

*Editorial***Genetic Variation and Cancer: Improving the Environment for Publication of Association Studies**

Timothy R. Rebbeck,¹ María Elena Martínez,² Thomas A. Sellers,³ Peter G. Shields,⁴ Christopher P. Wild,⁵ and John D. Potter⁶

¹Abramson Cancer Center, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; ²Arizona Cancer Center, University of Arizona, Tucson, Arizona; ³Moffitt Cancer Center and University of South Florida, Tampa, Florida; ⁴Lombardi Comprehensive Cancer Center, Georgetown University School of Medicine, Washington, District of Columbia; ⁵Molecular Epidemiology Unit, University of Leeds, Leeds, United Kingdom; and ⁶Fred Hutchinson Cancer Research Center, Seattle, Washington

The completion of a working draft of the human genome and the rapidly expanding knowledge base that has accompanied it are having a major impact on our understanding of the role of genes in human cancer. A large number of studies have appeared in the literature that report associations of low penetrance genetic variants with disease or disease-related phenotypes. It is likely that the majority of these variants act in concert with specific environmental exposures, which are often ignored, poorly measured, or not considered at all in genetic association studies. Similarly, it is likely that these genes do not act alone but interact with other genes or biomarkers in identifiable pathways. Therefore, genetic association studies may not result in clear inferences about the causative role of these genes with disease end points and are not always replicated in subsequent studies. Accordingly, a major need in molecular epidemiology is to promote research that identifies biologically plausible, replicable genotype-disease associations.

As a forum in which many molecular epidemiologic association studies are reported, *Cancer Epidemiology, Biomarkers & Prevention* will increasingly prioritize the publication of reports that are more likely to represent disease-causing events. To help in this process, editors and reviewers for *Cancer Epidemiology, Biomarkers & Prevention* are now being asked to evaluate genetic association papers using stringent criteria. These criteria include the following:

1. Consideration of Biology

Biological Plausibility: Publication priority will be influenced by data on the prior probability that a genotype or haplotype is associated with disease. For example, papers will be given higher priority if they study genetic variants with documented functional significance and when the reported association is consistent with a prior hypothesis based on functional data. This *a priori* information can come from the selection of candidate genes from regions of linkage peaks, the incorporation of functional information from laboratory investigations, the use of bioinformatics computational tools investigating

evolutionary conservation or predicted structural effects of genetic variants, or other relevant approaches. The editors expect the authors to provide data supporting function or a literature-based discussion about the hypothesized effect of a genetic variant. An exception to the requirement that biological plausibility be considered is the publication of genome-wide associations that, when using appropriate methods, may identify regions of the genome that contain loci of interest before those loci have been well characterized. As most low penetrance genes are expected to act through gene-gene or gene-environment interactions, studies that consider only main effects of a single gene will be considered with lower priority.

Pathways: For most commonly occurring cancers, it is likely that interactions of multiple proteins in pathways as well as interactions between genetic variants and exposures are involved in the causal etiology of the phenotype. Therefore, priority will be given to studies that consider biologically plausible interactions of multiple genes in a pathway as well as interactions of environmental exposures with genetic variants that are involved in the metabolism of those exposures. Single-gene studies will be considered increasingly less likely for publication when there are additional known biologically related and plausible genes or exposures to study. When there are unique circumstances in which single-gene effects are relevant, this should be explained by the authors in a cover letter and, preferably, in the research itself. Studies that aim to elucidate mechanistic pathways (e.g., linking genetic variants to intermediate end points, such as DNA adducts or mutations) will also be encouraged. Priority for publication will be given to reports of genes and/or exposures in pathways even when those associations are null.

Population Structure: Single variants in a candidate gene may be insufficient to characterize completely the genetic variability related to a phenotype. Therefore, studies that consider a more complete range of genetic variability at a locus will be given higher priority. These measures include assessment of linkage disequilibrium and assessment of associations that involve haplotypes. Evaluation and discussion of population genetic characteristics (e.g., Hardy-Weinberg proportions among controls) are strongly encouraged. While it remains unclear whether, and under what conditions, haplotype data may provide improved estimates of disease risk beyond that of

functional variants, papers that include haplotype information in addition to data on functional genetic variation are encouraged.

Etiologic Heterogeneity of Phenotypes: Cancer and related phenotypes have complex, multifactorial etiologies that result in heterogeneity in the phenotypes themselves. Therefore, analyses should consider etiologic heterogeneity by evaluating histopathologic or molecular data that subclassify disease to identify homogeneous subsets for analysis. This may include analyses stratified by histologic type, molecular subtype as defined by expression arrays or proteomics, demographic characteristics, or other phenotypes. However, a prior rationale for the classification is necessary, and careful attention should be paid to statistical power to avoid spurious findings that may result from subset analyses.

2. Methodologic Considerations

Statistical Power: Only studies that are adequately powered to detect associations of small magnitude will be published. Large sample sizes are required to detect associations with small magnitude effects, to detect these effects in population subsets, and to identify relevant interactions among genes and between genes and exposures that have been generated as the result of biologically plausible *a priori* hypotheses. Statements regarding prior statistical power are essential to avoid the publication of opportunistic interactions or associations found only in subsets that were not part of the study as originally designed.

Statistical Analysis: Because etiologic complexity (including interactions) may be a hallmark of many associations, creative, but appropriate, use of statistical methods should be employed to optimize the chances that the true association is revealed. This may include consideration of a full range of etiologic models. For example, analyses should not be limited to only dominant or recessive effects unless strongly dictated by functional data. Gene-dosage effects and appropriate interactions should also be considered. Methods should be employed that explore the joint associations with many genes and environmental factors involved in a common pathway or competing pathways rather than one at a time or in pairwise combinations. Authors are expected to acknowledge the full range of variables and models considered and should adjust for these variables appropriately. Although there is little evidence that it confers a strong bias in most association settings, consideration of potential confounding by ethnicity (i.e., population stratification) may be informative. Finally, corrections for multiple comparisons may be required, and authors should set appropriately stringent and justified levels of statistical significance when making inferences.

Laboratory Methods: Inconsistent or error-prone laboratory methods may contribute to lack of replication among studies of genetic variation and disease. Nondifferential genotype misclassification between cases and controls can result in bias toward the null, and differential genotype misclassification will confer unpredictable biases. Therefore, reports involving genotype associations should include information about laboratory quality control procedures. These may include (but are not limited to) descriptions of the use of internal blind duplicate repeats and the rate of genotype misclassification; use of positive and negative controls; genetic "barcoding" (e.g., using

microsatellite repeats) to monitor sample identity; evaluation of Mendelian inheritance to confirm genotype integrity when using family data; evaluation of deviations from Hardy-Weinberg proportions to identify potential systematic misclassification of specific genotypes; employment of a random genotyping order for cases and controls; and use of appropriate laboratory information management systems for data entry and management. In addition to the inclusion of data on quality-control methods, results of these approaches need to be provided so that the reader can evaluate the quality of the data presented. The editors also encourage all authors to make provisions for sample sharing for validation studies by others.

3. Data Interpretation

Consistency of Inferences: Although studies reporting a novel association are important, many molecular epidemiologic association studies have not been replicated in subsequent work. Therefore, although the quality of the individual study under review will be the primary consideration in determining publication, all papers should address issues of consistency or replication relative to previous reports. For example, papers will be prioritized if they replicate prior findings, refine the population subsets or exposed groups in which the association is primarily acting, or show that prior studies represent false-positive or false-negative findings. Similarly, reports may be prioritized if the association is replicated in two or more independent study samples, ethnic groups, or split samples of the same population. Contrary to the behavior of many journals, second or third reports that replicate previously published work may also be accorded priority, perhaps as Short Communications.

Post hoc Inferences: Not all gene-environment or gene-gene interactions evident in a data set may be based on *a priori* hypotheses. Although the journal does not wish to completely discourage such "hypothesis-generating" explorations, the authors have a responsibility to present them as such and interpret the results of these exploratory analyses accordingly. *Post hoc* justifications for statistically significant associations should be avoided and skepticism regarding the danger of false-positive findings is encouraged. This is particularly true when the prior probability of an association is low.

The criteria outlined here may be used to prioritize papers for publication, but these criteria specifically do not refer to strength of association. These criteria may apply whether the report is positive or null (see also specific instructions for the Null Reports in Brief section of this journal). Although it is not necessary to incorporate all of the above criteria in every paper reporting genetic associations, close consideration of these criteria by authors, reviewers, and editors should limit the publication of false-positive and false-negative reports. These criteria will be the guide for determining publication priorities in *Cancer Epidemiology, Biomarkers & Prevention*. The ability to implement these and other measures will allow genetic association studies to provide more meaningful information about disease etiology or outcome and thus allow this information to be further translated into cancer prevention, public health recommendations, clinical practice, or basic science studies that elucidate disease mechanisms. We look forward to responses from readers and reviewers on these guidelines and always welcome other suggestions regarding the quality of the journal.

BLOOD CANCER DISCOVERY

Genetic Variation and Cancer: Improving the Environment for Publication of Association Studies

Timothy R. Rebbeck, María Elena Martínez, Thomas A. Sellers, et al.

Cancer Epidemiol Biomarkers Prev 2004;13:1985-1986.

Updated version Access the most recent version of this article at:
<http://cebp.aacrjournals.org/content/13/12/1985>

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/13/12/1985>. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.