HMG Advance Access published September 6, 2011

Genome-wide association study of HPV seropositivity

Dan Chen[1], James D. McKay[1], Gary Clifford[1], Valérie Gaborieau[1], Amélie Chabrier[1], Tim Waterboer[2], David Zaridze[3], Jolanta Lissowska[4], Peter Rudnai[5], Eleonora Fabianova[6], Vladimir Bencko[7], Vladimir Janout[8], Lenka Foretova[9], Ioan Nicolae Mates[10], Neonila Szeszenia-Dabrowska[11], Maria Paula Curado[12], Sergio Koifman[13], Ana Menezes[14], Victor Wünsch-Filho[15], José Eluf-Neto[15], Leticia Fernández Garrote[16], Elena Matos[17], Diana Zelenika[18], Anne Boland[18], Paolo Boffetta[19,20], Michael Pawlita[2], Mark Lathrop[18,21], Paul Brennan[1*]

1. International Agency for Research on Cancer, Lyon, France.
2. Infection and Cancer Program, German Cancer Research Center (DKFZ), Heidelberg, Germany.
3. Department of Epidemiology and Prevention, N.N. Blokhin Cancer Research Center, Moscow, Russia.
4. Department of Cancer Epidemiology and Prevention. M. Sklodowska-Curie Cancer Center and Institute of Oncology, Warsaw, Poland
5. National Institute of Environmental Health, Budapest 1097, Hungary.
6. Regional Institute of Public Health, Banská Bystrica, Slovakia.
7. Institute of Hygiene and Epidemiology, First Faculty of Medicine and General University Hospital, Charles University in Prague, Czech Republic.
8. Palacky University, Olomouc, Czech Republic.
9. Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic.
10. University of Medicine and Pharmacy "Carol Davila", Bucharest, Romania.
11. Department of Epidemiology, Institute of Occupational Medicine, Lodz, Poland.
12. International Prevention Research Institute (IPRI), France
13. National School of Public Health/FIOCRUZ, Rio de Janeiro, Brazil.
14. Universidade Federal de Pelotas, Pelotas, Brazil.
15. Universidade de São Paulo, São Paulo, Brazil.
16. Institute of Oncology and Radiobiology, Havana, Cuba.
17. Angel H. Roffo Oncology Institute, University of Buenos Aires, Buenos Aires, Argentina.
18. Centre National de Génotypage, Institut Génomique, Commissariat à l'énergie Atomique, Evry, France.
19. The Tisch Cancer Institute Mount Sinai School of Medicine, New York, NY, USA.
20. International Prevention Research Institute, Lyon, France.
21. Fondation Jean Dausset-CEPH, Paris, France.

* To whom correspondence and reprints should be addressed: Paul Brennan. International Agency for Research on Cancer (IARC), 150 Cours Albert Thomas, 69372 Lyon CEDEX 08, France. Telephone: +33 472738533; Fax: +33 472738342; Email: gep@iarc.fr.

Abstract

High-risk α mucosal types of human papillomavirus (HPV) cause anogenital and oropharyngeal cancers, whereas β cutaneous HPV types (e.g. HPV8) have been implicated in non-melanoma skin cancer. Although antibodies against the capsid protein L1 of HPV are considered as markers of cumulative exposure, not all infected persons seroconvert. To identify common genetic variants that influence HPV seroconversion, we performed a two-stage genome-wide association study (GWAS). Genome-wide genotyping of 316,015 single nucleotide polymorphisms was carried out using the Illumina HumanHap300 BeadChip in 4,811 subjects from a Central European case-control study of lung, head and neck and kidney cancer that had serology data available on 13 HPV types. Only one association met GWA significance criteria, namely that between HPV8 seropositivity and rs9357152 (odds ratio (OR) =1.37, 95% confidence interval (CI) =1.24-1.50 for the minor allele G; $P=1.2\times10^{-10}$), a common genetic variant (minor allele frequency (MAF) =0.33) located within the major histocompatibility complex (MHC) II region at 6p21.32. This association was subsequently replicated in an independent set of 2,344 subjects from a Latin American case-control study of head and neck cancer (OR=1.35, 95%CI=1.18-1.56, $P=2.2\times10^{-5}$), yielding $P=1.3\times10^{-14}$ in the combined analysis (P-heterogeneity=0.89). No heterogeneity was noted by cancer status (controls/lung cancer cases/head and neck cancer cases/kidney cancer cases). This study provides a proof of principle that genetic variation plays a role in antibody reactivity to HPV infection.

Introduction

Human papillomaviruses (HPV) are a large and highly diverse group of DNA viruses that infect either mucosal or cutaneous epithelia. More than 100 known HPV types are phylogenetically classified into genera (sharing less than 60% nucleotide identity), species (60%-70%) and types (71%-89%) (1). Mucosal HPVs belong exclusively to the genus alpha (α) and can be divided into low-risk (LR) types (e.g. HPV6 and HPV11), that have the potential to induce warts and low-grade squamous intraepithelial lesions, and high-risk (HR) types (e.g. HPV16 and 18), that are causally involved in cervical, other anogenital and oropharyngeal cancers (2-4).

Cutaneous HPVs display a much greater heterogeneity and are found among five genera: alpha (α), beta (β), gamma (γ), mu (μ) and nu (ν). Cutaneous HPV types from genus α (e.g. HPV2 and 3), γ (e.g. HPV4), μ (e.g. HPV1) and ν (HPV41) are associated with benign warts (5-7). β cutaneous HPVs, however, have been implicated in non-melanoma skin cancer (NMSC), namely squamous cell carcinoma (SCC) and less frequently basal cell carcinoma in both immunosuppressed organ-transplant recipients (OTR) and immunocompetent individuals (8-11), even though this relationship has not yet been established as causal (4). The first evidence for this association was reported in patients with the rare genetic disorder epidemodysplasia verruciformis (EV), who suffer from defective cell-mediated immunity predisposing them to high-viral-load infection with β cutaneous HPVs (predominantly HPV5 and HPV8), and who frequently develop cutaneous SCC (12).

Mucosal and cutaneous HPVs share a similar genetic organization. In particular, the L1 open reading frame (ORF) encodes the major structural protein, which composes the viral capsid, and the two oncogenes E6 and E7 encode proteins that modulate the transformation process (13). However, the exclusively intraepithelial infectious cycle of HPVs, as well as the production of virus particles in the superficial layer of the epithelium only, gives rise to weak antigen presentation to the immune system, which in turn results in a weak production of antibody (14-16). The detection of antibodies against mucosal HPV L1 is associated with high viral load and persistent replication, and only about half of persons with active genital HPV infection seroconvert (15-18). Antibodies against the E6/E7 proteins of

mucosal HPVs, on the other hand, are markers of high-grade or malignant HPV related disease. Much less is known about the natural history and immune response for cutaneous HPVs.

Despite the causal role of HPVs in benign or malignant epithelial lesions and the increasing application of HPV serology as epidemiological tool, the genetic basis of serological immune response to HPV infections has never been studied. To identify common genetic variants that influence seroconversion to HPV L1 capsid protein, we performed a two-stage genome-wide association study (GWAS) for 13 HPV types representing different genera.

Results

Genome-wide association results

Subjects included in the discovery phase were from a Central European multi-center case-control study (CE) of lung, head and neck and kidney cancer that had serology data available on HPV types, as described in the Materials and Methods. Genome-wide genotyping was successfully performed in 1,413 lung cancer cases, 742 head and neck cancer cases, 853 kidney cancer cases and 2,191 cancer-free controls. Following stringent quality-control steps (see Materials and Methods for details), genome-wide statistical analysis was performed among 4,811 subjects (1,286 lung cancer cases, 679 head and neck cancer cases, 811 kidney cancer cases and 2,035 cancer-free controls) for associations between 316,015 single nucleotide polymorphisms (SNPs) and L1 seropositivity for the 13 HPV types with 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). Antibody seroprevalence and demographic characteristics of included subjects are presented in Table 1 and Supplementary Table 2, respectively. Highest seroprevalence was observed for β cutaneous HPV8 (28.04%), followed by α mucosal HPV6 (24.78%) and γ cutaneous HPV4 (24.03%). No serotype showed any evidence for systematic bias in the genome-wide association analysis (each genomic inflation factor $\lambda<1.02$).

Only one clear genome-wide significant association was identified (threshold for significance $P<1.2\times10^{-8}$), namely that between HPV8 seropositivity and rs9357152, a common genetic variant (minor allele frequency (MAF)=0.33) located on chromosome 6p21.32 in the major histocompatibility complex (MHC) class II region (odds ratio (OR)=1.37, 95% confidence interval (CI)=1.24-1.50 for the minor

allele G; $P=1.2\times10^{-10}$) (Figure 1, Figure 2, Table 2). This region contains genes that encode highly polymorphic human leukocyte antigen (HLA) class II molecules. Conditioning on rs9271366, a proxy of rs3135388 ($D'=1$, $r^2=0.97$ with rs3135388 in Hapmap CEU) which tagged *HLA* II haplotype *DRB1*1501-DQB1*0602* ($D'=1$, $r^2=0.97$ in CEU) (19), did not alter the association with rs9357152 (OR=1.36, 95%CI=1.23-1.50, $P=1.2\times10^{-9}$).

A borderline genome-wide significant association was observed between HPV31 seropositivity and rs9401090, a genetic variant located on chromosome 6q22.3 (OR=1.69, 95%CI=1.39-2.06 for the minor allele T; $P=2.1\times10^{-7}$) (Table 2, Supplementary Figure 1).

Replication and combined results

The association between rs9357152 and HPV8 seropositivity was independently replicated in the Latin American (LA) study (OR=1.35, 95%CI=1.18-1.56, $P=2.2\times10^{-5}$) with no evidence of heterogeneity between studies (P-heterogeneity=0.89). By contrast, no association (OR=0.87, 95%CI=0.66-1.16, $P=0.34$) was observed between rs9401090 and HPV31 seropositivity in the replication series (P-heterogeneity=2.0×10^{-4}) (Table 2).

After pooling the discovery and replication data, the ORs (95%CI) were 1.36 (1.22-1.52) and 1.86 (1.56-2.21) for heterozygous (genotype AG) and homozygous carriers (genotype GG) of rs9357152 minor allele G, respectively, yielding $P=1.3\times10^{-14}$ for trend in the combined analysis (Figure 3). No heterogeneity for the association of rs9357152 with HPV8 seropositivity was noted by cancer status (controls/lung cancer cases/head and neck cancer cases/kidney cancer cases), gender, age or smoking status. This variant allele G was also found to be positively associated with seropositivity for other β cutaneous types (HPV38 or 49) (OR=1.20, 95%CI=1.10-1.30, $P=1.9\times10^{-5}$) and γ cutaneous HPV4 (OR=1.16, 95%CI=1.07-1.26, $P=6.0\times10^{-4}$) as well as LR mucosal HPV types (HPV6 or 11) (OR=1.12, 95%CI=1.03-1.21, $P=5.2\times10^{-3}$) but not with seropositivity for α cutaneous HPV77, μ cutaneous HPV1 or HR mucosal HPV types (HPV16, 18, 31, 33, 35, 45, 52 or 58) (Figure 4). However, none of these associations remained statistically significant after adjusting for HPV8 seropositivity.

We also observed an inverse association between rs9357152 G allele and mucosal HPV E7 seropositivity (OR=0.83, 95%CI=0.74-0.94, $P=3.1\times10^{-3}$), which was not altered by adjustment for HPV8 seropositivity (data not shown), but found no association with mucosal HPV E6 seropositivity.

Imputation

When conditioning on rs9357152, residual association with HPV8 seropositivity was detected at some SNPs at 6p21.32, with the most significant signal occurring at rs3135363 ($P=3.0\times10^{-4}$) compared to $P=5.2\times10^{-7}$ in the unconditional analysis. Further conditional analysis on rs3135363 still left a residual association within this region (e.g. $P=1.5\times10^{-4}$ for rs1042337). Imputation of the genotypes for all SNPs in the entire MHC region using the 1000 Genomes data (20) (see Materials and Methods for details), identified a group of highly correlated SNPs rs115639952, chr6:32390621 and rs17202724 ($D'=1$, $r^2=1.0$ with each other) located 267~274 kb telomeric of rs9357152, with a stronger effect (OR=1.65, 95%CI=1.47-1.86 for rs115639952 minor allele T; $P=3.3\times10^{-16}$) (Figure 2). Conditional analysis on rs115639952 resulted in attenuation of the association with majority of the SNPs at this locus, although not completely (e.g. OR=1.15, 95%CI=1.03-1.29 for rs9357152 minor allele G; $P=0.02$).

The SNP rs115639952 was unable to be designed for Taqman assay. Nevertheless, rs114427648, a strong proxy of rs115639952 ($D'=1$, $r^2=1$ in the June 2010 release of the 1000 Genomes CEU data) was successfully genotyped in the CE and LA studies for validation. Concordance between imputed and direct genotyping data was greater than 95.7%. Comparable associations were noted between HPV8 seropositivity and this variant using imputed genotypes compared with direct genotyping in the discovery phase (CE in the 4,432 overlapping individuals: Direct genotyping OR=1.59, 95%CI=1.40-1.82, $P=5.5\times10^{-12}$ *vs* imputed genotypes OR=1.62, 95%CI=1.43-1.83, $P=6.6\times10^{-14}$ for the minor allele G). Similarly, a comparable association was noted within replication series (LA: OR=1.51, 95%CI=1.21-1.88 for the minor allele G; $P=3.0\times10^{-4}$) (Table 2).

Discussion

We have identified and replicated a genome-wide significant association between seropositivity against HPV8 L1 capsid protein and a common genetic variant rs9357152 in MHC II region at 6p21.32.

This association was subsequently replicated in an independent series from Latin America. This locus is located 5'-upstream of *HLA-DQB1* which encodes β chain of HLA class II molecule. To our knowledge, this study is the first to implicate role of common genetic variation in serological immune response to HPV infection.

Despite the evidence suggesting a role for β cutaneous HPVs (including HPV8) in NMSC, much less is known about their natural history and immune response in comparison to what is known for mucosal HPVs. β cutaneous HPV infection commences in early infancy (21), becoming ubiquitous by adulthood (22,23), when infections frequently persist (24, 25). Seroprevalence, on the other hand, appears to accumulate continuously with age (26), and is much lower than the prevalence of HPVs on the skin in the same population (often measured as HPV DNA in plucked eyebrow hairs) (9). It has been suggested that the replication of β cutaneous HPVs might be largely controlled in a latent cycle of infection, resulting in relatively little antigen presentation to the immune system (1,8,22), but that insufficient control of HPV infection may lead to higher viral loads, more antigen presentation and hence increased antibody production (8,26,27). Indeed, β HPV seroprevalence, mainly HPV5 and 8, is elevated in EV and immunosuppressed patients, as well as those with dermatological diseases or second degree burns (8,28-30). Given this evidence, β cutaneous HPV seropositivity in the adult population of present study may reflect a combination of cumulative exposure to HPV and subsequent capability of the host to mount an effective immune response (11,26,31,32).

The most significant association with HPV8 L1 seropositivity was identified for the genetic variants in the MHC II region containing the highly polymorphic *HLA* class II loci. As genetic polymorphisms are unlikely to be related to exposure to β cutaneous HPVs *per se* (i.e. the propensity to come into contact with these ubiquitous viruses), they must somehow influence the immune response, although neither the mechanism, nor nature of the causative allele(s), is clear. The expression of HLA class II molecules is restricted to specialized antigen-presenting cells (APC) which present antigenic peptides to $CD4^+$ T cells. The interaction between the HLA-peptide complex and the T cell receptor is an essential step in T cell activation and accumulating evidence has demonstrated that $CD4^+$ T cell regulated

cell-mediated immune response plays a critical role in the resolution and control of HPV infections (8,18,33).

The MHC region is one of the most gene-dense and extremely polymorphic regions. The identification of causal variant (s) is hindered by the complex pattern of LD that extends across multiple HLA and non-HLA genes in the MHC. Several studies in different populations have described HLA class II polymorphisms associated with susceptibility to acquisition and persistence of genital α HPV infection, suggesting that subjects with certain haplotypes are less or more likely to clear the viral infection. Among the most prominent is HLA II haplotype *DRB1*1501-DQB1*0602,* which has been consistently associated with risk of cumulative infection (34), persistency (35) and higher viral load (36) of HPV16. On the other hand, protection against HPV transient and persistent infection has been reported for *DRB1*0301-DQB1*0201* haplotype (37). Furthermore, *DRB1*13* was found to be associated with a higher risk of HPV infections, whereas *DQB1*03* confers a lower risk of HPV infection, even though the findings for them were not always consistent across different studies (38). Unfortunately, as classical *HLA* typing data was not available in our study, we were not able to directly assess the effect of different *HLA* class II alleles systematically and their correlation with the association signals. As found in Utah residents with European ancestry from the Centre d'Etude du Polymorphisme Humain collection (CEU), *DRB1*1501-DQB1*0602* is tagged by rs3135388 (D'=1, r^2=0.97) (19). In our study, a very modest association was found between rs9271366, a strong proxy for rs3135388 (D'=1, r^2=0.97 in Hapmap CEU), and HPV8 seropositivity (OR=0.86, 95%CI=0.75-0.99 for the minor allele G; P=0.037). However, this association was attenuated after adjusting for rs9357152 (OR=0.97, 95%CI=0.84-1.11, P=0.63). Within the constraints of adjustment using SNP rs9271366 as a proxy for carrying HLA haplotype *DRB1*1501-DQB1*0602*, as the association for HPV8 seropositivity with rs9357152 is little changed, it seems unlikely that this *HLA* haplotype can account for the association in our study. Ideally overlapping HLA typing and genotyping data in the same series is needed to tease out the role of HLA alleles and the association we have noted with rs9357152.

Significant association was also observed for rs9357152 with L1 seropositivity for other β cutaneous types, γ cutaneous HPV4 and LR mucosal HPV types. However, these associations were

almost completely attenuated after adjusting for HPV8 L1 serological status (data not shown), reflecting the correlation between seropositivity for HPV8 and for these other types. Multiple infections of β and γ cutaneous types have been reported previously, measured by PCR or the same multiplex assay (26,22-25,39). Multiple seropositivity may be explained in part by a general susceptibility to a wide range of HPV types and/or their shared routes of exposure. Alternatively, this correlation could be due to cross-reactivity between antibodies in the multiplex assay. However, a validation of the assay showed that for HPV8 there was very little correlation between double seropositivity and L1 amino acid sequence relatedness, suggesting the measured antibodies are mainly type-specific (26). Intriguingly, we also observed an inverse association of rs9357152 with mucosal HPV E7 seropositivity. However, it is worth emphasizing that given the small sample size of mucosal HPV E7 seropositive subjects and multiple tests performed, one should exercise caution in interpreting this result. Speculatively, the rs9357152 variant allele which is associated with higher antibody reactivity to HPV infections at early stage may confer a lower risk for progression to high-grade HPV related disease. Nevertheless, such finding needs to be further replicated in other studies before drawing any conclusion.

Imputation across different populations can be problematic, particularly in the context of the complex LD structure found within the MHC region. The countries participating in the Central European study are distinguishable from imputation series samples (40), making the 1000 Genomes series sub-optimal as a reference for imputation. Furthermore, the LD structure within the MHC region may differ between the discovery (Central Europe) and replication (Latin America) series. Nevertheless, imputation and subsequent validation by direct genotyping did identify a group of highly correlated variants with more important effects, and this association does appear consistent in both the discovery and replication series. Mutual adjustment seems to indicate that the majority of the association noted with rs9357152 may be accounted for by this group of correlated variants. Some residual association was noted with rs9357152 after adjustment. This could result from the play of chance or alternately indicate multiple associations or that both variants are indirectly tagging a third unknown variant. Further study of these variants, and particularly how they are correlated with *HLA* alleles as well as how they differ across populations, is needed to determine the causative alleles.

The fact that genome-wide significant findings were detected for HPV8 than for other types may be in part related to its higher population-wide seroprevalence, meaning that our study was maximally powered to detect variants associated with this HPV type. The power of our study may be more limited to detect variants associated with HPV types with more modest seroprevalences. Adequately powered studies with longitudinal design would help to elucidate the genetic basis of the immune response for less frequent, particularly HR mucosal HPV types. Furthermore, given the association of HPV8 seroreactivity with NMSC, the exploration of the association between 6p21.32 variants and NMSC is likely to yield further insights into the etiology of HPV8-related malignancy.

In conclusion, we have identified and replicated a genome-wide significant association between the 6p21.32 locus and HPV8 L1 seropositivity. This study provides a proof of principle that genetic variation plays a role in antibody reactivity to HPV infections.

Materials and Methods

Study population

Subjects included in the discovery phase were from 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. DNA of sufficient quality and quantity for genome-wide genotyping was available for 1,413 lung cancer cases, 742 head and neck cancer cases, 853 kidney cancer cases and 2,191 cancer-free controls.

The replication series consisted of 2,344 subjects (1,307 head and neck cancer cases and 1,037 cancer-free controls) from a Latin American multi-center head and neck cancer study who were recruited in Argentina, Cuba, and Brazil from 1998 to 2003. The structured questionnaire was similar to the CE study. Both studies were coordinated by the International Agency for Research on Cancer (IARC) using similar protocols across all centers. Details on study designs and subjects recruitment, have been described previously (41,42). 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 each study was approved by an ethical committee.

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) as well as early proteins E6 (α mucosal: 6, 11, 16, 18, 31, 33, 35, 45, 52 and 58) and E7 (α mucosal: 6, 11, 16, 18, 31, 33 and 35), 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 (43,44). Briefly, this high throughput method employs full length viral proteins fused with an N-terminal GST domain and a C-terminal peptide (tag) expressed in bacteria as antigens which allows the simultaneous determination of antibodies to a large number of HPV types in one 2-μl serum sample. Glutathione cross linked to casein was coupled to fluorescence-labeled polystyrene beads (Luminex, Austin, TX), and GST fusion proteins were affinity-purified on the beads directly in a one-step procedure. The differently labeled bead sets carrying different antigens were mixed and incubated with human sera. Beads with fused GST and tag without intervening viral antigen served for individual serum background determination. Antibodies bound to the beads via the viral antigens were then stained with biotinylated anti-human immunoglobulin and fluorescent reporter conjugate streptavidin-R-phycoerythrin. A Luminex analyser was used to identify the internal colour of the individual beads and thus the antigens carried by the beads. Antibodies bound to antigens on beads were quantified via the reporter fluorescence expressed as median R-phycoerythrin fluorescence intensity (MFI) of at least 100 beads of the same internal colour after subtraction of background reactivity. More information on quality control and data processing has been described elsewhere (26).

MFI values were dichotomized as antibody positive or negative. Seropositivity cut-offs for cutaneous HPV L1 antibodies were those defined previously (26). 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 (45). Seropositivity cut-offs for E6 and E7 antibodies were defined as previously (46). A bridging panel of approximately 200 sera

tested in these previous studies was re-tested side-by-side with the CE and LA serum samples, allowing normalization of the present data to the predefined cut-off values.

Genome-wide genotyping and quality control

The CE study was genotyped using the Illumina's Sentrix HumanHap300 BeadChip 317K at the Centre National de Genotypage (CNG Paris, France) as described previously (41).

We conducted systematic quality control steps on the raw Illumina HumanHap300 genotyping data (Supplementary Table 1). The exclusion criteria for SNPs were: minor allele frequency (MAF) ≤ 1%; genotyping call rate <95%; genotype frequency deviated from the Hardy-Weinberg Equilibrium (HWE) among cancer-free controls ($P < 1\times10^{-7}$); resulting in data on 316,015 SNPs. Individuals with the overall genotype completion rate lower than 95% were excluded. We also conducted further exclusions for those who showed sex discrepancies based on the heterozygosity rate from chromosome X. Unexpected duplicates and first-degree relatives were removed based on identity-by-state (IBS) estimates calculated in PLINK (47). Those genotyped were restricted to individuals of self–reported European ethnicity. To further increase the ethnic homogeneity of the series, we used the program STRUCTURE (48) to identify individuals of mixed ethnicity. Using a subseries of 12,898 genetic variants from the HumanHap 300 BeadChip panel evenly distributed across the genome and in low Linkage Disequilibrium (LD) ($r^2<0.004$) (49), we estimated the genetic profile of the study participants compared with individuals of known ethnic origins (the Caucasian, African and east-Asian individuals genotyped by the HapMap project). Nine individuals were excluded because of evidence of ethnic admixture. A total of 4,811 subjects were included in the final analysis.

Genome-wide statistical analysis

The association between each genetic variant and type-specific HPV seropositivity was estimated by the OR per allele and 95% CI using multivariate unconditional logistic regression assuming a log-additive genetic model with age, sex, country and cancer case/control status included in the model as covariates. The corresponding estimated genome-wide significance level after correction for testing 316,015 SNPs and 13 phenotypes is $P = 1.2\times 10^{-8}$. All analyses were conducted using PLINK (47).

The potential for population stratification not accounted for by adjustment for country was also investigated by principal components analysis (PCA) undertaken with the EIGENSTRAT package (50) using 12,898 markers in low LD (49). None of the eigenvectors was significantly associated with serological status of any HPV type hence not included in the final model.

Replication genotyping

To validate the findings from the whole-genome scan, genetic variants of rs9357152 and rs9401090 were genotyped for replication in an independent set of 2,344 participants with HPV serology data from the LA study. Genotyping was performed using the TaqMan genotyping platform at IARC. The information on assay conditions and the primers and probes is available upon request. A common series of 227 standard DNAs were genotyped to ensure the quality and comparability of the genotyping results. Concordance with the consensus genotype and the results produced for the standardized DNAs was above 99.5%. Genotype success rate was greater than 96.2% and genotype distributions were consistent with that expected by HWE for each SNP.

Replication and combined statistical analysis

The association between rs9357152 and HPV8 seropositivity as well as that between rs9401090 and HPV31 seropositivity was estimated by ORs and 95% CIs per allele (under log-additive model) and per genotype derived from multivariate unconditional logistic regression, with age, sex and country included in the model as covariates. Heterogeneity of ORs across the subgroups stratified by cancer status, smoking status, gender and age was assessed using the Cochran's Q test.

We further investigated whether rs9357152 was also associated with L1 seropositivity by phylogenetic groups of HPV types, as well as with E6 and E7 seropositivity for mucosal HPVs, in the pooled analysis of the CE and LA studies. All replication and combined analyses were conducted using SAS 9.1 software. *P* values were two sided.

Imputation and validation

Genotypes for genetic variants in the entire MHC region not genotyped on the Illumina HumanHap300 BeadChip, but identified by the 1000 Genomes Project, were imputed using the program MACH (51) with phased haplotypes in 283 subjects from the August 2010 release of the 1000 Genomes

data as a reference panel (20) which include 90 Utah residents (CEPH) with Northern and Western European ancestry (CEU), 92 Toscani in Italia (TSI), 43 British in England and Scotland (GBR), 36 Finnish in Finland (FIN), 17 Mexican in Los Angeles, CA (MXL) and 5 Puerto Rican in Puerto Rico (PUR). Unconditional logistic regression using posterior genotype probabilities (allele dosages) from MACH were carried out using ProbABEL (52) including age, sex, country of origin, and cancer case/control status in the regression as covariates. Association results for all SNPs with Rsq above 0.3 in the MHC region are shown in Figure 2. LD statistics (*D'* and r^2) were calculated using Haploview (53). To validate the imputed results, the SNP rs114427648 was genotyped in both the discovery (CE) and the replication series (LA) as a proxy for rs115639952 (*D'*=1, r^2=1 in the June 2010 release of the 1000 Genomes CEU data) due to availability of Taqman assays.

Acknowledgement

This work was supported by L'INCA, Fondo para la Investigacion Cientifica y Tecnologica (Argentina); Institut Municipal d'Investigacio Medica (Barcelona); and European Commission (grant no IC18-CT97-0222). We gratefully acknowledge all the staff in the collaborating centers and participants in this research.

Conflict of Interest Statement

None declared.

References

1. de Villiers, E.M., Fauquet, C., Broker, T.R., Bernard, H.U. and zur Hausen, H. (2004) Classification of papillomaviruses. *Virology*, **324,**17-27.
2. Muñoz, N., Bosch, F.X., de Sanjosé, S., Herrero, R., Castellsagué, X., Shah, K.V., Snijders, P.J., Meijer, C.J. and International Agency for Research on Cancer Multicenter Cervical Cancer Study Group. (2003) Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N. Engl. J. Med.,* **348,** 518–527.
3. Narechania, A., Chen, Z., DeSalle, R. and Burk, R.D. (2005) Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. *J. Virol.,* **79,** 15503–15510.
4. Bouvard, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., E.l. Ghissassi, F., Benbrahim-Tallaa, L., Guha, N., Freeman, C., Galichet, L., *et al*. (2009) A review of human carcinogens--Part B: biological agents. *Lancet. Oncol*., **10**, 321-322.
5. Jablonska, S., Majewski, S., Obalek, S. and Orth, G. (1997) Cutaneous warts. *Clin. Dermatol*., **15**, 309–319.
6. Gassenmaier, A., Fuchs, P., Schell, H. and Pfister, H. (1986) Papillomavirus DNA in warts of immunosuppressed renal allograft recipients. *Arch. Dermatol. Res*., **278**, 219–223.
7. Akgul, B., Cooke, J.C. and Storey, A. (2006) HPV-associated skin disease. *J. Pathol*., **208**, 165–175.
8. Nindl, I., Gottschling, M. and Stockfleth, E, (2007) Human papillomaviruses and nonmelanoma skin cancer: basic virology and clinical manifestations. *Dis. Markers.,* **23**, 247–259.
9. Masini. C., Fuchs, P.G., Gabrielli, F., Stark, S., Sera, F., Ploner, M., Melchi, C.F., Primavera, G., Pirchio, G., Picconi, O., *et al*. (2003) Evidence for the association of human papillomavirus infection and cutaneous squamous cell carcinoma in immunocompetent individuals. *Arch. Dermatol*., **139**, 890-894.
10. Karagas, M.R., Waterboer, T., Li, Z., Nelson, H.H., Michael, K.M., Bavinck, J.N., Perry, A.E., Spencer, S.K., Daling, J., *et al*. (2010) Genus beta human papillomaviruses and incidence of basal cell and squamous cell carcinomas of skin: population based case-control study. *B.M.J.*, **341**, c2986.
11. Feltkamp, M.C., Broer, R., di, Summa, F.M., Struijk, L., van der Meijden, E., Verlaan, B.P., Westendorp, R.G., ter, Schegget, J., Spaan, W.J., Bouwes, Bavinck, J.N. (2003) Seroreactivity to epidermodysplasia verruciformis-related human papillomavirus types is associated with nonmelanoma skin cancer. *Cancer. Res*., **63**, 2695-2700.
12. Jablonska, S. and Majewski, S. (1994) Epidermodysplasia verruciformis: immunological and clinical aspects. *Curr. Top. Microbiol. Immunol.,* **186**, 157–175.
13. Munger, K. and Howley, P.M. (2002) Human papillomavirus immortalization and transformation functions. *Virus. Res*., **89**, 213– 228.
14. Stanley, M. (2006) Immune responses to human papillomavirus. *Vaccine*., **24**, S16-22.
15. Wang, S.S., Schiffman, M., Herrero, R., Carreon, J., Hildesheim, A., Rodriguez, A.C., Bratti, M.C., Sherman, M.E., Morales, J., Guillen, D., *et al*. (2004) Determinants of human papillomavirus 16 serological conversion and persistence in a population-based cohort of 10 000 women in Costa Rica. *Br. J. Cancer*., **91**, 1269-1274.
16. Konya, J. and Dillner, J. (2001) Immunity to oncogenic human papillomaviruses. *Adv. Cancer Res*., **82**, 205-238.
17. Coursaget, P. (2003) Serology for human papillomavirus. *Salud. Publica. Mex*., **45**, S361-366.

18. Frazer, I.H. (2009) Interaction of human papillomaviruses with the host immune system: a well evolved relationship. *Virology*., **384**, 410-414.
19. de Bakker, P.I., McVean, G., Sabeti, P.C., Miretti, M.M., Green, T., Marchini, J., Ke, X., Monsuur, A.J., Whittaker, P., Delgado, M., *et al*. (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet*., **38**, 1166-1172.
20. 1000 Genomes Project Consortium., Durbin, R.M., Abecasis, G.R., Altshuler, D.L., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E. and McVean, G.A. (2010) A map of human genome variation from population-scale sequencing. *Nature*., **467**,1061-1073.
21. Antonsson, A., Karanfilovska, S., Lindqvist, P.G. and Hansson, B.G. (2003) General acquisition of human papillomavirus infections of skin occurs in early infancy. *J. Clin. Microbiol*., **41**, 2509-2514.
22. Antonsson, A., Forslund, O., Ekberg, H., Sterner, G. and Hansson, B.G. (2000) The ubiquity and impressive genomic diversity of human skin papillomaviruses suggest a commensalic nature of these viruses. *J. Virol*., **74**, 11636–11641.
23. Kohler, A., Forschner, T., Meyer, T., Ulrich, C., Gottschling, M., Stockfleth, E. and Nindl, I. (2007) Multifocal distribution of cutaneous human papillomavirus types in hairs from different skin areas. *Br. J. Dermatol*., **156**, 1078–1080.
24. Hazard, K., Karlsson, A., Andersson, K., Ekberg, H., Dillner, J. and Forslund, O. (2007) Cutaneous human papillomaviruses persist on healthy skin. *J. Invest. Dermatol*., **127**, 116–119.
25. de, Koning, M.N., Struijk, L., Bavinck, J.N., Kleter, B., ter, Schegget. J., Quint, W.G. and Feltkamp, M.C. (2007) Betapapillomaviruses frequently persist in the skin of healthy individuals. *J. Gen. Virol*., **88**, 1489–1495.
26. Michael, K.M., Waterboer, T., Sehr, P., Rother, A., Reidel, U., Boeing, H., Bravo, I.G., Schlehofer, J., Gärtner, B.C. and Pawlita, M. (2008) Seroprevalence of 34 human papillomavirus types in the German general population. *PLoS. Pathog.*, **4**, e1000091.
27. Steger, G., Olszewsky, M., Stockfleth, E. and Pfister, H. (1990) Prevalence of antibodies to human papillomavirus type 8 in human sera. *J. Virol*., **64**, 4399-4406.
28. Weissenborn, S.J., Nindl, I., Purdie, K., Harwood, C., Proby, C., Breuer, J., Majewski, S., Pfister, H. and Wieland, U. (2005) Human papillomavirus-DNA loads in actinic keratoses exceed those in non-melanoma skin cancers. *J. Inves . Dermatol.,* **125**, 93–97.
29. Favre, M., Majewski, S., Noszczyk, B., Maienfisch, F., Pura, A,. Orth, G. and Jablonska, S. (2000) Antibodies to human papillomavirus type 5 are generated in epidermal repair processes. *J. Invest. Dermatol*., **114**, 403-407.
30. Favre, M., Orth, G., Majewski, S., Baloul, S., Pura, A. and Jablonska, S. (1998) Psoriasis: A possible reservoir for human papillomavirus type 5, the virus associated with skin carcinomas of epidermodysplasia verruciformis. *J. Invest. Dermatol*., **110**, 311-317.
31. Dillner, J. (1999). The serological response to papillomaviruses. *Semin. Cancer. Biol*., **9**, 423–430.
32. Waterboer, T., Neale, R., Michael, K.M., Sehr, P., de, Koning, M.N., Weissenborn, S.J., Sampogna, F., Abeni, D., Green, A.C., Bouwes, Bavinck, J.N. and Pawlita, M. (2009) Antibody responses to 26 skin human papillomavirus types in the Netherlands, Italy and Australia. *J. Gen. Virol.,* **90**, 1986-1998.

33. Stanley, M.A. (2009) Immune responses to human papilloma viruses. *Indian. J. Med. Res*., **130**, 266-276.
34. Mahmud, S.M., Robinson, K., Richardson, H., Tellier, P.P., Ferenczy, A.S., Roger, M., Coutlee, F. and Franco, E.L. (2007) HLA polymorphisms and cervical human Papillomavirus infection in a cohort of Montreal University students. *J. Infect. Dis*., **196**, 82-90.
35. Beskow, A.H., Josefsson, A.M. and Gyllensten, U.B. (2001) HLA class II alleles associated with infection by HPV16 in cervical cancer in situ. *Int. J. Cancer*., **93**, 817-822.
36. Beskow, A.H. and Gyllensten, U.B. (2002) Host genetic control of HPV 16 titer in carcinoma in situ of the cervix uteri. *Int. J. Cancer*., **101**, 526-531.
37. Maciag, P.C., Schlecht, N.F., Souza, P.S., Rohan, T.E., Franco, E.L. and Villa, L.L. (2002) Polymorphisms of the human leukocyte antigen DRB1 and DQB1 genes and the natural history of human papillomavirus infection. *J. Infect. Dis*., **186**,164-172.
38. de Araujo Souza, P.S., Sichero, L. and Maciag, P.C. (2009) HPV variants and HLA polymorphisms: the role of variability on the risk of cervical cancer. *Future. Oncol*., 5, 359-370.
39. Iannacone, M.R., Michael, K.M., Giuliano, A.R., Waterboer, T., Pawlita, M. and Rollison, D.E. (2010) Risk factors for cutaneous human papillomavirus seroreactivity among patients undergoing skin cancer screening in Florida. *J. Infect. Dis.*, **201**, 760-769.
40. Heath, S.C., Gut, I.G., Brennan, P., McKay, J.D., Bencko, V., Fabianova, E., Foretova, L., Georges, M., Janout, V., Kabesch, M., *et al*. (2008) Investigation of the fine structure of European populations with applications to disease association studies. *Eur. J. Hum. Genet*., **16**,1413-1429.
41. Hung, R.J., McKay, J.D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D,, Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., *et al*. (2008) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*., **452**, 633–637.
42. Hashibe, M., McKay, J.D., Curado, M.P., Oliveira, J.C., Koifman, S., Koifman, R., Zaridze, D., Shangina, O., Wünsch-Filho, V., Eluf-Neto, J., *et al*. (2008) Multiple ADH genes are associated with upper aerodigestive cancers. *Nat. Genet*., **40**, 707-709.
43. Waterboer, T., Sehr, P., Michael, K.M., Franceschi, S., Nieland, J.D., Joos, T.O., Templin, M.F. and Pawlita, M. (2005) Multiplex human papillomavirus serology based on in situ-purified glutathione s-transferase fusion proteins. *Clin. Chem.*, **51**, 1845-1853.
44. Waterboer, T., Sehr, P. and Pawlita, M. (2006) Suppression of non-specific binding in serological Luminex assays. *J. Immunol. Methods*., **309**, 200-204.
45. Clifford, G.M., Shin, H.R., Oh, J.K., Waterboer, T., Ju, Y.H., Vaccarella, S., Quint, W., Pawlita, M. and Franceschi S. (2007) Serologic response to oncogenic human papillomavirus types in male and female university students in Busan, South Korea. *Cancer. Epidemiol. Biomarkers. Prev*., **16**, 1874-1879.
46. Ribeiro, K.B., Levi, J.E., Pawlita, M., Koifman, S., Matos, E., Eluf-Neto, J., Wunsch-Filho, V., Curado, M.P., Shangina, O., Zaridze, D., *et al*. (2011) Low human papillomavirus prevalence in head and neck cancer: results from two large case-control studies in high-incidence regions. *Int. J. Epidemiol*., **40**, 489-502.
47. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A. Bender, D., Maller, J., Sklar, P., de, Bakker, P.I., Daly, M.J, *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet*., **81**, 559–575.

48. Falush, D., Stephens, M. and Pritchard, J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*., **164**, 1567–1587.
49. Yu, K., Wang, Z., Li, Q., Wacholder, S., Hunter, D.J. Hoover, R.N., Chanock, S. and Thomas, G. (2008) Population substructure and control selection in genome-wide association studies. *PloS. ONE*., **3**, 2551.
50. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet*., **38**, 904–909.
51. http://wwwsphumichedu/csg/abecasis/mach/indexhtml.
52. Aulchenko, Y.S., Struchalin, M.V. and van Duijn, C.M. (2010) ProbABEL package for genome-wide association analysis of imputed data. *B.M.C. Bioinformatics*., **11**,134.
53. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*., **21**, 263-265.
54. Traherne, J.A. (2008) Human MHC architecture and evolution: implications for disease association studies. *Int. J. Immunogenet*., **35**,179-192.

Figure legends

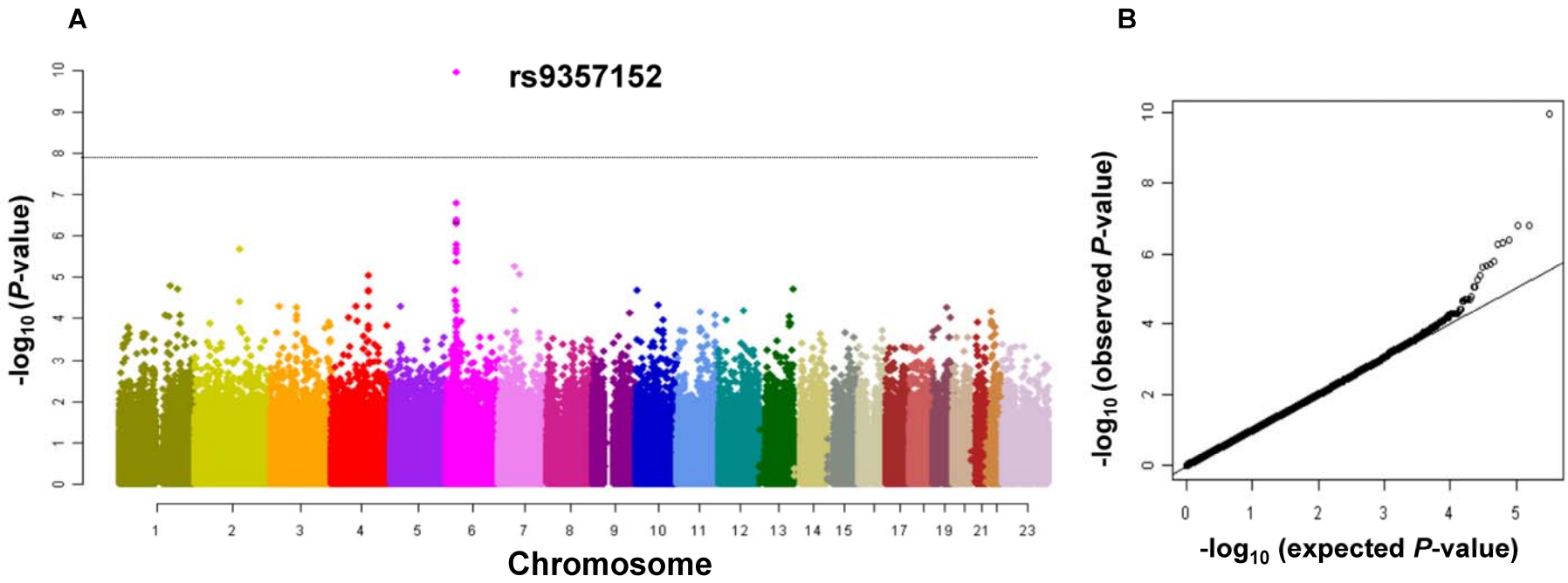
Figure 1. Genome-wide association results for HPV8 seropositivity. (**A**) Manhattan plot of *P*-values. (**B**) Quantile-quantile plot for *P*-values in -$\log_{10}$ scale.

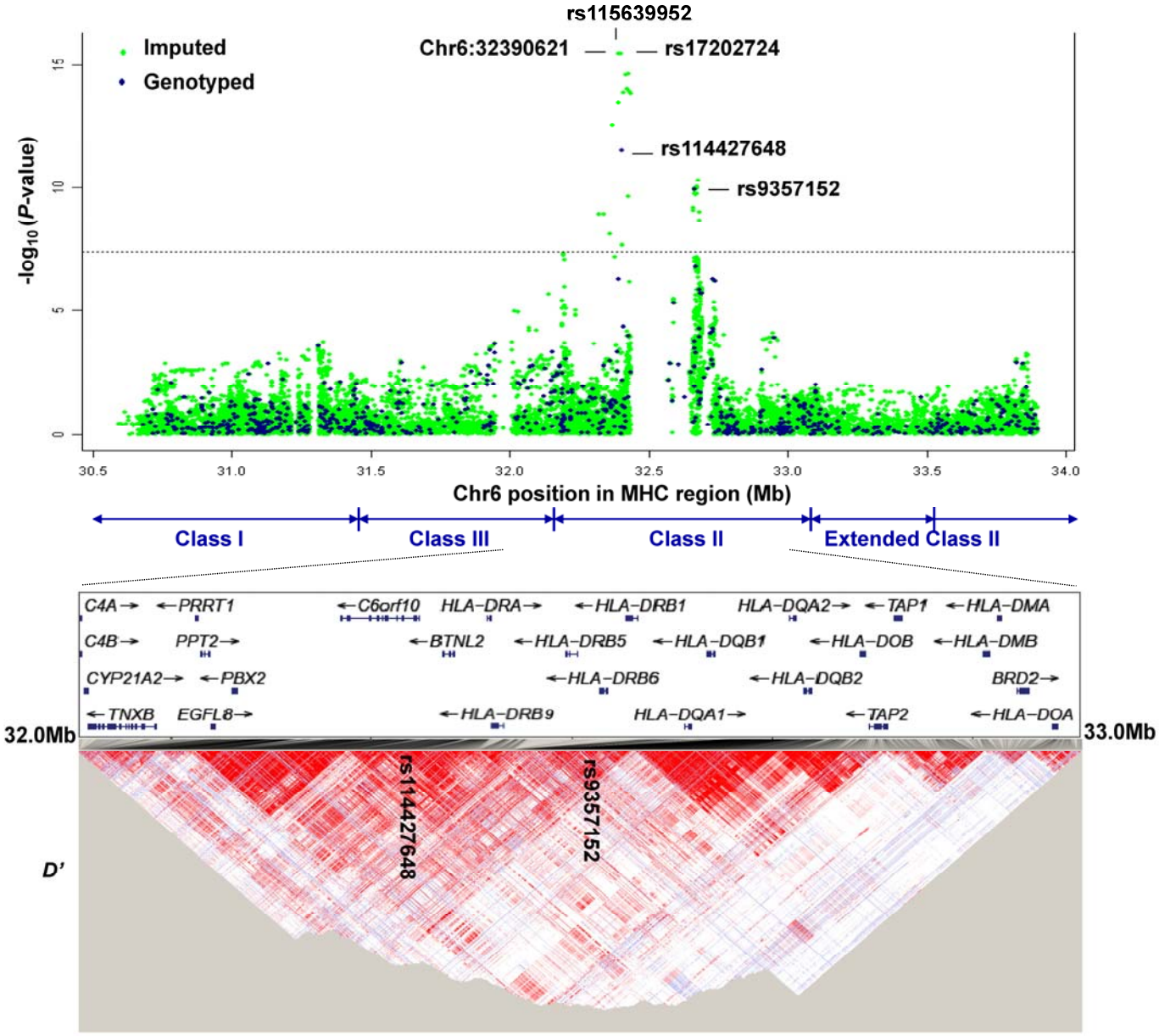
Figure 2. Regional association plot for HPV8 seropositivity in MHC region. Single marker association results for directly genotyped (blue) and imputed variants (green) with Rsq (a measure of genotype certainty) above 0.3. The gap in the MHC class II region corresponds to the hypervariable DRB region with gene copy number variations (CNVs) between *HLA-DRB1* (centromeric) and the pseudogene *HLA-DRB9* (telomeric) (54) which is underrepresented on the HumanHap300 Genotyping BeadChip 317K. Imputation was performed in 4,811 subjects based on the August 2010 release of 1000 Genomes data. rs114427648 was genotyped as a proxy of rs115639952 (D'=1, r^2=1) to validate the imputed results. *P*-values in -$\log_{10}$ scale are plotted against SNPs by their positions on chromosome 6 (Build 37.1). Genes in the region are represented with arrow heads indicating the direction of transcription. Lower panel: pair-wise LD estimates, increasing intensities of red indicate higher *D'* statistics. Blue colour indicates that the pair-wise comparison is not informative.

Figure 3. Association between rs9357152 and HPV 8 seropositivity in the combined series. HN: head and neck. Unless specified, the ORs, and 95% CIs were derived from the log-additive genetic model with adjustment for age, country and sex when appropriate. *P* for heterogeneity was derived from the Cochran's Q test.

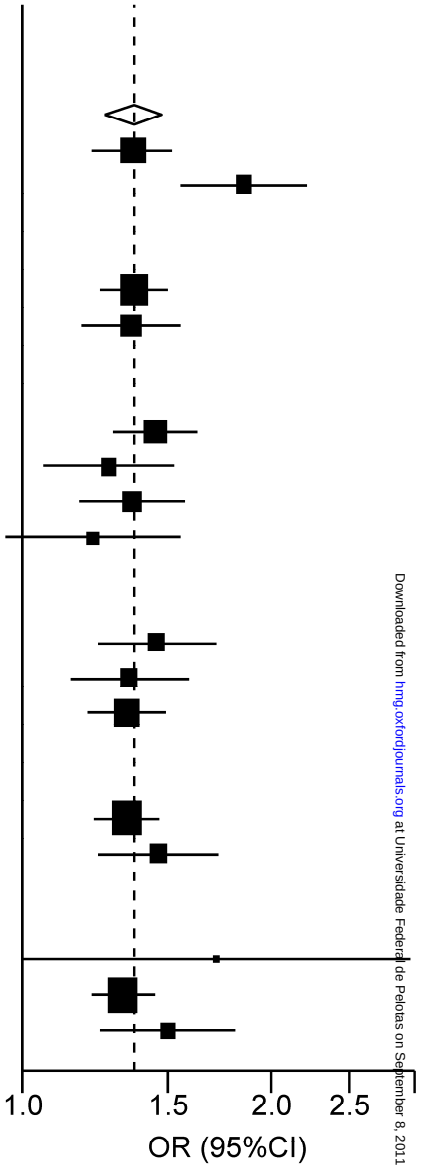
Figure 4. Association between rs9357152 and other serotypes in the combined series.

Unless specified, the ORs, and 95% CIs were derived from the log-additive genetic model with adjustment for age, country and sex when appropriate. *P* for heterogeneity was derived from the Cochran's Q test. [a] Being seropositive for L1 of HPV16, 18, 31, 33, 35, 45, 52 or 58. [b] Being seropositive for E6 of HPV6, 11, 16, 18, 31, 33, 35, 45, 52 or 58. [c] Being seropositive for E7 of HPV6, 11, 16, 18, 31, 33, or 35.





rs9357152	(+)	(-)	OR	95%CI
Log-additive	2040	4961	1.36	1.26-1.47
Heterozygous	908	1984	1.36	1.22-1.52
Homozygous	257	415	1.86	1.56-2.21
By phase (p-heterogeneity=0.91)				
GWAS	1333	3414	1.37	1.24-1.50
Replication	707	1547	1.35	1.18-1.56
By cancer status (p-heterogeneity=0.48)				
Controls	868	2138	1.45	1.29-1.64
Lung cancer	371	891	1.27	1.06-1.53
HN cancer	590	1333	1.36	1.17-1.57
Kidney cancer	211	599	1.21	0.95-1.55
By smoking status (p-heterogeneity=0.70)				
Never smokers	488	1100	1.45	1.23-1.72
Former smokers	478	1138	1.35	1.14-1.59
Current smokers	1071	2719	1.34	1.20-1.49
By gender (p-heterogeneity=0.35)				
Men	1601	3732	1.34	1.22-1.46
Women	439	1229	1.46	1.24-1.73
By age (p-heterogeneity=0.35)				
<40	46	146	1.72	1.00-2.96
40-<70	1622	4005	1.32	1.21-1.45
≥70	372	810	1.50	1.24-1.81



rs9357152	(+)	(-)	OR	95%CI
HPV8 L1	2040	4961	1.36	1.26-1.47
Cutaneous HPV L1				
a HPV77	704	6297	1.03	0.92-1.16
ß HPV38/49	1682	5319	1.20	1.10-1.30
γ HPV4	1627	5374	1.16	1.07-1.26
μ HPV1	960	6041	1.01	0.91-1.12
Mucosal HPV L1				
Low-risk HPV6/11	2073	4928	1.12	1.03-1.21
High-risk HPVs [a]	1977	5024	1.03	0.95-1.12
Mucosal HPV E6/E7				
E6 [b]	369	6632	0.97	0.82-1.14
E7 [c]	755	6246	0.83	0.74-0.94

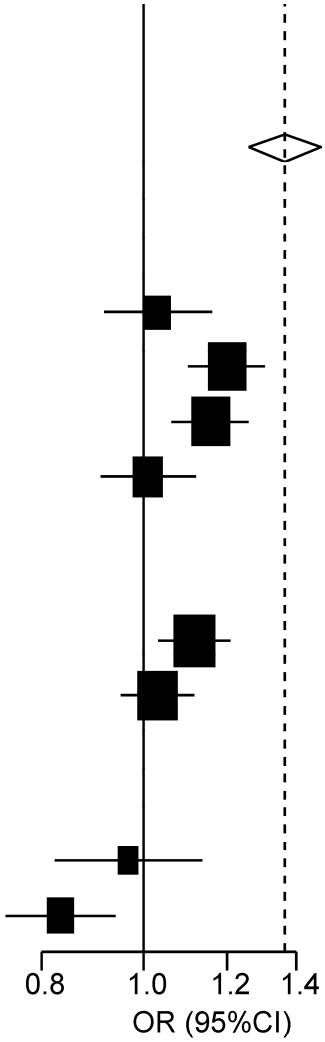


Table1. Distribution of type-specific HPV L1 antibody in the Central European study

Strain	Genus	Species	Seropositive for n (%) [a]	λ [b]
HPV6	α muc.(LR)	10	1192 (24.78)	1.002
HPV11	α muc.(LR)	10	556 (11.56)	1.011
HPV16	α muc.(HR)	9	330 (6.86)	1.007
HPV18	α muc.(HR)	7	569 (11.83)	1.012
HPV31	α muc.(HR)	9	264 (5.49)	1.002
HPV35	α muc.(HR)	9	397 (8.25)	1.003
HPV45	α muc.(HR)	7	336 (6.98)	0.997
HPV77	α cut.	2	494 (10.27)	1.007
HPV8	β cut.	1	1349 (28.04)	1.007
HPV38	β cut.	2	714 (14.84)	1.005
HPV49	β cut.	3	828 (17.21)	1.000
HPV4	γ cut.	1	1156 (24.03)	1.016
HPV1	μ cut.	1	705 (14.65)	1.010

muc: mucosal; LR: low-risk; HR: high-risk; cut: cutaneous,
[a] 4,811 subjects are included in the Central European study for genome-wide association analysis overall.
[b] Inflation factor in the genome-wide association analysis.

Table 2. List of SNPs followed-up with replication

							GWAS (4,811 [a])		Replication (2,344 [a])		
Locus	**SNP**	**Position [b]**	**Nearby gene**	**Major allele**	**Minor allele**	**MAF**	**OR (95%CI) [c]**	***P* [c]**	**OR (95%CI) [d]**	***P* [d]**	***P-het* [e]**
HPV8 L1											
6p21.32	rs9357152	32664960	*HLA-DQB1*	A	G	0.33	1.37 (1.24-1.50)	1.2×10^{-10}	1.35 (1.18-1.56)	2.2×10^{-5}	0.89
6p21.32	rs114427648 [f]	32407906	*HLA-DRA*	A	G	0.12	1.59 (1.40-1.82)	5.5×10^{-12}	1.51 (1.21-1.88)	3.0×10^{-4}	0.70
HPV31 L1											
6q22.3	rs9401090	119113317	*ASF1A*	C	T	0.19	1.69 (1.39-2.06)	2.1×10^{-7}	0.87 (0.66-1.16)	0.34	2.0×10^{-4}

MAF: minor allele frequency in cancer-free controls; OR: odds ratio; CI: confidence interval.
[a] Total number of study subjects for analysis.
[b] Build 37.1(GRCh37/hg19) Assembly.
[c] Derived from the log-additive genetic model with adjustment for age, sex and country. Further adjustment by cancer case/control status didn't alter the estimates. The potential for population stratification (PCA) not accounted by adjustment by country of recruitment was also investigated by principle components analysis (PCA) undertaken with the EIGENSTRAT package using 12,898 markers in low LD. None of the eigenvectors was significantly associated with serological status for any HPV type hence not included in the final model.
[d] Derived from the log-additive genetic model with adjustment for age, sex and country. Further adjustment by cancer case/control status didn't alter the estimates.
[e] Derived from the Cochran's Q test.
[f] rs114427648 was genotyped as a proxy of rs115639952 (D'=1, r^2=1)