Lung cancer is a leading cause of cancer death and disease burden in many countries. Understanding of the biological pathways involved in lung cancer aetiology is required to identify key biomolecules that could be of significant clinical value, either as predictive, prognostic or diagnostic markers, or as targets for the development of novel therapies to treat this disease, in addition to smoking avoidance strategies. Genome-wide association studies (GWAS) have enabled significant progress in the past 5 years in investigating genetic susceptibility to lung cancer. Large scale, multi-cohort GWAS of mainly Caucasian, smoking, populations have identified strong associations for lung cancer mapped to chromosomal regions 15q (nicotinic acetylcholine receptor (nAChR) subunits: CHRNA3, CHRNA5), 5p (TERT-CLPTM1L locus) and 6p (BAT3-MSHS). Some studies in Asian populations of smokers have found similar risk loci, whereas GWAS in never smoking Asian females have identified associations in other chromosomal regions, e.g., 3q (TP63), that are distinct from smoking-related lung cancer risk loci. GWAS of smoking behaviour have identified risk loci for smoking quantity at 15q (similar genes to lung cancer susceptibility: CHRNA3, CHRNA5) and 19q (CYP2A6). Other genes have been mapped for smoking initiation and smoking cessation. In chronic obstructive pulmonary disease (COPD), which is a known risk factor for lung cancer, GWAS in large cohorts have also found CHRNA3 and CHRNAS single nucleotide polymorphisms (SNPs) mapping at 15q as risk loci, as well as other regions at 4q31 (HHIP), 4q24 (FAM13A) and 5q (HTR4). The overlap in risk loci between lung cancer, smoking behaviour and COPD may be due to the effects of nicotine addiction; however, more work needs to be undertaken to explore the potential direct effects of nicotine and its metabolites in gene-environment interaction in these phenotypes. Goals of future genetic susceptibility studies of lung cancer should focus on refining the strongest risk loci in a wide range of populations with lung cancer, and integrating other clinical and biomarker information, in order to achieve the aim of personalised therapy for lung cancer.

KEY WORDS
Lung cancer; genetics; pulmonary disease; chronic obstructive; genome-wide association study (GWAS)
Genetic susceptibility to lung cancer–study designs

The heritability of lung cancer susceptibility has been clearly established in numerous studies, including analyses of familial risk (8) and segregation analyses (9). However genetic influence on lung cancer is moderate at best. For example, using the 9.6 million subject Swedish Family-Cancer database, Czene et al. estimated heritability at 8% (10) and in twin studies and a higher concordance for monozygotic than for dizygotic twins has been noted (11). With tobacco smoking being by far the strongest environmental cause, it is possible that the heritable effects of genes governing smoking behaviour, [given the high heritability of smoking habits, ~0.5 in twin studies (12)], rather those determining individual susceptibility to carcinogenesis may play a more important role. However Lorenzo Bermejo et al. estimated the relative risk of lung cancer attributable to smoking according to the extent to which smokers transmit their smoking habits to the offspring (heritability of smoking), the prevalence of smoking in the general population, and the risk of lung cancer for smokers compared with non-smokers (13). They showed that the relative risk of lung cancer for the offspring of lung cancer patients attributable to smoking was 1.19 when published data on smoking practice were modelled suggesting familial cases of lung cancer cannot be attributed to shared smoking habits.

Given that there are apparent genetic determinants of lung cancer, there are a number of alternative study approaches that can be utilised to determine the genetic determinants of disease susceptibility.

**Linkage analysis** involves proposing a model to explain the inheritance pattern of phenotypes and genotypes observed in a pedigree (14). Linkage is evident when a gene that produces a phenotypic trait and its surrounding markers are coinherited. In contrast, those markers not associated with the anomalous phenotype of interest will be randomly distributed among affected family members as a result of the independent assortment of chromosomes and crossing over during meiosis. **Association studies** do not examine inheritance patterns of alleles; rather, they are case-control studies based on a comparison of allele frequencies between groups of affected and unaffected individuals from a population. A particular allele is said to be associated with the trait if it occurs at a significantly higher frequency among affected individuals as compared with those in the control group. The odds ratio of the trait in individuals is then assessed as the ratio of the frequency of the allele in the affected population compared with the unaffected population. The greatest problem in association studies is the selection of a suitable control group to compare with the affected population group. Genome-wide linkage analysis in family cohorts resulted in the identification of highly penetrant, low-frequency susceptibility genes for cancers, such as **BRCA1** and **BRCA2** for breast cancer and **APC** for colorectal cancer.

For lung cancer, several studies have attempted to identify susceptibility loci using a genome-wide linkage approach. However, while a few genetic loci that potentially harbour susceptibility genes have been identified, e.g., linkage of lung, laryngeal, and pharyngeal cancer in families to a region on chromosome 6q23-25 (15), no causal gene has been identified and, as with subsequent GWAS (see below), there is considerable overlap between the result for lung cancer and those for COPD, and lung function (16).

**GWAS**

Advances in array-based SNP genotyping technologies and haplotype mapping of the human genome (17) have presented the possibility of simultaneously determining millions of SNPs throughout the genome of an individual and this has allowed extension of association study design to allow hypothesis independent assessment of association across the genome. GWAS have revolutionized the study of genetic factors in complex common disease (18,19). For more than 200 phenotypes, from common diseases to physiological measurements such as height and BMI and biological measurements such as circulating lipid levels, GWAS have provided compelling statistical associations for over hundreds of different loci in the human genome (www.gwascentral.org/). There are now clearly established approaches for GWAS including stringent genotype calling, quality control, population stratification (genomic controls) and statistical techniques (20). Due to the large number of statistical tests undertaken, carefully controlling for multiple testing using Bonferroni or false discovery rate (FDR) corrections is essential. A cluster of P-values below the 1% FDR from SNPs in one chromosomal location is defined as the region of ‘maximal association’ and is the first candidate gene region to examine further with analysis of secondary outcome measures, gene database searches, fine mapping to find the causal locus and replication in other cohorts/populations. It is unlikely that the SNP showing the strongest association will be the causal locus, as SNPs are chosen to provide maximal coverage of other variation in that region of the genome and not on biological function. Once a candidate region or gene is identified, gene expression can be compared between a selection of cases and
controls and within individuals of different genotypes to provide further evidence for the genes involvement in disease. If linkage disequilibrium prevents the identification of a specific gene in a haplotype block then it may be necessary to utilize different racial and ethnic populations to hone in on the causative candidate gene that accounts for the genetic signal in GWAS (21).

Valuable insights into lung cancer susceptibility genes have been identified using the GWAS approach; however, the loci identified account for an extremely small proportion of the familial risk. The finding that loci identified through GWAS studies for common disease fail to account for all heritability of these conditions (termed missing heritability) is a subject of much discussion. There are a number of possible reasons that may account for this observation. These include gene-gene interaction, gene-environment interaction, and other types of genetic variation such as rare variants and structural variation and epigenetic heritability. In the future, the analysis of genome-wide copy number variation and/or rare variants through exome- or whole-genome sequencing, as is being applied to other complex diseases, may identify further loci responsible for inherited susceptibility to lung cancer (22). However one approach is to utilise novel analytical approaches to identify weakly associated variants whose effect may be lost in the GWAS approach due to the stringent significance level after multiple comparison correction. For example, utilisation of a less stringent multiple correction followed by gene pathway analysis can highlight genes involved in common biological pathways in the ‘grey area’ of SNPs whose association with disease status lies below the conventional level of genome-wide significance. Using a similar approach Zhang et al. performed a two-stage pathway analysis in GWAS of lung cancer in Han Chinese using gene set enrichment analysis (GSEA) method. Four pathways (achPathway, metPathway, At1rPathway and rac1Pathway) were consistently significant and may provide new insights into the etiology of lung cancer (23).

**GWAS of lung cancer**

A number of GWAS have now been performed in a range of populations, to test genetic influences in susceptibility to lung cancer (Table 1).

An initial, relatively small GWAS was reported in an Italian population, showing association with SNPs in the KLF6 gene, but not in replication cohorts (24). Another relatively small study of patients with familial lung cancer found associations with SNPs at chromosomal region 15q (27). Larger scale GWAS have since been performed in Caucasian populations, with replication cohorts (25,26,28-30,32) and meta-analyses (31,32,42). These GWAS have found statistically significant associations with SNPs, particularly in chromosomal regions 15q, 5p and 6p (Table 1). GWAS have also been undertaken in Asian populations with lung cancer (35-37,39,41), identifying some similar SNPs as detected in the studies of Caucasian populations, but also finding other SNPs conferring lung cancer risk distinctly in Asian populations (44).

In many of these studies, the observed associations with key SNPs were independent of smoking status or smoking history (25,26,29,32), although in some studies, SNPs (e.g., on 15q) were related to smoking behaviour (28). In studies of smokers, population attributable risk (PAR) for lung cancer for these SNPs, where calculated, were modest, ranging between 14% (26) and 18% (28), compared to the overwhelming PAR of >80% from tobacco smoking.

**GWAS in specific lung cancer populations**

GWAS have also been undertaken in never smokers, and also to address other lung cancer-related questions. A GWAS of lung cancer in never smokers in the USA, with replication cohorts, identified an association with SNPs at the 13q region, in the GPC5 gene (34). Studies have also examined genome-wide association in never smokers in Asian populations (38,40,43), finding new susceptibility loci that were different to the loci found for Caucasian populations. SNPs in the TERT gene on 5p15 were associated with lung adenocarcinoma as a specific lung cancer histology (32), whereas SNPs at 9p21 are associated with risk for lung squamous cell carcinoma (SCC) (42). Genome-environment interaction was tested in a GWAS of lung cancer risk and self-reported asbestos exposure (45). Whilst this pilot study was not sufficiently powered to find significant differences, a suggested gene-asbestos exposure interaction was seen for SNPs in C7orf54 on 7q32. In addition, a number of GWAS have identified SNPs that predict response to chemotherapy in patients with SCLC (46) and NSCLC (47), and other GWAS have explored SNPs related to prognosis and survival in patients with lung cancer [e.g., (48)].

**Candidate genes for lung cancer susceptibility from GWAS**

The discovery and replication studies from the GWAS (Table 1), and other replication studies since, have identified emerging patterns in candidate genes for lung cancer susceptibility [reviewed in (6,7,42)]. Consistent candidate genes for Caucasian smoking populations have been the neuronal nicotinic acetylcholine receptor (nAChR) subunits (cholinergic receptor, nicotinic, alpha 3 and 5: CHRNA3 and CHRNA5) at 15q25. Neuronal nAChRs are activated by acetylcholine or nicotine, and comprise subunits (pentamers) of α and β subunits. In the lungs, nAChRs are expressed in neurons, and also non-neuronal cells, including bronchial epithelial cells and lung cancer cells.
<table>
<thead>
<tr>
<th>Study</th>
<th>Lung cancer cases (discovery set)</th>
<th>Controls (discovery set)</th>
<th>Arrays [nos. of SNPs]</th>
<th>Chromosomal regions and main associated genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spinola 2007 (24)</td>
<td>335 smokers</td>
<td>338 smokers</td>
<td>Affymetrix [116,204]</td>
<td>10p KLF6</td>
</tr>
<tr>
<td>Amos 2008 (25)</td>
<td>1,154 smokers</td>
<td>1,137 smokers</td>
<td>Illumina [317,498]</td>
<td>15q CHRNA3</td>
</tr>
<tr>
<td>Hung 2008 (26)</td>
<td>1,989 smokers</td>
<td>2,625 smokers</td>
<td>Illumina [317,139]</td>
<td>15q CHRNA3, CHRNA5</td>
</tr>
<tr>
<td>Liu 2008 (27)</td>
<td>194 with familial lung cancer</td>
<td>219 smokers and non-smokers</td>
<td>Affymetrix [500,568 and 906,703]</td>
<td>15q various genes</td>
</tr>
<tr>
<td>Thorgeirsson 2008 (28)</td>
<td>1,024 smokers</td>
<td>32,244 controls</td>
<td>Illumina [306,207]</td>
<td>15q CHRNA3</td>
</tr>
<tr>
<td>McKay 2008 (29)</td>
<td>3,259 smokers</td>
<td>4,159 smokers</td>
<td>Illumina [315,194]</td>
<td>5p TERT-CLPTM1L, 15q CHRNA3</td>
</tr>
<tr>
<td>Wang 2008 (30)</td>
<td>1,952 smokers</td>
<td>1,438 smokers</td>
<td>Illumina [511,919]</td>
<td>5p CLPTM1L, 6p BAT3-MSH5, 15q CHRNA3</td>
</tr>
<tr>
<td>Broderick 2009 (31)</td>
<td>1,978 smokers, and meta-analysis</td>
<td>1,438 smokers, and meta-analysis</td>
<td>Meta-analysis</td>
<td>5p TERT-CLPTM1L, 6p BAT3-MSH5, TNXB, 15q CHRNA3</td>
</tr>
<tr>
<td>Landi 2009 (32)</td>
<td>5,739 smokers</td>
<td>5,848 smokers</td>
<td>Illumina [515,922]</td>
<td>5p TERT-CLPTM1L, 15q CHRNA3</td>
</tr>
<tr>
<td>Hsiung 2010 (33)</td>
<td>584 cases (never smoking females with lung adenocarcinoma)</td>
<td>585 controls (never smoking females)</td>
<td>Illumina [610,901]</td>
<td>5p15 TERT-CLPTM1L</td>
</tr>
<tr>
<td>Li 2010 (34)</td>
<td>377 never smokers</td>
<td>377 never smokers</td>
<td>Illumina [373,397 and 592,532]</td>
<td>13q31.3 GPC5</td>
</tr>
<tr>
<td>Miki 2010 (35)</td>
<td>1,004 with lung adenocarcinoma</td>
<td>1,900 controls</td>
<td>Illumina [610,901]</td>
<td>3q28 TP63, 5p15 TERT</td>
</tr>
<tr>
<td>Yoon 2010 (36)</td>
<td>621 cases (smokers and never smokers)</td>
<td>1,541 controls (smokers and never smokers)</td>
<td>Affymetrix [500,568]</td>
<td>3q29 C3orf21, 5p TERT-CLPTM1L</td>
</tr>
<tr>
<td>Hu 2011 (37)</td>
<td>2,331 cases (smokers and never smokers)</td>
<td>3,077 controls (smokers and never smokers)</td>
<td>Affymetrix [906,703]</td>
<td>3q28 TP63, 5p15 TERT-CLPTM1L, 13q12, MPEP, TNSRF19, 22q12 MTMR3-HORMAD2-LIF</td>
</tr>
<tr>
<td>Ahn 2012 (38)</td>
<td>446 never smokers</td>
<td>497 healthy controls</td>
<td>Affymetrix [906,703]</td>
<td>18p11 FAM388</td>
</tr>
<tr>
<td>Dong 2012 (39)</td>
<td>833 cases with SCC</td>
<td>3,094 controls</td>
<td>Affymetrix [906,703]</td>
<td>12q23 SLC17A8-NR1H4</td>
</tr>
<tr>
<td>Lan 2012 (40)</td>
<td>5,510 never-smoking female lung cancer cases</td>
<td>4,544 controls</td>
<td>Various</td>
<td>3q28 TP63, 5p15, 6p21 HLA, 6q22 ROS1, DCBLD1, 1q25 VT11A, 17q24 BPTF</td>
</tr>
<tr>
<td>Shiraishi 2012 (41)</td>
<td>1,722 cases (smokers and never smokers)</td>
<td>5,846 controls (smokers and never smokers)</td>
<td>Illumina [709,857]</td>
<td>3q28 TP63, 5p15 TERT, 6p21 BTNL2, 17q24 BPTF</td>
</tr>
<tr>
<td>Timofeeva 2012 (42)</td>
<td>Meta-analysis: 14,900 cases (smokers and never smokers)</td>
<td>29,485 controls (smokers and never smokers)</td>
<td>Various</td>
<td>5p15, 6p21, 15q25 for NSCLC; 9p21 for SCC</td>
</tr>
<tr>
<td>Kim 2013 (43)</td>
<td>285 female never smokers with lung cancer</td>
<td>1,455 controls</td>
<td>Affymetrix [440,794]</td>
<td>2p16 NRXN1</td>
</tr>
</tbody>
</table>

* In Tables 1 to 3, discovery study details have been included, and replication study samples sizes have not been included. For details, see http://www.genome.gov/gwastudies.
Whilst SNPs in nAChRs may alter risk of lung cancer through smoking behaviour, these SNPs could also regulate direct effects of nicotine, through anti-apoptotic and proliferative effects, or effects of nicotine-derived carcinogens in tobacco smoke (6,7,49-51).

On the 5p15 locus, SNPs in the TERT and CLPTM1L genes have been associated in a number of GWAS of lung cancer and other cancers (52). The TERT gene encodes human telomerase reverse transcriptase, which is important in the maintenance of telomere length. CLPTM1L (cleft lip and palate transmembrane protein 1-like protein) may induce apoptosis in lung cells (29).

The identified SNPs in TERT have generally been intronic, and in linkage disequilibrium with SNPs in CLPTM1L. On 6p21, BAT3 and MSH5 have emerged as signals in a number of GWAS. BAT3 (renamed BAG6, BCL2-associated athanogene 6) encodes a nuclear protein involved in DNA damage-induced apoptosis and modulation of p53 in response to genotoxic stress (7). The MSH5 [mutS homolog 5 (E. coli)] gene is involved in DNA mismatch repair.

Whilst the 5p15 SNPs have demonstrated replication in both Caucasian and Asian populations, this has not been the case for the 15q SNPs (40,53). This discordance may represent differences in lung cancer aetiology between smoking and never smoking populations (particularly where indoor air pollution from biomass fuels may be the predominant carcinogen). In the never smokers, especially in Asian female populations, other candidate genes emerge, for example, that are involved in receptor tyrosine kinase activity [ROSI, c-ros oncogene 1, receptor tyrosine kinase (40)], vesicle transport [VTI1A, vesicle transport through interaction with t-SNAREs 1A (40)] and cell adhesion [NRXN1, neurexin 1 (43)]. Unexpectedly, the transcription factor TP63 (tumor protein p63) is also a candidate marker in the never smoking populations. TP63 encodes a protein which is often used as an immunohistochemical marker of squamous cell carcinoma, a cancer strongly associated with tobacco smoking (54,55).

Many of these associations are novel for lung cancer susceptibility, and were not detected in previous candidate gene studies of lung cancer, which focused on metabolising enzymes [e.g., CYP1A1 (56-58)], oxidative stress pathways and other DNA repair mechanisms (7).

GWAS and lung cancer pharmacogenetics

The GWAS approach is now being extended to examine other related phenotypes in lung cancer. For example, Han et al. recently undertook a GWAS of survival in small-cell lung cancer patients treated with irinotecan plus cisplatin chemotherapy, and identified candidate SNPs that may be predictive of the clinical outcome (59).

### GWAS of smoking behaviour

Of relevance to lung cancer aetiology, GWAS of smoking behaviour have been performed in large population cohorts (Table 2), and have focused on smoking initiation, smoking quantity (cigarettes per day) and success of smoking cessation. The interest for lung cancer susceptibility is not only for causation from smoking, but also for SNPs that are common to both smoking behaviour and lung cancer, that could have direct biological effects.

#### Smoking onset

Onset of smoking has been associated with the 11p region, with BDNF (brain-derived neurotrophic factor), a neurotropin, identified as a possible candidate gene (61). Other regions identified with age of smoking initiation in a study of patients with COPD were 6p21 (HLA) and 2q21 (intergenic region) (63).

#### Smoking quantity

The intensity of smoking (smoking quantity) has been consistently associated in GWAS with SNPs in the 15q25 region (60-65), at the loci of nAChR genes (especially CHRNA3 and
S459

CHRNA5). Other loci include 19q (CYP2A6, a cytochrome P450 nicotine metabolising enzyme) (61,63,66) and 8p (CHRN3B3, a nAChR subunit) (62).

Smoking cessation

A range of biopsychosocial factors influence an individual’s ability to abstain from smoking (67). Genetic factors for smoking cessation identified from GWAS have centred around SNPs in the 9q region, including DBH (dopamine beta-hydroxylase) (61,63), which is involved in the metabolism of dopamine.

GWAS of COPD

Lung cancer and COPD frequent coexist in at-risk smokers. Epidemiological evidence supports an association between the presence of COPD and increased risk of developing lung cancer. Common mechanisms for susceptibility to lung cancer and COPD, in addition to tobacco smoke, may involve biological processes such as inflammation, epithelial-mesenchymal transition, abnormal repair, oxidative stress, and cell proliferation. In addition, genomic and epigenomic changes are likely to alter biological pathways, leading to susceptibility to lung cancer and COPD (68,69). Therefore, studying genetic influences in COPD could yield greater insight into the shared pathogenesis of lung cancer and COPD. Importantly, genetic epidemiological principles should be considered when designing and interpreting studies of COPD and lung cancer, because of the shared aetiological and possibly genetic factors (70).

GWAS of susceptibility to COPD

A number of GWAS have been performed for COPD (Table 3), albeit a smaller number of studies than the lung cancer GWAS to date.

GWAS have so far been undertaken in the Framingham cohort (71); Bergen (Norway) COPD Cohort (with replication in other cohorts) (72); combined Bergen cohort with National Emphysema Treatment Trial (NETT) and Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) study subjects (73); combined four cohorts [ECLIPSE, Normative Aging Study, Bergen (Norway) COPD Cohort and COPD Gene] (74); and a combination of 15 cohorts (75).

Candidate genes for COPD from GWAS

Several regions significantly associated with COPD have emerged from the COPD GWAS to date (Table 3). At the 15q locus (similar to the GWAS for lung cancer and for smoking behaviour), SNPs in the nAChR subunit genes (CHRNA5, CHRNA3) were associated with COPD (72,75), possibly indicating a link with smoking intensity as an aetiological factor, although direct effects should also be considered (76). Similarly, association at 19q13 may be related to smoking behaviour [e.g., CYP2A6 (74)].

Other novel associations have been found for COPD in these GWAS. At 4q31, hedgehog interacting protein (HHIP) has been identified as a candidate in two studies (71,72). HHIP encodes a membrane glycoprotein that is an inhibitor of hedgehog signalling, which is involved in development processes. Gene expression studies in BEAS-2B bronchial epithelial cell lines implicate HHIP in extracellular matrix and cell proliferation (77). At 4q24, FAM13A (family with sequence similarity 13, member A) has been detected in two GWAS (73,74), and also in a genetic association in which SNPs were related to both COPD and lung cancer, indicating a possible shared genetic pathway (78). FAM13A contains a Rho GTPase-activating protein-binding domain, inhibits signal transduction and responds to hypoxia; however, its full function in the lung remains to be determined. At the 5q region, HTR4 [5-hydroxytryptamine (serotonin) receptor 4] was associated with COPD in smokers; its function in airways disease may involve regulation of cytokine release (75).

Table 3. GWAS of COPD vs. controls.

<table>
<thead>
<tr>
<th>Study</th>
<th>COPD cases (discovery)</th>
<th>Controls (discovery)</th>
<th>Arrays (nos. of SNPs)</th>
<th>Chromosomal regions and main associated genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilk 2009 (71)</td>
<td>7,691 Framingham study participants, plus replication cohort</td>
<td>Affymetrix (550,000)</td>
<td>4q31 HHIP</td>
<td></td>
</tr>
<tr>
<td>Pillai 2009 (72)</td>
<td>823 COPD</td>
<td>810 smokers</td>
<td>Illumina (561,466)</td>
<td>4q31 HHIP, 15q CHRNA3, CHRNA5</td>
</tr>
<tr>
<td>Cho 2010 (73)</td>
<td>2,940 COPD</td>
<td>1,380 smokers</td>
<td>Various (&gt;500,000)</td>
<td>4q24 FAM13A</td>
</tr>
<tr>
<td>Cho 2012 (74)</td>
<td>3,499 COPD</td>
<td>1,922 controls</td>
<td>Illumina (&gt;500,000)</td>
<td>4q24 FAM13A, 19q13 RAB4B, EGLN2, CYP2A6</td>
</tr>
<tr>
<td>Wilk 2012 (75)</td>
<td>3,368 COPD, plus replication cohort</td>
<td>29,507 controls</td>
<td>Various (&gt;500,000)–meta-analysis</td>
<td>5q HTR4, 15q AGPDH1, IREB2, CHRNA5, CHRNA3</td>
</tr>
</tbody>
</table>
Summary and clinical implications

Main findings to date from GWAS

Whilst lung cancer is predominantly caused by cigarette smoking, a genetic component to susceptibility is well-recognised in epidemiological studies. GWAS have now been completed in lung cancer and the related phenotypes of smoking behaviour and COPD. Large scale, multi-cohort GWAS of lung cancer in mainly Caucasian, smoking populations have identified strong associations for lung cancer mapped to chromosomal regions 15q (CHRNA3, CHRNA5), 5p (TERT-CLPTM1L locus) and 6p (BAT3-MSH5). Some studies in Asian populations of smokers have found similar risk loci, whereas other GWAS, particularly in never smoking Asian females, have identified associations in other chromosomal regions that are distinct from the smoking-related genetic loci. GWAS of smoking behaviour have identified risk loci for smoking quantity at 15q (similar genes to lung cancer susceptibility: CHRNA3, CHRNA6) and also at 19q (CYP2A6). Other genes have been mapped for smoking initiation and smoking cessation. In COPD, GWAS in large cohorts have also found NACHR SNPs mapping at 15q as risk loci, as well as other regions at 4q31 (HHIP), 4q24 (FAM13A) and 5q (HTR4). The overlap in risk loci between lung cancer, smoking behaviour and COPD may be due to the effects of nicotine addiction; However, more work needs to be undertaken to explore the potential direct effects of nicotine and its metabolites in gene-environment interaction in these phenotypes.

Applications and future directions

From the evidence presented to date, GWAS have been useful not only in addressing genetic influences in lung cancer susceptibility, but also gene-environment interaction in terms of smoking as causation, as well as COPD as a risk factor for lung cancer. The translation of findings from the lung cancer and related GWAS could in the future enable profiling of an individual’s risk of lung cancer, biomarkers for diagnosis, and markers for prognosis and therapy. The challenge now will be to combine genomics (79), epigenomics (80), transcriptomics (81-83) and proteomics profiling to improve the management of patients with lung cancer and related comorbidities (68). Future studies should include DNA sequencing of lung tumours and lung tissue, and network analysis of genomic information and clinical phenotypes (3, 84, 85). Goals of future genetic susceptibility studies of lung cancer should now be focused on refining the strongest risk loci in a wide range of populations with lung cancer, and integrating other clinical and biomarker information, in order to achieve the aim of personalised therapy for lung cancer (4), through enhanced diagnosis, prognosis, prevention and management.

Acknowledgements

This work was supported by Asthma, Allergy & Inflammation Research (AAIR) (JWH), NHMRC Career Development Fellowship (IY), NHMRC Practitioner Fellowship (KF), and project grants from NHMRC, Queensland Health and Medical Research, The Prince Charles Hospital Foundation, and Cancer Council Queensland.

Disclosure: The authors declare no conflict of interest.

References