An Overview of Genetic Polymorphism and Lung Cancer Risk

Kouya Shiraishi, Takayuki Honda and Takashi Kohno

Division of Genome Biology, National Cancer Center Research, Tokyo, Japan

Corresponding author: Kouya Shiraishi, Division of Genome Biology, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan, Tel: +81-3-3542-2511; Fax: +81-3-3542-0807; E-mail: kshirais@ncc.go.jp

Rec Date: Jan 20, 2016; Acc Date: Feb 22, 2016; Pub Date: Feb 29, 2016

Keywords: Lung cancer; Genome-wide association study; Single nucleotide polymorphism; Cancer susceptibility

Abstract

The incidence and mortality rates of lung cancer in Asia have increased over the past few decades. Owing to the development of molecular-targeted drugs, the prognosis for patients with advanced stage lung cancer has improved. Despite this improvement the 5-year survival rate is very low. Accordingly, it is necessary to identify groups at high risk for lung cancer for prevention. Both environmental and genetic factors are related to the development of lung cancer. For example, cigarette smoking, passive (secondhand) smoking, asbestos exposure, and air pollution are associated with lung cancer. Genetic polymorphisms, including Single Nucleotide Polymorphisms (SNPs), underlie inter-individual differences in cancer susceptibility, and genetic loci for lung cancer risk have been identified by Genome-Wide Association Studies (GWAS) in both European/American and Asian populations. Recent GWAS of lung cancer have identified genetic susceptibility variants on Chromosome 15q25 (CHRNA), 5p15 (TERT), 3q28 (TP63), 8q22 (DCBLD1), 6p21 (BAT3-MSH2 and BTNL2), 10q25 (VTI1A), and 17q.24 (BPTF); however, loci significantly associated with lung cancer risk differ among ethnicities. Here, we review previous GWAS and discuss genetic factors that may be useful for the early detection or prevention of lung cancer.

GWAS of lung cancer risk

The results of previous GWAS are summarized in the Table 1. Several studies report that three chromosomal loci, 15q24−25.1, 5p15.33, and 6p21, are associated with lung cancer risk in European and American populations [7–14], while four, 3q28, 5p15.33, 6p21, and 17q24.2, are associated with ADC risk in Japanese and/or Korean populations [10,11]. In addition, loci at 5q32, 10p14, 13q12.12, 20q13.2, and 22q12.2 are associated with lung cancer risk in the Chinese population [12,15] and loci at 10q25 and 6p21 are associated with susceptibility to lung cancer in females who have never smoked in the Asian population [16]. Loci at 12p13.33 and 12q23.1 are associated with SQC risk in individuals of European ancestry [17] and in the Chinese population [18]. However, the associations for some susceptibility loci were not validated in independent samples, and further verification is needed.

The chromosomal 15q24−25.1 region contains nicotinic acetylcholine receptor subunit genes, i.e., CHRNA3 (cholinergic receptor, nicotinic, alpha 3) and CHRNA5. These subunits are expressed in pulmonary epithelial cells and bind to nicotine and nitrosamines, including potential lung carcinogens in cigarette smoke and food [19,20]. The binding induces proliferation of cancer cells [20]. In Asia, associations between SNPs in these genes and lung cancer risk have been reported [21,22], but studies have yielded conflicting results [23]. Because the frequency of risk alleles in the Asian population is much lower than that in European populations, the conflicting results probably reflect the lower statistical power in these studies. At minimum, the contribution of the CHRNA risk alleles to lung cancer risk differs between Asian and European populations. Thus, it is necessary to investigate a cohort of subjects or large sample sets in Asian populations.
<table>
<thead>
<tr>
<th>Susceptibility locus</th>
<th>Reported Gene(s)</th>
<th>Strongest SNP-Risk Allele</th>
<th>Risk allele frequency*</th>
<th>Odds ratio</th>
<th>Histology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>EAS</td>
<td>EUR</td>
<td>SAS</td>
</tr>
<tr>
<td>(1) Common variants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3q28</td>
<td>TP63</td>
<td>rs4488809-C/</td>
<td>0.520/ 0.689</td>
<td>0.494/ 0.572</td>
<td>0.468/ 0.681</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs10937407-C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5p15.33</td>
<td>TERT</td>
<td>rs2736100-G/</td>
<td>0.415/ 0.364</td>
<td>0.499/ 0.413</td>
<td>0.605/ 0.579</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs2853677-G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5p15.33</td>
<td>CLPTM1L</td>
<td>rs31489-C/</td>
<td>0.810/ 0.681</td>
<td>0.591/ 0.559</td>
<td>0.832/ 0.805</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs401681-G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5q32</td>
<td>STK32A</td>
<td>rs2895680-C</td>
<td>0.338</td>
<td>0.233</td>
<td>0.429</td>
</tr>
<tr>
<td>6q22.1</td>
<td>DCBLD1</td>
<td>rs9387478-C</td>
<td>0.502</td>
<td>0.495</td>
<td>0.266</td>
</tr>
<tr>
<td></td>
<td>TRNAA-UGC</td>
<td>rs4324798-A</td>
<td>1.000</td>
<td>0.071</td>
<td>1.000</td>
</tr>
<tr>
<td>6p21.32</td>
<td>HLA class II</td>
<td>rs2395185-T</td>
<td>0.311</td>
<td>0.321</td>
<td>0.299</td>
</tr>
<tr>
<td></td>
<td>BTN2</td>
<td>rs3817963-G</td>
<td>0.231</td>
<td>0.276</td>
<td>0.333</td>
</tr>
<tr>
<td>6p21.33</td>
<td>BAT3-MSH5-APOM</td>
<td>rs3117582-C</td>
<td>0.000</td>
<td>0.075</td>
<td>0.000</td>
</tr>
<tr>
<td>10p14</td>
<td>Near GATA3</td>
<td>rs1663689-A</td>
<td>0.560</td>
<td>0.806</td>
<td>0.655</td>
</tr>
<tr>
<td>10q25.2</td>
<td>VT11A</td>
<td>rs7086603-A</td>
<td>0.294</td>
<td>0.028</td>
<td>0.050</td>
</tr>
<tr>
<td>12p13.33</td>
<td>RAD52</td>
<td>rs6489769-G</td>
<td>0.465</td>
<td>0.364</td>
<td>0.330</td>
</tr>
<tr>
<td>12q23.1</td>
<td>SLC17A8-NR1H4</td>
<td>rs12296850-A</td>
<td>0.743</td>
<td>0.940</td>
<td>0.806</td>
</tr>
<tr>
<td>13q12.12</td>
<td>MIPEP</td>
<td>rs75955-G</td>
<td>0.337</td>
<td>0.627</td>
<td>0.687</td>
</tr>
<tr>
<td>15q25.1</td>
<td>CHRNA3,CHRNA5,</td>
<td>rs8034191-C/</td>
<td>0.029/ 0.027</td>
<td>0.375/ 0.369</td>
<td>0.193/ 0.183</td>
</tr>
<tr>
<td></td>
<td>CHRNB4,PSM4A,AGPHD1</td>
<td>rs1051730-T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17q24.2</td>
<td>BPTF</td>
<td>rs7216068-A</td>
<td>0.692</td>
<td>0.214</td>
<td>0.269</td>
</tr>
<tr>
<td>20q13.2</td>
<td>CYP24A1</td>
<td>rs4809957-T</td>
<td>0.394</td>
<td>0.791</td>
<td>0.625</td>
</tr>
<tr>
<td>22q12.2</td>
<td>MTMR3</td>
<td>rs36600-A</td>
<td>0.093</td>
<td>0.273</td>
<td>0.221</td>
</tr>
<tr>
<td>(2) Rare variants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13q13.1</td>
<td>BRCAC2</td>
<td>rs11571823-T</td>
<td>0.000</td>
<td>0.011</td>
<td>0.0082</td>
</tr>
<tr>
<td>22q12.1</td>
<td>CHEK2</td>
<td>rs1787961-C</td>
<td>1.000</td>
<td>0.0050</td>
<td>1.000</td>
</tr>
<tr>
<td>6p21.33</td>
<td>BAT2</td>
<td>rs9469031-C</td>
<td>0.985</td>
<td>0.9960</td>
<td>0.989</td>
</tr>
<tr>
<td>6p21.33</td>
<td>FKBPL</td>
<td>rs200847762-C</td>
<td>0.9990</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>20q11.21</td>
<td>BPIFB1</td>
<td>rs6141383-A</td>
<td>0.016</td>
<td>1.000</td>
<td>0.0020</td>
</tr>
</tbody>
</table>

*EAS, East Asia in 1000 Genomes super population; EUR, Europe in 1000 Genomes super population; SAS, South Asia in 1000 Genomes super population.

Table 1: Summary of loci identified in previous GWAS for lung cancer risk.

The 5p15.33 region contains the TERT (Telomerase Reverse Transcriptase) and CLPTM1L (cleft lip and palate transmembrane protein 1) genes. TERT protein functions in telomere replication and maintenance and promotes epithelial cell proliferation [24]. The risk allele frequency of the TERT SNP rs2736100 is similar among various ethnic populations, and associations have been detected in Europeans, Americans, and Asians [9-14,23]. According to a recent meta-analysis of various lung cancer types, this SNP is more strongly associated with ADC than with SQC or SCC [25]. Similarly, the association between lung cancer and TERT SNPs is stronger (i.e., associations with greater odds ratios) in Asian than in Caucasian populations (Table 1). The frequency of driver mutations in ADC is different among Caucasians and Asians. EGFR is activated by inflame deletion mutations in exon 19 and a point mutation in exon 21 (e.g., L858R) in 40–55% of ADC cases in the East Asian population and in 5–15% of ADC cases in the American population [26]. The association for the TERT SNP rs2736100 SNP in TERT intron 2 was reported to be stronger in ADC with an EGFR mutation than in ADC without an EGFR mutation [27].

Therefore, genetic modifiers and/or environmental factors might contribute to differences among histological types. The rs2736100 SNP is associated with susceptibility to other cancer types, including cancers of the brain, bladder, prostate, uterine cervix, and skin, as well as testicular cancer and chronic lymphocytic leukemia [28,29]. There are conflicting results regarding the association between the TERT SNPs and telomere length in leucocytes [28,30]. However, variants in the TERT promoter region (rs2853669 and rs2735940) and intron 4 (rs10069690) are likely to affect telomere length in leucocytes or non-cancerous tissues [31,32]. The rs2853669 SNP is associated with ADC risk (rs2853669, odds ratio (OR) = 1.38); however, neither rs2853669 nor rs10069690 is statistically associated with ADC risk in the Japanese population (OR=1.06 and 1.07, respectively) [11]. To elucidate the effects of TERT SNPs, additional studies are needed to determine the relationships among TERT SNPs, ADC risk, and telomere length in non-cancerous or normal lung tissues.

CLPTM1L is located near TERT. This gene was identified by screening and ovarian cancer cell line for Cisplatin (CDDP) resistance-related genes. A recent meta-analysis suggested that the association between rs31489 located in CLPTM1L and lung cancer risk was stronger in a population of European ancestry than in Asians [29]. CLPTM1L is required for lung tumorigenesis in a conditional K-RasG12D transgenic mouse model [33]. The frequency of KRAS-mutated ADC is different among Caucasians and Asians. KRAS protein is activated by a single amino acid substitution (at codons 12 or 13) in 20–30% of ADC cases in the American population and in 8–10% of ADC cases in the East Asian population [26]. CLPTM1L SNPs might be preferentially associated with the risk of KRAS-mutated ADC.

The 3q28 region contains TP63; it 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 [34]. The risk allele frequencies of the rs10937405 and rs4488809 SNPs, which are located in intron 1 of TP63, are similar among ethnic populations, and previous GWAS show that TP63 SNPs are risk-conferencing factors for the development of non-small cell carcinoma, especially for ADC and East Asian populations [10,35]. The rs4488809 SNP was identified in an eQTL (expression quantitative trait loci) analysis of TP63 in lung tissue [36]. Therefore, inter-individual differences in the accumulation of DNA damage and the response to genotoxic stress might contribute to ADC development.

The 17q24.3 region contains BPTF (bromodomain PHD finger transcription factor), which encodes a nucleosome remodeling factor that regulates transcription via specific recognition of methylated histone proteins [37]. According to a GWAS, variants of BPTF are associated with ADC risk, and the risk allele frequency of the rs7216064 SNP is different among ethnic populations [11]. The association for this variant was validated in one study of female lung cancer patients with no history of smoking in Asian populations [16]. BPTF protein is required for c-MYC transcriptional activity [38,39] and somatic alterations in BPTF have been detected in a small percentage of ADC tissue samples [2,30,39]. However, further studies are needed to determine whether BPTF is responsible for susceptibility to ADC.

Associations between lung cancer and SNPs in the 6p21 region have been found in GWAS of Europeans and Asians [8-16]. This region contains BAT3 (HLA-B associated transcript 3) and MSH5 (mutS homolog 5). The BAT3 protein forms a complex with a histone acetyltransferase, p300, which acetylates the p53 protein in response to DNA damage, whereas MSHS is involved in DNA mismatch repair. Variants in BAT3-MSH5 are associated with lung cancer risk in European populations, irrespective of histological type; however, these BAT3 and MSH5 SNPs are mono-morphic in Asian populations. In our own GWAS, variants of BTNL2 (butyrophilin-like 2) and the HLA-DQA1 locus located in the HLA (Human Leukocyte Antigen)-class II region in 6p21 were significantly associated with ADC risk [11,40]. The risk allele frequency of BTNL2 SNPs is similar among various ethnic populations. In addition, BTNL2 SNPs are associated with the risk of lung cancer in females with no history of smoking [16]. It is possible that polymorphisms in HLA-class II genes confer lung cancer susceptibility owing to inter-individual differences in the immune response against tumor cells. The 6p21 region is highly polymorphic and contains most major histocompatibility complex genes. The construction of HLA reference panels of Caucasians and Japanese [41,42] has enabled GWAS via HLA genotype imputation, and therefore will enable the identification of HLA alleles that confer a cancer risk.

To identify rare variants that have large effects on lung cancer risk (e.g., OR > 2.0), it may be particularly useful to conduct GWAS via genotype imputation using the 1000 Genomes reference panel [43] or whole-genome and/or whole-exome sequencing. By a study using genotype imputation, such variants in BRCA2 (p.Lys3326X; rs11571833) and CHEK2 (p.Leu1577Trh; rs13314271) associated with SGC risk were found in populations of European ancestry [35]. Rare variants at 6p21.33 and 20q11.21 were found to be associated with lung cancer risk in the Chinese population [44]. Variants of PARK2 (p.Arg257Trp) and HER2 (p.Gly660Asp) have been detected in familial lung cancer by whole-exome sequencing [45,46]. These results indicate that genotype imputation and whole-genome and/or whole-exome sequencing may facilitate the identification of rare variants with large effects on lung cancer risk.

The limitations and applications of GWAS

Genetic factors affecting lung cancer risk have been identified by GWAS and other association studies. The results of these studies indicate that the effect sizes for risk variants differ among populations for a variety of reasons, including differences in allele frequency and interactive effects between variants and the genetic and environmental backgrounds. Additional studies are needed to clarify rare disease-causing variants in a range of ethnic groups, particularly studies that use larger numbers of Asian and European subjects, i.e., via international and/or Intra Consortiums (ILCCO, the International Lung Cancer Consortium [47]; FLCCA, the Female Lung Cancer Consortium in Asia; and ILCS, Japanese Lung Cancer Collaborative Study) [11,16]. Another point to be clarified is interaction between genetic factors and exogenous/endogenous factors on lung cancer risk. Interaction of SNPs identified by GWASs with other cancer-related factors, such as smoking and age, is unclear. Recently, a GWAS focusing on gene-smoking interaction of cigarette smoke with lung cancer risk was performed in a Chinese population, but it did not yield a conclusive result [48]. Therefore, GWASs that consider such interactions should be further performed in several populations and/or in larger scales.

GWAS focusing on specific lung cancer types will be a powerful approach to identify additional genetic factors that are associated with the risk of specific lung cancer types, such as SCC (the deadliest histological type of lung cancer), lung cancers of female never-smokers, those of patients of young ages, or those with specific
oncogene mutations, such as EGFR and KRAS mutations. Understanding the underlying genetic factors will help greatly in clarifying the disease etiology and in identifying high-risk individuals for targeted screening and/or prevention based on a combination of genetic and environmental factors.

Acknowledgments

This work was supported in part by the Practical Research for Innovative Cancer Control from the Japan Agency for Medical Research and Development (15ck01060960002 and 15Ach01061680001), the Princess Takamatsu Cancer Research Fund, the Suzukan Memorial Foundation, and the National Cancer Center Research and Development Fund (26-A-8, 25-A-1 and NCC Biobank).

References