Subjects and Methods

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

The NCCH set consisted of 2,343 lung cancer cases and 1,173 controls (**Table I**). The cases were 1,656 ADC, 390 SQC and 297 SCC cases. All ADC, SQC and SCC cases were enrolled in the National Cancer Center Hospitals (NCCH) from 1999 to 2008. All ADC, SQC and SCC cases, from whom informed consent as well as blood samples were obtained, were consecutively included in this study without any particular exclusion criteria. The participation rate was nearly 80%. All the cases were diagnosed by cytological and/or histological examinations according to WHO classification. The controls were 328 inpatients/outpatients of the NCCH; and 645 and 200 volunteers enrolled in Keio and Tokai Universities, respectively. The control NCCH subjects were selected with a criterion of no history of cancer from 1999 to 2007, while the 645 volunteers were the individuals with no known malignancies who offered blood on the occasion of a health check examination at Keio University in 2002 and 2003 [13]. The 200 volunteers in Tokai University were healthy individuals enrolled from 2001 to 2003 as control subjects in a previous case-control study [14].

The NNGH set were 84 ADC and 52 SQC cases and 145 controls who were enrolled in the National Nishi-Gunma Hospital (NNGH) from 1999 to

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

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

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

GWAS. The method of GWAS on microsatellite loci was previously described [14]. Equal amounts of DNAs from 200 lung ADC cases and from 200 controls enrolled in Tokai University were mixed for the first set of case and control DNA pools, respectively. The second set of DNA pools was also

prepared from another 200 ADC cases and 200 controls enrolled in Keio University. Fifty ng of pooled-DNA was amplified by 40 cycles of PCR in 96-well plates using a pair of PCR primers designed for amplifying fragments that include polymorphic microsatellite sequences. Allele frequencies in pooled-DNA were estimated from the height of peaks: the frequency of each allele was determined by dividing the height of each allele by the summed height of all alleles. The significance for difference in allelic distribution was evaluated by Fisher's exact test, with the use of 2×m (where m is the number of alleles).

The first set of case and control DNA pools was examined for differences in allelic distribution for 23,010 microsatellite markers, and the distribution for 1,328 (5.8%) markers were judged as being significantly different by the criteria of $p<0.05$ (1st stage of GWAS in **Table II**). The inflation factor calculated by dividing the mean of the lower 90% of $-\log_{10}(P)$ values by the mean of the lower 90% of the expected values [19] for this screening was 0.639, indicating a deflation in the statistical tests (**Supplementary Table I**). However, in this screening, deduction of allele frequencies was affected by an inevitable experimental bias of the pooled DNA typing, i.e., "shadow bands" in electrophoregrams due to slippages in the PCR reaction particularly for microsatellite markers containing repeat units of 2-bp, as previously reported [20]. In fact, inflation factors for microsatellite markers containing repeat units of 3~6-bp were 0.919~1.022 (0.955 in total), i.e., deviations were within ± 10% as have been observed in previous GWASs in which adequacy of the case-control matching (i.e., lack of a significant hidden population substructure) was indicated [4,8,9,19]. Thus, the adequacy of the case-control matching was also indicated in the present screening with microsatellite markers containing repeat

units of 3~6-bp. On the other hand, inflation factor for microsatellite markers containing repeat units of 2-bp was 0.520, therefore, the deflation described above was considered to be caused by mis-estimation in allele frequency in the screening with microsatellite markers containing repeat units of 2-bp. Therefore, among 1,328 markers selected in the 1st stage of GWAS, 431 microsatellite markers with 3~6-bp units were further subjected to the 2nd stage of GWAS.

The second set of DNA pools was examined for differences in allelic distribution for 431 microsatellite markers containing repeat units of 3~6-bp which passed the criteria of $p<0.05$ in the 1st stage of GWAS. The distribution for 17 (3.9%) markers were significantly different by the criteria of $p<0.05$ (2nd stage of GWAS in **Table II**). The inflation factor for the 2nd stage screening was 1.010, indicating the adequacy of the case-control matching as well as the lack of differential genotyping of cases and controls (**Supplementary Table I**).

Next, individual typing was done on the 17 markers, which passed the criteria for the 3rd stage, for 576 cases and 576 controls, consisting of 384 cases and 384 controls used in the first and second pooled-DNA screening and an additional 192 cases and 192 controls from NCCH (3rd stage of GWAS in **Table II**). These 384 cases and 384 controls were consisted of two sets of 192 subjects which were chosen from two sets of 200 subjects examined in the 1st and 2nd GWAS stages, respectively, by simple random sampling. These analyses led to the identification of six loci, including D6S0067i, with differences in allelic distributions between the cases and controls with P values less than 0.05 by the $\chi 2$ test. The D6S0067i locus showed a P value of 2.4×10^{-7}, while the other five showed P values of 0.012~0.0011. A level of $P<2.2 \times 10^{-6}$ was judged as significant

by applying Bonferroni correction for multiple test (i.e., $P<2.2\times10^{-6}=0.05/23{,}010$).

Genotyping of SNPs in the 6p21.31 locus. Five-hundred-twenty-five cases and 525 controls, which were respectively chosen from the 576 cases and 576 controls examined in the 3rd GWAS stage by simple random sampling, were subjected to SNP analysis. Twenty-nine SNPs were selected from the 450-kb region surrounding the D6S0067i locus based on the following criteria; 1) SNPs whose minor allele frequency in the Japanese population was > 0.01 in the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/), and 2) SNPs for which pre-designed or validated Taqman probes were available from Applied Biosystems (Foster City, CA). Three other SNPs, rs1794282, rs3129763 and rs2187668, which showed significant associations with lung cancer risk in Europeans [8], were also examined. Thirty-two SNPs, in total, were genotyped using the TaqMan method according to the protocol for the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems).

Twenty-four SNPs located in exon 2 of the *DRB1, DQA1* and *DQB1* genes, which enable allele discrimination for *DRB1, DQA1* and *DQB1* at high, low and high resolution levels, respectively, were genotyped by sequence-based typing methods recommended by the International Histocompatibility Working Group (Chapters 10-D, 11 and 12-A at http://www.ihwg.org/tmanual/TMcontents.htm). In brief, exon 2 of the *DRB1* and *DQB1* genes was amplified by PCR with mixtures of allele specific primers, while exon 2 of the *DQA1* gene was amplified with a set of common primers, and PCR products were directly sequenced using the ABI3700 DNA analyzer (Applied Biosystems). The location and alleles of the SNPs are described

according to the dbMHC database (http://www.ncbi.nlm.nih.gov/gv/mhc/). Based on the genotypes of 24 SNPs, alleles for *DRB1, DQA1, DQB1,* and *DR-DQ* were determined, and alleles with frequencies > 0.02 were subjected to the association study.

Statistical analyses. A Hardy-Weinberg equilibrium (HWE) test was performed using the SNPAlyze version 3 software (DYNACOM Co., Ltd), and SNPs with a *P* value for deviation > 0.01 were considered to be in HWE as described [7]. Calculation of the D′ and R^2 values between SNPs and allele/haplotype estimation was performed by the EM algorithm using the SNPAlyze version 3 software. The D6S0067i locus showed 19 polymorphic alleles in the same sets of cases and controls, and among them, alleles of 367-bp and 404-bp in sizes were significantly associated with an elevated risk for lung ADC (OR=1.60, $P= 9.9\times10^{-3}$ and OR=1.42, 4.9×10^{-5}, respectively). Therefore, for the calculation of the D′ and R^2 values, genotypes for the D6S0067i polymorphism was expressed by presence or absence of these two alleles (**Supplementary Table II**).

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

Genotyping of SNPs in the lung cancer susceptibility loci identified by previous GWASs. SNPs in the lung cancer susceptibility loci identified by previous GWASs were genotyped by the TaqMan method. Two intronic SNPs, rs2736100 and rs401681, in the *TERT* and *CLPTM1L* genes [4,21] were genotyped for the 5p15.33 locus against 2,343 cases and 1,173 controls (subjects of the NCCH set in **Table I**). Association results of the rs1051730 SNP in the *CHRNA3* gene for the 15q25.1 locus in a subset of the present study population were previously reported [22]. Therefore, in this study, 1,094 ADC cases and 236 controls which had not been examined in our previous study were genotyped [22]. Eight SNPs in the 6p21.33 locus, consisting of rs3117582 and seven SNPs in LD with this SNP in Europeans (D'=1 in the HapMap database), were genotyped for 525 ADC cases and 525 controls used for the mapping stage (**Table II**).

Results and Discussion

We performed a GWAS on the risk of lung ADC in a Japanese population for 23,010 polymorphic microsatellite loci. After a three stage screening against 576 ADC cases and 576 controls from the NCCH set (**Table I**), a locus, D6S0067i, at 6p21.31 was identified as being significantly different in allelic distribution after Bonferroni correction (i.e., P=2.4x10^{-7}, which is less than 0.05/23,010=2.2x10^{-6}) (details in **Subjects and Methods** and **Table II**).

The D6S0067i locus was mapped between two linkage disequilibrium (LD) blocks previously defined [23], one containing four HLA-class II genes, *HLA-DRA*, *-DRB1*, *-DQA1*, and *-DQB1*, and the other containing two pseudogenes, *HLA-DQA2* and *–DQB2* (**Figure 1**). Therefore, the locus of the strongest association was searched for in the 450-kb region containing these two LD blocks by analyzing 32 SNPs. Five-hundred-twenty-five cases and 525 controls, randomly selected from the GWAS subjects, were genotyped by the Taqman method (**Table III**). The rs1794282 SNP was monomorphic in the study subjects, while the other 31 were polymorphic. A SNP in the *DRA* gene, rs16822586, significantly deviated from the HWE in cases (P=0.001), while other SNPs did not deviate in either the cases or the controls, suggesting that SNPs in the regions examined in the present study normally segregated in the Japanese irrespective of lung cancer susceptibility. The 31 SNPs, which were polymorphic in our study population, comprised three LD blocks. The largest difference in allelic distribution between the cases and controls was observed at an intronic SNP in the *DQA1* gene, rs17426593 (OR=1.51, $P=4.2\times10^{-6}$) (**Figure 1**) in the block containing four HLA-class II genes (LD block 1 in **Table III**). The D6S0067i polymorphism was in LD (D′=0.516 in controls and D′=0.603 in cases) and showed a considerably high correlation coefficient (R^2=0.225 in controls and R^2=0.349 in cases) with the rs17426593 SNP (**Subjects and Methods** and **Supplementary Table II**). Therefore, we further examined associations of SNPs in this LD block with lung ADC risk.

Among the four HLA-class II genes, the *HLA-DRB1*, *-DQA1*, and *-DQB1* genes are known to comprise haplotypes carrying diverse non-synonymous SNPs and express polymorphic antigen proteins (HLA-class II alleles) [23]. Therefore, we genotyped the same set of case and control subjects for 24 SNPs in the coding exons of the *DRB1*, *DQA1* and *DQB1* genes that discriminate the HLA-class II alleles by the sequencing-based typing method (**Table III**). These 24 SNPs did not deviate from HWE in either the cases or the controls. These SNPs showed LD with the SNPs in LD block 1 (**Figure 1**), and patterns of LD were quite similar between the cases and controls (**Supplementary Figure 1**), indicating that distribution of 6p21.31 SNPs on chromosome DNA is not significantly different between these two populations. Many HLA-class II alleles, including those for each of the *DRB1*, *DQA1* and *DQB1* genes as well as those for contigs of the three genes (i.e., *DR-DQ* allele), determined by haplotypes for these exonic SNPs showed significantly different distributions between the cases and controls (**Table IV**). Among them, the DQA1*03 allele showed the largest difference with an OR of 1.50 (P=6.6x10^{-6}) and the DQA1*01 allele was the second largest (OR=0.69, P=2.8x10^{-5}). Accordingly, several *DR-DQ* alleles containing the DQA1*03 or DQA1*01 allele as well as several *DRB1* and *DQB1* alleles linked to the DQA1*03 or DQA1*01 allele also showed significantly different distributions (**Table IV**).

Discrimination of HLA alleles using intronic or intergenic SNPs is considered to be appropriate to analyze a large number of samples as an

alternative to conventional methods using exonic SNPs, due to rapidity and cost-effectiveness [23,24]. Two exonic SNPs in the *DQA1* gene, DQA1_2_145 and DQA1_2_150, which were genotyped by sequencing, were responsible for discrimination of the DQA1*01 and *03 alleles. These two SNPs showed high (R^2>0.98) correlation coefficients with two intronic SNPs in *DQA1*, rs17426593 and rs34843907, respectively (**Supplementary Table III**), which were genotyped by the Taqman method (**Supplementary Figure 2**). In fact, DQA1*03 and DQA1*01 alleles deduced by these two intronic SNPs showed high (R^2>0.97) correlation coefficients with those determined by two exonic SNPs (**Supplementary Table IV**). Thus, DQA1*03 and DQA1*01 alleles were discriminated by combined genotypes of two intronic SNPs, rs17426593 and rs34843907, and the association of *DQA1* alleles with lung ADC risk was further examined in a larger number of subjects by genotyping these two SNPs. Genotyping of an additional 1,131 ADC cases and 648 controls in the NCCH set enabled us to calculate combined ORs in 1,656 ADC cases and 1,173 controls (all subjects of the NCCH set in **Table I**), and the ORs of the DQA1*03 and DQA1*01 alleles were 1.36 (P=5.3x10^{-7}) and 0.77 (P=1.4x10^{-5}), respectively (**Figure 2a, Table II and Supplementary Table V**). Associations of these alleles with ADC risk were observed both in smokers and non-smokers; both male and female. A significant increase in OR of DQA1*03 for ADC risk was also observed in another set (NNGH set in **Table I)** of cases and controls **(Figure 2a, Supplementary Table V**), while a decrease in OR of DQA1*01 was insignificant. The DQA1*03 allele comprised the same haplotype with the risk allele of the intronic rs17426593 SNP (R^2=0.988), which showed the largest difference in allelic distribution between the cases and controls (**Table III**). Therefore,

DQA1*03 was defined as a risk allele in the 6p21.31 locus, although it is possible that intronic SNPs rather than exonic SNPs play a causal role in lung ADC susceptibility. Associations of the DQA1*03 allele with risks were further examined for SQC and SCC, two other major histological types of lung cancer to clarify whether the association is specific to ADC or not. Increases in ORs of DQA1*03 for SQC risk were significant both in the NCCH and NNGH sets, while ORs for SCC risk, calculated only for the NCCH set, were marginally increased for DQA1*03 (**Figure 2a, Supplementary Table V**). 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 a 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 7 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 the 6p21.31 locus containing four HLA-class II genes (**Supplementary Figure 3**). 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 1,656 ADC cases and 1,173 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)** as recently reported [11,21]. A 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 and Supplementary Table VII)**. 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=7.8x10^{-16}), reaching 4.76 for carriers of all four risk alleles **(Figure 2b and Supplementary Table VIII)**. 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)**. 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 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 COPD. A recent GWAS on COPD 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 1:1-4:1 [27], and this fact resulted in larger 95% confidence intervals 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**). The strength of this association was similar to those for SNPs in the *BAT3-MSH5* locus in those populations (**Supplementary Figure 3**). 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**). 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 approximately 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**), 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 no-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.

Disclosure of Potential Conflicts of Interest

The authors disclosed no potential conflicts of interest.

Acknowledgement

Grant support: Grants-in-Aid from the Ministry of Health, Labor and Welfare for the 3rd-term Comprehensive 10-year Strategy for Cancer Control (J. Yokota) and for Cancer Research, 19S-1 (T. Kohno) and the Ministry of Education, Culture, Sports, Science and Technology for Scientific Research on Priority Areas, KAKENHI 20014031 (T. Kohno).

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.

Figure legends

Figure 1 LD and association with lung cancer risk of polymorphisms in the 6p21.31 locus. The top panel shows association results for polymorphisms and the location of genes. The green square depicts the result for the D6S0067i microsatellite polymorphism, and red lozenges depict those for SNPs in the present study. Circles depict the results of GWASs on European and American populations. Blue circle: results for 1,989 cases and 2,625 controls in European countries [8]; yellow circle: 5,095 cases and 5,200 controls in European countries and USA [4]; purple circle: 2,971 cases and 3,746 controls in European countries, Canada and USA [5]. Results for ten SNPs commonly analyzed in the present and previous GWA studies (indicated in **Supplementary Table IX**) are depicted by bordered lozenges and circles. rs1794282 was monomorphic in the Japanese subjects. The bottom panel shows the LD structure for 55 SNPs in 525 control subjects. Boxes are shaded according to the pair-wise D′ values. Three LD blocks are indicated by bold black lines.

Figure 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% CI 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.** **(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% CI are shown. Detailed data, including the numbers of case and control subjects, are summarized in **Supplementary**

Tables VII and VIII.

References

1. Travis, W.D., Brambilla, E., Muller-Hermelink, H.K. and Harris, C.C. (2004) Pathology and Genetics, Tumours of Lung, Pleura, Thymus and Heart. In Travis, W.D., Brambilla, E., Muller-Hermelink, H.K. and Harris, C.C. (eds.), *World Health Organization Classification of Tumors:* , pp. 1-344.
2. Sun, S., Schiller, J.H. and Gazdar, A.F. (2007) Lung cancer in never smokers--a different disease. *Nat Rev Cancer*, **7**, 778-90.
3. Sobue, T., Yamamoto, S., Hara, M., Sasazuki, S., Sasaki, S. and Tsugane, S. (2002) Cigarette smoking and subsequent risk of lung cancer by histologic type in middle-aged Japanese men and women: the JPHC study. *Int J Cancer*, **99**, 245-51.
4. Wang, Y., Broderick, P., Webb, E., Wu, X., Vijayakrishnan, J., Matakidou, A., Qureshi, M., Dong, Q., Gu, X., Chen, W.V., Spitz, M.R., Eisen, T., Amos, C.I. and Houlston, R.S. (2008) Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet*, **40**, 1407-9.
5. McKay, J.D., Hung, R.J., Gaborieau, V., Boffetta, P., Chabrier, A., Byrnes, G., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., Bencko, V., Foretova, L., Janout, V., McLaughlin, J., Shepherd, F., Montpetit, A., Narod, S., Krokan, H.E., Skorpen, F., Elvestad, M.B., Vatten, L., Njolstad, I., Axelsson, T., Chen, C., Goodman, G., Barnett, M., Loomis, M.M., Lubinski, J., Matyjasik, J., Lener, M., Oszutowska, D., Field, J., Liloglou, T., Xinarianos, G., Cassidy, A., Vineis, P., Clavel-Chapelon, F., Palli, D., Tumino, R., Krogh, V., Panico, S., Gonzalez, C.A., Ramon Quiros, J., Martinez, C., Navarro, C., Ardanaz, E., Larranaga, N., Kham, K.T., Key, T., Bueno-de-Mesquita, H.B., Peeters, P.H., Trichopoulou, A., Linseisen, J., Boeing, H., Hallmans, G., Overvad, K., Tjonneland, A., Kumle, M., Riboli, E., Zelenika, D., Boland, A., Delepine, M., Foglio, M., Lechner, D., Matsuda, F., Blanche, H., Gut, I., Heath, S., Lathrop, M. and Brennan, P. (2008) Lung cancer susceptibility locus at 5p15.33. *Nat Genet*, **40**, 1404-6.
6. Liu, P., Vikis, H.G., Wang, D., Lu, Y., Wang, Y., Schwartz, A.G., Pinney, S.M., Yang, P., de Andrade, M., Petersen, G.M., Wiest, J.S., Fain, P.R., Gazdar, A., Gaba, C., Rothschild, H., Mandal, D., Coons, T., Lee, J., Kupert, E., Seminara, D., Minna, J., Bailey-Wilson, J.E., Wu, X., Spitz, M.R., Eisen, T., Houlston, R.S., Amos, C.I., Anderson, M.W. and You, M. (2008) Familial aggregation of common sequence variants on 15q24-25.1 in lung cancer. *J Natl Cancer Inst*, **100**, 1326-30.
7. Kohno, T., Kunitoh, H., Suzuki, K., Yamamoto, S., Kuchiba, A., Matsuno, Y., Yanagitani, N. and Yokota, J. (2008) Association of KRAS polymorphisms with risk for lung adenocarcinoma accompanied by atypical adenomatous hyperplasias. *Carcinogenesis*, **29**, 957-63.
8. Hung, R.J., McKay, J.D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., Bencko, V., Foretova, L., Janout, V., Chen, C., Goodman, G., Field, J.K., Liloglou, T., Xinarianos, G., Cassidy, A., McLaughlin, J., Liu, G., Narod, S., Krokan, H.E., Skorpen, F., Elvestad, M.B., Hveem, K., Vatten, L., Linseisen, J., Clavel-Chapelon, F., Vineis, P., Bueno-de-Mesquita, H.B., Lund, E., Martinez, C., Bingham, S., Rasmuson, T., Hainaut, P., Riboli, E., Ahrens, W.,

Benhamou, S., Lagiou, P., Trichopoulos, D., Holcatova, I., Merletti, F., Kjaerheim, K., Agudo, A., Macfarlane, G., Talamini, R., Simonato, L., Lowry, R., Conway, D.I., Znaor, A., Healy, C., Zelenika, D., Boland, A., Delepine, M., Foglio, M., Lechner, D., Matsuda, F., Blanche, H., Gut, I., Heath, S., Lathrop, M. and Brennan, P. (2008) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*, **452**, 633-7.

9. Amos, C.I., Wu, X., Broderick, P., Gorlov, I.P., Gu, J., Eisen, T., Dong, Q., Zhang, Q., Gu, X., Vijayakrishnan, J., Sullivan, K., Matakidou, A., Wang, Y., Mills, G., Doheny, K., Tsai, Y.Y., Chen, W.V., Shete, S., Spitz, M.R. and Houlston, R.S. (2008) Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet*, **40**, 616-22.
10. Broderick, P., Wang, Y., Vijayakrishnan, J., Matakidou, A., Spitz, M.R., Eisen, T., Amos, C.I. and Houlston, R.S. (2009) Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. *Cancer Res*, **69**, 6633-41.
11. Landi, M., Chatterjee, N., Kai Yu, K., Goldin, L., Goldstein, A., Rotunno, M., Mirabello, L., Jacobs, K., Wheeler, W., Yeager, M., Bergen, A., Li, Q., Consonni, D., Pesatori, A., Wacholder, S., Thun, M., Diver, R., Oken, M., Virtamo, J., Albanes, D., Wang, Z., Burdette, L., Doheny, K., Pugh, E., Laurie, C., Brennan, P., Hung, R., Gaborieau, V., McKay, J., Lathrop, M., McLaughlin, J., Wang, Y., Tsao, M., Spitz, M., Wang, Y., Krokan, H., Vatten, L., Skorpen, F., Arnesen, E., Benhamou, S., Bouchard, C., Metsapalu, A., Vooder, T., Nelis, M., Välk, K., Field, J., Chen, C., Goodman, G., Sulem, P., Thorleifsson, G., Rafnar, T., Eisen, T., Sauter, W., Rosenberger, A., Bickeböller, H., Risch, A., Chang-Claude, J., Wichmann, H., Stefansson, K., Houlston, R., Amos, C., Fraumeni, J., Savage, S., Bertazzi, P., Tucker, M., Chanock, S. and Caporaso, N. (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., Nakazono, K., Takahashi, A., Saito, S., Hosono, N., Kubo, M., Nakamura, Y. and Kamatani, N. (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**, 445-56.
13. Sakamoto, H., Yoshimura, K., Saeki, N., Katai, H., Shimoda, T., Matsuno, Y., Saito, D., Sugimura, H., Tanioka, F., Kato, S., Matsukura, N., Matsuda, N., Nakamura, T., Hyodo, I., Nishina, T., Yasui, W., Hirose, H., Hayashi, M., Toshiro, E., Ohnami, S., Sekine, A., Sato, Y., Totsuka, H., Ando, M., Takemura, R., Takahashi, Y., Ohdaira, M., Aoki, K., Honmyo, I., Chiku, S., Aoyagi, K., Sasaki, H., Ohnami, S., Yanagihara, K., Yoon, K.A., Kook, M.C., Lee, Y.S., Park, S.R., Kim, C.G., Choi, I.J., Yoshida, T., Nakamura, Y. and Hirohashi, S. (2008) Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat Genet*, **40**, 730-40.
14. Tamiya, G., Shinya, M., Imanishi, T., Ikuta, T., Makino, S., Okamoto, K., Furugaki, K., Matsumoto, T., Mano, S., Ando, S., Nozaki, Y., Yukawa, W., Nakashige, R., Yamaguchi, D., Ishibashi, H., Yonekura, M., Nakami, Y., Takayama, S., Endo, T., Saruwatari, T., Yagura, M., Yoshikawa, Y., Fujimoto, K., Oka, A., Chiku, S., Linsen, S.E., Giphart, M.J., Kulski, J.K., Fukazawa, T., Hashimoto, H., Kimura, M., Hoshina, Y., Suzuki, Y., Hotta, T., Mochida, J.,