Human genomics of the humoral immune response against polyomaviruses

F. Hodel,[1,2,‡] A.Y. Chong,[3] P. Scepanovic,[4] Z.M. Xu,[1,2] O. Naret,[1,2] C.W. Thorball,[1,5] S. Rüeger,[6] P. Marques-Vidal,[7] P. Vollenweider,[7,§] M. Begemann,[8] H. Ehrenreich,[8] N. Brenner,[9] N. Bender,[9] T. Waterboer,[9] A. J. Mentzer,[3,10,**] A.V.S. Hill,[3,11] C. Hammer,[11,12,13,†,††] and J. Fellay[1,2,5,*,†]

[1]Global Health Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, [2]Swiss Institute of Bioinformatics, Quartier UNIL-Sorge, CH-1015 Lausanne, Switzerland, [3]The Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, United Kingdom, [4]Roche Pharmaceutical Research and Early Development, F. Hoffmann-La Roche Ltd, Headquarters Grenzacherstrasse 124, CH-4070 Basel, Switzerland, [5]Precision Medicine Unit, Lausanne University Hospital and University of Lausanne, Rue du Bugnon 46, CH-1011 Lausanne, Switzerland, [6]Institute for Molecular Medicine Finland, Institute of Life Science HiLIFE, University of Helsinki, Haartmaninkatu 8, 00290 Helsinki, Finland, [7]Department of Medicine, Internal Medicine, Lausanne University Hospital and University of Lausanne, Rue du Bugnon 46, CH-1011 Lausanne, Switzerland, [8]Clinical Neuroscience, Max Planck Institute of Experimental Medicine, DFG Research Center for Nanoscale Microscopy and Molecular Physiology of the Brain, Hermann-Rein-Straße 3, 37075 Göttingen, Germany, [9]Infections and Cancer Epidemiology, German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany, [10]Big Data Institute, LiKa Shing Centre for Health Information and Discovery, University of Oxford, Old Road Campus, Oxford OX3 7LF, United Kingdom, [11]The Jenner Institute, University of Oxford, Old Road Campus Research Build, Roosevelt Dr, Oxford OX1 2JD, United Kingdom, [12]Department of Cancer Immunology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080, USA and [13]Department of Human Genetics, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080, USA

[†]Equal contribution.
[‡]http://orcid.org/0000-0001-7331-7357
[§]http://orcid.org/0000-0002-0765-896X
[**]http://orcid.org/0000-0002-4502-2209
[††]http://orcid.org/0000-0003-4548-7548
[*]Corresponding author: E-mail: jacques.fellay@epfl.ch

Abstract

Human polyomaviruses are widespread in humans and can cause severe disease in immunocompromised individuals. To identify human genetic determinants of the humoral immune response against polyomaviruses, we performed genome-wide association studies and meta-analyses of qualitative and quantitative immunoglobulin G responses against BK polyomavirus (BKPyV), JC polyomavirus (JCPyV), Merkel cellpolyomavirus (MCPyV), WU polyomavirus (WUPyV), and human polyomavirus 6 (HPyV6) in 15,660 individuals of European ancestry from three independent studies. We observed significant associations for all tested viruses: JCPyV, HPyV6, and MCPyV associated with human leukocyte antigen class II variation, BKPyV and JCPyV with variants in *FUT2*, responsible for secretor status, MCPyV with variants in *STING1*, involved in interferon induction, and WUPyV with a functional variant in *MUC1*, previously associated with risk for gastric cancer. These results provide insights into the genetic control of a family of very prevalent human viruses, highlighting genes and pathways that play a modulating role in human humoral immunity.

Key words: infection; human; genomics; polyomavirus; GWAS; meta-analysis

1. Introduction

The variability of the humoral response to antigenic stimulation has been documented for a long time (Grundbacher 1974). While it is partly due to demographic and environmental factors, a substantial part of this variation is heritable and can thus be attributed to differences in the host genome. For example, human genetic factors were estimated to account for 35 per cent (±6 per cent) and 21 per cent (±5 per cent) of the variance in *Helicobacter pylori* and *Toxoplasma gondii* antibody levels and 60 per cent (±11 per cent) and 57 per cent (±11 per cent) of the variance in cytomegalovirus and Epstein–Barr virus (EBV) serostatus, respectively (Rubicz et al. 2011). This high heritability makes serological phenotypes highly promising targets for human genomic studies. Indeed, specific genetic variants have already been associated with variation in immunoglobulin G (IgG) levels. Most variants identified to date are located within the major histocompatibility complex locus on chromosome 6 (Rubicz et al. 2013; Hammer et al. 2015; Scepanovic et al. 2018). This highly polymorphic complex notably encodes the class I and class II human leukocyte antigen (HLA) proteins, which present antigenic epitopes to effector cells of the immune system—a key role in the crossroad between innate and acquired immune responses.

Polyomaviruses are small (40–50 nm), double-stranded DNA viruses belonging to the *Polyomaviridae* family, with an average genome size of 5,000 base pairs (Gardner et al. 1971; Padgett et al. 1971). First discovered in rodents in the 1950s, the first human polyomavirus, namely, JC polyomavirus (JCPyV),

Table 1. Descriptive statistics of polyomavirus infections in the three cohorts

	CoLaus (N = 3,689)		GRAS (N = 2,363)		UKB (N = 9,351)	
	N seropositives (%)	IgG level in seropositives (log, mean ± SD)	N seropositives (%)	IgG level in seropositives (log, mean ± SD)	N seropositives (%)	IgG level in seropositives (log, mean ± SD)
BKPyV	3,112 (84)	7.63 ± 0.99	2,226 (94)	8.45 ± 0.98	8,911 (95)	7.96 ± 0.90
HPyV6	3,156 (86)	8.13 ± 0.88	1,938 (82)	8.30 ± 0.90	–	–
JCPyV	1,915 (52)	7.16 ± 0.90	1,268 (54)	7.15 ± 0.82	5,384 (58)	6.75 ± 0.88
MCPyV	–	–	1,871 (79)	8.01 ± 1.02	6,220 (67)	7.70 ± 1.01
WUPyV	3,536 (96)	7.85 ± 0.81	2,257 (96)	8.31 ± 0.81	–	–

–: Not measured.

was discovered in 1965 in a patient with progressive multifocal leukoencephalopathy (PML) (Gross 1953; Zurhein and Chou 1965). Serological surveys indicate that human polyomaviruses are ubiquitous, with most of the world's population infected by several of them during childhood (Cook 2016). In most cases, polyomavirus infections are asymptomatic. After acute infection, they silently persist in healthy individuals. However, they can lead to symptoms and even severe disease in immunocompromised individuals. As an example, reactivation of JCPyV can cause PML in patients with low $CD4^+$ T-cell counts, such as individuals with AIDS or on chronic immunosuppressive therapy, and BK polyomavirus (BKPyV) is known to cause renal disease in kidney-transplanted patients (Gardner et al. 1971; Padgett et al. 1971; Crum-Cianflone et al. 2007; Trofe-Clark and Sawinski 2016).

A better understanding of the genetic control of the human immune response against polyomaviruses may provide new insights into pathologic mechanisms. We therefore searched for associations between human genetic variation and both differences in IgG levels and serostatus against five human polyomaviruses: BKPyV, JCPyV, human polyomavirus 6 (HPyV6), Merkel cell polyomavirus (MCPyV), and WU polyomavirus (WUPyV), using genome-wide association study (GWAS) and meta-analysis in a total of 15,660 healthy individuals of European ancestry from three studies from Switzerland, the United Kingdom, and Germany.

2. Results

2.1 Seroprevalence of polyomaviruses

We included 3,689 individuals from CoLaus study (46 per cent males, mean age ± SD: 53.1 ± 10.6), 2,363 from GRAS Data Collection (63 per cent males, mean age ± SD: 39.1 ± 13.5), and 9,351 from the UKB (44 per cent males, mean age ± SD: 56.5 ± 8.2), resulting in a total of 15,660 individuals with available genotyping data, serological data, and covariates. All tested polyomaviruses had high seroprevalence in the three samples, with highest observed seroprevalence for WUPyV (96 per cent), followed by BKPyV (91 per cent), HPyV6 (84 per cent), MCPyV (73 per cent), and JCPyV (55 per cent) (Table 1). In all cohorts combined, the serological responses against several polyomaviruses correlated with age and sex, confirming previous observations (Egli et al. 2009; Viscidi et al. 2011; Hammer et al. 2015). We observed a significant decrease with age in antibody levels against BKPyV ($P < 2 \times 10^{-16}$), JCPyV ($P = 4.56 \times 10^{-03}$), MCPyV ($P = 5.77 \times 10^{-03}$), and WUPyV ($P < 2 \times 10^{-16}$) and lower IgG levels against HPyV6 ($P = 3.51 \times 10^{-15}$), JCPyV ($P < 2 \times 10^{-16}$), and MCPyV ($P = 1.07 \times 10^{-05}$) in female participants.

2.2 GWASs and meta-analyses

We performed a total of nine independent GWASs using either a case–control study design (serostatus: antibody-positive versus antibody-negative) or a continuous, quantitative approach (IgG levels in seropositive individuals) to search for human genetic determinants of the antibody response to BKPyV and JCPyV in the three cohorts, to HPyV6 and WUPyV in CoLaus and GRAS, and to MCPyV in GRAS and UKB (Fig. 1). We did not run a case–control analysis for WUPyV because of the very high seroprevalence of that virus (96 per cent). Meta-analyses were conducted using GWAS summary statistics. For each of the five phenotypes, meta-analytic inflation factors λ ranged from 0.99 to 1.01 across the tested cohorts, indicating proper control of population stratification (Supplementary Fig. S1). In total, we identified four genome-wide significant loci, mapping to the *HLA* (for HPyV6, JCPyV, and MCPyV), *FUT2* (for BKPyV and JCPyV), *STING1* (for MCPyV), and *MUC1* (for WUPyV) regions (Fig. 2).

2.3 Role of HLA variation in antibody response to polyomaviruses

Significant associations were observed in the HLA class II region on chromosome 6 for HPyV6, JCPyV, and MCPyV (Fig. 2), with the lead variants being rs9272624 ($P = 9.62 \times 10^{-14}$), rs112463084 ($P = 2.18 \times 10^{-20}$), and rs147658078 ($P = 8.09 \times 10^{-20}$), respectively. To further dissect the association signals, we imputed four-digit HLA alleles and tested them for association with anti-HPyV6, anti-JCPyV, and anti-MCPyV VP1 IgG levels. The regional association plots are shown in Supplementary Fig. S2, and the resulting association *P*-values are listed in Table 2. We observed strong associations with quantitative seroreactivity for HLA-DRB1 and HLA-DQA1 alleles for all three human polyomaviruses. The size and directionality of the effects were virus-specific.
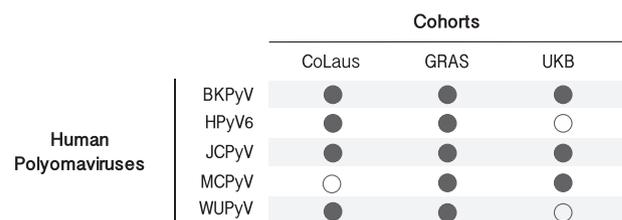

Figure 1. Visual representation of study data. Diagram representing data availability for each human polyomavirus by cohort. Black dots indicate that serological data are available in the cohort, and white dots indicate that data are not available.
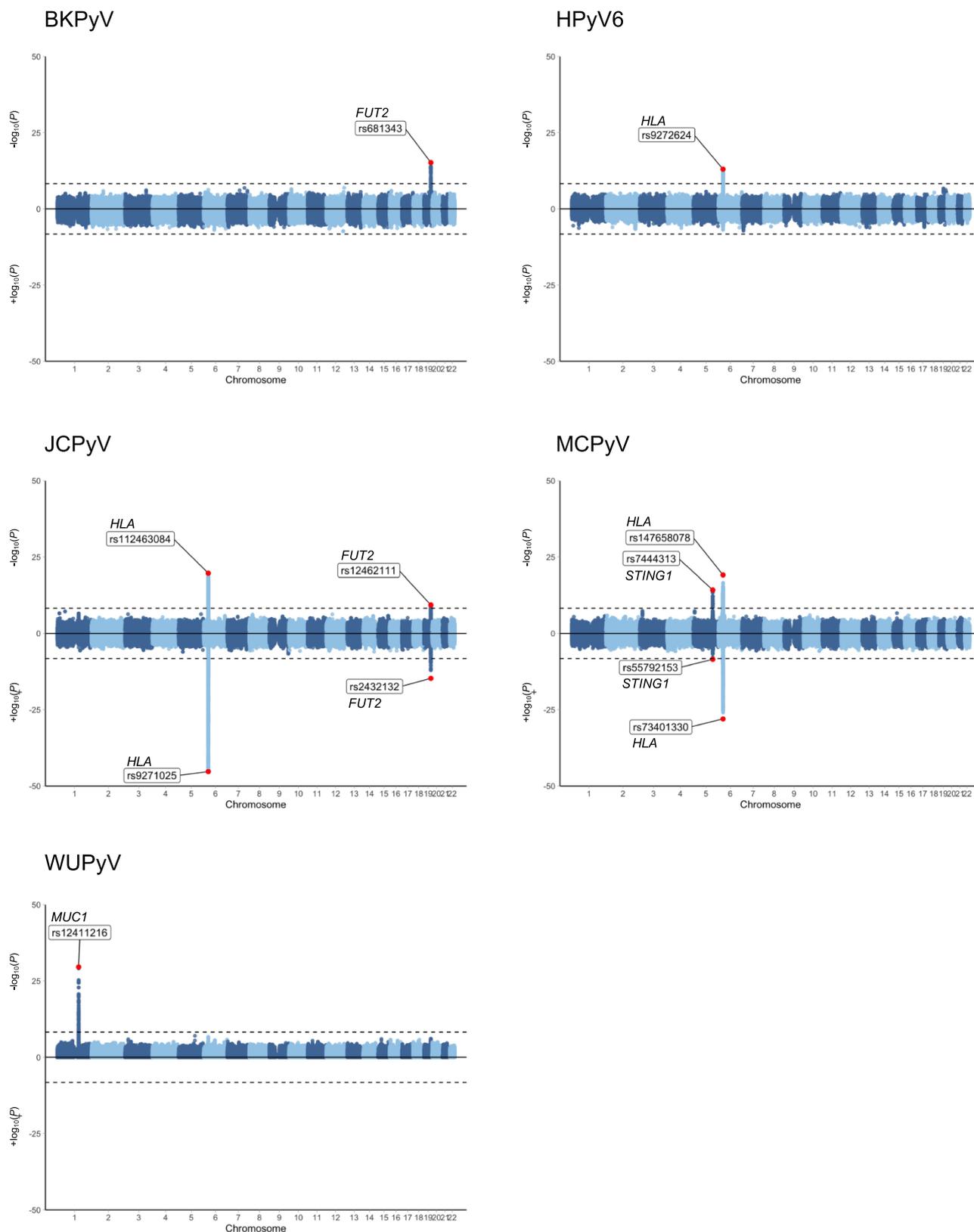
Figure 2. Chicago plots of genome-wide association results of the meta-analyses of antibody levels (top) and serostatus (bottom) against BKPyV, HPyV6, JCPyV, MCPyV, and WUPyV. Chicago plots showing the significance of association of all SNPs across chromosomes 1–22. SNPs are plotted on the x-axis according to their physical position on each chromosome, and the strength of the association with antibody levels is indicated on the y-axis (as $-\log 10$ P-value for IgG response and as $+\log 10$ P-value for serostatus). The dashed line marks the Bonferroni-corrected genome-wide significance threshold of 5.6×10^{-09}. For each locus, the SNP with the most significant association is plotted in red and labeled. Serostatus was not tested for WUPyV due to very high seroprevalence. Chromosome X is not shown since data were not available for all studies.

Table 2. Significant HLA association results

Antibody levels				Serostatus			
Polyomavirus (sample size)	Marker	Z-score	P-value	Polyomavirus (sample size)	Marker	Z-score	P-value
HPyV6 (N = 5,091)	DQA1*05:01	−6.11	1.01×10^{-09}	JCPyV (N = 15,400)	DQB1*06:02	15.97	2.08×10^{-57}
	DRB1*03:01	−4.92	8.81×10^{-07}		DQA1*01:02	15.39	2.07×10^{-53}
	DQB1*02:01	−4.84	1.28×10^{-06}		DRB1*15:01	14.13	2.58×10^{-45}
	DQA1*01:01	4.48	7.53×10^{-06}		B*07:02	8.10	5.45×10^{-16}
	DRB1*01:01	4.33	1.47×10^{-05}		C*07:02	6.92	4.69×10^{-12}
JCPyV (N = 8,567)	DRB1*15:01	9.10	9.19×10^{-20}		DRB1*13:01	−6.27	3.71×10^{-10}
	DQB1*06:02	9.05	1.43×10^{-19}		DQB1*06:03	−5.97	2.43×10^{-09}
	DQA1*01:02	8.70	3.30×10^{-18}		DQB1*02:01	−5.87	4.27×10^{-09}
	DQB1*03:01	−5.29	1.19×10^{-07}		DRB1*03:01	−5.65	1.57×10^{-08}
	B*07:02	4.46	8.12×10^{-06}		DQA1*05:01	−5.55	2.85×10^{-08}
	DRB1*03:01	−4.21	2.57×10^{-05}		DQB1*03:01	−5.39	7.01×10^{-08}
	DQB1*02:01	−4.16	3.17×10^{-05}		DQA1*01:03	−4.97	6.56×10^{-07}
	DQA1*05:01	−4.14	3.51×10^{-05}		B*08:01	−4.97	6.70×10^{-07}
MCPyV (N = 8,091)	DQA1*01:01	−9.17	4.96×10^{-20}		DQB1*05:02	4.88	1.07×10^{-06}
	DQB1*05:01	−9.13	7.12×10^{-20}		DRB1*16:01	4.74	2.13×10^{-06}
	A*29:02	8.64	5.85×10^{-18}		DRB1*04:01	−4.49	7.18×10^{-06}
	DRB1*15:01	8.58	9.68×10^{-18}		C*07:01	−4.44	8.94×10^{-06}
	DRB1*01:01	−7.64	2.20×10^{-14}	MCPyV (N = 11,714)	DQB1*06:02	13.65	1.91×10^{-42}
	DRB1*04:04	7.57	3.85×10^{-14}		DRB1*15:01	13.54	9.35×10^{-42}
	DQB1*06:02	7.54	4.72×10^{-14}		DQA1*01:02	11.28	1.59×10^{-29}
	DQA1*01:02	6.28	3.43×10^{-10}		B*07:02	9.98	1.93×10^{-23}
	DRB1*11:01	−5.92	3.24×10^{-09}		C*07:02	8.49	2.12×10^{-17}
	C*07:02	5.63	1.86×10^{-08}		DQA1*01:01	−7.40	1.40×10^{-13}
	B*07:02	5.22	1.76×10^{-07}		DQB1*05:01	−7.12	1.07×10^{-12}
	C*16:01	4.79	1.64×10^{-06}		DRB1*01:01	−6.63	3.29×10^{-11}
	DQB1*03:01	−4.61	4.12×10^{-06}		DQA1*05:01	−5.64	1.74×10^{-08}
	B*44:02	−4.24	2.22×10^{-05}		DRB1*11:01	−5.57	2.60×10^{-08}
	DPB1*06:01	4.23	2.35×10^{-05}		DRB1*04:04	5.40	6.86×10^{-08}
	B*44:03	4.18	2.95×10^{-05}		DQB1*03:02	4.50	6.81×10^{-06}

2.4 FUT2 non-secretor status is associated with stronger antibody responses to BKPyV and JCPyV

We observed significant associations in the *FUT2* gene on chromosome 19 for BKPyV and JCPyV (Fig. 2). The fucosyltransferase 2 protein, encoded by *FUT2*, is expressed in epithelial cells and adds fucose to type 1 glycoprotein chains to form soluble ABO blood group (ABH) antigens that can be found in secretions and fluids (Kelly et al. 1995; Rouquier et al. 1995). The top associated single-nucleotide polymorphisms (SNPs) for BKPyV (rs681343, $P = 6.40 \times 10^{-16}$) and JCPyV (rs12462111, $P = 4.84 \times 10^{-10}$) are in high linkage disequilibrium (LD, $r^2 > 0.7$), with the common *FUT2* nonsense variant rs601338 ($P = 6.81 \times 10^{-16}$ and $P = 1.89 \times 10^{-09}$, respectively). The homozygous genotype for the rs601338 minor allele (AA) is classically referred to as the 'non-secretor' variant, as individuals homozygous for this null allele do not secrete blood group antigens at epithelial surfaces (Kelly et al. 1995; Rouquier et al. 1995). Among the three cohorts, 24 per cent of individuals were FUT2 non-secretors (AA) and 76 per cent were FUT2 secretors (GG or GA). Overall, we observed that non-secretors had significantly higher IgG levels against both polyomaviruses in comparison to secretors (Fig. 3).

2.5 Genomic variation in STING1 is associated with antibody response to MCPyV

We observed genome-wide significant associations with anti-MCPyV IgG levels in the *STING1* gene on chromosome 5 (Fig. 2). *STING1* encodes the endoplasmic reticulum-resident membrane protein STING that functions as a major regulator of innate immunity. Upon viral infection and activation of systolic cyclic GMP-AMP synthase, STING mediates type I interferon (IFN-I) production by infected cells to protect them and neighboring cells from local infection (Barber 2011). The top associated variant in meta-analysis using GRAS and UKB was rs7444313 ($P = 5.95 \times 10^{-15}$), a strong eQTL for *STING1* in the GTEx data set, whose G allele was associated with a stronger IgG response against MCPyV VP1 (GTEx Consortium 2013).

2.6 Alternative splicing of MUC1 is associated with humoral immune response to WUPyV

We observed a genome-wide significant association with anti-WUPyV IgG levels on chromosome 1, mapping to the *MUC1* gene that encodes the Mucin1 protein (lead SNP rs12411216, $P = 2.70 \times 10^{-30}$; Fig. 2). This variant is in perfect LD ($r^2 = 1$), with the splicing variant rs4072037 ($P = 4.08 \times 10^{-30}$) located in the second exon of the *MUC1* gene. The rs4072037 T allele was associated with a stronger IgG response against WUPyV VP1. The same allele was previously associated with an increased risk of gastric cancer (Zheng et al. 2013; Liu et al. 2014). Because infections with *H. pylori* and EBV have been implicated in the pathogenesis of gastric cancers, we searched for potential associations of rs4072037 with the humoral response to these pathogens (Shibata and Weiss 1992; Tokunaga et al. 1993; Takada 2000; Li et al. 2003). Using GWAS data available in CoLaus, we performed serological analyses for six different *H. pylori* antigens and four EBV antigens but did not find evidence for an rs4072037-dependent differential antibody response (Supplementary Table S1).
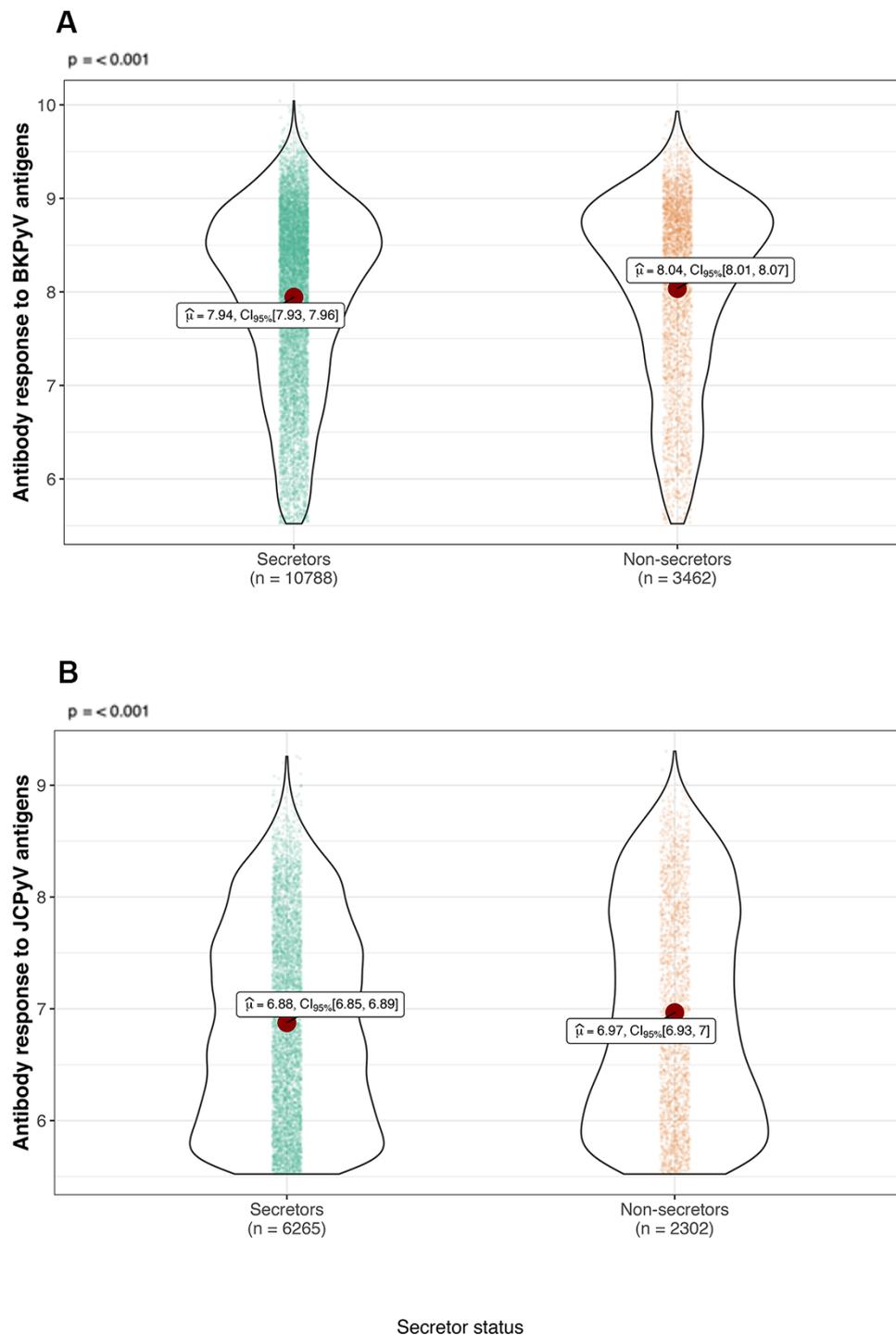
Figure 3. Serum IgG levels against BKPyV and JCPyV, by FUT2 genotype/secretor status. Violin plots show the log-transformed antibody levels (in MFI) against (A) BKPyV and (B) JCPyV. Antibody levels are categorized according to individuals, FUT2 secretor or non-secretor genotype. Red dots represent the mean. *P*-values based on t-test are shown.

2.7 Relative abundance of short MUC1 isoforms depends on rs4072037 allele status

The SNP rs4072037 has been functionally investigated previously and was shown to cause alternative splicing at the 5′ end of exon 2, leading to the absence or presence of nine amino acids in the signal peptide region (Zheng et al. 2013). Due to the difficulty in determining full exon connectivity from short-read sequencing data, the distribution of isoforms remained unclear. Therefore,

we have decided to use nanopore sequencing technology, allowing us to investigate sequencing long reads. We obtained twelve stomach RNA samples from GTEx: six from homozygous carriers of the rs4072037 major allele T and six from homozygous carriers of the minor allele C (GTEx Consortium 2013). We amplified the MUC1 region and sequenced all barcoded amplicons on a single flow cell on the MinION sequencing device. In total, we obtained 1,121 reads spanning the full-length MUC1 isoforms, i.e., having

assignable barcodes at both ends for downstream analysis (Supplementary Figs. S3–S4; Supplementary Table S2). We exclusively observed an isoform with a short exon 2 in the TT individuals, whereas a longer isoform was always present in the CC individuals, thereby confirming the modulating effect of rs4072037 genotype on MUC1 splicing.

2.8 Similar results using a case–control study design

We then investigated whether the genetic architecture of antibody levels against human polyomaviruses is similar to that of serostatus of the same polyomaviruses. Serological status was defined based on the manufacturers' defined cutoff of 250 MFI (see the 'Materials and methods' section). Individuals with levels of IgG antibodies above the threshold were considered seropositive (cases), while those below the threshold were seronegative (controls) (Table 1). Case–control GWASs using logistic regression were performed independently in the three cohorts and meta-analyzed (Willer, Li, and Abecasis 2010). Q–Q plots revealed no inflation of the test statistics for HPyV6 ($\lambda = 0.99$) and JCPyV ($\lambda = 0.98$), while MCPyV ($\lambda = 0.90$) and BKPyV ($\lambda = 0.72$) showed varying degrees of deflation (Supplementary Fig. S1). Chicago plots for the genome-wide significant loci are presented in Fig. 2. We again observed evidence of an involvement of HLA class II variation for JCPyV (rs9271025, $P = 4.31 \times 10^{-46}$) and MCPyV (rs73401330, $P = 8.49 \times 10^{-29}$) (Table 2). Strongest associations with seropositivity were observed for HLA-DQB1 allele. We also replicated the associations observed in FUT2 for JCPyV (rs2432132, $P = 1.70 \times 10^{-15}$) and STING1 for MCPyV (rs55792153, $P = 2.97 \times 10^{-09}$). Associations with FUT2 and HLA variants were not replicated for BKPyV and HPyV6, respectively. In addition, the case–control analyses did not reveal any additional associated region, suggesting that there are no SNPs purely associated with susceptibility.

3. Discussion

Human polyomaviruses are highly prevalent and can cause severe disease in immunocompromised individuals. They present a wide range of tissue tropisms and appear to cause neurologic, renal, and skin diseases (Feng et al. 2008; Ferenczy et al. 2012; Kuppachi et al. 2016). While their molecular mechanisms of infection are well-known, the potential impact of human genetic variation on their life cycle and pathogenicity remains understudied. For this study, we were interested in gaining insight into the genetic underpinning of the humoral immune responses to five polyomaviruses: BKPyV, JCPyV, HPyV6, MCPyV, and WUPyV. Their seroprevalence was high in all three cohorts included in our study (55–96 per cent), consistent with previously published surveys and confirming that human polyomavirus infections are common in the general population (Hammer et al. 2015; Kamminga et al. 2018).

Genes within the HLA region are prime candidates for susceptibility to viral infection and regulation of the immune response. HLA has been associated with multiple infectious diseases such as HIV, hepatitis B, hepatitis C, and tuberculosis (Sanchez-Mazas 2020). In this study, we confirm the involvement of HLA variation in the modulation of humoral immune response to two polyomaviruses, JCPyV and MCPyV (Hammer et al. 2015; Scepanovic et al. 2018; Kachuri et al. 2020). Specifically, we observed the strongest associations with quantitative antibody levels for HLA-DRB1 and HLA-DQA1 alleles and with seropositivity for HLA-DQB1 allele.

Our study also shows that FUT2 variation (rs601338 G>A) plays a role in the determination of antibody levels against BKPyV and JCPyV VP1. The FUT2 enzyme catalyzes the addition of a fucose residue to type 1 glycoprotein chains to form ABH blood group antigens that are released in secretions and at mucosal surfaces (Marionneau et al. 2001). Consistent with previous reports, 78 per cent of our study population carried at least one functional allele, which is sufficient to render an individual FUT2 secretor, whereas 22 per cent of individuals were found to be homozygous for the nonsense variant, leading to defective FUT2 and the absence of ABH antigens in secretions (FUT2 non-secretors) (Kelly et al. 1995; Azad, Wade, and Timpson 2018). Interestingly, non-secretors had significantly higher IgG titers against BKPyV and JCPyV compared to secretors, suggesting higher antigen exposure and/or stronger immune response. One hypothetic explanation is that the binding of polyomaviruses to host cell sialic acid receptors could be impaired in the presence of FUT2-dependent blood antigens on the mucosal surfaces. Consequently, non-secretors might be more permissive to viral propagation and replication than secretors. The same FUT2 variant is known to have an impact on several infections: non-secretors are protected against norovirus and rotavirus but are at higher risk of infection with other pathogens, including, e.g., *Candida albicans, Streptococcus pneumoniae*, and mumps (Thorven et al. 2005; Greenberg and Estes 2009; Ruvoën-Clouet, Belliot, and Le Pendu 2013; Imbert-Marcille et al. 2014; Tian et al. 2017; Azad, Wade, and Timpson 2018). The capacity to generate soluble ABH blood group antigens, dependent on the functionality of the FUT2 enzyme, represents an important modulator of infectious disease susceptibility in humans.

Our analyses confirm the implication of the STING1 gene in the determination of antibody levels against MCPyV VP1, which was recently reported by Kachuri et al. (2020). The variants most associated with anti-MCPyV IgG levels are strong regulators of the expression levels of STING1. The encoded protein, STING, plays a pivotal role in anti-viral innate immunity by activating IFN-I and proinflammatory cytokines. Individuals carrying the homozygous genotype for the rs7444313 minor allele (GG) had significantly higher anti-MCPyV IgG levels. A recent study by Kennedy *et al.* reported an association between the same variant and the intensity of the immune response following primary smallpox vaccination (Kennedy et al. 2020). The same study also showed that expression of the correlated non-synonymous rs1131769 variant (R232H) in the STING1 gene leads to a decrease in IFN-α expression levels. This result substantiates a previous finding showing reduced IFN-β transcription for carriers of the R232H variant (Zhang et al. 2013). Evidence suggests that this amino acid change may have a modifying effect on the function of STING. Due to a weaker binding between cGAMP and the H232 variant, STING might be less effective, resulting in a lower immune response. In turn, this might allow higher viral replication, resulting in increased anti-MCPyV IgG levels. Interestingly, rs7444313 has significantly different allele frequencies in European populations (ref. allele: $G = 0.28$) relative to African populations (ref. allele: $G = 0.83$), suggesting a possible role of infection-driven selection in this genomic region.

We observed a very strong association signal with anti-WUPyV IgG response in MUC1, the gene encoding the Mucin1 protein. There is no established role of WUPyV in disease thus far. However, it was detected in the trachea of immunocompromised children, in close proximity to MUC5AC-expressing cells (Siebrasse et al. 2016). Physiologically, cell surface mucins form a network that contributes to the mucosal barrier to infection (Hattrup and

Gendler 2008). MUC1, in particular, consists of a large extracellular O-glycosylated polypeptide backbone that extends 200–500 nm above the apical surface (Hollingsworth and Swanson 2004). The minor allele (T) of the top associated SNP, rs4072037, which associates with higher IgG levels, is known to alter the physiological functions of MUC1 because it leads to an alternative splicing of the 5′-region of exon 2. As a consequence, a shorter protein isoform is produced, which impairs the protective function of MUC1 in the gastric mucosa (Xu et al. 2009). The rs4072037 T/T genotype has been previously associated with an increased risk of gastric cancer (Xu et al. 2009; Jia et al. 2010), most likely due to a concomitant increase in the risk of gastritis due to *H. pylori* infection (Vinall et al. 2002). The newly discovered association with humoral response to WUPyV is most likely coincidental. Indeed, the absence of any trace of WUPyV in cancer samples suggests that this virus is unlikely to play any role in gastric carcinogenesis (Toptan et al. 2016). However, our results suggest once again that MUC1 plays an essential role in the protection of gastric mucosa and that a common, genetically encoded alteration of MUC1 has a deleterious impact on that important physiological function.

For human polyomaviruses with seroprevalence below 95 per cent, we investigated the relationship between the genetic contribution to serostatus and humoral immune response in seropositive individuals. Whereas the quantitative approach is purely looking at immune response in infected individuals, the case–control study design also has the potential to uncover genetic loci involved in susceptibility to infection. Our meta-analysis results suggest very similar genetic contributions to the continuous (quantitative) and binary (case–control) phenotypes, with comparable statistical strength. The same variable genetic loci in HLA class II, *FUT2* and *STING1*, were found to be associated with JCPyV6 and MCPyV serostatus and IgG levels. These similarities in the results could be explained by the impossibility of being able to differentiate between individuals who were never infected and individuals who were infected but had an extremely weak antibody response.

Here, we report the first meta-analyses aiming to identify human genetic determinants of the humoral immune response against multiple polyomaviruses. Statistically combining data from CoLaus, GRAS, and a subset of the UKB allowed us to run analyses on a total of 15,660 individuals of European ancestry. This provided greater statistical power and a more accurate estimate of the underlying effects, also reducing the risk of false-negative results. Some limitations need, however, to be considered. First, we did not have serological data for all five polyomaviruses in each cohort. Second, the meta-analyses relied on summary statistics derived from independently computed GWAS, which might result in biases due to differences in the design, analysis, and conduct of the individual studies. Third, our analyses were restricted to individuals of European ancestry, which limits the generalizability of our results. Finally, as mentioned above, individuals with very low IgG levels were excluded from the quantitative association analyses, due to the impossibility of distinguishing them from truly seronegative (i.e., never infected) study participants; this resulted in reduced statistical power.

In summary, we here report strong associations of HLA class II, *FUT2*, *STING1*, and *MUC1* genetic variants with the intensity of the humoral IgG response to multiple human polyomaviruses. Together, these results demonstrate the modulating contribution of host genetic variation to the individual response against some of the most prevalent human viruses.

4. Materials and methods
4.1 Study cohorts
4.1.1 *CoLaus study*
The *Cohorte Lausannoise* (CoLaus) is a population-based prospective study that started in 2003 in Lausanne, Switzerland (Firmann et al. 2008). It includes 6,188 individuals of European descent (47.5 per cent male) initially aged 35–75 years (mean ± SD: 51.1 ± 10.9), who were randomly selected from the general population and continuously undergo a follow-up every 5 years. Detailed phenotypic information was obtained from every study participant through questionnaires, a physical assessment, and biological measurement of blood and urine markers. The institutional Ethics Committee of the University of Lausanne, which afterward became the Ethics Commission of Canton Vaud (www.cer-vd.ch), approved the baseline CoLaus study (reference 16/03, decisions of 13 January and 10 February 2003), and written consent was obtained from all participants.

4.1.2 *The GRAS data collection*
The Göttingen Research Association for Schizophrenia (GRAS) data collection consists of 2,363 immunocompetent adults of European ancestry, comprising 1,147 anonymized blood donors (62.0 per cent male, mean age ± SD: 37.5 ± 13.2) and 1,216 individuals with psychiatric diagnoses (64.9 per cent male, mean age ± SD: 40.6 ± 13.5) (Begemann et al. 2010; Ribbe et al. 2010). All study participants provided informed consent, including genetic testing, and the GRAS data collection has been approved by the ethical committee of the Georg-August-University of Göttingen (master committee) as well as by the respective local regulatory/ethical committees of all collaborating centers (Ribbe et al. 2010).

4.1.3 *UK biobank*
The UK Biobank (UKB) is a population-based prospective study whose recruitment process has been described previously (Sudlow et al. 2015). Briefly, half a million men and women aged 40–69 years (45.6 per cent male, mean age ± SD: 56.5 ± 8.1) attended one of the twenty-two UKB assessment centers located throughout England, Scotland, and Wales between 2006 and 2010. All participants completed a touchscreen questionnaire and verbal interview and had a range of physical measurements and blood, urine, and saliva samples taken for long-term storage. A subset of 20,000 individuals attended a repeat assessment between 2012 and 2013. This project was conducted with approved access to UKB data under application numbers 50085 (PI: Fellay) and 43920. All UKB participants provided informed consent at recruitment. Ethics approval for the UKB study was obtained from the North West Centre for Research Ethics Committee (11/NW/0382).

4.2 Serological analyses
To assess humoral responses to polyomaviruses, serum samples were independently shipped by the three studies to the Infections and Cancer Epidemiology Division at the German Cancer Research Center (Deutsches Krebsforschungszentrum, DKFZ) in Heidelberg. Seroreactivity against BKPyV, HPyV6, JCPyV, MCPyV, and WUPyV viral capsid protein 1 (VP1) was measured at serum dilution 1:1,000 using multiplex serology based on glutathione-S-transferase–VP1 fusion capture immunosorbent assays combined with fluorescent bead technology (Waterboer et al. 2005). Antigen preparation and methodology, as well as robustness of seroprevalence estimates, have been described previously for human polyomaviruses (Gossai et al. 2016). For the quantitative

IgG level association analyses including only seropositive individuals, the results above the assay cutoff of 250 mean fluorescence intensity (MFI) were used. For the case–control study design, the results were expressed as a binary output (serostatus: anti-polyomavirus IgG-positive or IgG-negative) based on the same assay cutoff. Using the same assay, seroreactivity against six *H. pylori* antigens (HP1564 OMP, HP10 GroEL, HP547 CagA, HP887 VacA, HP73 UreaseA, and HP875 Catalase) and four EBV antigens (VCA p18, EBNA, Zebra, and EA-D) was measured in CoLaus.

4.3 DNA genotyping and statistical analyses

4.3.1 *CoLaus study*

DNA samples from 5,399 CoLaus participants were genotyped for 799,653 SNPs using the BB2 GSK-customized Affymetrix Axiom Biobank array. Quality control procedures described by Anderson *et al.* were applied (Anderson et al. 2010). A total of 4,781 individuals were included for further analyses. Variants were excluded if they were missing in >5 per cent of the subjects or if they deviated significantly from Hardy–Weinberg equilibrium (HWE, $P < 10 \times 10^{-7}$), resulting in 697,923 SNPs kept for phasing and imputation.

To statistically estimate the haplotypes, the genotypes were phased using EAGLE2 v2.0.5 software and the Haplotype Reference Consortium (HRC) r1.1 reference panel (Loh et al. 2016; McCarthy et al. 2016). Genotype imputation was performed using two independent reference panels: the HRC reference panel and the merged 1000 Genomes Phase 3 and UK10K reference panel (Birney and Soranzo 2015; Walter et al. 2015; Loh et al. 2016). Both phasing and imputation were performed using the Sanger Imputation Service (https://imputation.sanger.ac.uk). The imputed data set ($N = 89$ million SNPs) consisted of (1) SNPs imputed with HRC reference panel only, (2) SNPs imputed with the merged reference panel only, and (3) SNPs with the highest imputation information (INFO) score if imputed with both reference panels. To filter out poorly imputed or rare genotypes, SNPs with INFO score <0.8, minor allele frequency (MAF) <1 per cent, or significant deviation from HWE ($P < 10 \times 10^{-7}$) were excluded, resulting in a final number of 9,031,264 SNPs available for association analysis. We also imputed 111 four-digit classical HLA alleles using SNP2HLA and the T1DGC Immunochip/HLA reference panel (Jia et al. 2013).

Genome-wide association analyses were performed using mixed models in GCTA (Yang et al. 2011). HLA association analyses were performed using generalized linear models in PLINK v2 (Chang et al. 2015). We assumed an additive model for SNP associations and a dominant model for associations with HLA alleles. The top three principal components of ancestry, sex, and age were used as covariates in all analyses.

4.3.2 *The GRAS data collection*

Genome-wide SNP genotyping of GRAS study participants was performed using an Axiom® myDesign™ Genotyping Array (Affymetrix, Santa Clara, CA, USA). The quality control steps have been described previously (Hammer et al. 2015). Imputation of unobserved genotypes was performed using the 1000 Genomes Project Phase 1 v3 haplotypes as reference panel. Genotypes were pre-phased with MaCH v19 and subsequently imputed by Minimac (Howie et al. 2012). SNPs with a reported INFO score <0.8 or an MAF <5 per cent, as well as markers on sex chromosomes, were excluded from downstream analyses. SNPs were also filtered on the basis of missingness (excluded if <95 per cent genotyping rate) and marked deviation from HWE (excluded if $P < 5 \times 10^{-07}$). We then used linear regression models in PLINK v1.9 to test for association between ∼6 million SNPs and IgG responses using a continuous, quantitative approach (log-normalized IgG levels in seropositive samples) (Chang et al. 2015). The first three principal components of ancestry, calculated using GCTA (v1.24), as well as sex and age, were included as covariates in all analyses (Yang et al. 2011).

4.3.3 *UK Biobank*

Genotyping and imputation of UKB individuals have been fully described by Bycroft et al. (2018). Briefly, samples were genotyped on either the UK BiLEVE Axiom array (Affymetrix) or UKB Axiom array (Applied Biosystems). Genotypes were phased using SHAPEIT3 and the 1000 Genome Phase 3 data set as a reference and then imputed using IMUTE4 using the Haplotype Reference Consortium data, 1,000 Genomes Phase 3, and UK10K data as references. Serological markers of infection were measured for a random subset of 9,611 unrelated UKB participants (44.0 per cent male, mean age ± SD: 56.5 ± 8.2) (Mentzer et al. 2019). Association analyses were carried out using rank-based inverse normal transformed MFI values and linear mixed models as implemented in BOLT-LMM (Loh et al. 2015, 2018), using age at recruitment and genetic sex as covariates and variants with an MAF >0.01 and an INFO score >0.3.

4.4 Meta-analyses

Meta-analyses of the GWAS results obtained in the different cohorts were performed using a fixed-effect model in METAL (Willer, Li, and Abecasis 2010). For BKPyV and JCPyV, summary statistics of all three studies were used. For HPyV6 and WUPyV, only the results of CoLaus and GRAS studies were used, and for MCPyV, only the results of UKB and GRAS were used, as the data were not available in the UKB and CoLaus, respectively (Fig. 1). The meta-analyses followed the sample size-based strategy, considering *P*-value, sample size, and direction of effect. The Bonferroni-corrected significance level threshold of $P = 5.6 \times 10^{-09}$ ($P = 5 \times 10^{-08}$ divided by the eight GWAS performed) was applied. Similarly, for HLA alleles, we determined significance thresholds correcting for the number of alleles, the viruses tested, and the two study designs ($P = 4.3 \times 10^{-05}$).

4.5 Inference of FUT2 secretor status

FUT2 secretor status was determined from the rs601338 genotype, where the wild-type rs601338 (G) encodes the secretor allele, whereas A is the nonsense 'non-secretor' allele. Individuals with homozygous major (GG) or heterozygous (GA) genotypes were considered as secretors and compared with homozygous minor (AA) non-secretors (Kelly et al. 1995).

4.6 MUC1 full-length mRNA and nanopore sequencing

A total of twelve RNA samples from gastrointestinal tissue (200–225 ng each) were obtained from the Genotype-Tissue Expression (GTEx) project (GTEx Consortium 2013). Samples were selected based on expression values of the MUC1 gene in the GTEx expression data set and based on the genotype of the associated variant. Nine donors were male, and three donors were female. Retrotranscription was performed using SMARTer™ PCR cDNA Synthesis Kit from Clontech. We used the following forward and reverse primers to amplify the complementary (cDNA) of the MUC1 transcript: 5′-GCGCCTGCCTGAATCTGTTC-3′ and 5′-CCCACATGAGCTTCCACACAC-3′. For the polymerase

chain reaction, we used LongAmp™ Taq 2X Master Mix (Super-Script from Life Technologies/Q5 High-Fidelity from New England BioLabs) and the following protocol: 30 seconds at 94°C for initial denaturation, followed by 30 cycles at 94°C for 10 seconds, 55°C for 60 seconds, 65°C for 50 seconds, and final extension at 65°C for 10 minutes. To pool the twelve samples for sequencing, we used native EXP-PBC001 barcoding kit on R9.4 flow cell. The following adapter sequences for MinION sequencing were added to the forward and reverse MUC1 primers, respectively: 5′-TTTCTGTTGGTGCTGATATT-3′ and 5′-ACTTGCCTGTCGCTCTATCTTC-3′. We used the SQK-MAP006 protocol for library preparation for nanopore sequencing on MinION (Oxford Nanopore Technologies, ONT). Briefly, the combined end repairing and dA-tailing step was performed using the NEB-Next Ultra™ II End Repair/dA-Tailing Module (New England BioLabs). Samples were incubated for 5 minutes at 20°C followed by 5 minutes at 65°C. The reactions were purified by using Agencourt AMPure XP beads (Beckman Coulter) and further ligated to HP-adapter using Blunt/TA Ligase Master Mix (New England BioLabs) with 10 minutes of incubation at room temperature. Adapted DNA was purified using Dynabeads MyOne Streptavidin C1 (Life Technologies) and washed to remove unbound DNA. The captured cDNA library was eluted by resuspending the beads in ONT's Elution Buffer for 10 minutes at 37°C, and the beads were pelleted using a magnetic rack, leaving the supernatant containing the library. We used poretools (Loman and Quinlan 2014) to extract FASTA sequences from the ONT fast5 format, with a minimum length of 600 base pairs and forward and reverse strand sequence (2d reads). We then used nanocorrect (Loman, Quick, and Simpson 2015) for computational error correction and aligned all corrected reads to a list of MUC1 isoforms using the LAST aligner (Kiełbasa et al. 2011). Each read was assigned to the isoform with the best alignment score.

Data availability

Full meta-analyses association results are available for download from Zenodo (https://doi.org/10.5281/zenodo.4189621).

Supplementary data

Supplementary data is available at *Virus Evolution* online.

Funding

Conflict of interest: C.H. is an employee of Genentech Inc., a member of the Roche Group.

P.S. is an employee of the Roche Group.

The remaining authors declare that they have no competing interests.

Author contribution

C.H. and J.F. contributed to the conception of the study. F.H., C.H., and J.F. contributed to the design of the study. F.H., A.C., O.N., P.M.-V., P.V., M.B., H.E., N.B., N.B., T.W., A.M., A.H., C.H., and J.F. contributed to the acquisition of the data. F.H., A.C., P.S., and C.H. contributed to the analysis of the data. F.H., A.C., C.H., and J.F. contributed to the drafting of the manuscript. F.H., A.C., P.S., O.N., C.T., S.R., P.M.-V., H.E., N.B., N.B., T.W., A.M., C.H., and J.F. contributed to the revising of the manuscript. All the authors read and approved the final manuscript.

References

Anderson, C. A. et al. (2010) 'Data Quality Control in Genetic Case-control Association Studies', *Nature Protocols*, 5: 1564–73.

Azad, M. B., Wade, K. H., and Timpson, N. J. (2018) 'FUT2 Secretor Genotype and Susceptibility to Infections and Chronic Conditions in the ALSPAC Cohort [Version 2; Peer Review: 2 Approved]', *Wellcome Open Research*, 3: 65.

Barber, G. N. (2011) 'Innate Immune DNA Sensing Pathways: STING, AIMII and the Regulation of Interferon Production and Inflammatory Responses', *Current Opinion in Immunology*, 23: 10–20.

Begemann, M. et al. (2010) 'Modification of Cognitive Performance in Schizophrenia by Complexin 2 Gene Polymorphisms', *Archives of General Psychiatry*, 67: 879–88.

Birney, E., and Soranzo, N. (2015) 'Human Genomics: The End of the Start for Population Sequencing', *Nature*, 526: 52–3.

Bycroft, C. et al. (2018) 'The UK Biobank Resource with Deep Phenotyping and Genomic Data', *Nature*, 562: 203–9.

Chang, C. C. et al. (2015) 'Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets', *GigaScience*, 4: 7.

Cook, L. (2016) 'Polyomaviruses', *Diagnostic Microbiology of the Immunocompromised Host*, 197–216.

Crum-Cianflone, N. et al. (2007) 'BK Virus-associated Renal Failure among HIV Patients', *AIDS*, 21: 1501–2.

Egli, A. et al. (2009) 'Prevalence of Polyomavirus BK and JC Infection and Replication in 400 Healthy Blood Donors', *The Journal of Infectious Diseases*, 199: 837–46.

Feng, H. et al. (2008) 'Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma', *Science*, 319: 1096–100.

Ferenczy, M. W. et al. (2012) 'Molecular Biology, Epidemiology, and Pathogenesis of Progressive Multifocal Leukoencephalopathy, the JC Virus-Induced Demyelinating Disease of the Human Brain', *Clinical Microbiology Reviews*, 25: 471–506.

Firmann, M. et al. (2008) 'The CoLaus Study: A Population-Based Study to Investigate the Epidemiology and Genetic Determinants of Cardiovascular Risk Factors and Metabolic Syndrome', *BMC Cardiovascular Disorders*, 8: 6.

Gardner, S. D. et al. (1971) 'New Human Papovavirus (B.K.) Isolated from Urine after Renal Transplantation', *The Lancet*, 1: 1253–7.

Gossai, A. et al. (2016) 'Seroepidemiology of Human Polyomaviruses in a US Population', *American Journal of Epidemiology*, 183: 61–9.

Greenberg, H. B., and Estes, M. K. (2009) 'Rotaviruses: From Pathogenesis to Vaccination', *Gastroenterology*, 136: 1939–51.

Gross, L. (1953) 'A Filterable Agent, Recovered from Ak Leukemic Extracts, Causing Salivary Gland Carcinomas in C3H Mice', *Proceedings of the Society for Experimental Biology and Medicine*, 83: 414–21.

Grundbacher, F. J. (1974) 'Heritability Estimates and Genetic and Environmental Correlations for the Human Immunoglobulins G, M, and A', *American Journal of Human Genetics*, 26: 1–12.

GTEx Consortium. (2013) 'The Genotype-Tissue Expression (Gtex) Project', *Nature Genetics*, 45: 580–5.

Hammer, C. et al. (2015) 'Amino Acid Variation in HLA Class II Proteins Is a Major Determinant of Humoral Response to Common Viruses', *American Journal of Human Genetics*, 97: 738–43.

Hattrup, C. L., and Gendler, S. J. (2008) 'Structure and Function of the Cell Surface (Tethered) Mucins', *Annual Review of Physiology*, 70: 431–57.

Hollingsworth, M. A., and Swanson, B. J. (2004) 'Mucins in Cancer: Protection and Control of the Cell Surface', *Nature Reviews. Cancer*, 4: 45–60.

Howie, B. et al. (2012) 'Fast and Accurate Genotype Imputation in Genome-wide Association Studies through Pre-phasing', *Nature Genetics*, 44: 955–9.

Imbert-Marcille, B.-M. et al. (2014) 'A FUT2 Gene Common Polymorphism Determines Resistance to Rotavirus A of the P[8] Genotype', *The Journal of Infectious Diseases*, 209: 1227–30.

Jia, X. et al. (2013) 'Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens', *PLoS One*, 8: e64683.

Jia, Y. et al. (2010) 'A Comprehensive Analysis of Common Genetic Variation in MUC1, MUC5AC, MUC6 Genes and Risk of Stomach Cancer', *Cancer Causes and Control*, 21: 313–21.

Kachuri, L. et al. (2020) 'The Landscape of Host Genetic Factors Involved in Immune Response to Common Viral Infections', *Genome Medicine*, 12: 93.

Kamminga, S. et al. (2018) 'Seroprevalence of Fourteen Human Polyomaviruses Determined in Blood Donors', *PLoS One*, 13: e0206273.

Kelly, R. J. et al. (1995) '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', *The Journal of Biological Chemistry*, 270: 4640–9.

Kennedy, R. B. et al. (2020) 'Polymorphisms in STING Affect Human Innate Immune Responses to Poxviruses', *Frontiers in Immunology*, 11: 567348.

Kiełbasa, S. M. et al. (2011) 'Adaptive Seeds Tame Genomic Sequence Comparison', *Genome Research*, 21: 487–93.

Kuppachi, S. et al. (2016) 'BK Polyoma Virus Infection and Renal Disease in Non-renal Solid Organ Transplantation', *Clinical Kidney Journal*, 9: 310–8.

Li, H. et al. (2003) 'Helicobacter and Gastric Cancer Disease Mechanisms: Host Response and Disease Susceptibility', *Current Gastroenterology Reports*, 5: 459–67.

Liu, X. et al. (2014) 'MUC1 Gene Polymorphism Rs4072037 and Susceptibility to Gastric Cancer: A Meta-analysis', *SpringerPlus*, 3: 599.

Loh, P.-R. et al. (2016) 'Reference-based Phasing Using the Haplotype Reference Consortium Panel', *Nature Genetics*, 48: 1443–8.

—— (2018) 'Mixed-Model Association for Biobank-Scale Datasets', *Nature Genetics*, 50: 906–8.

—— (2015) 'Efficient Bayesian Mixed-model Analysis Increases Association Power in Large Cohorts', *Nature Genetics*, 47: 284–90.

Loman, N. J., Quick, J., and Simpson, J. T. (2015) 'A Complete Bacterial Genome Assembled De Novo Using Only Nanopore Sequencing Data', *Nature Methods*, 12: 733–5.

Loman, N. J., and Quinlan, A. R. (2014) 'Poretools: A Toolkit for Analyzing Nanopore Sequence Data', *Bioinformatics*, 30: 3399–401.

Marionneau, S. et al. (2001) 'ABH and Lewis Histo-blood Group Antigens, a Model for the Meaning of Oligosaccharide Diversity in the Face of a Changing World', *Biochimie*, 83: 565–73.

McCarthy, S. et al. (2016) 'A Reference Panel of 64,976 Haplotypes for Genotype Imputation', *Nature Genetics*, 48: 1279–83.

Mentzer, A. J. et al. (2019) 'Identification of Host-Pathogen-Disease Relationships Using a Scalable Multiplex Serology Platform in UK Biobank', *medRxiv*. 10.1101/19004960.

Padgett, B. L. et al. (1971) 'Cultivation of Papova-like Virus from Human Brain with Progressive Multifocal Leucoencephalopathy', *The Lancet*, 1: 1257–60.

Ribbe, K. et al. (2010) 'The Cross-sectional GRAS Sample: A Comprehensive Phenotypical Data Collection of Schizophrenic Patients', *BMC Psychiatry*, 10: 91.

Rouquier, S. et al. (1995) 'Molecular Cloning of a Human Genomic Region Containing the H Blood Group Alpha(1,2)fucosyltransferase Gene and Two H Locus-related DNA Restriction Fragments. Isolation of a Candidate for the Human Secretor Blood Group Locus', *The Journal of Biological Chemistry*, 270: 4632–9.

Rubicz, R. et al. (2011) 'Genetic Factors Influence Serological Measures of Common Infections', *Human Heredity*, 72: 133–41.

—— (2013) 'A Genome-wide Integrative Genomic Study Localizes Genetic Factors Influencing Antibodies against Epstein-Barr Virus Nuclear Antigen 1 (EBNA-1)', *PLoS Genetics*, 9: e1003147.

Ruvoën-Clouet, N., Belliot, G., and Le Pendu, J. (2013) 'Noroviruses and Histo-blood Groups: The Impact of Common Host Genetic Polymorphisms on Virus Transmission and Evolution', *Reviews in Medical Virology*, 23: 355–66.

Sanchez-Mazas, A. (2020) 'A Review of HLA Allele and SNP Associations with Highly Prevalent Infectious Diseases in Human Populations', *Swiss Medical Weekly*, 150: w20214.

Scepanovic, P. et al. (2018) 'Human Genetic Variants and Age Are the Strongest Predictors of Humoral Immune Responses to Common Pathogens and Vaccines', *Genome Medicine*, 10: 59.

Shibata, D., and Weiss, L. M. (1992) 'Epstein-Barr Virus-associated Gastric Adenocarcinoma', *The American Journal of Pathology*, 140: 769–74.

Siebrasse, E. A. et al. (2016) 'Multiorgan WU Polyomavirus Infection in Bone Marrow Transplant Recipient', *Emerging Infectious Diseases*, 22: 24–31.

Sudlow, C. et al. (2015) 'UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age', *PLoS Medicine*, 12: e1001779.

Takada, K. (2000) 'Epstein-Barr Virus and Gastric Carcinoma', *Molecular Pathology: MP*, 53: 255–61.

Thorven, M. et al. (2005) 'A Homozygous Nonsense Mutation (428G—>A) in the Human Secretor (FUT2) Gene Provides Resistance to Symptomatic Norovirus (GGII) Infections', *Journal of Virology*, 79: 15351–5.

Tian, C. et al. (2017) 'Genome-wide Association and HLA Region Fine-mapping Studies Identify Susceptibility Loci for Multiple Common Infections', *Nature Communications*, 8: 599.

Tokunaga, M. et al. (1993) 'Epstein-Barr Virus in Gastric Carcinoma', *The American Journal of Pathology*, 143: 1250–4.

Toptan, T. et al. (2016) 'Survey for Human Polyomaviruses in Cancer', *Journal of Clinical Investigation Insight*, 1(2): e85562.

Trofe-Clark, J., and Sawinski, D. (2016) 'BK and Other Polyomaviruses in Kidney Transplantation', *Seminars in Nephrology*, 36: 372–85.

Vinall, L. E. et al. (2002) 'Altered Expression and Allelic Association of the Hypervariable Membrane Mucin MUC1 in Helicobacter pylori Gastritis', *Gastroenterology*, 123: 41–9.

Viscidi, R. P. et al. (2011) 'Age-specific Seroprevalence of Merkel Cell Polyomavirus, BK Virus, and JC Virus', *Clinical and Vaccine Immunology I*, 18: 1737–43.

UK10K Consortium, Walter, K. et al. (2015) 'The UK10K Project Identifies Rare Variants in Health and Disease', *Nature*, 526: 82–90.

Waterboer, T. et al. (2005) 'Multiplex Human Papillomavirus Serology Based on in Situ-Purified Glutathione S-Transferase Fusion Proteins', *Clinical Chemistry*, 51: 1845–53.

Willer, C. J., Li, Y., and Abecasis, G. R. (2010) 'METAL: Fast and Efficient Meta-analysis of Genomewide Association Scans', *Bioinformatics*, 26: 2190–1.

Xu, Q. et al. (2009) 'Risk of Gastric Cancer Is Associated with the MUC1 568 A/G Polymorphism', *International Journal of Oncology*, 35: 1313–20.

Yang, J. et al. (2011) 'GCTA: A Tool for Genome-wide Complex Trait Analysis', *American Journal of Human Genetics*, 88: 76–82.

Zhang, X. et al. (2013) 'Cyclic GMP-AMP Containing Mixed Phosphodiester Linkages Is an Endogenous High-affinity Ligand for STING', *Molecular Cell*, 51: 226–35.

Zheng, L. et al. (2013) 'Functional Polymorphism Rs4072037 in MUC1 Gene Contributes to the Susceptibility to Gastric Cancer: Evidence from Pooled 6,580 Cases and 10,324 Controls', *Molecular Biology Reports*, 40: 5791–6.

Zurhein, G., and Chou, S. M. (1965) 'Particles Resembling Papova Viruses in Human Cerebral Demyelinating Disease', *Science*, 148: 1477–9.