

Editorial

The Beginning of the End for the Epidemiologic Focus on Gene-Environment Interactions?

Thomas A. Sellers

H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida

With our harried schedules in academic research environments, it seems there is less and less time available to have good open dialogue and debate on issues. During my postdoctoral fellowship, I was exposed to the British tradition of breaking for tea twice daily. The department chair and my advisor, Dr. Robert Elston, took this very seriously, and everyone knew that at 10:00 am and 3:00 pm you dropped what you were doing and gathered in the conference room for tea and lively discussion. I should probably interject at this time that I was the first epidemiologist to train in this department that was noted for biometry and statistical geneticists. Their focus was primarily on mapping genes, whereas I was more interested in studying how the susceptibility genes they were studying interacted with nongenetic factors to cause risk. This merging of perspectives turned out to be quite important for an ongoing project on lung cancer susceptibility. We managed to integrate measures of tobacco and occupational exposures in a segregation analysis that provided the first evidence for a major gene influence (1). After 15 years of additional research, the first report of genetic linkage for susceptibility to lung cancer was published (2). However, this meeting of the minds was not without some good-natured ribbing of the relative merits of epidemiology versus genetics. At one morning tea, Dr. Elston grabbed a piece of chalk, walked to the board, and proceeded to illustrate why epidemiologists were vastly inferior to the geneticists. While muttering about how silly our models were, he wrote $D = E + e$ (Disease = Environment + error). Genetics was nowhere to be found in the model (which was true back in 1988). He proudly produced a "far superior" model advocated by the geneticists: $P = G + E + e$ [Phenotype = Gene(s) + Environment(s) + error]. "But Robert, geneticists do not measure environment," I said. Laughing in response he agreed, adding "But at least we have a better model!" I was reminded of this while listening to a session at the most recent AACR annual meeting, as the differences in perspectives may still be relevant ~2 decades later.

Is Environment a Nuisance Parameter? Epidemiology has made several notable contributions toward the etiology and prevention of cancer (3). There is no need to belabor the overwhelming evidence from migrant studies and case-control and cohort studies that cancer risk has a significant nongenetic component. The question at hand is whether we can ignore the environment in efforts to identify the susceptibility genes. I do not pretend to know the answer to this seemingly simple question, but I feel quite confident that more healthy debate and discussion could help move the field forward. One perspective is that the search for susceptibility can only be successful if one considers environmental factors. Indeed, there is some evidence from analysis of simulated data that

genetic linkage efforts are enhanced when measured environment is considered (4). The other perspective is that unless there is evidence for a genetic signal (a main effect), then consideration of environment or fitting gene-environment interactions is moot. This has rarely been a problem in the past because geneticists (largely ignoring environment) were usually studying families, sib-pairs, or twins, whereas epidemiologists were studying unrelated individuals in case-control studies. Although some major gene influences on cancer risk have been identified through linkage analysis [the on-line version of Mendelian Inheritance in Man (<http://www.ncbi.nlm.nih.gov>) lists several hundred rare inherited disorders for which neoplasia is a major feature or complication, and many were identified in this manner], the growing suspicion that few high-penetrance mutations remain to be discovered has steered geneticists increasingly to the use of the epidemiologists' strategy: the case-control study (in genetics parlance an "association study"). Moreover, the appreciation that the associated risks of these low-penetrance alleles will be modest has signaled the need for large studies of well-characterized cases and controls (5, 6).

Are Answers in the Genes? Given that we are in the same scientific sandbox, it is important to carefully consider the perspectives and expertise of epidemiology and genetics in the design, conduct, and execution of these large studies (7). For example, the Haplotype Map project (<http://www.hapmap.org>) has generated considerable excitement in the promise to capture genetic variation in an efficient manner suitable for large-scale studies. The plan to analyze 270 individuals representing diverse geographic populations, with a sample size up to 45 per group, is sufficient to characterize single-nucleotide polymorphisms frequencies of the minor allele >5% and identify tag single-nucleotide polymorphisms that will capture up to 80% of the genetic variation in a chromosomal region. Terwilliger et al. (8) have raised concerns about the Haplotype Map project and the underlying statistical basis. Pritchard (9) has advocated consideration of rare alleles. From a genetics perspective, capturing 80% of the variability may be considered quite adequate. From an epidemiologic perspective, does that lead to unacceptable levels of misclassification? If some of the international variation in rates of cancer incidence reflects differences in genetics rather than environment, is it a mistake to focus only on variants that are common across populations?

Gene-Environment Interactions: Reading the Tea Leaves. There are many reasons why it is important that we make progress in further unraveling the etiology of cancer, particularly the interactions of genes and environment. Foremost are the 564,830 cancer deaths projected to occur in 2006 by the American Cancer Society (<http://www.cancer.org>), which translates to roughly one death per minute. In addition, as if the exceedingly stringent pay lines at the NIH were not critical enough, there is also the evidence for shifting priorities in research that signals we could be facing the beginning of the end. With a few clicks of the mouse, one can

review National Cancer Institute priorities since 2000. Ranging in titles from "Extraordinary Opportunities for Investment" to "Plans & Priorities" to "The Nation's Investment in Cancer Research," they provide a brief yet interesting summary of recent trends. In 2000 (<http://plan2000.cancer.gov>), the top priority was cancer genetics, including a specific objective to "establish approaches to study the interaction between genes and individual genetic variations and the environment to understand cancer risk." Genes and the environment remained the top priority in 2001, 2002, 2003, 2004, and 2005. By 2006, molecular epidemiology was listed as the seventh of seven strategic investments (<http://plan2006.cancer.gov>). The plan for 2007 identifies four areas where integrative approaches are being targeted for support. One of these, entitled "Integration of the Biological with the Population and Public Health Sciences," notes "we will support research that extends beyond molecular epidemiology to consider the roles of behavioral, sociocultural, and psychosocial factors in cancer susceptibility, utilization of recommended screening, and treatment outcomes." The reader is left to derive their own conclusions as to where gene-environment interactions fit in this priority.

Timing is Everything. The field of molecular/genetic epidemiology has received notable criticism for the perceived lack of success. It is not the intent of this commentary to review these or their validity. Although we all learn in grade school that life is not fair, it is really frustrating if we truly are facing the beginning of the end when the opportunities for progress are now vastly greater than ever. Looking back just over the past 5 to 10 years, one can appreciate the incredible advances in technology that afford ever faster, cheaper, and accurate genotyping. The bioinformatics tools now available greatly improve our ability to evaluate single-nucleotide polymorphisms based on their evolutionary conservation, type of change, tissue distribution, and pathways. We are increasingly aware that a focus on inherited variation may be insufficient to capture the complexity of cancer, given that it ignores other genetic alterations (loss of heterozygosity, somatic mutations, epigenetic silencing, etc; ref. 7). High-throughput platforms are now available or near release that will permit integration of data on both inherited and acquired genetics, along with measured environment on a scale appropriate for epidemiology. If timing is really everything, maybe we were pressured to start before we really had the necessary technologies and platforms to be successful out of the blocks? It looks like time to play catch up is now on us in a race to verify and justify our discipline.

A Cold Hard Look in the Mirror. Of course, an alternative reaction to the criticisms and apparent decreasing support are to get out of this business entirely. Are we only satisfying our intellectual curiosity by unraveling etiologic pathways? At the end of the day, will the ultimate conclusions be that we should not smoke, eat a healthy balanced diet, avoid weight gain, minimize harmful UV wave exposure, moderate intake of alcohol, and practice safe sex? (My apologies to those of you whose pet environment or lifestyle was omitted from, or included in, this incomplete list.) Many would argue that we should be following a healthy lifestyle because it improves quality of life on a daily basis not because of disease avoidance or prevention. Although I agree with this perspective, identification of high-risk individuals is really important, as the promise to derail the cancer process seems within the very real realm of possibility. Given that interventions are rarely entirely risk-free, targeting the appropriate population is

critical to tip the scales to an acceptable risk-benefit balance. Sadly, I do not think our discipline has had a fighting chance of identifying the subset of the population at exceptionally high risk because of the combination of risk alleles and environment. Regardless, it is not time to be disillusioned or to quit. Perhaps, our perspectives on the magnitude of possible effect associated with inherited variants were initially overly optimistic. Ever increasing sample sizes can certainly contribute (6) but it is not the only solution, as small studies have shown convincing evidence for genetic influences (10). In addition, we are starting to see pooled and meta-analyses that seem to suggest some of the reported gene-environment interactions reported over the past decade or so may be real (11-13). Creative designs that explore intermediate phenotypes, refinement of the cancer end point on molecular markers to identify more homogenous subsets, or improving signal-to-noise through control of the environment must also be considered.

In summary, the search for gene-environment interactions in cancer has gone from the glory child of the National Cancer Institute investment portfolio to seemingly an afterthought. Although our lack of success has been good for Dr. Shields and Null Results in Brief, I am certain he would delight in seeing fewer articles come his way in exchange for more success in this arena. Rapid changes in technology and bioinformatics have almost completely changed the playing field over our initial forays. We need to rapidly and aggressively realize these opportunities before the funding agencies decide the return has not been worth the investment. As a senior editor for the journal, I see increasing evidence for success in the quality of the articles submitted on the topic and the significance of the findings presented therein. We are all pulling for continuance of that apparent trend.

References

1. Sellers TA, Bailey-Wilson JE, Elston RC, Wilson AF, Elston GE, Rothschild HR. Evidence for Mendelian inheritance in the pathogenesis of lung cancer. *J Natl Cancer Inst* 1990;82:1272-9.
2. Bailey-Wilson JE, Amos CI, Pinney SM, et al. A major lung cancer susceptibility locus maps to chromosome 6q23-25. *Am J Hum Genet* 2004; 75:460-74.
3. Colditz G, Sellers TA, Trapido E. Contribution of epidemiology to the identification of causes of cancer and the subsequent reduction in the burden of cancer in the United States. *Nat Rev Cancer* 2006;6:75-83.
4. Arnett DK, Pankow JS, Atwood LD, Sellers TA. Impact of adjustments for intermediate phenotypes on the power to detect linkage. *Genet Epidemiol* 1997;14:749-54.
5. Houlston RS, Peto J. The search for low-penetrance cancer susceptibility alleles. *Oncogene* 2004;23:6471-6.
6. Potter JD. Toward the last cohort. *Cancer Epidemiol Biomarkers Prev* 2004; 13:895-7.
7. Potter JD. Epidemiology informing clinical practice: from bills of mortality to population laboratories. *Nat Clin Pract Oncol* 2005;2:625-34.
8. Terwilliger J, Hiekkalinna T. An utter refutation of the "Fundamental Theorem of the HapMap." *Eur J Hum Genet* 2005;14:426-37.
9. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001;69:124-37.
10. Goldstein AM, Landi MT, Tsang S, Fraser MXC, Munroe D, Tucker MA. Association of MC1R variants and risk of melanoma in melanoma-prone families with CDKN2A mutations. *Cancer Epidemiol Biomarkers Prev* 2005; 14:2208-12.
11. Hung RJ, Hall J, Brennan P, Boffetta P. Genetic polymorphisms in the base excision repair pathway and cancer risk: a HuGE review. *Am J Epidemiol* 2005;162:925-42.
12. Kiyohara C, Yoshimasu K, Takayama K, Nakanishi Y. EPHX1 polymorphisms and the risk of lung cancer: a HuGE review. *Epidemiology* 2006;17: 89-99.
13. Coles BF, Kadlubar FF. Human α class glutathione S-transferases: genetic polymorphism, expression, and susceptibility to disease. *Methods Enzymol* 2005;401:9-42.