Table I Oncogenes and Tumor Suppressor Genes Genetically Altered in Lung Cancer

Year[a]	Gene	Alteration	ADC	SQC	SCLC
1983	MYC[b]	Amp	+	+	++
1987	KRAS[b]	Mut	+[c]	−	−
1988	RB (RB1)[d]	Del, Mut	−	−	++[c]
1989	TP53[d]	Del, Mut	++	+++	+++
1994	p16 (CDKN2A)[d]	Del, Met, Mut	+[c]	+[c]	−
1998	PTEN[d]	Del, Met, Mut	+	+	++
2002	LKB1[d]	Del, Mut	++	−	−
2004	EGFR[b]	Small Del, Mut	++[c]	−	−
2005	PTPRD[d]	Del, Mut	+	+	+
2007	ALK[b]	Inv	+[c]	−	−

Abbreviations: ADC, adenocarcinoma; SQC, squamous cell carcinoma; SCLC, small cell lung cancer; Amp, amplification; Mut, mutation; Del, deletion; Met, methylation; Inv, inversion.
[a]Year of genetic alterations identified in lung cancer cells.
[b]Oncogene.
[c]Occurrence of genetic alterations in a mutually exclusive manner.
[d]Tumor suppressor gene.

and SCLC. RB and p16 are inactivated in a type-specific and mutually exclusive manner. Namely, RB is specifically inactivated in SCLC whereas p16 is specifically in ADC and SQC. KRAS, LKB1, EGFR, and ALK are specifically mutated in ADC. It is noted that KRAS, EGFR, and ALK are mutated in a mutually exclusive manner in ADC; therefore, in the light of accumulated genetic alterations, ADC can be further divided into at least four different types: KRAS-type, EGFR-type, ALK-type, and non-KRAS/EGFR/ALK-type (Govindan, 2010; Subramanian and Govindan, 2008; Travis *et al.*, 2004). EGFR-type is the major type of lung ADC in Asian people (30–50%) and the fraction of EGFR-type in Asian people is higher than that in American/European people (i.e., individuals of European decent; 10–20%), representing a geographic and/or ethnic difference in lung cancer development. Instead, the fraction of KRAS-type in American/European people (20–40%) is higher than that in Asian people (10%). The fraction of ALK-type is ~5% in both populations.

All three major types of lung cancer are associated with tobacco smoking; however, associations are much stronger in SQC and SCLC than ADC (Sobue *et al.*, 2002). Roughly speaking, more than 90% of patients with SQC or SCLC are smokers, whereas only ~50% of patients with ADC are smokers. Therefore, by considering the incidence of these three types of lung cancer, ~25% of lung cancer cases are not attributed to smoking. The proportion of patients with lung cancer in never-smokers (less than 100 cigarettes in their life time) is higher in Asian populations than American/European populations. In this subset of lung cancer, mostly classified into

ADC, environmental factors causing the accumulation of multiple genetic alterations have been poorly understood. In particular, EGFR-type and ALK-type are frequent in never-smokers, while KRAS-type and non-KRAS/EGFR/ALK-type are frequent in smokers.

LKB1 alterations preferentially accumulate in the KRAS-type, and both LKB1 and KRAS are genetically altered more frequently in American/European people than in Asian people. In contrast, EGFR mutations occur more frequently in Asian people than in American/European people, as described above. Therefore, even though both KRAS-type and EGFR-type are histologically classified into ADC, they are thought to be different diseases from each other (Govindan, 2010; Subramanian and Govindan, 2008; Travis *et al.*, 2004). From this point of view, it can be said that American/European populations are more susceptible to KRAS-type ADC, while Asian populations is more susceptible to EGFR-type ADC, due to the difference in either or both genetic and environmental factors.

III. CANDIDATE GENE ASSOCIATION STUDIES

Histological heterogeneity of lung cancer has been known for many years. However, lung cancers in never-smokers have not been classified into a different disease until recently. Therefore, in the last two decades, genetic susceptibility for tobacco-induced lung cancer has been extensively investigated by a candidate gene approach focusing on the metabolism of tobacco smoke carcinogens and the suppression of tobacco-induced genetic alterations. Lung cancer cells developed in smokers have been shown to have a unique mutation spectrum, with an excess of G:C to T:A transversions (Hollstein *et al.*, 1991; Le Calvez *et al.*, 2005). Therefore, associations of metabolic enzyme activities as well as DNA repair activities to induce or prevent G:C to T:A transversions have been a focus of genetic susceptibility studies in lung cancer (Govindan, 2010). Benzo[a]pyrene (BP) is a major polyaromatic hydrocarbon (PAH) in tobacco smoke, and benzopyrene-diol-epoxide (BPDE), a metabolite of BP (Alexandrov *et al.*, 2002; Rubin, 2001), forms a DNA adduct and induces G:C to T:A transversions at hot spot codons in the p53 gene in lung cancers of smokers (Le Calvez *et al.*, 2005). Cytochrome P450 (CYP)-related enzymes and glutathione-*S*-transferases (GSTs) are representative metabolic enzymes for tobacco smoke carcinogens because their polymorphisms have been extensively investigated in association with risk for lung cancer, particularly for SQC (Bartsch *et al.*, 2000). CYP1A1 bioactivates PAHs, such as BP, and a single nucleotide polymorphism (SNP) of Ile462Val in the CYP1A1 gene causes the difference in the enzymatic ability. The 462Val allele encodes a protein with a higher activity

to bioactivate PAHs than the 462Ile allele, and individuals carrying the 462Val allele have been shown to have higher risk to lung cancer than those carrying the 462Ile allele. In contrast, GSTs detoxify tobacco carcinogens such as PAH, and individuals lacking GSTM1 (null-type in an insertion/deletion polymorphism) have been shown to have an elevated risk for lung cancer. There have also been extensive works on the role of DNA repair genes as a determinant of inherited susceptibility to lung cancer (Schwartz *et al.*, 2007). For instance, in 1999, we first reported the possible contribution of OGG1 SNPs to lung SQC risk (Sugimura *et al.*, 1999). 8-oxo-deoxyguanosine (8-oxo-dG) is a major form of oxidative DNA damage induced by reactive free radicals and is highly mutagenic with frequent induction of G:C to T:A transversions both *in vitro* and *in vivo*. The OGG1 gene encodes an oxo-guanine DNA glycosylase that removes 8-oxo-dG from double-stranded DNA, thus preventing the occurrence of G:C to T:A transversions induced by 8-oxo-dG. The risk (326Cys) allele for the Ser326Cys SNP in the OGG1 gene encodes a DNA glycosylase with a weaker activity to repair 8-oxo-dG, in part produced by tobacco carcinogens, than the 326Ser allele (Kohno *et al.*, 1998; Yamane *et al.*, 2004). TP53 and MDM2 are also representative DNA repair genes associated with lung cancer risk (Bond and Levine, 2007; Imyanitov, 2009; Whibley *et al.*, 2009). The risk (72Pro) allele for the TP53-Arg72Pro SNP in the TP53 gene encodes a protein with a weaker apoptotic activity, thus allowing better survival of cells with DNA damages than the 72Arg allele. The risk (G) allele for a T/G SNP in the promoter region of the MDM2 gene (which is called MDM2 SNP309) allows a lower level of expression of MDM2 protein to suppress TP53 function than the T allele.

Table II summarizes lung cancer susceptibility genes identified to date by candidate gene association studies, and confirmed as being consistently associated with lung cancer risk by recent meta-analyses or pooled analyses of various studies (Dai *et al.*, 2009; Dong *et al.*, 2008; Kohno *et al.*, 2006; Li *et al.*, 2008; Wilkening *et al.*, 2007; Ye *et al.*, 2006). However, in most of these analyses, histological differences are not critically analyzed, probably because histological types were not available in some studies selected for meta-analyses. Therefore, the contribution of those polymorphisms to each histological type of lung cancers is not clear at present, although most polymorphisms are thought to be associated with the development of smoking-related lung cancer. In addition, results are different among several meta-analyses due to differences in the studies selected for meta-analyses. Such differences would be due to the quality of each study selected for the analysis. Therefore, to make uniform the quality among studies, several international consortiums have been established to date; thus, more reliable data will be available for various functional SNPs in the near future.

Table II Lung Cancer Susceptibility Genes Identified by Candidate Gene Association Studies (Meta-Analysis)

Gene	Gene product	Function	Polymorphism	SNP ID	Odd ratio	References
CYP1A1	Cytochrome P450	Phase I metabolism	Ile462Val	rs1048943	2.36	Dong *et al.* (2008)
mEH (EPHX1)	Epoxide hydrolase	Phase I metabolism	His113Tyr	rs1051740	0.70	Dong *et al.* (2008)
MPO	Myeloperoxidase	Phase I metabolism	G-463A	rs2333227	0.71	Dong *et al.* (2008)
GSTM1	Glutathione-*S*-transferase	Phase II metabolism	Presence/null	–	1.18	Ye *et al.* (2006)
GSTT1	Glutathione-*S*-transferase	Phase II metabolism	Presence/null	–	1.28	Dong *et al.* (2008)
XPA	Nucleotide excision repair protein	Mutation suppression	G-23A	rs1800975	0.73	Dong *et al.* (2008)
XPC	Nucleotide excision repair protein	Mutation suppression	Lys939Gln	rs2228001	1.30	Dong *et al.* (2008)
XPD	Nucleotide excision repair protein	Mutation suppression	Lys751Gln	rs1052559	1.30	Dong *et al.* (2008)
XRCC1	Base excision repair protein	Mutation suppression	Arg399Gln	rs25487	1.34	Dong *et al.* (2008)
OGG1	Base excision repair protein	Mutation suppression	Ser326Cys	rs1052133	1.32	Li *et al.* (2008)
OGG1 (ADC)					1.43	Kohno *et al.* (2006)
TP53	Transcription factor	Cell cycle/death regulation	Arg72Pro	rs1042522	1.20	Dai *et al.* (2009)
MDM2	Ubiquitine ligase	Cell cycle/death regulation	T309G	rs2279744	1.27	Wilkening *et al.* (2007)

IV. GENOME-WIDE ASSOCIATION STUDIES

Recent GWA studies have lead to the identification of a number of candidate lung cancer susceptibility genes (Table III). Three chromosomal loci, 15q24-25.1, 5p15.33, and 6p21, have been shown to be associated with lung cancer risk in Europeans and Americans (Amos *et al.*, 2008; Hung *et al.*, 2008; McKay *et al.*, 2008; Thorgeirsson *et al.*, 2008; Wang *et al.*, 2008). The chromosome 15q24-25.1 region contains the nicotinic acetylcholine receptor subunit genes, CHRNA3 and CHRNA5, and their products are expressed in pulmonary epithelial cells including neuroendocrine cells and bind to nicotine. Therefore, the association of this locus with lung cancer risk could be primarily mediated by nicotine dependence as described below. The 5p15.33 region contains the TERT (telomerase reverse transcriptase) gene and the CLPTM1L (cleft lip and palate transmembrane protein 1-like) gene. TERT is known to function in telomere replication and maintenance, and to promote epithelial cell proliferation. CLPTM1L was identified through screening for cisplatin (CDDP) resistance-related genes. Interestingly, this locus is associated with the risk for ADC but not for SQC or SCLC, suggesting the weak association of this locus with lung cancer risk in smokers (Landi *et al.*, 2009). Indeed, the 5p15.33 (TERT-CLPTM1L) genotypes were shown to be associated with lung ADC risk in never-smokers (Wang *et al.*, 2010). Associations of the 5p15.33 genotypes have been detected not only in lung cancer but also in various other types of cancers, including cancers of the brain, bladder, prostate, uterine cervix, and skin (Rafnar *et al.*, 2009; Stacey *et al.*, 2009). Therefore, it is likely that genotypes of this locus are associated with the development of a wide variety of cancers. Association with lung cancer risk of a SNP in the CHRNA3 gene at 15q24-25.1 was replicated in a Japanese population, although the frequency of the risk variant in the Japanese is much lower than that in Europeans and Americans (Kohno *et al.*, 2010). In a Chinese population, the association of SNPs in this locus with lung cancer risk was also replicated; however, risk variants seem to be different from Europeans and Americans (Wu *et al.*, 2009). Interestingly, in Asian populations, the associations of 15q24-25.1 SNPs with lung cancer risk were independent of smoking behavior. Associations with lung cancer risk of SNPs in the 5p15.33 region were validated in both Japanese and Chinese populations (Jin *et al.*, 2009; Shiraishi *et al.*, 2009), and the TERT gene was indicated to be a more likely target rather than the CLPTM1L gene.

The 6p21 region contains the BAT3 (HLA-B associated transcript 3) and MSH5 (mutS homolog 5) genes. BAT3 protein complexes with a histone acetyltransferase (HAT), p300, which acetylates p53 protein in response to DNA damage. MSH5 is a gene involved in DNA mismatch repair.

Table III Lung Cancer Susceptibility Genes Identified by Genome-Wide Association Studies (GWAS)

Chromosomal location	Gene	Risk allele	Geographic area	Frequency in control population	Allele OR (95% CI, P)	Case/control	References
15q25.1	*CHRNA3*[a]	rs1051730-T	USA and UK	0.33	1.32 (1.23–1.39, 7.0×10^{-18})	2013/3062	Amos *et al.* (2008)
			Central Europe	0.33	1.30 (1.19–1.43, 5.4×10^{-9})	1922/2520	Hung *et al.* (2008)
			Iceland	0.35	1.31 (1.19–1.44, 1.5×10^{-8})	1024/32,244	Thorgeirsson *et al.* (200
			Europe, USA and Canada	0.35	1.31 (1.27–1.36, 1.9×10^{-51})	13,300/19,666	Landi *et al.* (2009)
			Japan	0.02	1.79 (1.19–2.78, 0.0095)	2343/1173	Kohno *et al.* (2010)
5p15.33	*TERT*[b]	rs2736100-G	Europe and Canada	0.49	1.19 (1.11–1.27, 2×10^{-6})	2971/3746	McKay *et al.* (2008)
			Europe, USA and Canada	0.50	1.12 (1.08–1.16, 1.6×10^{-10})	13,300/19,666	Landi *et al.* (2009)
			USA and Europe	0.51	1.15 (1.10–1.20, 1×10^{-10})	9162/11,812	Truong *et al.* (2010)
			Asia, USA and Canada	0.39	1.23 (1.12–1.35, 2×10^{-5})	1686/2101	
			China	0.42	1.16 (1.03–1.30)	1221/1344	Jin *et al.* (2009)
			Japan	0.38	1.38 (1.23–1.56, 6.3×10^{-8})	2343/1173	Kohno *et al.* (2010)
	CLPTM1L[c]	rs402710-C	Europe and Canada	0.68	1.18 (1.12–1.24, 2×10^{-7})	2971/3746	McKay *et al.* (2008)
			USA and Europe	0.65	1.14 (1.09–1.19, 5×10^{-8})	8860/9198	Truong *et al.* (2010)
			Asia, USA and Canada	0.68	1.15 (1.04–1.27, 0.007)	1680/2117	
			China	0.69	1.09 (0.97–1.24)	1221/1344	Jin *et al.* (2009)
			Japan	0.65	1.10 (0.97–1.23, 0.15)	2343/1173	Kohno *et al.* (2010)
		rs401681-C	USA, UK and central Europe	0.55	1.15 (1.09–1.19, 7.9×10^{-9})	5095/5200	Wang *et al.* (2008)
			Europe, USA and Canada	0.56	1.12 (1.09–1.16, 6.7×10^{-11})	13,300/19,666	Landi *et al.* (2009)
			Japan	0.67	1.14 (1.01–1.28, 0.044)	2343/1173	Kohno *et al.* (2010)
6p21.33	*BAT3*[d]*–MSH5*[e]	rs3117582-C	USA, UK and central Europe	0.10	1.24 (1.16–1.33, 5.0×10^{-10})	5095/5200	Wang *et al.* (2008)
			Europe, USA and Canada	0.10	1.22 (1.15–1.29, 4.8×10^{-12})	13,300/19,666	Landi *et al.* (2009)
			Japan	0	–	525/525	Kohno *et al.* (2010)
6p21.31	*HLA-DQA1*[f]	*03	Japan	0.36	1.36 (1.20–1.54, 5.3×10^{-7})	1656/1173	Kohno *et al.* (2010)

[a] Cholinergic receptor, nicotinic, alpha 3.
[b] Telomerase reverse transcriptase.
[c] Cleft lip and palate transmembrane protein 1-like protein.
[d] HLA-B associated transcript 3.
[e] mutS homolog 5.
[f] Major histocompatibility complex, class II, DQ alpha 1.

Therefore, both genes are attractive candidates for lung cancer susceptibility genes; however, a recent pooled analysis from the international lung cancer consortium did not replicate the association of these SNPs with lung cancer risk (Truong *et al.*, 2010). The significance of SNPs at 6p21 on lung cancer risk of Asians has not been fully investigated; however, our recent GWA study on the Japanese using 23,000 microsatellite markers for the screening indicated that the HLA-DQA1 gene, encoding a HLA (human leukocyte antigen)-class II protein, mapped at 6p21.31 is the most significant region at 6p21 (Jin *et al.*, 2009). DQA1*03 of the HLA-DQA1 gene was defined as a risk allele with odds ratio (OR) of 1.36 (95%CI = 1.21–1.54, $P = 5.3 \times 10^{-7}$) by analysis of 1656 ADC cases and 1173 controls. The HLA-DQA1 locus was mapped 1-Mb proximal to the BAT3–MSH5 locus. Therefore, we further examined a SNP in the BAT3–MSH5 locus, rs3117582, which showed a significant association in Europeans and Americans. It was monomorphic for the protective allele in the Japanese. We therefore examined seven SNPs in linkage disequilibrium (LD) with this SNP in Europeans (i.e., $D' = 1$ in the HapMap data); however, associations of these SNPs in the BAT3–MSH5 locus were weaker than genotypes of the HLA-DQA1 locus, and these SNPs comprised a distinct LD block from the locus containing the HLA-DQA1 gene. Therefore, it was concluded that the 6p21.31 region containing the HLA-DQA1 locus is a lung ADC susceptibility locus distinct from the BAT3–MSH5 locus at 6p21.33. In a recent meta-analysis, the associations of the BAT3–MSH5 locus were shown to vary among studies (Broderick *et al.*, 2009). Thus, it is possible that the BAT3–MSH5 locus and also the HLA-DQA1 locus could be affected by the difference in population structure since it is located near/in the locus for major histocompatibility complex, a highly polymorphic locus in the human genome. Therefore, further investigation of this region is warranted to conclude whether and how genotypes in this region are associated with lung cancer risk.

As described above, the incidence of lung cancer in never-smokers is increasing, and lung cancers in never-smokers are now considered to be a different disease from lung cancers in smokers. Therefore, there have been a few GWA studies for the identification of loci associated with lung cancer risk specifically of never-smokers. One of the regions identified is chromosome 6q containing the RGS17 gene, which encodes a member of the regulator of G protein signaling (RGS) family (Amos *et al.*, 2010). RGS17 was identified as a major candidate for familial lung cancer susceptibility gene at chromosome 6q23-25 (Liu *et al.*, 2010a; You *et al.*, 2009), and proliferation and tumorigenesis of human lung tumor cells in nude mice were inhibited by knockdown of RGS17 expression levels. Never-smoking individuals with a risk haplotype of this locus were shown to have a 4.7-fold higher risk than those without risk haplotypes. Another region identified by a GWA study is chromosome 13q31.3 (Li *et al.*, 2010). This locus was identified by a four-stage association screening of lung cancers in never-smokers, and there was a strong correlation between genotypes of this locus and transcription levels of the GPC5 gene in normal lung tissues, with the high-risk allele linked with a lower level of transcription. Therefore, it was suggested that downregulation of GPC5 might contribute to the development of lung cancer in never-smokers. As described above, the TERT–CLPTM1L locus on chromosome 5p15.33 was also shown to be associated with lung cancer risk in never-smokers (Landi *et al.*, 2009; Wang *et al.*, 2010).

V. ASSESSMENT OF LUNG CANCER RISK IN EACH INDIVIDUAL BY COMBINED GENOTYPES (GENE–GENE INTERACTIONS)

For many years, gene–gene interaction has been investigated among candidate genes with functional polymorphisms. In particular, interactions among CYP-family genes and GST-family genes have been indicated by both molecular epidemiological studies and biological studies (Alexandrov *et al.*, 2002; Bartsch *et al.*, 2000; Schwartz *et al.*, 2007). Biologically, activities of CYP1A1 and GSTM1 are a critical determinant for the dose of carcinogenic BPDE and other DNA-reactive PAH; however, there has been no clear epidemiological evidence indicating the interaction between genotypes of CYP1A1 and GSTM1 on the risk for smoking-associated lung cancer risk, including SQC and SCLC. Since several lung cancer susceptibility genes have been identified by GWA studies, it is now very important to elucidate the interaction among their genotypes in the contribution to lung cancer risk. Then, we will be able to further develop a method to assess individual susceptibility to lung cancer based on the combined genotypes of several lung cancer susceptibility genes. However, up to the present, there have been only a few reports pursuing such an interaction. Here, we briefly summarize the results of three different studies investigating the effect of combined genotypes among genes identified by recent GWA studies on lung cancer risk.

In our recent study, we attempted to evaluate the combined effect among the HLA-DQA1, TERT, and CHRNA3 loci on lung ADC risk, because these three loci showed significant associations with lung ADC risk in the Japanese (Kohno *et al.*, 2010). However, the frequency of the susceptible haplotype in the CHRNA3 gene in the Japanese (0.02) was much lower than in European and American populations (0.4); therefore, interaction of CHRNA3 genotypes with HLA-DQA1 and TERT genotypes was unclear in this analysis. However, when ORs were calculated according to the number of risk alleles

for the HLA-DQA1 and TERT genes, there was an increasing trend with increasing number of risk alleles (per risk-allele OR = 1.43, $P = 7.8 \times 10^{-16}$), reaching up to OD = 4.76 for carriers of all four risk alleles. Namely, 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, $P = 4.2 \times 10^{-7}$). These two alleles independently (i.e., without a significant interaction) conferred the risk (P for interaction = 0.88). This result indicates that individuals highly susceptible to ADC can be defined by combined genotypes of HLA-DQA1 and TERT.

Recently, pooled analysis was performed for the replication of lung cancer susceptibility loci at chromosomes 15q24-25.1, 5p15.33, and 6p21 (Truong *et al.*, 2010). Associations between 15q24-25.1 variants and the risk for lung cancer were replicated in white ever-smokers; however, there were no such associations in never-smokers or in Asians. For the chromosome 5p15.33 region, statistically significant associations were confirmed in both whites and Asians. The 6p21 variants were not associated with the risk for lung cancer. Therefore, in this study, associations of the combined genotypes for the 15q24-25.1 locus (rs16969968) and the 5p15.33 locus (rs2736100 and rs402710) with the risk for lung cancer were further analyzed in whites. The OR of lung cancer risk for homozygotes of the three risk variants compared with individuals with no risk allele was 2.64 (95%CI = 1.86–3.74, $P=4 \times 10^{-8}$; per risk-allele OR = 1.15, $P = 1 \times 10^{-26}$). Liu *et al.* recently determined the cumulative association of four loci, 5p15.33, 6p21.33, 6q23-25, and 15q24-25.1, with familial lung cancer risk (Liu *et al.*, 2010a). The results indicate a stronger cumulative association of any combined genotype than any individual genotype with familial lung cancer. The risk for lung cancer was increased to 3- to 11-fold among those who had at least one copy of the risk allele at each locus in comparison with those who did not have any of the risk alleles.

The results of those three studies are consistent and indicate the cumulative effect of the SNPs in three chromosomal regions, 5p15.33, 6p21, and 15q24-25.1, on the genetic susceptibility to lung cancer, although interactions among those SNPs in lung cancer risk are unlikely.

VI. SMOKING-ASSOCIATED DIFFERENCES (GENE–ENVIRONMENT INTERACTIONS)

Cigarette smoking increases the risk for all three major histological types of lung cancers, although the risk is less for ADC than for SQC and SCLC (Govindan, 2010; Sobue *et al.*, 2002; Subramanian and Govindan, 2008; Sun *et al.*, 2007; Travis *et al.*, 2004). The smoking habit is largely attributed to nicotine dependence, because nicotine is addictive. Therefore, although nicotine itself is not carcinogenic, it has been assumed that nicotine dependence is indirectly associated with lung cancer risk by primarily causing the smoking habit and consequently resulting in the increase of tobacco carcinogen intake (Hecht, 2004). Recent GWA studies have identified an association of a common variant in the chromosome 15q24-25.1 region with lung cancer susceptibility (Amos *et al.*, 2008; Hung *et al.*, 2008; Thorgeirsson *et al.*, 2008). The region of ~200 kb in size with high LD contains six genes, and three of them encode nicotine acetylcholine receptor subunits, CHRNA5/A3/B4. This locus has been identified as being associated with nicotine dependence and smoking quantity by several studies (Lips *et al.*, 2010; Liu *et al.*, 2010b; Spitz *et al.*, 2008; Thorgeirsson *et al.*, 2008). Furthermore, this locus has been also identified as being associated with risk for several smoking-related diseases, such as chronic obstructive pulmonary disease (COPD) (Pillai *et al.*, 2009) and peripheral arterial disease (PAD) (Thorgeirsson *et al.*, 2008). Associations of this locus with smoking quantity and several smoking-related diseases, including lung cancer, support that the CHRNA genotypes are at least, in part, indirectly associated with lung cancer risk through smoking behavior. In contrast, associations with lung cancer risk in never-smokers as well as associations with an earlier age of lung cancer onset indicate the direct association of genotypes with lung cancer risk in a smoking behavior-independent manner. Associations with lung cancer risk after adjusting smoking habit also support the direct effect of genotypes on lung cancer risk. For this reason, associations of the CHRNA3/A5/B4 genotypes with lung cancer risk have been extensively and carefully investigated together with those with smoking behavior and nicotine dependence. However, several inconsistent results have been reported to date; thus, further studies are warranted.

Another example of smoking-associated differences is CYP family–GST family gene polymorphisms associated with smoking-related lung cancers, as described. However, to our knowledge, modifications by smoking behavior and/or smoking quantity of the associations between those genotypes and lung cancer risk have not yet been critically analyzed to date. Importantly, a recent study further indicated the association of other CHRNA genes, CHRNB3 and CHRNA6, on chromosome 8p11, with smoking and nicotine dependence as well as lung cancer risk (Thorgeirsson *et al.*, 2010). Interestingly, in their study, the association was also observed between the chromosome 19q13 region and smoking behavior as well as lung cancer risk. The 19q13 region contains the CYP2A6 gene, whose products have an enzyme activity to oxidize nicotine and to activate procarcinogenic nitrosamines.

VII. NECESSITY OF FURTHER ASSOCIATION STUDIES

To obtain more conclusive information on the genetic basis for susceptibility to lung cancer, we will have to analyze all the polymorphic sequences in the human genome for association with susceptibility. Various SNP array platforms have been developed to date, and the numbers of SNPs analyzable in one platform have been increasing year by year. In 2010, over a million SNPs can be analyzed by a single SNP array. However, it has been assumed that there are at least 10 million SNPs with a minor allele frequency (MAF) >1% and 5 million SNPs with a MAF >10% (Chung *et al.*, 2010; Frazer *et al.*, 2009). Therefore, although recent GWA studies have lead to the identification of several lung cancer susceptibility genes, it is still possible that there are several additional SNPs involved in the susceptibility in the human genome. In particular, several functional polymorphisms which have been identified by candidate gene association studies to date are not mounted on major SNP array platforms used in previous GWA studies, such as Affymetrix 500 K/1 M and Illumina HumanHap 300/550. Therefore, at present, it is not possible to obtain association data for those functional polymorphisms together with those for SNPs identified by GWA studies using SNP arrays. For this reason, we recently performed an association study of lung SQC for genes identified by GWA studies (CHRNA3, TERT, and HLA-DQA1) and genes identified by candidate gene association studies (TP53, MDM2, OGG1, CYP1A1, and GSTM1), because associations of these candidate gene polymorphisms were not investigated in recent GWA studies due to the lack of probes to discriminate these polymorphisms in the platforms used for GWA studies (Kohno *et al.*, in press). Genotypes for the TP53 and OGG1 genes showed significant associations with SQC risk in addition to those for the CHRNA3 and HLA-DQA1 genes to similar extents. Therefore, it will be necessary to reevaluate the significance of polymorphisms identified only by candidate gene association studies in several populations together with SNPs identified by GWA studies. In addition, it has been assumed that rare variants with frequencies less than 1% would play much more important role than common SNPs with a MAF >10% for the susceptibility to various diseases (Ioannidis *et al.*, 2010; Knerr *et al.*, 2010; McClellan and King, 2010). Therefore, a further technological advancement is absolutely required for the assessment of the role of rare variants in cancer susceptibility. Genetic polymorphisms include not only SNPs but also structural variations and copy number variations (CNVs) (Feero *et al.*, 2010; Frazer *et al.*, 2009). However, structural variations and CNVs are not yet easily analyzable at the genome-wide level. A CNV at 1q21.1 was recently shown to be associated with neuroblastoma susceptibility (Diskin *et al.*, 2009); therefore, development of a novel and easy

analytical method for CNVs throughout the human genome will be also necessary to finally identify all types of genetic polymorphisms associated with lung cancer risk.

Allele frequencies of several lung cancer susceptibility genes are different among different ethnic and geographic groups. Therefore, contribution of each susceptibility gene to lung cancer risk, represented by OR, is also considerably different among different ethnic/geographic groups. Figure 3 shows the differences in the frequencies of risk alleles for representative lung cancer susceptibility genes among Japanese, Chinese, Europeans, and Africans. Risk alleles for the CYP1A1 and OGG1 genes are more frequent in Asians than Europeans and Africans. In contrast, the risk allele for the CHRNA5 gene is more frequent in Europeans than Asians. Accordingly, comparative studies of polymorphisms with different allele frequencies and ORs among different ethnic/geographic groups will enable us to clarify the differences in lung cancer susceptibility

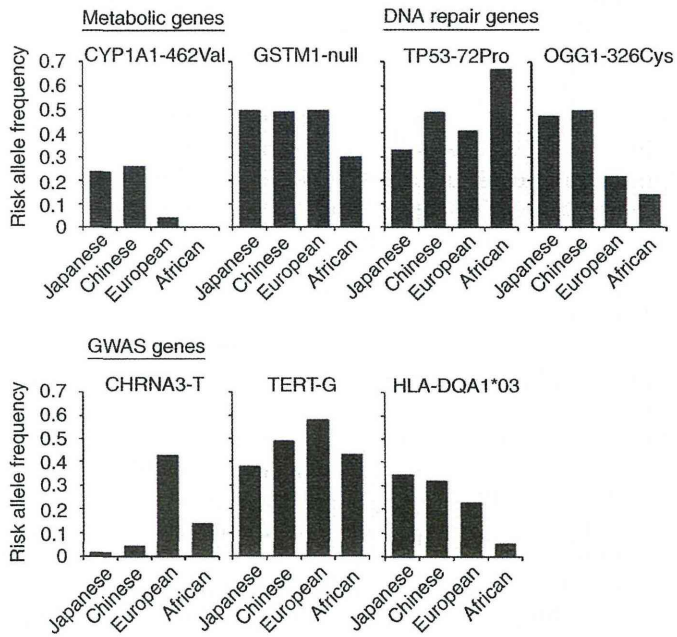

Fig. 3 Frequencies of risk alleles for lung cancer among different ethnic groups. As representatives, allele frequencies of seven genes in Japanese, Chinese, Europeans, and Africans are shown. Allele frequencies determined by the HapMap project, by the International Histocompatibility Working Group projects, by Ye *et al.* (2006), and by us (Kohno *et al.*, 2010; Shiraishi *et al.*, 2009) are combined in each column.

among different populations. For instance, lung ADCs with EGFR mutations in female nonsmokers are more common in Asian than in Americans and Europeans. However, it is still unknown whether or not such a difference is due to the difference in the distribution of risk alleles of some lung cancer susceptibility genes among these populations.

VIII. FUTURE DIRECTIONS

Recent GWA studies have identified three lung cancer susceptibility gene loci at chromosomes 15q24-25.1, 5q15.33, and 6p21. The 15q24-25.1 locus is associated not only with lung cancer but also with smoking behavior and other smoking-related diseases. Associations of the 15q24-25.1 genotypes with lung cancer risk in never-smokers and with lung ADC risk have been inconsistently observed among studies. In addition, the frequency of the risk allele is markedly different among ethnic groups. Therefore, further genetic studies as well as biological studies will be necessary to conclude whether the 15q24-25.1 genotypes play a direct or indirect role in the development of lung cancer, and how commonly/differentially the 15q24-25.1 genotypes contribute to lung cancer risk among different ethnic groups. The 5p15.33 locus is associated with risks not only for lung cancer but also for a variety of cancers, and the risk allele is prevalent among different ethnic groups. Therefore, this locus is likely to be associated with risks in general for a wide variety of cancers, irrespective of ethnic groups. Association of the 6p21 locus, containing the BAT3, MSH5, and HLA-DQA1 genes, with lung cancer has not yet been well reproduced by other genome-wide scale association studies. Therefore, further studies are necessary to obtain more convincing information for this locus in the association with lung cancer risk. Reevaluation of functional polymorphisms identified by candidate gene association studies will also be important for the assessment of individual risk for lung cancer (Wilkening *et al.*, 2009).

Lung cancers in never-smokers have been considered to be a different disease from those in ever-smokers. Associations have been observed between the 5p15.33, 6q23-25, and 13q31.3 genotypes and lung cancer risks in never-smokers. However, lung cancers in never-smokers are more common in women than in men, and also more common in Asian populations than in American and European populations (Govindan, 2010; Reid *et al.*, 2008; Subramanian and Govindan, 2008; Sun *et al.*, 2007). Therefore, it will be very important to elucidate interactions among genotypes, gender, and ethnicity/geography on lung cancer risk in never-smokers. The most frequent type of lung cancer in never-smokers is ADC; however, lung ADC is now considered to be a heterogeneous disease with respect to accumulated genetic alterations in cancer cells. EGFR-types are more frequent in Asian

populations, while KRAS-types are more frequent in American and European populations. However, it is still unknown whether such a difference is due to genetic differences or environmental differences. Identification of lung ADC susceptibility genes in never-smokers will facilitate the identification of environmental factors by subsequent functional analyses of identified genes. The elucidation of associations among such genetic factors, environmental factors other than smoking, and acquired genetic alterations in cancer cells will help us understand the molecular mechanisms underlying lung carcinogenesis in never-smokers and develop methods of its prevention.

Lastly, we point out here that none of polymorphisms have been identified yet to specifically define the risk for tobacco-induced lung cancer, such as SCC and SCLC. The 15q24-25.1 region containing three CHRNA genes is associated with smoking behavior as well as lung cancer susceptibility; however, it is still unclear whether this locus is associated with lung cancer risk among heavy smokers or not. Since only one in 10 smokers is estimated to develop lung cancer (Reid *et al.*, 2008), individual risks for lung cancer by smoking would be different due to genotype differences. Polymorphisms in genes for metabolism of tobacco smoke carcinogens and those for repair of carcinogen-induced genetic alterations have been considered as being candidates for many years. However, their significance is still unclear at present. Therefore, association studies of those genotypes with lung cancer risk by considering the smoking quantity, such as the number of cigarettes smoked per day (CPD), will be also important in assessing the individual risk for lung cancer in smokers. For this reason, the recently identified CYP2A6 gene locus will be another candidate to define lung cancer susceptibility in smokers (Thorgeirsson *et al.*, 2010).

ACKNOWLEDGMENTS

This study was supported in part by a Grant-in-Aid from the Ministry of Health, Labor, and Welfare for the third-term Comprehensive 10-year Strategy for Cancer Control and a Grant-in-Aid for the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio), Japan.

REFERENCES

Alexandrov, K., Cascorbi, I., Rojas, M., Bouvier, G., Kriek, E., and Bartsch, H. (2002). CYP1A1 and GSTM1 genotypes affect benzo[a]pyrene DNA adducts in smokers' lung: Comparison with aromatic/hydrophobic adduct formation. *Carcinogenesis* **23**, 1969–1977.

Amos, C. I., Wu, X., Broderick, P., Gorlov, I. P., Gu, J., Eisen, T., Dong, Q., Zhang, Q., Gu, X., Vijayakrishnan, J., Sullivan, K., Matakidou, A., *et al.* (2008). Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat. Genet.* **40**, 616–622.

Amos, C. I., Pinney, S. M., Li, Y., Kupert, E., Lee, J., de Andrade, M. A., Yang, P., Schwartz, A. G., Fain, P. R., Gazdar, A., Minna, J., Wiest, J. S., *et al.* (2010). A susceptibility locus on chromosome 6q greatly increases lung cancer risk among light and never smokers. *Cancer Res.* **70,** 2359–2367.

Bartsch, H., Nair, U., Risch, A., Rojas, M., Wikman, H., and Alexandrov, K. (2000). Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. *Cancer Epidemiol. Biomarkers Prev.* **9,** 3–28.

Bond, G. L., and Levine, A. J. (2007). A single nucleotide polymorphism in the p53 pathway interacts with gender, environmental stresses and tumor genetics to influence cancer in humans. *Oncogene* **26,** 1317–1323.

Broderick, P., Wang, Y., Vijayakrishnan, J., Matakidou, A., Spitz, M. R., Eisen, T., Amos, C. I., and Houlston, R. S. (2009). Deciphering the impact of common genetic variation on lung cancer risk: A genome-wide association study. *Cancer Res.* **69,** 6633–6641.

Chung, C. C., Magalhaes, W. C., Gonzalez-Bosquet, J., and Chanock, S. J. (2010). Genome-wide association studies in cancer—Current and future directions. *Carcinogenesis* **31,** 111–120.

Dai, S., Mao, C., Jiang, L., Wang, G., and Cheng, H. (2009). P53 polymorphism and lung cancer susceptibility: A pooled analysis of 32 case-control studies. *Hum. Genet.* **125,** 633–638.

Diskin, S. J., Hou, C., Glessner, J. T., Attiyeh, E. F., Laudenslager, M., Bosse, K., Cole, K., Mossé, Y. P., Wood, A., Lynch, J. E., Pecor, K., Diamond, M., *et al.* (2009). Copy number variation at 1q21.1 associated with neuroblastoma. *Nature* **459,** 987–991.

Dong, L. M., Potter, J. D., White, E., Ulrich, C. M., Cardon, L. R., and Peters, U. (2008). Genetic susceptibility to cancer: The role of polymorphisms in candidate genes. *JAMA* **299,** 2423–2436.

Feero, W. G., Guttmacher, A. E., and Collins, F. S. (2010). Genomic medicine—An updated primer. *N. Engl. J. Med.* **362,** 2001–2011.

Frazer, K. A., Murray, S. S., Schork, N. J., and Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. *Nat. Rev. Genet.* **10,** 241–251.

Govindan, R. (2010). Lung cancer in never smokers: A new hot area of research. *Lancet Oncol.* **11,** 304–305.

Hecht, S. S. (2004). Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat. Rev. Cancer* **3,** 733–744.

Hollstein, M., Sidransky, D., Vogelstein, B., and Harris, C. C. (1991). p53 mutations in human cancers. *Science* **253,** 49–53.

Hung, R. J., McKay, J. D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., *et al.* (2008). A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* **452,** 633–637.

Imyanitov, E. N. (2009). Gene polymorphisms, apoptotic capacity and cancer risk. *Hum. Genet.* **125,** 239–246.

Ioannidis, J. P., Castaldi, P., and Evangelou, E. (2010). A compendium of genome-wide associations for cancer: Critical synopsis and reappraisal. *J. Natl. Cancer Inst.* **102,** 846–858.

Jin, G., Xu, L., Shu, Y., Tian, T., Liang, J., Xu, Y., Wang, F., Chen, J., Dai, J., Hu, Z., and Shen, H. (2009). Common genetic variants on 5p15.33 contribute to risk of lung adenocarcinoma in a Chinese population. *Carcinogenesis* **30,** 987–990.

Kim, C. F., Jackson, E. L., Woolfenden, A. E., Lawrence, S., Babar, I., Vogel, S., Crowley, D., Bronson, R. T., and Jacks, T. (2005). Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* **121,** 823–835.

Knerr, S., Ramos, E., Nowinski, J., Dixon, K., and Bonham, V. L. (2010). Human difference in the genomic era: Facilitating a socially responsible dialogue. *BMC Med. Genomics* **3,** 20.

Kohno, T., Shinmura, K., Tosaka, M., Tani, M., Kim, S. R., Sugimura, H., Nohmi, T., Kasai, H., and Yokota, J. (1998). Genetic polymorphisms and alternative splicing of the hOGG1 gene, that is involved in the repair of 8-hydroxyguanine in damaged DNA. *Oncogene* **16,** 3219–3225.

Kohno, T., Kunitoh, H., Toyama, K., Yamamoto, S., Kuchiba, A., Saito, D., Yanagitani, N., Ishihara, S., Saito, R., and Yokota, J. (2006). Association of the OGG1-Ser326Cys polymorphism with lung adenocarcinoma risk. *Cancer Sci.* **97,** 724–728.

Kohno, T., Kunitoh, H., Shimada, Y., Shiraishi, K., Ishii, Y., Goto, K., Ohe, Y., Nishiwaki, Y., Kuchiba, A., Yamamoto, S., Hirose, H., Oka, A., *et al.* (2010). Individuals susceptible to lung adenocarcinoma defined by combined HLA-DQA1 and TERT genotypes. *Carcinogenesis* **31,** 834–841.

Kohno, T., Kunitoh, H., Mimaki, S., Shiraishi, K., Kuchiba, A., Yamamoto, S., and Yokota, J. Contribution of the TP53, OGG1, CHRNA3 and HLA-DQA1 genes to the risk for lung squamous cell carcinoma. *J. Thorac. Oncol.* (In Press).

Landi, M. T., Chatterjee, N., Yu, K., Goldin, L. R., Goldstein, A. M., Rotunno, M., Mirabello, L., Jacobs, K., Wheeler, W., Yeager, M., Bergen, A. W., Li, Q., *et al.* (2009). A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Hum. Genet.* **85,** 679–691.

Le Calvez, F., Mukeria, A., Hunt, J. D., Kelm, O., Hung, R. J., Tanière, P., Brennan, P., Boffetta, P., Zaridze, D. G., and Hainaut, P. (2005). TP53 and KRAS mutation load and types in lung cancers in relation to tobacco smoke: Distinct patterns in never, former, and current smokers. *Cancer Res.* **65,** 5076–5083.

Li, H., Hao, X., Zhang, W., Wei, Q., and Chen, K. (2008). The hOGG1 Ser326Cys polymorphism and lung cancer risk: A meta-analysis. *Cancer Epidemiol. Biomarkers Prev.* **17,** 1739–1745.

Li, Y., Sheu, C. C., Ye, Y., de Andrade, M., Wang, L., Chang, S. C., Aubry, M. C., Aakre, J. A., Allen, M. S., Chen, F., Cunningham, J. M., Deschamps, C., *et al.* (2010). Genetic variants and risk of lung cancer in never smokers: A genome-wide association study. *Lancet Oncol.* **11,** 321–330.

Lips, E. H., Gaborieau, V., McKay, J. D., Chabrier, A., Hung, R. J., Boffetta, P., Hashibe, M., Zaridze, D., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., *et al.* (2010). Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17000 individuals. *Int. J. Epidemiol.* **39,** 563–577.

Liu, P., Vikis, H. G., Lu, Y., Wang, Y., Schwartz, A. G., Pinney, S. M., Yang, P., de Andrade, M., Gazdar, A., Gaba, C., Mandal, D., Lee, J., *et al.* (2010a). Cumulative effect of multiple loci on genetic susceptibility to familial lung cancer. *Cancer Epidemiol. Biomarkers Prev.* **19,** 517–524.

Liu, J. Z., Tozzi, F., Waterworth, D. M., Pillai, S. G., Muglia, P., Middleton, L., Berrettini, W., Knouff, C. W., Yuan, X., Waeber, G., Vollenweider, P., Preisig, M., *et al.* (2010b). Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat. Genet.* **42,** 436–440.

McClellan, J., and King, M. C. (2010). Genetic heterogeneity in human disease. *Cell* **141,** 210–217.

McKay, J. D., Hung, R. J., Gaborieau, V., Boffetta, P., Chabrier, A., Byrnes, G., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., *et al.* (2008). Lung cancer susceptibility locus at 5p15.33. *Nat. Genet.* **40,** 1404–1406.

Pillai, S. G., Ge, D., Zhu, G., Kong, X., Shianna, K. V., Need, A. C., Feng, S., Hersh, C. P., Bakke, P., Gulsvik, A., Ruppert, A., Lødrup Carlsen, K. C., *et al.* (2009). A genome-wide association study in chronic obstructive pulmonary disease (COPD): Identification of two major susceptibility loci. *PLoS Genet.* **5,** e1000421.

Rafnar, T., Sulem, P., Stacey, S. N., Geller, F., Gudmundsson, J., Sigurdsson, A., Jakobsdottir, M., Helgadottir, H., Thorlacius, S., Aben, K. K., Blöndal, T., Thorgeirsson, T. E., *et al.* (2009). Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat. Genet.* **41,** 221–227.

Reid, M. E., Santella, R., and Ambrosone, C. B. (2008). Molecular epidemiology to better predict lung cancer risk. *Clin. Lung Cancer* **9,** 149–153.

Rubin, H. (2001). Synergistic mechanisms in carcinogenesis by polycyclic aromatic hydrocarbons and by tobacco smoke: A bio-historical perspective with updates. *Carcinogenesis* **22,** 1903–1930.

Schwartz, A. G., Prysak, G. M., Bock, C. H., and Cote, M. L. (2007). The molecular epidemiology of lung cancer. *Carcinogenesis* **28,** 507–518.

Shiraishi, K., Kohno, T., Kunitoh, H., Watanabe, S., Goto, K., Nishiwaki, Y., Shimada, Y., Hirose, H., Saito, I., Kuchiba, A., Yamamoto, S., and Yokota, J. (2009). Contribution of nicotine acethylchorine receptor polymorphisms to lung cancer risk in a smoking-independent manner in the Japanese. *Carcinogenesis* **30,** 65–70.

Sobue, T., Yamamoto, S., Hara, M., Sasazuki, S., Sasaki, S., and Tsugane, S. (2002). Cigarette smoking and subsequent risk of lung cancer by histologic type in middle-aged Japanese men and women: The JPHC study. *Int. J. Cancer* **99,** 245–251.

Spitz, M. R., Amos, C. I., Dong, Q., Lin, J., and Wu, X. (2008). The CHRNA5-A3 region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer. *J. Natl. Cancer Inst.* **21,** 1552–1556.

Stacey, S. N., Sulem, P., Masson, G., Gudjonsson, S. A., Thorleifsson, G., Jakobsdottir, M., Sigurdsson, A., Gudbjartsson, D. F., Sigurgeirsson, B., Benediktsdottir, K. R., Thorisdottir, K., Ragnarsson, R., *et al.* (2009). New common variants affecting susceptibility to basal cell carcinoma. *Nat. Genet.* **41,** 909–914.

Subramanian, J., and Govindan, R. (2008). Molecular genetics of lung cancer in people who have never smoked. *Lancet Oncol.* **9,** 676–682.

Sugimura, H., Kohno, T., Wakai, K., Nagura, K., Genka, K., Igarashi, H., Morris, B. J., Baba, S., Ohno, Y., Gao, C., Li, Z., Wang, J., *et al.* (1999). hOGG1 Ser326Cys polymorphism and lung cancer susceptibility. *Cancer Epidemiol. Biomarkers Prev.* **8,** 669–674.

Sun, S., Schiller, J. H., and Gazdar, A. F. (2007). Lung cancer in never smokers—A different disease. *Nat. Rev. Cancer* **7,** 778–790.

Thorgeirsson, T. E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K. P., Manolescu, A., Thorleifsson, G., Stefansson, H., Ingason, A., Stacey, S. N., Bergthorsson, J. T., *et al.* (2008). A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452,** 638–642.

Thorgeirsson, T. E., Gudbjartsson, D. F., Surakka, I., Vink, J. M., Amin, N., Geller, F., Sulem, P., Rafnar, T., Esko, T., Walter, S., Gieger, C., Rawal, R., *et al.* (2010). Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat. Genet.* **42,** 448–453.

Travis, W. D., Brambilla, E., Muller-Hermelink, H. K., and Harris, C. C. (2004). Pathology and genetics, tumours of lung, pleura, thymus and heart. *In* World Health Organization Classification of Tumors, (W. D. Travis, E. Brambilla, H. K. Muller-Hermelink, and C. C. Harris, Eds.), pp. 1–344. IARC Press, Lyon.

Truong, T., Hung, R. J., Amos, C. I., Wu, X., Bickeböller, H., Rosenberger, A., Sauter, W., Illig, T., Wichmann, H. E., Risch, A., Dienemann, H., Kaaks, R., *et al.* (2010). Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: A pooled analysis from the International Lung Cancer Consortium. *J. Natl. Cancer Inst.* **102,** 959–971.

Wang, Y., Broderick, P., Webb, E., Wu, X., Vijayakrishnan, J., Matakidou, A., Qureshi, M., Dong, Q., Gu, X., Chen, W. V., Spitz, M. R., Eisen, T., *et al.* (2008). Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat. Genet.* **40,** 1407–1409.

Wang, Y., Broderick, P., Matakidou, A., Eisen, T., and Houlston, R. S. (2010). Role of 5p15.33 (TERT-CLPTM1L), 6p21.33 and 15q25.1 (CHRNA5-CHRNA3) variation and lung cancer risk in never-smokers. *Carcinogenesis* **31,** 234–238.

Whibley, C., Pharoah, P. D., and Hollstein, M. (2009). p53 polymorphisms: Cancer implications. *Nat. Rev. Cancer* **9,** 95–107.

Wilkening, S., Bermejo, J. L., and Hemminki, K. (2007). MDM2 SNP309 and cancer risk: A combined analysis. *Carcinogenesis* **28,** 2262–2267.

Wilkening, S., Chen, B., Bermejo, J. L., and Canzian, F. (2009). Is there still a need for candidate gene approaches in the era of genome-wide association studies? *Genomics* **93,** 415–419.

Wu, C., Hu, Z., Yu, D., Huang, L., Jin, G., Liang, J., Guo, H., Tan, W., Zhang, M., Qian, J., Lu, D., Wu, T., *et al.* (2009). Genetic variants on chromosome 15q25 associated with lung cancer risk in Chinese populations. *Cancer Res.* **69,** 5065–5072.

Yamane, A., Kohno, T., Ito, K., Sunaga, N., Aoki, K., Yoshimura, K., Murakami, H., Nojima, Y., and Yokota, J. (2004). Differential ability of polymorphic OGG1 proteins to suppress mutagenesis induced by 8-hydroxyguanine in human cell in vivo. *Carcinogenesis* **25,** 1689–1694.

Ye, Z., Song, H., Higgins, J. P., Pharoah, P., and Danesh, J. (2006). Five glutathione s-transferase gene variants in 23, 452 cases of lung cancer and 30, 397 controls: Meta-analysis of 130 studies. *PLoS Med.* **3,** e91.

You, M., Wang, D., Liu, P., Vikis, H., James, M., Lu, Y., Wang, Y., Wang, M., Chen, Q., Jia, D., Liu, Y., Wen, W., *et al.* (2009). Fine mapping of chromosome 6q23-25 region in familial lung cancer families reveals RGS17 as a likely candidate gene. *Clin. Cancer Res.* **15,** 2666–2674.

A Catalog of Genes Homozygously Deleted in Human Lung Cancer and the Candidacy of *PTPRD* as a Tumor Suppressor Gene

Takashi Kohno,[1] Ayaka Otsuka,[1] Luc Girard,[2] Masanori Sato,[1] Reika Iwakawa,[1] Hideaki Ogiwara,[1] Montse Sanchez-Cespedes,[3] John D. Minna,[2] and Jun Yokota[1]*

[1]Biology Division, National Cancer Center Research Institute, Tokyo, Japan
[2]Hamon Center for Therapeutic Oncology Research, University of Texas Southwestern Medical Center, Dallas, USA
[3]Genes and Cancer Group, Programa Epigenetica y Biologia del Cancer (PEBC), IDIBELL Hospital Duran i Reynals Av. Gran Via del'Hospitalet, Barcelona, Spain

A total of 176 genes homozygously deleted in human lung cancer were identified by DNA array-based whole genome scanning of 52 lung cancer cell lines and subsequent genomic PCR in 74 cell lines, including the 52 cell lines scanned. One or more exons of these genes were homozygously deleted in one (1%) to 20 (27%) cell lines. These genes included known tumor suppressor genes, e.g., *CDKN2A/p16*, *RB1*, and *SMAD4*, and candidate tumor suppressor genes whose hemizygous or homozygous deletions were reported in several types of human cancers, such as *FHIT*, *KEAP1*, and *LRP1B/LRP-DIP*. *CDKN2A/p16* and *p14ARF* located in 9p21 were most frequently deleted (20/74, 27%). The *PTPRD* gene was most frequently deleted (8/74, 11%) among genes mapping to regions other than 9p21. Somatic mutations, including a nonsense mutation, of the *PTPRD* gene were detected in 8/74 (11%) of cell lines and 4/95 (4%) of surgical specimens of lung cancer. Reduced *PTPRD* expression was observed in the majority (>80%) of cell lines and surgical specimens of lung cancer. Therefore, *PTPRD* is a candidate tumor suppressor gene in lung cancer. Microarray-based expression profiling of 19 lung cancer cell lines also indicated that some of the 176 genes, such as *KANK* and *ADAMTS1*, are preferentially inactivated by epigenetic alterations. Genetic/epigenetic as well as functional studies of these 176 genes will increase our understanding of molecular mechanisms behind lung carcinogenesis. © 2010 Wiley-Liss, Inc.

INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths in the world (Herbst et al., 2008). The majority of lung cancers are comprised of four major histological types, which are small cell lung carcinoma (SCLC) and three nonsmall cell lung carcinoma (NSCLC) types; adenocarcinoma (ADC), squamous cell carcinoma (SQC), and large cell carcinoma (LCC). Lung cancer develops through the acquisition of alterations in oncogenes, such as *EGFR* (10–40% of ADC) and *KRAS* (10–30% of ADC), and tumor suppressor genes, such as *TP53* (~90% of SCLC; 50% of NSCLC), *RB1* (~90% of SCLC; ~20% of NSCLC), *CDKN2A/p16* (~50% of NSCLC), and *LKB1/STK11* (20–30% of NSCLC) (Minna et al., 2002; Herbst et al., 2008). The *EGFR*, *KRAS*, and *TP53* genes have been subjected to diagnostic and therapeutic applications (Toloza et al., 2006; Herbst et al., 2008); therefore, identification of more genes involved in lung carcinogenesis will be highly applicable to further improve the diagnosis and therapy of lung cancer. Allelic imbalance (AI) studies on lung cancer have identified several chromosome arms frequently hemizygously deleted, such as 1p, 4q, 5q, 6q, 8p, 11q, 12q, 13q, 17q, and 21q (Shiseki et al., 1996; Kawanishi et al., 1997; Virmani et al., 1998; Girard et al., 2000). Our recent comparative genome-wide AI study of noninvasive and invasive lung adenocarcinomas (ADCs) further suggested that AI on each chromosome arm has different roles in the development and progression of lung cancer (Nakanishi et al., 2009). Therefore, chromosomal deletions and inactivation of corresponding tumor suppressor genes are thought to play multiple roles in the development and/or progression of lung cancer. However, responsible tumor suppressor genes for most of these chromosomal deletions have not yet been identified.

Homozygous deletion (i.e., deletion of both alleles) is a genetic event causing inactivation of tumor suppressor genes (Minna et al., 2002; Yokota and Kohno, 2004), and has played an important role as a tool in identifying several tumor suppressor genes, such as *CDKN2A/p16*, *PTEN*, and *SMAD4* (Kamb et al., 1994; Nobori et al., 1994; Hahn et al., 1996; Li et al., 1997; Steck et al., 1997). Up to the present, DNA array analyses have been performed by several groups, including ours, to find homozygously deleted regions in lung cancer genomes (Sato et al., 2005; Tonon et al., 2005; Zhao et al., 2005; Garnis et al., 2006; Imoto et al., 2006; Nagayama et al., 2007; Weir et al., 2007), and tens of genomic regions with homozygous deletions have been identified. However, only a few genes located in some of the homozygously deleted regions were focused on and investigated.

In this study, genes whose exons were removed by homozygous deletions were comprehensively searched for by a DNA array-based whole genome scanning of 52 human lung cancer cell lines followed by genomic PCR analyses. Notably, several well-known tumor suppressor genes, such as *RB1*, *DCC*, and *BRCA2*, have been identified from a single or a few cases of homozygous deletions detected in a large number of cancer cases analyzed (Dryja et al., 1986; Fearon et al., 1990; Wooster et al., 1995). The results indicated the significance of homozygous deletions irrespective of their frequencies for the identification of novel tumor suppressor genes. Therefore, in the present study, all genes deduced to be mapped in homozygously deleted regions were examined, even if the deletions were detected only in a single lung cancer case. Lung cancer cell lines were used for two reasons: First, the presence of a homozygous deletion can be easily validated by genomic PCR due to the lack of noncancerous cell contamination that hampers detection of homozygous deletions; second, frequencies of copy number changes in the genome were shown to be similar in cell lines and surgical specimens in our previous study (Ogiwara et al., 2008). Hence homozygous deletions detected in cell lines can be considered to have occurred mostly in vivo, and not during their establishment and cultivation in vitro. In total, 176 genes located in 45 genomic loci on 17 chromosomes were identified as genes whose exons were homozygously deleted (Supporting Information Table 1). One of the 176 genes, *PTPRD*, was subjected to mutation and expression analyses in surgical specimens of lung cancer as well as lung cancer cell lines to address the authenticity of this gene as a lung tumor suppressor gene.

MATERIALS AND METHODS

Human Lung Cancer Cell Lines and Surgical Specimens for Lung Cancer

Forty-three lung cancer cell lines were previously subjected to a SNP array analysis at a 100-kb resolution using an Affymetrics Mapping 100-k array (Affymetrix, Inc., Santa Clara, California), and they were 11 SCLCs, 21 ADCs, 7 SQCs, and 4 LCCs (Nagayama et al., 2007). In the present study, 27 ADC cell lines consisting of 18 lines (II-18, A549, Ma17, Ma24, H23, H322, H1395, H1437, H2009, H2087, H2122, H2347, PC3, PC7, PC9, PC14, RERF-LCMS, and VMRC-LCD) analyzed in the previous analysis (Nagayama et al., 2007) and 9 cell lines (ABC1, Ma10, Ma12, Ma26, Ma29, HCC44, HCC78, HCC193, and HCC515) prepared for this study were subjected to an array-CGH analysis at a 30-kb resolution using a Human CGH 185-k array (Agilent Technologies, Santa Clara, California). Therefore, 52 cell lines in total were scanned for homozygous deletions by using one or two DNA-array methods at 30 and 100-kb resolutions. To validate homozygous deletion, 74 cell lines consisting of 52 cell lines subjected to these array analyses and an additional 22 lung cancer cell lines consisting of 11 SCLCs (H526, H774, H1339, H1450, H1607, H1819, NCI-H1963, H2195, HCC33, Lu24, and Ms18), 3 ADCs (H2126, H1703, and RERF-LCOK), 3 SQCs (HCC95, Sq-5, and PC10), 3 LCCs (Lu99, Ma2, and Ma25), and 2 adenosquamous carcinomas (ASCs) (H596 and HCC366) were analyzed. Details of H- and HCC-series cell lines have been described elsewhere (Burbee et al., 2001). PC-, Lu-, Ma-series, and II-18 cell lines were provided by Drs. Y. Hayata (Tokyo Medical University, Tokyo, Japan), T. Terasaki (Kanagawa Institute of Technology, Kanagawa, Japan) and S. Hirohashi (National Cancer Center Research Institute, Tokyo, Japan), M. Takada (National Hospital Organization Kinki-chuo Chest Medical Center, Osaka, Japan), and K. Hagiwara (Saitama Medical University, Saitama, Japan), respectively. Cell

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

Supported by: Grants-in-Aid from the Ministry of Health, Labor and Welfare of Japan for the 3rd-term Comprehensive 10-year Strategy for Cancer Control and for Cancer Research (16-1); from the program for promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio); National Cancer Institute Lung Cancer SPORE Grant P50CA70907 and the DOD VITAL grant.

*Correspondence to: Jun Yokota, Biology Division, National Cancer Center Research Institute, Tokyo, Japan.
E-mail: jyokota@ncc.go.jp

Received 16 October 2009; Accepted 2 December 2009
DOI 10.1002/gcc.20746
Published online 13 January 2010 in Wiley InterScience (www.interscience.wiley.com).

© 2010 Wiley-Liss, Inc.