International Journal of *Molecular Sciences*

MDPI

Type of the Paper (Communication)

HLA-B*44 and C*01 prevalence correlates with Covid19 spreading across Italy

Pierpaolo Correale[*1], Luciano Mutti[*2], Francesca Pentimalli[3], Giovanni Baglio[4], Rita Emilena Saladino[5], Pierpaolo Sileri[4], Antonio Giordano[2,6,#]

1 Unit of Medical Oncology, Oncology Department, Grand Metropolitan Hospital 'Bianchi Melacrino Morelli' Reggio Calabria I-89124, Italy; correalep@yahoo.it

2 Sbarro Institute for Cancer Research and Molecular Medicine, Center for Biotechnology, College of Science and Technology, Temple University, Philadelphia, Pennsylvania, PA 19122, USA; luciano.mutti@hotmail.it

3 Cell Biology and Biotherapy Unit, Istituto Nazionale Tumori-IRCCS-Fondazione G. Pascale, I-80131, Napoli, Italy; f.pentimalli@istitutotumori.na.it

4 Ministry of Health, Italy; g.baglio@sanita.it; piersileri@yahoo.com

5 Tissue Typing Unit Grand Metropolitan Hospital 'Bianchi Melacrino Morelli' Reggio Calabria I-89124, Italy; ritaemilena.saladino@gmail.com

6 Department of Medical Biotechnologies, University of Siena, Italy; antonio.giordano@unisi.it

* Dr. Pierpaolo Correale and Prof. Luciano Mutti have to be equally considered as first author.

Correspondence: Prof. Antonio Giordano, Sbarro Institute for Cancer Research and Molecular Medicine, Center for Biotechnology, College of Science and Technology, Temple University, BioLife Science Bldg. Suite 333, 1900 North 12th Street, Philadelphia, PA 19122, USATel:001 215-204 9520 Fax: 001- 215-204 9522 Email: giordano@temple.edu

Received: date; Accepted: date; Published: date

Abstract: Covid-19 spreading is showing huge, unexplained, differences between northern and southern Italy. We hypothesized that the regional prevalence of specific class I HLA alleles, which shape the anti-viral immune-response, might partly underlie these differences.

Through an ecological approach, w~~W~~e analyzed whether a set of HLA alleles (A, B, C), known to be involved in the immune response against infections, correlates with Covid-19 incidence. Covid-19 data were provided by the National Civil Protection Department, whereas HLA allele prevalence was retrieved through the Italian Bone-Marrow Donors Registry. Among all the alleles HLA-A*25, B*08, B*44, B*15:01, B*51, C*01, and C*03 showed a positive log-linear correlation with Covid-19 incidence rate fixed at the National outbreak peak on April 9th 2020 (Pearson's coefficients between 0.50 and 0.70, p-value<0.0001), whereas HLA-B*14, B*18, and B*49 showed an inverse log-linear correlation (Pearson's coefficients between -0.47 and -0.59, p-value<0.0001).

When alleles were examined simultaneously using a multiple regression model to control for confounding factors, HLA-B*44 and C*01 still resulted positively and independently associated with Covid-19: a growth rate of 16% (95%CI: 0.1%-35%) per 1% point increase in B*44 prevalence; and of 19% (95%CI: 1%-41%) per 1% point increase in C*01 prevalence.

Our epidemiologic analysis, despite the limits of the ecological approach, is strongly suggestive of ~~unravels~~ a permissive role of HLA-C*01 and B*44 towards SARS-CoV-2 infection, which warrants further investigation in case-control studies. This study opens a new potential avenue for the identification of sub-populations at risk, which could provide Health Services a tool to define more targeted clinical management strategies and priorities in vaccination campaigns.

Keywords: SARS-Cov2; coronavirus; Covid-19; HLA class I; viral infection susceptibility.

1. Introduction

Covid-19 has been declared a pandemic by the WHO [1]. Italy showed an explosive and apparently unrestrainable evolution throughout the country rapidly achieving one of the highest infection and mortality rates worldwide since the first case diagnosed in the province of Lodi in Lombardy, on February, 21st, 2020 [2]. Italian authorities are strictly monitoring the outbreak and report a large gradient of frequency that decreases from the northern to the southern and the Islands, across the twenty regions of the country [3]. So far this gradient has not been significantly modified even though the epidemic had the possibility to spread all along the peninsula due to massive migratory fluxes of individuals escaping from the high risk regions to return in their native landscapes [4] and a delayed restrictive response by the different regional authorities until March 2020. The incidence of Covid-19 cases reported by the national authorities also indicates relevant differences in the infection spreading within single provincial areas composing some of the most affected Italian regions (Figure 1a,b). Many socio-political as well as environmental hypotheses have been proposed to explain inter- and intra-regional differences and the reasons for such an aggressive spreading throughout northern Italy but so far no clear demonstration has been provided.

Emerging data are showing that both T cell and humoral response to Covid-19 infection may equally contribute to virus clearance and protective memory [2,5], however in a minority of infected patients an inappropriate immune-response may lead to virus spread from the oropharyngeal district to lung and other tissues including kidney and the central nervous system (CNS) [2,6,7]. Moreover an exaggerated cell mediated response to the virus in the alveolar tissue may be responsible of the dreaded cytokines storm and interstitial pneumonitis leading to a fatal acute respiratory distress syndrome (ARDS) [2,8–10].

Considering the crucial role played by Class I/II Human leukocyte antigen (HLA) molecules in triggering the anti-viral immune-response, it has been hypothesized that different HLA alleles may define an individual susceptibility to Covid-19 infection and spreading as reported for other viruses as well as the two different corona-viruses responsible for Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) [5]. In this context a number of studies is searching for selected HLAs with very efficient ability to present viral-antigen derived epitope peptides to cytotoxic T cells. The identification of highly immunogenic peptide epitopes recognized by specific T Cell Receptors (TCRs) might in fact, provide potential candidates for vaccine development [11–13]. By triggering and sustaining of the human host immune-defences to the virus, specific class I HLA alleles may also be involved in the occurrence of other symptoms, morbidity or lethality. Indeed, a study in a small cohort of Covid-19 Chinese patients suggested that specific HLA alleles might correlate with disease occurrence [14]. So, we set out to perform an exploratory epidemiological analysis, through an ecological approach, aimed to investigate whether known

differences existing in HLA-A, B, and C allele distribution among the Italian population could be correlated to Covid-19 incidence and spreading throughout the peninsula.

2. Results

Correlation between HLA-A, B, and C allele frequency and Covid-19 incidence in Italian Provinces

We extrapolated HLA-A, B, and C allele frequency within the different Italian regions and relative intraregional provinces from the Italian Bone Marrow Donors Registry (IBMDR) report published on February, 2010 [15]. This is the largest published national database of bone marrow healthy donors and includes a cohort of 370,000 individuals. The regional distribution of the HLA, A, B, C alleles was also evaluated in a more recent IBMDR high-definition-analysis database including a further 120,926-individual sample size typed with high resolution method, which did not show significant changes compared to the previous report [16]. These data extrapolated from healthy donors represent a reliable surrogate of the real HLA-allele frequency scenario existing within the Italian population inhabiting different geographic areas of the country.

We examined allele prevalence across the Italian regions and detected a higher frequency of HLA-A*25, B*08, B*44, B*15:01, B*51, and C*01, and C*03 alleles in the northern regions compared with that recorded in the southern regions. A reverse situation was instead observed for HLA-B*14, B*18, and B*49 alleles, which frequency is higher in the southern regions (Figure 1c, d and supplementary data). All of the other HLA A, B, and C alleles in the database did not show substantial inter and intraregional differences (data from IBMDR database) [15].

We subsequently analyzed all the selected alleles ~~examined and classified~~ (HLA-A*25, B*08, B*44, B*15:01, B*51, B*14, B*18, B*49, C*01, and C*03) ~~allele frequency for each province of the twenty Regions~~ and ~~then~~ compared them with Covid-19 incidence, as reported by the Italian Department of Civil Protection. Our results showed the existing correlation between HLA allele frequency and Covid-19 incidence rate fixed at the National outbreak peak on April 9th 2020, suggesting that the shape of the relationship is non-linear: as HLA prevalence increases, Covid-19 incidence varies according to an exponential trend (Figure 2~~a~~). In particular, HLA-A*25, B*08, B*44, B*15:01, B*51, C*01, and C*03 alleles showed a positive log-linear correlation with Covid-19 incidence rate. On the other hand, HLA-B*14, B*18, and B*49 alleles, whose frequency is higher in the southern regions, showed an inverse log-linear correlation (Figure 2~~a~~).

Pearson's coefficient (r) was calculated as a measure of the correlation between the logarithm of the Covid-19 incidence and the prevalence of different HLA alleles, in accordance with the exponential model. The complete correlation matrix with Pearson's coefficients and corresponding p-value for each couple of variables are shown in Table 1. Such HLA alleles that were positively correlated to Covid-19 infection revealed Pearson's coefficient values between 0.50 and 0.70 ($p<0.0001$), and those inversely correlated between -0.47 and -0.59 ($p<0.0001$).

In order to control for mutual confounding (also including in the model the regions as confounders), the above mentioned HLA-alleles were examined simultaneously by performing a multivariable

regression analysis whose results showed that only HLA-B*44 and C*01 alleles maintained a positive and independent association with Covid-19 incidence (Table 2).

The exponential of the regression coefficient (growth rate) allowed us to quantify an increase of 16% (95%CI: 0.1%-35%) in Covid-19 incidence per 1 percentage point increase in B*44 prevalence (Table 2). Considering that the range of B*44 prevalence varies in Italy from 4% to 12%, the risk of developing Covid-19 for the highest level of prevalence can be estimated 3 times greater than that for the lowest one (Table 2).

Similarly, for C*01 the growth rate was 19% (95%CI: 1%-41%): considering a range of prevalence among provinces from 1% to 9%, the risk of developing Covid-19 for the highest level of prevalence is 4 times higher.

In order to provide further proof of evidence concerning the correlation of permissive HLA allele prevalence and Covid-19 incidence, we focused on two regions (Emilia Romagna and Marche) where the prevalence of B*44 and C*01 alleles is unevenly distributed among the different provinces (Figure 3~~2b~~).

Remarkably, in these regions the identified correlation seems to account for the intra-regional differences that are currently unexplained, such as the low incidence of Covid-19 in the province of Ferrara, compared with the other Emilian provinces, which are highly affected by the virus. Similarly, Pesaro-Urbino, which is the most affected province in the Marche Region, is also one with the highest prevalence of HLA-B*44 (Figure 1e, f). In the latter case, the exponential regression curve explains almost all the variance of the data ($r^2 = 0.9172$) and the prevalence of B*44 can almost predict exactly the incidence of Covid-19 (Figure 3~~2b~~).

3. Discussion

Our epidemiologic analysis, through a geographical ecological approach, identified putative permissive class I alleles that are potentially unable to trigger an efficient immune-response unable to counteract SARS-Cov-2 infection~~.~~.

In particular, we selected from the widest national genetic study that reports HLA data (in terms of allele prevalence) from almost 500,000 bone marrow donors representing the population from the whole national territory, those that presented stable inter- and intra-regional differences in prevalence to examine whether they could underlie the geographic differences in Covid-19 incidence. By univariate analysis we found that HLA-A*25, B*08, B*44, B*15:01, B*51, and C*01, and C*03 alleles showed a positive correlation with Covid-19 incidence rate, whereas ~~found HLA-A*25, B*08, B*44 and C*01 allele prevalence in the healthy population as geographically correlated with the incidence of Covid-19 in Italy, while the prevalence of~~ HLA-B*14, B*18, and B*49 showed an opposite trend.

Then, we tested the association between Covid-19 incidence and HLA alleles independently of each other using a multivariable regression analysis. Importantly, as an alternative approach to stratified

analysis, the Italian regions were included as covariates in the model to control for the confounding effect of the geographical context, and at the same time to verify the association of interest regardless of the North-South gradient.

Interestingly, ~~specific HLA assets in the Italian epidemiologic scenario seems to explain, at least partly, the dramatic differences in the infection rate recorded in nearby provinces within the same regions. Overall, the inter- and intra regional infection rate clearly seems to mirror the frequency of HLA-A*25, B*08, B*44, B*15:01, B*51, and C*01, and C*03 alleles in the healthy inhabitants. In our analysis, we also identified other HLA alleles showing significant differences in frequency among the twenty Italian regions but~~ none of the selected alleles could be independently associated to Covid-19 incidence if not associated to the permissive HLA B*44 and C*01.

Our results are not surprising considering that class I HLA molecules have the specific task of binding and presenting antigen-derived epitope peptides to the TCR of epitope peptide specific T cells. With this mechanisms class I HLA molecules are critically involved in both CTL replication and ability to recognizse and destroy virus infected target cells. These 9-10 mere epitope peptides derive from the intracellular processing of protein antigens operated by the proteasome system prior being complexed with HLA molecules and translated on the membrane to be exposed to the TCR of the immune-effectors [17–20]. Peptide ability to bind HLA molecules is allele specific and is restricted by specific amino-acidic consensus motifs that allows their anchorage to different HLA molecules [21,22]. In this context, two individuals carrying the same antigen but different HLA profile may give rise to a completely different T cell mediated immune-response since they may have completely different amount of HLA specific antigen derived epitopes. This hypothesis has been confirmed in several studies concerning a number of different viruses as well as tumour antigens and autoimmune models [11,23–29].

Our model suggests that healthy individuals carrying HLA-B*44 and/or C*01 and in less extent HLA-A*25, HLA-B*08 alleles may be more susceptible to SARS-CoV-2 infection; indeed, they could be unable to present a sufficient amount of immune-dominant virus derived epitope peptides and consequently, they would be unable to mount a fast and efficient anti-viral immune response. It can be hypothesized that, in these patients, the virus may freely spread from the oropharyngeal mucosae starting a more efficient replication. Consistently, both HLA-B*44 and C*01 alleles, that we identified as possibly permissive to SARS-CoV-2 infection in Italy, have also been associated to known inflammatory autoimmune diseases [30–34], a fact that highlights their ability to trigger non-proficient and often inappropriate immunological reactions. The latter finding deserve to be explored in direct experimental approaches aimed to investigate whether the expression of these HLA allele also correlates with more aggressive disease outcomes and development of interstitial pneumonitis. Interestingly, inheritance of HLA-B*44 was shown to underlie susceptibility to recurrent sinopulmonary infection [35] A further consideration stems from the knowledge that the HLA-C*01 allele, which resulted as the most permissive to SARS-CoV-2 infection in our study, also represents the specific ligand of Killer cell Immunoglobulin like receptors (KIRs), KIR2DL2 and

KIR2DL3 [36–38]. These receptors are able to inhibit the activity of Natural killer cells, which represent the first line of host defence to the infection before the occurrence of a more specific T cell response [39]. An hypothesis that deserves further and more accurate investigation.

In the present study we could not assess whether a correlation existed among HLA alleles and Covid-19-associated morbidity (Hospitalization) and lethality rates owing to the absence of reliable data and the presence of significant confounding factors, including delay in hospital records transmission, saturation of emergency rooms and presence of co-morbidities. However, our data concerning the number of Covid-19 cases might be biased towards more severe outcomes because the initial national screening was mostly limited to symptomatic or hospitalized individuals, therefore implying in our correlation analyses an association with severity of the disease.

Although~~Despite the limits of our~~ this type of ecological~~, epidemiologic~~ approach has intrinsic limits, it also has the advantage of considering a large number of cases which are readily available through public-access data sets. Indeed, geographical ecological studies are often the first to identify risk factors for a variety of diseases, which are then verified through subsequent studies [40].

O~~o~~ur observational study identifies HLA-C*01 and B*44 alleles as potential genetic determinants for the identification of individuals at risk, which warrants to be further investigated in case-control studies. To this purpose we are currently investigating the expression of different HLA alleles in pauci-symptomatic patients affected by Covid-19 and those with severe interstitial pneumonitis. We suppose that Class I and II HLA genotyping in Covid-19 patients could be easily achieved and cost-effective and could provide the bases for the identification of individuals with high risk of interstitial pneumonitis and cytokines storm who should be immunised first as the vaccine will become available.

Overall our results, by identifying the potential relevance of HLA-C*01 and B*44 alleles in developing Covid-19 open new avenues of investigation not only to understand the diffusion and the physio-pathogenesis of the disease, but also to inform future vaccination campaign priorities and clinical management strategies while promoting the research of other potential permissive alleles and high risk population worldwide.

4. Materials and Methods

4.1 Data source and population sample

HLA is a highly polymorphic genetic system. The frequency of specific HLA alleles varies significantly among the different populations inhabiting the twenty geographic regions that compose the Italian Republic. We retrieved the HLA allele frequency data, which were recorded within the different regions and ~~each~~ relative intra-regional provinces, from the database of the IBMDR. We referred to an IBMDR database Analysis, published on February, 1st, 2010 [15], containing data that were collected in a twenty year interval from a cohort of 370,000 volunteer donors with known provincial and regional birthplace origin. These allele frequencies, expressed in % as calculated through the Arlequin Software, were also compared to those reported in an updated version of the IBMDR database, including data collected on a further 120,926 volunteer donor cohort [16]. Samples from the more

recent cohort were typed with high resolution method [16]; frequency data were grouped in regions and the same allele nomenclature version was used (according to the Immuno Polymorphism Database(IPD)-International ImMunoGeneTics project (IMGT/HLA) Database Release 3.32, April 2018; http://www.ebi.ac.uk/ipd/imgt/hla/stats.html). The same allele frequency distribution across Italy was confirmed.

Data concerning the total number of individuals infected by SARS-CoV-2 per province (updated to April 9th, 2020) were provided by the Italian Department of Civil Protection, the institution under the Presidency of the Council of Ministers that manages the emergency at national level. Data were provided as aggregate numbers, in an anonymized manner.

4.2 Statistical analysis

The relationship between SARS-CoV-2 infection and the frequency of HLA alleles was explored as part of an ecological approach aimed at assessing the degree of correlation between the incidence of Covid-19 and the prevalence of HLA alleles, both measured on a geographical basis (taking each Italian province as the unit of observation).

For each allele, values were preliminarily plotted in a scatter diagram and the curve with the best fit (corresponding to exponential curve) was selected using the least squares method, which is the most widely used procedure for developing estimates of the model parameters. The estimated regression equations are indicated at the top of each graph (Figures 2, 3).

~~Pearson's coefficient was used as a measure of the correlation between the logarithm of Covid-19 incidence and the prevalence of different HLA alleles, taking each Italian province as the unit of observation. Data were expressed as percentages of total population.~~

Consistent with the exponential model, Pearson's coefficient (r) was calculated as a measure of the correlation between the logarithm of the Covid-19 incidence and the prevalence of different HLA alleles. For each value of r, the corresponding p-value was also considered, in order to assess the statistical significance of the correlation (with respect to the null hypothesis of no log-linear correlation).

~~Multivariable regression analysis was also performed in order to assess the association between SARS-CoV-2 infection and each HLA allele independently from the others, also including regions in the model as confounders.~~

Finally, the association between the logarithm of the Covid-19 incidence (considered as dependent variable) and HLA alleles independently of each other was tested using a multivariable regression analysis. Furthermore, the Italian regions were included as covariates in the model (which is an alternative approach to stratified analysis), in order to control for confounding of geographical context, and at the same time to verify the association of interest regardless of the North-South gradient.

All statistical analyses were conducted using STATA software 11.0 version (StataCorp LLC, Texas, USA). Microsoft Excel was used to draw maps.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/xxx/s1.

Author Contributions: Conceptualization, P.C., L.M., A.G.; methodology, F.P., G.B., R.E.S.; formal analysis, P.C. F.P. G.B. R.E.S. P.S.; data curation, F.P. G.B. R.E.S.; writing—original draft preparation, P.C., L.M.; writing—review and editing, P.C., L.M., F.P., G.B., R.E.S.; supervision, P.S., A.G. All authors have read and agreed to the submitted version of the manuscript.

Acknowledgments: We are grateful to Dr Nicoletta Sacchi, Director of the Italian Bone Marrow Donors Registry (IBMDR), for providing access to the database and to the Tissue Typing Unit of Grand Metropolitan Hospital in Reggio Calabria in the persons of its director Dr Giuseppa Romeo responsible of Calabrian Regional Bone Marrow Donors' Registry and Dr Marina Francone for the precious contribution provided.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. WHO Virtual press conference on COVID-19. Available online: https://www.who.int/docs/default-source/coronaviruse/transcripts/who-audio-emergencies-coronavirus-press-conference-full-and-final-11mar2020.pdf?sfvrsn=cb432bb3_2.
2. Mutti, L.; Pentimalli, F.; Baglio, G.; Maiorano, P.; Saladino, R.E.; Correale, P.; Giordano, A. Coronavirus Disease (Covid-19): What Are We Learning in a Country With High Mortality Rate? *Front. Immunol.* **2020**, *11*, 1208, doi:10.3389/fimmu.2020.01208.
3. Livingston, E.; Bucher, K. Coronavirus Disease 2019 (COVID-19) in Italy. *JAMA* **2020**, doi:10.1001/jama.2020.4344.
4. Leaked coronavirus plan to quarantine 16m sparks chaos in Italy | World news | The Guardian Available online: https://www.theguardian.com/world/2020/mar/08/leaked-coronavirus-plan-to-quarantine-16m-sparks-chaos-in-italy.
5. Li, X.; Geng, M.; Peng, Y.; Meng, L.; Lu, S. Molecular immune pathogenesis and diagnosis of COVID-19. *J. Pharm. Anal.* **2020**, doi:https://doi.org/10.1016/j.jpha.2020.03.001.
6. Cheng, Y.; Luo, R.; Wang, K.; Zhang, M.; Wang, Z.; Dong, L.; Li, J.; Yao, Y.; Ge, S.; Xu, G. Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int.* **2020**, doi:10.1016/j.kint.2020.03.005.
7. Moriguchi, T.; Harii, N.; Goto, J.; Harada, D.; Sugawara, H.; Takamino, J.; Ueno, M.; Sakata, H.; Kondo, K.; Myose, N.; et al. A first Case of Meningitis/Encephalitis associated with SARS-Coronavirus-2. *Int. J. Infect. Dis.* **2020**, doi:10.1016/j.ijid.2020.03.062.
8. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265–269, doi:10.1038/s41586-020-2008-3.
9. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506, doi:10.1016/S0140-6736(20)30183-5.
10. Xu, Z.; Shi, L.; Wang, Y.; Zhang, J.; Huang, L.; Zhang, C.; Liu, S.; Zhao, P.; Liu, H.; Zhu, L.; et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet. Respir. Med.* **2020**, doi:10.1016/S2213-2600(20)30076-X.
11. Nguyen, A.; David, J.K.; Maden, S.K.; Wood, M.A.; Weeder, B.R.; Nellore, A.; Thompson, R.F. Human leukocyte antigen susceptibility map for SARS-CoV-2. *J. Virol.* **2020**,

doi:10.1128/jvi.00510-20.

12. Hyun-Jung Lee, C.; Koohy, H. In silico identification of vaccine targets for 2019-nCoV. *F1000Research* **2020**, *9*, 145, doi:10.12688/f1000research.22507.1.
13. Lurie, N.; Saville, M.; Hatchett, R.; Halton, J. Developing Covid-19 Vaccines at Pandemic Speed. *N. Engl. J. Med.* **2020**, doi:10.1056/nejmp2005630.
14. Wang, W.; Wei, Z.; Zhang, J.; He, J.; Zhu, F. Distribution of HLA allele frequencies in 82 Chinese individuals with coronavirus disease-2019. *HLA* **2020**, doi:10.1111/tan.13941.
15. Amoroso, A.; Ferrero, N.; Rendine, S.; Sacchi, N. *Le caratteristiche HLA della popolazione Italiana: analisi di 370.000 volontari iscritti all'IBMDR*; Analysis 1.; Sindacato nazionale dirigenti sanitari SSN e ARPA (SDS-SNABI), 2010;
16. Sacchi, N.; Castagnetta, M.; Miotti, V.; Garbarino, L.; Gallina, A. High-resolution analysis of the HLA-A, -B, -C and -DRB1 alleles and national and regional haplotype frequencies based on 120 926 volunteers from the Italian Bone Marrow Donor Registry. *HLA* **2019**, *94*, 285–295, doi:10.1111/tan.13613.
17. Wölfel, T.; Klehmann, E.; Müller, C.; Schütt, K.H.; Meyer zum Büschenfelde, K.H.; Knuth, A. Lysis of human melanoma cells by autologous cytolytic T cell clones. Identification of human histocompatibility leukocyte antigen A2 as a restriction element for three different antigens. *J. Exp. Med.* **1989**, *170*, 797–810, doi:10.1084/jem.170.3.797.
18. Crowley, N.J.; Darrow, T.L.; Quinn-Allen, M.A.; Seigler, H.F. MHC-restricted recognition of autologous melanoma by tumor-specific cytotoxic T cells. Evidence for restriction by a dominant HLA-A allele. *J. Immunol.* **1991**, *146*, 1692–9.
19. Hunt, D.F.; Henderson, R.A.; Shabanowitz, J.; Sakaguchi, K.; Michel, H.; Sevilir, N.; Cox, A.L.; Appella, E.; Engelhard, V.H. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science (80-.).* **1992**, *255*, 1261–1263, doi:10.1126/science.1546328.
20. McDonnell, A.M.; Robinson, B.W.S.; Currie, A.J. Tumor antigen cross-presentation and the dendritic cell: where it all begins? *Clin. Dev. Immunol.* **2010**, *2010*, 539519, doi:10.1155/2010/539519.
21. Falk, K.; Rötzschke, O.; Stevanović, S.; Jung, G.; Rammensee, H.G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **1991**, *351*, 290–296, doi:10.1038/351290a0.
22. Gross, G.; Margalit, A. Targeting tumor-associated antigens to the MHC class I presentation pathway. *Endocr. Metab. Immune Disord. Drug Targets* **2007**, *7*, 99–109, doi:10.2174/187153007780832064.
23. Simmonds, M.J.; Gough, S.C.L. Genetic insights into disease mechanisms of autoimmunity. *Br. Med. Bull.* **2004**, *71*, 93–113, doi:10.1093/bmb/ldh032.
24. Li, S.; Jiao, H.; Yu, X.; Strong, A.J.; Shao, Y.; Sun, Y.; Altfeld, M.; Lu, Y. Human leukocyte antigen class I and class II allele frequencies and HIV-1 infection associations in a Chinese cohort. *J. Acquir. Immune Defic. Syndr.* **2007**, *44*, 121–31, doi:10.1097/01.qai.0000248355.40877.2a.
25. Vejbaesya, S.; Thongpradit, R.; Kalayanarooj, S.; Luangtrakool, K.; Luangtrakool, P.; Gibbons, R. V; Srinak, D.; Ngammthaworn, S.; Apisawes, K.; Yoon, I.-K.; et al. HLA Class I Supertype Associations With Clinical Outcome of Secondary Dengue Virus Infections in Ethnic Thais. *J.*

Infect. Dis. **2015**, *212*, 939–47, doi:10.1093/infdis/jiv127.

26. Hudson, L.E.; Allen, R.L. Leukocyte Ig-like receptors - A Model for MHC class i disease associations. *Front. Immunol.* **2016**, *7*, doi:10.3389/fimmu.2016.00281.
27. Rallón, N.; Restrepo, C.; Vicario, J.L.; Del Romero, J.; Rodríguez, C.; García-Samaniego, J.; García, M.; Cabello, A.; Górgolas, M.; Benito, J.M. Human leucocyte antigen (HLA)-DQB1*03:02 and HLA-A*02:01 have opposite patterns in their effects on susceptibility to HIV infection. *HIV Med.* **2017**, *18*, 587–594, doi:10.1111/hiv.12494.
28. Falfán-Valencia, R.; Narayanankutty, A.; Reséndiz-Hernández, J.M.; Pérez-Rubio, G.; Ramírez-Venegas, A.; Nava-Quiroz, K.J.; Bautista-Félix, N.E.; Vargas-Alarcón, G.; Castillejos-López, M.D.J.; Hernández, A. An Increased Frequency in HLA Class I Alleles and Haplotypes Suggests Genetic Susceptibility to Influenza A (H1N1) 2009 Pandemic: A Case-Control Study. *J. Immunol. Res.* **2018**, *2018*, 3174868, doi:10.1155/2018/3174868.
29. Correale, P.; Saladino, R.E.; Nardone, V.; Giannicola, R.; Agostino, R.; Pirtoli, L.; Caraglia, M.; Botta, C.; Tagliaferri, P. Could PD-1/PDL1 immune checkpoints be linked to HLA signature? *Immunotherapy* 2019, *11*, 1523–1526.
30. Sanders, P.A.; Thomson, W.; Dyer, P.A.; Grennan, D.M. Haplotypes bearing HLA-A, -B, and -DR: Bf and C4 genes in rheumatoid arthritis families. *Tissue Antigens* **1989**, *33*, 21–9, doi:10.1111/j.1399-0039.1989.tb01673.x.
31. Orchard, T.R.; Thiyagaraja, S.; Welsh, K.I.; Wordsworth, B.P.; Gaston, J.S.H.; Jewell, D.P. Clinical phenotype is related to HLA genotype in the peripheral arthropathies of inflammatory bowel disease. *Gastroenterology* **2000**, *118*, 274–278, doi:10.1016/S0016-5085(00)70209-5.
32. Grams, S.E.; Moonsamy, P. V.; Mano, C.; Oksenberg, J.R.; Begovich, A.B. Two new HLA-B alleles, B*4422 and B*4704, identified in a study of families with autoimmunity. *Tissue Antigens* **2002**, *59*, 338–340, doi:10.1034/j.1399-0039.2002.590417.x.
33. Ueta, M.; Kannabiran, C.; Wakamatsu, T.H.; Kim, M.K.; Yoon, K.-C.; Seo, K.Y.; Joo, C.-K.; Sangwan, V.; Rathi, V.; Basu, S.; et al. Trans-ethnic study confirmed independent associations of HLA-A*02:06 and HLA-B*44:03 with cold medicine-related Stevens-Johnson syndrome with severe ocular surface complications. *Sci. Rep.* **2014**, *4*, 5981, doi:10.1038/srep05981.
34. Jung, E.S.; Cheon, J.H.; Lee, J.H.; Park, S.J.; Jang, H.W.; Chung, S.H.; Park, M.H.; Kim, T.-G.; Oh, H.-B.; Yang, S.-K.; et al. HLA-C*01 is a Risk Factor for Crohn's Disease. *Inflamm. Bowel Dis.* **2016**, *22*, 796–806, doi:10.1097/MIB.0000000000000693.
35. Johnston, D.T.; Mehaffey, G.; Thomas, J.; Young, K.R.; Wiener, H.; Li, J.; Go, R.C.P.; Schroeder, H.W. Increased frequency of HLA-B44 in recurrent sinopulmonary infections (RESPI). *Clin. Immunol.* **2006**, *119*, 346–50, doi:10.1016/j.clim.2006.02.001.
36. Fadda, L.; Körner, C.; Kumar, S.; van Teijlingen, N.H.; Piechocka-Trocha, A.; Carrington, M.; Altfeld, M. HLA-Cw*0102-restricted HIV-1 p24 epitope variants can modulate the binding of the inhibitory KIR2DL2 receptor and primary NK cell function. *PLoS Pathog.* **2012**, *8*, e1002805, doi:10.1371/journal.ppat.1002805.
37. Mori, M.; Wichukchinda, N.; Miyahara, R.; Rojanawiwat, A.; Pathipvanich, P.; Miura, T.; Yasunami, M.; Ariyoshi, K.; Sawanpanyalert, P. Impact of HLA allele-KIR pairs on disease outcome in HIV-infected Thai population. *J. Acquir. Immune Defic. Syndr.* **2018**, *78*, 356–361, doi:10.1097/QAI.0000000000001676.

399 38. Pende, D.; Falco, M.; Vitale, M.; Cantoni, C.; Vitale, C.; Munari, E.; Bertaina, A.; Moretta, F.;
400 Del Zotto, G.; Pietra, G.; et al. Killer Ig-Like Receptors (KIRs): Their Role in NK Cell
401 Modulation and Developments Leading to Their Clinical Exploitation. *Front. Immunol.* **2019**,
402 *10*, 1179, doi:10.3389/fimmu.2019.01179.
403 39. Vitale, M.; Cantoni, C.; Della Chiesa, M.; Ferlazzo, G.; Carlomagno, S.; Pende, D.; Falco, M.;
404 Pessino, A.; Muccio, L.; De Maria, A.; et al. An Historical Overview: The Discovery of How
405 NK Cells Can Kill Enemies, Recruit Defense Troops, and More. *Front. Immunol.* **2019**, *10*, 1415,
406 doi:10.3389/fimmu.2019.01415.
407 40. Grant, W.B. The role of geographical ecological studies in identifying diseases linked to UVB
408 exposure and/or vitamin D. *Dermatoendocrinol.* 2016, *8*.
409

410

411

412

413

414

Figure legends.

Figure 1 | Covid-19 incidence, HLA-B*44 and C*01 prevalence in Italian Provinces | A. The graphical map shows the twenty Italian regions each constituted by various provinces. **B.** The graphical map shows quintiles of Covid-19 incidence across Italian provinces. Incidence data were calculated as the number of laboratory-confirmed Covid-19 cases up to 04/09/2020 divided by the number of residents, according to the official National data (supplementary data). **C.** and **D.** The graphical maps show B*44 and C*01 prevalence (%) in Italian Provinces. **E.** and **F.** The graphical maps show Covid-19 incidence and B*44 prevalence (%) in the provinces of Emilia Romagna and Marche. Geographical maps were built through Microsoft Excel. All Covid-19 incidence and HLA prevalence values are reported as Supplementary data.

Figure 2 | Correlation between Covid-19 incidence rate and HLA prevalence | The graphs show the correlation between Covid-19 incidence and the prevalence of HLA-A*25, B*08, B*44, B*15:01, B*51, B*14, B*18, B*49, C*01, and C*03, expressed as percentages, for all the available Italian provinces. For each correlation the R-squared value is provided at the top ~~right of~~ the graph along with the estimated regression equations. The R and p values are reported in table 1.

Figure 3 | Correlation between Covid-19 incidence rate and HLA-B*44 prevalence in Emilia Romagna and Marche provinces | The graphs show the correlation between Covid-19 incidence and the prevalence of HLA-B*44 prevalence, both expressed as percentages, for all the provinces of Emilia Romagna (top panel) and all the available provinces of Marche (bottom panel). For each correlation the R-squared value is provided at the top ~~right~~ of the graph along with the estimated regression equations. For Emilia Romagna: R=0,6813 and p value=0.0628; for Marche R=0,9577 and p value=0,0423.