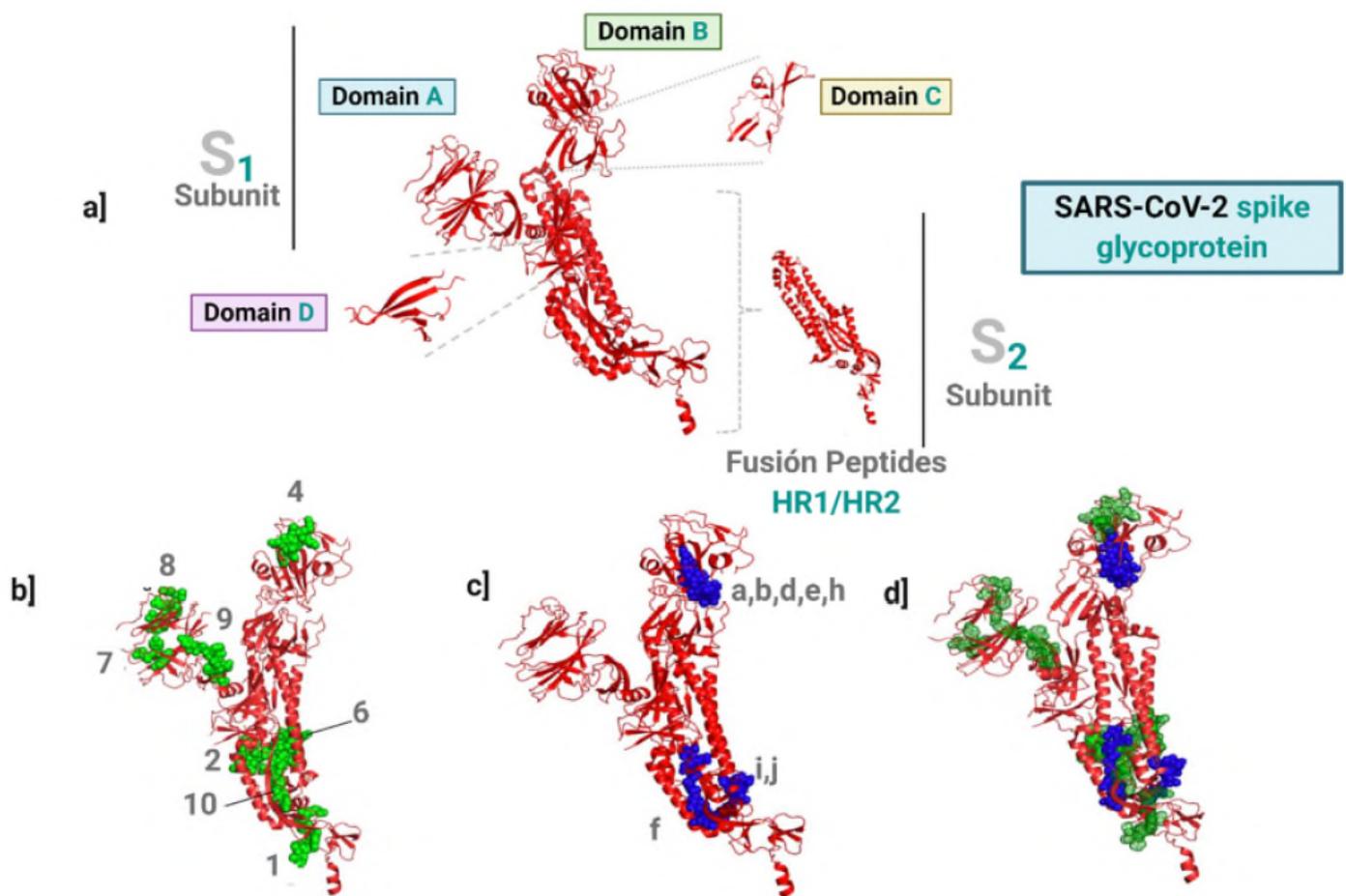


Figure 1

Localization analysis of immunogenic peptides of SARS-CoV-2 Spike Glycoprotein by 3D modeling. (A) Structure of the three full SARS-CoV-2 spike glycoprotein monomers with subunits S1-S2. The S1 domains consist of A, B, C, and D. The S2 subunit consists

of the fusion peptides and domains HR1 and HR2. (B) The predicted epitopes for HLA class I are shown in green (C) and the suggested peptides for HLA-class II in blue. (D) Merge of class I and II peptides within the protein structure. The peptides are marked individually listed from 1-10 for class I and a-j for class II, corresponding to the immunogenicity tables 3 and 4.

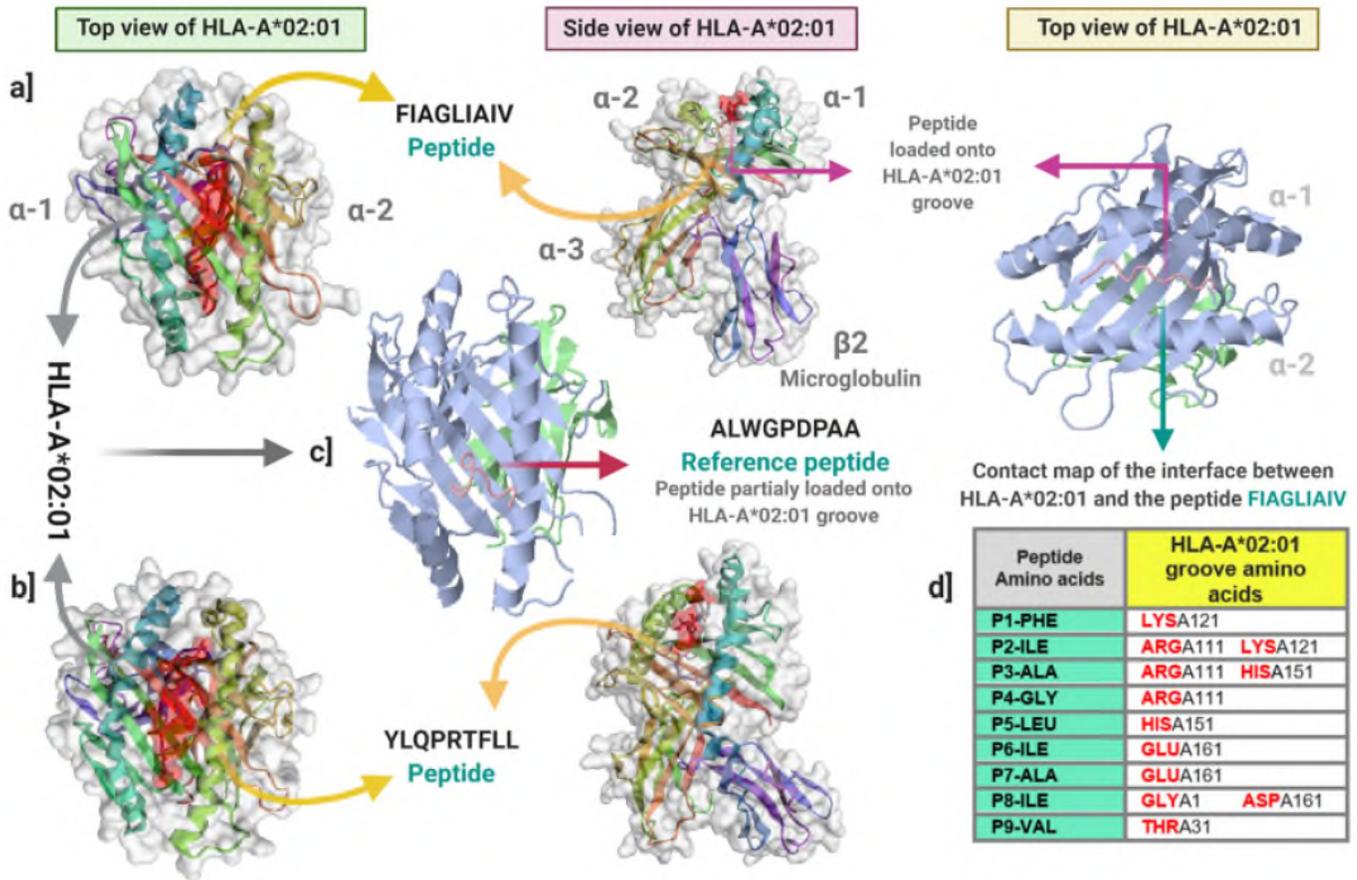


Figure 2

Modeling of HLA-A*02:01-peptide interactions predicted by full flexible docking analysis of two immunogenic SARS-CoV-2 peptides from spike protein. The selected peptides FIAGLIAIV (a) and YLQPRTFLL (b) compared against a reference ligand, ALWGPDPAA (c). The docking analysis images are representative of the most probable interaction models. Despite differences between the affinity of the two A chain SARS-CoV-2 peptides, both have joined onto the groove of the HLA, maintaining their unfolded conformation. In comparison, the reference peptide is partially loaded, showing less affinity than the SARS-CoV2 peptides and might have a lower possibility of being presented. d) Contact map of the interface between each aminoacid of the FIAGLIAIV peptide with the aminoacids of HLA-A*02*01 groove, at 3.1 Å of contact cut off, showing that every aminoacid of the peptide interacts with at least one aminoacid in the HLA pocket.

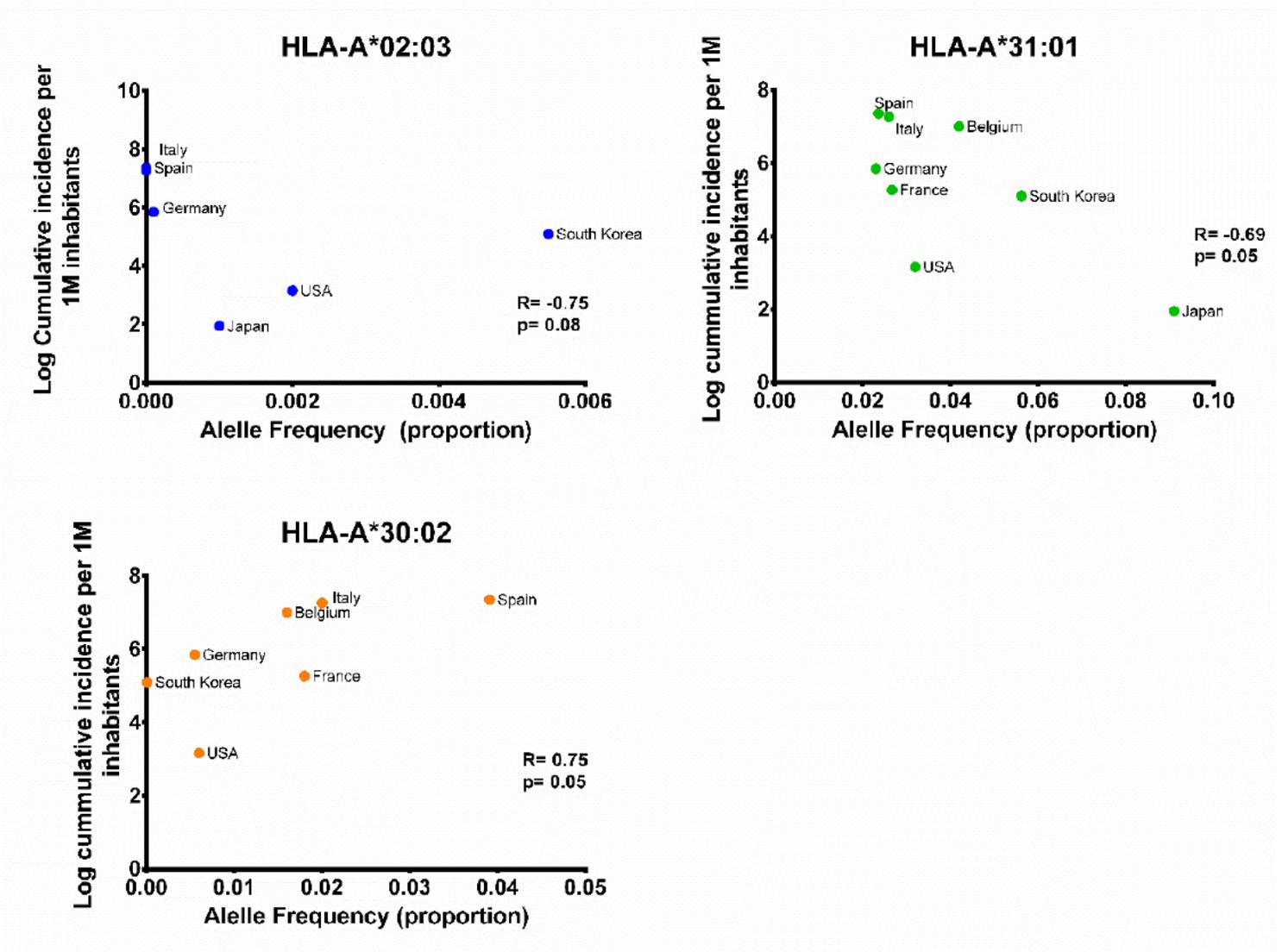


Figure 3

Spearman's correlation of HLA-A frequency and cumulative incidence per country. The correlation is shown as dot plots graphs.