Multiple genetic control of anti-COVID-19 vaccine response by HLA locus

Francesca Colombo (francesca.colombo@itb.cnr.it)
National Research Council - Institute for Biomedical Technologies https://orcid.org/0000-0003-2015-4317

Martina Esposito
National Research Council - Institute for Biomedical Technologies

Francesca Minnai
National Research Council - Institute for Biomedical Technologies

Massimiliano Copetti
IRCSS Casa Sollievo della Sofferenza

Giuseppe Miscio
Fondazione IRCCS Casa Sollievo della Sofferenza

Rita Perna
Fondazione IRCCS Casa Sollievo della Sofferenza

Ada Piepoli
Fondazione IRCCS Casa Sollievo della Sofferenza

Gabriella De Vincentis
Fondazione IRCCS Casa Sollievo della Sofferenza

Mario Benvenuto
Fondazione IRCCS Casa Sollievo della Sofferenza

Paola D'Addetta
Fondazione IRCCS Casa Sollievo della Sofferenza

Susanna Croci
University of Siena, Italy

Margherita Baldassarri
Azienda Ospedaliera Universitaria Senese

Mirella Bruttini
University of Siena, Italy

Chiara Fallerini
University of Siena, Italy

Raffaella Brugnoni
Fondazione IRCCS Istituto Neurologico Carlo Besta

Fulvio Baggi
Fondazione IRCCS Istituto Neurologico Carlo Besta https://orcid.org/0000-0002-4618-0102

Elena Corsini
Fondazione IRCCS Istituto Neurologico Carlo Besta
Emilio Ciusani
Fondazione IRCCS Neurological Instituite Carlo Besta
Francesca Andreetta
Fondazione IRCCS Istituto Neurologico Carlo Besta https://orcid.org/0000-0001-8239-8830
Tommaso Dragani
Aspidia s.r.l.
Maddalena Fratelli
Mario Negri Institute for Pharmacological Research
Massimo Carella
IRCSS Casa Sollievo della Sofferenza, Opera di San Pio da Pietrelcina
Renato Mantegazza
IRCCS FONDAZIONE ISTITUTO NEUROLOGICO CARLO BESTA
Alessandra Renieri
Medical Genetics, University of Siena, Siena, Italy https://orcid.org/0000-0002-0846-9220

Article

Keywords:

Posted Date: July 17th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3121965/v1

Abstract

Since the beginning of the anti-COVID-19 vaccination campaign, it has become evident that vaccinated subjects exhibit considerable inter-individual variability in the response to the vaccine that could be partly explained by host genetic factors. A recent study reported that the immune response elicited by the Oxford-AstraZeneca vaccine in individuals from the United Kingdom was influenced by a specific allele of the human leukocyte antigen gene *HLA-DQB1*. We performed a genome-wide association study to investigate the genetic determinants of the antibody response to the Pfizer-BioNTech vaccine in an Italian cohort of 1,351 subjects. We confirmed the involvement of the HLA locus and observed significant associations with variants in *HLA-A* gene. In particular, the HLA-A*03:01 was the most significantly associated with serum levels of anti-SARS-CoV-2 antibodies. These results support the hypothesis that HLA genes modulate the response to anti-COVID-19 vaccines and highlight the need for genetic studies in diverse populations.

Introduction

The response to the vaccine against COVID-19 is highly variable among vaccinated individuals, as reflected by the levels of antibodies detected in their serum after vaccination [1]. This phenotypic variability may be partly influenced by host genetic factors, as shown for other types of vaccines [2].

A recent genome-wide association study (GWAS) by Mentzer AJ et al. reported that the immune response triggered by the ChAdOx1 nCoV-19 (AZD1222, Oxford-AstraZeneca) vaccine in UK individuals was associated with a specific allele of the HLA-DQB1 gene [3], which encodes a human leukocyte antigen (HLA) molecule.

The HLA locus is highly polymorphic and the HLA allele frequencies differ substantially across populations, suggesting that the genetics of the response to anti-COVID-19 vaccine may vary depending on the population background. Moreover, the investigation of individuals receiving a different type of vaccine may reveal novel genetic factors involved in the production of anti-SARS-CoV-2 antibodies in response to vaccination.

In this study, we performed a GWAS of anti-SARS-CoV-2 antibody levels in 1,351 Italian subjects who received two doses of the BNT162b2 (Pfizer-BioNTech) vaccine.

Results

In this study, we included 1,351 individuals who received the COVID-19 vaccine and were recruited in three Italian hospitals: Fondazione IRCCS Istituto Neurologico Carlo Besta in Milan (n = 306), Azienda Ospedaliero-Universitaria Senese in Siena (n = 689), and Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG) (n = 356; Table 1). The recruitment period spanned from December 27th 2020 and May 15th 2021. Participants with European origin, as determined by principal component analysis (PCA, **Supplementary Fig. 1**) were included in the genetic analyses. The cohort

consisted of Italian vaccinees who were primarily hospital workers, with a predominant female representation (66.5%), and a median age of 48 years (range: 19–84). The measurement of IgG levels was performed at a median time of 40 days after the administration of the second vaccine dose (interquartile range, IQR = 68). IgG levels ranged from 12.64 to 6056 BAU/ml, with a median of 801.5 BAU/ml.

Table 1
Personal and clinical characteristics of vaccinated subjects included in the genetic analyses.

Characteristic	Total (= 1351)	Milan (= 306)	Siena (= 689)	SGR (= 356)
Age (years), median (range)	48 (19–84)	47 (25–84)	43 (19–78)	54 (21–67)
Sex, n (%)				
male	452 (33.5)	86 (28.1)	218 (31.6)	148 (41.6)
female	899 (66.5)	220 (71.9)	471 (68.4)	208 (58.4)
Days between vaccination and serological test, median (IQR)	40 (68)	30 (5)	97 (37)	30 (0)
Serum Ab anti-SARS-CoV2 (BAU/ml), median (range)	801.5 (12.64–6056)	1343 (41.21–5680)	440.5 (13.63–5680)	1519 (12.64–6056)

Normalized IgG values inversely correlated with age at vaccination (beta=-0.012, SE = 0.0019, *P*-value = $6.5x10^{-11}$) and time (in days) passed between vaccination and serum collection for antibody measurement (beta=-0.016, SE = 0.00099, *P*-value < $2.0x10^{-16}$). Additionally, we observed lower levels of IgG in individuals from Siena than those recruited in the other two cohorts (beta=-0.18, SE = 0.078, *P*-value = 0.024); the mean IgG levels of Siena subjects differed of 1,020 and 1,101 BAU/ml from the mean of Milan and San Giovanni Rotondo individuals, respectively. Antibody quantity was not significantly different between females and males (*P*-value = 0.19).

We carried out a genome-wide association analysis between normalized IgG levels and the genotypes of 521,859 variants, including in the linear regression model sex, age at vaccination, recruiting center, the first 5 principal components (PCs), and the time interval between the second vaccine dose and the serological as potential confounders. The results are reported in the Manhattan plot, shown in Fig. 1 (and listed in **Supplementary Table 1**). A statistically significant locus was identified on chromosome 6, in the HLA locus, with 40 variants associated with a nominal *P*-value < $5.0x10^{-8}$. These variants spanned a region from 29.8 Mbp to 33.5 Mbp and the lead variant, rs2499, mapped in the 3'UTR of *HLA-A* gene (beta = 0.31, SE = 0.045, *P*-value = $1.19x10^{-11}$).

Zooming in the locus, we observed that there was not a single association signal (Fig. 2A). Indeed, there were other regions, in addition to the one led by rs2499 at position 29,945,765, although this locus has the highest number of significantly associated variants, spanning from 29.8 Mbp to 30.1 Mbp. An

additional peak of association (beta = 0.27, SE = 0.048, *P*-value = $4.5x10^{-8}$) was led by rs28366135 (at position 31,396,328), that maps less than 20 kbp upstream the *HLA-B* gene. Then, we observed other two variants (rs454875, beta=-0.23, SE = 0.042, *P*-value = $4.10x10^{-8}$; rs28688207, beta=-0.28, SE = 0.052, *P*-value = $5.25x10^{-8}$) in positions 32,245,231 and 32,660,883, respectively. The latter was a splice acceptor variant of *HLA-DQB1* gene. We calculated the linkage disequilibrium (LD) between rs2499 and these variants, and we observed that they were not in LD (D'=0.0030 with rs28366135, D'=0.28 with rs454875, and D'=0.0043 with rs28688207), suggesting that the three loci are independent. Indeed, in a linear regression analysis with the genotype of rs2499 as an additional covariate, we observed that the other two loci (led by rs28366135 and rs28688207, respectively) remained statistically significant (Fig. 2B).

Looking at four-digits HLA haplotypes (n = 204), we observed that in our analysis the HLA-A*03:01 was the most significantly associated with IgG levels (beta = 0.29, SE = 0.047, *P*-value = $5.79x10^{-10}$), followed by HLA-C*12:02 (beta=-0.48, SE = 0.10, *P*-value = $2.81x10^{-6}$), HLA-DQB1*06:01 (beta=-0.47, SE = 0.10, *P*-value = $2.85x10^{-6}$), HLA-DRB1*15:02 (beta=-0.46, SE = 0.10, *P*-value = $8.50x10^{-6}$), and HLA-B*52:01 (beta=-0.44, SE = 0.10, *P*-value = $1.62x10^{-5}$). Considering the 107 two-digits HLA haplotypes, the top significant one was HLA-A*03 (beta = 0.28, SE = 0.044, *P*-value = $2.43x10^{-10}$), and then we found HLA-B*52 (beta=-0.44, SE = 0.10, *P*-value = $1.62x10^{-5}$), and HLA-DQB1*05 (beta=-0.15, SE = 0.035, *P*-value = $3.11x10^{-5}$). The HLA-DQB1*06, identified by Mentzer et al. [3], was not significantly associated with IgG levels in our sample (P-value = 0.43). All results are reported in **Supplementary Table 2**.

Discussion

Our results independently validated the finding by Mentzer et al.[3] of an important role of HLA locus in the modulation of levels of anti-spike IgG after vaccination. Of note, our results were obtained from Italian patients underwent to a different type of vaccine (two doses of the BNT162b2, Pfizer-BioNTech) and, most interestingly, we found different HLA haplotypes, significantly associated with anti-spike IgG levels, from that identified in the UK population, suggesting the involvement of multiple genes, in this same HLA locus, in the modulation of immunogenic response against the anti-COVID-vaccine. Several studies have reported association of HLA variants and genotypes with different COVID-19 outcomes, even though with some discrepancies, and differences in the immune responses against SARS-CoV-2 infection (as reviewed in [4,5]). Several molecular mechanisms underlying HLA-mediated modulation of vaccine-induced immunity have been hypothesized and are currently under investigation. Interestingly, a previous GWAS reported a significant association between the HLA-A*03:01 and adverse events after vaccination with BNT162b2 Pfizer-BioNTech vaccine in a series of more than 3,500 Americans individuals [6]. That finding, together with ours, point to a pivotal role of HLA-A*03:01 in the response to Pfizer-BioNTech vaccine, both in terms of efficacy and toxicity.

Overall, our results provide further evidence for a genetic regulation of the response to anti-COVID-19 vaccines, mediated by the HLA locus, in two independent European populations, and strengthen the importance to investigate these associations in individuals of different origin. The identification of

specific HLA alleles conferring different ability to produce anti-spike IgG after vaccination with anti-COVID vaccine can be of clinical utility for tailoring vaccination campaign, especially in most fragile subjects. In addition to COVID-19, these results may stimulate geneticists to explore the genetics of the response to other type of vaccines, against different diseases, in view of a precision vaccination medicine supported by vaccinogenomics.

Materials and Methods

Study cohort and ethical statement

For this study, individuals who received the BNT162b2 COVID-19 vaccine were recruited in three Italian hospitals: Fondazione IRCCS Istituto Neurologico Carlo Besta in Milan (n = 344), Azienda Ospedaliero-Universitaria Senese in Siena (n = 802), and Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG) (n = 384). The recruitment period spanned from December 27th 2020 and May 15th 2021. The study was performed in accordance with the Declaration of Helsinki. The research was approved by the ethics committees of recruiting hospitals, namely, the University Hospital (Azienda ospedaliero-universitaria Senese) ethical review board, Siena, Italy, the Ethics Committee of IRCCS Istituto Tumori “Giovanni Paolo II”, Bari at Fondazione Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG), Italy, and the Ethics Committee Regione Lombardia, Sezione Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy. All participants provided written informed consent to take part in the study and granted permission to use their biological samples and clinical data for genetic research purposes.

Sample and clinical data collection

Study data from Azienda Ospedaliero-Universitaria Senese in Siena, and Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo were collected and managed using REDCap electronic data capture tools [7] hosted at Azienda Ospedaliero-Universitaria Senese. Data from subjects from Fondazione IRCCS Istituto Neurologico Carlo Besta were independently collected and stored in a dedicated database. Peripheral blood samples were collected for automated genomic DNA extraction, while serum samples were obtained for measurement of antibody levels, independently in each of the three recruiting centers. The quantification of anti-SARS-CoV-2-spike antibodies (IgG) was performed, as single measurement, by Abbott (at Azienda Ospedaliero-Universitaria Senese and Fondazione IRCCS Istituto Neurologico Carlo Besta) and Siemens (at Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo) tests, and the measurement units were converted in binding antibody units (BAU)/ml. IgG measurement was done at a median time of 40 days after the administration of the second vaccine dose.

Genome-wide genotyping, data quality control and HLA haplotypes imputation

Genomic DNA from 1,509 samples was genotyped using the Axiom Human Genotyping SARS-CoV-2 Research Array (Thermo Fisher Scientific, CA, USA) at the Functional Genomics facility of the Instituto de Investigaciones Biomédicas August Pi i Sunyer (IDIBAPS, Barcelona, Spain). Genotype calling was

performed using Axiom Analysis Suite software (Thermo Fisher Scientific) following the best practice workflow (with the modified average call rate threshold ≥ 97) and data of passed samples (n = 1,474) were exported for subsequent steps. Quality control of genotype was carried out using PLINK2 software [8], as well as principal component analysis (PCA). In detail, variants with a genotyping call rate < 95%, minor allele frequency (MAF) < 1%, and a Hardy-Weinberg equilibrium test *P*-value < $1.0x10^{-10}$, were filtered out. We removed samples with call rate < 98% (n = 20), with sex inconsistencies both for PLINK2 and Axiom Analysis Suite software (n = 42), with excess of heterozygosity (n = 2), and duplicates or related individuals up to the third degree of relatedness (n = 28). In addition, four samples of non-European origin were excluded after PCA (**Supplementary Fig. 1**), as well as individuals with no full phenotypic data available (i.e., IgG levels or any of the covariates; n = 22), individuals treated with immunosuppressive drugs and outliers with very low IgG values (n = 5). A flow diagram with selected subjects for genetic analyses is shown in **Supplementary Fig. 2**.

HLA haplotypes were imputed using Minimac4 [9] on the Michigan imputation server (https://imputationserver.sph.umich.edu/index.html#!pages/home), using the Four-digit Multi-ethnic HLA v2 reference panel and phasing data with Eagle v2.4.

Statistical analyses

IgG values were inverse-normalized with the qqnorm () function in R environment (**Supplementary Fig. 3**). Linear regression between normalized IgG values and sex, age at vaccination, center (coded as dummy variable), and time between vaccination and serological test was carried out with glm () function in R. To investigate the association between genetic variants and anti-spike IgG levels (normalized BAU/ml values), genome-wide linear regression was carried out with PLINK. The analysis included 521,859 variants, and sex, age at vaccination, the first 5 principal components (PCs), and the time interval between the second vaccine dose and the serological test served as covariates. Linear regressions between imputed HLA genotypes and haplotypes and the normalized IgG values were carried out with glm() function in PLINK2 (following the pipeline described in [10], using sex, age at vaccination, centre (coded as dummy variable), and the time interval between the second vaccine dose and serological test, as covariates. Multiple test correction, using the Benjamini-Hochberg method [11], was applied to calculate the false discovery rate (FDR). Genome-wide standard significance threshold was set at *P*-value < $5.0x10^{-8}$. The library qqman in R was used to draw Manhattan and QQ plots. Zoom plot of the selected region of chromosome 6 was done with locus.zoom () function in R.

Declarations

Code availability

All analyses were performed using the cited software, packages and pipelines, whose codes were publicly available.

Data availability

Raw genotyping data are not openly available to preserve individuals' privacy under the European General Data Protection Regulation. They are available from the corresponding author upon reasonable request. Data are located in controlled access data storage at Institute for Biomedical Technologies of the National Research Council.

Acknowledgements

Istituto Buddista Italiano Soka Gakkai funded the project with the 8x1000 funds (ID n. 2020-2016_RIC_3). "PAT-COVID: Host genetics and pathogenetic mechanisms of COVID-19". The funding organization had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript. This study was a sub-project of the GEN-COVID consortium (https://sites.google.com/dbm.unisi.it/gen-covid; full list of GEN-COVID contributors is available as Supplementary Information). Specimens collected within the GEN-COVID consortium were provided by the COVID-19 Biobank of Siena, which is part of the Genetic Biobank of Siena, member of BBMRI-IT, Telethon Network of Genetic Biobanks (project no. GTB18001), EuroBioBank, and RD-Connect.

Author contributions

Conceptualization, F.C., M. Carella, M.F., and T.A.D.; formal analysis, M.E., F.M., M. Copetti. and F.C.; resources, A.R., S.C., M.Bruttini, M.Baldassarri, C.F., M.Carella, G.M., R.P., A.P., G.D.V., M.B., P.D.A, R.B., F.B, E.M.G.C, E.C, F.A., and R.E.M.; data curation, M.Baldassarri, M. Copetti, R.P., R.B., F.M, M.E., and F.C; writing—original draft preparation, F.C. and T.A.D.; writing—review and editing, T.A.D., F.C., M.F., C.F., and M. Copetti; funding acquisition, F.C., M.F, A.R., and M. Carella. All authors have read and agreed to the published version of the manuscript.

Competing interest statement

The authors declare no competing interests.

References

1. Wei, J. *et al.* Antibody responses to SARS-CoV-2 vaccines in 45,965 adults from the general population of the United Kingdom. *Nat Microbiol* **6**, 1140–1149 (2021).
2. Newport, M. J. *et al.* Genetic regulation of immune responses to vaccines in early life. *Genes Immun* **5**, 122–129 (2004).
3. Mentzer, A. J. *et al.* Human leukocyte antigen alleles associate with COVID-19 vaccine immunogenicity and risk of breakthrough infection. *Nat Med* **29**, 147–157 (2023).
4. Augusto, D. G. & Hollenbach, J. A. HLA variation and antigen presentation in COVID-19 and SARS-CoV-2 infection. *Curr Opin Immunol* **76**, 102178 (2022).
5. Wolday, D. *et al.* HLA Variation and SARS-CoV-2 Specific Antibody Response. *Viruses* **15**, 906 (2023).

6. Bolze, A. *et al.* HLA-A∗03:01 is associated with increased risk of fever, chills, and stronger side effects from Pfizer-BioNTech COVID-19 vaccination. *Human Genetics and Genomics Advances* **3**, 100084 (2022).
7. Harris, P. A. *et al.* Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* **42**, 377–381 (2009).
8. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575 (2007).
9. Luo, Y. *et al.* A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. *Nature Genetics 2021 53:10* **53**, 1504–1516 (2021).
10. Sakaue, S. *et al.* A statistical genetics guide to identifying HLA alleles driving complex disease. *bioRxiv* 2022.08.24.504550 (2022) doi:10.1101/2022.08.24.504550.
11. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).

Figures

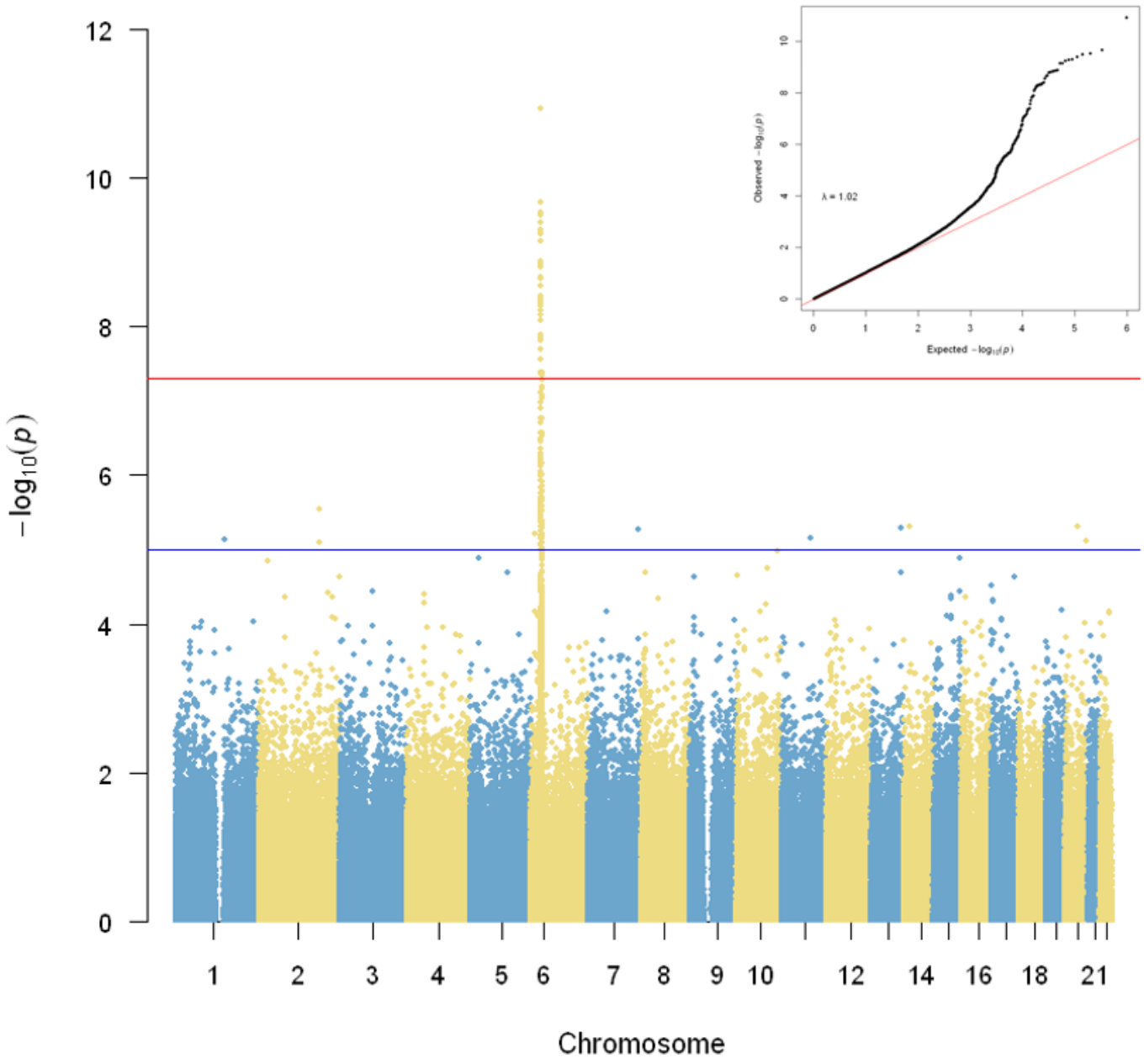

Figure 1

A locus on chromosome 6 is strongly associated with anti-spike IgG levels. Manhattan plot of the results of the GWAS between anti-spike inverse-normalized IgG values and 521,859 variants, teste in a linear regression model, using sex, age at vaccination, the first 5 principal components (PCs), and the time interval between the second vaccine dose and the serological as covariates. SNPs are plotted on the x-axis according to their genomic position (GChr38, hg38 release), and *P*-values ($-\log_{10}P$) for their association with IgG levels on the y-axis. The horizontal red line represents the threshold of genome-wide significance (*P*-value<5.0×10^{-8}). In the up-right corner is shown the Q-Q plot of observed and expected *P* values. Genomic inflation factor (λ) is reported.

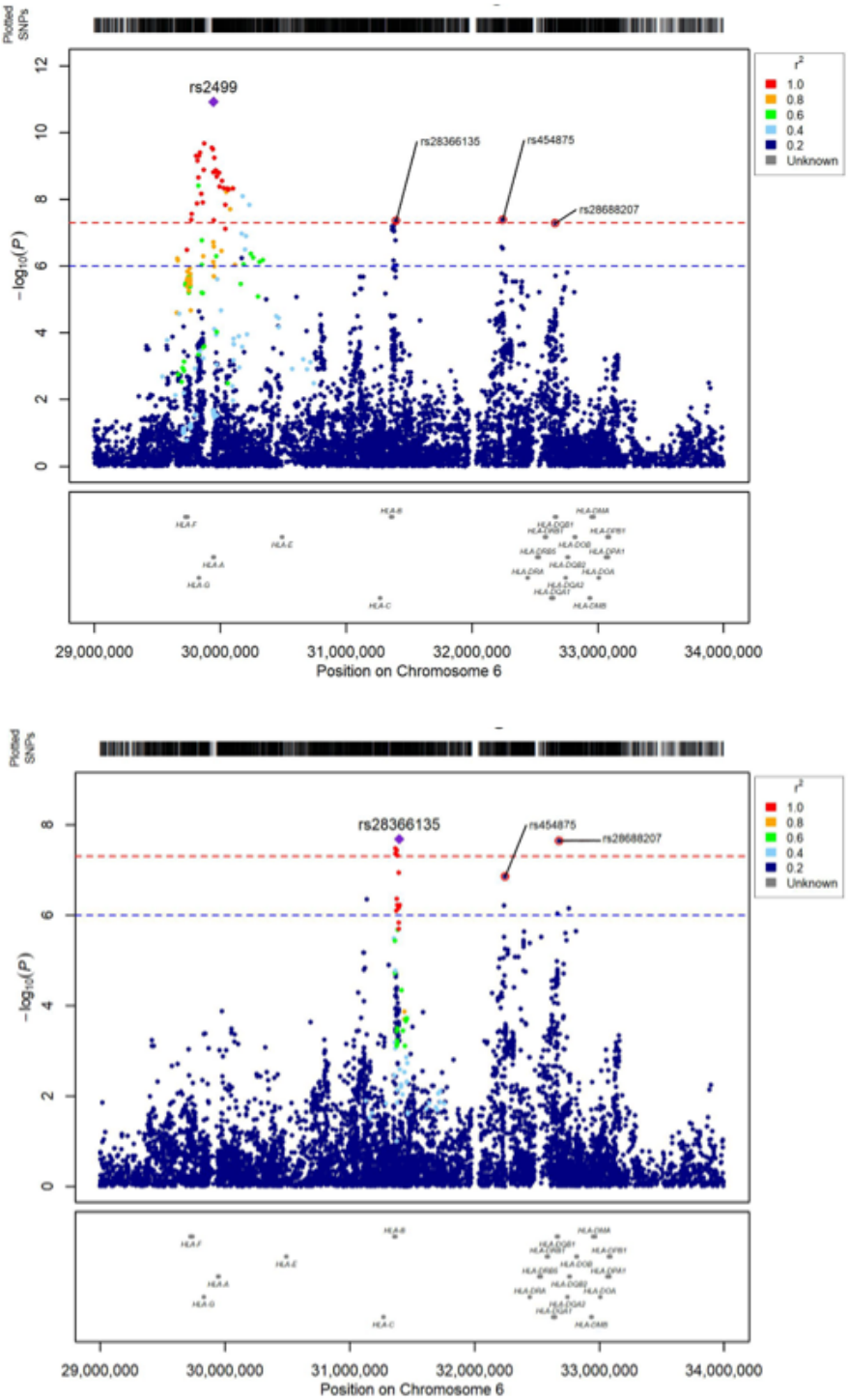

Figure 2

Multiple signals of association in the HLA locus suggest a multi-gene control of IgG production after vaccination. Zoom plot of the locus on chromosome 6 identified in the GWAS (top panel) and after adjustment for rs2499 (bottom panel). Plots span the region from 29 Mbp to 33 Mbp, containing HLA genes and al the top significant variants identified. SNPs are plotted on the x-axis according to their position on chromosome 6, and *P*-values ($-\log_{10}P$) for their association with IgG levels are plotted on the

y-axis. Horizontal red dashed line represents the threshold of significance (*P*-value<$5.0x10^{-8}$) whereas the blue one represents a suggestive threshold (*P*-value<$1.0x10^{-6}$). Dot color represents the level of linkage disequilibrium, expressed as r^2 between each SNP and the lead variant (purple diamond; top panel: rs2499; bottom panel:rs28366135).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplTab3.xlsx
- EspositoMSupplFigures.docx
- EspositoMSupplInfoNatComm.docx
- newalldatachr122linearresults.assoc.txt