

<sup>a</sup>Frequency in controls.

**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<sup>+</sup> 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

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, Ikuro 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

- 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.
- Sun,S. *et al.* (2007) Lung cancer in never smokers—a different disease. *Nat. Rev. Cancer*, **7**, 778–790.
- 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.
- Wang,Y. *et al.* (2008) Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat. Genet.*, **40**, 1407–1409.
- McKay,J.D. *et al.* (2008) Lung cancer susceptibility locus at 5p15.33. *Nat. Genet.*, **40**, 1404–1406.
- 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.
- Kohno,T. *et al.* (2008) Association of KRAS polymorphisms with risk for lung adenocarcinoma accompanied by atypical adenomatous hyperplasias. *Carcinogenesis*, **29**, 957–963.
- 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.
- 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.
- 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.
- 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.
- 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.
- Sakamoto,H. *et al.* (2008) Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat. Genet.*, **40**, 730–740.
- Tamiya,G. *et al.* (2005) Whole genome association study of rheumatoid arthritis using 27 039 microsatellites. *Hum. Mol. Genet.*, **14**, 2305–2321.
- Sunaga,N. *et al.* (2002) Contribution of the NQO1 and GSTT1 polymorphisms to lung adenocarcinoma susceptibility. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 730–738.
- 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.
- Kohno,T. *et al.* (2006) Association of the OGG1-Ser326Cys polymorphism with lung adenocarcinoma risk. *Cancer Sci.*, **97**, 724–728.
- Kohno,T. *et al.* (2006) Association of polymorphisms in the MTH1 gene with small cell lung carcinoma risk. *Carcinogenesis*, **27**, 2448–2454.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Knutson,K.L. *et al.* (2005) Tumor antigen-specific T helper cells in cancer immunity and immunotherapy. *Cancer Immunol. Immunother.*, **54**, 721–728.
- 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.
- 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.
- 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



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

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

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

<sup>12</sup>Present address: Department of Respiratory Medicine, Mitsui Memorial Hospital, Kanda-Izumicho 1, Chiyoda-ku, Tokyo 101-8643, Japan

<sup>13</sup>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.

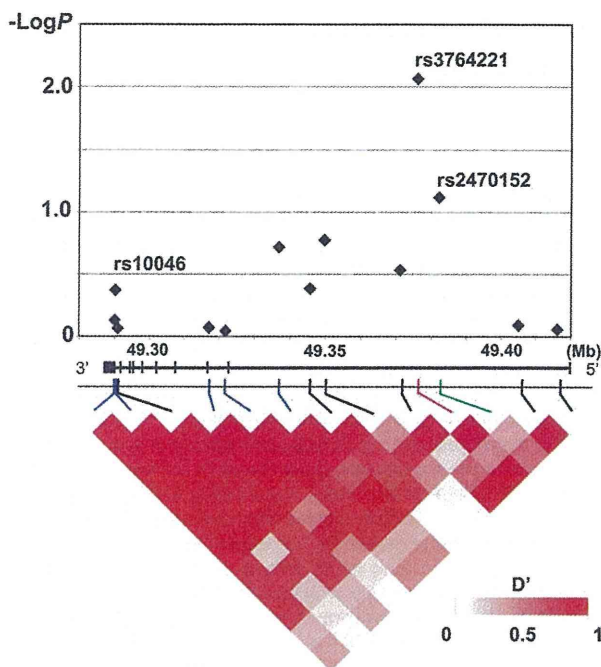
### Introduction

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)  $\beta$  is expressed in bronchiolar epithelial cells (4). ER $\beta$  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). ER $\beta$  expression was also detected in atypical adenomatous hyperplasia (AAH) (4), a possible precancerous lesion for BAC (6). Growth of lung ADC cells with ER $\beta$  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 ER $\beta$  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, (TTA) $_n$ , 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).





**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 $\pm$ SD)	Smoker (%)	
GGO	Case	100	45 (45)	57 $\pm$ 9	37 (37)	
	Control	424	194 (46)	57 $\pm$ 9	197 (46)	
Lung ADC	Case	359	193 (54)	59 $\pm$ 9	187 (52)	
	AAH accompaniment					
	Present	81	39 (48)	60 $\pm$ 7	42 (52)	
	Absent	278	154 (55)	58 $\pm$ 9	145 (52)	
	BAC components					
	Present	151	70 (46)	59 $\pm$ 8	66 (44)	
Absent	21	11 (52)	61 $\pm$ 11	8 (38)		
Control		330	186 (56)	62 $\pm$ 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  $\geq$ 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%



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*<sup>2</sup> 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 ≤ *P* < 0.1 was judged as marginal.

## Results

### Association of a *CYP19A1* 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, <i>N</i> (%)	Case, <i>N</i> (%)	OR <sup>a</sup> (95% CI)	<i>P</i>	
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 <sup>b</sup>	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	

<sup>a</sup>Adjusted for age, sex and smoking.

<sup>b</sup>ORs according to the accompaniment of AAH were assessed by the multinomial logistic regression model.



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 heterozygotes 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 <sup>a</sup>	Ref	0.26 <sup>a</sup>
		CT	120	+4.8%		+1.0%	
		TT	17	+16.0%		+13.4%	
	Japanese	CC	86	Ref	0.11 <sup>a</sup>	Ref	0.30 <sup>a</sup>
		CT	86	+6.6%		+1.2%	
		TT	12	+17.1%		+17.0%	
rs10046	All	GG	116	Ref	0.92 <sup>a</sup>	Ref	0.36 <sup>a</sup>
		GA	193	+0.04%		+1.1%	
		AA	54	+0.69%		+5.3%	
	Japanese	GG	61	Ref	0.83 <sup>a</sup>	Ref	0.43 <sup>a</sup>
		GA	93	+0.8%		+2.5%	
		AA	31	–2.3%		+6.1%	
	Caucasian <sup>b</sup>	GG	835	Ref	$2.9 \times 10^{-9}$	Ref	$1.1 \times 10^{-8}$
		GA	1691	+5.7%		+5.4%	
		AA	799	+12.8%		+11.7%	

<sup>a</sup>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.

<sup>b</sup>Data from Haimann *et al.* (26).



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 the rs3764221 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

- 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.
- 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.
- Siegfried, J.M. *et al.* (2009) Estrogen receptor signaling in lung cancer. *Semin. Oncol.*, **36**, 524–531.
- 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.
- 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.
- 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.
- 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.
- Hammoud, Z. *et al.* (2008) Estrogen promotes tumor progression in a genetically defined mouse model of lung adenocarcinoma. *Endocr. Relat. Cancer*, **15**, 475–483.
- 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.
- Greiser, C.M. *et al.* (2010) Menopausal hormone therapy and risk of lung cancer—Systematic review and meta-analysis. *Maturitas*, **65**, 198–204.
- 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.
- 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.
- 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.
- Wu-Williams, A.H. *et al.* (1990) Lung cancer among women in north-east China. *Br. J. Cancer*, **62**, 982–987.
- Taioli, E. *et al.* (1994) Re: endocrine factors and adenocarcinoma of the lung in women. *J. Natl Cancer Inst.*, **86**, 869–870.
- 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.
- 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.
- Kreuzer, M. *et al.* (2003) Hormonal factors and risk of lung cancer among women? *Int. J. Epidemiol.*, **32**, 263–271.
- 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.
- 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.
- Chen, D. *et al.* (2009) Regulation of breast cancer-associated aromatase promoters. *Cancer Lett.*, **273**, 15–27.
- Ma, C.X. *et al.* (2005) Human aromatase: gene resequencing and functional genomics. *Cancer Res.*, **65**, 11071–11082.
- Twoogor, S.S. *et al.* (2004) Association of CYP17, CYP19, CYP1B1, and COMT polymorphisms with serum and urinary sex hormone



- concentrations in postmenopausal women. *Cancer Epidemiol. Biomarkers Prev.*, **13**, 94–101.
24. Bugano, D.D. *et al.* (2008) Genetic polymorphisms, the metabolism of estrogens and breast cancer: a review. *Eur. J. Gynaecol. Oncol.*, **29**, 313–320.
  25. Gennari, L. *et al.* (2004) A polymorphic CYP19 TTTA repeat influences aromatase activity and estrogen levels in elderly men: effects on bone metabolism. *J. Clin. Endocrinol. Metab.*, **89**, 2803–2810.
  26. Haiman, C.A. *et al.* (2007) Genetic variation at the CYP19A1 locus predicts circulating estrogen levels but not breast cancer risk in postmenopausal women. *Cancer Res.*, **67**, 1893–1897.
  27. Nakata, M. *et al.* (2003) Prospective study of thoracoscopic limited resection for ground-glass opacity selected by computed tomography. *Ann. Thorac. Surg.*, **75**, 1601–1605; discussion, 1605–1606.
  28. Kakinuma, R. *et al.* (2004) Progression of focal pure ground-glass opacity detected by low-dose helical computed tomography screening for lung cancer. *J. Comput. Assist. Tomogr.*, **28**, 17–23.
  29. Park, C.M. *et al.* (2006) CT findings of atypical adenomatous hyperplasia in the lung. *Korean J. Radiol.*, **7**, 80–86.
  30. Godoy, M.C. *et al.* (2009) Subsolid pulmonary nodules and the spectrum of peripheral adenocarcinomas of the lung: recommended interim guidelines for assessment and management. *Radiology*, **253**, 606–622.
  31. Koga, T. *et al.* (2002) Lung adenocarcinoma with bronchioloalveolar carcinoma component is frequently associated with foci of high-grade atypical adenomatous hyperplasia. *Am. J. Clin. Pathol.*, **117**, 464–470.
  32. Hamashima, C. *et al.* (2006) Comparison of observed and expected numbers of detected cancers in the research center for cancer prevention and screening program. *Jpn. J. Clin. Oncol.*, **36**, 301–308.
  33. Gomi, S. *et al.* (2008) Low-dose CT screening for lung cancer with automatic exposure control: phantom study. *Radiol. Phys. Technol.*, **1**, 244–250.
  34. Otani, T. *et al.* (2006) Serum triglycerides and colorectal adenoma in a case-control study among cancer screening examinees (Japan). *Cancer Causes Control*, **17**, 1245–1252.
  35. Kohno, T. *et al.* (2008) Association of KRAS polymorphisms with risk for lung adenocarcinoma accompanied by atypical adenomatous hyperplasias. *Carcinogenesis*, **29**, 957–963.
  36. Yokose, T. *et al.* (2001) Atypical adenomatous hyperplasia of the lung in autopsy cases. *Lung Cancer*, **33**, 155–161.
  37. Yokose, T. *et al.* (2000) High prevalence of atypical adenomatous hyperplasia of the lung in autopsy specimens from elderly patients with malignant neoplasms. *Lung Cancer*, **29**, 125–130.
  38. Iwasaki, M. *et al.* (2009) Isoflavone, polymorphisms in estrogen receptor genes and breast cancer risk in case-control studies in Japanese, Japanese Brazilians and non-Japanese Brazilians. *Cancer Sci.*, **100**, 927–933.
  39. Iwasaki, M. *et al.* (2009) Dietary isoflavone intake and breast cancer risk in case-control studies in Japanese, Japanese Brazilians, and non-Japanese Brazilians. *Breast Cancer Res. Treat.*, **116**, 401–411.
  40. Shimada, N. *et al.* (2009) Genetic polymorphisms in estrogen metabolism and breast cancer risk in case-control studies in Japanese, Japanese Brazilians and non-Japanese Brazilians. *J. Hum. Genet.*, **54**, 209–215.
  41. Breslow, N.E. *et al.* (1980) Statistical methods in cancer research. Volume I—The analysis of case-control studies. *IARC Sci. Publ.*, **32**, 5–338.
  42. Takashima, S. *et al.* (2003) CT findings and progression of small peripheral lung neoplasms having a replacement growth pattern. *AJR Am. J. Roentgenol.*, **180**, 817–826.
  43. Travis, W.D. *et al.* (2006) Bronchioloalveolar carcinoma and lung adenocarcinoma: the clinical importance and research relevance of the 2004 World Health Organization pathologic criteria. *J. Thorac. Oncol.*, **1**, S13–S19.
  44. Sterner, D.J. *et al.* (1997) Prevalence of pulmonary atypical alveolar cell hyperplasia in an autopsy population: a study of 100 cases. *Mod. Pathol.*, **10**, 469–473.
  45. Carreau, S. *et al.* (2004) The promoter(s) of the aromatase gene in male testicular cells. *Reprod. Biol.*, **4**, 23–34.
  46. Lambard, S. *et al.* (2005) Aromatase in testis: expression and role in male reproduction. *J. Steroid Biochem. Mol. Biol.*, **95**, 63–69.
  47. Baron, J.A. (1996) Beneficial effects of nicotine and cigarette smoking: the real, the possible and the spurious. *Br. Med. Bull.*, **52**, 58–73.
  48. Tanko, L.B. *et al.* (2004) An update on the antiestrogenic effect of smoking: a literature review with implications for researchers and practitioners. *Menopause*, **11**, 104–109.
  49. Parazzini, F. *et al.* (1995) Smoking and risk of endometrial cancer: results from an Italian case-control study. *Gynecol. Oncol.*, **56**, 195–199.
  50. Bremnes, Y. *et al.* (2007) Different measures of smoking exposure and mammographic density in postmenopausal Norwegian women: a cross-sectional study. *Breast Cancer Res.*, **9**, R73.
  51. Zhou, B. *et al.* (2008) Cigarette smoking and the risk of endometrial cancer: a meta-analysis. *Am. J. Med.*, **121**, 501–508 e503.
  52. Low, Y.L. *et al.* (2010) Multi-variant pathway association analysis reveals the importance of genetic determinants of estrogen metabolism in breast and endometrial cancer susceptibility. *PLoS Genet.*, **6**, e1001012.

Received June 1, 2010; revised July 6, 2010; accepted July 8, 2010



# Genetic Basis for Susceptibility to Lung Cancer: Recent Progress and Future Directions

Jun Yokota, Kouya Shiraishi, and Takashi Kohno

Biology Division, National Cancer Center Research Institute,  
Tsukiji, Chuo-ku, Tokyo, Japan

- I. Introduction: Overview of Studies on Genetic and Environmental Factors Involved in Lung Cancer Susceptibility
- II. Differences in the Process of Lung Cancer Development Between Smokers and Never-Smokers
- III. Candidate Gene Association Studies
- IV. Genome-Wide Association Studies
- V. Assessment of Lung Cancer Risk in Each Individual by Combined Genotypes (Gene–Gene Interactions)
- VI. Smoking-Associated Differences (Gene–Environment Interactions)
- VII. Necessity of Further Association Studies
- VIII. Future Directions
- References

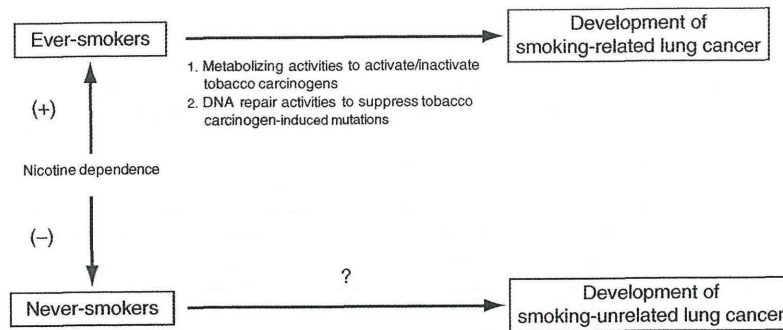
Lung cancer is the leading cause of cancer death worldwide, and cigarette smoking is the major environmental factor for its development. To elucidate the genetic differences in the susceptibility to lung cancer among individuals, genetic factors involved in tobacco-induced lung cancers have been extensively investigated and a number of genetic polymorphisms have been identified to date as candidates. Most of the polymorphisms identified are of genes encoding proteins associated with the activity to metabolize tobacco smoke carcinogens and to suppress mutations induced by those carcinogens, and functional significances have been elucidated for some of these polymorphisms. However, the significance of these polymorphisms in the contribution to lung cancer development still remains unclear. Recently, several novel lung cancer susceptibility genes, including those on chromosomes 5p15.33, 6p21, and 15q24-25.1, have been identified by large-scale genome-wide association (GWA) studies. The 15q25 region contains three nicotine acetylcholine receptor subunit genes, and their polymorphisms have been also reported as being associated with nicotine dependence. The 5p15.33 region is associated with risks specifically for lung adenocarcinoma, the commonest histological type and weakly associated with smoking. This locus has been shown to be associated with risks for a wide variety of cancers, including lung adenocarcinoma. Associations of the 6q21 region have not been consistently replicated among studies. The 6q23-25 and 13q31.3 regions were also identified by recent GWA studies as being associated with risk for lung cancer, particularly in never-smokers. However, contributions of genetic differences on these five loci to the susceptibility to overall lung cancer seem to be small. There are several molecular pathways for the development of lung adenocarcinomas, and environmental factors for their development are still unclear,

especially those in never-smokers. In addition, geographic differences as well as gender differences in lung cancer risk have been indicated. Furthermore, various genes identified by candidate gene association studies have not been reevaluated for their significance together with genes identified by GWA studies in the same population. Therefore, further studies will be necessary to assess the individual susceptibility to lung cancer based on the combination of polymorphisms in multiple genes, and to establish a novel way of evaluating the individual risk for lung cancer for its prevention. © 2010 Elsevier Inc.

## I. INTRODUCTION: OVERVIEW OF STUDIES ON GENETIC AND ENVIRONMENTAL FACTORS INVOLVED IN LUNG CANCER SUSCEPTIBILITY

Lung cancer is the leading cause of cancer death worldwide (Sun *et al.*, 2007). Therefore, identification of genetic factors as well as environmental factors is very important in developing novel methods of lung cancer prevention. Since cigarette smoking is the major environmental risk factor for the development of lung cancer, genetic factors for tobacco-induced lung cancer have been extensively investigated by candidate gene association studies for many years. Genes involved in the metabolism of tobacco smoke carcinogens and genes involved in the repair of genetic alterations induced by those carcinogens have been the major targets of those investigations. In contrast, recent advances in molecular technology and knowledge of the distribution of genetic polymorphisms in the human genome have made it possible to identify genetic factors responsible for the development of common polygenic diseases, including lung cancer, by a genome-wide approach. Indeed, several loci containing candidate lung cancer susceptibility genes have been identified in recent years by genome-wide association (GWA) studies. One of the chromosomal loci identified was 15q24-25.1, and this region contained three genes encoding nicotinic acetylcholine receptor subunit genes. Since this locus has been also suggested to be associated with nicotine dependence, genetic susceptibility for nicotine addiction has come to be the major genetic factor for the development of lung cancer. However, epidemiologically, the incidence of lung cancers has been increasing in never-smokers, in women, and in Asian population, in recent years; therefore, lung cancers in smokers and those in never-smokers are now considered to be different diseases from each other. For this reason, identification of genetic factors as well as environmental factors for the development of lung cancers in never-smokers has also arisen as a major topic for prevention of these lung cancers. Accordingly, GWA studies for the identification of lung cancer susceptibility genes without association with smoking behavior are now also being extensively conducted. Therefore, when we discuss the genetic basis for susceptibility to lung cancer, three different critical points should be considered, as summarized in Fig. 1, in association with their functional significance in the





**Fig. 1** Three different types of lung cancer susceptibility genes involved in the process of lung cancer development. Details are described in Section 1.

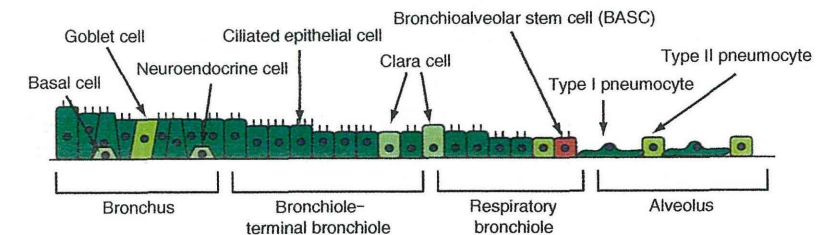
susceptibility. The first point is interindividual differences in nicotine dependence, the second point is susceptibility to tobacco-induced (smoking-related) lung cancer, and the third point is susceptibility to lung cancer in never-smokers (smoking-unrelated). Nicotine dependence should be associated with the quantity of the intake of tobacco smoke carcinogens. In the development of smoking-related lung cancers, interindividual differences in metabolizing activities to activate/inactivate tobacco smoke carcinogens as well as DNA repair activities to suppress tobacco smoke carcinogen-induced mutations would play a major role. Since environmental factors for the development of lung cancer in never-smokers are largely unclear at present, GWA studies will be an effective approach to identify responsible genes for their development. In addition, we should accumulate the knowledge for the difference between lung cancers in smokers and those in never-smokers, from the viewpoint of molecular processes for their development. For this reason, in this review chapter, the differences between lung cancers in smokers and those in never-smokers are summarized first, recent progresses in lung cancer susceptibility gene studies are summarized second, and future directions of this field of science are discussed last.

## II. DIFFERENCES IN THE PROCESS OF LUNG CANCER DEVELOPMENT BETWEEN SMOKERS AND NEVER-SMOKERS

Lung cancers are divided into the two major categories of small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) from clinicopathological aspects (Sun *et al.*, 2007). NSCLCs are further divided into three

major histological types, adenocarcinoma (ADC), squamous cell carcinoma (SQC), and large cell carcinoma (LCC). However, LCC is thought to be poorly or undifferentiated forms of and more heterogeneous than the other three types of lung cancer, and only limited information is available at present on genetic susceptibility to LCC. Therefore, in this review chapter, LCC is not specifically taken up as a subject for discussion. ADC, SQC, and SCLC are thought to be different in their origins (Govindan, 2010; Subramanian and Govindan, 2008; Travis *et al.*, 2004). Lung epithelial cells consist of monolayer columnar glandular epithelial cells (Fig. 2). Basal cells and neuroendocrine cells in the bronchi are thought to be precursors of SQC and SCLC, respectively. Clara cells in the bronchioles and/or type II pneumocytes in the alveoli are thought to be precursors of ADC. In mice, bronchioalveolar stem cells (BASC) have been identified as being a candidate precursor of ADC (Kim *et al.*, 2005); however, corresponding cells in the human have not yet been identified. Therefore, three major types of lung cancer are thought to be originated from different precursor epithelial cells in the lungs.

It is now widely accepted that cancer is attributed to accumulation of multiple genetic alterations in targeted precursor cells. Therefore, when we discuss genetic susceptibility to lung cancer, it is important to understand the differences and similarities in molecular pathways of cancer development among the three major types. Indeed, accumulated genetic alterations during their development are considerably different among them. Table I summarizes the accumulated genetic alterations in ADC, SQC, and SCLC (Govindan, 2010; Subramanian and Govindan, 2008; Sun *et al.*, 2007; Travis *et al.*, 2004). Ten genes identified to date with frequent genetic alterations in lung cancer cells are chosen as representatives. MYC, p53, PTEN, and PTPRD are genetically altered commonly among ADC, SQC,



**Fig. 2** Component epithelial cells in the pulmonary system. Precursor (progenitor) cells of small cell carcinoma, squamous cell carcinoma, and adenocarcinoma are thought to be neuroendocrine cells, basal cells, and Clara cells/type II pneumocytes, respectively. In mice, bronchioalveolar stem cells (BASCs) were identified as being precursor cells of adenocarcinoma; however, corresponding human cells have not identified to date.