Editorial

Integrative Study Designs—Next Step in the Evolution of Molecular Epidemiology?

Neil E. Caporaso

Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, NIH, Rockville, Maryland

The ascendancy of the molecular epidemiology approach (1, 2), simply defined as including biomarkers in population-based study designs, is clear by any survey of current population studies. Most scientists would agree that embedding advanced technologies in molecular epidemiology designs will be key to achieving breakthroughs. Given the increasingly central role of molecular epidemiology in unraveling major questions in cancer, there is a compelling argument to expand its scope to include two areas that are often absent, namely, behavior and outcome. Integrative epidemiology refers to population-based study designs incorporating information or biomarkers that add the behavior and outcome 'wings' to the traditional/familiar molecular epidemiology paradigm that extends from exposure to disease (see Fig. 1, integrative epidemiology).

First, with regard to behavior, it is well established that the exposures that account for the majority of human cancer, such as tobacco (3), alcohol (4), and diet/energy balance (5), have strong hereditary components. Yet, although even the very earliest molecular epidemiology studies emphasized investigation of genes that process exposures as potentially influencing cancer (6), the study of genes that contribute to the key exposures responsible for cancer in the population has often been left to behavioral geneticists who focus on more extreme clinical samples (i.e., alcoholics and the morbidly obese). The lack of attention by cancer epidemiologists seems shortsighted because the genetic contribution to the exposures themselves is at least as strong as the hereditary component of risk of the individual cancers (7). A comprehensive understanding of the genetic architecture of cancer, not to mention cancer etiology, will be deficient without understanding the genes that influence the likelihood of the exposure itself. Moreover, selected genes influence both the likelihood of exposure and the downstream effects [i.e., CYP2A6 for tobacco (8) and the alcohol and acetaldehyde hydrogenase gene families for alcohol (9)]. Understanding such pleiotropic effects will be critical to unraveling the complex interplay of genes and environment. Complementing the role of genes, and considering tobacco as an example, behavioral measures can include dependency (i.e., Fagerstrom test for nicotine dependence; ref. 10), psychological characteristics (i.e., depression and anxiety), traits (introversion and neuroticism), other substance use, psychiatric disorders (schizophrenia and eating disorders), or exposure markers (i.e., cotinine). As behaviors associated with tobacco smoking, obesity, and alcohol consumption are resistant to modification, enhanced behavioral data on populations relevant to cancer may yield new prevention or therapeutic strategies or allow existing approaches to be better tailored to individuals. Just as expression/proteomic/genetic characterization of disease is expected to redefine disease categories and provide insights into molecular etiology, combined behavior and exposure information may result in more precise definitions of tobacco phenotypes (11).

Second, with respect to outcome, molecular epidemiology emphasizes cancer incidence with the investigation of outcomes relegated to clinical trials, where unfortunately, an emphasis on exposure, genetics, and biomarkers has, until recently, often been absent. Heredity influences treatment response and toxicity, and at least a few genes influence current therapy (12) and specifically cancer therapy (12). It is widely anticipated that genomics research (integrated with proteomics and expression) will eventually enhance risk stratification by genetic or biomarker defined subsets, with the potential for targeted prevention and therapy (i.e., personalized medicine; ref. 13). Spitz has used the term integrative epidemiology (14) to advocate this approach, emphasizing that selected genes/markers may influence both risk and outcome, highlighting glutathione _S_-transferase family genes, cyclin D1, and matrix metalloproteinase 1, as examples. How common such associations are is unknown, but as we currently lack a comprehensive understanding of the common modifier genes for any major human cancer (15), it would seem clear that if both disease and outcome can be investigated using the same technology, biospecimens, and study platform, such an effort could be both informative and efficient.

There are two central advantages of integrative epidemiology. The first is that a given study platform can accommodate a greatly expanded group of investigations. Epidemiologists traditionally emphasize well-defined exposure and disease measures in their studies, along with the collection of biospecimens. Given this baseline commitment, adding the relevant behavior and outcome components is efficient, as often minor adjustments to existing designs (i.e., questionnaires, biospecimens, collection protocols, informed consents, etc.) may capture large new classes of information. Even when more substantial resources are required, the overall effort will be far less then fielding a _de novo_ study. Epidemiologists traditional focus on data quality will enhance interpretation. The application of diverse high-technology approaches (genomics, proteomics, etc.) along with clinical and epidemiologic data, and tissue resources is expected to provide the molecular tools to both elucidate mechanism and refine risk and response (16, 17).

The second advantage is that integrative epidemiology allows investigation of characteristics (genes, biomarkers, etc.) that span more than one study feature (i.e., a gene with pleiotropic effects related to both disease and outcome as noted above) or behavior and exposure. Tobacco smoking is related to lung cancer and smokers who persist after diagnosis have increased second tumors and a poorer outcome (18). Isolated investigations will not reveal the actual complex roles of pleiotropic genes relevant to cancer without integrative epidemiology studies that include the relevant study domains (i.e., disease, outcome, etc.). The additional data will also permit more sophisticated analysis to sort out causal pathways in the presence of mediator variables (19) and interaction (20).

Arguments against such approaches generally cite cost and complexity, in part, because integrative studies require participation of diverse disciplines with resulting large study

Requests for reprints: Neil E. Caporaso, Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, NIH, 6120 Executive Boulevard, EPS 7116, Rockville, MD 20892. E-mail: caporasn@exchange.nih.gov
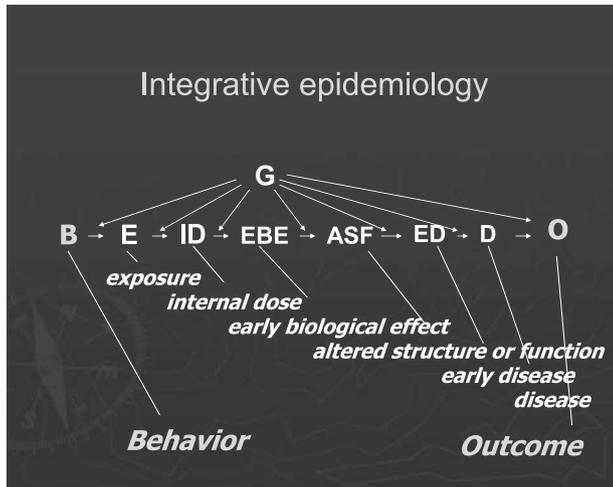
Figure 1. Integrative epidemiology is simply the familial molecular epidemiology paradigm with the 'wings', behavior, and outcome added.

teams. Both meta-analyses and recent findings from large scans (21) indicate that relative risks of susceptibility genes for common cancers are modest, so substantial sample size is mandatory to detect weak genetic signals, to achieve power to detect gene-environment and gene-gene effects, and to investigate subgroups. Accordingly, consortia have formed or are planned for virtually every major disease (22) and it is clear that team science featuring interdisciplinary teams will be an increasingly prominent trend (23). It follows that the increased cost will mandate that available efficiencies be applied to maximize the science. If we consider each study domain (behavior, outcome, exposure, genetics, disease, etc.) to be a 'node', and connections between nodes reflect potential areas for scientific investigation, adding behavior and outcome clearly enhances scientific opportunities (i.e., increases the number of nodes) for a given study platform. Efficiency also improves with larger studies [i.e., the absolute cost of an integrative epidemiology study is greater but the marginal cost (per unit of information) is lower].

In conclusion, integrated designs that incorporate behavior and outcome data into population-based studies will help fully realize the benefits anticipated for molecular epidemiology. As we commit large resources to conducting high-technology studies, platforms that incorporate these elements will provide an enhanced scientific payoff.

References

1. Perera F, Weinstein I. Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation. J Chronic Dis 1982;35:581–600.
2. Rothman N, Stewart W, Schulte P. Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. Cancer Epidemiol Biomarkers Prev 1995;4:301–11.
3. Li MD, Cheng R, Ma JZ, Swan GE. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. Addiction 2003;98:23–31.
4. Oroszi G, Goldman D. Alcoholism: genes and mechanisms. Pharmacogenomics 2004;5:1037–48.
5. Rankinen T, Zuberi A, Chagnon YC, Weisnagel SJ, et al. The human obesity gene map: the 2005 update. Obesity 2006;14:529–644.
6. Lower GM, Nilsson T, Nelson CE, Wolf H, Gamsky TE, Bryan GT. N-acetyltransferase phenotype and risk in urinary bladder cancer: approaches in molecular epidemiology. preliminary results in Sweden and Denmark. Environ Health Perspect 1979;29:71–9.
7. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78–85.
8. Malaiyandi V, Sellers EM, Tyndale RF. Implications of CYP2A6 genetic variation for smoking behaviors and nicotine dependence. Clin Pharmacol Ther 2005;77:145–58.
9. Brennan P, Lewis S, Hashibe M, et al. Pooled analysis of alcohol dehydrogenase genotypes and head and neck cancer: A HuGE review. Am J Epidemiol 2004;159:1–6.
10. Fagerstrom KO. Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. Addict Behav 1978;3:235–41.
11. Furberg H, Sullivan PF, Maes H, et al. The types of regular cigarette smokers: a latent class analysis. Nicotine Tob Res 2005;7:351–60.
12. Daly AK. Individualized drug therapy. Curr Opin Discov Devel 2007;10:29–36.
13. Herbst RS, Lippman SM. Molecular signatures of lung cancer—toward personalized therapy. N Engl J Med 2007;336:76–8.
14. Spitz MR, Wu X, Mills G. Integrative epidemiology: from risk assessment to outcome prediction. J Clin Oncol 2005;23:267–75.
15. Caporaso N. Genetic modifiers of cancer risk. In cancer epidemiology and prevention. 3rd ed. Oxford University Press; 2006. p. 577–602.
16. Chen HY, Yu S, Chen CH, et al. A five-gene signature and clinical outcome in non-small cell lung cancer. N Engl J Med 2007;356:11–20.
17. Lossos IS, Czerwinski DK, Alizadeh AA, et al. Prediction of survival in diffuse large-B-cell lymphoma based on expression of six genes. N Engl J Med 2004;350:1828–37.
18. Johnson BE. Second lung cancers in patients after treatment for an initial lung cancer. J Natl Cancer Inst 1998;90:1335–45.
19. Cole SR, Hernan MA. Fallibility in estimating direct effects. Int J Epidemiol 2002;31:163–5.
20. Chatterjee N, Kalaylioglu Z, Moshlehi R, Peters U, Wacholder S. Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. Am J Hum Genet 2006;79:1002–16.
21. Herbert A, Gerry NP, McQueen MB, et al. A common genetic variant is associated with adult and childhood obesity. Science 2006;312:279–83.
22. Ioannidis JP, Bernstein J, Boffetta P, et al. A network of investigator networks in human genome epidemiology. Am J Epidemiol 2005;162:302–3.
23. Sellers TA, Caporaso N, Lapidus S, Petersen GM, Trent J. Opportunities and barriers in the age of team science: strategies for success. Cancer Causes Control 2006;17:229–37.

Integrative Study Designs––Next Step in the Evolution of Molecular Epidemiology?

Neil E. Caporaso

Updated version	Access the most recent version of this article at: http://cebp.aacrjournals.org/content/16/3/365

Cited articles	This article cites 19 articles, 3 of which you can access for free at: http://cebp.aacrjournals.org/content/16/3/365.full#ref-list-1
Citing articles	This article has been cited by 5 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/16/3/365.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/16/3/365. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.