The landscape of host genetic factors involved in infection to common viruses and SARS-CoV-2

Linda Kachuri[1]*, Stephen S. Francis[1,2,3,4]*, Maike Morrison[5,6], Yohan Bossé[7], Taylor B. Cavazos[8], Sara R. Rashkin[1,9], Elad Ziv[3,10,11], John S. Witte[1,3,11,12]

* Authors contributed equally to this work

Affiliations:

1. Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, USA
2. Department of Neurological Surgery, University of California San Francisco, San Francisco, USA
3. Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, USA
4. Weill Institute for Neurosciences, University of California San Francisco, San Francisco, USA
5. Summer Research Training Program, Graduate Division, University of California San Francisco, San Francisco, USA
6. Department of Mathematics, The University of Texas at Austin, Austin, USA
7. Institut universitaire de cardiologie et de pneumologie de Québec, Department of Molecular Medicine, Université Laval, Quebec City, Canada
8. Program in Biological and Medical Informatics, University of California San Francisco, San Francisco, USA
9. Center for Applied Bioinformatics, St. Jude Children's Research Hospital, Memphis, USA
10. Department of Medicine, University of California, San Francisco, San Francisco, USA
11. Institute for Human Genetics, University of California San Francisco, San Francisco, USA
12. Department of Urology, University of California San Francisco, San Francisco, USA

Corresponding Authors:

Stephen S. Francis

Department of Neurological Surgery
Helen Diller Family Comprehensive Cancer Center
University of California, San Francisco
1450 3rd Street, Room 482, San Francisco, CA 94158
Email: stephen.francis@ucsf.edu

John S. Witte

Department of Epidemiology and Biostatistics
Helen Diller Family Comprehensive Cancer Center
University of California, San Francisco
1450 3rd Street, Room 388, San Francisco, CA 94158
Email: jwitte@ucsf.edu

ABSTRACT

Introduction: Humans and viruses have co-evolved for millennia resulting in a complex host genetic architecture. Understanding the genetic mechanisms of immune response to viral infection provides insight into disease etiology and informs public health interventions.

Methods: We conducted a comprehensive study linking germline genetic variation and gene expression with antibody response to 28 antigens for 16 viruses using serological data from 7924 participants in the UK Biobank cohort. Using test results from 2010 UK Biobank subjects, we also investigated genetic determinants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection.

Results: Signals in human leukocyte antigen (HLA) class II region dominated the landscape of viral antibody response, with 40 independent loci and 14 independent classical alleles, 7 of which exhibited pleiotropic effects across viral families. Genome-wide association analyses discovered 7 novel genetic loci associated with viral antibody response ($P<5.0\times10^{-8}$), including *FUT2* (19q13.33) for human polyomavirus BK (BKV), *STING1* (5q31.2) for Merkel cell polyomavirus (MCV), as well as *CXCR5* (11q23.3) and *TBKBP1* (17q21.32) for human herpesvirus 7. Transcriptome-wide association analyses identified 114 genes associated with response to viral infection, 12 outside of the HLA region, including *ECSCR*: $P=5.0\times10^{-15}$ (MCV), *NTN5*: $P=1.1\times10^{-9}$ (BKV), and *P2RY13*: $P=1.1\times10^{-8}$ (Epstein-Barr virus nuclear antigen). We also demonstrated pleiotropy between viral response genes and complex diseases, such as *C4A* expression in varicella zoster virus and schizophrenia. Finally, our analyses of SARS-CoV-2 revealed the first genome-wide significant infection susceptibility signal in *EHF*, an epithelial-specific transcriptional repressor implicated in airway disease. Targeted analyses of expression quantitative trait loci suggest a possible role for tissue-specific *ACE2* expression in modifying SARS-CoV-2 susceptibility.

Conclusions: Our study confirms the importance of the HLA region in host response to viral infection and elucidates novel genetic determinants of host-virus interaction. Our results may have implications for complex disease etiology and COVID-19.

KEY WORDS

Infection, virus, serology, antigen, antibody, immune response, human leukocyte antigen (HLA), polyomavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), genome-wide association study (GWAS), transcriptome-wide association study (TWAS)

INTRODUCTION

Viruses have been infecting cells for a half a billion years[1]. During our extensive co-evolution viruses have exerted significant selective pressure on humans; overtly during fatal outbreaks, and covertly through cryptic immune interaction when a pathogen remains latent. The recent pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) highlights the paramount public health need to understand human genetic variation in response to viral challenge. Clinical variation in COVID-19 severity and symptomatic presentation may be due to differences host genetic factors relating to immune response[2]. Furthermore, many common infections are cryptically associated with a variety of complex illnesses, especially those with an immunologic component, from cancer to autoimmune and neurologic conditions[3-5]. Despite their broad health relevance, few large-scale genome-wide association studies (GWAS) have been conducted on viral infections[6-10]. Understanding the genetic architecture of immunologic response to viruses may therefore provide new insight into etiologic mechanisms of diverse complex diseases.

Several common viruses exert a robust cell mediated and humoral immune response that bi-directionally modulate the balance between latent and lytic infection. Previous studies have identified associations between host polymorphisms in genes relating to cell entry, cytokine production, and immune response and a variety of viruses[11]. Other studies have demonstrated a strong heritable component (32-48%) of antibody response[12]. Previous investigations have implicated human leucocyte antigen (HLA) class I and II in the modulation of immune response to diverse viral antigens[7,13]. Furthermore, during the 2003 severe acute respiratory syndrome (SARS) outbreak, caused by a betacoronavirus related to SARS-CoV-2, an association with infection severity was reported for the HLA-B*46:01 allele in East Asian patients[14].

In this study we utilize data from the UK Biobank (UKB) cohort[15] to evaluate the relationship between host genetics and antibody response to 28 antigens for 16 viruses that have been linked to cancer and neurodegenerative diseases[16], as well as SARS-CoV-2 infection. We conduct integrative genome-wide and transcriptome-wide analyses of antibody response and positivity to viral antigens, which elucidate novel genetic underpinnings of viral infection and immune response.

METHODS

Study Population and Phenotypes

The UK Biobank (UKB) is a population-based prospective cohort of over 500,000 individuals aged 40-69 years at enrollment in 2006-2010 who completed extensive questionnaires, physical assessments, and provided blood samples[15]. Analyses were restricted to individuals of predominantly European ancestry based on self-report and after excluding samples with any of the first two genetic ancestry principal components (PCs) outside of 5 standard deviations (SD) of the population mean. We removed samples with discordant self-reported and genetic sex, samples with call rates <97% or heterozygosity >5 SD from the mean, and one sample from each pair of first-degree relatives identified using KING[17].

Of the 413,810 European ancestry individuals available for analysis, a total of 7948 had serological measures. A multiplex serology panel was performed over a 2-week period using previously developed methods[18,19] that have been successfully applied in epidemiological studies[7,20]. Details of the serology methods and assay validation performance are described in Mentzer et al.[16] Briefly, multiplex serology was performed using a bead-based glutathione S-transferase (GST) capture assay with glutathione-casein coated fluorescence-labelled polystyrene beads and pathogen-specific GST-X-tag fusion proteins as antigens[16]. Each antigen was loaded onto a distinct bead set and the beads were simultaneously presented to primary serum antibodies at serum dilution 1:1000[16]. Immunocomplexes were quantified using a Luminex 200 flow cytometer, which produced Median Fluorescence Intensities (MFI) for each antigen. The serology assay showed adequate performance, with a median coefficient of variation (CV) of 17% across all antigens and 3.5% among seropositive samples only[16].

A total of 5356 records for SARS-CoV-2 test results between March 16, 2020 and May 3, 2020 was provided by Public Health England for 3002 UKB participants. Results based on serum, skin swabs, or specimens of unknown origin (799 records) were excluded due to potentially unreliable detection[2]. Using the remaining 4557 records (2622 participants), cases were classified as individuals with at least one positive SARS-CoV-2 test result based on respiratory specimens. After restricting to individuals of predominantly European ancestry, 676 cases and 1334 controls results remained.

Genome-Wide Association Analysis

We evaluated the relationship between constitutive genetic variants across the genome and antigens or SARS-CoV-2 status using PLINK 2.0 (October 2017 version). Participants were genotyped on the Affymetrix Axiom UK Biobank array (89%) or the UK BiLEVE array (11%)[15] with genome-wide imputation performed using the Haplotype Reference Consortium data and the merged UK10K and 1000 Genomes phase 3 reference panels[15]. We excluded variants out of Hardy-Weinberg equilibrium at $p<1\times10^{-5}$, call rate <95% (alternate allele dosage required to be within 0.1 of the nearest hard call to be non-missing), imputation quality INFO<0.30, and MAF<0.01.

Seropositivity for each antigen was determined using established cut-offs based on prior validation work[16]. The primary antigen GWAS focused on those with seroprevalence of ≥20%, and investigated continuous phenotypes (MFI values), corresponding to the magnitude of antibody response or seroreactivity among seropositive individuals. MFI values were transformed to standard normal distributions using ordered quantile normalization[21].

The antigen GWAS entailed linear regression of MFI z-scores on genotype. These regression models adjusted for age at enrollment, sex, body-mass index (BMI), socioeconomic status (Townsend deprivation index), the presence of any autoimmune and/or inflammatory conditions, genotyping array, serology assay date, quality control flag indicating sample spillover or an extra freeze/thaw cycle, and the top 10 genetic ancestry principal components (PC's). Autoimmune and chronic inflammatory conditions were identified using the following primary and secondary diagnostic ICD-10 codes in Hospital Episode Statistics. Individuals diagnosed with any immunodeficiency (ICD-10 D80-89, n=24) were excluded from all analyses. For all antigens with at least 100 seropositive (or seronegative for pathogens with ubiquitous exposure) individuals, GWAS of discrete seropositivity phenotypes was undertaken using logistic regression, adjusting for the same covariates listed above.

The functional relevance of the lead GWAS loci for antibody response was assessed using in-silico functional annotation analyses based on Combined Annotation Dependent Depletion (CADD)[22] scores and RegulomeDB 2.0[23], and by leveraging external datasets, such as GTEx v8, DICE (Database of Immune Cell Expression)[24], and the Human Plasma Proteome Atlas[25,26].

Pleiotropic Associations with Disease

We explored pleiotropic associations between lead variants influencing antibody levels and several chronic diseases with known or hypothesized viral risk factors. Associations with selected cancers were obtained from a cancer pleiotropy meta-analysis of the UK Biobank and Genetic Epidemiology Research on Aging cohorts[27]. Summary statistics for the schizophrenia GWAS of 33,640 cases and 43,456 controls by Lam et al.[28] were downloaded from the Psychiatric Genomics Consortium. Association p-values were obtained from the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site for the GWAS by Jun et al.[29], which included 17,536 cases and 53,711 controls. Associations with $p<7.3\times10^{-4}$ were considered statistically significant after correction for the number of variants and phenotypes tested.

HLA Regional Analysis

For phenotypes displaying a genome-wide significant signal in the HLA region, independent association signals were ascertained using two complementary approaches: clumping and conditional analysis. Clumping was performed on all variants with $P<5\times10^{-8}$ for each phenotype, as well as across phenotypes. Clumps were formed around sentinel variants with the lowest p-value and all other variants with LD $r^2>0.05$ within a ±500kb window were assigned to that variant's clump.

Next, we conducted conditional analyses using a forward stepwise strategy to identify statistically independent signals within each type of variant (SNP/indel or classical HLA allele). A total of 38,655 SNPs/indels on chromosome 6 (29,600,000 – 33,200,000 bp) were extracted to conduct regional analyses. Classical HLA alleles were imputed for UKB participants at 4-digit resolution using the HLA*IMP:02 algorithm applied to diverse population reference panels[15]. Imputed dosages were available for 362 classical alleles in 11 genes: *HLA-A*, *HLA-B*, and *HLA-C* (class I); *HLA-DRB5*, *HLA-DRB4*, *HLA-DRB3*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, and *HLA-DPB1* (class II). Analyses were restricted to 101 common alleles (frequency ≥ 0.01) in 413,810 European ancestry participants. Linear regression models were adjusted for the same set of covariates as the GWAS.

For each antigen response phenotype, we identified genetic variants (SNPs/indels or classical HLA alleles) with the lowest p-value and performed forward iterative conditional regression to identify other independent

signals, until no associations with a conditional p-value (P_{cond})<5×10^{-8} remained. Conditional analyses of classical HLA alleles were restricted to Bonferroni-significant associations corrected for 101 alleles ($P<5\times10^{-4}$) in the unconditional analysis. We also assessed the independence of associations across different types of genetic variants by including conditionally independent HLA alleles as covariates in the SNP-based analysis.

Genetic Associations With SARS-CoV-2 Status

GWAS of SARS-CoV-2 status was conducted using logistic regression with adjustment for age at assessment, sex, BMI, Townsend index, specimen origin (confirmed inpatient vs. unknown), genotyping array, the top 10 PC's, and smoking status. We also examined associations with 101 common classical HLA alleles. Based on recent evidence that SARS-CoV-2 uses the receptor encoded by the *ACE2* gene for cell entry[30], we evaluated associations between significant *ACE2* eQTLs in any tissue ($q_{FDR}<0.05$) identified in GTEx and SARS-CoV-2 test status. The overall relationship between *ACE2* expression and SARS-CoV-2 was quantified using a linear regression model with log(OR) for testing positive as the outcome and eQTL effect size as the predictor, clustered by tissue type. We further investigated this relationship using genotyping data and *ACE2* expression in lung tissues from 409 subjects that underwent lung cancer surgery at the Institut universitaire de cardiologie et de pneumologie de Québec (IUCPQ)[31].

Transcriptome-Wide Association Analysis

Gene transcription levels were imputed and analyzed using the MetaXcan approach[32], applied to GWAS summary statistics for quantitative antigen phenotypes. For imputation, we used biologically informed MASHR-M prediction models[33] based on GTEx v8 with effect sizes computed using MASHR (Multivariate Adaptive Shrinkage in R)[34] for variants fine-mapped with DAP-G (Deterministic Approximation of Posteriors)[35,36]. An advantage of this approach is that MASHR effect sizes are smoothed by taking advantage of the correlation in cis-eQTL effects across tissues. For each antigen, we performed a transcriptome-wide association study (TWAS) using gene expression levels in whole blood. Statistically significant associations for each gene were determined based on Bonferroni correction for the number of genes tested.

We also examined gene expression profiles in tissues that represent known infection targets or related pathologies. Human herpesviruses and polyomaviruses are neurotropic and have been implicated in several neurological conditions[37,38], therefore we considered gene expression in the frontal cortex. For Epstein-Barr virus (EBV) antigens additional models included EBV-transformed lymphocytes. Merkel cell polyomavirus (MCV) is a known cause of Merkel cell carcinoma[39], a rare but aggressive type of skin cancer, therefore we examined transcriptomic profiles in skin tissues for MCV only.

Pathways represented by genes associated with antibody response to viral antigens were summarized by conducting enrichment analysis using curated Reactome gene sets and by examining protein interaction networks using the STRING database[40]. Significantly associated TWAS genes were grouped by virus family (herpesviruses vs. polyomaviruses) and specificity of association (multiple antigens vs. single antigen). Protein interaction analyses were restricted to genes associated with more than one antigen. We considered unidirectional functional interactions with confidence scores ≥400 (medium).

RESULTS

A random sample of the participants representative of the full UKB cohort was assayed using a multiplex serology panel[16]. We analyzed data from 7924 participants of predominantly European ancestry, described in **Supplementary Table 1**. Approximately 90% of individuals were seropositive for herpes family viruses with ubiquitous exposure: EBV (EBV EA-D: 86.2% to ZEBRA: 91.2%), Human Herpesvirus 7 (HHV7 94.8%), and Varicella Zoster Virus (VZV 92.3%). Seroprevalence was somewhat lower for cytomegalovirus (CMV), ranging between 56.5% (CMV pp28) and 63.3% (CMV pp52), and Herpes Simplex virus-1 (HSV1 69.3%). Human polyomavirus BKV was more prevalent (95.3%) compared to other polyomaviruses, Merkel cell polyoma virus (MCV 66.1%) and polyomavirus JC (JCV) (56.6%). Less common infections included HSV-2 (15.2%), HPV16 (E6 and E7 oncoproteins: 4.7%), HPV18 (2.4%), Human T-cell lymphotropic virus type 1 (HTLV1, 1.6%), Hepatitis B (HBV, 1.6%), and Hepatitis C (HCV, 0.3%).

Baseline characteristics of 3002 UKB participants with SARS-CoV-2 test results up to May 3, 2020 are described in **Supplementary Table 2**. Compared to the full UKB cohort, individuals who were tested for SARS-CoV-2 were more likely to be ever smokers (51.9% vs. 45.2%), had higher mean BMI (28.5 vs. 27.4

kg/m^2), higher levels of deprivation based on the Townsend index, and a higher burden of comorbidities based on the Charlson index (1 or more: 45.8% vs. 29.8%). Comparing within participants who were tested, those with at least one positive SARS-CoV-2 test result included a higher proportion of men (54.1% vs. 40.6%), former smokers (41.6% vs. 37.0%), and self-identified as non-white (12.2% vs. 8.1%). The two groups had similar distributions of other health-related characteristics. The final analytic dataset included 676 cases and 1334 controls of European ancestry with test results from respiratory samples.

Genetic Determinants of Response to Viral Infection

Results from our GWAS of antibody response phenotypes were dominated by signals in the HLA region, which were detected for all EBV antigens (EA-D, EBNA, p18, ZEBRA), CMV pp52, HSV1, HHV7, VZV, JCV and MCV (**Table 1; Supplementary Figure 1**). Most of the top-ranking HLA variants for each antigen were independent of those for other antigens (**Supplementary Figure 2**). Exceptions were moderate LD between lead variants for EBV ZEBRA and HSV1 (r^2=0.45), EBV EBNA and JCV (r^2=0.45), and HHV7 and MCV (r^2=0.44). Outside of the HLA region, genome-wide significant associations with seroreactivity were detected for: MCV at 3p24.3 (rs776170649, *LOC339862*: $P=1.7\times10^{-8}$) and 5q31.2 (rs7444313, *TMEM173* (also known as *STING1*): $P=2.4\times10^{-15}$); BKV at 19q13.3 (rs681343, *FUT2*: $P=4.7\times10^{-15}$) (**Figure 1**); EBV EBNA at 3q25.1 (rs67886110, *MED12L*: $P=1.3\times10^{-9}$); HHV-7 at 11q23.3 (rs75438046, *CXCR5*: $P=1.3\times10^{-8}$) and 17q21.3 (rs1808192, *TBKBP1*: $P=9.8\times10^{-9}$); and HSV-1 at 10q23.3 (rs11203123: $P=3.9\times10^{-8}$).

GWAS of discrete seropositivity phenotypes identified associations in HLA for EBV EA-D (rs2395192: OR=0.66, $P=4.0\times10^{-19}$), EBV EBNA (rs9268848: OR=1.60, $P=1.2\times10^{-18}$), EBV ZEBRA (rs17211342: 0.63, $P=1.6\times10^{-15}$), VZV (rs3096688: OR=0.70, $P=3.7\times10^{-8}$), JCV (rs9271147: OR=0.54, $P=1.3\times10^{-42}$), and MCV (rs17613347: OR=0.61, $P=1.2\times10^{-26}$) (**Supplementary Figure 1; Supplementary Table 3**). An association with susceptibility to MCV infection was also observed at 5q31.2 (rs1193730215, *ECSCR*: OR=1.26, $P=7.2\times10^{-9}$), with high LD (r^2=0.95) between seroreactivity and seropositivity lead variants.

Several significant associations were observed for antigens with <20% seroprevalence, which were not included in the GWAS of antibody response due to inadequate sample size (**Supplementary Table 3**). Significant associations with infection susceptibility were observed for HSV2 in 17p13.2 (rs2116443:

OR=1.28, $P=4.5\times10^{-8}$; *ITGAE*); HPV16 E6 and E7 oncoproteins in 6p21.32 (rs601148: OR=0.60, $P=3.3\times10^{-9}$; *HLA-DRB1*) and 19q12 (rs144341759: OR=0.383, $P=4.0\times10^{-8}$; *CTC-448F2.6*); and HPV18 in 14q24.3 (rs4243652: OR=3.13, $P=7.0\times10^{-10}$). Associations were also detected for Kaposi's sarcoma-associated herpesvirus (KSHV), HTLV1, HBV and HCV, including a variant in the *MERTK* oncogene (HCV Core rs199913364: OR=0.25, $P=1.2\times10^{-8}$).

Functional Characterization of GWAS Findings

In-silico functional analyses of the lead 17 GWAS variants identified enrichment for multiple regulatory elements (summarized in **Supplementary Table 4**). Three variants were predicted to be in the top 10% of deleterious substitutions in GRCh37 based on CADD scores >10: rs776170649 (MCV, CADD=15.61), rs139299944 (HHV7, CADD=12.15), and rs9271525 (JCV, CADD=10.73). Another HHV7-associated variant, rs1808192 (RegulomeDB rank: 1f), an eQTL and sQTL for *TBKBP1*, mapped to 44 functional elements for multiple transcription factors, including IKZF1, a critical regulator of lymphoid differentiation frequently mutated in B-cell malignancies.

Eleven sentinel variants were eQTLs and 8 were splicing QTLs in GTEx, with significant (FDR<0.05) effects across multiple genes and tissues (**Supplementary Figure 3**). The most common eQTL and sQTL targets included *HLA-DQA1*, *HLA-DQA2*, *HLA-DQB1*, *HLA-DQB2*, *HLA-DRB1*, and *HLA-DRB6*. Outside of HLA, rs681343 (BKV), a synonymous *FUT2* variant was an eQTL for 8 genes, including *FUT2* and *NTN5*. MCV variant in 5q31.2, rs7444313, was an eQTL for 7 genes, with concurrent sQTL effects on *TMEM173*, also known as *STING1* (stimulator of interferon response cGAMP interactor 1) and *CXXC5*. Gene expression profiles in immune cell populations from DICE[24] identified several cell-type specific effects that were not observed in GTEx. An association with *HLA-DQB1* expression in CD4+ T_H2 cells was observed for rs9273325, 6:31486158_GT_G was an eQTL for *ATP6V1G2* in naïve CD4+ T cells, and rs1130420 influenced the expression of 8 HLA class II genes in naïve B-cells and CD4+ T_H17 cells.

We identified 7 significant ($p<5.0\times10^{-8}$) protein quantitative trait loci (pQTL) for 38 proteins (**Supplementary Table 5**). Most of the pQTL targets were components of the adaptive immune response, such as the complement system (C4, CFB), chemokines (CCL15, CCL25), and defensin processing (Beta-defensin 19,

Trypsin-3). The greatest number and diversity of pQTL targets (n=16) was observed for rs681343, including BPIFB1, which plays a role in antimicrobial response in oral and nasal mucosa[41]; FUT3, which catalyzes the last step of Lewis antigen biosynthesis; and FGF19, part of the PI3K/Akt/MAPK signaling cascade that is dysregulated in cancer and neurodegenerative diseases[42].

Pleiotropic associations with disease outcomes

To contextualize the relevance of genetic loci involved in infection response, we explored associations with selected cancers, schizophrenia, and that have a known or suspected viral etiology (**Supplementary Table 6**). The strongest secondary signal was observed for rs9273325, which was negatively associated with VZV antibody response and positively associated with schizophrenia susceptibility (OR=1.13, $P=4.3\times10^{-15}$). Other significant (Bonferroni $P<7.4\times10^{-4}$) associations with schizophrenia were detected for HSV1 (rs1130420: OR=1.06, $P=1.8\times10^{-5}$), EBV EA-D (rs2647006: OR=0.96, $P=2.7\times10^{-4}$), BKV, (rs681343: OR=0.96, $P=2.5\times10^{-4}$) and JCV (rs9271525: OR=1.06, $P=6.8\times10^{-5}$). Inverse associations with hematologic cancers were observed for HSV1 (rs1130420: OR=0.89, $P=3.5\times10^{-6}$), VZV (rs9273325: OR=0.88, $P=4.4\times10^{-5}$), and EBV EBNA (rs9269233: OR=0.88, $P=2.7\times10^{-4}$) variants. HSV1 antibody response was also linked to Alzheimer's disease (rs1130420: $P=1.2\times10^{-4}$).

Regional HLA Associations

Associations within the HLA region were refined by identifying independent (LD $r^2<0.05$ within ±500kb) sentinel variants with $P<5.0\times10^{-8}$ for each antigen response phenotype (**Supplementary Table 7**). Clumping seropositivity associations with respect to lead antibody response variants did not retain any loci, suggesting non-independence in signals for infection and reactivity for the same antigen. For this reason, all subsequent analyses focus on seroreactivity phenotypes. Clumping across phenotypes to assess the independence of HLA associations for different antigens identified 40 independent sentinel variants: EBV EBNA (12), VZV (11), EBV ZEBRA (8), EBV p18 (5), MCV (3), and EBV EA-D (1) (**Supplementary Table 9**). No LD clumps were anchored by variants detected for CMV pp52, HHV7, HSV1, or JCV, suggesting that the HLA signals for these antigens are captured by lead loci for other phenotypes. The largest region with the lowest p-value was anchored by rs9274728 ($P=4.7\times10^{-67}$) near *HLA-DQB1*, originally detected for

EBV ZEBRA. Of the 11 VZV-associated variants, the largest clump was formed around rs4990036 (P=4.5×10^{-26}) in *HLA-B*.

Iterative conditional analyses adjusting for the HLA SNP/indel with the lowest p-value were performed until no variants remained with P_{cond}<5.0×10^{-8}. Additional independent variants were identified for EBV EBNA (rs139299944, rs6457711, rs9273358, rs28414666, rs3097671), EBV ZEBRA (rs2904758, rs35683320, rs1383258), EBV p18 (rs6917363, rs9271325, rs66479476), and MCV (rs148584120, rs4148874) (**Figure 2; Supplementary Table 8**). For CMV pp52, HHV7, HSV1, JCV, and VZV, the regional HLA signal was captured by the top GWAS variant (**Figure 2; Supplementary Table 8**).

Next, we tested 101 classical HLA alleles and performed analogous iterative conditional analyses. Significant associations across viruses were predominantly observed for class II HLA alleles. Five statistically independent signals were identified for antibody response to EBV ZEBRA (DRB4*99:01: P=1.4×10^{-46}; DQB1*04:02: P_{cond}=1.0×10^{-19}; DRB1*04:04: P_{cond}=1.1×10^{-18}; DQA1*02:01: P_{cond}=1.1×10^{-10}, A*03:01: P_{cond}=1.9×10^{-8}) and EBV EBNA (DRB5*99:01: P=8.7×10^{-30}; DRB3*02:02: P_{cond}=6.8×10^{-30}; DQB1*02:01: P_{cond}=3.6×10^{-12}; DRB4*99:01: P_{cond}=8.3×10^{-17}; DPB1*03:01: P_{cond}=4.7×10^{-14}) (**Figure 3; Supplementary Tables 10-11**). Fewer independent alleles were observed for EBV p18 (DRB5*99:01: P=1.7×10^{-22}; DRB1*04:04: P_{cond}=1.3×10^{-18}) (**Figure 3; Supplementary Tables 12**).

DQB1*02:01 was the only independently associated allele for EBV EA-D (β=-0.154, P=8.4×10^{-11}) and HSV1 (β=0.145, P=2.8×10^{-8}), although its effects were in opposite directions for each antigen (**Supplementary Table 13**). For VZV, associations with 16 classical alleles were accounted for by DRB1*03:01 (P=7.3×10^{-26}). JCV shared the same lead allele as EBV EBNA and EBV p18 (DRB5*99:01: P=1.2×10^{-21}) (**Supplementary Table 13**). Four conditionally independent signals were identified for MCV (DQA1*01:01: P=1.1×10^{-15}; DRB1*04:04: P_{cond}=3.0×10^{-11}; A*29:02: P=1.0×10^{-11}; DRB1*15:01: P=3.7×10^{-12}) (**Figure 3; Supplementary Table 14**). Lastly, we integrated associations across variant types by including conditionally independent HLA alleles as covariates in the SNP-based analysis. With the exception of EBV antigens and HHV7, the classical alleles captured all genome-wide significant SNP signals (**Supplementary Figure 5**).

Genetic Associations With SARS-CoV-2 Status

We detected a significant association with having a positive SARS-CoV-2 test for rs286914 (OR=1.52, $P=2.0\times10^{-8}$) in *EHF* (ETS homologous factor) on 11p13, a gene part of the ETS transcription factor family characterized by epithelial-specific expression (**Table 2; Supplementary Table 4**). Sensitivity analyses using remaining UKB subjects (n=411,795) further demonstrate that rs286914 confers an increased risk of SARS-CoV-2 infection (OR=1.22, $P=5.0\times10^{-4}$) and reduced likelihood of having a negative test (OR=0.81, $P=7.2\times10^{-6}$). However, rs286914 was not associated with being tested for SARS-CoV-2 ($P=0.10$). We also identified two classical HLA alleles inversely associated with testing positive for SARS-CoV-2 (**Table 2**): DQA1*03:01 (OR=0.80, $P=0.012$) and A*31:01 (OR=0.53, $P=0.019$). The former was also associated with antibody response to five viral antigens (Bonferroni-corrected $P<5\times10^{-4}$).

We observed tissue-dependent and inconsistent associations between SARS-CoV-2 susceptibility and *ACE2* expression. A total of 288 significant ($q_{FDR}<0.05$) eQTLs were available in GTEx v8 for 9 tissue types (adipose, artery, brain, breast, nerve, muscle, pituitary, prostate, testis), but no significant eQTLs were available for lung. After restricting to independent eQTLs (LD $r^2<0.10$) available in our GWAS (n=10), we identified two variants inversely associated SARS-CoV-2: rs11798628 (OR=0.88, $P=5.1\times10^{-3}$) and rs4830974 (OR=0.77, $P=0.026$) (**Table 2**). There was negative overall association (-0.293, $P=1.3\times10^{-3}$) between *ACE2* expression and testing positive for SARS-CoV-2 (**Supplementary Figure 6**). However, based on eQTLs in lung tissue (LD $r^2<0.10$; $q_{FDR}<0.05$), a weak positive relationship was observed (0.098, $P=0.13$) with infection status. None of the significant lung eQTLs were associated with testing positive for SARS-CoV-2, but we identified a potential risk variant (rs5934251: OR=2.12, $P=5.3\times10^{-4}$) among nominally associated lung *ACE2* eQTLs (**Table 2**). We also note limited overlap in *ACE2* eQTLs across tissues, with only 6 GTEx eQTLs from other tissues being associated with lung *ACE2* expression at $P<0.05$.

TWAS of Genes Involved in Antibody Response

Based on known targets of infection or related pathologies, we considered expression in the frontal cortex (**Supplementary Table 15**), EBV-transformed lymphocytes for EBV antigens (**Supplementary Table 16**), and skin for MCV (**Supplementary Table 17**). Concordance across tissues was summarized using Venn

316 diagrams (**Figure 4; Supplementary Figure 5**). TWAS identified 114 genes significantly associated
317 (P_{TWAS}<4.2×10^{-6}) with antibody response in at least one tissue, 54 of which were associated with a single
318 phenotype, while 60 influenced seroreactivity to multiple antigens. We also include results for 92 additional
319 suggestively (P_{TWAS}<4.5×10^{-5}) associated genes.

320 The TWAS results included a predominance of associations in HLA class II genes. Some of the strongest
321 overall associations were observed for *HLA-DRB5* (EBV ZEBRA: P_{cortex}=4.2×10^{-45}) and *HLA-DRB1* (EBV
322 EBNA: P_{cortex}=6.7×10^{-39}; EBV ZEBRA: P_{cortex}=3.3×10^{-33}; JCV: P_{cortex}=6.5×10^{-14}; EBV p18: P_{cortex}=2.2×10^{-12}).
323 Increased expression of *HLA-DQB2* was positively associated with antibody response to EBV ZEBRA
324 (P_{blood}=7.6×10^{-19}), JCV (P_{blood}=9.9×10^{-10}), VZV (P_{blood}=7.0×10^{-9}), HHV7 (P_{blood}=7.3×10^{-8}), and HSV1
325 (P_{blood}=3.3×10^{-7}), but negatively associated with EBV EBNA (P_{blood}=3.6×10^{-34}) and EBV p18 (P_{blood}=2.1×10^{-8}), in a consistent manner across tissues. The opposite was observed for *HLA-DQB1*, with positive effects
327 on EBV EBNA and EBV p18 and inverse associations with EBV ZEBRA, JCV, VZV, HHV7, and HSV1.

328 The TWAS analyses also identified a number of significant associations in the HLA class III region that
329 were not detected in other analyses. The top-ranking VZV associated gene was *APOM* (P_{blood}=7.5×10^{-27},
330 P_{cortex}=1.1×10^{-25}). Interestingly, opposite directions of effect for *C4A* and *C4B* gene expression. Increased
331 *C4A* expression was positively associated with all EBV antigens (**Supplementary Table 16**), but negatively
332 associated with VZV (P_{blood}=2.3×10^{-24}) and HSV1 (P_{cortex}=1.8×10^{-5}) antibody levels (**Supplementary Table**
333 **15**). On the other hand, increased *C4B* expression was inversely associated with EBV phenotypes, but
334 positively associated with VZV (P_{blood}=8.1×10^{-25}) and HSV1 (P_{blood}=1.1×10^{-5}). A similar pattern was also
335 observed for *CYP21A2* and *C2*, with positive effects on antibody response to VZV and HSV1, and negative
336 effects for all EBV antigens. Other novel TWAS findings were detected for HHV7 in 22q13.2 (*CTA-223H9.9*:
337 P_{TWAS}=2.5×10^{-6}; *CSDC2*: P_{TWAS}=3.0×10^{-6}; *TEF*: P_{TWAS}=3.1×10^{-6}) and 1q31.2 (*RGS1*: P_{TWAS}=3.3×10^{-6}).

338 The TWAS recapitulated several GWAS-identified loci: 3q25.1 for EBV EBNA (*P2RY13*: P_{cortex}=1.1×10^{-8};
339 *P2RY12*: P_{blood}=3.3×10^{-8}) and 19q13.33 for BKV (*FUT2*: P_{TWAS}=8.1×10^{-13}; *NTN5*: P_{TWAS}=1.1×10^{-9}).
340 Transcriptomic profiles in skin tissues provided supporting evidence for the role of multiple genes in 5q31.2
341 in modulating MCV antibody response (**Figure 4; Supplementary Table 17**). The strongest signal was
342 observed in for *ECSCR* (skin sun unexposed: P_{TWAS}=5.0×10^{-15}; skin sun exposed: P_{TWAS}=4.2×10^{-13}),

followed by *PROB1* (sun unexposed: $P_{TWAS}=1.5\times10^{-11}$). *ECSCR* expression was also associated based on expression in the frontal cortex, while *PROB1* exhibited a significant, but attenuated effect in whole blood. *VWA7* was the only gene associated across all four tissues for MCV and was also associated with antibody response to several EBV antigens.

Comparison of results for seroreactivity and seropositivity revealed a number of genes implicated in both steps of the infection process (**Supplementary Table 18**). Associations with HLA DQA and DQB genes in whole blood and HLA-DRB genes in the frontal cortex were observed for EBV antigens, JCV, and MCV. For MCV, the strongest seropositivity signals were observed for HLA class III genes *AGER* ($P_{cortex}=9.0\times10^{-21}$) and *EHMT2* ($P_{blood}=5.8\times10^{-18}$), which were also among the top-ranking genes for seroreactivity. Increased *ECSCR* expression conferred an increased susceptibility to MCV infection ($P_{cortex}=1.8\times10^{-8}$), mirroring its effect on seroreactivity. In contrast to antibody response, no significant associations with any HLA genes were observed for VZV seropositivity.

Analyses using the Reactome database identified significant ($q_{FDR}<0.05$) enrichment for TWAS-identified genes in pathways involved in initiating antiviral responses, such as MHC class II antigen presentation, TCR signaling, and interferon (IFN) signaling (**Supplementary Figure 7**). Pathways unique to herpesviruses included folding, assembly and peptide loading of class I MHC ($q=3.2\times10^{-7}$) and initial triggering of complement ($q=9.8\times10^{-3}$). Polyomaviruses were associated with the non-canonical nuclear factor (NF)-κB pathway activated by tumor necrosis factor (TNF) superfamily ($q=1.9\times10^{-3}$). Protein interaction networks among genes associated with more than antigen identified 21 significant ($q_{FDR}<0.05$) nodes, most of which were centered around HLA-DRB1 and HLA-DQB1, and involved interactions with other MHC class II molecules, C4, and TNF (**Supplementary Figure 8**).

DISCUSSION

We performed genome-wide and transcriptome-wide association studies for serological phenotypes for 16 common viruses in a well-characterized, population-based cohort. We discovered novel genetic determinants of viral antibody response beyond the HLA region for BKV, MCV, HHV7, EBV EBNA, as well as SARS-CoV-2 infection status. Our comprehensive HLA analyses demonstrate that class II and III genes

369 are crucial host genetic factors involved in regulating immune response to diverse viral antigens[7]. Taken
370 together, the findings of this work provide a resource for further understanding the complex interplay
371 between viruses and the human genome, as well as a first step towards understanding germline
372 determinants of SARS-CoV-2 infection.

373 One of our main findings is the discovery of 5q31.2 as a susceptibility locus for MCV infection and MCV
374 antibody response, implicating two main genes: *TMEM173* (or *STING1*) and *ECSCR*. The former encodes
375 STING (stimulator of interferon genes), an endoplasmic reticulum (ER) protein that controls the transcription
376 of host defense genes and plays a critical role in response to DNA and RNA viruses[43]. STING is activated
377 by cyclic GMP-AMP synthase (cGAS), a cytosolic DNA sensor that mounts a response to invading
378 pathogens by inducing IFN1 and NF-κB signalling[44,45]. Polyomaviruses penetrate the ER membrane during
379 cell entry, a process that may be unique to this viral family[46], which may trigger STING signaling in a distinct
380 manner from other viruses[46]. Multiple cancer-causing viruses, such as KSHV, HBV, and HPV18, encode
381 oncoproteins that disrupt cGAS-STING activity, which illustrates the evolutionary pressure on DNA tumor
382 viruses to develop functions against this pathway and its importance in carcinogenesis[44]. Furthermore,
383 cGAS-STING activation has been shown to trigger antitumor T-cell responses, a mechanism that can be
384 leveraged by targeted immunotherapies [47-49]. Several studies suggest STING agonists may be effective
385 against tumors resistant to PD-1 blockade, as well as promising adjuvants in cancer vaccines[50-52].

386 *ECSCR* expression in skin and brain tissues was associated with MCV antibody response and infection.
387 This gene encodes an endothelial cell-specific chemotaxis regulator, which plays a role in angiogenesis
388 and apoptosis[53]. *ECSCR* is a negative regulator of PI3K/Akt signaling by enhancing membrane localization
389 of *PTEN* and operates in tandem with VEGFR-2 and other receptor tyrosine kinases[54]. In addition to 5q31.2,
390 another novel MCV seroreactivity associated region was identified in 3p24.3, anchored by rs776170649,
391 which has been linked to platelet phenotypes[55]. These findings align with a role of platelet activation in
392 defense against infections via degranulation-mediated release of chemokines and β-defensin[56].

393 Genetic variation within Fucosyltransferase 2 (*FUT2*) has been studied extensively in the context of human
394 infections; however, its effect on BKV seroreactivity is novel. Homozygotes for the nonsense mutation
395 (rs601338 G>A) that inactivates the FUT2 enzyme are unable to secrete ABO(H) histo-blood group

antigens or express them on mucosal surfaces[57,58]. The allele which confers increased BKV antibody response (rs681343-T) is in LD (r^2=1.00) with rs601338-A, the non-secretor allele, which confers resistance to norovirus[59,60], rotavirus[61], H. pylori[62], childhood ear infection, mumps, and common colds[13]. However, increased susceptibility to other pathogens, such as meningococcus and pneumococcus[63] has also been observed in non-secretors. Isolating the underlying mechanisms for BKV response is challenging because *FUT2* is a pleiotropic locus associated with diverse phenotypes, including autoimmune and inflammatory conditions[64,65], serum lipids[66], B vitamins[58,67], alcohol consumption[68], and even certain cancers[69]. In addition to *FUT2* in 19q13.33, *NTN5* (netrin 5) suggests a possible link between BKV and neurological conditions. NTN5 is primarily expressed in neuroproliferative areas, suggesting a role in adult neurogenesis, which is dysregulated in glioblastoma and Alzheimer's disease[70,71].

We also report the first GWAS of serological phenotypes for HHV7. Genetic determinants of HHV7 antibody response in 6p21.32 were predominantly localized in *HLA-DQA1* and *HLA-DQB1*, with associations similar to other herpesviruses. In 11q23.3, rs75438046 maps to the 3′ UTR of *CXCR5*, which controls viral infection in B-cell follicles[72], and *BCL9L*, a translocation target in acute lymphoblastic leukemia[73] and transcriptional activator of the Wnt/β-catenin cancer signaling pathway[74]. In 17q21.32, *TBKBP1* encodes an adaptor protein that binds to TBK1 and is part of the TNF/NF-κB interaction network, where it regulates immune responses to infectious triggers, such as IFN1 signaling[75]. Interestingly, a protein interactome map recently revealed that SARS-CoV-2 nonstructural protein 13 (Nsp13) includes TBK1-TBKBP1 among its targets [76]. Other functions of the TBK1-TBKBP1 axis relate to tumor growth and immunosuppression through induction of PD-L1[77].

Several additional genes involved in HHV7 immune response were identified in TWAS. *TEF* in 22q13.2 is an apoptotic regulator of hematopoietic progenitors with tumor promoting effects mediated by inhibition of G1/S cell cycle transition and Akt/FOXO signaling[78]. *RGS1* in 1q31.2 has been linked to multiple autoimmune diseases, including multiple sclerosis[79], as well as poor prognosis in melanoma and diffuse large B cell lymphoma mediated by inactivation of Akt/ERK[80,81].

Other genes outside of the HLA region associated with viral infection response were detected for EBV EBNA in 3q25.1. The lead variant (rs67886110) is an eQTL for *MED12L* and *P2RY12* genes, which have

been linked to neurodegenerative conditions[82,83]. *P2RY12* and *P2RY13*, identified in TWAS, are purinergic receptor genes that regulate microglia homeostasis and have been implicated in Alzheimer's susceptibility via inflammatory and neurotrophic mechanisms[83].

Considering genetic variation within the HLA region, our results not only confirm its pivotal role at the interface of host pathogen interactions, but also highlight the overlap in variants, classical alleles, and genes that mediate these interactions across virus families and antigens. We identified 40 independent SNPs/indels associated with EBV (EBNA, EA-D, VCA p18, and ZEBRA), VZV, and MCV antibody response that accounted for all significant HLA associations for other phenotypes. Of the 14 conditionally independent, genome-wide significant classical alleles identified for 10 antigens, 7 were associated with multiple phenotypes. The most commonly shared HLA alleles were DRB5*99:01, DRB1*04:04, an know rheumatoid arthritis risk allele[84], and DQB1*02:01, which is implicated in susceptibility to celiac disease[85]. Furthermore, nodes anchored by *HLA-DRB1* and *HLA-DQB1* were at the center of the protein interaction network identified for TWAS genes that were implicated in antibody response to multiple antigens. Despite the predominance of association in HLA class II, one notable association in HLA class I was detected for A*29:02 and MCV seroreactivity, which is consistent with downregulation of MHC I as a potential mechanism through which Merkel cell tumors evade immune surveillance[86].

Comparison with other studies of host genetics and viral infection susceptibility shows that our results align with previously reported findings. We replicated ($P<5\times10^{-8}$) many associations from a GWAS of humoral immune response by Hammer et al.[7], specifically HLA signals for MCV (rs1049130, DQB1*06:02), JCV (rs9269910, DQA1*01:02), and EBV EBNA (DRB1*07:01, HLA-DRB1*03:01, also reported by Scepanovic et al[8]). Associations with JCV seroreactivity replicated ($P<5\times10^{-8}$) class II HLA alleles (DQB1*03:01, DQB1*06:02, and DQA1*01:02) linked to JCV infection in a cohort of multiple sclerosis patients and population-based controls[87]. We also replicated two *HLA-DRB1* variants (rs477515, rs2854275: $P<5\times10^{-19}$) associated with EBV EBNA antibody levels in a Mexican American population[9]. Our GWAS of HPV16 L1 replicated a variant previously linked to HPV8 seropositivity (rs9357152, $P=0.008$)[6]. Some of our findings contrast with Tian et al [13], although we confirmed selected associations, such as A*02:01 (shingles) with VZV ($P=4.1\times10^{-8}$) and rs2596465 (mononucleosis) with EBV EBNA ($P=3.3\times10^{-9}$) and EBV p18 ($P=1.0\times10^{-}$

[12]). These differences may be partly accounted for by self-reported disease status in Tian et al. which may be an imprecise indicator of infection with certain viruses.

One of the most striking findings in SNP-based HLA analyses was the genome-wide significant association between rs9273325, sentinel VZV antibody response variant, and risk of schizophrenia. This observation is consistent with the established role the HLA region, including *HLA-DQB1*, in schizophrenia etiology[88,89], and is further supported by previously reported associations for rs9273325 with blood cell traits[55] and immunoglobulin A deficiency[90], as well as its role as an eQTL for *HLA-DQB1* in CD4+ T_2h cells. An inverse association with schizophrenia has been reported for DRB1*03:01[88], the lead HLA allele associated with increased VZV antibody response. Two other strongly associated schizophrenia alleles, HLA-DQB1:02 and HLA-B*08:01, were also among the top VZV-associated variants ($P<5\times10^{-25}$) in the unconditional analysis. Enhanced complement activity has been proposed as the mechanism mediating the synaptic loss and excessive pruning which is a hallmark of schizophrenia pathophysiology[91]. Complement component 4 (C4) alleles were found to increase risk of schizophrenia proportionally to their effect on increasing *C4A* expression in brain tissue[91]. Using gene expression models in whole blood and the frontal cortex we demonstrated that increased *C4A* expression is negatively associated with VZV antibody response. We also observed associations with *C4A* and *C4B* in EBV and HSV-1, but not other viruses. Taken together, these findings delineate a potential mechanism through which aberrant immune response to VZV infection, and potentially HSV-1 and EBV infection, may increase susceptibility to schizophrenia.

The main finding of our genetic associations analyses of SARS-CoV-2 infection status was the discovery of a genome-wide significant signal in *EHF* (rs286914), which encodes a transcriptional repressor that plays a critical role in lung inflammation and response to injury[92] and modifies disease severity in cystic fibrosis[93]. Since EHF occupies the relatively common ETS motif, allowing it to up- or down-regulate gene expression, the mechanism of action most relevant for SARS-CoV-2 infection or COVID-19 symptom severity may depend on its interactions with other transcription factors. We also identified nominal inverse associations with SARS-CoV-2 infection for HLA A*31:01 allele, which has been implicated in immune-mediated adverse reactions to anticonvulsant carbamazepine[94], and DQA1*03:01, which is a part of a known type I diabetes susceptibility haplotype[95]. We were unable to assess the role of B*46:01, previously linked to 2003 SARS

in an East Asian population, due to the low frequency (3.2×10^{-5}) of this allele in our data. This illustrates the need for studies in large, genetically diverse populations to elucidate the role of HLA in SARS-CoV-2 infection dynamics.

The relationship between genetically predicted *ACE2* gene expression and SARS-CoV-2 infection appears to be complex and tissue-dependent. *ACE2* is expressed at high levels in intestinal organs, adipose tissue, kidney, heart, and testis and has multiple significant eQTLs in different brain structures. Based on expression profiles in these diverse tissues, we observed an inverse association between genetically predicted *ACE2* levels and having a positive SARS-CoV-2 test. However, considering *ACE2* expression in lung tissue, we observed a trend towards higher expression conferring increased SARS-CoV-2 susceptibility, which is consistent with increased expression facilitating increased receptor availability. SARS-CoV-2 uses the *ACE2* receptor for cell entry[30], yet studies of betacoronavirus SARS have shown that *ACE2* is downregulated upon viral establishment[96], which is believed to be the molecular basis of COVID-19 respiratory distress. The role of *ACE2* expression in susceptibility to infection remains unclear and may be modulated by tissue-specific post-translational regulation. Further research is needed to elucidate the pathways through which *ACE2* may influence susceptibility to SARS-CoV-2 infection as well as COVID-19 symptom severity.

While the reported associations for SARS-CoV-2 were based on a limited sample size, we believe these findings may be informative when considered in the context of our results for other viral infections, and contribute to the global, pressing need to understand COVID-19, which motivated our rapid analysis of these data. However, care must be taken with the interpretation of genetic associations with SARS-CoV-2 status, which should be regarded as preliminary and exploratory given the small number of tests conducted in a biased sample, confounding by factors related to symptom recognition, access to testing, and limited statistical power.

Several additional limitations of this work should be noted. First, the UK Biobank is unrepresentative of the general UK population due to low participation resulting in healthy volunteer bias[97]. However, since the observed pattern of seroprevalence is consistent with previously published estimates[16] we believe the impact of this bias is likely to be minimal on genetic associations with serological phenotypes. Second, our

analyses were restricted to participants of European ancestry due to limited serology data for other ancestries, which limits the generalizability of our findings to diverse populations. Third, we were unable to conduct formal statistical replication of novel GWAS and TWAS signals in an independent sample due to the lack of such a population. Nevertheless, our successful replication of multiple previously reported variants and, combined with the observation that newly discovered genes and variants are part of essential adaptive and innate immunity pathways, support the credibility of our findings. Lastly, we also stress caution in the interpretation of GWAS results for non-ubiquitous pathogens, such as HBV, HCV, and HPV, due to a lack of information on exposure, as well as low numbers of seropositive individuals.

Our study also has distinct advantages. The large sample size of the UK Biobank facilitated more powerful genetic association analyses than previous studies, particularly in a population-based cohort unselected for disease status. Our detailed HLA analysis shows independent effects of specific HLA alleles and pleiotropic effects across multiple viruses. Analyses of genetic associations in external datasets further demonstrate a connection between host genetic factors influencing immune response to infection and susceptibility to cancers and neurological conditions. Finally, integration of SARS-CoV-2 test results with findings for serology measures for common viruses provides previously unknown context for putative associations.

The results of this work highlight widespread genetic pleiotropy between pathways involved in regulating humoral immune response to novel and common viruses, as well as complex diseases. Understanding the interplay between host genetic factors and immune response has implications for public health and may facilitate the discovery of novel therapeutics including vaccines.

DATA AVAILABILITY

The UK Biobank in an open access resource, available at https://www.ukbiobank.ac.uk/researchers/. This research was conducted with approved access to UK Biobank data under application number 14105 (PI: Witte).

WEB RESOURCES

PLINK 2.0: https://www.cog-genomics.org/plink/2.0/

R packages for pathway analysis: https://bioconductor.org/packages/release/bioc/html/ReactomePA.html and https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html

R package for protein network interaction analysis: http://xgr.r-forge.r-project.org

ACKNOWLEDGEMENTS

This research was supported by funding from the National Institutes of Health (US NCI R25T CA112355 and R01 CA201358; PI: Witte). Maike Morrison was funded by the University of California San Francisco's Amgen Scholars Program.

The lung eQTL study at Laval University was supported by the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec and the Canadian Institutes of Health Research (MOP - 123369). Y.B. holds a Canada Research Chair in Genomics of Heart and Lung Diseases.

COMPETING INTERESTS

The authors declare no competing interests.

References

1. Aiewsakun, P. & Katzourakis, A. Marine origin of retroviruses in the early Palaeozoic Era. *Nat Commun* **8**, 13954 (2017).
2. Wang, W. *et al.* Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA* (2020).
3. Moore, P.S. & Chang, Y. Why do viruses cause cancer? Highlights of the first century of human tumour virology. *Nat Rev Cancer* **10**, 878-89 (2010).
4. Engdahl, E. *et al.* Increased Serological Response Against Human Herpesvirus 6A Is Associated With Risk for Multiple Sclerosis. *Front Immunol* **10**, 2715 (2019).
5. Readhead, B. *et al.* Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus. *Neuron* **99**, 64-82 e7 (2018).
6. Chen, D. *et al.* Genome-wide association study of HPV seropositivity. *Hum Mol Genet* **20**, 4714-23 (2011).
7. 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* **97**, 738-43 (2015).
8. Scepanovic, P. *et al.* Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. *Genome Med* **10**, 59 (2018).
9. Rubicz, R. *et al.* A genome-wide integrative genomic study localizes genetic factors influencing antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1). *PLoS Genet* **9**, e1003147 (2013).
10. Liu, S. *et al.* Genomic Analyses from Non-invasive Prenatal Testing Reveal Genetic Associations, Patterns of Viral Infections, and Chinese Population History. *Cell* **175**, 347-359 e14 (2018).
11. Kenney, A.D. *et al.* Human Genetic Determinants of Viral Diseases. *Annu Rev Genet* **51**, 241-263 (2017).
12. Besson, C. *et al.* Strong correlations of anti-viral capsid antigen antibody levels in first-degree relatives from families with Epstein-Barr virus-related lymphomas. *J Infect Dis* **199**, 1121-7 (2009).
13. Tian, C. *et al.* Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun* **8**, 599 (2017).
14. Lin, M. *et al.* Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med Genet* **4**, 9 (2003).
15. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
16. Mentzer, A.J. *et al.* Identification of host-pathogen-disease relationships using a scalable Multiplex Serology platform in UK Biobank. *medRxiv*, 19004960 (2019).
17. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-73 (2010).
18. Waterboer, T. *et al.* Multiplex human papillomavirus serology based on in situ-purified glutathione s-transferase fusion proteins. *Clin Chem* **51**, 1845-53 (2005).

19. Waterboer, T., Sehr, P. & Pawlita, M. Suppression of non-specific binding in serological Luminex assays. *J Immunol Methods* **309**, 200-4 (2006).
20. Kreimer, A.R. *et al.* Kinetics of the Human Papillomavirus Type 16 E6 Antibody Response Prior to Oropharyngeal Cancer. *J Natl Cancer Inst* **109**(2017).
21. Peterson, R.A. & Cavanaugh, J.E. Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*, 1-16 (2019).
22. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894 (2019).
23. Dong, S. & Boyle, A.P. Predicting functional variants in enhancer and promoter elements using RegulomeDB. *Hum Mutat* **40**, 1292-1298 (2019).
24. Schmiedel, B.J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701-1715 e16 (2018).
25. Sun, B.B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).
26. Yao, C. *et al.* Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat Commun* **9**, 3268 (2018).
27. Rashkin, S.R. *et al.* Pan-Cancer Study Detects Novel Genetic Risk Variants and Shared Genetic Basis in Two Large Cohorts. *bioRxiv*, 635367 (2019).
28. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet* **51**, 1670-1678 (2019).
29. Jun, G. *et al.* A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol Psychiatry* **21**, 108-17 (2016).
30. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
31. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **8**, e1003029 (2012).
32. Barbeira, A.N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* **9**, 1825 (2018).
33. Barbeira, A.N. *et al.* Widespread dose-dependent effects of RNA expression and splicing on complex diseases and traits. *bioRxiv*, 814350 (2019).
34. Urbut, S.M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat Genet* **51**, 187-195 (2019).
35. Wen, X., Lee, Y., Luca, F. & Pique-Regi, R. Efficient Integrative Multi-SNP Association Analysis via Deterministic Approximation of Posteriors. *Am J Hum Genet* **98**, 1114-1129 (2016).
36. Lee, Y., Luca, F., Pique-Regi, R. & Wen, X. Bayesian Multi-SNP Genetic Association Analysis: Control of FDR and Use of Summary Statistics. *bioRxiv*, 316471 (2018).

37. Steiner, I., Kennedy, P.G. & Pachner, A.R. The neurotropic herpes viruses: herpes simplex and varicella-zoster. *Lancet Neurol* **6**, 1015-28 (2007).
38. Khalili, K., Del Valle, L., Otte, J., Weaver, M. & Gordon, J. Human neurotropic polyomavirus, JCV, and its role in carcinogenesis. *Oncogene* **22**, 5181-91 (2003).
39. Feng, H., Shuda, M., Chang, Y. & Moore, P.S. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* **319**, 1096-100 (2008).
40. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-D613 (2019).
41. Shin, O.S. *et al.* LPLUNC1 modulates innate immune responses to Vibrio cholerae. *J Infect Dis* **204**, 1349-57 (2011).
42. Shafi, O. Inverse relationship between Alzheimer's disease and cancer, and other factors contributing to Alzheimer's disease: a systematic review. *BMC Neurol* **16**, 236 (2016).
43. Chen, Q., Sun, L. & Chen, Z.J. Regulation and function of the cGAS-STING pathway of cytosolic DNA sensing. *Nat Immunol* **17**, 1142-9 (2016).
44. Kwon, J. & Bakhoum, S.F. The Cytosolic DNA-Sensing cGAS-STING Pathway in Cancer. *Cancer Discov* **10**, 26-39 (2020).
45. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z.J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786-91 (2013).
46. Inoue, T. & Tsai, B. How viruses use the endoplasmic reticulum for entry, replication, and assembly. *Cold Spring Harb Perspect Biol* **5**, a013250 (2013).
47. Woo, S.R. *et al.* STING-dependent cytosolic DNA sensing mediates innate immune recognition of immunogenic tumors. *Immunity* **41**, 830-42 (2014).
48. Demaria, O. *et al.* STING activation of tumor endothelial cells initiates spontaneous and therapeutic antitumor immunity. *Proc Natl Acad Sci U S A* **112**, 15408-13 (2015).
49. Ohkuri, T. *et al.* STING contributes to antiglioma immunity via triggering type I IFN signals in the tumor microenvironment. *Cancer Immunol Res* **2**, 1199-208 (2014).
50. Fu, J. *et al.* STING agonist formulated cancer vaccines can cure established tumors resistant to PD-1 blockade. *Sci Transl Med* **7**, 283ra52 (2015).
51. Corrales, L. *et al.* Direct Activation of STING in the Tumor Microenvironment Leads to Potent and Systemic Tumor Regression and Immunity. *Cell Rep* **11**, 1018-30 (2015).
52. Ohkuri, T., Ghosh, A., Kosaka, A., Sarkar, S.N. & Okada, H. Protective role of STING against gliomagenesis: Rational use of STING agonist in anti-glioma immunotherapy. *Oncoimmunology* **4**, e999523 (2015).
53. Ikeda, K. *et al.* Identification of ARIA regulating endothelial apoptosis and angiogenesis by modulating proteasomal degradation of cIAP-1 and cIAP-2. *Proc Natl Acad Sci U S A* **106**, 8227-32 (2009).

54. Verma, A. *et al.* Endothelial cell-specific chemotaxis receptor (ecscr) promotes angioblast migration during vasculogenesis and enhances VEGF receptor sensitivity. *Blood* **115**, 4614-22 (2010).
55. Astle, W.J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429 e19 (2016).
56. Assinger, A. Platelets and infection - an emerging role of platelets in viral infection. *Front Immunol* **5**, 649 (2014).
57. Kelly, R.J., Rouquier, S., Giorgi, D., Lennon, G.G. & Lowe, J.B. Sequence and expression of a candidate for the human Secretor blood group alpha(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. *J Biol Chem* **270**, 4640-9 (1995).
58. Hazra, A. *et al.* Common variants of FUT2 are associated with plasma vitamin B12 levels. *Nat Genet* **40**, 1160-2 (2008).
59. Carlsson, B. *et al.* The G428A nonsense mutation in FUT2 provides strong but not absolute protection against symptomatic GII.4 Norovirus infection. *PLoS One* **4**, e5593 (2009).
60. Ruvoen-Clouet, N., Belliot, G. & Le Pendu, J. Noroviruses and histo-blood groups: the impact of common host genetic polymorphisms on virus transmission and evolution. *Rev Med Virol* **23**, 355-66 (2013).
61. Imbert-Marcille, B.M. *et al.* A FUT2 gene common polymorphism determines resistance to rotavirus A of the P[8] genotype. *J Infect Dis* **209**, 1227-30 (2014).
62. Ikehara, Y. *et al.* Polymorphisms of two fucosyltransferase genes (Lewis and Secretor genes) involving type I Lewis antigens are associated with the presence of anti-Helicobacter pylori IgG antibody. *Cancer Epidemiol Biomarkers Prev* **10**, 971-7 (2001).
63. Blackwell, C.C. *et al.* Non-secretion of ABO antigens predisposing to infection by Neisseria meningitidis and Streptococcus pneumoniae. *Lancet* **2**, 284-5 (1986).
64. de Lange, K.M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* **49**, 256-261 (2017).
65. Ellinghaus, D. *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* **48**, 510-8 (2016).
66. Hoffmann, T.J. *et al.* A large electronic-health-record-based genome-wide study of serum lipids. *Nat Genet* **50**, 401-413 (2018).
67. Tanaka, T. *et al.* Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations. *Am J Hum Genet* **84**, 477-82 (2009).
68. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* **51**, 237-244 (2019).
69. McKay, J.D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet* **49**, 1126-1132 (2017).

70. Batista, C.M. *et al.* Adult neurogenesis and glial oncogenesis: when the process fails. *Biomed Res Int* **2014**, 438639 (2014).
71. Yamagishi, S. *et al.* Netrin-5 is highly expressed in neurogenic regions of the adult brain. *Front Cell Neurosci* **9**, 146 (2015).
72. Leong, Y.A. *et al.* CXCR5(+) follicular cytotoxic T cells control viral infection in B cell follicles. *Nat Immunol* **17**, 1187-96 (2016).
73. Willis, T.G. *et al.* Molecular cloning of translocation t(1;14)(q21;q32) defines a novel gene (BCL9) at chromosome 1q21. *Blood* **91**, 1873-81 (1998).
74. Deka, J. *et al.* Bcl9/Bcl9l are critical for Wnt-mediated regulation of stem cell traits in colon epithelium and adenocarcinomas. *Cancer Res* **70**, 6619-28 (2010).
75. Pilli, M. *et al.* TBK-1 promotes autophagy-mediated antimicrobial defense by controlling autophagosome maturation. *Immunity* **37**, 223-34 (2012).
76. Gordon, D.E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* (2020).
77. Zhu, L. *et al.* TBKBP1 and TBK1 form a growth factor signalling axis mediating immunosuppression and tumourigenesis. *Nat Cell Biol* **21**, 1604-1614 (2019).
78. Yang, J. *et al.* Thyrotroph embryonic factor is downregulated in bladder cancer and suppresses proliferation and tumorigenesis via the AKT/FOXOs signalling pathway. *Cell Prolif* **52**, e12560 (2019).
79. International Multiple Sclerosis Genetics, C. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214-9 (2011).
80. Sun, M.Y. *et al.* Critical role for nonGAP function of Galphas in RGS1mediated promotion of melanoma progression through AKT and ERK phosphorylation. *Oncol Rep* **39**, 2673-2680 (2018).
81. Carreras, J. *et al.* Clinicopathological characteristics and genomic profile of primary sinonasal tract diffuse large B cell lymphoma (DLBCL) reveals gain at 1q31 and RGS1 encoding protein; high RGS1 immunohistochemical expression associates with poor overall survival in DLBCL not otherwise specified (NOS). *Histopathology* **70**, 595-621 (2017).
82. Mukherjee, S. *et al.* Genetic data and cognitively defined late-onset Alzheimer's disease subgroups. *Mol Psychiatry* (2018).
83. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290 e17 (2017).
84. Jawaheer, D. *et al.* Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis. *Am J Hum Genet* **71**, 585-94 (2002).
85. Vader, W. *et al.* The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. *Proc Natl Acad Sci U S A* **100**, 12390-5 (2003).
86. Paulson, K.G. *et al.* Downregulation of MHC-I expression is prevalent but reversible in Merkel cell carcinoma. *Cancer Immunol Res* **2**, 1071-9 (2014).

87. Sundqvist, E. *et al.* JC polyomavirus infection is strongly controlled by human leucocyte antigen class II variants. *PLoS Pathog* **10**, e1004084 (2014).
88. International Schizophrenia, C. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-52 (2009).
89. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).
90. Bronson, P.G. *et al.* Common variants at PVT1, ATG13-AMBRA1, AHI1 and CLEC16A are associated with selective IgA deficiency. *Nat Genet* **48**, 1425-1429 (2016).
91. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177-83 (2016).
92. Fossum, S.L. *et al.* Ets homologous factor (EHF) has critical roles in epithelial dysfunction in airway disease. *J Biol Chem* **292**, 10938-10949 (2017).
93. Wright, F.A. *et al.* Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. *Nat Genet* **43**, 539-46 (2011).
94. McCormack, M. *et al.* HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. *N Engl J Med* **364**, 1134-43 (2011).
95. Hu, X. *et al.* Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet* **47**, 898-905 (2015).
96. Kuba, K. *et al.* A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat Med* **11**, 875-9 (2005).
97. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* **186**, 1026-1034 (2017).

Table 1: Lead genome-wide significant variants ($P<5.0\times10^{-8}$) for continuous antibody response phenotypes for antigens with at least 20% seroprevalence.

Antigen		N	Chr	Position	Variant	Alleles		EAF	Beta[2]	(SE)	P	Function	Nearest Gene
						Effect	Other						
CMV	pp52	5000	6	32301427	rs115378818	C	T	0.978	0.633	(0.095)	2.9×10^{-11}	intronic	*TSBP1*
EBV	EA-D	6806	6	32665840	rs34825357	T	TC	0.409	-0.114	(0.017)	2.0×10^{-11}	intergenic	*MTCO3P1*
EBV	EBNA	7003	3	151114852	rs67886110*	G	T	0.596	0.103	(0.017)	1.3×10^{-9}	intronic	*MED12L*
			6	32451762	rs9269233	A	C	0.249	0.315	(0.019)	3.5×10^{-61}	intergenic	*HLA-DRB9*
EBV	VCA p18	7492	6	31486158	6:31486158	GT	G	0.245	0.197	(0.018)	7.1×10^{-27}	intergenic	*PPIAP9*
EBV	ZEBRA	7197	6	32637772	rs9274728	A	G	0.718	-0.315	(0.018)	4.7×10^{-67}	intergenic	*HLA-DQB1*
HHV6	IE1A	6077	7	139985625	rs2429218	T	C	0.615	0.106	(0.019)	1.4×10^{-8}	downstream	*RP5-1136G2.1*
HHV7	U14	7481	6	32602665	rs139299944	C	CT	0.655	0.114	(0.017)	1.5×10^{-11}	intronic	*HLA-DQA1*
			11	118767564	rs75438046	G	A	0.970	0.280	(0.049)	1.3×10^{-8}	3'-UTR	*CXCR5* / *BCL9L*
			17	45794706	rs1808192	A	G	0.331	-0.099	(0.017)	9.8×10^{-9}	intergenic	*TBKBP1*
HSV1	1gG	5468	6	32627852	rs1130420	G	A	0.583	-0.122	(0.019)	2.5×10^{-10}	3'-UTR	*HLA-DQB1*
			10	91189187	rs11203123*	A	C	0.988	0.512	(0.093)	3.9×10^{-8}	intergenic	*SLC16A12*
VZV	gE/Ig[1]	7289	6	32623193	rs9273325	G	A	0.831	-0.232	(0.021)	8.2×10^{-28}	intergenic	*HLA-DQB1*
BKV	VP1	7523	19	49206462	rs681343	C	T	0.491	-0.125	(0.016)	4.7×10^{-15}	synonymous	*FUT2*
JCV	VP1	4471	6	32589842	rs9271525	G	A	0.163	-0.318	(0.031)	3.9×10^{-24}	intergenic	*HLA-DQA1*
MCV	VP1	5219	3	18238783	rs776170649	CT	C	0.790	-0.134	(0.024)	1.7×10^{-8}	intergenic	*LOC339862*
			5	138865423	rs7444313	G	A	0.263	0.169	(0.021)	2.4×10^{-15}	intergenic	*TMEM173*
			6	32429277	rs9268847	A	G	0.750	-0.195	(0.022)	2.4×10^{-19}	intronic	*HLA-DRB9*

[1] VZV antigens gE and gI were co-loaded onto the same Luminex bead set

[2] Regression coefficients were estimated per 1 standard deviation increase in normalized MFI value z-scores with adjustment for age at enrollment, sex, body mass index, socioeconomic status (Townsend deprivation index), the presence of any autoimmune conditions, genotyping array, serology assay date, quality control flag and the top 10 genetic ancestry principal components

* Multi-allelic variants: rs67886110 (G/T and G/C) and rs11203123 (A/C and A/AC)

Table 2: Genetic variants associated with laboratory-confirmed SARS-Cov-2 in a subset of UK Biobank participants with available test results. Odds ratios (OR) for testing positive were estimated in 2010 individuals of predominantly European ancestry, restricting to test results based on respiratory specimens.

Genome-wide association analysis

Variant	Frequency	OR[1]	(95% CI)	P	Function	Gene
rs286914 - A (11:34653124)	0.298	1.52	(1.32 – 1.77)	2.0×10^{-8}	intronic	*EHF*

ACE2 gene expression

Variant	Frequency	OR[1]	(95% CI)	P	eQTL effect size	Tissue
rs11798628 - T (X:16020859)	0.135	0.78	(0.65 – 0.93)	5.1×10^{-3}	0.19	Nerve – tibial
rs4830974 - A (X:15676417)	0.479	0.88	(0.79 – 0.99)	0.026	0.60	Brain – frontal cortex (5 other tissues)[2]
rs5934251 - A (X: 15637513)	0.020	2.12	(1.39 – 3.25)	5.3×10^{-4}	0.40	Lung

Classical HLA alleles

Allele	Frequency	OR[1]	(95% CI)	P	Associated phenotypes[3]
DQA1*03:01	0.201	0.80	(0.67 – 0.95)	0.012	EBV EA-D, EBV EBNA, EBV ZEBRA, HHV6 IE1B, JCV
A*31:01	0.026	0.53	(0.31 – 0.90)	0.019	EBV ZEBRA

[1] Odds ratios were estimated using logistic regression with adjustment for age at enrollment, sex, body mass index, Townsend deprivation index, smoking status (never, former, or current smoker), specimen origin (confirmed inpatient or acute care setting vs. unknown), genotyping array, and the top 10 genetic ancestry principal components

[2] For variants that are eQTLs in multiple tissues, the effect size is reported for the tissue with the lowest p-value

[3] Other associated phenotypes include any antigens for which that specific allele had an association with antibody response levels at the Bonferroni-corrected significance threshold of $P<5\times10^{-4}$

Figure 1: Results from genome-wide and regional association analyses of continuous antibody response phenotypes (MFI z-scores) among individuals seropositive for human polyomaviruses BKV, JCV, and Merkel cell (MCV). The lower two panels depict the association signal and linkage disequilibrium (LD) structure in the HLA region for JCV and MCV.

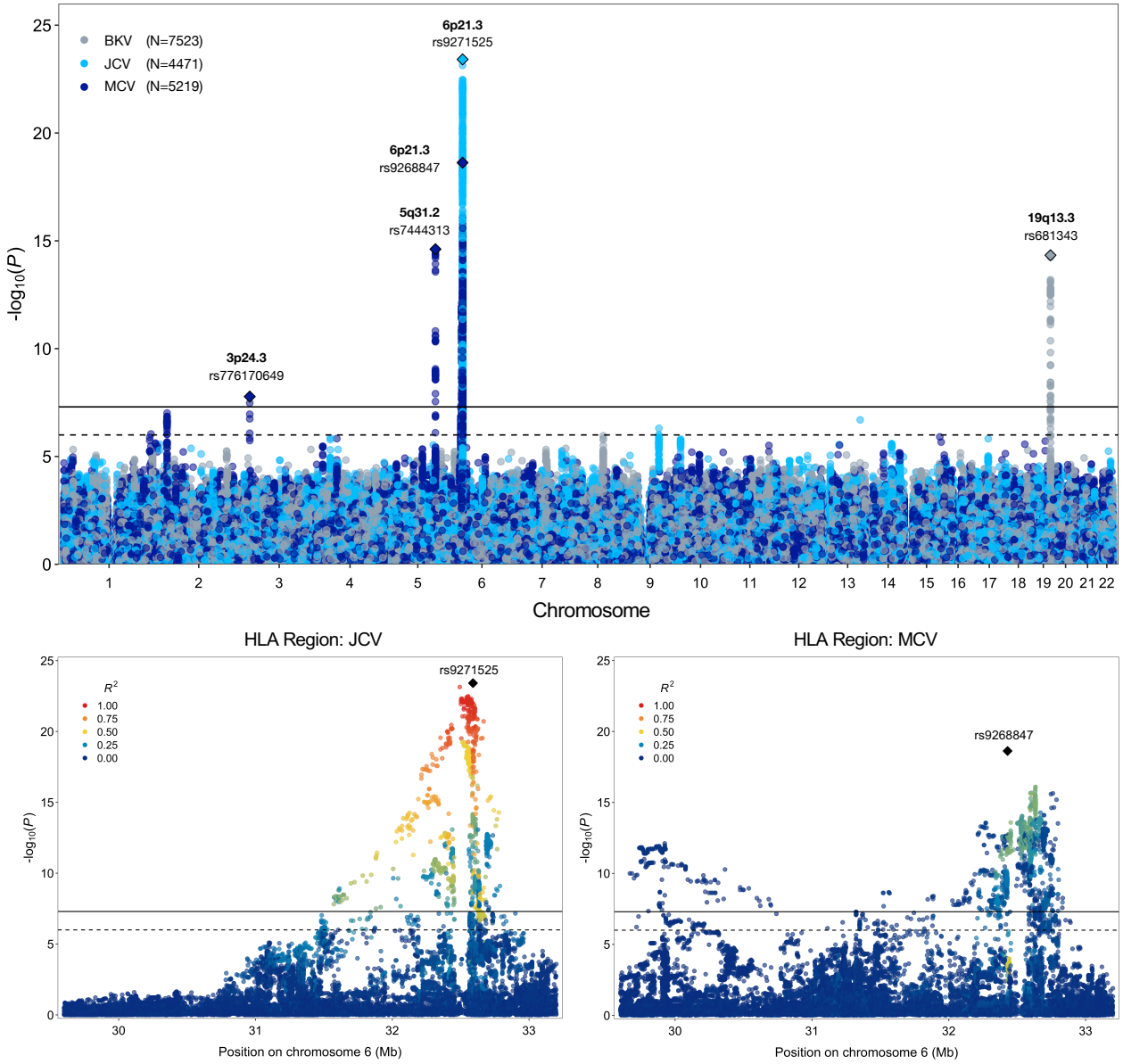


Figure 2: Regional association plots for conditionally independent HLA genetic variants that were significantly (P<5.0×10^{-8}, solid black line) associated with each continuous antibody response phenotype. The suggestive significance threshold corresponds to P<1.0×10^{-6} (dotted black line).

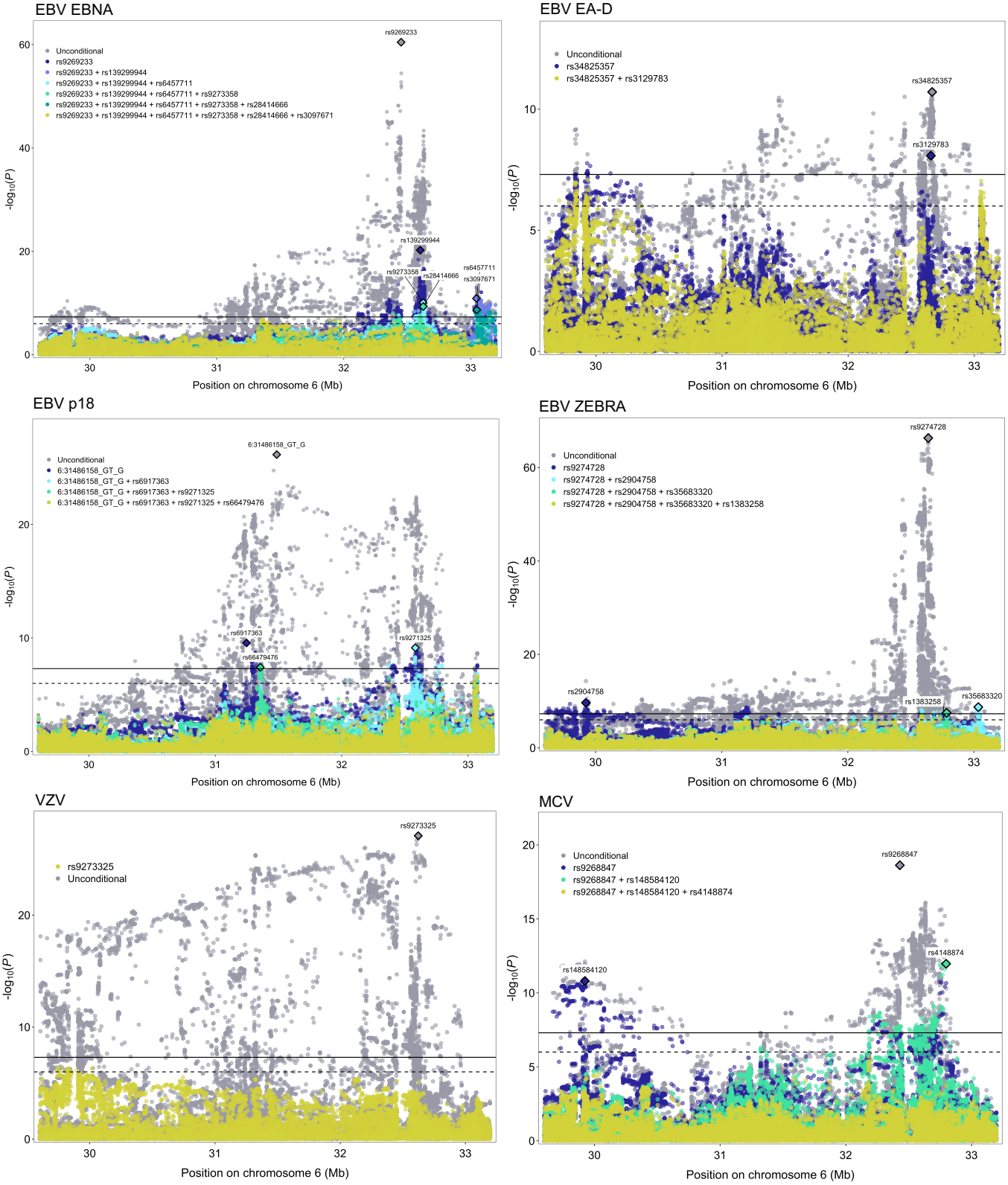
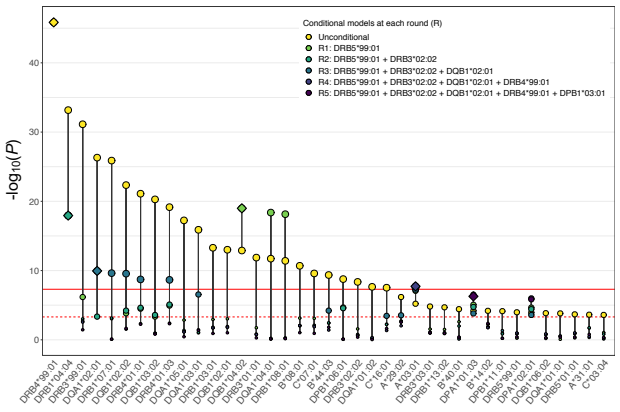


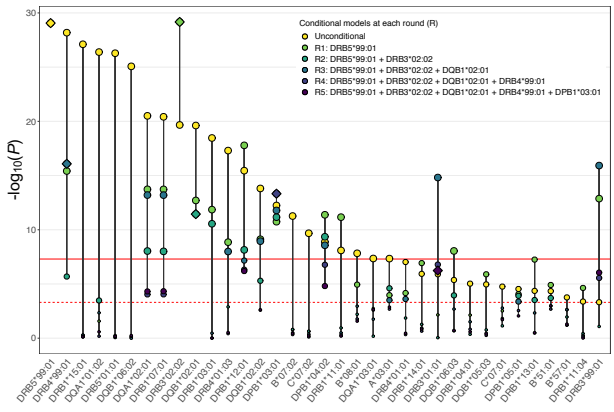
Figure 3: Conditionally independent classical HLA alleles significantly (P_{cond}<5.0×10^{-8}, solid red line) associated with each continuous antibody response phenotype. Only classical alleles that surpassed the Bonferroni-corrected significance threshold (P<5.0×10^{-4}, dotted red line) were included in conditional analyses.

EBV ZEBRA



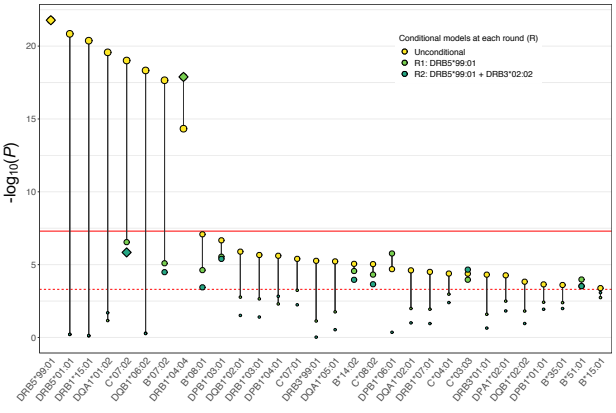
Allele	Frequency	Beta	(SE)	P_{cond}
DRB4*99:01	0.652	-0.246	(0.017)	1.4×10^{-46}
DQB1*04:02	0.020	0.504	(0.055)	1.0×10^{-19}
DRB1*04:04	0.041	0.376	(0.042)	1.1×10^{-18}
DQA1*02:01	0.145	0.187	(0.029)	1.1×10^{-10}
A*03:01	0.144	0.129	(0.023)	1.9×10^{-8}

EBV EBNA



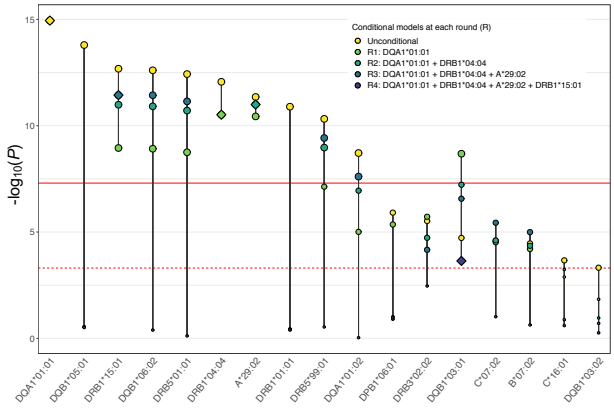
Allele	Frequency	Beta	(SE)	P_{cond}
DRB5*99:01	0.845	-0.255	(0.022)	8.7×10^{-30}
DRB3*02:02	0.136	0.276	(0.024)	6.8×10^{-30}
DQB1*02:01	0.149	-0.164	(0.023)	3.6×10^{-12}
DRB4*99:01	0.652	0.176	(0.021)	8.3×10^{-17}
DPB1*03:01	0.100	-0.220	(0.029)	4.7×10^{-14}

EBV p18



Allele	Frequency	Beta	(SE)	P_{cond}
DRB5*99:01	0.845	-0.210	(0.021)	1.7×10^{-22}
DRB1*04:04	0.041	0.357	(0.040)	1.3×10^{-18}

MCV



Allele	Frequency	Beta	(SE)	P_{cond}
DQA1*01:01	0.141	0.215	(0.027)	1.1×10^{-15}
DRB1*04:04	0.041	-0.362	(0.054)	3.0×10^{-11}
A*29:02	0.042	-0.350	(0.051)	1.0×10^{-11}
DRB1*15:01	0.144	-0.203	(0.029)	3.7×10^{-12}

Figure 4: TWAS associations with continuous antigen response phenotypes. Two Manhattan plots depicting the transcriptome-wide associations for genes with a positive direction of effect (increased expression leads to higher antibody response) and genes with a negative direction of effect (increased expression is associated with a reduced antibody response).