A systematic investigation of the contribution of genetic variation within the MHC region to HPV seropositivity

Dan Chen1,2,* Valérie Gaborieau3 Yao Zhao5 Amélie Chabrier4 Huibo Wang6 Tim Waterboer7 David Zardze9 Jolanta Lissowska10 Peter Rudnai11 Eleonora Fabianova12 Vladimir Bencko13 Vladimir Janout14 Lenka Foretova15 Ioan Nicolae Mates16 Neonila Szeszenia-Dabrowska17 Paolo Boffetta18 Michael Pawlita8 Mark Lathrop19 Ulf Gyllensten2 Paul Brennan3,* and James D. McKay4

1Ministry of Education and Shanghai Key Laboratory of Children’s Environmental Health, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, 2Department of Immunology, Genetics and Pathology, Science for Life Laboratory Uppsala, Uppsala University, Uppsala, Sweden, 3Genetic Epidemiology Group, 4Genetic Cancer Susceptibility Group, International Agency for Research on Cancer (IARC), Lyon, France, 5Department of Neurosurgery, Huashan Hospital, Shanghai Medical School, Fudan University, Shanghai, China, 6Department of Neurosurgery, First Affiliated Hospital of Nanjing Medical University, Nanjing, China, 7Infections and Cancer Epidemiology Group, Division of Genome Modifications and Carcinogenesis, 8Virus-Host Interactions of Polyoma and Papilloma Viruses Group, Division of Genome Modifications and Carcinogenesis, German Cancer Research Center (DKFZ), Heidelberg, Germany, 9Institute of Carcinogenesis, Cancer Research Centre, Moscow, Russia, 10Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland, 11National Institute of Environmental Health, Budapest, Hungary, 12Regional Authority of Public Health, Banská Bystrica, Slovakia, 13Institute of Hygiene and Epidemiology, 1st Faculty of Medicine, Charles University, Prague, Czech Republic, 14Palacky University, Olomouc, Czech Republic, 15Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic, 16St Mary General and Esophageal Surgery Clinic, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania, 17Department of Epidemiology, Institute of Occupational Medicine, Lodz, Poland, 18Mount Sinai Hospital, Icahn Medical Institute, New York, USA and 19Centre D’innovation Genôme Québec et Université McGill, Montréal, Canada

*To whom correspondence should be addressed at: Ministry of Education and Shanghai Key Laboratory of Children’s Environmental Health, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, 1665 Kongjiang Road, 200092 Shanghai, China. Email: simplesandan1981@163.com; dan.chen@igp.uu.se (D.C.); Genetic Epidemiology Group, International Agency for Research on Cancer (IARC), 150 Cours Albert Thomas, 69372 Lyon CEDEX 08, France. Email: brennanp@iarc.fr (P.B.)

Abstract

High-risk mucosal types of human papillomavirus (HPV) cause anogenital and oropharyngeal cancers, whereas cutaneous types (e.g. HPV8 and 77) are suspected to be involved in non-melanoma skin cancer. The antibody response to HPVs is a key
determinant of protective immunity, but not all infected individuals seroconvert. Genetic variability of the host may have large impact on seroconversion. A previous genome-wide association study (GWAS) has identified a susceptibility locus (rs41270488) for HPV8 seropositivity within the major histocompatibility complex (MHC) region. To further study this locus, we imputed alleles at classical leukocyte antigen (HLA) loci using HLA*IMP:02 with a reference panel from the HapMap Project and the 1958 Birth Cohort, and conducted an integrated analysis among 4811 central European subjects to assess the contribution of classical HLA alleles and gene copy number variation (CNV) at the hypervariable DRB locus within the MHC region to HPV seropositivity at both the individual HPV type level and the phylogenetic species level. Our study provides evidence that the association noted between rs41270488 and HPV8 seropositivity is driven by two independent variants, namely DQB1*0301 (odds ratio (OR) = 1.51, 95% confidence interval (CI) = 1.36 – 1.68, \( P = 1.0 \times 10^{-14} \)) and DRB1*1101 (OR = 1.89, 95%CI = 1.57 – 2.28, \( P = 1.5 \times 10^{-11} \)) within the HLA class II region. Additionally, we identified two correlated alleles DRB1*0701 (OR = 1.67, 95%CI = 1.41 – 1.98, \( P = 2.6 \times 10^{-7} \)) and DQA1*0201 (OR = 1.67, 95%CI = 1.38 – 1.93, \( P = 1.7 \times 10^{-5} \)), to be associated with HPV77 seropositivity. Comparable results were observed through imputation using SNP2HLA with another reference panel from the Type 1 diabetes Genetics Consortium. This study provides support for an important role of HLA class II alleles in antibody response to HPV infection.

**Introduction**

The human papillomavirus (HPV) family comprises over 150 types that are grouped in different genera according to their DNA sequence homology (1). All HPVs infect cells in the basal layer of stratified epithelia at a variety of anatomical sites, e.g. mucosa of the genital tract and skin (2) and thus can be broadly subdivided into mucosal and cutaneous types on the basis of their tissue tropism. Although the majority of HPV types are associated with benign wart-like lesions, certain mucosal HPVs of the genus alpha (\( \alpha \)), referred to as high-risk (HR) mucosal HPV types (e.g. HPV16 and 18), are causally involved in cervical, other anogenital and oropharyngeal cancers (3). Many HPV types (e.g. HPV6 and 11) are causally involved in the development of psoriasis and non-melanoma skin cancer (NMSC), although the causal nature of such associations has been disputed (6–8). The cutaneous HPV type 77 that belongs to genus \( \alpha \) has been detected in NMSC of renal transplant recipients (9–11).

Despite HPV’s ability to evade host defense, a successful immune response to HPV infections is established in most cases. This seems to be characterized by cell-mediated immunity that is associated with the generation of serum neutralizing antibodies (12). Such antibody responses are generated in most, but not all, infected individuals (13). They are directed against conformational epitope(s) on the major capsid protein L1 displayed on the outer surface of the intact virus particle and is HPV type-specific (12). Antibody-mediated humoral immunity clears free virus particles from body fluids and can prevent viral reinfection (12,13). Serum neutralizing antibody levels following natural HPV infections, even at peak titters, are low (14). This is probably due to the exclusively intraepithelial infectious cycle, as well as the production of virus particles in the superficial layer of the epithelium, distant from antigen-presenting cells (APCs) and patrolling macrophages (12). Possibly, a large infection may produce more viral protein that may more effectively induce an antibody response (13). It is also expected that seroconversions should occur in response to infections sufficiently persistent to be exposed to the immune system (13). Therefore, HPV seropositivity may reflect a combination of cumulative exposure to HPV and subsequent capability of the host to mount an effective antibody response.

Genetic variability of the host may have impact on the outcome of an HPV infection, especially genetic factors that control the immune response (15). The major histocompatibility complex (MHC) region is the most important region in the vertebrate genome with respect to infection, and is crucial in innate and adaptive immunity, where leukocyte antigen (HLA) genes and many other immune-related genes are located (16). In particular, HLA molecules are responsible for the presentation of foreign antigens to the immune system (17), and they may play a central role in the immune recognition and subsequent serological response. Classical HLA alleles, defined by genetic variation within the peptide-binding site of HLA genes, are believed to influence the ability to bind and present different peptide antigens (18). Variation at classical HLA alleles has been implicated in susceptibility to a wide range of infections (18). In addition, there is a hypervariable DRB locus located in the MHC class II region displaying gene copy number variation (CNV) between the HLA-DRB9 pseudogene and HLA-DRB1 (19).

We have previously performed a genome-wide association study (GWAS) of seroreactivity to 13 HPV types within a central European population and identified a signal (rs41270488 or rs114427648) for HPV8 seropositivity within the MHC class II region (odds ratio (OR) = 1.51, 95% confidence interval (CI) = 1.36 – 1.68, \( P = 1.0 \times 10^{-14} \)) and DRB1*1101 (OR = 1.89, 95%CI = 1.57 – 2.28, \( P = 1.5 \times 10^{-11} \)). This finding was further replicated in another independent series (20). However, given the complex linkage disequilibrium (LD) pattern in the MHC region as well as lack of data on classical HLA alleles and the CNV at DRB locus, it is not clear whether the association with rs41270488 is driven by the neighboring classical HLA alleles or CNV. It is also unknown whether any classical HLA allele or CNV within the MHC region contribute to seropositivity to other HPV types. Detailed studies of the MHC region have been hampered by the lack of methods for high throughput genotyping. Both direct high-resolution HLA genotyping and next-generation sequencing (NGS) typing do not easily scale for large cohorts. Recently, a method that allows the imputation of alleles at key classical HLA loci based on the MHC single-nucleotide polymorphism (SNP) data has become available for analysis of large-scale data sets, with accuracy that exceeds 90% at the 4-digit level (21). Using existing genotype data from the previous GWAS of HPV seropositivity, we have performed an integrated analysis among 4811 central European subjects to assess the contribution of classical HLA alleles and CNV(s) at the hypervariable DRB locus within the MHC region to HPV seroreactivity at both the individual HPV type level and the phylogenetic species level.

**Results**

Subjects included in the present study were from a previous GWAS conducted in a central European case–control study (CE) of lung, head and neck and kidney cancer that had serology...
data available on HPV types (22), as described in the Materials and Methods section. Following rigorous quality control (QC) steps, genome-wide SNP data were available for 4811 subjects (1286 lung cancer patients, 679 head and neck cancer patients, 811 kidney cancer patients and 2035 cancer-free subjects). L1 seropositivity for 13 HPV types had seroprevalence above 5% (**α** mucosal: HPV 6, 11, 16, 18, 31, 35 and 45; **α** cutaneous: HPV77; **β** cutaneous: HPV8, 38 and 49; **γ** cutaneous: HPV4; and **μ** cutaneous: HPV1). The highest seroprevalence was observed for **β** cutaneous HPV8 (28.04%), followed by **α** mucosal HPV6 (24.78%) and **γ** cutaneous HPV4 (24.03%) (Supplementary Material, Table S1).

Detailed information for these 13 HPV types with the corresponding case (seropositive) and control (seronegative) counts is shown in Supplementary Material, Table S1. There are more than one HPV type for phylogenetic species alpha-7 (HPV 18 and HPV 45), alpha-9 (HPV 16, HPV 31 and HPV 35) and alpha-10 (HPV 6 and HPV 10), while there is only one HPV type for other species in our study.

Imputation of classical HLA alleles at the 4-digit level was successfully performed at three class I loci (HLA-A, HLA-B and HLA-C) and four class II loci (HLA-DRB1, HLA-DQA1, HLA-DQB1 and HLA-DPB1) with HLA*IMP:02 (21) using a reference panel from the HapMap Project (23,24) and the 1958 Birth Cohort (25). As shown in Figure 1, the gap in the MHC class II region, which is underrepresented in SNP genotyping, corresponds to the hypervariable DRB locus with CNV(s) between the HLA-DRB9 pseudogene (telomeric) and HLA-DRB1 (centromeric), where HLA class II haplotypes carry one or no additional functional HLA-DRB gene (HLA-DRB3, HLA-DRB4 or HLA-DRB5) and one, two or no additional copies of HLA-DRB pseudogene (HLA-DRB2, HLA-DRB6, HLA-DRB7 or HLA-DRB8) (Supplementary Material, Fig. S1) (19). Imputation was also performed for classical HLA alleles at the additional functional

Figure 1. (A) Regional association plot for HPV8 seropositivity in the MHC region. (B) Regional association plot for HPV77 seropositivity in the MHC region. Single-marker association results for directly genotyped (blue) and imputed SNPs (green) with Rsq (a measure of genotype certainty) ≥0.6, classical HLA alleles (red) and gene copy number variation (CNV) (brown) at the hypervariable DRB locus between the HLA-DRB9 pseudogene (telomeric) and HLA-DRB1 (centromeric). P-values in -log10 scale for each SNP (Y-axis) are plotted against their position (X-axis) on chromosome 6 (hg19). The horizontal dotted line represents the significance threshold level (P = 1.0 × 10^{-6}). Genes in the region are represented with arrow heads indicating the direction of transcription. The gap in the MHC class II region corresponds to the hypervariable DRB locus with gene copy number variation (CNV) between the HLA-DRB9 pseudogene (telomeric) and HLA-DRB1 (centromeric) (19) which is underrepresented on the HumanHap300 Genotyping BeadChip.
genes of HLA-DRB3, HLA-DRB4 and HLA-DRB5. In total, 129 classical HLA alleles were observed in the CE study, the information of which is shown in Supplementary Material, Table S2. Statistical analysis was performed among 4811 subjects for association between seropositivity of 13 HPV types with seroprevalence above 5% and 3 phylogenetic species (alpha-9, alpha-7 and alpha-10) and 1092 genotyped and 2818 imputed SNPs as well as 129 imputed classical HLA alleles and 3 types of gene CNV(s) (HLA-DRB3, HLA-DRB4 or HLA-DRB5). The association results of the classical HLA alleles and gene CNV(s) across HPV types and phylogenetic species are shown in Supplementary Material, Table S3. Given the strong LD between the variants in the MHC region, a significance threshold of $P = 1.0 \times 10^{-6}$ was used.

Four HLA class II alleles showed evidence of association with HPV8 seropositivity at the $P = 1.0 \times 10^{-6}$ threshold in the unconditional logistic regression analysis, namely, DQB1*0301 (OR = 1.51, 95%CI = 1.36–1.68, $P = 1.0 \times 10^{-14}$), DQA1*0505 (OR = 1.67, 95%CI = 1.44–1.93, $P = 7.1 \times 10^{-12}$), DRB3*0200 (OR = 1.44, 95%CI = 1.30–1.60, $P = 9.8 \times 10^{-12}$) and DRB1*1101 (OR = 1.89, 95%CI = 1.57–2.28, $P = 1.5 \times 10^{-11}$) (Table 1) (Fig. 1A). All of these alleles confer increased probability of antibody response against HPV8. Of the gene CNV(s) across the variable DRB locus, the haplotype carrying an additional HLA-DRB3 was significantly associated with HPV8 seropositivity (OR = 1.27, 95%CI = 1.16–1.39, $P = 7.0 \times 10^{-7}$).

The LD pattern between rs41270488 and the above HLA alleles is shown in Supplementary Material, Table S4. To evaluate the independence of associations, we conducted stepwise logistic regression analysis. As shown in Table 2, when conditioning on DQB1*0301, residual association was still detected at DRB1*1101 (OR = 1.54, 95%CI = 1.13–2.10) and DRB3*0200 (OR = 1.55, 95%CI = 1.17–2.02). Further conditioning on DRB1*1101 left little evidence for residual association at rs41270488 (OR = 1.22, 95%CI = 0.96–1.55, $P = 0.10$), DQA1*0505 (OR = 1.40, 95%CI = 0.94–2.09, $P = 0.10$) and DRB3*0200 (OR = 1.10, 95%CI = 0.93–1.29, $P = 0.27$). Collectively, these data provide evidence that the association we noted with rs41270488 is better explained by two independently acting classical HLA alleles for HPV8 seropositivity, defined by DQB1*0301 and DRB1*1101.

To validate the imputation results generated by HLA*IMP-02, the HLA alleles that were found to be significantly associated with seropositivity to HPV8 (DQB1*0301 and DRB1*1101) and HPV77 (DRB1*0701 and DQA1*0201) were also imputed using SNP2HLA using another reference panel from the Type 1 diabetes Genetics Consortium (T1DGC) (26). As shown in Table 4, comparable allele frequencies and association results for these HLA alleles were observed using SNP2HLA, with higher concordance observed for those with higher imputation quality, as measured with the most significant signal occurring at DRB1*1101 (OR = 1.55, 95%CI = 1.57–2.28, $P = 1.0 \times 10^{-11}$) in the unconditional analysis compared with OR = 1.89, 95%CI = 1.57–2.28, $P = 1.5 \times 10^{-11}$ in the unconditional analysis). Further conditioning on DRB1*1101 left little evidence for residual association at rs41270488 (OR = 1.22, 95%CI = 0.96–1.55, $P = 0.10$), DQA1*0505 (OR = 1.40, 95%CI = 0.94–2.09, $P = 0.10$) and DRB3*0200 (OR = 1.10, 95%CI = 0.93–1.29, $P = 0.27$). Collectively, these data provide evidence that the association we noted with rs41270488 is better explained by two independently acting classical HLA alleles for HPV8 seropositivity, defined by DQB1*0301 and DRB1*1101.

Table 1. Association between HPV8 seropositivity and classical HLA alleles imputed by HLA*IMP-02

<table>
<thead>
<tr>
<th>HLA alleles</th>
<th>Frequency Casesa</th>
<th>Controlsb</th>
<th>Number Casesa</th>
<th>Controlsb</th>
<th>AssociationOR (95%CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQB1*0301</td>
<td>0.28</td>
<td>0.21</td>
<td>826</td>
<td>826</td>
<td>1.51 (1.36–1.67)</td>
<td>1.0 x 10^{-14}</td>
</tr>
<tr>
<td>DQA1*0505</td>
<td>0.17</td>
<td>0.12</td>
<td>456</td>
<td>144</td>
<td>1.67 (1.44–1.99)</td>
<td>7.1 x 10^{-12}</td>
</tr>
<tr>
<td>DRB3*0200</td>
<td>0.28</td>
<td>0.21</td>
<td>740</td>
<td>740</td>
<td>1.44 (1.30–1.60)</td>
<td>9.8 x 10^{-12}</td>
</tr>
<tr>
<td>DRB1*1101</td>
<td>0.09</td>
<td>0.05</td>
<td>619</td>
<td>619</td>
<td>1.89 (1.57–2.28)</td>
<td>1.5 x 10^{-11}</td>
</tr>
<tr>
<td>CNV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.27 (1.16–1.39)</td>
<td>7.0 x 10^{-7}</td>
</tr>
</tbody>
</table>

OR, odds ratio; CI, confidence interval.

---

Table 2. Evaluation of the independence of associations for HPV8 seropositivity

<table>
<thead>
<tr>
<th>SNP/HLA alleles</th>
<th>Conditioning on DQB1*0301a</th>
<th>Conditioning on DQB1<em>0301 and DRB1</em>1101b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95%CI)</td>
<td>P-value</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>DRB1*1101</td>
<td>1.55 (1.24–1.94)</td>
<td>1.0 x 10^{-4}</td>
</tr>
<tr>
<td>rs41270488</td>
<td>1.34 (1.12–1.60)</td>
<td>1.4 x 10^{-3}</td>
</tr>
<tr>
<td>DQA1*0505</td>
<td>1.54 (1.13–2.10)</td>
<td>5.8 x 10^{-3}</td>
</tr>
<tr>
<td>DRB3*0200</td>
<td>1.18 (1.02–1.36)</td>
<td>0.02</td>
</tr>
<tr>
<td>DRB3</td>
<td>1.09 (0.97–1.20)</td>
<td>0.14</td>
</tr>
</tbody>
</table>

OR, odds ratio; CI, confidence interval.

---

*Derived from logistic regression for HLA alleles imputed by HLA*IMP-02 conditioning on DQB1*0301 assuming a log-additive model with adjustment for age, sex, country and cancer type (control, lung cancer, kidney cancer and head and neck cancer).

**Derived from logistic regression for HLA alleles imputed by HLA*IMP-02 conditioning on both DQB1*0301 and DRB1*1101 assuming a log-additive model with adjustment for age, sex, country and cancer type (control, lung cancer, kidney cancer and head and neck cancer).
ciated with HPV77 seropositivity, which are nearly perfectly correlated with each other, to be asso-
ciation with spontaneous clearance of the hepatitis B virus,

Discussion

In the present study, we have systematically examined the rela-
tionship between classical HLA alleles and CNV(s) at the hyper-
variable DRB locus within the MHC and HPV L1 seropositivity,
both the individual HPV type level and the phylogenetic species
level. The large data set and dense set of informative markers in
our study have allowed us to detect independent effects among
SNPs (both genotyped and imputed), classical HLA alleles and
and the CNV(s) at the hypervariable DRB locus within the MHC region.

Through imputation by HLA*IMP:02 using a reference panel from
the HapMap Project (23, 24) and the 1958 Birth Cohort (25), our
study provides evidence that the association of rs41270488 with
HPV8 seropositivity noted earlier (20), is driven by two indepen-
dently acting HLA class II alleles, defined by the DRB1*0301 and
DRB1*1101 alleles. We also identified DRB1*0701 and DQA1*0201,
which are nearly perfectly correlated with each other, to be asso-
ciated with HPV77 seropositivity, a finding that was not detected
using traditional SNP tests of association. This illustrates the
power of HLA imputation for identifying new variants associated
with HPV seropositivity. Comparable results were observed
through imputation using SNP2HLA with another reference
panel from the TIDGC.

The function of the HLA’s is to bind peptide fragments that are
produced as a result of intracellular protein degradation. HLA
then displays these fragments on the cell surface where peptides
are presented endogenously to cytotoxic CD8+ T cells, which kill
the presenting cells. Class II HLA molecules are expressed on
APCs (e.g. dendritic cells and macrophages) and present exogenous
antigens to CD4+ T cells. CD4+ T cell activation results in the
secretion of a variety of small proteins, or cytokines, that help
and regulate other cells (12). One major subset of CD4+ T cells
is known as Th2 cells, which help antigen-primed B lymphocytes
differentiate into plasma cells and secrete antibodies, the effect-
or molecules of humoral response (12). Therefore, HLA class II
molecules may play a more important role in the antibody im-

Table 3. Association between HPV77 seropositivity and classical HLA alleles imputed by HLA*IMP:02

<table>
<thead>
<tr>
<th>HLA alleles</th>
<th>Frequency Casesb</th>
<th>Controlsc</th>
<th>Number Casesb</th>
<th>Controlsc</th>
<th>Associationd OR (95%CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRB1*0701</td>
<td>0.20</td>
<td>0.13</td>
<td>492</td>
<td>4299</td>
<td>1.67 (1.41–1.98)</td>
<td>2.6 × 10−9</td>
</tr>
<tr>
<td>DQA1*0201</td>
<td>0.20</td>
<td>0.13</td>
<td>487</td>
<td>4262</td>
<td>1.63 (1.38–1.93)</td>
<td>1.7 × 10−8</td>
</tr>
</tbody>
</table>

OR, odds ratio; CI, confidence interval.

a Derived from unconditional logistic regression assuming the log-additive model with adjustment for age, sex, country and cancer type (control, lung cancer, kidney cancer and head and neck cancer).

b Subjects that do not have antibody to L1 of HPV77 in the serum who carry the specific HPV77 variant.

c Subjects that have antibody to L1 of HPV77 in the serum who carry the specific HPV77 variant.

d OR, odds ratio; CI, confidence interval.

by Rsq, a measure of imputation quality generated by SNP2HLA,
which estimates the squared correlation between imputed and
ture genotypes.
### Materials and Methods

#### Study population

The CE study is a large multi-center case–control study of lung, head and neck and kidney cancer, conducted from 1998 to 2002 in six central European countries including Czech Republic, Hungary, Poland, Romania, Russia and Slovakia. Briefly, for cancer subjects, only patients newly diagnosed with histopathologically confirmed cancer were consecutively recruited. Blood samples were collected before radiotherapy and chemotherapy. Informed consent was obtained from all subjects and this study was approved by an Ethical Committee. Details on study designs and subjects recruitment have been described previously (22). Genotyping was performed on DNA samples of 1413 lung cancer cases, 742 head and neck cancer cases, 853 kidney cancer cases and 2191 cancer-free controls (1286 lung cancer cases, 679 head and neck cancer cases, 811 kidney cancer cases and 2035 cancer-free controls) were available for 316 015 SNPs with an overall call rate of 98.94%.

#### Detection of HPV serology

Serum samples from study participants were analyzed for antibodies to the HPV major capsid protein L1 (α mucosal: 6, 11, 16, 18, 31, 33, 35, 45, 52, 58; α cutaneous: 77; beta cutaneous: 8, 38 and 49; γ cutaneous: 4; and μ cutaneous: 1) at the German Cancer Research Center, Heidelberg, Germany. The multiplex serology method was applied based on a glutathione S-transferase capture (GST) enzyme linked immunosorbent assay (ELISA) in combination with fluorescent bead technology, as described previously (20,29,30). More information on QC and data processing has been described elsewhere (31).

Median R-phycocerythrin fluorescence intensity (MFI) values were dichotomized as antibody positive or negative. Seropositivity cut-offs for mucosal HPV L1 were determined using serum samples of 371 female, HPV DNA negative, self-reported virgins from a cross-sectional study among Korean students, as reported previously (32). For assays of antibodies to the cutaneous types the definition of a mostly uninfected and thus mostly seronegative group is not possible. All antigens were identically constructed L1 fusion proteins expressed in the same bacterial expression system. The full-length L1 fusion protein density on the beads for different HPV types was very similar, since MFI values obtained after staining of the carboxy (C)-terminal tag power to detect associations of smaller magnitude for some HPV types or variants associated with HPV types with more modest seroprevalence. Larger studies with longitudinal design are needed to elucidate the genetic basis of the serological response for less frequent, particular HR mucosal HPVs.

In summary, our study provides for the first time an investigation of the contribution of classical HLA alleles and the CNV(s) at the hypervariable DRB locus to HPV seropositivity at both the individual HPV type level and the phylogenetic species level within the MHC region. We have identified two independent HLA alleles, namely DQB1*0301 and DRB1*1101, that are associated with HPV8 seropositivity. We also found DRB1*0701 and/or DQA1*0201 to be associated with HPV77 seropositivity. This study provides evidence for an important role of HLA class II alleles in antibody response to HPV infections.

### Table 4: Comparison of the imputation results of associated HLA alleles imputed by HLA*IMP:02 and SNP2HLA

<table>
<thead>
<tr>
<th>HLA Alleles</th>
<th>Number Cases</th>
<th>Frequency Cases</th>
<th>P-value</th>
<th>OR (95%CI)</th>
<th>P-het</th>
<th>Rsq</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV8</td>
<td>1329 (0.282)</td>
<td>1.54 (1.39–1.71)</td>
<td>1.0 × 10⁻¹¹</td>
<td>1.04 (0.99–1.10)</td>
<td>9.7 × 10⁻¹²</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQB1*0301</td>
<td>3420 (0.020)</td>
<td>1.51 (1.36–1.68)</td>
<td>0.09</td>
<td>1.44 (1.16–1.79)</td>
<td>2.6 × 10⁻⁹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPV77</td>
<td>494 (0.197)</td>
<td>1.65 (1.39–1.93)</td>
<td>0.04</td>
<td>1.63 (1.38–1.93)</td>
<td>1.7 × 10⁻⁹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQA1*0201</td>
<td>429 (0.130)</td>
<td>1.63 (1.38–1.93)</td>
<td>0.09</td>
<td>1.63 (1.38–1.93)</td>
<td>1.7 × 10⁻⁹</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Notes

- **Imputation was performed by HLA*IMP:02 using HapMap Project (24,25) and the 1958 Birth Cohort (26) as reference panel. Downstream analysis was restricted to subjects who carry the specified HLA alleles with each genotype having a posterior probability over 0.7.**

- **Imputation was performed by SNP2HLA using the Type 1 diabetes Genetics Consortium (T1DGC) (27) as reference panel.**

- **Number of subjects that have and do not have detectable antibody to L1 of HPV8 or HPV77 in the serum, respectively.**

- **Frequency of each HLA allele in cases and controls, respectively.**

- **Heterogeneity in effect between the imputation by HLA*IMP:02 and SNP2HLA was evaluated using Cochran’s Q-test statistic.**

- **Rsq, a measure of imputation quality generated by SNP2HLA, which estimates the squared correlation between imputed and true genotypes.**
epitope with a monoclonal antibody varied less than 2-fold. Thus, given the similar properties of the antigens the use of a uniform cut-off appears justified. To avoid false-positivity by low-level cross-reactivity and to increase type specificity of the seroprevalence values, a uniform cut-off of 200 MFI well above background levels was chosen for cutaneous HPV L1 antibodies, as described previously (31). A bridging panel of approximately 200 sera tested in these previous studies was re-tested side-by-side with the CE serum samples, allowing normalization of the present data to the predefined cut-off values.

**Statistical analysis for genotyped SNPs**

For single SNP analysis, we considered the MHC to be defined by a 4.5-Mb region bordered by the Ring finger protein (RFP) and motif (MLN) genes (rs209130 at 28 867 800 bp and rs1547668 at 33 775 446 bp, respectively) at the telomeric and centromeric ends of 6p21, respectively, which include 1092 genotyped SNPs in the CE study. All positions are with respect to build 104.0 of the Single Nucleotide Polymorphism database (dbSNP) and GRCh37.p10/hg19 assembly of the Human Genome. The association between the integration of classical HLA alleles and statistical analysis was performed for HLA-A (n = 2474 reference samples), HLA-B (n = 3090), HLA-C (n = 2022), HLA-DQα1 (n = 175), HLA-DQB1 (n = 2629), HLA-DRB1 (n = 2665), HLA-DRB1 (n = 74), HLA-DRB3 (n = 282), HLA-DRB4 (n = 282) and HLA-DRB5 (n = 282). For each person and locus, the imputation methodology of HLA*IMP:02 returns probabilities associated with the sample having each genotype, these probabilities have been shown to be reasonably well calibrated (21). Imputation accuracy was assessed through a cross-validation analysis of the training data. Training was performed on two-thirds of the reference panel and tested on the remaining third at the SNPs chosen for imputation in the current study. Thresholding calls at a posterior probability of 0.7 provided accuracy of between 0.96 (HLA-DRB1) and 0.99 (HLA-DQA1, HLA-DRB4 and HLA-DRB5). To ensure the accuracy of prediction, we restricted our case–control analysis to the subjects who carry the specified HLA alleles with each genotype having a posterior probability over 0.7. Relaxing the posterior probability criterion made no substantial difference to the study findings (data not shown).

The association between each HLA allele and HPV seropositivity was estimated by OR and 95% CI per allele using unconditional logistic regression assuming a log-additive model with adjustment for age, sex, country and cancer type (control, lung cancer, head and neck cancer and kidney cancer) when appropriate. To validate the imputation results, HLA alleles that were found to be significantly associated with seropositivity to any HPV types were also imputed by SN2PHLA using the T1DGC as the reference panel which imputed 5225 unrelated individuals (10 450 haplotypes) (26). Unconditional logistic regression using posterior genotype probabilities (allele dosages) for each HLA allele from SN2PHLA were carried out using PLINK (38) with adjustment for age, sex, country and cancer type (control, lung cancer, head and neck cancer and kidney cancer) when appropriate. Unless specified, all analyses were conducted using SAS 9.3 software. All statistical tests were two-sided.

**Supplementary Material**

Supplementary Material is available at HMG online.

**Acknowledgements**

The authors gratefully acknowledge all the staff in the collaborating centers and participants in this research.

**Conflict of Interest statement.** None declared.

**Funding**

This work was supported by European Commission (grant no. IC18-CT97-0222).

**References**


