Genome-wide association study of HPV seropositivity

Dan Chen¹, James D. McKay¹, Gary Clifford¹, Valérie Gaborieau¹, Amélie Chabrier¹, Tim Waterboer², David Zaridze³, Jolanta Lissowska⁴, Peter Rudnai⁵, Eleonora Fabianova⁶, Vladimir Bencko⁷, Vladimir Janout⁸, Lenka Foretova⁹, Ioan Nicolae Mates¹⁰, Neonila Szeszenia-Dabrowska¹¹, Maria Paula Curado¹², Sergio Koifman¹³, Ana Menezes¹⁴, Victor Wünsch-Filho¹⁵, José Eluf-Neto¹⁵, Leticia Fernández Garrote¹⁶, Elena Matos¹⁷, Diana Zelenika¹⁸, Anne Boland¹⁸, Paolo Boffetta¹²,¹⁹, Michael Pawlita², Mark Lathrop¹⁸,²⁰ and Paul Brennan¹,*

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

Received May 20, 2011; Revised July 30, 2011; Accepted August 23, 2011

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. Genome-wide genotyping of 316 015 single nucleotide polymorphisms was carried out using the Illumina HumanHap300 BeadChip in 4811 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 genome-wide 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 = 1.3 × 10⁻¹⁴], a common genetic variant (minor allele frequency = 0.33) located within the major histocompatibility complex (MHC) II region at 6p21.32. This association was subsequently replicated in an independent set of 2344 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 × 10⁻⁵), yielding P = 1.3 × 10⁻¹⁴ in the combined analysis (P-heterogeneity = 0.87). No heterogeneity was noted by cancer status.

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

© The Author 2011. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
INTRODUCTION

Human papillomaviruses (HPVs) 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 <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 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 epidermodysplasia 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 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 the 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 section. Genome-wide genotyping was successfully performed in 1413 lung cancer cases, 742 head and neck cancer cases, 853 kidney cancer cases and 2191 cancer-free controls. Following stringent quality-control steps (see Materials and Methods section for details), genome-wide statistical analysis was performed among 4811 subjects (1286 lung cancer cases, 679 head and neck cancer cases, 811 kidney cancer cases and 2035 cancer-free controls) for associations between 316015 single nucleotide polymorphisms (SNPs) and L1 seropositivity for the 13 HPV types with seroprevalence >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 Material, Table S2, 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 λ < 1.02).

Only one clear genome-wide significant association was identified (threshold for significance \( P < 1.2 \times 10^{-8} \)), 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; \( P = 1.2 \times 10^{-10} \)) (Figs 1 and 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 DBPI1*1501-DQBI1*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, \( P = 1.2 \times 10^{-8} \)).

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; \( P = 2.1 \times 10^{-7} \)) (Table 2, Supplementary Material, Fig. S1).

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, \( P = 2.2 \times 10^{-5} \)) with no evidence of heterogeneity between studies (\( P \)-heterogeneity = 0.87). In contrast, no association...
muc, mucosal; LR, low-risk; HR, high-risk; cut, cutaneous.

A total of 4811 subjects are included in the central European study for genome-wide association analysis overall.

Inflation factor in the genome-wide association analysis.

(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 (allele AG) and homozygous carriers (allele GG) of rs9357152 minor allele G, respectively, yielding P = 1.3 × 10^{-14} for trend in the combined analysis (Fig. 3). No heterogeneity for the association of rs9357152 with HPV8 seropositivity was noted by cancer status (control/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 × 10^{-5}) and γ cutaneous HPV4 (OR = 1.16, 95% CI = 1.07–1.26, P = 6.0 × 10^{-4}) as well as LR mucosal HPV types (HPV6, 11, 16, 18, 31, 33, 35, 39, 45, 52 or 58) (Fig. 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 × 10^{-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 × 10^{-4} compared with P = 5.2 × 10^{-7} in the unconditional analysis). Further conditional analysis on rs3135363 still left a residual association within this region (e.g. P = 1.5 × 10^{-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 section for details) identified a group of highly correlated SNPs rs115639952, chr6:32390621 and rs17207274 (D′ = 1, r² = 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 × 10^{-10}) (Fig. 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, r² = 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 >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 4432 overlapping individuals: direct genotyping OR = 1.59, 95% CI = 1.40–1.82, P = 5.5 × 10^{-12} versus imputed genotypes OR = 1.62, 95% CI = 1.43–1.83, P = 6.6 × 10^{-18} 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 × 10^{-7}) (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 the 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 the 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 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...
Table 2. List of SNPs followed-up with replication

<table>
<thead>
<tr>
<th>Locus</th>
<th>SNP</th>
<th>Position</th>
<th>Nearby gene</th>
<th>Major allele</th>
<th>Minor allele</th>
<th>MAF</th>
<th>GWAS (4811 a) OR (95% CI)</th>
<th>Replication (2344 a) OR (95% CI)</th>
<th>P&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P&lt;sup&gt;b&lt;/sup&gt;-het&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV8 L1</td>
<td>rs9357152</td>
<td>32664960</td>
<td>HLA-DQB1</td>
<td>A</td>
<td>G</td>
<td>0.33</td>
<td>1.37 (1.24–1.56)</td>
<td>1.35 (1.18–1.56)</td>
<td>10&lt;sup&gt;−10&lt;/sup&gt;</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>rs114427648&lt;sup&gt;d&lt;/sup&gt;</td>
<td>32407906</td>
<td>HLA-DRA</td>
<td>A</td>
<td>G</td>
<td>0.12</td>
<td>1.59 (1.40–1.82)</td>
<td>1.51 (1.21–1.88)</td>
<td>10&lt;sup&gt;−12&lt;/sup&gt;</td>
<td>0.70</td>
</tr>
<tr>
<td>HPV31 L1</td>
<td>6q22.3</td>
<td>119113317</td>
<td>ASF1A</td>
<td>C</td>
<td>T</td>
<td>0.19</td>
<td>1.69 (1.39–2.06)</td>
<td>0.87 (0.66–1.16)</td>
<td>10&lt;sup&gt;−7&lt;/sup&gt;</td>
<td>0.34</td>
</tr>
</tbody>
</table>

MAF, minor allele frequency in cancer-free controls; OR, odds ratio; CI, confidence interval.

*Total number of study subjects for analysis.

<sup>a</sup>Build 37.2 (GRCh37/hg19) Assembly.

<sup>b</sup>Derived from the log-additive genetic model with adjustment for age, sex and country. Further adjustment by cancer case/control status did not 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.

<sup>c</sup>Derived from the log-additive genetic model with adjustment for age, sex and country. Further adjustment by cancer case/control status did not alter the estimates.

<sup>d</sup>Derived from the Cochran’s Q test.

rs114427648 was genotyped as a proxy of rs115639952 (\(D' = 1, r^2 = 1\)).
acid sequence relatedness, suggesting that 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 CE 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 to be 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 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 1413 lung cancer cases, 742 head and neck cancer cases, 853 kidney cancer cases and 2191 cancer-free controls. The replication series consisted of 2344 subjects (1307 head and neck cancer cases and 1037 cancer-free controls) from a LA 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 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, USA), 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 analyzer was used to identify the internal color 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 color 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 cutoffs for cutaneous HPV L1 antibodies were those defined previously (26). Seropositivity cutoffs for mucosal HPV L1 were determined using serum samples of 371 females, HPV DNA negative, self-reported virgins from a cross-sectional study among Korean students, as reported previously (45). Seropositivity cutoffs for E6 and E7 antibodies were defined as previously (46). A bridging panel of ~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 cutoff values.

Genome-wide genotyping and quality control

The CE study was genotyped using the Illumina’s Sentrix HumanHap300 BeadChip 317Kc 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 Material, Table S1). The exclusion criteria for SNPs were: MAF ≤1%; genotyping call rate <95%; genotype frequency deviated from the Hardy-Weinberg equilibrium (HWE) among cancer-free controls (P < 1 × 10^-5); resulting in data on 316 015 SNPs. Individuals with the overall genotype completion rate <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 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 subsample of 12898 genetic variants from the HumanHap 300 BeadChip panel evenly distributed across the genome and in low 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 4811 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 × 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 12898 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 2344 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 >99.5%. The genotype success rate was >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 (http://www.sph.umich.edu/csg/abecasis/MACH/index.html, last accessed date on 30 June 2011) 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 (51), including age, sex, country of origin, and cancer case/control status in the regression as covariates. Association results for all SNPs with $r^2 > 0.3$ in the MHC region are shown in Figure 2. LD statistics (D' and r2) were calculated using Haploview (52). To validate the imputed assays.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgement

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 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.

References
