Recent advances in genome analysis technologies have provided a detailed genome-wide view of cancerous and non-cancerous cells. Lung cancer is largely caused by tobacco smoking, but several studies have implicated inherited genetic factors in disease etiology. Genome-wide association studies (GWASs) using DNA chips have identified loci/genes with polymorphisms that underlie interindividual differences in cancer susceptibility, including single nucleotide polymorphisms (SNPs). CHRNA (cholinergic receptor, nicotinic, alpha), TERT (telomerase reverse transcriptase) and TP63 (tumor protein p63) loci have been linked to lung cancer susceptibility by GWASs. SNPs in TERT and TP63 are preferentially associated with the risk of adenocarcinoma, the commonest histological type of lung cancer affecting both smokers and non-smokers, whereas those in CHRNA are associated with lung cancer risk irrespective of histological type. An association of functional polymorphisms in DNA repair/metabolic genes with the risk of squamous cell carcinoma, a major histological type developed in smokers, has been suggested, but it remains inconclusive. It was also suggested that an SNP in the TP53 tumor suppressor gene influences the response to platinum-doublet chemotherapy in lung cancer patients. However, analyses have shown that only a subset of SNPs is involved in lung carcinogenesis/therapy. Further GWASs are needed to translate the information on genetic variations into cancer prevention and clinical practice by focusing on specific subtypes of lung cancers or therapeutic responses.

Key words: lung cancer, genome-wide association studies (GWASs), single nucleotide polymorphisms (SNPs), cancer susceptibility, DNA repair genes, metabolic genes

GWAS

Genetic polymorphisms responsible for cancer susceptibility have been investigated in case-control (association) studies where polymorphisms with different distributions between cancer cases and non-cancer controls have been identified (9) (Fig. 1A). Genetic factors involved in cancer susceptibility were previously studied in the polymorphisms of genes encoding proteins with the ability to metabolize carcinogens or suppress mutations induced by carcinogens. However, 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 diseases using GWASs, including lung cancer (http://gwas.lifesciencedb.jp/cgi-bin/gwasdb/gwas_top.cgi) (10). Several cancer susceptibility genes/loci have been identified by GWASs. For example, gene polymorphisms in the ABO blood group gene (11) and the ALDH2 (aldehyde dehydrogenase 2) gene (12) were associated with susceptibility to pancre-
GWASs of Lung Cancer Risk

Three chromosomal loci, 15q24-25.1, 5p15.33 and 6p21, were found to be associated with lung cancer risk in GWASs of European/American populations (2–4), while locus 3q28 was associated with lung ADC risk in a GWAS of Japanese/Korean populations (7). The chromosomal 15q24-25.1 region contains nicotinic acetylcholine receptor subunit genes, i.e., CHRNA3 (cholinergic receptor, nicotinic, alpha 3) and CHRNA5, and their products are expressed in pulmonary epithelial cells and bind to nicotine, an addictive compound found in cigarette smoke, and nitrosamines, which are potential lung carcinogens in cigarette smoke and food (16,17). The associated lung cancer risk of intronic SNPs in 15q24-25.1 was replicated in Asian populations (18,19). However, the frequency of risk alleles was much lower than that in populations of European descent (Fig. 2A, left). Thus, the CHRNA risk alleles make a smaller subset of individuals more susceptible to lung cancer in Asians compared with European and American populations. Interestingly, the association with lung cancer risk was found irrespective of smoking and histological types in many studies, so CHRNA SNPs might contribute to the risk in a (active) smoking-independent manner (Fig. 2A, right). The involvement of other factors should be further investigated, such as food intake and passive smoking. On the other hand, the contribution of CHRNA SNPs to lung cancer risk via tobacco addiction has been also suggested (20). Thus, further studies need to investigate a cohort of subjects where detailed data are available on lung cancer development, food intake, active and passive smoking exposure, nicotine dependence, and the duration and intensity of smoking to elucidate how CHRNA SNPs contribute to lung cancer risk. The 5p15.33 region contains the TERT (telomerase reverse transcriptase) gene. TERT is known to function in telomere replication and maintenance, and it promotes epithelial cell proliferation (21). The risk allele frequency of the landmark SNP (rs2736100) is similar among ethnic populations (Fig. 2B, left), and associations have been detected in among Europeans, American and Asians (22,23) (Fig. 2B, right). Interestingly, this SNP is associated with the risk of ADC, but not SQC or SCLC (5,6,22–24), suggesting a preferential association of this locus with lung cancer risk in never-smokers (Fig. 2B, right). Risk associations of the
Fig. 2. Increased lung cancer risk with SNPs identified by GWASs according to populations, smoking behavior, and histological types. Categories where statistically significant association was observed are underlined. (A) rs8034191 (CHRNA5) at 15q24–25.1. (B) rs2736100 (TERT) at 5p15.33. (C) rs10937405 (TP63) at 3q28. (D) HLA-DQA1*03 (HLA-DQA1) at 6p21. The left graphs show frequencies of chromosomes for risk allele in all chromosomes in each population. Frequencies were determined by the HapMap project or the International Histocompatibility Working Group are shown on the left.

5p15.33 genotype have been detected in lung cancer and other types of cancers, including cancers of the brain, bladder, prostate, uterine cervix, and skin (25). The rs2736100 SNP is located in intron 2 of TERT, which is a putative regulatory region. It was previously suggested that risk alleles in the TERT gene are associated with shorter telomeres (25). Therefore, these variants may lead to an increase in the gradual shortening of telomeres over time, leading to genomic instability driving carcinogenesis in many organs. Thus, functional analyses of TERT SNPs are warranted. The 3q28 region contains the TP63 gene that encodes a member of the tumor suppressor TP53 (also known as p53) gene family, which is involved in development, differentiation, and the cellular stress response (26). The risk allele frequency of the landmark SNP (rs10937405) was similar among ethnic populations (Fig. 2C, left) and association with lung SQC risk were detected in both Asians and Europeans (7, 27) (Fig. 2C, right). However, the association appears to be stronger in Asians than Europeans (27). In Europeans, there is a weaker association with lung SQC risk, while the association with SQC risk is unknown in Asians. SNPs associated with lung cancer risk are located in intron 1 of TP63 and it has been suggested that they have a functional role in the regulation of TP63 gene expression. TP63 is induced after the exposure of cells to DNA damage. Therefore, inter-individual differences in the accumulation of DNA damage and the response to genotoxic stress might contribute to lung cancer susceptibility. Associations of SNPs in the 6p21 region were found in a GWAS of Europeans (4). This region contains the BAT3 (HLA-B associated transcript 3) and MSH5 (mutS homolog 5) genes. The BAT3 protein complexes with a histone acetyltransferase, p300, which acetylates p53 protein in response to DNA damage, while MSH5 is involved in DNA mismatch repair. However, a recent pooled analysis by the International Lung Cancer Consortium failed to replicate the association of these SNPs with lung cancer risk (24). Our GWAS on Japanese lung ADC indicated that the HLA-DQA1 locus encoding a HLA (human leukocyte antigen)-class II protein (1 Mb proximal to the BAT3-MSH5 locus) is a significant region in 6p21. The DQA1*03 allele of the HLA-DQA1 gene was associated with an increased risk of the development of all major histological types of lung cancer (23) (Fig. 2D). It is possible that the HLA-DQA1 polymorphism confers lung cancer susceptibility due to inter-individual
Fig. 3.  Lung ADC risk of combined genotypes. Left: ORs of combined TERT (rs2736100) and TP63 (rs10937405) genotypes. Right: ORs of combined TERT (rs2736100), CHRNA3 (rs105173), and HLA-DQA1 (*03) genotypes.

Genetic Polymorphisms in Lung Cancer

Fig. 4. Increased lung SQC risk with functional SNPs in the TP53 (rs1042522), OGG1 (rs1052133), and CYP1A1 (rs1048943) genes. (A) Frequencies of risk alleles in each population determined by the HapMap project. (B) ORs of TP53, OGG1, and CYP1A1 genotypes. TP53-72Pro, OGG1-326Cys and CYP1A1-462Val alleles were shown to be statistically significantly associated with lung SQC risk. Genotypes showing statistically significant association are underlined.

326Ser allele (36,37). The risk (462Val) allele of the Ile462Val SNP in the CYP1A1 gene encodes a metabolic protein with a higher activity in bio-activating the major tobacco procarcinogens, polycyclic aromatic hydrocarbons (PAHs), than the 462Ile allele (38). However, the associations of these functional polymorphisms were not investigated in previous GWASs due to the lack of probes for discriminating these polymorphisms in the platforms used by GWASs (http://www.ncbi.nlm.nih.gov/snp), so they remain unconfirmed. We previously demonstrated the association of these functional SNPs with lung SQC risk in a population where polymorphisms of the GWAS genes show associations (39). Genotypes for two DNA repair genes, TP53 and OGG1, and a metabolic gene, CYP1A1, showed significant associations with SQC risk along with those for CHRNA3 and HLA-DQA1 (Fig. 4B). Based on these results, there is a need to analyze various functional polymorphisms together with millions of GWAS marker SNPs. This will provide a powerful method for analyzing these polymorphisms in populations that were tested in recent GWASs.

SNPs in DNA repair genes and therapeutic responses: Genetic polymorphisms underlie inter-individual differences in terms of susceptibility to disease and also the therapeutic response, although published association studies in this area lack sufficient case numbers (40). Agents that damage DNA or that disrupt differences in the immune response against tumor cells. However, 6p21 is a highly polymorphic region containing major histocompatibility complex genes, therefore, the observed associations might have simply reflect differences in the population substructure (6). Further investigation of this region is warranted.

Lung ADC risk based on combined genotypes: Miki et al. and we estimated the risk of lung ADC caused by combined genotypes with multiple lung ADC susceptibility loci (7,23) (Fig. 3). These loci were suggested to independently confer risk, and carriers of all risk alleles, i.e., <10% of the population, had an odds ratio (OR) of >4.0 compared with those having no risk alleles. The results are significant when we consider that the OR of smoking for lung ADC risk is <2.0 in the Japanese population (28). The relative risk of carrying these variants should now be assessed in a cohort study to the relative ratios of these combined genotypes and to identify high risk individuals in near future.

Functional polymorphisms in DNA repair and metabolic genes: Studies of DNA adducts/damage, including that produced by tobacco carcinogens and their repair processes, have led to the identification of various metabolic and DNA repair genes with functional polymorphisms, which might possibly produce inter-individual differences in the rate of somatic mutation and the susceptibility to tobacco-related cancers (29). Representative SNPs in TP53, OGG1, and CYP1A1 have been indicated to be associated with lung cancer risk for a long time (30–34) and the risk allele frequencies of those SNPs are shown to be different among ethnic populations. (Fig. 4A). The risk (72Pro) allele of the TP53-Arg72Pro SNP in the TP53 gene encodes a protein with a weaker apoptotic activity that better allows the survival of DNA-damaged cells compared with the 72Arg allele (35), while the risk (326Cys) allele of the Ser326Cys SNP in the OGG1 gene encodes a DNA glycosylase with a weaker activity in repairing an oxidatively damaged promutagenic base produced by tobacco carcinogens, 8-hydroxyguanine, compared with the
protein has a greater activity in inducing apoptosis than the p53-72Pro protein, as described above (35), however, the opposite relationship was reported for mutant p53 proteins. p73 is a p53-related protein that plays a role in the apoptosis of cancer cells carrying TP53 mutations via anticancer agents, and its function is abrogated by mutant p53 proteins. This abrogating activity is greater in mutant p53 proteins with an Arg residue at codon 72 compared with those with a Pro residue (42) and, therefore, the TP53-72Pro allele may lead to a better response to platinum-based doublet chemotherapy in NSCLC patients (Fig. 5B). This study indicates the potential utility of SNP as predictive markers for responses to chemotherapy.

**Future Directions**

Genetic factors affecting lung cancer risk have been identified by GWASs and other association studies. The results indicate that risk variants confer different magnitudes of increased risk in different populations for a variety of reasons, including differences in allele frequency and the genetic and environmental backgrounds that interact with the variants. Given the statistical power applied to the detection of associations in GWASs to date, there are unlikely to be many additional SNPs (tagged by commercially available DNA chips) with similar effects on alleles with high frequencies (>0.2) in populations of European ancestry (5,6). Thus, several different GWAS approaches are required to identify additional genetic factors underlying lung cancer risk. This should include GWASs of Asian populations. The fraction of lung ADC patients who are never-smokers and female is considerably higher in Asians than Europeans/Americans (43,44), suggesting that the former experience a greater or more distinct effect of genetic factors than the latter. Therefore, GWASs of Asians would be highly useful. In addition, GWASs focusing on specific lung cancer types would be worthwhile, because some genetic factors might be specifically associated with the risk of a specific type of lung cancer, such as SCLC, lung cancers in female never-smokers, or lung cancers with defined gene mutations. Lung ADC is known to develop via several carcinogenic pathways defined by oncogenic driver mutations in EGFR, KRAS, HER2, ALK, and RET genes, and the etiological factors are suggested to be different among those pathways (1,45,46). Finally, low frequency variants and common SNPs that have not been tagged by the DNA chips used in previous studies might also be involved in lung cancer risk (47,48). Thus, efforts to expand the scale of GWASs in terms of both sample size and SNP coverage, and to increase the number of SNPs taken forward to large-scale replication, may also lead to the identification of additional variants for lung cancer. Understanding the remaining genetic factors will help greatly in clarifying the disease etiology.
and also in identifying high risk individuals for targeted screening and/or prevention. SNPs in DNA repair genes are associated with the response to platinum-doublet therapy. SNPs can be examined using blood cells and they may provide useful biomarkers in the clinical decision-making process for patients with advanced NSCLC who do not received surgery, which makes the molecular analysis of cancer cells difficult. No large-scale GWASs have been performed to investigate links with the response to lung cancer therapy or drug toxicities to the best of our knowledge, and these will also be useful for identifying biomarkers with clinical applications.

Acknowledgments: This work was supported in part by Grants-in-Aid from the Ministry of Health, Labor and Welfare for Research on Applying Health Technology, the Japan Society for the Promotion of Science for Research Activity Start-up (23800073), and the Ministry of Education, Culture, Sports, Science and Technology of Japan for Scientific Research on Innovative Areas (22131006) and the National Cancer Center Research and Development Fund.

References
25 Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson


