Genome-wide Association Studies and Human Disease: From Trickle to Flood

Peter M. Visscher; Grant W. Montgomery

JAMA. 2009;302(18):2028-2029 (doi:10.1001/jama.2009.1643)

http://jama.ama-assn.org/cgi/content/full/302/18/2028

Genome-wide Association Studies and Human Disease

From Trickle to Flood

Peter M. Visscher, PhD
Grant W. Montgomery, PhD

MANY COMMON COMPLEX DISEASES SUCH AS hypertension, diabetes, coronary heart disease, psychiatric disorders, and some cancers have a genetic etiology. Despite enormous efforts over the last few decades, little real progress was made in finding the genes and causal variants involved. Genome-wide association studies, in which hundreds of thousands of DNA markers are tested (usually in a case-control design) for association with disease, provide the first effective approach to search for genetic variants that contribute to the complex etiology of common human diseases.

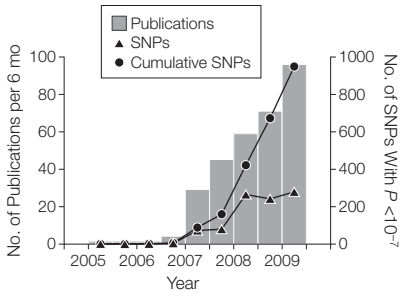
In the last 3 years, almost 1000 variants associated with a range of human traits and common diseases have been identified using genome-wide association methods (FIGURE).[1,2] To date, most of these studies have been in populations of European descent.

Genome-wide search strategies developed from advances in genotyping technology, greater understanding of the structure of common variation in the human genome, and continued advances in computing power and software tools. Commercial genotyping platforms can type as many as 1 million single-nucleotide polymorphisms (SNPs) on a single chip, capturing (tagging) most variation between individuals in a single experiment. Instead of genotyping per sample the 10 million to 15 million common SNPs that segregate in the population for each sample, a much smaller subset of approximately 500 000 SNPs is sufficient to cover common variation in the genome. The flipside of this redundancy is that SNPs that are statistically associated with disease are unlikely to be causal and will be correlated with an ungenotyped causal variant.

Genome-wide association studies have provided insights about disease, in particular: (1) for almost any disease that has been investigated, there are SNP variants common in the population (with an allele frequency >5%) that are robustly associated with disease; (2) most of these variants are in genes that contribute to biological pathways that were previously not known to be involved in disease or are nowhere near a known protein-coding gene; (3) the effect sizes of associated SNPs are typically small with odds ratios of risk alleles in the range of approximately 1.1 to 1.5; (4) for any particular disease, accumulating the effects of many different SNPs associated with a disease usually explains only a small fraction of the familial risk (or heritability); and (5) not all diseases and traits are alike in genetic architecture. For example, in age-related macular degeneration, approximately 50% of genetic variation has been accounted for by only 6 loci,[3] whereas for adult height, only 6% of genetic variation has been accounted for by approximately 50 loci.[4] In the iron homeostasis pathway, several common SNPs have been reported that each explain 5% or more of genetic variation.[5]

Failure to account for much of the genetic variation or "missing heritability"[6] has divided the community with respect to the success or failure of genome-wide association studies. Since the common goal is to understand pathways to disease and develop improved methods of prevention, diagnosis, and treatment, it is important to understand what the current flood of associated SNPs reveals about disease biology and gene regulation, why so little genetic variation has been accounted for, and what experimental approaches might lead the identification of causal variants and mechanisms.

Figure. The Genome-wide Association Revolution: From Trickle to Flood



SNPs indicates single-nucleotide polymorphisms.
Data are adapted from the National Human Genome Research Institute[1] at http://www.genome.gov/gwastudies/.

Author Affiliations: Department of Statistical Genetics (Dr Visscher) and Department of Molecular Epidemiology (Dr Montgomery), Queensland Institute of Medical Research, Brisbane, Australia.
Corresponding Author: Peter M. Visscher, PhD, Queensland Statistical Genetics, Queensland Institute of Medical Research, 300 Herston Rd, Brisbane, Australia 4006 (peter.visscher@qimr.edu.au).

The main hypotheses that explain the missing heritability are that either there are many variants with odds ratios so small they do not reach formal statistical significance despite large genome-wide association studies, or causal variants are not tagged well by SNPs on the commercial chips (eg, because these variants occur at lower frequency or are in areas of the genome in which it is difficult to develop SNP assays).

If most causal variants have small odds ratios, many loci must be contributing to genetic variation for disease risk and therefore, all individuals must carry their own, probably unique, portfolio of risk variants. Mathematical polygenic models for disease are well developed and are consistent with the observation that some cases appear to be sporadic whereas other cases have a family history. A recent genome-wide association analysis of schizophrenia showed that the data were consistent with a substantial proportion of disease risk that is polygenic,[7] suggesting that hundreds or even thousands of variants contribute to risk—similar to the findings for human height. Another consequence of many genes being involved is that there must be widespread pleiotropy, ie, comorbidity due to shared genetic factors. Indeed, genome-wide association analysis of schizophrenia and bipolar disorder revealed a strong overlap in risk.[7] Thus, genetic analysis can lead to reassessment of diagnostic criteria.

Resequencing of genomic regions is uncovering much new variation. Emerging evidence points to both common and rare variants contributing to disease risk. For example, Nejentsev et al[8] resequenced a candidate gene for type 1 diabetes and detected 4 new variants at approximately 1% frequency that in total contributed more to variation in risk in the population than a single common variant in the same gene detected by a previous genome-wide association study. Causal variants at lower frequencies must have very large odds ratios to contribute substantially to variation in disease risk. If not, combined resequencing and genotyping projects to detect such variants will need to be even larger than current genome-wide association studies.

Whatever the source of the missing heritability, genome-wide association results have provided novel insights into the pathogenesis of many diseases. Variants that increase the risk of type 2 diabetes influence β-cell development and function, and focus attention on insulin secretion in the development of disease.[9] Discoveries in inflammatory bowel disease have highlighted the importance of the autophagy pathway in disease development. Better understanding of disease pathogenesis has already opened up new avenues for research likely to provide opportunities for drug discovery in the future.

In addition to providing new insights into disease biology, another consideration is whether and how genome-wide association studies will affect clinical outcome. Commercial companies are offering personalized genomic risk profiles based on genome-wide association results directly to consumers. However, since for most diseases, at present, only a small proportion of familial risk has been accounted for, most risk profiles closely follow the population average and the value of these profiles is questionable. If a substantial proportion of genetic variation has been accounted for by genetic markers, then risk predictors could be constructed that are more informative than current predictors based on family history. Risk prediction profiles can also be used to investigate genetic comorbidity and to evaluate use of current diagnostic criteria in closely related conditions.

Whole genome resequencing will uncover new variation but will be constrained by sample size in detecting association between rare variants and disease. Ever larger genome-wide association studies will continue to provide insight into the genetic architecture of complex diseases. Although genome-wide association studies have facilitated critical discovery in understanding the genetics of disease, much more remains to be done to translate the flood of knowledge into clinical practice.

Financial Disclosures: None reported.
Funding/Support: This work was supported by the Australian National Health and Medical Research Council.
Role of the Sponsors: The funding organization had no role in the preparation, review, or approval of the manuscript.

REFERENCES

1. National Human Genome Research Institute. A Catalog of Genome-Wide Association Studies. National Institutes of Health Web site. http://www.genome.gov/gwastudies. Accessed: July 21, 2009.
2. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106(23):9362-9367.
3. Seddon JM, Reynolds R, Maller J, Fagerness JA, Daly MJ, Rosner B. Prediction model for prevalence and incidence of advanced age-related macular degeneration based on genetic, demographic, and environmental variables. *Invest Ophthalmol Vis Sci*. 2009;50(5):2044-2053.
4. Visscher PM. Sizing up human height variation. *Nat Genet*. 2008;40(5):489-490.
5. Benyamin B, McRae AF, Zhu G, et al. Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels. *Am J Hum Genet*. 2009;84(1):60-65.
6. Maher B. Personal genomes: the case of the missing heritability. *Nature*. 2008;456(7218):18-21.
7. Purcell SM, Wray NR, Stone JL, et al; International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748-752.
8. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science*. 2009;324(5925):387-389.
9. Majithia AR, Florez JC. Clinical translation of genetic predictors for type 2 diabetes. *Curr Opin Endocrinol Diabetes Obes*. 2009;16(2):100-106.