HLA Alleles, COVID-19 Vaccine Antibody Response and Real-World Breakthrough Outcomes

Daniel Prieto-Alhambra (daniel.prietoalhambra@ndorms.ox.ac.uk)
Pharmaco- and Device Epidemiology Group, CSM, NDORMS, University of Oxford
https://orcid.org/0000-0002-3950-6346

Junqing Xie
Centre for Statistics in Medicine and NIHR Biomedical Research Centre Oxford, NDORMS, University of Oxford https://orcid.org/0000-0002-0040-0042

Beatriz Mothe
Infectious Diseases Department & IrsiCaixa AIDS Research Institute https://orcid.org/0000-0001-9975-407X

Marta Herraiz
Centre for Statistics in Medicine and NIHR Biomedical Research Centre Oxford, NDORMS, University of Oxford, Oxford, UK

Chunxiao Li
Medical Research Council Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, UK

Yu Xu
Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Annika Jödicke
Pharmaco- and Device Epidemiology Group, CSM, NDORMS, University of Oxford

Yaqing Gao
Nuffield Department of Population Health, Big Data Institute, University of Oxford, Old Rd Campus, Oxford, UK

Shuo Feng
Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford, UK
https://orcid.org/0000-0001-7855-0991

Jia Wei
University of Oxford

Zhuoyao Chen
University of Oxford https://orcid.org/0000-0003-4201-4649

Shenda Hong
Peking University https://orcid.org/0000-0001-7521-5127

Yeda Wu

Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, 4072, Australia
https://orcid.org/0000-0002-5977-1526

Binbin Su

School of Population Medicine and Public Health, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing, 100730, China

Xiaoying Zheng

School of Population Medicine and Public Health, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing, 100730, China

Catherine Cohet

Real-World Evidence Workstream, Data Analytics and Methods Task Force, European Medicines Agency: Amsterdam, Noord-Holland, NL

Raghib Ali

Medical Research Council Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, UK

Nicholas Wareham

University of Cambridge https://orcid.org/0000-0003-1422-2993

Article

Keywords:

Posted Date: July 18th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3178189/v1

Additional Declarations: There is **NO** Competing Interest.

Title: HLA Alleles, COVID-19 Vaccine Antibody Response and Real-World Breakthrough Outcomes

Authors: Junqing Xie[1], Beatriz Mothe[2], Marta Alcalde Herraiz[1], Chunxiao Li[3], Yu Xu[4-6], Annika M. Jödicke[1], Yaqing Gao[7], Shuo Feng[8], Jia Wei[9], Zhuoyao Chen[10], Shenda Hong[11,12], Yeda Wu[13], Binbin Su[14], Xiaoying Zheng[14*], Catherine Cohet[15*], Raghib Ali[3,16*], Nick Wareham[3*], Daniel Prieto Alhambra[1,17#]

Affiliations:

1. Centre for Statistics in Medicine and NIHR Biomedical Research Centre Oxford, NDORMS, University of Oxford, Oxford, UK
2. Infectious Diseases Department, IrsiCaixa AIDS Research Institute, Hospital Universitari Germans Trias i Pujol, Badalona, Spain
3. Medical Research Council Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, UK
4. Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
5. British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
6. Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge, UK
7. Nuffield Department of Population Health, Big Data Institute, University of Oxford, Old Rd Campus, Oxford, UK
8. Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford, UK
9. Nuffield Department of Medicine, Big Data Institute, University of Oxford, Oxford, UK
10. Centre for Medicines Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK
11. National Institute of Health Data Science, Peking University, Beijing, 100191, China
12. Institute of Medical Technology, Peking University Health Science Center, Beijing, 100191, China
13. Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, 4072, Australia
14. School of Population Medicine and Public Health, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing, 100730, China
15. Real-World Evidence Workstream, Data Analytics and Methods Task Force, European Medicines Agency: Amsterdam, Noord-Holland, NL
16. Public Health Research Center, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
17. Department of Medical Informatics, Erasmus University Medical Centre, Rotterdam, The Netherlands

*Senior authors

Corresponding author: Daniel Prieto Alhambra

daniel.prietoalhambra@ndorms.ox.ac.uk

Abstract: The rapid development, approval, and global distribution of COVID-19 vaccines represent an unprecedented intervention in public health history, with over 13 billion doses administered worldwide in two years. However, our understanding of the HLA genetic underpinnings of COVID-19 vaccine-induced antibody responses and their clinical implications for breakthrough outcomes remain limited. To bridge this knowledge gap, we designed and performed a series of genetic and epidemiological analyses among 368,098 vaccinated individuals, and a subset of 194,371 participants who had antibody serology tests. Firstly, we corroborated earlier findings that SNPs associated with antibody response were predominantly located in Major Histocompatibility Complex region, and that the expansive HLA-DQB1*06 allele family was linked to better antibody responses. However, our findings contest the claim that DQB1*06 alleles alone significantly impact breakthrough risks. Additionally, our results suggest that the specific DQB1*06:04 subtype could be the true causal allele, as opposed to the previously reported DQB1*06:02. Secondly, we identified and validated six new functional HLA alleles that independently contribute to vaccine-induced antibody responses. Moreover, we unravelled additive effects of variations across multiple HLA genes that, concurrently, change the risk of clinically relevant breakthrough COVID-19 outcomes. Finally, we detangled the overall vaccine effectiveness and showed that antibody positivity accounts for approximately 20% protection against breakthrough infection and 50% against severe outcomes. These novel findings provide robust population evidence demonstrating how variations within HLA genes strongly, collectively, and causally influence vaccine-induced antibody responses, and the risk of COVID-19 breakthrough infection and related outcomes, with implications for subsequent functional research and personalised vaccination.

Background

The COVID-19 pandemic has led to an unprecedented development, approval, and rollout of vaccines in the history of vaccinology. As of June 8, 2023, over 13 billion doses have been delivered worldwide, with over 70% of the global population receiving at least one dose.[1] The "one-size-fits-all" vaccination strategy[2,3] has shown remarkable variability in impact across different subpopulations.[4–6] For COVID-19 vaccine, this heterogeneity has led to a considerable number of fully vaccinated individuals remaining susceptible to COVID-19, requiring further booster doses.

Historically, immunogenetic variations, particularly the human leukocyte antigens (HLA) genes, have been recognized as significant influencers of adaptive immune responses to various vaccines, including hepatitis B, measles, and influenza.[7–9] However, the role of HLA in the context of COVID-19 vaccines, a novel exogenous antigen, remains largely elusive.[10] A recent genome-wide association study (GWAS)[11] focusing on post-vaccination antibodies against COVID-19 unveiled a number of genome-wide significant single nucleotide polymorphisms (SNPs) and pinpointed HLA-DQB1:06, DQB1:06:02 subtype in particular, as the potential causal alleles. Because of the complex structure in the major histocompatibility complex (MHC) region, including intense gene density, high polymorphism, and long linkage disequilibrium, the statistically fine-mapping of GWAS output proximally near this region is subject to critical methodological challenges and limitations.[12,13] As a result, the predicted DQB1:06 alleles likely represent a fraction of HLA allelic variations that influence heterogenous antibody response to COVID-19 vaccines, offering a glimpse into the larger, more comprehensive picture of its genetic basis. Furthermore, in the same study, HLA-DQB1:06 alleles were found to confer over 30% lower risk of breakthrough infection among a cohort of trials participants. The extent to which this association can be generalised to a broader general population is however unclear.[14]

This population-based research consisted of three interconnected analyses to fill existing knowledge gaps. Firstly, we aimed to replicate prior findings on the effect of HLA-DQB1:06 alleles on enhancing vaccine-induced antibody responses and reducing COVID-19 breakthrough infection risk. Secondly, we aimed to discover and validate novel genetic associations between HLA alleles and antibody response, and to investigate their clinical significance. Finally, we used Mendelian randomisation analyses to estimate the portion of effectiveness against SARS-CoV-2 infection and against severe COVID-19 attributable to vaccine-induced antibodies.

Results

Figure 1 summarizes the impact of eligibility criteria for participants of each analysis. Out of the 194,371 individuals who enrolled in the SARS-CoV-2 antibody and infection seroprevalence study, 181,812 who self-reported having received either one or two doses of COVID-19 vaccines at the time of antibody testing and had no previous COVID-19 infection as confirmed by the presence of antibodies against SARS-CoV-2 nucleocapsid antigen were included for the cross-sectional (CS) cohort. Baseline characteristics of the overall, discovery, and validation CS cohorts are detailed in **Table 1**. Participants were stratified according to whether they had received one dose (CS-1-dose) or two doses of COVID-19 vaccine (CS-2-dose), and further analyses were conducted accordingly. Following a median interval of 52 days (interquartile range: 36 to 64 days) after the first dose of vaccination, 28.4% of recipients tested positive for anti-spike SARS-CoV-2 antibodies. This proportion escalated to 65.5% among individuals who had the antibody test a median of 20 days (interquartile range: 8 to 33 days) after administration of the second vaccine dose.

In the prospective cohort (PS) of 357,806 individuals with recorded COVID-19 vaccination in the linked primary care database, the mean age was 69.3 years (SD: 8.0) with 55.3% being female (baseline demographics are shown in **Extended Table 1**). During the one-year follow-up period post-first dose vaccination, a total of 17,068 individuals experienced a breakthrough infection as confirmed by polymerase chain reaction test, and 1,353 were hospitalized with or died from COVID-19.

Verification and extension of previous key findings

In the GWAS of the first-dose antibody positivity, a number of genome-wide significant associations were observed. The lead SNP, rs2150392827 (Position: 6:32451297;OR, 0.865; P, $1.27*10^{-25}$), was located in the chromosome 6 MHC region with HLA genes to be nearest one (**Figure a**). Similar GWAS signals were found for the second-dose antibody positivity, with the lead SNP rs11490315 (Position: 6:32663047; OR, 0.863; P, $3.89*10^{-11}$), also nearest to HLA genes (**Figure e**).

In the candidate HLA-gene based analysis, 42.1% of people in the overall CS cohort carried at least one HLA-DQB1:06 alleles, a proportion similar to the previous report of 44.1% in trial's participants. The distribution of temporal duration between the date of vaccination and the date of antibody testing was similar between carriers and non-carriers of the HLA-DQB1:06 alleles, which verified our assumption that genetic variations were independent of this critical determinant for the antibody positivity (**Extended Table 2**). On average, the antibody response post-first dose was positive in 29.5% of HLA-DQB1:06 alleles carriers, compared to 27.4% in non-carriers. This difference of antibody positivity remained consistent over time and was observed up to eight weeks after the initial vaccination (**Figure b**). A higher but attenuated rate of antibody positivity was observed post-second dose in people carrying the DQB1*06 alleles (66.1%) than those without carrying the DQB1*06 alleles (65.2%) (**Figure f**).

After adjusting for potential confounding factors, a significant positive association was found between HLA-DQB1*06 alleles and the antibody response in the CS-1-dose cohort (odds ratio OR 1.13, 95% CI 1.10 to 1.17) and in the CS-2-dose cohort (OR 1.04, 95% CI 1.00 to 1.07). When considering the five specific subtypes of DQB1:06 alleles (**Figure c**), the DQB1:06:04 subtype showed a markedly stronger association with vaccine-induced antibodies, compared to the DQB1:06:02 subtype, which was previously hypothesized to be primarily responsible for binding to the SARS-CoV-2 spike protein more effectively.

Furthermore, the frequency of DQB1:06 alleles shown no association with COVID-19 outcomes after vaccination (therefore breakthrough COVID-19) (**Figure d**), with adjusted OR corresponding to 0.98 (95% CI 0.94 to 1.02) for infection, and 1.00 (95% CI of 0.88 to 1.13) for severity, defined as hospitalisation with or death from COVID-19.

Discovery and validation of novel HLA allele associations with antibody response

We then proceeded to identify novel HLA associations with antibody response by conducting a gene-based analysis across 213 classic alleles in 11 HLA genes. In general, the associations were more pronounced in relation to the first-dose (**Figure 2a** left) than the second-dose (**Figure 2a** right) antibody response.

In total, 15 alleles of 2 class I and 6 class II HLA genes (A, C, DQA1, DQB1, DRB1, DRB3, DRB5, DPB1) retained statistical significance in both the discovery and validation dataset after FDR-correction (**Extended Table 3**). Among them, 14 HLA alleles were associated with first-dose antibody response, 2 with second-dose antibody response, and 1 with both (**Figure 2b)**. 9 HLA alleles were linked to a higher antibody positivity (enhancer), while 6 alleles linked to lower antibody positivity (suppressor).

The HLA-DQB1*06:04 allele was the most potent enhancer of antibody response (β = 0.213 in discovery and 0.384 in validation) following first-dose vaccination. However, no effect was detected for this allele after the second dose. In contrast, DQA1*01:01 consistently functioned as an antibody suppressor either following the first (β = -0.122 in discovery and -0.180 in validation) or second vaccination (β = -0.137 in discovery and -0.165 in validation). Several haplotypes or clusters of HLA alleles were identified among all validated alleles (**Figure 2c**): Cluster 1: DQA1*01:02; DRB5*01:01; DRB1*15:01, Cluster 2: DRB3*01:01; DRB1*03:01; DQB1*02:01, Cluster 3: DRB1*01:01; DQB1*05:01; DQA1*01:01, Cluster 4: DQB1*06:04; DRB3*03:01; DRB1*13:02.

After removing highly correlated alleles from each cluster, 7 independent HLA alleles remained significantly associated with antibody response (DQB1*06:04, DQA1*01:02, DRB3*01:01, C*16:01, DPB1*10:01, A*03:01, DQA1*01:01). All alleles, except for A*03:01, appeared to have an additive effect on antibody response (**Figure 2d**).

We constructed genetic scores (GS) of antibody response to assess its combined impact by aggregating all HLA allelic variations with different statistical thresholding (**Methods**). The distribution of each GS is provided in **Supplement 1**. The results showed a significant association of the GS with antibody response. More specifically, for GS1, an adjusted OR of 1.14 (95% CI: 1.13 to 1.16) per SD increase was found for antibody positivity following first-dose vaccination, and an OR of 1.10 (95% CI: 1.08 to 1.11) for 2-dose antibody positivity. For GS2, the corresponding ORs were 1.16 (95% CI: 1.14 to 1.18) and 1.12 (95% CI: 1.10 to 1.14) respectively. GS3 presented the strongest association (1.17 [95% CI: 1.15 to 1.18] for the first dose, and 1.14 [95% CI: 1.12 to 1.15] for the second dose) and was selected for subsequent analyses (**Supplement 2**).

Extrapolation of HLA effects on COVID-19 breakthrough outcomes

Among the 7 validated HLA alleles with associations with either first-dose or second-dose antibody response, none of them was individually associated with clinical outcomes of breakthrough COVID-19 (**Extended Table 4**). However, a clear dose-dependent correlation between antibody response and COVID-19 outcome predisposed by each HLA allele was observed (**Figure 3a-b**). Individuals who carried an allele that enhanced antibody response had a numerically (but not significant) lower risk of breakthrough COVID-19 outcomes, with HLA-DQB1:06:04 exhibiting most strongest effects on both

breakthrough infection (Hazard ratio HR 0.94 95% 0.88 to 1.01) and severe COVID-19 (HR 0.90 95% 0.70 to 1.15). In general, the severe COVID-19 was more correlated with antibody response conferred by each HLA allele than the breakthrough infection outcome, as suggested by a steeper slope **Figure 3a-b**.

In comparison with the GS of first-dose antibody response, GS of second-dose antibody response showed a stronger association with breakthrough COVID-19 outcomes, including infection (HR per GS unit increase 0.94, 95% CI 0.90 to 0.98) and hospitalization or death (HR 0.87, 0.76 to 0.99). Furthermore, we identified 5% of participants who had a 26% (HR 1.26 95% CI 1.01 to 1.58) or a 34% (HR 1.34 95% CI 1.08 to 1.67) increased risk of severe COVID-19 based on first-dose or second-dose GS respectively (**Figure 3c-d, Extended Table 4**).

Finally, we estimated the causal effect of vaccine-induced antibody positivity on breakthrough COVID-19 outcomes. Based on the 7 validated independent alleles used as instrumental variables in Mendelian randomisation analyses, we demonstrated an effectiveness of antibody positivity ranging from 16.84% (95% CI 6.39% to 26.12%) to 21.68% (5.99% to 34.76%) in preventing breakthrough infection, and 46.44% (95% CI -1.49% to 71.73%) to 49.94% (95% CI 15.52% to 70.34%) in preventing severe COVID-19 (**Table 2**). Sensitivity analyses produced consistent results (**Supplement 3**).

Discussion

The present study is the largest and most comprehensive to investigate the genetic determinants of antibody response to COVID-19 vaccines and their impact on subsequent clinical outcomes, with a particular focus on the HLA genes. Firstly, we confirmed earlier findings that the HLA-DBQ1*06 allele family enhances the antibody response to initial COVID-19 vaccination. However, we refuted the notion that it can meaningfully impact clinical outcomes of breakthrough COVID-19 in isolation. Additionally, we found that the specific HLA-DBQ1*06:04 allele is likely the driving subtype, as opposed to the previously hypothesized HLA-DBQ1*06:02.

Secondly, we identified and validated six new HLA alleles that independently influence antibody response after COVID-19 vaccines. The cumulative effect of multiple HLA variants, represented in the form of a genetic score, was associated with both the susceptibility to, and the severity of breakthrough COVID-19.

Thirdly, we estimated that vaccine-induced antibody response alone confers approximately 20% protection against SARS-CoV-2 infection, and to a much greater extent (around 50%) against severe COVID-19 outcomes.

Findings in context

Few population-based studies have been undertaken to understand the relationship between HLA alleles and antibody responses to COVID-19 vaccines. The majority of these have been constrained by limited sample size, typically including dozens of individuals recruited from convenient populations, such as healthcare workers or clinical trial participants. The first significant HLA allele was reported by a Spanish study of 87 healthcare workers,[15] showing a positive association between HLA-DRB1*07:01 and antibody levels 30 days post-receipt of the second dose of the mRNA-1273 vaccine. This finding was partially supported by our data, with an elevated antibody positivity post-second dose vaccination amongst HLA-DRB107:01 carriers. However, this association did not meet statistical significance. Other studies involving healthcare workers from Japan (100 participants),[16] Italy (56 participants),[17] and the UK (251 participants)[18] have all failed to detect any significant associations between HLA alleles and anti-spike antibody response or seroprevalence following vaccination, except for the last UK study supporting DRB1*15:01 linked to stronger spike T-cell responses. In our study, DRB1*15:01 is one of the validated alleles associated with increased antibody rate, but it was highly correlated with DQA1*01:02 and DRB5*01:01 alleles.

In addition to vaccine antibody response, a GWAS among 17,440 vaccinated participants reported a robust association between HLA-A03:01 and vaccine-related side effects, such as chills and fever.[19] In our study, the HLA-A03:01 allele was also associated with an enhanced antibody response. This observation may reflect the epidemiological evidence of a correlation between reactogenicity and immunogenicity following COVID-19 vaccination[20] and suggests a potential genetic colocalization of stronger antibody responses and higher risk side effects via the HLA-A03:01 allele. However, further mechanistic research is warranted to confirm this hypothesis.

To date, there is a paucity of data on the impact of HLA variations on the risk of post-vaccination breakthrough COVID-19 outcomes. One previous study based on pivotal trial data found a substantial reduction (over 30%) in breakthrough infection among HLA-DQB1:06 allele carriers.[11] However, we were unable to replicate this finding in our larger and more extensive community-based population-based cohort. This divergence could

be attributed to immunosenescence,[21,22] considering the relatively younger (mean age of 37 years) and healthier population in the prior study, and the design, as trial participants exclusively received the ChAdOx1 vaccine.

There is a clinical agreement that vaccination is efficacious and offers superior protection against severe COVID-19 than infection. This consensus is based on evidence gathered from initial vaccine trials as well as routinely collected health data. Our research represents the first application of genetic methodologies to estimate vaccine effectiveness. Our findings can further improve confidence in COVID-19 vaccines and tackle vaccine hesitancy. Moreover, our estimations provide quantitative evidence of the proportion of protection attributable to antibody-mediated immunity following vaccination, informing the additional protection conferred by cellular immunity. Notably, a recent meta-analysis of 15 studies modelled that detectable neutralizing antibodies may provide approximately 20% protection against infection and 50% protection against severe COVID-19.[23] This figure was surprisingly aligned with our estimation.

Implications

While current vaccines have shown substantial effectiveness, cases of non-responsiveness and breakthrough infections continue to persist, even following booster doses. Our study has produced a comprehensive map of HLA associations with vaccine-induced antibody responses. These most significant alleles should be prioritized in mechanistic research to elucidate their functional implications, thereby informing the development of novel vaccines. Compared to previous genetic risk scores that predicted COVID-19 outcomes by aggregating numerous SNPs from COVID-19 GWAS,[24] our newly developed HLA genetic score is biologically understandable. This GS could be implemented through a tailored HLA-based genotyping array.

Limitations

Although our gene allele-based investigation was better positioned to identify functional HLA variants that have potentially been overlooked in previous SNP-based GWAS studies, experimental data are still needed to support the associations found. Antibody response is a dynamic and intricate process. The genetic profile for binary anti-S antibody seropositivity may not fully encapsulate other immune attributes such as peak antibody titre, durability, and specificity. Also, it is imperative to explore these associations in individuals who have received booster doses in future research. We used routine linked public health data collection rather than active surveillance to capture COVID-19 events. The resulting potential outcome misclassification may lead to underestimation of genetic effects on asymptomatic breakthrough infection, but less likely for severe cases.

Using an extensive dataset incorporating host genetics, immunological biomarkers, vaccination statuses, and disease outcomes, alongside advanced methodologies and multidisciplinary expertise within our team, we successfully identified six previously unreported HLA alleles that regulate the COVID-19 vaccine antibody response. Furthermore, we demonstrated that variations within the HLA strongly, collectively, and causally impact critical COVID-19 outcomes within a vaccinated population. These novel can inform vaccine developments by substantially advancing understanding of genetic mechanisms of vaccine immunogenicity and clinical effectiveness, which my open new paths for personalised vaccination in the foreseeable future.

Tables and Figures

Figure 1 Study flowchart

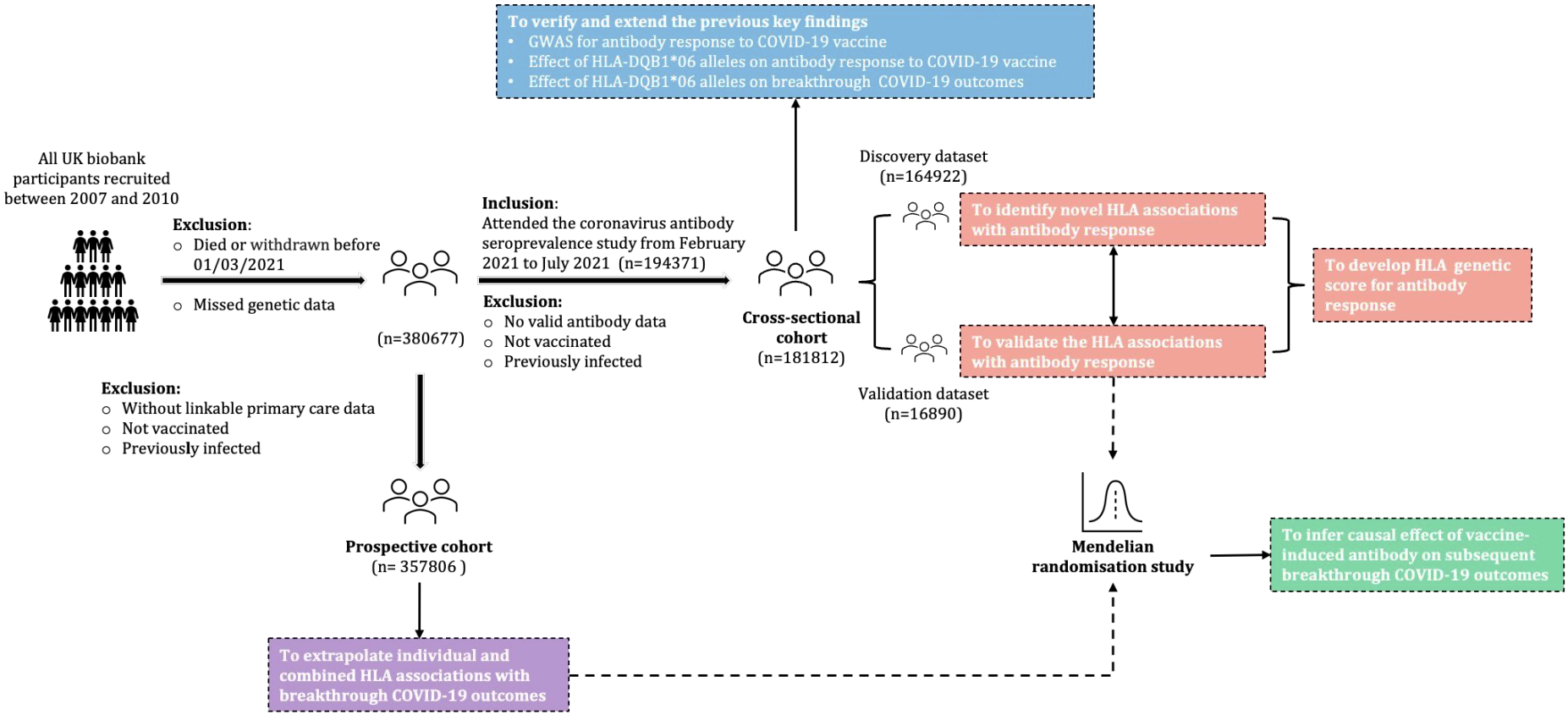
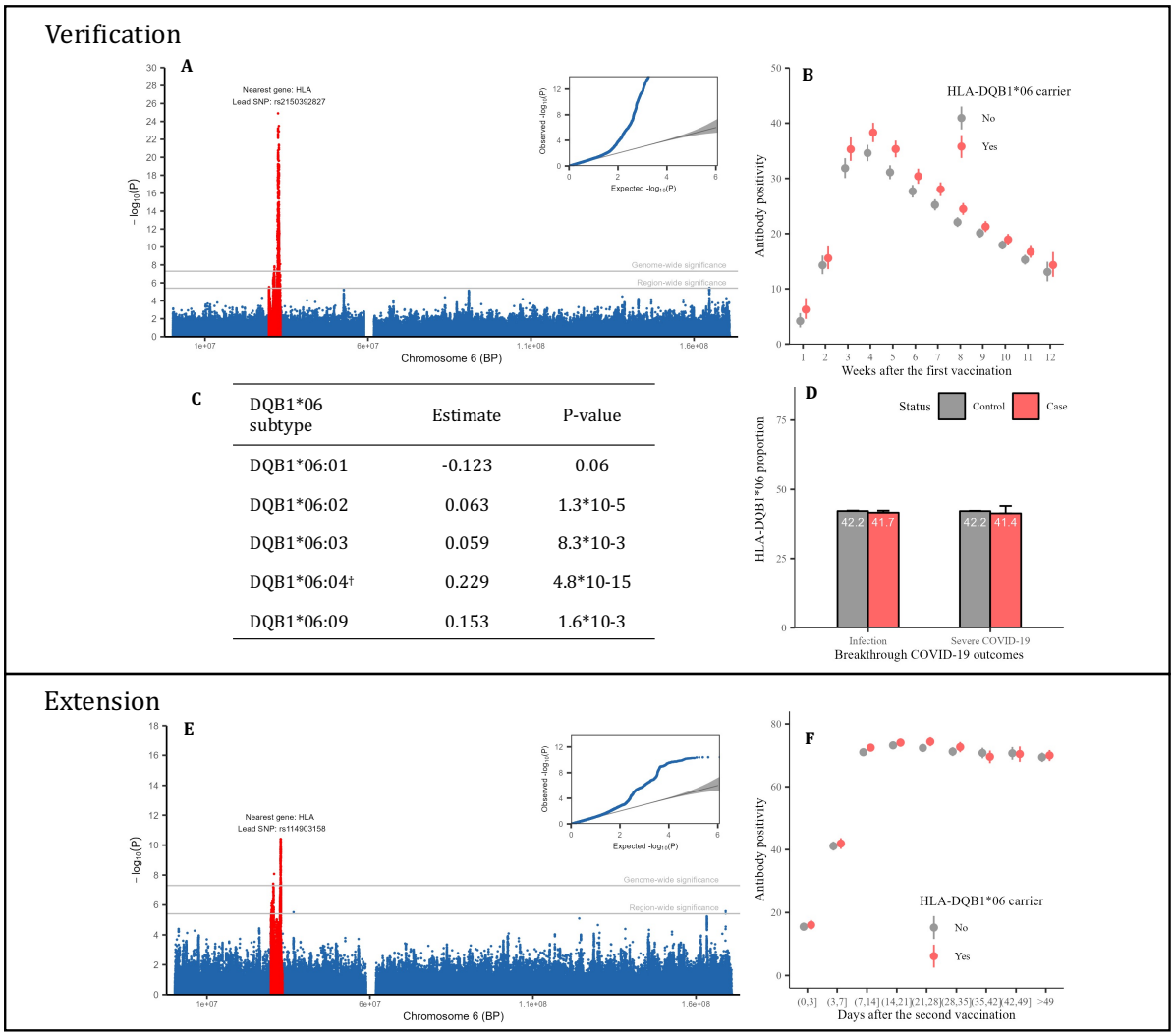


Table 1 Baseline characteristics of eligible participants with the antibody test in the cross-sectional cohort

	One-dose vaccine recipients			Two-dose vaccine recipients		
	All	Discovery dataset	Validation dataset	All	Discovery dataset	Validation dataset
Number	107175	97097	10078	67825	61013	6812
Age, mean (SD)	67.27 (7.59)	67.23 (7.60)	67.61 (7.53)	71.22 (6.56)	71.21 (6.57)	71.29 (6.50)
Age category, n (%)						
50-59 years	42872 (40.0)	38671 (39.8)	4201 (41.7)	45931 (67.7)	41302 (67.7)	4629 (68.0)
60-69 years	19970 (18.6)	18223 (18.8)	1747 (17.3)	5085 (7.5)	4582 (7.5)	503 (7.4)
>=70 years	44333 (41.4)	40203 (41.4)	4130 (41.0)	16809 (24.8)	15129 (24.8)	1680 (24.7)
Sex, n (%)						
Female	59761 (55.8)	54584 (56.2)	5177 (51.4)	39417 (58.1)	35733 (58.6)	3684 (54.1)
Male	47414 (44.2)	42513 (43.8)	4901 (48.6)	28408 (41.9)	25280 (41.4)	3128 (45.9)
Ethnicity						
White	90511 (84.5)	81825 (84.3)	8686 (86.2)	56814 (83.8)	50933 (83.5)	5881 (86.3)
Others	16664 (15.5)	15272 (15.7)	1392 (13.8)	11011 (16.2)	10080 (16.5)	931 (13.7)
Weeks since the latest vaccination, n (%)						
1 week	1677 (1.6)	1513 (1.6)	164 (1.6)	13222 (19.5)	11941 (19.6)	1281 (18.8)
2 weeks	2957 (2.8)	2683 (2.8)	274 (2.7)	12911 (19.0)	11624 (19.1)	1287 (18.9)
3 weeks	4609 (4.3)	4173 (4.3)	436 (4.3)	10715 (15.8)	9675 (15.9)	1040 (15.3)
4 weeks	7053 (6.6)	6408 (6.6)	645 (6.4)	8939 (13.2)	7963 (13.1)	976 (14.3)
5 weeks	9188 (8.6)	8361 (8.6)	827 (8.2)	7018 (10.3)	6295 (10.3)	723 (10.6)
6 weeks	10489 (9.8)	9558 (9.8)	931 (9.2)	4989 (7.4)	4483 (7.3)	506 (7.4)
7 weeks	12350 (11.5)	11171 (11.5)	1179 (11.7)	3426 (5.1)	3068 (5.0)	358 (5.3)
8 weeks	14840 (13.8)	13424 (13.8)	1416 (14.1)	2438 (3.6)	2211 (3.6)	227 (3.3)
9 weeks	15613 (14.6)	14088 (14.5)	1525 (15.1)	1718 (2.5)	1543 (2.5)	175 (2.6)
10 weeks	14655 (13.7)	13263 (13.7)	1392 (13.8)	1275 (1.9)	1153 (1.9)	122 (1.8)
11 weeks	11319 (10.6)	10272 (10.6)	1047 (10.4)	733 (1.1)	650 (1.1)	83 (1.2)
12 weeks	2425 (2.3)	2183 (2.2)	242 (2.4)	441 (0.7)	407 (0.7)	34 (0.5)
Vaccine types, n (%)[1] received, n (%)[1]						
ChAdOx1	39948 (37.3)	36366 (37.5)	3582 (35.5)	18578 (27.4)	16721 (27.4)	1857 (27.3)
BNT162b2	15188 (14.2)	13690 (14.1)	1498 (14.9)	17587 (25.9)	15886 (26.0)	1701 (25.0)
Unknown	52039 (48.6)	47041 (48.4)	4998 (49.6)	31660 (46.7)	28406 (46.6)	3254 (47.8)

[1] Calculated based on a subgroup of participants who had complete linkage to primary care records to obtain information on COVID-19 vaccine types

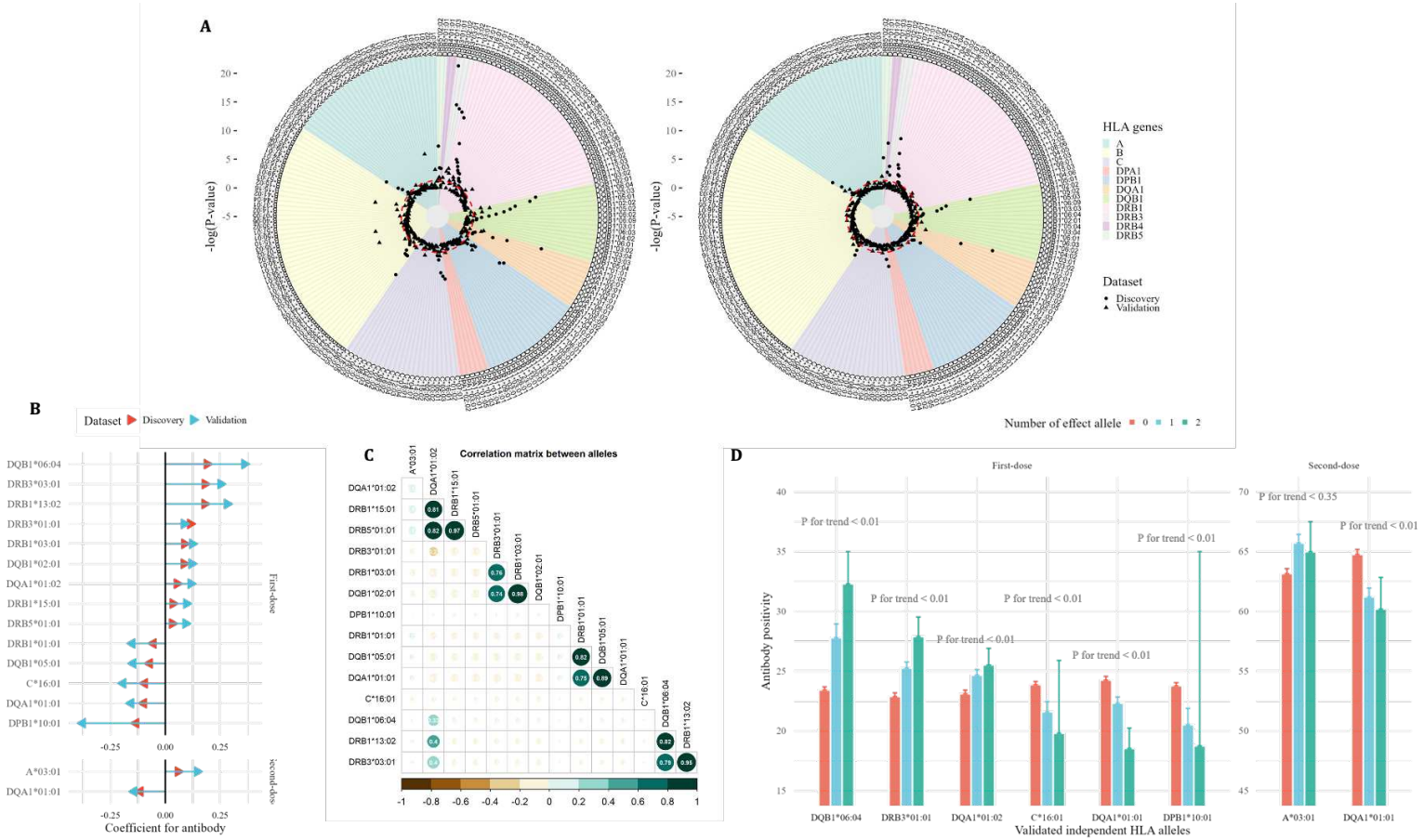
Figure 2 Verification and extension of previous key findings on HLA-DQB1*06 alleles



DQB1*06 subtype	Estimate	P-value
DQB1*06:01	-0.123	0.06
DQB1*06:02	0.063	1.3*10-5
DQB1*06:03	0.059	8.3*10-3
DQB1*06:04†	0.229	4.8*10-15
DQB1*06:09	0.153	1.6*10-3

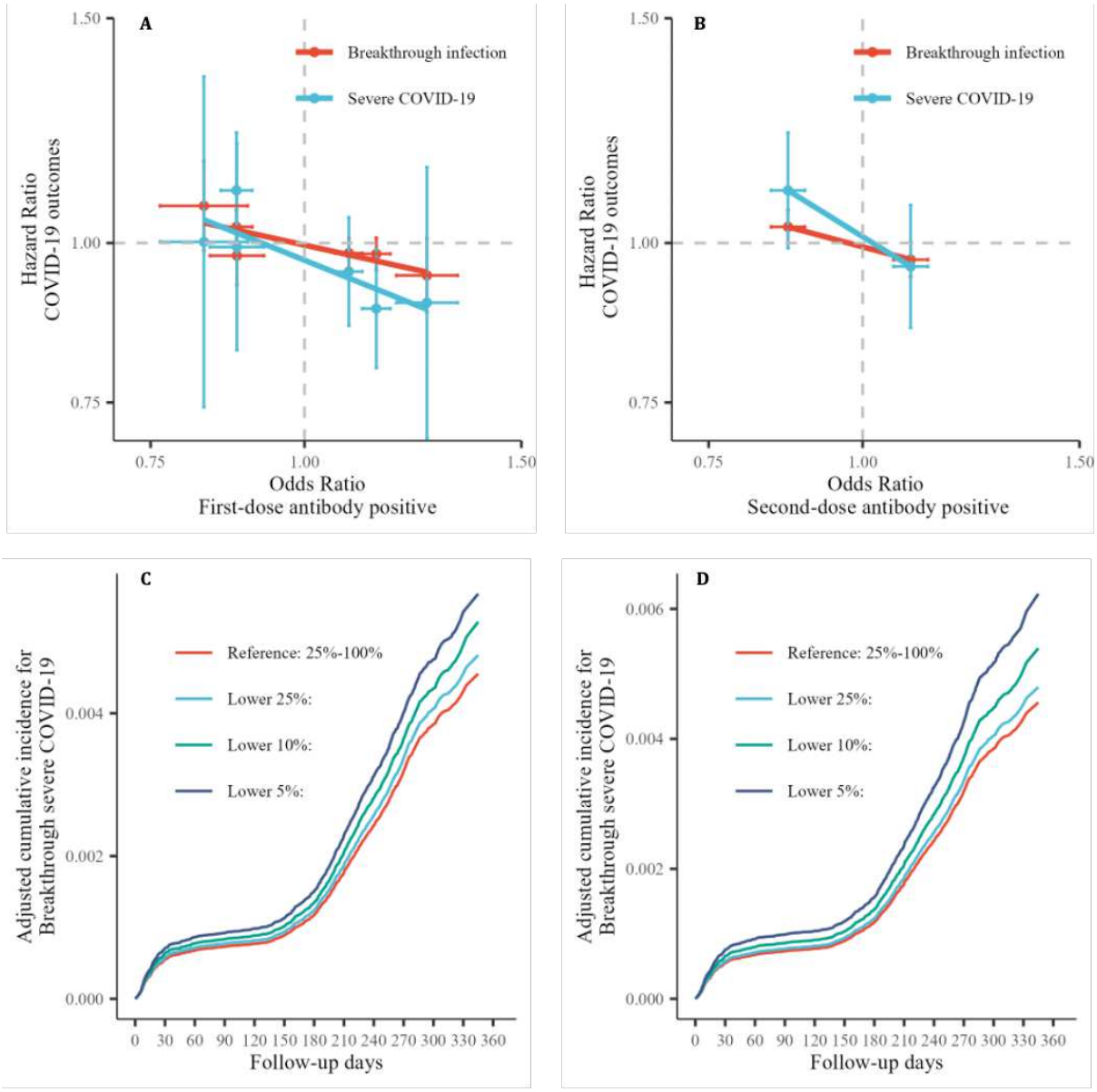
The verification and extension results are presented in panels a-d and e-f, respectively. (**A**) Manhattan plot of GWAS in chromosome 6 on antibody response to first-dose COVID-19 vaccine. SNPs within the MHC region are labelled in red colour, with the nearest gene to the lead SNP presented. (**B**) The proportion of individuals with a positive antibody response stratified by their HLA-DQB1*06 status in the CS-1 cohort. Individuals with one or two DQB1*06 alleles are classified as carriers, whereas those with zero DQB1*06 alleles are as non-carriers. The maximum observation period is to eight weeks following the first vaccination. (**C**) The association of five specific HLA-DQB1*06 subtypes with the antibody response. The estimate presents the coefficient of each allele in the logistic regression model. The p-value is presented raw and does not account for multiple test corrections. † indicates the strongest association among the five DQB1*06 alleles subtype (**D**) The proportion of individuals carrying HLA-DQB1*06 alleles, comparing COVID-19 cases with controls (those not infected). Specific numerical data is labelled with white text within each bar. (**E**) Manhattan plot of GWAS in chromosome 6 on antibody response to second-dose COVID-19 vaccine. SNPs within the MHC region are labelled in red colour, with the nearest gene to the lead SNP presented. (**F**) The proportion of individuals with a positive antibody response stratified by their HLA-DQB1*06 status in the CS-2 cohort. Individuals with one or two DQB1*06 alleles are classified as carriers, whereas those with zero DQB1*06 alleles are as non-carriers. The maximum observation period is to eight weeks following the second vaccination. The x-axis is grouped slightly different from the panel B due to the different time distribution (see **Table 1**).

Figure 2 Association of 203 HLA alleles with COVID-19 vaccine antibody response



(**A**) Circular Manhattan plot of HLA-gene based genetic analysis for the first-dose antibody response (left panel) and second-dose antibody response (right panel) in the discovery dataset. The red dashed circle delineating the y-axis indicates a P-value of 0.05. (**B**) Association strength and direction of HLA alleles that significant in both the discovery and validation datasets following the False Discovery Rate correction of total number of alleles tested. The upper portion of the panel represents the results for the first-dose antibody response, while the lower portion represents the second-dose antibody response. (**C**) Pairwise correlation plot between all significant HLA alleles. (**D**) Antibody positivity rate among individuals with varying numbers of risk alleles (0, 1, 2) at each locus. The left portion of the panel represents the results for the first-dose antibody response, while the right portion for the second-dose antibody response.

Figure 3 Extrapolation of individual and combined HLA effects



(**A**) Correlation between the effect of six independent HLA alleles on first-dose antibody response and their subsequent effects on breakthrough COVID-19 outcomes. (**B**) Correlation between the effect of two independent HLA alleles on second-dose antibody response and their subsequent effects on breakthrough COVID-19 outcomes. Age-sex adjusted cumulative incidence curve of severe COVID-19 stratified by the first-dose (**C**) and second-dose (**D**) genetic scores.

Table 2 Effect of vaccine-induced antibody on breakthrough COVID-19 outcomes

Exposure	Breakthrough outcomes	Number of HLA alleles	Relative Risk	Effectiveness (%)
First-dose antibody	Infection	6	0.83 (0.74 to 0.94)	16.84 (6.39 to 26.12)
	Severe COVID-19	6	0.55 (0.36 to 0.84)	49.94 (15.52 to 70.34)
Second-dose antibody	Infection	2	0.78 (0.65 to 0.94)	21.68 (5.99 to 34.76)
	severe COVID-19	2	0.54 (0.28 to 1.01)	46.44 (-1.49 to 71.73)

Extended Table 1 Baseline characteristics of participants at the time of receiving the first COVID-19 vaccine dose in the prospective cohort

	All participants
Number of recipients	357,806
Age, mean (SD)	69.28 (7.99)
Sex, n (%)	
Female	197824 (55.3)
Male	159982 (44.7)
Ethnicity	
White	328557 (91.8)
Other ethnic groups	29249 (8.2)
Vaccine types, n (%)[1]	
ChAdOx1	135237 (37.8)
BNT162b2	83202 (23.3)
Unknown	139367 (39.0)

Extended Table 2 Baseline characteristics of HLA DQB1*06 carriers and non-carriers in the cross-sectional cohort

	One-dose vaccine recipients			Two-dose vaccine recipients		
	DQB1*06 carriers	DQB1*06 non-carriers	SMD	DQB1*06 carriers	DQB1*06 non-carriers	SMD
Number	61984	45191		39317	28508	
Age, mean (SD)	67.26 (7.59)	67.28 (7.59)	0.002	71.19 (6.56)	71.26 (6.56)	0.011
Age category, n (%)						
50-59 years	24865 (40.1)	18007 (39.8)	0.007	26592 (67.6)	19339 (67.8)	0.009
60-69 years	11567 (18.7)	8403 (18.6)		2985 (7.6)	2100 (7.4)	
>=70 years	25552 (41.2)	18781 (41.6)		9740 (24.8)	7069 (24.8)	
Sex, n (%)						
Female	34590 (55.8)	25171 (55.7)	0.002	22879 (58.2)	16538 (58.0)	0.004
Male	27394 (44.2)	20020 (44.3)		16438 (41.8)	11970 (42.0)	
Ethnicity						
White	52372 (84.5)	38139 (84.4)	0.003	33007 (84.0)	23807 (83.5)	0.012
Others	9612 (15.5)	7052 (15.6)		6310 (16.0)	4701 (16.5)	
Weeks since the latest vaccination, n (%)						
1 week	988 (1.6)	689 (1.5)	0.021	7609 (19.4)	5613 (19.7)	0.024
2 weeks	1709 (2.8)	1248 (2.8)		7448 (18.9)	5463 (19.2)	
3 weeks	2613 (4.2)	1996 (4.4)		6205 (15.8)	4510 (15.8)	
4 weeks	4051 (6.5)	3002 (6.6)		5198 (13.2)	3741 (13.1)	
5 weeks	5286 (8.5)	3902 (8.6)		4074 (10.4)	2944 (10.3)	
6 weeks	6040 (9.7)	4449 (9.8)		2899 (7.4)	2090 (7.3)	
7 weeks	7192 (11.6)	5158 (11.4)		2030 (5.2)	1396 (4.9)	
8 weeks	8611 (13.9)	6229 (13.8)		1447 (3.7)	991 (3.5)	
9 weeks	8984 (14.5)	6629 (14.7)		1006 (2.6)	712 (2.5)	
10 weeks	8552 (13.8)	6103 (13.5)		729 (1.9)	546 (1.9)	
11 weeks	6511 (10.5)	4808 (10.6)		433 (1.1)	300 (1.1)	
12 weeks	1447 (2.3)	978 (2.2)		239 (0.6)	202 (0.7)	
Vaccine types, n (%)[1]						
ChAdOx1	23039 (37.2)	16909 (37.4)	0.012	10824 (27.5)	7754 (27.2)	0.022
BNT162b2	8893 (14.3)	6295 (13.9)		10316 (26.2)	7271 (25.5)	
Unknown	30052 (48.5)	21987 (48.7)		18177 (46.2)	13483 (47.3)	

SMD, standardised mean difference

[1] Calculated based on a subgroup of participants who had complete linkage to primary care records to obtain information on COVID-19 vaccine types.

Extended Table 3 Significant HLA alleles for COVID-19 vaccine antibody response in both discovery and validation datasets

	Discovery dataset			Validation dataset			
Significant alleles[1]	**Beta**	**P-value**	**FDR-corrected P-value**	**Beta**	**P-value**	**FDR-corrected P-value**	**Direction of effects**
Antibody response after the first-dose vaccination							
DQB1*06:04	0.213	4.2*10^-12	1.1*10^-10	0.384	5.5*10^-05	2.6*10^-03	Positive
DRB1*13:02	0.203	1.9*10^-14	7.5*10^-13	0.305	2.5*10^-04	5.2*10^-03	Positive
DRB3*03:01	0.204	6.4*10^-15	4.3*10^-13	0.275	9.5*10^-04	7.4*10^-03	Positive
DRB1*03:01	0.110	1.4*10^-13	4.7*10^-12	0.146	1.6*10^-03	9.4*10^-03	Positive
DQB1*02:01	0.108	3.2*10^-13	9.1*10^-12	0.144	1.9*10^-03	9.7*10^-03	Positive
DQA1*01:02	0.077	1.3*10^-08	2.1*10^-07	0.139	7.8*10^-04	7.3*10^-03	Positive
DRB1*15:01	0.058	1.6*10^-04	1.1*10^-03	0.119	1.1*10^-02	4.6*10^-02	Positive
DRB5*01:01	0.055	3.1*10^-04	1.9*10^-03	0.116	1.4*10^-02	4.7*10^-02	Positive
DRB3*01:01	0.137	2.8*10^-22	5.6*10^-20	0.109	1.4*10^-02	4.7*10^-02	Positive
DQB1*05:01	-0.095	1.6*10^-08	2.3*10^-07	-0.170	1.4*10^-03	9.3*10^-03	Negative
DRB1*01:01	-0.078	3.8*10^-05	2.9*10^-04	-0.175	4.0*10^-03	1.9*10^-02	Negative
DQA1*01:01	-0.122	9.3*10^-15	4.7*10^-13	-0.180	3.3*10^-04	5.2*10^-03	Negative
C*16:01	-0.117	1.4*10^-05	1.1*10^-04	-0.216	1.2*10^-02	4.7*10^-02	Negative
DPB1*10:01	-0.157	3.1*10^-04	1.9*10^-03	-0.519	6.4*10^-04	7.3*10^-03	Negative
Antibody response after the second-dose vaccination							
A*03:01	0.081	3.0*10^-06	7.7*10^-05	0.168	1.2*10^-03	1.5*10^-02	Positive
DQA1*01:01	-0.137	7.6*10^-16	1.5*10^-13	-0.165	1.2*10^-03	1.5*10^-02	Negative

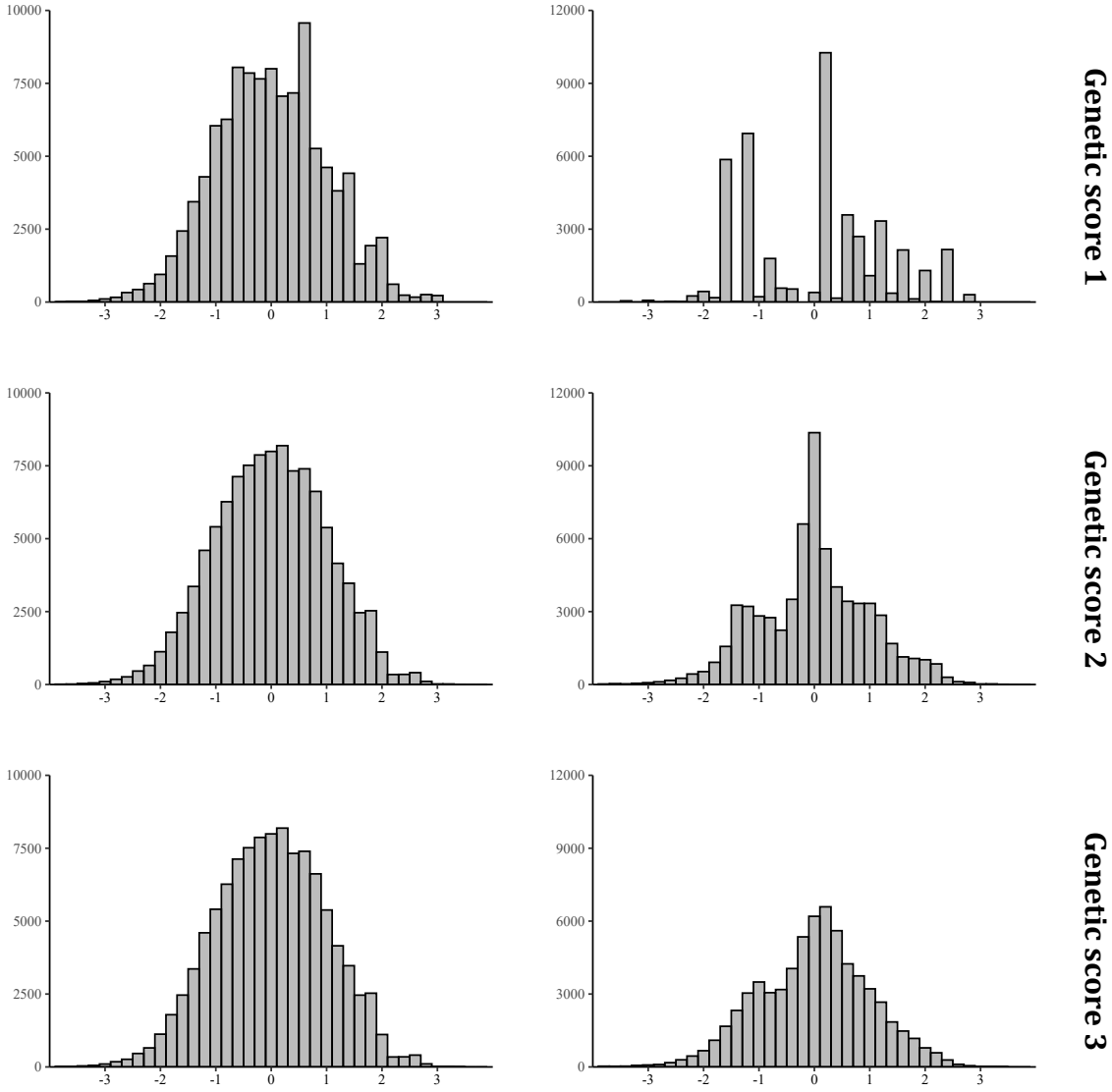
[1]The alleles is ordered by the magnitude of associations as measured by the beta value of the logistic regression model.

Extended Table 4 Hazard ratios for the association between individual HLA alleles, composite HLA genetic score, and breakthrough COVID-19 outcomes.

	Breakthrough infection	Breakthrough severe COVID-19
7 validated HLA alleles		
A*03:01	0.97 (0.94 to 1.00)	0.96 (0.86 to 1.07)
C*16:01	0.98 (0.93 to 1.03)	0.99 (0.82 to 1.20)
DPB1*10:01	1.07 (0.99 to 1.16)	1.00 (0.74 to 1.35)
DQA1*01:01	1.03 (1.00 to 1.06)	1.10 (0.99 to 1.22)
DQA1*01:02	0.98 (0.96 to 1.01)	0.95 (0.86 to 1.05)
DQB1*06:04	0.94 (0.88 to 1.01)	0.90 (0.70 to 1.15)
DRB3*01:01	0.98 (0.95 to 1.01)	0.91 (0.80 to 1.00)
Continuous first-dose GS	1.02 (0.99 to 1.05)	0.89 (0.82 to 0.98)*
First-dose GS subgroups		
Lower 25% vs Higher 25%-100%	NC	1.06 (0.94 to 1.20)
Lower 10% vs Higher 25%-100%	NC	1.16 (0.97 to 1.37)*
Lower 5% vs Higher 25%-100%	NC	1.26 (1.01 to 1.58) *
Continuous second-dose GS	0.94 (0.90 to 0.98)*	0.87 (0.76 to 0.99)*
Second-dose GS subgroups		
Lower 25% vs Higher 25%-100%	1.04 (1.01 to 1.08)*	1.05 (0.93 to 1.19)
Lower 10% vs Higher 25%-100%	1.05 (1.00 to 1.11)*	1.16 (0.98 to 1.38)*
Lower 5% vs Higher 25%-100%	1.05 (0.98 to 1.12)	1.34 (1.08 to 1.67)*

GS, genetic score. The asterisk denotes statistical significance at P<0.05 without correction. NC, Not calculated due to the lack of statistical significance for the continuous form of genetic score.

Supplement 1 Distribution of different HLA genetic scores



Supplement 2 Association of different HLA genetic scores with antibody response

Category	First-dose antibody			Second-dose antibody		
	GS 1	GS 2	GS 3	GS 1	GS 2	GS 3
Continuous*	1.14 (1.13 to 1.16)	1.16 (1.14 to 1.18)	1.17 (1.15 to 1.18)	1.10 (1.08 to 1.11)	1.12 (1.10 to 1.14)	1.14 (1.12 to 1.15)
Quintile 1	1 ref	1 ref	1 ref	1 ref	1 ref	1 ref
Quintile 2	1.19 (1.14 to 1.25)	1.17 (1.12 to 1.23)	1.16 (1.11 to 1.22)	1.11 (1.06 to 1.17)	1.13 (1.07 to 1.18)	1.09 (1.04 to 1.14)
Quintile 3	1.26 (1.21 to 1.32)	1.28 (1.22 to 1.34)	1.30 (1.24 to 1.36)	1.14 (1.09 to 1.20)	1.18 (1.12 to 1.24)	1.21 (1.15 to 1.27)
Quintile 4	1.37 (1.30 to 1.43)	1.37 (1.31 to 1.44)	1.38 (1.32 to 1.45)	1.29 (1.23 to 1.36)	1.26 (1.20 to 1.32)	1.26 (1.20 to 1.32)
Quintile 5	1.47 (1.41 to 1.54)	1.52 (1.45 to 1.59)	1.54 (1.47 to 1.61)	1.28 (1.22 to 1.34)	1.38 (1.32 to 1.46)	1.39 (1.32 to 1.46)

Supplement 3 Sensitivity Mendelian randomisation estimation

Exposure	Methods	Number of HLA alleles	Breakthrough infection (effectiveness)	Breakthrough severe COVID-19 (effectiveness)
First-dose antibody	MR Egger	24	11.19 (-5.39 to 25.16)	52.82 (14.14 to 74.08)
	Weighted median	24	14.69 (5.95 to 22.62)	46.90 (25.50 to 62.15)
	Inverse-variance weighted	24	14.51 (8.38 to 20.23)	41.13 (24.64 to 54.00)
	Simple mode	24	17.73 (4.69 to 28.98)	46.10 (11.97 to 67.00)
	Weighted mode	24	15.20 (3.64 to 25.38)	47.39 (20.86 to 65.03)
Second-dose antibody	MR Egger	4	-2.06 (-139.77 to 56.56)	64.88 (-622.50 to 98.29)
	Weighted median	4	20.08 (5.35 to 32.52)	49.51 (10.11 to 71.64)
	Inverse-variance weighted	4	21.34 (9.29 to 31.78)	46.56 (12.09 to 67.51)
	Simple mode	4	20.46 (0.81 to 36.22)	48.91 (-12.29 to 76.75)
	Weighted mode	4	19.58 (2.14 to 33.92)	50.07 (2.28 to 74.49)

Methods

Study design and participants

UK Biobank study

The UK Biobank (UKBB) is a longitudinal population-based cohort of over 500,000 individuals aged 40 to 69 years at the time of recruitment between 2006 and 2010, recruited from England, Scotland, and Wales of the United Kingdom. Its study design and participant characteristics have been previously described in detail elsewhere.[25]

On enrolment, participants responded to socio-demographic, lifestyle, and health-related questionnaires, undertook physical assessments, and provided electronically signed consent. Biological samples of blood, urine, and saliva were collected and stored to facilitate the use of diverse assays, such as genotyping data, as used in this investigation. Participants granted permission to follow up their health outcomes over an extended period through linkage to electronic health records, including data related to the COVID-19 pandemic.

Nested SARS-CoV-2 coronavirus antibody seroprevalence study in UKBB

Alive UKBB participants were re-invited for a SARS-CoV-2 coronavirus antibody study from February 2021 to July 2021, when the UK's vaccination program was being rapidly implemented. At the design stage of the self-test antibody epidemiological study, potential participants of both sexes and all age groups who met the predetermined inclusion and exclusion criteria were eligible for recruitment. These individuals were invited to participate through an email containing a brief overview of the study, links to the information sheet, an instructional video, a list of frequently asked questions, and the online consent form. A concerted effort was made to ensure that as many people as possible participated in this study. For instance, interested individuals were instructed to confirm their contact information and consent to receive a lateral flow self-test kit at their residence. Non-participants were also encouraged to confirm their non-participation via the same website. Upon consenting, participants received an acknowledgement email confirming their participation and providing details on when to expect their antibody testing kit. Participants were notified by email one to three days before kit shipment, and their addresses were securely transferred to a third-party mailing house and shipping company. A reminder email was sent to participants who had not returned a test result one week after their kit was dispatched.

Participant recruitment was done in two phases. The first phase targeted individuals who had previously attended a UK Biobank imaging assessment centre, with approximately ~34,713, ~22,390, and ~21,405 participants sequentially invited. The second phase invited the remaining ~371,985 participants who were not eligible for inclusion in phase 1 (More details on the study design, participant inclusion and exclusion criteria are provided in the online document: https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=998).

Nested COVID-19 infection seroprevalence study in UKBB

As the lateral flow test (LFT) device used in the SARS-CoV-2 antibody seroprevalence study could not distinguish between antibodies produced in response to infection and those generated by vaccination, a follow-on COVID-19 infection seroprevalence study was conducted. Individuals who had previously participated in the self-test antibody study (phase 1 or phase 2) and had reported a "positive" antibody test result were re-

invited to provide a capillary blood sample for laboratory analysis of specific antibodies (i.e., nucleocapsid) that are solely produced in response to COVID-19 infection. The recruitment of participants for the COVID-19 infection study was similar to that of the antibody study and is outlined in detail online (https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=997).

Secondary data linkage

Multiple electronic health records databases have been linked at the individual level to enable following up on the health and disease status of UKBB participants. The pertinent databases used in this study included primary care records (prescriptions and diagnoses), hospital inpatient admissions (diagnoses), cause-specific death registrations, and national infectious diseases surveillance data (COVID-19 test results).[26] Our research team and other investigators have comprehensively scrutinized the data linkage methodology for COVID-19 research, assuring its quality and validity.[27,28]

Analytic cohort curation

We generated a cross-sectional (CS) and prospective (PS) cohort for current analyses. The cross-sectional cohort was composed of participants from the SARS-CoV-2 antibody and infection study, which was stratified according to number of vaccine doses received at the time of testing antibody: the CS-1-dose cohort, consisting of individuals who received one dose of the COVID-19 vaccine, and the CS-2-dose cohort, consisting of those who received two doses. Participants were excluded from the CS-1-dose cohort if they received the first vaccine dose on the day of or more than 84 days (12 weeks) prior to the antibody test. Similarly, participants were excluded from the CS-2-dose cohort if they received the second vaccine dose on the day of or more than 84 days (12 weeks) before the antibody test. Individuals with evidence of prior infection, as confirmed by the presence of nucleocapsid antibody, were excluded from both cohorts. To ensure the robustness and reliability of the identified genetic associations, we partitioned participants into discovery (90%) and validation (10%) subsets in a non-random manner. The discovery subset performed genotype calling using the UK Biobank Axiom array, while the validation subset was processed using the UK BiLEVE Axiom array. This methodological strategy can maximize the independence across both subsets.

The prospective cohort (PS) included all participants who had a record of COVID-19 vaccination in their linked primary care data between December 1, 2020, and September 30, 2021. The follow-up for the PS cohort began on the date of receipt of the first vaccine dose. Participants with a previous positive PCR result for COVID-19 before vaccination were excluded from the cohort.

Genotyping and HLA imputation

Genotyping and initial quality control of the genetic dataset for UKBB participants have previously been documented.[24] Briefly, genotyping of UK Biobank participants was undertaken in 2 phases using two custom-built genome-wide arrays that share 95% of over 820,000 SNP marker content. SNP genotyping was used to impute classical HLA types with four-digit resolution at the MHC class I and the class II regions. The HLA imputation was conducted using a modified HLA*IMP:02 model, which was designed to operate on a multi-population reference panel.[29] Imputation algorithms, tools, and quality control process can be found online (https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/HLA_imputation.pdf).

This imputation model performed reasonably well in the entire UK Biobank sample, with a 4-digit accuracy of HLA alleles ranging from 93.9% for HLA-DRB1 gene to 99.5 for HLA-DPA1 gene among European populations with a posterior probability call threshold of 0.7 (94% of the UKBB individuals self-reported as White). The utility of the HLA imputation was also confirmed in a previous study by replicating signals of known associations between HLA alleles and 11 self-reported immune-mediated diseases.[30]

We grouped people with any of six specific HLA alleles (HLA-DQB1*06:01, HLA-DQB1*06:02, HLA-DQB1*06:03, HLA-DQB1*06:04, HLA-DQB1*06:05, and HLA-DQB1*06:09) as the carrier of HLA-DQB1*06 alleles subtype. Finally, we examined 203 alleles at the four-digit resolution with frequency >= 1/1000 across 11 HLA genes, including HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5, HLA-DQA1, HLA-DQB1, HLA-DPA1, and HLA-DPB1.

Ascertainment of antibody positive and COVID-19 outcomes

The presence of detectable antibodies against SARS-COV-2 was defined as antibody positive. The LFT testing assays shown a good clinical performance in detecting IgG antibodies, with a sensitivity of 98.4% and 98.0% and a specificity of 99.8% and 99.5% for the Fortress Fast COVID-19 Device and AbC-19TM Rapid Test, respectively (refer to the online documents for further technical specification: https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=4513, https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=4520).

Breakthrough infection was defined through a combination of data sources, including a positive result on PCR testing, hospital admission with a COVID-19 related diagnosis (as indicated by the ICD-10 codes U07.1 and U07.2), or a death certificate in which COVID-19 was listed as the cause of death (using the same ICD-10 code). Severe cases were defined as those for whom breakthrough infection required hospitalization or resulted in death.

Statistical approach

Characterisation of vaccinated cohorts

We evaluated characteristics of HLA-DQB1*06 carriers and HLA-DQB1*06 non-carriers at baseline, which was when the antibody test was performed for the cross-sectional cohort and when prospective cohort participants received their first dose of COVID-19 vaccine. Factors that were potentially confounders for the genetic analysis were specified based on prior knowledge, including age, sex, vaccine types (ChAdOx1 and BNT162b2), and time since the latest vaccine dose by week.

Genome-wide association analysis

The GWAS of antibody response following COVID-19 vaccination was performed using REGENIE, controlling for population stratification, relatedness and case-control imbalance, and adjusted for baseline age (at the date of antibody testing), sex, genetic batch, and first ten genetic principal components (PCs). We retain only high-quality variants satisfying the following criteria (1) missing call rates $\leqslant$ 1%; (2) minor allele frequency $\geqslant$1%; (3) minor allele count $\geqslant$ 20; and (4) Hardy-Weinberg equilibrium $\geqslant$ $1*10^{-15}$. To further minimize the influence of population stratification, our analysis was restricted to participants to those of the Caucasian ethnic group.

HLA gene-based association analysis

We examined the proportion of antibody positivity among different HLA genotype groups, stratified by time. To quantify the association between HLA alleles and antibody response in the CS cohort, we used a multivariable logistic regression model with each allele of interest as the independent variable (number of copy: 0, 1, 2) and the presence of antibodies (positive or negative) as the dependent variable. The model for the primary analysis was adjusted for age, sex, ethnicity, genotyping arrays, and the first ten genetic PCs derived from the entire UK Biobank population. The significance of HLA alleles in the logistic regression model was tested using the Wald test while accounting for multiple comparisons with FDR correction. Associations between HLA alleles and vaccine antibody response were estimated separately in the CS-1-dose and CS-2-dose cohorts. Only alleles with FDR-correct *P*-value below 0.05 in both discovery and validation dataset were considered as significant and subsequently analyzed.

To investigate the associations between HLA alleles affecting antibody response and the risk of breakthrough infection and severe COVID-19 in real-world settings, we conducted a survival analysis in the PS cohort using the Cox proportional hazard model. Potential confounders were adjusted including age, sex, ethnicity, genotyping arrays, and 10 PCs. In our primary analysis, we started follow-up from the first vaccination date until death, outcome of interest or the end of the study (30th November 2021 before the Omicron outbreak in the UK), whichever occur first.

Genetic score

We constructed genetic score to measure the combined effect of all HLA allelic variations on the antibody positivity phenotype. The GS was calculated by aggregating the product of the effect size for an allele and the quantity of allele copies across all HLA genes. Effect size estimates for each allele were coefficients (Beta) derived from the logistic regression model and subsequently modified using three approaches: (GS1) forcing the effect size estimate to zero for all alleles with an FDR-correct *P*-value below 0.05, (GS2) forcing the effect size estimate to zero for all alleles with an uncorrected *P*-value below 0.05, and (GS3) directly applying the original estimate regardless of the statistical significance. This process of shrinking some coefficients to zero is equivalent to exclude corresponding alleles from the GS calculation.

Drug/vaccine-target Mendelian randomisation analysis

We used a Mendelian randomization analysis to study the causal effect between the exposure of antibody positivity and COVID-19 outcomes. We selected HLA alleles that were statistically significant with antibody response and used them as genetic instrumental variable. The effect size of the instrumental variables with exposures (antibody positivity following the initial vaccine dose and the subsequent dose) and with outcomes (incident breakthrough outcomes) were obtained from our analyses of CS and PS cohorts, respectively. In the primary analysis, we only used independent alleles as instrumental variables. For the sensitivity analysis, all significant alleles were used regardless of their intercorrelation. The analysis was performed using the R packages TwoSampleMR.

Acknowledgements

Mr Xie is funded through Jardine-Oxford Graduate Scholarship and a titular Oxford Clarendon Fund Scholarship. Professor Prieto-Alhambra 's research group has received funding form the European Medicines Agency and Innovative Medicines Initiative. The research was partially supported by the Oxford National Institute for Health and Care Research (NIHR) Biomedical Research Centre. DPA is funded through a NIHR Senior Research Fellowship (Grant number SRF-2018-11-ST2-004). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the department of Health.

The authors acknowledge English-language editing by Jennifer A. de Beyer, DPhil (Centre for Statistics in Medicine, University of Oxford), and constructive insights from Dr. Hongxiang Zeng and Professor Qizhou Lian on improving the study. The authors express sincere gratitude to all individuals who generously participated in the UK Biobank study, providing an invaluable resource to advance scientific research. The authors also extend appreciation to the UK Biobank management team for their dedication and administrative efforts. This study was conducted under the auspices of Project 65397.

Author contributions

Conceptualization (Junqing Xie, Chunxiao Li, Xiaoying Zheng, Daniel Prieto Alhambra); data curation (Junqing Xie); statistical analysis (Junqing Xie, Marta Alcalde Herraiz, Yaqing Gao, Shuo Feng, Shenda Hong, Yeda Wu); investigation (Chunxiao Li, Shuo Feng, Jia Wei, Zhuoyao Chen, and Binbin Su); supervision (Annika Jodicke, Xiaoying Zheng, Raghib Ali, Nick Wareham, Daniel Prieto Alhambra); interpretation of data (Junqing Xie, Beatriz Mothe, Yeda Wu, Cohet Catherine, Daniel Prieto Alhambra); drafting of the manuscript (Junqing Xie, Chunxiao Li, Yaqing Gao, Shuo Feng, Jia Wei); and critical revision of the manuscript (Beatriz Mothe, Yu Xu, Cohet Catherine, Daniel Prieto Alhambra). All authors reviewed and approved the final version.

Conflict of Interest statement:

The views expressed in this article are the personal views of the author(s) and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties. DPA's department has received grant/s from Amgen, Chiesi-Taylor, Lilly, Janssen, Novartis, and UCB Biopharma. His research group has received consultancy fees from Astra Zeneca and UCB Biopharma. Amgen, Astellas, Janssen, Synapse Management Partners and UCB Biopharma have funded or supported training programmes organised by DPA's department.

Reference

1. Mathieu E, Ritchie H, Rodés-Guirao L, et al. Coronavirus Pandemic (COVID-19). *Our World Data*. Published online March 5, 2020. Accessed June 14, 2023. https://ourworldindata.org/covid-vaccinations

2. Poland GA, Ovsyannikova IG, Kennedy RB. Personalized vaccinology: A review. *Vaccine*. 2018;36(36):5350-5357. doi:10.1016/j.vaccine.2017.07.062

3. Poland GA, Ovsyannikova IG, Jacobson RM. Personalized vaccines: the emerging field of vaccinomics. *Expert Opin Biol Ther*. 2008;8(11):1659-1667. doi:10.1517/14712598.8.11.1659

4. Falahi S, Kenarkoohi A. Host factors and vaccine efficacy: Implications for COVID-19 vaccines. *J Med Virol*. 2022;94(4):1330-1335. doi:10.1002/jmv.27485

5. Antonelli M, Penfold RS, Merino J, et al. Risk factors and disease profile of post-vaccination SARS-CoV-2 infection in UK users of the COVID Symptom Study app: a prospective, community-based, nested, case-control study. *Lancet Infect Dis*. 2022;22(1):43-55. doi:10.1016/S1473-3099(21)00460-6

6. Lee ARYB, Wong SY, Chai LYA, et al. Efficacy of covid-19 vaccines in immunocompromised patients: systematic review and meta-analysis. *BMJ*. 2022;376:e068632. doi:10.1136/bmj-2021-068632

7. Pulendran B. Immunology taught by vaccines. *Science*. 2019;366(6469):1074-1075. doi:10.1126/science.aau6975

8. Pulendran B, Davis MM. The science and medicine of human immunology. *Science*. 2020;369(6511). doi:10.1126/SCIENCE.AAY4014

9. Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. *Nat Rev Immunol*. 2018;18(5):325-339. doi:10.1038/nri.2017.143

10. Smatti MK, Alkhatib HA, Al Thani AA, Yassine HM. Will Host Genetics Affect the Response to SARS-CoV-2 Vaccines? Historical Precedents. *Front Med*. 2022;9:802312. doi:10.3389/fmed.2022.802312

11. Mentzer AJ, O'Connor D, Bibi S, et al. Human leukocyte antigen alleles associate with COVID-19 vaccine immunogenicity and risk of breakthrough infection. *Nat Med*. 2023;29(1):147-157. doi:10.1038/s41591-022-02078-6

12. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet*. 2018;19(8):491-504. doi:10.1038/s41576-018-0016-z

13. Wang QS, Huang H. Methods for statistical fine-mapping and their applications to auto-immune diseases. *Semin Immunopathol*. 2022;44(1):101-113. doi:10.1007/s00281-021-00902-8

14. Pairo-Castineira E, Rawlik K, Bretherick AD, et al. GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19. *Nature*. 2023;617(7962):764-768. doi:10.1038/s41586-023-06034-3

15. Gutiérrez-Bautista JF, Sampedro A, Gómez-Vicente E, et al. HLA Class II Polymorphism and Humoral Immunity Induced by the SARS-CoV-2 mRNA-1273 Vaccine. *Vaccines*. 2022;10(3):402. doi:10.3390/vaccines10030402

16. Khor SS, Omae Y, Takeuchi JS, et al. An Association Study of HLA with the Kinetics of SARS-CoV-2 Spike Specific IgG Antibody Responses to BNT162b2 mRNA Vaccine. *Vaccines*. 2022;10(4):563. doi:10.3390/vaccines10040563

17. Ragone C, Meola S, Fiorillo PC, et al. HLA Does Not Impact on Short-Medium-Term Antibody Response to Preventive Anti-SARS-Cov-2 Vaccine. *Front Immunol*. 2021;12:734689. doi:10.3389/fimmu.2021.734689

18. Astbury S, Reynolds CJ, Butler DK, et al. HLA-DR polymorphism in SARS-CoV-2 infection and susceptibility to symptomatic COVID-19. *Immunology*. 2022;166(1):68-77. doi:10.1111/imm.13450

19. Bolze A, Neveux I, Schiabor Barrett KM, et al. HLA-A∗03:01 is associated with increased risk of fever, chills, and stronger side effects from Pfizer-BioNTech COVID-19 vaccination. *Hum Genet Genomics Adv*. 2022;3(2):100084. doi:10.1016/j.xhgg.2021.100084

20. Bauernfeind S, Salzberger B, Hitzenbichler F, et al. Association between Reactogenicity and Immunogenicity after Vaccination with BNT162b2. *Vaccines*. 2021;9(10):1089. doi:10.3390/vaccines9101089

21. Sadarangani M, Marchant A, Kollmann TR. Immunological mechanisms of vaccine-induced protection against COVID-19 in humans. *Nat Rev Immunol*. 2021;21(8):475-484. doi:10.1038/s41577-021-00578-z

22. Aw D, Silva AB, Palmer DB. Immunosenescence: emerging challenges for an ageing population. *Immunology*. 2007;120(4):435-446. doi:10.1111/j.1365-2567.2007.02555.x

23. Cromer D, Steain M, Reynaldi A, et al. Predicting vaccine effectiveness against severe COVID-19 over time and against variants: a meta-analysis. *Nat Commun*. 2023;14(1):1633. doi:10.1038/s41467-023-37176-7

24. Horowitz JE, Kosmicki JA, Damask A, et al. Genome-wide analysis provides genetic evidence that ACE2 influences COVID-19 risk and yields risk scores associated with severe disease. *Nat Genet*. 2022;54(4):382-392. doi:10.1038/s41588-021-01006-7

25. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779

26. Armstrong J, Rudkin JK, Allen N, et al. Dynamic linkage of COVID-19 test results between Public Health England's Second Generation Surveillance System and UK Biobank. *Microb Genomics*. 2020;6(7):mgen000397. doi:10.1099/mgen.0.000397

27. Xie J, Feng S, Li X, Gea-Mallorquí E, Prats-Uribe A, Prieto-Alhambra D. Comparative effectiveness of the BNT162b2 and ChAdOx1 vaccines against Covid-19 in people over 50. *Nat Commun*. 2022;13(1):1519. doi:10.1038/s41467-022-29159-x

28. Xie J, Prats-Uribe A, Feng Q, et al. Clinical and Genetic Risk Factors for Acute Incident Venous Thromboembolism in Ambulatory Patients With COVID-19. *JAMA Intern Med*. 2022;182(10):1063-1070. doi:10.1001/jamainternmed.2022.3858

29. Dilthey A, Leslie S, Moutsianas L, et al. Multi-Population Classical HLA Type Imputation. *PLOS Comput Biol*. 2013;9(2):e1002877. doi:10.1371/journal.pcbi.1002877

30. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z