decision-making process for patients with advanced NSCLC. Reports on the association of SNPs in several other DNA repair genes with prognosis of patients with NSCLC who received chemotherapy also suggested their associations with the outcome of the patients.[7,9-14] However, sample sizes were small (50 to 250 patients), and only four to 15 genomic polymorphisms were investigated in those studies. In addition, the data in each trial were not confirmed by an independent validation set. Therefore, clinical importance of these SNPs still remains unclear.

We previously searched for nonsynonymous (ie, associated with amino acid changes) SNPs in 36 DNA repair genes involved in diverse intracellular processes that maintain genome integrity and identified 29 SNPs in 26 DNA repair genes, whose minor allele frequencies were more than 5% in Japanese patients[15] (Table 1). Thus, in this study, we conducted a single-hospital–based retrospective analysis of 640 patients with NSCLC to elucidate associations of these 29 SNPs and the ERCC1 SNP above[8] with the patients' outcome after platinum-based doublet chemotherapy. To minimize type I errors, the significance of candidate SNPs picked up by the first discovery set were validated by using the second independent validation set. We chose the response evaluated by the Response Evaluation Criteria in Solid Tumors (RECIST)[16] as the primary end point of outcome to search for predictive factors for the primary effect of chemotherapy.

Table 1. 30 SNPs in DNA Repair Genes

Pathway	Gene	SNP (rs number)	Amino Acid/ Nucleotide Change	Minor Allele Frequency				
				Japanese*	Japanese†	Chinese†	European†	African†
29 Nonsynonymous SNPs (associated with amino acid change)								
Base excision repair								
	PARP1	rs1805412	Val762Ala	0.40	0.46	0.48	0.17	0.01
		rs1136471	Lys940Arg	0.05	—	—	—	—
	APEX	rs1130409	Asp148Glu	0.38	0.32	0.46	0.51	0.28
	MBD4	rs140693	Glu346Lys	0.35	0.41	0.27	0.00	0.03
	MTH1	rs4866	Val83Met	0.09	—	—	—	—
	OGG1	rs1052133	Ser326Cys	0.48	0.52	0.50	0.22	0.14
	XRCC1	rs1799782	Arg194Trp	0.33	0.28	0.24	0.09	0.08
		rs25489	Arg280His	0.09	—	—	0.03	0.03
		rs25487	Arg399Gln	0.25	0.28	0.27	—	0.10
Nucleotide excision repair								
	XPG	rs17655	His1104Asp	0.42	0.48	0.56	0.73	0.46
	CSB	rs2228528	Gly399Asp	0.45	0.46	0.40	0.19	0.22
	XPC	rs2228001	Lys939Gln	0.40	0.34	0.38	0.41	0.26
	XPD	rs13181	Lys751Gln	0.05	0.08	0.06	0.33	0.18
Mismatch repair								
	MLH3	rs175080	Pro844Leu	0.18	0.14	0.13	0.43	0.41
	MSH3	rs26279	Thr1045Ala	0.24	0.22	0.37	0.22	0.40
	MSH6	rs1042821	Gly39Glu	0.32	—	—	—	—
DNA double-strand break repair								
	BRCA2	rs144848	Asn372His	0.22	0.31	0.21	0.29	0.13
	SNM1	rs3750898	His317Asp	0.26	0.26	0.10	0.27	0.74
	NBS1	rs1805794	Gln185Glu	0.50	0.46	0.49	0.28	0.16
	XRCC3	rs861539	Thr241Met	0.09	0.15	0.07	0.42	0.24
DNA damage response								
	TP53	rs1042522	Arg72Pro	0.33	0.23	0.49	0.41	0.67
DNA polymerase								
	POLD1	rs1726801	Arg119His	0.20	0.22	0.18	0.06	0.35
	POLI	rs5305	Thr731Ala	0.25	0.28	0.29	0.26	0.00
	REV1	rs3087386	Phe257Ser	0.33	0.30	0.37	0.50	0.30
	POLZ	rs462779	Thr1224Ile	0.35	0.43	0.49	0.82	0.38
Other pathways								
	BLM	rs28384991	Thr298Met	0.09	—	—	—	—
	FANCA	rs2239359	Ser501Gly	0.17	0.16	0.21	0.62	0.33
	FANCG	rs2237857	Thr297Ile	0.12	0.13	0.01	0.00	0.14
	WRN	rs1346044	Cys1367Arg	0.09	0.07	0.08	0.23	0.15
One synonymous SNP (not associated with amino acid change)								
Nucleotide excision repair								
	ERCC1	rs11615	C118T	—	0.29‡	0.22	0.65	0.02

Abbreviation: SNP, single nucleotide polymorphism.
*Frequency in Japanese determined by Sakiyama et al.[15]
†Frequency determined by the HapMap project.
‡Frequency in Japanese (T. Kohno, unpublished data).

PATIENTS AND METHODS

Selection of Study Population and Acquisition of Clinical Information

In total, 987 patients with NSCLC with clinical stages IIIA, IIIB, and IV tumors, who had not received prior platinum-based chemotherapy, were given platinum-based chemotherapy at the National Cancer Center Hospital in Tokyo, Japan, from 2000 to 2008 (Fig 1A). Clinical information was obtained by attending physicians and nurses. Of the 987 patients, 640 were eligible for the study according to the following criteria: they were not indicated for definitive chemoradiotherapy; they received a platinum-based doublet but not single or triplet chemotherapy; and their tumor response was evaluable according to RECIST[16] on the basis of data from computed tomography scans. However, those with clinical or radiologic evidence of early progression, such as emergence of new lesions, were included as patients with progressive disease (PD) in the analysis, even when unaccompanied by corresponding computed tomography scans, according to the definition in RECIST.[16] All patients were Japanese and were diagnosed with adenocarcinoma (ADC), squamous cell carcinoma (SQC), or other histologic types of NSCLC according to WHO classification[17,18] (Table 2).

Written informed consent was obtained from all patients for the use of blood cells for the analysis of genetic polymorphisms in association with clinical findings, including response to chemotherapy. Thus, 201 patients in the discovery set received therapy from 2000 to 2004, and 439 patients in the validation set received therapy from 2004 to 2008. Information on response in a subset of patients was obtained from the data in clinical trials conducted at the National Cancer Center Hospital.[3,19,20] This study was approved by the institutional review boards of the National Cancer Center. Smoking habit was recorded by pack-years. Patients with pack-years > 0 were defined as smokers, including both former and current smokers. Patients who report no smoking history (ie, pack-years = 0) were defined as never-smokers.

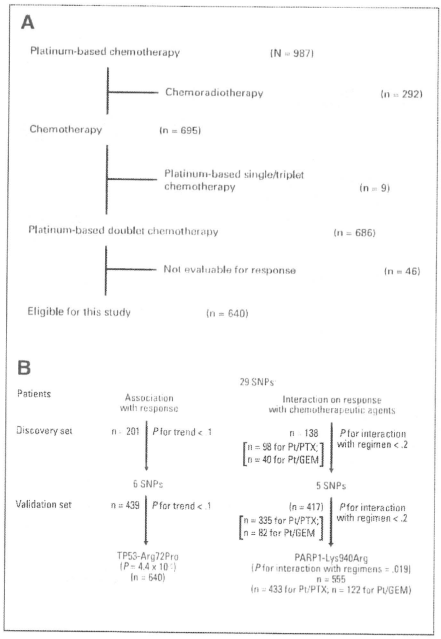

Fig 1. Patients and strategy. (A) Selection of eligible cases. (B) A two-phase screening of single nucleotide polymorphisms (SNPs) associated with responses to platinum-based doublet chemotherapy. Pt, platinum; PTX, paclitaxel; GEM, gemcitabine.

Chemotherapy

Patients were treated with one of the following regimens: (1) paclitaxel 200 mg/m^2 followed by cisplatin 80 mg/m^2, carboplatin at a dose calculated to produce an area under the serum concentration-time curve of 6.0 min · mg/mL, or nedaplatin 100 mg/m^2 on day 1, repeated every 3 weeks; (2) docetaxel 60 mg/m^2 followed by cisplatin 80 mg/m^2 on day 1, repeated every 3 weeks; (3) vinorelbine 25 mg/m^2 on days 1 and 8 and cisplatin 80 mg/m^2 on day 1, repeated every 3 weeks; (4) gemcitabine 1,000 mg/m^2 on days 1 and 8 and cisplatin 80 mg/m^2 or carboplatin to area under the serum concentration-time curve of 5.0 min · mg/mL on day 1, repeated every 3 weeks; or (5) irinotecan 60 mg/m^2 on days 1, 8, and 15 and cisplatin 80 mg/m^2 on day 1, repeated every 4 weeks. Each treatment was repeated for two or more cycles unless the patient met the criteria for PD or experienced unacceptable toxicity. Chemotherapy dosage was modified by toxicities in subsequent courses.

Genetic Analysis

A 20 mL whole-blood sample was obtained from each patient, and genomic DNA was extracted from whole-blood cells.[15] Genotyping for 30 SNPs in 27 genes was performed by pyrosequencing or TaqMan methods as previously described.[15,21]

Statistical Analysis

Patients were divided into two categories: responders were those with complete response and partial response, and nonresponders were those with stable disease and PD. Odds ratios (ORs) and 95% CIs for the response (ie, responder *v* nonresponder) according to genotypes were calculated as a measure of difference in the response rate against therapy. ORs were calculated by adjusting sex (male *v* female), age (increase by 10 years), performance status (0 *v* 1 to 2), smoking status (never-smoker *v* smoker), stage (III *v* IV), and chemotherapy (platinum plus a DNA-damaging agent *v* platinum plus a microtubule-targeting agent) by using an unconditional logistic regression analysis.[22] *P* value by the trend test was also calculated by using an unconditional logistic regression analysis under the same adjustments as above. Differences in the response between two chemotherapeutic regimens according to genotypes were examined by calculating *P* values for interaction with the regimens on the trend of OR for response.

A two-phase screening was used to search for SNPs associated with the response to chemotherapy (Fig 1B). In the first phase, 29 SNPs were examined for associations with the response and differences in the association according to regimens in 201 and 138 patients (for whom paclitaxel or gemcitabine therapy was used, respectively) in the discovery set. In the second phase, SNPs that showed *P* values < .1 by the trend test for association with the response and *P* values < .2 for interaction with the regimen were subjected to genotyping of 439 and 417 patients (for whom paclitaxel or gemcitabine was used, respectively) in the validation set. SNPs that showed *P* values < .1 for association with the response and *P* values < .2 for interaction with the regimen in patients in the validation set were further subjected to analysis in all 640 and 555 patients, respectively. Progression-free survival (PFS) was defined as the period from the first day of chemotherapy to the date of documentation of disease progression by RECIST and overall survival (OS) was defined as the period from the first day of chemotherapy to death. Hazard ratios (HRs) for PFS and OS and 95% CIs were calculated by using multivariate Cox proportional hazards models with adjustment for sex, age, histology, performance status, smoking status, clinical stage, and treatment as above. Statistical analyses were performed using JMP version 8.0 software (SAS Institute, Cary, NC). A level of $P < .05$ was considered significant, whereas a level of $P < .10$ was considered marginal.

Table 2. Patient Characteristics

Variant	All			Discovery Set*		Validation Set	
	No.	%	95% CI	No.	%	No.	%
Total patients	640			201		439	
Age, years							
Mean	57.9			57.2		58.2	
Range	22-78			22-78		26-74	
± Standard deviation	9.2			10.0		9.1	
Sex							
Male	402	62.8		136	67.7	266	60.6
Female	238	37.2		65	32.3	173	39.4
ECOG performance status							
0	218	34.1		46	22.9	172	39.2
1	402	62.8		153	76.1	249	56.7
2	20	3.1		2	1.0	18	4.1
Histologic cell type							
Adenocarcinoma	549	85.8		167	83.1	382	87.0
Squamous cell carcinoma	84	13.1		34	16.9	50	11.4
Others	7	1.1		0	0.0	7	1.6
Smoking habit							
Never-smoker	233	36.4		74	36.8	159	36.2
Smoker	407	63.6		127	63.2	280	63.8
Pack-years of smokers							
Mean	46.3			45.9		46.5	
± Standard deviation	29.6			29.4		29.7	
Stage							
III	172	26.9		60	29.9	112	25.5
IIIA	24	3.8		12	6.0	12	2.7
IIIB	148	23.1		48	23.9	100	22.8
IV	468	73.1		141	70.1	327	74.5
Tumor response							
Responder	231	36.1		74	36.8	157	35.8
CR	4	0.6		0	0.0	4	0.9
PR	227	35.5		74	36.8	153	34.9
Non-responder	409	63.9		127	63.2	282	64.2
SD	232	36.3		70	34.8	162	36.9
PD	177	27.7		57	28.4	120	27.3
Platinum-based regimens							
Platinum + a microtubule-targeted agent	476	74.4		129	64.2	347	79.0
Paclitaxel†	433	67.7		98	48.8	335	76.3
Docetaxel‡	8	1.3		2	1.0	6	1.4
Vinorelbine§	35	5.5		29	14.4	6	1.4
Platinum + a DNA-damaging agent	164	25.6		72	35.8	92	21.0
Gemcitabine¶	122	19.1		40	19.9	82	18.7
Irinotecan‖	42	6.6		32	15.9	10	2.3
PFS, median month							
Platinum + Paclitaxel	4.7		4.2 to 5.3				
Platinum + Gemcitabine	4.6		3.8 to 5.4				
Responder	6.1		5.7 to 6.4				
Nonresponder	3.0		2.7 to 3.3				

Abbreviations: ECOG, Eastern Cooperative Oncology Group; CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; PFS, progression-free survival.

*Genotype for 29 nonsynonymous DNA repair gene single nucleotide polymorphisms were determined by Sakiyama et al.[15]

†Cisplatin or carboplatin or nedaplatin + paclitaxel.

‡Cisplatin + docetaxel.

§Cisplatin + vinorelbine.

¶Cisplatin or carboplatin + gemcitabine.

‖Cisplatin + irinotecan.

RESULTS

Association of a TP53-Arg72Pro SNP With Response to Platinum-Based Doublet Chemotherapy

Among 987 patients with NSCLC who were treated with platinum-based chemotherapy, 640 were eligible for this study (Fig 1A). Characteristics of these patients are summarized in Table 2. Genotypes for the 29 nonsynonymous SNPs in 26 DNA repair genes had been determined in 201 of the 640 patients in our previous study[15] (the discovery set in Table 2). Therefore, associations of these 29 SNPs with responses to chemotherapy were first investigated in these patients (Fig 1B). Six of the 29 SNPs fulfilled the criteria described above ($P < .1$ by the trend test; Appendix Table A1, online only); thus, they were further genotyped in the remaining 439 patients (the validation set in Table 2). Only one SNP, TP53-Arg72Pro, reproducibly showed an association that met the criteria ($P < .1$; Fig 1B and Appendix Table A1). In the analysis of all 640 patients, TP53-72Pro, the minor allele, was associated with a better response ($P = 9.5 \times 10^{-5}$ by the trend test; Table 3), and response rates increased according to the increase in the number of minor alleles (Fig 2A). Minor allele homozygotes showed a better response rate (54.3%) than major allele homozygotes (29.1%; $P = 4.4 \times 10^{-5}$). The association remained significant after Bonferroni correction (ie, $< 0.05/29 = 1.7 \times 10^{-3}$). Response rates of heterozygotes and homozygotes for the TP53-72Pro allele were higher in SQC than in ADC (Fig 2A and Table 3).

In the Cox proportional hazard model, minor allele homozygotes showed a significantly longer PFS than major allele homozygotes (HR, 0.85; 95% CI, 0.74 to 0.98; $P = .020$). The HR for progression of these homozygotes in SQC (HR, 0.67; 95% CI, 0.45 to 0.98; $P = .041$) was lower than that in ADC (HR, 0.89; 95% CI, 0.76 to 1.03; $P = .13$). Minor allele homozygotes showed a significantly longer OS than major allele homozygotes (HR, 0.86; 95% CI, 0.74 to 0.99; $P = .039$). The HR for death of these homozygotes in SQC (HR, 0.66; 95% CI, 0.43 to 0.98; $P = .037$) was lower than that in ADC (HR, 0.87; 95% CI, 0.74 to 1.02; $P = .13$).

SNP rs11615 (C118T) in the *ERCC1* gene was reported to be associated with response to platinum-based chemotherapy of NSCLC[8]; thus, it was also examined for association with response in all 640 patients. Minor allele homozygotes for the *ERCC1* SNP showed a higher response rate than others, consistent with a recent report[8]; however, the association was not statistically significant (Appendix Table A2, online only).

Differential Response According to Chemotherapeutic Regimens by PARP1 Genotypes

We next investigated whether or not SNPs in DNA repair genes affect responses differentially according to chemotherapeutic agents. Paclitaxel (433 patients; 68%) and gemcitabine (122 patients; 19%) were the most and second-most commonly used drugs in the platinum-based regimens (other drugs were also used but less frequently [< 10%; Table 2]). Therefore, differences in the response among the

Table 3. Association of *TP53* Genotypes With Response to Chemotherapy in 640 Patients With NSCLC

		Nonresponders		Responders						
NSCLC	Genotype	No.	%	No.	%	Response Rate (%)*	OR	95% CI	*P*	*P* by Trend Test
All	Arg/Arg	175	42.8	72	31.2	29.1	Reference			9.5×10^{-5}
	Arg/Pro	197	48.2	115	49.8	36.9	1.38	0.96 to 1.99	.082†	
	Pro/Pro	37	9.0	44	19.0	54.3	3.02	1.77 to 5.18	4.4×10^{-5}†	
	Dominant						1.63	1.15 to 2.30	.0053†	
	Recessive						2.48	1.54 to 4.04	2.1×10^{-4}†	
Adenocarcinoma	Arg/Arg	152	42.2	64	33.9	29.6	Reference			.0024
	Arg/Pro	176	48.9	90	47.6	33.8	1.19	0.81 to 1.77	.38‡	
	Pro/Pro	32	8.9	35	18.5	52.2	2.67	1.60 to 4.81	8.7×10^{-4}‡	
	Dominant						1.42	0.98 to 2.07	.062‡	
	Recessive						2.44	1.44 to 4.15	9.2×10^{-4}‡	
Squamous cell carcinoma	Arg/Arg	21	46.7	7	17.9	25.0	Reference			.0032
	Arg/Pro	19	42.2	23	59.0	54.8	3.63	1.10 to 13.5	.033‡	
	Pro/Pro	5	11.1	9	23.1	64.3	8.71	1.64 to 62.5	.010‡	
	Dominant						4.62	1.52 to 16.3	.0062‡	
	Recessive						3.85	1.02 to 17.6	.047‡	
Smoker	Arg/Arg	98	39.5	44	27.7	31.0	Reference			.0084
	Arg/Pro	124	50.0	88	55.3	41.5	1.52	0.97 to 2.41	.069§	
	Pro/Pro	26	10.5	27	17.0	50.9	2.31	1.19 to 4.50	.013§	
	Dominant						1.65	1.07 to 2.57	.023§	
	Recessive						1.78	0.99 to 3.23	.056§	
Never-smoker	Arg/Arg	77	47.0	28	38.9	26.7	Reference			.0052
	Arg/Pro	73	44.5	27	37.5	27.0	1.06	0.55 to 2.02	.87§	
	Pro/Pro	11	6.7	17	23.6	60.7	5.31	2.00 to 15.3	6.8×10^{-3}§	
	Dominant	161					1.56	0.86 to 2.86	.14§	
	Recessive						4.76	2.02 to 11.8	3.6×10^{-3}§	

Abbreviations: NSCLC, non–small-cell lung cancer; OR, odds ratio.
*Fraction of responder.
†OR for responder against nonresponder adjusted for sex, age, histology, smoking status, clinical stage, performance status, and treatment.
‡OR for responder against nonresponder adjusted for sex, age, smoking status, clinical stage, performance status, and treatment.
§OR for responder against nonresponder adjusted for sex, age, histology, clinical stage, performance status, and treatment.

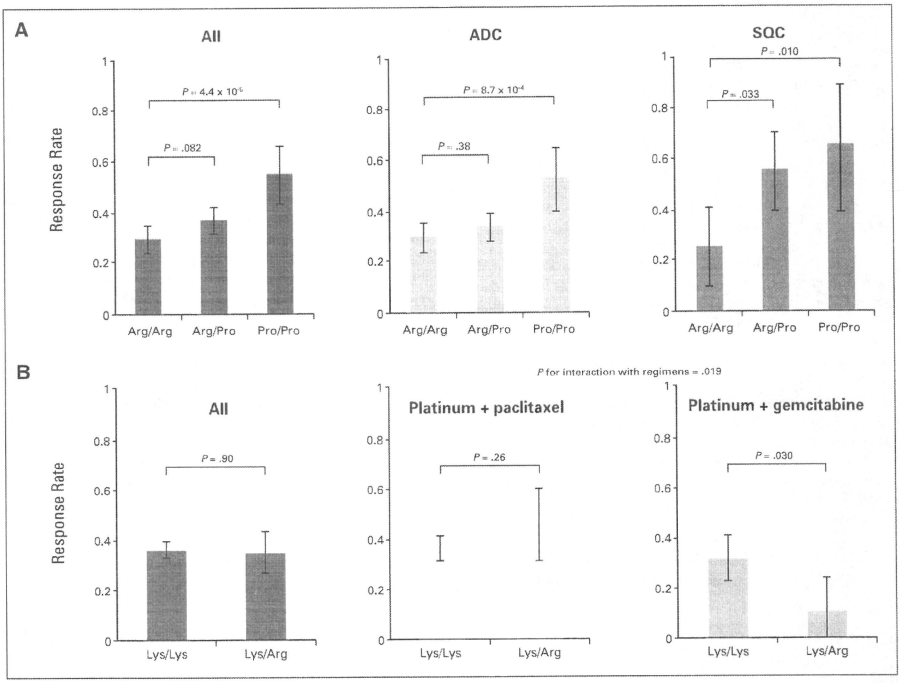

Fig 2. (A) Response rates according to *TP53* genotypes in (left) all patients and those with (middle) aenocarcinoma (ADC) and (right) squamous cell carcinoma (SQC). (B) Response rates according to *PARP1* genotypes in (left) all patients and those treated with (middle) platinum plus paclitaxel or (right) platinum plus gemcitabine. Response rate is shown with its sampling variations estimated by 95% CI.

agents according to genotypes were investigated in 555 patients who received chemotherapy with either of these two regimens.

Among 201 patients in the discovery set, 138 received chemotherapy with regimens using paclitaxel (98 patients) or gemcitabine (40 patients; Fig 1B). Five of the 29 SNPs met the criteria in these 138 patients ($P < .2$ for interaction). Therefore, these five SNPs were further genotyped for 417 patients who received chemotherapy with regimens using paclitaxel (335 patients) or gemcitabine (82 patients) among 439 patients in the validation set. Only one SNP, poly (ADP-ribose) polymerase 1 (PARP1) -Lys940Arg, reproducibly showed $P < .2$ for interaction (Appendix Table A3, online only). This SNP showed a statistically significant interaction with the regimens on the response when analyzed in all 555 patients ($P = .019$ for interaction; Fig 1B, Appendix Table A4, online only), although the association did not remain significant after Bonferroni correction (ie, > 0.05 of 29 SNPs tested $= 1.7 \times 10^{-3}$). Heterozygotes for this SNP showed a better response rate to the paclitaxel regimen (45.8%) than to the gemcitabine regimen (10.5%; Fig 2B). There were no minor allele homozygotes for this SNP in this population.

PFS according to the PARP1-Lys940Arg genotype was compared between the two regimens. In the Cox proportional hazard model, the risk for progression of major allele homozygotes with the platinum/paclitaxel treatment was similar to that with the platinum/gemcitabin treatment (HR, 0.97; 95% CI, 0.86 to 1.09; $P = .60$). Conversely, the risk of heterozygotes with the platinum/paclitaxel treatment was smaller than that with the platinum/gemcitabine treatment, although it was not statistically significant (HR, 0.82; 95% CI, 0.59 to 1.17; $P = .27$). SNPs in *TP53* and *ERCC1* did not show differential associations according to regimens (Appendix Table A4).

DISCUSSION

An SNP in the *TP53* genes was shown to be associated with the response to platinum-based doublet chemotherapy. In this study, association results obtained by the discovery set were confirmed by using an independent validation set. The association of the p53-72Pro allele with a better response to platinum-based doublet chemotherapy

retained statistical significance after Bonferroni correction. Therefore, the results strongly indicate the importance of p53-Arg72Pro SNP as a determinant for the response to platinum-based chemotherapy.

TP53 is a tumor suppressor gene somatically mutated in 40% to 70% of NSCLCs.[23] p53-72Arg protein has a greater activity to induce apoptosis than p53-72Pro protein[24]; however, the relationship was reported as being the reverse in mutant p53 proteins.[25,26] p73, a p53-related protein, plays a role in apoptosis in anticancer agents for cancer cells carrying *TP53* mutations; however, its function is abrogated by mutant p53 proteins. The abrogating activity is greater in mutant p53 proteins with the Arg residue at codon 72 than in those with the Pro residue.[25,26] In an analysis of 25 patients with head and neck cancer, those with a *TP53* mutation on the 72Pro allele showed a better response than those with a mutation on the 72Arg allele with cisplatin-based chemoradiotherapy.[25] Similarly, in this study, the TP53-72Pro allele appeared to confer a better response to platinum-based doublet chemotherapy in patients with NSCLC (Fig 2A). In a previous study,[11] patients with NSCLC who carry the TP53-72Pro allele also showed a better OS after cisplatin-gemcitabine treatment, although the association did not reach statistical significance. These results indicate that p53 mutants with the Pro residue at codon 72 only weakly inhibit the function of p73 protein in NSCLC cells and therefore efficiently induce apoptosis of NSCLC cells treated with platinum and other anticancer agents. In fact, the effect of this SNP was more apparent in patients with SQC than in patients with ADC (Fig 1A), consistent with the fact that *TP53* mutations are more frequent in SQC than in ADC.[27] Since tumor specimens for examination of somatic *TP53* mutations were not available for these patients, *TP53* status in their tumor cells could not be determined. Therefore, we could not conclude whether this differential association was really due to differences in *TP53* mutations. An association study of patients with NSCLC informative for somatic *TP53* mutation will provide a more complete picture of the role of *TP53* SNP in chemotherapeutic responses.

The PARP1-Lys940Arg genotype was suggested to differentially affect the response according to chemotherapeutic agents (Fig 2B), although the association was not significant after Bonferroni correction and needs validation. The *PARP1* gene encodes poly (ADP-ribose) polymerase 1, which regulates multiple processes for DNA repair, such as DNA strand break repair.[28] It is noted that suppression of PARP activity has been recognized as a method of tumor suppression in breast and other cancers[29] and that a PARP inhibitor enhanced the cytotoxic activity of gemcitabine.[30] The biologic significance of the PARP1-Lys940Arg SNP is unknown at present; however, the lysine-arginine residue at codon 940 is located in the catalytic domain of the PARP1 protein.[31] Therefore, this polymorphism may cause differences in the activity of PARP1 protein that affect the response to some chemotherapeutic agents, in particular to DNA-damaging agents.

Interestingly, the frequencies of the TP53-72Pro allele are known to be different among ethnic populations, although those of the PARP1-940Arg allele in other ethnic populations are unknown at present (Table 1). Therefore, examination of these two SNPs in NSCLC patient populations other than Japanese will also help elucidate the mechanism of interethnic differences in the outcome of patients after chemotherapy, as recently discussed.[32]

Identification of polymorphisms associated with drug toxicities is also important to develop customized chemotherapies. For instance, the *UGT1A1* gene polymorphisms are known to be associated with the toxicity of irinotecan, such as neutropenia.[33] In this study, the *TP53* and *PARP1* SNPs were not associated with grade 4 hematologic toxicities, including neutropenia (data not shown). Therefore, genetic factors responsible for response are likely to be different from those for toxicity. In addition, associations of these two SNPs with responses were not significantly different according to smoking habit ($P > .05$ for interaction with smoking; for *TP53*, see Table 3); therefore, these SNPs are likely to contribute to the response irrespective of smoking.

Our study has several limitations. This is a single-institution retrospective study with various therapeutic regimens. Therefore, the effects of SNPs on differential responses according to chemotherapeutic agents were only preliminarily investigated. The results should be confirmed by a larger, preferably prospective, cohort using a defined set of agents. More extensive analyses of interaction between SNPs and responses to chemotherapeutic agents will also be worth performing. Another limitation of this study is that, although the *TP53* polymorphism was significantly associated with response to chemotherapy, differences in PFS and OS were only modest. We chose the response as the primary end point of efficacy to pick up subgroups for which chemotherapy does work. Although this information would be potentially valuable, clinical response alone would be inadequate to improve the outcome of patients with advanced NSCLC. Therefore, investigation of polymorphisms in other genes might provide more information for individually optimized chemotherapy. Indeed, a few other SNPs in DNA repair genes have been reported to be associated with prognosis of patients with NSCLC.[7,9-14] In addition to ERCC1-118T, the APEX-148Asp, XRCC1-399Arg, and XPD-751Gln alleles, which had been reported to be associated with favorable prognosis of patients,[9,13,14] were consistently more frequent in responders than in nonresponders in our study population (Appendix Table A2), although these SNPs did not fulfill the criteria as validated predictive factors in this study.

In conclusion, our extensive analysis of 30 SNPs in 27 DNA repair genes identified the *TP53* and *PARP1* SNPs as strong candidates for defining inter-individual differences in the response to platinum-based chemotherapy of NSCLC. Our results indicate the significance of SNPs in DNA repair genes in the outcome of patients with NSCLC and also imply the utility of these SNPs as predictive markers for responses to chemotherapy. Further investigation is warranted.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: Takashi Kohno, Jun Yokota, Hideo Kunitoh
Financial support: Takashi Kohno, Tomohide Tamura, Jun Yokota
Provision of study materials or patients: Chiharu Tanai, Yasushi Goto, Hiroshi Nokihara, Noboru Yamamoto, Ikuo Sekine, Yuichiro Ohe, Tomohide Tamura, Hideo Kunitoh
Collection and assembly of data: Kouya Shiraishi, Chiharu Tanai, Yasushi Goto, Koji Tsuta
Data analysis and interpretation: Kouya Shiraishi, Takashi Kohno, Aya Kuchiba, Seiichiro Yamamoto, Jun Yokota, Hideo Kunitoh
Manuscript writing: Kouya Shiraishi, Takashi Kohno, Jun Yokota, Hideo Kunitoh

Final approval of manuscript: Kouya Shiraishi, Takashi Kohno, Chiharu Tanai, Yasushi Goto, Aya Kuchiba, Seiichiro Yamamoto, Koji Tsuta, Hiroshi Nokihara, Noboru Yamamoto, Ikuo Sekine, Yuichiro Ohe, Tomohide Tamura, Jun Yokota, Hideo Kunitoh

REFERENCES

1. Parkin DM, Bray F, Ferlay J, et al: Global cancer statistics, 2002. CA Cancer J Clin 55:74-108, 2005

2. Goffin J, Lacchetti C, Ellis PM, et al: First-line systemic chemotherapy in the treatment of advanced non-small cell lung cancer: A systematic review. J Thorac Oncol 5:260-274, 2010

3. Ohe Y, Ohashi Y, Kubota K, et al: Randomized phase III study of cisplatin plus irinotecan versus carboplatin plus paclitaxel, cisplatin plus gemcitabine, and cisplatin plus vinorelbine for advanced non-small-cell lung cancer: Four-Arm Cooperative Study in Japan. Ann Oncol 18:317-323, 2007

4. Scagliotti GV, De Marinis F, Rinaldi M, et al: Phase III randomized trial comparing three platinum-based doublets in advanced non-small-cell lung cancer. J Clin Oncol 20:4285-4291, 2002

5. Schiller JH, Harrington D, Belani CP, et al: Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. N Engl J Med 346:92-98, 2002

6. Olaussen KA, Dunant A, Fouret P, et al: DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. N Engl J Med 355:983-991, 2006

7. Camps C, Sirera R, Iranzo V, et al: Gene expression and polymorphisms of DNA repair enzymes: Cancer susceptibility and response to chemotherapy. Clin Lung Cancer 8:369-375, 2007

8. Wei SZ, Zhan P, Shi MQ, et al: Predictive value of ERCC1 and XPD polymorphism in patients with advanced non-small cell lung cancer receiving platinum-based chemotherapy: A systematic review and meta-analysis. Med Oncol (epub ahead of print on February 9, 2010)

9. Gurubhagavatula S, Liu G, Park S, et al: XPD and XRCC1 genetic polymorphisms are prognostic factors in advanced non-small-cell lung cancer patients treated with platinum chemotherapy. J Clin Oncol 22:2594-2601, 2004

10. Isla D, Sarries C, Rosell R, et al: Single nucleotide polymorphisms and outcome in docetaxel-cisplatin-treated advanced non-small-cell lung cancer. Ann Oncol 15:1194-1203, 2004

11. de las Peñas R, Sanchez-Ronco M, Alberola V, et al: Polymorphisms in DNA repair genes modulate survival in cisplatin/gemcitabine-treated non-small-cell lung cancer patients. Ann Oncol 17:668-675, 2006

12. Giachino DF, Ghio P, Regazzoni S, et al: Prospective assessment of XPD Lys751Gln and XRCC1 Arg399Gln single nucleotide polymorphisms in lung cancer. Clin Cancer Res 13:2876-2881, 2007

13. Matakidou A, el Galta R, Webb EL, et al: Genetic variation in the DNA repair genes is predictive of outcome in lung cancer. Hum Mol Genet 16:2333-2340, 2007

14. Wu X, Lu C, Ye Y, et al: Germline genetic variations in drug action pathways predict clinical outcomes in advanced lung cancer treated with platinum-based chemotherapy. Pharmacogenet Genomics 18:955-965, 2008

15. Sakiyama T, Kohno T, Mimaki S, et al: Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. Int J Cancer 114:730-737, 2005

16. Therasse P, Arbuck SG, Eisenhauer EA, et al: New guidelines to evaluate the response to treatment in solid tumors: European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 92:205-216, 2000

17. Travis W, Colby, TV, Corrin, B, et al: Histological Typing of Lung and Pleural Tumors (ed 3). Heidelberg, Germany, Springer-Verlag, 1999

18. Brambilla E, Travis WD, Colby TV, et al: The new World Health Organization classification of lung tumours. Eur Respir J 18:1059-1068, 2001

19. Kawaishi M, Fujiwara Y, Fukui T, et al: Circulating endothelial cells in non-small cell lung cancer patients treated with carboplatin and paclitaxel. J Thorac Oncol 4:208-213, 2009

20. Sekine I, Nokihara H, Horiike A, et al: Phase I study of cisplatin analogue nedaplatin (254-S) and paclitaxel in patients with unresectable squamous cell carcinoma. Br J Cancer 90:1125-1128, 2004

21. Shiraishi K, Kohno T, Kunitoh H, et al: Contribution of nicotine acetylcholine receptor polymorphisms to lung cancer risk in a smoking-independent manner in the Japanese. Carcinogenesis 30:65-70, 2009

22. Breslow NE, Day NE: Statistical methods in cancer research: Volume I—The analysis of case-control studies. IARC Scientific Publication No. 32, 1980, pp 5-338

23. Weston A, Perrin LS, Forrester K, et al: Allelic frequency of a p53 polymorphism in human lung cancer. Cancer Epidemiol Biomarkers Prev 1:481-483, 1992

24. Dumont P, Leu JI, Della Pietra AC 3rd, et al: The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. Nat Genet 33:357-365, 2003

25. Bergamaschi D, Gasco M, Hiller L, et al: p53 polymorphism influences response in cancer chemotherapy via modulation of p73-dependent apoptosis. Cancer Cell 3:387-402, 2003

26. Vikhanskaya F, Siddique MM, Kei Lee M, et al: Evaluation of the combined effect of p53 codon 72 polymorphism and hotspot mutations in response to anticancer drugs. Clin Cancer Res 11:4348-4356, 2005

27. The International Agency for Research on Cancer: IARC TP53 Database. http://www-p53.iarc.fr/

28. Amé JC, Spenlehauer C, de Murcia G: The PARP superfamily. Bioessays 26:882-893, 2004

29. Iglehart JD, Silver DP: Synthetic lethality: A new direction in cancer-drug development. N Engl J Med 361:189-191, 2009

30. Jacob DA, Bahra M, Langrehr JM, et al: Combination therapy of poly (ADP-ribose) polymerase inhibitor 3-aminobenzamide and gemcitabine shows strong antitumor activity in pancreatic cancer cells. J Gastroenterol Hepatol 22:738-748, 2007

31. Cao WH, Wang X, Frappart L, et al: Analysis of genetic variants of the poly(ADP-ribose) polymerase-1 gene in breast cancer in French patients. Mutat Res 632:20-28, 2007

32. Gandara DR, Kawaguchi T, Crowley J, et al: Japanese-US common-arm analysis of paclitaxel plus carboplatin in advanced non-small-cell lung cancer: A model for assessing population-related pharmacogenomics. J Clin Oncol 27:3540-3546, 2009

33. Bosch TM: Pharmacogenomics of drug-metabolizing enzymes and drug transporters in chemotherapy. Methods Mol Biol 448:63-76, 2008

Appendix

Table A1. Association of Six DNA Repair Gene SNPs With Response to Platinum-Doublet Chemotherapy

			Discovery Set (n = 201)									Validation Set (n = 438)								
			Non-responders		Responders							Non-responders		Responders						
Gene	SNP	Genotype	No.	%	No.	%	Response Rate (%)*	OR†	95% CI	*P*	Trend *P*	No.	%	No.	%	Response Rate %*	OR†	95% CI	*P*	Trend *P*
TP53	Arg72Pro	Arg/Arg	54	42.5	23	31.1	29.9	Reference			.084	121	42.9	49	31.2	28.8	Reference			8.2×10^{-2}
		Arg/Pro	59	46.5	36	48.6	37.9	1.22	0.61 to 2.44	.58		138	48.9	79	50.3	36.4	1.42	0.92 to 2.20	.12	
		Pro/Pro	14	11.0	15	20.3	51.7	2.31	0.80 to 6.66	.063		23	8.2	29	18.5	55.8	3.16	1.64 to 6.21	6.1×10^{-4}	
		Dominant						1.48	0.79 to 2.83	.22							1.66	1.10 to 2.53	.018	
		Recessive						2.10	0.88 to 5.02	.094							2.59	1.43 to 4.75	.0017	
ERCC6	Gly399Asp	Gly/Gly	26	20.5	22	29.7	45.8	Reference			.049	85	30.1	60	38.2	41.4	Reference			.17
		Gly/Asp	77	60.6	43	58.1	35.8	0.54	0.26 to 1.11	.094		141	50.0	68	43.3	32.5	0.68	0.43 to 1.06	.087	
		Asp/Asp	24	18.9	9	12.2	27.3	0.32	0.085 to 0.94	.038		56	19.9	29	18.5	34.1	0.73	0.41 to 1.28	27	
		Dominant						0.51	0.25 to 1.02	.062							0.69	0.45 to 1.05	.081	
		Recessive						0.66	0.24 to 1.39	.23							0.91	0.56 to 1.50	.72	
POLD1	Arg119His	Arg/Arg	91	74.0	39	51.3	28.8	Reference			.0006	179	63.5	91	58.0	33.7	Reference			.20
		Arg/His	29	22.8	34	45.0	54.0	3.55	1.81 to 7.15	2.1×10^{-4}		90	31.9	56	35.7	38.4	1.24	0.81 to 1.90	.32	
		His/His	4	3.1	2	2.7	33.3	1.76	0.15 to 7.51	.61		13	4.6	10	6.4	42.5	1.56	0.63 to 2.75	.33	
		Dominant						3.21	1.68 to 6.29	4.1×10^{-4}							1.28	0.85 to 1.92	.23	
		Recessive						0.78	0.089 to 4.52	.79							1.43	0.60 to 3.38	.43	
APEX1	Asp148Glu	Asp/Asp	51	40.2	24	32.4	32.0	Reference			.075	13	40.1	85	41.4	36.6	Reference			.85
		Asp/Glu	62	48.8	37	50.0	37.4	1.38	0.71 to 2.75	.35		129	45.7	68	43.3	34.5	0.87	0.56 to 1.34	.52	
		Glu/Glu	14	11.0	13	17.6	48.1	2.34	0.89 to 6.31	.088		40	14.2	24	15.3	37.5	0.98	0.52 to 1.83	.95	
		Dominant						1.55	0.82 to 2.99	.17							0.91	0.61 to 1.37	.66	
		Recessive						2.00	0.84 to 4.83	.12							1.07	0.60 to 1.87	.81	
MSH3	Thr1046Ala	Thr/Thr	67	52.8	47	63.5	41.2	Reference			.041	150	56.0	94	59.9	37.3	Reference			.79
		Thr/Ala	49	37.8	25	33.8	34.2	0.70	0.36 to 1.34	.28		110	39.9	63	33.8	32.5	0.97	0.57 to 1.33	.53	
		Ala/Ala	12	9.4	2	2.7	14.3	0.25	0.050 to 1.11	.071		14	5.0	10	6.4	41.7	1.14	0.46 to 2.72	.77	
		Dominant						0.60	0.32 to 1.13	.11							0.69	0.59 to 1.33	.57	
		Recessive						0.25	0.037 to 1.01	.057							1.26	0.53 to 2.93	.59	
FANCA	Ser501Gly	Ser/Ser	94	74.0	48	64.8	33.8	Reference			.071	197	69.9	108	68.8	35.4	Reference			.96
		Ser/Gly	33	26.0	25	33.8	43.1	1.64	0.85 to 3.19	.14		76	27.0	44	28.0	36.7	1.01	0.65 to 1.58	.95	
		Gly/Gly	0	0	1	1.4	100	—		.088		9	3.2	5	3.2	35.7	0.95	0.28 to 2.86	.92	
		Dominant						1.72	0.89 to 3.32	.10							1.02	0.66 to 1.56	.94	
		Recessive						—		.11							0.98	0.29 to 2.93	.97	

Abbreviations: SNP, single nucleotide polymorphism; OR, odds ratio.
*Fraction of responders.
†OR for responders against nonresponders adjusted for sex, age, histology, smoking status, clinical stage, performance status, and treatment.

Table A2. Association of Genotypes for DNA Repair Genes With Response to Chemotherapy of 640 Patients With NSCLC

Gene	SNP	SNP ID	Genotype	Nonresponders No.	Nonresponders %	Responders No.	Responders %	Response Rate (%)*	OR†	95% CI	*P*	*P* by Trend Test
ERCC1‡	C118T (Asn3Asn)	rs11615	C/C	212	51.8	127	55.0	37.5	Reference			.87
			C/T	174	42.5	86	37.2	33.1	0.83	0.58 to 1.16	.27	
			T/T	23	5.6	18	7.8	43.9	1.36	0.69 to 2.64	.37	
			Dominant						0.89	0.64 to 1.23	.47	
			Recessive						1.45	0.74 to 2.72	.29	
APEX1‡	Asp148Glu	rs1130409	Asp/Asp	164	40.1	89	38.5	35.2	Reference			.36
			Asp/Glu	191	46.7	105	45.5	35.5	1.01	0.71 to 1.45	.95	
			Glu/Glu	54	13.2	37	16.0	40.7	1.29	0.77 to 2.14	.33	
			Dominant						1.08	0.77 to 1.51	.66	
			Recessive						1.32	0.82 to 2.09	.25	
XRCC1§	Arg399Gln	rs25487	Arg/Arg	69	54.3	44	59.5	38.9	Reference			.22
			Arg/Gln	45	35.4	25	33.8	35.7	0.89	0.45 to 1.74	.74	
			Gln/Gln	13	10.2	5	6.8	27.8	0.38	0.10 to 1.22	.11	
			Dominant						0.78	0.41 to 1.45	.43	
			Recessive						0.43	0.12 to 1.37	.14	
XPD§	Lys751Gln	rs13181	Lys/Lys	116	91.3	65	87.8	35.9	Reference			.32
			Lys/Gln	11	8.7	9	12.2	45.0	1.68	0.60 to 4.60	.32	
			Gln/Gln	0	0	0	0	—	—			
			Dominant						1.68	0.60 to 4.60	.32	
			Recessive						—			

Abbreviations: NSCLC, non–small-cell lung cancer; SNP, single nucleotide polymorphism; ID, identification; OR, odds ratio.
*Fraction of responder.
†OR for responder against nonresponder adjusted for sex, age, histology, smoking status, clinical stage, performance status, and treatment.
‡Result in the discovery and validation set patients.
§Result in the discovery set patients.

Table A3. Differences in Responses to Chemotherapeutic Agents According to Genotypes for Five DNA Repair Genes

			Discovery Set (n = 138)									
			Platinum + Paclitaxel					Platinum + Gemcitabine				
			Nonresponders		Responders			Nonresponders		Responders		
Gene	SNP	Genotype	No.	%	No.	%	Response Rate (%)†	No.	%	No.	%	Response Rate (%)†
PARP1	Lys940Arg	Lys/Lys	55	91.7	29	76.3	34.5	26	86.7	10	100	27.8
		Lys/Arg	5	8.3	9	23.7	64.3	4	13.3	0	0	0
		Arg/Arg	0	0	0	0	—	0	0	0	0	—
ERCC2	Lys751Gln	Lys/Lys	57	95.0	32	84.2	36.0	26	86.7	10	100	27.8
		Lys/Gln	3	5.0	6	15.8	66.7	4	13.3	0	0	0
		Gln/Gln	0	0	0	0	—	0	0	0	0	—
BRCA2	Asn372His	Asn/Asn	38	63.3	20	52.6	34.5	22	73.3	8	80.0	26.7
		Asn/His	21	35.0	13	34.2	38.2	7	23.3	2	20.0	22.2
		His/His	1	1.7	5	13.2	83.3	1	3.3	0	0	0
REV1	Phe257Ser	Phe/Phe	23	38.3	17	44.7	42.5	12	40.0	3	30.0	20.0
		Phe/Ser	31	51.7	18	47.4	36.7	16	53.3	5	50.0	23.8
		Ser/Ser	6	10.0	3	7.9	33.3	2	6.7	2	20.0	50.0
REV3L	Thr1224Ile	Thr/Thr	25	41.7	16	42.1	39.0	16	53.3	3	30.0	15.8
		Thr/Ile	28	46.7	19	50.0	40.4	13	43.3	5	50.0	27.8
		Ile/Ile	7	11.7	3	7.9	30.0	1	3.3	2	20.0	66.7

			Validation Set (n = 417)									
			Platinum + Paclitaxel					Platinum + Gemcitabine				
			Nonresponders		Responders			Nonresponders		Responders		
Gene	SNP	Genotype	No.	%	No.	%	Response Rate (%)†	No.	%	No.	%	Response Rate (%)†
PARP1	Lys940Arg	Lys/Lys	189	90.0	112	89.6	37.2	44	77.2	23	92.0	34.3
		Lys/Arg	21	10.0	13	10.4	38.2	13	22.8	2	8.0	13.3
		Arg/Arg	0	0	0	0	—	0	0	0	0	—
ERCC2	Lys751Gln	Lys/Lys	191	91.0	115	92.0	37.6	54	94.7	24	96.0	30.8
		Lys/Gln	19	9.0	10	8.0	34.5	3	5.3	1	4.0	25.0
		Gln/Gln	0	0	0	0	—	0	0	0	0	—
BRCA2	Asn372His	Asn/Asn	127	60.5	78	62.4	38.0	34	59.6	16	64.0	32.0
		Asn/His	74	35.2	42	33.6	36.2	22	38.6	7	28.0	24.1
		His/His	9	4.3	5	4.0	35.7	1	1.8	2	8.0	66.7
REV1	Phe257Ser	Phe/Phe	107	51.0	52	41.6	32.7	29	50.9	13	52.0	31.0
		Phe/Ser	87	41.4	60	48.0	40.8	25	43.9	10	40.0	28.6
		Ser/Ser	16	7.6	13	10.4	44.8	3	5.3	2	8.0	40.0
REV3L	Thr1224Ile	Thr/Thr	86	41.0	62	49.6	41.9	21	36.8	9	36.0	30.0
		Thr/Ile	97	46.2	53	42.4	35.3	28	49.1	14	56.0	33.3
		Ile/Ile	27	12.9	10	8.0	27.0	8	14.0	2	8.0	20.0

Gene	SNP	*P* for Interaction†
PARP1	Lys940Arg	.021
ERCC2	Lys751Gln	.021
BRCA2	Asn372His	.14
REV1	Phe257Ser	.18
REV3L	Thr1224Ile	.080

Gene	SNP	*P* for Interaction*
PARP1	Lys940Arg	.16
ERCC2	Lys751Gln	.98
BRCA2	Asn372His	.81
REV1	Phe257Ser	.56
REV3L	Thr1224Ile	.64

Abbreviation: SNP, single nucleotide polymorphism.
*Interaction with chemotherapeutic regimens on response.
†Fraction of responders.

Cancer Science
The official journal of the Japanese Cancer Association JCA

Prevalence of human papillomavirus 16/18/33 infection and p53 mutation in lung adenocarcinoma

Reika Iwakawa,[1] Takashi Kohno,[1] Masato Enari,[1] Tohru Kiyono[2] and Jun Yokota[1,3]

Divisions of [1]Biology and [2]Virology, National Cancer Center Research Institute, Tokyo, Japan

(Received March 11, 2010/Revised April 28, 2010/Accepted April 30, 2010/Accepted manuscript online May 19, 2010/Article first published online June 14, 2010)

Human papillomavirus (HPV) infection is a causative event for the development of uterine cervical carcinoma. Human papillomavirus (HPV) 16, 18, and 33 DNA has been also detected frequently in lung adenocarcinomas (AdCs) in East Asian countries; however, its prevalence in Japan remains unclear. We therefore screened for HPV 16/18/33 DNA in 297 lung AdCs in a Japanese population by multiplex PCR with type-specific primers. As reported previously, HPV 16 DNA was detected in two cervical cancer cell lines, CaSki and SiHa, while HPV 18 DNA was detected in HeLa cells, and 0.1–1.0 copies of HPV-DNA per cell were detectable by this method. However, with this method, none of the 297 lung AdCs showed positive signals for HPV 16/18/33 DNA, indicating that HPV-DNA is not or is very rarely integrated in lung AdC genomes in the Japanese. Furthermore, none of the lung AdCs showed positive signals by nested PCR with HPV 16/18 type-specific primers. Therefore, we further attempted to detect HPV 16/18/33 DNA in 91 lung cancer cell lines, including 40 AdC cell lines. Among them, 30 have been established in Japan and the remaining 61 in the USA. No HPV signals were obtained in any of the 91 cell lines by either multiplex or nested PCR, while the *p53* gene was mutated in 81 of them including 35 of the 40 AdC cell lines. These results indicate that HPV 16/18/33 infection does not play a major role in the development of lung AdC in Japan nor in the USA. (*Cancer Sci* 2010; 101: 1891–1896)

Infection with human papillomavirus (HPV) is a critical event for the development of uterine cervical cancer.[1] E6 protein, encoded by HPV, binds the host cellular tumor suppressor protein p53, and triggers its degradation through the ubiquitin pathway.[2,3] Therefore, the biological significance of continuous p53 degradation by HPV-E6 protein in cervical carcinoma is thought to be equivalent to that of p53 inactivation by genetic alterations in various other types of cancers in human carcinogenesis. The *p53* gene is frequently inactivated in lung adenocarcinoma (AdC) by mutations and/or deletions of both alleles, and the prevalence of p53 mutations in lung AdC is approximately 50% with a higher incidence in smokers.[4,5] However, p53 is not genetically altered in the other half of lung AdCs. Therefore, it is possible that p53 is inactivated by other mechanisms in lung AdC cells without p53 mutations. For this reason, there have been many reports investigating the involvement of HPV in lung AdC development. However, the prevalence of HPV infection in lung AdCs varies drastically among the reports.[5–7] Recently, reasons for a wide variation in the prevalence of HPV infection in lung cancer were investigated by two systematic surveys of a large number of publications.[6,7] A higher prevalence in Asia than in Europe was pointed out by these two investigations,[6,7] and a higher prevalence in studies using HPV type-specific primers than in those using consensus HPV primers was also pointed out in the latter investigation.[7] In East Asia (Supplementary Table S1), a high incidence of HPV infection in lung AdC was reported from Taiwan (92.8%), China (46.9%), and Korea (55.1%).[8–10] In particular, a high prevalence of HPV 16 and 18 infections was reported from Taiwan and China and of HPV 33 infection from Korea. In Japan, the incidence of HPV infection (0–19.4%) has been reported to be not as high as in other East Asian countries, but is still high enough to consider its involvement in lung AdC development.[11–14]

Taiwan, China, and Korea are geographically close to Japan and the people in these countries are ethnically also close to the Japanese. Therefore, in this study, we aimed to elucidate whether or not HPV 16, 18, and 33 are also involved in the development of lung AdC in Japan, as in Taiwan, China, and Korea. We applied a multiplex PCR method as well as a nested PCR method using type-specific primers for detection of HPV 16, 18, and 33 DNA in 275 primary and 22 metastatic lung AdCs in Japanese, and also in 91 lung cancer cell lines established in either Japan or the USA. To validate the specificity and sensitivity of HPV detection methods, three cervical carcinoma cell lines were analyzed by the same methods. In 91 cell lines, the status of p53 mutations was comprehensively analyzed and the results were compared with several p53 databases to evaluate accurately the prevalence of p53 inactivation in lung cancers.

Materials and Methods

Patients and tissues. A total of 275 primary lung AdCs and 22 metastatic lung AdCs to the brain were obtained at surgery from patients treated at the National Cancer Center Hospital, Tokyo, and at Saitama Medical University Hospital. The tumors were pathologically diagnosed according to the tumor-node-metastasis classification of malignant tumors[15] (Table 1). Tumor tissues were stored at −80°C until DNA extraction, and genomic DNA was extracted as previously described.[16] This study was undertaken under the approval of the Institutional Review Board of the National Cancer Center.

Cell line DNA. DNA from 91 lung cancer cell lines[17,18] was screened for HPV-DNA in its genome. These cell lines consisted of 40 AdCs, 11 squamous cell carcinomas (SqCs), two adenosquamous carcinomas (ASCs), nine large-cell carcinomas (LCCs), 27 small-cell lung carcinomas (SCLCs), and two others (one carcinoid tumor and one neuroendocrine tumor), as listed in Table 2. Detailed information will be provided upon request. DNA from three cervical carcinoma cell lines, CaSki, SiHa, and HeLa, and HPV 33 containing plasmid DNA, was used as positive controls for detection of HPV-DNA.

Multiplex PCR with HPV type-specific primers. Sequences for the E1 and L2 regions of HPV 16 and for the E1 region of HPV 18 and 33, together with the aminolevulinate, delta-, synthase 1 (ALAS1) gene segment as an internal positive control, were simultaneously amplified by multiplex PCR in a single tube, as reported.[19] The primer sequences are shown in Supplementary Table S2. Multiplex PCR was performed with Takara Taq (Takara, Shiga, Japan) with a volume of 50 μL containing 1×

[3]To whom correspondence should be addressed.
E-mail: jyokota@ncc.go.jp

doi: 10.1111/j.1349-7006.2010.01622.x

Table 1. Clinicopathologic characteristics of lung adenocarcinomas

	PCR	Primary tumor		Brain metastasis
		Multiplex (%)	Nested (%)	Both (%)
No. of cases	–	275	138	22
Gender	Male	161 (59)	81 (59)	15 (68)
	Female	114 (41)	57 (41)	7 (32)
Age (years)	Mean	60.7	62.0	57.3
	Range	30–84	30–84	48–74
Pathological stage	I	201 (73)	124 (90)	–
	II	27 (10)	6 (4)	–
	III	45 (16)	8 (6)	–
	IV	2 (1)	0 (0)	–
Smoking history	Smoker	71 (55)	69 (58)	15 (68)
	Nonsmoker	57 (45)	51 (43)	7 (32)
	Unknown	147	18	0
p53 mutation	+	34 (32)	34 (33)	16 (73)
	–	72 (68)	70 (67)	6 (27)
	ND	169	34	0

ND, not determined.

PCR buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.025 U Taq polymerase, 3 nM primers, and 10 ng template DNA. Amplifications were performed with the following cycling profiling using a GeneAmp PCR system 9700 apparatus (Applied Biosystems, Foster City, CA, USA): Taq polymerase activation by incubation at 95°C for 1 min, followed by 40 cycles of denaturation at 94°C for 30 s, annealing at 70°C for 90 s, and elongation at 72°C for 60 s. Five micro liters of the amplicons were analyzed by electrophoresis on 3% agarose gels and ethidium bromide staining.

Nested PCR with HPV type-specific primers. Sequences from the upstream regulatory region (URR) to the E7 region of HPV 16 and HPV 18 were first amplified by PCR with outer primers, and the HPV 16 E6/E7 and HPV 18 E6 regions were secondly amplified by nested PCR with inner primers, as reported previously.[20] The primer sequences are shown in Supplementary Table S2. The first round of PCR was performed under the following conditions: Taq polymerase activation at 95°C for 1 min, followed by 35 cycles of denaturation at 95°C for 1 min, annealing at 60°C for 1 min, and elongation at 72°C for 1 min. The second round of PCR was performed as follows: 95°C for 1 min, followed by 20 cycles of denaturation for 1 min at 95°C, 1 min of annealing at 60°C, and 1 min of elongation at 72°C. Polymerase chain reaction (PCR) was performed with a Takara Taq with a volume of 20 μL containing 1× PCR buffer, 0.2 mM dNTPs, 0.05 U Taq polymerase, 2 nM of primers, and 10 ng of template DNA for the first round PCR and 1 μL of the first round PCR products for the second round PCR using a GeneAmp PCR system 9700 apparatus (Applied Biosystems).

Mutation analysis of the *p53* gene. A total of 106 of the 275 primary lung AdCs and all of the 22 metastatic lung AdCs were previously examined for mutations in exons 4–8 of the *p53* gene by genomic PCR and direct sequencing.[21,22] All of the 91 lung cancer cell lines were examined for mutations in exons 2–11 covering all the coding sequences of the p53 gene by genomic PCR and direct sequencing as previously described.[18,23] Sequence data for the cell lines obtained in this study were compared with those of the Catalogue of Somatic Mutations in Cancer (COSMIC) (http://www.sanger.ac.uk/cosmic/).[24]

Results

Detection of HPV 16, 18, and 33 DNA by multiplex PCR. Recently, Nishiwaki *et al.* developed a rapid and sensitive multiplex PCR-based HPV genotyping method that allows the simultaneous amplification of 16 different HPV genotypes in a single tube reaction.[19] This method is based on the amplification of multiple HPV-DNA sequences with a set of HPV type-specific primers, and the HPV types are visually distinguished by the size of amplified fragments after separation by gel electrophoresis. Since DNA for HPV 16, 18, and 33 types has been frequently detected in lung AdC cells in East Asia, four primer sets for these three HPV types, in addition to a primer set for a control genome sequence, were used in this study. Two sets of primers were prepared for the amplification of the HPV 16 DNA[19] because of a possible high prevalence of HPV 16 DNA integration in lung AdC genomes.

The sensitivity and specificity of this method was validated using genomic DNA from three cervical cancer cell lines, CaSki, SiHa, and HeLa, and a lung cancer cell line, A549. Human papillomavirus (HPV) 16 has been shown to be integrated into chromosomal DNA in the CaSki and SiHa cell lines, while HPV 18 is integrated in the HeLa cell line.[25–27] A cell line with integration of HPV 33 was not available; therefore, HPV 33 containing plasmid DNA was mixed with A549 cell DNA as a ratio of one copy of HPV 33 DNA per diploid human genome. Specific DNA fragments for HPV 16, 18, and 33 of different sizes from each other were successfully amplified with the control genomic DNA fragment (Genome in Fig. 1) in CaSki, SiHa, and HeLa cells, as well as A549 cells mixed with HPV 33 DNA (Fig. 1). Two bands for HPV 16 DNA (HPV16-U and HPV16-L) were detected in CaSKi and SiHa cell DNA, while a band for HPV 18 DNA was detected in HeLa cell DNA. Human papillomavirus (HPV) 33-specific DNA was amplified from the mixture of plasmid DNA and A549 cell DNA, while no HPV-specific DNA was amplified from A549 cell DNA. Therefore, by this method, three different HPV types were successfully identified and distinguished by the difference in the sizes of amplified DNA. To determine the sensitivity of this method, each sample was serially diluted and mixed with A549 cell DNA to obtain genomic DNA with 0.1–1.0 copies of each HPV-DNA. Approximately 600 copies of HPV 16 DNA are integrated in CaSki cells, one to two copies of HPV 16 DNA are integrated in SiHa cells, and 20–50 copies of HPV 18 DNA are integrated in HeLa cells.[25–27] As shown in Figure 1, 0.1–1.0 copies of the HPV-DNA sequence per cell were detected by this method. Therefore, this method allowed us to detect one copy of HPV 16, 18, and/or 33 DNA integrated in chromosomal DNA of human cells. Further validation of this method was performed using DNA isolated from 18 primary cervical cancers because the presence of the HPV 16/18 DNA in these tumors was previously determined by Southern blot analysis.[28,29] Human papillomavirus (HPV) types detected by multiplex PCR analysis were completely the same as those by Southern blot analysis, and the sensitivity of multiplex PCR analysis for detection of HPV 16 DNA was higher than that of Southern blot analysis. Four cases negative for HPV 16 DNA by Southern blot analysis were positive by multiplex PCR analysis (data not shown). Therefore, we concluded that the sensitivity of the multiplex PCR analysis is higher than that of Southern blot analysis for detection of HPV 16 DNA in cancer cells.

We then applied this method for detection of HPV 16/18/33 DNA in 275 primary lung AdCs and 22 metastatic lung AdCs to the brain (Table 1). However, HPV-specific DNA was not amplified in any of these 297 lung AdCs. Thus, it was strongly suggested that HPV 16/18/33 DNA is not integrated in the chromosomal DNA of these lung AdCs.

Detection of HPV 16 and 18 DNA by nested PCR. It was reported that only a part of HPV-DNA, from the URR to the E6/E7 region, is commonly integrated in chromosomal DNA of cervical cancer cells, and that deletions of other regions occur in the course of viral DNA integration into host cell DNA.[25,26,30]

doi: 10.1111/j.1349-7006.2010.01622.x

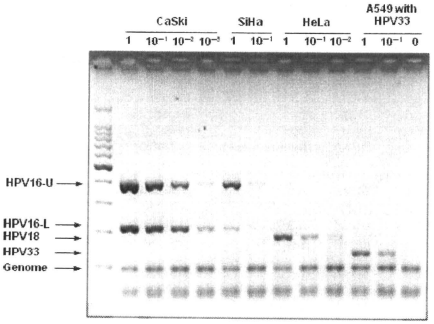

Fig. 1. Detection of human papillomavirus (HPV) 16, 18, and 33 DNA in cervical cancer cell lines by multiplex PCR analysis. Specificity and sensitivity for detection of HPV 16, 18, and 33 DNA. Polymerase chain reaction (PCR) was performed using DNA from CaSki (~600 copies of HPV 16 integrated), SiHa (1–2 copies of HPV 16 integrated), HeLa (20–50 copies of HPV 18 integrated), and A549 cells with/without HPV 33 containing plasmid DNA. Each sample was serially diluted with A549 cell DNA up to the copy number of 0.1–1.0 per cell for HPV-DNA. Five micro liters of the amplicons were analyzed by electrophoresis on 3% agarose gels and ethidium bromide staining. 100 bp DNA Ladder (Takara, Shiga, Japan) was used as a size marker.

multiplex PCR analysis were subjected to nested PCR analysis (Table 1). However, none of them showed positive signals for the E6/E7 regions of the HPV 16 or 18. The results of multiplex PCR analysis as well as those of nested PCR analysis strongly indicated that HPV 16 and 18 are not integrated in lung AdCs developed in Japan, at least in the Tokyo area.

Absence of HPV 16, 18, and 33 DNA sequences in lung cancer cell lines. We next attempted to detect HPV 16, 18, and 33 DNA in a panel of 91 human lung cancer cell lines established in either Japan or the USA. Among the 91 cell lines, 30 originated from Japanese, 42 from Caucasians, and five from African-Americans. Detailed information was not available for the remaining 14 cell lines. Forty cell lines were derived from AdC and the remaining 51 were from other histological types. Both multiplex PCR and nested PCR were performed on all of these cell lines. However, no HPV-specific signals were obtained in any of these cell lines. Therefore, HPV 16, 18, and 33 DNA is not integrated in the 91 lung cancer cell lines established in Japan and the USA. Eleven cell lines were derived from SqC, and 27 cell lines were derived from SCLC; therefore, HPV 16/18/33 integration was not evident in any major histological types of lung cancer.

Status of p53 mutations in lung cancer cell lines. We previously examined for p53 mutations in 106 of the 275 primary tumors and all 22 brain metastases,[21,22] and the mutations were detected in 34 of the 106 primary tumors (32%) and 16 of the 22 brain metastases (73%) (Table 1). We recently reported the status of p53 mutations in 87 of the 91 cell lines analyzed in the present study.[18] In that study, mutation data of several cell lines were obtained not only by direct sequencing of the p53 coding regions but also from the COSMIC database, and the mutations were detected in 70 of the 87 cell lines (80%). However, during this study, we noticed that data for p53 mutations are not the same among three major databases, COSMIC, UMD_TP53 database (http://p53.free.fr), and IARC p53 database (http://www-p53.iarc.fr).[24,31,32] Absence of HPV 16/18/33 integration as well as p53 mutations in 17 lung cancer cell lines prompted us to re-investigate the status of p53 mutations in these cell lines. Therefore, the p53 mutation status in all the 91 cell lines was determined by direct sequencing of all the coding exons, from exon 2 to exon 11, together with exon–intron boundaries of these exons (Table 2). If mutations were detected in the exon–intron boundaries, a possible occurrence of splicing abnormalities due to the mutations was examined by direct sequencing of p53 cDNA products from the corresponding cell lines. Point mutations were detected in 64 of the 91 cell lines, small insertions/deletions in six of them, and large deletions in two of them. Splice-site mutations were detected in nine cell lines, in which shifts of open reading frames due to either exon skipping or intron retention were confirmed. Accordingly, only 10 cell lines were shown to carry the wild-type *p53* gene and express normal p53 protein, including five of the 40 AdC cell lines.

The status of 36 cell lines was not available in COSMIC and thus was defined by our studies (Supplementary Table S3-1) ([18,21–23], this study). The status of 45 cell lines was concordant between our data and COSMIC data (Supplementary Table S3-2), whereas that of the remaining 10 cell lines was discordant (Supplementary Table S3-3). Therefore, although 10 of the 91 lung cancer cell lines carry the wild-type *p53* gene, HPV 16, 18, or 33 are not integrated in these cell lines.

Discussion

To detect HPV-DNA in lung cancer cells, we applied two different PCR methods with HPV type-specific primers, one-step multiplex PCR[19] and nested PCR,[20] because PCR with type-specific primers was reported to be more sensitive than PCR with consensus primers to detect HPV-DNA sequences in human cell DNA.[7] The prevalence and genotype distribution of HPV in cervical cancer precursor lesions defined by one-step multiplex PCR was reported to be compatible with several previous data.[19] In addition, by using these methods, HPV 16 and 18 DNA was distinguishably and efficiently amplified from three cervical cancer cell lines. Therefore, the lack of HPV 16, 18, and 33 DNA in primary lung AdC as well as in lung cancer cell lines would not be due to the low sensitivity of this method for HPV detection. Accordingly, it was concluded from this study that HPV 16, 18, and 33 are not (or are rarely) integrated in lung AdC genomes in the Japanese, particularly those living in the Tokyo area. Lung cancer cell lines analyzed in this study have been established in either Japan or the USA, and consist of all major histological types of lung cancer. Absence of HPV 16/18/33 infection in primary lung AdCs in the US population and lung cancer cell lines established in the USA was previously reported.[33–35] Therefore, the results in the cell lines are consistent with the results in primary AdCs in both Japan and the USA. Indeed, we further attempted to detect HPV-associated DNA sequences in these cell lines by PCR under several low stringent conditions using a set of consensus primers for HPV 16, 18, and 33. However, no HPV-specific signals were detected in any of the 91 lung cancer cell lines examined (data not shown). Therefore, we concluded that no HPV 16/18/33 DNA is integrated in these cell lines. Accordingly, HPV infection seems not to play an important role in the development of lung cancer in Japan nor in the USA, although it is still possible that other HPV types play some role in its development.

A Taiwanese study reported that female never-smokers with lung cancer who were older than 60 years of age had a significantly higher prevalence of HPV 16/18 infections.[8] However, in Korean lung cancer patients, HPV 16/18/33 infections were not associated with gender, smoking status, and histological type.[10] In a study in China, HPV 16/18 infections were not correlated with any clinicopathological parameter, including

doi: 10.1111/j.1349-7006.2010.01622.x

Table 2. Status of the *p53* gene in 91 lung cancer cell lines

No.	Cell line	Hist.	Amino acid	Nucleotide
Point mutation				
1	ABC1	AdC	p.P278S	c.832C>T
2	CALU-3	AdC	p.M237I	c.711G>T
3	HCC44	AdC	p.S94X+ p.R175L	c.281C>G+ c.524G>T
4	HCC78	AdC	p.S241F	c.722C>T
5	HCC193	AdC	p.R248Q	c.743G>A
6	HCC515	AdC	p.L194F	c.580C>T
7	Ma10	AdC	p.G245V	c.734G>T
8	Ma17	AdC	p.Y126C	c.377A>G
9	Ma24	AdC	p.R337C	c.1009C>T
10	H23	AdC	p.M246I	c.738G>C
11	H441	AdC	p.R158L	c.473G>T
12	H820	AdC	p.T284P	c.850A>C
13	H1437	AdC	p.R267P	c.800G>C
14	H1975	AdC	p.R273H	c.818G>A
15	H2009	AdC	p.R273L	c.818G>T
16	H2087	AdC	p.V157F	c.469G>T
17	H2122	AdC	p.Q16L+ p.C176F	c.527G>T+ c.47A>T
18	H2126	AdC	p.E62X	c.184G>T
19	PC3	AdC	p.R282W	c.844C>T
20	PC7	AdC	p.H214R	c.641A>G
21	PC9	AdC	p.R248Q	c.743G>A
22	PC14	AdC	p.R248W	c.742C>T
23	RERF-LCMS	AdC	p.R248L	c.743G>T
24	RERF-LC-OK	AdC	p.F113C	c.338T>G
25	VMRC-LCD	AdC	p.R175H	c.524G>A
26	II-18	AdC	p.K164X	c.490A>T
27	H322	AdC	p.R248L	c.743G>T
28	EBC1	SqC	p.E171X	c.511G>T
29	LC1/Sq	SqC	p.M237I	c.711G>T
30	LK2	SqC	p.V272M	c.814G>A
31	HCC15	SqC	p.D259V	c.776A>T
32	H520	SqC	p.W146X	c.438G>A
33	SK-MES-1	SqC	p.E298X	c.892G>T
34	PC10	SqC	p.G245C	c.733G>T
35	HCC366	ASC	p.Y220C	c.659A>G
36	H596	ASC	p.G245C	c.733G>T
37	Lu65	LCC	p.E11Q	c.31G>C
38	Ma2	LCC	p.R175H	c.524G>A
39	Ma25	LCC	p.M237I	c.711G>T
40	H661	LCC	p.R158L+ p.S215I	c.473G>T+ c.644G>T
41	H1155	LCC	p.R273H	c.818G>A
42	PC13	LCC	p.G334V	c.1001G>T
43	HCC33	SCLC	p.C242Y	c.725G>A
44	Lu134	SCLC	p.P278L	c.833C>T
45	Lu135	SCLC	p.G244C	c.730G>T
46	Lu139	SCLC	p.V157F	c.469G>T
47	N417	SCLC	p.E298X	c.892G>T
48	H69	SCLC	p.E171X	c.511G>T
49	H128	SCLC	p.E62X	c.184G>T
50	H345	SCLC	p.Y236C	c.707A>G
51	H446	SCLC	p.Q154V	c.461G>T
52	H841	SCLC	p.C242S	c.724A>T
53	H1184	SCLC	p.G334V	c.1001G>T
54	H1450	SCLC	p.L194R	c.581T>G
55	H1607	SCLC	p.P151H	c.452C>A
56	H1963	SCLC	p.V147D+ p.H214R	c.440T>A+ c.641A>G
57	H2107	SCLC	p.K101X	c.301A>T
58	H2141	SCLC	p.R209X	c.625A>T
59	H2171	SCLC	p.Q144X	c.430C>T
60	H2195	SCLC	p.V157F	c.469G>T
61	H1618	SCLC	p.R248L	c.743G>T
62	H187	SCLC	p.S241C	c.722C>G
63	H510	SCLC	p.R282G	c. 844C>G
64	H1770	Neuroendocrine	p.R248W	c.741-742CC>TT
Small insertion/deletion (≤9 nucleotides)				
1	Ma29	AdC	p.V121fs	c.363delT
2	H522	AdC	p.P191fs	c.572delC
3	H1648	AdC	p.L35fs	c.103-104insT
4	HCC95	SqC	p.G334fs	c.1000(-1003) 1G del
5	H157	SqC	p.L35fs+ p.E298X	c.103-104insT+ c.892G>T
6	H727	Carcinoid	p.Q165-S166 insYKQ	c.496-497ins9
Large deletion				
1	H358	AdC	p?	Large deletion
2	H1299	LCC	p?	Large deletion
Splicing-site mutation				
1	H1703	AdC	p.G262fs	g. Ivs8 +1g>t
2	H1819	AdC	p.A307fs	g. Ivs9 +1g>t
3	H2347	AdC	p.Y126fs	g.375G>A
4	H1650	AdC	p.V225fs	g.Ivs6 -2a>g
5	Sq1	SqC	p.Y126fs	g. Ivs4 +2t>c
6	H82	SCLC	p.Y126fs	g.375G>T
7	H209	SCLC	p.V225fs	g. Ivs6 -2a>t
8	H526	SCLC	p.S33fs	g. Ivs3 -1g>c
9	H1339	SCLC	p.I332fs	g.Ivs9 +1g>t
Wild type				
1	A427	AdC	—	—
2	A549	AdC	—	—
3	Ma12	AdC	—	—
4	Ma26	AdC	—	—
5	H1395	AdC	—	—
6	H226	SqC	—	—
7	Lu99A	LCC	—	—
8	H460	LCC	—	—
9	Lu24	SCLC	—	—
10	Ms18	SCLC	—	—

p, c, and g indicate protein, cDNA, and genomic DNA.
AdC, adenocarcinoma; ASC, adenosquamous carcinoma; LCC, large-cell carcinoma;
SCLC, small-cell lung carcinoma; SqC, squamous cell carcinoma.

The primers for HPV 16, 18, and 33 in the above multiplex PCR analysis were designed to amplify the E1 or L2 region (Supplementary Table S2).[19] Therefore, it was possible that multiplex PCR analysis failed to detect the HPV-DNA sequences because of integration of truncated HPV genomes without the E1 and L2 regions into host cell DNA. To pursue the possible integration of HPV 16 and 18 DNA in lung AdC cells, we performed a nested PCR analysis for the E6 and E7 regions of HPV 16 and 18. The URR to the E7 region of both HPV 16 and 18 genomes was first amplified using outer primers, then, the E6 to E7 region of the HPV 16 DNA and the E6 region of the HPV 18 DNA were amplified using inner primers (Supplementary Table S2), respectively, according to the method previously described.[20] As in the multiplex PCR analysis, HPV 16- and 18-specific DNA fragments were successfully amplified from the CaSki, SiHa, and HeLa cell lines, but not from A549. Next, 138 of the 275 primary AdCs and all of the 22 metastatic AdCs used for

age, gender, smoking status, and histological type, either.[9] In this study, 41% (121/297) and 43% (64/150) of AdC patients were female and non-smokers, respectively (Table 1). Therefore, the etiological role of HPV 16/18 in lung carcinogenesis in non-smokers seems to be restricted to certain geographic areas, and in Japan, HPV 16/18 infection does not play a causative role in the development of lung AdC in female non-smokers.

An inverse correlation of HPV 16/18 E6 protein expression with p53 expression was also reported in Taiwanese lung tumors.[36] However, in a study in China, there was a relationship between the presence of HPV 16/18 DNA and abnormal p53 protein accumulation.[37] Therefore, association of HPV infection with p53 inactivation is still unclear in lung cancer. We previously examined for p53 mutations in 128 of 297 lung AdCs analyzed in this study, and the mutations were detected in 50 cases (39%) (Table 1); therefore, it was possible that HPV is infected in another 78 cases. However, none of the 78 lung AdCs carried HPV 16/18/33 DNA in their genomes. Accordingly, HPV 16/18/33 infections appear to play a limited role in the development of lung AdC in Japan. These results prompted us to analyze comprehensively the status of the *p53* gene in a large panel of lung cancer cell lines. The *p53* gene is inactivated not only by mutations in the coding regions, but also by splicing abnormalities caused by mutations in the exon–intron boundaries and homozygous deletions, and the incidence of p53 genetic alterations in total was 89% (81/91). Therefore, although 10 of the 91 cell lines were shown to carry the wild-type *p53* gene, no HPV 16/18/33 DNA was detected in these cell lines. Since the status of the *p53* gene in these cell lines was not consistent among several databases and reports, the results provided here will be highly informative to diverse scientists using these cell lines for molecular and biological studies.

In Japan, HPV-DNA has been detected in <10% of lung AdC in Chiba and Hokkaido, and ~20% in Okinawa (Supplementary Table S1). Therefore, we cannot totally rule out the involvement of HPV infection in the etiology of lung AdC in Japan. However, the present results strongly indicate that HPV infection plays only a limited role, if any, in the development of lung AdC in Japan.

Acknowledgments

This work was supported by Grants-in-Aid from the Ministry of Health, Labor and Welfare for the 3rd-Term Comprehensive 10-year Strategy for Cancer Control and for Cancer Research (16-1) and a Grant-in-Aid for the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio). We are grateful to Drs R. Nishikawa and K. Mishima of the Saitama Medical University Hospital for preparation of metastatic lung adenocarcinoma specimens.

Disclosure Statement

The authors have no conflict of interest.

References

1 zur Hausen H. Papillomaviruses in the causation of human cancers - a brief historical account. *Virology* 2009; **384**: 260–5.
2 Werness BA, Levine AJ, Howley PM. Association of human papillomavirus types 16 and 18 E6 proteins with p53. *Science* 1990; **248**: 76–9.
3 Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* 1990; **63**: 1129–36.
4 Nakanishi H, Matsumoto S, Iwakawa R *et al*. Whole genome comparison of allelic imbalance between noninvasive and invasive small-sized lung adenocarcinomas. *Cancer Res* 2009; **69**: 1615–23.
5 Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers–a different disease. *Nat Rev Cancer* 2007; **7**: 778–90.
6 Klein F, Kotb WF, Petersen I. Incidence of human papilloma virus in lung cancer. *Lung Cancer* 2009; **65**: 13–8.
7 Srinivasan M, Taioli E, Ragin CC. Human papillomavirus type 16 and 18 in primary lung cancers–a meta-analysis. *Carcinogenesis* 2009; **30**: 1722–8.
8 Cheng YW, Chiou HL, Sheu GT *et al*. The association of human papillomavirus 16/18 infection with lung cancer among nonsmoking Taiwanese women. *Cancer Res* 2001; **61**: 2799–803.
9 Fei Y, Yang J, Hsieh WC *et al*. Different human papillomavirus 16/18 infection in Chinese non-small cell lung cancer patients living in Wuhan, China. *Jpn J Clin Oncol* 2006; **36**: 274–9.
10 Park MS, Chang YS, Shin JH *et al*. The prevalence of human papillomavirus infection in Korean non-small cell lung cancer patients. *Yonsei Med J* 2007; **48**: 69–77.
11 Hirayasu T, Iwamasa T, Kamada Y, Koyanagi Y, Usuda H, Genka K. Human papillomavirus DNA in squamous cell carcinoma of the lung. *J Clin Pathol* 1996; **49**: 810–7.
12 Miyagi J, Kinjo T, Tsuhako K *et al*. Extremely high Langerhans cell infiltration contributes to the favourable prognosis of HPV-infected squamous cell carcinoma and adenocarcinoma of the lung. *Histopathology* 2001; **38**: 355–67.
13 Kinoshita I, Dosaka-Akita H, Shindoh M *et al*. Human papillomavirus type 18 DNA and E6-E7 mRNA are detected in squamous cell carcinoma and adenocarcinoma of the lung. *Br J Cancer* 1995; **71**: 344–9.
14 Hiroshima K, Toyozaki T, Iyoda A *et al*. Ultrastructural study of intranuclear inclusion bodies of pulmonary adenocarcinoma. *Ultrastruct Pathol* 1999; **23**: 383–9.
15 Sobin LH, Wittekind CH, eds. *TNM classification of malignant tumours*, 6th ed. New York: Wiley-Liss; 2002: p. 99–103.
16 Sakamoto H, Mori M, Taira M *et al*. Transforming gene from human stomach cancers and a noncancerous portion of stomach mucosa. *Proc Natl Acad Sci U S A* 1986; **83**: 3997–4001.
17 Kohno T, Otsuka A, Girard L *et al*. A catalog of genes homozygously deleted in human lung cancer and the candidacy of PTPRD as a tumor suppressor gene. *Genes Chromosomes Cancer* 2010; **49**: 342–52.
18 Blanco R, Iwakawa R, Tang M *et al*. A gene-alteration profile of human lung cancer cell lines. *Hum Mutat* 2009; **30**: 1199–206.
19 Nishiwaki M, Yamamoto T, Tone S *et al*. Genotyping of Human Papillomaviruses by a Novel One-Step Typing Method with Multiplex PCR and Clinical Applications. *J Clin Microbiol* 2008; **46**: 1161–8.
20 Wu EQ, Zhang GN, Yu XH *et al*. Evaluation of high-risk human papillomaviruses type distribution in cervical cancer in Sichuan province of China. *BMC Cancer* 2008; **8**: 202.
21 Tomizawa Y, Kohno T, Fujita T *et al*. Correlation between the status of the p53 gene and survival in patients with stage I non-small cell lung carcinoma. *Oncogene* 1999; **18**: 1007–14.
22 Iwakawa R, Kohno T, Anami Y *et al*. Association of p16 homozygous deletions with clinicopathological characteristics and EGFR/KRAS/p53 mutations in lung adenocarcinoma. *Clin Cancer Res* 2008; **14**: 3746–53.
23 Matsumoto S, Iwakawa R, Takahashi K *et al*. Prevalence and specificity of LKB1 genetic alterations in lung cancers. *Oncogene* 2007; **26**: 5911–8.
24 Forbes SA, Tang G, Bindal N *et al*. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res* 2010; **38**(Database issue): D652–7.
25 Schwarz E, Freese UK, Gissmann L *et al*. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature* 1985; **314**: 111–4.
26 Baker CC, Phelps WC, Lindgren V, Braun MJ, Gonda MA, Howley PM. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. *J Virol* 1987; **61**: 962–71.
27 Yee C, Krishnan-Hewlett I, Baker CC, Schlegel R, Howley PM. Presence and expression of human papillomavirus sequences in human cervical carcinoma cell lines. *Am J Pathol* 1985; **119**: 361–6.
28 Yokota J, Tsukada Y, Nakajima T *et al*. Loss of heterozygosity on the short arm of chromosome 3 in carcinoma of the uterine cervix. *Cancer Res* 1989; **49**: 3598–601.
29 Kohno T, Takayama H, Hamaguchi M *et al*. Deletion mapping of chromosome 3p in human uterine cervical cancer. *Oncogene* 1993; **8**: 1825–32.
30 Pett M, Coleman N. Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? *J Pathol* 2007; **212**: 356–67.
31 Berglind H, Pawitan Y, Kato S, Ishioka C, Soussi T. Analysis of p53 mutation status in human cancer cell lines: a paradigm for cell line cross-contamination. *Cancer Biol Ther* 2008; **7**: 699–708.
32 Petitjean A, Mathe E, Kato S *et al*. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat* 2007; **28**: 622–9.

33 Yousem SA, Ohori NP, Sonmez-Alpan E. Occurrence of human papillomavirus DNA in primary lung neoplasms. *Cancer* 1992; **69**: 693–7.
34 Wistuba II, Behrens C, Milchgrub S *et al.* Comparison of molecular changes in lung cancers in HIV-positive and HIV-indeterminate subjects. *JAMA* 1998; **279**: 1554–9.
35 Shimizu E, Coxon A, Otterson GA *et al.* RB protein status and clinical correlation from 171 cell lines representing lung cancer, extrapulmonary small cell carcinoma, and mesothelioma. *Oncogene* 1994; **9**: 2441–8.
36 Cheng YW, Wu MF, Wang J *et al.* Human papillomavirus 16/18 E6 oncoprotein is expressed in lung cancer and related with p53 inactivation. *Cancer Res* 2007; **67**: 10686–93.
37 Wang Y, Wang A, Jiang R *et al.* Human papillomavirus type 16 and 18 infection is associated with lung cancer patients from the central part of China. *Oncol Rep* 2008; **2**: 333–9.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Prevalence of human papillomavirus (HPV) 16, 18, and 33 in lung adenocarcinomas in East Asia.

Table S2. Primer sequences for detection of human papillomavirus (HPV) DNA in cancer cell DNA.

Table S3-1. p53 status defined in our studies but not registered in the COSMIC database.

Table S3-2. Concordance of p53 status defined in our studies and registered in the COSMIC database.

Table S3-3. Discordance of p53 status defined in our studies and registered in the COSMIC database.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

doi: 10.1111/j.1349-7006.2010.01622.x

Carcinogenesis vol.31 no.5 pp.834–841, 2010
doi:10.1093/carcin/bgq003
Advance Access publication January 8, 2010

Individuals susceptible to lung adenocarcinoma defined by combined *HLA-DQA1* and *TERT* genotypes

Takashi Kohno[1], Hideo Kunitoh[2], Yoko Shimada[1], Kouya Shiraishi[1], Yuko Ishii[1], Koichi Goto[3], Yuichiro Ohe[2], Yutaka Nishiwaki[3], Aya Kuchiba[4], Seiichiro Yamamoto[5], Hiroshi Hirose[6], Akira Oka[7], Noriko Yanagitani[8], Ryusei Saito[8], Hidetoshi Inoko[7] and Jun Yokota[1,*]

[1]Biology Division, National Cancer Center Research Institute, Tokyo 104-0045, Japan, [2]Thoracic Oncology Division, National Cancer Center Hospital, Tokyo 104-0045, Japan, [3]Division of Thoracic Oncology, National Cancer Center Hospital East, Chiba 277-8577, Japan, [4]Genetics Division, National Cancer Center Research Institute, Tokyo 104-0045, Japan and [5]Statistics and Cancer Control Division, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo 104-0045, Japan, [6]Health Center, Keio University School of Medicine, Tokyo 160-8582, Japan, [7]Department of Molecular Life Science, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, Kanagawa 259-1193, Japan and [8]First Department of Internal Medicine, Gunma University School of Medicine, Gunma 371-8511, Japan

*To whom correspondence should be addressed. Tel/Fax: +81 3 35 42 08 07; Email: jyokota@ncc.go.jp

Adenocarcinoma (ADC) is the commonest histological type of lung cancer, and its weak association with smoking indicates the necessity to identify high-risk individuals for targeted screening and/or prevention. By a genome-wide association study (GWAS), we identified an association of polymorphisms in the 6p21.31 locus containing four human leukocyte antigen (HLA) class II genes with lung ADC risk. DQA1*03 of the *HLA-DQA1* gene was defined as a risk allele with odds ratio (OR) of 1.36 [95% confidence interval (CI) = 1.21–1.54, $P = 5.3 \times 10^{-7}$] by analysis of 1656 ADC cases and 1173 controls. DQA1*03 and the minor allele for a polymorphism, rs2736100, in *TERT*, another lung cancer susceptibility locus identified in recent GWASs on Europeans and Americans, were indicated to independently contribute to ADC risk with per allele OR of 1.43 (95% CI = 1.31–1.56, $P = 7.8 \times 10^{-16}$). Individuals homozygous both for the *DQA1*03* and minor *TERT* alleles were defined as high-risk individuals with an OR of 4.76 (95% CI = 2.53–9.47, $P = 4.2 \times 10^{-7}$). The present results indicated that individuals susceptible to lung ADC can be defined by combined genotypes of *HLA-DQA1* and *TERT*.

Introduction

Lung cancer is the leading cause of cancer-related deaths in the world. Adenocarcinoma (ADC) is the commonest histological type comprising ~40% of lung cancer cases among European, North American and Asian countries and is increasing in incidence (1). Development of ADC is more weakly associated with smoking than those of two other major histological types of cancer, squamous cell carcinoma (SQC) and small cell carcinoma (SCC) (1–3). Therefore, identification of high-risk individuals for lung ADC and targeted screening and/or prevention for these individuals will be a powerful way to reduce the number of lung cancer deaths in the world.

Recent genome-wide association studies (GWASs) with single-nucleotide polymorphism (SNP) array methodology have led to the identification of three loci associated with lung cancer risk, *CHRNA3/5* at chromosome 15q25.1, *TERT* and *CLPTM1L* at 5p15.33 and *BAT3-MSH5* at 6p21.33 (4–10). Among these loci, 5p15.33 was revealed as being a locus specifically associated with risk of ADC among major histological types of lung cancer (11). However, loci associated with lung ADC risk in Asians remain obscure. Here, we performed a GWAS on the risk of lung ADC in a Japanese population for 23 010 polymorphic microsatellite loci and identified *HLA-DQA1* at 6p21.31 as a novel locus associated with lung ADC risk. We further examined whether or not individuals susceptible to ADC can be defined by combined genotypes of *HLA-DQA1* and other lung cancer susceptibility loci described above.

Abbreviations: ADC, adenocarcinoma; CI, confidence interval; GWAS, genome-wide association study; HLA, human leukocyte antigen; HWE, Hardy–Weinberg equilibrium; LD, linkage disequilibrium; NCCH, National Cancer Center Hospitals; NNGH, National Nishi-Gunma Hospital; OR, odds ratio; PCR, polymerase chain reaction; SCC, small cell carcinoma; SNP, single-nucleotide polymorphism; SQC, squamous cell carcinoma.

Subjects and methods

Subjects

All the case and control subjects were Japanese and were enrolled in institutions in the Kanto area of Japan, an ~200 km diameter region containing Tokyo. This region is located in the middle of the main island in Japan, where homogeneity of the genetic background of individuals with several common diseases, including lung cancer, has been shown by a recent GWAS on population structure of Japanese (12).

The National Cancer Center Hospitals (NCCH) set consisted of 2343 lung cancer cases and 1173 controls (Table I). The cases were 1656 ADC, 390 SQC and 297 SCC cases. All ADC, SQC and SCC cases were enrolled in the NCCH from 1999 to 2008. All ADC, SQC and SCC cases, from whom informed consent as well as blood samples were obtained, were consecutively included in this study without any particular exclusion criteria. The participation rate was nearly 80%. All the cases were diagnosed by cytological and/or histological examinations according to WHO classification. The controls were 328 inpatients/outpatients of the NCCH, and 645 and 200 volunteers enrolled in Keio and Tokai Universities, respectively. The control NCCH subjects were selected with a criterion of no history of cancer from 1999 to 2007, whereas the 645 volunteers were the individuals with no known malignancies who offered blood on the occasion of a health check examination at Keio University in 2002 and 2003 (13). The 200 volunteers in Tokai University were healthy individuals enrolled from 2001 to 2003 as control subjects in a previous case–control study (14).

The National Nishi-Gunma Hospital (NNGH) sets were 84 ADC and 52 SQC cases and 145 controls who were enrolled in the NNGH from 1999 to 2003 (Table I). All ADC and SQC cases, from whom informed consent as well as blood samples were obtained, were consecutively included in this study without any particular exclusion criteria. The participation rate was nearly 80%. Controls were randomly selected from inpatients and outpatients with no history of cancer. Most of the controls had diseases other than lung cancer such as chronic obstructive pulmonary disease, pulmonary tuberculosis, bronchitis/pneumonia. Their characteristics were described in our previous studies (14–18).

Smoking histories of the subjects were obtained via interview using a questionnaire. Smokers were defined as those who had smoked at least one cigarette per day for 12 months or longer at any time in their life, whereas non-smokers were defined as those who had not. There were no individuals who had smoked less than one cigarette per day and/or for <12 months. Smoking exposure was represented by pack years, which was defined as the number of cigarette packs smoked daily multiplied by years of smoking.

Genomic DNA was extracted from whole-blood cells using a Blood Maxi Kit (Qiagen, Tokyo, Japan) according to the supplier's instructions. Genomic DNAs for 645 and 200 volunteers enrolled in Keio and Tokai Universities, respectively, were extracted from Epstein-Barr virus-transformed B-lymphocytes derived from the collected whole-blood cells (14,16).

Genome-wide association studies

The method of GWAS on microsatellite loci was described previously (14). Equal amounts of DNAs from 200 lung ADC cases and from 200 controls enrolled in Tokai University were mixed for the first set of case and control DNA pools, respectively. The second set of DNA pools was also prepared from another 200 ADC cases and 200 controls enrolled in Keio University. Fifty nanograms of pooled DNA was amplified by 40 cycles of polymerase chain reaction (PCR) in 96-well plates using a pair of PCR primers designed for amplifying fragments that include polymorphic microsatellite sequences. Allele frequencies in pooled DNA were estimated from the height of peaks:

Table I. Characteristics of study subjects

Category	Group	No	Age (mean ± SD)	Sex (% male)	Smoking habit (%)		Pack years of smokers (mean ± SD)
					Non-smoker	Smoker	
NCCH set	Case	2 3443	59 ± 9	65	34	66	51 ± 30
	ADC	1656	58 ± 9	56	46	54	43 ± 27
	SQC	390	62 ± 7	91	3	97	61 ± 29
	SCC	52	70 ± 9	90	6	94	62 ± 32
NNGH set	Case	136	68 ± 10	74	27	73	55 ± 29
	ADC	84	67 ± 10	64	39	61	48 ± 25
	SQC	52	70 ± 9	90	6	94	62 ± 32
	Control	145	64 ± 14	71	33	67	45 ± 35

the frequency of each allele was determined by dividing the height of each allele by the summed height of all alleles. The significance for difference in allelic distribution was evaluated by Fisher's Exact test, with the use of $2m$ (where m is the number of alleles).

The first set of case and control DNA pools was examined for differences in allelic distribution for 23 010 microsatellite markers, and the distribution for 1328 (5.8%) markers were judged as being significantly different by the criteria of $P < 0.05$ (first stage of GWAS in Table II). The inflation factor calculated by dividing the mean of the lower 90% of $-\log10$ (P) values by the mean of the lower 90% of the expected values (19) for this screening was 0.639, indicating a deflation in the statistical tests (supplementary Table I is available at *Carcinogenesis* Online). However, in this screening, deduction of allele frequencies was affected by an inevitable experimental bias of the pooled DNA typing, i.e. 'shadow bands' in electrophoregrams due to slippages in the PCR reaction particularly for microsatellite markers containing repeat units of 2 bp, as reported previously (20). In fact, inflation factors for microsatellite markers containing repeat units of 3–6 bp were 0.919–1.022 (0.955 in total), i.e. deviations were within ±10% as have been observed in previous GWASs in which adequacy of the case–control matching (i.e. lack of a significant hidden population substructure) was indicated (4,8,9,19). Thus, the adequacy of the case–control matching was also indicated in the present screening with microsatellite markers containing repeat units of 3–6 bp. On the other hand, inflation factor for microsatellite markers containing repeat units of 2 bp was 0.520; therefore, the deflation described above was considered to be caused by misestimation in allele frequency in the screening with microsatellite markers containing repeat units of 2 bp. Therefore, among 1328 markers selected in the first stage of GWAS, 431 microsatellite markers with 3–6 bp units were further subjected to the second stage of GWAS.

The second set of DNA pools was examined for differences in allelic distribution for 431 microsatellite markers containing repeat units of 3–6 bp that passed the criteria of $P < 0.05$ in the first stage of GWAS. The distribution for 17 (3.9%) markers were significantly different by the criteria of $P < 0.05$ (second stage of GWAS in Table II). The inflation factor for the second stage screening was 1.010, indicating the adequacy of the case–control matching as well as the lack of differential genotyping of cases and controls (supplementary Table I is available at *Carcinogenesis* Online).

Next, individual typing was done on the 17 markers, which passed the criteria for the third stage, for 576 cases and 576 controls, consisting of 384 cases and 384 controls used in the first and second pooled DNA screening and an additional 192 cases and 192 controls from NCCH (third stage of GWAS in Table II). These 384 cases and 384 controls were consisted of two sets of 192 subjects, which were chosen from two sets of 200 subjects examined in the first and second GWAS stages, respectively, by simple random sampling. These analyses led to the identification of six loci, including D6S0067i, with differences in allelic distributions between the cases and controls with P-values <0.05 by the χ^2 test. The D6S0067i locus showed a P-value of 2.4×10^{-7}, whereas the other five showed P-values of 0.012–0.0011. A level of $P < 2.2 \times 10^{-6}$ was judged as significant by applying Bonferroni correction for multiple test (i.e. $P < 2.2 \times 10^{-6} = 0.05/23\,010$).

Genotyping of SNPs in the 6p21.31 locus

Five hundred and twenty-five cases and 525 controls, which were, respectively, chosen from the 576 cases and 576 controls examined in the third GWAS stage by simple random sampling, were subjected to SNP analysis. Twenty-nine SNPs were selected from the 450 kb region surrounding the D6S0067i locus based on the following criteria: (i) SNPs whose minor allele frequency in the Japanese population was >0.01 in the database of single nucleotide polymorphism (http://www.ncbi.nlm.nih.gov/projects/SNP/) and (ii) SNPs for which pre-designed or validated TaqMan probes were available from Applied Biosystems (Foster City, CA). Three other SNPs, rs1794282, rs3129763 and rs2187668, which showed significant associations with lung cancer risk in Europeans (8), were also examined. Thirty-two SNPs, in total, were genotyped using the TaqMan method according to the protocol for the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems).

Twenty-four SNPs located in exon 2 of the *DRB1*, *DQA1* and *DQB1* genes, which enable allele discrimination for *DRB1*, *DQA1* and *DQB1* at high-, low- and high-resolution levels, respectively, were genotyped by sequence-based typing methods recommended by the International Histocompatibility Working Group (http://www.ihwg.org/). In brief, exon 2 of the *DRB1* and *DQB1* genes was amplified by PCR with mixtures of allele-specific primers, whereas exon 2 of the *DQA1* gene was amplified with a set of common primers, and PCR products were directly sequenced using the ABI3700 DNA analyzer (Applied Biosystems). The location and alleles of the SNPs are described according to the database of major histocompatibility complex (http://www.ncbi.nlm.nih.gov/gv/mhc/). Based on the genotypes of 24 SNPs, alleles for *DRB1*, *DQA1*, *DQB1* and *DR-DQ* were determined, and alleles with frequencies >0.02 were subjected to the association study.

Statistical analyses

A Hardy–Weinberg equilibrium (HWE) test was performed using the SNPAlyze version 3 software (DYNACOM Co., Ltd, Chiba, Japan), and SNPs with a P-value for deviation >0.01 were considered to be in HWE as described (7). Calculation of the D' and R^2 values between SNPs and allele/haplotype estimation was performed by the expectation-maximization algorithm using the SNPAlyze version 3 software. The D6S0067i locus showed 19 polymorphic alleles in the same sets of cases and controls, and among them, alleles of 367 and 404 bp in sizes were significantly associated with an elevated risk for lung ADC [odds ratio (OR) = 1.60, $P = 9.9 \times 10^{-3}$ and OR = 1.42, 4.9×10^{-5}, respectively]. Therefore, for the calculation of the D' and R^2 values, genotypes for the D6S0067i polymorphism was expressed by presence or absence of these two alleles (supplementary Table II is available at *Carcinogenesis* Online).

Associations of SNPs/alleles with risks were digitized as crude ORs and ORs adjusted for gender, age and smoking with 95% confidence intervals (CIs) by unconditional logistic regression analysis using the JMP version 6.0 software (SAS Institute, Cary, NC). Variables used for adjustment in each test are described in the footnotes to supplementary Tables are available at *Carcinogenesis* Online. A level of $P < 0.05$ for an OR was judged as significant and that of $0.05 \leq P < 0.1$ was judged as marginal in association studies other than GWAS.

Genotyping of SNPs in the lung cancer susceptibility loci identified by previous GWASs

SNPs in the lung cancer susceptibility loci identified by previous GWASs were genotyped by the TaqMan method. Two intronic SNPs, rs2736100 and rs401681, in the *TERT* and *CLPTM1L* genes (4,21) were genotyped for the 5p15.33 locus against 2343 cases and 1173 controls (subjects of the NCCH set in Table I). Association results of the rs1051730 SNP in the *CHRNA3* gene for the 15q25.1 locus in a subset of the present study population were reported previously (22). Therefore, in this study, 1094 ADC cases and 236 controls that had not been examined in our previous study were genotyped (22). Eight SNPs in the 6p21.33 locus, consisting of rs3117582 and seven SNPs in linkage disequilibrium (LD) with this SNP in Europeans ($D' = 1$ in the HapMap database), were genotyped for 525 ADC cases and 525 controls used for the mapping stage (Table II).

Results and discussion

We performed a GWAS on the risk of lung ADC in a Japanese population for 23 010 polymorphic microsatellite loci. After a three-stage

Table II. GWAS and validation studies to identify loci associated with lung ADC risk

Stage	Subject (no.)		Polymorphic loci analyzed	Genotyping method	Result				
	Case	Control			Loci/allele	Crude OR (95% CI)	*P*-value	Adjusted OR[a] (95% CI)	*P*-value
GWAS									
first[b]	Patients of NCCH (200)	Volunteers enrolled in Tokai University (200)	23 010 microsatellites	Pooled DNA typing	1328 loci	ND	$P\ 2m < 0.05$	ND	ND
Second[c]	Patients of NCCH (200)	Volunteers enrolled in Keio University (200)	431 microsatellites	Pooled DNA typing	17 loci	ND	$P\ 2m < 0.05$	ND	ND
Third[d]	Patients of NCCH (576) consisting of 192 first set. 192 second set and 192 another subjects	Tokai and Keio volunteers and non-cancer NCCH patients (576) consisting f 192 first and 192 non-cancer NCCH patients	17 microsatellites	Individual DNA typing	D6S0067i and other five loci	ND	$P\ 2m = 2.4 \times 10^{-7}$ $P\ 2m =$ 0.012–0.0011 at five other loci	ND	ND
SNP analysis[e]	Randomly selected third stage GWAS subjects (525)	Randomly selected third stage GWAS subjects (525)	56 SNPs	24 SNPs by sequencing 32 SNPs by TaqMan PCR	rs17426593 DQA1*03 DQA1*01	1.51 (1.27–1.80) 1.50 (1.26–1.79) 0.69 (0.58–0.82)	4.2×10^{-6} 6.6×10^{-6} 2.8×10^{-5}	ND	ND
	Same as above	Same as above	rs17426593 and rs34843907	—	DQA1*03 DQA101	1.52 (1.27–1.81) 0.69 (0.58–0.82)	3.4×10^{-6} 2.3×10^{-5}	ND	ND
Validation	Other patients of NCCH (1131)	Other subjects (648) consisting of 9 Tokai and 478 Keio volunteers and 161 non-cancer NCCH patients	Same as above	TaqMan PCR	DQA1*03 DQA1*01	1.27 (1.11–1.45) 0.86 (0.75–0.99)	5.6×10^{-4} 0.030	ND	ND
(Combined analysis)	Patients of NCCH (1656)	Subjects (1173) consisting of 200 Tokai and 635 Keio volunteers and 338 non-cancer	Same as above	—	DQA1*03 DQA1*01	1.35 (1.21–1.51) 0.78 (0.70–0.87)	5.6×10^{-8} 6.0×10^{-6}	1.36 (1.20–1.54) 0.77 (0.68–0.87)	5.3×10^{-7} 1.4×10^{-5}
Validation in another set	Patients of NNGH (84)	Non-cancer patients of NNGH (145)	Same as above	TaqMan PCR	DQA1*03 DQA1*01	1.57 (1.07–2.30) 0.77 (0.52–1.13)	0.022 0.18	1.70 (1.14–2.53) 0.68 (0.49–1.09)	0.0087 0.12

ND, not determined.

[a]Adjusted for sex, age and smoking.

[b]Microsatellite loci (23 010) containing repeat units of 2–6 bp were examined, and 1328 loci showed significant differences in allele distribution.

[c]Among 1328 loci selected in the first GWAS stage, 431 loci that contained repeat units 3–6 bp were examined, and 17 loci showed significant differences in allele distribution.

[d]Seventeen loci selected in the second GWAS stage were examined, and six loci showed significant differences in allele distribution. Only the D6S0067i locus was identified as being significantly different after Bonferroni correction (i.e. $P < 0.05/23\ 010 = 2.2 \times 10^{-6}$). In this state, 576 subjects consisted of two sets of 192 subjects that were chosen from two sets of 200 subjects examined in the first and second GWAS stages, respectively, and another 192 subjects were examined.

[e]In this stage, 525 cases and 525 controls, which were randomly chosen from the 576 cases and 576 controls examined in the third GWAS stage, were examined for 56 SNPs.

screening against 576 ADC cases and 576 controls from the NCCH set (Table I), a locus, D6S0067i, at 6p21.31 was identified as being significantly different in allelic distribution after Bonferroni correction (i.e. $P = 2.4 \times 10^{-7}$, which is $<0.05/23\,010 = 2.2 \times 10^{-6}$) (details in Subjects and Methods and Table II).

The D6S0067i locus was mapped between two LD blocks previously defined (23), one containing four human leukocyte antigen (HLA) class II genes, *HLA-DRA*, *-DRB1*, *-DQA1* and *-DQB1*, and the other containing two pseudogenes, *HLA-DQA2* and *-DQB2* (Figure 1). Therefore, the locus of the strongest association was searched for in the 450 kb region containing these two LD blocks by analyzing 32 SNPs. Five hundred and twenty-five cases and 525 controls, randomly selected from the GWAS subjects, were genotyped by the TaqMan method (Table III). The rs1794282 SNP was monomorphic in the study subjects, whereas the other 31 were polymorphic. An SNP in the *DRA* gene, rs16822586, significantly deviated from the HWE in cases ($P = 0.001$), whereas other SNPs did not deviate in either the cases or the controls, suggesting that SNPs in the regions examined in the present study normally segregated in the Japanese irrespective of lung cancer susceptibility. The 31 SNPs, which were polymorphic in our study population, comprised three LD blocks. The largest difference in allelic distribution between the cases and controls was observed at an intronic SNP in the *DQA1* gene, rs17426593 (OR = 1.51, $P = 4.2 \times 10^{-6}$) (Figure 1), in the block containing four HLA class II genes (LD block 1 in Table III). The D6S0067i polymorphism was in LD ($D' = 0.516$ in controls and $D' = 0.603$ in cases) and showed a considerably high correlation

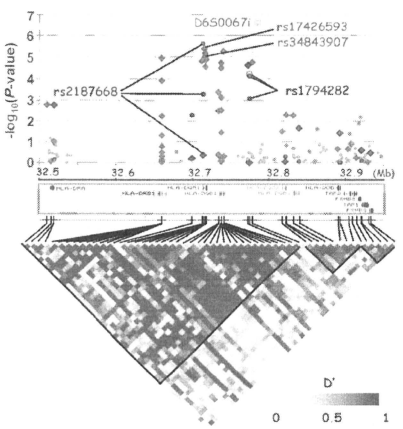

Fig. 1. LD and association with lung cancer risk of polymorphisms in the 6p21.31 locus. The top panel shows association results for polymorphisms and the location of genes. The green square depicts the result for the D6S0067i microsatellite polymorphism, and red lozenges depict those for SNPs in the present study. Circles depict the results of GWASs on European and American populations. Blue circle: results for 1989 cases and 2625 controls in European countries (8); yellow circle: 5095 cases and 5200 controls in European countries and USA (4); purple circle: 2971 cases and 3746 controls in European countries, Canada and USA (5). Results for 10 SNPs commonly analyzed in the present and previous GWASs (indicated in supplementary Table IX is available at *Carcinogenesis* Online) are depicted by bordered lozenges and circles. rs1794282 was monomorphic in the Japanese subjects. The bottom panel shows the LD structure for 55 SNPs in 525 control subjects. Boxes are shaded according to the pairwise D' values. Three LD blocks are indicated by bold black lines.

coefficient ($R^2 = 0.225$ in controls and $R^2 = 0.349$ in cases) with the rs17426593 SNP (Subjects and Methods, supplementary Table II is available at *Carcinogenesis* Online). Therefore, we further examined associations of SNPs in this LD block with lung ADC risk.

Among the four HLA class II genes, the *HLA-DRB1*, *-DQA1* and *-DQB1* genes are known to comprise haplotypes carrying diverse non-synonymous SNPs and express polymorphic antigen proteins (HLA class II alleles) (23). Therefore, we genotyped the same set of case and control subjects for 24 SNPs in the coding exons of the *DRB1*, *DQA1* and *DQB1* genes that discriminate the HLA class II alleles by the sequencing-based typing method (Table III). These 24 SNPs did not deviate from HWE in either the cases or the controls. These SNPs showed LD with the SNPs in LD block 1 (Figure 1), and patterns of LD were quite similar between the cases and controls (supplementary Figure 1 is available at *Carcinogenesis* Online), indicating that distribution of 6p21.31 SNPs on chromosome DNA is not significantly different between these two populations. Many HLA class II alleles, including those for each of the *DRB1*, *DQA1* and *DQB1* genes as well as those for contigs of the three genes (i.e. *DR-DQ* allele), determined by haplotypes for these exonic SNPs showed significantly different distributions between the cases and controls (Table IV). Among them, the DQA1*03 allele showed the largest difference with an OR of 1.50 ($P = 6.6 \times 10^{-6}$) and the DQA1*01 allele was the second largest (OR = 0.69, $P = 2.8 \times 10^{-5}$). Accordingly, several *DR-DQ* alleles containing the DQA1*03 or DQA1*01 allele as well as several *DRB1* and *DQB1* alleles linked to the DQA1*03 or DQA1*01 allele also showed significantly different distributions (Table IV).

Discrimination of HLA alleles using intronic or intergenic SNPs is considered to be appropriate to analyze a large number of samples as an alternative to conventional methods using exonic SNPs due to rapidity and cost effectiveness (23,24). Two exonic SNPs in the *DQA1* gene, DQA1_2_145 and DQA1_2_150, which were genotyped by sequencing, were responsible for discrimination of the DQA1*01 and DQA1*03 alleles. These two SNPs showed high ($R^2 > 0.98$) correlation coefficients with two intronic SNPs in *DQA1*, rs17426593 and rs34843907, respectively (supplementary Table III is available at *Carcinogenesis* Online), which were genotyped by the TaqMan method (supplementary Figure 2 is available at *Carcinogenesis* Online). In fact, DQA1*03 and DQA1*01 alleles deduced by these two intronic SNPs showed high ($R^2 > 0.97$) correlation coefficients with those determined by two exonic SNPs (supplementary Table IV is available at *Carcinogenesis* Online). Thus, DQA1*03 and DQA1*01 alleles were discriminated by combined genotypes of two intronic SNPs, rs17426593 and rs34843907, and the association of *DQA1* alleles with lung ADC risk was further examined in a larger number of subjects by genotyping these two SNPs. Genotyping of an additional 1131 ADC cases and 648 controls in the NCCH set enabled us to calculate combined ORs in 1656 ADC cases and 1173 controls (all subjects of the NCCH set in Table I), and the ORs of the DQA1*03 and DQA1*01 alleles were 1.36 ($P = 5.3 \times 10^{-7}$) and 0.77 ($P = 1.4 \times 10^{-5}$), respectively (Figure 2a, Table II; supplementary Table V is available at *Carcinogenesis* Online). Associations of these alleles with ADC risk were observed both in smokers and non-smokers and both male and female. A significant increase in OR of DQA1*03 for ADC risk was also observed in another set (NNGH set in Table I) of cases and controls (Figure 2a, supplementary Table V is available at *Carcinogenesis* Online), whereas a decrease in OR of DQA1*01 was insignificant. The DQA1*03 allele comprised the same haplotype with the risk allele of the intronic rs17426593 SNP ($R^2 = 0.988$), which showed the largest difference in allelic distribution between the cases and controls (Table III). Therefore, DQA1*03 was defined as a risk allele in the 6p21.31 locus, although it is possible that intronic SNPs rather than exonic SNPs play a causal role in lung ADC susceptibility. Associations of the DQA1*03 allele with risks were further examined for SQC and SCC, two other major histological types of lung cancer to clarify whether the association is specific to ADC or not. Increases in ORs of DQA1*03 for SQC risk were significant both in the NCCH and NNGH sets, whereas ORs for SCC risk, calculated only for the NCCH set, were marginally