In Response:

Statnikov et al. reanalyzed our original case-control study (1) and concluded that (*a*) the single-nucloetide polymorphisms (SNP) identified were false positives and (*b*) the performance of our classifier was overestimated. Their conclusions reflect some misunderstandings about the purpose, design, and analytic approach in our study. Most importantly, they failed to appreciate that the primary purpose of the original study, a pilot effort, was to generate leads for replication in other independent samples of cases and controls, and that validity of the selected SNPs can only be adequately judged by appropriate replication studies.

As background relevant to design, our original study (1) was part of a larger project with 600 esophageal squamous cell carcinoma cases and 600 controls. In the entire study population, 20% of cases are family history positive for esophageal cancer compared with just 11% of controls. Thus, family history is a strong risk factor for esophageal squamous cell carcinoma in this study as it has been in all previous studies conducted in this high-risk area (2, 3). Our pilot study enriched cases for family history by design (to 50%) purposefully to improve our likelihood of identifying genetic loci related to family history in an area where multiplex families are common; use of family history as a covariate in the analysis was simply to match its use in the design. Such enrichment by family history has been done in many genome-wide association studies (4).

Whether the SNPs we identified represent true-positive or false-positive signals can only be determined in hindsight after replication studies. We recently conducted just such a replication study in which we evaluated the 38 SNPs identified in our original study in a new sample of 300 esophageal squamous cell carcinoma cases and 300 matched controls and determined that SNPs in four genes, *EPHB1, PGLYRP2, PIK3C3,* and *SLC9A9*, were also significant in the replication sample. Although further replication is needed, results of this initial replication experiment suggest that a subset of the SNPs identified in our original study may represent true-positive signals.

It can readily be shown algebraically that for the purpose of selecting a set of SNPs for further evaluation, our model (GLM1) and Statnikov's model (GLM2) are the same because the rank order of SNPs by GLM1 and that by GLM2 are identical. The difference in approaches is that Statnikov et al. focused on the absolute *P* values whereas our interest was in the rank order of the *P* values, regardless of absolute value, to select a set of SNPs for further study.

Statnikov et al. make a reasonable point on the performance of our classifier. Nevertheless, our 10-fold cross-validation analysis using selected SNPs showed that our classifier was significantly better than a random guess with $P = 1.9 \times 10^{-23}$. Although our classifier provided supporting evidence for the validity of the selected SNPs, we recognize that cross-validation based on a single data set is insufficient to validate the results. As with the issue of false positives, the better approach is not cross-validation in the same sample, but replication, replication, replication.

To conclude as we began, the validity of the selected SNPs can only be effectively judged by experimental validation from replication studies of genome-wide association study findings.

Howard H. Yang
Ying Hu
Maxwell P. Lee
Laboratory of Population Genetics,
Center for Cancer Research,
Bethesda, Maryland

Nan Hu
David Ng
Alisa M. Goldstein
Chaoyu Wang
Philip R. Taylor
Division of Cancer Epidemiology and Genetics,
National Cancer Institute, NIH,
Bethesda, Maryland

References

1. Hu N, Wang C, Hu Y, et al. Genome-wide association study in esophageal cancer using GeneChip mapping 10K array. Cancer Res 2005;65:2542–6.
2. Chang-Claude J, Becher H, Blettner M, Qiu S, Yang G, Wahrendorf J. Familial aggregation of oesophageal cancer in a high incidence area in China. Int J Epidemiol 1997;26:1159–65.
3. Hu N, Dawsey SM, Wu M, Taylor PR. Family history of oesophageal cancer in Shanxi Province, China. Eur J Cancer 1991;27:1336.
4. Easton DF, Pooley KA, Dunning AM. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087–93.

Howard H. Yang, Ying Hu, Maxwell P. Lee, et al.

Cancer Res 2008;68:3075.

Updated version	Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/68/8/3075

Cited articles	This article cites 4 articles, 1 of which you can access for free at: http://cancerres.aacrjournals.org/content/68/8/3075.full#ref-list-1
Citing articles	This article has been cited by 1 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/68/8/3075.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://cancerres.aacrjournals.org/content/68/8/3075. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.