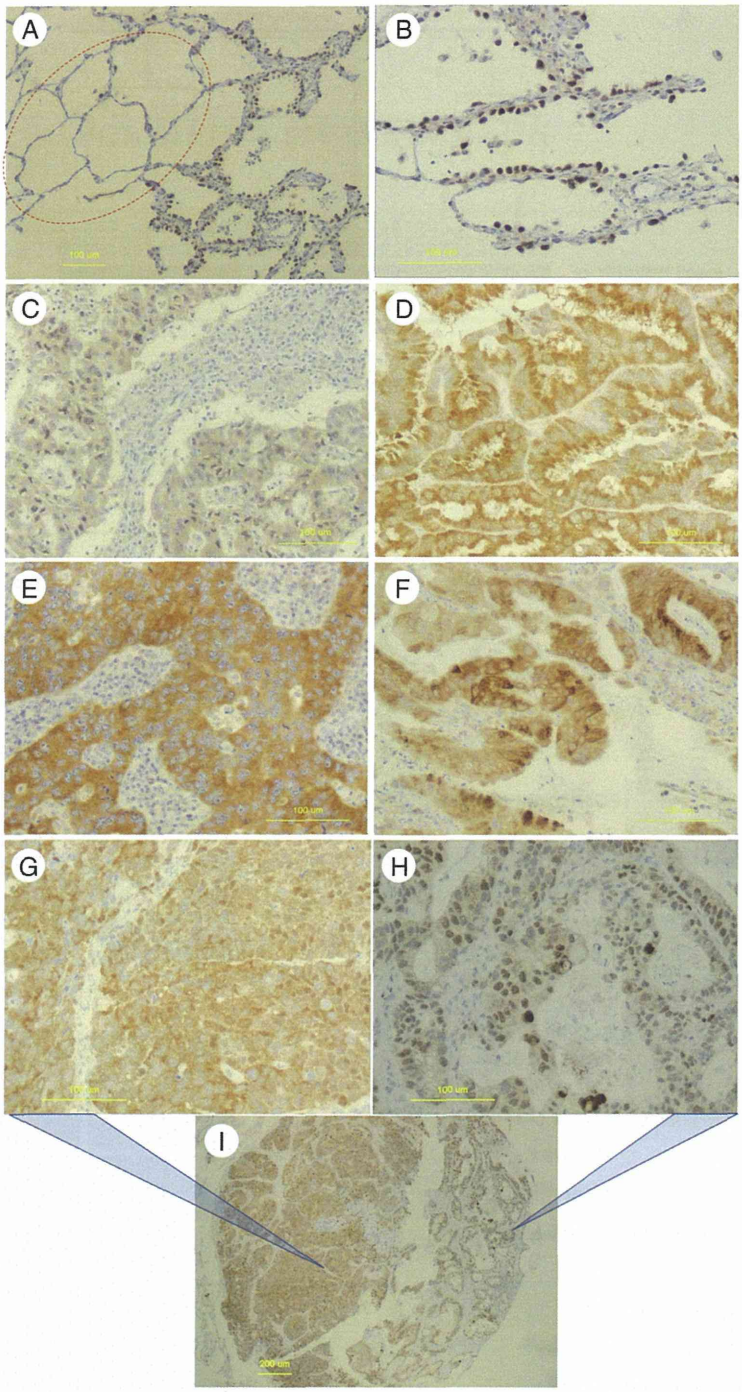

Fig. 2 PRMT5 expression in primary lung adenocarcinoma. A, The well-differentiated adenocarcinoma component (lepidic growth component) was positive for PRMT5 (right side), whereas the expression of PRMT5 was not observed in normal lung alveolar epithelia (surrounded by a red dotted line). B, A magnified view of the lepidic growth component showing the nuclear expression of PRMT5. C, Moderately differentiated adenocarcinoma (acinar and papillary adenocarcinoma) sometimes showed the nuclear expression of PRMT5. D, Moderately differentiated adenocarcinoma (acinar and papillary adenocarcinoma) sometimes showed the cytoplasmic expression of PRMT5. E, Poorly differentiated adenocarcinoma (solid adenocarcinoma) frequently showed the cytoplasmic expression of PRMT5. F, Invasive mucinous adenocarcinoma also frequently showed the cytoplasmic expression of PRMT5. G to I, Heterogeneous PRMT5 expression in lung adenocarcinoma. I shows a low-power field of invasive adenocarcinoma with mixed subtypes. The left side shows solid adenocarcinoma components, and right side shows acinar adenocarcinoma and lepidic growth components. G shows a high-power field of solid adenocarcinoma components, which showed the cytoplasmic expression of PRMT5. H shows a high-power field of acinar adenocarcinoma and lepidic growth components, which showed the nuclear expression of PRMT5. Original magnification ×100 (A), ×200 (B-H), ×40 (I).

sections by 2 pathologists (D. M. and A. G.). Each histologic subtype of lung adenocarcinoma was classified into 3 grades, by referring to the histopathologic grading described previously by Yoshizawa et al [23] with a slight modification. Detailed information on histopathologic subtyping and grading is provided in the Supplementary methods.

2.7. Immunohistochemistry and evaluation

Formalin-fixed, paraffin-embedded tumor specimens were analyzed by immunohistochemistry using antibodies to PRMT5, E-cadherin, CK7, MUC1, and TTF-1. Staining procedures and evaluation methods are given in the Supplementary methods.

2.8. Bioinformatic analyses and statistics

Details are shown in the Supplementary methods.

3. Results

3.1. *PRMT5*, a candidate gene involved in EMT, among histone methyltransferases and demethylases, depending on oligonucleotide array analysis of 40 cell lines.

We extracted expression profile data for histone methyltransferases and demethylases and examined the relative expression levels of these genes in the 40 lung cancer cell lines examined to identify histone methyltransferases and demethylases that correlated with EMT. The comprehensive data set for the expression profiles of these genes is shown in Supplementary Table S2. We then calculated the correlation coefficients of vimentin and E-cadherin for each gene and selected the genes that met the following requirements: (correlation coefficient with vimentin − correlation coefficient with E-cadherin) × ½ > 0.3 or < −0.3. The results are shown in Supplementary Table S3. We focused on *PRMT5* as the best suitable candidate correlated with EMT. We performed hierarchical cluster analysis of the 40 NSCLC cell lines, based on *PRMT5*, *TTF-1*, *MUC1*, *CK7*, *E-cadherin*, and *vimentin* gene expression. We found that *PRMT5* was correlated with *vimentin* and highly expressed in cell lines that expressed high levels of *vimentin* and low levels of *E-cadherin* and the other bronchial epithelial markers (*TTF-1*, *CK7*, and *MUC1*) (Fig. 1A). *EGFR* mutations were frequently observed in *PRMT5*-low cell lines, whereas *v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS)* mutations appeared in both *PRMT5*-high and *PRMT5*-low cell lines (Fig. 1A).

3.2. Protein expression of PRMT5 in mesenchymal-like and bronchial epithelial phenotype cell lines by Western blotting and immunocytochemistry

We performed Western blot analysis using 6 lung adenocarcinoma cell lines that contained 3 mesenchymal-like phenotypes: H522, H1651, and A549 and 3 bronchial epithelial phenotypes: HCC4006, H1650, and PC3, to compare the protein expression of PRMT5 between the 2 phenotypes. Fig. 1B summarizes the following: (i) the genetic status of *EGFR* and *KRAS* (upper panel); (ii) gene expression levels of *PRMT5*, *vimentin*, *E-cadherin*, *TTF-1*, *CK7*, and *MUC1* (middle panel); and (iii) protein expression levels of *PRMT5*, *vimentin*, *E-cadherin*, *TTF-1*, *CK7*, and *MUC1* (lower panel) in the 6 cell lines. The expression of the PRMT5 protein was higher in mesenchymal-like cell lines in which the expression of vimentin was high and that of bronchial epithelial markers was low.

We then performed immunocytochemical analysis using these 6 cell lines and found that PRMT5 expression was predominant in the cytoplasm in the mesenchymal-like phenotype, whereas it was predominant in the nucleus but faint in the cytoplasm in the bronchial epithelial phenotype with *EGFR* mutations (Fig. 1C). This result suggests that the cytoplasmic expression of PRMT5 may be associated with EMT and/or wild-type *EGFR*.

3.3. Immunohistochemical expression of PRMT5 in primary lung adenocarcinoma tissues

We used TMA sections of primary lung adenocarcinoma cases (n = 130) to examine the immunohistochemical expression patterns of PRMT5. Forty-three of 130 cases showed the high cytoplasmic expression and low nuclear expression of PRMT5, 30 showed the low cytoplasmic expression and high nuclear expression, and 53 showed low

Table 1 Correlations between PRMT5 expression levels in the cytoplasm (C) and nucleus (N) and histopathologic subtypes of primary lung adenocarcinomas

Subtypes		C+	C−
Lepidic growth component	N+	0	6
	N−	1	10
Acinar adenocarcinoma component	N+	0	4
	N−	9	12
Papillary adenocarcinoma component	N+	1	20
	N−	11	24
Solid adenocarcinoma component	N+	3	0
	N−	20	6
Invasive mucinous adenocarcinoma component	N+	0	0
	N−	2	1
	Total	47	83

Table 2 Correlations between PRMT5 expression levels in the cytoplasm (C) and nucleus (N) and histopathologic grades of primary lung adenocarcinomas

Histologic grades	C+	C−	*P*	N+	N−	*P*
Low grade	1	16	<.0001	6	11	.0444
Intermediate grade	21	60		25	56	
High grade	25	7		3	29	

cytoplasmic expression and low nuclear expression. Although both cytoplasmic and nuclear expression levels of PRMT5 were high in 4 cases, an inverse correlation was observed between the cytoplasmic and nuclear expression of PRMT5 in 130 cases ($P = .0002$). We then examined the expression levels of PRMT5 in the cytoplasm and nucleus of each histopathologic subtype. Normal alveolar epithelia were negative for PRMT5 (Fig. 2A), whereas the nuclear expression of PRMT5 was high in the well-differentiated adenocarcinoma component, that is, lepidic growth component (6 of 17, 35%), which less frequently showed the high cytoplasmic expression of PRMT5 (1 of 17, 6%) (Fig. 2A and B; Table 1). Moderately differentiated adenocarcinoma components, that is, acinar or papillary adenocarcinomas, showed the high nuclear expression of PRMT5 in 25 of 81 cases (31%) and high cytoplasmic expression of PRMT5 in 21 of 81 cases (26%) (Fig. 2C and D; Table 1). The poorly differentiated adenocarcinoma component, that is, solid adenocarcinoma with mucin, frequently showed the high cytoplasmic expression of PRMT5 (23 of 29, 79%) and less frequently showed the high nuclear expression of PRMT5 (3 of 29, 10%) (Fig. 2E and Table 1). Cytoplasmic predominance was also seen in mucinous adenocarcinoma (2 of 3, 66%) (Fig. 2F and Table 1). PRMT5 sometimes showed heterogeneous staining pattern, typically showing nuclear positive staining in well- to moderately differentiated adenocarcinoma components and cytoplasmic staining in poorly differentiated adenocarcinoma components (Fig. 2G-I).

Histologic progression to higher grade with loss of bronchial epithelial phenotype is involved in the process of EMT in our previous report [2]. Here, we examined correlations between histologic grades (low grade, intermediate grade, and high grade) and PRMT5 expression levels in the cytoplasm and nucleus, respectively (Table 2). The

Table 3 Correlation between the grades of lung adenocarcinoma and groups defined by the PRMT5 expression pattern

Histologic grades	C+ group	C−N+ group	C−N− group	*P*
Low grade	1	6	10	C+ *vs* C−N+: <.0001
Intermediate grade	21	24	36	C−N+ *vs* C−N−: .3780
High grade	25	0	7	C+ *vs* C−N−: <.0001

Table 4 Correlations between the cytoplasmic expression levels of PRMT5 and (i) clinicopathological factors, (ii) *EGFR* mutations, and (iii) the expression of bronchial epithelial markers

	Cytoplasmic PRMT5		
	High	Low	*P*
Pathologic stage [a]			.0887
Stage IA	14	38	
Stage IB-IV	32	45	
T stage			.0680
T1	16	42	
T2,T3,T4	31	41	
Nodal involvement [b]			.8858
Positive	12	21	
Negative	34	56	
Lymphatic invasion			.5064
Positive	12	17	
Negative	35	66	
Vessel invasion			.0376
Positive	24	27	
Negative	23	56	
Pleural invasion			.1489
Positive	26	35	
Negative	21	48	
Dissemination			.2835
Positive	0	2	
Negative	47	81	
Tumor size			.3640
<3 cm	16	22	
>3 cm	31	61	
Pulmonary metastasis			.5916
Positive	4	5	
Negative	43	78	
Smoking index [c]			.7410
<600	21	34	
>600	24	44	
EGFR mutations [d]			.0880
Positive	12	31	
Negative	25	31	
E-cadherin			.0138
High level	42	82	
Low level	5	1	
TTF-1			<.0001
High level	30	78	
Low level	17	5	
CK7			.0140
High level	34	74	
Low level	13	9	
MUC1 (membranous)			.0006
High level	21	62	
Low level	26	21	
MUC1 (depolarized)			.0002
High level	12	3	
Low level	35	80	

[a] Pathologic N factors were not determined for 7 cases.
[b] Six of 7 cases were more than stage IA and included 2 cases of stage IV patients with pleural dissemination.
[c] The smoking index was not determined for 7 cases.
[d] The *EGFR* genetic status was not determined for 31 cases.

prevalence of high-grade tumors was significantly higher in cases that expressed high levels of PRMT5 in the cytoplasm than in those that expressed low levels in the cytoplasm ($P < .0001$), whereas the prevalence of low-grade tumors was significantly higher in cases that expressed high levels of PRMT5 in the nucleus than in those that expressed low levels in the nucleus ($P = .0444$). However, among 4 cases that expressed high levels of PRMT5 in both the cytoplasm and nucleus, 3 (75%) showed high-grade subtype (solid adenocarcinoma), and 1 (25%) showed intermediate-grade subtype (papillary adenocarcinoma), which suggested to us that cytoplasmic PRMT5 expression would be more closely correlated with histologic grades than nuclear PRMT5 expression. Next, we classified 130 cases into 3 groups according to their PRMT5 expression patterns to verify the significance of PRMT5 cytoplasmic expression (Table 3): a C+ group, comprising cases that expressed high levels in the cytoplasm with or without expression in the nucleus (n = 47); C−N+ group, comprising cases that expressed low levels in the cytoplasm and high levels in the nucleus (n = 30); and C−N− group, comprising cases that expressed low levels in both the cytoplasm and nucleus (n = 53). We compared histologic grades and this group classification and showed that grades in the C+ group were significantly higher than those in the C−N+ and C−N− groups ($P < .0001$); however, no significant difference was observed between the C−N+ and C−N− groups ($P = .3780$) (Table 3). These results suggested that the nuclear localization of PRMT5 may not be correlated with the maintenance of a differentiated phenotype, whereas the cytoplasmic localization of PRMT5 appears to be significant in the processes of EMT.

We examined the correlation between the cytoplasmic expression of PRMT5 and clinicopathological factors, *EGFR* status, and bronchial epithelial markers (TTF-1, CK7, MUC1, and E-cadherin) (Table 4). We found that high cytoplasmic PRMT5 expression was correlated with low expression levels of TTF-1, CK7, MUC1 (membranous), and E-cadherin and high expression levels of depolarized MUC1 (Fig. 3A-E; Table 4). The prevalence of wild-type *EGFR* was slightly higher in cases that expressed high levels of PRMT5 in the cytoplasm ($P = .0880$). Fig. 3F to J shows the typical immunohistochemical expression patterns of TTF-1, CK7, E-cadherin, and MUC1 in lepidic growth components with the high nuclear expression of PRMT5.

3.4. Prognostic significance of PRMT5 expression

Survival curves based on histologic grades are shown in Fig. 4A. Low-grade cases showed the best prognosis (the 5-

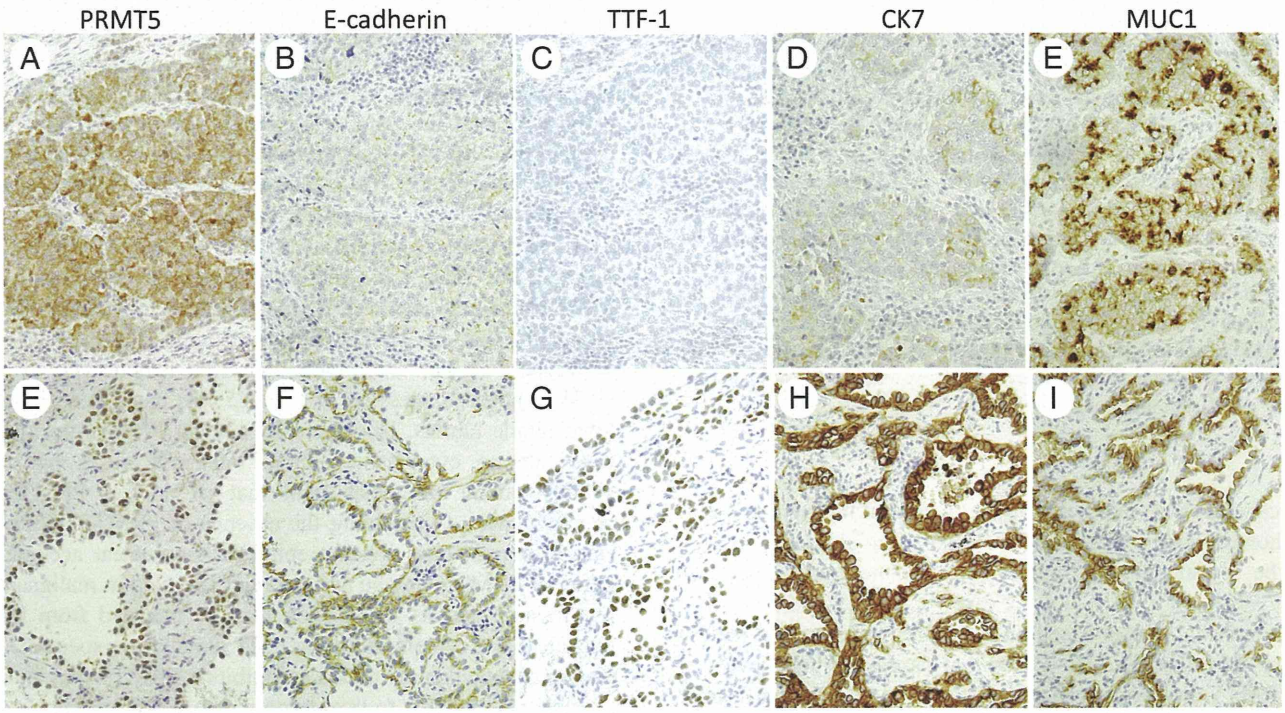

Fig. 3 Immunohistochemical expression of PRMT5, E-cadherin, TTF-1, CK7, and MUC1 in the same tissue samples. A to E, Case of solid adenocarcinoma with mucin (high-grade subtype). A shows the cytoplasmic expression of PRMT5. B shows the low level of E-cadherin expressed in the membrane. C shows the low level of TTF-1 expressed in the nucleus. D shows the low level of CK7 expressed in the membrane and cytoplasm. E shows the low level of membranous MUC1 expressed and the depolarized (cytoplasmic) expression pattern of MUC1. F to J, Case of adenocarcinoma in situ (low-grade subtype). F shows the nuclear expression of PRMT5. G shows the high level of E-cadherin expressed in the membrane. H shows the high level of TTF-1 expressed in the nucleus. I shows the high level of CK7 expressed in the membrane and cytoplasm. J shows the high level of MUC1 expressed in the apical membrane (Original magnification ×200).

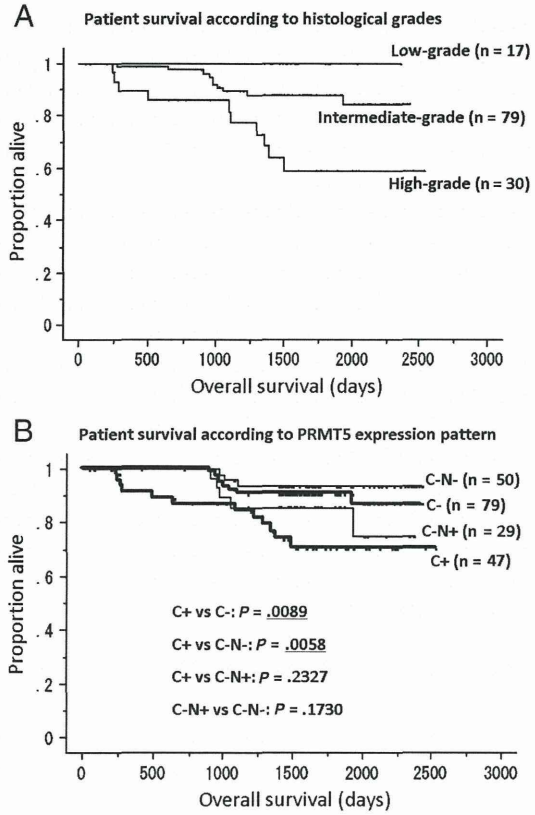

Fig. 4 Overall survival curves according to (A) histologic grades and (B) PRMT5 expression patterns. A, Patient survival curves according to histologic grades. Patients were classified into 3 groups according to their histologic grades: low grade (n = 17), intermediate grade (n = 79), and high grade (n = 30). B, Patient survival curves according to PRMT5 expression patterns. In this figure, the C+ group comprised cases that expressed high levels of PRMT5 in the cytoplasm with or without its expression in the nucleus (n = 47); C− group, cases that expressed low levels of PRMT5 in the cytoplasm with or without its expression in the nucleus (n = 79); C−N+ group, cases that expressed low levels of PRMT5 in the cytoplasm and high levels in the nucleus (n = 29); and C−N− group, cases that expressed low levels of PRMT5 in both the cytoplasm and nucleus (n = 50).

year survival rates = 100%), whereas high-grade cases showed worst prognosis (the 5-year survival rates = 58.9%). The 5-year survival rate of intermediate-grade cases was 87.9%.

Survival curves based on the PRMT5 expression pattern are shown in Fig. 4B. Cases that expressed high levels of PRMT5 in the cytoplasm (C+ group) had significantly poorer survival rates than those that expressed low levels in the cytoplasm (C− group) (P = .0089) (Fig. 4B). We also compared prognoses among the aforementioned groups: the C+ group, C+N− group, and C−N− group. The C+ group had the worst prognosis, whereas the C−N+ group had a slightly poorer prognosis than that of the C−N− group (P = .1730) (Fig. 4B).

4. Discussion

We here demonstrated the high cytoplasmic expression of PRMT5 in mesenchymal-like phenotype cell lines and that high cytoplasmic PRMT5 expression was closely related to the high-grade subtypes of primary lung adenocarcinomas with the loss of E-cadherin and other bronchial epithelial markers (TTF-1, CK7, and MUC1) and a poor prognosis.

Shilo et al [24] recently reported a correlation between cytoplasmic PRMT5 and the histologic high grade in NSCLC, except for lung adenocarcinoma. The frequency of cytoplasmic expression in adenocarcinomas was higher in our study (36%) than in the study of Shilo et al [24] (8%). These discrepancies have been attributed to differences in the evaluation methods used because we set a high value for predominant invasive lesions in histologic grading and the evaluation of PRMT5 expression (as shown in the Supplementary methods). We considered our histologic grading of the TMA cores to be accurate because the results obtained closely correlated with patient prognosis, which was consistent with the findings by Yoshizawa et al [23]. We speculated that the cytoplasmic expression of PRMT5 may contribute to a poor prognosis by promoting high-grade transformation and EMT. However, how cytosolic PRMT5 induces EMT remains unknown. PRMT5 is known to methylate splice some proteins SmD1, SmD3, and SmB/B′ in the cytoplasm, which are involved in premessenger RNA splicing [25]. PRMT5 may affect the expression of some EMT-related genes when this epigenetic cytoplasmic role is considered. EMT has also been associated with the gain of stem cell properties [26], and cytoplasmic PRMT5 of embryonic stem cells is important to maintain pluripotency through the methylation of cytosolic histone H2A during mouse development [27]. This finding also justifies the accumulation of PRMT5 in the cytoplasm during EMT. There was a slightly inverse correlation between *EGFR* mutations and cytoplasmic PRMT5 expression. We speculated that this result will reflect the high frequency of wild-type *EGFR* in lung tumors with EMT features [1].

The nuclear expression of PRMT5 was more frequent in lower grade tumors. However, the nuclear PRMT5-positive cases among cytoplasmic PRMT5-negative cases had a slightly poorer prognosis than that of nuclear PRMT5-negative cases. Considering the absence of PRMT5 expression in normal lung alveolar epithelia, the nuclear accumulation of PRMT5 may be an important first step in malignant progression, and its localization may be changed from the nucleus to the cytoplasm during EMT. Finally, we speculated that epigenetic therapy aimed at inhibiting PRMT5 may be a possible new therapy to treat tumors with EMT features.

Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.humpath.2014.02.013.

Acknowledgments

The authors thank Ms Makiko Naka Mieno for her advice on the statistical analysis.

References

[1] Matsubara D, Kishaba Y, Ishikawa S, et al. Lung cancer with loss of BRG1/BRM, shows epithelial mesenchymal transition phenotype and distinct histologic and genetic features. Cancer Sci 2013;104:266-73.

[2] Yauch RL, Januario T, Eberhard DA, et al. Epithelial versus mesenchymal phenotype determines in vitro sensitivity and predicts clinical activity of erlotinib in lung cancer patients. Clin Cancer Res 2005;11:8686-98.

[3] Matsubara D, Ishikawa S, Sachiko O, Aburatani H, Fukayama M, Niki T. Co-activation of epidermal growth factor receptor and c-MET defines a distinct subset of lung adenocarcinomas. Am J Pathol 2010;177:2191-204.

[4] Takeuchi T, Tomida S, Yatabe Y, et al. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J Clin Oncol 2006;24:1679-88.

[5] Jones S, Wang TL, Shih IM, et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. Science 2010;330:228-31.

[6] Gui Y, Guo G, Huang Y, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. Nat Genet 2011;43:875-8.

[7] Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 2011;469:539-42.

[8] Zang ZJ, Cutcutache I, Poon SL, et al. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. Nat Genet 2012;44:570-4.

[9] Dacwag CS, Ohkawa Y, Pal S, Sif S, Imbalzano AN. The protein arginine methyltransferase Prmt5 is required for myogenesis because it facilitates ATP-dependent chromatin remodeling. Mol Cell Biol 2007;27:384-94.

[10] Dacwag CS, Bedford MT, Sif S, Imbalzano AN. Distinct protein arginine methyltransferases promote ATP-dependent chromatin remodeling function at different stages of skeletal muscle differentiation. Mol Cell Biol 2009;29:1909-21.

[11] Mallappa C, Hu YJ, Shamulailatpam P, Tae S, Sif S, Imbalzano AN. The expression of myogenic microRNAs indirectly requires protein arginine methyltransferase (Prmt)5 but directly requires Prmt4. Nucleic Acids Res 2011;39:1243-55.

[12] Scoumanne A, Zhang J, Chen X. PRMT5 is required for cell-cycle progression and p53 tumor suppressor function. Nucleic Acids Res 2009;37:4965-76.

[13] Cho EC, Zheng S, Munro S, et al. Arginine methylation controls growth regulation by E2F-1. EMBO J 2012;31:1785-97.

[14] Ren J, Wang Y, Liang Y, Zhang Y, Bao S, Xu Z. Methylation of ribosomal protein S10 by protein-arginine methyltransferase 5 regulates ribosome biogenesis. J Biol Chem 2010;285:12695-705.

[15] Zhou Z, Sun X, Zou Z, et al. PRMT5 regulates Golgi apparatus structure through methylation of the golgin GM130. Cell Res 2010;20: 1023-33.

[16] Kim JM, Sohn HY, Yoon SY, et al. Identification of gastric cancer-related genes using a cDNA microarray containing novel expressed sequence tags expressed in gastric cancer cells. Clin Cancer Res 2005;11:473-82.

[17] Pal S, Baiocchi RA, Byrd JC, Grever MR, Jacob ST, Sif S. Low levels of miR-92b/96 induce PRMT5 translation and H3R8/H4R3 methylation in mantle cell lymphoma. EMBO J 2007;26:3558-69.

[18] Wang L, Pal S, Sif S. Protein arginine methyltransferase 5 suppresses the transcription of the RB family of tumor suppressors in leukemia and lymphoma cells. Mol Cell Biol 2008;28:6262-77.

[19] Powers MA, Fay MM, Factor RE, Welm AL, Ullman KS. Protein arginine methyltransferase 5 accelerates tumor growth by arginine methylation of the tumor suppressor programmed cell death 4. Cancer Res 2011;71:5579-87.

[20] Gu Z, Li Y, Lee P, Liu T, Wan C, Wang Z. Protein arginine methyltransferase 5 functions in opposite ways in the cytoplasm and nucleus of prostate cancer cells. PLoS One 2012;7:e44033.

[21] Gu Z, Gao S, Zhang F, Wang Z, Ma W, Davis RE. Protein arginine methyltransferase 5 is essential for growth of lung cancer cells. Biochem J 2012;446:235-41.

[22] Matsubara D, Kanai Y, Ishikawa S, et al. Identification of CCDC6-RET fusion in a human lung adenocarcinoma cell line, LC-2/ad. J Thorac Oncol 2012;7:1872-6.

[23] Yoshizawa A, Motoi N, Riely GJ, et al. Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases. Mod Pathol 2011;24:653-64.

[24] Shilo K, Wu X, Sharma S, et al. Cellular localization of protein arginine methyltransferase-5 correlates with grade of lung tumors. Diagn Pathol 2013;8:201.

[25] Friesen WJ, Paushkin S, Wyce A, et al. The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. Mol Cell Biol 2001;21:8289-300.

[26] Abell AN, Jordan NV, Huang W, et al. MAP3K4/CBP-regulated H2B acetylation controls epithelial-mesenchymal transition in trophoblast stem cells. Cell Stem Cell 2011;8:525-37.

[27] Tee WW, Pardo M, Theunissen TW, et al. Prmt5 is essential for early mouse development and acts in the cytoplasm to maintain ES cell pluripotency. Genes Dev 2010;24:2772-7.

Carcinogenesis Advance Access published February 26, 2014
Carcinogenesis vol.00 no.00 p.1 of 9, 2014
doi:10.1093/carcin/bgu038
Advance Access publication February 12, 2014

HPV-associated lung cancers: an international pooled analysis

Camille Ragin[1,*,†], Monisola Obikoya-Malomo[1,†], Sungjin Kim[2], Zhengjia Chen[2], Rafael Flores-Obando[3], Denise Gibbs[1], Chihaya Koriyama[4], Francisco Aguayo[5], Jill Koshiol[6], Neil E.Caporaso[7], Giovanna E.Carpagnano[8], Marco Ciotti[9], Hirotoshi Dosaka-Akita[10], Masashi Fukayama[11], Akiteru Goto[11], Demetrios A.Spandidos[12], Vassilis Gorgoulis[13], Daniëlle A.M.Heideman[14], Robert A.A.van Boerdonk[14], Kenzo Hiroshima[15], Reika Iwakawa[16], Nikolaos G.Kastrinakis[13], Ichiro Kinoshita[10], Suminori Akiba[4], Maria T.Landi[7], H.Eugene Liu[17], Jinn-Li Wang[17], Ranee Mehra[18], Fadlo R.Khuri[19], Wan-Teck Lim[20], Taofeek K.Owonikoko[19], Suresh Ramalingam[19], Emmanuela Sarchianaki[12], Kari Syrjanen[21,22], Ming-Sound Tsao[23], Jenna Sykes[24], Siew Wan Hee[25], Jun Yokota[16], Apostolos Zaravinos[12] and Emanuela Taioli[26]

[1]Cancer Prevention and Control Program, Fox Chase Cancer Center, Philadelphia, PA, USA, [2]Biostatistics & Bioinformatics, Winship Cancer Institute, Emory University, Atlanta, GA, USA, [3]Department of Cell Biology, State University of New York, Downstate Medical Center, Brooklyn, NY, USA, [4]Department of Epidemiology and Preventative Medicine, Kagoshima University, 8-35-1 Sakuragaoka, Kagoshima, Japan, [5]Virology Program, Instituto de Ciencias Biomédicas (ICBM), Faculty of Medicine, University of Chile, Santiago, Chile, [6]Infections and Immunoepidemiology Branch and [7]Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA, [8]Department of Medical and Surgical Sciences, Faculty of Medicine, University of Foggia, Foggia, Italy, [9]Laboratory of Molecular Virology, Foundation Polyclinic Tor Vergata, Rome, Italy, [10]Department of Medical Oncology, Hokkaido University, Graduate School of Medicine, North 15, West 7, Kita-ku, Sapporo 060-8638, Japan, [11]Department and Cellular and Organ Pathology, Graduate School of Medicine, Akita University, 1-1-1 Hondo, Akita 010-8543, Japan, [12]Faculty of Medicine, Laboratory of Clinical Virology, 5D10, University of Crete, Vassilika Voutes, 71110 Heraklion, Crete, Greece, [13]Laboratory of Histology & Embryology, School of Medicine, University of Athens, Athens, Greece, [14]Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands, [15]Department of Pathology, Tokyo Women's Medical University Yachiyo Medical Center, 477-96 Owada-Shinden, Yachiyo-shi, Chiba 276-8524, Japan, [16]Division of Multistep Carcinogenesis, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104-0045, Japan, [17]Graduate Institute of Clinical Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan, [18]Department of Medical Oncology and Developmental Therapeutics Program, Fox Chase Cancer Center, Philadelphia, PA, USA, [19]Department of Hematology and Medical Oncology, Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA, USA, [20]Department of Medical Oncology, National Cancer Center Singapore, 11 Hospital Drive, Singapore 169610, Republic of Singapore, [21]Department of Oncology & Radiotherapy, Turku University Hospital, Turku, Finland, [22]Fundação Pio XII-Hospital de Câncer de Barretos, Barretos, Brazil, [23]Department of Pathology and [24]Department of Biostatistics, University Health Network, Princess Margaret Cancer Centre, Toronto, Ontario M5G2M9, Canada, [25]Division of Health Sciences, Warwick Medical School, The University of Warwick, Coventry, UK and [26]Department of Population Health, North Shore Long Island Jewish Health System, Hofstra Medical School, The Feinstein Institute for Medical Research, 350 Community Drive, Manhasset, NY, USA

*To whom correspondence should be addressed. Cancer Prevention and Control Program, Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111, USA. Tel: +1 215-728-1148; Fax: +1 215 214 1622; Email: camille.ragin@fccc.edu

Abbreviations: AdjPr, adjusted prevalence; CI, confidence interval; FFPE, formalin fixed paraffin embedded; HPV, human papillomavirus; OR, odds ratio; SCC, squamous cell carcinoma.

†These authors contributed equally to this work and should be considered co-first authors.

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* = 3249 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, 4 and 3%, 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 among female 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 (1,2). 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 (4), ~22% of the cases analyzed contained HPV. This meta-analysis was updated in 2007 and reported a prevalence of 24.5%, but considerable heterogeneity was observed between studies (5). In our meta-analysis in 2009 (6), we also identified a wide variation in the prevalence of HPV in lung cancer tissues, despite the fact that only studies utilizing PCR-based methods for HPV detection were included in order to reduce the heterogeneity between studies. Nevertheless, HPV16 and 18 were the two most common genotypes detected in lung tumors, and a higher prevalence of HPV16 and 18 was noted in Asian populations compared with European populations. In 2012, the most recent meta-analysis was published and reported that the variability of HPV prevalence may be due to differences in geographical study origin and histological types of lung cancer rather than the method of HPV detection (7). However, meta-analyses findings to date have limitations because important characteristics such as

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 (8,9,11,13,14,26,34,38,45–50,53–60). 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

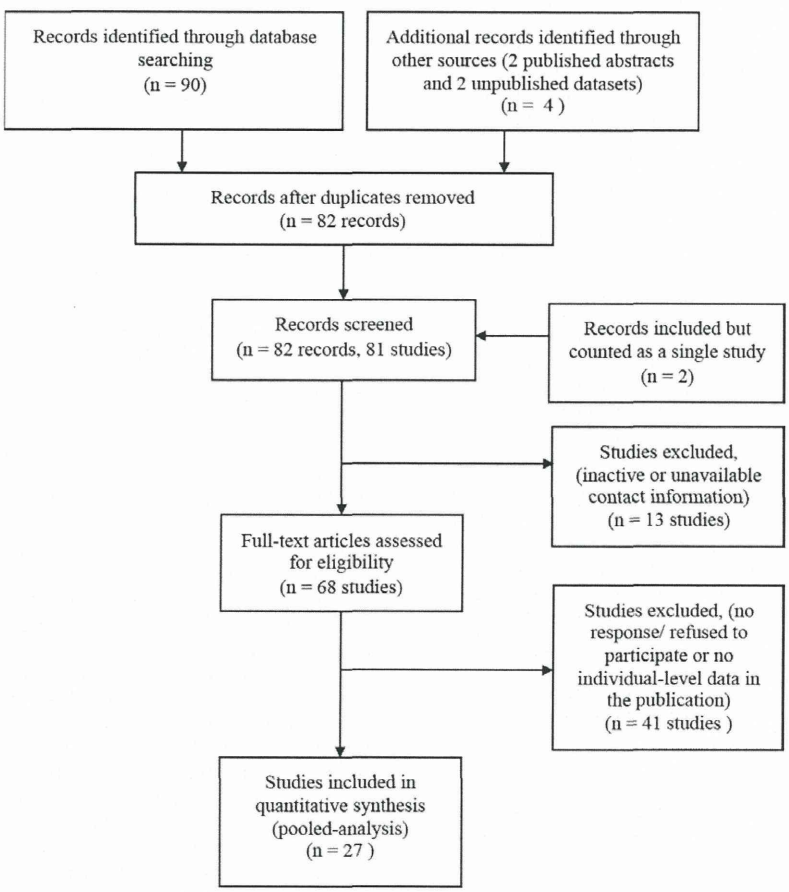

Fig. 1. Flow diagram summarizing the identification and selection process of eligible studies for inclusion in the pooled analysis.

(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 (13,57,59,60). 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 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 A.Zaravinos *et al.* (unpublished results; $n = 17$), 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 and 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 and adenocarcinoma) diagnosed at Fox Chase Cancer Center (FCCC) from 1993 to 2010. FFPE tissues and corresponding demographic and clinical information were obtained from the FCCC Biosample Repository. DNA from tumor tissue was extracted and tested for HPV, and physical status was determined for all HPV16-positive tumor tissues. Data from Pillai *et al.* (31) ($n = 208$), 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 amplicon 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 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 9 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 positive samples were 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 LiPA 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 (6,11,40,43,44,54,70) 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 QIAmp DNA minikit (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, pH 8.0, in a 1.5 ml eppendorf tube. Samples were then incubated at 60°C for 2 days. Fresh proteinase K was added three times daily. 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. Specimens 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 subtypes. The samples were initially examined for the presence of non-type-specific HPV DNA using the general HPV primers, GP5+/GP6+. The following PCR amplification cycling conditions were used: 94°C for 4 min; 40 cycles, 94°C for 1 min, 40°C for 2 min and 72°C for 1–5 min. The last cycle was extended by a 4 min elongation at 72°C. Appropriate negative and positive controls were included in each PCR reaction in order to exclude contamination events and to establish the specificity of primer-directed amplification. Recombinant plasmids carrying HPV type-specific sequences served as positive controls for HPV6, -11, -16, -18 and -33 genomes detection. For the general screening of HPV DNA, HeLa 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 5× Green GoTaq reaction buffer, 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleotide triphosphate (dNTPs), 0.6 U 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′-AACGAAGTATCC TCTCCTGAAATTATTAG-3′ (3361–3389 nt); reverse 5′-CCAAGGCG ACGGCTTTG-3′ (3427–3443 nt), as well as 81 bp of the *E6* gene, primers forward 5′-GAGAACTGCAA TGTTTCAGGACC-3′ (94–116 nt); reverse 5′-TGTATAAGTTGTTTGCAGC TCTGTGC-3′ (150–169 nt), in the presence of specific hybridization probes for *E2*-(FAM-CACCC CGCCGCGACCCATA-TAMRA) (3406–3424 nt) and *E6*-(FAM-C AGGAGCGAC CCAGAAAGTTACCACAGTT-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 15 s and 60°C for 60 s for a total of 40 cycles. A standard 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 DNA 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) with a significant 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 respect. 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

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

Study	Tissue source	Method of HPV detection	Histological type	HPV types detected
Asia (N = 1312)				
Baba *et al.* (N = 77) (49)	FFPE	PCR (SPF10, Inno-LIPA)	AC, SCC	HPV 6, 16, 18, 33
Kinoshita *et al.* (N = 34) (57)	FF	PCR (TS: HPV 16, 18, 33)	AC, SCC, ASq[a], LCC[a], SmCC[a]	HPV 18
Iwakawa *et al.* (N = 275) (53)	FF	PCR (TS: HPV 16, 18, 33)	AC	HPV negative
Hiroshima *et al.* (N = 49) (52)	FFPE	PCR (TS: HPV 16, 18, 33)	AC, SCC	HPV 16
Goto *et al.* (N = 296) (51)	FFPE	PCR GP5+/GP6+	AC, SCC, CSrc[a]	HPV 6, 11, 16, 18
Lim *et al*, 2009 (N = 99) (56)	FFPE	ISH (HPV6, 11, 16, 18, 31, 33, 35, 45, 51, 52, 56, 58, 66)	AC	HPV negative
Miyagi *et al.* (N = 176) (8,9)	FFPE	PCR (TS: HPV 6, 11, 16, 18)	AC, SCC	HPV 6, 11, 16, 18
Hirayasu *et al.* (N = 73) (26)	FFPE	PCR (TS: HPV 6, 11, 16, 18)	SCC	HPV 6, 16, 18
Wang *et al.* (N = 210) (58)	FFPE, Bx[b], PE[b]	PCR (MY09/MY11 and TS: HPV 16, 18)	AC	HPV16, HPV18
Tsuhako *et al.* (N = 23) (47)	FFPE	PCR (TS: HPV 6, 11, 16, 18)	ASq	HPV 6, 11, 16, 18
Adjusted HPV 16/18 prevalence (95% CI)[c] = 4.60% (3.48, 5.73)				
Adjusted HPV 16 prevalence (95% CI)[c] = 1.49% (0.86, 2.11)				
Adjusted HPV 18 prevalence (95% CI)[c] = 1.09% (0.66, 1.52)				
Europe (N = 1100)				
Syrjanen *et al.* (N = 131) (14)	FFPE	ISH (HPV6, 11, 16, 18, 30)	SCC	HPV 16, X[d]
Syrjanen *et al.* (N = 77) (55)	FFPE	PCR (MY09/MY11/GP5+/GP6+)	AC, SCC, other	HPV 6, 16
Giuliani *et al.* (N = 77) (13)	FFPE and FF	PCR (MY09/MY11/GP5+)	AC, SCC, ASq[a], LCC[a], SmCC[a], NSCLC[a], other[a]	HPV 6, 16, 18, 31, 53,
Gorgoulis *et al.* (N = 68) (11)	FFPE and FF	PCR (MY09/MY11/GP5+/GP6+)	AC, SCC, LCC[a]	HPV negative
Zaravinos *et al.* (unpublished results, N = 17)	FFPE	PCR (GP5+/GP6+)	AC[e], SCC[e], LCC[a], NSCLC[a]	HPV negative
Koshiol *et al.* (N = 399) (54)	FFPE	PCR (TS: HPV 16, 18; SPF10 primers (n = 92)	AC, SCC, LCC, SmCC, other	HPV 16, 18
Carpagnano *et al.* (N = 89) (34)	FFPE and BB[b]	PCR (consensus HPV primers)	AC, SCC, SmCC	HPV 16, 30, 31, 39,
Nuorva *et al.* (N = 22) (46)	FFPE	PCR (MY09/MY11)	AC	HPV 6, 11, 16, 18, 31, 33
van Boerdonk *et al.* (N = 220) (59)	FFPE	PCR (GP5+/GP6+)	AC, SCC, LCC, NSCLC[e]	HPV negative
Adjusted HPV 16/18 prevalence (95% CI)[f] = 3.03% (2.76, 3.30)				
Adjusted HPV 16 prevalence (95% CI)[f] = 2.94% (2.68, 3.21)				
Adjusted HPV 18 prevalence (95% CI)[f] = 0.82%(0.73, 0.92)				
South and Central America (N = 105)				
Castillo *et al.* (N = 36) (50)	FFPE	PCR (GP5+6+) and SB	AC, SCC, SmCC[e]	HPV 16, 18, 33
Aguayo *et al.* (N = 69) (48)	FFPE	PCR (GP5+/GP6+) and SB	AC, SCC	HPV 6, 16, 18, 31, 45
Adjusted HPV 16/18 prevalence (95% CI)[f]= 21.90% (19.61, 24.20)				
Adjusted HPV 16 prevalence (95% CI)[f] = 19.18% (16.88, 21.49)				
Adjusted HPV 18 prevalence (95% CI)[f] = 7.78% (6.61, 8.95)				
North America (N = 732)				
Taioli *et al.* (unpublished results, N = 69)	FFPE	PCR (PGMY09/PGMY11/GP5+/6+)	AC, SCC, LCC[a], NSCLC[a], other[e]	HPV 51, X
Pillai *et al.* (N = 208) (31)	FFPE	PCR (SPF10, Inno-LIPA)	AC, SCC, ASq[e], LCC, other[e]	HPV 6, 16, 18, 39, 53, X
Mehra *et al.* (N = 62) (32)	FFPE	PCR (SPF10, Inno-LIPA)	AC, SCC[e], ASq[a], LCC[a],	HPV 6, 16, 18, 52, 53, 44, 68, 39, 74, 82, X
Joh *et al.* (N = 29) (38)	FF	PCR (GP5+/GP6+)	AC, SCC[e], LCC[a], NSCLC[a]	HPV 11, 16
Yanagawa *et al.* (N = 330) (60)	FFPE	PCR (GP5+/GP6+) and ISH	AC, SCC	HPV negative
Bohlmeyer *et al.* (N = 34) (45)	FFPE	PCR (MY09/MY11)	SCC	HPV 18
Adjusted HPV 16/18 prevalence (95% CI)[f] = 3.78% (3.35, 4.22)				
Adjusted HPV 16 prevalence (95% CI)[f] = 2.03% (1.68, 2.39)				
Adjusted HPV 18 prevalence (95% CI)[f] = 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.

[a]Number of cases ≤6.

[b]Cancer cells present and confirmed by pathologist.

[c]Study is included in the model as a random effect.

[d]Type not defined; HPV+ (for the HPV6/11/16/18/30 panel).

[e]Number of cases 7–10.

[f]Study is included as a covariate in the model.