ONLINE FIRST

# The Promise and Limitations of Genome-wide Association Studies

Christine Klein, MD
Katja Lohmann, PhD
Andreas Ziegler, PhD

As of May 30, 2012, the catalog of published genome-wide association studies (GWAS)[1] lists an impressive 1269 GWAS, covering a broad spectrum of conditions including Alzheimer disease, breast cancer, and human immunodeficiency virus susceptibility.[2] The catalog also contains studies on common traits such as height and freckles, as well as responses to drugs for various medical conditions. It is difficult to discuss GWAS without sounding megalomaniacal. Considerably more than 1000 published GWAS, replication studies, and meta-analyses have been conducted in an unprecedented global research effort in only 7 years. Most GWAS have included hundreds or even thousands of patients and controls, have hundreds of thousands of participants worldwide, and although genotyping costs have plummeted in recent years, hundreds of millions of research dollars have been spent on GWAS since 2005.

A 2007 fact sheet released by the National Human Genome Research Institute, in the early days of GWAS, raised expectations that personalized medicine, including individual risk prediction, disease prevention, and specific treatment, was just around the corner. "With the first GWAS published in 2005, . . . health professionals will be able to use such tools to provide patients with individualized information about their risks of developing certain diseases . . . to tailor prevention programs to each person's unique genetic makeup . . . to select the treatments most likely to be effective and least likely to cause adverse reactions. . . ."[3] Has the promise of GWAS[2] been realized 5 years later? Although there is no simple answer to this question, it is helpful to consider 3 important and closely intertwined features of GWAS, ie, sample size, characterization of probands (samples), and effect size.

## Sample Size: The Larger, the Better?

One of the best examples of limited success of GWAS is the detection of an association of age-related macular degeneration with a complement factor H polymorphism in only 96 cases and 50 control participants published in 2005.[4] In contrast, many of the recent GWAS have been very large, including thousands of patients and control participants. What contributes to the success of some small GWAS? Even though small GWAS can only detect large effects, as illustrated by the example of age-related macular degeneration with an odds ratio (OR) of approximately 7, the promise of large GWAS lies in the discovery of long lists of susceptibility loci contributing to disease risk, albeit usually with small effect sizes. Meta-analyses are an ideal way to increase sample size. However, larger sample sizes do not result in larger effect sizes, but rather in increasing the number of variants conferring small effects, as indicated by ORs hovering around 1.

Accordingly, even very well-established associations, such as polymorphisms in the *SNCA* or *MAPT* genes with Parkinson disease, have limited clinical-diagnostic implications, especially when focused on developing screening tools for at-risk individuals or even unselected populations. For example, the first meta-analysis of Parkinson disease GWAS, which included 12 386 cases and 21 583 controls, reported an almost 2.5-times increased odds of Parkinson disease in carriers of selected risk variants.[5] Based on the prevalence of Parkinson disease of 0.14%, this association translates into a lifetime risk for developing the disease of only 0.35% even in the highest-risk group. These considerations lead to 2 important conclusions: risk prediction for an individual usually cannot be derived even from large-scale GWAS data,[6] and sample size is not a quality marker of GWAS per se, especially in terms of clinical relevance.

## Characterization of Samples: Comparing Apples and Oranges?

A common problem with multicenter studies of very large sample size, especially in the absence of measurable diagnostic parameters, is inclusion of heterogeneous groups of probands. An instructive example is genetic risk for progressive supranuclear palsy (PSP), a form of parkinsonism that is strongly associated with genetic variants in the *MAPT* gene region, which has a very small *P* value ($P = 1.5 \times 10^{-116}$; OR, 5.11; 95% CI, 4.43-5.91).[7] Intriguingly, variants in a same gene have also been identified as the "top hit" genetic

**Author Affiliations:** Section of Clinical and Molecular Neurogenetics at the Department of Neurology (Drs Klein and Lohmann), and Institute of Medical Biometry and Statistics (Dr Ziegler), University of Lübeck, Lübeck, Germany.
**Corresponding Author:** Christine Klein, MD, Section of Clinical and Molecular Neurogenetics at the Department of Neurology, University of Lübeck, Maria-Goeppert-Strasse 1, 23562 Lübeck, Germany (christine.klein@neuro.uni-luebeck.de).

risk factor for idiopathic Parkinson disease,[8] however, with much smaller effect sizes (OR, 1.28; 95% CI, 1.25-1.33). Because PSP and Parkinson disease are often difficult to distinguish on clinical grounds, especially at early disease stages, it may be possible that a proportion of individuals categorized as Parkinson disease patients in the Parkinson disease GWAS represent undiagnosed PSP cases, thereby contributing to the *MAPT* association seen in Parkinson disease.

Recently, a new approach combines GWAS data with endophenotypes, ie, hereditary characteristics that are usually associated with the disease under study but not a direct symptom and usually not visible with the unaided eye. The promise of the concept rests on the idea that the endophenotype is more directly determined by the genotype than the overt disease phenotype, thereby providing simpler clues to genetic underpinnings, and thus patient stratification.

## Effect Size: A Measure of Clinical Significance?

When it comes to potential personalized medicine, effect size appears to be clearly the most important aspect of GWAS. Indeed, a number of large-effect candidate-gene associations were found decades before GWAS and have shaped biological understanding and even treatment of complex diseases, as for multiple sclerosis. Of the more than 1200 published GWAS, only 86 studies (6.8%) found ORs of greater than 3.0 at a *P* value of less than $10^{-5}$. Notably, approximately half of those studies (n=46) were conducted with 300 patients or fewer, including 12 studies in the past 12 months.[1] Thus, paradoxically, the very large-size GWAS may be of less medical interest than those identifying larger effects and, therefore, simply do not require high numbers of case and control participants.

## First Success Stories: Future Promises and Challenges

Although medical science is still far from the GWAS-based personalized medicine promised in the 2007 fact sheet, at least 3 important considerations fuel legitimate hope that genetics will continue to become an integral part of a modern medicine more specifically tailored to individual patients. First, important discoveries, for example in the field of pharmacogenomics, have already changed medical practice and resulted in medical policy codes for some treatments such as genetic testing for warfarin dose.

Second, the first genetic interaction studies are starting to provide useful data. For example, high-density lipoprotein cholesterol levels (HDL-C), one of the most important risk factors for coronary heart disease, are significantly influenced by the interplay of the *HMGCR* (hydroxy-3-methylglutaryl-CoA reductase; NG_011449.1) and *LIPC* (lipase hipatic; NG_011465.1) genes. The effect of the gene-gene interaction on HDL-C levels is twice as pronounced as that predicted by the sum of the marginal effects of the 2 loci.[9]

Third, it is important to consider that GWAS are based on common variants that are frequently in linkage disequilibrium with the actual causative variant, which may be associated with larger effect sizes than the common variant included in the GWAS. For instance, fine mapping of loci associated with low-density lipoprotein cholesterol (LDL-C) identified a rare nonsynonymous variant in the *PCSK9* (proprotein convertase subtilisin/kexin type 9; NG_009061.1) gene that exhibited a significantly higher effect on LDL-C levels (–12.9 mg/dL vs –3.7 mg/dL) and explained 5 times more of the contributed variance than the initial GWAS finding.[10]

In this context, genome-wide sequencing has emerged as an even more accurate and powerful tool than GWAS to elucidate the relationship between genetics and (common) diseases. However, even high-resolution genetic variation will only explain a fraction of the heritability of human diseases and traits. Thus, the search is still ongoing for future promise beyond simple genetics with gene-gene and gene-environment interactions, as well as epigenetic effects as important but complex targets.

**Published Online:** November 8, 2012. doi:10.1001/2012.jama.10823
**Conflict of Interest Disclosures:** All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Klein reports receipt of grants from the Schilling Foundation, the German Research Foundation, and the German Ministry of Education and Reaearch (SFB936). Dr Lohmann reports receipt of a grant from the German Research Foundation. Dr Ziegler reports receipt of grants to the institution from the German Ministry of Education and Research, the European Union, and the German Research Foundation.
**Funding/Support:** This study was funded by intramural funding from the University of Lübek (focused research program on genetics).
**Role of the Sponsor:** The University of Lübek had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

**REFERENCES**

**1.** National Human Genome Research Institute. A catalog of published genome-wide association studies. http://www.genome.gov/gwastudies. Accessed May 30, 2012.
**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.** National Human Genome Research Institute. Genome-wide association studies. http://www.genome.gov/20019523. Accessed May 30, 2012.
**4.** Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005;308(5720):385-389.
**5.** Nalls MA, Plagnol V, Hernandez DG, et al; International Parkinson Disease Genomics Consortium. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*. 2011;377(9766):641-649.
**6.** Klein C, Ziegler A. From GWAS to clinical utility in Parkinson's disease. *Lancet*. 2011;377(9766):613-614.
**7.** Höglinger GU, Melhem NM, Dickson DW, et al; PSP Genetics Study Group. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat Genet*. 2011;43(7):699-705.
**8.** Lill CM, Roehr JT, McQueen MB, et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. *PLoS Genet*. 2012;8(3):e1002548.
**9.** Ma L, Brautbar A, Boerwinkle E, Sing CF, Clark AG, Keinan A. Knowledge-driven analysis identifies a gene-gene interaction affecting high-density lipoprotein cholesterol levels in multi-ethnic populations. *PLoS Genet*. 2012;8(5):e1002714.
**10.** Sanna S, Li B, Mulas A, et al. Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS Genet*. 2011;7(7):e1002198.