Human papillomavirus (HPV) is the etiologic risk factor for cervical cancer. Some studies have suggested an association with a subset of lung tumors, but the etiologic link has not been firmly established. We performed an international pooled analysis of cross-sectional studies (27 datasets, n = 3,249 patients) to evaluate HPV DNA prevalence in lung cancer and to investigate viral presence according to clinical and demographic characteristics. HPV16/18 were the most commonly detected, but with substantial variation in viral prevalence between geographic regions. The highest prevalence of HPV16/18 was observed in South and Central America, followed by Asia, North America and Europe (adjusted prevalence rates = 22, 5 and 4%, respectively). Higher HPV16 prevalence was noted in each geographic region compared with HPV18, except in North America. HPV16/18-positive lung cancer was less likely observed among White race (adjusted odds ratio [OR] = 0.33, 95% confidence interval [CI] = 0.12–0.90), whereas no associations were observed with gender, smoking history, age, histology or stage. Comparisons between tumor and normal lung tissue show that HPV was more likely to be present in lung cancer rather than normal lung tissues (OR = 3.86, 95% CI = 2.87–5.19). Among a subset of patients with HPV16-positive tumors, integration was primarily observed in male patients (93%, 13/14), while the physical status in male cases (N = 14) was inconsistent. Our findings confirm that HPV DNA is present in a small fraction of lung tumors, with large geographic variations. Further comprehensive analysis is needed to assess whether this association reflects a causal relationship.

Introduction

Lung cancer is the third most common cancer in men and women of all races and the leading cause of cancer death in the United States and worldwide. Cigarette smoking is the primary risk factor and accounts for ~85% of all lung cancer cases. Other less prevalent risk factors are genetic factors, family history of lung cancer and exposures to radon, asbestos, arsenic, diesel exhaust and some forms of silica and chromium. Human papillomavirus (HPV) is a small non-enveloped DNA tumor virus of the family Papillomaviridae and is the established etiological agent of genital warts, cervical cancer and a proportion of cancers of the vulva, vagina, penis, anus and oropharynx (1). More than 100 different genotypes have been identified, and types 16 and 18 are the most common oncogenic types, leading to the development of ~70% of all cervical carcinomas. In 1979, Syrjänen first hypothesized that certain types of HPV are responsible for causing cancer in the lung (3). In the last three decades, the number of reports suggesting an association between HPV and lung cancer has increased tremendously. A review of HPV in 2468 lung cancer cases was first published in 2002 and showed that, using morphological, immunohistochemical and HPV DNA detection methods, ~22% of the cases analyzed contained HPV. This meta-analysis was updated in 2007 and reported a prevalence of 8.5%, with HPV generally present in non-small cell lung cancer (NSCLC) compared with small cell lung cancer (SCLC). In 2009, a pooled meta-analysis of 2468 lung cancer cases confirmed that HPV DNA was more likely to be present in lung cancer rather than normal lung tissue (OR = 3.86, 95% CI = 2.87–5.19). Among a subset of patients with HPV16-positive tumors, integration was primarily observed in male patients (93%, 13/14), while the physical status in male cases (N = 14) was inconsistent. Our findings confirm that HPV DNA is present in a small fraction of lung tumors, with large geographic variations. Further comprehensive analysis is needed to assess whether this association reflects a causal relationship.
gender, tumor stage and smoking status of HPV-positive lung cancer patients have not been taken into account. To address these limitations, we have conducted an international, multi-institutional, pooled analysis of studies to further analyze the characteristics of cases observed with HPV16/18 DNA with particular emphasis on race, gender, smoking status, tumor stage as well as histology of the cancers. The most recent meta-analysis has been updated (7); prevalence rates of HPV in lung tumors by geographic region, adjusted for age, gender, smoking status and tumor stage are presented. Because the majority of studies to date have been conducted in Asian and European populations, comparisons of demographic and clinical characteristics of HPV-positive lung cancer cases were also made between North American cases and the rest of the world.

Materials and methods

Literature review and data extraction
A flow diagram that summarizes the literature review process and selection of datasets for inclusion in the pooled analysis is described in Figure 1. A bibliographic search was carried out in the PubMed database to identify the studies that evaluated HPV status of primary lung tumors published up to July 2013. The search strategy used was: (HPV OR HPV) AND (lung OR bronchogenic) AND (cancer OR carcinoma). A manual review of the bibliographic references cited in the selected papers was also undertaken to retrieve papers that might have been missed in the initial search. From this search, 90 publications were identified. When the data were reported from the same cohort (8–30), only the most recent publication or one that had the larger cohort of lung cancer cases was included (10,13,14,17,20,21,25,26,30). Two publications had overlapping squamous cell carcinoma (SCC) cases (8,9) but only one reported data for adenocarcinoma cases (9). Therefore, although counted as a single study, both publications were included. To ensure that the larger cohort was included, the SCC data were extracted from one publication (8) and the adenocarcinoma data from the other (9). The majority of studies to date were conducted in Asian and European populations. Data from two published abstracts (31,32) and two unpublished datasets (E.Taioli et al. and A.Zaravinos et al., unpublished results) have also been included in this pooled analysis. These cases were recruited from three medical institutions in the United States and one from Greece. This resulted in 79 published studies and 2 unpublished studies, which include 7440 lung cancer cases. Sixteen studies also reported the prevalence of HPV in non-tumor lung tissues (i.e. in adjacent normal tissues from lung cancer cases or in lung tissues from non-cancer cases) (13,17,20,25,33–44). These data are included as a subset analysis comparing HPV prevalence between tumor and normal lung tissue. In order to describe the characteristics of all eligible studies, the following information were extracted for each study: first author, year of publication, geographic region, study size, method of HPV detection, HPV types detected and HPV status overall and when available, according to gender, smoking status, tumor stage and histology.

Pooled analysis data collection
Following Institutional Review Board approval, invitations to participate in the pooled analysis were prepared for the principal investigators of the 81 eligible studies. Invitations were not successfully sent to investigators of 13 studies due to inactive or unavailable contact information. Although no response or a refusal to participate was received from 47 studies, only 41 were excluded because 6 of these studies reported individual-level data in the publications. Datasets with age, gender, tumor stage, smoking status, histological type and HPV results were created for each of these studies and were included in the pooled analysis (8,9,26,38,45–47). In total, 27 studies were included in the pooled analysis (8,9,11,13,14,26,31,32,34,38,45–60) including the studies of E.Taioli et al. and A.Zaravinos et al. (unpublished results).

Verification of cancer diagnosis was reported in the corresponding publications for 21 of the 27 studies (8,9,11,13,14,26,31,32,34,38,45–50,53–60). If not explicitly stated in the published article, authors were contacted to clarify what, if any, efforts were made to avoid contamination. Efforts to avoid contamination were reported for 22 of the 27 studies.
HPV and lung cancer

(8,9,11,13,26,31,32,34,38,45–48,50–52,54–56,58,59) and for E.Taioli et al. and A.Zaravinos et al. (unpublished results), and all studies reported the inclusion of internal quality controls and/or positive and negative HPV controls. From the pooled dataset, 65 cases were excluded: 51 because of inadequate DNA based on quality control checks performed by the submitting investigator (54), 2 because they were duplicate samples (A.Zaravinos et al., unpublished results) and 12 because they were metastasis to the lung (13,57,50,60). The pooled analytic dataset included 3249 cases.

Published abstracts and unpublished studies

For E.Taioli et al. (unpublished results; n = 69), all cases were histologically confirmed lung cancer patients diagnosed at the Department of Surgical Pathology of the University of Crete from 2007 to 2011. About 5–10 serial tissue sections of 10 µm were cut from each FFPE block, which was stained with hematoxylin and eosin (H&E) before microscopic examination. When the proportion of tumor cells was >70%, the FFPE block was subjected to DNA extraction. Data from Mehra et al. (32) (n = 62) were published in the 2013 Proceedings of the American Association for Cancer Research. All cases were a convenience sample of randomly selected histologically confirmed lung cancer patients (enriched for non-smoking status) with adenocarcinoma (in situ and invasive) diagnosed at Fox Chase Cancer Center (Philadelphia, PA) from 1993 to 2010. FFPE tissue sections and corresponding demographic and clinical information were obtained from the FCCCC Biosample Repository. DNA from tumor tissue was extracted and tested for HPV, and physical status was determined for all HPV16-positive samples. For A.Zaravinos et al. (unpublished results; n = 17), all cases were histologically confirmed lung cancer patients diagnosed at New York University (NYU)-Bellevue Hospital from 1993 to 1999. The central pathology registry at NYU-Bellevue Hospital was used to identify all consecutive African American cases; clinical charts were then retrieved using the pathology report number, and personal/behavioral information was collected. The pathologist reviewed all the tissue blocks to make sure that a block containing material from the primary lung tumor was used for testing. DNA from formalin-fixed paraffin-embedded (FFPE) tissues was extracted and tested for HPV, and physical status was determined for all HPV16-positive samples. For E.Taioli et al. (unpublished results; n = 208), were published in the 2013 Proceedings of the American Society of Clinical Oncology. All cases were histologically confirmed surgically resected lung cancer patients diagnosed at WellStar Health Systems, Atlanta, GA, from 2002 to 2008. The samples were consecutively acquired paraffin-embedded tissues, archived in a temperature controlled storage area until ready for analysis. Details of the tissue acquisition have been described previously (61). DNA from FFPE tissues was extracted and tested for HPV, and physical status was evaluated for all HPV16-positive tissues.

HPV detection and genotyping

The HPV testing for the published abstracts and unpublished dataset (E.Taioli et al., unpublished results) was performed in a single laboratory. DNA was extracted from FFPE lung cancer tissues using ArchivePure DNA purification kit (5 Prime). DNA concentration and quality checks were performed by evaluating 260-280 ratios as well as PCR amplification of a β-globin amplion using the RS 42 and KM 29 primers. For the E.Taioli et al. (unpublished results) samples, HPV status was determined by nested HPV PCR reactions using the PGMY09/11, followed by GP5+/GP6+ primers. Briefly, 20 µl PCR amplifications were performed using a GeneAmp PCR System© 9700 (Applied Biosystems, Life Technologies) at 95°C for 90 min, 40 cycles of 95°C for 30 s, 55°C for 1 min and 72°C for 10 min with a final extension of 72°C for 5 min. The nested PCR reaction was performed by prediluting 1:100 PCR products from the initial reaction in a final reaction volume of 25 µl. The cycling conditions were 95°C for 10 min, 40 cycles of 95°C for 1 min, 55°C for 1 min and 72°C for 1 min, followed by a final extension at 72°C for 5 min. Negative and positive controls were included in each PCR reaction. The PCR products were separated in 2% agarose gel by electrophoresis and HPV genotyping of samples was performed using the Linear Array HPV Genotyping kit (Roche Diagnostics) according to the manufacturer’s instructions. For the Mehra et al. and Pillai et al. samples, HPV testing was performed using the INNO-LiPA genotyping Extra Amplification and Genotyping Extra kits (Innogenetics, Belgium). The SPF10 PCR primers are capable of identifying a broad spectrum of HPV genotypes by amplifying a 65 bp target in the HPV L1 sequence. Genotyping of positive HPV samples was performed using the LI-PA genotyping protocol that involves a reverse hybridization line probe assay to detect 18 high-risk types (16,18,26,31,33,35,39,45,51–53,56,58,59,66,68,73,82), 7 low-risk types (6,11,40,43,44,50,70) and additional types (69,71,74). Negative and positive controls (HPV6, as well as an internal control (HLA-DPB1 gene), to confirm DNA quality and the absence PCR inhibitors, were included in the assay. HPV testing for the second unpublished study was performed by the submitting investigator (A.Zaravinos et al., unpublished results). Lung cancer tissue sections (FFPE) were deparaffinized with xylene and ethanol washes, treated with protease and then DNA from tumor tissue was extracted using the QIAamp DNA mini kit (Qiagen, Valencia, CA) according to the manufacturer’s instructions. Briefly, tissue sections were digested with 0.1 mg/ml proteinase K (Promega, Madison, WI) and 400 µl of digestion buffer containing 150 mM NaCl, 400 mM Tris-HCl, 60 mM ethylenediaminetetraacetic acid and 15% sodium dodecyl sulfate. Specimens were then incubated at 60°C for 2 days. Fresh protease K was added three days later. The samples were extracted once with phenol/chloroform and once with chloroform. DNA was precipitated by the addition of 20 µl of 5 M NaCl and 1 ml of ethanol, recovered by centrifugation for 15 min, washed once with cold 70% ethanol and resuspended in 50 µl of double distilled water. DNA concentration was calculated using the NanoDrop™ 1000 Spectrophotometer. Sequences were examined for the presence of amplifiable DNA using a set of primers for the β2-microglobulin gene. Amplification of HPV DNA in the specimens was performed by PCR using specific primer pairs for each E6 gene region of the HPV6-, -11,-16,-18 and -33 genomes detection. For the general screening of HPV DNA, HEaL cells transfected with conserved L1 sequences among HPV strains were used as the positive control. PCR was carried out in a total volume of 25 µl containing 5 mM of 5x Green GoTaq reaction buffer, 1.5 mM MgCl2, 0.2 mM of each deoxynucleotide triphosphate (dNTPs), 0.6 µl of GoTaq Flexi DNA polymerase (Promega) and 200 ng of genomic DNA. The PCR products were examined by electrophoresis on a 2% agarose gel and photographed on an ultraviolet light transilluminator. The sensitivity of the PCR assay was determined by applying a serial-dilution amplification assay of viral-positive control DNA.

HPV physical status

The physical status of HPV16 was assessed in HPV16-positive lung cancer tissues for three published studies that were included in the pooled analysis (32,48,49) using the same quantitative PCR assay as described previously (62). Each HPV16 DNA-positive sample was amplified for 76 bp of the E2 gene using the following primers: forward 5′-AAGAAGATGTTCC TCTCTGAAATTTAG-3′ (3361–3389 nt); reverse 5′-CCAGGGCG ACGGCTTGT-3′ (3427–3443 nt), as well as 81 bp of the E6 gene, primers forward 5′-GAGAACCTGCA AGTTTACGAC-3′ (94–116 nt); reverse 5′-TGTATAAGTGGTTCGAC TCTGTGC-3′ (150–169 nt), in the presence of specific hybridization probes for E2-(FAM-CACCC CGCGGTACCCACAT-TAMRA) (3406–3424 nt) and E6-(FAM-C AGGGACGC CCAGAAGTACCAACGTT-TAMRA) (119–147 nt). The cycling conditions were 2 min at 50°C, 10 min at 95°C, and a two-step cycle of 95°C for 45 s and 60°C for 1 min for a total of 40 cycles. The resulting curve was generated from 8-fold serial dilution of p1203 PML2d HPV16 (Addgene plasmid 10869, deposited by Peter Howley, MD; Addgene, Cambridge, MA). Both HPV16-positive cell lines, SiHa and Caski cell lines were included as controls. The assay was performed in duplicate for each sample. Physical status was determined by calculating the ratios of E2 to E6 where a ratio <1.0 indicates a predominance of integrated HPV genomes, >1.0 indicates a predominance of episomal genomes. The results were recorded as copy numbers per 20 ng of DNA.

Statistical analysis

All statistical analyses were carried out using SAS 9.3 (SAS Institute, Cary, NC). A level of 0.05 and evaluation of publication bias was performed using the Comprehensive Meta Analysis Version 2 software (Biostat, Englewood, NJ). The adjusted HPV prevalence of each region was calculated after adjusting for age, gender, smoking history, tumor stage and study and reported as an estimated probability of HPV-positive lung cancer patients with 95% confidence intervals (CIs). Random effects logistic regression models were used to determine the association between HPV16/18 status and clinical and demographic variables. Study was used as a random effect to allow deviation caused by different study conditions. In the international pooled dataset, one out of three of the subjects was missing either smoking history or tumor stage. Both variables, we believe, are important in understanding the nature of HPV16/18 prevalence. However, exclusion of subjects with missing values is inefficient and can lead to biased results if those dropped are atypical in some respects. Therefore, multiple imputations were performed prior to regression analyses using IVEware. Ten imputation datasets were generated and final model estimates were combined using SAS 9.3. For each imputed dataset, the missing values were drawn from other observed data (63). The uncertainty

1269

Downloaded from http://carcin.oxfordjournals.org/ at University of Alberta on October 7, 2014
about the correct model for non-response was captured by the variance across multiple imputed datasets. Subjects with one or part of the covariates missing were included in the analysis and hence, no information was lost due to exclusion.

Results

Twenty-seven of the 81 eligible studies elected to participate in this pooled analysis and the resulting pooled dataset included 3249 cases, whereas the remaining 54 studies had 4199 cases. Detailed characteristics of each study in the pooled dataset are summarized in Table I. The majority of cases in the pooled dataset were from Asia (40%, 1312 cases) and Europe (34%, 1100 cases). North/South American studies represented 26% of all cases in the pooled dataset (n = 837). For the non-included studies, the majority of cases were also from Asia (55%, 2332 cases) and Europe (33%, 1404 cases), whereas 7% of cases (n = 299) were from North/South America and 4% (n = 164 cases) from other geographic regions (Australia and the Middle East). The size of studies in the pooled dataset varied from 17 to 399 cases, whereas the non-included study sizes ranged from 5 to 319 cases.

Table I. Characteristics of studies included in the pooled analysis and adjusted HPV prevalence

<table>
<thead>
<tr>
<th>Study</th>
<th>Tissue source</th>
<th>Method of HPV detection</th>
<th>Histological type</th>
<th>HPV types detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia (N = 1312)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baba et al. (N = 77) (49)</td>
<td>FFPE</td>
<td>PCR (SPF10, Inno-LIPA)</td>
<td>AC, SCC</td>
<td>HPV 6, 16, 18, 33</td>
</tr>
<tr>
<td>Kinoshita et al. (N = 34) (57)</td>
<td>FF</td>
<td>PCR (TS: HPV 16, 18, 33)</td>
<td>AC, SCC, ASq, LCC, SmCC</td>
<td>HPV 18</td>
</tr>
<tr>
<td>Iwakawa et al. (N = 275) (53)</td>
<td>FF</td>
<td>PCR (TS: HPV 16, 18, 33)</td>
<td>AC</td>
<td>HPV negative</td>
</tr>
<tr>
<td>Hiroshima et al. (N = 49) (52)</td>
<td>FFPE</td>
<td>PCR (TS: HPV 16, 18, 33)</td>
<td>AC, SCC</td>
<td>HPV 16</td>
</tr>
<tr>
<td>Goto et al. (N = 296) (51)</td>
<td>FFPE</td>
<td>PCR GP5+/GP6+</td>
<td>AC, SCC, CSrc</td>
<td>HPV 16, 11, 16, 18</td>
</tr>
<tr>
<td>Lim et al. (2009 N = 99) (56)</td>
<td>FFPE</td>
<td>ISH (HPV6, 11, 16, 18, 31, 33, 35, 45, 51, 52, 56, 58, 66)</td>
<td>AC</td>
<td>HPV negative</td>
</tr>
<tr>
<td>Miyagi et al. (N = 176) (8, 9)</td>
<td>FFPE</td>
<td>PCR (TS: HPV 6, 11, 16, 18)</td>
<td>AC, SCC</td>
<td>HPV 6, 11, 16, 18</td>
</tr>
<tr>
<td>Hirayasu et al. (N = 73) (26)</td>
<td>FFPE</td>
<td>PCR (TS: HPV 6, 11, 16, 18)</td>
<td>SCC</td>
<td>HPV 6, 11, 16, 18</td>
</tr>
<tr>
<td>Wang et al. (N = 210) (58)</td>
<td>FFPE, Bx, PE</td>
<td>PCR (MY09/11 and TS: HPV 16, 18)</td>
<td>AC</td>
<td>HPV16, HPV18</td>
</tr>
<tr>
<td>Tsuchako et al. (N = 23) (47)</td>
<td>FFPE</td>
<td>PCR (TS: HPV 6, 11, 16, 18)</td>
<td>ASq</td>
<td>HPV 6, 11, 16, 18</td>
</tr>
<tr>
<td>Europe (N = 1100)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syrjanen et al. (N = 131) (14)</td>
<td>FFPE</td>
<td>ISH (HPV6, 11, 16, 18, 30)</td>
<td>SCC</td>
<td>HPV 6, 16, 34,</td>
</tr>
<tr>
<td>Syrjanen et al. (N = 77) (55)</td>
<td>FFPE</td>
<td>PCR (MY09/11/GP5+/GP6+)</td>
<td>AC, SCC, other</td>
<td>HPV 6, 16, 18</td>
</tr>
<tr>
<td>Giuliani et al. (N = 77) (13)</td>
<td>FFPE and FF</td>
<td>PCR (MY09/11/GP5+)</td>
<td>HPV negative</td>
<td></td>
</tr>
<tr>
<td>Gorgoulis et al. (N = 68) (11)</td>
<td>FFPE and FF</td>
<td>PCR (MY09/11/GP5+/GP6+)</td>
<td>HPV negative</td>
<td></td>
</tr>
<tr>
<td>Zaravinos et al. (unpublished results, N = 17)</td>
<td>FFPE</td>
<td>PCR (GP5+/GP6+)</td>
<td>HPV negative</td>
<td></td>
</tr>
<tr>
<td>Koshiol et al. (N = 399) (54)</td>
<td>FFPE</td>
<td>PCR (TS: HPV 16, 18; SPF10 primers (n = 92))</td>
<td>AC, SCC, LCC, SmCC, other</td>
<td>HPV 16, 18</td>
</tr>
<tr>
<td>Carpagnano et al. (N = 89) (34)</td>
<td>FFPE and BB</td>
<td>PCR (consensus HPV primers)</td>
<td>AC, SCC, SmCC</td>
<td>HPV 16, 30, 31, 39</td>
</tr>
<tr>
<td>Nuovrta et al. (N = 22) (46)</td>
<td>FFPE</td>
<td>PCR (MY09/11)</td>
<td>AC</td>
<td>HPV 6, 11, 16, 18, 31, 33</td>
</tr>
<tr>
<td>van Boerdonk et al. (N = 220) (59)</td>
<td>FFPE</td>
<td>PCR (GP5+/GP6+)</td>
<td>HPV negative</td>
<td></td>
</tr>
<tr>
<td>South and Central America (N = 105)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Castillo et al. (N = 36) (50)</td>
<td>FFPE</td>
<td>PCR (GP5 +6+) and SB</td>
<td>AC, SCC, SmCC</td>
<td>HPV 16, 18, 33</td>
</tr>
<tr>
<td>Aguayo et al. (N = 69) (48)</td>
<td>FFPE</td>
<td>PCR (GP5+/GP6+) and SB</td>
<td>AC, SCC</td>
<td>HPV 6, 16, 18, 31, 45</td>
</tr>
<tr>
<td>North America (N = 732)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taioli et al. (unpublished results, N = 69)</td>
<td>FFPE</td>
<td>PCR (PGMY09/PGMY11/</td>
<td>HPV 51, X</td>
<td></td>
</tr>
<tr>
<td>Pillai et al. (N = 208) (31)</td>
<td>FFPE</td>
<td>PCR (GP5+/6+)</td>
<td>HPV 51, X</td>
<td></td>
</tr>
<tr>
<td>Mehra et al. (N = 62) (32)</td>
<td>FFPE</td>
<td>PCR (SPF10, Inno-LIPA)</td>
<td>HPV 6, 16, 18, 39, 53, X</td>
<td></td>
</tr>
<tr>
<td>Joh et al. (N = 29) (38)</td>
<td>FF</td>
<td>PCR (GP5+/GP6+)</td>
<td>HPV 6, 16, 18, 32, 53, 44, 68, 39, 74, 82, X</td>
<td></td>
</tr>
<tr>
<td>Yangawal et al. (N = 330) (60)</td>
<td>FFPE</td>
<td>PCR (GP5+/GP6+) and ISH</td>
<td>AC, SCC, LCC, NSCLC</td>
<td>HPV 11, 16</td>
</tr>
<tr>
<td>Bohlmeyer et al. (N = 34) (45)</td>
<td>FFPE</td>
<td>PCR (MY09/11)</td>
<td>HPV 18</td>
<td></td>
</tr>
</tbody>
</table>

Adjusted HPV 16/18 prevalence (95% CI) = 2.19% (1.45, 3.33)
Adjusted HPV 16 prevalence (95% CI) = 1.68, 2.39
Adjusted HPV 18 prevalence (95% CI) = 2.49% (2.23, 2.75)

Other: histological type, not otherwise specified. AC, adenocarcinoma; ASq, adenosquamous carcinoma; BB, bronchial brushing; Bx, biopsy tissue; CSrc, carcinosarcoma; FF, fresh frozen; ISH, in situ hybridization; LCC, large cell carcinoma; NSCLC, non-small-cell lung carcinoma, not otherwise specified; PE, pleural effusion; SB, Southern blot; SmCC, small cell carcinoma; TS, type-specific primers.

1Number of cases 56.
2Cancer cells present and confirmed by pathologist.
3Study is included in the model as a random effect.
4Type not defined; HPV+ (for the HPV6/11/16/18/30 panel).
5Number of cases 7–10.
6Study is included as a covariate in the model.
All studies included in the pooled analysis performed HPV DNA testing on surgical tissue (FFPE and/or frozen) or biopsy specimens. Similarly, the non-included studies performed HPV DNA testing on surgical or biopsy tissue specimens except for one study which utilized bronchial aspirates from lung cancer patients (64). Most of the studies in the pooled analysis as well as the non-included studies used PCR-based methods for HPV DNA detection, 93 and 80%, respectively. Although adenocarcinoma and SCC were the most common histological types included in the pooled dataset (56%, 1808 cases and 38%, 1220 cases, respectively) and in the non-included studies (33%, 1404 cases and 57%, 2379 cases, respectively), the pooled analysis consisted of a larger proportion of adenocarcinoma cases compared with the non-included studies. Similar proportions of other histological types were represented in both the pooled dataset and non-included studies (7%, 214 cases and 8%, 350 cases, respectively). Histology was not specified for 7 (0.2%) cases in the pooled dataset and for 68 (1.6%) cases in the non-included studies.

The predominant high-risk HPV types found in lung cancer patients were HPV 16 and HPV 18 (Table I). Other high-risk HPV types (HPV31, 33, 35, 45, 51, 52, 68 and 82) were also detected. For all geographic regions (Asia, Europe as well as North America) and with the exception of South and Central America, large-to-moderate heterogeneity was observed between the studies with no evidence of publication bias (data not shown). In order to address the heterogeneity, HPV prevalence rates were calculated after adjusting for age, gender, smoking history and tumor stage; study was included as a random effect where appropriate (Table I). Significant differences in HPV16/18 prevalence were noted between geographic regions. HPV 16/18 was most prevalent in South and Central America (adjusted prevalence [AdjPr] = 21.90%, 95% CI = 19.61–24.20), followed by Asia (AdjPr = 4.60%, 95% CI = 3.48–5.73), North America (AdjPr = 3.78%, 95% CI = 3.35–4.22) and Europe (AdjPr = 3.03%, 95% CI = 2.76–3.30). For each geographic region, the prevalence of HPV16 was significantly higher than HPV18, with the exception of South and Central America, whereas HPV 18 was also highest in South and Central America, followed by North America and Asia. There was virtually no HPV 18 detected in European patients (AdjPr = 0.82%, 95% CI = 0.73–0.92).

Clinical and demographic characteristics of HPV16/18-positive lung cancer cases

The mean age of all lung cancer patients in the pooled dataset was 65.0±10.1 (SD) years. The demographic and clinical characteristics for all cases are presented in Table II; comparisons were made between North American cases and the rest of the world. For the studies conducted in North America as well as the rest of the world, a higher proportion of the patients were male (53 and 72%, respectively), ever smokers (84 and 76%, respectively) and most of the lung cancers were not SCC (64 and 62%, respectively). North American patients were mostly White (89%) in contrast to the rest of the world where the majority were non-White (57%). Early-stage lung cancers were more represented in the North American studies (63%) in contrast to only 50% in studies from the rest of the world. For each geographic subgroup, there were no noted differences in the proportion of HPV16/18 patients according clinical and demographic characteristics with the exception of race. The majority of patients in North America were White and a higher proportion of HPV 16/18-positive lung cancers were also noted among White patients (72%). However, for the rest of the world, although a near equal proportion of White and non-White patients were included, the majority of HPV 16/18-positive lung cancers were observed in non-White patients (88%).

No statistically significant associations were observed between HPV16/18-positive lung tumors and gender, age, smoking history, stage and histology (Table III). However, HPV 16/18-positive lung cancer cases were less likely to be observed among White patients compared with non-White patients (adjusted odds ratio [Adj OR] = 0.33, 95% CI = 0.12–0.90). For larger studies, (≥100 cases, 4 studies with a minimum of 131 cases and a maximum of 399 cases), the association of HPV 16/18 with race persisted (White: Adj OR = 0.08, 96% CI = 0.02–0.33). For smaller studies (<100 cases, 16 studies with a minimum of 22 cases and a maximum of 99 cases), the HPV 16/18-positive

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total (N = 2500)</th>
<th>HPV16/18 positive (N = 233, 9.3%)</th>
<th>HPV16/18 negative (N = 2267, 90.7%)</th>
<th>P valuea</th>
<th>North America (N = 663)</th>
<th>HPV16/18 positive (N = 23, 3.5%)</th>
<th>HPV16/18 negative (N = 640, 96.5%)</th>
<th>P valueb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1749 (71.92)</td>
<td>154 (66.96)</td>
<td>1595 (72.43)</td>
<td>0.157</td>
<td>349 (52.64)</td>
<td>10 (43.48)</td>
<td>339 (52.97)</td>
<td>0.632</td>
</tr>
<tr>
<td>Female</td>
<td>683 (28.08)</td>
<td>76 (33.04)</td>
<td>607 (27.57)</td>
<td></td>
<td>314 (47.36)</td>
<td>13 (56.52)</td>
<td>301 (47.03)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤66</td>
<td>1212 (51.33)</td>
<td>92 (42.4)</td>
<td>1120 (52.24)</td>
<td>0.529</td>
<td>315 (47.51)</td>
<td>10 (43.48)</td>
<td>305 (47.66)</td>
<td>0.557</td>
</tr>
<tr>
<td>&gt;66</td>
<td>1149 (48.67)</td>
<td>125 (57.6)</td>
<td>1024 (47.76)</td>
<td></td>
<td>348 (52.49)</td>
<td>13 (56.52)</td>
<td>335 (52.34)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td>0.059</td>
<td></td>
<td></td>
<td></td>
<td>0.102</td>
</tr>
<tr>
<td>White</td>
<td>1083 (43.32)</td>
<td>27 (11.59)</td>
<td>1056 (46.58)</td>
<td></td>
<td>239 (88.52)</td>
<td>13 (72.22)</td>
<td>226 (89.68)</td>
<td></td>
</tr>
<tr>
<td>Othersc</td>
<td>1417 (56.68)</td>
<td>206 (88.41)</td>
<td>1211 (53.42)</td>
<td></td>
<td>31 (11.48)</td>
<td>5 (27.78)</td>
<td>26 (10.32)</td>
<td></td>
</tr>
<tr>
<td>Smoking history</td>
<td></td>
<td></td>
<td></td>
<td>0.762</td>
<td></td>
<td></td>
<td></td>
<td>0.256</td>
</tr>
<tr>
<td>Never smoked</td>
<td>602 (24.06)</td>
<td>67 (28.93)</td>
<td>534 (23.56)</td>
<td></td>
<td>104 (15.69)</td>
<td>3 (13.04)</td>
<td>101 (15.78)</td>
<td></td>
</tr>
<tr>
<td>Ever smoked</td>
<td>1898 (75.94)</td>
<td>166 (71.07)</td>
<td>1733 (76.44)</td>
<td></td>
<td>559 (84.31)</td>
<td>20 (86.96)</td>
<td>539 (84.22)</td>
<td></td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
<td></td>
<td>0.536</td>
<td></td>
<td></td>
<td></td>
<td>0.345</td>
</tr>
<tr>
<td>SCC</td>
<td>953 (38.12)</td>
<td>107 (45.92)</td>
<td>846 (37.32)</td>
<td></td>
<td>239 (36.43)</td>
<td>5 (22.73)</td>
<td>234 (36.91)</td>
<td></td>
</tr>
<tr>
<td>Othersc</td>
<td>1547 (61.88)</td>
<td>126 (54.08)</td>
<td>1421 (62.68)</td>
<td>0.614</td>
<td>417 (63.57)</td>
<td>17 (77.27)</td>
<td>400 (63.09)</td>
<td>0.099</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oes</td>
<td>1240 (49.60)</td>
<td>108 (46.52)</td>
<td>1132 (49.92)</td>
<td></td>
<td>422 (63.63)</td>
<td>15 (65.65)</td>
<td>407 (63.59)</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>460 (18.41)</td>
<td>36 (15.41)</td>
<td>424 (18.72)</td>
<td></td>
<td>106 (16.01)</td>
<td>1 (5.59)</td>
<td>105 (16.39)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>479 (19.14)</td>
<td>43 (18.37)</td>
<td>436 (19.22)</td>
<td></td>
<td>109 (16.43)</td>
<td>3 (13.91)</td>
<td>106 (16.53)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>321 (12.84)</td>
<td>46 (19.70)</td>
<td>275 (12.14)</td>
<td>0.26</td>
<td>26 (3.93)</td>
<td>4 (16.52)</td>
<td>22 (3.48)</td>
<td></td>
</tr>
</tbody>
</table>

Data are presented as number of patients (%).

aP value was calculated by a logistic regression model with study as a random effect.
bOthers include Black, Hispanic, Asian and race not specified.
cOthers include adenocarcinoma and histology not specified.

Includes four cases with stage 0 (in situ).
patients were more likely to be female (Adj OR = 1.61, 95% CI = 1.00–2.59), older than 66 years (Adj OR = 1.65, 95% CI = 1.05–2.59) and diagnosed with stage IV tumors than stage I/II (Adj OR = 2.55, 95% CI = 1.08–6.04). There were no differences in the association of HPV with stage IV tumors when we compared studies that reported verification of cancer diagnosis to those that did not report verification of cancer diagnosis in their publication. Similarly, for studies that reported efforts to avoid contamination, there was no difference in the association of HPV16/18 with stage IV tumors when we compared studies that reported verification of cancer diagnosis to those that did not report verification of cancer diagnosis (data not shown). When North American studies were evaluated separately, there was no association of HPV 16/18 with stage IV tumors (data not shown) were not statistically significantly different (log-rank P value = 0.263). The Kaplan–Meier curves according to HPV16/18 status showed that 2-year survival rates for HPV16/18-positive patients were 71.4 and 64.3%, respectively, whereas the 2-year and 5-year overall survival rates for HPV16/18-positive patients were 71.4 and 64.3%, respectively. The Kaplan–Meier curves according to HPV16/18 status (data not shown) were not statistically significantly different (log-rank P value = 0.263).

### Discussion

To date, >100 studies (including case reports) have analyzed HPV DNA in lung cancers (7). Although the majority of these studies have reported the prevalence of HPV16 and/or 18 DNA and a few publications have shown evidence of HPV oncogene (E6 and E7) expression (12,57), definite evidence of a causal relationship is still missing. Previously published meta-analyses of HPV in lung cancer show that HPV16 and HPV18 are the two most common genotypes detected in lung tumors worldwide (5–7); however, the characteristics of patients with these tumors were not determined. This study is the first international pooled analysis of individual data received from various research groups and was focused to define the demographic, behavioral and clinical characteristics of lung cancer cases with HPV16 and/or 18 DNA. We have also for the first time reported and compared published data from five North American studies with other parts of the world. Similar to previous reports, as expected, in our study (5–7), HPV16 and/or 18 were the most prevalent high-risk genotypes...
A significant association of HPV 16/18 DNA was also observed with race, where non-White lung cancer patients were more likely to have HPV-positive tumors compared with White lung cancer patients. There may be a number of reasons for this observed difference in HPV prevalence some of which might include differences in sexual practices or differences in susceptibility. Further investigation of the potential reasons for the racial difference in the association of HPV16/18 is needed.

Park et al. (65) evaluated HPV status in 112 non-small-cell lung cancer patients from Korea and reported that HPV16 was more common in younger lung cancer patients and HPV18 was more common in patients diagnosed with advanced stage. Although we categorized HPV16 and HPV18-positive tumors into a single group, we found no associations with age or stage, neither for gender, smoking history or histology. However, for small studies only, HPV16/18 was associated with old age (>66 years), female gender and advanced stage. It is possible that patients with metastatic disease (by definition, stage IV) could include metastases from other or mixed tumors, such as head and neck or cervical cancers (59), and the resulting misclassification could account for some of these findings. Therefore, the association with older females with stage IV tumors warrants further investigation in a single, larger and more representative study.

Although HPV-positive oropharyngeal cancers have been found to be associated with non-smokers, these tumors are also associated with patients having lower number of pack-years of cigarettes compared with heavy smokers (66). In vitro studies also show that tobacco smoke carcinogens have been shown to increase HPV16 and 18 viral synthesis as well as interact with HPV16 E6/E7 oncoproteins to increase lung cell proliferation (67). Furthermore, smoking suppresses the host innate immunity including functional and structural changes in the respiratory ciliary epithelium, lung surfactant protein and immune cells in the lung (68), thus may facilitate HPV infection and persistence in the lung in a subset of tumors. Mechanistically, it is not yet clear how HPV might promote lung carcinogenesis; however, the cooperation between HPV and tobacco smoke carcinogens for lung carcinogenesis is plausible. In this study, the association of HPV with smoking status was not conclusive. Therefore, further studies are needed to investigate the relationship between HPV and smoking status among patients with HPV-positive lung tumors.

Our findings suggest an association of HPV DNA with a small fraction of lung cancers and with heterogeneity among studies.

It is plausible that differences in the geographic location, study size, demographic and clinical makeup of each study could contribute to heterogeneity in HPV prevalence between studies. Furthermore, the majority of studies included in this pooled analysis were small (<100 cases) and may be less representative samples, thus contributing to heterogeneity in HPV prevalence. Therefore, for each geographic region, the prevalence of HPV 16/18 was adjusted for each study as well as potential confounders yet variations in HPV 16/18 prevalence remained. The highest prevalence of HPV16/18-positive lung cancers was observed in South and Central American, whereas North American and Asian studies had similar prevalence and European studies had significantly lower prevalence of HPV16/18. Similar observations were made for HPV 16 and HPV 18 independently. However, within each geographic region, distinct differences in the prevalence of HPV 16 and HPV18 genotypes were noted with the exception of Asia and North America. In North America, although the prevalence of each genotype was low, the prevalence of HPV 18 was slightly higher than that of HPV 16. In contrast, for Asia, Europe and South and Central America, if any HPV DNA detected in lung cancer, HPV 16 appeared to be more prominent than HPV18. In Europe, the prevalence of HPV18 was near zero. Factors that may be responsible for variability in HPV prevalence could be sexual behaviors of participants, genetics and possible environmental contamination as argued in the study by Koshiol et al. (54).

In addition to geographic location, the recent meta-analysis by Syrjanen et al. (7) reported that the heterogeneity between studies might be related to differences in histologic type. In this study, there were no differences in the proportion of histologic types in North America compared with the rest of the world and we observed no association of HPV 16/18 status with histology. Although South and Central American studies had the highest HPV prevalence, there were only two studies from this region; therefore, this finding will need to be confirmed by additional studies. Second to South and Central America, we observed that Asia had the highest prevalence of HPV 16 and 18.
carcinogenesis in non-cervical sites, such as the head and neck and penile cancer, where HPV-positive cancer patients have an improved overall and disease-free survival compared with HPV-negative patients (70–72). This observation needs to be verified in a much larger number of cases/events and adjustment for known prognostic factors including a history of HPV-related disease as well as treatment for previous HPV-related disease.

The availability of individual data from each study included in the analysis reduces the potential errors associated with aggregate data. However, a number of limitations are inherent in this study design as well. One of these limitations is the lack of a uniform gold standard in the HPV DNA detection methods leading to variability in HPV prevalence. Only 27 of the 81 eligible studies participate in this analysis. We have compared the characteristics of the cases in the pooled dataset and with that of the non-included studies and remain confident that both subsets are similar in characteristics. Nevertheless, possible selection bias cannot completely be ruled out. Although the overall sample size was large, some of the subgroup analyses included small numbers. Although this study has good representation of populations from the Americas, Europe as well as Asia, China, Korea and Taiwan were not as well represented. In addition, the North American dataset was limited, in that, the majority of cases were of European descent; a larger US study with broad ethnic diversity is needed to allow for further investigation related to HPV and race. Although the present analysis provides evidence in favor of an association of HPV DNA with a small subset of lung cancer, these findings need to be validated and further comprehensive analysis is needed to examine a potential causal relationship. A detailed analysis of a larger and more representative case series, including the testing of all criteria that were postulated to prove a causal involvement of HPV in carcinogenesis and excluding pulmonary metastases, needs to be performed. Most importantly, the preliminary evidence on the potential prognostic role of HPV in lung cancer needs confirmation because of its potential clinical implications, e.g. in therapy of lung cancer.

Funding
National Cancer Institute (R03 CA141483 to C.R. and P30 CA006927); Commonwealth of Pennsylvania.

Acknowledgements
The authors would like to thank Chaka Peters, Helina Tadesse and Justin Wong for their contributions in coordinating the collection of datasets included in this analysis. Special thanks to Dr Wen-Chi Chang for assisting with the translation of Chinese publications and to laboratory personnel Elizabeth Blackman for her contribution in this work. We thank Ms JoEllen Weaver and the Fox Chase Cancer Center Biosample Repository for their support of this study.

Conflict of Interest Statement: None declared.

References

Received April 1, 2013; revised December 6, 2013; accepted December 18, 2013