Commentary

Publication Environment and Broad Investigation of the Genome

Sholom Wacholder

Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland

The editorial (1) on *CEBP* publication criteria is timely and important. However, the days when most studies report results on a single gene, or even several genes, are almost over. *CEBP* editors, therefore, need to decide how to handle reports from broad investigations of the vast number of genes about which we know very little from the laboratory or from epidemiology. Although any one of these genes is unlikely to be related to disease, the total attributable risk from the vast number of unstudied genes may be greater than the total attributable risk from the few genes with far higher prior probabilities, which have been the focus of researchers until now (2). The editorial (1) does not consider how *CEBP* publication policies can best encourage the most rational, cost-effective, and timely strategies for exploring the genome broadly to identify causes of cancer.

The editorial (1) says that ''*CEBP* will increasingly prioritize the publication of reports that are more likely to represent disease-causing events (1).'' In my view, this policy would give results of a study too much weight in decisions about publication; as a consequence, readers will still be misled too often by authors' exaggeration of the importance of their findings.

There is an ongoing debate about appropriate strategies, including ''corrections for multiple comparisons'' (1, 3), and Bayesian (4) and quasi-Bayesian methods (2), for providing protection against false-positive and overinflated findings from large-scale genomic investigation. Whichever analytic approach is taken, fear of criticism, particularly from *CEBP* editors, for carrying out *too many* tests should not deter investigators from exploration of more genes. Neither should *CEBP* policy or practice create incentives for investigators to publish in ''least publishable units,'' nor for selective or incomplete reporting of results, which can lead to publication bias and subsequent wasted efforts by other scientists.

Concern in the editorial (1) on the danger from post hoc inferences and opportunistic interactions might also lead to unjustified limitations on the scope of investigation. Realistic accounting for the prior probability, followed by explicit consideration of the chance that any claim of a finding from these explorations is a false-positive (2), can help editors and readers evaluate whether authors' conclusions are warranted.

The editorial (1) says, ''Publication priority will be influenced by data on the prior probability that a genotype or haplotype is associated with disease'' and ''These criteria specifically do not refer to strength of association.'' To the contrary, data providing evidence of association that is strong enough to overwhelm a very low prior probability should get the *highest* priority for publication. Bayesian and quasi-Bayesian approaches can explicitly incorporate both prior probability and strength of association.

Beyond formal considerations of false-positives and false-negatives from random variation, the editors need to take a rigorous methodologic perspective to help readers evaluate the chance of distortion of results of a study. Concerns should include not only laboratory quality control and population stratification, mentioned in the editorial (1), but, perhaps more importantly, effects of definition, recruitment, and determinants of participation in cases and controls.

The editorial (1) calls for articles to ''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.'' The editors should particularly encourage those replication studies that extend the work of the original publication by also studying environmental factors and other genes possibly involved in the same pathway as the gene in the original finding.

Two challenges now facing researchers in the molecular epidemiology of cancer community are how best to exploit available biospecimens and data sets and how to make results of these analyses available. Quick reporting of solid results, especially when surprising, is certainly desirable. In the long run, however, dissemination of the information to guide future studies is at least as important as the immediate interpretation of the results of individual studies as positive or negative findings.

References

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.
2. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 2004;96:434–42.
3. Sabatti C, Service S, Freimer N. False discovery rate in linkage and association genome screens for complex disorders. Genetics 2003;164:829–33.
4. Thomas DC, Clayton DG. Betting odds and genetic associations. J Natl Cancer Inst 2004;96:421–3.

Cancer Epidemiol Biomarkers Prev 2005;14(6):1361