nature medicine

https://doi.org/10.1038/s41591-022-02078-6

Accelerated Article Preview

Human leukocyte antigen alleles associate with COVID-19 vaccine immunogenicity and risk of breakthrough infection

Received: 24 December 2021

Accepted: 7 October 2022

Accelerated Article Preview
Published online: 13 October 2022

Cite this article as: Mentzer, A. J. et al. Human leukocyte antigen alleles associate with COVID-19 vaccine immunogenicity and risk of breakthrough infection. *Nature Medicine* https://doi.org/10.1038/s41591-022-02078-6 (2022).

Alexander J. Mentzer, Daniel O'Connor, Sagida Bibi, Irina Chelysheva, Elizabeth A. Clutterbuck, Tesfaye Demissie, Tanya Dinesh, Nick J. Edwards, Sally Felle, Shuo Feng, Amy L. Flaxman, Eleanor Karp-Tatham, Grace Li, Xinxue Liu, Natalie Marchevsky, Leila Godfrey, Rebecca Makinson, Maireid B. Bull, Jamie Fowler, Bana Alamad, Tomas Malinauskas, Amanda Y. Chong, Katherine Sanders, Robert H. Shaw, Merryn Voysey, Matthew D. Snape, Andrew J. Pollard, Teresa Lambe & Julian C. Knight

Human leukocyte antigen alleles associate with COVID-19 vaccine immunogenicity and risk of breakthrough infection

Alexander J Mentzer[1]*, Daniel O'Connor[2,3]*, Sagida Bibi[2,3], Irina Chelysheva[2,3], Elizabeth A Clutterbuck[2,3], Tesfaye Demissie[2,3], Tanya Dinesh[2,3], Nick J Edwards[4], Sally Felle[2,3], Shuo Feng[2,3], Amy L Flaxman[2,3], Eleanor Karp-Tatham[1], Grace Li[2,3], Xinxue Liu[2,3], Natalie Marchevsky[2,3], Leila Godfrey[2,3], Rebecca Makinson[4,], Maireid B Bull[2,5], Jamie Fowler[4], Bana Alamad[1], Tomas Malinauskas[6], Amanda Y Chong[1], Katherine Sanders[2,3], Robert H Shaw[2,3], Merryn Voysey[2,3], Oxford COVID Vaccine Trial Genetics Study Team Group†, Matthew D Snape[2,3], Andrew J Pollard[2,3]*, Teresa Lambe[2,3,5]*, Julian C Knight[1,3,5]*

*Contributed equally

†Members are listed in Appendix 1

[1]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK; [2]Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford, UK; [3]NIHR Oxford Biomedical Research Centre and Oxford University Hospitals NHS Foundation Trust, Oxford, UK; [4]The Jenner Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK; [5]Chinese Academy of Medical Science (CAMS) Oxford Institute, University of Oxford, Oxford, UK; [6]Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK.

Abstract

SARS-CoV-2 vaccine immunogenicity varies between individuals, and immune responses correlate with vaccine efficacy. Using data from 1,076 participants enrolled in ChAdOx1 nCov-19 vaccine efficacy trials in the United Kingdom, we find that inter-individual variation in normalised antibody responses against SARS-CoV-2 spike (S) and its receptor binding domain (RBD) at 28 days following first vaccination shows genome-wide significant association with major histocompatibility complex (MHC) class II alleles. The most statistically significant association with higher levels of anti-RBD antibody was HLA-DQB1*06 (P=3.2 x 10^{-9}), which we replicate in 1,677 additional vaccinees. Individuals carrying HLA-DQB1*06 alleles were less likely to experience PCR-confirmed breakthrough infection during the ancestral SARS-CoV-2 virus and subsequent Alpha-variant waves compared with non-carriers (HR 0.63, 0.42-0.93, P=0.02). We identify a distinct S-derived peptide that is predicted to bind differentially to HLA-DQB1*06 compared with other similar alleles, and find evidence of increased spike-specific memory B-cell responses in HLA-DQB1*06 carriers at 84 days following first vaccination. Our results demonstrate association of HLA type with COVID-19 vaccine antibody response and risk of breakthrough infection, with implications for future vaccine design and implementation.

Introduction

Since its emergence in late 2019, SARS-CoV-2 has caused a global pandemic with estimates of between 6.5-15 million deaths up to September 2022[1,2]. Vaccines targeting, predominantly, the spike (S) antigen of SARS-CoV-2 have demonstrated high efficacy against severe disease in phase 3 trials, eliciting high levels of binding and neutralising antibodies, as well as T cell responses, with over 12 billion doses administered worldwide[1]. Two of the earliest developed vaccines, BNT162b2

(Pfizer-BioNTech)[3] and ChAdOx1 nCoV-19 (AZD1222; Oxford-AstraZeneca)[4] are estimated to have population effectiveness against a positive PCR test for the earliest variants of SARS-CoV-2 of 79% and 80% respectively when assessed at least 21 days following the second dose of vaccination in a community-based household survey from the United Kingdom (1 December 2020 to 8 May 2021)[5], together with 88-91% effectiveness against hospital admission for coronavirus disease (COVID-19)[5], although lower effectiveness is reported with more recent variants of concern[6]. Despite the success of vaccines at reducing mortality and morbidity in the population with effectiveness against severe disease and hospitalisation currently remaining high, vaccine breakthrough infections, while predominantly mild, are increasingly reported[7–9] .

Significant variation in immune responses, including antibody levels and T cell responses, has been reported among vaccinated individuals[10]. Neutralising antibody levels show association with vaccine efficacy in animal challenge studies[11] and humans[12,13], and risk of symptomatic COVID-19 has been shown to reduce with increasing levels of both anti-spike (anti-S IgG) and antibodies against receptor binding domain (RBD) antigenic sites on the viral spike (anti-RBD IgG) following vaccination with ChAdOx1 nCoV-19[13]. The reasons for inter-individual variation in total or neutralising antibody responses are incompletely understood[10,14]. Community-based surveys have provided some epidemiological insight into this question among individuals with no prior history of SARS-CoV-2 infection in the UK general population: a low anti-spike IgG antibody responder group following vaccination was identified and found to be more commonly male, elderly (over 75 years of age) and with long term health conditions[10].

We sought to investigate the contribution of genetic factors to the observed variation in response to vaccination with ChAdOx1 nCoV-19. Antibody responses following vaccination show evidence of heritability[15] with genetic variation in HLA within the Major Histocompatibility Complex (MHC) on chromosome 6 (position p21.3) associated with responses to hepatitis B[16–19], tetanus[20], and measles[21,22] vaccines. For these infections, the relevance for vaccine failure has not been robustly

demonstrated[23,24]. To date, genetic studies in COVID-19 have focused on risk of severe disease with replicated associations implicating antiviral defence mechanisms (notably involving interferon signalling), mediators of inflammatory organ damage, leucocyte differentiation and blood type antigen secretor status, but limited evidence to date for HLA[25–27]. Here we use data from five clinical trials of ChAdOx1 nCoV-19 to demonstrate association of HLA-DQB1*06 with higher antibody responses against the RBD of S antigen and lower risk of breakthrough infections, which we propose involves altered HLA-peptide binding influencing memory B-cell responses.

Results

Genome-wide association study of antibody responses 28 days after ChAdOx1 nCoV-19 vaccination

We hypothesised that genetic factors contribute to inter-individual variation in COVID-19 vaccine responses. To investigate this, we first performed a discovery analysis testing for genetic association with vaccine responses in participants enrolled in the phase 1/2 (COV001) and phase 2/3 (COV002) randomized single-blind ChAdOx1 nCov-19 (AZD1222) vaccine efficacy trials, conducted within the United Kingdom, and in whom humoral immune responses were measured post-vaccination. **Fig. 1** summarises participant inclusion. DNA from 1,222 ChAdOx1 nCov-19 trial participants was genotyped on the Affymetrix AxiomTM HGCoV2 1 array. After quality control (Methods), 667,496 variants in 1,190 individuals were available for single nucleotide polymorphism variant (SNP) imputation (**Supplementary Fig. 1**). Following imputation, 9,325,058 high quality SNPs were tested for association with normalised antibody responses against S and RBD (**Extended Data Fig. 1**) in 1,076 of the 1190 genotyped individuals who had received ChAdOx1 nCoV-19 vaccine, with antibody measures available at 28 days following first vaccination (baseline demographics are shown in **Table 1**). We performed the association analysis adjusting for age, sex, prior SARS-CoV-2 exposure based on anti-nucleocapsid (anti-N) IgG concentrations (N=128, 11.9%), and antibody assay type (all as fixed-effect covariates), for all individuals irrespective of ancestry (**Extended Data Fig. 2**) including a genetic relatedness matrix as a random effect covariate. The mixed model regression analysis

revealed genome-wide significant associations (P<5x10^{-8}) for both anti-S (index variant rs9271374, P=2.6x10^{-8}, beta -0.14, SE 0.03) and anti-RBD (rs1130456, P=4.4x10^{-10}, beta -0.26, SE 0.04) IgG antibody levels. rs9271374 and rs1130456 are SNPs located within 10 kilobases (kb) of *HLA-DQ* genes (**Fig. 2**) and in linkage disequilibrium (LD) within our multi-ancestry cohort (r^2=0.65). The distribution of *P*-values (**Extended Data Fig. 3A**) and beta coefficients (**Extended Data Fig. 3B**) for all genotyped and imputed variants across this locus show a clear correlation in genetic architecture between these two antibody responses (Spearman's rho coefficient 0.90 and 0.93 for *P*-values and beta coefficients respectively) correlated through LD (measured through r^2).

These genetic association signals may be falsely observed as a result of two important factors. Firstly, although every effort was made to normalise the antibody levels whilst acknowledging the different platforms for antibody measurement, the final distributions still deviated from normality, which could increase the risk of detecting a genetic association signal by chance (**Extended Data Fig. 1**). Therefore, we performed a further round of inverse normal transformation on the pooled RBD-specific antibody distributions (**Extended Data Fig. 1F**) to create a merged, normalised distribution (**Extended Data Fig. 4A**) and reran the association analysis for anti-RBD antibodies. The rs1130456 association was still present and the most significant association (P=4.7x10^{-9}, **Extended Data Fig. 4B-D**). Secondly, given the diverse ancestry of individuals included, it is possible that the association could be a result of confounding due to population structure. The genomic inflation factor (λ) from our primary RBD association analysis was 1.023. When the extended MHC region was excluded, λ was 1.007 suggesting much of the observed inflation was a result of the large number of variants associated with MHC (**Extended Data Fig. 5A** for Manhattan and **5B** for QQ). Furthermore, given the low levels of natural exposure to SARS-CoV-2 in our population at the time of sampling early in the pandemic, we used our data for anti-N IgG concentrations to again test for excessive inflation using N as a negative control. No associations of genome-wide significance were observed and λ was estimated at 1.017 (**Extended Data Fig. 5C-D**). To further explore the effect of population structure we re-ran our association analyses for RBD including the first ten genetic principal components (PCs)

(derived from the entire genotyped dataset) as additional fixed effect covariates using our mixed-model approach (**Extended Data Fig. 5E-F**). λ using this approach was 0.985, suggesting a degree of overfitting, but again the same variant (rs1130456) remained most significantly associated with RBD-specific antibodies, albeit with a marginally attenuated *P*-value ($1.3x10^{-9}$), as to be expected when including multiple additional covariates in the model.

HLA imputation and fine-mapping of associated variants

We proceeded to test for evidence of association with S- and RBD-specific IgG antibodies at the level of HLA gene and protein variation. Imputation (see Methods) identified 640 HLA alleles and 4513 amino acid (AA) changes (of which 81 alleles and 3027 AA substitutions were present in our dataset at a minor allele frequency of ≥ 0.05). To undertake fine-mapping we identified 1,023 individuals with identity-by-descent (IBD) values of 0.185 or less, and those of self-reported, and PC analysis (PCA)-derived European ancestry (using PC1 and PC2 cut-offs as shown in **Extended Data Fig. 2** inset). Of all HLA and AA alleles tested for association with S and RBD antibody levels, the HLA allele with the most significant association was HLA-DQB1*06 with anti-RBD antibodies ($P=3.2 \times 10^{-9}$, beta 0.27, SE 0.04; **Supplementary Table 1**). An AA variant had a *P*-value identical to that of HLA-DQB1*06 (3.2×10^{-9}), but the exact inverse beta coefficient (-0.27). This AA variant (DQB1-125A/S) denotes the presence of either an alanine or serine at position 125 of the HLA-DQB1 protein according to international ImMunoGeneTics project (IMGT) coordinates. HLA-DQB1*06 has a glycine at position 125 whereas other alleles common in our genotyped population possess either alanine (HLA-DQB1*02 and *04 alleles) or serine (HLA-DQB1*05). Thus, this AA variant is synonymous with the presence of HLA-DQB1*06 in our dataset. The index-associated variant from the primary analysis (rs1130456) was equally associated with the anti-RBD titres in this analysis (beta 0.27, se 0.04). Other variants imputed using the specific HLA imputation algorithm were identified as being marginally more significantly associated than rs1130456, with the new lead being rs9273817 ($P=2.4x10^{-9}$, beta=0.27, SE 0.04).

158 To further understand the relationship between the top associated variants we performed stepwise
159 forward regression analysis incorporating the full set of SNP, AA and HLA allele variants from HLA
160 imputation in the set of individuals restricted by IBD, PCs and self-reported ethnicity (**Fig. 3A and B**).
161 Adjusting for the new top SNP (rs9273817) in the first round of conditional analysis, there was a
162 complete abolition of the signals for both rs1130456 (P=0.65) and HLA-DQB1*06 (P=0.65),
163 supporting the likelihood that these variants are all tagging the causal variant, most plausibly HLA-
164 DQB1*06 (**Fig. 3A and 3B** middle rows). A second, likely independent, signal of association was
165 observed in HLA-DRB1 with the index variant being the presence of a glutamate or arginine at
166 position 71 (according to IMGT) of HLA-DRB1 (DRB1-71E/R, $P_{conditional}$=2.7x10^{-4}, beta=-0.14, SE=0.04,
167 **Fig. 3A** middle row). After conditioning on this variant no further independent signals with a P<1x10^{-2}
168 were observed in the class II MHC region. Before proceeding with further fine-mapping, we
169 confirmed that both the HLA-DQB1*06 and DRB1-71E/R associations were robust to statistical
170 inflation by performing a Monte Carlo exact test with 10^8 permutations. The likelihood of both
171 associations occurring by chance was still less than the number of permutations (i.e., P<2x10^{-8})
172 limited only by computational time for testing. Furthermore, we tested for evidence of both the
173 HLA-DQB1*06 and DRB1-71E/R associations in the different population strata (including 928
174 European vs 148 non-European) individuals and observed evidence of the same trend of association
175 in both groups supporting further that this association is unlikely to be spurious due to population
176 structure (**Supplementary Fig. 2**).

177 Given our findings from the stepwise conditional analysis we next sought evidence to further
178 substantiate the effect being driven by the DQ locus and the relationship of the two independent
179 associations to each other. Firstly, HLA-DQB1*06 is frequently inherited as part of a common
180 haplotype with HLA-DQA1*01 (with DQA1*01:02 being associated with anti-RBD antibody levels in
181 our discovery dataset with a P=1.3x10^{-8}, beta 0.28, SE 0.05) and HLA-DRB1*15 (DRB1*15:01
182 P=1.8x10^{-8}, beta 0.31, SE 0.05). These three alleles were most significantly associated with anti-spike
183 antibody levels (P=2.0x10^{-7}, 7.8x10^{-8}and 4.6x10^{-8} respectively; **Supplementary Table 2**). We assessed

the accuracy of HLA allele imputation by analysing 60 individuals from the genotyped COV001 and COV002 vaccinees (120 alleles) who also underwent classical HLA typing (**Supplementary Notes**). The agreement between 2-digit calls was 99.2% and 96.7% for HLA-DQB1 and HLA-DRB1 loci respectively (**Supplementary Table 3**). The accuracy of calling the specific HLA-DQB1*06 group of alleles to 4-digit resolution was 97.0%, and 100% for HLA-DRB1*15 alleles (**Supplementary Table 4**).

In order to confirm that the HLA-DQB1*06 allele was the most likely primary gene locus associated with the antibody response, we used Bayesian Information Criterion modelling with phased data. We found variation in RBD antibody levels was best described using HLA-DQ alleles (HLA-DQA1*01/HLA-DQB1*06, BIC 2715.2) rather than HLA-DRB1 (HLA-DRB1*15, BIC 2719.1) supporting that the primary association was likely linked to HLA-DQ, rather than HLA-DR, variation (**Supplementary Fig. 3**). We next investigated whether there was evidence of interaction between the HLA-DQB1*06 and independent DRB1-71E/R associations. Using a likelihood ratio test (LRT) comparing the linear and interaction terms we found no evidence of a complex inter-dependence between these two variants (P=0.44, **Supplementary Fig. 4**). We therefore compared models describing variation in a simple linear additive model (i.e. normalised anti-RBD antibody levels ~ HLA-DQB1*06 + DRB1-71E/R) compared with a model where we compared individuals grouped into the presence of one variant in the absence of the other (**Fig. 3C**) and found that the latter was more parsimonious after adjusting for age, sex, 5 PCs and anti-RBD antibody measurement assay (BIC 2965.42 vs 2689.65 respectively). Thus, using this combined description of variation we next tested for association of HLA-DQB1*06 with increased anti-RBD antibody levels accounting for DRB1-71E/R over the time course of the ChAdOx1 nCoV-19 trial. We found significant differences between the opposing DQB1*06 and DRB1-71E/R carrier groups seen at day of second dose (P=2.7x10^{-7} using the Student's t-test), day 28 post-second dose (P=2.6x10^{-7}) and at day 90 post-second dose (P=0.01, **Fig. 4A**). There was no significant difference observed at day 182 post-second dose. A summary of the baseline demographics of individuals stratified by HLA allele group is provided in **Supplementary Table 5**.

Replication of genetic association signals with COVID-19 vaccination antibody responses

In order to provide further evidence that the observed genetic associations were robust and not restricted to the COV001 and COV002 cohorts, we aimed to test for replication of the associations in a series of independently recruited cohorts (**Extended Data Fig. 6**). Three trials were coordinated after COV001 and COV002 to address questions regarding safety and immunogenicity of heterologous dosing of COVID-19 vaccines (COMCOV, COMCOV2) and to assess immunogenicity and safety in children (COV006). DNA from 1,722 individuals (638 COMCOV, 876 COMCOV2 and 208 COV006) were genotyped on the same Affymetrix AxiomTM HGCoV2 1 array with identical QC and imputation pipelines applied as to the COV001 and COV002 cohorts. After quality control, data from 1,677 (627 COMCOV, 847 COMCOV2 and 203 COV006) individuals were available (**Supplementary Table 6**). These replication cohorts differed substantially from the ChAdOx1 nCoV-19 trials in respect to age, sex proportions, and timing and nature of vaccination regimens and only data for S antibody levels (and not RBD) were available. Moreover, COMCOV was enriched for non-White reported ethnicities. Nevertheless, when using the same HLA-DQB1*06 and DRB1-71E/R classifications, we observed statistical evidence of association with anti-S antibody levels measured at the time of second vaccination in the same direction as observed for COV001 and COV002 when looking both at those individuals receiving ChAdOx1 nCoV-19 or BNT162b2 as their first vaccine. Both unadjusted and adjusted models were used to compare the groups as follows.

When including all individuals irrespective of the first vaccine received (all individuals genotyped from COMCOV, COMCOV2 and COV006 received either ChAdOx1 nCoV-19 or BNT162b2 as their first vaccine), there was a significant difference in anti-S antibody levels measured on the day of second vaccine (median of 67 days with a range of 28-184 days following the first vaccine) when comparing the group “carrying DRB1-71E/R with no DQB1*06” against both those “carrying DQB1*06 with no DRB1-71E/R” (P_t=5.4x10^{-3}; *P*-value derived from the t-test), and the “remainder” of individuals (P_t=0.02, **Supplementary Fig. 5A**). Significant differences in anti-S antibody levels were also observed

for those individuals given a first dose of BNT162b2 comparing the groups "carrying DRB1-71E/R with no DQB1*06" and those "carrying DQB1*06 with no DRB1-71E/R" (P_t=0.01), and the "remainder" of individuals (P_t=0.04, **SupplementaryFig. 5B**), and for ChAdOx1 nCoV-19 with significance only observed for the comparison between "Carrying DRB1-71E/R with no DQB1*06" and "Carrying DQB1*06 with no DRB1-71E/R" groups (P_t=0.04, **Supplementary Fig. 5C**). Using a linear model adjusting for age, sex, interval between vaccination and sampling, first vaccine, study and self-reported ethnicity there was a significant difference in S antibody levels observed for all individuals when testing the effect of HLA-DQB1*06 accounting for DRB1-71E/R (P_{LRT}=5.2x10^{-3} using the LRT to compare with the null), and for those individuals primed with BNT162b2 (P_{LRT}=0.02), with a trend towards significance for individuals primed with ChAdOx1 nCoV-19 (P_{LRT}=0.1). Similarly, we observed a significant effect in the same direction when performing the same analysis comparing individuals grouped by either HLA-DQB1*06 carriage alone (P_{LRT}=0.04) or DRB1-71E/R carriage alone (P_{LRT}=7.8x10^{-4}). All together, we have found further evidence that carriage of HLA-DQB1*06 and the presence of either a glutamate or arginine at position 71 in the HLA-DRB1 protein is associated with differential antibody responses to S across individuals, irrespective of the nature of first vaccine, and across ages and both self-reported White and non-White ethnicities. Given the correlation in antibody levels against S and RBD in our discovery set, it is likely that our observed genetic signal would also be observed for RBD even though RND was not measured directly in the replication cohorts.

Testing for association of genetic variants on immune response over time and risk of breakthrough infection

Given the observed association between HLA variants with variable immunogenicity, we next investigated whether there was a relationship with risk of breakthrough infection. At a median of 494 days (interquartile range (IQR) 479-535) of follow-up from the first dose of vaccine in participants from the trials used in the discovery analysis (COV001 and COV002), 112 episodes of breakthrough infection (individuals with a SARS-CoV-2 nucleic acid amplification test (NAAT)-positive

swab at least 22 days after a first dose of vaccine, with the prevalent ancestral virus and Alpha variants) had been recorded in genotyped individuals who had received the ChAdOx1 nCoV-19 vaccine (**Table 1** and **Supplementary Table 7**). We found that HLA-DQB1*06 was present in 33.9% of individuals experiencing breakthrough infection compared with 45.6% of those who did not have breakthrough infection (X^2 P=0.02). We subsequently found that individuals carrying HLA-DQB1*06 alleles were less likely to experience breakthrough infection over time compared with those who did not carry HLA-DQB1*06 after adjustment for age, sex, the first two genetic PCs (representing ethnicity), healthcare worker status, BMI and chronic disease status (adjusted HR 0.63, 0.42-0.93, P=0.02, **Fig. 4B** and **Extended Data Fig. 7**). We performed sample re-weighting for dose and interval between first and second dose of vaccination (using inverse probability weighting) to ensure our analyses were as consistent with prior correlates analyses as possible[13]. This significance persisted even after adjusting for whether individuals were likely to have been naturally exposed to SARS-CoV2 (determined using N measurements) and based on whether they were related to each other (IBD<0.185) or not (P=0.02). A similar effect was observed when describing individuals using our overall HLA status definition (i.e., carrying HLA-DQB1*06 alleles accounting for DRB1-71E/R), although significance was lost (adjusted HR for the group "carrying DRB1*06 with no DRB1-71E/R" 0.54, 0.26-1.1, P=0.09; **Fig. 4C** and **Extended Data Fig. 8**). The lower frequency of HLA-DQB1*06 in individuals experiencing breakthrough infection was observed both in the 41 individuals meeting the definition of primary symptomatic breakthrough infection (31.7% carrying HLA-DQB1*06 amongst those with the primary definition of breakthrough infection), and the 66 who were asymptomatic (28.8%), but not in the 9 individuals who did not meet the primary definition of symptomatic breakthrough (66.7%) (**Supplementary Table 7**). To further substantiate our finding, we explored whether we could find any evidence of an equivalent effect in the tested replication cohorts, acknowledging that the cohorts differed significantly from ChAdOx1 nCoV-19 not only in regards to age, ethnicity, comorbidities and the predominant circulating SARS-CoV-2 variant (Alpha and Delta), but also because breakthrough infection was defined through self-report rather than active

surveillance. In the subset of individuals of self-reported White ethnicity and less than or equal to 55 years of age (thus enriching for individuals more representative of the COV001 and COV002 trial , n=401), there were 29 individuals experiencing breakthrough and 372 individuals with no breakthrough reported after a median of 280 days of follow-up (IQR 244-332). HLA-DQB1*06 was observed in 37.9% of individuals experiencing breakthrough infection and 40.3% in individuals with no breakthrough infection. After adjustment for age, sex, first vaccine received (ChAdOx1 nCoV-19 or BNT162b2), and booster received (viral vector (ChAdOx1 nCoV-19), mRNA (BNT162b2 or mRNA-1273) or nanoparticle (NVXCoV2373)) with reweighting calculated based on days between first and second doses of vaccine, carriage of HLA-DQB1*06 had an adjusted HR of 0.87 (0.41-1.80, P=0.73) of risk of breakthrough infection (**Extended Data Fig. 9**).

Structural insights into HLA-Spike peptide binding

Given the observed immunological and clinical impact of HLA-DQB1*06 on vaccine response and effectiveness, we next tested for structural evidence of binding of S peptides by the associated HLA-DQB1*06 allele. We tested the hypothesis that HLA-DQB1*06:02 could bind peptides from SARS-CoV-2 Spike more effectively than an alternative HLA-DQB1 allele that was both common in the population and linked with another HLA-DQA1*01 allele. Using the COV001/COV002 data we identified HLA-DQB1*05:01 as an allele that would act as a suitable comparator (frequency in COV001/COV002 12%, beta for association with 28-day RBD levels -0.14, se 0.05, *P*=0.01, acknowledged to commonly pair with HLA-DQA1*01:01[28]). Other HLA alleles that were common in our merged COV001/COV002 dataset were less suitable for this analysis. DQB1*03, for example, (frequency 34%) pairs more frequently with DQA1*03 or *05 alleles, whereas DQB1*02 (23%) pairs with DQA1*02 or *05). The only available crystal structure for DQB1*06 alleles is the HLA-DQA1*01:02–HLA-DQB1*06:02 in complex with a hypocretin peptide (LPSTKVSWAAV)[29]. The key side chains of Ser (position (P) 3), Thr (P4) and Val (P6) of the peptide are buried in the centre of the groove formed by two HLA molecules (**Extended Data Fig. 10A**). Thus, we searched for a hypocretin-like peptide motif (Ser/Cys) ThrXVal in Spike protein (where X is any amino acid with its side chain

pointing away from the groove; Ser and Cys differ in one atom only). Spike residues Val615-NCTEVPVAI-His625 could be aligned with a hypocretin peptide and thus enabled us to model a complex of HLA-DQA1*01:02–HLA-DQB1*06:02 bound to the Spike peptide using AlphaFold[30] (**Fig. 5A**). Both the AlphaFold-based model and the crystal structure support DQB1*06:02 interacting differently with any peptide compared with DQB1*05:01. DQB1*05:01 differs from DQB1*06:02 by at least three key residues forming hydrogen bonds with the bound hypocretin peptide (**Extended Data Fig10B-D**), making analogous DQB1*05:01–peptide interactions impossible. Our analysis identifies specific residues of DQB1*05:01 and DQB1*06:02 responsible for different peptide recognition and subsequent recognition by T-cell receptors.

Immunological implications of the observed MHC associations

To further support these observations, we used peripheral blood mononuclear cells (PBMCs) available from a small number of participants from COV001 and COV002 to compare anti-S specific memory B-cell responses at day 84 following the first vaccine in 10 individuals carrying HLA-DQB1*06, and 10 not carrying HLA-DQB1*06. We observed an increase in anti-S memory B-cell responses in the individuals carrying HLA-DQB1*06 (P=0.05 using a one-tailed Wilcoxon-rank test) at Day 84 that was not apparent at Day 0 (**Fig. 5B**). We then searched for similar signals of association in the intermediate components of the MHC-T-B-antibody axis. We observed a difference in overall CD4 proliferation in response to stimulation with S1 (that includes the RBD domain and the putative Val615-NCTEVPVAI-His625 peptide, P=0.01 **Fig. 5C**), but not against S2 (cleaved away before residue 686). We did not observe an equivalent signal with antigen specific T-cell activation (using the antigen inducible marker (AIM) assay, **Fig. 5D**).

Discussion

Our findings show that individuals carrying HLA-DQB1*06 alleles have higher antibody responses against SARS-CoV-2 spike protein and the RBD following vaccination with both ChAdOx1 nCoV-19 and BNT162b2 vaccines than individuals not carrying this allele. HLA-DQB1*06 is also associated with a reduced risk of breakthrough infection based on PCR positivity after a median 494 days of follow

up after receiving their first dose of vaccine. To our knowledge, this is the first report of an HLA association with antibody responses following immunisation with SARS-COV-2 vaccines and of a genetic association with risk of SARS-CoV-2 breakthrough infection[15–18,22–24,31]. We further provide a working mechanistic hypothesis for the primary HLA-DQB1*06 association of potentially distinct peptide binding that may lead to improved CD4+ T-cell proliferation and memory B-cell activation. Our study design comprised of an infection- and vaccine-naïve population in a clinical trial setting with appropriate blinding, detailed immune phenotyping and patient follow up at defined timepoints, and we further substantiated our findings in a large replication dataset with preliminary follow-up functional experiments.

The global evidence of breakthrough infections following vaccination, of changes in immune correlates of protection over time, and of new SARS-CoV-2 variants highlight the importance of subsequent dosing of vaccination and understanding how this can be optimally deployed.[9,13,32] Our study demonstrates that there is a heritable component to observed inter-individual variation in antibody response at day 28 post-first dose but also throughout time post-vaccination, across vaccinee age and type of first vaccine. Given that these effects do persist over time, with some change in effect size, and have clinical relevance in terms of risk of breakthrough infection, the observed HLA associations raise the potential utility of prioritising at-risk populations based, for example, on HLA-DQB1*06 allele frequencies, among whom more intensive booster vaccination may be warranted. Variable HLA-DQB1*06 allele frequencies are reported across diverse populations[33,34] but there is not yet robust epidemiological evidence of the extent of breakthrough infections in such populations and this would require further investigation before implementation of such an intervention. The observed reduction in effect size on log10 transformed RBD-specific antibody levels in the ChAdOx1 nCoV-19 vaccinees over time from 0.38 at day 28, to 0.32 at boost, 0.17 at day 28 post-boost and 0.12 at day 90 may represent a true reduction in effect of genetic variation over time, or could also be a result of the limit of detection and dynamic range of the antibody assay.

Further re-analyses using recalibrated assay detection systems would be necessary to resolve this issue.

Although we have provided preliminary evidence for our mechanistic hypothesis for HLA-DQB1*06, further studies to understand the structure-function relationships based on the specific allele/peptide predictions and T cell biology will be required. Previous genetic studies for non-SARS-CoV-2 vaccines implicating the MHC suggest both HLA-peptide binding[23,37] and non-HLA effects relating, for example, to differential gene expression or complement activation[24], as potential underlying mechanisms for observed genetic associations. For SARS-CoV-2 vaccine response, consideration of HLA genotype has been advocated in vaccine design based on predicted antigens presented to T cells across different ethnic groups in order to maximise efficacy based on T cell immunity, with potential utility as a booster agent to strengthen immune responses[38].

Limitations of the study include the need for further replication of the genetic association in other studies and populations, and the representativeness of the trial population to the wider UK and global population. We propose that there is an urgent need to investigate these associations further in diverse ethnic groups and individuals of varying comorbidity to maximise insights and potential utility of the observed associations. There is also a requirement for mechanistic studies to further understand the functional basis of the association, and the relationship with specific SARS-CoV-2 variants. A further limitation is the extent to which fine mapping the association to specific variants and modulated genes was possible, reflecting the high level of sequence and structural polymorphism, sequence homologies and complex LD in the MHC[39]. Only two antibody responses were analysed, with a greater antibody repertoire and T cell immune response assays and other aspects of cell mediated immunity important to include in future studies.

We propose that to inform vaccine design and implementation against COVID-19 and other vaccine-preventable diseases with products either established or in development, an understanding of the

impact of human genetics should be prioritised to deliver translational outputs for the long-term benefit of populations world-wide.

Acknowledgements

The study is funded by the University of Oxford COVID-19 Research Response Fund and UK Research and Innovation. The ChAdOx1 nCoV-19 (AZD1222) vaccine efficacy trials were funded by the UK National Institute for Health Research, UK Research and Innovation, the Bill & Melinda Gates Foundation, the Lemann Foundation, Rede D'OR, the Brava and Telles Foundation, and the South African Medical Research Council. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The authors are grateful to the volunteers who participated in this study. AstraZeneca reviewed the data from the study and the final manuscript prior to submission, but the authors retained editorial control. JCK is supported by a Wellcome Trust Investigator Award (grant number 204969/Z/16/Z) and the NIHR Oxford Biomedical Research Centre. AJM is an NIHR Academic Clinical Lecturer. The research was supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z with additional support from the NIHR Oxford BRC and an Academy of Medical Sciences Starter Grants awarded to AJM (SGL024\1096)

Author contributions

Conceptualisation (JCK, DOC, AJM, AJP, TL), data curation (AJM, DOC), formal analysis (AJM, DOC), funding acquisition (JCK, AJP), investigation (DOC, AJM, JCK, TL, SB, IC, EAC, TDe, TDi, NJE, SF, SF, ALF, EKT, GL, XL, NM, LG, RM, MBB, JF, BA, TM, AYC, KS, RHS, MV), project administration (JCK, AJP, TL), supervision (JCK, AJP, TL), validation (AJM, DOC, JCK), visualisation (AJM, DOC, JCK), writing – original draft (JCK, AJM, DOC, AJP, TL), and writing – review & editing (JCK, AJM, DOC, AJP, TL, SF, MV). All authors critically reviewed and approved the final version.

Ethics declarations

Competing Interests

The University of Oxford has entered into a partnership with AstraZeneca for further development of ChAdOx1 nCoV-19. TL is named as an inventor on a patent application covering this SARS-CoV-2 vaccine and was a consultant to Vaccitech for an unrelated project. AJP is Chair of the UK DHSC Joint Committee on Vaccination & Immunisation but does not participate in discussions on COVID-19 vaccines, and is a member of the Strategic Advisory Group of Experts on Immunization to the WHO (World Health Organization). AJP is an NIHR senior investigator. The remaining authors declare no competing interests. The views expressed in this article do not necessarily represent the views of the DHSC, JCVI, NIHR, or WHO.

Tables and Figures

Table 1

Baseline characteristics for 1,076 ChAdOx1 nCoV-19 participants in the Phase 1/2 (COV001) and Phase 2/3 (COV002) trials who received ChAdOx1 nCoV-19 vaccine with genotype data available, and for the 1,069 individuals who either experienced a breakthrough infection after 21 days or who did not experience a breakthrough up to time of censoring. Seven individuals were not included in the breakthrough analysis as they experienced breakthrough infection within 22 days of receiving their first vaccine. Statistical differences were tested across groups using X^2 tests or exact Fisher tests if any group contained less than 5 observations.

Characteristic (range where applicable)	**Genotyped cohort (ChAdOx1 vaccinated) N=1,076**	**No reported breakthrough infection (N=957)**	**Breakthrough infection reported (N=112)**	**Evidence of difference between breakthrough groups (*P*)**
Age at recruitment	37.0 (30.0-47.0)	37.0 (30.0-47.0)	37.0 (28.8-48.0)	NS
18-55yrs, number (%)	1076 (100)	957 (100)	112 (100)	
Missing, number (%)	0 (0)	0 (0)	0 (0)	NS
Sex, number (%)				
Female	572 (53.2)	514 (53.7)	56 (50.0)	
Male	504 (46.8)	443 (46.3)	56 (50.0)	NS
BMI				
Median (IQR)	24.7 (22.6-27.5)	24.7 (22.5-27.5)	25.2 (22.7-28.3)	NS
Ethnic group, number (%)				
White	974 (90.5)	865 (90.4)	102 (91.1)	
Asian	50 (4.7)	44 (4.6)	6 (5.3)	
Black	10 (0.9)	9 (0.9)	1 (0.9)	
Mixed	29 (2.7)	26 (2.7)	3 (2.7)	
Other	13 (1.2)	13 (1.4)	0 (0)	
Not reported / missing	0 (0)	0 (0)	0 (0)	NS
Health and social care setting workers (HCW), number (%)				
Not HCW	590 (54.8)	539 (56.3)	48 (42.9)	
HCW unknown COVID contacts	101 (9.4)	91 (9.5)	10 (8.9)	
HCW with ≤1 COVID contacts	267 (24.8)	226 (23.6)	40 (35.7)	
HCW with >1 COVID contacts	118 (11.0)	101 (10.6)	14 (12.5)	0.02
Comorbidities, number (%)				
Cardiovascular	35 (3.3)	29 (3.0)	6 (5.2)	NS
Respiratory	73 (6.8)	59 (6.1)	14 (12.5)	0.02
Diabetes	9 (0.8)	7 (0.7)	2 (1.8)	NS
Interval between first and second vaccine, number (%)				
<6 weeks	12 (1.1)	11 (1.1)	1 (0.9)	
6-8 weeks	151 (14.0)	133 (13.9)	17 (15.2)	
9-11 weeks	257 (23.9)	220 (23.0)	33 (29.5)	
≥12 weeks	556 (51.7)	501 (52.4)	53 (47.3)	
none	100 (9.3)	92 (9.6)	8 (7.1)	NS
Vaccines received prior to infection occurring, number (%)				
SD	499 (46.4)	463 (48.4)	35 (31.3)	
SD / SD	400 (37.2)	342 (35.7)	54 (48.2)	
LD / SD	167 (15.5)	143 (15.0)	22 (19.6)	
SD / LD	10 (0.9)	9 (0.9)	1 (0.9)	0.005
HLA carrier, number (%)				
HLA-DQB1*06	474 (44.1)	436 (45.6)	38 (33.9)	0.02
DRB1-71E/R	881 (81.9)	784 (81.9)	92 (82.1)	NS
HLA status, number (%)				
Carrying DRB1-71E/R with no DQB1*06	532 (49.4)	464 (48.5)	63 (56.3)	
Remainder	419 (38.9)	377 (39.4)	40 (35.7)	
Carrying DQB1*06 with no DRB1-71E/R	125 (11.6)	116 (12.1)	9 (8.0)	NS

437

438 Figure 1

439 Flow diagram of participants selected for analysis from the Phase 1/2 (COV001) and Phase 2/3
440 (COV002) vaccine trials. For breakthrough infections, symptomatic individuals had primary
441 symptoms of COVID-19 (cough, shortness of breath, fever, anosmia or ageusia); if they presented
442 with symptoms other than the five primary COVID-19 symptoms, they were categorised as
443 nonprimary symptomatic cases.* samples selected to maintain investigator blinding during sample
444 selection.

445

Figure 2

Genome-wide and MHC SNP associations with RBD antibody level in 1,076 ChAdOx1 nCoV-19 vaccine recipients from the COV001 and COV002 vaccine trials. The association results for all tested autosomal and X chromosome variants with linear regression in a mixed-model framework are shown on the left in a Manhattan plot with the red horizontal line representing the nominal threshold for GWAS significance (P=5x10^{-8}) to account for the multiple tests performed. The QQ plot in the inset with expected P-values shown on the X axis and observed on the Y axis. A magnified view of a portion of the MHC locus is shown on the right of the Figure. All points represent SNPs positioned by build 37 of the human genome coordinates and coloured on the right by linkage disequilibrium (r^2) with relevant gene coordinates provided on the lower panel.

Figure 3

Fine-mapping the likely causal variants associated with day 28 post-prime anti-RBD antibody levels (normalised within immunoassay performed at MSD and PPD laboratories) in COV001 and COV002. A) and B) Stepwise conditional analyses using linear regression were performed in 1,023 individuals restricted by self-reported White ethnicity and PCA axes, and with IBD values less than or equal to 0.185. The primary unconditional association analysis across the class II MHC region (A) and HLA-DQB1 (B) locus is shown in the top rows with points shaped by variant type (amino acid (AA), HLA allele (HLA), insertion-deletion (INDEL) or single nucleotide variant (SNP) and coloured by linkage disequilibrium (r^2) with the index variant (rs9273817). The key variants of interest (rs9273817, rs1130456 and HLA-DQB1*06) are highlighted. The middle and bottom rows represent the same points after adjustment for rs9273817 (middle row) and also DRB1-71E/R (bottom row) using the same shape and colour definitions as the first row. (C) 1,076 individuals from COV001/COV002 grouped by carriage of either DQB1*06 or DRB1-71E/R in absence of the other demonstrate the most significant differences between groups tested using the two-sided Student's t-test as shown by violin plots overlain by box plots. The box plot center line represents the median; the box limits, the upper and lower quartiles; and the whiskers are the 1.5x interquartile range. *** P<0.001.

Figure 4

The effect of HLA-DQB1*06 on anti-RBD antibody accounting for DRB1-71E/R persists over time and influences risk of breakthrough infection in COV001 and COV002 in genotyped vaccine recipients. A) Where PPD-measured anti-RBD antibody levels were available in COV001 and COV002, the differences in vaccine responses by HLA type persisted over time. Differences were tested between the categories “Carrying DRB1-71E/R with no DQB1*06” and “Carrying DQB1*06 with no DRB1-71E/R” using the two-tailed Student’s t-test. Times of sampling are either after first or second (post-boost (PB)) vaccine doses. B) and C) Adjusted Cox regression curves with risk of breakthrough infection over time in 1,069 individuals stratified by carriage of (B) HLA-DQB1*06 alleles and (C) HLA-DQB1*06 accounting for DRB1-71E/R status in COV001 and COV002 vaccine recipients adjusted for age, sex, reported ethnicity, healthcare worker status, BMI and chronic disease status and including sample weighting for dose and interval between prime and boost vaccination. Included individuals had breakthrough infection at least 22 days after first vaccination. Box plot center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range ** $P<0.01$; * $P<0.05$; NS not significant; PB: post-boost.

Figure 5

The clinical implications and mechanisms of the HLA associations with differential Spike / RBD antibody levels. A) AlphaFold-based model of HLA-DQA1:01:02–HLA-DQB1:06:02–Spike peptide. The peptide is shown in orange. Residue numbering corresponds to UniProt ID P0DTC2. B) Memory B-cell, C) CD4+ T-cell proliferation and D) antigen inducible marker (AIM) CD4+ T-cell responses using biologically independent samples from 20 individuals from COV001 and COV002 stratified by carriage of HLA-DQB1*06 allele carriage sampled at days 0 and 84 following first vaccine with significant differences tested for using a one-sided Wilcoxon-rank test. Statistical differences were seen between HLA carriage groups for the memory B-cell responses (B, P=0.05) and S1 proliferation response (C, P=0.01) at day 84. Box plot center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. * P<0.05.

References

1. COVID-19 Map - Johns Hopkins Coronavirus Resource Center. Available at: https://coronavirus.jhu.edu/map.html. (Accessed: 26th September 2022)

2. Taylor, L. Covid-19: True global death toll from pandemic is almost 15 million, says WHO. *BMJ* **377**, o1144 (2022).

3. Polack, F. Safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).

4. Voysey, M. *et al.* Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* **397**, 99–111 (2021).

5. Pritchard, E. *et al.* Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. *Nat. Med. 2021 278* **27**, 1370–1378 (2021).

6. Lopez Bernal, J. *et al.* Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *N. Engl. J. Med.* **385**, 585–594 (2021).

7. Vasileiou, E. *et al.* Interim findings from first-dose mass COVID-19 vaccination roll-out and COVID-19 hospital admissions in Scotland: a national prospective cohort study. *Lancet* **397**, 1646–1657 (2021).

8. Bergwerk, M. *et al.* Covid-19 Breakthrough Infections in Vaccinated Health Care Workers. *N. Engl. J. Med.* **385**, 1474–1484 (2021).

9. Bar-On, Y. *et al.* Protection of BNT162b2 Vaccine Booster against Covid-19 in Israel. *N. Engl. J. Med.* **385**, 1393–1400 (2021).

10. 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. 2021* 1–10 (2021). doi:10.1038/s41564-021-00947-3

11. Baum, A. *et al.* REGN-COV2 antibodies prevent and treat SARS-CoV-2 infection in rhesus macaques and hamsters. *Science* **370**, 1110–1115 (2020).

12. Khoury, D. *et al.* Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat. Med.* **27**, 1205–1211 (2021).

13. Feng, S. *et al.* Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. *Nat. Med. 2021* 1–9 (2021). doi:10.1038/s41591-021-01540-1

14. Collier, D. A. *et al.* Age-related immune response heterogeneity to SARS-CoV-2 vaccine BNT162b2. *Nat. 2021* 1–6 (2021). doi:10.1038/s41586-021-03739-1

15. Newport, M. J. *et al.* Genetic regulation of immune responses to vaccines in early life. *Genes Immun.* **5**, 122–129 (2004).

16. Milich, D. R. & Leroux-Roels, G. G. Immunogenetics of the response to HBsAg vaccination. *Autoimmun Rev* **2**, 248–257 (2003).

17. Png, E. *et al.* A genome-wide association study of hepatitis B vaccine response in an Indonesian population reveals multiple independent risk variants in the HLA region. *Hum. Mol. Genet.* **20**, 3893–3898 (2011).

18. Zhang, Z. *et al.* Host Genetic Determinants of Hepatitis B Virus Infection. *Front. Genet.* **10**, 696 (2019).

19. Nishida, N. *et al.* Key HLA-DRB1-DQB1 haplotypes and role of the BTNL2 gene for response to a hepatitis B vaccine. *Hepatology* **68**, 848–858 (2018).

20. O'Connor, D. *et al.* Common Genetic Variations Associated with the Persistence of Immunity following Childhood Immunization. *Cell Rep.* **27**, 3241-3253.e4 (2019).

21. Ovsyannikova, I. G., Pankratz, V. S., Vierkant, R. A., Jacobson, R. M. & Poland, G. A. Consistency of HLA associations between two independent measles vaccine cohorts: a replication study. *Vaccine* **30**, 2146–2152 (2012).

22. Clifford, H. D. *et al.* Polymorphisms in key innate immune genes and their effects on measles vaccine responses and vaccine failure in children from Mozambique. *Vaccine* **30**, 6180–6185 (2012).

23. Suzuki, T., Yamauchi, K., Kuwata, T. & Hayashi, N. Characterization of hepatitis B virus surface antigen-specific CD4+ T cells in hepatitis B vaccine non-responders. *J. Gastroenterol. Hepatol.* **16**, 898–903 (2001).

24. Hohler, T. *et al.* C4A deficiency and nonresponse to hepatitis B vaccination. *J Hepatol* **37**, 387–392 (2002).

25. Mapping the human genetic architecture of COVID-19. *Nat. 2021* 1–8 (2021). doi:10.1038/s41586-021-03767-x

26. Kousathanas, A. *et al.* Whole-genome sequencing reveals host factors underlying critical COVID-19. *Nat. 2022 6077917* **607**, 97–103 (2022).

27. Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19. *Nat. 2020 5917848* **591**, 92–98 (2020).

28. Gonzalez-Galarza, F. F. *et al.* Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res* **43**, D784-8 (2015).

29. Siebold, C. *et al.* Crystal structure of HLA-DQ0602 that protects against type 1 diabetes and confers strong susceptibility to narcolepsy. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1999–2004 (2004).

30. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

31. Haralambieva, I. H. *et al.* Genome-wide associations of CD46 and IFI44L genetic variants with neutralizing antibody response to measles vaccine. *Hum Genet* **136**, 421–435 (2017).

32. Milne, G. *et al.* Does infection with or vaccination against SARS-CoV-2 lead to lasting immunity? *Lancet Respir. Med.* **0**, (2021).

33. Gonzalez-Galarza, F. F. *et al.* Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res.* **48**, D783–D788 (2020).

34. Hurley, C. *et al.* Common, intermediate and well-documented HLA alleles in world populations: CIWD version 3.0.0. *HLA* **95**, 516–531 (2020).

35. Hammer, C. *et al.* Amino acid variation in HLA class II proteins is a major determinant of humoral response to common viruses. *Am. J. Hum. Genet.* (2015). doi:10.1016/j.ajhg.2015.09.008

36. Dendrou, C. A., Petersen, J., Rossjohn, J. & Fugger, L. HLA variation and disease. *Nature Reviews Immunology* **18**, 325–339 (2018).

37. Nishida, N. *et al.* Importance of HBsAg recognition by HLA molecules as revealed by responsiveness to different hepatitis B vaccines. *Sci. Rep.* **11**, (2021).

38. Somogyi, E. *et al.* A Peptide Vaccine Candidate Tailored to Individuals' Genetics Mimics the Multi-Targeted T Cell Immunity of COVID-19 Convalescent Subjects. *Front. Genet.* **12**, (2021).

39. Trowsdale, J. & Knight, J. C. Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet* **14**, 301–323 (2013).

40. Folegatti, P. M. *et al.* Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against

SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet* **396**, 467–478 (2020).

41. Li, G. *et al.* Safety and immunogenicity of the ChAdOx1 nCoV-19 (AZD1222) vaccine in children aged 6-17 years: a preliminary report of COV006, a phase 2 single-blind, randomised, controlled trial. *Lancet (London, England)* **399**, 2212–2225 (2022).

42. Stuart, A. S. V. *et al.* Immunogenicity, safety, and reactogenicity of heterologous COVID-19 primary vaccination incorporating mRNA, viral-vector, and protein-adjuvant vaccines in the UK (Com-COV2): a single-blind, randomised, phase 2, non-inferiority trial. *Lancet (London, England)* **399**, 36–49 (2022).

43. Liu, X. *et al.* Safety and immunogenicity of heterologous versus homologous prime-boost schedules with an adenoviral vectored and mRNA COVID-19 vaccine (Com-COV): a single-blind, randomised, non-inferiority trial. *Lancet (London, England)* **398**, 856–869 (2021).

44. 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).

45. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279–1283 (2016).

46. Luo, Y. *et al.* A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. *Nat. Genet.* **53**, 1504–1516 (2021).

47. Stephens, M., Smith, N. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).

48. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

49. Trück, J. *et al.* The zwitterionic type I Streptococcus pneumoniae polysaccharide does not

induce memory B cell formation in humans. *Immunobiology* **218**, 368–372 (2013).

50. Clutterbuck, E. A. *et al.* Serotype-specific and age-dependent generation of pneumococcal polysaccharide-specific memory B-cell and antibody responses to immunization with a pneumococcal conjugate vaccine. *Clin. Vaccine Immunol.* **15**, 182–193 (2008).

51. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76–82 (2011).

52. Fay, M. P. & Shaw, P. A. Exact and Asymptotic Weighted Logrank Tests for Interval Censored Data: The interval R Package. *J. Stat. Softw.* **36**, 1–34 (2010).

53. Therneau, T. M. & Grambsch, P. M. A Package for Survival Analysis in S, version 2.38. *Model. Surviv. Data Extending Cox Model* (2000).

54. van der Wal, W. M. & Geskus, R. B. ipw: An R Package for Inverse Probability Weighting. *J. Stat. Softw.* **43**, 1–23 (2011).

55. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).

Methods

Study design and participants

Discovery cohort

The participants were enrolled in phase 1/2 (COV001) or phase 2/3 (COV002) randomized single-blind ChAdOx1 nCoV-19 (AZD1222) vaccine multi-centre efficacy trials conducted across multiple sites within the United Kingdom[4,40] (NCT04324606, NCT04400838). In brief, following written informed consent, adults aged 18 years and older were randomly assigned to receive either intramuscular ChAdOx1 nCoV-19 (AZD1222) or a control vaccine (MenACWY), to assess the safety and efficacy of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2[4,40]. All individuals included in the analyses agreed to being included in genetic studies as part of the vaccine trial consent, with the opportunity to opt out. The trials were conducted according to the principles of Good Clinical Practice and approved by the South Central Berkshire Research Ethics Committee (20/SC/0145 and 20/SC/0179) and the UK regulatory agency (the Medicines and Healthcare products Regulatory Agency). Individuals with humoral immune responses measured post-vaccination were selected for this genotyping study. To maintain blinding of investigators to vaccine allocation — prior to vaccine trial reporting — participants who received the control vaccine were also genotyped, at a ratio of 1:10.

Breakthrough infections. Participants received weekly reminders to report any primary symptoms of COVID-19 (cough, shortness of breath, fever, anosmia or ageusia), and if symptomatic a SARS-CoV-2 nucleic acid amplification test (NAAT) was performed on a nose and throat swab. Participants were also asked to return a weekly nose and throat swab for NAAT for the duration of the study. A breakthrough infection was defined as SARS-CoV-2 NAAT-positive swab at least 22 days after a first dose of vaccine. If participants were NAAT-positive but had symptoms other than the five primary COVID-19 described above, they were categorised as nonprimary symptomatic cases but still included in the final breakthrough analyses reported herein.

Replication cohort

The replication cohort was comprised of participants from three COVID-19 vaccine trials conducted across several sites within the United Kingdom[41–43]. Two of these trials (COMCOV and COMCOV2) were in adults aged 50 years and older, randomised to receive homologous or heterologous two-dose schedules of either intramuscular ChAdOx1 nCoV-19, mRNA vaccines (BNT162b2 or mRNA-1273) or a nanoparticle vaccine (NVX-CoV2373)[42,43]. The other trial (COV006) was in children aged 6-17 who were randomised to receive either intramuscular ChAdOx1 nCoV-19 or a control vaccine (capsular group B meningococcal vaccine, 4CMenB)[41]. All participants included in this manuscript consented (or their parents/guardians) to being included in genetic studies as part of the vaccine trial consent, with the opportunity to opt out in the consent form.

Breakthrough infections. Adult participants in the replication cohort self-reported breakthrough COVID-19 (no active COVID-19 testing as part of study) and associated symptoms. Parents/guardians reported PCR-confirmed or lateral flow assay-confirmed SARS-CoV-2 infections for the childhood participants. We defined a breakthrough infection for this cohort as a self-reported case of COVID-19 at least 22 days after a first dose of vaccine.

Antibody concentrations

Discovery cohort

Blood samples for serological testing were taken at baseline, day 28 after the first dose of vaccine, prior to the second vaccine, and then at day 28, 90 and 182 after the second dose of vaccine. Day 28 post-first vaccine responses were available on all participants, and the variance of response across the cohort was substantial and together this influenced the choice for the discovery GWAS. Humoral immune responses were assessed using Meso Scale Discovery (MSD) multiplexed immunoassay against SARS-CoV-2 S and RBD as well as the nucleocapsid (N) proteins — in the entire phase 1/2 UK cohort (n= 585) and 15% of the phase 2/3 UK cohort (n=637). Sample selection from the phase 2/3 trial was based on samples processed for the initial application for emergency use, which required 15% of samples included in the efficacy trial to be processed on validated assays. In addition,

serological testing was also performed on samples from NAAT-positive individuals — missing data were assumed to be missing at random.

Anti- S, RBD and N IgG levels were measured by a multiplex immunoassay using the MSD platforms at two laboratories; the phase 1/2 samples were performed at MSD and the phase 2/3 samples were performed at PPD Laboratories. Ninety-seven samples were assessed by both laboratories, to evaluate inter-assay agreement.

Replication cohort

The replication cohort had blood samples for serological testing taken at baseline participation in the study, which was 28–84 days following the first dose of vaccine and on the day of receiving the second dose of vaccine. Adult participants in the replication cohort had their SARS-CoV-2 anti-S IgG levels measured by enzyme-linked immunosorbent assay (ELISA) at Nexelis (Laval, Canada). In the childhood replication study, anti- S IgG levels were measured using MSD at PPD Laboratories prior to receipt of any further vaccine dose.

DNA extraction

DNA was extracted from either blood clots remaining following serum separation by centrifugation or whole blood collected in EDTA tubes. In brief, clots were dispersed by centrifugation through clots spin baskets (Qiagen, catalogue No. 158932). ATL buffer (Qiagen, catalogue No. 1014758) was added to the centrifuged clot and vortexed. Proteinase K (Qiagen, catalogue No. 19131) was added, vortexed thoroughly and incubated in shaking incubator at 56°C until clot was completely lysed (overnight). Following lysis, AL buffer was added (Qiagen, catalogue No. 1038826) and vortexed thoroughly. Lysate or whole blood was then transferred to the QIAsymphony 2.0 and extracted using the QIAsymphony DSP DNA Midi kit (Qiagen, catalogue No 937255).

Genotyping

Standardised DNA was sent to the National Institute for Health Research UK BioCentre and genotyped using their established pipelines on the Affymetrix AxiomTM HGCoV2 1 array. Raw CEL

719 files were transferred back to Oxford for file conversion into build 38 using the Axiom Analysis Suite
720 Best Practice Workflow. Individual samples and single nucleotide polymorphism (SNP) variants were
721 exported and underwent further quality control using PLINK[44] (version 1.9).

722 Individuals were excluded if more than 3% of SNPs were classed as missing, or if derived genetic sex
723 did not match reported sex, or if levels of heterozygosity lay more than 3 times the standard
724 deviation from the mean of individuals stratified by self-reported ethnicity. Identity by descent (IBD)
725 was calculated for all pairs of individuals and individuals were excluded if they had estimates ≥0.9,
726 excluding the individual with the highest SNP missingness rate from each pair preferentially. Hardy-
727 Weinberg estimates were calculated for each SNP within individuals classified as founders (with
728 IBD<0.05) and SNPs were excluded if exhibiting extreme deviations from equilibrium (using a
729 threshold of $P<1x10^{-50}$ calculated in PLINK 1.9).

730 Principal components were calculated for all individuals both within the genotyped COVID-19
731 vaccine cohorts and merged with individuals from the 1000 Genomes project derived from Great
732 Britain (GBR), Han individuals from China (CHS) and African individuals from Kenya (LWK) and Nigeria
733 (YRI). Before merging vaccinee data with 1000 Genomes individuals, SNPs were first lifted over from
734 Build 38 to Build 37 coordinates using LiftOver https://genome.sph.umich.edu/wiki/LiftOver). Plots
735 were inspected to detect samples with extreme outlier values (>3 standard deviations from any
736 expected cluster). A European cluster was defined by including only individuals falling within a
737 defined cluster with GBR individuals and self-reporting as White ethnicity. Quality controlled SNPs
738 and individuals were taken forwards for genotype imputation which was undertaken on the
739 Michigan Imputation server using the TopMed reference panel using recommended settings[45]. Files
740 were converted to MACH format using DosageConverter (version 1.0.4,
741 https://genome.sph.umich.edu/wiki/DosageConvertor).

742 HLA imputation was performed using the Multi-Ethnic HLA reference panel (version 1.0 2021)
743 available on the Michigan imputation server[46] using recommended settings. Phasing of multi-allelic

744 HLA alleles was undertaken using PHASE (version 2.1.1)[47]. HLA typing was also performed using PCR-
745 sequence specific primers (SSP) at the Histogenetic laboratory, Oxford University NHS Foundation
746 Trust, Oxford, UK.

747 **Structural analyses.**

748 The ternary HLA-DQA1*01:02–HLA-DQB1*06:02–Spike peptide complex was modelled using
749 AlphaFold[30] available on Google Colaboratory server[48]. Structures were analysed and displayed using
750 the PyMOL Molecular Graphics System, version 2.3.2. from Schrödinger, LLC

751 **Follow up functional assays.**

752 Samples at baseline (D0) and 28 days post-boost (D84) from 20 healthy adult volunteers
753 participating in COV001 and COV002 who had received a two dose ChAdOx1 nCoV-19 vaccine
754 schedule of either two standard doses, or one standard dose and one low dose were chosen for
755 further functional work. Samples were evenly stratified by dose of second vaccine across individuals
756 who either carried HLA-DQB1*06 (with no HLA-DQB1*05) and HLA-DQB1*05 carriers (not carrying
757 HLA-DQB1*06).

758 For the proliferation assay, cryopreserved PBMC were washed twice with sterile DPBS and incubated
759 with CellTrace Violet (Life Technologies), a free amine binding dye, at a final concentration of 2.5 µM
760 in PBS for 10 minutes at RT, protected from light. To quench any dye remaining in solution, cells
761 were incubated with FBS for 5 minutes at 4°C. Cells were then resuspended in RPMI-1640 media
762 supplemented with 1% L-glutamine, 1% penicillin streptomycin and 10% human AB serum (Sigma)
763 and 250,000 cells were seeded per condition in U bottom 96-well tissue culture plates. Cells were
764 stimulated with 15-mer peptides overlapping by 10 amino acids, comprising the length of the S1 or
765 the S2 domain of the SARS-CoV-2 spike protein (ProImmune) at a final concentration of 1 µg/ml. 2
766 µg/ml of PHA served as a positive control and cells incubated with 0.13% DMSO (Sigma) in cell media
767 was used as a negative control. Cells were then incubated for 7 days at 37°C, with 5% CO_2 and 95%

humidity and the media was changed on day 4. At the end of the incubation period, cells were stained with anti–human CD3-FITC (OKT3, Invitrogen), CD4-APC (RPA-T4, Invitrogen), CD8-PE-Cyanine7 (OKT8, Invitrogen), and Live/Dead near Infra-red stain (Invitrogen) prior to acquisition on a five-laser LSRFortessa X-20 flow cytometer (BD Biosciences) using FACSDiva v8.02 (BD Biosciences), and data were analysed in FlowJo v10.7. A hierarchical gating structure was applied, and data is presented as background subtracted.

For the T-cell activation induced marker assay, cryopreserved PBMCs from the same individuals were stimulated overnight (18 hours, 37°C, 5% CO2) at $2x10^6$ cells per tube in rounded 5ml polystyrene U-bottom tubes (Falcon). Samples were stimulated with anti-CD28 (CD28.2, 1µg/ml, Invitrogen) and anti-CD49d (9F10, 1µg/ml, Invitrogen) alongside either a SARS-CoV-2 S1/S2 peptide megapool (134 peptides, 2µg/ml per peptide, ProImmune) or anti-CD3 (OKT3, Tonbo Biosciences) as a positive control.

Samples were stained with αCCR7-PE-Cy7 (G043H7, Biolegend) for 10 minutes at 37°C prior to the addition of the following antibodies and incubated for 30mins at room temperature: αCD3-BV711 (UCHT1, BD Biosciences), αCD4-PerCP-Cy5.5 (RPA-T4, Biolegend), αCD8-BV650 (RPA-T8, Biolegend), αCD14-APC-Cy7 (HCD14, Biolegend), αCD19-APC-Cy7 (HIB19, Biolegend), LIVE/DEAD Fixable Near IR (Invitrogen), αCD137-PE-Cy5 (4B4-1, Biolegend), αOX40-APC (Ber-ACT, Biolegend), αCD69-BUV395(FN50, BD Biosciences), αCCR6-PE (G034E3, Biolegend), αCXCR5-PE Dazzle 594 (J252D4, Biolegend), PD1-BV510 (EH 12.1, BD Biosciences), CXCR3-FITC (G025H7, Biolegend), CD45RA-AF700 (H1100, Biolegend), CCR7-PE-Cy7 (G043H7, Biolegend). Samples were acquired using a BD LSRFortessa Cell Analyzer.

Memory B cells were differentiated into antibody-secreting plasma cells for the detection of IgG responses using Enzyme-linked immunospot (ELISpot) according to the following steps. Cryopreserved PBMCS from the same individuals were isolated from heparinized whole blood and resuspended at a final concentration of $2x10^6$ cells/ml in Complete Medium (CM). CM was prepared

by combining RPMI (450ml, R-5886, Merck-Sigma), FBS-HI (50ml, F9665, Merck-Sigma), penicillin/streptomycin (5ml, P-4458, Merck-Sigma), L-Glutamine (5ml, G-7513, Merck-Sigma), NEAA (5ml, 11140035, Life Technologies), Sodium Pyruvate (5ml, 11360039, Life Technologies), and 2-Mercaptoethanol (5002µl, 31350010, Life Technologies). 10ml of polyclonal stimulants antigen mix (CpG (tlrl-2006-5, InvivGen SAS)+ SAC (507861-50, VWR International Ltd)+ PWM (L-9370, Sigma)) was distributed in 1x96 well plate (650180, Greiner) at 100 µl/well. 100 µl/well of cell suspension was added to each well (giving $2x10^5$ cells/well) and incubated at 37°C/ 5% CO2/ 95% humidity for 5-6 days. Memory B cells were harvested from the 96-well plate by gentle re-suspension into a 30ml universal container (201150, Greiner), and washed 3 times. The final pellet was re-suspended in 1ml wash buffer for cell counting and resuspended in CM at $2x10^6$ cells/ml. ELISpot was performed as previously described[49,50].

Statistical analyses

Discovery cohort

Antibody responses were log-transformed and density distribution plots inspected for overlap in density distributions between laboratories (MSD or PPD) and paired correlation between assays using results available from both laboratories performed on samples from the same individual. If the paired correlation was less than 0.7 or the density distributions did not overlap, traits were tested for association within assay type then meta-analysed, but if the paired correlation was greater than 0.7 and the density distributions overlapped, traits were quintile normalised (using the qqnorm() function in R) within assay platform groups and tested in a pooled analysis including assay type as a fixed effect covariate. Samples taken at Day 28 post-first dose were used for the primary analysis.

A linear mixed model was used to maximise power and account for the diverse population structure and potential unrecognised close (cryptic) relatedness between study participants. Age, sex, antibody assay platform and nucleocapsid positivity were included as fixed effect covariates for each association. Genotype and HLA-wide association analyses were performed using the 'mlma' function

in GCTA (version 1.24.4)[51] after generating a genetic relatedness matrix (GRM) using non-pruned genotyped SNPs. The GRM was modelled as a random effect covariate with age (in years), sex, antibody assay laboratory and likelihood of natural exposure (based on anti-nucleocapsid protein antibodies (anti-N) >1000) coded as a binary variable included as fixed effect covariates for the primary genome-wide association study.

Sensitivity analyses for the genetic components included a further round of association analyses incorporating the first ten principal components for all individuals calculated within-cohort, and a further round of normalisation on the within-assay normalised RBD distributions.

Comparisons between imputed and classically typed HLA alleles were undertaken at the 4-digit (i.e. 2-field) level of resolution. If a classically typed available call at a single allele locus included several potential higher resolution alleles (i.e. a list of potential ambiguities) only the first available allele call (adhering to a CWD priority) was used for comparison. If types were available to 6- or 8- digit resolution, the calls were reduced to 4-digit resolution for comparison between call types. The classical types were treated as the 'truth' set. By comparing each individual allele in turn it was possible to define calls of the imputed (or 'test') set that were:

- True positives (TP)
- False positives (FP); called by the test as that allele when it was in fact another allele according to the truth)
- False negatives (FN; called by the test as another allele when it was in fact this allele according to the truth)
- True negatives (TN).

Thus at the level of an individual allele various metrics could be calculated. Sensitivity was defined as:

TP / (TP + FN)

Specificity was defined as:

TN / (TN + FP)

Positive predictive value (PPV) was defined as:

TP / (TP + FP)

Negative predictive value (NPV) was defined as:

TN / (TN + FN)

Accuracy was defined as:

(TP + TN) / (TP + FP + FN + TN)

Concordance was calculated at the level of the locus. For every pair of chromosomes with data available in both truth and test sets the number of identical allele calls between platforms was calculated and divided by the total number of alleles, equivalent to the positive predictive value (PPV). Any individual with missing alleles on either or both chromosomes on either platform were excluded from these calculations.

Univariate *P*-values were calculated using the Wilcoxon-rank test for continuous variables and Chi-squared Test for nominal variables. Permutation of Wilcoxon rank test of associating HLA alleles with RBD response was performed using the 'perm' package in R[52]. Adjusted Cox proportional hazards regression analyses for the breakthrough analyses were performed using the 'survival'[53] package in R and plots generated using 'survminer' and 'ggplot2'. The breakthrough infection definition above for the discovery cohorts was used as the outcome in the proportional hazards model with individuals censored at either time of breakthrough, date of withdrawal or the 10 October 2021, whichever came first. The effect of carriage of HLA-DQB1*06 was estimated after adjustment for age (in years measured at baseline), healthcare worker status (defining whether individuals were healthcare workers and whether they cared for one, or more than one patient with COVID-19 per week), BMI

(less than 30 kg/m^2, or equal to or greater than 30 kg/m^2), ancestry (using the first two principal components of genetic variation to capture genetic structure), chronic health condition (presence of none, or one or more chronic respiratory, cardiovascular or diabetic health conditions). Using analyses undertaken on understanding the correlates of protection of ChAdOx1 nCoV-19[13] we aimed to perform an identical modelling exercise and thus samples were reweighted based on the interval between first and second vaccines (no second dose, <6, 6-8, 9-11 and ≥12 weeks) and dose arm of trial (standard dose (SD) alone, SD/SD, low dose (LD)/SD, SD/LD) distributions of individuals within the entire genotyped set with antibody data available using inverse probability weighting calculated using the 'ipw' package in R[54]. All analyses were undertaken in R version 4.1.1, except estimation of the genomic inflation factor (λ) which was undertaken in R version 3.6.2 using the GenABEL package[55].

Replication cohort

Individuals from all three replication cohorts were stratified by HLA-DQB1*06 and DRB1-71E/R status and tested for association with $\log_{10}$-transformed anti-S antibody levels measured following the first vaccine dose (and before the second vaccine) using linear regression adjusting for age, sex, self-reported ethnicity, priming vaccine, study and interval between prime and blood sample in days. Individuals were either analysed together or stratified according to first vaccine dose received (ChAdOx1 nCoV-19 or BNT162b2). Survival analyses were performed using the same methods as for the discovery cohorts but including only age, sex, and first vaccine type received as covariates with reweighting performed using interval between first and second vaccines in days. Censoring was undertaken at point of breakthrough infection, withdrawal from study or date of database locking (21 January 2022 for COMCOV/COMCOV2 or 29 November 2021 for COV006), whichever came soonest. Healthcare worker status, BMI and chronic health condition information was not available for COV006 and so the variables were not included in the final replication analysis. All replication analyses were performed in R version 4.1.1.

Data availability

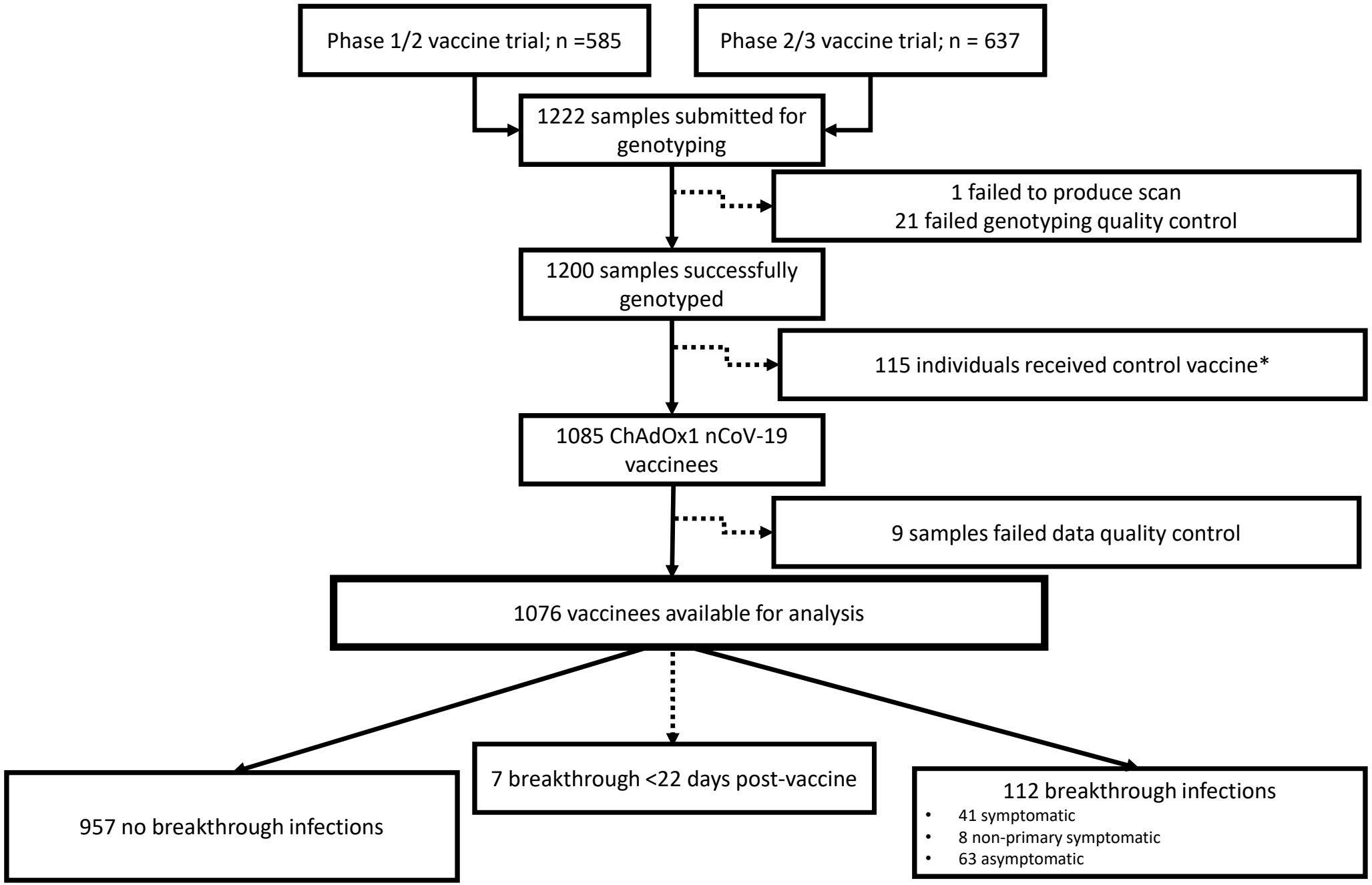
The University of Oxford is committed to providing access to anonymized data for non-commercial research. Participants genotype and phenotype data will be deposited on European Genome Phenome Archive and will be made available for non-commercial use only (DUO:0000046).

Oxford COVID Vaccine Trial Genetics Study Team Group

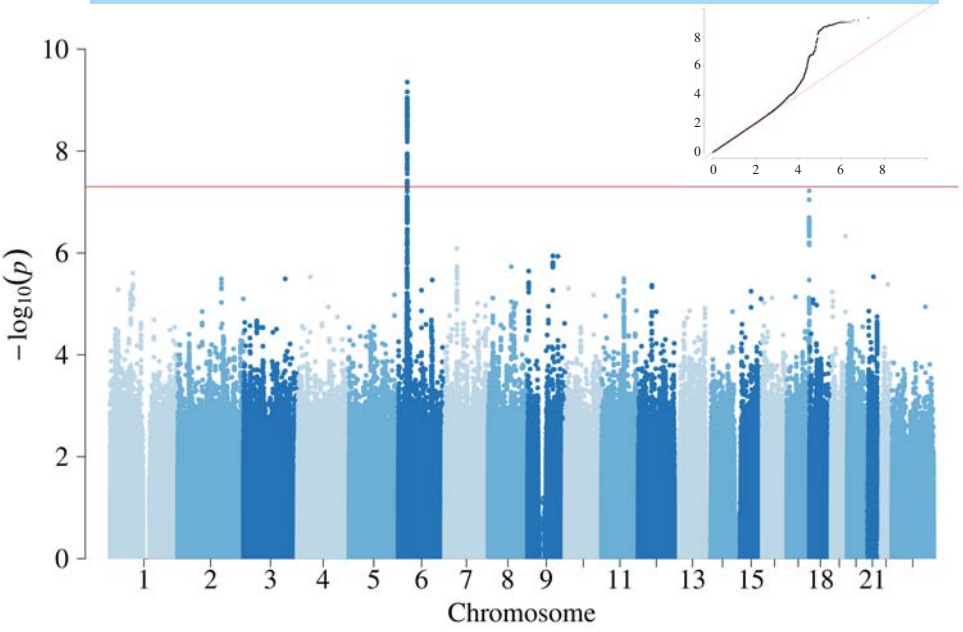
Ana Cavey[7], Angela Minassian[2,8], Arabella Stuart[2], Baktash Khozoee[4], Brama Hanumunthadu[2], Brian Angus[4], Catherine C Smith[2], Iain Turnbull[9], Jonathan Kwok[10], Katherine R W Emary[2], Liliana Cifuentes[11], Maheshi N Ramasamy[2], Paola Cicconi[4], Adam Finn[12], Alastair C McGregor[13], Andrea M Collins[14], Andrew Smith[15], Anna L Goodman[16,17], Christopher A Green[18], Christopher J A Duncan[19,20,21], Christopher J A Williams[22], Daniela M Ferreira[2,14], David P J Turner[23], Emma C Thomson[24], Helen Hill[14], Katrina Pollock[25], Mark Toshner[26], Patrick J Lillie[27], Paul Heath[28,29], Rajeka Lazarus[30], Rebecca K Sutherland[31], Ruth O Payne[32], Saul N Faust[33,34], Tom Darton[32], Vincenzo Libri[35], Rachel Anslow[2], Samuel Provtsgaard-Morys[2], Thomas Hart[2], Amy Beveridge[2], Syed Adlou[2]

[7]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[8]Department of Biochemistry, University of Oxford, Oxford UK
[9]Nuffield Department of Population Health, University of Oxford, Oxford, UK
[10]Nuffield Department of Medicine, University of Oxford, Oxford, UK
[11]Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, The University of Oxford, UK
[12]School of Population Health Sciences, University of Bristol and University Hospitals Bristol and Weston NHS Foundation Trust, UK
[13]London Northwest University Healthcare, Harrow, UK
[14]Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK
[15]College of Medical, Veterinary & Life Sciences, Glasgow Dental Hospital & School, University of Glasgow, Glasgow, UK
[16]Guy's and St Thomas' NHS Foundation Trust, London, UK
[17]MRC Clinical Trials Unit at UCL, London
[18]NIHR/Wellcome Trust Clinical Research Facility, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK
[19]Department of Infection and Tropical Medicine, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK
[20]NIHR Newcastle Clinical Research Facility, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK
[21]Translational and Clinical Research Institute, Immunity and Inflammation Theme, Newcastle University, UK
[22]Public Health Wales, Cardiff, Wales and Aneurin Bevan University Health Board, Wales
[23]School of Life Sciences, University of Nottingham and Nottingham University Hospitals NHS Trust, Nottingham, UK
[24]MRC - University of Glasgow Centre for Virus Research & Department of Infectious Diseases, Queen Elizabeth University Hospital, Glasgow, UK
[25]NIHR Imperial Clinical Research Facility and NIHR Imperial Biomedical Research Centre, London, UK
[26]Heart Lung Research Institute, Department of Medicine, Cambridge
[27]Hull University Teaching Hospitals NHS Trust, UK
[28]St George's Vaccine Institute, St George's, University of London, London, UK
[29]Department of Infection, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
[30]Severn Pathology, North Bristol NHS Trust, Bristol, UK
[31]Clinical Infection Research Group, Regional Infectious Diseases Unit, NHS Lothian, Edinburgh, UK
[32]Department of Infection, Immunity and Cardiovascular Diseases, University of Sheffield, Sheffield, UK
[33]NIHR Southampton Clinical Research Facility and Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK

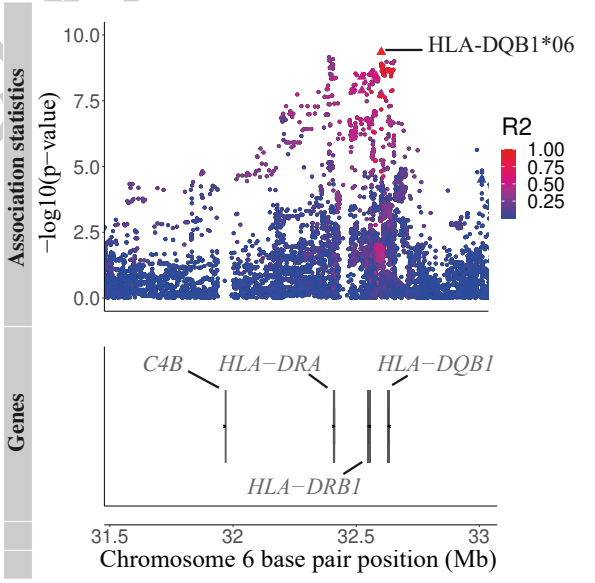
949 [34]Faculty of Medicine and Institute for Life Sciences, University of Southampton, Southampton, UK
950 [35]NIHR UCLH Clinical Research Facility and NIHR UCLH Biomedical Research Centre, London, UK
951


Phase 1/2 vaccine trial; n =585
Phase 2/3 vaccine trial; n = 637
1222 samples submitted for genotyping
1 failed to produce scan
21 failed genotyping quality control
1200 samples successfully genotyped
115 individuals received control vaccine*
1085 ChAdOx1 nCoV-19 vaccinees
9 samples failed data quality control
1076 vaccinees available for analysis
957 no breakthrough infections
7 breakthrough <22 days post-vaccine
112 breakthrough infections
41 symptomatic
8 non-primary symptomatic
63 asymptomatic

Anti-RBD Chromosomal Manhattan and QQ



MHC Regional Association Plot



A

B

Unconditional RBD in EUR

Conditional on HLA−DQB1*06

Conditional on HLA−DQB1*06 and DRB1−71E/R

−log10(p−value)

HLA-DQB1*06

rs1130456

rs9273817

DRB1-71E/R

R2

1.00
0.75
0.50
0.25

Variant type

AA
HLA
INDEL
SNP

HLA-DRB1

HLA-DQB1

32.5
32.6
32.7

32.625
32.63
32.635

Chromosome 6 base pair position (Mb)

C

Normalised anti−RBD titre

***
***
***

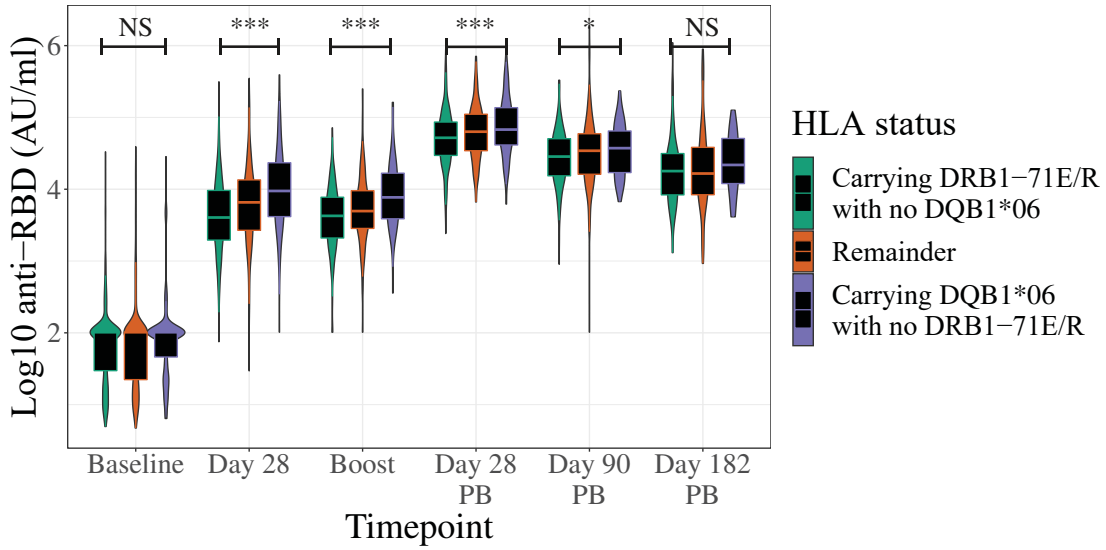
Carrying DRB1−71E/R with no DQB1*06

Remainder

Carrying DQB1*06 with no DRB1−71E/R

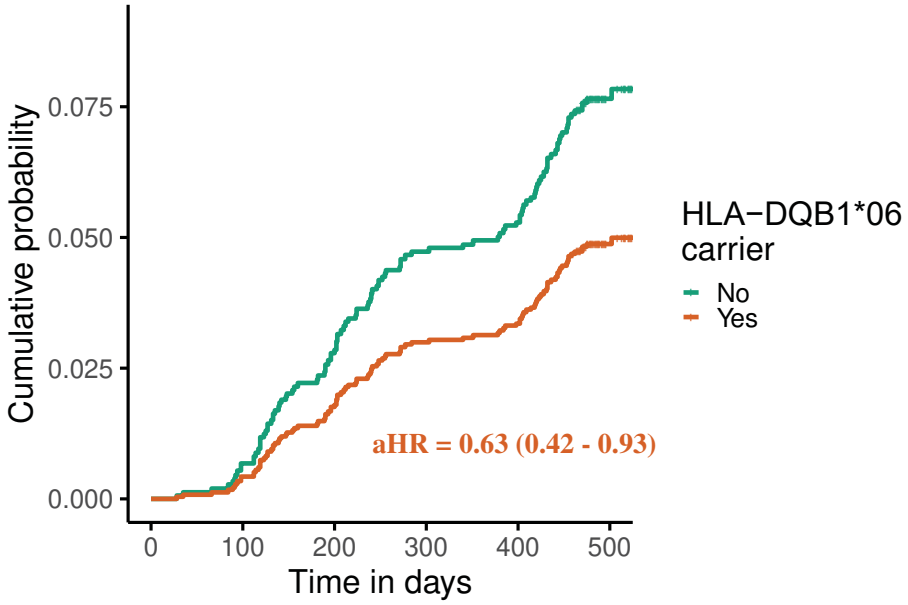
HLA−DQB1 and DRB1 status

A

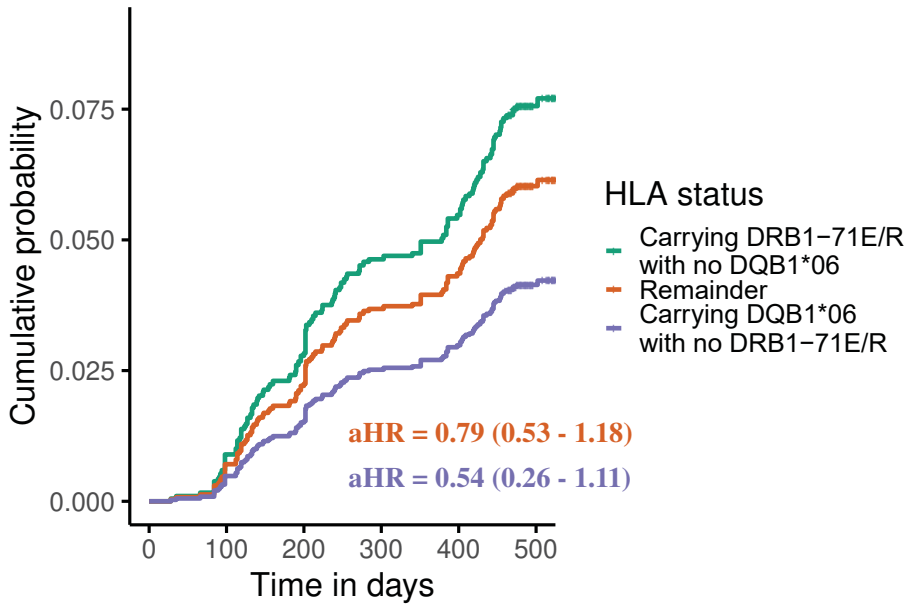


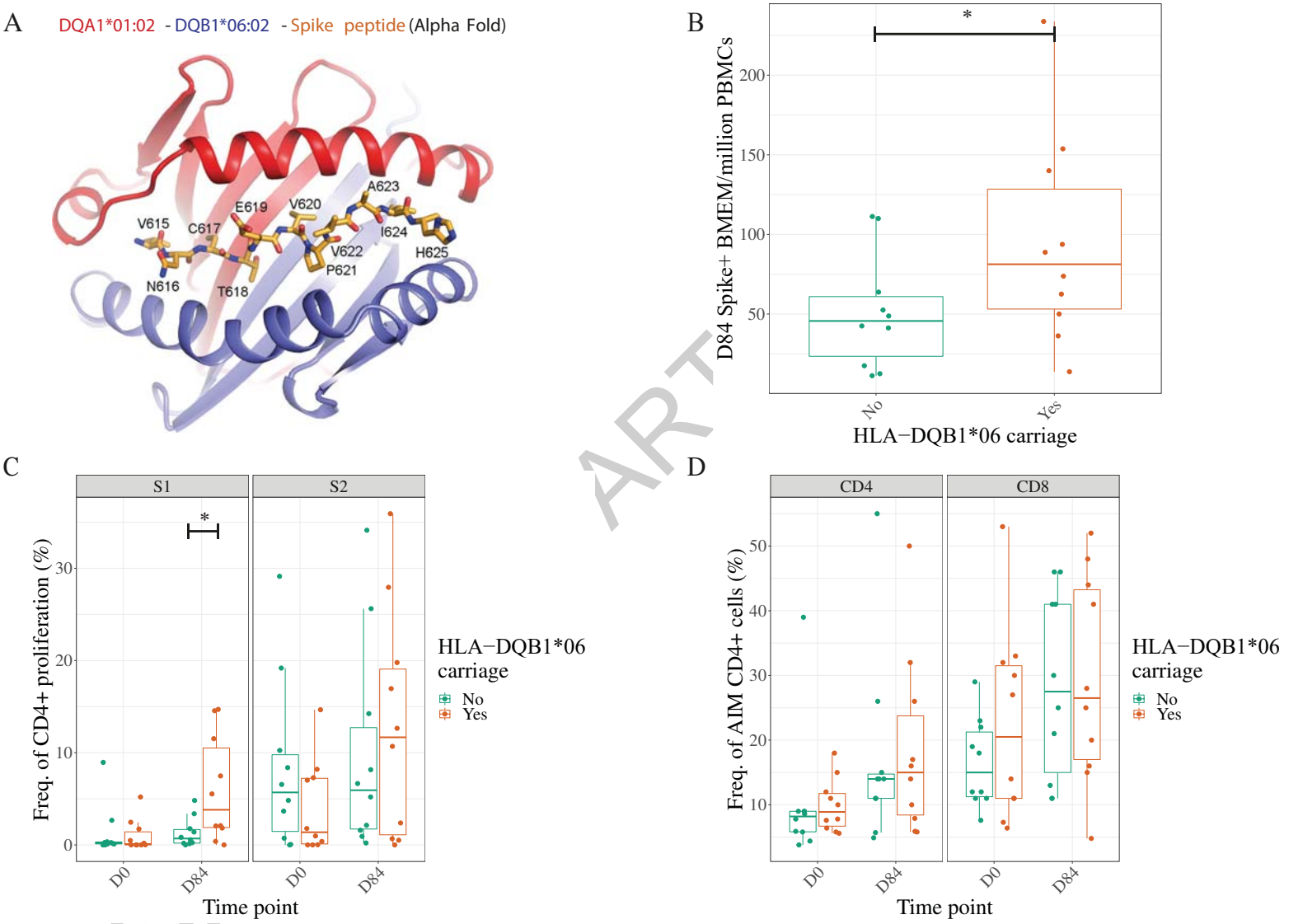
Carrying DRB1-71E/R with no DQB1*06	533	533	326	331	196	124
Carrying neither DQB1*05 or DQB1*06	419	419	251	353	163	108
Carrying DQB1*06 with no DQB1*05	125	125	81	82	54	32

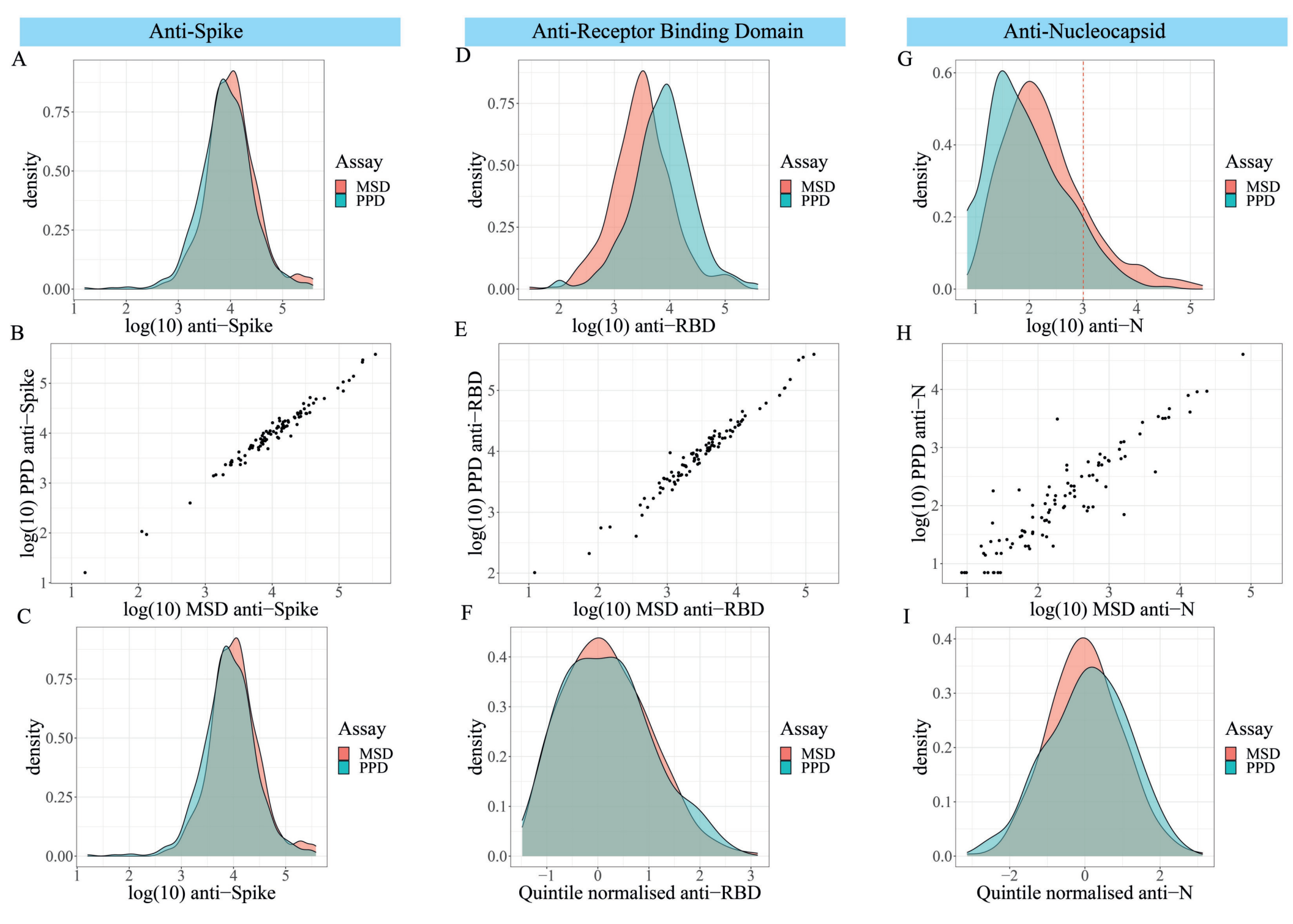
B

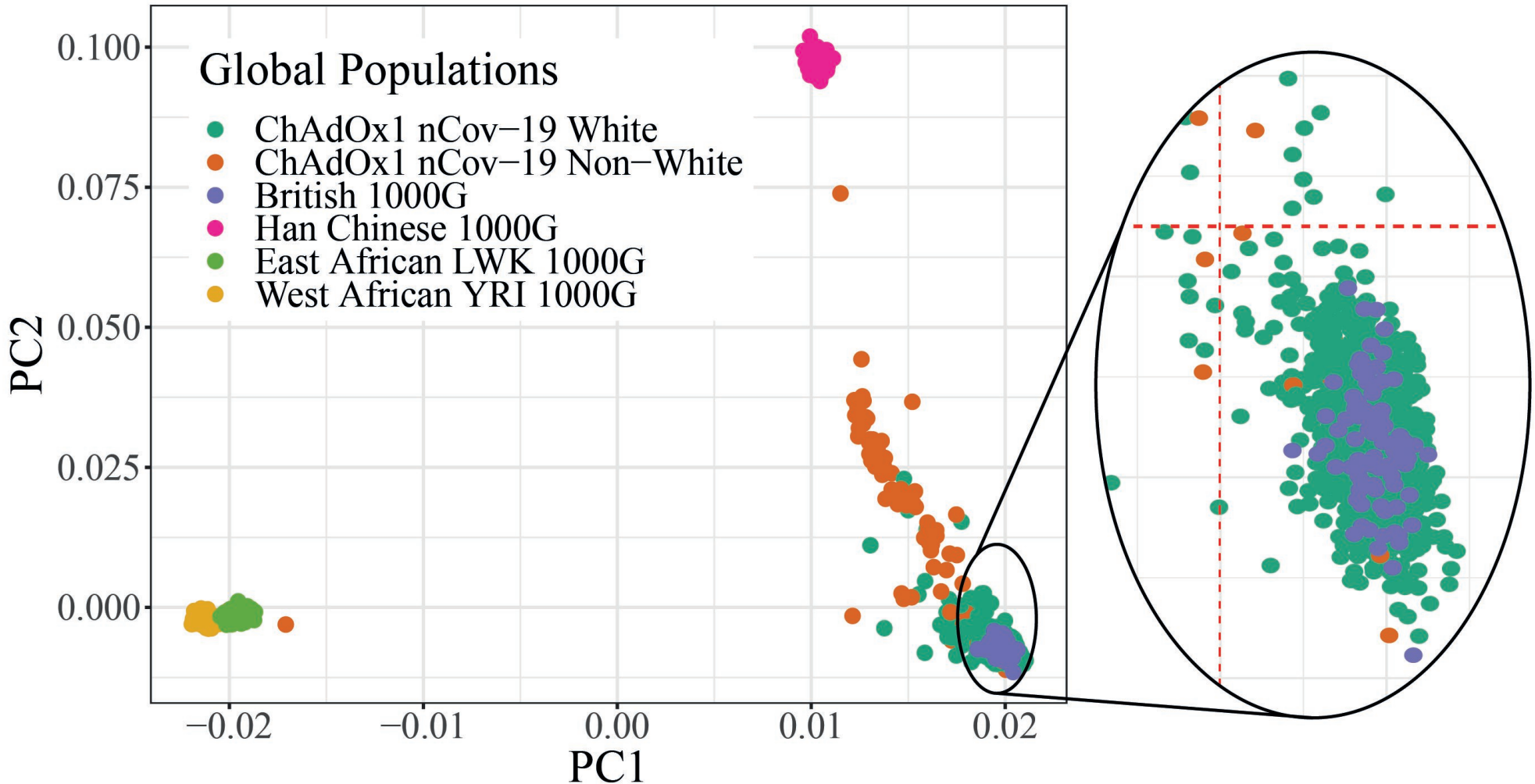


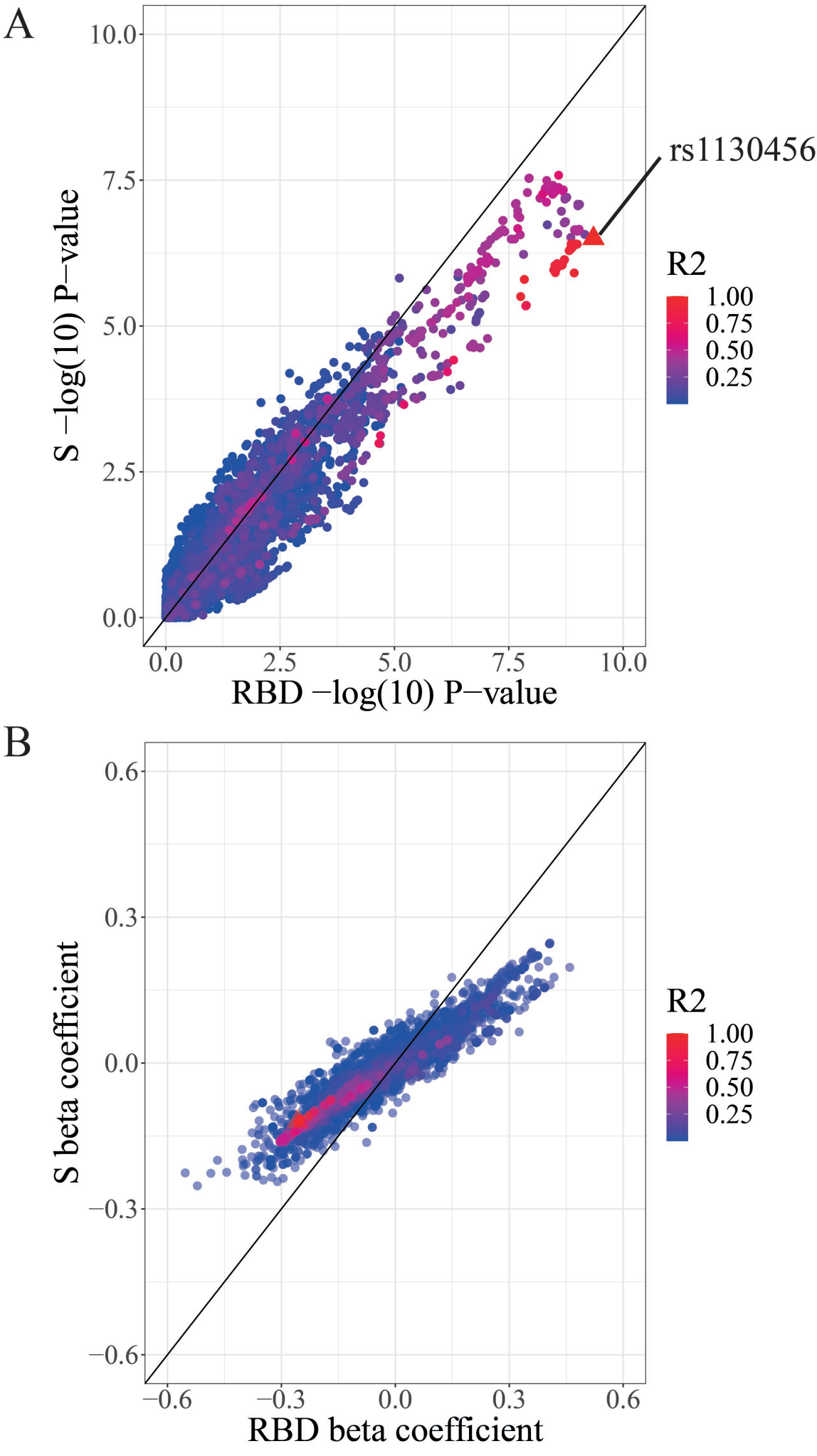
C

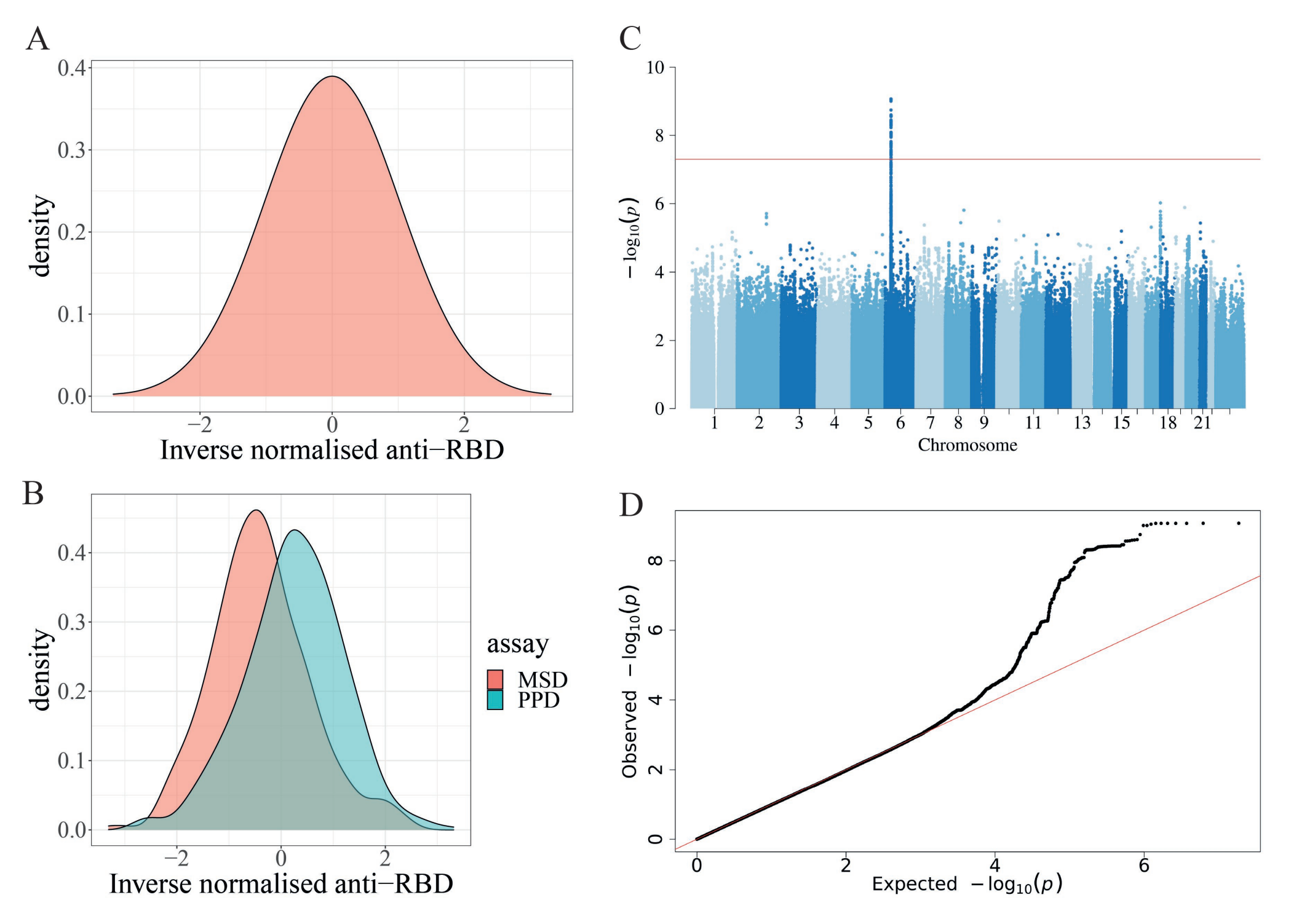


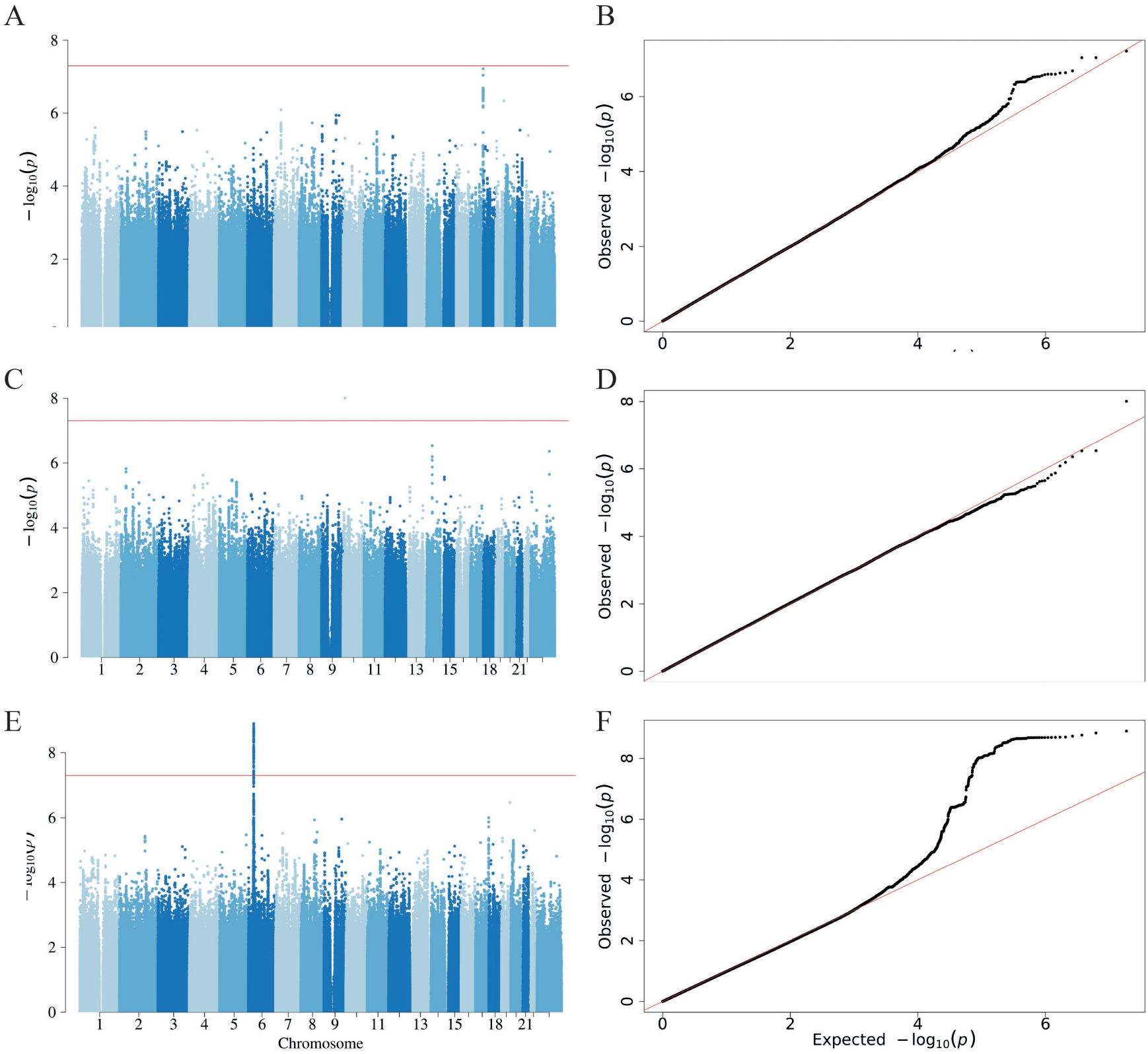


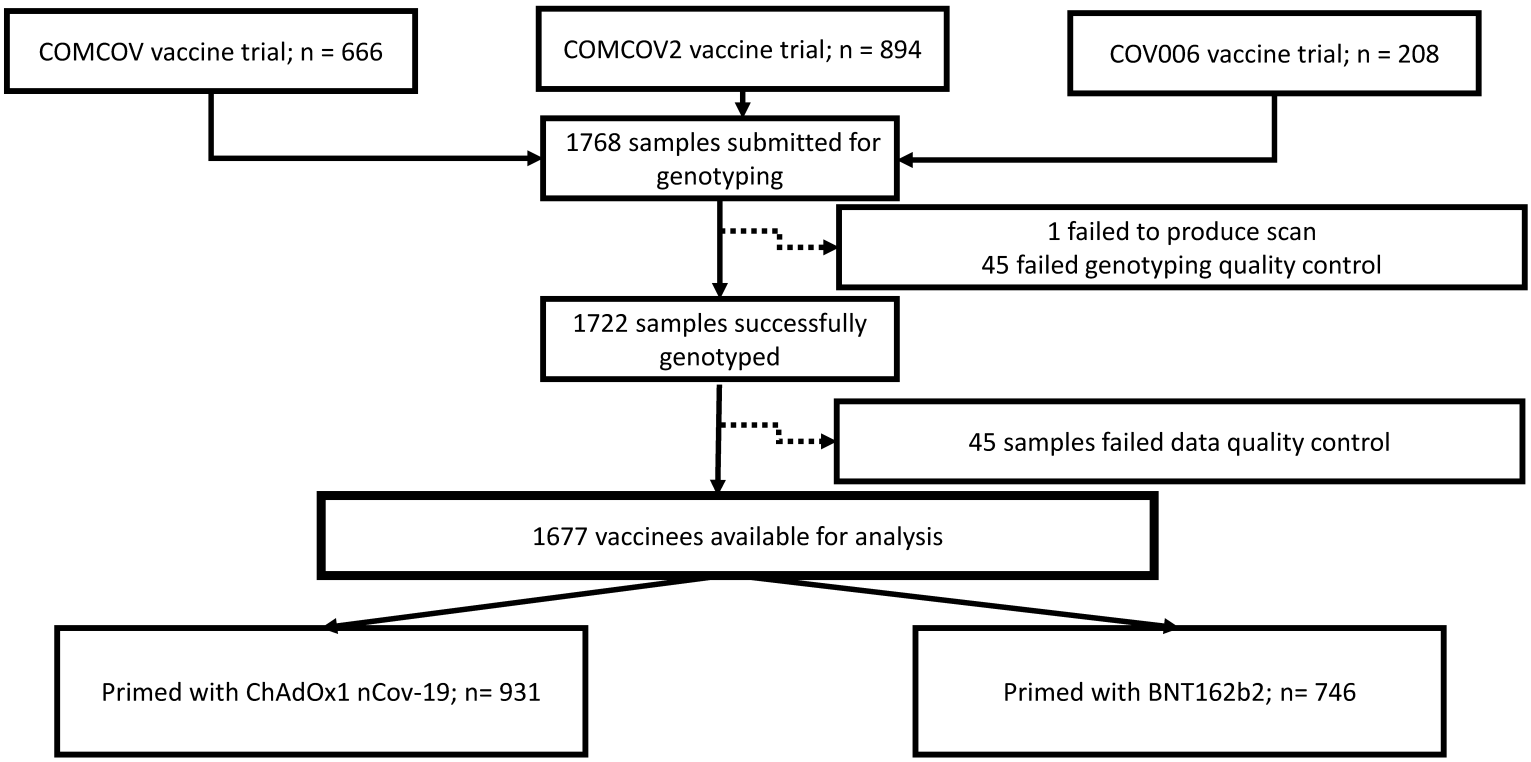

Anti-Spike
Anti-Receptor Binding Domain
Anti-Nucleocapsid
A
B
C
D
E
F
G
H
I
density
log(10) anti−Spike
log(10) PPD anti−Spike
log(10) MSD anti−Spike
log(10) anti−RBD
log(10) PPD anti−RBD
log(10) MSD anti−RBD
Quintile normalised anti−RBD
log(10) anti−N
log(10) PPD anti−N
log(10) MSD anti−N
Quintile normalised anti−N
Assay
MSD
PPD

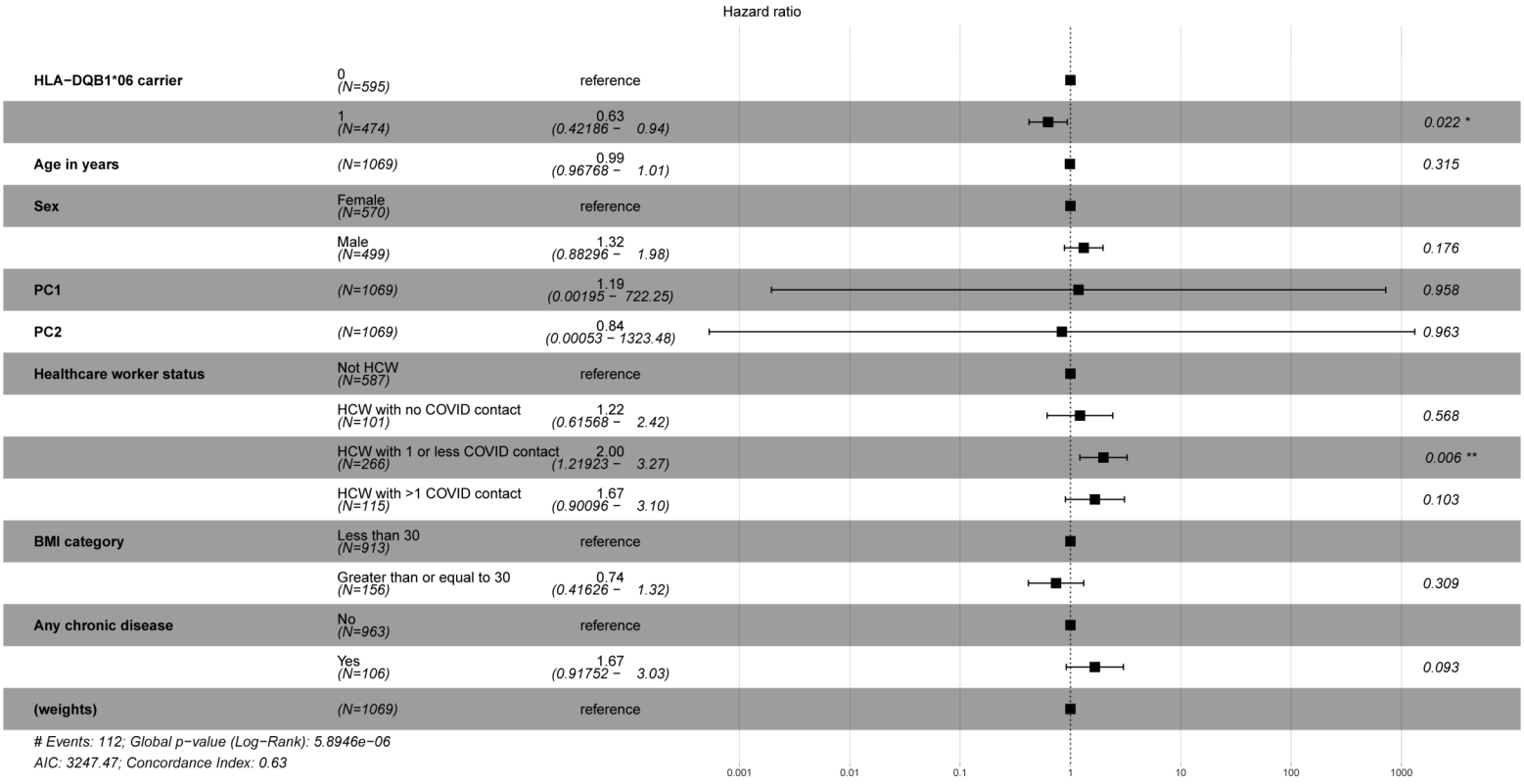

Global Populations
ChAdOx1 nCov−19 White
ChAdOx1 nCov−19 Non−White
British 1000G
Han Chinese 1000G
East African LWK 1000G
West African YRI 1000G
0.100
0.075
0.050
0.025
0.000
PC2
−0.02
−0.01
0.00
0.01
0.02
PC1

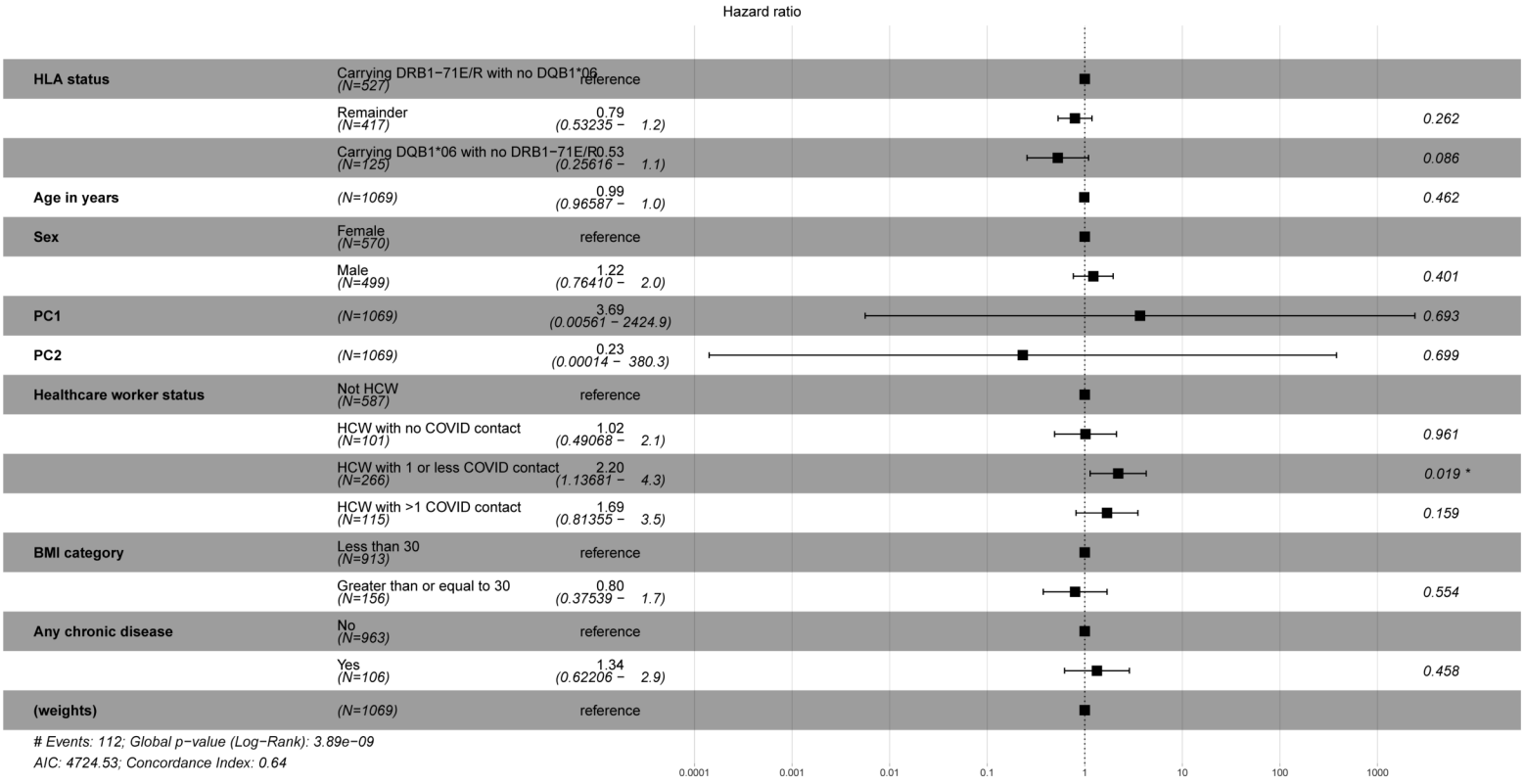

A
10.0
7.5
5.0
2.5
0.0
S −log(10) P−value
0.0
2.5
5.0
7.5
10.0
RBD −log(10) P−value
rs1130456
R2
1.00
0.75
0.50
0.25
B
0.6
0.3
0.0
−0.3
−0.6
S beta coefficient
−0.6
−0.3
0.0
0.3
0.6
RBD beta coefficient
R2
1.00
0.75
0.50
0.25

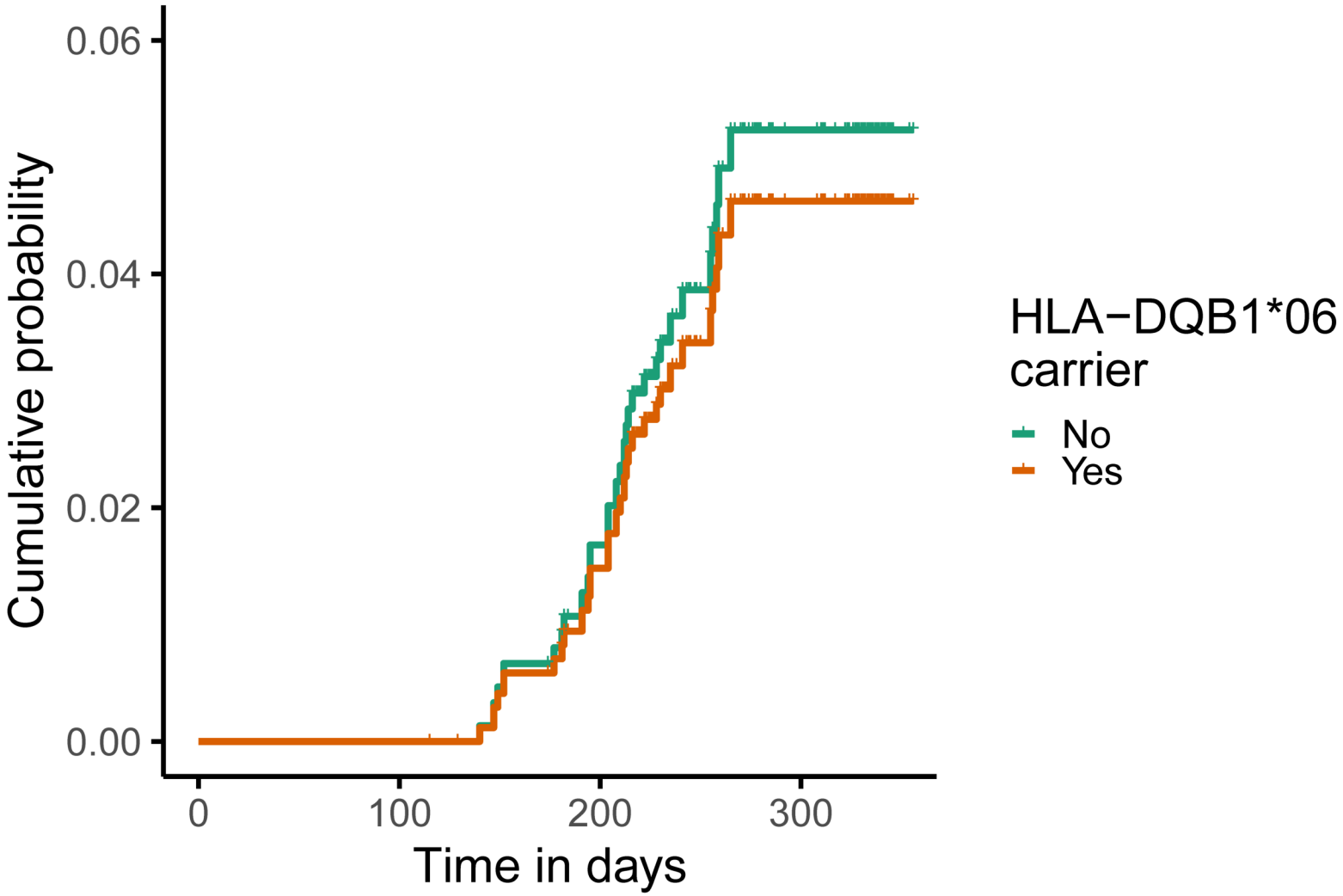

A
density
Inverse normalised anti−RBD
B
density
Inverse normalised anti−RBD
assay
MSD
PPD
C
$-\log_{10}(p)$
Chromosome
D
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$


A
B
C
D
E
F
$-\log_{10}(p)$
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
Chromosome
1 2 3 4 5 6 7 8 9 11 13 15 18 21

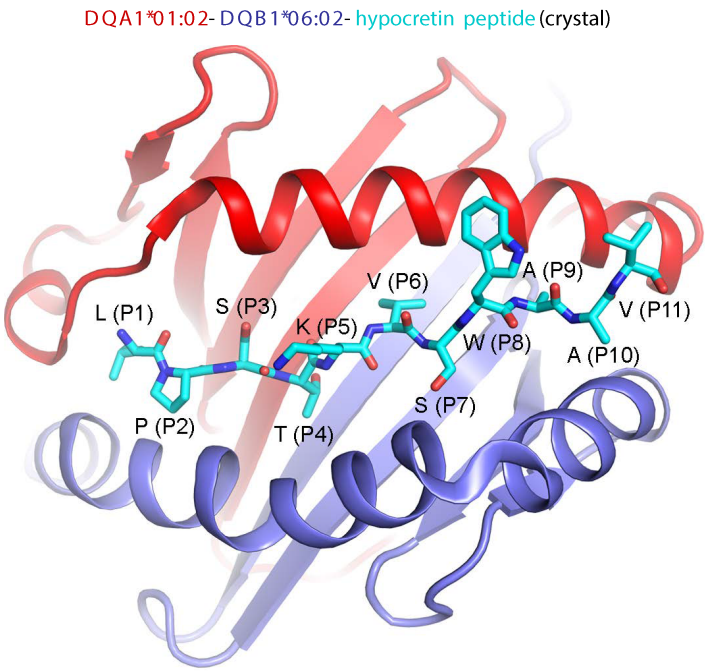

COMCOV vaccine trial; n = 666
COMCOV2 vaccine trial; n = 894
COV006 vaccine trial; n = 208
1768 samples submitted for genotyping
1 failed to produce scan
45 failed genotyping quality control
1722 samples successfully genotyped
45 samples failed data quality control
1677 vaccinees available for analysis
Primed with ChAdOx1 nCov-19; n= 931
Primed with BNT162b2; n= 746


Hazard ratio
HLA−DQB1*06 carrier
0 (N=595) reference
1 (N=474) 0.63 (0.42186 − 0.94) 0.022 *
Age in years (N=1069) 0.99 (0.96768 − 1.01) 0.315
Sex
Female (N=570) reference
Male (N=499) 1.32 (0.88296 − 1.98) 0.176
PC1 (N=1069) 1.19 (0.00195 − 722.25) 0.958
PC2 (N=1069) 0.84 (0.00053 − 1323.48) 0.963
Healthcare worker status
Not HCW (N=587) reference
HCW with no COVID contact (N=101) 1.22 (0.61568 − 2.42) 0.568
HCW with 1 or less COVID contact (N=266) 2.00 (1.21923 − 3.27) 0.006 **
HCW with >1 COVID contact (N=115) 1.67 (0.90096 − 3.10) 0.103
BMI category
Less than 30 (N=913) reference
Greater than or equal to 30 (N=156) 0.74 (0.41626 − 1.32) 0.309
Any chronic disease
No (N=963) reference
Yes (N=106) 1.67 (0.91752 − 3.03) 0.093
(weights) (N=1069) reference
Events: 112; Global p−value (Log−Rank): 5.8946e−06
AIC: 3247.47; Concordance Index: 0.63
0.001
0.01
0.1
1
10
100
1000

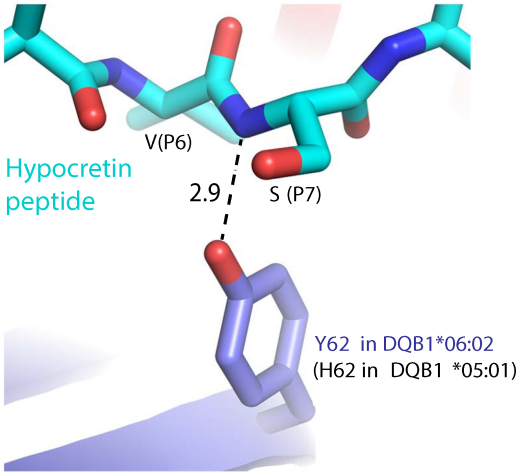

Hazard ratio
HLA status
Carrying DRB1−71E/R with no DQB1*06 (N=527) reference
Remainder (N=417) 0.79 (0.53235 − 1.2) 0.262
Carrying DQB1*06 with no DRB1−71E/R (N=125) 0.53 (0.25616 − 1.1) 0.086
Age in years (N=1069) 0.99 (0.96587 − 1.0) 0.462
Sex
Female (N=570) reference
Male (N=499) 1.22 (0.76410 − 2.0) 0.401
PC1 (N=1069) 3.69 (0.00561 − 2424.9) 0.693
PC2 (N=1069) 0.23 (0.00014 − 380.3) 0.699
Healthcare worker status
Not HCW (N=587) reference
HCW with no COVID contact (N=101) 1.02 (0.49068 − 2.1) 0.961
HCW with 1 or less COVID contact (N=266) 2.20 (1.13681 − 4.3) 0.019 *
HCW with >1 COVID contact (N=115) 1.69 (0.81355 − 3.5) 0.159
BMI category
Less than 30 (N=913) reference
Greater than or equal to 30 (N=156) 0.80 (0.37539 − 1.7) 0.554
Any chronic disease
No (N=963) reference
Yes (N=106) 1.34 (0.62206 − 2.9) 0.458
(weights) (N=1069) reference
Events: 112; Global p−value (Log−Rank): 3.89e−09
AIC: 4724.53; Concordance Index: 0.64
0.0001
0.001
0.01
0.1
1
10
100
1000


0.06
0.04
0.02
0.00
Cumulative probability
0
100
200
300
Time in days
HLA−DQB1*06
carrier
No
Yes

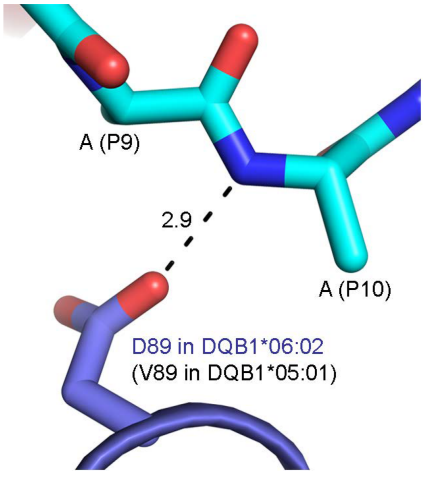
A


DQA1*01:02- DQB1*06:02- hypocretin peptide (crystal)
L (P1)
S (P3)
V (P6)
A (P9)
V (P11)
K (P5)
W (P8)
A (P10)
S (P7)
P (P2)
T (P4)

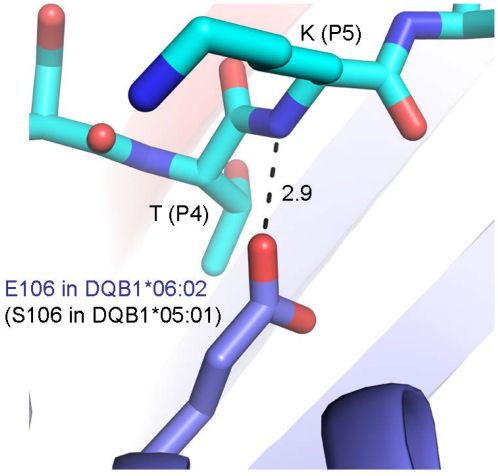
B


V(P6)
Hypocretin peptide
2.9
S (P7)
Y62 in DQB1*06:02
(H62 in DQB1 *05:01)

C


A (P9)
2.9
A (P10)
D89 in DQB1*06:02
(V89 in DQB1*05:01)

D


K (P5)
2.9
T (P4)
E106 in DQB1*06:02
(S106 in DQB1*05:01)

nature portfolio

Corresponding author(s): Mentzer

Last updated by author(s): Sep 23, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	PLINK (version 1.9, https://www.cog-genomics.org/plink/) Multi-Ethnic HLA reference panel (version 1.0, 2021, https://imputationserver.readthedocs.io/en/latest/reference-panels/)
Data analysis	PLINK (version 1.9, https://www.cog-genomics.org/plink/) DosageConverter (version 1.0.4, https://genome.sph.umich.edu/wiki/DosageConvertor) PHASE (version 2.1.1, https://stephenslab.uchicago.edu/phase/download.html) AlphaFold (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) PyMOL Molecular Graphics System (version 2.3.2, https://pymol.org/2/) GCTA (version 1.24.4, https://yanglab.westlake.edu.cn/software/gcta/#Overview) R (version 4.1.1; packages: ipw, ggplot2, survminer, survival, perm,) R (version 3.6.2; GenABEL package)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	The biological attribute of sex was used in this study. The justification for this is that biological sex was used as sample quality control metric, and the sex attribute used in analysis was derived from the genetic data.
Population characteristics	**Discovery cohort** The participants were enrolled in phase 1/2 (COV001) or phase 2/3 (COV002) randomized single-blind ChAdOx1 nCoV-19 (AZD1222) vaccine multi-centre efficacy trials conducted across multiple sites within the United Kingdom (NCT04324606, NCT04400838). In brief, following written informed consent adults aged 18 years and older were randomly assigned to receive either intramuscular ChAdOx1 nCoV-19 (AZD1222) or a control vaccine (MenACWY), to assess the safety and efficacy of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2. ** Replication cohort ** The replication cohort was comprised of participants from three COVID-19 vaccine trials conducted across several sites within the United Kingdom. Two of these trials (COMCOV and COMCOV2) were in adults aged 50 years and older, randomised to receive homologous or heterologous two-dose schedules of either intramuscular ChAdOx1 nCoV-19, mRNA vaccines (BNT162b2 or mRNA-1273) or a nanoparticle vaccine (NVX-CoV2373). The other trial (COV006) was in children aged 6-17 who were randomised to receive either intramuscular ChAdOx1 nCoV-19 or a control vaccine (capsular group B meningococcal vaccine, 4CMenB),
Recruitment	The participants were recruited from COVID-19 vaccine trials conducted across several sites within the United Kingdom. They were randomised into vaccine groups.
Ethics oversight	The trials were conducted according to the principles of Good Clinical Practice and approved by the South Central Berkshire Research Ethics Committee (20/SC/0145, 20/SC/0179, 21/SC/0119, 21/SC/0022 and 21/SC/0054) and the UK regulatory agency (the Medicines and Healthcare products Regulatory Agency). We obtained informed consent from participants (or parent/guardians) for genetic data to be analysed, this was an optional consent in their clinical study consent forms.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No formal sample-size calculation were conducted for this genetic analysis, rather sample size was based on the opportunistic sample availability from COVID-19 vaccine trials (genetic analysis was an exploratory endpoint in these studies).
Data exclusions	Genetic samples were removed if they had high genotyping missingness, genetic sex did not match reported sex or they had high heterozygosity — these are metrics of poor sample or genotyping quality. Individuals were excluded if they had estimates ≥0.9 identity by descent, excluding the individual with the highest SNP missingness rate from each pair preferentially — to remove duplicated samples.
Replication	An independent cohort was used to replicate findings.
Randomization	Participants were randomised into vaccine groups.

Blinding	Participants were blinded to their vaccine allocation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NCT04324606; NCT04400838; ISRCTN registry, 69254139 (EudraCT 2020–005085–33); ISRCTN Number: 27841311 (EudraCT Number: 2021-001275-16), ISRCTN Number: 15638344 (EudraCT number:2020-005765-13)
Study protocol	COV001 (https://clinicaltrials.gov/ct2/show/NCT04324606); COV002 (https://clinicaltrials.gov/ct2/show/NCT04400838); COMCOV (https://comcovstudy.org.uk/files/com-covprotocolv9220-sept-2021cleanpdf); COMCOV2 (https://comcovstudy.org.uk/files/com-cov2protocolv6122-sep-2021cleanpdf); COV006 (https://www.isrctn.com/ISRCTN15638344)
Data collection	The participants were recruited from COVID-19 vaccine trials conducted across several sites within the United Kingdom. Data between April 2020 and January 2022 were included in this manuscript.
Outcomes	Genetic analyses were included as exploratory outcome in the clinical trials from which these data were generated.