EDITORIAL

GENOME-WIDE ASSOCIATION STUDIES IN GENOMIC MEDICINE - ARE WE THERE YET?

M. Ilyas Kamboh

Almost all human diseases have a genetic component, small in some and larger in others. Rare disorders like cystic fibrosis or sickle cell disease are largely determined by single gene mutations with high penetrance. On the other hand, common diseases like coronary heart disease or type 2 diabetes result from a complex interaction of multiple genes and environmental factors where individual genes contribute small but cumulative effects to disease susceptibility. Identification of underlying genes for these diseases is a critical first step towards devising potential therapeutic and preventive measures.

Traditionally, geneticists have used two main approaches to identify genes for complex diseases: family-based linkage studies or case-control association studies. In linkage studies, coinheritance of the disease is traced with chromosomal regions among family members using highly polymorphic genetic markers. In association studies, cases and controls are used to find differences in frequencies of genetic variants between the two groups. Linkage studies have proved to be a bonanza for identifying the genetic basis of a number of rare single gene disorders because of the large contribution of a given gene to a phenotype. However, linkage studies lack the power to detect genes with modest to weak effects, which are responsible for common diseases. Genetic association studies are more powerful than linkage studies to detect modest effect sizes.[1] The added advantage is that a large number of unrelated cases and controls can be collected with relative ease compared to family samples. Moreover, for late-onset diseases, such as Alzheimer's disease, it is not feasible to collect samples from older family members who have already succumbed to the disease. In the past researchers have tried the 'candidate gene' approach (biologically plausible genes or genes located under linkage peaks, called positional candidate genes) where they looked for associations between single nucleotide polymorphisms (SNPs) and various common diseases. This approach has produced, for the most part, dismal results partly because most of the efforts for identifying genes were focused on screening a few SNPs per candidate gene. However, this approach is inadequate as it fails to cover the entire variation in a gene. The recent efforts by the Human Genome Project and the International HapMap Project to provide a dense SNP map of genes have provided the opportunity to thoroughly examine the role of biological or positional candidate genes.[2,3] However, a mere focus on known candidate genes would not capture the full spectrum of variation responsible for a common disease because of the potential existence of yet to be discovered biological pathways for a given disease. For this reason, a new approach was envisioned where an association study would be carried on hundreds of thousands of SNPs covering the entire genome, giving the concept of genome-wide association studies (GWAS). [4] This approach is hypothesis free and conceptually would identify all known and unknown causative genetic variation underlying a disease.

The new era of gene hunting began in 2005 when a GWAS reported the identification of a new gene for age-related macular degeneration using a chip array consisting of about 100,000 SNPs.[5] Since then, a plethora of GWAS have identified many new genes for several common and complex diseases, especially over the last two years. A ground-breaking large GWAS was reported in 2007 by a consortium of more than 50 British groups, called the Wellcome Trust Case Control Consortium (WTCCC),[6] where they used a panel of 500,000 SNPs to genotype 17,000 subjects for seven common diseases: bipolar disorder, coronary heart disease, Crohn's disease (the most common form of inflammatory disease), hypertension, rheumatoid arthritis, and type 1 & type 2 diabetes. They used 2,000 cases for each disease and 3,000 shared controls for each disease and identified strong statistical evidence of association with 24 variants. Several GWAS on additional diseases have been reported in 2008, including among others systemic lupus erythematosus, lung

Professor and Chairman, Department of Human Genetics, Graduate School of Public Health
University of Pittsburgh, PA, USA.
Correspondence: - M. Ilyas Kamboh
Professor and Chairman, Department of Human Genetics, Graduate School of Public Health
University of Pittsburgh, PA, USA.
Email: kamboh@pitt.edu
Received: September 24, 2008; accepted: November 27, 2008

cancer, prostate cancer, colorectal cancer, breast cancer and schizophrenia (http://www.genome.gov/26525384). The published GWAS available at the afore-mentioned site also include studies on continuous traits, such as plasma lipid and glucose levels, body mass index, hair color, skin pigmentation and height.

Thus, the initial GWAS are providing important information about the complex genetic architecture of common diseases as well as continuous traits and revealing potential novel biological pathways for new drug targets. This success is culminated by several factors, notably the availability of dense SNP maps, the use of state-of-the-art hardware and software in manufacturing the SNP chip arrays, the availability of large and well characterized clinical samples for a number of diseases and improved statistical tools for data analyses.

Since the ultimate purpose of genetic research is to translate the discoveries from bench to bedside, an obvious question arises about the clinical implications of new discoveries discerned by GWAS. Could this information be used in 'personalized genomic medicine' in order to lower the disease risk through therapeutic or preventive measures or tailor the drug therapy based on particular genetic background (pharmacogenomic) or used as a diagnostic tool? Although some genome companies like deCODEme, 23andMe, Knome and Navigenics have already started offering some of the new genetic tests directly to the public, most researchers feel that it is premature, at least now, to use this information because much more needs to be learned about how these genetic variants cause disease and how they interact with each other and environmental factors to modify disease risk. Furthermore, the variants identified thus far may not be functional and, more importantly, they account for only a fraction of the total variation responsible for a given disease because the currently used SNP panels have used far less coverage of the human genome compared to the HapMap-based estimates.[7] It has been suggested that less common variants (frequency <5%) will have a larger effect on the disease outcome as compared to the identified common variants having modest effects (odds ratios ranging from 1.2 to 1.5).[8,9] Since the commercially available genotyping arrays do not carry less common or rare variants, a concerted effort is needed to identify these variants by deep sequencing. The recently announced 1000 Genomes Project is precisely aimed towards achieving this goal where 1000 individuals will be sequenced as part of an international collaborative effort to produce a comprehensive catalog of human genetic variants occurring at a frequency of 1% or higher.[10] The new DNA sequencing technologies are being built to reduce the cost and time of sequencing such that the initial sequencing cost of about US$1 million over a few months for an individual genome would be cut down to US$1,000 over a few hours, in the not too distant future. This is evident from two recent individual genome sequences that were completed for a cost of US$ 250,000 and US$ 500,000,[11,12] respectively, but would cost less than half if done today. When a more complete picture emerges about the causative common and rare functional variants it will not only increase their predictive power of the disease outcome, but will also help to understand the disease mechanisms underlying the associations.

Finally, in order to translate the GWAS discoveries into clinical practice it would also be necessary to carry out large scale, prospective, population-based cohort studies in order to evaluate the predictive and diagnostic values of new genetic variants and how they interact with environmental factors to modulate the disease risk. Manipulation of the modifiable environment of genetically susceptible individuals for common diseases offers the best opportunity to intervene for prevention of disease process.

REFERENCES

1. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996; 273: 1516-7.

2. The International HapMap Consortium. The International HapMap Project. Nature 2003; 426: 789-96.

3. The International HapMap Consortium. A second generation human haplotype map of over 3 million SNPs. Nature 2007; 449: 851-61.

4. Kruglyak L. The road to genome-wide association studies. Nat Rev 2008; 9: 314-8.

5. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science 2005; 308: 385-9.

6. The Wellcome Trust Case Control Consortium. Genome-wide association studies of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007; 447: 661-78.

7. Bhangale TR, Reider MJ, Nickerson A. Estimating coverage and power for genetic association studies using near-complete variation data. Nat Genet 2008; 40: 841-3.

8. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 2008; 40: 695-701.

9. McCarthy MI, Abecasis GR, Gardon LR et al. Genome-wide association studies for complex traits: consensus, uncertainty, and challenges. Nat Rev 2008; 9: 356-69.

10. McGuire AL. 1000 genomes: on the road to personalized medicine. Personalized Med 2008; 5: 195-7.

11. Bentlely DR, Balasubramanilan S, Swerdlow HP et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008; 456: 53-9,

12. Wang J, Wang W, Li R et al. The diploid genome sequence of an Asian individual. Nature 2008; 456: 60-5.