During the “candidate gene era,” many studies with a small number of cases investigated a small number of variants in a limited number of genes chosen on the basis of partial knowledge of their function. These studies established only a few genetic risk factors for cancer (1) and other diseases. Remarkably, recent genome-wide association (GWA) studies have produced strong evidence that variation in a number of genomic regions affects risk of disease (2). Successful GWA studies most often used an agnostic approach to interrogating as many of the more common genetic variants as feasible (up to a million now) and large numbers of cases. They adopted a de facto criterion for genome-wide statistical significance of P values below $10^{-7}$ (3), thus guaranteeing a low chance of reporting a false positive even for quite low prior probabilities of association and weak effects (4), and they used statistical methods, such as principal components analysis (5), which protects not only against population stratification but also against poor control selection (6).

GWA studies are the culmination of advances in laboratory methodologies for detecting genetic variation, genomics, statistics, and informatics, yet they leave important scientific gaps (7). They can establish the presence of association but, by themselves, can identify neither the causal variants nor their function. Currently, they do not have power to capture the effects of the same variants that are most difficult to identify in family studies: variants with very small effects and low minor allele frequencies.

GWA studies of lung cancer are an instructive model for the study of gene the variant should not affect risk of lung cancer in those variants with small effects and low minor allele frequencies and behavioral determinants of disease because of the dominant role of smoking in the etiology of this disease. Like GWA studies of cancers of the breast and prostate, the first three GWA studies of lung cancer (8–10) identify several genetic variants strongly associated with disease. But a GWA study of lung cancer without smoking information cannot distinguish among three possibilities

Intermediacy and Gene–Environment Interaction: The Example of CHRNA5-A3 Region, Smoking, Nicotine Dependence, and Lung Cancer

Sholom Wacholder, Nilanjan Chatterjee, Neil Caporaso
for a genetic variant found to be strongly associated with disease but about which nothing was previously known: 1) the variant increases risk of lung cancer solely through effects on smoking behavior; 2) the variant induces carcinogenesis through a molecular mechanism relevant to lung cancer without affecting smoking behavior; or 3) the variant affects both carcinogenesis and smoking behavior. Indeed, the essence of the controversy over the role of the chromosome 15q24/25.1 or CHRN45-A3 region in lung cancer risk is the role of smoking: Does the association of 15q24/25.1 with nicotine dependence (11,12) arise because smoking is an intermediate factor, sometimes called a mediator or endophenotype, in the causal path between the genetic variant and disease? Does 15q24/25.1 itself affect lung carcinogenesis directly? Are both possibilities correct?

With the knowledge about the possible role of 15q24/25.1 on smoking behavior and the smoking data (11,12), the authors of the GWA studies (8–10) offer differing etiologic interpretations of the finding that 15q24/25 was associated with lung cancer. Now, following the spirit of the advice of Chanock and Hunter (13), Spitz et al. (14) use new data to argue for the third choice mentioned above, that variant rs1051370 in the 15q24/25.1 region affects both smoking behavior and risk of lung cancer. Their analysis focuses on direct effects of the variant on measured smoking phenotypes in approximately 3600 case and control subjects “from the same source as their GWA population.” They find strong evidence for a small effect on smoking intensity, as measured by cigarettes smoked per day, and provide further support for an effect on nicotine dependence (9,10,15), as measured by the Fagerström Test of Nicotine Dependence, but only weak or no supporting evidence for an association with other smoking phenotypes: age at initiation of smoking, duration of smoking, and sustained smoking cessation. The lack of association between the locus and bladder and renal cancers in a combined case set (14) and head and neck cancers (8), without adjustment for smoking, is consistent with a specific effect on lung carcinogenesis, but provides no further support for a more general effect on cancers related to tobacco. Overall, the receptor’s ability to bind nicotine and, possibly, downstream carcinogens (8,16) may account in part for the seeming inconsistency of the results for various endpoints in Spitz (14) and earlier studies (8–10). For example, some laboratory and population data suggest both central (an impact of smoking mediated at least in part by an effect on nicotine dependency) and peripheral (an effect on lung carcinogenesis), mechanisms as evidenced by expression of the nicotinic acetylcholine receptors in bronchial cells (17) and that the receptors are ligands for tobacco-specific carcinogens (18).

If the only effect of a variant on disease is through nicotine dependence, there should be no effect of the variant on lung cancer in those without exposure to smoking. In never smokers, Spitz et al. (14) saw no evidence of increased risk in carriers of the variant associated with lung cancer in smokers. Unmeasured risk factors for smoking and for lung cancer, however, can distort the stratified analysis (19). Measurement errors in identifying the functional variants and in reports on smoking are perhaps more important sources of bias than confounding (20). For instance, carriers and noncarriers of a variant may metabolize tobacco differently or behave differently, so that the carcinogenic dose from smoking varies by carrier status due perhaps to depth of inhalation, even if the reports of history of cigarettes smoked per day were perfectly accurate. Also, a GWA study cannot determine the specific variant or variants in the region of the tagging SNP associated with smoking and whether the same SNP or SNPs, or one or more additional proximal variant are associated directly with disease, especially in a region of high LD like 15q24/25.

Despite impressive work on 15q24/25.1 and lung cancer (8–10,14), we are not close to understanding the precise mechanisms underlying the genotypic association. The lack of statistically significant associations of the region with other cancers (8,14) and with lung cancer in nonsmokers (14) does not establish lack of association, just as a statistically significant result does not prove association. Even a convincing demonstration that there is no effect of a variant on lung cancer in nonsmokers does not imply that the variant acts on disease solely by its effect on addiction to smoking; we expect the noncausal effect of a consequence of smoking, like tobacco-stained teeth, on lung cancer would also to disappear on adjustment for smoking. Variants in a gene encoding a metabolic activity that affects carcinogenicity of tobacco components also will have no effect in nonsmokers, under the assumption that there is no effect from passive smoking or exposure to other carcinogens that bind to the receptors. Or a metabolic variant might affect behavior by providing smokers with pleasant or unpleasant feedback—like ALDH2 variants for alcohol (21)—rather than be related to addiction.

Table 1 lists several other examples where there are similar fundamental questions about the role of genes and behavioral or endogenous factors in cancer etiology. In each example, intermediary, where factor A greatly influences factor B, itself a cause of outcome C, but A does not affect C except through B, is plausible. Although intermediary is a special case of interaction, which addresses how risk factors act together to cause disease, it is not always considered in theoretical discussions (23,24).

Statistical approaches (25–28) may help distinguish among potential mechanisms. They can help articulate the assumptions and define precisely what can be estimated from standard multivariate analyses as well as alternative methods to address these questions. Both epidemiological (19,22,28,29) and clinical trial (30) literature discuss the statistical problem of whether and how much of the causal effect of an exposure or treatment on an outcome is mediated through an intermediate variable. Causal modeling frameworks, including theory of counterfactual outcomes (31,32) and directed acyclic graphs (19,22,33), can examine the complex interplay among an exposure, a possible intermediate and an outcome. These models can also be useful to understand the precise assumptions needed and the pitfalls, for example, unmeasured confounders, in the standard stratified or multivariate regression methods to measure the “direct” effect of an exposure on an outcome that is not mediated through an intermediate. We anticipate that in future such causal modeling framework will provide more insight into the interrelationship between 15q, smoking, and risk of lung cancer and other examples in Table 1.

The agnostic approach underpinning the design and analysis of GWA studies is integral to their success. Now we face the challenge of unraveling the meaning of the associations we have discovered and leveraging their findings. For the chromosome 15q–smoking–lung cancer association, inter-disciplinary teams will require high-quality information on environmental causes of
disease; careful use of perhaps unfamiliar statistical methods; and more gritty, perhaps hypothesis-based investigation of molecular and behavioral mechanisms.

References


