SNP309) allows a lower level of expression of MDM2 protein to suppress *TP53* function than the T allele.[10] The risk (326Cys) allele for the Ser326Cys SNP in the *OGG1* gene encodes a DNA glycosylase with a weaker activity to repair 8-hydroxyguanine, an oxidatively damaged promutagenic base, in part, produced by tobacco carcinogens than the 326Ser allele.[7]

Nevertheless, associations of these functional polymorphisms were not investigated in the GWASs above[1–3] because of the lack of probes to discriminate these polymorphisms in the platforms used for GWASs (http://www.ncbi.nlm.nih.gov/snp). Therefore, significance of functional polymorphisms on lung SQC risk and their interactions with GWAS genes in the risk is still unknown. In this study, polymorphisms identified by GWASs and those functional polymorphisms were investigated together for their significance in the risk for lung SQC in the same population by conducting a hospital-based case-control study.

PATIENTS AND METHODS

Case-Control Study

All cases and controls were Japanese and were admissions to National Cancer Center Hospital from 1999 to 2007. Cases were individuals diagnosed with lung SQC by cytological and histological examinations according to World Health Organization classification,[8] whereas controls were those without lung and other cancers and with no history of cancers. All cases and controls, from whom informed consent and blood samples were obtained, were consecutively included in this study without any exclusion criteria. The participation rate was approximately 80%. From each individual, a 20 ml whole-blood sample was obtained. Information on these subjects has been described elsewhere.[5,12] This study was approved by the institutional review boards of the National Cancer Center. Smoking habit was expressed by pack-years, which was defined as the number of cigarette packs smoked daily multiplied by years of smoking. Nonsmokers were defined as individuals of the pack-years = 0. Smokers were defined as individuals of pack-years more than 0 and included both former and current smokers.

Genotyping for Polymorphisms

Genomic DNA from whole blood cells was subjected to genotyping. All 377 cases and 325 controls were previously genotyped for the *CHRNA3*, *TERT*, and *HLA-DQA1* genes.[5] One hundred eighty-eight cases and 203 controls had been genotyped previously for two SNPs, TP53-Arg72Pro and OGG1-Ser326Cys[12]; therefore, the remaining 189 cases and 122 controls were newly genotyped in this study by pyrosequencing as described.[12] All cases and controls were genotyped for the *CYP1A1*, *GSTM1*, and *MDM2* genes. Genotypes for the *CYP1A1* gene were determined by pyrosequencing using primers 5′-GCCTGTCCTCTATCCTTTGGG-3′ and 5′-GCTCCATCAGCATCTATGTGGC-3′ for polymerase chain reaction (PCR) amplification and 5′-GGAAGTGTATCGGTGAGACC-3′ for sequencing, whereas those for the *MDM2* gene were determined by pyrosequencing using primers 5′-GGGAGTTCAGGGTAAAGGTC-3′ and 5′-CACACTAGTGACCCGACAGG-3′ for PCR amplification and 5′-CTGCGGGGCCGCT-3′ for sequencing. The GSTM1-presence or absence polymorphism was analyzed by multiplex PCR.[13]

Statistical Analysis

The strength of association of genotypes with risk was measured as odds ratios (ORs) adjusted for gender, age (≥49, 50–59, 60–69, and ≤70 years), and smoking dosage (0, 1–50, and ≤50 pack-years) with 95% confidence intervals by unconditional logistic regression analysis.[14] Adjustment of smoking was performed with lifetime pack-years for former smokers. ORs for each genotype and those in dominant (i.e., AA versus Aa + aa; A: major allele, a: minor allele) or recessive (i.e., AA + Aa versus aa) modes were calculated. To find pairs of SNPs that interact with each other, associations between genotypes for two polymorphisms were examined in the case subjects. The statistical analyses were performed using the JMP version 6.0 software (SAS Institute Inc., Cary, NC). A level of p less than 0.05 was considered as statistically significant, whereas a level of $0.05 \leq p \leq 0.10$ was considered as marginally significant.

RESULTS AND DISCUSSION

We conducted a case-control study consisting of 377 SQC cases and 325 controls enrolled in the National Cancer Center Hospital, Tokyo, Japan (Table 1). Smoking rates of cases and controls were similar to the ones in a previous cohort study of a Japanese population,[15] and therefore, the cases showed a higher fraction of smokers than the controls. Polymorphisms for two genes identified by GWASs, *CHRNA3* and *HLA-DQA1*, showed significant associations with SQC risk, but a SNP in the *TERT* gene did not show significant associations in our previous study[5] (Table 2). Association of the *TERT* SNP with risk not for SQC but for adenocarcinoma, another major histological type of lung cancer, was also shown by other studies.[3,4]

All the cases and controls were examined for associations with SQC risk of polymorphisms in three DNA repair genes, *TP53*, *MDM*, and *OGG1*; and two metabolic genes, *CYP1A1* and *GSTM1* (Table 2). Significant increases in ORs were observed in one or more genotypes or modes for three

TABLE 1. Subjects for Case-Control Study

		Gender (%)			Smoking (%)	
Subject	No.	Male	Female	Age (Mean ± SD)	Nonsmoker	Smoker
Case	377	340 (90)	37 (10)	62.7 ± 7.6	13 (3)	364 (97)
Control	325	185 (57)	140 (43)	62.5 ± 11.3	179 (55)	146 (45)

Copyright © 2011 by the International Association for the Study of Lung Cancer

Copyright © 2011 by the International Association for the Study of Lung Cancer.

TABLE 2. Association of Polymorphisms in Eight Lung Cancer Susceptibility Genes with Lung SQC Risk

Gene	Polymorphism/ Allele	rs Number	Genotype/ Mode[a]	Control (%)	Case (%)	OR[b]	(95% CI, *p* Value)
GWAS genes							
CHRNA3	C/T	rs1051730	C/C	314 (97)	349 (93)	Reference	
			C/T	10 (3)	25 (7)	2.57	(1.03–6.87, 0.04996)
			T/T	0 (0)	0 (0)	—	(—)
			Dominant			2.57	(1.03–6.87, 0.04996)
			Recessive			—	(—)
TERT	T/G	rs2736100	T/T	116 (36)	142 (38)	Reference	
			T/G	165 (52)	175 (47)	0.74	(0.49–1.11, 0.15)
			G/G	39 (12)	53 (14)	1.22	(0.66–2.29, 0.53)
			Dominant			0.83	(0.56–1.23, 0.36)
			Recessive			1.54	(0.88–2.72, 0.13)
HLA-DQA1	DQA1*03	—	Others/others	121 (37)	120 (32)	Reference	
			Others/*03	169 (52)	194 (51)	0.97	(0.64–1.46, 0.88)
			*03/*03	35 (11)	63 (17)	1.74	(0.94–3.27, 0.080)
			Dominant			1.10	(0.74–1.63, 0.63)
			Recessive			1.78	(1.03–3.13, 0.043)
DNA repair genes							
TP53	Arg72Pro	rs1042522	Arg/Arg	152 (47)	127 (34)	Reference	
			Arg/Pro	139 (43)	191 (51)	1.73	(1.15–2.63, 0.0093)
			Pro/Pro	34 (10)	59 (16)	1.85	(1.03–3.38, 0.042)
			Dominant			1.73	(1.17–2.56, 0.0058)
			Recessive			1.36	(0.79–2.39, 0.27)
MDM2	T/G in intron 1 (SNP309)	rs2279744	T/T	79 (24)	68 (18)	Reference	
			T/G	151 (46)	183 (49)	1.59	(0.96–2.63, 0.071)
			G/G	95 (29)	126 (33)	1.54	(0.90–2.65, 0.12)
			Dominant			1.55	(0.98–2.47, 0.064)
			Recessive			1.09	(0.72–1.63, 0.69)
OGG1	Ser326Cys	rs1052133	Ser/Ser	98 (30)	115 (31)	Reference	
			Ser/Cys	164 (50)	162 (43)	0.86	(0.55–1.33, 0.49)
			Cys/Cys	63 (19)	100 (27)	1.50	(0.90–2.50, 0.12)
			Dominant			1.04	(0.69–1.57, 0.84)
			Recessive			1.65	(1.06–2.57, 0.027)
Metabolic genes							
CYP1A1	Ile462Val	rs1048943	Ile/Ile	188 (58)	196 (52)	Reference	
			Ile/Val	118 (36)	160 (42)	1.51	(1.02–2.24, 0.042)
			Val/Val	19 (6)	21 (6)	0.86	(0.40–1.89, 0.71)
			Dominant			1.40	(0.96–2.04, 0.081)
			Recessive			0.73	(0.34–1.61, 0.43)
GSTM1	Presence/absence	—	Presence	159 (50)	174 (47)	Reference	
			Absence	158 (50)	200 (53)	0.93	(0.64–1.36, 0.72)

[a] ORs in the dominant (ie AA vs Aa + aa, where A, major allele; a, minor allele) and recessive (ie AA + Aa vs aa) modes are shown.
[b] Adjusted for sex, age (≥49, 50–59, 60–69, ≤70 yr), and smoking dosage (0, 1–50, ≤50 pack-years).
SQC, squamous cell carcinoma; OR, odds ratio; CI, confidence interval.

genes, *TP53*, *OGG1*, and *CYP1A1* ($p < 0.05$), consistent with previous reports that their minor alleles are associated with increased risk.[7,9,11] The OR of heterozygotes for the *CYP1A1* SNP was significantly increased, but that of minor homozygotes did not increase. Accordingly, increase in the OR in the dominant mode was only marginal. Genotypes containing the MDM2-G allele showed increased ORs as reported[10]; however, their increases were only marginally significant. Therefore, it was concluded that these two SNPs are marginally significantly associated with lung SQC risk in this study population. The GSTM1-absence genotype did not show an increase in OR. Thus, polymorphisms in two DNA repair genes, *TP53* and *OGG1*, were significantly and those in a DNA repair gene, *MDM2*, and a metabolic gene, *CYP1A1*, were marginally significantly associated with lung SQC risk in a population, in which associations of polymorphisms in the *CHRNA3* and *HLA-DQA1* genes identified by GWASs had been observed.

Copyright © 2011 by the International Association for the Study of Lung Cancer

Copyright © 2011 by the International Association for the Study of Lung Cancer.

TABLE 3. Combined Genotypes for the *TP53* and *MDM2* Genes and Lung SQC Risk

TP53 Genotype	*MDM2* Genotype	Control (%)	Case (%)	OR[a]	(95% CI, *p* Value)	*p* for Interaction[a]
Arg/Arg	T/T	41 (13)	27 (7)	Reference		1.00
Arg/Arg	T/G + G/G	111 (34)	100 (27)	1.48	(0.72–3.10, 0.29)	
Arg/Pro + Pro/Pro	T/T	37 (11)	41 (11)	1.68	(0.68–4.22, 0.26)	
Arg/Pro + Pro/Pro	T/G + G/G	136 (42)	209 (55)	2.53	(1.27–5.11, 0.0081)	

[a] Adjusted for sex, age (≥49, 50–59, 60–69, ≤70 yr), and smoking dosage (0, 1–50, ≤50 pack-years).
SQC, squamous cell carcinoma; OR, odds ratio; CI, confidence interval.

We next examined associations between genotypes for every six pairs of four genes, *CHRNA3*, *HLA-DQA1*, *TP53*, and *OGG1*, whose significant associations with lung SQC risk were observed, in the case subjects to find pairs of polymorphisms that interact with each other. Interaction between these four genes and two genes with marginally significant associations was also examined. In this analysis, four polymorphisms, CHRNA3-C/T, TP53-Arg72Pro, MDM2-T/G, and CYP1A1-Ile462Val, were examined in the dominant mode, whereas the remaining two, HLA-DQA1*03 and OGG1-Ser326Cys, were examined in the recessive mode, based on the results of association with SQC risk (Table 2). Significant or marginally significant association was not observed for any pairs of polymorphisms ($p > 0.1$; Supplementary Table 1, http://links.lww.com/JTO/A33). Previously, polymorphisms of the *MDM2* and *TP53* genes were suggested to interact with each other, because they cooperate in apoptosis.[10] Nevertheless, the ORs of individuals with risk genotypes both for *MDM2* and *TP53* (2.53) were similar to the OR deduced from ORs of individuals with a single-risk genotype for *MDM2* and *TP53* (1.48 [for *MDM2*] × 1.68 [for *TP53*] = 2.49; Table 3). Therefore, *TP53* and *MDM2* would not interactively contribute to SQC risk as reported.[16] Therefore, these genes were suggested to independently contribute to SQC risk, although it remains possible that interaction between polymorphisms were overlooked because of the small number of study subjects.

In this study, association of a few functional polymorphisms in DNA repair and metabolic genes with lung SQC risk was suggested in a population, in which polymorphisms in the GWAS genes showed associations. Therefore, significance of those functional polymorphisms as genetic factors for lung SQC risk was reindicated. Nevertheless, our study has limitations. The present results were obtained from a hospital-based case-control study with a small number of case and control subjects, and frequencies of critical factors involved in lung SQC risk, that is, gender and smoking, are largely different between the cases and controls. Therefore, the statistical power for detection of association in this study was lower than that in a study using a larger number of case and control subjects and was also lower than that using control subjects that were matched to case subjects by gender and smoking. Therefore, associations of some polymorphisms might have been overlooked because of the lack of statistical power. In addition, associations of polymorphisms with lung SQC risk could be confounded by gender or smoking because of insufficient adjustments in statistical tests. Nevertheless, ORs for risk genotypes only in males or smokers were similar to those in all subjects (Supplementary Table 2, http://links.lww.com/JTO/A33). Therefore, it is unlikely that the associations observed in this study were confounded by these factors. On the basis of these results, we would like to emphasize the importance of analyzing various functional polymorphisms together with millions of GWAS marker SNPs. It will be a powerful way for this purpose to analyze those polymorphisms in the populations that were subjected to recent GWASs.

ACKNOWLEDGMENTS

Supported by Grants-in-Aid from the Ministry of Health, Labor, and Welfare for the third-term Comprehensive 10-year Strategy for Cancer Control and for Cancer Research, 19S-1 and 19-9, and the Ministry of Education, Culture, Sports, Science and Technology for Scientific Research on Priority Areas, KAKENHI 20014031.

The authors thank Dr. Daizo Saito for collection of control subjects.

REFERENCES

1. McKay JD, Hung RJ, Gaborieau V, et al. Lung cancer susceptibility locus at 5p15.33. *Nat Genet* 2008;40:1404–1406.
2. Wang Y, Broderick P, Webb E, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet* 2008;40:1407–1409.
3. Landi MT, Chatterjee N, Yu K, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet* 2009;85:679–691.
4. Jin G, Xu L, Shu Y, et al. Common genetic variants on 5p15.33 contribute to risk of lung adenocarcinoma in a Chinese population. *Carcinogenesis* 2009;30:987–990.
5. Kohno T, Kunitoh H, Shimada Y, et al. Individuals susceptible to lung adenocarcinoma defined by combined HLA-DQA1 and TERT genotypes. *Carcinogeneis* 2010;31:834–841.
6. Broderick P, Wang Y, Vijayakrishnan J, et al. Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. *Cancer Res* 2009;69:6633–6641.
7. Bartsch H, Haugen A, Risch A, et al. Genetic susceptibility. In WD Travis, E Brambilla, HK Muller-Hermelink, et al. (Eds.), World Health Organization Classification of Tumors: Pathology and Genetics, Tumours of Lung, Pleura, Thymus and Heart. Lyon: IARC Press, 2004. Pp. 24–25.
8. Hammar SP, Bramilla C, Pugatch B, et al. Squamous cell carcinoma. In WD Travis, E Brambilla, HK Muller-Hermelink, et al. (Eds.), World Health Organization Classification of Tumors: Pathology and Genetics, Tumours of Lung, Pleura, Thymus and Heart. Lyon, IARC Press, 2004. Pp. 26–30.
9. Kiyohara C, Yoshimasu K, Shirakawa T, et al. Genetic polymorphisms and environmental risk of lung cancer: a review. *Rev Environ Health* 2004;19:15–38.

Copyright © 2011 by the International Association for the Study of Lung Cancer

Copyright © 2011 by the International Association for the Study of Lung Cancer.

10. Wilkening S, Bermejo JL, Hemminki K. MDM2 SNP309 and cancer risk: a combined analysis. *Carcinogenesis* 2007;28:2262–2267.
11. Li H, Hao X, Zhang W, et al. The hOGG1 Ser326Cys polymorphism and lung cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2008;17:1739–1745.
12. 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* 2005;114:730–737.
13. Arand M, Mühlbauer R, Hengstler J, et al. A multiplex polymerase chain reaction protocol for the simultaneous analysis of the glutathione S-transferase GSTM1 and GSTT1 polymorphisms. *Anal Biochem* 1996; 236:184–186.
14. Breslow NE, Day NE. Statistical methods in cancer research. Volume I—the analysis of case-control studies. *IARC Sci Publ* 1980:5–338.
15. Sobue T, Yamamoto S, Hara M, et al. Cigarette smoking and subsequent risk of lung cancer by histologic type in middle-aged Japanese men and women: the JPHC study. *Int J Cancer* 2002;99:245–251.
16. Zhang X, Miao X, Guo Y, et al. Genetic polymorphisms in cell cycle regulatory genes MDM2 and TP53 are associated with susceptibility to lung cancer. *Hum Mutat* 2006;27:110–117.

Copyright © 2011 by the International Association for the Study of Lung Cancer

Copyright © 2011 by the International Association for the Study of Lung Cancer.

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:

© The Author 2010. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

Downloaded from carcin.oxfordjournals.org at National Cancer Centre (JMLA) on September 29, 2010

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

Downloaded from carcin.oxfordjournals.org at National Cancer Centre (JMLA) on September 29, 2010

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.

Downloaded from carcin.oxfordjournals.org at National Cancer Centre (JMLA) on September 29, 2010

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

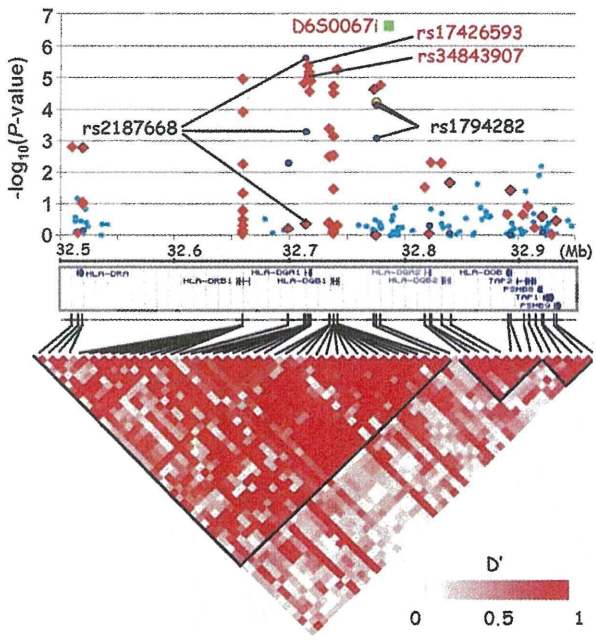

Fig. 1. LD and association with lung cancer risk of polymorphisms in the 6p21.31 locus. The top panel shows association results for polymorphisms and the location of genes. The green square depicts the result for the D6S0067i microsatellite polymorphism, and red lozenges depict those for SNPs in the 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 nonsynonymous 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 nonsmokers 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

Downloaded from carcin.oxfordjournals.org at National Cancer Centre (JMLA) on September 29, 2010

Table III. Differences in the allele distribution of 56 SNPs in the 450 kb region surrounding the D6S0067i locus between 525 lung ADC cases and 525 controls

SNP	Genome location	Gene	Position	Allele	Minor allele frequency		*P*-value	OR	95% CI	Deviation from HWE (*P*)		LD block
					Control	Case				Control	Case	
First 32 SNPs												
rs7773756	32,510,442			T/C	0.375	0.443	0.0016	1.33	1.11–1.58	0.66	0.83	1
rs16822586	32,515,751	*HLA-DRA*	Exon 1	G/C	0.076	0.074	0.86	0.97	0.70–1.34	0.20	0.001	1
rs2239806	32,519,285	*HLA-DRA*	Intron 3	G/A	0.208	0.179	0.089	0.83	0.67–1.03	0.09	0.31	1
rs7192[a]	32,519,624	*HLA-DRA*	Exon 4	G/T	0.444	0.376	0.0016	0.76	0.63–0.90	1.00	0.19	1
rs3129763[a]	32,698,903			G/A	0.064	0.070	0.60	1.10	0.78–1.55	0.91	0.44	1
rs9272346	32,712,350			G/A	0.469	0.564	1.5×10^{-5}	1.46	1.23–1.74	0.11	0.95	1
rs2187668[a]	32,713,862	*HLA-DQA1*	Intron 1	G/A	0.033	0.040	0.42	1.21	0.77–1.91	0.54	0.09	1
rs17426593	32,716,055	*HLA-DQA1*	Intron 1	T/C	0.344	0.442	4.2×10^{-6}	1.51	1.27–1.80	0.55	0.18	1
rs3483907	32,718,037	*HLA-DQA1*	Intron 3	A/C	0.470	0.563	1.3×10^{-6}	1.45	1.22–1.72	0.11	0.91	1
rs28584179	32,734,097			C/T	0.106	0.063	4.1×10^{-4}	0.57	0.41–0.78	0.68	0.50	1
rs17205373	32,734,188			C/G	0.054	0.046	0.42	0.85	0.57–1.26	0.66	0.92	1
rs6906021	32,734,289			T/C	0.377	0.358	0.39	0.92	0.77–1.10	0.82	0.28	1
rs28672722	32,734,515			G/T	0.416	0.353	0.0032	0.77	0.64–0.91	0.59	0.86	1
rs28746825	32,741,450	*HLA-DQB1*	Intron 1	A/G	0.365	0.463	5.4×10^{-6}	1.50	1.26–1.79	0.22	0.26	1
rs34692792	32,741,519	*HLA-DQB1*	Intron 1	T/C	0.218	0.205	0.47	0.93	0.75–1.14	0.47	0.29	1
rs2647012[a]	32,772,436			G/A	0.214	0.143	2.3×10^{-5}	0.61	0.49–0.77	0.37	0.22	1
rs1794282[a]	32,774,504			G/A	0.000	0.000	—	—	—	—	—	—
rs2856717	32,778,286			C/T	0.219	0.147	1.7×10^{-6}	0.61	0.49–0.77	0.19	0.33	1
rs2051600	32,817,287	*HLA-DQA2*	Intron 1	C/T	0.188	0.153	0.030	0.78	0.62–0.98	0.62	0.33	2
rs2239800[a]	32,821,245	*HLA-DQA2*	Intron 2	T/C	0.289	0.292	0.88	1.01	0.84–1.23	0.92	0.47	2
rs2071798	32,822,570	*HLA-DQA2*	3′UTR	T/C	0.360	0.302	0.0048	0.77	0.64–0.92	0.33	0.92	2
rs9276558	32,832,039			G/A	0.358	0.301	0.0051	0.77	0.64–0.92	0.28	0.94	2
rs1573649[a]	32,839,236			T/C	0.426	0.377	0.021	0.81	0.68–0.92	0.11	0.77	2
rs2071475	32,890,365	*HLA-DOB*	Intron 2	C/T	0.238	0.261	0.21	1.14	0.93–1.38	0.53	0.24	2
rs2071469[a]	32,892,761	*HLA-DOB*	5′UTR	G/A	0.424	0.469	0.037	1.20	1.01–1.43	0.13	0.66	2
rs241455	32,903,997	*TAP2*	3′UTR	G/T	0.314	0.339	0.21	1.12	0.94–1.35	0.35	0.17	2
rs1800454	32,908,390	*TAP2*	Exon 6	G/A	0.126	0.150	0.12	1.22	0.95–1.56	0.51	0.20	2/3
rs2071552	32,914,439	*TAP2*	5′UTR	T/C	0.418	0.406	0.56	0.95	0.80–1.13	0.53	0.22	3
rs2071463	32,920,506	*PSMB8*	5′UTR	G/A	0.380	0.404	0.25	1.11	0.93–1.32	0.74	0.18	3
rs1057373[a]	32,921,257	*TAP1*	3′UTR	G/T	0.105	0.121	0.23	1.18	0.90–1.55	0.72	0.18	3
rs2071480	32,929,837			G/T	0.339	0.341	0.95	1.01	0.84–1.20	0.16	0.23	3
rs17587[a]	32,933,068	*PSMB9*	Exon 3	G/A	0.243	0.225	0.33	0.90	0.74–1.11	0.47	0.71	3
Additional 24 SNPs for HLA allele discrimination												
DRB1_2_244	32,659,890	*HLA-DRB1*	Exon 2	G/T	0.269	0.271	0.88	1.01	0.84–1.23	0.49	0.88	1
DRB1_2_160	32,659,974	*HLA-DRB1*	Exon 2	C/A	0.029	0.023	0.69	0.79	0.47–1.33	0.52	0.56	1
DRB1_2_156	32,659,978	*HLA-DRB1*	Exon 2	G/A	0.204	0.259	0.0055	1.37	1.12–1.66	0.83	1.00	1
DRB1_2_127	32,660,007	*HLA-DRB1*	Exon 2	A/T	0.418	0.325	1.1×10^{-5}	0.67	0.56–0.80	0.16	0.28	1
DRB1_2_106	32,660,028	*HLA-DRB1*	Exon 2	T/A	0.003	0.004	0.70	1.35	0.30–6.03	0.95	0.93	1
DRB1_2_84	32,660,050	*HLA-DRB1*	Exon 2	A/C	0.191	0.266	1.2×10^{-4}	1.53	1.26–1.87	0.86	0.57	1
DRB1_2_81	32,660,053	*HLA-DRB1*	Exon 2	T/C	0.224	0.207	0.30	0.90	0.74–1.10	0.47	0.66	1
DRB1_2_64	32,660,070	*HLA-DRB1*	Exon 2	T/A	0.130	0.154	0.047	1.22	0.97–1.55	0.60	0.79	1
DRB1_2_61	32,660,073	*HLA-DRB1*	Exon 2	G/A	0.002	0.000	0.16	0.00	—	0.96	1.00	1
DRB1_2_33	32,660,101	*HLA-DRB1*	Exon 2	C/T	0.172	0.174	0.91	1.01	0.82–1.26	0.10	0.47	1
DRB1_2_136	32,717,200	*HLA-DQA1*	Exon 2	A/T	0.464	0.568	8.4×10^{-6}	1.52	1.28–1.79	0.16	0.84	1
DRB1_2_141	32,717,205	*HLA-DQA1*	Exon 2	A/C	0.348	0.450	9.7×10^{-6}	1.54	1.30–1.83	0.91	0.27	1
DRB1_2_145	32,717,209	*HLA-DQA1*	Exon 2	A/G	0.348	0.444	6.7×10^{-6}	1.50	1.26–1.79	0.51	0.26	1
DRB1_2_150	32,717,304	*HLA-DQA1*	Exon 2	G/A	0.470	0.562	2.8×10^{-5}	1.44	1.22–1.71	0.11	0.83	1
DRB1_2_156	32,740,667	*HLA-DQB1*	Exon 2	G/A	0.029	0.031	0.90	1.07	0.65–1.76	0.49	0.47	1
DRB1_2_145	32,740,678	*HLA-DQB1*	Exon 2	A/T	0.471	0.560	1.8×10^{-5}	1.43	1.21–1.69	0.04	0.87	1
DRB1_2_134	32,740,689	*HLA-DQB1*	Exon 2	G/A	0.065	0.059	0.72	0.90	0.64–1.27	0.53	0.39	1
DRB1_2_131	32,740,692	*HLA-DQB1*	Exon 2	G/C	0.092	0.058	0.0029	0.61	0.44–0.84	0.19	0.86	1
DRB1_2_121	32,740,702	*HLA-DQB1*	Exon 2	G/A	0.097	0.099	0.81	1.03	0.77–1.36	0.73	0.15	1
DRB1_2_101	32,740,722	*HLA-DQB1*	Exon 2	G/A	0.315	0.355	0.034	1.20	1.00–1.43	0.71	0.96	1
DRB1_2_100	32,740,723	*HLA-DQB1*	Exon 2	C/T	0.335	0.323	0.55	0.95	0.79–1.13	0.43	0.51	1
DRB1_2_68	32,740,755	*HLA-DQB1*	Exon 2	G/A	0.114	0.065	3.2×10^{-5}	0.54	0.40–0.73	0.42	0.50	1
DRB1_2_55	32,740,768	*HLA-DQB1*	Exon 2	G/T	0.103	0.153	7.3×10^{-5}	1.57	1.22–2.03	0.72	0.61	1
DRB1_2_27	32,740,796	*HLA-DQB1*	Exon 2	A/C	0.121	0.116	0.80	0.96	0.74–1.24	0.57	0.64	1

[a]SNPs examined in other GWASs (4,5,8).

increased for DQA1*03 (Figure 2a, supplementary Table V is available at *Carcinogenesis* Online). Therefore, involvements of *HLA-DQA1* not only in ADC risk but also in other histological types of lung cancer were suggested.

The 6p21.31 locus maps 1 Mb proximal to *BAT3-MSH5*, another lung cancer susceptibility locus at 6p21.33 identified by a GWAS on Europeans and Americans (4). Therefore, we next examined an SNP in this region, rs3117582, which showed a significant association in that study (4), in a set of 525 ADC cases and 525 controls (subjects for the SNP analysis stage, Table II). It was monomorphic for the protective allele in these subjects. We therefore examined seven SNPs in LD with this SNP in Europeans (i.e. $D' = 1$ in the HapMap data); however, associations of these loci were weaker than those of the 6p21.31 locus, and these SNPs comprised a distinct LD block from

Downloaded from carcin.oxfordjournals.org at National Cancer Centre (JMLA) on September 29, 2010

Table IV. Differences in the distribution of the HLA class II alleles between 525 lung ADC cases and 525 controls

Gene	No.	Allele[a]	Frequency		OR	(95% CI)	*P*-value by χ^2 test
			Control	Case			
DRB1	1	DRB1*1502	0.148	0.125	0.82	(064–105)	0.12
	2	DRB1*0901	0.117	0.157	1.40	(1.09–1.81)	0.0079
	3	DRB1*0405	0.106	0.154	1.53	(1.18–1.98)	0.0012
	4	DRB1*1302	0.097	0.058	0.58	(0.42–0.81)	0.0011
	5	DRB1*1501	0.091	0.055	0.58	(0.41–0.81)	0.0013
	6	DRB1*0803	0.073	0.075	1.02	(0.74–1.42)	0.90
	7	DRB1*0101	0.057	0.033	0.56	(0.37–0.86)	0.0077
	8	DRB1*1401/1405/1406/1412/1429	0.057	0.069	1.24	(0.87–1.77)	0.23
	9	DRB1*0802	0.052	0.037	0.70	(0.46–1.07)	0.10
	10	DRB1*0403/0404/0406	0.044	0.073	1.73	(1.19–2.52)	0.0040
	11	DRB1*1201/1202	0.040	0.050	1.25	(0.82–1.89)	0.30
	12	DRB1*1101	0.026	0.022	0.81	(0.46–1.42)	0.47
	13	DRB1*0410	0.021	0.016	0.78	(0.41–1.48)	0.44
	Total		0.929	0.924			
DQA1	1	DQA1*01	0.530	0.438	0.69	(0.58–1.14)	2.8×10^{-5}
	2	DQA1*03	0.348	0.444	1.50	(1.26–1.79)	6.6×10^{-6}
	3	DQA1*04/05/06	0.122	0.118	0.88	(0.68–1.14)	0.43
	Total		1.000	1.000			
DQB1	1	DQB1*0601	0.218	0.205	0.92	(0.75–1.14)	0.46
	2	DQB1*0303	0.139	0.164	1.22	(0.96–1.54)	0.11
	3	DQB1*0401	0.104	0.153	1.54	(1.19–2.00)	0.0010
	4	DQB1*0604	0.097	0.058	0.57	(0.41–0.79)	7.4×10^{-4}
	5	DQB1*0301	0.093	0.093	1.01	(0.75–1.35)	0.97
	6	DQB1*0602	0.086	0.053	0.59	(0.42–0.84)	0.0028
	7	DQB1*0302	0.081	0.092	1.15	(0.85–1.57)	0.35
	8	DQB1*0501	0.061	0.059	0.96	(0.67–1.37)	0.82
	9	DQB1*0402	0.042	0.051	1.21	(0.80–1.82)	0.36
	10	DQB1*0502	0.030	0.027	0.90	(0.54–1.51)	0.70
	11	DQB1*0503	0.028	0.028	1.03	(0.61–1.72)	0.93
	Total		0.979	0.983			
DR-DQ	1	DRB1*1502-DQA1*01-DQB1*0601	0.150	0.130	0.85	(0.66–1.09)	0.19
	2	DRB1*0901-DQA1*03-DQB1*0303	0.119	0.153	1.34	(1.05–1.73)	0.021
	3	DRB1*0405-DQA1*03-DQB1*0401	0.101	0.144	1.51	(1.16–1.96)	0.0022
	4	DRB1*0302-DQA1*01-DQB1*0604	0.094	0.057	0.58	(0.42–0.81)	0.0013
	5	DRB1*1501-DQA1*01-DQB1*0602	0.082	0.051	0.60	(0.42–0.85)	0.0042
	6	DRB1*0803-DQA1*01-DQB1*0601	0.070	0.072	1.02	(0.73–1.43)	0.89
	7	DRB1*0101-DQA1*01-DQB1*0501	0.047	0.034	0.72	(0.46–1.12)	0.14
	8	DRB1*0403/0404/0406/-DQA1*03-DQB1*0302	0.040	0.068	1.72	(1.17–2.54)	0.0058
	9	DRB1*1201/1202-DQA1*04/05/06-DQB1*0301	0.032	0.032	0.99	(0.61–1.61)	0.97
	10	DRB1*1401/1405/1406/1412/1429-DQA1*01-DQB1*0503	0.028	0.031	1.13	(0.68–1.88)	0.64
	11	DRB1*1101-DQA1*04/05/06-DQB1*0301	0.025	0.017	0.68	(0.37–1.25)	0.21
	12	DRB1*0802-DQA1*04/05/06-DQB1*0402	0.024	0.026	1.10	(0.64–1.91)	0.73
	Total		0.812	0.815			

[a]*DRB1* and *DQB1* alleles linked to the DQA1*03 or DQA1*01 alleles and DR-DQ alleles containing the DQA1*03 or DQA1*01 alleles, which were significantly associated with lung ADC risk, are underlined.

the 6p21.31 locus containing four HLA class II genes (supplementary Figure 3 is available at *Carcinogenesis* Online). Therefore, we concluded that 6p21.31 is a novel lung ADC susceptibility locus on chromosome 6p.

Next, we examined associations of SNPs in other lung cancer susceptibility loci (4–9) in 1656 ADC cases and 1173 controls (all subjects of the NCCH set in Table I). Two SNPs, rs2736100 and rs401681, were examined for the 5p15.33 locus, and the former located in intron 2 of the *TERT* gene showed a stronger association than the latter. The association was observed only in ADC, but not in SQC and SCC (supplementary Table VI is available at *Carcinogenesis* Online) as reported recently (11,21). An SNP in the *CHRNA3* gene at 15q25.1, rs1051730, showed a significant association with risks for ADC, SQC and SCC in our previous study (22).

Therefore, combined effects among the *HLA-DQA1*, *TERT* and *CHRNA3* loci with lung ADC risk were further investigated. Genotypes with risk alleles for each locus showed significantly increased ORs of 1.32–2.21, except for homozygotes for the minor allele of *CHRNA3* (Figure 2b, supplementary Table VII is available at *Carcinogenesis* Online). When ORs were calculated according to the number of risk alleles for two of these three genes, *HLA-DQA1* and *TERT*, there was an increasing trend with increasing number of risk alleles (per-risk allele OR = 1.43, $P = 7.8 \times 10^{-16}$), reaching 4.76 for carriers of all four risk alleles (Figure 2b, supplementary Table VIII is available at *Carcinogenesis* Online). These two alleles independently conferred the risk (*P* for interaction = 0.88). The present results indicated that individuals susceptible to ADC can be defined by combined genotypes of *HLA-DQA1* and *TERT*. There was also an increasing trend for the *TERT* and *CHRNA3* combination with a per-risk allele OR of 1.48. OR reached 4.27 for carriers of three or four risk alleles when heterozygotes and homozygotes for the *CHRNA3* risk were combined due to a small number of homozygotes (supplementary Table VIII is available at *Carcinogenesis* Online). Increases in OR by the combination of *HLA-DQA1* and *CHRNA3* were not evident, and a negative interaction was suggested ($P = 0.083$). However, it might be due to the small number of homozygotes for the *CHRNA3* risk allele. Accordingly, when

Downloaded from carcin.oxfordjournals.org at National Cancer Centre (NMLA) on September 29, 2010