

Genetic Association Studies of Cancer: Where Do We Go from Here?

Timothy R. Rebbeck,¹ Muin J. Khoury,² and John D. Potter³

¹Abramson Cancer Center, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; ²National Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, Georgia; and ³Fred Hutchinson Cancer Research Center, Seattle, Washington

"A foolish consistency is the hobgoblin of little minds, adored by little statesmen and philosophers and divines."

—Ralph Waldo Emerson

A goal of genetic association studies is to identify disease-causing genes. True positive associations between genetic variants and disease traits have the potential both to elucidate biologic principles and to lead to translation of research into clinical or preventive practice. For instance, an environmental effect on cancer risk may vary with genotype; or response to a treatment can differ based on information contained in the genotype. However, the reputation of genetic association studies has suffered because of the difficulty identifying reproducible and biologically interpretable results. Although criteria for identifying truly causative events in disease etiology have been proposed in numerous settings (1-7), we need to advance both conception and methods to improve the execution, interpretation, and integration of genetic association studies.

Studies on genetic associations in cancer have been published at an increasing pace. The CDC-curated published literature database on genetic associations (8) reveals the tremendous growth of this literature (Table 1). The annual number of reports of genetic associations involving cancer as an outcome has more than tripled over the past 6 years. These publications include an increasing number of articles on gene-gene and gene-environment interactions, as well as systematic reviews and meta-analyses that integrate the literature on genetic variants with that on cancer risk. This upward trend is likely to continue over the next few years, particularly with the advent of genome-wide association platforms in which hundreds of thousands of genetic variants will be studied simultaneously in large-scale consortium studies. So where is the field going? And how can *CEBP* help to bring order to this increasing flood of literature?

A fundamental conceptual requirement for molecular epidemiology has been the insistence that findings of a genetic association study be replicable. Nonetheless, critics of molecular epidemiologic associations note, sometimes properly, that many such studies are not replicated (9); inconsistencies are regularly explained, at least in part, by inadequate approaches to study design and analysis. Recent data suggest that population genetics structure, such as deviation from Hardy-

Weinberg equilibrium or linkage disequilibrium between the SNP under study and a causative SNP (10, 11), can affect association study results. Thus, some inconsistent associations can be explained by a more extensive consideration of the genomic or environmental context of the variants under study. Scientists are appropriately wary of incorrect inferences because they can lead subsequent research in the wrong direction, wasting valuable resources and time. Some journals have suggested that studies cannot be published unless evidence of replication is provided at the time of the initial publication (7). Taking this even further, others have suggested that the lack of consistency in associations raises serious concerns about the value of molecular epidemiology as a discipline.

Nonetheless, many studies are replicated (9), and many unreplicated associations could ultimately have biological or clinical relevance. Thus, a simple-minded requirement for replicating studies of genetic associations is in danger of becoming the hobgoblin of human genetics research. Overemphasis on the apparent inconsistency of association-study results fails to acknowledge the complexity and heterogeneity that underlie all common diseases, including cancer. For example, an association may genuinely be present (or detectable) in some populations or subpopulations, but not others. An association may be observed only in a specific context (e.g., under exposure to a relevant agent), but such contexts are not always understood nor are the relevant exposures well measured. Complexity and heterogeneity are always ignored in the design, analysis, execution, or interpretation of association studies. Most recognize how unlikely it is that a single nucleotide variant (no matter how functionally relevant) has a simple relationship with the risk of developing complex diseases (12). It is far more likely that many genetic variants, in conjunction with environmental exposures, will explain variation in disease etiology, drug response, and prognosis. A correlate of this higher-order complexity is that the marginal effects of single genetic variants may be obscured, and studies that do not properly account for this complexity are likely to miss relevant etiologic events. Therefore, studies need to be designed, executed, and interpreted in recognition of this complexity. A lack of consistency of associations may not always reflect an absence of causative association in disease but may instead be the consequence of true heterogeneity in the effects of genetic variants, exposures, or their interactions.

To avoid having overly simplistic notions of consistency overtake and hamper scientific progress, a sophisticated framework is required to identify biologically or clinically meaningful genetic associations. The Human Genome Epidemiology (HuGENet) Network⁴ and *Cancer Epidemiology, Biomarkers & Prevention (CEBP)* have provided separate forums

Cancer Epidemiol Biomarkers Prev 2007;16(5):864-5

Received 3/28/07; accepted 3/28/07.

Grant support: The Robert Wood Foundation, the Cancer Center Support Grant, National Cancer Institute grant 5P30 CA016672, and National Cancer Institute grant 5 R01 CA097893-02.

Requests for reprints: Timothy Rebbeck, Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, 904 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021.

Phone: 215-898-1793; Fax: 215-4409355. E-mail: trebbeck@cceb.med.upenn.edu

Copyright © 2007 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-07-0289

⁴ <http://www.cdc.gov/genomics/hugenet/default.htm>

Table 1. Trends in the published literature on genetic associations in cancer, 2001-2006

	Type of publication*			Total
	Genetic associations	Gene-gene and gene-environment interactions	Meta-analyses and HuGE reviews	
2001	333	139	5	383
2002	497	175	17	579
2003	555	165	14	644
2004	699	210	21	797
2005	983	319	33	1,090
2006	1,095	375	38	1,219

NOTE: Data are from analysis of the CDC HuGE Published Literature Database on February 12, 2007 (ref. 8).

*Categories are not mutually exclusive.

for the publication, interpretation, and synthesis of association studies. Now *CEBP* and HuGENet will team up to make sense of this evolving field. *CEBP* has been a leader in publishing genetic association studies in cancer but recognizes the limitations of the literature in attempting to identify truly causative associations with biologic, clinical, or public health relevance. HuGENet is an open global collaboration of individuals and groups interested in advancing the integration of knowledge and applications of genetic information in medicine and public health. HuGENet sponsors methodologic workshops, hosts a published literature database, promotes systematic reviews of genetic associations-HuGE reviews (13), and convenes collaborations among consortia and networks through its "network of networks" (14). *CEBP* and HuGENet have individually and jointly proposed guidelines by which an improved environment for the publication of association studies can be achieved (2, 15-17). These guidelines encompass biological plausibility; population genetics structure (including factors related to race or ethnicity); appropriate laboratory and questionnaire measurements of exposures; consideration of etiologic complexity and heterogeneity; appropriate study design; and the correct inferential methods that consider etiologic complexity.

Current guidelines, however, remain inadequate and require further development. HuGENet serves as a focus for this effort and provides continuing guidance and direction for the field (18, 19). *CEBP* recognizes the merit of the structured approach that HuGENet can provide to evaluate association study results. *CEBP* therefore encourages the publication of methodological research that enhances the field of association studies, and applied association studies that appropriately consider etiologic complexity and replication, as well as meta-analyses that summarize the existing literature. *CEBP* further

encourages authors interested in integrating information on genetic associations to use the HuGENet handbook for systematic reviews as a framework for providing meaningful information to the scientific literature (20). More than ever, as the pace of publication continues to increase, this field needs a collective effort to establish and apply the best set of guidelines for the analysis, interpretation, and synthesis of genetic associations.

References

- Weiss ST. Association studies in asthma genetics. *Am J Respir Crit Care Med* 2001;164:2014-5.
- Ioannidis JP. Commentary: grading the credibility of molecular evidence for complex diseases. *Int J Epidemiol* 2006;35:572-8; discussion 593-6.
- Cooper DN, Nussbaum RL, Krawczak M. Proposed guidelines for papers describing DNA polymorphism-disease associations. *Hum Genet* 2002;110:207-8.
- Huizinga TW, Pisetsky DS, Kimberly RP. Associations, populations, and the truth: recommendations for genetic association studies in Arthritis & Rheumatism. *Arthritis Rheum* 2004;50:2066-71.
- Rebeck TR, Spitz M, Wu X. Assessing the function of genetic variants in candidate gene association studies. *Nat Rev Genet* 2004;5:589-97.
- Freimer NB, Sabatti C. Guidelines for association studies in Human Molecular Genetics. *Hum Mol Genet* 2005;14:2481-3.
- Framework for a fully powered risk engine. *Nat Genet* 2005;37:1153.
- Lin BK, Clyne M, Walsh M, et al. Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. *Am J Epidemiol* 2006;164:1-4.
- Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;33:177-82.
- Lin PI, Vance JM, Pericak-Vance MA, Martin ER. No gene is an island: the flip-flop phenomenon. *Am J Hum Genet* 2007;80:531-8.
- Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet* 2005;76:967-86.
- Sing CF, Haviland MB, Reilly SL. Genetic architecture of common multifactorial diseases. *Ciba Found Symp* 1996;197:211-29; discussion 229-32.
- Little J, Khoury MJ, Bradley L, et al. The human genome project is complete. How do we develop a handle for the pump? *Am J Epidemiol* 2003;157:667-73.
- Ioannidis JP, Gwinn M, Little J, et al. A road map for efficient and reliable human genome epidemiology. *Nat Genet* 2006;38:3-5.
- Rebeck TR, Martinez ME, Sellers TA, Shields PG, Wild CP, Potter JD. Genetic variation and cancer: improving the environment for publication of association studies [see comment]. *Cancer Epidemiol Biomarkers Prev* 2004;13:1985-6.
- Ioannidis JP. Journals should publish all "null" results and should sparingly publish "positive" results. *Cancer Epidemiol Biomarkers Prev* 2006;15:186.
- Rebeck TR, Ambrosone CB, Bell DA, et al. SNPs, haplotypes, and cancer: applications in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2004;13:681-7.
- Seminara D, Khoury MJ, O'Brien TR, et al. The emergence of networks in human genome epidemiology: challenges and opportunities. *Epidemiology* 2007;18:1-8.
- Ioannidis JP, Bernstein J, Boffetta P, et al. A network of investigator networks in human genome epidemiology. *Am J Epidemiol* 2005;162:302-4.
- Little J, Higgins J. HuGENet HuGE review handbook, version 1.0. c2006 [cited 2007 Mar 9]. Available from: http://www.genesens.net/_intranet/doc_nouvelles/HuGE%20Review%20Handbook%20v11.pdf.