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 (JMLA) on September 29, 2010

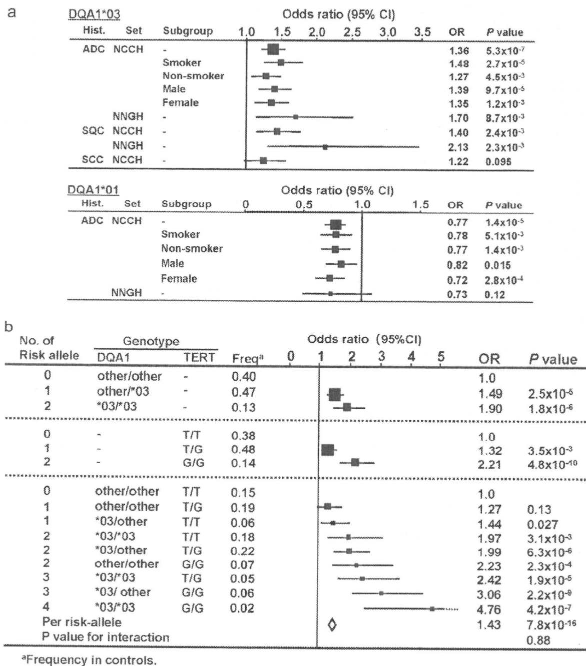

Fig. 2. Forest plot representing risk for lung cancer. (**a**) Risk of the DQA1*03 and DQA1*01 alleles for lung cancer. ORs of the alleles adjusted for age, sex, smoking habit and/or hospital and 95% CIs are shown. Detailed data, including the numbers of case and control subjects and variables for adjustments for each test, are summarized in supplementary Table V is available at *Carcinogenesis* Online. (**b**) Risk of combined *HLA-DQA1* and *TERT* genotypes for lung ADC. ORs of the alleles adjusted for age, sex and smoking habit and 95% CIs are shown. Detailed data, including the numbers of case and control subjects, are summarized in supplementary Tables VII and VIII are available at *Carcinogenesis* Online.

compared for all three genes, there was also an increasing trend with a per-risk allele OR of 1.45; however, only ORs for carriers of up to four risk alleles could be calculated.

The present study indicated *HLA-DQA1* at 6p21.31 as a novel locus associated with lung cancer risk and genotypes for this locus are useful for identification of individuals susceptible to lung ADC. It has been considered that immune surveillance systems conferred by HLA class I and II proteins are involved in the elimination of tumor cells *in vivo* (25). HLA class I proteins are expressed in most nucleated cells and present tumor-specific antigens for cytotoxic CD8+ T cells to recognize and lyse tumor cells. In addition, the immune response requires the presentation of antigenic peptides to T cells by class II molecules expressed on antigen-presenting cells, i.e. the heterodimer of HLA-DQA1 and -DQB1 proteins and of HLA-DRA and -DRB1 proteins. Therefore, it might be that polymorphisms of *HLA-DQA1* (and also those of *HLA-DQB1* and *-DRB1* that are in LD with those of *HLA-DQA1*) gene confers lung cancer susceptibility by causing inter-individual differences in the ability of HLA class II molecule to bind peptides produced in lung cancer cells and to cause immune response. However, we should consider that the present results were obtained by performing a number of association tests against smaller numbers of subjects than those of recent GWASs (4–11). In addition, control subjects from NNGH used for validation of association had lung diseases, including chronic obstructive pulmonary disease. A recent GWAS on chronic obstructive pulmonary disease has shown the same susceptibility loci as lung cancer, such as 15q25.1, suggesting that lung cancer and other lung diseases share the same genetic etiology (26). Therefore, it remains possible that associations observed in the present study were under- or over-represented. The number of control subjects in the present study was 30% less than that of ADC cases (combined analysis in Table II), although optimal ratios of control subjects to case subjects have been considered as being 1:1 to 4:1 (27), and this fact resulted in larger 95% CIs of OR than analyzing optimal number of control subjects. Thus, further case–control studies will be needed to validate the association of the 6p21.31 locus with lung ADC risk. Notably, synonymous SNPs in the 6p21.31 locus, such as rs2187668 and rs1794282, also showed significant differences in allelic distributions in Europeans and Americans (Figure 1, supplementary Table IX is available at *Carcinogenesis* Online). The strength of this association was similar to those for SNPs in the *BAT3-MSH5* locus in those populations (supplementary Figure 3 is available at *Carcinogenesis* Online). Therefore, it was strongly indicated that 6p21.31 is a lung ADC susceptibility locus

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

not only in Japanese but also in Europeans and Americans. However, at present, it remains unknown whether SNPs/alleles associated with risk are different among populations, since only a few SNPs have been examined for associations in Europeans and Americans (Figure 1, supplementary Figure 3 and supplementary Table IX are available at *Carcinogenesis* Online). In addition, LD among SNPs in the HLA class II locus is known to be different among different ethnic populations (23). Thus, studies on multiple populations will give us more critical information on the roles of polymorphisms in the 6p21.31 locus and their interaction with other lung cancer susceptibility loci in lung ADC susceptibility.

The present GWAS on ADC risk was performed against 23 010 microsatellite loci spaced at ~130 kb intervals in the human genome. However, two other lung cancer susceptibility loci, 15q25.1 and 5p15.33, whose SNPs showed associations with risk in the population analyzed in the present study (supplementary Table VI is available at *Carcinogenesis* Online), were not detected in the present GWAS using microsatellites. Therefore, several lung ADC susceptibility loci were likely to be overlooked in the present GWAS probably due to insufficient statistical power and a sparse marker density. Thus, a GWAS on lung ADC risk, in which hundreds of thousands SNPs are analyzed, is underway in our laboratory to comprehensively identify lung ADC susceptibility loci. Finally, in spite of facts that ADC is the commonest histological type of lung cancer in non-smokers and that ADC of non-smokers is showing an increasing trend (2,28), loci specifically associated with ADC risk of non-smokers have not been identified. Therefore, GWASs focusing on lung ADC risk of non-smokers would be also worth investigating to identify additional lung ADC susceptibility loci. In addition, case–control studies on subjects that were carefully chosen to represent cases and controls in the same population, such as a nested case–control study designated in a large-scale cohort study, will be critical to validate the significance of susceptibility loci on lung carcinogenesis for the application to targeted screening and/or prevention of lung ADC in future.

Funding

Grants-in-Aid from the Ministry of Health, Labor and Welfare for the third-term Comprehensive 10 year Strategy for Cancer Control (J.Y.) and for Cancer Research, 19S-1 (T.K.); Ministry of Education, Culture, Sports, Science and Technology for Scientific Research on Priority Areas, KAKENHI 20014031 (T.K.).

Supplementary material

Supplementary material can be found at http://carcin.oxfordjournals.org/

Acknowledgements

We thank Drs Takashi Eguchi, Ikuo Saito, Matsuhiko Hayashi, Keiichi Hirao, Teruhiko Yoshida, Hiromi Sakamoto, Kimio Yoshimura and Shunpei Ohnami for their help in collecting blood samples of the control subjects. We thank Drs Gen Tamiya, Tokuki Sakiyama and Tomohisa Chiku for their help in GWAS. We also thank Ms Satoko Matsuyama, Sachiyo Mimaki, Rumie Sasaki and Yuki Takei for technical assistance. K.S. is an awardee of a Research Resident Fellowship from the Foundation for Promotion of Cancer Research in Japan.

Conflict of Interest Statement: The authors disclosed no potential conflicts of interest.

References

1. Travis,W.D. *et al.* (2004) Pathology and genetics, tumours of lung, pleura, thymus and heart. In Travis,W.D., Brambilla,E., Muller-Hermelink,H.K. and Harris,C.C. (eds.) *World Health Organization Classification of Tumors.* IARC Press, pp. 1–344.
2. Sun,S. *et al.* (2007) Lung cancer in never smokers—a different disease. *Nat. Rev. Cancer*, **7**, 778–790.
3. Sobue,T. *et al.* (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.
4. Wang,Y. *et al.* (2008) Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat. Genet.*, **40**, 1407–1409.
5. McKay,J.D. *et al.* (2008) Lung cancer susceptibility locus at 5p15.33. *Nat. Genet.*, **40**, 1404–1406.
6. Liu,P. *et al.* (2008) Familial aggregation of common sequence variants on 15q24-25.1 in lung cancer. *J. Natl Cancer Inst.*, **100**, 1326–1330.
7. Kohno,T. *et al.* (2008) Association of KRAS polymorphisms with risk for lung adenocarcinoma accompanied by atypical adenomatous hyperplasias. *Carcinogenesis*, **29**, 957–963.
8. Hung,R.J. *et al.* (2008) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*, **452**, 633–637.
9. Amos,C.I. *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.
10. Broderick,P. *et al.* (2009) Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. *Cancer Res.*, **69**, 6633–6641.
11. Landi,M. *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**, 1–13.
12. Yamaguchi-Kabata,Y. *et al.* (2008) Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am. J. Hum. Genet.*, **83**, 445–456.
13. Sakamoto,H. *et al.* (2008) Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat. Genet.*, **40**, 730–740.
14. Tamiya,G. *et al.* (2005) Whole genome association study of rheumatoid arthritis using 27 039 microsatellites. *Hum. Mol. Genet.*, **14**, 2305–2321.
15. Sunaga,N. *et al.* (2002) Contribution of the NQO1 and GSTT1 polymorphisms to lung adenocarcinoma susceptibility. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 730–738.
16. Sakiyama,T. *et al.* (2005) Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. *Int. J. Cancer*, **114**, 730–737.
17. Kohno,T. *et al.* (2006) Association of the OGG1-Ser326Cys polymorphism with lung adenocarcinoma risk. *Cancer Sci.*, **97**, 724–728.
18. Kohno,T. *et al.* (2006) Association of polymorphisms in the MTH1 gene with small cell lung carcinoma risk. *Carcinogenesis*, **27**, 2448–2454.
19. Tomlinson,I.P. *et al.* (2008) A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.*, **40**, 623–630.
20. Hauge,X.Y. *et al.* (1993) A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. *Hum. Mol. Genet.*, **2**, 411–415.
21. Jin,G. *et al.* (2009) Common genetic variants on 5p15.33 contribute to risk of lung adenocarcinoma in a Chinese population. *Carcinogenesis*, **30**, 987–990.
22. Shiraishi,K. *et al.* (2009) Contribution of nicotine acetylcholine receptor polymorphisms to lung cancer risk in a smoking-independent manner in the Japanese. *Carcinogenesis*, **30**, 65–70.
23. de Bakker,P.I. *et al.* (2006) A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.*, **38**, 1166–1172.
24. Miretti,M.M. *et al.* (2005) A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am. J. Hum. Genet.*, **76**, 634–646.
25. Knutson,K.L. *et al.* (2005) Tumor antigen-specific T helper cells in cancer immunity and immunotherapy. *Cancer Immunol. Immunother.*, **54**, 721–728.
26. Pillai,S.G. *et al.* (2009) A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet.*, **5**, e1000421.
27. Breslow,N.E. *et al.* (1980) 1.5 planning. In Davis,W. (ed.) Statistical Methods in Cancer Research, Volume 1. The Analysis of Case-Control Studies. IARC Scientific Publication No. 32, IARC, Lyon, pp. 23–32.
28. Subramanian,J. *et al.* (2007) Lung cancer in never smokers: a review. *J. Clin. Oncol.*, **25**, 561–570.

Received November 8, 2009; revised December 26, 2009; accepted January 4, 2010

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

GENES, CHROMOSOMES & CANCER 49:342–352 (2010)

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

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.

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

TABLE 1. Homozygous Deletion and Expression of 17 Genes in Cancerous and Noncancerous Cultured Lung Cells (Extracted from Supplementary Table 1)

No.	Gene[a]	No. of cell lines with HD	Expression[b] T/N ratio	Expression[b] P value by t-test	Region no.	Chr band
3	LRP1B/[illegible]	7	1.19	0.30	3	2q22
16	FHIT	4	0.94	[illegible]	8	3p14
43	KANK1	1	0.81	[illegible]	19	9p24
47	PTPRD	8	1.26	0.0036	20	9p23
58	hsa-mir-491*	3			23	9p21
60	PTPLAD2	6	0.72	[illegible]		
78	hsa-mir-31*	10				
80/81	CDKN2A(p16/p14ARF)[e]	21	0.74	[illegible]		
82	CDKN2B(p15)	19	0.59	[illegible]		
131	RB1	1	0.80	[illegible]	34	13q14
155	SMAD4	1	0.90	[illegible]	41	18q21
157	KEAP1	2	0.97	0.46	42	19p13
167	hsa-mir-99a*	2			44	21q11
168	hsa-let-7c*	2				
169	hsa-mir-125b-2*	2				
175	ADAMTS1	1	0.39	[illegible]	45	21q11

Homozygous deletion / expression[c] (cell lines): ADC: A427, A549, ABC1, HCC193, Ma17, Ma24, H322, H1437, H1648, H2009, H2122, H2126, H2347, PC3, PC14, LC-MS, LC-OK, LCD, II-18, EBC-1; SQC: HCC95, LC1 Sq, H157, H520; ASC: HCC366; LCC: Lu99, Ma2, Ma25, H1299; SCC: N417, H841; Non-cancerous: NHBEC, SAEC, HBEC2-E, HBEC2-KT, HBEC3-ET, HBEC3-KT, HBEC4-KT, HBEC5-KT

[a]Genes deleted in multiple cases are in RED. MicroRNA genes are marked by *.

[b]Marked in BLUE and RED when T/N ratio was <1 and p value by t-test was <0.05 and 0.0001, respectively.

[c]Homozygous deletion: ■

[d]Expression examined by Affymetrix chips: Low – High: Presence call; □: Absence call

[e]p14ARF (Exon 1b) but not p16 is deleted in EBC1.

lines were also obtained from the American Type Culture Collection (Manassas, Virginia), the Japanese Collection of Research Bioresources (Tokyo, Japan), and the RIKEN BioResource Center (Tsukuba, Japan). Genomic DNA and poly A RNA were extracted by standard protocols.

Macro-dissected and micro-dissected cancerous and noncancerous lung cells were obtained from patients who were treated at the National Cancer Center Hospital, Tokyo, Japan. Details of these materials were described previously (Matsumoto et al., 2006; Nakamura et al., 2006). Genomic DNA and total RNA were extracted by standard protocols. This study was performed under the approval of the Institutional Review Board of National Cancer Center.

Detection of Homozygous Deletions by Array CGH Analysis

Copy number changes in genomic DNAs of 27 lung cancer cell lines were assessed using a Human CGH 185-k array covering 181,988 loci and Agilent CGH Analytics Software (Version 3.3) (Agilent Technologies). Genomic DNAs from these cell lines and 10 lymphoblastoid cell lines were analyzed according to the manufacturer's protocol using a human normal genomic DNA mix (Promega) as a reference. First, data for probes that were located in copy number variable regions deposited in the UCSC genome database and those that showed log2 ratios < -2.5 or >2.5 in 10 lymphoblastoid cell line DNAs in the present analyses were removed to mask copy number variable regions. Next, the copy number along the genome of 27 lung cancer cell lines was inferred by the Aberration Detection Method 2 (ADM2) algorithm. Autosomal regions encompassed by ≥ 2 consecutive probes with log2 ratios < -2.5 were defined as candidates for homozygously deleted regions. As the probes were placed at a mean interval of 15-kb, the present homozygous deletion search was undertaken at a 30-kb resolution.

Validation of Homozygous Deletions

One or more set(s) of PCR primers was designed to amplify genomic fragments that encompass an exon of genes deduced to be located in homozygously deleted regions by referring to the information in the UCSC database (http://genome.ucsc.edu/) and was subjected to multiplex PCR using the *IRF1* locus as a reference (Kishimoto et al., 2005). At least one exon was examined for each gene. Primer sequences are listed in Supporting Information Table 2. PCR products were separated by electrophoresis

TABLE 2. 34 Genes Mapped to Regions Other Than 9p21 and Homozygously Deleted in Two or More Lung Cancer Cell Lines

Chromosomal location	Gene	Gene product	No. of cell lines with homozygous deletion	(%)
9p23	PTPRD	Protein tyrosine phosphatase, receptor type D	8	(11)
2q21	LRP1B /LRP-DIP	Low density lipoprotein-related protein 1B	7	(9)
3p14	FHIT	Dinucleosidetriphosphatase	4	(5)
2q24	GRB14	Growth factor receptor-bound protein 14	2	(3)
2q24	COBLL1	COBL-like 1	2	(3)
2q24	SLC38A11	Solute carrier family 38, member 11	2	(3)
2q24	SCN3A	Sodium channel type III, alpha subunit	2	(3)
2q24	SCN2A	Sodium channel type II, alpha subunit	2	(3)
2q24	FAM130A2	TGF-beta induced apoptosis protein 2	2	(3)
2q24	GALNT3	UDP-N-acetyl-alpha-D-galactosamine	2	(3)
5q11	PDE4D	Phosphodiesterase 4D	2	(3)
5q31	CTNNA1	Alpha-catennin	2	(3)
7q35	CNTNAP2	Contactin associated protein-like 2	2	(3)
10p11	PARD3	Par-3 partitioning defective 3 homolog	2	(3)
18q21	ME2	Malate dehydrogenase 2	2	(3)
18q21	ELAC1	elaC homolog 1	2	(3)
19p13	KEAP1	Cytosolic inhibitor of Nrf2	2	(3)
19q13	MZF1	Myeloid zinc finger 1	2	(3)
19q13	MGC2752	Hypothetical LOC65996	2	(3)
21q11	LIPI	Membrane-associated phospholipase A1 beta	2	(3)
21q11	RBM11	RNA binding motif protein 11	2	(3)
21q11	STCH	Stress 70 protein chaperone	2	(3)
21q11	SAMSN1	SAM domain, SH3 domain and nuclear localization signals 1	2	(3)
21q11	NRIP1	Nuclear receptor interacting protein 1	2	(3)
21q11	USP25	Ubiquitin specific peptidase 25	2	(3)
21q11	C21orf34	Chromosome 21 open reading frame 34	2	(3)
21q11	hsa-mir-99a	miRNA	2	(3)
21q11	hsa-let-7c	miRNA	2	(3)
21q11	hsa-mir-125b-2	miRNA	2	(3)
21q11	CXADR	Coxsackie virus and adenovirus receptor	2	(3)
21q11	BTG3/ANA	BTG/Tob family protein	2	(3)
21q11	C21orf91	Chromosome 21 open reading frame 91	2	(3)
21q11	CHODL	Transmembrane protein MT75	2	(3)
21q11	PRSS7	Enterokinase	2	(3)

on 3% agarose gel and visualized by staining with ethidium bromide. When no PCR product was detected, such an exon was judged as being homozygously deleted.

Expression Analysis of the 176 Genes Homozygously Deleted in Lung Cancer

Information on the expression levels for 23,583 genes of 15 lung cancer cell lines (A549, H157, H322, H1299, H1437, H1648, H2009, H2122, H2126, H2347, HCC95, HCC193, HCC366, and HCC515) was previously obtained by analysis using Affymetrix Gene Chips HG-U133A and HG-U133B (Zhou et al., 2006). Information on those of four other lung cancer cell lines (H223, H209, H841, and H2141) and cultured noncancerous lung epithelial cells (Ramirez et al., 2004) was obtained for the present study. Expression data for 160 of the 176 genes homozygously deleted in lung cancer cell lines were available, and these 160 genes were assessed by 281 probes (Supporting Information Table 1). Differences in expression levels between lung cancer cell lines and noncancerous lung epithelial cells were examined by t test, and probes with $P < 0.05$ were judged as significantly different. In addition, probes with $P < 0.00018$ were judged as significantly different after Bonferroni correction for multiple tests (i.e., $0.05/281 = 0.00018$).

Mutation Analysis of the *PTPRD* Gene

All coding exons of the *PTPRD* gene were amplified from 10 ng of DNA from 74 cell lines and 95 surgical specimens of lung cancer by PCR

using 44 sets of primers. PCR products from the cell lines were directly sequenced using a Big Dye Terminator Sequencing kit and an ABI Prism 3700 Genetic Analyzer (Applied Biosystems, Foster City, California, USA). PCR products from the surgical specimens were subjected to WAVE analysis according to the manufacturer's protocol (Transgenomic, Omaha, Nebraska, USA). PCR products with different mobilities in the WAVE analysis were purified and directly sequenced.

Quantitative Real-time Reverse Transcription PCR (QRT-PCR) Analysis

Expression levels of the *PTPRD* gene were evaluated by QRT-PCR using ABI Prism 7900HT (Applied Biosystems). A Taqman probe (5′-AGGATCAATATCAGTTTTCCTA-3′) and a set of primers (5′-TGTTAAGAACACAACGACCAGCTAT-3′ and 5′- TCAAAGCTGCCCAGGTACTCTAGT-3′) were used as previously described (Sato et al., 2005). PCR was performed in a single tube in duplicate. Results were expressed as the average of these two independent tests.

RESULTS AND DISCUSSION

Identification of Genes Homozygously Deleted in 52 Human Lung Cancer Cell Lines

Homozygously deleted regions in 43 lung cancer cell lines were previously searched for at a 100-kb resolution by a SNP array analysis, and 113 genes were deduced to map to homozygously deleted regions in one or more cell lines (Nagayama et al., 2007). These 113 genes consisted of three genes, *CDKN2A/p16*, *p14ARF* (a gene sharing the same exons with *CDKN2A/p16* but encoding a different protein) (Stone et al., 1995) and *CDKN2B/p15*, which are considered target tumor suppressor genes for homozygous deletions at chromosome band 9p21 (Hamada et al., 2000), and 110 genes located in regions other than 9p21 (Fig. 1).

In this study, homozygous deletions were further searched for by an array CGH analysis at a 30-kb resolution in 27 lung cancer cell lines consisting of 18 cell lines previously analyzed (Nagayama et al., 2007) and 9 cell lines newly prepared (Fig. 1). Among 113 genes found to be deleted in the previous study (Nagayama et al., 2007), 111 were verified in this study. Two genes, *THSD4* and *C20orf133*, considered to be homozygously deleted in the previous study, were not found to be deleted here. Therefore, these two genes were excluded from further analyses. In addition, 83 genes not indicated in the previous study were deduced to be homozygously deleted in the present 30-kb resolution analysis. These 83 genes consisted of 37 genes in the 9p21 region and 46 genes in other regions. Thus, in total, 194 genes (111 + 83) were considered to be homozygously deleted in 52 lung cancer cell lines by using two different DNA array analyses.

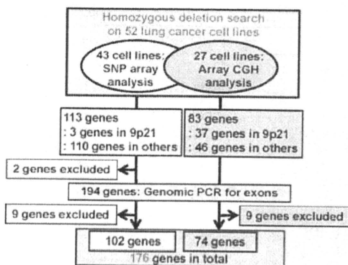

Figure 1. Strategy to identify genes homozygously deleted in lung cancer. A previous search on 43 lung cancer cell lines led to the identification of 113 genes deduced to be homozygously deleted (Nagayama et al., 2007). The present search on 27 lung cancer cell lines (18 cell lines overlaped) led to the identification of an additional 83 genes, and excluded two genes from the 113 genes above. Genomic PCR for exons of 194 genes led to the validation of homozygous deletions in 176 genes.

To confirm the homozygous deletions of these 194 genes, genomic PCR against DNA fragments encompassing an exon located in a homozygously deleted region was performed for all of the 194 genes (primer information in Supporting Information Table 2). One hundred seventy-six (91%) of the 194 genes showed homozygous deletions of at least one exon (Fig. 1). These 176 genes included *CDKN2A/p16* and *CDKN2B/p15*, and the results for deletions of these two genes determined by genomic PCR were consistent with the results determined by Southern blot analysis in our previous study (Okamoto et al., 1995). On the other hand, homozygous deletions of exons were not detected in the remaining 18 genes, probably due to the fact that only intronic or intergenic sequences were deleted or the deletions deduced were spurious ones caused by experimental errors. Therefore, these 18 genes were excluded from the remaining part of this study. Homozygous deletion of one or more genes was detected in 20 (74%) of the 27 cell lines analyzed

TABLE 3. PTPRD Mutations in Lung Cancer

Sample	Histological type	No. of exon	(nucleotide change)	Predicted effect	mRNA level[a]
Cell line					
Ma29	ADC	11	(C1184T: Homo)	Ala395Val	0.0015
H23	ADC	11	(C1201T: Homo)	Arg401Trp	0.053
Sq-5	SQC	20	(C3299T: Hetero)	Thr1100Met	0.00068
H1155	LCC	14	(C2057T: Hetero)	Thr686Ile	0.035
PC13	LCC	20	(A3164G: Hetero)	Asp1055Gly	0.010
Lu65	LCC	32	(G5258T: Homo)	Gly1753Cys	0.00047
H2171[b]	SCC	5	(G460T: Homo)	Asp154Tyr	0.025
H526	SCC	17	(A2443G: Hetero)	Lys815Glu	0.70
Surgical specimen					
Na68T[b]	SQC	4	(G235T: Hetero)	Gly79STOP	0.10
Na182T[b]	SQC	21	(TG3472-3473AT: Hetero)	Trp1158Met	0.27
S171T[b]	SCC	15	(G2206T: Hetero)	Val1736Leu	Not tested
1662T[b]	SCC	17	(C IVS17−16 T: Hetero)	Unknown	Not tested

[a]Relative expression level to noncancerous lung tissues.
[b]Validated to be somatic mutation.

by the present array CGH analysis, but not in the remaining seven cell lines. Among a total of 52 cell lines subjected to the present and/or previous homozygous deletion scanning, homozygous deletion of one or more genes was detected in 36 (69%) cell lines.

Characteristics and Genomic Status of 176 Genes Homozygously Deleted in Lung Cancer Cell Lines

The 176 genes with verified homozygous deletions are listed in Table 1 and Supporting Information Table 1. They consisted of 171 protein-encoding genes and five miRNA genes (genes 58, 78, 167–169). These 176 genes were located in 45 regions on 17 chromosomes (Supporting Information Fig. 1). They included known tumor suppressor genes, *CDKN2A/p16*, *p14ARF*, *CDKN2B/p15*, *RB1*, and *SMAD4* (genes 80–82, 131, and 155 in Table 1) (Futreal et al., 2004), as well as candidate tumor suppressor genes shown to be hemizygously or homozygously deleted in several types of human cancers, such as *LRP1B/LRP-DIP*, *FHIT*, *PTPRD* and *KEAP1* (genes 3, 18, 47, and 157 in Table 1) (Sozzi et al., 1996; Liu et al., 2000; Sonoda et al., 2004; Sato et al., 2005; Singh et al., 2006; Stallings et al., 2006; Ohta et al., 2008).

Frequencies of homozygous deletion for these 176 genes were examined in 74 lung cancer cell lines, consisting of 52 cell lines used for the array analyses and 22 additional cell lines, by genomic PCR using the same sets of primers as described above (Supporting Information Table 2). One to 75 of these genes were homozygously deleted in 44 of the 74 cell lines analyzed. No gene was deleted in the remaining 30 cell lines (Supporting InformationTable 1). Homozygous deletion of each gene was detected in one (1%) to 20 (27%) of the 74 cell lines. Sixty-four (36%) of the 176 genes were deleted in two or more cell lines (Supporting Information Table 3), while the other 112 (64%) were deleted in a single cell line. The *CDKN2A/p16* and *p14ARF* genes (genes 80–81) in the 9p21 region were most frequently deleted (20/74, 27%). Thirty-four genes deleted in two or more cell lines were located in regions other than 9p21 (Table 2). A candidate tumor suppressor gene, *PTPRD* (gene 47), was most frequently deleted (8/74, 11%) among them. Other known candidate tumor suppressors, *LRP1B* (gene 3), *FHIT* (gene 18), and *KEAP1* (gene 157), were also included in these 34 genes. Therefore, other genes listed in Table 2 will be also strong candidates for lung tumor suppressors.

Expression Status of the 176 Genes Homozygously Deleted in Lung Cancer Cell Lines

Nineteen of the 44 cell lines with homozygous deletion were available for information on expression levels of 23,583 genes obtained by microarray analysis. These cell lines included all four major histological types of lung cancers. Information on expression levels was available for 160 (91%) of the 176 genes. Most of these genes showed nonsignificant signals (i.e., absent call) in cell lines with homozygous deletion of the corresponding genes, and such genes were *LRP1B* (gene 3), *PTPRD* (gene 47), *CDKN2A/p16*, and *p14ARF* (genes 80–81 assessed by the same probes), *CDKN2B* (gene 82), *RB1* (gene 131) and *KEAP1* (gene 157) (Table 1). On the other hand,

some genes with deletions of parts of genes, such as *FHIT* (gene 18), showed significant signals (i.e., present call) in cell lines with homozygous deletions. As for the *FHIT* gene, transcripts lacking exons, which are homozygously deleted, were previously shown to be expressed in lung cancer cells (Sozzi et al., 1996).

Expression data on eight cultured noncancerous lung epithelial cells were also available for the same set of genes as in lung cancer cell lines (Zhou et al., 2006). Therefore, we searched for genes whose expression was significantly lower in lung cancer cells compared to noncancerous lung epithelial cells. In 55 (31%) genes, at least one probe showed a level of expression significantly lower than that in noncancerous lung epithelial cells (T/N ratio <1 and $P < 0.05$ by t test, marked in blue in Table 1 and Supporting Information Table 1). Expression levels in 52 (95%) of these 55 genes remained significantly lower after removing lung cancer cases with homozygous deletion. The differences in expression of 18 genes (10%) remained significant after Bonferroni correction for multiple tests (i.e., $P < 0.00018$, marked in red in Table 1 and Supporting Information Table 1), and that of 13 (72%) genes remained significant after removing cases with homozygous deletion. These genes included the *KANK* and *ADAMTS1* (genes 131 and 175) candidate tumor suppressor genes whose down-regulation by epigenetic alterations rather than genetic alterations in renal and lung cancers, respectively, were reported (Sarkar et al., 2002; Choi et al., 2008). It was noted that homozygous deletions of these three genes were detected only in one cell line, respectively. The results suggest that the present 176 genes include genes preferentially inactivated in lung cancer cells by epigenetic alterations rather than homozygous deletions.

PTPRD Alterations in Human Lung Cancer

Homozygous deletions and mutations in the *PTPRD* gene in human lung cancer and other cancers, as well as the ability of PTPRD protein to inhibit growth and to cause apoptosis have indicated that *PTPRD* is a tumor suppressor gene (Sjoblom et al., 2006; Weir et al., 2007; Ding et al., 2008; Solomon et al., 2008; Veeriah et al., 2009). Thus, we searched for mutations in the *PTPRD* gene in both cell lines and surgical specimens of lung cancer. Sequencing of all coding exons in 74 lung cancer cell lines revealed that eight cell lines (11%) had nonsynonymous (i.e., associated with amino acid change) nucleotide substitutions that were not deposited in the dbSNP database (Table 3). The substitution in H2171 cells was validated to be a somatic mutation, by the absence of this substitution in the corresponding lymphoblastoid cell line (Supporting Information Fig. 2A). The other seven substitutions detected in the remaining seven cell lines were also likely to be somatic mutations because these substitutions were not detected in noncancerous cells of 95 different individuals (see below), and each of them was detected in only one of the 74 lung cancer cell lines and in none of the 95 primary tumors. Among the 95 surgical specimens analyzed for *PTPRD* mutations, four cases (4%) were concluded as having somatic mutations because nucleotide substitutions were detected only in cancer cells and not in the corresponding noncancerous cells (Table 3). One was a nonsense mutation, two were missense mutations, and the remaining one was a mutation in an intronic sequence. By RT-PCR and sequencing, mutant alleles were shown to be expressed in all of the eight cell lines with *PTPRD* mutations and a surgical specimen whose RNA was available for analysis (Supporting Information Fig. 2A).

The *PTPRD* mutations detected in this study were dispersed through the PTPRD protein as previously observed in lung and others cancers (Fig. 2A) (Sjoblom et al., 2006; Weir et al., 2007; Solomon et al., 2008; Veeriah et al., 2009). It was noted that the same mutations were not present among the mutations detected in human cancers up to the present, and hot spots for mutations were not obvious (Supporting Information Fig. 2B and 2C). A recent study indicated that several mutant *PTPRD* proteins have lower abilities than the wild-type protein to inhibit growth and to cause apoptosis in cells (Solomon et al., 2008). In addition, a subset of mutations, including Gly79X detected in the present study, were nonsense mutations causing a production of truncated PTPRD proteins lacking the whole or a part of protein tyrosine phosphatase catalytic domains. These results indicate that somatic *PTPRD* mutations are a genetic event causing functional inactivation of the *PTPRD* gene in human cancer cells.

We next examined the expression of the *PTPRD* gene in both cell lines and surgical specimens of lung cancer. We previously reported that the majority (>90%) of lung cancer cell lines, including eight cell lines with homozygous *PTPRD* deletions showed lower expression levels

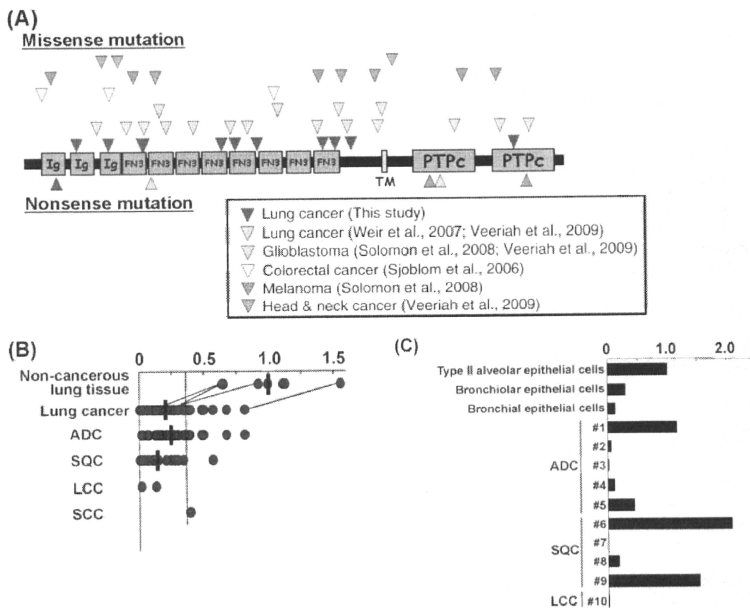

Figure 2. *PTPRD* mutation and expression in lung cancer. (A) Location of missense and nonsense mutations detected in the present and previous studies. Ig, immunoglobulin-like C2-type domain; FN3, fibronectin type III domain; TM, transmembrane domain; PTPc, protein tyrosine phosphatase catalytic domain. (B) Expression in macro-dissected cancerous and noncancerous lung cells. Values for four paired noncancerous and cancerous lung tissues are connected. Expression levels are indicated after adjusting the mean for the levels of expression in seven cases of noncancerous lung tissues to 1. Mean values are indicated by horizontal bars if the group has three or more samples. The threshold level to judge as reduced expression is indicated by a dashed line. (C) Expression in micro-dissected cancerous and noncancerous lung cells. The levels of *PTPRD* expression relative to those of *GAPDH* expression are shown after adjusting the level of *PTPRD* expression in type II alveolar epithelial cells to 1.

than noncancerous lung cells (Sato et al., 2005). In this study, eight cell lines with *PTPRD* mutations were also shown to express lower levels compared to normal lung tissue (Table 3). Sixty surgical specimens of lung tumors were also subjected to QRT-PCR expression analysis. Fifty-one (85%) specimens showed lower levels of *PTPRD* expression than the mean-2SD for seven noncancerous lung tissues, therefore, these samples were judged as having a reduced *PTPRD* expression (Fig. 2B). Two specimens with *PTPRD* mutations for which RNA was available also showed reduced *PTPRD* expression (Table 3). Differences in *PTPRD* levels between noncancerous lung tissues and all lung cancers, ADCs or SQCs were significant ($P < 0.05$ by t test). Two LCCs also showed reduced *PTPRD* expression. A SCC case examined also showed a lower level of *PTPRD*, however, was not judged as having a reduced *PTPRD* expression according to the criteria above.

We also examined *PTPRD* expression in noncancerous lung component cells and cancerous cells which were obtained by micro-dissection of surgical specimens (Nakamura et al., 2006). *PTPRD* expression was detected in noncancerous lung component cells with the highest expression in type II alveolar epithelial cells, candidate precursor cells for lung ADC (Otto, 2002) (Fig. 2C). The levels of expression in five lung cancer

samples were lower than those of any component cells (cases 2, 3, 4, 7, and 10 in Fig. 2C). These results suggested that reduced *PTPRD* expression commonly occurs in lung carcinogenesis. Recently, reduced expression of *PTPRD* was shown to be a frequent event in human glioblastoma, and to be caused by hypermethylation of the promoter region of the *PTPRD* gene (Veeriah et al., 2009). Therefore, it was strongly suggested that *PTPRD* is silenced by a promoter hypermethylation also in lung cancer, although methylation status of the *PTPRD* gene was not examined in this study.

CONCLUSION

We identified 176 genes homozygously deleted in human lung cancer. These genes included known tumor suppressor genes and candidate tumor suppressor genes, whose hemizygous or homozygous deletions as well as intragenic mutations had been reported in several types of human cancers. Furthermore, these 176 genes include genes preferentially inactivated by epigenetic alterations, such as *KANK* and *ADAMTS1*. Indeed, one of these candidates, *PTPRD*, was shown to be genetically and/or epigenetically altered in a considerable fraction of lung cancer. Therefore, this set of genes will be informative to identify novel lung tumor suppressor genes. The Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) suggests that genes with specific functions or involved in specific signaling pathways are not significantly enriched among these 176 genes. Thus, it was suggested that genes involved in a variety of biological processes could function as lung tumor suppressors. Further genetic/epigenetic as well as functional studies of these 176 genes will help understanding of molecular mechanism of lung carcinogenesis. In fact, the present homozygous deletion scanning was performed on a set of 52 lung cancer cell lines including all four major histological types of lung cancer. However, fractions of ADC (30/52, 58%) and SQC (7/52, 13%) were larger and smaller than those in the population of lung cancer patients (Parkin et al., 2004), respectively, therefore, scanning of other sets of lung cancer that are predominant for SQC might provide additional genes.

One to 75 genes were homozygously deleted in 44 lung cancer cell lines analyzed, while no genes was deleted in the remaining 30 cell lines. The result might imply that the intrinsic genomic stability against homozygous deletion is different among lung cancer cases. We recently reported genetic/epigenetic alteration profiles of known oncogenes and tumor suppressor genes in the cell lines used in this study (Medina et al., 2008; Blanco et al., 2009). Therefore, relationships between these alterations and homozygous deletions of those 176 genes were examined. Interestingly, numbers of genes with homozygous deletions are significantly or marginally significantly different according to alterations of tumor suppressor genes, *TP53*, *CDKN2A/p16*, *LKB1*, and *PTEN* (Supporting Information Table 4). Multivariate analysis indicated that only the *CDKN2A/p16* alteration among them was independently associated with the number of genes homozygously deleted. The result suggests that *CDKN2A/p16* alteration is involved in genomic instability inducing homozygous deletions as this gene is critical for the maintenance of genome integrity in human cells (McDermott et al., 2006). Interestingly, the cell lines with homozygous *CDKN2A/p16* deletions carried deletions of significantly larger number of genes than those with promoter hypermethylation and mutation of the *CDKN2A/p16* gene; and those without (Supporting Information Fig. 3). Therefore, homozygous *CDKN2A/p16* deletion can be a marker of intrinsic instability for homozygous deletion. More detailed analysis of genetic/epigenetic interactions as well as functional interactions among genes altered in lung cancer cells will further provide insights into the molecular mechanism of lung carcinogenesis.

ACKNOWLEDGMENTS

The authors thank Kaho Minoura and Yayoi Fukuoka of Agilent Technologies Japan for technical assistance in array-CGH analysis.

REFERENCES

Blanco R, Iwakawa R, Tang M, Kohno T, Angulo B, Pio R, Montuenga LM, Minna JD, Yokota J, Sanchez-Cespedes M. 2009. A gene-alteration profile of human lung cancer cell lines. Hum Mutat 30:1199–1206.

Burbee DG, Forgacs E, Zochbauer-Muller S, Shivakumar L, Fong K, Gao B, Randle D, Kondo M, Virmani A, Bader S, Sekido Y, Latif F, Milchgrub S, Toyooka S, Gazdar AF, Lerman MI, Zabarovsky E, White M, Minna JD. 2001. Epigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression. J Natl Cancer Inst 93:691–699.

Choi JE, Kim DS, Kim EJ, Chae MH, Cha SI, Kim CH, Jheon S, Jung TH, Park JY. 2008. Aberrant methylation of ADAMTS1 in non-small cell lung cancer. Cancer Genet Cytogenet 187:80–84.

Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, Fulton L, Fulton RS, Zhang Q, Wendl MC, Lawrence MS,

Larson DE, Chen K, Dooling DJ, Sabo A, Hawes AC, Shen H, Jhangiani SN, Lewis LR, Hall O, Zhu Y, Mathew T, Ren Y, Yao J, Scherer SE, Clerc K, Metcalf GA, Ng B, Milosavljevic A, Gonzalez-Garay ML, Osborne JR, Meyer R, Shi X, Tang Y, Koboldt DC, Lin L, Abbott R, Miner TL, Pohl C, Fewell G, Haipek C, Schmidt H, Dunford-Shore BH, Kraja A, Crosby SD, Sawyer CS, Vickery T, Sander S, Robinson J, Winckler W, Baldwin J, Chirieac LR, Dutt A, Fennell T, Hanna M, Johnson BE, Onofrio RC, Thomas RK, Tonon G, Weir BA, Zhao X, Ziaugra L, Zody MC, Giordano T, Orringer MB, Roth JA, Spitz MR, Wistuba, II, Ozenberger B, Good PJ, Chang AC, Beer DG, Watson MA, Ladanyi M, Broderick S, Yoshizawa A, Travis WD, Pao W, Province MA, Weinstock GM, Varmus HE, Gabriel SB, Lander ES, Gibbs RA, Meyerson M, Wilson RK. 2008. Somatic mutations affect key pathways in lung adenocarcinoma. Nature 455:1069–1075.

Dryja TP, Rapaport JM, Joyce JM, Petersen RA. 1986. Molecular detection of deletions involving band q14 of chromosome 13 in retinoblastomas. Proc Natl Acad Sci USA 83:7391–7394.

Fearon ER, Cho KR, Nigro JM, Kern SE, Simons JW, Ruppert JM, Hamilton SR, Preisinger AC, Thomas G, Kinzler KW, Vogelstein B. 1990. Identification of a chromosome 18q gene that is altered in colorectal cancers. Science 247:49–56.

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. 2004. A census of human cancer genes. Nat Rev Cancer 4:177–183.

Garnis C, Lockwood WW, Vucic E, Ge Y, Girard L, Minna JD, Gazdar AF, Lam S, Macaulay C, Lam WL. 2006. High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH. Int J Cancer 118:1556–1564.

Girard L, Zochbauer-Muller S, Virmani AK, Gazdar AF, Minna JD. 2000. Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering. Cancer Res 60:4894–4906.

Hahn SA, Schutte M, Hoque AT, Moskaluk CA, da Costa LT, Rozenblum E, Weinstein CL, Fischer A, Yeo CJ, Hruban RH, Kern SE. 1996. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science 271:350–353.

Hamada K, Kohno T, Takahashi M, Yamazaki M, Tashiro H, Sugawara C, Ohwada S, Sekido Y, Minna JD, Yokota J. 2000. Two regions of homozygous deletion clusters at chromosome band 9p21 in human lung cancer. Genes Chromosomes Cancer 27:308–318.

Herbst RS, Heymach JV, Lippman SM. Lung cancer. N Engl J Med 359:1367–1380. 2008.

Imoto I, Izumi H, Yokoi S, Hosoda H, Shibata T, Hosoda F, Ohki M, Hirohashi S, Inazawa J. 2006. Frequent silencing of the candidate tumor suppressor PCDH20 by epigenetic mechanism in non-small-cell lung cancers. Cancer Res 66:4617–4626.

Kamb A, Gruis NA, Weaver-Feldhaus J, Liu Q, Harshman K, Tavtigian SV, Stockert E, Day RS, III, Johnson BE, Skolnick MH. 1994. A cell cycle regulator potentially involved in genesis of many tumor types. Science 264:436–440.

Kawanishi M, Kohno T, Otsuka T, Adachi J, Sone S, Noguchi M, Hirohashi S, Yokota J. 1997. Allelotype and replication error phenotype of small cell lung carcinoma. Carcinogenesis 18:2057–2062.

Kishimoto M, Kohno T, Okudela K, Otsuka A, Sasaki H, Tanabe C, Sakiyama T, Hirama C, Kitabayashi I, Minna JD, Takenoshita S, Yokota J. 2005. Mutations and deletions of the CBP gene in human lung cancer. Clin Cancer Res 11:512–519.

Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, Puc J, Miliaresis C, Rodgers L, McCombie R, Bigner SH, Giovanella BC, Ittmann M, Tycko B, Hibshoosh H, Wigler MH, Parsons R. 1997. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 275:1943–1947.

Liu CX, Musco S, Lisitsina NM, Forgacs E, Minna JD, Lisitsyn NA. 2000. LRP-DIT, a putative endocytic receptor gene, is frequently inactivated in non-small cell lung cancer cell lines. Cancer Res 60:1961–1967.

Matsumoto S, Iwakawa R, Kohno T, Suzuki K, Matsuno Y, Yamamoto S, Noguchi M, Shimizu E, Yokota J. 2006. Frequent EGFR mutations in noninvasive bronchioloalveolar carcinoma. Int J Cancer 118:2498–2504.

McDermott KM, Zhang J, Holst CR, Kozakiewicz BK, Singla V, Tlsty TD. 2006. p16(INK4a) prevents centrosome dysfunction and genomic instability in primary cells. PLoS Biol 4:e51.

Medina PP, Romero OA, Kohno T, Montuenga LM, Pio R, Yokota J, Sanchez-Cespedes M. 2008. Frequent BRG1/SMARCA4-inactivating mutations in human lung cancer cell lines. Hum Mutat 29:617–622.

Minna JD, Roth JA, Gazdar AF. 2002. Focus on lung cancer. Cancer Cell 1:49–52.

Nagayama K, Kohno T, Sato M, Arai Y, Minna JD, Yokota J. 2007. Homozygous deletion scanning of the lung cancer genome at a 100-kb resolution. Genes Chromosomes Cancer 46:1000–1010.

Nakamura N, Kobayashi K, Nakamoto M, Kohno T, Sasaki H, Matsuno Y, Yokota J. 2006. Identification of tumor markers and differentiation markers for molecular diagnosis of lung adenocarcinoma. Oncogene 25:4245–4255.

Nakanishi H, Matsumoto S, Iwakawa R, Kohno T, Suzuki K, Tsuta K, Matsuno Y, Noguchi M, Shimizu E, Yokota J. 2009a. Whole genome comparison of allelic imbalance between noninvasive and invasive small-sized lung adenocarcinomas. Cancer Res 69:1615–1623.

Nobori T, Miura K, Wu DJ, Lois A, Takabayashi K, Carson DA. 1994. Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers. Nature 368:753–756.

Ogiwara H, Kohno T, Nakanishi H, Nagayama K, Sato M, Yokota J. 2008. Unbalanced translocation, a major chromosome alteration causing loss of heterozygosity in human lung cancer. Oncogene 27:4788–4797.

Ohta T, Iijima K, Miyamoto M, Nakahara I, Tanaka H, Ohtsuji M, Suzuki T, Kobayashi A, Yokota J, Sakiyama T, Shibata T, Yamamoto M, Hirohashi S. 2008. Loss of Keap1 function activates Nrf2 and provides advantages for lung cancer cell growth. Cancer Res 68:1303–1309.

Okamoto A, Hussain SP, Hagiwara K, Spillare EA, Rusin MR, Demetrick DJ, Serrano M, Hannon GJ, Shiseki M, Zariwala M, Xiong Y, Beach DH, Yokota J, Harris CC. 1995. Mutations in the p16INK4/MTS1/CDKN2, p15INK4B/MTS2, and p18 genes in primary and metastatic lung cancer. Cancer Res 55:1448–1451.

Otto WR. 2002. Lung epithelial stem cells. J Pathol 197:527–535.

Parkin M, Je T, Boffetta P, Samet J, Shields PG, Caporaso NE. 2004. Lung cancer epidemiology and etiology. In: Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, editors. World Health Organization Classification of Tumors: Pathology and Genetics, Tumours of Lung, Pleura, Thymus and Heart. Lyon, France: IARC Press. pp. 31–34.

Ramirez RD, Sheridan S, Girard L, Sato M, Kim Y, Pollack J, Peyton M, Zou Y, Kurie JM, Dimaio JM, Milchgrub S, Smith AL, Souza RF, Gilbey L, Zhang X, Gandia K, Vaughan MB, Wright WE, Gazdar AF, Shay JW, Minna JD. 2004. Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. Cancer Res 64:9027–9034.

Sarkar S, Roy BC, Hatano N, Aoyagi T, Gohji K, Kiyama R. 2002. A novel ankyrin repeat-containing gene (Kank) located at 9p24 is a growth suppressor of renal cell carcinoma. J Biol Chem 277:36585–36591.

Sato M, Takahashi K, Nagayama K, Arai Y, Ito N, Okada M, Minna JD, Yokota J, Kohno T. 2005. Identification of chromosome arm 9p as the most frequent target of homozygous deletions in lung cancer. Genes Chromosomes Cancer 44:405–414.

Shiseki M, Kohno T, Adachi J, Okazaki T, Otsuka T, Mizoguchi H, Noguchi M, Hirohashi S, Yokota J. 1996. Comparative allelotype of early and advanced stage non-small cell lung carcinomas. Genes Chromosomes Cancer 17:71–77.

Singh A, Misra V, Thimmulappa RK, Lee H, Ames S, Hoque MO, Herman JG, Baylin SB, Sidransky D, Gabrielson E, Brock MV, Biswal S. 2006. Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer. PLoS Med 3:e420.

Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE. 2006. The consensus coding sequences of human breast and colorectal cancers. Science 314:268–274.

Solomon DA, Kim JS, Cronin JC, Sibenaller Z, Ryken T, Rosenberg SA, Ressom H, Jean W, Bigner D, Yan H, Samuels Y, Waldman T. 2008. Mutational inactivation of PTPRD in glioblastoma multiforme and malignant melanoma. Cancer Res 68:10300–10306.

Sonoda I, Imoto I, Inoue J, Shibata T, Shimada Y, Chin K, Imamura M, Amagasa T, Gray JW, Hirohashi S, Inazawa J. 2004.

Frequent silencing of low density lipoprotein receptor-related protein 1B (LRP1B) expression by genetic and epigenetic mechanisms in esophageal squamous cell carcinoma. Cancer Res 64:3741–3747.

Sozzi G, Veronese ML, Negrini M, Baffa R, Cotticelli MG, Inoue H, Tornielli S, Pilotti S, De Gregorio L, Pastorino U, Pierotti MA, Ohta M, Huebner K, Croce CM. 1996. The FHIT gene 3p14.2 is abnormal in lung cancer. Cell 85:17–26.

Stallings RL, Nair P, Maris JM, Catchpoole D, McDermott M, O'Meara A, Breatnach F. 2006. High-resolution analysis of chromosomal breakpoints and genomic instability identifies PTPRD as a candidate tumor suppressor gene in neuroblastoma. Cancer Res 66:3673–3680.

Steck PA, Pershouse MA, Jasser SA, Yung WK, Lin H, Ligon AH, Langford LA, Baumgard ML, Hattier T, Davis T, Frye C, Hu R, Swedlund B, Teng DH, Tavtigian SV. 1997. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. Nat Genet 15:356–362.

Stone S, Jiang P, Dayananth P, Tavtigian SV, Katcher H, Parry D, Peters G, Kamb A. 1995. Complex structure and regulation of the P16 (MTS1) locus. Cancer Res 55:2988–2994.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102:15545–15550.

Toloza EM, Morse MA, Lyerly HK. 2006. Gene therapy for lung cancer. J Cell Biochem 99:1–22.

Tonon G, Wong KK, Maulik G, Brennan C, Feng B, Zhang Y, Khatry DB, Protopopov A, You MJ, Aguirre AJ, Martin ES, Yang Z, Ji H, Chin L, Depinho RA. 2005. High-resolution genomic profiles of human lung cancer. Proc Natl Acad Sci USA 102:9625–9630.

Veeriah S, Brennan C, Meng S, Singh B, Fagin JA, Solit DB, Paty PB, Rohle D, Vivanco I, Chmielecki J, Pao W, Ladanyi M, Gerald WL, Liau L, Cloughesy TC, Mischel PS, Sander C, Taylor B, Schultz N, Major J, Heguy A, Fang F, Mellinghoff IK, Chan TA. 2009. The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers. Proc Natl Acad Sci USA 106:9435–9440.

Virmani AK, Fong KM, Kodagoda D, McIntire D, Hung J, Tonk V, Minna JD, Gazdar AF. 1998. Allelotyping demonstrates common and distinct patterns of chromosomal loss in human lung cancer types. Genes Chromosomes Cancer 21:308–319.

Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja A, Johnson LA, Shah K, Sato M, Thomas RK, Barletta JA, Borecki IB, Broderick S, Chang AC, Chiang DY, Chirieac LR, Cho J, Fujii Y, Gazdar AF, Giordano T, Greulich H, Hanna M, Johnson BE, Kris MG, Lash A, Lin L, Lindeman N, Mardis ER, McPherson JD, Minna JD, Morgan MB, Nadel M, Orringer MB, Osborne JR, Ozenberger B, Ramos AH, Robinson J, Roth JA, Rusch V, Sasaki H, Shepherd F, Sougnez C, Spitz MR, Tsao MS, Twomey D, Verhaak RG, Weinstock GM, Wheeler DA, Winckler W, Yoshizawa A, Yu S, Zakowski MF, Zhang Q, Beer DG, Wistuba, II, Watson MA, Garraway LA, Ladanyi M, Travis WD, Pao W, Rubin MA, Gabriel SB, Gibbs RA, Varmus HE, Wilson RK, Lander ES, Meyerson M. 2007. Characterizing the cancer genome in lung adenocarcinoma. Nature 450:893–898.

Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. 1995. Identification of the breast cancer susceptibility gene BRCA2. Nature 378:789–792.

Yokota J, Kohno T. 2004. Molecular footprints of human lung cancer progression. Cancer Sci 95:197–204.

Zhao X, Weir BA, LaFramboise T, Lin M, Beroukhim R, Garraway L, Beheshti J, Lee JC, Naoki K, Richards WG, Sugarbaker D, Chen F, Rubin MA, Janne PA, Girard L, Minna J, Christiani D, Li C, Sellers WR, Meyerson M. 2005. Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. Cancer Res 65:5561–5570.

Zhou BB, Peyton M, He B, Liu C, Girard L, Caudler E, Lo Y, Baribaud F, Mikami I, Reguart N, Yang G, Li Y, Yao W, Vaddi K, Gazdar AF, Friedman SM, Jablons DM, Newton RC, Fridman JS, Minna JD, Scherle PA. 2006. Targeting ADAM-mediated ligand cleavage to inhibit HER3 and EGFR pathways in non-small cell lung cancer. Cancer Cell 10:39–50.

Carcinogenesis vol.31 no.10 pp.1794–1799, 2010
doi:10.1093/carcin/bgq159
Advance Access publication August 5, 2010

Association of *CYP19A1* polymorphisms with risks for atypical adenomatous hyperplasia and bronchioloalveolar carcinoma in the lungs

Takashi Kohno, Ryutaro Kakinuma[1], Motoki Iwasaki[2], Taiki Yamaji[2], Hideo Kunitoh[3,12], Kenji Suzuki[4,13], Yoko Shimada, Kouya Shiraishi, Yoshio Kasuga[5], Gerson Shigeaki Hamada[6], Koh Furuta[7], Koji Tsuta[8], Hiromi Sakamoto[9], Aya Kuchiba[9], Seiichiro Yamamoto[10], Yae Kanai[11], Shoichiro Tsugane[2] and Jun Yokota*

Biology Division, National Cancer Center Research Institute, Tokyo 104-0045, Japan, [1]Cancer Screening Technology and Development Division, National Cancer Center, Tokyo 104-0045, Japan, [2]Epidemiology and Prevention Division, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo 104-0045, Japan, [3]Thoracic Oncology Division and [4]Thoracic Surgery Division, National Cancer Center Hospital, Tokyo 104-0045, Japan, [5]Department of Surgery, Nagano Matsushiro General Hospital, Nagano 381-1231, Japan, [6]Nikkei Disease Prevention Center, São Paulo 04122-000, Brazil, [7]Clinical Support Laboratory and [8]Clinical Laboratory Division, National Cancer Center Hospital, Tokyo 104-0045, Japan, [9]Genetics Division, National Cancer Center Research Institute, Tokyo 104-0045, Japan, [10]Cancer Information Services and Surveillance Division, Center for Cancer Control and Information Services, National Cancer Center, Tokyo 104-0045, Japan and [11]Pathology Division, National Cancer Center Research Institute, Tokyo 104-0045, Japan

[12]Present address: Department of Respiratory Medicine, Mitsui Memorial Hospital, Kanda-Izumicho 1, Chiyoda-ku, Tokyo 101-8643, Japan

[13]Present address: Department of General Thoracic Surgery, Juntendo University School of Medicine, Hongo 3-1-3, Bunkyo-ku, Tokyo 113-8431, Japan

*To whom correspondence should be addressed. Tel: +81 3 3547 5272; Fax: +81 3 3542 0807;
Email: jyokota@ncc.go.jp

Estrogen has been indicated to play an etiological role in the development of lung adenocarcinoma (ADC), particularly bronchioloalveolar carcinoma (BAC), a type of ADC that develops from a benign adenomatous lesion, atypical adenomatous hyperplasia (AAH). Polymorphisms in the *CYP19A1* gene cause interindividual differences in estrogen levels. Here, 13 *CYP19A1* single-nucleotide polymorphisms (SNPs) were examined for associations with lung AAH risk. AAH is detected as ground-glass opacity (GGO) by computed tomography (CT) examination, and this study consisted of 100 individuals diagnosed with GGO in their lungs among 3088 CT-based cancer screening examinees and 424 without. Minor allele carriers for the rs3764221 SNP showed an elevated risk for GGO [odds ratio (OR) = 1.72, P = 0.017]. Associations of this SNP with risks for lung AAH and BAC in the lungs were next examined using 359 ADC cases whose resected lung lobes were subjected to a histological examination for AAH accompaniment and the presence of BAC components and 330 controls without cancer. The ORs were also increased for lung ADC accompanied by AAH (OR = 1.74, P = 0.029) as well as lung ADC with BAC components (OR = 1.41, P = 0.091). The minor allele was associated with an increased circulating estradiol level (P = 0.079) in a population of 363 postmenopausal women without cancer. These results indicate that *CYP19A1* polymorphisms are involved in the risk for lung AAH and BAC in the lungs by causing differences in estrogen levels.

Abbreviations: AAH, atypical adenomatous hyperplasia; ADC, adenocarcinoma; BAC, bronchioloalveolar carcinoma; CI, confidence interval; CT, computed tomography; ER, estrogen receptor; GGO, ground-glass opacity; HWE, Hardy–Weinberg equilibrium; LD, linkage disequilibrium; OR, odds ratio; SNP, single-nucleotide polymorphism; SQC, squamous cell carcinoma.

Introduction

Adenocarcinoma (ADC) is the commonest histological type of lung cancer, 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 lung cancer, squamous cell carcinoma (SQC) and small-cell lung carcinoma. Thus, effective ways of preventing ADC are being searched for. Recent studies indicate that estrogen plays a role in the growth of lung ADC cells (2,3). Estrogen receptor (ER) β is expressed in bronchiolar epithelial cells (4). ERβ expression was detected in >75% of lung ADC being more frequent than SQC and small-cell lung carcinoma, and the expression was preferentially observed in bronchioloalveolar carcinoma (BAC) (4), a differentiated type of lung ADC developed in the peripheral lung (5). ERβ expression was also detected in atypical adenomatous hyperplasia (AAH) (4), a possible precancerous lesion for BAC (6). Growth of lung ADC cells with ERβ expression was enhanced by estrogen, whereas it was suppressed by antagonizing estrogen (2,4). Therefore, estrogen is probably to play an essential role in the growth of lung ADC cells. In fact, an ER antagonist, fulvestrant, is being examined for its utility in the treatment of lung ADC (7).

Estrogen treatment significantly increased the development of adenoma and ADC in the lungs of ovariectomized female and male mice, therefore, estrogen is a risk factor for the development of lung ADC in mice (8). In a cohort study of 44 667 lifelong never-smoking women in Japan, women of either early age menarche or late age menopause showed significant increase in the risk for lung cancer, and involvement of the use of hormone replacement therapy in the risk for lung cancer of postmenopausal women was also suggested (9). Since ADC comprised >85% of lung cancer cases in this study, estrogen is a candidate risk factor for lung ADC also in the human. However, the involvement of endogenous and exogenous estrogen in the etiology of lung cancer of women has been inconsistent in other populations (10–18). In addition, the significance of estrogen on lung cancer risk of men has not been reported to our knowledge, although men have similar levels of circulating estrogen to postmenopausal women (19) and ERβ expression was detected in lung ADC both of men and women (4,20). Therefore, estrogen is a possible target for prevention of lung ADC, and the significance of estrogen on its etiology should be further investigated.

Polymorphisms in genes involved in estrogen metabolism have been suggested to be associated with circulating estrogen levels (19). Particularly, polymorphisms in the *CYP19A1* gene, encoding an aromatase responsible for the final step in the biosynthesis of estrogens, estradiol (E2) and estrone (E1) (21), have been most intensively investigated (22). A tandem repeat polymorphism, (TTTA)n, in intron 4 and a single-nucleotide polymorphism (SNP), rs10046, in the 3′-untranslated region of exon 10 were reported to be associated with circulating estrogen levels in postmenopausal women (23,24). The tandem repeat polymorphism was also associated with circulating estrogen levels in men (25). Recently, by a large-scale association study, in which >3000 postmenopausal women of European descent were analyzed for 103 SNPs dispersed in the *CYP19A1* gene, SNPs located in the 3′ region (i.e. exons 2–10) of the *CYP19A1* gene, such as rs10046 and four other SNPs (marked by blue lines in Figure 1), were defined as most strongly associated with serum E2 and E1 levels (26). Therefore, it was indicated that polymorphisms in the 3′ region of the *CYP19A1* gene are responsible for interindividual differences in circulating estrogen levels. On the other hand, in a recent association study involving 1068 men from Sweden and 2568 men from the USA, SNPs in the 3′ region of the *CYP19A1* gene, including rs10046, also showed associations with serum E2 and E1 levels in men (19).

© 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 October 13, 2010

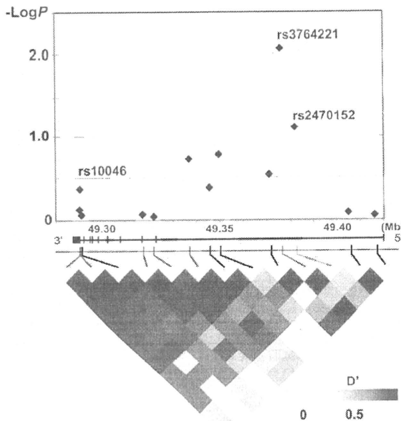

Fig. 1. Association of 13 SNPs in the *CYP19A1* gene with GGO risk. The top panel shows association results by trend test for SNPs and the location of SNPs. Black lozenges depict the results for risks for GGO. The bottom panel shows the LD structure in 424 control subjects. Boxes are shaded according to the pair-wise D' values.

However, the rs2470152 SNP located in intron 1 of the *CYP19A1* gene (marked by a green line in Figure 1) showed a stronger association than the SNPs in the 3′ region (19). Therefore, it was indicated that polymorphisms in intron 1 of the *CYP19A1* gene also affect the estrogen levels. Since the rs2470152 SNP was not examined in the association study above in postmenopausal women (26), polymorphisms responsible for estrogen levels remain unclear. However, these studies strongly indicate that *CYP19A1* polymorphisms are a critical determinant of interindividual differences in the serum estrogen levels both in men and women.

We investigated here the significance of *CYP19A1* SNPs on risks for AAH and ADC by conducting four independent association studies to further obtain information on estrogen in the etiology of lung ADC. (i) AAH in the lungs is detected as a ground-glass opacity (GGO) by helical computed tomography (CT) examination (6,27–30). Therefore, the first study was to examine association of *CYP19A1* SNPs with GGO risk in the lungs among examinees admitted to a single cancer screening center. This study consisted of 100 cancer screening examinees diagnosed with GGOs by a thin-section (high resolution) CT examination among 3088 examinees and 424 examinees without GGO who were matched to the GGO cases in sex and age categories. (ii) AAH is an incidental histologic finding detected in 16–35% of lungs bearing primary lung ADC (6,31). Therefore, the second study was to examine association of *CYP19A1* SNPs with the risk for ADC accompanied with AAH(s) in the lungs among patients admitted to hospital. This study consisted of 81 cases diagnosed with lung ADC accompanied with AAH(s) among 359 lung ADC cases who received lobectomy followed by a histological examination of resected lobes serially sliced at intervals of 5 mm and 330 controls without cancer. (iii) AAH has been considered as a precancerous lesion that particularly develops to BAC, a type of ADC. Therefore, the third study was to examine association with risk for BAC. This study consisted of 151 cases diagnosed with lung ADC containing BAC components among 172 cases diagnosed with small-sized ADC, which include BAC as the majority, and 330 controls without cancer.

Table I. Study subjects

Set	Subject	Category		All	Male (%)	Age (mean ± SD)	Smoker (%)
GGO	Case			100	45 (45)	57 ± 9	37 (37)
	Control			424	194 (46)	57 ± 9	197 (46)
Lung ADC	Case	All		359	193 (54)	59 ± 9	187 (52)
		AAH accompaniment					
			Present	81	39 (48)	60 ± 7	42 (52)
			Absent	278	154 (55)	58 ± 9	145 (52)
		BAC components					
			Present	151	70 (46)	59 ± 8	66 (44)
			Absent	21	11 (52)	61 ± 11	8 (38)
	Control			330	186 (56)	62 ± 11	154 (47)

(iv) Finally, *CYP19A1* SNPs that were associated with GGO, AAH and BAC risks were examined for association with circulating estrogen levels of 363 postmenopausal women without cancer.

Subjects and methods

Subjects for association study on GGO risk

Study subjects were Japanese and consisted of examinees who underwent helical CT examination of the lungs from 2005–07 as a cancer screening program provided by the Research Center for Cancer Prevention and Screening of the National Cancer Center, Japan. Details of the screening program have been described elsewhere (32). All examinees gave written informed consent to allow their data and materials collected through the screening program to be used for the purpose of medical research. The study protocol was approved by the institutional review board of the National Cancer Center, Tokyo, Japan. Eligible examinees were individuals who underwent helical CT examination of the lungs. Details of the CT screening method were described previously (33). Examinees diagnosed with lung cancer or with a history of malignancies were considered ineligible. In a consecutive series of 3088 examinees aged from 40 to 79, 2822 fulfilled the necessary conditions above. One hundred and five examinees were defined as GGO cases because they had at least one GGO ≥5 mm in diameter by a screening CT examination followed by validation by a thin-section (high resolution) CT examination. Four hundred and forty examinees were chosen as control subjects from examinees without GGO by frequency matching to these GGO cases in sex and four age categories (ages 40–49, 50–59, 60–69 and 70–79 years). Genomic DNAs were available for 100 cases and 424 controls of these subjects for this study (the GGO set, Table I). One hundred and two examinees were chosen by a simple random sampling method from 512 examinees diagnosed as having at least one GGO <5 mm in diameter by screening and/or high-resolution CT examinations and were examined as a population containing GGO cases as a subset.

Before undergoing the screening program, examinees completed a self-administered questionnaire concerning medical history and lifestyle characteristics, including smoking habit. The composition of the questionnaire has been detailed elsewhere (32,34). The questionnaire inquired about smoking habits by first determining smoking status (current, former and never) and then expressing lifetime exposure to cigarette smoking among current- and former-smokers by pack-years, with one pack-year defined as the smoking of 20 cigarettes every day for 1 year. Both current- and former-smokers were expressed as smokers in this study.

Subjects for association study on lung ADC risk

All 359 cases and 330 controls were Japanese and were admissions to the National Cancer Center Hospital from 1999 to 2004. Cases were admissions who were diagnosed with lung ADC by histological examinations according to World Health Organization classification (5) and received lobectomies at National Cancer Center Hospital. Controls were admissions who were not diagnosed with lung cancers and had no history of cancers (the lung ADC set, Table I). They were individuals who had been suspected to carry lung or gastric cancer in other hospitals and were not diagnosed with these cancers in National Cancer Center Hospital by CT, endoscope examinations, etc. All cases and controls, from whom informed consent as well as blood samples were obtained, were consecutively included in this study without any exclusion criteria. The participation rate was nearly 80%. From each individual, a 20 ml whole-blood sample was obtained.

All 359 ADC cases were subjected to pathological search for AAH in the resected lobes as described (35). Briefly, resected lungs were inflated with 10%

Downloaded from carcin.oxfordjournals.org at National Cancer Centre (JMLA) on October 13, 2010

formalin through bronchial cut ends, and after fixation for a few days were serially sliced at intervals of 5 mm, and each cut surface was macroscopically examined. Sliced lungs containing a lesion(s) suspected for AAH were further examined microscopically. Even in cases without macroscopic lesions, at least one tissue block was prepared from all sliced lungs and subjected to microscopic examination. The criteria for AAH were as follows and as described previously (36,37): (i) a localized lesion with well-defined boundaries; (ii) an alveolar wall slightly thickened with mild infiltration of inflammatory cells but without scar formation; (iii) proliferating atypical epithelial cells abutting each other but not as compact as in ADC; (iv) atypical epithelial cells that were cuboidal to low columnar or peg-shaped in appearance, resembling either type II pneumocytes or non-ciliated bronchiolar epithelial cells (Clara cells) and (v) the presence of some atypical cells with two or more nuclei, most of which had relatively smaller and smoother contours than those of ADC. These criteria are compatible with those described in the reference of World Health Organization classification of lung tumors as a proposal (6). In the lobes, AAH lesions were detected in 81 cases (23%), whereas no AAH lesion was detected in the remaining 278 cases (77%) (Table I). The 359 ADC cases included 172 cases of small-sized ADC (i.e. <2 cm in maximum diameter), and the information on the presence of BAC components in the tumor was available (Table I). One hundred and fifty-one cases (87%) contained BAC components in the tumor, whereas the remaining 21 cases (13%) did not.

The study protocol was approved by the institutional review board of the National Cancer Center, Tokyo, Japan. Smoking histories of the case and control subjects were obtained via interview using a questionnaire. The definitions of never-smokers and smokers are described above.

Subjects for association with estrogen levels

Postmenopausal women who participated as controls in multicenter hospital-based case–control studies of breast cancer (38–40) were analyzed in the present study. This study was designed to determine lifestyle factors and genetic susceptibility to the risk for breast cancer and to compare potential risk factors among Japanese living in Nagano, Japan and Japanese Brazilians and non-Japanese Brazilians living in São Paulo, Brazil. Written informed consent was obtained from all these subjects. This study was approved by Comissão Nacional de Ética em Pesquisa (CONEP, National Committee of Ethics in Research), Brasília, Brazil and by the institutional review board of the National Cancer Center, Tokyo, Japan.

Estrogen (E2 and E1) levels in serum for Nagano and in plasma for São Paulo were determined by radioimmunoassay by Mitsubishi Kagaku Bio-Clinical Laboratories (Tokyo, Japan). Both the hormone levels and genomic DNA from peripheral blood cells of 185 Japanese, 44 Japanese Brazilians and 134 non-Japanese Brazilians were available for the present study.

SNP analysis

Genomic DNA was extracted from whole-blood cells using a Blood Maxi Kit (QIAGEN, Tokyo, Japan) according to the supplier's instructions. Thirteen SNPs located in the *CYP19A1* gene were selected. Five SNPs, rs4646, rs10046, rs2414096, rs727479 and rs1008805, were chosen since significant associations with serum estrogen levels of postmenopausal women were reported (26). rs2470152 was chosen since a significant association with serum estrogen levels of men was reported (19). The other seven SNPs were chosen based on the fact that their minor allele frequencies in the Japanese population were >0.1 in the GEMDBJ SNP database (https://gemdbj.nibio.go.jp/dgdb/). Genotyping of GGO set subjects for six SNPs, rs4646, rs10046, rs2414096, rs727479, rs1008805 and rs3764221 was performed by the Goldengate assay (Illumina, San Diego, CA) and that for the remaining seven SNPs was performed by the Taqman assay (Applied Biosystems, Foster City, CA) according to the supplier's instructions. Genotyping of lung ADC set subjects for the rs3764221 SNP and genotyping of the subjects for association of the rs3764221 and rs10046 SNPs with serum estrogen levels was performed by the Taqman assay.

Statistical analyses

A Hardy–Weinberg equilibrium (HWE) test was performed using the SNPAlyze version 3 software (DYNACOM, Chiba, Japan), and SNPs with a P value for deviation >0.05 were considered to be in HWE. Calculation of the D' and R^2 values between SNPs was performed by the expectation–maximization algorithm using the SNPAlyze version 3 software.

Associations of 13 SNPs in the *CYP19A1* gene with GGO risk were examined by a trend test adjusted for gender, age (≤49, 50–59, 60–69 and ≥70) and smoking (never-smoker versus smoker). Associations of the rs3764221 SNP with GGO and ADC risks were digitized as odds ratios (ORs) adjusted for gender, age (-49, 50–59, 60-69 and 70-) and smoking (never-smoker versus smoker) with 95% confidence intervals (CIs) by unconditional logistic regression analysis (41). ORs for ADC risk according to the accompaniment of AAH were assessed by the multinomial logistic regression model. These analyses were performed using the JMP version 6.0 software (SAS Institute, Cary, NC). Linear trends for estrogen levels according to increases in the number of minor alleles for the rs3764221 and rs10046 SNPs were tested in a multivariate regression model using SAS software version 9.1 (SAS Institute). Variables used for adjustment in each test are described in the footnotes to Tables II and III. A level of $P < 0.05$ in a test was judged as significant and that of $0.05 \leq P < 0.1$ was judged as marginal.

Results

Association of a CYP19 SNP with lung GGO risk

Thirteen SNPs dispersed in the *CYP19A1* gene region were examined for association with GGO risk in a case–control study that consisted of 100 examinees with GGO and 424 without (GGO set in Table I). All 13 SNPs were in HWE both in cases and controls. Significant association with GGO risk was observed for an SNP, rs3764221, located in intron 1 of the *CYP19A1* gene (P by trend test = 0.0085) (Figure 1; supplementary Table I is available at *Carcinogenesis* Online).

Five SNPs associated with estrogen levels in postmenopausal women of European descent (indicated by blue lines in Figure 1) were in strong linkage disequilibrium (LD) with each other (D' = 0.85–1.0) as reported (26). These five SNPs also showed LD with rs3764221 (D' = 0.75–0.92), however, none of them showed significant associations with GGO risk (supplementary Table I is available at *Carcinogenesis* Online). The rs2470152 SNP associated with estrogen levels in men from Sweden and the USA (indicated by a green line in Figure 1) were in a complete LD (D' = 1.0) with rs3764221, and this SNP showed a marginal association (P = 0.076) with GGO risk (supplementary Table I is available at *Carcinogenesis* Online).

Heterozygotes and homozygotes for the minor allele of the rs3764221 SNP showed increased ORs for the GGO risk (Table II), and the increase in the homozygotes was statistically significant. The OR in the dominant mode (C/T + T/T versus C/C) also

Table II. Association of CYP19A1 (rs3764221) genotypes with lung ADC risk

Category	Genotype	Control, N (%)	Case, N (%)	OR[a] (95% CI)	P
GGO	C/C	262 (62)	47 (47)	Reference	
	C/T	138 (33)	42 (42)	1.59 (0.99–2.56)	0.057
	T/T	24 (6)	11 (11)	2.47 (1.09–5.28)	0.030
	Dominant			1.72 (1.10–2.70)	0.017
	Recessive			2.03 (0.92–4.23)	0.077
ADC	C/C	187 (57)	184 (51)	Reference	
	C/T	123 (37)	145 (40)	1.21 (0.88–1.67)	0.24
	T/T	20 (6)	30 (8)	1.47 (0.80–2.77)	0.22
	Dominant			1.25 (0.92–1.70)	0.16
	Recessive			1.37 (0.76–2.53)	0.30
AAH accompaniment[b]					
Present	C/C		35 (43)	Reference	
	C/T		38 (47)	1.69 (1.01–2.85)	0.047
	T/T		8 (10)	2.05 (0.79–4.93)	0.12
	Dominant			1.74 (1.06–2.86)	0.029
	Recessive			1.66 (0.66–3.82)	0.26
Absent	C/C		149 (54)	Reference	
	C/T		107 (38)	1.10 (0.78–1.55)	0.60
	T/T		22 (8)	1.33 (0.68–2.60)	0.40
	Dominant			1.13 (0.81–1.57)	0.46
	Recessive			1.29 (0.68–2.47)	0.44
BAC components					
Present	C/C		72 (48)	Reference	
	C/T		64 (42)	1.34 (0.88–2.03)	0.17
	T/T		15 (10)	1.87 (0.88–3.90)	0.10
	Dominant			1.41 (0.95–2.09)	0.091
	Recessive			1.65 (0.80–3.35)	0.17

[a]Adjusted for age, sex and smoking.
[b]ORs according to the accompaniment of AAH were assessed by the multinomial logistic regression model.

Downloaded from carcin.oxfordjournals.org at National Cancer Centre (JMLA) on October 13, 2010

showed a statistically significant increase [OR = 1.72 (1.10–2.70) P = 0.017] (Table II; supplementary Figure 1 is available at *Carcinogenesis* Online). The OR in the dominant mode was also calculated against 102 examinees with GGO < 5 mm in diameter by screening and/or high-resolution CT examinations. An increase in OR in the dominant mode was also observed [OR = 1.42 (0.90–2.23)]; however, the increase did not reach a statistical significance (P = 0.13).

Association of a CYP19 SNP with lung ADC risk

Association of the rs3764221 SNP with lung ADC risk was examined in a case–control study consisting of 359 lung ADC cases and 330 controls (Lung ADC set in Table I). This SNP was in HWE both in cases and controls. ORs of heterozygotes and homozygotes for the minor allele and those in both the dominant and recessive modes for the lung ADC risk were increased; however, the increases were not statistically significant (Table II; supplementary Figure 1 is available at *Carcinogenesis* Online).

All 359 lung ADC cases were informative for the presence of AAH in the lung lobe with primary ADC (Table I). Eighty-one (23%) cases had AAHs with primary ADC, consistent with previous reports that AAHs were detected in 16–35% of lungs with primary ADC (6,31). The ORs of heterozygotes and homozygotes for the minor allele and those in the dominant and recessive modes were higher for the risk for ADC with AAH than for ADC without AAH, although their 95% CIs overlapped (Table II; supplementary Figure 1 is available at *Carcinogenesis* Online). ORs of hetrozygotes and in the dominant mode for the risk for ADC with AAH were statistically significant.

Among the 359 cases, 172 cases had small-sized ADC (i.e. <2cm in maximum diameter) and were informative whether their tumors contained BAC components or not (Table I). Tumors of 151 cases were diagnosed as containing BAC components. The ORs of heterozygotes and homozygotes for the minor allele and those in the dominant and recessive modes were higher for ADC with BAC components than for overall ADC, although their 95% CIs overlapped (Table II; supplementary Figure 1 is available at *Carcinogenesis* Online). ORs in the dominant mode for the risk for ADC with BAC components were marginally significant. The number of ADC cases without BAC components was small; therefore, ORs for ADC without BAC components were not calculated.

Association of the rs3764221 SNP with estrogen level

Association of the rs3764221 SNP with GGO and ADC risks prompted us to examine whether this SNP is associated with estrogen levels or not. For this purpose, we examined the allele distribution of this SNP in 363 postmenopausal women, consisting of 185 Japanese, 44 Japanese Brazilians and 134 non-Japanese Brazilians, whose information on circulating E2 and E1 levels was available (38–40). We also examined the allele distribution of the rs10046 SNP because the E2 and E1 levels in heterozygotes and homozygotes for the minor allele had been shown previously to be significantly higher than those in major allele homozygotes (Caucasian in Table III) (26). Heterozygotes and homozygotes for the minor allele for the rs3764221 SNP in all subjects showed higher E2 and E1 levels as for rs10046 in the previous report (Table III) (26). The increase in the E2 level according to increases in the number of minor alleles in all subjects was marginally significant (P = 0.078), whereas that in the E1 level was not significant. Heterozygotes and homozygotes for Japanese subjects also showed higher E2 and E1 levels, although the differences were not statistically significant. On the other hand, heterozygotes and homozygotes for the minor allele for the rs10046 SNP showed only slightly increased levels of E1 and E2 in this study population.

Discussion

In this study, the rs3764221 SNP in the *CYP19A1* gene was shown to be associated with risk for GGO (Table II). AAHs are usually detected as GGOs by CT examinations and a subset of these AAHs progress to ADC, including BAC (28,30,42). Therefore, this SNP was suggested to be involved in the risk for the development of AAH and also of lung ADC, particularly of BAC in the lungs. This suggestion was supported by the following two findings. First, the rs3764221 SNP showed a significant association with the risk for ADC accompanied by AAH but not for ADC not accompanied by AAH (Table II). Second, this SNP showed a marginal association with the risk for ADC containing BAC components, and the association in this subset of ADC was more evident than that in overall lung ADC (Table II). This result is consistent with the concept that AAH is a precancerous lesion of ADC, preferentially of BAC (5,6,43). The frequency of having AAH in the lungs has been shown to be considerably higher in ADC patients than in individuals without cancer (6,31,36,37,44). Therefore, the susceptibility to the development of AAH is probably to be associated with that of ADC in the lungs. Thus, the rs3764221 SNP might confer lung ADC risk by affecting the susceptibility to the development of AAH that progress to ADC, preferentially BAC.

In the present study, the minor allele for the rs3764221 SNP was marginally associated with a higher estrogen level in postmenopausal women. Notably, rs3764221 was in complete LD (D' = 1) with rs2470152, whose association with serum estrogen levels in men had been reported (19). Accordingly, the rs2470152 SNP also showed a marginally significant association with risk for GGO (Figure 1).

Table III. Association of CYP19A1 SNPs with circulating estrogen levels SNP

	Population	Genotype	No. of subjects	Increase in estradiol (E2)	P for trend	Increase in estrone (E1)	P for trend
rs3764221	All	CC	220	Ref	0.078[a]	Ref	0.26[a]
		CT	120	+4.8%		+1.0%	
		TT	17	+16.0%		+13.4%	
	Japanese	CC	86	Ref	0.11[a]	Ref	0.30[a]
		CT	86	+6.6%		+1.2%	
		TT	12	+17.1%		+17.0%	
rs10046	All	GG	116	Ref	0.92[a]	Ref	0.36[a]
		GA	193	+0.04%		+1.1%	
		AA	54	+0.69%		+5.3%	
	Japanese	GG	61	Ref	0.83[a]	Ref	0.43[a]
		GA	93	+0.8%		+2.5%	
		AA	31	+2.3%		+6.1%	
	Caucasian[b]	GG	835	Ref	2.9×10^{-9}	Ref	1.1×10^{-8}
		GA	1691	+5.7%		+5.4%	
		AA	799	+12.8%		+11.7%	

[a]Adjusted for age, ethnic group, age at menarche, age at menopause, number of births, age at first birth, height, body mass index, smoking, alcohol drinking and physical activity in the past 5 years.
[b]Data from Haimann *et al.* (26).

Downloaded from carcin.oxfordjournals.org at National Cancer Centre (JMLA) on October 13, 2010

Interestingly, intron 1 of the *CYP19A1* gene contains 10 tissue specific promoters, which have been indicated to play regulatory roles in *CYP19A1* gene expression differentially among diverse tissues (21,22). rs2470152 and rs3764221 SNPs are located, respectively, in and 3′ to the I.4 promoter, which enables *CYP19A1* expression in skin, testis and adipose tissues (21,45,46). Therefore, genetic variations in the region spanning these two SNPs might be responsible for differential *CYP19A1* expression among individuals, and this might cause interindividual differences in estrogen levels. In contrast to previous reports (26), the rs10046 SNP did not show association with estrogen levels in the present study. Such an inconsistency might have come from ethnic differences of subjects examined. Since the minor allele frequency for the3764221 SNP is considerably lower in Europeans (<0.05) than in Asians (>0.2) (http://www.ncbi.nlm.nih.gov/projects/SNP/), this SNP was not examined in previous association studies of Europeans (19,26). The rs3764221 SNP is in LD with SNPs located in the 3′ region of the *CYP19A1* gene, including rs10046, therefore, it is also possible that SNP(s) critical for estrogen levels is located in this 3′ region.

Interaction of *CYP19A1* genotypes with smoking and gender was also investigated. The ORs for the risks for GGO and lung ADC were consistently higher in never-smokers than in smokers, although their 95% CIs overlapped (supplementary Table II is available at *Carcinogenesis* Online). This result went along with the result of meta-analysis showing that hormone replacement therapy particularly increases lung ADC risk of never-smokers (10). This stronger association of *CYP19A1* genotypes with GGO and lung ADC risks in never-smokers than smokers might be due to the anti-estrogenic effect of smoking (47,48). Smoking has been indicated to be associated with low levels of estrogen and with decreased risks for estrogen-dependent cancers, such as endometrial cancers (49–51). On the other hand, risks for GGO and lung ADC were not consistently associated with gender (supplementary Table III is available at *Carcinogenesis* Online); therefore, the interaction of *CYP19A1* genotypes with gender remains unclear.

The present study proposes that *CYP19A1* polymorphisms are involved in the risk for AAH and BAC in the lungs by causing differences in estrogen levels. Association studies of a single population among *CYP19A1* genotypes, estrogen levels and the risk for AAH and BAC, by taking gender and smoking into account, will further authenticate the present results. The contribution of *CYP19A1* polymorphisms to cancer risks has been investigated in estrogen-dependent cancers, such as ADCs of breast and endometrium. The contribution has been indicated to be possible but remains inconclusive due to inconsistent results among studies (22,26,52). Studies of *CYP19A1* polymorphisms on risks for ADCs of a variety of organs, including the lungs, breast and endometrium, will further elucidate the significance of these polymorphisms and estrogen levels on cancer risks.

Supplementary material

Supplementary Figure 1 and Tables I–III can be found at http://carcin.oxfordjournals.org/

Funding

Grants-in-Aid from the Ministry of Health, Labor and Welfare for the Third-term Comprehensive 10 year Strategy for Cancer Control, Cancer Research (19S-1), Research on Risk of Chemical Substances; Ministry of Education, Culture, Sports, Science and Technology for Scientific Research on Priority Areas (20014031 and 17015049); Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation.

Acknowledgements

We thank Chisato Hamashima, Shizuka Sasazuki, Norie Kurahashi, Masahiro Kaneko, Masahiko Kusumoto, Takaaki Tsuchida and Hiroshi Saito for their help in data and sample collection for the GGO set; Daizo Saito and Sachiyo Mimaki for their help in data and sample collection for the lung ADC set. K.S. is an awardee of a Research Resident Fellowship from the Foundation for Promotion of Cancer Research in Japan. We are also grateful to the participants in the São Paulo-Japan Breast Cancer Study and the Nagano Breast Cancer Study.

Conflict of Interest Statement: None declared.

References

1. Travis,W.D. *et al.* (2004) Pathology and genetics, tumours of lung, pleura, thymus and heart. In Travis,W.D., Brambilla,E., Muller-Hermelink,K. and Harris,C.C. (eds) *World Health Organization Classification of Tumors.* IARC Press, Lyon, France, pp. 1–344.
2. Marquez-Garban,D.C. *et al.* (2009) Targeting aromatase and estrogen signaling in human non-small cell lung cancer. *Ann. N. Y. Acad. Sci.*, **1155**, 194–205.
3. Siegfried,J.M. *et al.* (2009) Estrogen receptor signaling in lung cancer. *Semin. Oncol.*, **36**, 524–531.
4. Omoto,Y. *et al.* (2001) Expression, function, and clinical implications of the estrogen receptor beta in human lung cancers. *Biochem. Biophys. Res. Commun.*, **285**, 340–347.
5. Colvy,T. *et al.* (2004) Adenocarcinoma. In Travis,W.D., Brambilla,E., Muller-Hermelink,H.K. and Harris,C.C. (eds) *World Health Organization Classification of Tumors: Pathology and Genetics, Tumours of Lung, Pleura, Thymus and Heart.* pp. 35–44.
6. Kerr,K.M. *et al.* (2004) Atypical adenomatous hyperplasia. In Travis,W.D., Brambilla,E., Muller-Hermelink,H.K. and Harris,C.C. (eds) *World Health Organization Classification of Tumors: Pathology and Genetics, Tumours of Lung, Pleura, Thymus and Heart.* pp. 73–75.
7. Traynor,A.M. *et al.* (2009) Pilot study of gefitinib and fulvestrant in the treatment of post-menopausal women with advanced non-small cell lung cancer. *Lung Cancer*, **64**, 51–59.
8. Hammoud,Z. *et al.* (2008) Estrogen promotes tumor progression in a genetically defined mouse model of lung adenocarcinoma. *Endocr. Relat. Cancer*, **15**, 475–483.
9. Liu,Y. *et al.* (2005) Reproductive factors, hormone use and the risk of lung cancer among middle-aged never-smoking Japanese women: a large-scale population-based cohort study. *Int. J. Cancer*, **117**, 662–666.
10. Greiser,C.M. *et al.* (2010) Menopausal hormone therapy and risk of lung cancer-Systematic review and meta-analysis. *Maturitas*, **65**, 198–204.
11. Slatore,C.G. *et al.* (2010) Lung cancer and hormone replacement therapy: association in the vitamins and lifestyle study. *J. Clin. Oncol.*, **28**, 1540–1546.
12. Chlebowski,R.T. *et al.* (2009) Oestrogen plus progestin and lung cancer in postmenopausal women (Women's Health Initiative trial): a post-hoc analysis of a randomised controlled trial. *Lancet*, **374**, 1243–1251.
13. Wu,A.H. *et al.* (1988) Personal and family history of lung disease as risk factors for adenocarcinoma of the lung. *Cancer Res.*, **48**, 7279–7284.
14. Wu-Williams,A.H. *et al.* (1990) Lung cancer among women in north-east China. *Br. J. Cancer*, **62**, 982–987.
15. Taioli,E. *et al.* (1994) Re: endocrine factors and adenocarcinoma of the lung in women. *J. Natl Cancer Inst.*, **86**, 869–870.
16. Seow,A. *et al.* (2002) Diet, reproductive factors and lung cancer risk among Chinese women in Singapore: evidence for a protective effect of soy in nonsmokers. *Int. J. Cancer*, **97**, 365–371.
17. Brenner,A.V. *et al.* (2003) Menstrual and reproductive factors and risk of lung cancer among Chinese women, Eastern Gansu Province, 1994–1998. *J. Epidemiol.*, **13**, 22–28.
18. Kreuzer,M. *et al.* (2003) Hormonal factors and risk of lung cancer among women? *Int. J. Epidemiol.*, **32**, 263–271.
19. Eriksson,A.L. *et al.* (2009) Genetic variations in sex steroid-related genes as predictors of serum estrogen levels in men. *J. Clin. Endocrinol. Metab.*, **94**, 1033–1041.
20. Wu,C.T. *et al.* (2005) The significance of estrogen receptor beta in 301 surgically treated non-small cell lung cancers. *J. Thorac. Cardiovasc Surg.*, **130**, 979–986.
21. Chen,D. *et al.* (2009) Regulation of breast cancer-associated aromatase promoters. *Cancer Lett.*, **273**, 15–27.
22. Ma,C.X. *et al.* (2005) Human aromatase: gene resequencing and functional genomics. *Cancer Res.*, **65**, 11071–11082.
23. Tworoger,S.S. *et al.* (2004) Association of CYP17, CYP19, CYP1B1, and COMT polymorphisms with serum and urinary sex hormone

Downloaded from carcin.oxfordjournals.org at National Cancer Centre (JMLA) on October 13, 2010