



Human genome epidemiology, progress and future

Hongbing Shen^{a,b,✉}, Guangfu Jin^{a,b}

^aDepartment of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu 210029, China;

^bSection of Clinical Epidemiology, Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Cancer Center, Nanjing Medical University, Nanjing, Jiangsu 210029, China.

Received 10 April 2013, Accepted 16 April 2013, Epub 25 May 2013

Human genome epidemiology (HuGE) uses systematic applications of epidemiologic methods to assess the impact of human genetic variation on health and disease. In the past ten years, human genome epidemiology has made great progresses along with advances in genomics technologies, which make it possible for the examination of genetic variants in a large sample size at a sufficiently low cost. Genetic association study in population provides a powerful approach to identify variants or genes associated with disease of interest by comparing distributions of genetic variants between affected and unaffected individuals. A critical question remains that these findings from genetic association studies are not consistently reproducible, possibly due to false positive, false negative or true variability in association among different populations. Before 2006, only few low penetrance variants outside the HLA locus had been discovered to be reproducibly associated with disease susceptibility based on the candidate gene approach^[1]. With the application of high throughput genotyping technology, genome-wide association studies (GWAS) have emerged as a powerful tool for investigating the genetic architecture of complex diseases without a prior hypothesis about a particular gene or locus. In the GWAS approach, several hundred thousand to more than a million single nucleotide polymorphisms (SNPs) are assayed across the whole genome in a large sample size of thousands of individuals^[2]. To date, GWAS have led

to the discovery of thousands of loci that are associated with different kinds of human diseases or traits. These findings have advanced our current knowledge of genetics of complex diseases, and provided new insights into identification of therapeutic targets and developing targeted interventions^[3]. Given the advance in association study of complex diseases, especially for cancer, in the current issue, two reviews have been invited to comment on the progress of HuGE on head and neck cancer^[4] and bladder cancer^[5].

The SNP chips used for GWAS are designed to provide better coverage of common SNPs, which, however, may be incompetent to interrogate all common variants for certain regions and have very limited potential to capture rare and low frequency variants^[6]. Such variants may hardly be focused on using available genome-wide genotyping technologies. To date, only a small fraction of disease heritability could be interpreted by known susceptibility loci. For example, more than 40 prostate cancer susceptibility loci have been reported; however, these only account for approximately 25% of the familial risk of the disease^[7]. In this regard, genetic association studies should direct more efforts at the variants around genes/regions already implicated in disease pathogenesis. Functional variants in such genes/regions represent particularly impressive candidates. Several reports in this issue demonstrate the necessity and importance of association studies based on functionally candidate genes for variants associated with

This work was supported by National Natural Science Foundation of China (81230067, 81071715, and 81001276), and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

✉ Corresponding author: Hongbing Shen, MD, Ph.D, Department

of Epidemiology and Biostatistics, School of public health, Nanjing Medical University, 140 Hanzhong Road, Nanjing, Jiangsu 210029, China. Tel (Fax): +86-25-86868439, E-mail: hbshen@njmu.edu.cn.

The authors reported no conflict of interests.

disease susceptibility or prognosis^[8-10]. Of course, these findings, in the absence of replication, need to be validated in other studies with independent samples. Nevertheless, we suggest that more efforts should be taken on regions identified as relevant to diseases through GWAS using a targeted region approach, which has been shown as an efficient and cost-effective screening for rare and low-frequency polymorphisms to expand the disease variance explained^[11].

In addition to the issues discussed above, rare variants, epistasis, epigenetics and genotype-environment interactions may also contribute to the missing heritability^[12]. As compared with the identified common variants by GWAS (in general, minor allele frequency is more than 5%), rarer variants that are poorly detected by available genotyping arrays. Structural variation, including copy number variants (CNVs), inversions, translocations, microsatellite, repeat expansions, insertions of new sequence, complex rearrangements, and short insertions or deletions (indels), may also account for some of the unexplained heritability and are poorly captured by existing arrays^[13]. In the current issue, Wang et al^[14] explored the landscape and impact of a new form of genomic variation, run of homozygosity (ROH), on lung cancer. ROH is a continuous or uninterrupted stretch of a genomic sequence without heterozygosity in the diploid state, which is poorly investigated to date. Using an existing GWAS dataset including 1473 lung cancer cases and 1962 controls, they identified a new region at 14q23.1 that was consistently associated with lung cancer risk in Chinese population^[14], suggesting that ROHs may be also responsible for the unexplained familial risk of diseases. Nevertheless, the epistasis or gene-environment interaction may contribute to a large percentage of disease variance. However, it is still difficult to detect them in the current status of epidemiological study design, exposure assessment, and methods of analysis. Long term environmental exposure assessment will be more important in the future epidemiological study design.

Beyond doubt, genetic association studies, especially GWAS, have made a great progress in elucidating genetic factors underlying complex diseases. However, there are still several barriers to overcome. GWAS-identified SNPs may point to functional variants but are unlikely themselves to be the causative variants, since there will often be several variants in strong linkage disequilibrium (LD) that show more or less equivalent evidence of association for any given signal of association. Extensive sequencing of the identified region followed by a well-designed fine-mapping study in multiple populations may be helpful to narrow an association signal to potentially causa-

tive variants^[15]. Furthermore, much additional work is needed to determine the functional basis for the observed associations.

Particularly, the potential for variants identified in GWAS to predict the risk of complex diseases has been anticipated, but the usefulness of translating these fundamental genetic findings into the bedside remains debatable^[16]. Nevertheless, there are already a number of benefits of such genetic prediction over classical non-genetic models. For instance, genetic risk prediction is more stable over time than traditional risk factors, as a person's genetic sequence is absolutely constant throughout their life. Recently, Sun et al.^[17] reported that genetic score calculated by genetic variants discovered through association study is an objective and better measurement of inherited risk of prostate cancer than family history, which can be obtained without a laboratory test but influenced by family size, age and survival status of male relatives, recall ability, family communication, and prevalence of the disease in populations. With increasing numbers of discovered genetic variants that can be used as biomarkers in future genetic risk prediction, we believe that identification of a proportion of a high risk population may be feasible for target diagnostic, and preventive and therapeutic interventions for complex disorders.

References

- [1] Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003; 33: 177-82.
- [2] Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med* 2009; 360: 1759-68.
- [3] Klein C, Lohmann K, Ziegler A. The Promise and Limitations of Genome-wide Association Studies. *JAMA* 2012 Nov 8; 1-2. doi: 10.1001/2012.jama.10823. [Epub ahead of print].
- [4] Wang M, Chu H, Zhang Z, Wei Q. Molecular epidemiology of DNA repair and head and neck cancer. *J Biomed Res* 2013 May. doi:10.7555/JBR.27.20130034. [Epub ahead of print].
- [5] Chu H, Wang M, Zhang Z. Bladder Cancer Epidemiology and Genetic Susceptibility. *J Biomed Res* 2013 May. doi: 10.7555/JBR.27.20130026. [Epub ahead of print].
- [6] Zeggini E, Rayner W, Morris AP, Hattersley AT, Walker M, Hitman GA, et al. An evaluation of HapMap sample size and tagging SNP performance in large-scale empirical and simulated data sets. *Nat Genet* 2005; 37: 1320-2.
- [7] Kote-Jarai Z, Olama AA, Giles GG, Severi G, Schleutker J, Weischer M, et al. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat Genet* 2011; 43: 785-91.
- [8] Zhao J, Liu L, Zhang AQ, Chen Q, Fang W, Zeng L, Lu J.

- Effect of EME1 exon variant Ile350Thr on risk and early onset of breast cancer in southern Chinese women. *J Biomed Res* 2013 May. doi: 10.7555/JBR.27.20130013. [Epub ahead of print]
- [9] Liu L, Liu Y, Liu J, Zhai X, Wen J, Xie K, Shen H, Hu Z, Fan Z. Genetic variants in pseudogene E2F3P1 confer risk for HBV-related hepatocellular carcinoma in a Chinese population. *J Biomed Res* 2013 May. doi: 10.7555/JBR.27.20130019. [Epub ahead of print].
- [10] Cao S, Wang C, Huang X, Dai J, Hu L, Liu Y, Chen J, Ma H, Jin G, Hu Z, Xu L, Shen H. Prognostic Assessment of Apoptotic Gene Polymorphisms on Non-small Cell Lung Cancer in Chinese. *J Biomed Res* 2013 May. doi: 10.7555/JBR.27.20130014. [Epub ahead of print].
- [11] Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 2011; 43: 1066-73.
- [12] Gibson G. Hints of hidden heritability in GWAS. *Nat Genet* 2010; 42: 558-60.
- [13] Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009; 461: 747-53.
- [14] Wang C, Xu Z, Jin G, Hu Z, Da J, Ma H, Jiang Y, Hu L, Chu M, Cao S, Shen H. Genome-Wide Analysis of Runs of Homozygosity (ROHs) Identifies New Susceptibility Regions of Lung Cancer in Han Chinese. *J Biomed Res* 2013 May. doi: 10.7555/JBR.27.20130017. [Epub ahead of print].
- [15] Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, et al. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* 2011; 43: 513-8.
- [16] Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010; 363: 166-76.
- [17] Sun J, Na R, Hsu FC, Zheng SL, Wiklund F, Condeary LD, et al. Genetic score is an objective and better measurement of inherited risk of prostate cancer than family history. *Eur Urol* 2013; 63: 585-7.

RECEIVE IMMEDIATE NOTIFICATION FOR EARLY
RELEASE ARTICLES PUBLISHED ONLINE

To be notified by e-mail when *Journal* early release articles are
published online, sign up at **JBR-pub.org**.