

Commentary

The Reliable Identification of Disease-Gene Associations

Paul D.P. Pharoah,¹ Alison M. Dunning,¹ Bruce A.J. Ponder,¹ and Douglas F. Easton²

¹Department of Oncology and ²Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, United Kingdom

A recent *Cancer Epidemiology, Biomarkers & Prevention* editorial proposed guidelines for prioritizing publication of genetic association studies in cancer (1). In their Introduction, they indicate that the journal will "... increasingly prioritize the publication of reports that are more likely to represent disease-causing events." We completely agree that the reliable identification of true disease-gene associations should be a major priority. However, we have concerns about the specific criteria by which these will be judged.

The editors propose that publication priority be determined to an extent by "biological plausibility." Thus, associations with variants with "documented functional significance" will be given more weight. In our view, this exaggerates the reliability of functional arguments. Many investigators will concentrate on genes and variants that they believe to be of functional relevance. Whereas this may be a reasonable approach to selecting variants for study, using functional arguments to judge the evidence for association (and hence the worthiness of publication) seems more dangerous.

Of course, functional arguments become central once an association is established, and they may be essential to determine which of the multiple variants in linkage disequilibrium is truly causative. However, many variants will have effects of one sort or another in biological assays, the relevance of which in relation to cancer will be quite unknown. Moreover, many reported functional assays are not robust or replicable (false positives), and reports that a variant has no functional effect may be false negatives—the appropriate phenotype may not be tested or the assay may be insensitive. Therefore, to use such a criterion of function to bolster the evidence for association may be very misleading. It may turn out that certain general genomic features, such as degree of evolutionary conservation, are strongly predictive of disease association, but the general applicability of these arguments also remains to be shown.

Furthermore, a need for functional data precludes the study of genes with no functional data on known variants. It also precludes the use of "tagging" approaches (which provide an empirical and robust approach to assessing associations with any common variants within a gene) as these rely on correlation between potentially neutral markers and risk alleles. The history of genetic epidemiology has shown that the empirical discovery of susceptibility alleles often precedes an understanding of biological function and the development of robust functional assays. In our current state of ignorance, we believe that the primary emphasis should be on the statistical evidence for association.

Our second area of concern surrounds interactions. It is axiomatic that susceptibility to common cancers will involve many genes, and that determining the combined effects of these genes and nongenetic risk factors on overall cancer risk is an important goal. The question is how one arrives at that goal.

It is well documented that analysis of interactions suffer from two serious problems: (a) because there are many more possible interactions than main effects, there is a much more serious issue of multiple testing and (b) the power to distinguish between different models of interaction is usually extremely limited. The result is that, in practice, inclusion of interactions in association studies is often a recipe for confusion. There is a virtually unlimited number of possible interactions with almost no way of selecting a model to test as the prior hypothesis. Restricting attention to "plausible" interactions may not be that helpful. For example, it is difficult to argue convincingly whether combined effect of two polymorphisms, each influencing estrogen metabolism, should be additive, multiplicative, or something else. Equally, it is completely unclear whether one should focus on interactions between genes that are thought to act in the same pathway or, perhaps just as importantly, on interactions between different pathways. In our view, consideration of interactions should primarily be undertaken for variants where a clear main effect has been established.

Similar considerations apply to "gene-environment" interactions, with the added difficulty of investigating lifestyle/environment in the context of a case-control study. There are circumstances in which a very clear dichotomy into exposed and unexposed individuals may be possible and sensible (e.g., certain occupational exposures that are strongly related to risk). Many exposures of current interest (diet, exercise, infection, etc.) are, however, ubiquitous, not easy to measure accurately, and uncertainly associated with the risk. In these circumstances, analysis of interactions in the absence of a clear disease-gene association seem to us to be of doubtful value.

It is widely accepted that there is important individual genetic variation in the risk of most, if not all, common cancers. Equally, it is clear that very few of the relevant genetic variants have been identified, and that the large majority of associations reported in the literature cannot be replicated. A major emphasis should now be on conducting association studies, with well-matched cases and controls that are sufficiently large to establish or rule out moderate associations. The journal can play an important role by encouraging publications of such studies, whether positive or negative.

Acknowledgments

The authors are funded by Cancer Research UK.

Reference

1. Rebbeck TR, Martinez ME, Sellers TA, Shields PG, Wild CP, Potter JD. Genetic variation and cancer: improving the environment for publication of association studies. *Cancer Epidemiol Biomarkers Prev* 2004;13:1985–6.

The Reliable Identification of Disease-Gene Associations

Paul D.P. Pharoah, Alison M. Dunning, Bruce A.J. Ponder, et al.

Cancer Epidemiol Biomarkers Prev 2005;14:1362.

Updated version Access the most recent version of this article at:
<http://cebp.aacrjournals.org/content/14/6/1362>

Cited articles This article cites 1 articles, 1 of which you can access for free at:
<http://cebp.aacrjournals.org/content/14/6/1362.full#ref-list-1>

Citing articles This article has been cited by 4 HighWire-hosted articles. Access the articles at:
<http://cebp.aacrjournals.org/content/14/6/1362.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/14/6/1362>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.