The majority of lung cancers are attributable to tobacco usage and other behavioral and environmental risk factors. Tobacco-induced malignancies are fueled by multiple powerful carcinogens and nicotine, a potent addictive agent. Thus, lung cancer is often regarded as being primarily determined by behavioral and environmental factors. However, because only a minority of heavy smokers develops lung cancer, genetic differences in susceptibility to tobacco carcinogens must exist, and lung cancers must arise from complex
gene–environment interactions (1). Up to 25% of lung cancers arise in lifetime never-smokers and are more frequent in patients of female sex or East Asian ethnicity or with adenocarcinoma histology (2). Environmental tobacco smoke is a weak carcinogen and cannot account for the majority of these cases [reviewed in (2)]. Thus, lung cancers in ever-smokers and never-smokers may have both common and distinct risk factors and gene–environment interactions.

The Holy Grail for medical genetics is to have a single, simple, relatively inexpensive genetic test for risk assessment of multiple diseases, including cancer. Molecular epidemiology arose as a scientific discipline to study gene–environment interrelationships by comparing affected subjects with control subjects (3). Initially, these case–control studies were hypothesis driven and used functional assays of suspect genes, often in surrogate tissues. However, many such studies have yielded little useful information about low- and moderate-penetrance genes (4), indicating the need for improved search techniques. Approximately 100 genes cause a smaller number of cancer syndromes with a Mendelian inheritance pattern, but these genes only explain a minor part of the familial clustering of the common cancers (5). Linkage analyses of high-risk families may identify other rare high-penetrance genes, and such studies have identified a lung cancer susceptibility locus on chromosome 6q (6). It is now believed that alleles with high frequency (typically >1%) and low penetrance (typically less than twofold increased lifetime risk) contribute substantially to susceptibility to many diseases, including cancers (5).

The completion of the Human Genome Project in 2003 and the subsequent International Haploid Mapping (HapMap) Project (http://hapmap.ncbi.nlm.nih.gov) provided a huge bonanza for risk association studies. These projects cataloged a large number of single-nucleotide polymorphisms (SNPs) in the human genome. SNPs are DNA sequence variations that occur when a single nucleotide differs between individuals at minor allele frequencies greater than 1%. Typically, a SNP occurs every few hundred base pairs along the 3-billion-bp genome. The National Human Genome Research Institute of the National Institutes of Health (http://www.genome.gov) estimates that there are 10 million or more SNPs in the human genome, and some SNPs demonstrate ethnic differences. Sets of nearby SNPs tend to be inherited in combination (linkage disequilibrium). The HapMap is a map of these nonrandomly distributed haplotypes, and the specific SNPs that identify the haplotypes are called tag SNPs. The HapMap permits the entire genome to be examined for associations with a phenotype by using approximately 500 000 tag SNPs. Starting in 2005, SNP-based genome-wide association studies have provided a powerful tool to the molecular epidemiologist because the whole genome could be scanned with high-density SNP microarrays, thus allowing previously unsuspected associations to be identified; in other words, genome-wide association studies are agnostic in terms of prior assumptions and hypothesis generating as opposed to being hypothesis driven (7). Previous genome-wide association studies identified lung cancer susceptibility loci at 15q25, 6p21, and 5p15.33, providing powerful evidence of a genetic contribution to lung cancer [as discussed in (8)]. In this issue of the Journal, Truong et al. (9) report a large, coordinated study they conducted to independently confirm the findings of these earlier genome-wide association studies.

However, SNP studies may yield misleading or contradictory results because of a poor design, a lack of knowledge about the biology of the experimental system, being unpowered, publication bias, a lack of ethnic diversity among subjects, or other reasons (3). Thus, results of genome-wide association studies need to be confirmed either by meta-analyses of the published literature or, preferably, by large independent studies. The National Cancer Institute has just established the Transdisciplinary Cancer Genomics Research Post-Genome Wide Association Initiative. The Initiative will consist of five research centers that will collaborate to rapidly move forward promising leads from initial cancer genome-wide association studies by replicating and extending prior findings, by better pinpointing genomic regions that cause cancer, by conducting functional studies, and by understanding gene–environment interactions to set the stage for translating these findings into clinical and prevention applications.

In the confirmatory study by Truong et al. (9) for lung cancer, six tag SNPs (two each from the three previously identified susceptibility loci) were studied in more than 26 000 lung cancer case and control subjects. This study has resulted in several important observations that drive home many of the lessons that can be learned from genome-wide association studies, with specific relevance for lung cancer. First, for confirmatory genome-wide association studies, such as that by Truong et al. (9), linkage disequilibrium permits the selection of only the most statistically significant and robustly reproducible tag SNPs. Second, the necessity to study large numbers of case and control subjects almost always requires data pooled from many cooperating centers. Truong et al. (9) pooled data from 21 case–control studies involving more than 11 000 lung cancer patients and nearly 15 000 control subjects. They used resources made available from the International Lung Cancer Consortium (http://ilcco.iarc.fr), an organization that was established for fostering such studies. Third, findings from genome-wide association studies have to be confirmed. For specific subgroups of subjects, the well-described association with 15q25 was validated, as was the more recently identified association with 5p15, but not the link with 6p21 (9). Fourth, inclusion of both sexes and other ethnicities is important. Most genome-wide association studies of lung cancer have been performed on Caucasian populations (10); however, lung cancer in East Asians is a very different disease (2). The confirmatory study (9) was well balanced with regard to sex, and 15% of the study subjects were East Asians. The study failed to reveal a statistically significant association between 15q25 and the risk of lung cancer in East Asian subjects. Therefore, despite the large overall sample size, the percentage of Asian subjects might not have been large enough to detect a weak association in this ethnic group, possibly because of the high proportion of never-smokers and ethnic differences in allele frequency (10).

Fifth, the finding of genes located within the risk-associated regions that may play a role in cancer pathogenesis offers a rational basis for lung cancer susceptibility. For example, within the 5p15.33 region (which contains two of the SNPs examined by Truong et al.) lie two genes, one of which is the telomerase reverse transcriptase gene, TERT. TERT is activated in most cancers and
is frequently amplified in some, including lung adenocarcinomas. An endless replicative capacity (via telomerase activation) is one of the hallmarks of cancer (11), and thus, it is not surprising that this region is associated with risk for multiple cancer types (12), including lung cancer in smokers and never-smokers. As a result, the search for lung cancer genes by genome-wide association studies has generated results with immediate relevance to lung carcinogenesis and potentially to new therapeutic approaches. This has not been the case in genome-wide association studies for most other cancers, in which many variants of unknown biological significance have been identified (13).

Sixth, for complex heterogeneous cancers such as lung cancer, analysis by pathological subtype is important. The major forms of lung cancer (small cell, squamous cell, and adenocarcinoma) demonstrate important differences related to sex, geographic distribution, and smoke exposure. Truong et al. (9) found that the association between the 5q15 loci and adenocarcinoma was strongest, as have previous studies (14).

Seventh, as discussed previously (2), lung cancers arising in ever- and never-smokers are very different diseases. Except in earlier, possibly underpowered, studies (15–17), the association between 15q25 and the risk of lung cancer applied only to smokers. This region contains a cluster of genes involved in nicotine addiction. In addition to predisposing to nicotine dependence and increased tobacco consumption, variants at 15q25 may play a direct role in lung carcinogenesis (9,18). Thus, they may affect both the end organ (lung epithelium) and nicotinic receptors in the brain. Although SNPs at this locus were not found to be associated with lung cancer risk in never-smokers, they are associated with risk of other smoking-associated cancers and diseases (9,19). Of interest, a recent study found that genetic variants at 13q31.3 are associated with susceptibility to lung cancer in never-smokers and alter the expression of the glypican 5 (GPC5) gene, a gene involved in cell division and growth regulation (20), providing further evidence that the genetic factors for risk in smokers and never-smokers may be different.

The very interesting findings of Truong et al. (9) about susceptibility to lung cancer represent a major step in our understanding of the genetics of lung cancer. However, we need additional well-designed studies before the findings can be translated to yield clinical benefits. There is a complex relationship between genotype and phenotype (21), and this relationship needs to be investigated for suspected susceptibility genes. The variants identified thus far probably explain only a small proportion of the heritability of a complex disease such as lung cancer, and methods to combine hypothesis-driven pathway-based analyses with agnostic genome-wide association studies need to be explored (22). Until recently, SNPs were believed to represent the vast majority of human genetic variation. However, structural variants, especially copy number variants (ie, insertions, deletions, and duplications), are important contributors to human diversity (23). Studies that have examined copy number variation, although modest in number, have demonstrated associations with several disease states (24), and such studies should be extended to cancers.

The widespread application of genome-wide association studies has opened new horizons for exploration and has highlighted the complex genomic architecture of disease susceptibility (13). Although SNP assays using dense microarrays are technically simple to perform and relatively cheap (<$1000), genome-wide association studies require sophisticated biostatistics for correct interpretation and must be complemented by comprehensive data about lifestyle and environmental exposures and uniform procurement, processing, and assaying of specimens. Perhaps global genomic interrogation, such as by SNP analyses that are possibly combined with copy number estimates, may eventually lead to the Holy Grail.

Between 80 and 90 years ago, James Ewing proposed two somewhat controversial concepts [as quoted in (18, 25)] that cancer was linked to hereditary and that increasing lung cancer rates were attributable to smoking. Subsequent investigations have confirmed both of these insightful beliefs and, more remarkably, have linked them.

References


**Notes**

Although A. F. Gazdar and P. Boffetta are members of the International Lung Cancer Consortium and have interacted or collaborated with several of the authors of the study by Truong et al. (9), they played no part in the design, performance, data interpretation, or writing of that report.

We thank John Minna for helpful comments.

**Affiliations of authors:** Hamon Center for Therapeutic Oncology Research and Department of Pathology, UT Southwestern Medical Center, Dallas, TX (AFG); The Tisch Cancer Institute, Mount Sinai School of Medicine, New York City, NY (PB).