Interpretation of Genome-Wide Association Study Results

June 15, 2010
By Dezheng Huo, PhD [1] and Olufunmilayo I. Olopade, MD [2]

As genome-wide association studies (GWAS) have opened the door to systematic discovery of genetic factors for complex diseases, including cancers, the clinical utility of the findings remains to be determined. This is elegantly discussed in the article in this issue of ONCOLOGY by Stadler et al. The authors rightfully caution against the use of “personal genomic tests” based on cancer GWAS results for personal cancer risk prediction.

While GWAS have provided new insights into genetic risk factors for cancer, the hundreds of genetic variants that are associated with cancers identified through these studies only explain a small proportion of estimated heritability. For example, under a polygenic model, seven reproducible genetic susceptibility alleles for breast cancer were estimated to explain about 5% of breast cancer heritability.[1] The large unidentified heritability can be partially explained by the design and philosophy of current GWAS, which is based upon common disease-common variant philosophy and genotyping of several hundred thousand to one million single nucleotide polymorphisms (SNP). Through linkage disequilibrium, these SNPs were selected as tag SNPs that cover up to 80% of all common SNPs (> 5% frequency in population) in the genome. Therefore, rare variants are poorly represented and less likely to be covered by current GWA genotyping arrays. Structural variants, either rare or common, have also not been completely mapped out in the human genome, so they could explain some of the missing heritability.

The interpretation of GWAS results involves two related concepts: false positive and false negative. The importance of replication studies cannot be overstated for GWAS because only a limited number of variants are genuine risk alleles. The first step of replication should test the same index SNPs at the same direction in a similar population with the purpose of ruling out false positivity. Replication in different populations is complicated because of differences in linkage disequilibrium pattern across populations. Failure to replicate index SNPs in different populations is quite common and may indicate that the original SNPs identified from GWAS are neither causal variants nor in linkage disequilibrium with causal variants in the replication population. Regarding false-negative results, a single GWAS usually does not have the power to detect all but the biggest effects and only the strongest signals in the discovery stage are further tested in the replication stage. Several approaches can identify additional causal variants, and some have been implemented, including multimaker analysis (haplotype-based and imputation methods) and meta-analysis of several GWAS.

It is laudatory that GWAS data can be retrieved from publicly available databases such as dbGaP. This public effort helps to reduce publication bias and allow the entire scientific community to apply bioinformatics techniques to discover additional genetic variants and gene-gene interactions beyond traditional statistical validation.

Thus far, the overwhelming majority of GWAS have been limited to populations of European ancestry. This raises at least two questions: whether novel cancer susceptibility genes can be identified in other populations and whether causal variants in European populations affect cancer risk in other ethnic populations in the same way. The answer to the first question is yes, as suggested by a GWAS in Chinese populations that identified ESR1 as a breast cancer susceptibility gene, but this locus had been missed by several GWAS in Caucasian populations.[2] Because genetic variation is greatest in populations of recent African ancestry, GWAS in indigenous Africans and African Americans are warranted and have the potential to generate novel insight into the genetic architecture of cancer.

The answer to the second question is probably no, but studies of causal variants in non-European populations are urgently needed. Stadler and colleagues soberly concluded that it is premature to
translate current findings from GWAS to preventive oncology practice as the results are not yet
generalizable. Causal variants need to be identified and their frequencies need to be characterized in
different populations before a genetic prediction model based on GWAS can be useful in diverse
clinical settings. If and when all causal variants are identified, the prediction of cancer risk can, in
theory, be improved significantly.[3]
Lastly, the contribution of environmental risk factors should not be ignored in the genome era. Twin
studies demonstrate the that environment has the primary role in causing all common cancers
examined, whereas heritable genetic factors account for a nonignorable minority for malignancies of
the prostate, colon, and breast (> 25%).[4]
One needs to understand that familial cancer is not equal to inherited cancer and sporadic cancer is
not equal to cancer caused by environmental exposure. On one hand, familial cancer is often
considered as resulting from inherited factors, but shared environmental factors can also contribute
to familial aggregation of cancers. On the other hand, not all inherited diseases exhibit familial
aggregation. The majority of carriers of BRCA1 and BRCA2 mutations, two highly penetrant breast
cancer susceptibility genes, have no family history of breast cancer.[5] Similarly, low-penetrant
common variants identified from recent GWAS of prostate cancer can predict prostate cancer risk in
individuals with a family history.[6]
As reviewed by Stadler and colleagues, findings from the first wave of GWAS reflect known
underlying biology of some cancers such as melanomas and shed light on new pathways that are
unknown or less emphasized in other cancers. GWAS data can also be utilized to better understand
the carcinogenesis process induced by environmental factors. Future efforts should focus on
gene-environment interaction; only then can the promise of GWAS studies be translated to
preventive oncology practice.

Financial Disclosure: The authors have no significant financial interest or other relationship with
the manufacturers of any products or providers of any service mentioned in this article.

References:

1. Pharoah PD, Antoniou AC, Easton DF, et al: Polygenes, risk prediction, and targeted prevention of


of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland.

cumulative risk of breast cancer for a defined set of protein-truncating mutations in BRCA1 and


Source URL:
y-results

Links: