

The landscape of host genetic factors involved in immune response to common viral infections

Linda Kachuri^{1*}, Stephen S. Francis^{1,2,3,4*}, Maïke Morrison^{5,6}, George A. Wendt², Yohan Bossé⁷, Taylor B. Cavazos⁸, 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

1 **ABSTRACT**

2 **Introduction:** Humans and viruses have co-evolved for millennia resulting in a complex host genetic
3 architecture. Understanding the genetic mechanisms of immune response to viral infection provides insight
4 into disease etiology and therapeutic opportunities.

5 **Methods:** We conducted a comprehensive study including genome-wide and transcriptome-wide
6 association analyses to identify genetic loci associated with immunoglobulin G antibody response to 28
7 antigens for 16 viruses using serological data from 7924 European ancestry participants in the UK Biobank
8 cohort.

9 **Results:** Signals in human leukocyte antigen (HLA) class II region dominated the landscape of viral
10 antibody response, with 40 independent loci and 14 independent classical alleles, 7 of which exhibited
11 pleiotropic effects across viral families. We identified specific amino acid (AA) residues that are associated
12 with seroreactivity, the strongest associations presented in a range of AA positions within DR β 1 at
13 positions 11, 13, 71, and 74 for Epstein-Barr Virus (EBV), Varicella Zoster Virus (VZV), Human Herpes
14 virus 7, (HHV7) and Merkel cell polyomavirus (MCV). Genome-wide association analyses discovered 7
15 novel genetic loci outside the HLA associated with viral antibody response ($P < 5.0 \times 10^{-8}$), including *FUT2*
16 (19q13.33) for human polyomavirus BK (BKV), *STING1* (5q31.2) for MCV, as well as *CXCR5* (11q23.3)
17 and *TBKBP1* (17q21.32) for HHV7. Transcriptome-wide association analyses identified 114 genes
18 associated with response to viral infection, 12 outside of the HLA region, including *ECSCR*: $P = 5.0 \times 10^{-15}$
19 (MCV), *NTN5*: $P = 1.1 \times 10^{-9}$ (BKV), and *P2RY13*: $P = 1.1 \times 10^{-8}$ EBV nuclear antigen. We also demonstrated
20 pleiotropy between viral response genes and complex diseases; from autoimmune disorders to cancer to
21 neurodegenerative and psychiatric conditions.

22 **Conclusions:** Our study confirms the importance of the HLA region in host response to viral infection and
23 elucidates novel genetic determinants beyond the HLA that contribute to host-virus interaction.

24 **KEY WORDS**

25 Infection, virus, serology, antigen, antibody, immunoglobulin G, immune response, human leukocyte
26 antigen (HLA), polyomavirus, genome-wide association study (GWAS), transcriptome-wide association
27 study (TWAS)

28 INTRODUCTION

29 Viruses have been infecting cells for a half a billion years¹. During our extensive co-evolution viruses have
30 exerted significant selective pressure on humans and vice versa; overtly during fatal outbreaks, and covertly
31 through cryptic immune interaction when a pathogen remains latent. The recent pandemic of severe acute
32 respiratory syndrome coronavirus 2 (SARS-CoV-2) highlights the paramount public health need to
33 understand human genetic variation in response to viral challenge. Clinical variation in COVID-19 severity
34 and symptomatic presentation may be due to differences host genetic factors relating to immune response².
35 Furthermore, many common infections are cryptically associated with a variety of complex illnesses,
36 especially those with an immunologic component, from cancer to autoimmune and neurologic conditions³⁻
37 ⁵. Despite their broad health relevance, few large-scale genome-wide association studies (GWAS) have
38 been conducted on serological response phenotypes⁶⁻¹⁰. Understanding the genetic architecture of
39 immunologic response to viruses may therefore provide new insight into etiologic mechanisms of diverse
40 complex diseases.

41 Several common viruses exert a robust cell mediated and humoral immune response that bi-directionally
42 modulate the balance between latent and lytic infection. Studies have demonstrated a strong heritable
43 component (32-48%) of antibody response¹¹ and identified associations between host polymorphisms in
44 genes relating to cell entry, cytokine production, and immune response and a variety of viruses¹². The
45 predominance of previously reported associations with have implicated genetic variants in human leucocyte
46 antigen (HLA) class I and II genes in the modulation of immune response to diverse viral antigens^{7,13}.

47 In this study we utilize data from the UK Biobank (UKB) cohort¹⁴ to evaluate the relationship between host
48 genetics and immunoglobulin G antibody response to 28 antigens for 16 viruses. Immunoglobulin G (IgG)
49 antibody is the most common antibody in blood, which serves as a stable biomarker of lifetime exposure to
50 common viruses. High levels of specific IgG's can be the result of chronic infection, while low levels may
51 indicate poor immunity. Viruses assayed in the UKB multiplex serology panel were previously chosen based
52 on putative links to chronic diseases including cancer, autoimmune, and neurodegenerative conditions¹⁵.
53 We conduct integrative genome-wide and transcriptome-wide analyses of antibody response and positivity

54 to viral antigens (**Figure 1**), which elucidate novel genetic underpinnings of viral infection and immune
55 response.

56 **METHODS**

57 *Study Population and Phenotypes*

58 The UK Biobank (UKB) is a population-based prospective cohort of over 500,000 individuals aged 40-69
59 years at enrollment in 2006-2010 who completed extensive questionnaires, physical assessments, and
60 provided blood samples¹⁴. Analyses were restricted to individuals of predominantly European ancestry
61 based on self-report and after excluding samples with any of the first two genetic ancestry principal
62 components (PCs) outside of 5 standard deviations (SD) of the population mean (**Supplementary Figure**
63 **1**). We removed samples with discordant self-reported and genetic sex, samples with call rates <97% or
64 heterozygosity >5 SD from the mean, and one sample from each pair of first-degree relatives identified
65 using KING¹⁶.

66 Of the 413,810 European ancestry individuals available for analysis, a total of 7948 had serological
67 measures. A multiplex serology panel (IgG) was performed over a 2-week period using previously
68 developed methods^{17,18} that have been successfully applied in epidemiological studies^{7,19}. Details of the
69 serology methods and assay validation performance are described in Mentzer et al.¹⁵ Briefly, multiplex
70 serology was performed using a bead-based glutathione S-transferase (GST) capture assay with
71 glutathione-casein coated fluorescence-labelled polystyrene beads and pathogen-specific GST-X-tag
72 fusion proteins as antigens¹⁵. Each antigen was loaded onto a distinct bead set and the beads were
73 simultaneously presented to primary serum antibodies at serum dilution 1:1000¹⁵. Immunocomplexes were
74 quantified using a Luminex 200 flow cytometer, which produced Median Fluorescence Intensities (MFI) for
75 each antigen. The serology assay showed adequate performance, with a median coefficient of variation
76 (CV) of 17% across all antigens and 3.5% among seropositive samples only¹⁵.

77 *Genome-Wide Association Analysis*

78 We evaluated the relationship between genetic variants across the genome and serological phenotypes
79 using PLINK 2.0 (October 2017 version). Participants were genotyped on the Affymetrix Axiom UK Biobank

80 array (89%) or the UK BiLEVE array (11%)¹⁴ with genome-wide imputation performed using the Haplotype
81 Reference Consortium data and the merged UK10K and 1000 Genomes phase 3 reference panels¹⁴. We
82 excluded variants out of Hardy-Weinberg equilibrium at $p < 1 \times 10^{-5}$, call rate $< 95\%$ (alternate allele dosage
83 within 0.1 of the nearest hard call to be non-missing), imputation quality $INFO < 0.30$, and $MAF < 0.01$.

84 Seropositivity for each antigen was determined using established cut-offs based on prior validation work¹⁵.
85 The primary GWAS focused on continuous phenotypes (MFI values), which measure the magnitude of
86 antibody response, also referred to as seroreactivity. These analyses were conducted among seropositive
87 individuals only for antigens with seroprevalence of $\geq 20\%$ ($n=1500$) based on 80% power to detect only
88 common variants with large effect sizes at this sample size (**Supplementary Figure 2**). MFI values were
89 transformed to standardized, normally distributed z-scores using ordered quantile normalization²⁰.

90 Seroreactivity GWAS was conducted using linear regression with adjustment for age at enrollment, sex,
91 body-mass index (BMI), socioeconomic status (Townsend deprivation index), the presence of any
92 autoimmune and/or inflammatory conditions, genotyping array, serology assay date, quality control flag
93 indicating sample spillover or an extra freeze/thaw cycle, and the top 10 genetic ancestry principal
94 components (PC's). Autoimmune and chronic inflammatory conditions were identified using the following
95 primary and secondary diagnostic ICD-10 codes (E10, M00-03, M05-M14, M32, L20-L30, L40, G35, K50-
96 52, K58, G61) in Hospital Episode Statistics. Individuals diagnosed with any immunodeficiency (ICD-10
97 D80-89, $n=24$) were excluded from all analyses.

98 For all antigens with at least 100 seropositive (or seronegative for pathogens with ubiquitous exposure)
99 individuals, GWAS of discrete seropositivity phenotypes was undertaken using logistic regression, adjusting
100 for the same covariates listed above.

101 The functional relevance of the lead GWAS loci for antibody response was assessed using in-silico
102 functional annotation analyses based on Combined Annotation Dependent Depletion (CADD)²¹ scores and
103 RegulomeDB 2.0²², and by leveraging external datasets, such as GTEx v8, DICE (Database of Immune
104 Cell Expression)²³, and the Human Plasma Proteome Atlas^{24,25}.

105

106 *Cross-Trait Associations with Disease*

107 We explored pleiotropic associations between lead variants influencing antibody levels and several chronic
108 diseases with known or hypothesized viral risk factors. Associations with selected cancers were obtained
109 from a cancer pleiotropy meta-analysis of the UK Biobank and Genetic Epidemiology Research on Aging
110 cohorts²⁶. Summary statistics for the schizophrenia GWAS of 33,640 cases and 43,456 controls by Lam et
111 al.²⁷ were downloaded from the Psychiatric Genomics Consortium. Association p-values were obtained
112 from the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site for the GWAS by
113 Jun et al.²⁸, which included 17,536 cases and 53,711 controls. Associations with $p < 7.3 \times 10^{-4}$ were
114 considered statistically significant after correction for the number of variants and phenotypes tested.

115 *HLA Regional Analysis*

116 For phenotypes displaying a genome-wide significant signal in the HLA region, independent association
117 signals were ascertained using two complementary approaches: clumping and conditional analysis.
118 Clumping is a post-processing step applied to GWAS summary statistics to identify independent association
119 signals by grouping variants based on LD within specific windows. Clumping was performed on all variants
120 with $P < 5 \times 10^{-8}$ for each phenotype, as well as across phenotypes. Clumps were formed around index
121 variants with the lowest p-value and all other variants with LD $r^2 > 0.05$ within a ± 500 kb window were
122 considered non-independent and assigned to that variant's clump.

123 Next, we conducted conditional analyses using a forward stepwise strategy to identify statistically
124 independent signals within each type of variant (SNP/indel or classical HLA allele). Unlike clumping,
125 conditional analyses involve fitting a new model that includes specific variants as covariates, thereby
126 directly accounting for LD and providing association estimates that are adjusted for other relevant SNP
127 effects. A total of 38,655 SNPs/indels on chromosome 6 (29,600,000 – 33,200,000 bp) were extracted to
128 conduct regional analyses. Classical HLA alleles were imputed for UKB participants at 4-digit resolution
129 using the HLA*IMP:02 algorithm¹⁴, with modified settings to accommodate the addition of diverse samples
130 from population reference panels described by Motyer et al.²⁹. Details of the HLA imputation procedure are
131 described in UKB Resource 182. Imputed dosages were available for 362 classical alleles in 11 genes:

132 *HLA-A*, *HLA-B*, and *HLA-C* (class I); *HLA-DRB5*, *HLA-DRB4*, *HLA-DRB3*, *HLA-DRB1*, *HLA-DQA1*, *HLA-*
133 *DQB1*, *HLA-DPA1*, and *HLA-DPB1* (class II). Allele names with “99:01” for DRB3/4/5, which denote copy
134 number absence, were renamed as “00:00” to avoid confusion with traditional HLA nomenclature. We also
135 used SNP2HLA³⁰ to impute HLA alleles and corresponding amino acid sequences at a four-digit resolution
136 in *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1* and *HLA-DPB1* using the Type
137 1 Diabetes Genetics Consortium (T1DGC) reference panel comprised of 2,767 unrelated individuals of
138 European descent. T1DGC was also among several reference datasets used by HLA*IMP:02.

139 Analyses were restricted to common HLA alleles and amino acid sequences (frequency ≥ 0.01) with
140 imputation quality scores >0.30 , for a total of 1081 markers (101 alleles + 980 amino acid residues). Linear
141 regression models were adjusted for the same set of covariates as the GWAS. Associations for each marker
142 were considered statistically significant if $P < 4.6 \times 10^{-5}$ based on Bonferroni correction for 1081 tests.

143 For each antigen response phenotype, we identified SNPs/indels or classical HLA alleles with the lowest
144 p-value, among variants that achieved Bonferroni-significant associations ($P < 4.6 \times 10^{-5}$), and performed
145 forward iterative conditional regression to identify other independent signals, until no associations with a
146 conditional p-value ($P_{\text{cond}} < 5 \times 10^{-8}$) remained. We also assessed the independence of associations across
147 different types of genetic variants by including conditionally independent HLA alleles as covariates in the
148 SNP-based analysis.

149 For amino acid positions with >2 possible residues (alleles), we applied the haplotype omnibus test to obtain
150 an overall p-value for jointly testing all possible substitutions at that specific position. The omnibus test was
151 applied to all amino acid residues at a given position, even if not all substitutions achieved the Bonferroni-
152 corrected threshold ($P < 4.6 \times 10^{-5}$) in the single-marker analysis. The frequency of amino acid substitutions
153 at specific HLA alleles was determined using European ancestry reference populations part of the Allele
154 Frequency Net Database (AFND 2020)³¹.

155

156

157

158 *Transcriptome-Wide Association Analysis*

159 Gene transcription levels were imputed and analyzed using the MetaXcan approach³², applied to GWAS
160 summary statistics for quantitative antigen phenotypes. For imputation, we used biologically informed
161 MASHR-M prediction models³³ based on GTEx v8 with effect sizes computed using MASHR (Multivariate
162 Adaptive Shrinkage in R)³⁴ for variants fine-mapped with DAP-G (Deterministic Approximation of
163 Posteriors)^{35,36}. An advantage of this approach is that MASHR effect sizes are smoothed by taking
164 advantage of the correlation in cis-eQTL effects across tissues. For each antigen, we performed a
165 transcriptome-wide association study (TWAS) using gene expression levels in whole blood. Statistically
166 significant associations for each gene were determined based on Bonferroni correction for the number of
167 genes tested.

168 We also examined gene expression profiles in tissues that represent known infection targets or related
169 pathologies. Human herpesviruses and polyomaviruses are neurotropic and have been implicated in
170 several neurological conditions^{37,38}, therefore we considered gene expression in the frontal cortex. For
171 Epstein-Barr virus (EBV) antigens additional models included EBV-transformed lymphocytes. Merkel cell
172 polyomavirus (MCV) is a known cause of Merkel cell carcinoma³⁹, a rare but aggressive type of skin cancer,
173 therefore we examined transcriptomic profiles in skin tissues for MCV only.

174 Pathways represented by genes associated with antibody response to viral antigens were summarized by
175 conducting enrichment analysis using curated Reactome gene sets and by examining protein interaction
176 networks using the STRING database⁴⁰. Significantly associated TWAS genes were grouped by virus family
177 (herpesviruses vs. polyomaviruses) and specificity of association (multiple antigens vs. single antigen).

178 **RESULTS**

179 A random sample of the participants representative of the full UKB cohort was assayed using a multiplex
180 serology panel¹⁵. We analyzed data from 7924 participants of predominantly European ancestry, described
181 in **Supplementary Table 1**. Approximately 90% of individuals were seropositive for herpes family viruses
182 with ubiquitous exposure: EBV (EBV EA-D: 86.2% to ZEBRA: 91.2%), Human Herpesvirus 7 (HHV7
183 94.8%), and Varicella Zoster Virus (VZV 92.3%). Seroprevalence was somewhat lower for cytomegalovirus

184 (CMV), ranging between 56.5% (CMV pp28) and 63.3% (CMV pp52), and Herpes Simplex virus-1 (HSV1
185 69.3%). Human polyomavirus BKV was more prevalent (95.3%) compared to other polyomaviruses, Merkel
186 cell polyoma virus (MCV 66.1%) and polyomavirus JC (JCV) (56.6%). Less common infections included
187 HSV-2 (15.2%), HPV16 (E6 and E7 oncoproteins: 4.7%), HPV18 (2.4%), Human T-cell lymphotropic virus
188 type 1 (HTLV1, 1.6%), Hepatitis B (HBV, 1.6%), and Hepatitis C (HCV, 0.3%).

189 *Genetic Determinants of Response to Viral Infection*

190 Results from our GWAS of antibody response phenotypes were dominated by signals in the HLA region,
191 which were detected for all EBV antigens (EA-D, EBNA, p18, ZEBRA), CMV pp52, HSV1, HHV7, VZV, JCV
192 and MCV (**Table 1; Supplementary Figure 3**). Most of the top-ranking HLA variants for each antigen were
193 independent of those for other antigens based on r^2 but not D' (**Supplementary Figure 4**). Exceptions were
194 moderate LD between lead variants for EBV ZEBRA and HSV1 ($r^2=0.45$), EBV EBNA and JCV ($r^2=0.45$),
195 and HHV7 and MCV ($r^2=0.44$). However, based on the complex LD structure and effect sizes, we cannot
196 rule out that these linked to rare haplotypes. Outside of the HLA region, genome-wide significant
197 associations with seroreactivity were detected for: MCV at 3p24.3 (rs776170649, *LOC339862*: $P=1.7\times 10^{-8}$)
198 and 5q31.2 (rs7444313, *TMEM173* (also known as *STING1*): $P=2.4\times 10^{-15}$); BKV at 19q13.3 (rs681343,
199 *FUT2*: $P=4.7\times 10^{-15}$) (**Figure 2**); EBV EBNA at 3q25.1 (rs67886110, *MED12L*: $P=1.3\times 10^{-9}$); HHV-7 at
200 11q23.3 (rs75438046, *CXCR5*: $P=1.3\times 10^{-8}$) and 17q21.3 (rs1808192, *TBKBP1*: $P=9.8\times 10^{-9}$); and HSV-1 at
201 10q23.3 (rs11203123: $P=3.9\times 10^{-8}$). However, the loci outside of HLA identified for HHV7 and HSV1 were
202 not statistically significant considering a more stringent significance threshold corrected for the number of
203 seroreactivity phenotypes tested ($P < 5.0\times 10^{-8}/16 = 3.1\times 10^{-9}$).

204 GWAS of discrete seropositivity phenotypes identified associations in HLA for EBV EA-D (rs2395192:
205 OR=0.66, $P=4.0\times 10^{-19}$), EBV EBNA (rs9268848: OR=1.60, $P=1.2\times 10^{-18}$), EBV ZEBRA (rs17211342: OR=0.63,
206 $P=1.6\times 10^{-15}$), VZV (rs3096688: OR=0.70, $P=3.7\times 10^{-8}$), JCV (rs9271147: OR=0.54, $P=1.3\times 10^{-42}$), and MCV
207 (rs17613347: OR=0.61, $P=1.2\times 10^{-26}$) (**Supplementary Figure 2; Supplementary Table 3**). An association
208 with susceptibility to MCV infection was also observed at 5q31.2 (rs1193730215, *ECSCR*: OR=1.26,
209 $P=7.2\times 10^{-9}$), with high LD ($r^2=0.95$) between seroreactivity and seropositivity lead variants.

210 Several genome-wide significant associations were observed for antigens with <20% seroprevalence,
211 which were not included in the GWAS of antibody response due to inadequate sample size
212 (**Supplementary Table 3**). Infection susceptibility variants were identified for HSV2 in 17p13.2 (rs2116443:
213 OR=1.28, $P=4.5\times 10^{-8}$; *ITGAE*); HPV16 E6 and E7 oncoproteins in 6p21.32 (rs601148: OR=0.60, $P=3.3\times 10^{-$
214 9 ; *HLA-DRB1*) and 19q12 (rs144341759: OR=0.383, $P=4.0\times 10^{-8}$; *CTC-448F2.6*); and HPV18 in 14q24.3
215 (rs4243652: OR=3.13, $P=7.0\times 10^{-10}$). Associations were also detected for Kaposi's sarcoma-associated
216 herpesvirus (KSHV), HTLV1, HBV and HCV, including a variant in the *MERTK* oncogene (HCV Core
217 rs199913364: OR=0.25, $P=1.2\times 10^{-8}$). After correcting for 28 serostatus phenotypes tested ($P<1.8\times 10^{-9}$),
218 the only statistically significant associations remained for EBV EA-D (rs2395192), EBV EBNA (rs9268848),
219 EBV ZEBRA (rs17211342), JCV (rs9271147), MCV (rs17613347), and HPV18 (rs4243652).

220 *Functional Characterization of GWAS Findings*

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

228 Eleven sentinel variants were eQTLs and 8 were splicing QTLs in GTEx, with significant (FDR<0.05) effects
229 across multiple genes and tissues (**Supplementary Figure 5**). The most common eQTL and sQTL targets
230 included *HLA-DQA1*, *HLA-DQA2*, *HLA-DQB1*, *HLA-DQB2*, *HLA-DRB1*, and *HLA-DRB6*. Outside of HLA,
231 rs681343 (BKV), a synonymous *FUT2* variant was an eQTL for 8 genes, including *FUT2* and *NTN5*. MCV
232 variant in 5q31.2, rs7444313, was an eQTL for 7 genes, with concurrent sQTL effects on *TMEM173*, also
233 known as *STING1* (stimulator of interferon response cGAMP interactor 1) and *CXXC5*. Gene expression
234 profiles in immune cell populations from DICE²³ identified several cell-type specific effects that were not
235 observed in GTEx. An association with *HLA-DQB1* expression in CD4+ T_H2 cells was observed for

236 rs9273325, 6:31486158_GT_G was an eQTL for *ATP6V1G2* in naïve CD4+ T cells, and rs1130420
237 influenced the expression of 8 HLA class II genes in naïve B-cells and CD4+ T_H17 cells.

238 We identified 7 significant ($p < 5.0 \times 10^{-8}$) protein quantitative trait loci (pQTL) for 38 proteins (**Supplementary**
239 **Table 5**). Most of the pQTL targets were components of the adaptive immune response, such as the
240 complement system (C4, CFB), chemokines (CCL15, CCL25), and defensin processing (Beta-defensin 19,
241 Trypsin-3). The greatest number and diversity of pQTL targets ($n=16$) was observed for rs681343, including
242 BPIFB1, which plays a role in antimicrobial response in oral and nasal mucosa⁴¹; FUT3, which catalyzes
243 the last step of Lewis antigen biosynthesis; and FGF19, part of the PI3K/Akt/MAPK signaling cascade that
244 is dysregulated in cancer and neurodegenerative diseases⁴².

245 *Cross-trait associations with disease outcomes*

246 To contextualize the relevance of genetic loci involved in infection response, we explored associations with
247 selected cancers, schizophrenia, and that have a known or suspected viral etiology (**Supplementary Table**
248 **6**). The strongest secondary signal was observed for rs9273325 (*HLA-DQB1*), which was negatively
249 associated with VZV antibody response and positively associated with schizophrenia susceptibility
250 ($OR=1.13$, $P=4.3 \times 10^{-15}$). Other significant (Bonferroni $P < 7.4 \times 10^{-4}$) associations with schizophrenia were
251 detected for HSV1 (rs1130420: $OR=1.06$, $P=1.8 \times 10^{-5}$), EBV EA-D (rs2647006: $OR=0.96$, $P=2.7 \times 10^{-4}$), JCV
252 (rs9271525: $OR=1.06$, $P=6.8 \times 10^{-5}$) and BKV (rs681343: $OR=0.96$, $P=2.5 \times 10^{-4}$), with the latter being the
253 only pleiotropic signal outside of HLA. Inverse associations with hematologic cancers were observed for
254 HSV1 (rs1130420: $OR=0.89$, $P=3.5 \times 10^{-6}$), VZV (rs9273325: $OR=0.88$, $P=4.4 \times 10^{-5}$), and EBV EBNA
255 (rs9269233: $OR=0.88$, $P=2.7 \times 10^{-4}$) variants. HSV1 antibody response was also linked to Alzheimer's
256 disease (rs1130420: $P=1.2 \times 10^{-4}$).

257 *Regional HLA Associations*

258 Associations within the HLA region were refined by identifying independent (LD $r^2 < 0.05$ within ± 500 kb)
259 index variants with $P < 5.0 \times 10^{-8}$ for each antigen response phenotype (**Supplementary Table 7**). Clumping
260 seropositivity associations with respect to lead antibody response variants did not retain any loci,

261 suggesting non-independence in signals for infection and reactivity for the same antigen. For this reason,
262 all subsequent analyses focus on seroreactivity phenotypes. Clumping across phenotypes to assess the
263 independence of HLA associations for different antigens identified 40 independent index variants: EBV
264 EBNA (12), VZV (11), EBV ZEBRA (8), EBV p18 (5), MCV (3), and EBV EA-D (1) (**Supplementary Table**
265 **9**). No LD clumps were anchored by variants detected for CMV pp52, HHV7, HSV1, or JCV, suggesting
266 that the HLA signals for these antigens are captured by lead loci for other phenotypes. The largest region
267 with the lowest p-value was anchored by rs9274728 ($P=4.7\times 10^{-67}$) near *HLA-DQB1*, originally detected for
268 EBV ZEBRA. Of the 11 VZV-associated variants, the largest clump was formed around rs4990036
269 ($P=4.5\times 10^{-26}$) in *HLA-B*.

270 Iterative conditional analyses adjusting for the HLA SNP/indel with the lowest p-value were performed until
271 no variants remained with $P_{\text{cond}} < 5.0\times 10^{-8}$. Additional independent variants were identified for EBV EBNA
272 (rs139299944, rs6457711, rs9273358, rs28414666, rs3097671), EBV ZEBRA (rs2904758, rs35683320,
273 rs1383258), EBV p18 (rs6917363, rs9271325, rs66479476), and MCV (rs148584120, rs4148874) (**Figure**
274 **3; Supplementary Table 8**). For CMV pp52, HHV7, HSV1, JCV, and VZV, the regional HLA signal was
275 captured by the top GWAS variant (**Figure 2; Supplementary Table 8**).

276 Next, we tested 101 classical HLA alleles and performed analogous iterative conditional analyses for
277 significantly associated variants ($P < 4.6\times 10^{-5}$). To help with the interpretation of our results, we depict the
278 LD structure for HLA alleles in class II genes in **Supplementary Figure 5**. Significant associations across
279 viruses were predominantly observed for class II HLA alleles. Five statistically independent signals were
280 identified for antibody response to EBV ZEBRA (DRB4*00:00: $\beta = -0.246$, $P = 1.4\times 10^{-46}$; DQB1*04:02:
281 $\beta_{\text{cond}} = 0.504$, $P_{\text{cond}} = 1.0\times 10^{-19}$; DRB1*04:04: $\beta_{\text{cond}} = 0.376$, $P_{\text{cond}} = 1.1\times 10^{-18}$; DQA1*02:01: $\beta_{\text{cond}} = 0.187$,
282 $P_{\text{cond}} = 1.1\times 10^{-10}$, A*03:01: $\beta_{\text{cond}} = 0.129$, $P_{\text{cond}} = 1.9\times 10^{-8}$) (**Figure 3; Supplementary Table 11**). DRB4*00:00
283 represents copy number absence, which co-occurs with DRB1*04 and DRB1*07 alleles⁴³. This is consistent
284 with the magnitude and direction of unconditional associations observed for DRB1*07:01 ($\beta = 0.251$,
285 $P = 1.3\times 10^{-26}$) and DRB4*04:01 ($\beta = 0.293$, $P = 7.9\times 10^{-22}$). Five conditionally independent alleles were also
286 identified for EBV EBNA: DRB5*00:00: $\beta = -0.246$, $P = 8.7\times 10^{-30}$; DRB3*02:02: $\beta_{\text{cond}} = 0.276$, $P_{\text{cond}} = 6.8\times 10^{-30}$;
287 DQB1*02:01: $\beta_{\text{cond}} = -0.164$, $P_{\text{cond}} = 3.6\times 10^{-12}$; DRB4*00:00: $\beta = 0.176$, $P_{\text{cond}} = 8.3\times 10^{-17}$; DPB1*03:01: $\beta_{\text{cond}} = -$

288 0.220, $P_{\text{cond}}=4.7\times 10^{-14}$ (**Figure 3; Supplementary Table 11**). DRB5*00:00 denotes a copy number deletion
289 that sits on an common haplotype comprised of DRB1*15:01, DQB1*06:02, DQA1*01:02⁴³, which may also
290 include DRB5*01:01⁴⁴ (**Supplementary Figure 6**). The presence of the DRB1*15:01-DQB1*06:02-
291 DQA1*01:02 haplotype was associated with increased EBV EBNA seroreactivity ($\beta=0.330$, $P=2.5\times 10^{-28}$).
292 Fewer independent alleles were observed for EBV p18 (DRB5*00:00: $\beta=-0.210$, $P=1.7\times 10^{-22}$; DRB1*04:04:
293 $\beta_{\text{cond}}=0.357$, $P_{\text{cond}}=1.3\times 10^{-18}$) (**Figure 3; Supplementary Tables 12**).

294 DQB1*02:01 was the only independently associated allele for EBV EA-D ($\beta=-0.154$, $P=8.4\times 10^{-11}$) and HSV1
295 ($\beta=0.145$, $P=2.8\times 10^{-8}$), although its effects were in opposite directions for each antigen (**Supplementary**
296 **Table 13**). For VZV, associations with 16 classical alleles were accounted for by DRB1*03:01 ($\beta=0.236$,
297 $P=7.3\times 10^{-26}$). JCV shared the same lead allele as EBV EBNA and EBV p18 (DRB5*00:00: $\beta=0.350$,
298 $P=1.2\times 10^{-21}$) (**Supplementary Table 13**). Four conditionally independent signals were identified for MCV
299 (DQA1*01:01: $\beta=0.215$, $P=1.1\times 10^{-15}$; DRB1*04:04: $\beta_{\text{cond}}=-0.362$, $P_{\text{cond}}=3.0\times 10^{-11}$; A*29:02: $\beta_{\text{cond}}=-0.350$,
300 $P=1.0\times 10^{-11}$; DRB1*15:01: $\beta_{\text{cond}}=-0.203$, $P=3.7\times 10^{-12}$) (**Figure 3; Supplementary Table 14**). Lastly, we
301 integrated associations across variant types by including conditionally independent HLA alleles as
302 covariates in the SNP-based analysis. With the exception of EBV antigens and HHV7, classical HLA alleles
303 captured all genome-wide significant SNP signals (**Supplementary Figure 7**).

304 Finally, we tested 980 HLA amino acid substitutions (**Supplementary Tables 15-24**), followed by omnibus
305 haplotype tests at each position that had a significant amino acid and more than two possible alleles. The
306 strongest allele-specific and haplotype associations were found at different positions in the same protein
307 for EBV p18 (DR β 1 Ala -17: $\beta=-0.194$, $P=1.0\times 10^{-21}$; DR β 1 (13): $P_{\text{omni}}=4.6\times 10^{-22}$), MCV (DQ β 1 Leu-26: $\beta=-$
308 0.173, $P=7.0\times 10^{-18}$; DQ β 1 (125): $P_{\text{omni}}=2.0\times 10^{-17}$), HHV7 (DQ β 1 His-30: $\beta=-0.111$, $P=1.2\times 10^{-8}$; DQ β 1 (57):
309 $P_{\text{omni}}=5.6\times 10^{-9}$), and HHV6 IE1B at (DR β 1 Ile-67: $\beta=0.131$, $P=1.6\times 10^{-8}$; DR β 1 (13): $P_{\text{omni}}=1.1\times 10^{-5}$).

310 The strongest residue-specific and haplotype associations mapped to the same amino acid position for four
311 phenotypes: EBV ZEBRA (**Supplementary Table 18**), HHV6 IE1A (**Supplementary Table 19**), HSV1
312 (**Supplementary Table 21**), and JCV (**Supplementary Table 23**). Amino acid residues at DQ α 1 (175) were
313 associated with antibody response to EBV ZEBRA (Glu: $\beta=0.279$, $P=1.1\times 10^{-61}$; $P_{\text{omni}}=8.3\times 10^{-62}$). Glu-175 is

314 present in DQA1*02:01 ($P=4.9\times 10^{-27}$), DQA1*03:01 ($P=1.3\times 10^{-16}$), DQA1*04:01 ($P=1.9\times 10^{-12}$), and seems
315 to better summarize the EBV ZEBRA signal at this locus. Substitutions in DR β 1 (96) contained the strongest
316 predictors of JCV seroreactivity (His or Tyr: $\beta=0.325$, $P=1.6\times 10^{-25}$; $P_{\text{omni}}=7.7\times 10^{-23}$). His-96/Tyr-96 are in
317 high LD ($r^2=0.92$) with DRB5*00:00, the top JCV-associated allele. However, this might mask the signal for
318 Gln-96 ($\beta=-0.310$, $P=9.0\times 10^{-23}$), which is part of the DRB1*15:01 sequence ($\beta=-0.309$, $P=9.0\times 10^{-21}$; LD
319 $r^2=0.94$). The lead signal for HSV1 mapped to DQ β 1 (57) (Ala: $\beta=0.123$, $P=2.2\times 10^{-10}$; $P_{\text{omni}}=6.5\times 10^{-9}$), which
320 aligns with the association for the lead HSV1-allele DQB1*02:01.

321 For EBV EBNA the strongest haplotype association was in DR β 1 (37) ($P_{\text{omni}}=1.1\times 10^{-55}$), while the residue
322 with the lowest p-value was DQ β 1 Ala-57 ($\beta=-0.237$, $P=1.4\times 10^{-42}$) (**Supplementary Table 16**). Ala-57 maps
323 to multiple DQB1 alleles and achieved a stronger signal for EBV EBNA than any classical HLA allele. Asp-
324 9 in HLA-B showed the strongest association with antibody response to EBV EA-D ($\beta=-0.146$, $P=1.8\times 10^{-9}$;
325 **Supplementary Table 15**) and VZV ($\beta=0.237$, $P=9.7\times 10^{-25}$; **Supplementary Table 22**). This amino acid
326 sequence is part of B*08:01, which had analogous effects on both phenotypes (EBV EA-D: $\beta=-0.144$,
327 $P=2.7\times 10^{-9}$; VZV: $\beta=0.238$, $P=4.7\times 10^{-25}$). Haplotypes with the lowest overall p-values were found in DQ β 1
328 (71) for VZV ($P_{\text{omni}}=9.8\times 10^{-19}$) and DR β 1 (11) for EBV EA-D ($P_{\text{omni}}=1.7\times 10^{-10}$).

329 *TWAS of Genes Involved in Antibody Response*

330 Based on known targets of infection or related pathologies, we considered expression in the frontal cortex
331 (**Supplementary Table 25**), EBV-transformed lymphocytes for EBV antigens (**Supplementary Table 26**),
332 and skin for MCV (**Supplementary Table 27**). Concordance across tissues was summarized using Venn
333 diagrams (**Figure 4; Supplementary Figure 8**). TWAS identified 114 genes significantly associated
334 ($P_{\text{TWAS}}<4.2\times 10^{-6}$) with antibody response in at least one tissue, 54 of which were associated with a single
335 phenotype, while 60 influenced seroreactivity to multiple antigens. We also include results for 87 additional
336 suggestively ($P_{\text{TWAS}}<4.2\times 10^{-5}$) associated genes.

337 The TWAS results included a predominance of associations in HLA class II genes. Some of the strongest
338 overall associations were observed for *HLA-DRB5* (EBV ZEBRA: $P_{\text{cortex}}=4.2\times 10^{-45}$) and *HLA-DRB1* (EBV
339 EBNA: $P_{\text{cortex}}=6.7\times 10^{-39}$; EBV ZEBRA: $P_{\text{cortex}}=3.3\times 10^{-33}$; JCV: $P_{\text{cortex}}=6.5\times 10^{-14}$; EBV p18: $P_{\text{cortex}}=2.2\times 10^{-12}$).

340 Increased expression of *HLA-DQB2* was positively associated with antibody response to EBV ZEBRA
341 ($P_{\text{blood}}=7.6\times 10^{-19}$), JCV ($P_{\text{blood}}=9.9\times 10^{-10}$), VZV ($P_{\text{blood}}=7.0\times 10^{-9}$), HHV7 ($P_{\text{blood}}=7.3\times 10^{-8}$), and HSV1
342 ($P_{\text{blood}}=3.3\times 10^{-7}$), but negatively associated with EBV EBNA ($P_{\text{blood}}=3.6\times 10^{-34}$) and EBV p18 ($P_{\text{blood}}=2.1\times 10^{-8}$), in a consistent manner across tissues. The opposite was observed for *HLA-DQB1*, with positive effects
343 on EBV EBNA and EBV p18 and inverse associations with EBV ZEBRA, JCV, VZV, HHV7, and HSV1.
344

345 The TWAS analyses also identified a number of significant associations in the HLA class III region that
346 were not detected in other analyses. The top-ranking VZV associated gene was *APOM* ($P_{\text{blood}}=7.5\times 10^{-27}$,
347 $P_{\text{cortex}}=1.1\times 10^{-25}$). Interestingly, opposite directions of effect were observed for *C4A* and *C4B* gene
348 expression. Increased *C4A* expression was positively associated with all EBV antigens (**Supplementary**
349 **Table 26**), but negatively associated with VZV ($P_{\text{blood}}=2.3\times 10^{-24}$) and HSV1 ($P_{\text{cortex}}=1.8\times 10^{-5}$) antibody levels
350 (**Supplementary Table 25**). On the other hand, increased *C4B* expression was inversely associated with
351 EBV phenotypes, but positively associated with VZV ($P_{\text{blood}}=8.1\times 10^{-25}$) and HSV1 ($P_{\text{blood}}=1.1\times 10^{-5}$). A similar
352 pattern was also observed for *CYP21A2* and *C2*, with positive effects on antibody response to VZV and
353 HSV1, and negative effects for all EBV antigens. Other novel TWAS findings were detected for HHV7 in
354 22q13.2 (*CTA-223H9.9*: $P_{\text{TWAS}}=2.5\times 10^{-6}$; *CSDC2*: $P_{\text{TWAS}}=3.0\times 10^{-6}$; *TEF*: $P_{\text{TWAS}}=3.1\times 10^{-6}$) and 1q31.2
355 (*RGS1*: $P_{\text{TWAS}}=3.3\times 10^{-6}$).

356 The TWAS recapitulated several GWAS-identified loci: 3q25.1 for EBV EBNA (*P2RY13*: $P_{\text{cortex}}=1.1\times 10^{-8}$;
357 *P2RY12*: $P_{\text{blood}}=3.3\times 10^{-8}$) and 19q13.33 for BKV (*FUT2*: $P_{\text{TWAS}}=8.1\times 10^{-13}$; *NTN5*: $P_{\text{TWAS}}=1.1\times 10^{-9}$).
358 Transcriptomic profiles in skin tissues provided supporting evidence for the role of multiple genes in 5q31.2
359 in modulating MCV antibody response (**Figure 4**; **Supplementary Table 27**). The strongest signal was
360 observed in for *ECSCR* (skin sun unexposed: $P_{\text{TWAS}}=5.0\times 10^{-15}$; skin sun exposed: $P_{\text{TWAS}}=4.2\times 10^{-13}$),
361 followed by *PROB1* (sun unexposed: $P_{\text{TWAS}}=1.5\times 10^{-11}$). *ECSCR* expression was also associated based on
362 expression in the frontal cortex, while *PROB1* exhibited a significant, but attenuated effect in whole blood.
363 *VWA7* was the only gene associated across all four tissues for MCV and was also associated with antibody
364 response to several EBV antigens.

365 Comparison of results for seroreactivity and seropositivity revealed a number of genes implicated in both
366 steps of the infection process (**Supplementary Table 28**). Associations with HLA DQA and DQB genes in

367 whole blood and HLA-DRB genes in the frontal cortex were observed for EBV antigens, JCV, and MCV.
368 For MCV, the strongest seropositivity signals were observed for HLA class III genes *AGER* ($P_{\text{cortex}}=9.0\times 10^{-21}$) and *EHMT2* ($P_{\text{blood}}=5.8\times 10^{-18}$), which were also among the top-ranking genes for seroreactivity.
369
370 Increased *ECSCR* expression conferred an increased susceptibility to MCV infection ($P_{\text{cortex}}=1.8\times 10^{-8}$),
371 mirroring its effect on seroreactivity. In contrast to antibody response, no significant associations with any
372 HLA genes were observed for VZV seropositivity.

373 Analyses using the Reactome database identified significant ($q_{\text{FDR}}<0.05$) enrichment for TWAS-identified
374 genes in pathways involved in initiating antiviral responses, such as MHC class II antigen presentation,
375 TCR signaling, and interferon (IFN) signaling (**Supplementary Figure 9**). Pathways unique to
376 herpesviruses included folding, assembly and peptide loading of class I MHC ($q=3.2\times 10^{-7}$) and initial
377 triggering of complement ($q=9.8\times 10^{-3}$). Polyomaviruses were associated with the non-canonical nuclear
378 factor (NF)- κ B pathway activated by tumor necrosis factor (TNF) superfamily ($q=1.9\times 10^{-3}$).

379 **DISCUSSION**

380 We performed genome-wide and transcriptome-wide association studies for serological phenotypes for 16
381 common viruses in a well-characterized, population-based cohort. We discovered novel genetic
382 determinants of viral antibody response beyond the HLA region for BKV, MCV, HHV7, EBV EBNA.
383 Consistent with previous studies^{7,8} we detected strong signals for immune response to diverse viral antigens
384 in the HLA region, with a predominance of associations observed for alleles and amino acids in *HLA-DRB1*
385 and *HLA-DQB1*, as well as transcriptome-level associations for multiple class II and III HLA genes. Taken
386 together, the findings of this work provide a resource for further understanding the complex interplay
387 between viruses and the human genome, as well as a first step towards understanding genetic
388 determinants of reactivity to common infections.

389 One of our main findings is the discovery of 5q31.2 as a susceptibility locus for MCV infection and MCV
390 antibody response, implicating two main genes: *TMEM173* (or *STING1*) and *ECSCR*. The former encodes
391 STING (stimulator of interferon genes), an endoplasmic reticulum (ER) protein that controls the transcription
392 of host defense genes and plays a critical role in response to DNA and RNA viruses⁴⁵. STING is activated

393 by cyclic GMP-AMP synthase (cGAS), a cytosolic DNA sensor that mounts a response to invading
394 pathogens by inducing IFN1 and NF- κ B signalling^{46,47}. Polyomaviruses penetrate the ER membrane during
395 cell entry, a process that may be unique to this viral family⁴⁸, which may trigger STING signaling in a distinct
396 manner from other viruses⁴⁸. Multiple cancer-causing viruses, such as KSHV, HBV, and HPV18, encode
397 oncoproteins that disrupt cGAS-STING activity, which illustrates the evolutionary pressure on DNA tumor
398 viruses to develop functions against this pathway and its importance in carcinogenesis⁴⁶. Furthermore,
399 cGAS-STING activation has been shown to trigger antitumor T-cell responses, a mechanism that can be
400 leveraged by targeted immunotherapies⁴⁹⁻⁵¹. Several studies suggest STING agonists may be effective
401 against tumors resistant to PD-1 blockade, as well as promising adjuvants in cancer vaccines⁵²⁻⁵⁴.

402 *ECSCR* expression in skin and brain tissues was associated with MCV antibody response and infection.
403 This gene encodes an endothelial cell-specific chemotaxis regulator, which plays a role in angiogenesis
404 and apoptosis⁵⁵. *ECSCR* is a negative regulator of PI3K/Akt signaling by enhancing membrane localization
405 of *PTEN* and operates in tandem with VEGFR-2 and other receptor tyrosine kinases⁵⁶. In addition to 5q31.2,
406 another novel MCV seroreactivity associated region was identified in 3p24.3, anchored by rs776170649,
407 which has been linked to platelet phenotypes⁵⁷. These findings align with a role of platelet activation in
408 defense against infections via degranulation-mediated release of chemokines and β -defensin⁵⁸.

409 Genetic variation within Fucosyltransferase 2 (*FUT2*) has been studied extensively in the context of human
410 infections; however, its effect on BKV seroreactivity is novel. Homozygotes for the nonsense mutation
411 (rs601338 G>A) that inactivates the *FUT2* enzyme are unable to secrete ABO(H) histo-blood group
412 antigens or express them on mucosal surfaces^{59,60}. The allele which confers increased BKV antibody
413 response (rs681343-T) is in LD ($r^2=1.00$) with rs601338-A, the non-secretor allele, which confers resistance
414 to norovirus^{61,62}, rotavirus⁶³, *H. pylori*⁶⁴, childhood ear infection, mumps, and common colds¹³. However,
415 increased susceptibility to other pathogens, such as meningococcus and pneumococcus⁶⁵ has also been
416 observed in non-secretors. Isolating the underlying mechanisms for BKV response is challenging because
417 *FUT2* is a pleiotropic locus associated with diverse phenotypes, including autoimmune and inflammatory
418 conditions^{66,67}, serum lipids⁶⁸, B vitamins^{60,69}, alcohol consumption⁷⁰, and even certain cancers⁷¹. In addition
419 to *FUT2* in 19q13.33, *NTN5* (netrin 5) suggests a possible link between BKV and neurological conditions.

420 NTN5 is primarily expressed in neuroproliferative areas, suggesting a role in adult neurogenesis, which is
421 dysregulated in glioblastoma and Alzheimer's disease^{72,73}.

422 We also report the first GWAS of serological phenotypes for HHV7. Genetic determinants of HHV7 antibody
423 response in 6p21.32 were predominantly localized in *HLA-DQA1* and *HLA-DQB1*, with associations similar
424 to other herpesviruses. In 11q23.3, rs75438046 maps to the 3' UTR of *CXCR5*, which controls viral infection
425 in B-cell follicles⁷⁴, and *BCL9L*, a translocation target in acute lymphoblastic leukemia⁷⁵ and transcriptional
426 activator of the Wnt/ β -catenin cancer signaling pathway⁷⁶. In 17q21.32, *TBKBP1* encodes an adaptor
427 protein that binds to TBK1 and is part of the TNF/NF- κ B interaction network, where it regulates immune
428 responses to infectious triggers, such as IFN1 signaling⁷⁷. Interestingly, a protein interactome map recently
429 revealed that SARS-CoV-2 nonstructural protein 13 (Nsp13) includes TBK1-TBKBP1 among its targets⁷⁸.
430 Other functions of the TBK1-TBKBP1 axis relate to tumor growth and immunosuppression through induction
431 of PD-L1⁷⁹.

432 Several additional genes involved in HHV7 immune response were identified in TWAS. *TEF* in 22q13.2 is
433 an apoptotic regulator of hematopoietic progenitors with tumor promoting effects mediated by inhibition of
434 G1/S cell cycle transition and Akt/FOXO signaling⁸⁰. *RGS1* in 1q31.2 has been linked to multiple
435 autoimmune diseases, including multiple sclerosis⁸¹, as well as poor prognosis in melanoma and diffuse
436 large B cell lymphoma mediated by inactivation of Akt/ERK^{82,83}.

437 Other genes outside of the HLA region associated with viral infection response were detected for EBV
438 EBNA in 3q25.1. The lead variant (rs67886110) is an eQTL for *MED12L* and *P2RY12* genes, which have
439 been linked to neurodegenerative conditions^{84,85}. *P2RY12* and *P2RY13*, identified in TWAS, are purinergic
440 receptor genes that regulate microglia homeostasis and have been implicated in Alzheimer's susceptibility
441 via inflammatory and neurotrophic mechanisms⁸⁵.

442 Considering genetic variation within the HLA region, our results confirm its pivotal role at the interface of
443 host pathogen interactions and highlight the extensive sharing of HLA variants that mediate these
444 interactions across virus families and antigens. Genes in this region code for cell-surface proteins that
445 facilitate antigenic peptide presentation to immune cells that regulate responses to invading pathogens.

446 This region is critical for adaptive immune response but also has significant overlap with susceptibility
447 alleles for autoimmune diseases. We identified 40 independent SNPs/indels associated with EBV (EBNA,
448 EA-D, VCA p18, and ZEBRA), VZV, and MCV antibody response that accounted for all significant HLA
449 associations for other phenotypes. Of the 14 conditionally independent, genome-wide significant classical
450 alleles identified for 10 antigens, 7 were associated with multiple phenotypes. The most commonly shared
451 HLA alleles were DRB5*00:00, DRB1*04:04, an known rheumatoid arthritis risk allele⁸⁶, and DQB1*02:01,
452 associated with celiac disease risk⁸⁷. Copy number deletion represented by DRB5*00:00 may itself have a
453 functional role in altering response by the absence of these alleles. DRB5*00:00 also summarizes signals
454 from multiple HLA loci, including the extended DRB5*01:01-DRB1*15:01-DQB1*06:02-DQA1*01:02
455 haplotype that has been implicated in the etiology of multiple autoimmune diseases and EBV EBNA IgG
456 levels. DRB1*15:01-DQB1*06:02-DQA1*01:02 is protective for type 1 diabetes⁸⁸, while DRB5*01:01-
457 DRB15:01 confers the strongest risk for developing multiple sclerosis⁸¹. Amino acid residues in DRβ1 at
458 positions 11, 13, 71, and 74 and in DQβ1 codon 57 represent established susceptibility loci for rheumatoid
459 arthritis⁸⁹, type 1 diabetes⁹⁰, and multiple sclerosis⁹¹ that exhibited strong associations with IgG levels for
460 EBV, HHV7, VZV, JCV, and MCV antigens, and in some cases harbored the top signal of all HLA variants.
461 Further research is needed to delineate shared genetic pathways that invoke autoimmunity and influence
462 viral response.

463 Despite the predominance of association in HLA class II, several notable associations in HLA class I were
464 detected. A*29:02 conferred reduced MCV seroreactivity and its sequence overlaps with amino acid
465 residues in the A α1 domain (Thr-9, Leu-62, Gln-63, Asn-77, and Met-97) that were also significantly
466 associated with decreased MCV antibody response. This is consistent with downregulation of MHC I as a
467 potential mechanism through which Merkel cell tumors evade immune surveillance⁹². The strongest
468 residue-specific signal for EBV EA-D and VZV mapped to B-Asp-9, which is located in the peptide binding
469 groove and tags the B*08:01 allele, part of the HLA 8.1 ancestral haplotype. There is extensive evidence
470 linking HLA 8.1, and B*08:01 specifically, with autoimmune diseases⁹³ and certain cancers^{94,95}, which may
471 be attributed to its high cell-surface stability and increased probability of CD8+ T cell activation.

472 Comparison with other studies of host genetics and viral infection susceptibility shows that our results align
473 with previously reported findings^{7-9,96} (**Supplementary Table 29**). We replicated most associations from
474 two of the largest GWAS of humoral immune response in European ancestry subjects by Hammer et al.⁷
475 (n=2363) and Scepanovic et al.⁸ (n=1000), including HLA SNPs, alleles, amino acids, and haplotypes linked
476 to EBV EBNA IgG, MCV IgG and serostatus, and JCV serostatus. We also replicated two *HLA-DRB1*
477 variants (rs477515, rs2854275) associated with EBV EBNA antibody levels in a Mexican American
478 population⁹. GWAS of HPV16 L1 replicated a variant previously linked to HPV8 seropositivity (rs9357152,
479 $P=0.008$)⁶. Some of our findings contrast with Tian et al.¹³, although we confirmed selected associations,
480 such as A*02:01 (shingles) with VZV ($P=4.1\times 10^{-8}$) and rs2596465 (mononucleosis) with EBV EBNA
481 ($P=3.3\times 10^{-9}$) and EBV p18 ($P=1.0\times 10^{-12}$). These differences may be partly accounted for by self-reported
482 disease status in Tian et al. which is likely to reflect symptom severity and may be an imprecise indicator
483 of infection with certain viruses or the magnitude of antibody response to infection.

484 One of the most striking findings in SNP-based HLA analyses was the genome-wide significant association
485 between rs9273325, index VZV antibody response variant, and risk of schizophrenia. Previous
486 epidemiologic and serologic studies have linked infections to schizophrenia, although the underlying
487 mechanisms remain to be elucidated⁹⁷. Viruses are plausible etiologic candidates for schizophrenia due to
488 their ability to invade the central nervous system and disrupt neurodevelopmental processes by targeting
489 specific neurons, as well as the potential for latent infection to negatively impact plasticity and neurogenesis
490 via pro-inflammatory and aberrant immune signaling^{97,98}. These observations are consistent with the
491 established role the HLA region, including *HLA-DQB1*, in schizophrenia etiology^{99,100}, and is further
492 supported by previously reported associations for rs9273325 with blood cell traits⁵⁷ and immunoglobulin A
493 deficiency¹⁰¹, as well as its role as an eQTL for *HLA-DQB1* in CD4+ T_h cells. Schizophrenia susceptibility
494 alleles DRB1*03:01⁹⁹, DQB1*02:01, and B*08:01 were also the top three alleles associated with VZV
495 antibody response in the unconditional analysis. Enhanced complement activity has been proposed as the
496 mechanism mediating the synaptic loss and excessive pruning which is a hallmark of schizophrenia
497 pathophysiology¹⁰². Complement component 4 (C4) alleles were found to increase risk of schizophrenia
498 proportionally to their effect on increasing *C4A* expression in brain tissue¹⁰². Using gene expression models
499 in whole blood and the frontal cortex we demonstrated that increased *C4A* expression is negatively

500 associated with VZV antibody response. We also observed associations with *C4A* and *C4B* in EBV and
501 HSV-1, but not other viruses. Taken together, these findings delineate a potential mechanism through which
502 aberrant immune response to VZV infection, and potentially HSV-1 and EBV, may increase susceptibility
503 to schizophrenia. However, cautious interpretation is warranted due to significant pleiotropy between HLA
504 loci associated with viral infection and broad immune function.

505 Several limitations of this work should be noted. First, the UK Biobank is unrepresentative of the general
506 UK population due to low participation resulting in healthy volunteer bias¹⁰³. However, since the observed
507 pattern of seroprevalence is consistent with previously published estimates¹⁵ we believe the impact of this
508 bias is likely to be minimal on genetic associations with serological phenotypes. Second, our analyses were
509 restricted to participants of European ancestry due to limited serology data for other ancestries, which limits
510 the generalizability of our findings to diverse populations. Third, we were unable to conduct formal statistical
511 replication of novel GWAS and TWAS signals in an independent sample due to the lack of such a
512 population. Nevertheless, our successful replication of multiple previously reported variants and, combined
513 with the observation that newly discovered genes and variants are part of essential adaptive and innate
514 immunity pathways, support the credibility of our findings. Lastly, we also stress caution in the interpretation
515 of GWAS results for non-ubiquitous pathogens, such as HBV, HCV, and HPV, due to a lack of information
516 on exposure, as well as low numbers of seropositive individuals.

517 Our study also has distinct advantages. The large sample size of the UK Biobank facilitated more powerful
518 genetic association analyses than previous studies, particularly in a population-based cohort unselected for
519 disease status. Our detailed HLA analysis shows independent effects of specific HLA alleles and pleiotropic
520 effects across multiple viruses. Analyses of genetic associations in external datasets further demonstrate
521 a connection between host genetic factors influencing immune response to infection and susceptibility to
522 cancers and neurological conditions.

523 The results of this work highlight widespread genetic pleiotropy between pathways involved in regulating
524 humoral immune response to novel and common viruses, as well as complex diseases. The complex
525 evolutionary relationship between viruses and humans is not dictated simply by infection and acute
526 sickness, it is a complex nuanced architecture of initial challenge tempered with tolerance of viral latency

527 over time. Yet it is that architecture that is evolutionarily optimized to maximize fitness early in life, the result
528 of which may be increased risk for complex diseases later in life. Understanding this complex interplay
529 through both targeted association studies and functional investigations between host genetic factors and
530 immune response has implications for complex disease etiology and may facilitate the discovery of novel
531 therapeutics in a wide range of diseases.

532 **DATA AVAILABILITY**

533 The UK Biobank is an open access resource, available at <https://www.ukbiobank.ac.uk/researchers/>.
534 This research was conducted with approved access to UK Biobank data under application number
535 14105 (PI: Witte).

536 **WEB RESOURCES**

537 PLINK 2.0: <https://www.cog-genomics.org/plink/2.0/>
538 PLINK 1.07 conditional haplotype module: <https://zzz.bwh.harvard.edu/plink/whap.shtml>
539 R packages for pathway analysis: <https://bioconductor.org/packages/release/bioc/html/ReactomePA.html>
540 and <https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>
541 Database of HLA allele frequencies and amino acid substitutions:
542 http://www.allelefrequencies.net/hla9001a.asp?type_analysis=by_pops

543

544 **ACKNOWLEDGEMENTS**

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

548

549 **COMPETING INTERESTS**

550 The authors declare no competing interests.

551 **References**

- 552 1. Aiewsakun, P. & Katzourakis, A. Marine origin of retroviruses in the early Palaeozoic Era. *Nat*
553 *Commun* **8**, 13954 (2017).
- 554 2. Wang, W. *et al.* Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA* (2020).
- 555 3. Moore, P.S. & Chang, Y. Why do viruses cause cancer? Highlights of the first century of human
556 tumour virology. *Nat Rev Cancer* **10**, 878-89 (2010).
- 557 4. Engdahl, E. *et al.* Increased Serological Response Against Human Herpesvirus 6A Is Associated
558 With Risk for Multiple Sclerosis. *Front Immunol* **10**, 2715 (2019).
- 559 5. Readhead, B. *et al.* Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of
560 Molecular, Genetic, and Clinical Networks by Human Herpesvirus. *Neuron* **99**, 64-82 e7 (2018).
- 561 6. Chen, D. *et al.* Genome-wide association study of HPV seropositivity. *Hum Mol Genet* **20**, 4714-23
562 (2011).
- 563 7. Hammer, C. *et al.* Amino Acid Variation in HLA Class II Proteins Is a Major Determinant of Humoral
564 Response to Common Viruses. *Am J Hum Genet* **97**, 738-43 (2015).
- 565 8. Scepanovic, P. *et al.* Human genetic variants and age are the strongest predictors of humoral
566 immune responses to common pathogens and vaccines. *Genome Med* **10**, 59 (2018).
- 567 9. Rubicz, R. *et al.* A genome-wide integrative genomic study localizes genetic factors influencing
568 antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1). *PLoS Genet* **9**, e1003147
569 (2013).
- 570 10. Liu, S. *et al.* Genomic Analyses from Non-invasive Prenatal Testing Reveal Genetic Associations,
571 Patterns of Viral Infections, and Chinese Population History. *Cell* **175**, 347-359 e14 (2018).
- 572 11. Besson, C. *et al.* Strong correlations of anti-viral capsid antigen antibody levels in first-degree
573 relatives from families with Epstein-Barr virus-related lymphomas. *J Infect Dis* **199**, 1121-7 (2009).
- 574 12. Kenney, A.D. *et al.* Human Genetic Determinants of Viral Diseases. *Annu Rev Genet* **51**, 241-263
575 (2017).
- 576 13. Tian, C. *et al.* Genome-wide association and HLA region fine-mapping studies identify susceptibility
577 loci for multiple common infections. *Nat Commun* **8**, 599 (2017).
- 578 14. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**,
579 203-209 (2018).
- 580 15. Mentzer, A.J. *et al.* Identification of host-pathogen-disease relationships using a scalable Multiplex
581 Serology platform in UK Biobank. *medRxiv*, 19004960 (2019).
- 582 16. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies.
583 *Bioinformatics* **26**, 2867-73 (2010).
- 584 17. Waterboer, T. *et al.* Multiplex human papillomavirus serology based on in situ-purified glutathione
585 s-transferase fusion proteins. *Clin Chem* **51**, 1845-53 (2005).
- 586 18. Waterboer, T., Sehr, P. & Pawlita, M. Suppression of non-specific binding in serological Luminex
587 assays. *J Immunol Methods* **309**, 200-4 (2006).

- 588 19. Kreimer, A.R. *et al.* Kinetics of the Human Papillomavirus Type 16 E6 Antibody Response Prior to
589 Oropharyngeal Cancer. *J Natl Cancer Inst* **109**(2017).
- 590 20. Peterson, R.A. & Cavanaugh, J.E. Ordered quantile normalization: a semiparametric
591 transformation built for the cross-validation era. *Journal of Applied Statistics*, 1-16 (2019).
- 592 21. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. & Kircher, M. CADD: predicting the
593 deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894
594 (2019).
- 595 22. Dong, S. & Boyle, A.P. Predicting functional variants in enhancer and promoter elements using
596 RegulomeDB. *Hum Mutat* **40**, 1292-1298 (2019).
- 597 23. Schmiedel, B.J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression.
598 *Cell* **175**, 1701-1715 e16 (2018).
- 599 24. Sun, B.B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).
- 600 25. Yao, C. *et al.* Genome-wide mapping of plasma protein QTLs identifies putatively causal genes
601 and pathways for cardiovascular disease. *Nat Commun* **9**, 3268 (2018).
- 602 26. Rashkin, S.R. *et al.* Pan-Cancer Study Detects Novel Genetic Risk Variants and Shared Genetic
603 Basis in Two Large Cohorts. *bioRxiv*, 635367 (2019).
- 604 27. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European
605 populations. *Nat Genet* **51**, 1670-1678 (2019).
- 606 28. Jun, G. *et al.* A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol*
607 *Psychiatry* **21**, 108-17 (2016).
- 608 29. Motyer, A. *et al.* Practical Use of Methods for Imputation of HLA Alleles from SNP Genotype Data.
609 *bioRxiv*, 091009 (2016).
- 610 30. Jia, X. *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* **8**,
611 e64683 (2013).
- 612 31. Gonzalez-Galarza, F.F. *et al.* Allele frequency net database (AFND) 2020 update: gold-standard
613 data classification, open access genotype data and new query tools. *Nucleic Acids Res* **48**, D783-
614 D788 (2020).
- 615 32. Barbeira, A.N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression
616 variation inferred from GWAS summary statistics. *Nat Commun* **9**, 1825 (2018).
- 617 33. Barbeira, A.N. *et al.* Widespread dose-dependent effects of RNA expression and splicing on
618 complex diseases and traits. *bioRxiv*, 814350 (2019).
- 619 34. Urbut, S.M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating
620 and testing effects in genomic studies with multiple conditions. *Nat Genet* **51**, 187-195 (2019).
- 621 35. Wen, X., Lee, Y., Luca, F. & Pique-Regi, R. Efficient Integrative Multi-SNP Association Analysis via
622 Deterministic Approximation of Posteriors. *Am J Hum Genet* **98**, 1114-1129 (2016).
- 623 36. Lee, Y., Luca, F., Pique-Regi, R. & Wen, X. Bayesian Multi-SNP Genetic Association Analysis:
624 Control of FDR and Use of Summary Statistics. *bioRxiv*, 316471 (2018).

- 625 37. Steiner, I., Kennedy, P.G. & Pachner, A.R. The neurotropic herpes viruses: herpes simplex and
626 varicella-zoster. *Lancet Neurol* **6**, 1015-28 (2007).
- 627 38. Khalili, K., Del Valle, L., Otte, J., Weaver, M. & Gordon, J. Human neurotropic polyomavirus, JCV,
628 and its role in carcinogenesis. *Oncogene* **22**, 5181-91 (2003).
- 629 39. Feng, H., Shuda, M., Chang, Y. & Moore, P.S. Clonal integration of a polyomavirus in human Merkel
630 cell carcinoma. *Science* **319**, 1096-100 (2008).
- 631 40. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage,
632 supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**,
633 D607-D613 (2019).
- 634 41. Shin, O.S. *et al.* LPLUNC1 modulates innate immune responses to *Vibrio cholerae*. *J Infect Dis*
635 **204**, 1349-57 (2011).
- 636 42. Shafi, O. Inverse relationship between Alzheimer's disease and cancer, and other factors
637 contributing to Alzheimer's disease: a systematic review. *BMC Neurol* **16**, 236 (2016).
- 638 43. Gragert, L., Madbouly, A., Freeman, J. & Maiers, M. Six-locus high resolution HLA haplotype
639 frequencies derived from mixed-resolution DNA typing for the entire US donor registry. *Hum*
640 *Immunol* **74**, 1313-20 (2013).
- 641 44. Degenhardt, F. *et al.* Construction and benchmarking of a multi-ethnic reference panel for the
642 imputation of HLA class I and II alleles. *Hum Mol Genet* **28**, 2078-2092 (2019).
- 643 45. Chen, Q., Sun, L. & Chen, Z.J. Regulation and function of the cGAS-STING pathway of cytosolic
644 DNA sensing. *Nat Immunol* **17**, 1142-9 (2016).
- 645 46. Kwon, J. & Bakhoun, S.F. The Cytosolic DNA-Sensing cGAS-STING Pathway in Cancer. *Cancer*
646 *Discov* **10**, 26-39 (2020).
- 647 47. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z.J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor
648 that activates the type I interferon pathway. *Science* **339**, 786-91 (2013).
- 649 48. Inoue, T. & Tsai, B. How viruses use the endoplasmic reticulum for entry, replication, and assembly.
650 *Cold Spring Harb Perspect Biol* **5**, a013250 (2013).
- 651 49. Woo, S.R. *et al.* STING-dependent cytosolic DNA sensing mediates innate immune recognition of
652 immunogenic tumors. *Immunity* **41**, 830-42 (2014).
- 653 50. Demaria, O. *et al.* STING activation of tumor endothelial cells initiates spontaneous and therapeutic
654 antitumor immunity. *Proc Natl Acad Sci U S A* **112**, 15408-13 (2015).
- 655 51. Ohkuri, T. *et al.* STING contributes to antiglioma immunity via triggering type I IFN signals in the
656 tumor microenvironment. *Cancer Immunol Res* **2**, 1199-208 (2014).
- 657 52. Fu, J. *et al.* STING agonist formulated cancer vaccines can cure established tumors resistant to
658 PD-1 blockade. *Sci Transl Med* **7**, 283ra52 (2015).
- 659 53. Corrales, L. *et al.* Direct Activation of STING in the Tumor Microenvironment Leads to Potent and
660 Systemic Tumor Regression and Immunity. *Cell Rep* **11**, 1018-30 (2015).

- 661 54. Ohkuri, T., Ghosh, A., Kosaka, A., Sarkar, S.N. & Okada, H. Protective role of STING against
662 gliomagenesis: Rational use of STING agonist in anti-glioma immunotherapy. *Oncoimmunology* **4**,
663 e999523 (2015).
- 664 55. Ikeda, K. *et al.* Identification of ARIA regulating endothelial apoptosis and angiogenesis by
665 modulating proteasomal degradation of cIAP-1 and cIAP-2. *Proc Natl Acad Sci U S A* **106**, 8227-
666 32 (2009).
- 667 56. Verma, A. *et al.* Endothelial cell-specific chemotaxis receptor (ecscr) promotes angioblast migration
668 during vasculogenesis and enhances VEGF receptor sensitivity. *Blood* **115**, 4614-22 (2010).
- 669 57. Astle, W.J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common
670 Complex Disease. *Cell* **167**, 1415-1429 e19 (2016).
- 671 58. Assinger, A. Platelets and infection - an emerging role of platelets in viral infection. *Front Immunol*
672 **5**, 649 (2014).
- 673 59. Kelly, R.J., Rouquier, S., Giorgi, D., Lennon, G.G. & Lowe, J.B. Sequence and expression of a
674 candidate for the human Secretor blood group alpha(1,2)fucosyltransferase gene (FUT2).
675 Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-
676 secretor phenotype. *J Biol Chem* **270**, 4640-9 (1995).
- 677 60. Hazra, A. *et al.* Common variants of FUT2 are associated with plasma vitamin B12 levels. *Nat*
678 *Genet* **40**, 1160-2 (2008).
- 679 61. Carlsson, B. *et al.* The G428A nonsense mutation in FUT2 provides strong but not absolute
680 protection against symptomatic GII.4 Norovirus infection. *PLoS One* **4**, e5593 (2009).
- 681 62. Ruvoen-Clouet, N., Belliot, G. & Le Pendu, J. Noroviruses and histo-blood groups: the impact of
682 common host genetic polymorphisms on virus transmission and evolution. *Rev Med Virol* **23**, 355-
683 66 (2013).
- 684 63. Imbert-Marcille, B.M. *et al.* A FUT2 gene common polymorphism determines resistance to rotavirus
685 A of the P[8] genotype. *J Infect Dis* **209**, 1227-30 (2014).
- 686 64. Ikehara, Y. *et al.* Polymorphisms of two fucosyltransferase genes (Lewis and Secretor genes)
687 involving type I Lewis antigens are associated with the presence of anti-Helicobacter pylori IgG
688 antibody. *Cancer Epidemiol Biomarkers Prev* **10**, 971-7 (2001).
- 689 65. Blackwell, C.C. *et al.* Non-secretion of ABO antigens predisposing to infection by Neisseria
690 meningitidis and Streptococcus pneumoniae. *Lancet* **2**, 284-5 (1986).
- 691 66. de Lange, K.M. *et al.* Genome-wide association study implicates immune activation of multiple
692 integrin genes in inflammatory bowel disease. *Nat Genet* **49**, 256-261 (2017).
- 693 67. Ellinghaus, D. *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations
694 and highlights disease-specific patterns at shared loci. *Nat Genet* **48**, 510-8 (2016).
- 695 68. Hoffmann, T.J. *et al.* A large electronic-health-record-based genome-wide study of serum lipids.
696 *Nat Genet* **50**, 401-413 (2018).

697 69. Tanaka, T. *et al.* Genome-wide association study of vitamin B6, vitamin B12, folate, and
698 homocysteine blood concentrations. *Am J Hum Genet* **84**, 477-82 (2009).

699 70. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic
700 etiology of tobacco and alcohol use. *Nat Genet* **51**, 237-244 (2019).

701 71. McKay, J.D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci
702 and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet* **49**, 1126-1132
703 (2017).

704 72. Batista, C.M. *et al.* Adult neurogenesis and glial oncogenesis: when the process fails. *Biomed Res*
705 *Int* **2014**, 438639 (2014).

706 73. Yamagishi, S. *et al.* Netrin-5 is highly expressed in neurogenic regions of the adult brain. *Front Cell*
707 *Neurosci* **9**, 146 (2015).

708 74. Leong, Y.A. *et al.* CXCR5(+) follicular cytotoxic T cells control viral infection in B cell follicles. *Nat*
709 *Immunol* **17**, 1187-96 (2016).

710 75. Willis, T.G. *et al.* Molecular cloning of translocation t(1;14)(q21;q32) defines a novel gene (BCL9)
711 at chromosome 1q21. *Blood* **91**, 1873-81 (1998).

712 76. Deka, J. *et al.* Bcl9/Bcl9l are critical for Wnt-mediated regulation of stem cell traits in colon
713 epithelium and adenocarcinomas. *Cancer Res* **70**, 6619-28 (2010).

714 77. Pilli, M. *et al.* TBK-1 promotes autophagy-mediated antimicrobial defense by controlling
715 autophagosome maturation. *Immunity* **37**, 223-34 (2012).

716 78. Gordon, D.E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing.
717 *Nature* (2020).

718 79. Zhu, L. *et al.* TBKBP1 and TBK1 form a growth factor signalling axis mediating immunosuppression
719 and tumorigenesis. *Nat Cell Biol* **21**, 1604-1614 (2019).

720 80. Yang, J. *et al.* Thyrotroph embryonic factor is downregulated in bladder cancer and suppresses
721 proliferation and tumorigenesis via the AKT/FOXOs signalling pathway. *Cell Prolif* **52**, e12560
722 (2019).

723 81. International Multiple Sclerosis Genetics, C. *et al.* Genetic risk and a primary role for cell-mediated
724 immune mechanisms in multiple sclerosis. *Nature* **476**, 214-9 (2011).

725 82. Sun, M.Y. *et al.* Critical role for nonGAP function of Galphas in RGS1mediated promotion of
726 melanoma progression through AKT and ERK phosphorylation. *Oncol Rep* **39**, 2673-2680 (2018).

727 83. Carreras, J. *et al.* Clinicopathological characteristics and genomic profile of primary sinonasal tract
728 diffuse large B cell lymphoma (DLBCL) reveals gain at 1q31 and RGS1 encoding protein; high
729 RGS1 immunohistochemical expression associates with poor overall survival in DLBCL not
730 otherwise specified (NOS). *Histopathology* **70**, 595-621 (2017).

731 84. Mukherjee, S. *et al.* Genetic data and cognitively defined late-onset Alzheimer's disease
732 subgroups. *Mol Psychiatry* (2018).

- 733 85. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of
734 Alzheimer's Disease. *Cell* **169**, 1276-1290 e17 (2017).
- 735 86. Jawaheer, D. *et al.* Dissecting the genetic complexity of the association between human leukocyte
736 antigens and rheumatoid arthritis. *Am J Hum Genet* **71**, 585-94 (2002).
- 737 87. Vader, W. *et al.* The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude
738 and breadth of gluten-specific T cell responses. *Proc Natl Acad Sci U S A* **100**, 12390-5 (2003).
- 739 88. Erlich, H. *et al.* HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the
740 type 1 diabetes genetics consortium families. *Diabetes* **57**, 1084-92 (2008).
- 741 89. Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association
742 between MHC and seropositive rheumatoid arthritis. *Nat Genet* **44**, 291-6 (2012).
- 743 90. Hu, X. *et al.* Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR
744 molecules drive type 1 diabetes risk. *Nat Genet* **47**, 898-905 (2015).
- 745 91. Patsopoulos, N.A. *et al.* Fine-mapping the genetic association of the major histocompatibility
746 complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet* **9**, e1003926 (2013).
- 747 92. Paulson, K.G. *et al.* Downregulation of MHC-I expression is prevalent but reversible in Merkel cell
748 carcinoma. *Cancer Immunol Res* **2**, 1071-9 (2014).
- 749 93. Candore, G., Lio, D., Colonna Romano, G. & Caruso, C. Pathogenesis of autoimmune diseases
750 associated with 8.1 ancestral haplotype: effect of multiple gene interactions. *Autoimmun Rev* **1**, 29-
751 35 (2002).
- 752 94. Ferreiro-Iglesias, A. *et al.* Fine mapping of MHC region in lung cancer highlights independent
753 susceptibility loci by ethnicity. *Nat Commun* **9**, 3927 (2018).
- 754 95. Abdou, A.M. *et al.* Human leukocyte antigen (HLA) A1-B8-DR3 (8.1) haplotype, tumor necrosis
755 factor (TNF) G-308A, and risk of non-Hodgkin lymphoma. *Leukemia* **24**, 1055-8 (2010).
- 756 96. Sundqvist, E. *et al.* JC polyomavirus infection is strongly controlled by human leucocyte antigen
757 class II variants. *PLoS Pathog* **10**, e1004084 (2014).
- 758 97. Khandaker, G.M. *et al.* Inflammation and immunity in schizophrenia: implications for
759 pathophysiology and treatment. *Lancet Psychiatry* **2**, 258-270 (2015).
- 760 98. Dickerson, F. *et al.* Schizophrenia is Associated With an Aberrant Immune Response to Epstein-
761 Barr Virus. *Schizophr Bull* **45**, 1112-1119 (2019).
- 762 99. International Schizophrenia, C. *et al.* Common polygenic variation contributes to risk of
763 schizophrenia and bipolar disorder. *Nature* **460**, 748-52 (2009).
- 764 100. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108
765 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).
- 766 101. Bronson, P.G. *et al.* Common variants at PVT1, ATG13-AMBRA1, AHI1 and CLEC16A are
767 associated with selective IgA deficiency. *Nat Genet* **48**, 1425-1429 (2016).
- 768 102. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature*
769 **530**, 177-83 (2016).

770 103. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank
771 Participants With Those of the General Population. *Am J Epidemiol* **186**, 1026-1034 (2017).
772

Table 1: Lead genome-wide significant variants ($P < 5.0 \times 10^{-8}$) 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	<i>TSBP1</i>
EBV EA-D	6806	6	32665840	rs34825357	T	TC	0.409	-0.114	(0.017)	2.0×10^{-11}	intergenic	<i>MTCO3P1</i>
EBV EBNA	7003	3	151114852	rs67886110*	G	T	0.596	0.103	(0.017)	1.3×10^{-9}	intronic	<i>MED12L</i>
		6	32451762	rs9269233	A	C	0.249	0.315	(0.019)	3.5×10^{-61}	intergenic	<i>HLA-DRB9</i>
EBV VCA p18	7492	6	31486158	6:31486158	GT	G	0.245	0.197	(0.018)	7.1×10^{-27}	intergenic	<i>PPIAP9</i>
EBV ZEBRA	7197	6	32637772	rs9274728	A	G	0.718	-0.315	(0.018)	4.7×10^{-67}	intergenic	<i>HLA-DQB1</i>
HHV6 IE1A	6077	7	139985625	rs2429218	T	C	0.615	0.106	(0.019)	1.4×10^{-8}	downstream	<i>RP5-1136G2.1</i>
		6	32602665	rs139299944	C	CT	0.655	0.114	(0.017)	1.5×10^{-11}	intronic	<i>HLA-DQA1</i>
HHV7 U14	7481	11	118767564	rs75438046	G	A	0.970	0.280	(0.049)	1.3×10^{-8}	3'-UTR	<i>CXCR5 / BCL9L</i>
		17	45794706	rs1808192	A	G	0.331	-0.099	(0.017)	9.8×10^{-9}	intergenic	<i>TBKBP1</i>
HSV1 1gG	5468	6	32627852	rs1130420	G	A	0.583	-0.122	(0.019)	2.5×10^{-10}	3'-UTR	<i>HLA-DQB1</i>
		10	91189187	rs11203123*	A	C	0.988	0.512	(0.093)	3.9×10^{-8}	intergenic	<i>SLC16A12</i>
VZV gE/Ig ¹	7289	6	32623193	rs9273325	G	A	0.831	-0.232	(0.021)	8.2×10^{-28}	intergenic	<i>HLA-DQB1</i>
BKV VP1	7523	19	49206462	rs681343	C	T	0.491	-0.125	(0.016)	4.7×10^{-15}	synonymous	<i>FUT2</i>
JCV VP1	4471	6	32589842	rs9271525	G	A	0.163	-0.318	(0.031)	3.9×10^{-24}	intergenic	<i>HLA-DQA1</i>
		3	18238783	rs776170649	CT	C	0.790	-0.134	(0.024)	1.7×10^{-8}	intergenic	<i>LOC339862</i>
MCV VP1	5219	5	138865423	rs7444313	G	A	0.263	0.169	(0.021)	2.4×10^{-15}	intergenic	<i>TMEM173</i>
		6	32429277	rs9268847	A	G	0.750	-0.195	(0.022)	2.4×10^{-19}	intronic	<i>HLA-DRB9</i>

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)

Figure 1: Flow chart describing the main serological phenotypes and association analyses

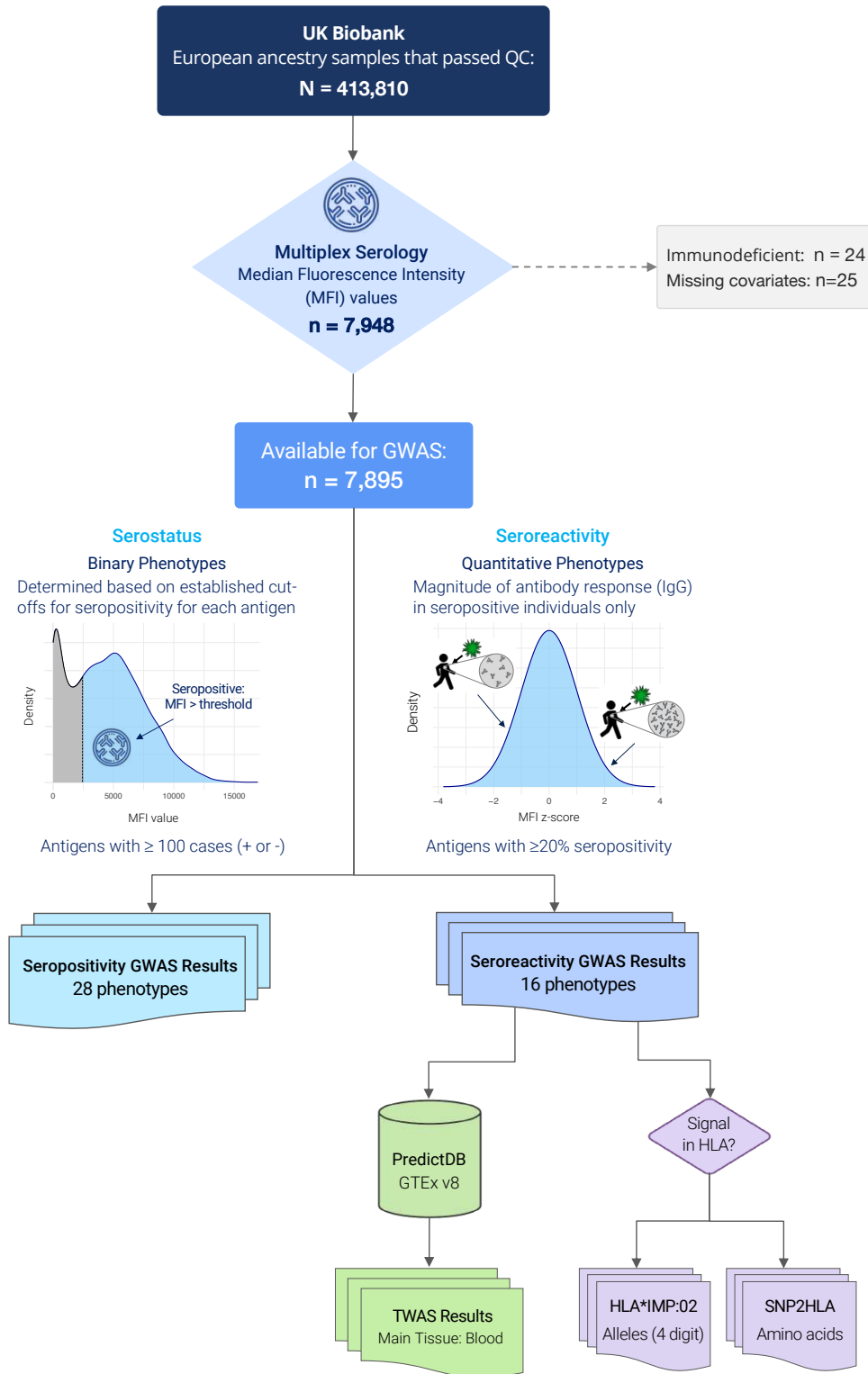


Figure 2: Results from genome-wide and regional association analyses of continuous antibody response phenotypes (MFI z-scores) among individuals seropositive for human polyomaviruses BKV, JCV, and Merkel cell (MCV). The lower two panels depict the association signal and linkage disequilibrium (LD) structure in the HLA region for JCV and MCV.

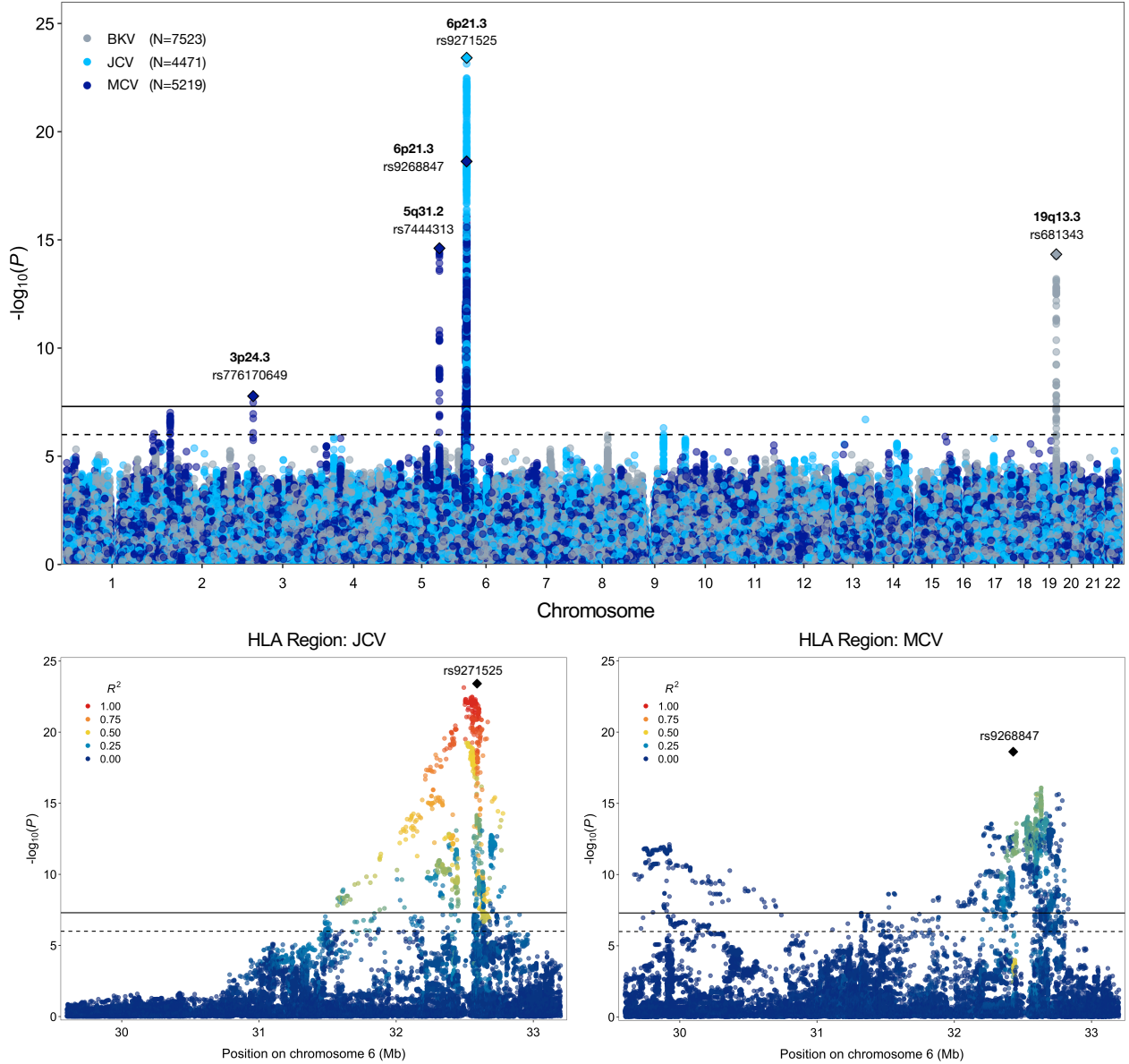


Figure 3: Regional association plots for conditionally independent HLA genetic variants that were significantly ($P < 5.0 \times 10^{-8}$, solid black line) associated with each continuous antibody response phenotype. The suggestive significance threshold corresponds to $P < 1.0 \times 10^{-6}$ (dotted black line).

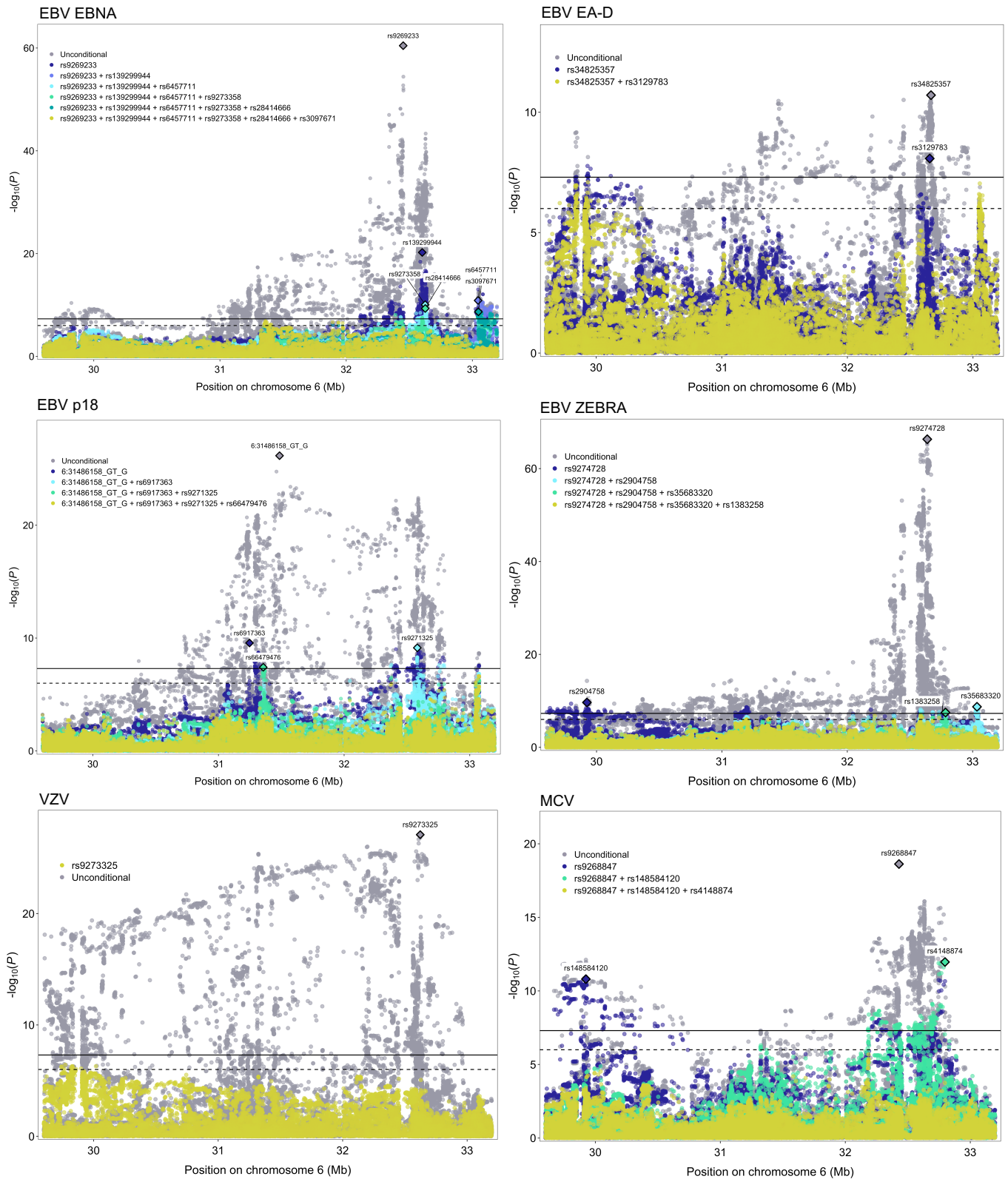
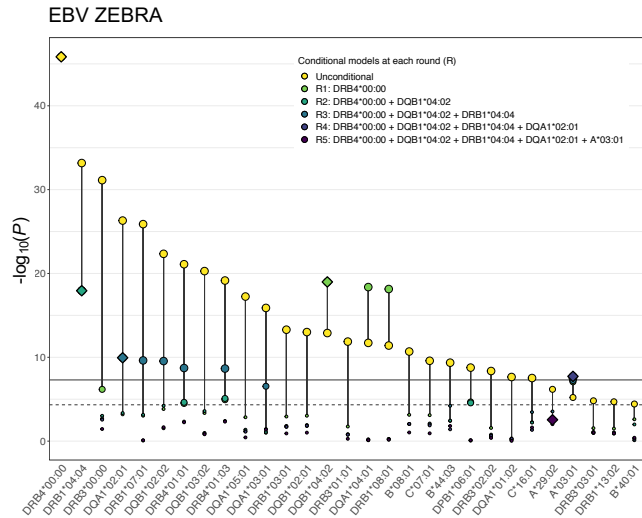
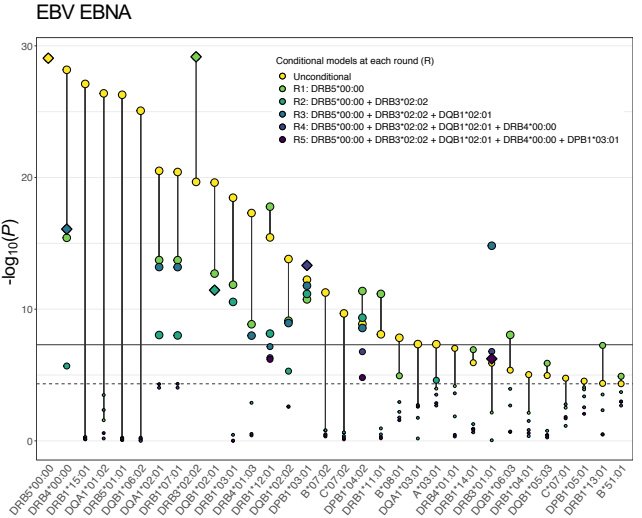


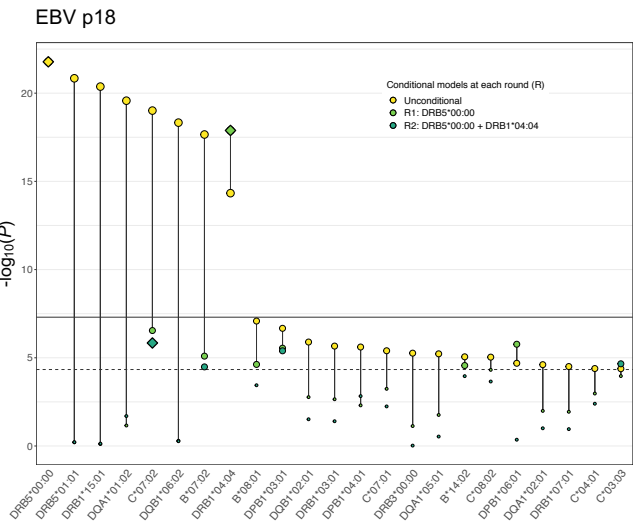
Figure 4: Conditionally independent classical HLA alleles significantly ($P_{\text{cond}} < 5.0 \times 10^{-8}$, solid line) associated with each continuous antibody response phenotype. Only classical alleles that surpassed the Bonferroni-corrected significance threshold ($P < 4.6 \times 10^{-5}$, dotted line) were included in conditional analyses.



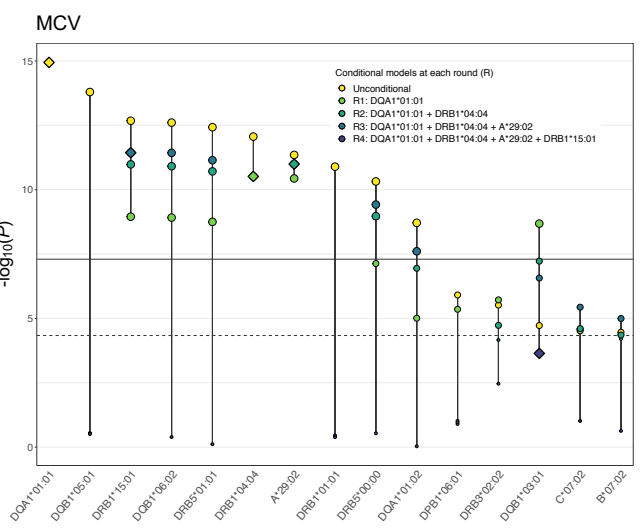
Allele	Frequency	Beta	(SE)	P_{cond}
DRB4*00:00	0.652	-0.246	(0.017)	1.4×10^{-46}
DQB1*04:02	0.020	0.504	(0.055)	1.0×10^{-19}
DRB1*04:04	0.041	0.376	(0.042)	1.1×10^{-18}
DQA1*02:01	0.145	0.187	(0.029)	1.1×10^{-10}
A*03:01	0.144	0.129	(0.023)	1.9×10^{-8}



Allele	Frequency	Beta	(SE)	P_{cond}
DRB5*00:00	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}



Allele	Frequency	Beta	(SE)	P_{cond}
DRB5*00:00	0.845	-0.210	(0.021)	1.7×10^{-22}
DRB1*04:04	0.041	0.357	(0.040)	1.3×10^{-18}



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 5: 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).

